A Fast Analog Mismatch Analysis by an Incremental and Stochastic Trajectory Piecewise Linear Macromodel *

Hao Yu ‡, Xuexin Liu †, Hai Wang †, Sheldon X.-D. Tan †
‡School of EEE, Nanyang Technological University, Singapore 639798
†Department of Electrical Engineering, University of California, Riverside, CA 92521

ABSTRACT

To cope with an increasing complexity when analyzing analog mismatch in sub-90nm designs, this paper presents a fast non-Monte-Carlo method to calculate mismatch in time domain. The local random mismatch is described by a noise source with an explicit dependence on geometric parameters, and is further expanded by stochastic orthogonal polynomials (SOPs). This forms a stochastic differential-algebra-equation (SDAE). To deal with large-scale problems, the SDAE is linearized at a number of snapshots along the nominal transient trajectory, and hence is naturally embedded into a trajectory-piecewise-linear (TPWL) macromodeling. The TPWL is improved with a novel incremental aggregation of subspaces identified at those snapshots. Experiments show that the proposed method, isTPWL, is hundreds of times faster than Monte-Carlo method with a similar accuracy. In addition, our macromodel further reduces runtime by up to 25X, and is faster to build and more accurate to simulate compared to existing approaches.

1. INTRODUCTION

Transistor mismatch is the primary obstacle to reach a high-yield rate for analog designs in sub-90nm technologies. For example, due to an inverse-square-root-law dependence with the transistor area, the mismatch of CMOS devices nearly doubles for every process generation less than 90nm [1,2]. Since the traditional worst-case or corner-case based analysis is too pessimistic to sacrifice the speed, power, and area, the statistical approach [2–6] thereby becomes a trend to estimate the analog mismatch. Same as the process variation, there are two types of mismatch. One is systematic (or global spatial variation) and the other is stochastic (or local random variation). This paper is focused on the stochastic one. Analog circuit designers usually perform a Monte-Carlo (MC) analysis to analyze the stochastic mismatch and predict the statistical functionality of their designs. As MC analysis requires a large number of repeated circuit simulations, its computational cost is expensive. Moreover, the pseudo-random generator in MC introduces numerical noises that may lead to errors.

There are many non-Monte-Carlo methods [2, 4, 6] developed recently for the stochastic mismatch analysis. [4] first calculated *dc* sensitivities with respect to small device-parameter perturbations and scaled them as desired mismatches. [2] extended [4] by modeling *dc* mismatches as *ac* noise sources. The mismatch, defined in a transient simulation, is converted back from the power spectral density (PSD) in frequency-domain. The speed of these equivalent mismatch simulations is hundred times faster than the Monte-Carlo approaches but the accuracy remains a concern.

Recently, SiSMA [6] studied the mismatch within the framework of the stochastic differential-algebra-equation (SDAE) similar to deal with the transient noise [7]. Due to the introduction of the random variable into the DAE, it is unknown if the derivative is still continuous. Moreover, designers' top interest is the mismatch of the channel current in CMOS transistors. SiSMA thereby modeled the mismatch as a stochastic current-source and formed a SDAE. As such, by assuming that the magnitude of the stochastic mismatch is much smaller than the nominal case, SiSMA linearized the nominal SDAE at *dc* with the stochastic current source. The obtained *dc* solution is used as an initial condition (*ic*) for the transient analysis. As the stochastic current source is only included during *dc*, this assumption may not hold to accurately describe the mismatch during the transient simulation. Moreover, to avoid an expensive Monte-Carlo simulation, SiSMA calculated the mismatch by the extraction and analysis of a covariance matrix. It would be slow to analyze the covariance matrix for thousands of devices. In addition, the entire circuit is analyzed twice and is computationally expensive for large-scale problems. Therefore, there is still a need to develop a fast transient mismatch analysis that requires improvements in two-fold: a different non-Monte-Carlo method and an efficient macromodel by the nonlinear model order reduction.

To this end, this paper introduces a fast non-Monte-Carlo mismatch analysis by an incremental and stochastic trajectory piecewise linear macromodel, namely *isTPWL method*. First, we develop a transient mismatch model and its macromodeling. We linearize the SDAE along a number of snapshots on a nominal transient trajectory, and add the stochastic current source (for mismatch) at each snapshot as a perturbation. This is more accurate than considering the mismatch through an *ic* condition [6]. Along the snapshots of the nominal transient trajectory, we further apply an improved trajectory-piecewise-linear (TPWL) model order reduction [8–10] to generate a stochastic nonlinear macromodel, and apply it for a fast transient mismatch analysis along the full transient trajectory. Our approach applies an incremental aggregation on those local tangent subspaces, linearized at snapshots. It reduces the computational complexity of [10] yet improves the accuracy of [8]. As shown by experiments, our isTPWL method is 5 times more accurate than the work in [8] and is 20 times faster than the work in [10] on average. In addition, utilizing nonlinear macromodels reduces the runtime by up to 25 times compared to the use of the full model during the mismatch analysis.

Next, in order to efficiently solve the SDAE without applying the Monte-Carlo iterations or analyzing the expensive co-variance matrix [6], we describe the stochastic variation by a spectral stochastic method based on stochastic orthogonal polynomials (SOPs) and form an according SDAE [11]. The SOPs have been applied to deal with the linear interconnection variation in [12]. Our paper is the first to study how to apply SOPs for nonlinear analog circuits during a non-Monte-Carlo mismatch analysis. Experiments show that compared to the Monte-Carlo method, our method is 1000 times faster with a similar accuracy.

