A Fast Non-Monte-Carlo Yield Analysis and Optimization by Stochastic Orthogonal Polynomials

FANG GONG, University of California at Los Angeles, US
XUEXIN LIU, University of California at Riverside, US
HAO YU, Nanyang Technological University, Singapore
SHELDON X.D. TAN, University of California at Riverside, US
JUNYAN REN, Fudan University, China
LEI HE, University of California at Los Angeles, US

Performance failure has become a significant threat to the reliability and robustness of analogue circuits. In this paper, we first develop an efficient non-Monte-Carlo (NMC) transient mismatch analysis, where transient response is represented by stochastic orthogonal polynomial (SOP) expansion under PVT variations and probabilistic distribution of transient response is solved. We further define performance yield and derive stochastic sensitivity for yield within the framework of SOP, and finally develop a gradient-based multi-objective optimization to improve yield while satisfying other performance constraints. Extensive experiments show that compared to Monte Carlo based yield estimation, our NMC method achieves up to $700X$ speedup and maintains 98% accuracy. Furthermore, multi-objective optimization not only improves yield by up to 95.3% with performance constraints, it also provides better efficiency than other existing methods.

Categories and Subject Descriptors: B.7.2 [**Hardware**]: Integrated Circuits—*Design Aids*

General Terms: Design, Algorithms, Performance

Additional Key Words and Phrases: Yield analysis, Circuit simulation, Monte Carlo, Yield optimization.

ACM Reference Format:
Gong, F., Liu, X., Yu, H., Tan, X., Ren, J., and He, L. 2010. A Fast Non-Monte-Carlo Yield Analysis and Optimization by Stochastic Orthogonal Polynomials. ACM Trans. Embedd. Comput. Syst. 9, 4, Article 39 (March 2010), 22 pages.
DOI = 10.1145/0000000.0000000 http://doi.acm.org/10.1145/0000000.0000000

1. INTRODUCTION

A robust design beyond 90nm is challenging due to PVT (Process, Voltage and Temperature) variations [Cox et al. 1985; Pelgrom et al. 1989; Lampaert et al. 1995; McAndrew et al. 1997; Schenkel et al. 2001; Drennan and McAndrew 2003; Biagetti et al. 2004; Vrudhula et al. 2006; Kim et al. 2007; Pileggi et al. 2008; Nassif and Nowka 2010; Liu et al. 2010; Gong et al. 2011; Wang et al. 2009; Gong et al. 2010a; Gong et al. 2010b; Gong et al. 2009]. The sources of process variation can come from etching, lithography, polishing, and stress, etc. For example, the proximity effect caused by

This work was partially supported by grants: (a) UC Discovery Grant program and UCLA; (b) NRF2010NRF-POC001-001 (Singapore); (c) MOE ACRF Tier-1 (Singapore).
Author's address: Fang Gong and Lei He, Electrical Engineering Department, University of California, Los Angeles, CA, 90095. Hao Yu, Electrical and Electronic Engineering, Nanyang Technological University, Singapore. Xuexin Liu and Sheldon X.D. Tan, Electrical Engineering Department, University of California Riverside, CA, 92521. Junyan Ren, School of Microelectronics, Fudan University, Shanghai, China.
Some initial results of this paper are published at ASPDAC'10[Yu et al. 2010].

DOI 10.1145/0000000.0000000 http://doi.acm.org/10.1145/0000000.0000000

stress from shallow-trench-isolation regions affect the stress in the channel of nearby transistors and therefore affect carrier mobility and threshold voltage. Process variation (or mismatch) significantly threatens not only the timing closure of digital circuits but also the functionality of analog circuits.

To ensure the robustness in terms of a high yield-rate, in addition to performance, a fast engine for yield estimation and optimization is needed to verify designs beyond 90nm. Note that there are two types of variations: systematic global variation, and s-tochastic local variation. The stochastic variation such as analog mismatch is the most difficult one. One either performs Monte Carlo [Swidzinski and Chang 2000] and its variants (e.g. Quasi Monte Carlo[Niederreiter 1992], Latin hypercube sampling (LHS) [Jaffari and Anis 2009b][Jaffari and Anis 2011], and Importance Sampling [Jaffari and Anis 2009a]) with thousand times of Monte-Carlo (MC) runs consuming engineering resources, or uses pessimistic process corners provided from the foundry. Since corners are usually pessimistic for yield estimation and Monte-Carlo is too painful for verification, the stochastic engine with a non-Monte-Carlo (NMC) approach is currently required for yield estimation and optimization.

To this end, the development of fast variation (mismatch) analysis to estimate yield is the first priority. Many NMC methods [Oehm and Schumacher 1993; Biagetti et al. 2004; Kim et al. 2007] have been developed recently for stochastic variation (mismatch) analysis. [Oehm and Schumacher 1993] first calculated *dc* sensitivities with respec-t to small device-parameter perturbations and scaled them as desired mismatches. [Kim et al. 2007] extended [Oehm and Schumacher 1993] by modeling *dc* mismatch as *ac* noise source. The speed of these equivalent mismatch simulations is hundred times faster than the Monte-Carlo simulations but accuracy remains a concern. SiS-MA [Biagetti et al. 2004] studied mismatch within the framework of the stochastic differential-algebra-equation (SDAE). The stochastic variational source is mapped into a noise current source introduced at *dc*, and the SDAE is solved similarly to deal with the transient noise [Demir et al. 1994] by analyzing the correlation. However, as [Biagetti et al. 2004] introduced random variables into the DAE, it is unknown whether the derivative of SDAE is still continuous. Moreover, SiSMA only included stochastic current source during *dc* based on the assumption that magnitude of the stochastic mismatch is much smaller than the nominal case. This may not hold to accurately describe the mismatch during a transient simulation. Therefore, a fast yet accurate transient mismatch analysis is still needed.

In this paper, we develops a fast NMC mismatch analysis by introducing the noise current sources along the linearized transient trajectory to model the PVT variations. By further representing the noise current source by the stochastic orthogonal polynomials (SOPs) [Xiu and Karniadakis 2002; Vrudhula et al. 2006], one can efficiently calculate the stochastic variation (mismatch) and its variance in only one transient simulation. This can result in a huge speedup with a similar accuracy when compared to the MC method. In addition, one needs to improve or optimize the yield by tuning parameters at nominal conditions to ensure a robust design. An efficient approach is to derive a gradient-based optimization method using the stochastic sensitivity of yield with respect to design parameters. Unfortunately, it is unknown how to calculate the stochastic sensitivity in the framework of the SOPs [Xiu and Karniadakis 2002; Vrudhula et al. 2006].

Our paper is the first to discuss the stochastic sensitivity analysis under SOP, which can be effectively deployed in any gradient-based optimization such as the sequential linear or quadratic programming. Moreover, it is necessary, even imperative, to optimize the objective function while satisfying other performance constraints (i.e. power consumption, area and etc.) [Sawaragi et al. 1985; Deb 2002]. To do so, we formulate a stochastic optimization problem and develop a multi-objective optimization algorithm, which can tune the design parameters along their gradient directions to reach a more robust design automatically.

Experiments show that fast mismatch analysis can achieve up to $700X$ speedup and maintain 98% accuracy when compared with Monte Carlo based methods; meanwhile, our multi-objective optimization can not only improve the yield rate up to 95.3% and satisfy other performance constraints, but also provide better efficiency than other existing methods.

In Section II, the NMC mismatch analysis is first developed as the foundation to calculate the yield. In Section III, the parametric yield is defined and the according yield problem formulation is presented. Section IV shows how to apply the fast NMC mismatch analysis to calculate the yield. Section V shows how to further obtain the yield and leverage the according stochastic sensitivity to optimize yield rate. In Section VI, the validity and efficiency of the proposed method is demonstrated by three different circuits: an operational amplifier, a schmitt trigger, and a SRAM-cell. The paper concludes in Section VII.

2. BACKGROUND

2.1. Stochastic Orthogonal Polynomial

We will first introduce stochastic orthogonal polynomial (SoP)[Vrudhula et al. 2006] or polynomial chaos in this section, which has been applied to the nanometer scale integrated circuit analysis [Xiu and Karniadakis 2002] in past a few years. Based on the *Askey scheme*, any stochastic random variable can be represented by stochastic orthogonal polynomials (SoPs), and the random variable with different probability distribution type is associated with different type of SoP.

For example, for white noise current source with random variable ψ, the Gaussian distribution of $f(\psi)$ can be spanned by Hermite polynomials $\Phi(\psi) = [1, \psi, \psi^2 - 1, \cdots]^T$ as follows

$$f(\psi) = \alpha_0\Phi_0 + \alpha_1\Phi_1 + \alpha_2\Phi_2 + \cdots = \sum_{i=0}^{n} \alpha_i\Phi_i. \tag{1}$$

Note that SoPs satisfy the following orthogonal property under so-called point-collocation:

$$\langle \Phi_i(\psi), \Phi_j(\psi) \rangle = \left\langle \Phi_i^2(\psi) \right\rangle \cdot \delta_{ij} \tag{2}$$

where δ_{ij} is the Kronecker delta and $\langle *, * \rangle$ denotes an inner product.

