Application of Krylov-type Parametric Model Order Reduction in Efficient Uncertainty Quantification of Electro-thermal Circuit Models

Y. Yue[1], **L. Feng**[1], **P. Meuris**[2], **W. Schoenmaker**[2], **and P. Benner**[1]

[1]Max Planck Institute for Dynamics of Complex Technical Systems
Sandtorstr. 1, Magdeburg 39106, Germany
[2]MAGWEL NV, Martelarenplein 13, Leuven B-3000, Belgium

Abstract— As CMOS devices scale down to the nanoscale regime, it becomes increasingly more desired to design systems robust to device variations due to the fabrication process. In robust system design, uncertainty quantification plays an indispensable role. This paper considers uncertainty quantification of power-MOS devices used in energy harvesting. Uncertainty quantification of such a system is usually computationally demanding because it requires either simulating the high-order system at many sampling points, or simulating an even larger system. This paper uses parametric model order reduction techniques to accelerate uncertainty quantification of electro-thermal systems. We embed the reduced model into two uncertainty quantification methods, namely a Latin hypercube sampling method and a stochastic collocation method. Numerical results show that for both methods, uncertainty quantification based on a reduced model not only yields accurate results, but also achieves a significant speedup.

1. INTRODUCTION

In the nanoscale era, we have to consider uncertainty in circuit design because uncertainty is unignorable and unavoidable. Therefore, uncertainty quantification (UQ), which quantifies the uncertainties of the system outputs propagated from the process variations, serves as a useful tool for robust design [11]. This paper studies efficient UQ of power-MOS devices. These devices are commonly used in energy harvesting, where energy from external sources like light and environmental heat are collected in order to power small devices such as implanted biosensors [10]. A major problem in UQ of power-MOS devices is the high computational cost. A finite element model of a power-MOS device, which is derived from a fine 3D mesh of a multi-layered structure consisting of dielectric, vias, contacts, and metal interconnection, is normally of an extremely high order. Therefore, simulation analysis of electro-thermal systems is already computationally demanding. The computational cost of its UQ is even higher since it requires simulating the model at many parameter samples, or simulating a coupled system with a much higher dimension.

The goal of this paper is to use parametric model order reduction (pMOR) techniques [4], which prove to be efficient in many application fields such as circuit simulation, acoustics, and structural vibrations, to reduce the high computational cost in UQ of electro-thermal systems. A pMOR method builds a parametric reduced order model (pROM) that can capture the system dynamics regardless of parameter changes within a certain range. In this work, we use a Krylov-type pMOR method [3], whose goal is to match the (cross-)moments of the state vector, because of its modest requirements on system properties and low computational complexity. To reduce the computational cost of non-intrusive UQ methods, we employ pROM-based methods, which replace the full-order model (FOM) with the pROM in non-intrusive UQ methods. In this paper, we use pROM-based methods to accelerate two UQ methods, namely a Latin hypercube sampling (LHS) method [7] and a stochastic collocation (SC) method [9]. In our numerical test, the system of interest is a nonlinear electro-thermal system with one-way coupling from the electrical part to the thermal part. Numerical results show that a pROM of a low order exhibits high accuracy over a very large parameter range in approximating the original high-order model, and for both UQ methods, pROM-based UQ not only computes accurate results, but also achieves a significant speedup.

2. MATHEMATICAL FORMULATION

The model of interest in this paper is a Power-MOS device shown in Figure 1(a). The device has three contacts: the drain, the source, and the back contact. Here we study the system behavior within the time interval $t \in [0\,\mathrm{s},\ 10^{-6}\,\mathrm{s}]$, where $10^{-6}\,\mathrm{s}$ is the rise time of the source voltage. The