The rest of the paper is organized in the following manner. In Section 2, we present the background of the mismatch model and the nonlinear model order reduction. In Section 3, we discuss a transient mismatch analysis in SDAE, including a perturbation analysis and a non-Monte-Carlo analysis by the SOP expansions.

*The work at UC-Riverside is funded in part by NSF CAREER Award No. CCF-0448534, in part by NSF grant under No. OISE-0623038 and in part by National Natural Science Foundation of China (NSFC) grant under No. OISE-0929699. Address comments to haoyu@ntu.edu.sg and stan@ee.ucr.edu.

978-1-4244-5766-3/10/$26.00 ©2010 IEEE

We develop an incremental and stochastic TPWL model order reduction for mismatch in Section 4. We present numerical results in Section 5, and conclude the paper in Section 6.

2. BACKGROUND

2.1 Mismatch Model

The mismatch model and analysis is the key to a precision analog circuit design such as ADC/DACs. There are global mismatch and local mismatch. The local mismatch is the most difficult one to analyze and hence it is the focus of this paper.

The local mismatch is stochastic and process-parameter dependent. Most CMOS mismatch models are based on the Pelgrom's work [3] that relates the local mismatch variance of one electrical parameter (such as the channel current I_d) with geometrical parameters (such as the area A) by a geometrical dependence equation

$$\sigma_{I_d} = \frac{\kappa^\beta}{\sqrt{A}} \tag{1}$$

for two devices closely laid out, i.e. a local variation. Note that $A = W \cdot L$ is the area of a width W and length L, and κ^β is an extracted constant depending on the operating region β.

For other transistors such as diode, BJT and etc. and to consider process parameters other than the geometry, a more general purposed mismatch model can be derived through a so-called backward propagation of variance (BPV) method [5]. For example, the base-current I_b depends on the emitter area, sheet resistance and base current density. The BPV model then relates the local mismatch of an electrical property e with those process parameters p_l by a first-order sensitivity equation

$$\sigma_e = \sum_l (\frac{\partial e}{\partial p_l})\sigma_{p_l}. \tag{2}$$

Based on the above mismatch model, we introduce a non-Monte-Carlo transient mismatch analysis for a large number of transistors in Section 3.

2.2 Nonlinear Model Order Reduction

The nominal nonlinear circuit is described by the following differential-algebra-equation (DAE)

$$f(x, \dot{x}, t) = \mathcal{B}u(t), \tag{3}$$

where x ($\dot{x} = dx/dt$) is the state variable including nodal voltage and branch current, $f(x, \dot{x}, t)$ is to describe the nonlinear $i - v$ relation, and $u(t)$ is the external sources with a topology matrix $\mathcal{B}$ describing how to add them into the circuit. The time to solve (3) comes from three parts: device evaluation, matrix factorization, and time-step control and integration. When the circuit size is large or when devices are latent in most of time, the portion of runtime mainly comes from the matrix factorization. Under this condition, the use of model order reduction to reduce the circuit size is effective to reduce the overall runtime and hence can be applied in a transient mismatch analysis as well.

Model order reduction is basically to find a small dimensioned subspace that can represent the original state space with a preserved system response. This can be usually realized in the view of a coordinate transformation. For linear circuits, the coordinate transformation can be described by a linear mapping

$$z = V^T x, \quad x = Vz,$$

where a small dimensioned projection matrix V ($\in N \times q$, $q << N$) can be constructed from the first few dominant bases spanning a space of moments (or derivatives of transfer functions) [13, 14].

There are many model order reductions [8–10, 15] developed for nonlinear circuits as well. Similarly, there can be a nonlinear mapping defined by a function ϕ

$$z = \phi(x), \quad x = \phi^{-1}(z).$$

For the simplicity of illustration, we assume an ordinary differential equation (ODE) form below

$$\dot{x} = f(x, t) + \mathcal{B}u(t)$$

for the DAE in (3). Since

$$\dot{z} = \frac{d\phi}{dx}\frac{dx}{dt} = (\frac{d\phi}{dx}f(x, t)) + (\frac{d\phi}{dx}\mathcal{B})u(t),$$

we have

$$\dot{z} = \hat{f}(z, t) + \hat{\mathcal{B}}u(t),\ \hat{f}(z, t) = \left[\frac{d\phi}{dx}f(x, t)\right]|_{x=\phi^{-1}(z)},\ \hat{\mathcal{B}} = \frac{d\phi}{dx}\mathcal{B}. \tag{4}$$

As such, if we can find a lower-dimensioned mapping function ϕ ($\in N \times q$), the original nonlinear system can be reduced within a tangent subspace spanned by $d\phi/dx$ (or called manifold).

The work in [10] related the above nonlinear mapping function ϕ with a trajectory piecewise linear (TPWL) method [8]. It leads to a local two-dimensional (2D) projection [10]. Since such a local 2D-projection is constructed from local tangent subspaces, it maintains a high accuracy. However, it could be computationally expensive to project and store when the number of local tangent subspaces is large. On the other hand, the TPWL method [8] approximated the nonlinear mapping function ϕ by aggregating those local tangent subspaces with the use of a global singular-value-decomposition (SVD). This results in a global one-dimensional (1D) projection. Obviously, the global 1D-projection leads to an efficient projection and runtime. On the other hand, the accuracy of the TPWL model order reduction is limited because the information in the dominant bases of each local tangent subspace are lost during the global SVD [10]. In Section 4, we introduce an incremental aggregation that can balance the speed and accuracy. In addition, we also extend the nonlinear model order reduction to consider the stochastic mismatch.

