

# Uncertainty Quantification of Electronic and Photonic ICs with Non-Gaussian Correlated Process Variations

Chunfeng Cui and Zheng Zhang

Department of Electrical & Computer Engineering, University of California, Santa Barbara, CA 93106  
chunfengcui@ucsb.edu; zhengzhang@ece.ucsb.edu

## ABSTRACT

Since the invention of generalized polynomial chaos in 2002, uncertainty quantification has impacted many engineering fields, including variation-aware design automation of integrated circuits and integrated photonics. Due to the fast convergence rate, the generalized polynomial chaos expansion has achieved orders-of-magnitude speedup than Monte Carlo in many applications. However, almost all existing generalized polynomial chaos methods have a strong assumption: the uncertain parameters are mutually independent or Gaussian correlated. This assumption rarely holds in many realistic applications, and it has been a long-standing challenge for both theorists and practitioners.

This paper propose a rigorous and efficient solution to address the challenge of non-Gaussian correlation. We first extend generalized polynomial chaos, and propose a class of smooth basis functions to efficiently handle non-Gaussian correlations. Then, we consider high-dimensional parameters, and develop a scalable tensor method to compute the proposed basis functions. Finally, we develop a sparse solver with adaptive sample selections to solve high-dimensional uncertainty quantification problems. We validate our theory and algorithm by electronic and photonic ICs with 19 to 57 non-Gaussian correlated variation parameters. The results show that our approach outperforms Monte Carlo by 2500 $\times$  to 3000 $\times$  in terms of efficiency. Moreover, our method can accurately predict the output density functions with multiple peaks caused by non-Gaussian correlations, which is hard to handle by existing methods.

Based on the results in this paper, many novel uncertainty quantification algorithms can be developed and can be further applied to a broad range of engineering domains.

## 1. INTRODUCTION

Uncertainties are unavoidable in almost all engineering fields, and they should be carefully quantified and managed in order to improve design reliability and robustness. In semiconductor chip design, a major source of uncertainty is the fabrication process variations. Process variations are significant in deeply scaled electronic integrated circuits (ICs) [1] and MEMS [2], and they have also become a major concern in emerging design technologies such as integrated photonics [3]. A popular uncertainty quantification method is Monte Carlo [4], which is easy to implement but has a low convergence rate. In recent years, various stochastic spectral methods (e.g., stochastic Galerkin [5], stochastic testing [6] and stochastic collocation [7]) have been developed and have achieved orders-of-magnitude speedup than Monte Carlo in vast applications. These methods represent a stochastic solution as the linear combination of some basis functions (e.g., generalized polynomial chaos [8]), and they can obtain highly accurate solutions at a low computational cost when the parameter dimensionality is not high.

Stochastic spectral methods have been successfully ap-

plied in the variation-aware modeling and simulation of many devices and circuits, including (but not limited to) VLSI interconnects [9–12], nonlinear ICs [6, 13, 14], MEMS [2, 15] and photonic circuits [16]. A major challenge of stochastic spectral methods is the curse of dimensionality: a huge number of basis functions and simulation samples may be required as the number of random parameters becomes large. In recent years, there has been significant improvement to address this challenge. Representative techniques include (but are not limited to) compressive sensing [17, 18], analysis of variance [15, 19], stochastic model order reduction [20], hierarchical methods [15, 21, 22] and tensor computation [22–24].

**Major Challenge.** Despite their great success, existing stochastic spectral methods are limited by a long-standing challenge: the generalized polynomial-chaos basis functions require all random parameters to be **mutually independent** [8]. This is a very strong assumption, and it fails in many realistic cases. For instance, a lot of device geometric or electrical parameters are highly correlated because they are influenced by the same fabrication steps; many circuit-level performance parameters used in system-level analysis depend on each other due to the network coupling and feedback. Data-processing techniques such as principal or independent component analysis [25, 26] can handle Gaussian correlations, but they cause huge errors in general **non-Gaussian correlated** cases. A modified and non-smooth chaos representation was proposed in [27], and it was applied to the uncertainty analysis of silicon photonics [16]. However, the method in [27] does not converge well, and designers cannot easily extract statistical information (e.g., mean value and variance) from the solution.

**Paper Contributions.** This paper proposes a novel and rigorous solution to handle the challenging non-Gaussian correlated process variations in electronic ICs and integrated photonics. The specific contributions include:

- Derivation and implementation of a class of basis functions for non-Gaussian correlated random parameters. The proposed basis functions can overcome the theoretical limitations of [27]: they are smooth and can provide highly accurate solutions for non-Gaussian correlated cases (like the standard generalized polynomial chaos [8] does for independent parameters), and it can provide closed-form expressions for the mean values and variance of a stochastic solution. In order to make our methods scalable, we also propose a highly efficient functional tensor-train method to compute the basis functions for many non-Gaussian correlated random parameters equipped with a Gaussian-mixture density function.
- An adaptive-sampling sparse solver. In order to apply our method to electronic and photonic ICs with non-Gaussian correlated variations, we develop an  $\ell_0$ -minimization framework to compute the sparse coefficients of our basis functions. Our *main contribution* is an adaptive sampling approach: instead of setting up a compressive-sensing prob-

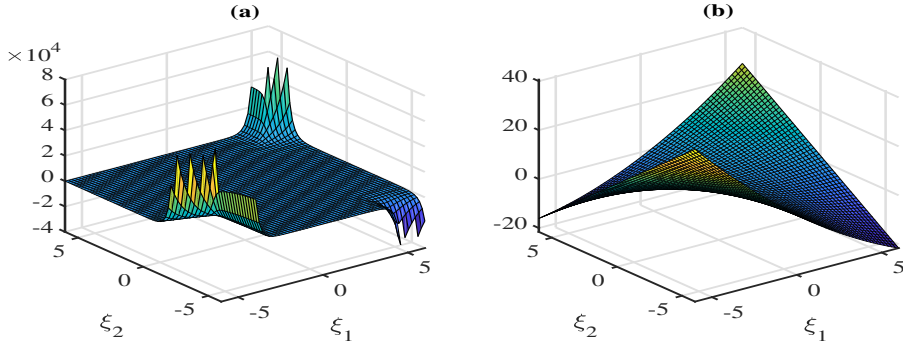


Figure 1: (a): A two-variable basis function by [27]; (b): a basis function obtained by our proposed method.

lem using random samples (as done in [17]), we select the most informative samples via a rank-revealing QR procedure and use a D-optimal method to add new samples and to update the solution.

- Validation on electronic and photonic ICs. We demonstrate the effectiveness of our framework on electronic and photonic IC examples with 19 to 57 non-Gaussian correlated process variations. Our method can accurately predict the statistical information (e.g., multi-peak probability density function and mean value) of the circuit performance, and it is faster than Monte Carlo by about 3000 $\times$  when the similar level of accuracy is required.

Our uncertainty quantification framework has the following two excellent features simultaneously: it does not need any error-prone de-correlation step such as independent component analysis [26], and it has the similarly high performance for non-Gaussian correlated uncertainties as generalized polynomial chaos [8] does for independent uncertainties.