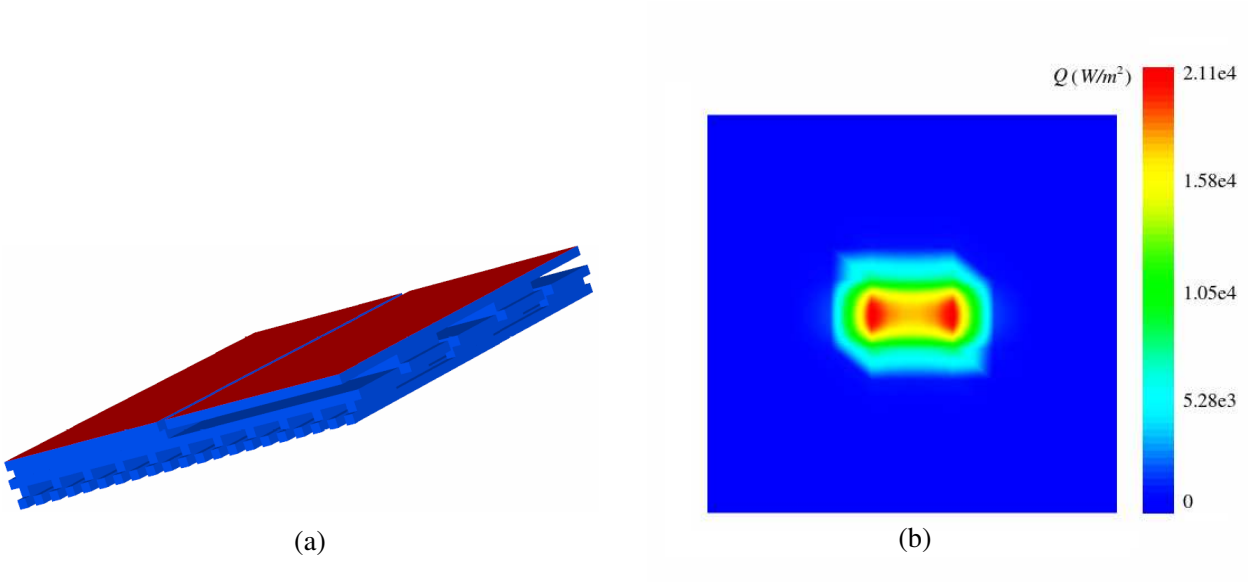

Figure 1: The Power-MOS device and its heat flux density on the back contact at $t = 10^{-6}$ s. (a) Power-MOS device (the back contact is not shown). (b) Heat flux density on the back contact at $t = 10^{-6}$ s.

inputs of the system include the voltages (V) and temperatures (T) of the three contacts, namely:

$$\begin{aligned} V_{\text{drain}}(t) &= V_{\text{back}}(t) = 0\,(V), \quad V_{\text{source}}(t) = 10^7 t\,(V), \quad (t \in [0\,\text{s},\ 10^{-6}\,\text{s}]) \\ T_{\text{drain}}(t) &= T_{\text{source}}(t) = T_{\text{back}}(t) = 300\,(K). \end{aligned}$$

As the voltage on the source rises, the chip is heated up as is shown in Figure 1(b). The outputs we consider here include the currents (I) and thermal fluxes (ϕ) of the three contacts. The electro-thermal model of the power-MOS device that we consider in this paper is

$$A_E(p)x_E(p,t) = -B_E(p)u(t), \quad \text{(Electrical Part)}, \tag{1a}$$
$$E_T(p)\dot{x}_T(p,t) = A_T(p)x_T(p,t) + B_T(p)u(t) + F(p) \times_2 x_E(p) \times_3 x_E(p), \quad \text{(Thermal Part)}, \tag{1b}$$
$$x_T(p,0) = x_T^0, \quad x_E(p,0) = x_E^0, \quad \text{(Initial Conditions)}, \tag{1c}$$
$$y(p,t) = C_E(p)x_E(p,t) + C_T(p)x_T(p,t) + D(p)u, \quad \text{(Output)}. \tag{1d}$$

For the whole system, p represents parameter(s), $u(t) \in \mathcal{R}^l$ is the input vector and $y(p) \in \mathcal{R}^m$ is the output vector. In the electrical part governed by algebraic Equation (1a), $A_E(p) \in \mathcal{R}^{n_E \times n_E}$ is the system matrix, $B_E(p) \in \mathcal{R}^{n_E \times l}$ is the input matrix and $x_E \in \mathcal{R}^{n_E}$ is the state vector. In the thermal part governed by ordinary differential Equation (1b), which is a set of ordinary differential equations, $B_T(p) \in \mathcal{R}^{n_T \times l}$ is the input matrix, $x_T(p) \in \mathcal{R}^{n_T}$ is the state vector, $A_T(p), E_T(p) \in \mathcal{R}^{n_T \times n_T}$ are system matrices. The tensor $F(p) \in \mathcal{R}^{n_T \times n_E \times n_E}$, which can be considered as n_T slices of n_E by n_E matrices $F_i(p) \in \mathcal{R}^{n_E \times n_E}$, $i = 1, \ldots, n_T$, represents the nonlinear coupling from the electrical part to the thermal part. Denoting the i-mode tensor-matrix product by $\times_i$ [2, 8], the product $F(p) \times_2 x_E(p) \times_3 x_E(p)$ is a vector of length n_T, whose i-th component is the standard vector-matrix-vector product $x_E(p)^T F_i(p) x_E(p)$. In the output part (1d), $D \in \mathcal{R}^{m \times l}$ represents the feed through, and $C_E(p) \in \mathcal{R}^{m \times n_E}$ and $C_T(p) \in \mathcal{R}^{m \times n_T}$ represent the output matrices corresponding to the electrical part and the thermal part, respectively. In this formulation, some weak factors are ignored, e.g., the Joule heating related to the input variables and the coupling from the thermal part to the electrical part. System (1b) shows that the system has a one-way coupling from the electrical part to the thermal part through the tensor $F(p)$, and the coupling the other way round is ignored. All parametric matrices in system (1) are of the form