3. STOCHASTIC TRANSIENT MISMATCH ANALYSIS

3.1 Stochastic Mismatch Current

Directly adding the stochastic mismatch ξ as a parameter into the state variable x of (3) would lead to a difficulty that $f(x, \dot{x}, \xi)$ may not be differentiable. Similar to SiSMA [6], we model the mismatch as a current source $\mathbf{i}(x, \xi)$ added at the right-hand-side (rhs) of (3)

$$f(x, \dot{x}, t) = \mathcal{F}\mathbf{i}(x, \xi) + \mathcal{B}u(t). \tag{5}$$

Here, $\mathcal{F}$ is the topology matrix describing how to connect $\mathbf{i}$ into the circuit.

Based on the BPV equation (2), the stochastic current source $\mathbf{i}$ has a form of

$$\mathbf{i}(x, \xi) = n(x)\sum_l g^\beta(p_l)\xi_l. \tag{6}$$

Here, ξ_l is a random variable associated with a stochastic distribution $W(\xi_l)$ for the parameter p_l. $n(x)$ describes the biasing-dependent condition (depending on $x, \dot{x}$), provided from a nominal transient simulation. $g^\beta(p_l)$ is a constant for the parameter p_l at operating region β. For example, for one CMOS transistor with respect to the parameter area A, ξ_A is one Gaussian random variable, $g^\beta(A)$ is $\kappa^\beta/\sqrt{A}$ and $n(x)$ becomes I_d. In general, $g^\beta(p_l)$ can be either derived based on the analytical device equations or can be practically characterized from measurements [5].

3.2 Perturbation Analysis

Assuming that the impact of the local mismatch is small, (5) can be solved by treating the rhs-term for mismatch as a perturbation to the nominal trajectory $x^{(0)}(t)$ of the circuit. Here, $x^{(0)}(t)$ is the nominal state variable or solution of the nonlinear circuit equation below

$$f(x^{(0)}, \dot{x}^{(0)}, t) = \mathcal{B}u(t). \tag{7}$$

With a first-order Taylor expansion of $f(x, \dot{x}, t)$ in (5), it leads to

$$f(x^{(0)}, \dot{x}^{(0)}, t) + \frac{\partial f(x, \dot{x}, t)}{\partial x}(x - x^{(0)}) + \frac{\partial f(x, \dot{x}, t)}{\partial \dot{x}}(\dot{x} - \dot{x}^0)$$
$$= \mathcal{F}\mathbf{i}_n(x^{(0)}, \xi) + \mathcal{B}u(t). \tag{8}$$

Or

$$G(x^{(0)},\dot{x}^{(0)})x_m + C(x^{(0)},\dot{x}^{(0)})\dot{x}_m = \mathcal{F}\mathbf{i}_n(x^{(0)},\xi), \quad (9)$$

where

$$G(x^{(0)},\dot{x}^{(0)}) = \frac{\partial f(x,\dot{x},t)}{\partial x}\mid_{x=x^{(0)},\dot{x}=\dot{x}^{(0)}}$$
$$C(x^{(0)},\dot{x}^{(0)}) = \frac{\partial f(x,\dot{x},t)}{\partial \dot{x}}\mid_{x=x^{(0)},\dot{x}=\dot{x}^{(0)}} \quad (10)$$

are the linearized conductive and capacitive components stamped by the companion models in SPICE, and $x_m = x - x^{(0)}$ is the first-order perturbed mismatch response. Recall that $x^{(0)}(t)$ and $\dot{x}^{(0)}(t)$ are a number of time-dependent biasing points along the transient trajectory.

3.3 Non-MC Analysis by SOP Expansion

Next, instead of performing the expensive Monte-Carlo or the correlation analysis, the perturbed SDAE (9) with the random variable ξ is solved through an expansion of the stochastic orthogonal polynomial (SOP) [11, 12]. Different random processes are related to the different orthogonal polynomial. The 'Homogeneous Chaos' can be used as the span of Hermite polynomial functionals for a Gaussian process [11, 12]. In this paper, we assume that the random process parameters for the local mismatch have a Gaussian distribution. Therefore, an according Hermite polynomial (one variable)

$$\Phi(\xi) = [\Phi_1(\xi),\Phi_2(\xi),\Phi_3(\xi),...,]^T = [1,\xi,\xi^2-1,...,]^T \quad (11)$$

is used to construct the expansion basis to calculate the mean and the variance of $x_m(t)$.

The stochastic state variable $x_m(t)$ is first expanded by

$$x_m(t) = \sum_i \alpha_i(t)\Phi_i(\xi). \quad (12)$$

Note that for different random processes, many other orthogonal polynomials can be selected as well based on a so-called *Askey scheme* [11].

Then, when applying the inner-product of the residue error

$$\begin{aligned}\Delta(\xi) &= G(x^{(0)},\dot{x}^{(0)})\sum_i \alpha_i(t)\Phi_i(\xi) + C(x^{(0)},\dot{x}^{(0)})\sum_i \dot{\alpha_i}(t)\Phi_i(\xi)\\ &- \mathcal{F}n(x^{(0)})\sum_l g^\beta(p_l)\xi_l\end{aligned}$$

by the orthogonal basis $\Phi_j(\xi)$, it results in

$$< \Delta(\xi),\Phi_j(\xi) >= \int_\xi \Delta(\xi)\Phi_j(\xi)W(\xi)d\xi = 0. \quad (13)$$

Here, $W(\xi)$ is the probability distribution of the random variable ξ. We assumed a Gaussian distribution of $W(\xi)$ for all parameters in this paper.