## 2. PRELIMINARIES

### 2.1 Generalized Polynomial Chaos

Let  $\xi = [\xi_1, \dots, \xi_d] \in \mathbb{R}^d$  denote  $d$  random parameters with a joint probability density function  $\rho(\xi)$ , and  $y(\xi) \in \mathbb{R}$  be a parameter-dependent performance metric (e.g., the power consumption or frequency of a chip). When  $y(\xi)$  is smooth and has a bounded variance, stochastic spectral methods aim to approximate  $y(\xi)$  via a truncated generalized polynomial-chaos expansion [8]:

$$y(\xi) \approx \sum_{|\alpha|=0}^p c_\alpha \Psi_\alpha(\xi), \quad (1)$$

where  $c_\alpha$  is the coefficient, and  $\{\Psi_\alpha(\xi)\}$  are orthonormal polynomials satisfying

$$\mathbb{E}[\Psi_\alpha(\xi) \Psi_\beta(\xi)] = \begin{cases} 1, & \text{if } \alpha = \beta; \\ 0, & \text{otherwise.} \end{cases} \quad (2)$$

Here the operator  $\mathbb{E}$  denotes expectation;  $\alpha = [\alpha_1, \dots, \alpha_d] \in \mathbb{N}^d$  is a vector, with each element  $\alpha_i$  being the highest polynomial order in terms of  $\xi_i$ . The total polynomial order  $|\alpha| = |\alpha_1| + \dots + |\alpha_d|$  is bounded by  $p$ , and thus the total number of basis functions is  $N = (p+d)!/(p!d!)$ . The unknown coefficients  $c_\alpha$ 's can be computed via various numerical solvers such as stochastic Galerkin [5], stochastic testing [6] and stochastic collocation [7]. Once  $c_\alpha$ 's are computed, the mean value, variance and density function of  $y(\xi)$  can be easily obtained.

The generalized polynomial-chaos theory [8] assumes that all random parameters are mutually independent. In other

words, if  $\rho_k(\xi_k)$  denotes the marginal density of  $\xi_k$ , then the joint density is  $\rho(\xi) = \prod_{k=1}^d \rho_k(\xi_k)$ . Under this assumption, a multivariate basis function has a product form:

$$\Psi_\alpha(\xi) = \prod_{k=1}^d \phi_{k,\alpha_k}(\xi_k). \quad (3)$$

Here  $\phi_{k,\alpha_k}(\xi_k)$  is a univariate degree- $\alpha_k$  orthonormal polynomial of parameter  $\xi_k$ , and it is adaptively chosen based on  $\rho_k(\xi_k)$  via the three-term recurrence relation [28].

### 2.2 Existing Solutions for Correlated Cases

The random parameters  $\xi$  are rarely guaranteed to be independent in realistic cases. It is easy to de-correlate Gaussian correlated random parameters via principal or independent component analysis [25, 26], but de-correlating non-Gaussian correlated parameters can be error-prone. In [27], Soize and Ghanem suggested the following basis function:

$$\Psi_\alpha(\xi) = \left( \frac{\prod_{k=1}^d \rho_k(\xi_k)}{\rho(\xi)} \right)^{\frac{1}{2}} \prod_{k=1}^d \phi_{k,\alpha_k}(\xi_k). \quad (4)$$

The modified basis functions are guaranteed to be orthonormal even if  $\rho(\xi) \neq \prod_{k=1}^d \rho_k(\xi_k)$ , but they have two limitations as shown by the numerical results in [16]:

- Firstly, the basis functions are very non-smooth and numerically unstable due to the first part on the right-hand side of (4). This is demonstrated in Fig. 1 (a). As a result, the modified basis functions have a much slower convergence rate compared with the standard method in [8].
- Secondly, the basis functions in (4) do not allow an explicit expression for the expectation and variance of  $y(\xi)$ . This is caused by the fact the basis function indexed by  $\alpha = 0$  is not a constant.

### 2.3 Background: Tensor Train Decomposition

A tensor is a generalization of a vector and a matrix. A vector  $\mathbf{a} \in \mathbb{R}^n$  is a one-way data array; a matrix  $\mathbf{A} \in \mathbb{R}^{n_1 \times n_2}$  is two-way data array; a tensor  $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times \dots \times n_d}$  is a  $d$ -way data array. We refer readers to [24, 29] for detailed tensor notations and operations, and its application in EDA [24].

A high-way tensor has  $O(n^d)$  elements, leading to a prohibitive computation and storage cost. Fortunately, realistic data can often be factorized using tensor decomposition techniques [29]. Tensor train decomposition [30] is

very suitable for factorizing high-way tensors, and it only needs  $O(dr^2n)$  elements to represent a high-way data array. Specifically, given a  $d$ -way tensor  $\mathcal{A}$ , the tensor-train decomposition represents each element  $a_{i_1 i_2 \dots i_d}$  as

$$a_{i_1 i_2 \dots i_d} = \mathbf{A}_1(i_1) \mathbf{A}_2(i_2) \dots \mathbf{A}_d(i_d), \forall i_k = 1, 2, \dots, n_k, \quad (5)$$

where  $\mathbf{A}_k(i_k)$  is an  $r_{k-1} \times r_k$  matrix, and  $r_0 = r_d = 1$ . Given two tensors  $\mathcal{A}$  and  $\mathcal{B}$  and their tensor train decompositions. If we want to compute the tensor train decomposition of their Hadamard (element-wise) product

$$\mathcal{C} = \mathcal{A} \circ \mathcal{B} \iff c_{i_1 i_2 \dots i_d} = a_{i_1 i_2 \dots i_d} b_{i_1 i_2 \dots i_d},$$

then the result can be directly obtained via

$$\mathbf{C}_k(i_k) = \mathbf{A}_k(i_k) \otimes \mathbf{B}_k(i_k). \quad (6)$$

Here  $\otimes$  denotes a matrix Kronecker product.

### 3. BASIS FUNCTIONS FOR NON-GAUSSIAN CORRELATED CASES

When process variations are non-Gaussian correlated, the joint probability density of  $\boldsymbol{\xi}$  is not the product of  $\rho_k(\xi_k)$ 's, and the basis functions in (3) cannot be employed. This section derives a set of multivariate polynomial basis functions. These basis functions can be obtained if a multivariate moment computation framework is available. A broad class of non-Gaussian correlated parameters are described by Gaussian mixture models. For these cases, we propose a fast functional tensor-train method to compute the desired multivariate basis functions.

#### 3.1 Proposed Multivariate Basis Functions

We aim to generate a set of multivariate orthonormal polynomials with respect to the joint density  $\rho(\boldsymbol{\xi})$ , such that they have the excellent properties of the generalized polynomial chaos [8] even for non-Gaussian correlated cases. Several orthogonal polynomials exist for a few specific density functions [31]. In general, one may construct multivariate orthogonal polynomials via the three-term recurrence in [32] or [33]. However, the theories in [32, 33] either are hard to implement or can only guarantee weak orthogonality.

Inspired by [34], we present a simple yet efficient method for computing a set of multivariate orthonormal polynomial basis functions. Let  $\boldsymbol{\xi}^\alpha = \xi_1^{\alpha_1} \xi_2^{\alpha_2} \dots \xi_d^{\alpha_d}$  be a monomial indexed by  $\boldsymbol{\alpha}$ , then the corresponding moment is

$$m_\alpha = \mathbb{E}[\boldsymbol{\xi}^\alpha] := \int \boldsymbol{\xi}^\alpha \rho(\boldsymbol{\xi}) d\boldsymbol{\xi}. \quad (7)$$

We intend to construct  $N = \frac{(p+d)!}{p!d!}$  multivariate orthonormal polynomials  $\{\Psi_\alpha(\boldsymbol{\xi})\}$  with their total degrees  $|\boldsymbol{\alpha}| \leq p$ . For convenience, we resort all monomials in the graded lexicographic order, denoted as  $\mathbf{b}(\boldsymbol{\xi}) = [b_1(\boldsymbol{\xi}), \dots, b_N(\boldsymbol{\xi})]^T$ . We further denote the multivariate moment matrix as  $\mathbf{M}$

$$\mathbf{M} = \mathbb{E}(\mathbf{b}(\boldsymbol{\xi})\mathbf{b}^T(\boldsymbol{\xi})) \iff m_{ij} = \mathbb{E}[b_i(\boldsymbol{\xi})b_j(\boldsymbol{\xi})]. \quad (8)$$

For instance, if  $d = 2$  and  $p = 2$ , then the monomials are ordered as

$$\mathbf{b}(\boldsymbol{\xi}) = [1, \xi_1, \xi_2, \xi_1^2, \xi_1\xi_2, \xi_2^2]^T.$$

The total number of monomials is  $N = 6$ , and the corresponding  $\mathbf{M}$  is a 6-by-6 matrix.