As such, when the SoP representation is available, the mean and variance of $f(\psi)$ can be obtained from one-time calculation using collocation (up to the second order expansion) by:

$$\begin{array}{l} E(f(\psi)) = \alpha_0 \\ Var(f(\psi)) = {\alpha_1}^2 + 2{\alpha_2}^2 \end{array}. \tag{3}$$

In this paper, we show how to apply the SoP technique for the non-Monte-Carlo mismatch analysis and yield estimation.

2.2. Stochastic Mismatch Analysis

We further review the existing works of mismatch analysis [Pelgrom et al. 1989; M-cAndrew et al. 1997; Drennan and McAndrew 2003; Biagetti et al. 2004]. Notice that we focus on the stochastic variation, or referred as local mismatch in this paper. We illustrate the stochastic mismatch analysis using the CMOS transistors, but a similar approach can be extended to other types of transistors by the so-called propagation of variance (POV) method[McAndrew et al. 1997; Drennan and McAndrew 2003].

The mismatch of one MOS transistor is usually modeled by Pelgrom's model [Pelgrom et al. 1989], which relates the local mismatch variance of one electrical parameter with geometrical parameters by

$$\sigma = \frac{\kappa^{\beta}}{\sqrt{W \cdot L}} \tag{4}$$

where κ^{β} is the additional fitting parameter.

To consider the local mismatch during circuit simulation without running Monte-Carlo, SiSMA [Biagetti et al. 2004] models the random local mismatch of a MOS transistor by a stochastic noise current source ζ, coupled with the nominal drain current I_D in parallel. ζ can be expressed by

$$\zeta = I_D^{\beta} t_m(W, L) \gamma(x, y). \tag{5}$$

Here, the I_D^{β} is determined by the operating region of MOS transistors; $t_m(W, L)$ considers the geometry of the device active area

$$t_m(W, L) = 1 + \frac{\kappa^{\beta}}{\sqrt{W \cdot L}}; \tag{6}$$

and $\gamma(x, y)$ refers to the sources of all the variations that depend on the device position, which can include the spatial correlation [Biagetti et al. 2004]. Here, $\gamma(x, y) = 1$ because all parameters are decoupled after the principal component analysis (PCA).

Note that the random variable in the stochastic current source can be expanded by the stochastic orthonormal polynomial (SOPs) [Xiu and Karniadakis 2002; Vrudhula et al. 2006]. For example, let's use the channel length L of one MOS transistor as the variation source. Assuming the variation of L is small, one can expand $t_m(W, L)$ around its nominal value $W_{(0)}$ and $L_{(0)}$ with Taylor expansion by

$$\begin{aligned} t_m(W, L) &= 1 + \tfrac{\kappa^{\beta}}{\sqrt{WL}} \\ &= 1 + \tfrac{\kappa^{\beta}}{\sqrt{W_{(0)}}} \left[\tfrac{1}{\sqrt{L_{(0)}}} - \tfrac{1}{2\sqrt{\left(L_{(0)}\right)^3}} \left(L - L_{(0)}\right) \right]. \\ &= 1 + \tfrac{\kappa^{\beta}}{\sqrt{W_{(0)}}} \left[\tfrac{1}{\sqrt{L_{(0)}}} - \tfrac{1}{2\sqrt{\left(L_{(0)}\right)^3}} \xi \right] \end{aligned} \tag{7}$$

Here, ξ is the random variable for the variation of the channel length L. One can describe ξ by the stochastic orthogonal polynomials (SOPs). Based on the *Askey scheme* [Xiu and Karniadakis 2002], a Gaussian distribution of ξ can be expanded using Hermite polynomials Φ_i $(i = 0, ..., n)$ by

$$\xi = \sum_{i=0}^{n} g_i' \Phi_i \tag{8}$$

where g_i' is the SOP expansion coefficient.

As such, one can summarize the expression of the stochastic current source ζ as

$$\begin{aligned} \zeta \ &= \ I_D^{\beta} \left[1 + \frac{\kappa^{\beta}}{\sqrt{W}_{(0)}} \left(\frac{1}{\sqrt{L_{(0)}}} - \frac{1}{2\sqrt{L_{(0)}^3}} \sum_{i=1}^{n} g_i' \Phi_i \right) \right] \\ &= \sum_{i=0}^{n} g_i \Phi_i, \end{aligned} \tag{9}$$

where g_i is the new expression of the expanded coefficients but with geometry-dependence. Knowing the expression of ζ for one parameter variation source, multiple process parameters p_i ($i = 1, \cdots, m$) can be considered by a vector of stochastic current source $\vec{\zeta}(t)$.

On the other hand, any integrated circuit is composed of passive and active devices described by a number of terminal-branch equations. According to KCL's law, one can obtain a differential-algebraic equation (DAE) as below

$$\frac{d}{dt}q(\vec{x}(t)) + f(\vec{x}(t), t) + B \cdot \vec{u}(t) = 0. \tag{10}$$

Here, $x(t)$ is vector of state variables consisting of node voltages and branch currents. $q(x(t), t)$ contains active components such as charges and fluxes. Also, $f(x(t), t)$ describes passive components, and $\vec{u}(t)$ denotes input sources. B describes how to connect sources into the circuit which is determined by circuit topology.

Similar to [Biagetti et al. 2004], one can add $\vec{\zeta}(t)$, representing the mismatch, to the right-hand-side (rhs) of the Differential Algebraic Equation (DAE)

$$\frac{dq(\vec{x}(t))}{dt} + f(\vec{x}(t)) + B \cdot \vec{u}(t) = T \cdot \vec{\zeta}(t), \tag{11}$$

which describes the circuit and system under stochastic variations. Note that T is the topology matrix describing how to connect $\vec{\zeta}(t)$ into the circuit and one can have

$$T \cdot \vec{\zeta}(t) = \sum_{i=1}^{m} T_{p_i} \zeta_{p_i} \tag{12}$$

for multiple parameters. For example, ζ_{p_i} is the mismatch current source for i-th parameter variation, which can be expanded using SOP shown in (9).

3. PROBLEM FORMULATION

In this section, we formulate the yield optimization problem based on the important observation: the parameter vector $\vec{p}$ can change the performance metric f_m (i.e. delay, output swing and etc), and further lead to circuit failure as well as yield loss. Noted that the parametric yield $Y(\vec{p})$ is defined as the percentage of manufactured circuits that can satisfy the performance constraints.

We can consider one output voltage that discharges from high to low as an example. Under process variation, the variable parameters in $\vec{p}$ can deviate from their nominal values and lead to transient variation (mismatch) waveform shown in Fig.(1).

The performance constraint $h(\vec{p})$ in this case can be defined as

$$h(\vec{p}) = f_m(t_{\max}) - f_{m_{threshold}} \leq 0. \tag{13}$$

This means that those discharge curves below $v_{threshold}$ at t_{max} belong to successful samples. In addition, one can plot the distribution of output voltages at t_{max} shown in Fig.(2). It is clear that samples located at the left of the performance constraint are successful ones, while those at the right are failures.

As such, parametric yield can be defined as:

$$Y(\vec{p}) = \int_S pdf(f_m(\vec{p}; t))dS \tag{14}$$

where S is the successful region and $pdf(f_m(\vec{p}; t))$ is the probability density function (PDF) of the performance metric $f_m(\vec{p}; t)$ of interest.

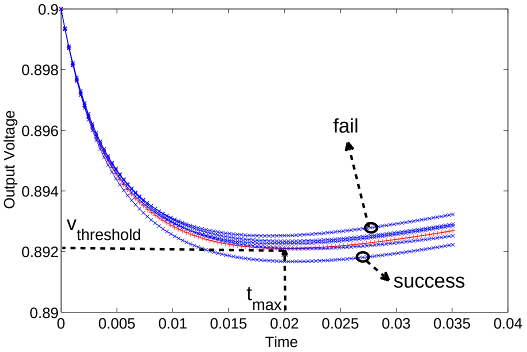


Fig. 1. Example of the stochastic transient variation or mismatch

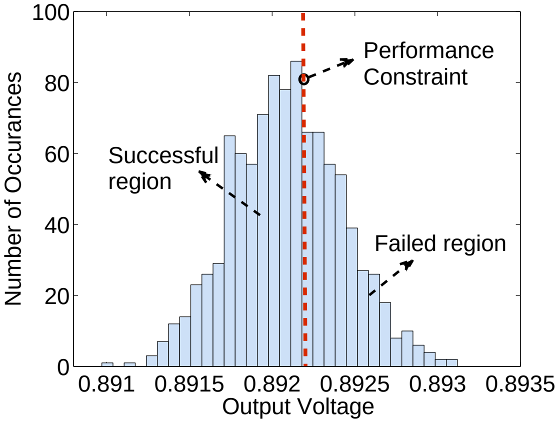


Fig. 2. Distribution of output voltage at t_{max}

In order to further increase the yield rate, one can tune the nominal values of variable parameters in order to enable more parameters satisfy the given performance constraints. Also, it is practical to consider multiple constraints (e.g. power consumption, area, bandwidth and etc.) which yields the stochastic multi-objective optimization problem as detailed below:

$$\begin{array}{cl} \text{maximize} & Y(\vec{p}) \\ \text{minimize} & P_c(\vec{p}) \\ & \\ \text{subject to} & \begin{array}{l} Y(\vec{p}) \geq \bar{Y} \\ P_c(\vec{p}) \leq \bar{P}_c \\ \mathcal{F}(\vec{p}) \leq \mathcal{F}_{max} \\ \vec{p}_{\min} \leq \vec{p}_0 \leq \vec{p}_{\max} \end{array} \end{array}$$

Here, $Y(\vec{p})$ is the parametric yield associated with the parameter vector $\vec{p}$ and $P_c(\vec{p})$ is the power consumption. $\mathcal{F}(\vec{p})$ denotes other performance metrics (such as area A and etc.), which define the feasible design-space. Moreover, $\bar{Y}$ and $\bar{P}_c$ are the minimum yield-rate and maximum power consumption (or targeted values) that can be accepted, respectively. Notice that the lower bound ($\bar{Y}$) and the upper bound ($\bar{P}_c$) are used for final verification.