$$Y(p) = Y_c + pY_v, \quad Y \in \{A_E, A_T, B_E, B_T, C_E, C_T, D, E_T, F\}. \tag{2}$$

In our numerical tests, p represents a single parameter σ, the conductivity of the third metal layer.

3. KRYLOV-TYPE PMOR METHODS

Krylov-type pMOR methods belong to projection-type pMOR methods. To reduce a general parametric linear system of order n:

$$(G_0 + p_1 G_1 + p_2 G_2 + \ldots + p_q G_q)\, X = Pu, \tag{3}$$

where $G_i \in \mathcal{R}^{n\times n}$ $(0 \le i \le q)$, $P \in \mathcal{R}^{n\times l}$ and $u \in \mathcal{R}^l$, a projection-type pMOR method generates a basis $V \in \mathcal{R}^{n\times k}$ for an order-n system, with which we can build an order-k pROM of the form

$$\left(\hat{G}_0 + p_1 \hat{G}_1 + p_2 \hat{G}_2 + \ldots + p_l \hat{G}_q\right) \hat{X} = \hat{P}u, \tag{4}$$

where $\hat{G}_i = V^T G_i V$ and $\hat{P} = V^T P$ [4]. Different projection-type pMOR methods vary in the way of constructing V. In this paper, we use the Krylov-type method proposed in [3], which matches the dominant multi-moments to a certain order implicitly in a numerically stable manner.

To reduce the electro-thermal system (1), we need to conduct pMOR on both (1a) and (1b). Since system (1a) can easily be rearranged into form (3), Krylov-type pMOR methods can be directly applied to compute a basis V_{E} and the corresponding pROM. To reduce system (1b), however, we have to conduct the Laplace transform first to obtain a frequency domain representation. Following the approach presented in [6], we first ignore the nonlinear part in system (1b) and conduct the Laplace transform to obtain its frequency domain representation

$$(A_1 + \sigma A_2 - sE_1 - (\sigma s)E_2)\, X = \left[B_1\ B_2\ A_1 x_T^0\ A_2 x_T^0\right] \left[-U^T\ -\sigma U^T\ \frac{-1}{s}\ \frac{-\sigma}{s}\right]^T, \tag{5}$$

where X and U represent the Laplace transforms of the state vector $x_T(p,t)$ and the input vector u, respectively, and s and σ denote the radial frequency and the conductivity, respectively. System (5) can be reduced by the pMOR method in [3] with $G_0 \leftarrow A_1$, $G_1 \leftarrow A_2$, $G_3 \leftarrow -E_1$, $G_4 \leftarrow -E_2$, $p_1 \leftarrow \sigma$, $p_2 \leftarrow s$, $p_3 \leftarrow \sigma s$, and $P = [B_1\ B_2\ A_1 x_T^0\ A_2 x_T^0]$. Denote the basis built for (5) by V_T. To obtain a pROM for (1b), we approximate x_E by $V_E \hat{x}_E$ and x_T by $V_T \hat{x}_T$, and then force the approximation error to be orthogonal to the range of V_T. The resulting pROM is

$$\hat{E}_T(p)\dot{\hat{x}}_T(p,t) = \hat{A}_T(p)\hat{x}_T(p,t) + \hat{B}_T(p)u + \hat{F}(p) \times_2 \hat{x}_E(p) \times_3 \hat{x}_E(p), \tag{6}$$