Without the loss of generality, for one random variable ξ of one geometrical parameter p, it is easy to verify that (13) leads to

$$\alpha_0 = 0, \quad \alpha_2 = 0$$
$$G(x^{(0)},\dot{x}^{(0)})\alpha_1(t) + C(x^{(0)},\dot{x}^{(0)})\dot{\alpha_1}(t) = \mathcal{F}n(x^{(0)})g^\beta(p) \quad (14)$$

with a second-order expansion of $x_m(\xi)$. The according standard-deviation is thereby given by

$$Var < x_m(\xi) >= \alpha_1^2 Var(\xi) + \alpha_2^2 Var(\xi^2-1) = \alpha_1^2.$$

The first-order SOP coefficient $\alpha_1(t)$ in (14) can be solved by a Backward-Euler integration

$$(G_k + \frac{1}{h}C_k)\alpha_1(t_k) = \frac{1}{h}C_k\alpha_1(t_k-h) + \mathcal{F}\mathbf{i}_k \quad (15)$$

where

$$G_k = G(x_k^{(0)},\dot{x}_k^{(0)}), \quad C_i = C(x_k^{(0)},\dot{x}_k^{(0)}), \quad \mathbf{i}_k = n(x_k)\sum_l g^\beta(p_l)$$

are Jacobians and the mismatch current-source at the k-th time-instant along the nominal trajectory $x^{(0)}$.

Clearly, a native application of the above perturbation-based mismatch analysis is still slow, since G_k, C_k and $\mathbf{i}_k$ have to be evaluated during every time-step along the nominal trajectory. Instead of linearizing along the full nominal trajectory, in Section 4, only K snapshots along the nominal trajectory are used in the frame of a macromodeling.

3.4 One CMOS Transistor Example

For one CMOS transistor with a geometric parameter A and the according Gaussian random variable ξ_A, (14) becomes

$$(G_k + \frac{1}{h}C_k)\alpha_1(t_k) = \frac{1}{h}C_k\alpha_1(t_k-h) + \frac{\kappa^\beta}{\sqrt{A}}\cdot(I_d)_k$$

at the k-th time-step. Recall that G_k, C_k and $(I_d)_k$ are the nominal conductance (g_{ds}), capacitance (c_{ds}) and channel current I_d evaluated at t_k, $g^\beta(A)$ is $\kappa^\beta/\sqrt{A}$ and $n(x)$ becomes I_d. Note that κ^β is the extracted constant from Pelgrom's model.

As such, the transient mismatch voltage ($x_m = \alpha_1(t)\Phi_1(\xi_A)$) of this transistor has a time-varying standard variance $\alpha_1(t)^2$, solved from the above perturbation equation. Usually, $\kappa^\beta/\sqrt{A}$ is about few percentages of the nominal channel current I_d. More importantly, for thousands of different typed transistors, we can simultaneously solve the transient mismatch vector using (14) with a generally characterized $g^\beta(p_l)$ by the BPV model [5].

4. MACROMODELING FOR MISMATCH

Instead of performing a full simulation for the nominal transient and transient mismatch, we can take K snapshots along a nominal transient trajectory and find subspaces, or macromodels, from the K snapshots with respect to right-hand-sides of the nominal input and stochastic current-source, respectively. Afterwards, efficient transient and transient mismatch analysis can be performed along the full transient trajectory using macromodels. In the following, for the nominal transient, we first introduce an incremental TPWL method to balance the accuracy and efficiency when generating the macromodel. We then extend this approach to handle the stochastic mismatch.

4.1 Incremental TPWL

4.1.1 Local Tangent Subspace

Given a set of typical inputs, we take K snapshots $\{x_1^{(0)},...,x_K^{(0)}\}$ along a nominal transient trajectory $x^{(0)}(t)$, and linearize the DAE (3) at K snapshots (or biasing points), with the first snapshot x_1 taken at the *ic* point. The linearized DAE at k-th ($k = 1,...,K$) snapshot is

$$G_k(x - x_k^{(0)}) + C_k(\dot{x} - \dot{x}_k^{(0)}) = \delta_k,\ \delta_k = \mathcal{B}u(t_k) - f(x_k^{(0)},\dot{x}_k^{(0)},t_k)$$

where δ_k represents the rhs-source and the 'non-equilibrium' update. In frequency domain, $x_k^{(0)}$ at the k-th snapshot is contained by a subspace of moments $\{A_k, A_kR_k, A_k^2R_k,...,\}$ expanded at a frequency-point s_0, where

$$A_k = (G_k + s_0C_k)^{-1}C_k,\ R_k = (G_k + s_0C_k)^{-1}\delta_k$$

are two moments matrices.

With the use of a block-Arnoldi orthonormalization [14], a q'-th order projection matrix V_k ($\in N \times q'$) with q' bases

$$V_k = [v_k^1, v_k^2, ..., v_k^{q'}]$$

can be constructed locally to represent that local subspace. We call v_k^i ($k = 1,...,K; i = 1,...,q'$) as the first-$q'$ *dominant bases* of one V_k. Here, for each v_k^i the subscript describes the index of the local subspace, and the superscript describes the index of the order of the dominant base. It finds a linear coordinate transformation V_k that maintains $||z - z_k^{(0)}|| \approx ||x - x_k^{(0)}||$. Moreover, as discussed in the following part, those V_ks span a subspace for $d\phi/dx$, the tangent (or manifold) of the mapping function ϕ introduced in Section 2. As such, we call the space spanned by V_ks as *local tangent subspace*.