Specifically, when the optimization is converged, the yield rate and power consumption under the optimal design point will be checked with $\bar{Y}$ and $\bar{P}_c$, and unsatisfied results will be rejected. Meanwhile, other constraints defined by $\mathcal{F}(\vec{p})$ should be satisfied.

Moreover, $\vec{p}$ is a vector of the process parameters with variations, and can be expressed as $\vec{p} = \vec{p}_0 + \delta\vec{p}$, where $\delta\vec{p}$ depends on $\vec{p}_0$ and needs to be updated for every feasible $\vec{p}_0$. Also, $\vec{p}_0$ is a vector of the nominal values assigned in the design stage, and $\delta\vec{p}$ consists of parameter variations with zero-mean Gaussian distributions. In addition, all nominal values of process parameters $\vec{p}_0$ are assumed to be limited within the feasible parameter space ($\vec{p}_{\min}, \vec{p}_{\max}$) and can be tuned for better yield rate.

One effective solution for this optimization is the gradient-based approach, which requires the calculation of the sensitivity in the stochastic domain. As discussed later, this paper develops a stochastic sensitivity analysis, which can be embedded into one sequential linear programming (SLP) to solve this optimization problem efficiently.

4. FAST NMC MISMATCH ANALYSIS

In this section, one NMC transient mismatch analysis is developed. Instead of performing the expensive Monte-Carlo or the correlation analysis, the perturbed SDAE (18) with the random variable ξ is solved through an expansion of the stochastic orthogonal polynomials (SOP) [Xiu and Karniadakis 2002; Vrudhula et al. 2006].

4.1. Transient Mismatch by Trajectory Perturbation

For illustration purpose, we can denote the $\frac{d}{dt}q(\vec{x}(t)) + f(\vec{x}(t), t) + \mathcal{B}u(t)$ as $f(x, \dot{x}, t)$, because all terms are functions of time t, state variable $x(t)$ and its derivatives $\dot{x}(t)$. As such, equation (10) can be rewritten as:

$$f(x, \dot{x}, t) = 0. \tag{15}$$

Assuming that the impact of the local mismatch is small, (11) can be solved by treating the rhs-term for mismatch as a perturbation to the nominal trajectory $x^{(0)}(t)$ of the circuit. Here, $x^{(0)}(t)$ are the nominal values or solution of the nonlinear circuit equation

$$f(x^{(0)}, \dot{x}^{(0)}, t) = \mathcal{F}\mathbf{i}_n(x^{(0)}, \xi). \tag{16}$$

Where $\mathbf{i}_n(x^{(0)}, \xi)$ is the vector of mismatch current sources that can model the process variations. And $\mathcal{F}$ describes how to connect sources into the circuit which is determined by circuit topology.

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

$$\begin{aligned} & f(x^{(0)}, \dot{x}^{(0)}, t) + \left.\frac{\partial f(x, \dot{x}, t)}{\partial x}\right|_{x=x^{(0)}, \dot{x}=\dot{x}^{(0)}} (x - x^{(0)}) + \left.\frac{\partial f(x, \dot{x}, t)}{\partial \dot{x}}\right|_{x=x^{(0)}, \dot{x}=\dot{x}^{(0)}} (\dot{x} - \dot{x}^{0)}) \\ & = \mathcal{F}\mathbf{i}_n(x^{(0)}, \xi). \end{aligned} \tag{17}$$

Or

$$G(x^{(0)}, \dot{x}^{(0)})x_m + C(x^{(0)}, \dot{x}^{(0)})\dot{x}_m = \mathcal{F}\mathbf{i}_n(x^{(0)}, \xi), \tag{18}$$

where

$$\begin{aligned} G(x^{(0)},\dot{x}^{(0)}) &= \frac{\partial q(x,\dot{x},t)}{\partial x}\ |_{x=x^{(0)},\dot{x}=\dot{x}^{(0)}} = \left.\frac{\partial q(x(t))}{\partial x}\right|_{x=x^{(0)},\dot{x}=\dot{x}^{(0)}} \\ C(x^{(0)},\dot{x}^{(0)}) &= \frac{\partial f(x,\dot{x},t)}{\partial \dot{x}}\ |_{x=x^{(0)},\dot{x}=\dot{x}^{(0)}} = \left.\frac{\partial f(x(t))}{\partial x}\right|_{x=x^{(0)},\dot{x}=\dot{x}^{(0)}} \end{aligned} \tag{19}$$

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.

With a perturbation analysis, the parameter variations can be considered as a perturbation to the nominal transient trajectory ($x = x_{(0)},; \dot{x} = \dot{x}_{(0)}$) of the SDAE in (11). This leads to a linearized SDAE

$$\begin{aligned} &G_{(0)} \cdot \hat{x} + C_{(0)}\dot{\hat{x}} = T \cdot \zeta(t) \\ &G_{(0)} = \frac{\partial f(x,\dot{x},t)}{\partial x}, \quad C_{(0)} = \frac{\partial f(x,\dot{x},t)}{\partial \dot{x}} \end{aligned} \tag{20}$$

where $\hat{x}(t) = x - x_{(0)}$ is the state variable for the stochastic mismatch. When the perturbation is large, the high-order expansion can be performed and the derivation below still holds. However, we focus on small variations, and assume $G_{(0)}$ and $C_{(0)}$ are both constant at each time-step which only depend on nominal values of parameters.

4.2. NMC by SOP Expansion

Next, we can introduce the SOP to transient mismatch analysis so that the mean and variance of transient mismatch can be computed with only one-time simulation. Note that different distributions types are associated with different orthogonal polynomials. In this paper, we assume that the random variables for the local mismatch follow Gaussian distributions and thus Hermite polynomial functions can be used for their SOP expansions [Xiu and Karniadakis 2002; Vrudhula et al. 2006]. For example, Hermite polynomial functions with one variable ξ can be shown as:

$$\Phi(\xi) = [\Phi_1(\xi), \Phi_2(\xi), \Phi_3(\xi), ...,]^T = [1, \xi, \xi^2 - 1, ...,]^T \tag{21}$$

is used to construct the expansion basis to calculate the mean and the variance of $x_m(t)$. Note that variable ξ follows the standard Gaussian distribution $N(0,1)$.

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

$$x_m(t) = \sum_i \alpha_i(t)\Phi_i(\xi). \tag{22}$$

Note that for different random processes, many other orthogonal polynomials can be selected as well, based on a so-called *Askey scheme* [Xiu and Karniadakis 2002].

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

$$< \Delta(\xi), \Phi_j(\xi) >= \int_\xi \Delta(\xi)\Phi_j(\xi)W(\xi)d\xi = 0. \tag{23}$$

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 (23) leads to

$$\begin{aligned}&\alpha_0 = 0, \quad \alpha_2 = 0\\ &G(x^{(0)}, \dot{x}^{(0)})\alpha_1(t) + C(x^{(0)}, \dot{x}^{(0)})\dot{\alpha_1}(t) = \mathcal{F}n(x^{(0)})g^{\beta}(p),\end{aligned} \tag{24}$$

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 (24) can be solved by a Backward-Euler integration

$$(G_k + \frac{1}{h}C_k)\alpha_1(t_k) = \frac{1}{h}C_k\alpha_1(t_k - h) + \mathcal{F}\mathbf{i}_k, \tag{25}$$

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 $x^{(0)}$.

Expanding $\zeta(t)$ ($\sum_i g_i\Phi_i$) and $\hat{x}(t)$ ($\sum_i \alpha_i(t)\Phi_i$) by Hermite polynomials, one can have

$$G_{(0)} \cdot \sum_i \alpha_i(t)\Phi_i + C_{(0)} \cdot \sum_i \dot{\alpha}_i(t)\Phi_i = T \cdot \sum_i g_i\Phi_i, \tag{26}$$

Where g_i and α_i are the coefficients of SoP expansion of stochastic current sources and state variables, respectively.