Because  $\mathbf{M}$  is a symmetric positive definite matrix, a lower-triangular matrix  $\mathbf{L}$  is calculated via the Cholesky factorization  $\mathbf{M} = \mathbf{L}\mathbf{L}^T$ . Finally, we define our basis functions as

$$\boldsymbol{\Psi}(\boldsymbol{\xi}) = \mathbf{L}^{-1}\mathbf{b}(\boldsymbol{\xi}). \quad (9)$$

Here the  $N$ -by-1 functional vector  $\boldsymbol{\Psi}(\boldsymbol{\xi})$  stores all basis functions  $\{\Psi_\alpha(\boldsymbol{\xi})\}$  in the graded lexicographic order.

**Properties of the Basis Functions.** Our proposed basis functions have the following excellent properties:

- 1). Our basis functions are smooth (c.f. Fig. 1 (b)). In fact, the basis function  $\Psi_\alpha(\boldsymbol{\xi})$  is a multivariate polynomial of a total degree  $|\boldsymbol{\alpha}|$ . It differs from the standard generalized polynomial chaos [8] in the sense that our basis functions are not the product of univariate polynomials.
- 2). All basis functions are orthonormal to each other. This can be easily seen from

$$\mathbb{E}(\boldsymbol{\Psi}(\boldsymbol{\xi})\boldsymbol{\Psi}^T(\boldsymbol{\xi})) = \mathbf{L}^{-1}\mathbf{M}\mathbf{L}^{-T} = \mathbf{I}.$$

This property is important for the sparse approximation and for extracting the statistical information of  $y(\boldsymbol{\xi})$ .

- 3). Due to the orthonormality of  $\{\Psi_\alpha(\boldsymbol{\xi})\}$ , there exist closed-form formula for the expectation and variance of  $y(\boldsymbol{\xi})$ :

$$\mathbb{E}[y(\boldsymbol{\xi})] \approx \sum_{|\boldsymbol{\alpha}|=0}^p c_\alpha \mathbb{E}[\Psi_\alpha(\boldsymbol{\xi})] = c_0, \quad (10)$$

$$\text{var}[y(\boldsymbol{\xi})] = \mathbb{E}(y^2(\boldsymbol{\xi})) - \mathbb{E}^2(y(\boldsymbol{\xi})) \approx \sum_{|\boldsymbol{\alpha}|=1}^p c_\alpha^2. \quad (11)$$

A key step in constructing our basis functions is to compute a set of moments  $\{m_\alpha\}$  for all  $|\boldsymbol{\alpha}| \leq 2p$ . In general, this can be done with the help of the Rosenblatt transform [35]. Suppose  $\rho(\boldsymbol{\xi}) = \sum_{i=1}^n w_i \rho_i(\boldsymbol{\xi})$ . Based on the joint cumulative density function, the Rosenblatt formulate a function transformation  $\boldsymbol{\xi} = \mathcal{T}_i(\boldsymbol{\eta})$  such that  $\tilde{\rho}_i(\boldsymbol{\eta}) = \beta_i \rho_i(\mathcal{T}_i(\boldsymbol{\eta}))$  is the density function of the mutually independent parameters  $\boldsymbol{\eta}$ . Here,  $\beta_i$  is a coefficient to ensure that the integral of  $\tilde{\rho}_i(\boldsymbol{\eta})$  is one. With the Rosenblatt transform, we have

$$\mathbb{E}[\boldsymbol{\xi}^\alpha] = \sum_{i=1}^n w_i \int_{-\infty}^{\infty} \boldsymbol{\xi}^\alpha \rho_i(\boldsymbol{\xi}) d\boldsymbol{\xi} = \sum_{i=1}^n w_i \int_{-\infty}^{\infty} (\mathcal{T}_i(\boldsymbol{\eta}))^\alpha \tilde{\rho}_i(\boldsymbol{\eta}) d\boldsymbol{\eta}.$$

Then we can use the sparse grid technique [36] to numerically compute  $\mathbb{E}[\boldsymbol{\xi}^\alpha]$ . We can also compute a high-dimensional integration via tensor trains as has been done in [22].

#### 3.2 Moments for Gaussian-Mixture Models

In practice, semiconductor foundries usually have a lot of measurement data about process variations, and they generate a joint density function  $\rho(\boldsymbol{\xi})$  to fit the measurement data set. An excellent choice for this data-driven modeling flow is the Gaussian-mixture model. A Gaussian mixture model describes the joint density function as

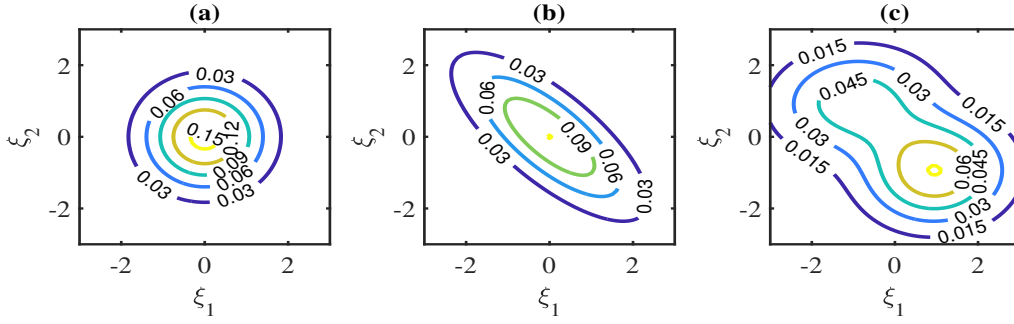
$$\rho(\boldsymbol{\xi}) = \sum_{i=1}^n w_i \mathcal{N}(\boldsymbol{\xi} | \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i), \text{ with } w_i > 0, \sum_{i=1}^n w_i = 1. \quad (12)$$

Here  $\mathcal{N}(\boldsymbol{\xi} | \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$  is multi-variate Gaussian density function with mean  $\boldsymbol{\mu}_i \in \mathbb{R}^d$  and a positive definite covariance matrix  $\boldsymbol{\Sigma}_i \in \mathbb{R}^{d \times d}$ . Please note that a Gaussian mixture model describes a non-Gaussian correlated joint density function, as shown in Fig. 2. Now the moment is

$$m_\alpha = \sum_{i=1}^n w_i q_{\alpha,i}, \text{ with } q_{\alpha,i} = \int_{-\infty}^{\infty} \boldsymbol{\xi}^\alpha \mathcal{N}(\boldsymbol{\xi} | \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i) d\boldsymbol{\xi}, \quad (13)$$

and we need to compute  $q_{\alpha,i}$  for  $i = 1, 2, \dots, n$ .