4.1.2 Local and Global Projection

As discovered in [10], one approach to approximate the nonlinear mapping function ϕ introduced in Section 2 is

$$x = \phi^{-1}(z) \approx \sum_{k=1}^{K} w_k[x_k + V_k(z - z_k^{(0)})] \quad (16)$$

and

$$z = \phi(x) \approx \sum_{k=1}^{K} w_k[z_k + V_k^T(x - x_k^{(0)})] \quad (17)$$

where w_k ($\sum_{k=1}^{K} w_k = 1$) is the weighted kernel function, which depends on the distance between a point on the trajectory and a linearization point [8].

Based on the equations (4), (16) and (17), a nonlinear model order reduction is derived in terms of a local two-dimensional (2D) projection

$$\sum_{l=1}^{K}\sum_{k=1}^{K} w_l w_k \left[V_l^T G_k V_k(z - z_k^{(0)}) + V_l^T C_k V_k(\dot{z} - \dot{z}_k^{(0)})\right] = \sum_{l=1}^{K} w_l V_l^T \delta_k, \quad (18)$$

where we assume that all V_ks are reduced to the same order q'. For circuits with a sharp transition (input) or strong nonlinearity (device), the number of sampled snapshots is required to be large to maintain a high accuracy. For such kind of circuits, our experiments show that the number of sampled snapshots (or neighbors) has to be large to produce a good accuracy. As such, the computational cost would be prohibitive by the local 2D-projection (18) in [10].

On the other hand, the TPWL method in [8] approximates the nonlinear mapping function ϕ by aggregating the local subspace V_k ($\in N \times q'$) into an unified global subspace $span\{V_1, V_2, ..., V_K\}$, which is further compressed into a lower-dimensioned subspace $\mathcal{V}$ ($\in N \times q$, $q << N$) by a singular-value-decomposition (SVD)

$$\mathcal{V} = SVD_q\left([V_1, V_2, ..., V_K]\right).$$

We call this procedure as *global aggregation*. A global aggregation results in a global one-dimensional (1D) projection by

$$\sum_{k=1}^{K} w_k \left[\mathcal{V}^T G_k \mathcal{V}(z - z_k^{(0)}) + \mathcal{V}^T C_k \mathcal{V}(\dot{z} - \dot{z}_k^{(0)})\right] = \sum_{k=1}^{K} w_k \mathcal{V}^T \delta_k. \quad (19)$$

Clearly, such a global 1D-projection has a smaller projection time and storage than the local 2D-projection. However, at the same time, since the dominant bases of those local V_ks are interpolated by the global aggregation, the global 1D-projection usually requires a higher-order q to achieve an accuracy similar to the local 2D-projection with the order q' ($q' < q$) [10].

4.1.3 Incremental Aggregation

The local 2D-projection (18) requires a longer runtime and larger storage compared to the global 1D-projection (19). On the other hand, the global 1D-projection (19) by $\mathcal{V}$ is less accurate than the local 2D-projection (18). As such, we need a procedure that can balance both of the accuracy and efficiency. This paper shows the following observation.

The $d\phi/dx$ (manifold) is covered by those local tangent subspaces $\{V_1, V_2, ..., V_K\}$ along the trajectory, where each V_k is further composed of different orders of dominant bases: $\{v_k^1, v_k^2, ..., v_k^{q'}\}$. An effective aggregation thereby needs to consider the order or the dominance of those bases. This motivates us to first decompose the space spanned by those local tangent subspaces according to the order. As such, (16) becomes

$$\begin{aligned} x = \phi^{-1}(z) &\approx \sum_{k=1}^{K} w_k x_k + \sum_{k=1}^{K} w_k \sum_{p=1}^{q'} v_k^p (z - z_k^{(0)}) \\ &= \sum_{k=1}^{K} w_k x_k + \sum_{p=1}^{q'}\sum_{k=1}^{K} v_k^p w_k (z - z_k^{(0)}) \\ &= \sum_{k=1}^{K} w_k x_k + \left[v_1^1 w_1 (z - z_1^{(0)}) + ... + v_K^1 w_K (z - z_K^{(0)})\right] \\ &+ ... + \left[v_1^q w_1 (z - z_1^{(0)}) + ... + v_K^q w_K (z - z_K^{(0)})\right]. \end{aligned} \quad (20)$$

Therefore, we can form a *global tangent subspace* in the order of the dominant bases by

$$span\{v_1^1, v_2^1, ..., v_K^1\},\ ...,\ span\{v_1^{q'}, v_2^{q'}, ..., v_K^{q'}\}.$$

A global projection matrix $\mathcal{V}$ is accordingly constructed below in a fashion of an *incremental aggregation*. We first aggregate each global tangent subspace by orders

$$\mathbf{V}_1 = SVD_q\left([v_1^1, ..., v_K^1]\right),\ ...,\ \mathbf{V}_{q'} = SVD_q\left([v_1^{q'}, ..., v_K^{q'}]\right). \quad (21)$$

I.e., we identify a $\mathbf{V}_p$ ($p = 1, ..., q'$) to represent the p-th order global tangent subspace.