By further taking the inner-product with Φ_j for the collocation at the two sides, one can have

$$G_{(0)} \cdot \alpha_i(t) + C_{(0)} \cdot \dot{\alpha}_i(t) = T \cdot g_i(t). \tag{27}$$

The above equation can be solved with a Backward-Euler method

$$\left(G_{(0)}^k + \frac{1}{h}C_{(0)}^k\right)\alpha_i(t_k) = \frac{1}{h}C_{(0)}^k\alpha_i(t_k - h) + T \cdot g_i(t_k). \tag{28}$$

As a result, one can obtain the mean $\mu_{\hat{x}}(t) = \alpha_0(t)$ and the variance $(\sigma_{\hat{x}}(t))^2 = \alpha_1(t)^2$ for the stochastic transient variation at the time-step t_k.

4.3. One CMOS Transistor Example

For illustration purpose, one CMOS transistor is presented as an example to show the NMC mismatch analysis. The variable channel length L_{eff} is considered as an independent variation source and the variation of L_{eff} can be mapped into a noise current source as:

$$\zeta = I_D^{\beta} t_m(W, L)\gamma(x, y) = I_D^{\beta}\left(1 + \frac{\kappa^{\beta}}{\sqrt{W_{(0)}}}\left[\frac{1}{\sqrt{L_{(0)}}} - \frac{1}{2\sqrt{\left(L_{(0)}\right)^3}}\xi\right]\right)\gamma(x, y). \tag{29}$$

where the $t_m(W, L)$ can be expanded with first-order Taylor expansion as shown in equation (7) and ξ is a Gaussian random variable for the variation of channel length L_{eff}. Also, $\gamma(x, y) = 1$ because ξ is assumed to be an independent variable.

As such, the DAE system with noise current sources ζ becomes:

$$G(x^{(0)}, \dot{x}^{(0)})x_m + C(x^{(0)}, \dot{x}^{(0)})\dot{x}_m = \zeta(\xi). \tag{30}$$

Moreover, the variables in above equation can be expanded with Hermite polynomials Φ_i $(i = 0, ..., n)$ as:

$$\begin{aligned} &G(x^{(0)}, \dot{x}^{(0)}) \textstyle\sum_i \alpha_i(t)\Phi_i(\xi) + C(x^{(0)}, \dot{x}^{(0)}) \textstyle\sum_i \dot{\alpha_i}(t)\Phi_i(\xi) = \\ &I_D^\beta \left(1 + \frac{\kappa^\beta}{\sqrt{W_{(0)}}}\left[\frac{1}{\sqrt{L_{(0)}}} - \frac{1}{2\sqrt{\left(L_{(0)}\right)^3}}\sum_{i=0}^{n} g_i'\Phi_i\right]\right). \end{aligned} \tag{31}$$

Without the loss of generality, the first-order SoPs expansion $(n = 1)$ can be considered. When applying inner-product with the orthogonal basis Φ_i on both sides, we can obtain:

$$\begin{aligned} &\alpha_0 = 0, \quad \alpha_2 = 0 \\ &G(x^{(0)}, \dot{x}^{(0)})\alpha_1(t) + C(x^{(0)}, \dot{x}^{(0)})\dot{\alpha_1}(t) = -I_D^\beta(t)\frac{\kappa^\beta}{\sqrt{W_{(0)}}}\frac{1}{2\sqrt{\left(L_{(0)}\right)^3}}g_1'. \end{aligned}$$

Note that g_1 is known because the variation of channel length is given. Thereby, the first-order SOP coefficient α_1 can be solved with a Backward-Euler integration as:

$$(G_k + \frac{1}{h}C_k)\alpha_1(t_k) = \frac{1}{h}C_k\alpha_1(t_k - h) - (I_D^\beta)_k\frac{\kappa^\beta}{\sqrt{W_{(0)}}}\frac{1}{2\sqrt{\left(L_{(0)}\right)^3}}g_1'. \tag{32}$$

at the k-th time-step. Recall that G_k, C_k and $(I_D^\beta)_k$ are the nominal conductance (g_{ds}), capacitance (c_{ds}) and channel current I_d evaluated at t_k. 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.

More importantly, for large-scale problems with a large number of transistors (e.g. Operational Amplifier and Schmidt Trigger in the experiment), we can simultaneously solve the transient mismatch vector by adding all noise current sources into the DAE system with equation (11).

5. STOCHASTIC YIELD ESTIMATION AND OPTIMIZATION

In this section, we will discuss how to estimate the parametric yield and further optimize it by tuning parameters automatically. We will first show how to estimate the parametric yield with the stochastic variation (mismatch) $(\mu_{f_m;t}, \sigma_{f_m;t})$ obtained from the above NMC mismatch analysis.

5.1. Fast Yield Calculation

First, we construct the performance distribution at one time-step t_k by $(\mu_{f_m}(t_k)$, $\sigma_{f_m}(t_k))$, shown as the solid curve from $\mu - 3\sigma$ to $\mu + 3\sigma$ in Figure(3). Then, the performance constraint is given as

$$h(\vec{p}; t_k) = f_m(\vec{p}; t_k) - f_{m_{threshold}} \leq 0. \tag{33}$$

With the constraints, the boundary separating success region from failure region can be plotted as the straight line $h(\vec{p}; t_k) = 0$ in following figure.

As a result, the performance $f_m(t_k)$ located at the left of $h(\vec{p}; t_k) = 0$ (shown as the shaded region) can satisfy the constraint in (33) and thus belong to the successful region $\hat{S}$. Hence, the parametric yield can be estimated with the area ratio by

$$Y(\vec{p}) = \frac{\hat{S}}{S_{f_m}}. \tag{34}$$

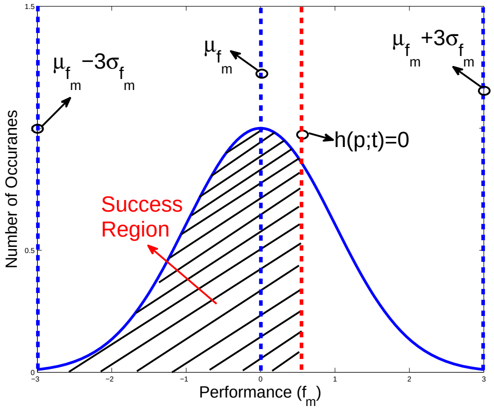


Fig. 3. Parametric yield estimation based on SOP based stochastic variation analysis

When denoting the entire region area $S_{f_m} = 1$, $Y(\vec{p})$ becomes $\hat{S}$ and is determined by the integration below

$$Y(\vec{p}) = \int\limits_{\hat{S}} pdf(f_m(\vec{p}; t_k))dS = \int\limits_{\hat{S}} pdf(\mu_{f_m}, \sigma_{f_m})dS \tag{35}$$

where $pdf(f_m)$ is the probability-distribution-function (PDF) of the performance merit of interest, characterized by μ_{f_m} and σ_{f_m} at the time-step t_k.

5.2. Stochastic Sensitivity Analysis

In order to improve the yield rate, most optimization engines need sensitivity information to identify and further tune those critical parameters. However, with the emerging process variations beyond 90nm, traditional sensitivity analysis becomes inefficient: either use the worst-case scenario or conduct Monte Carlo simulations [Lampaert et al. 1995; Schenkel et al. 2001; Liu et al. 2010]. Therefore, an efficient NMC-based *stochastic sensitivity analysis* is needed for this purpose.

With all parameter variations calculated from the fast mismatch analysis in Section 4, one can further explore the impact or contribution from the parameter variation $\sigma_{\xi_{p_i}}$ to the performance variation $\sigma_{\xi_{f_m}}$. This can be utilized to perform optimization procedure for better performance merits. In this section, we develop an approach to evaluate the sensitivity of transient variation (mismatch) with respect to each parameter variation.

We start from the definition of *stochastic sensitivity*. Expressing the relationship between the performance metric variation ξ_{f_m}. From now on, we note $\xi_{f_m}(t) = f_m(\vec{\xi_p}; t))$ for illustration purpose and assume the random parameter vector $\vec{\xi_p}$ $(\in \mathbb{R}^m)$. As such, the *stochastic sensitivity* can be defined by

$$s_{p_i}(t) = \frac{\partial f_m(\vec{\xi_p}; t)}{\partial \xi_{p_i}}, \quad i = 1, \cdots, m \tag{36}$$

where $s_{p_i}(t)$ is the derivative of the performance variation ξ_{f_m} with respect to the i-th random parameter variable ξ_{p_i} at one time-instant t. Depending on the problem or circuit under study, the performance f_m can be output voltage, period and power, etc.; and the parameter can be transistor width, length and oxide thickness, etc. Such a so-called stochastic sensitivity can be also understood based on the propagation of variance (POV) relationship [McAndrew et al. 1997; Drennan and McAndrew 2003] below

$$\sigma^2_{\xi_{f_m}} = \sum_i \left(\frac{\partial f_m(\vec{\xi_p};t)}{\partial \xi_{p_i}} \right)^2 \sigma^2_{\xi_{p_i}}. \tag{37}$$

Here, $\sigma^2_{\xi_{p_i}}$ is the parameter variance and $\sigma^2_{\xi_{f_m}}$ is the performance variance.