For simplicity, we ignore the index  $i$  in  $\boldsymbol{\mu}_i$ ,  $\boldsymbol{\Sigma}_i$  and  $q_{\alpha,i}$ . Let  $\mathbf{A}$  be the lower triangular matrix from the Cholesky decomposition of  $\boldsymbol{\Sigma}$  (i.e.,  $\boldsymbol{\Sigma} = \mathbf{A}\mathbf{A}^T$ ), and let  $\boldsymbol{\xi} = \mathbf{A}\boldsymbol{\eta} + \boldsymbol{\mu}$ ,



**Figure 2: Joint density functions for different cases. (a): independent Gaussian; (b): correlated Gaussian; (c): correlated non-Gaussian (e.g., a Gaussian-mixture distribution).**

then  $\boldsymbol{\eta}$  is a vector with standard a Gaussian distribution. Consequently,  $q_\alpha$  can be calculated via the integral of  $\boldsymbol{\eta}$ :

$$\begin{aligned} q_\alpha &= \int_{-\infty}^{\infty} \boldsymbol{\xi}^\alpha \mathcal{N}(\boldsymbol{\xi}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) d\boldsymbol{\xi} \\ &= \int_{-\infty}^{\infty} (\mathbf{A}\boldsymbol{\eta} + \boldsymbol{\mu})^\alpha \frac{\exp(-\boldsymbol{\eta}^T \boldsymbol{\eta})}{\sqrt{(2\pi)^d}} d\boldsymbol{\eta}. \end{aligned} \quad (14)$$

In this formulation,  $(\mathbf{A}\boldsymbol{\eta} + \boldsymbol{\mu})^\alpha$  is not the product of univariate functions of each  $\eta_i$ , therefore, the above integration is still hard to compute. We show that  $q_\alpha$  can be computed exactly with an efficient functional tensor-train method.

### 3.3 Functional Tensor-Train Implementation

In this subsection, we show that there exists a matrix  $\mathbf{G}_0 \in \mathbb{R}^{1 \times r_0}$  and a set of univariate functional matrices  $\mathbf{G}_i(\eta_i) \in \mathbb{R}^{r_{i-1} \times r_i}$  for  $i = 1, \dots, d$  and with  $r_d = 1$ , such that

$$(\mathbf{A}\boldsymbol{\eta} + \boldsymbol{\mu})^\alpha = \mathbf{G}_0 \mathbf{G}_1(\eta_1) \mathbf{G}_2(\eta_2) \dots \mathbf{G}_d(\eta_d). \quad (15)$$

As a result, we have the following cheap computation

$$q_\alpha = \mathbf{G}_0 \mathbb{E}[\mathbf{G}_1(\eta_1)] \mathbb{E}[\mathbf{G}_2(\eta_2)] \dots \mathbb{E}[\mathbf{G}_d(\eta_d)].$$

#### 3.3.1 Formula for (15) with $|\alpha| = 1$

Recall that from  $\boldsymbol{\xi} = \mathbf{A}\boldsymbol{\eta} + \boldsymbol{\mu}$ , we have

$$\xi_j = a_{j1}\eta_1 + a_{j2}\eta_2 + \dots + a_{jd}\eta_d + \mu_j, \forall j = 1, \dots, d. \quad (16)$$

Here  $a_{kj}$  denotes the  $(k, j)$ -th element of  $\mathbf{A}$ .

**Theorem 3.1** [Theorem 2, [37]] *For any function written as the summation of univariate functions:*

$$f(x_0, \dots, x_d) = \omega_0(x_0) + \dots + \omega_d(x_d),$$

it holds that

$$\begin{aligned} f(x_0, x_1, \dots, x_d) &= \begin{pmatrix} \omega_0(x_0) & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ \omega_1(x_1) & 1 \end{pmatrix} \\ &\dots \begin{pmatrix} 1 & 0 \\ \omega_{d-1}(x_{d-1}) & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ \omega_d(x_d) & 1 \end{pmatrix}. \end{aligned}$$

Applying Theorem 3.1 to (16), we can derive a functional tensor train decomposition for  $\xi_j$ :

$$\xi_j = (\mu_j \ 1) \begin{pmatrix} 1 & 0 \\ a_{j1}\eta_1 & 1 \end{pmatrix} \dots \begin{pmatrix} 1 & 0 \\ a_{j(d-1)}\eta_{d-1} & 1 \end{pmatrix} \begin{pmatrix} 1 \\ a_{jd}\eta_d \end{pmatrix}. \quad (17)$$

Then the expectation equals to

$$\mathbb{E}[\xi_j] = (\mu_j \ 1) \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \dots \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \mu_j. \quad (18)$$

---

**Algorithm 1:** A functional tensor-train method for computing the basis functions of Gaussian mixtures

---

**Input:** The mean value  $\boldsymbol{\mu}_i$ , covariance  $\boldsymbol{\Sigma}_i$  and weight  $w_i$  for Gaussian-mixtures, and the order  $p$ .

**for**  $i = 1, \dots, n$  **do**

Calculate the Cholesky factor  $\mathbf{A}$  via  $\boldsymbol{\Sigma}_i = \mathbf{A}\mathbf{A}^T$ ;  
 Calculate the functional tensor trains for the first-order and high-order monomials via (17) and (19), respectively;  
 Obtain the moments via (18) and (20).

Assemble the multivariate moment matrix  $\mathbf{M}$  in (8);  
 Compute the basis functions via (9).

**Output:** The multivariate basis functions  $\{\Psi_\alpha(\boldsymbol{\xi})\}$ .

---

The obtained functional tensor trains (17) can be reused to compute high-order moments.

#### 3.3.2 Recurrence Formula for $1 < |\alpha| \leq 2p$

For each  $\alpha$  with  $1 < |\alpha| \leq 2p$ , there exist  $\alpha_1$  and  $\alpha_2$  with  $|\alpha_1|, |\alpha_2| \leq p$ , such that

$$\boldsymbol{\xi}^\alpha = \boldsymbol{\xi}^{\alpha_1} \cdot \boldsymbol{\xi}^{\alpha_2}, \text{ where } \alpha = \alpha_1 + \alpha_2.$$

According to (6), the tensor-train representation of  $\boldsymbol{\xi}^\alpha$  can be obtained as the Hadamard product of the tensor trains of  $\boldsymbol{\xi}^{\alpha_1}$  and  $\boldsymbol{\xi}^{\alpha_2}$ . Suppose that  $\boldsymbol{\xi}^{\alpha_1} = \mathbf{E}_0(\boldsymbol{\mu})\mathbf{E}_1(\eta_1) \dots \mathbf{E}_d(\eta_d)$  and  $\boldsymbol{\xi}^{\alpha_2} = \mathbf{F}_0(\boldsymbol{\mu})\mathbf{F}_1(\eta_1) \dots \mathbf{F}_d(\eta_d)$ , then

$$\boldsymbol{\xi}^\alpha = \mathbf{G}_0(\boldsymbol{\mu})\mathbf{G}_1(\eta_1) \dots \mathbf{G}_d(\eta_d), \text{ with } \mathbf{G}_i = \mathbf{E}_i \otimes \mathbf{F}_i, \forall i = 0, \dots, d. \quad (19)$$

Because  $\eta_i$ 's are mutually independent, finally we have

$$\mathbb{E}[\boldsymbol{\xi}^\alpha] = \mathbf{G}_0(\boldsymbol{\mu})\mathbb{E}[\mathbf{G}_1(\eta_1)] \dots \mathbb{E}[\mathbf{G}_d(\eta_d)]. \quad (20)$$

The moments therefore can be easily computed via small-size matrix products.