Then, the global projection matrix $\mathcal{V}$ is further aggregated

$$\mathcal{V} = SVD_q\left([\mathbf{V}_1, \mathbf{V}_2, ..., \mathbf{V}_{q'}]\right) \quad (22)$$

by those global tangent subspace in an order of dominance. As shown by experiments, usually we can choose a much lower q' ($q' << q$) for each local tangent subspace V_k, and the order q depends on the number of snapshots. For circuits with the sharp transition (input) or strong nonlinearity (device), the number of snapshots is large and so does q.

Clearly, as the local tangent subspace is incrementally aggregated according to their ordered bases, the information of those dominant bases at low orders are preserved. As shown by experiments, when compared to the previous TPWL method [8], this incremental aggregation results in a better accuracy yet with a similar computational cost in the projection time and memory storage. Moreover, our incremental aggregation can also consider more sampled biasing (linearizion) points than the approach in [10], whereas the computational cost of of the local 2D-projection would increase dramatically.

4.2 Stochastic Extension for Mismatch

We further extend the above discussion to build the TPWL macromodel for the stochastic mismatch. Instead of linearizing the DAE (3), we linearize the SDAE (14) at K snapshots along the nominal trajectory similarly, and construct the local tangent subspace V_k by

$$A_k' = (G_k + s_0 C_k)^{-1} C_k,\ R_k' = (G_k + s_0 C_k)^{-1} \delta_k'$$

where δ_k' is determined by the non-equilibrium correction associated with $\mathcal{F}\mathbf{i}_k$. Then, we can build the similar incrementally aggregated mapping $\mathcal{V}$ through (21) and (22).

Afterwards, the global macromodel is constructed from a set of weighted local macromodels by

$$\sum_{k=1}^{K} w_k \cdot \left[\mathcal{V}^T G_k \mathcal{V}\alpha_1(t) + \mathcal{V}^T C_k \mathcal{V}\dot{\alpha_1}(t) - \mathcal{V}^T \mathcal{F} \cdot \mathbf{i}_k\right] = 0 \quad (23)$$

to calculate the transient mismatch. We call such a macromodeling as *isTPWL method*. Using such a macromodel sampled from K snapshots, we can then efficiently perform a transient mismatch analysis for the full trajectory.

5. EXPERIMENTAL RESULTS

The proposed mismatch algorithm is implemented in C and Matlab. A modernized Spice3 (http://ngspice.sourceforge.net/) is used to generate the K snapshots of a nominal trajectory and to extract the mismatch current model. The SOP expansion, Backward-Euler, and incremental and stochastic TPWL (isTPWL) are implemented in Matlab. For the comparison, the TPWL method and maniMOR method are implemented exactly following the procedure described in [8] and [10], respectively. For example, as for the TPWL method [8], the state variables at snapshots are added to have a 'richer' information during the global aggregation. The flow under the Monte-Carlo analysis is also implemented as the baseline with 1000 iterations. All experimental results are measured on an Intel duel-core 2.0GHZ PC with 2GB memory. We compared the accuracy and study

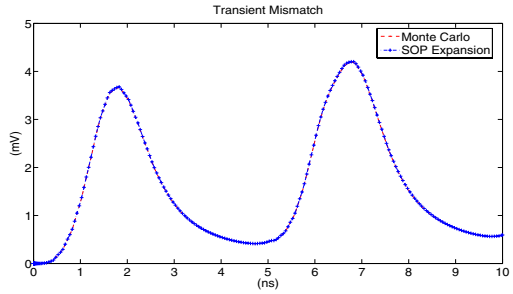

Figure 1: Transient mismatch (the time-varying standard deviation) comparison at output of a bjt-mixer with distributed inductor: the exact by Monte-Carlo and the exact by SOP-expansion.

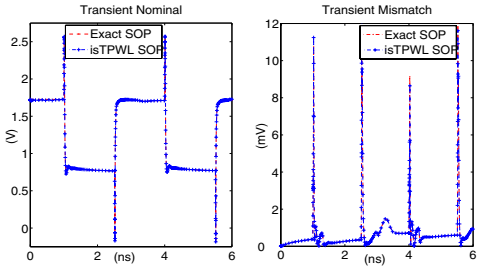

Figure 2: Transient nominal ($x^{(0)}(t)$) (a) and transient mismatch ($\alpha_1(t)$) (b) for one output of a coms-comparator by the exact SOP and the isTPWL.

the scalability of our method with four industrial analog/RF circuits. They contain different transistors such as diode, bjt and cmos. The circuits also include the extracted parasitics so the matrix time is dominant. For the characterization of $g^{\beta}(p_l)$, we use Pelgrom's model for CMOS transistors and BPV model for diode and bjt, all resulted in $\sim 10\%$ variation from the nominal bias $n(x)$ (for example, I_d for CMOS transistor). In addition, we measure the waveform error by taking the averaged difference of two waveforms. Three waveforms are measured at each time-step: one is the transient nominal ($x^{(0)}(t)$), the other is the transient mismatch ($\alpha_1(t)$, the time-varying standard deviation), and the last one is the transient ($x(t)$, the nominal plus the standard deviation).

5.1 Comparison of Mismatch Waveform-Error and Runtime

We first compare the waveform accuracy of the transient mismatch between the MC method (1000 iterations) and the exact SOP expansion, and further compare the accuracy with the isTPWL macromodel. In addition, we also compare the waveform of the transient mismatch and the waveform by adding mismatch as one initial condition as SiSMA [6] does. Finally, we summarize the runtime and waveform error in a table.