Note that the performance variation ξ_{f_m} is mainly determined by α_1[Xiu and Karniadakis 2002] in (28) at time-step t_k as derived in Section 4.3, while α_2 has little impact on the performance variation. As such, one can truncate the SoP expansions to the first order for the calculation of mean and variance, and experiments show that the first order expansion can provide adequate accuracy. Therefore, α_1 is the dominant moment for ξ_{f_m} while α_2 can be truncated to simplify calculation. Therefore, we have the following:

$$\alpha_1(t_k) = c_1 + c_0 T \cdot g(t_k), \tag{38}$$

where

$$c_0 = \left(G^k_{(0)} + \frac{1}{h} C^k_{(0)} \right)^{-1},$$
$$c_1 = c_0 \cdot \left(\frac{1}{h} C^k_{(0)} \alpha_1(t_k - h) \right).$$

As such, one can further calculate the stochastic sensitivity $\partial f_m(\vec{\xi_p};t) \Big/ \partial \xi_{p_i}$ using

$$s_{p_i}(t_k) \ = \ \frac{\partial f_m(\vec{\xi_p};t)}{\partial \xi_{p_i}} = (c_0 T_{p_i}) \times \frac{\partial g(t_k)}{\partial p_i} \tag{39}$$

which can be utilized in any gradient-based optimization to improve the yield-rate.

5.3. Multi-Objective Optimization

Next, we make use of sensitivities s_{p_i} to improve parametric yield. Meanwhile, since power is also a primary design concern, we treat power consumption reduction as an extra objective and solve a multi-objective optimization problem defined in Section 3. Note that other performance merits can be treated as objectives of optimization in a similar way. As such, by tuning nominal process parameters along gradient directions, we enable more parameters containing process variations to satisfy the performance constraints. This is a importance feature for a robust design. In this section, we demonstrate this requirement by a sequential linear programming (SLP).

At the beginning of each optimization iteration, the nonlinear objective functions $Y(\vec{p})$ and $P_c(\vec{p})$ can be approximated by linearization

$$\begin{aligned} Y(\vec{p}) &= Y(\vec{p}_{(0)}) + \nabla_p Y(\vec{\xi}_{p_{(0)}})^T (\vec{p} - \vec{p}_{(0)}) \\ P_c(\vec{p}) &= P_c(\vec{p}_{(0)}) + \nabla_p P_c(\vec{\xi}_{p_{(0)}})^T (\vec{p} - \vec{p}_{(0)}), \end{aligned} \tag{40}$$

where $p_{(0)}$ represents the nominal design parameters while $\vec{p}$ contains the process variations of these parameters. Note that (31) is a first-order Taylor expansion of parametric yield $Y(\vec{p})$ defined in (35) and power consumption $P_c(\vec{p})$, around the nominal parameter region $\vec{p}_{(0)}$. Thus, $\nabla_p Y(\vec{\xi}_{p_{(0)}})$ is a vector consisting of $\partial Y(\vec{\xi_p})/\partial \xi_{p_i}$. The same is true for power consumption $\nabla_p P_c(\vec{\xi}_{p_{(0)}})$. Therefore, the nonlinear objective functions can be transformed into a series of linear optimization sub-problems. The optimization terminates when the convergence criterion is achieved.

As such, the stochastic multi-objective yield optimization problem in Section 3 can be re-formulated as

$$\begin{array}{ll}\text{maximize} & Y(\vec{p}) = Y(\vec{p}_{(0)}) + \nabla_p Y(\vec{\xi}_{p_{(0)}})^T(\vec{p} - \vec{p}_{(0)}) \\ \text{minimize} & P_c(\vec{p}) = P_c(\vec{p}_{(0)}) + \nabla_p P_c(\vec{\xi}_{p_{(0)}})^T(\vec{p} - \vec{p}_{(0)}) \\ & \\ \text{subject to} & \begin{array}{l} Y(\vec{p}) \geq \bar{Y} \\ P_c(\vec{p}) \leq \bar{P}_c \\ \mathcal{F}(\vec{p}) \leq \mathcal{F}_{max} \\ \vec{p}_{\min} \leq \vec{p} \leq \vec{p}_{\max} \end{array}\end{array}$$

where $\delta\vec{p} = \vec{p} - \vec{p}_0$ is the step size. Within each iteration, the sensitivity vector $\nabla_p Y(\vec{\xi}_{p_{(0)}})$, $\nabla_p P_c(\vec{\xi}_{p_{(0)}})$ and $\delta\vec{p}$ should be updated.

However, the stochastic sensitivity analysis in Section 5 can only calculate $\partial F(\vec{\xi}_p;t)\Big/\partial\xi_{p_i}$ rather than $\partial Y(\vec{\xi}_p)\Big/\partial\xi_{p_i}$. To obtain $\partial Y(\vec{\xi}_p)\Big/\partial\xi_{p_i}$, we start from (35) with the following derivation

$$\begin{aligned}\frac{\partial Y(\vec{\xi}_p)}{\partial\xi_{p_i}} &= \int\limits_{\hat{S}} \frac{\partial pdf(F(\vec{\xi}_p;t))}{\partial\xi_{p_i}} dS \\ &= \int\limits_{\hat{S}} \frac{\partial pdf(F)}{\partial F} \cdot \frac{\partial F(\vec{\xi}_p;t)}{\partial\xi_{p_i}} dS.\end{aligned} \tag{41}$$

As a result, $\partial Y(\vec{\xi}_p)/\partial\xi_{p_i}$ can be obtained with $\partial F(\vec{\xi}_p;t)\Big/\partial\xi_{p_i}$ calculated from the stochastic sensitivity analysis. Note that the PDF of the performance variation and the integral region $\hat{S}$ are both given from the yield estimation in (35).

We illustrate our optimization procedure for yield objective function $Y(\vec{p})$ through Fig.(4). With the parametric yield estimation using the NMC mismatch analysis, the distribution of performance f_m for nominal parameters $\vec{p}_0$ can be plotted as a solid curve, which has a mean-value $\mu_{f_m}(p_0)$. With the performance constraint $h(\vec{p};t) \leq 0$ in (33), the shaded area located at the left of the constraint-line is the desired successful region.

One yield optimization procedure needs to move the performance distribution to left side so that the shaded area can be maximized. Therefore, the problem here is how to change the process parameters $\vec{p}$ in order to move the performance distribution for an enhanced yield rate.

Moreover, power consumption can be estimated by

$$P_c(\vec{p}) = -[Vdd \cdot \bar{i}_{Vdd}], \tag{42}$$

where Vdd is the power supply voltage source and $\bar{i}_{Vdd}$ is the average value of current through the voltage source. The power consumption optimization can be explained as shown in Fig.(5). The initial design generates the current i_{Vdd} denoted as the black curve and leads to high power consumption P_c.

According to (42), P_c can be reduced by lowering the average value of i_{Vdd}. To do so, we move the minimum point on the current trajectory close to zero and obtain the optimal design with minimum P_c as the red curve shown in Fig.(5). As such, the power

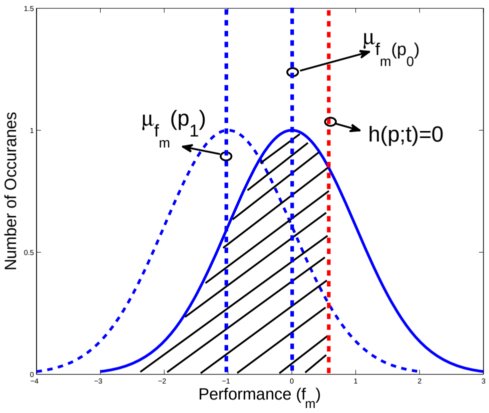


Fig. 4. Stochastic Yield Optimization

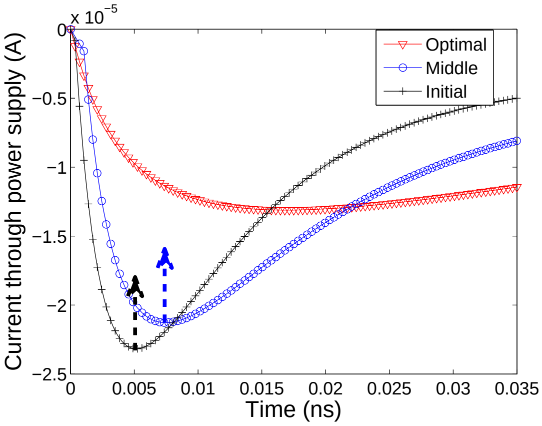


Fig. 5. Power Consumption Optimization

optimization requires us to change $\vec{p}$ in order to move the minimum point of i_{Vdd} close to zero for smaller power consumption. To solve this problem, the parametric yield-rate $Y(\vec{p}_0)$ is first calculated from (35) and the performance distribution is constructed accordingly, similar to the one in Fig.(4). Then, the targeted yield-rate $\bar{Y}$ is used to compare with $Y(\vec{p}_0)$ by

$$\Delta Y(\vec{p}_0) = \bar{Y} - Y(\vec{p}_0). \tag{43}$$

Next, the NMC stochastic sensitivity analysis is performed to find $\partial F(\vec{\xi_p};t) \Big/ \partial \xi_{p_i}$ and thus $\partial Y(\vec{\xi_p})/\partial \xi_{p_i}$ in the (42). As a result, with the first-order Taylor expansion in SLP (41), one can determine the parameter incremental $\delta \vec{p}_{yield} = \vec{p} - \vec{p}_{(0)}$ in order to reach $Y(\vec{p}) = \bar{Y}$ by

$$\delta \vec{p}_{yield} = \frac{\bar{Y} - Y(\vec{p}_{(0)})}{\nabla_p Y(\vec{p}_{(0)})} = \frac{\Delta Y(\vec{p}_{(0)})}{\nabla_p Y(\vec{p}_{(0)})}. \tag{44}$$