where $\hat{E}_T(p) = V_T^T E_T(p) V_T$, $\hat{A}_T(p) = V_T^T A_T(p) V_T$, $\hat{B}_T(p) = V_T^T B_T(p)$, $\hat{F}(p) = F(p) \times_1 V_T \times_2 V_E \times_3 V_E$. To obtain the reduced tensor $\hat{F}(p)$, we first approximate $x_E(p)$ in the range of V_E, and then project the approximation onto the test subspace V_T, i.e., the tensor product $\hat{F}(p) \times_2 \hat{x}_E(p) \times_3 \hat{x}_E(p)$ equals $V_T^T [F(p) \times_2 (V_E \hat{x}_E(p)) \times_3 (V_E \hat{x}_E(p))]$. The advantage of the tensor formulation for the ROM is that using the reduced tensor, evaluating the ROM does not require computations with quantities of the order of the FOM. In our actual computations, the parametric matrices in the ROM are computed by

$$\hat{Y}(p) = \hat{Y}_c + p\hat{Y}_v, \quad \hat{Y} \in \left\{\hat{A}_T, \hat{B}_T, \hat{C}_T, \hat{E}_T, \hat{F}\right\}, \tag{7}$$

where $\hat{Y}_c$ and $\hat{Y}_v$ are pre-computed at the construction of the pROM.

4. PROM-BASED UQ

UQ methods can be categorized into non-intrusive methods and intrusive methods [9]. Non-intrusive methods conduct UQ by solving the original deterministic system, e.g., system (1), at various parameter points. Intrusive methods, however, require building a large-scale coupled system, which is often of a much higher order than the original deterministic system. In this paper, we focus on non-intrusive methods since the pROM of the original deterministic model (FOM) can be directly used to replace the FOM in UQ. We embed our pROMs into two UQ methods, namely the Latin hypercube sampling (LHS) method and the stochastic collocation (SC) method.

- LHS [7]. To obtain n samples, LHS divides the input distribution into n intervals of equal probability, and selects one sample randomly in each interval. The mean and standard deviation of the samples are used to approximate those of the original continuous model. Compared to the standard Monte-Carlo sampling, LHS ensures a set of evenly distributed samples.

- SC [9]. Since the mean and the standard deviation can be computed via numerical integration, SC uses a quadrature rule to present the relevant integrals as a weighted sum of the corresponding function values at the collocation points. As we conduct UQ on a single normally-distributed random variable, we use the Gauss-Hermite quadrature rule.

The computationally dominant part of both LHS and SC is the simulation of the high-order FOM at all sampled points p_i. Since our pROMs are highly accurate for these simulations as we will show in §5, pROM-based UQ replaces the FOM (1) with pROMs for these simulations to achieve a significant speedup.

5. NUMERICAL TESTS

In this section, we test the effectiveness and efficiency of pROM-based UQ. All codes are implemented in MATLAB®. For the SC method, we use the SGMGA library [5] to compute the quadrature rule. For tensor computations, we use the Tensor Toolbox [1, 2].

First, we check the quality of the pROMs in a simulation analysis. For (1), we build an order-2 pROM for the order-1660 FOM of the electrical part (1a), and an order-50 pROM for the order-11556 FOM of the thermal part (1b). Both pROMs are built at the interpolation point $\sigma = 3 \times 10^7\,\mathrm{S/m}$ with the radial frequency shifts $s_i = \frac{(i-1)\times 10^8}{9}\,\mathrm{rad/s}$ $(i = 1, 2, \ldots, 10)$. Figure 2 shows the relative error for thermal fluxes for $\sigma = 10\,\mathrm{S/m}$ and $\sigma = 10^{11}\,\mathrm{S/m}$, which are extreme points far away from the interpolation point $\sigma = 3 \times 10^7\,\mathrm{S/m}$. When the system starts, the relative errors are high because: 1) the thermal fluxes are close to zero since the system is hardly heated up; 2) the thermal outputs are dominated by modeling error and numerical error at the starting stage:

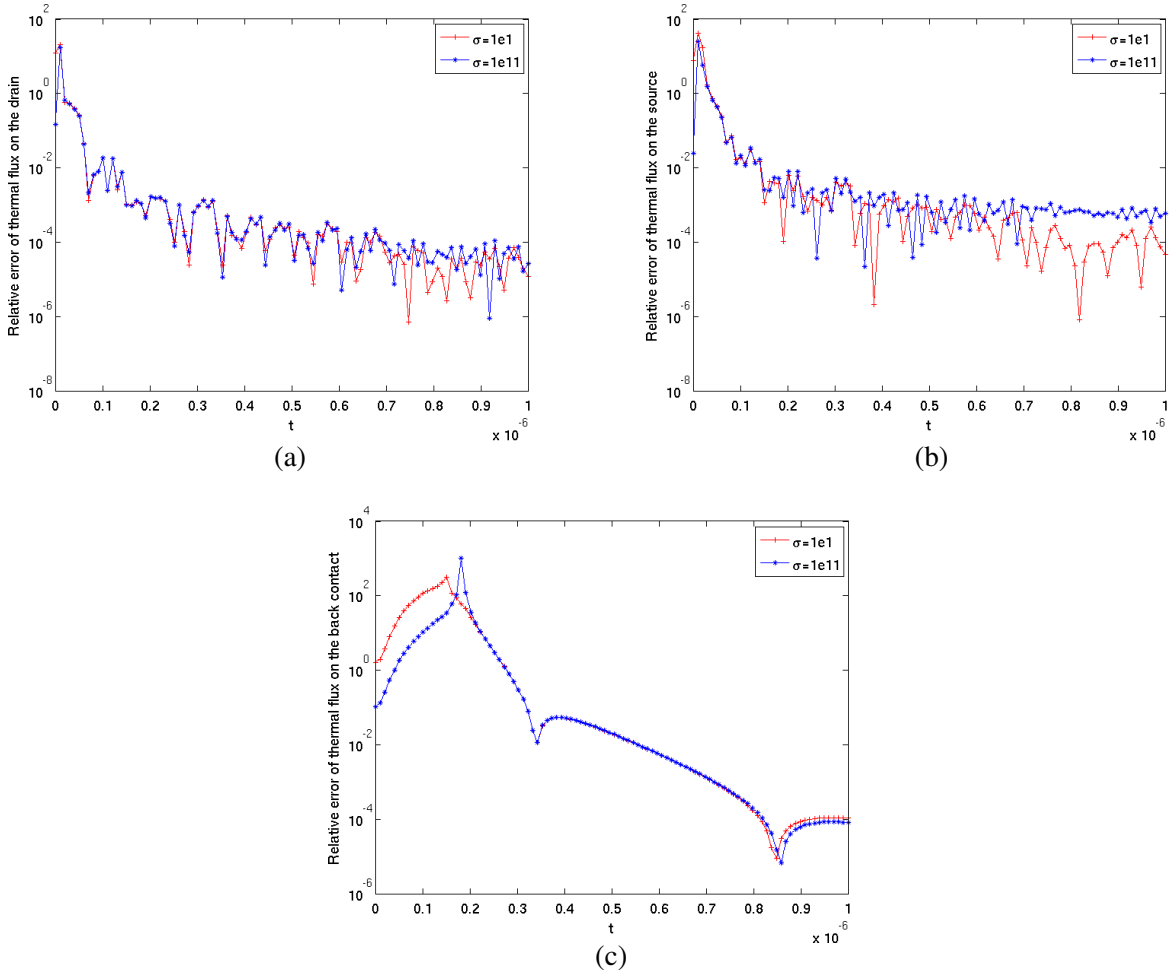

Figure 2: The evolution of the relative errors of the thermal outputs for extreme σ's at $t = 10^{-6}\,\mathrm{s}$. (a) Relative error of the thermal flux on the drain. (b) Relative error of the thermal flux on the source. (c) Relative error of the thermal flux on the back contact.

the outputs of the FOM also oscillate slightly around zero even when the temperature is the same everywhere and no voltage excitation is imposed. However, as time elapses, the relative errors decrease to the order of 10^{-4} and therefore, the dominant dynamics are accurately captured over a large parameter range.

Now we apply the pROMs to UQ analysis of the electro-thermal system (1). Here we conduct UQ on the outputs at $t = 10^{-6}$ s. We assume that the conductivity obeys the normal distribution $\mathcal{N}(3 \times 10^7, (10^7)^2)$. The numerical results in Table 1 show that for both UQ methods, pROM-based UQ computes highly accurate means ($E(\cdot)$). For the nontrivial electrical outputs I_{drain} and I_{source}, which are sensitive to the change in the conductivity with the same coefficient of variation (CV) of 33.23%, the standard deviations ($\sigma(\cdot)$) are also computed with high accuracy. The thermal outputs ϕ_{drain}, ϕ_{source} and ϕ_{back}, however, are insensitive to the change in the conductivity with CV's of 3.77e-07, 4.4138e-07 and 2.3483e-08, respectively. Although the standard deviations to these insensitive thermal outputs are not of so high accuracy, the relative orders are correct.