The basis construction framework is summarized in Alg. 1

## 4. AN ADAPTIVE SPARSE SOLVER

With the proposed basis functions  $\{\Psi_\alpha(\boldsymbol{\xi})\}$  for non-Gaussian correlated cases, now we proceed to compute  $\{c_\alpha\}$  and express  $y(\boldsymbol{\xi})$  as the form in (1). The standard stochastic spectral methods [5–7] cannot be directly applied, therefore we improve a sparse regression method via adaptive sampling.

Again, we resort all basis functions and their weights  $c_\alpha$  based on the graded lexicographic order, and denote them as  $\Psi_j(\boldsymbol{\xi})$  and  $c_j$  for  $j = 1, 2, \dots, N$ . Given  $M$  pairs of parameter samples and simulation samples  $\{\boldsymbol{\xi}_i, y(\boldsymbol{\xi}_i)\}$  for  $i = 1, 2, \dots, M$ , we have a linear equation system

$$\Phi \mathbf{c} = \mathbf{y}, \text{ with } \Phi_{ij} = \Psi_j(\boldsymbol{\xi}_i), y_i = y(\boldsymbol{\xi}_i), \quad (21)$$

---

**Algorithm 2:** An adaptive sparse solver

---

**Input:** Input a set of candidate samples  $\Xi_0$  and basis functions  $\Psi_1, \dots, \Psi_N$ .

Choose the initial sample set  $\Xi \subset \Xi_0$  via the rank-revealing QR factorization, with  $|\Xi| \ll N$ .

Call the simulator to calculate  $y(\xi)$  for all  $\xi \in \Xi$ .

**for** Outer iteration  $T = 1, 2, \dots$  **do**

    Solve the  $\ell_0$  minimization problem (22) to obtain  $\mathbf{c}$ , such that the sparsity  $s$  is less than  $|\Xi|$ .

**for** Inner iteration  $t = 1, 2, \dots, t_{\max}$  **do**

        Fix the indices of the nonzero elements in  $\mathbf{c}$ ;

        Choose a new sample  $\xi$  from  $\Xi_0 \setminus \Xi$  by formula

        (27), and update the sample set  $\Xi = \Xi \cup \{\xi\}$ ;

        Call the simulator to get  $y(\xi)$  at the sample  $\xi$ ;

        Update the nonzero elements of  $\mathbf{c}$  via (28).

**if** the stopping criterion is satisfied **then**

            Stop

**Output:** The basis function coefficient  $\mathbf{c}$ .

---

where  $\Phi \in \mathbb{R}^{M \times N}$  store the value of  $N$  basis functions at  $M$  parameter samples. In practice, computing each output sample  $y(\xi_i)$  requires calling a computationally expensive device or circuit level simulator. Consequently, it is highly desirable to solve (21) when  $M \ll N$ .

### 4.1 Why Do Sparse Solvers Work?

When  $M \ll N$ , there are infinitely many solutions to (21). A popular method to overcome this issue is to seek for a sparse solution by solving the  $\ell_0$ -minimization problem

$$\min_{\mathbf{c} \in \mathbb{R}^N} \|\mathbf{c}\|_0 \quad \text{s.t.} \quad \Phi \mathbf{c} = \mathbf{y}, \quad (22)$$

where  $\|\mathbf{c}\|_0$  denotes the number of nonzero elements.

The success of an  $\ell_0$ -minimization relies on some conditions. Firstly, the solution  $\mathbf{c}$  should be really sparse. This is generally true for high-dimensional uncertainty quantification problems. Secondly, the exact recovery of a sparse  $\mathbf{c}$  requires the matrix  $\Phi$  to have the restricted isometry property [38]: there exists a positive value  $\delta_s$  such that

$$(1 - \delta_s) \|\mathbf{c}\|_2^2 \leq \|\Phi \mathbf{c}\|_2^2 \leq (1 + \delta_s) \|\mathbf{c}\|_2^2 \quad (23)$$

holds for any  $\|\mathbf{c}\|_0 \leq s$ . Here,  $\|\cdot\|_2$  is the Euclidean norm. Intuitively, this requires that all columns of  $\Phi$  are nearly orthogonal to each other. If the  $M$  samples are chosen randomly, then for any  $k \neq j$ , the inner product of the  $k$ -th and  $j$ -th columns of the matrix  $\Phi$  is

$$\frac{1}{M} \sum_{i=1}^M (\Psi_k(\xi_i) \Psi_j(\xi_i)) \approx \mathbb{E}[\Psi_k(\xi) \Psi_j(\xi)] = 0. \quad (24)$$

In other words, the restricted isometry property holds with a high probability due to the orthonormal property of our basis functions. Consequently, the formulation (22) can provide an accurate solution with a high probability.

Once the above conditions hold, (22) can be solved via various numerical solvers. We employ COSAMP [39], because it can significantly enhance the sparsity of  $\mathbf{c}$ .

### 4.2 Improvement via Adaptive Sampling

Previous publications mainly focused on how to solve a under-determined linear system (21), and the system equations were set up by simply using random samples [17]. We note that some random samples are informative, yet others are not. Therefore, the performance of a sparse solver can be improved if we select a subset of “important” samples to set up the linear equation. Here we present an adaptive sampling method: it uses a rank-revealing QR decomposition

to pick “important” initial samples, and uses a D-optimal criteria to select subsequent samples.

**Initial Sampling.** Given a pool of randomly generated candidate samples for  $\xi$ , we first select a small number of initial samples from this pool. By evaluating all basis functions on the candidate samples, we form a matrix  $\Phi$  whose  $j$ -th rows stores the values of  $N$  basis functions at the  $j$ -th candidate sample. In order to choose the  $r$  most informative rows, we perform a rank-revealing QR factorization [40]:

$$\Phi^T \mathbf{P} = \mathbf{Q} \begin{bmatrix} \mathbf{R}_{11} & \mathbf{R}_{12} \\ \mathbf{0} & \mathbf{R}_{22} \end{bmatrix}. \quad (25)$$

Here  $\mathbf{P}$  is a permutation matrix;  $\mathbf{Q}$  is an orthogonal matrix;  $\mathbf{R}_{11} \in \mathbb{R}^{r \times r}$  is a upper triangular matrix with nonnegative diagonal elements. Usually,  $\mathbf{P}$  is chosen such that the minimal singular value of  $\mathbf{R}_{11}$  is large, and the maximal singular value of  $\mathbf{R}_{22}$  is sufficiently small. Consequently, the first  $r$  columns of  $\mathbf{P}$  indicate the most informative  $r$  rows of  $\Phi$ . We keep these rows and the associated parameter samples.

**Adding New Samples and Solution Update.** The above initial sampling has generated a  $r$ -by- $N$  matrix  $\Phi$ . Now we further add new informative samples. Our idea is motivated by the D-optimal sampling update in [41], but differs in the numerical implementation.