On the other hand, we perform the same procedure to optimize the power consumption. Similarly to (39), we calculate the sensitivity of power consumption w.r.t. process parameters at i_{Vdd} with a minimum current value:

$$\frac{\partial P_c(\vec{p})}{\partial p_i} = -[V_{dd} \cdot \left.\frac{\partial i_{Vdd}}{\partial p_i}\right|_{i_{Vdd}=Minimum}]. \tag{45}$$

The according parameter increments can be computed as

$$\delta\vec{p}_{power} = \frac{\bar{P}_c - P_c(\vec{p}_{(0)})}{\nabla_p P_c(\vec{p}_{(0)})} = \frac{\Delta P_c(\vec{p}_{(0)})}{\nabla_p P_c(\vec{p}_{(0)})}. \tag{46}$$

In this way, the total changes to the process parameters are the weighted summation below

$$\delta\vec{p}_{total} = \lambda_1 \cdot \delta\vec{p}_{yield} + \lambda_2 \cdot \delta\vec{p}_{power}, (\lambda_1, \lambda_2 \in [0, 1]), \tag{47}$$

where λ_1 and λ_2 are weights for yield and power consumption. Also, λ_1 and λ_2 can be updated dynamically and weight λ should be larger for the performance merit that is farther from the target value.

Therefore, one can update $\vec{p}$ with the new parameter $\vec{p}_0 + \delta\vec{p}_{total}$. Moreover, the N-MC mismatch analysis is conducted to update the performance distribution, which is denoted by a dashed-curve shown in Fig.(4). With the updated new parameters and performance distribution, all performance constraints $\mathcal{F}(\vec{p}) \leq \mathcal{F}_{max}$ are checked for violations. If they are still valid, $\vec{p}$ becomes the new design point and this procedure is repeated again to enhance the yield-rate.

6. EXPERIMENTAL RESULTS

6.1. Algorithm Overview

For illustration purpose, we summarize the proposed algorithm in Algorithm(1). The optimization procedure involves several optimization iterations; each of them contains three major steps: stochastic yield estimation, stochastic sensitivity analysis, and stochastic yield optimization. The optimal design point can be achieved by tuning nominal parameters along their gradient directions.

ALGORITHM 1: Proposed Yield Optimization Algorithm

```
Input: Circuit topology, distribution of variable parameters $p_i$ ($\mu_{p_i}, \sigma_{p_i}$), and performance
       constraints $h(\vec{p}; t)$.
Output: The optimal design point ($\mu_{p_i}^{optimal}$).
/* Yield Optimization Loop */;
repeat
    /* Step 1: Yield Estimation */
    Use stochastic mismatch analysis to calculate the transient mismatch ($\mu_{f_m;t}, \sigma_{f_m;t}$);
    Build the distribution of $f_m$;
    Calculate the parametric yield $Y(\vec{p})$ with $h(\vec{p})$ using equation(14);

    /* Step 2: Sensitivity Analysis */
    Calculate the sensitivities of $Y(\vec{p})$ as $s_i(t) = \partial Y(\vec{\xi}_p)/\partial \xi_{p_i}$ and that of $P_c$ in (45);

    /* Step 3: Yield Enhancement */
    Move the nominal parameters $\mu_{p_i}$ along tangent direction to improve $Y(\vec{p})$ and reduce $P_c(\vec{p})$;

until Yield $Y(\vec{p})$ cannot be maximized and Power Consumption $P_c(\vec{p})$ is satisfied.;
return the optimal design point ($\mu_{p_i}^{optimal}$)
```

6.2. Settings

We have implemented the proposed non-Monte-Carlo (NMC) algorithms for NMC mismatch analysis, yield estimation, and optimization in a Matlab based circuit simulator. All experiments are performed on a Linux server with a 2.4GHz Xeon processor and 4GB memory. In our experiment, we take the widths of MOSFETs as process variable parameters for illustration.

We first use an Operational Amplifier (OPAM) to study the accuracy and efficiency of our NMC mismatch analysis by comparing it against Monte Carlo simulations. Then, a Schmitt Trigger is used to verify our proposed parametric yield estimation and stochastic yield analysis. Finally, we demonstrate the efficiency of our yield optimization method using a 6-transistor SRAM cell.

6.3. NMC Mismatch Analysis

The operational amplifier (OPAM) is shown in Fig.(6), which consists of eight MOS transistors. We introduce the channel width variation (Gaussian distributions with 10% perturbation from their nominal values) to all MOSFETs. Notice that we consider the matching design requirements for the input pair devices, such as the same nominal width ($Wp_1 = Wp_2$, $Wn_3 = Wn_4$, $Wp_5 = Wp_7 = Wp_8$) and the fixed width ratio ($Wn_6 = kWn_3$).

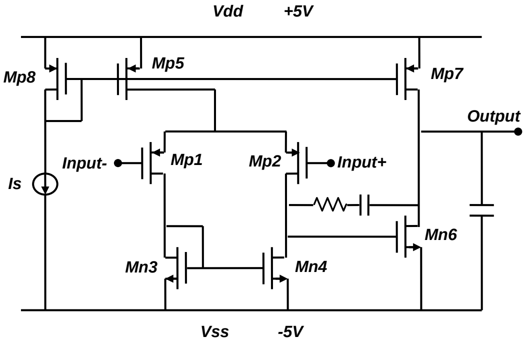


Fig. 6. Schematic of Operational Amplifier

We first perform 1000 times MC simulations with a high confidence level to find the variational trajectories at the output node. Then, we apply the developed NMC mismatch analysis to OPAM and locate the 3σ boundaries (i.e. $\mu - 3\sigma$, $\mu + 3\sigma$) of perturbed trajectories with a one-time run of transient circuit simulation. The results are shown in Fig.(7), where blue lines denotes the MC simulations and the two black lines are results from our mismatch analysis. We observe that our approach can capture the transient stochastic variation (mismatch) as accurately as Monte Carlo results.

We further compare the accuracy and efficiency between NMC mismatch analysis and the Monte Carlo method in the Table(I). From this table, we can see that NM-C mismatch analysis can not only achieve 2% accuracy of MC, but also gains $680X$ speedup over Monte-Carlo method.

6.4. Stochastic Yield Estimation

We further use the Schmitt Trigger example shown in Fig.(8) to investigate the stochastic yield estimation. Similarly, we assume the widths of all MOSFETs have 10% variations from their nominal values and follow Gaussian distributions. Moreover, we study the lower switching threshold V_{TL} as the performance metric for the parametric

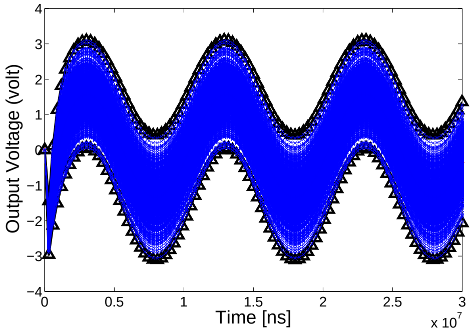


Fig. 7. NMC mismatch analysis vs. Monte Carlo for Operational Amplifier Case

Table I. Comparison of Accuracy and Runtime

Operational Amplifier Example		
Runtime (seconds)	Proposed	1.33
	Monte-Carlo	905.06
Mean value (μ) Unit: volt	Proposed	0.35493
	Monte-Carlo	0.34724
Std. value (σ) Unit: volt	Proposed	0.57032
	Monte-Carlo	0.56272

yield estimation, which can be perturbed due to MOSFET width variations. In other words, the performance constraint can be defined as follows: when the input V_{TL} is $1.8V$ and the output is initially set to $Vdd = 5V$, the output V_{OUH} should be greater than 4.2V.