Table 1: UQ results for the outputs at $t = 10^{-6}$ s.

	LHS using FOM	LHS using ROM	SC using FOM	SC using ROM
$E(I_{\text{drain}})$	7.4621e-04	7.4621e-04	7.4602e-04	7.4602e-04
$\sigma(I_{\text{drain}})$	2.4794e-04	2.4794e-04	2.4867e-04	2.4867e-04
$E(I_{\text{source}})$	−7.4621e-04	−7.4621e-04	−7.4602e-04	−7.4602e-04
$\sigma(I_{\text{source}})$	2.4794e-04	2.4794e-04	2.4867e-04	2.4867e-04
$E(I_{\text{back}})$	0	0	0	0
$\sigma(I_{\text{back}})$	0	0	0	0
$E(\phi_{\text{drain}})$	5.8479e-04	5.8478e-04	5.8479e-04	5.8479e-04
$\sigma(\phi_{\text{drain}})$	1.5838e-10	1.5677e-10	1.5985e-10	1.5719e-10
$E(\phi_{\text{source}})$	4.1977e-04	4.1975e-04	4.1977e-04	4.1977e-04
$\sigma(\phi_{\text{source}})$	1.8528e-10	9.1986e-11	4.6370e-11	9.2124e-11
$E(\phi_{\text{back}})$	6.6781e-07	6.6773e-07	6.6781e-07	6.6781e-07
$\sigma(\phi_{\text{back}})$	1.5682e-14	1.7778e-14	1.1199e-14	1.6189e-14
Number of sampled points	100	100	11	11
CPU time	6001.14 s	94.19 s	733.64 s	30.51 s

ACKNOWLEDGMENT

This work is funded by the Collaborative Project nanoCOPS, Nanoelectronic COupled Problems Solutions, which is supported by the European Union in the FP7-ICT-2013-11 Program under Grant Agreement Number 619166. Website: http://www.fp7-nanocops.eu.

REFERENCES

1. Bader, B. W. and T. G. Kolda, "Efficient MATLAB computations with sparse and factored tensors," *SIAM Journal on Scientific Computing*, Vol. 30, 205–231, 2007.
2. Bader, B. W., T. G. Kolda, et al., "Matlab tensor toolbox version 2.6," available online, Feb. 2015.
3. Benner, P. and L. Feng, "A robust algorithm for parametric model order reduction based on implicit moment matching," *Reduced Order Methods for Modeling and Computational Reduction*, A. Quarteroni and G. Rozza, eds., Vol. 9 of *MS&A — Modeling, Simulation and Applications*, 159–185, Springer International Publishing, 2014.
4. Benner, P., S. Gugercin, and K. Willcox, "A survey of model reduction methods for parametric systems," Max Planck Institute Magdeburg, Preprint MPIMD/13-14, MPI-Magdeburg, Aug. 2013, available from http://www.mpi-magdeburg.mpg.de/preprints/.
5. Burkardt, J., "SGMGA: Sparse grid mixed growth anisotropic rules," http://people.sc.fsu.edu/∼jburkardt/m_src/sgmga/sgmga.html.
6. Chen, Y., "Model order reduction for nonlinear systems," Master's Thesis, Massachusetts Institute of Technology, 1999.

7. Helton, J. and F. Davis, "Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems," *Reliability Engineering & System Safety*, Vol. 81, 23–69, 2003.
8. Kolda, T. G. and B. W. Bader, "Tensor decompositions and applications," *SIAM Review*, Vol. 51, 455–500, 2009.
9. Le Maître, O. and O. Knio, *Spectral Methods for Uncertainty Quantification: With Applications to Computational Fluid Dynamics, Scientific Computation*, Springer, Netherlands, 2010.
10. Spirito, P., G. Breglio, V. d'Alessandro, and N. Rinaldi, "Thermal instabilities in high current power MOS devices: Experimental evidence, electro-thermal simulations and analytical modeling," *23rd International Conference on Microelectronics, MIEL*, Vol. 1, 23–30, 2002.
11. Zhang, Z., I. Elfadel, and L. Daniel, "Uncertainty quantification for integrated circuits: Stochastic spectral methods," *IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*, 803–810, 2013.