Assume that our sparse solver has computed  $\mathbf{c}$  based on the available samples, then we fix the indices of the nonzero elements in  $\mathbf{c}$  and update the samples and solution sequentially. Specifically, denote the locations of nonzero coefficients as  $\mathcal{S} = \{i_1, \dots, i_s\}$  with  $r > s$ , and denote  $\Phi_1 \in \mathbb{R}^{r \times s}$  as the sub-matrix of  $\Phi$  generated by extracting the columns of  $\mathcal{S}$ . The next most informative sample  $\xi$  associated with the row vector  $\mathbf{x} = [\Psi_{i_1}(\xi) \dots, \Psi_{i_s}(\xi)]$  can be found via solving the following problem:

$$\max_{\mathbf{x} \in \Omega} \det(\Phi_1^T \Phi_1 + \mathbf{x}^T \mathbf{x}), \quad (26)$$

where  $\Omega$  is value of basis functions in  $\mathcal{S}$  for all candidate samples. In practice, we do not need to compute the above determinant for each sample. Instead, the matrix determinant lemma [42] shows  $\det(\Phi_1^T \Phi_1 + \mathbf{x}^T \mathbf{x}) = (\det(\Phi_1^T \Phi_1))(1 + \mathbf{x}(\Phi_1^T \Phi_1)^{-1} \mathbf{x}^T)$ , therefore, (26) can be solved via

$$\max_{\mathbf{x} \in \Omega} \mathbf{x}(\Phi_1^T \Phi_1)^{-1} \mathbf{x}^T. \quad (27)$$

After getting the new sample, we update the matrix  $\Phi_1 := \begin{bmatrix} \Phi_1 \\ \mathbf{x} \end{bmatrix}$ , update  $(\Phi_1^T \Phi_1)^{-1}$  via the Sherman-Morrison formula [43], and recompute the  $s$  nonzero elements of  $\mathbf{c}$  by

$$\mathbf{c}_1 = (\Phi_1^T \Phi_1)^{-1} \Phi_1^T \mathbf{y}. \quad (28)$$

Inspired by [44], we stop the procedure if  $\mathbf{c}_1$  is very close to the value of the previous step or the maximal iteration number is reached. The whole framework is summarized in Alg. 2.

## 5. NUMERICAL RESULTS

In this section, we validate our algorithms by two real-world examples: a photonic bandpass filter and 7-stage CMOS ring oscillator. All codes are implemented in MATLAB and run on a desktop with a 3.40-GHz CPU and a 8-GB memory. For each example, we adaptively select a small number of samples from a pool of 1000 candidate samples, and we use 9000 different samples for accuracy validation. Given random samples, we define the relative error based on (21):

$$\epsilon = \|\Phi \mathbf{c} - \mathbf{y}\|_2 / \|\mathbf{y}\|_2. \quad (29)$$

We call  $\epsilon$  as a testing error if the samples are those used in our sparse solver, and as a training error if the new set of 9000 samples are used to verify the predictability.

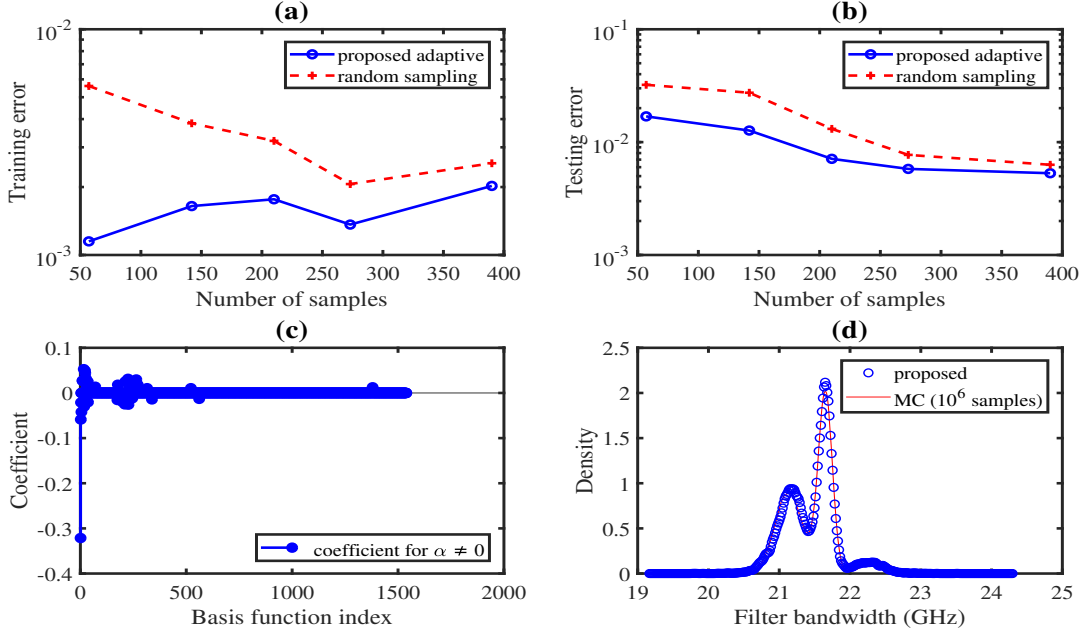


Figure 3: Results for the phonic bandpass filter. (a) training error on 390 samples; (b) testing error on 9000 new samples; (c) calculated coefficients/weights of our proposed basis functions; (d) probability density functions of the 3-dB bandwidth obtained with our proposed method and Monte Carlo (MC), respectively.

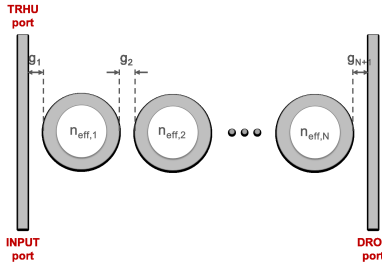


Figure 4: A bandpass filter with 9 ring resonators.

Table 1: Accuracy comparison on the photonic bandpass filter. The underscores indicate precision.

method	Proposed	Monte Carlo		
# samples	390	$10^2$	$10^4$	$10^6$
mean (GHz)	21.4717	21. <u>5297</u>	21. <u>4867</u>	21.4782

### 5.1 Photonic Bandpass Filter (19 Parameters)

Firstly we consider the photonic bandpass filter in Fig. 4. This photonic IC has 9 ring resonators, and it was originally designed to have a 3-dB bandwidth of 20 GHz, a 400-GHz free spectral range, and a  $1.55\text{-}\mu\text{m}$  operation wavelength. A total of 19 random parameters are used to describe the variations of the effective phase index ( $n_{\text{eff}}$ ) of each ring, as well as the gap ( $g$ ) between adjacent rings and between the first/last ring and the bus waveguides. These non-Gaussian correlated random parameters are described by a Gaussian-mixture joint probability density function.

We approximate the 3-dB bandwidth  $f_{3\text{dB}}$  at the DROP port using our proposed basis functions with a total order bounded by  $p = 3$ . The numerical results are shown in Fig. 3. Fig. 3 (b) clearly shows that the adaptive sampling

method leads to significantly lower testing (i.e., prediction) errors when a few samples are used, because it chooses important samples. Finally, we use 390 samples to assemble a linear system and solve it by  $\ell_0$  minimization, and obtain the sparse coefficients of our basis functions in Fig. 3 (c). Although a third-order expansion involves more than 1000 basis functions, only a few dozens are important. Fig. 3 (d) shows the predicted probability density function of the filter's 3-dB bandwidth, and it matches well with the result from Monte Carlo. More importantly, it is clear that our algorithm can capture accurately the multiple peaks in the output density function, and these peaks can be hardly predicted using existing stochastic spectral methods.

In order to demonstrate the effectiveness of our method in detail, we compare the computed mean value of  $f_{3\text{dB}}$  from our methods and from Monte Carlo in Table 1. Our method provides a closed-form expression for the mean value. Monte Carlo method converges very slowly, and requires  $2564 \times$  more simulation samples to achieve the similar level of accuracy (with 2 accurate fractional digits).

### 5.2 CMOS Ring Oscillator (57 Parameters)

We continue to consider the 7-stage CMOS ring oscillator in Fig. 6. This circuit has 57 random parameters describing the variations of threshold voltages, gate-oxide thickness, and effective gate length/width. We use Gaussian mixtures to describe the strong non-Gaussian correlations.