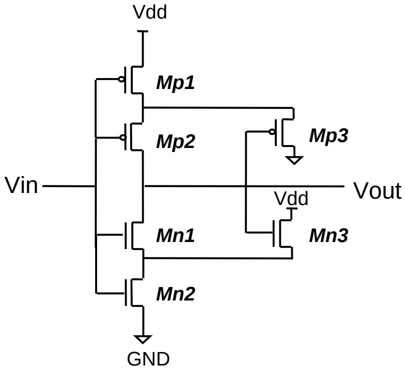
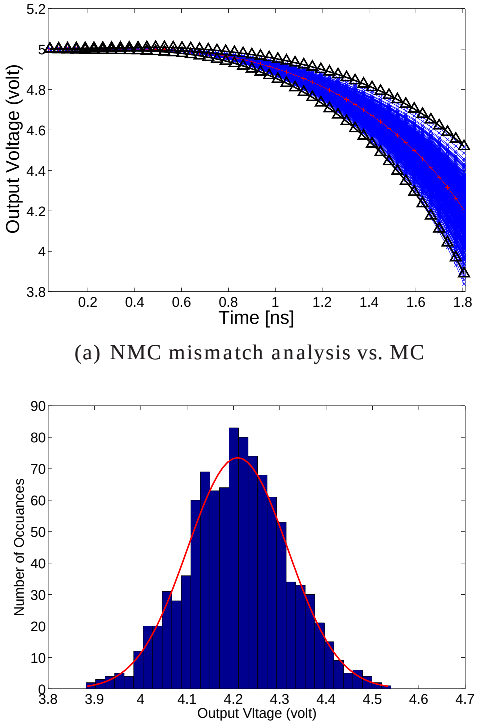


Fig. 8. Schematic of Schmitt Trigger

We first conduct 1000 times MC simulations and compare it with the result from NMC stochastic variation analysis shown in Fig.(9(a)). Then, the output distribution from the MC simulation at the time-step when input voltage equals to $1.8V$ is plotted in Fig.(9(b)). In addition, the PDF estimation by our NMC mismatch analysis (defined by mean μ and standard deviation σ) is compared with MC simulations in the same figure. The two distributions coincide with each other very well so that this experiment can validate the accuracy of proposed stochastic yield estimation.



(a) NMC mismatch analysis vs. MC

(b) Output distributions from NMC mismatch analysis and Monte Carlo

Fig. 9. Comparison of Schmitt Trigger example

Then, the yield rate can be calculated with one estimated PDF from our NMC mismatch analysis efficiently. We list the mean (μ), standard deviation (σ), and yield estimation results from our approach and those by MC simulations in Table(II). With the accurate estimation of output distribution, our method can calculate the yield rate with 2.7% accuracy of MC as well as $756X$ speedup when compared to the MC method.

Table II. Comparison of Accuracy and Runtime

Schmitt Trigger Example		
Runtime (seconds)	Proposed	1.06
	Monte-Carlo	801.84
Mean value (μ) Unit: volt	Proposed	4.2043
	Monte-Carlo	4.1993
Std. value (σ) Unit: volt	Proposed	0.10487
	Monte-Carlo	0.094346
Yield rate	Proposed	0.48357
	Monte-Carlo	0.47059

6.5. Stochastic Sensitivity Analysis

Furthermore, we apply the proposed stochastic sensitivity analysis to the Schmitt Trigger example, which can find the contribution of each variation source to the output variation. Note that we are interested in the lower switching threshold V_{TL}, where input increases from zero and output decreases from Vdd. As a illustration, the sensitivity of output voltage variation ξ_{output} with respect to all MOSFET widths variations ξ_{p_i} at the time-step when input voltage equals to $1.8V$ are shown in Table(III). We can observe that widths of $Mp1$, $Mp2$, and $Mn3$ transistors are more critical than other MOSFETs.

Table III. Sensitivity of ξ_{output} with respect to each MOSFET width variation ξ_{p_i}

Parameter	Mn1 width	Mn2 width	Mn3 width
Sensitivity	2.4083e-4	2.4083e-4	4.8069e-3

Parameter	Mp1 width	Mp2 width	Mp3 width
Sensitivity	2.4692e-2	2.4692e-2	0

6.6. Stochastic Yield Optimization

To further validate the yield optimization method, we deploy a 6-T SRAM cell in Fig.(10) and apply the proposed optimization method to improve the yield rate under reading accessing failure. During the reading operation, both BL_B and BL are pre-charged to Vdd, while Q_B stores zero and Q stores one. When reading the SRAM cell, BL_B starts to discharge from Vdd and produces a voltage difference ΔV between itself and BL. The time it takes BL_B to produce a large enough voltage difference $\Delta V_{threshold}$ is called *access time*. If the access time is larger than the threshold value at the time-step $t_{threshold}$, this leads to an *access time failure*. In our experiment, we assume that $t_{threshold} = 0.04ns$ and $\Delta V_{threshold} = 0.8662V$.

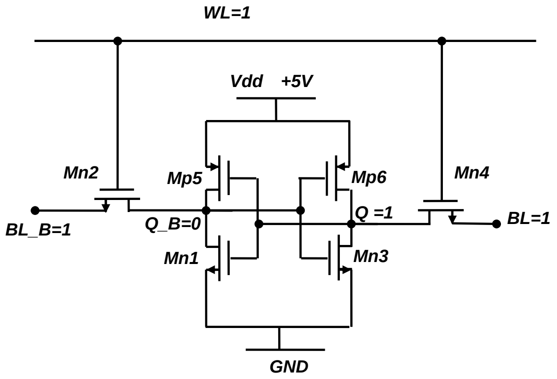


Fig. 10. Schematic of SRAM 6-T Cell

In this example, all channel widths of MOSFETs are considered as the variable parameters which follow Gaussian distributions with 12% perturbation from nominal values. As such, when the access time differs from the nominal value due to variations in channel width, access time failure occurs and thus yield loss may happen. In order to relieve the issue, we first perform NMC mismatch analysis to find the voltage distribution of BL_B at $t_{threshold}$, which is shown in Fig.(11). Also, 1000 times MC simulations have been performed as a baseline for comparison which can provide the variational transient waveforms of BL_B in Fig.(12). The comparison can validate the accuracy of our NMC mismatch analysis.

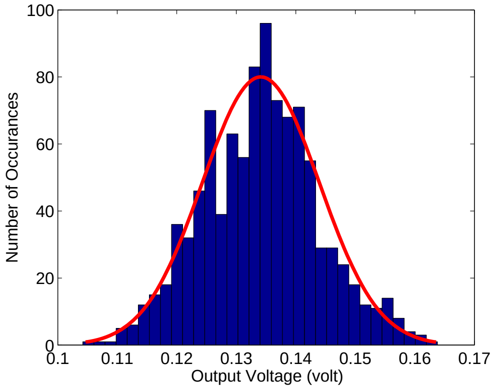


Fig. 11. Voltage distribution at BL_B node

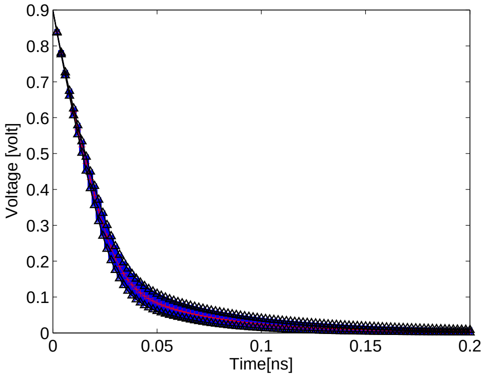


Fig. 12. NMC mismatch analysis vs. MC

Then, the sensitivity analysis developed in this paper is used to find the $\partial\xi_{v_{BL\_B}}/\partial\xi_{p_i}$ and $\partial\xi_{power}/\partial\xi_{p_i}$ where ξ_{p_i} is the width variation of i-th MOS transistor and ξ_{power} is the variation of power supply voltage source. Results are shown in Table(IV). From this table, we can see that only $Mn1$, $Mn2$, and $Mp6$ can have influence on the access time and power variations in our experiment setup; also, we can see that their nominal values can be tuned to reduce access time failure for better parametric yield rate and to lower the power consumption simultaneously due to different gradient directions.

Table IV. Sensitivity of $\xi_{v_{BL\_B}}$ and ξ_{power} with respect to each MOSFET width variation ξ_{p_i}

Parameter	Mn1 width	Mn2 width	Mp6 width
Sensitivity ($\xi_{v_{BL\_B}}$)	1.3922e-3	2.0787e-3	7.0941e-2
Sensitivity (ξ_{power})	3.7888e-4	5.7816e-4	-5.8871e-4

Finally, we apply the multi-objective yield optimization to improve yield. For comparison purpose, two other algorithms have been implemented: (1) baseline, the generic gravity-directed method in [Soin and Spence 1980] which moves the nominal parameters to the gravity of successful region, and (2) the single-objective optimization which only improves the yield. The results from all optimization methods are shown in Table (V). From this table, it can be observed that all methods can improve the parametric yield to be around or even more than 95% compared with the initial design. According nominal values can be used as better initial design parameters. Meanwhile, the area is smaller than the maximum acceptable area criterion $A \leq 1.2A_{initial}$.