The first example is a bjt-mixer circuit including an extracted distributed inductor with 238 state variables. We compare the waveforms by solving the perturbed SDAE (9) with use of the Monte-Carlo (MC) analysis and the SOP expansion, respectively. We apply a MC with Gaussian distribution 1000 times at one time-step and calculate the time-varying standard deviation. It takes 348 seconds for the transient mismatch by the MC analysis, and 0.20 second (more than 1000 times speedup) for the exact SOP expansion up to the second order with error less than 0.18%. Clearly, as shown in Fig.1, the two waveforms of transient mismatches are virtually identical.

Next, we show a further speed improvement by macromodeling. The second example is a cmos-comparator including an extracted

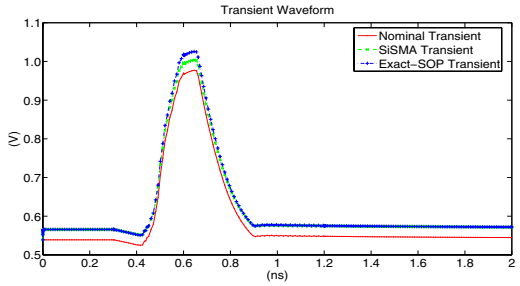

Figure 3: Transient waveform comparison at output of a diode-chain: the transient nominal, the transient with mismatch by SiSMA (adding mismatch at ic only), the transient with mismatch by our method (adding mismatch at transient trajectory).

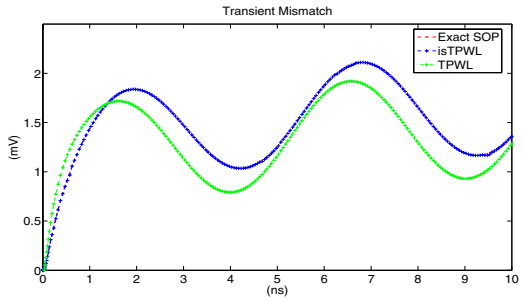

Figure 4: Transient mismatch ($\alpha_1(t)$, the time-varying standard deviation) comparison at output of a bjt-mixer with distributed substrate: the exact by SOP-expansion, the macromodel by TPWL (order 45), and the macromodel by isTPWL (order 45). The waveform by isTPWL is visually identical to the exact SOP.

power supply with 654 state variables. We compare waveforms of the exact SOP expansion and the one further reduced by isTPWL. Fig.2 (a) shows the comparison of the transient nominal, and Fig.2 (b) shows the comparison of the transient mismatch. 75 snapshots are used to generate the macromodel, and the original model is reduced to a macromodel with the order of 60. For a short-transient with 228 time-steps, it takes 0.39s for the exact and 0.08s for the isTPWL (5 times speedup). The waveforms error by isTPWL is 0.62%.

We further compare the transient mismatch waveforms on how to add the mismatch. The first is to add the stochastic mismatch only as the ic condition, the procedure used in SiSMA [6]. The second is our approach by adding the stochastic mismatch during every time-step. A diode-chain with 802 state variables is used. Fig.4 shows one waveform of the transient nominal, and two waveforms with mismatches added differently. Clearly, the waveform with mismatch added at ic shows a non-negligible difference.

Finally, we summarize the runtime and error of four different analog/RF circuits in Table 1. The waveform error here is defined as the relative difference between the exact and the macromodel. The runtime here is the total simulation time. Column 1–5 summarize the circuit type, size, time-steps in transient simulation, snapshot points, and reduced order. Column 6 shows the runtime of MC by 1000 iterations, Column 7–8 show the runtime to simulate the exact model with the SOP expansion and the error compared to MC, Column 9–10 show the runtime and error of macromodels reduced by isTPWL. We find that the SOP expansion reduces the runtime by 1000 times yet with an error of 0.23% on average. Moreover, the macromodel by isTPWL further reduces the runtime up to 25 times (diode chain) yet with an er-

Table 1: Scalability comparison of runtime and error for the exact model with MC, the exact model with SOP and the isTPWL macromodel with SOP.

1	2	3	4	5	6	7	8	9	10
Ckt	# of nodes	# of steps	# of snapshots	# of orders	MC	Exact SOP		SOP+isTPWL	
					time(s)	time(s)	error	time(s)	error
diode-chain	802	225	24	25	520.1	0.53	0.41%	0.02	0.43%
bjt-mixer-1	238	135	25	25	338.0	0.34	0.29%	0.02	0.36%
bjt-mixer-2	1248	219	83	45	348.0	0.20	0.18%	0.04	0.24%
cmos-comparator	654	228	75	60	412.1	0.39	0.41%	0.08	0.62%

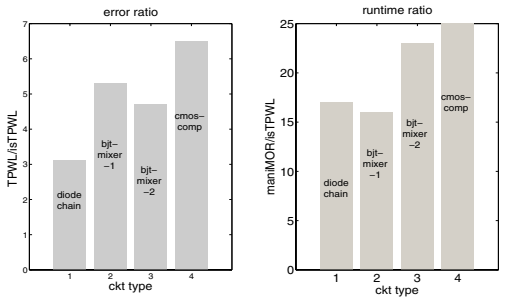

Figure 5: (a) Comparison of the ratio of the waveform error by TPWL and by isTPWL under the same reduction order; (b) Comparison of the ration of the reduction runtime by maniMOR and by isTPWL under the same reduction order. In both cases, isTPWL are used as the baseline.

ror up to 0.43%. This demonstrates the efficiency and accuracy of our isTPWL method for the transient mismatch analysis.