We use a 2nd-order expansion of our basis functions to model the oscillator frequency. The simulation samples are obtained by calling a periodic steady-state simulator repeatedly. The detailed results are shown in Fig. 5. Similar to the previous example, our adaptive sparse solver produces a sparse and highly accurate stochastic solution with better prediction behaviors than the standard compressive sensing does. The proposed basis functions can well capture the multiple peaks of the output probability density function caused by the strong non-Gaussian correlation.

Table 2 compares our method with Monte Carlo. Our

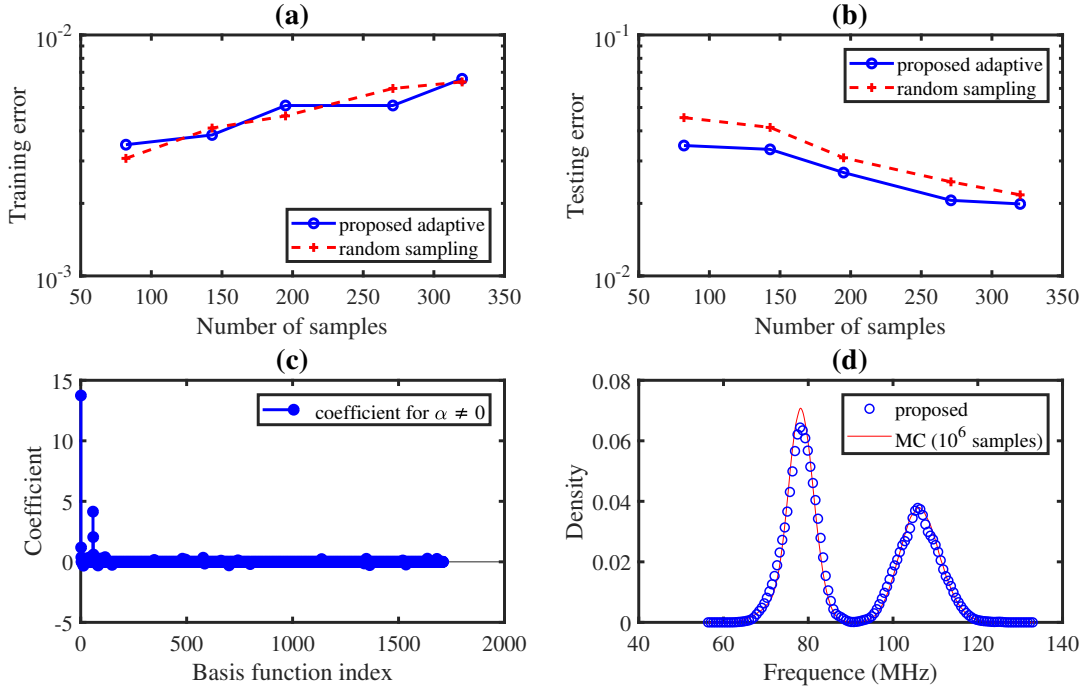


Figure 5: Results for the CMOS ring oscillator. (a) training error on 320 samples; (b) testing error on 9000 new samples; (c) calculated coefficients/weights of our proposed basis functions; (d) probability density functions of the oscillator frequency obtained with our proposed method and Monte Carlo (MC), respectively.

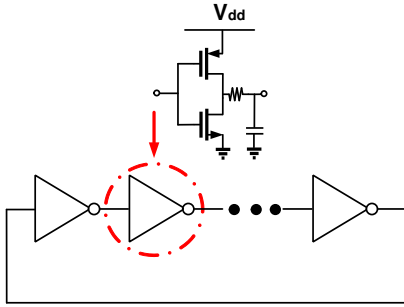


Figure 6: Schematic of a CMOS ring oscillator.

Table 2: Accuracy comparison on the CMOS ring oscillator. The underscores indicate precision.

method	Proposed	Monte Carlo		
# samples	320	$10^2$	$10^4$	$10^6$
mean (MHz)	90.5441	89.7795	<u>90.4945</u>	<u>90.5253</u>

method is about  $3125\times$  faster than Monte Carlo to achieve a precision of one fractional digit for the mean value.

## 6. CONCLUSION

This paper has presented some theories and algorithms for the fast uncertainty quantification of electronic and photonic ICs with non-Gaussian correlated process variations. We have proposed a set of basis functions for non-Gaussian correlated cases. We have also presented a functional tensor train method to efficiently compute the high-dimensional basis functions. In order to reduce the computational time

of analyzing process variations, we have proposed an adaptive sampling sparse solver, and this algorithm only uses a small number of important simulation samples to predict the uncertain output. The proposed approach has been verified with two electronic and photonic ICs. On these benchmarks, our method has achieved very high accuracy in predicting the multi-peak output probability density function and output mean value. Our method has achieved  $2500\times$  to  $3100\times$  speedup over Monte Carlo to achieve the similar level of accuracy. To our best knowledge, this is the first non-Monte-Carlo uncertainty quantification approach that can handle non-Gaussian correlated process variations without any error-prone de-correlation steps. Many novel algorithms can be further developed based on our results.

## Acknowledgment

This work was supported by NSF-CCF Award No. 1763699, the UCSB start-up grant and a Samsung Gift Funding.

## 7. REFERENCES

- [1] D. S. Boning, "Variation," *IEEE Trans. Semiconductor Manufacturing*, vol. 21, no. 1, pp. 63–71, Feb 2008.
- [2] N. Agarwal and N. R. Aluru, "Stochastic analysis of electrostatic MEMS subjected to parameter variations," *Journal of Microelectromechanical Systems*, vol. 18, no. 6, pp. 1454–1468, 2009.
- [3] W. A. Zortman, D. C. Trotter, and M. R. Watts, "Silicon photonics manufacturing," *Optics express*, vol. 18, no. 23, pp. 23 598–23 607, 2010.
- [4] S. Weinzierl, "Introduction to Monte Carlo methods," NIKHEF, Theory Group, The Netherlands, Tech. Rep. NIKHEF-00-012, 2000.
- [5] R. G. Ghanem and P. D. Spanos, "Stochastic finite element method: Response statistics," in *Stochastic finite elements: a spectral approach*. Springer, 1991, pp. 101–119.
- [6] Z. Zhang, T. A. El-Moselhy, I. A. M. Elfadel, and L. Daniel, "Stochastic testing method for transistor-level