However, optimal designs from baseline (gravity-directed) method and single-objective optimization require $2.75X$ and 21% more power consumption when compared with initial design, respectively. Proposed method can lead to optimal design with only 7% more power requirement. Therefore, it can be demonstrated that proposed multi-objective optimization can not only improve the yield rate but also suppress the power penalty simultaneously. Moreover, the proposed optimization procedure only needs 6 iteration runs to achieve the shown results within 15.21 seconds. Notice that the parametric yield $Y(\vec{p})$ can be further improved with a higher target yield $\bar{Y}$ at the cost of more optimization iterations.

Table V. Comparison of Different Yield Optimization Algorithms for SRAM Cell

Parameter	First-cut	Baseline	Single-objective	Multi-objective
Mn1 width (m)	1e-5	2.872e-5	2.7841e-5	3.577e-5
Mn2 width (m)	1e-5	2.3282e-5	2.2537e-5	2.7341e-5
Mp6 width (m)	3e-5	1.5308e-5	1.6296e-5	9.7585e-6
Power (W)	1.0262e-005	3.0852e-5	1.2434e-5	1.0988e-5
Area (m^2)	2.4e-11	2.81e-11	2.8e-11	2.88e-11
Yield	49.32%	94.23%	95.49%	95.31%
Runtime (seconds)	2.42	32.384	27.226	15.21
Iterations	1	12	10	6

7. CONCLUSION

In this paper, we have developed one fast non-Monte-Carlo (NMC) yield estimation and optimization approach. It first models the PVT variation sources as stochastic current sources, and expands them by stochastic orthogonal polynomials (SOPs). Then, the probabilistic distribution of transient mismatch can be calculated from one-time simulation and the yield rate can be computed under given performance constraints. Moreover, we further derive the stochastic sensitivity of yield within the context of SoPs, and develop a gradient-based multi-objective optimization which can efficiently improve the yield rate and satisfy other performance constraints at the same time. The extensive experiments on a number of circuits demonstrate that proposed method can achieve up to 98% accuracy and $700X$ speedup when compared with Monte Carlo simulations. Also, the optimization procedure can not only improve the yield rate up to 95.3% and satisfy other performance constraints, but also provide the best efficiency when compared with other existing methods.

REFERENCES

BIAGETTI, G., ORCIONI, S., TURCHETTI, C., CRIPPA, P., AND ALESSANDRINI, M. 2004. SiSMA-a tool for efficient analysis of analog CMOS integrated circuits affected by device mismatch. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems 23,* 2, 192 – 207.

COX, P., YANG, P., MAHANT-SHETTI, S., AND CHATTERJEE, P. 1985. Statistical modeling for efficient parametric yield estimation of MOS VLSI circuits. *IEEE Journal of Solid-State Circuits 20,* 1, 391 – 398.

DEB, K. 2002. *Multi-objective optimization using evolutionary algorithms*, Wiley Publishing, Hoboken, NJ.

DEMIR, A., LIU, E. W. Y., AND SANGIOVANNI-VINCENTELLI, A. L. 1994. Time-domain non-Monte Carlo noise simulation for nonlinear dynamic circuits with arbitrary excitations. In *Proceedings of the 1994 IEEE/ACM international conference on Computer-aided design*. ICCAD '94. 598–603.

DRENNAN, P. AND MCANDREW, C. 2003. Understanding MOSFET mismatch for analog design. *IEEE Journal of Solid-State Circuits 38,* 3, 450 – 456.

GONG, F., YU, H., AND HE, L. 2009. PiCAP: a parallel and incremental capacitance extraction considering stochastic process variation. In *Proc. ACM/IEEE Design Automation Conf. (DAC)*. 764–769.

GONG, F., YU, H., AND HE, L. 2011. Stochastic analog circuit behaviour modelling by point estimation method. In *Proc. ACM Int. Sym. Physical Design (ISPD)*. 175–182.

GONG, F., YU, H., SHI, Y., KIM, D., REN, J., AND HE, L. 2010a. QuickYield: an efficient global-search based parametric yield estimation with performance constraints. In *Proc. ACM/IEEE Design Automation Conf. (DAC)*. 392–397.

GONG, F., SHI, Y., YU, H., AND HE, L. 2010b. Parametric yield estimation for SRAM cells: concepts, algorithms and challenges. In *ACM/IEEE Design Automation Conf. (DAC), Knowledge Center*.

JAFFARI, J. AND ANIS, M. 2009a. Adaptive sampling for efficient failure probability analysis of SRAM cells. In *Proc. IEEE/ACM Int. Conf. Computer-aided-design (ICCAD)*. 623 –630.

JAFFARI, J. AND ANIS, M. 2009b. Timing yield estimation of digital circuits using a control variate technique. In *International Symposium on Quality of Electronic Design*. 382 –387.

JAFFARI, J. AND ANIS, M. 2011. On efficient LHS-based yield analysis of analog circuits. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems 30,* 1, 159 –163.

KIM, J., JONES, K., AND HOROWITZ, M. 2007. Fast, non-monte-carlo estimation of transient performance variation due to device mismatch. In *Proc. ACM/IEEE Design Automation Conf. (DAC)*. 440 –443.

LAMPAERT, K., GIELEN, G., AND SANSEN, W. 1995. Direct performance-driven placement of mismatch-sensitive analog circuits. In *Proc. ACM/IEEE Design Automation Conf. (DAC)*. 445 –449.

LIU, B., FERNANDEZ, F. V., AND GIELEN, G. 2010. An accurate and efficient yield optimization method for analog circuits based on computing budget aladdress and memetic search technique. In *Proc. IEEE/ACM Design, Automation, and Test in Europe (DATE)*. 1106 –1111.

MCANDREW, C., BATES, J., IDA, R., AND DRENNAN, P. 1997. Efficient statistical BJT modeling, why beta; is more than Ic /Ib. In *Proceedings of the Bipolar/BiCMOS Circuits and Technology Meeting*. 28 –31.

NASSIF, S. R. AND NOWKA, K. J. 2010. Physical design challenges beyond the 22nm node. In *Proc. ACM Int. Sym. Physical Design (ISPD)*. 13–14.

NIEDERREITER, H. 1992. *Random number generation and quasi-Monte Carlo methods*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA.

OEHM, J. AND SCHUMACHER, K. 1993. Quality assurance and upgrade of analog characteristics by fast mismatch analysis option in network analysis environment. *IEEE Journal of Solid-State Circuits 28,* 7, 865 –871.

PELGROM, M., DUINMAIJER, A., AND WELBERS, A. 1989. Matching properties of MOS transistors. *IEEE Journal of Solid-State Circuits 24,* 5, 1433 – 1439.

PILEGGI, L., KESKIN, G., LI, X., MAI, K., AND PROESEL, J. 2008. Mismatch analysis and statistical design at 65nm and below. In *Proc. IEEE Custom Integrated Circuits Conf.* 9–12.

SAWARAGI, Y., NAKAYAMA, H., AND TANINO, T. 1985. *Theory of multiobjective optimization*, Academic Press Inc., Orlando, (1985).

SCHENKEL, F., PRONATH, M., ZIZALA, S., SCHWENCKER, R., GRAEB, H., AND ANTREICH, K. 2001. Mismatch analysis and direct yield optimization by spec-wise linearization and feasibility-guided search. In *Proc. ACM/IEEE Design Automation Conf. (DAC)*. 858 – 863.

SOIN, R. AND SPENCE, R. 1980. Statistical exploration approach to design centring. *IEEE Proceedings of Electronic Circuits and Systems 127,* 6, 260 –269.

SWIDZINSKI, J. AND CHANG, K. 2000. Nonlinear statistical modeling and yield estimation technique for use in Monte Carlo simulations [microwave devices and ics]. *IEEE Transactions on Microwave Theory and Techniques 48,* 12, 2316 –2324.

VRUDHULA, S., WANG, J. M., AND GHANTA, P. 2006. Hermite polynomial based interconnect analysis in the presence of process variations. *IEEE Tran. on Computer-aided-design (TCAD)*, 2001–2011.

WANG, H., YU, H., AND TAN, S. X.-D. 2009. Fast analysis of nontree-clock network considering environmental uncertainty by parameterized and incremental macromodeling. In *Proc. IEEE/ACM Asia South Pacific Design Automation Conf. (ASPDAC)*. 379–384.

XIU, D. AND KARNIADAKIS, G. E. 2002. The Wiener–Askey polynomial chaos for stochastic differential equations. *SIAM Journal of Scientific Computing 24*, 619–644.

YU, H., LIU, X., WANG, H., AND TAN, S.-D. 2010. A fast analog mismatch analysis by an incremental and stochastic trajectory piecewise linear macromodel. In *Proc. IEEE/ACM Asia South Pacific Design Automation Conf. (ASPDAC)*. 211 –216.