5.2 Comparison of TPWL Macromodel

In this part of experiment, we further show the accuracy and runtime improvement by isTPWL. First, Fig.4 presents the transient-mismatch waveform comparison for a bjt-mixer including the distributed substrate with total 1248 state variables. 83 snapshots are used and both TPWL and isTPWL reduce the original model to a macromodel with the order of 45. We find that the waveform by isTPWL is visually identical to the exact SOP but the waveform by TPWL [8] shows a non-negligible waveform error 4.5 times larger than the one by our isTPWL. Fig.5 further summarizes the comparison by the four circuits used before. Fig.5 (a) is the comparison of the ratio (TPWL vs. isTPWL) of waveform-errors for simulated macromodels by TPWL [8] and by isTPWL under the same reduction order. Fig.5 (b) is the comparison of the ratio (maniMOR vs. isTPWL) of the reduction time for reduced macromodels by maniMOR [10] and by isTPWL under the same reduction order. In both cases, isTPWL are used as the baseline when calculating the ratio. We find that our isTPWL method is 5 times more accurate than TPWL [8] and is 20 times faster than maniMOR [10] on average. This clearly demonstrates the advantage to use the incremental aggregation.

6. CONCLUSIONS

It is challenging to analyze the stochastic mismatch during the transient analysis. This paper presents a fast non-Monte-Carlo mismatch analysis. It models the mismatch by a current source associated with a random variable and forms a stochastic differential-algebra-equation (SDAE). The random variable in SDAE is expanded by stochastic orthogonal polynomials (SOPs). This leads to an efficient solution without using the Monte-Carlo or correlation analysis. Moreover, the SDAE is solved by an improved trajectory-piecewise-linear (TPWL) mode order reduction, called isTPWL. An incremental aggregation is introduced to balance the efficiency and accuracy when generating the macromodel. Experiments show that when compared to the Monte-Carlo method, our SOP expansion is 1000 times faster with a similar accuracy. Moreover, on average our isTPWL method is 5 times more accurate than the work in [8] and is 20 times faster than the work in [10]. In addition, the use of a reduced macromodel reduces the runtime by up to 25 times when compared to the use of a full model. Future study will show how to identify the subspace if there is a large nonlinear mismatch.

7. REFERENCES

[1] H. Masuda, S. Ohkawa, A. Kurokawa, and M. Aoki, "Challenge: Variability characterization and modeling for 65- to 90-nm processes," in *Proc. IEEE CICC*, 2005.

[2] J. Kim, K. Jones, and M. Horowitz, "Fast, non-monte-carlo estimation of transient performance variation due to device mismatch," in *Proc. ACM/IEEE DAC*, 2007.

[3] M. Pelgrom, A. Duinmaijer, and A. Welbers, "Matching properties of mos transistors," *IEEE JSSC*, pp. 1433–1439, 1989.

[4] J. Oehm and K. Schumacher, "Quality assurance and upgrade of analog characteristics by fast mismatch analysis option in network analysis environment," *IEEE JSSC*, pp. 865–871, 1993.

[5] C. McAndrew, J. Bates, R. Ida, and P. Drennan, "Efficient statistical BJT modeling, why beta is more than ic/ib," in *Proc. IEEE Bipolar/BiCMOS Circuits and Tech. Meeting*, 1997.

[6] G. Biagetti, S. Orcioni, C. Turchetti, P. Crippa, and M. Alessandrini, "SiSMA: A tool for efficient analysis of analog cmos integrated circuits affected by device mismatch," *IEEE Trans. on CAD*, pp. 192–207, 2004.

[7] A. Demir, E. Liu, and A.Sangiovanni-Vincentelli, "Time-domain non-monte carlo noise simulation for nonlinear dynamic circuits with arbitrary excitations," *IEEE Trans. on CAD*, pp. 493–505, 1996.

[8] M. Rewienski and J. White, "A trajectory piecewise-linear approach to model order reduction and fast simulation of nonlinear circuits and micromachined devices," *IEEE Trans. on CAD*, pp. 155–170, 2003.

[9] S.K.Tiwary and R.A.Rutenbar, "Faster, parametric trajectory-based macromodels via localized linear reductions," in *Proc. IEEE/ACM ICCAD*, 2006.

[10] C.J.Gu and J.Roychowdhury, "Model reduction via projection onto nonlinear manifolds, with applications to analog circuits and biochemical systems," in *Proc. IEEE/ACM ICCAD*, 2008.

[11] D. Xiu and G. Karniadakis, "The Wiener-Askey polynomial chaos expansion for stochastic differential equations," *SIAM J. Scientific Computing*, pp. 619–644, 2002.

[12] S. Vrudhula, J. M. Wang, and P. Ghanta, "Hermite polynomial based interconnect analysis in the presence of process variations," *IEEE Trans. on CAD*, pp. 2001–2011, 2006.

[13] P. Feldmann and R. W. Freund, "Efficient linear circuit analysis by pade approximation via the lanczos process," *IEEE Trans. on CAD*, pp. 639–649, 1995.

[14] A. Odabasioglu, M. Celik, and L. Pileggi, "PRIMA: Passive reduced-order interconnect macro-modeling algorithm," *IEEE Trans. on CAD*, pp. 645–654, 1998.

[15] J. Roychowdhury, "Reduced-order modelling of time-varying systems," in *Proc. ASPDAC*, 1999.