- uncertainty quantification based on generalized polynomial chaos,” *IEEE Trans. CAD Integrated Circuits Syst.*, vol. 32, no. 10, Oct. 2013.
- [7] D. Xiu and J. S. Hesthaven, “High-order collocation methods for differential equations with random inputs,” *SIAM J. Sci. Comp.*, vol. 27, no. 3, pp. 1118–1139, 2005.
- [8] D. Xiu and G. E. Karniadakis, “The Wiener-Askey polynomial chaos for stochastic differential equations,” *SIAM J. Sci. Comp.*, vol. 24, no. 2, pp. 619–644, Feb 2002.
- [9] T. Moselhy and L. Daniel, “Stochastic integral equation solver for efficient variation aware interconnect extraction,” in *Proc. Design Auto. Conf.*, Jun. 2008, pp. 415–420.
- [10] J. Wang, P. Ghanta, and S. Vrudhula, “Stochastic analysis of interconnect performance in the presence of process variations,” in *Proc. DAC*, 2004, pp. 880–886.
- [11] R. Shen, S. X.-D. Tan, J. Cui, W. Yu, Y. Cai, and G. Chen, “Variational capacitance extraction and modeling based on orthogonal polynomial method,” *IEEE Trans. VLSI*, vol. 18, no. 11, pp. 1556–1565, Nov. 2010.
- [12] D. V. Ginste, D. D. Zutter, D. Deschrijver, T. Dhaene, P. Manfredi, and F. Canavero, “Stochastic modeling-based variability analysis of on-chip interconnects,” *IEEE Trans. Comp. Pack. Manufact. Tech.*, vol. 2, no. 7, pp. 1182–1192, Jul. 2012.
- [13] K. Strunz and Q. Su, “Stochastic formulation of SPICE-type electronic circuit simulation with polynomial chaos,” *ACM Trans. Modeling and Computer Simulation*, vol. 18, no. 4, pp. 15:1–15:23, Sep 2008.
- [14] J. Tao, X. Zeng, W. Cai, Y. Su, D. Zhou, and C. Chiang, “Stochastic sparse-grid collocation algorithm (SSCA) for periodic steady-state analysis of nonlinear system with process variations,” in *Proc. Asia and South Pacific Design Automation Conference*, 2007, pp. 474–479.
- [15] Z. Zhang, X. Yang, G. Marucci, P. Maffezzoni, I. M. Elfadel, G. Karniadakis, and L. Daniel, “Stochastic testing simulator for integrated circuits and MEMS: Hierarchical and sparse techniques,” in *Proc. Custom Integrated Circuit Conf. CA*, Sept. 2014, pp. 1–8.
- [16] T.-W. Weng, Z. Zhang, Z. Su, Y. Marzouk, A. Melloni, and L. Daniel, “Uncertainty quantification of silicon photonic devices with correlated and non-Gaussian random parameters,” *Optics Express*, vol. 23, no. 4, pp. 4242–4254, Feb 2015.
- [17] X. Li, “Finding deterministic solution from underdetermined equation: large-scale performance modeling of analog/RF circuits,” *IEEE Trans. CAD*, vol. 29, no. 11, pp. 1661–1668, Nov 2011.
- [18] J. Hampton and A. Doostan, “Compressive sampling of polynomial chaos expansions: Convergence analysis and sampling strategies,” *Journal of Computational Physics*, vol. 280, pp. 363–386, 2015.
- [19] X. Ma and N. Zabaras, “An adaptive high-dimensional stochastic model representation technique for the solution of stochastic partial differential equations,” *Journal Comp. Physics*, vol. 229, no. 10, pp. 3884–3915, 2010.
- [20] T. El-Moselhy and L. Daniel, “Variation-aware interconnect extraction using statistical moment preserving model order reduction,” in *Proc. DATE*, 2010, pp. 453–458.
- [21] Z. Zhang, T. A. El-Moselhy, I. M. Elfadel, and L. Daniel, “Calculation of generalized polynomial-chaos basis functions and Gauss quadrature rules in hierarchical uncertainty quantification,” *IEEE Trans. CAD of Integrated Circuits and Systems*, vol. 33, no. 5, pp. 728–740, 2014.
- [22] Z. Zhang, X. Yang, I. V. Oseledets, G. E. Karniadakis, and L. Daniel, “Enabling high-dimensional hierarchical uncertainty quantification by anova and tensor-train decomposition,” *IEEE Trans. CAD of Integrated Circuits and Systems*, vol. 34, no. 1, pp. 63–76, 2015.
- [23] Z. Zhang, T.-W. Weng, and L. Daniel, “Big-data tensor recovery for high-dimensional uncertainty quantification of process variations,” *IEEE Trans. Components, Packaging and Manufacturing Tech.*, vol. 7, no. 5, pp. 687–697, 2017.
- [24] Z. Zhang, K. Batselier, H. Liu, L. Daniel, and N. Wong, “Tensor computation: A new framework for high-dimensional problems in eda,” *IEEE Trans CAD Integr. Circuits Syst.*, vol. 36, no. 4, pp. 521–536, 2017.
- [25] S. Wold, K. Esbensen, and P. Geladi, “Principal component analysis,” *Chemometrics and intelligent laboratory systems*, vol. 2, no. 1-3, pp. 37–52, 1987.
- [26] J. Singh and S. Sapatnekar, “Statistical timing analysis with correlated non-gaussian parameters using independent component analysis,” in *Proc. DAC*, 2006, pp. 155–160.
- [27] C. Soize and R. Ghanem, “Physical systems with random uncertainties: chaos representations with arbitrary probability measure,” *SIAM Journal on Scientific Computing*, vol. 26, no. 2, pp. 395–410, 2004.
- [28] W. Gautschi, “On generating orthogonal polynomials,” *SIAM J. Sci. Stat. Comput.*, vol. 3, no. 3, pp. 289–317, Sept. 1982.
- [29] T. G. Kolda and B. W. Bader, “Tensor decompositions and applications,” *SIAM Rev.*, vol. 51, no. 3, pp. 455–500, 2009.
- [30] I. V. Oseledets, “Tensor-train decomposition,” *SIAM Journal Sci. Comp.*, vol. 33, no. 5, pp. 2295–2317, 2011.
- [31] M. E. Ismail and R. Zhang, “A review of multivariate orthogonal polynomials,” *Journal of the Egyptian Mathematical Society*, vol. 25, no. 2, pp. 91–110, 2017.
- [32] Y. Xu, “On multivariate orthogonal polynomials,” *SIAM Journal Math. Analysis*, vol. 24, no. 3, pp. 783–794, 1993.
- [33] R. Barrio, J. M. Pena, and T. Sauer, “Three term recurrence for the evaluation of multivariate orthogonal polynomials,” *Journal of Approximation Theory*, vol. 162, no. 2, pp. 407–420, 2010.
- [34] G. H. Golub and J. H. Welsch, “Calculation of gauss quadrature rules,” *Math. Comp.*, vol. 23, pp. 221–230, 1969.
- [35] M. Rosenblatt, “Remarks on a multivariate transformation,” *The annals of mathematical statistics*, vol. 23, no. 3, pp. 470–472, 1952.
- [36] T. Gerstner and M. Griebel, “Numerical integration using sparse grids,” *Numer. Algor.*, vol. 18, pp. 209–232, 1998.
- [37] I. Oseledets, “Constructive representation of functions in low-rank tensor formats,” *Constructive Approximation*, vol. 37, no. 1, pp. 1–18, 2013.
- [38] E. Candes, J. Romberg, and T. Tao, “Stable signal recovery from incomplete and inaccurate measurements,” *Comm. Pure Appl. Math.*, vol. 59, no. 8, pp. 1207–1223, 2006.
- [39] D. Needell and J. A. Tropp, “CoSaMP: Iterative signal recovery from incomplete and inaccurate samples,” *Appl. Comp. Harm. Analysis*, vol. 26, no. 3, pp. 301–321, 2009.
- [40] M. Gu and S. C. Eisenstat, “Efficient algorithms for computing a strong rank-revealing QR factorization,” *SIAM J. Sci. Comp.*, vol. 17, no. 4, pp. 848–869, 1996.
- [41] P. Diaz, A. Doostan, and J. Hampton, “Sparse polynomial chaos expansions via compressed sensing and  $D$ -optimal design,” *arXiv preprint arXiv:1712.10131*, 2017.
- [42] D. A. Harville, *Matrix algebra from a statistician’s perspective*. Springer, 1997, vol. 1.
- [43] J. Sherman and W. J. Morrison, “Adjustment of an inverse matrix corresponding to a change in one element of a given matrix,” *The Annals of Mathematical Statistics*, vol. 21, no. 1, pp. 124–127, 1950.
- [44] D. M. Malioutov, S. R. Sanghavi, and A. S. Willsky, “Sequential compressed sensing,” *IEEE J. Selected Topics in Signal Processing*, vol. 4, no. 2, pp. 435–444, 2010.