High-Dimensional Uncertainty Quantification of Electronic and Photonic IC with Non-Gaussian Correlated Process Variations

Chunfeng Cui and Zheng Zhang, *Member, IEEE*

Abstract—Uncertainty quantification based on generalized polynomial chaos has been used in many applications. It has also achieved great success in variation-aware design automation. However, almost all existing techniques assume that the parameters are mutually independent or Gaussian correlated, which is rarely true in real applications. For instance, in chip manufacturing, many process variations are actually correlated. Recently, some techniques have been developed to handle non-Gaussian correlated random parameters, but they are time-consuming for high-dimensional problems. We present a new framework to solve uncertainty quantification problems with many non-Gaussian correlated uncertainties. Firstly, we propose a set of smooth basis functions to well capture the impact of non-Gaussian correlated process variations. We develop a tensor approach to compute these basis functions in a high-dimension setting. Secondly, we investigate the theoretical aspect and practical implementation of a sparse solver to compute the coefficients of all basis functions. We provide some theoretical analysis for the exact recovery condition and error bound of this sparse solver in the context of uncertainty quantification. We present three adaptive sampling approaches to improve the performance of the sparse solver. Finally, we validate our methods by synthetic and practical electronic/photonic ICs with 19 to 57 non-Gaussian correlated variation parameters. Our approach outperforms Monte Carlo by thousands of times in terms of efficiency. It can also accurately predict the output density functions with multiple peaks caused by non-Gaussian correlations, which are hard to capture by existing methods.

Index Terms—High dimensionality, uncertainty quantification, electronic and photonic IC, non-Gaussian correlation, process variations, tensor, sparse solver, adaptive sampling.

I. INTRODUCTION

U NCERTAINTIES are unavoidable in almost all engineering fields. In semiconductor chip design, a major source of uncertainty is the fabrication process variations. For instance, in deeply scaled electronic integrated circuits (ICs) [2] and MEMS [3], process variations have become a major concern in emerging design technologies such as integrated photonics [4]. One of the traditional uncertainty quantification methods is Monte Carlo [5], which is easy to implement but has a low convergence rate. In recent years, various stochastic spectral methods (e.g., stochastic Galerkin [6], stochastic testing [7] and stochastic collocation [8]) have been developed and have achieved orders-of-magnitude speedup compared with Monte Carlo in vast applications, including (but not limited to) the modeling and simulation of VLSI interconnects [9]–[14], nonlinear ICs [7], [15]–[20], MEMS [3], [21], photonic circuits [22]–[24], and computer architecture [25].

The key idea of stochastic spectral method is to represent the stochastic solution as the linear combination of basis

functions. It can obtain highly accurate solutions at a low computational cost when the parameter dimensionality is not high (e.g., less than 20). Despite their great success, stochastic spectral methods are limited by a long-standing challenge: the generalized polynomial chaos basis functions assume that all random parameters are mutually independent [26]. This assumption fails in many realistic cases. For instance, a lot of device-level 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 [27], [28] can handle Gaussian correlations, but they cause huge errors in general *non-Gaussian correlated* cases. Soize and Ghanem [29] proposed to modify the basis functions to a non-smooth chaos formulation, which was applied to the uncertainty analysis of silicon photonics [22]. It was found that the method in [29] does not converge well, and designers cannot easily extract mean value and variance from the solution. Recently, we proposed a novel approach to handle non-Gaussian correlated process variations in [30], [31]. We constructed the basis functions via a Gram-Schmidt formula, and then built the surrogate model via an optimization-based stochastic collocation approach. Our basis functions inherit three important properties of the independent counterparts: smoothness, orthonormality, and the capability of providing closed-form mean value and variance of a stochastic solution. In [31], some theoretical results about the numerical error and complexity were provided, and thousands of times of speedup than Monte Carlo were achieved in electronic and photonic ICs with a few non-Gaussian correlated random parameters. A later paper [32] presented a similar method to solve the same kind of problems. However, how to handle high-dimensional non-Gaussian correlated uncertain parameters remains an open question, despite significant progress in high-dimensional uncertainty quantification with independent random parameters [21], [33]–[40].

Contributions. This paper presents a framework to quantify the uncertainties of electronic and photonic ICs with *high-dimensional and non-Gaussian correlated* process variations. Our approach has two excellent features: it efficiently computes high-dimensional basis functions that well capture the impact of non-Gaussian correlated uncertainties; it can also automatically choose informative parameter samples to reduce the numerical simulation cost in a high-dimension setting. The specific contributions of this paper include:

- We derive a class of basis functions for non-Gaussian correlated random parameters. Our method is based on the Cholesky factorization and can overcome the theoretical limitations of [29]. For high-dimensional problems that cannot be handled by [30], [31], we construct the basis functions via a functional tensor train method when the

Chufeng Cui and Zheng Zhang are with the Department of Electrical and Computer Engineering, University of California, Santa Barbara, CA 93106, USA (e-mail: chunfengcui@ucsb.edu, zhengzhang@ece.ucsb.edu).
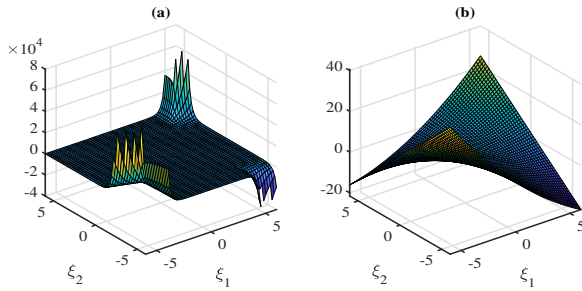
Fig. 1. (a): A two-variable basis function by [29]; (b): a basis function obtained by our proposed method.

random parameters are equipped with a Gaussian mixture density function. We also present a theoretical analysis about the expressive power of our basis functions;
- In order to apply our method to high-dimensional problems, we investigate the theoretical aspect and implementation of sparse solver with ℓ_0-minimization. Our contributions are twofold. Firstly, we provide the theoretical justification for this ℓ_0-minimization and an error bound for the resulting surrogate model in the context of uncertainty quantification. Secondly, we improve its performance by adaptive sampling. Instead of using random simulation samples (as done in [33]), we select the most informative samples via a rank-revealing QR factorization and adaptive optimal sampling criteria including the D-optimal, R-optimal, and E-optimal.

Compared with our conference paper [1], this extended journal manuscript presents the following additional results:
- We prove that our basis functions are complete in the polynomial subspace, and that our expression is able to approximate any square-integrable function;
- We show the theoretical conditions to obtain an accurate sparse stochastic approximation and the error bounds of the resulting sparse stochastic surrogate model;
- We proposed two additional approaches, i.e., R-optimal and E-optimal, to select informative samples;
- We add more examples to verify the theoretical properties and performance of our proposed approach.

II. PRELIMINARY

A. Generalized Polynomial Chaos

Let $\boldsymbol{\xi} = [\xi_1, \ldots, \xi_d] \in \mathbb{R}^d$ be d random parameters describing the process variations. Assume that $\boldsymbol{\xi}$ has a joint probability density function $\rho(\boldsymbol{\xi})$, and let $y(\boldsymbol{\xi}) \in \mathbb{R}$ be an uncertain performance metric (e.g., the bandwidth or frequency of a circuit). Suppose $y(\boldsymbol{\xi})$ is smooth and has a bounded variance. Stochastic spectral method approximates $y(\boldsymbol{\xi})$ via a truncated generalized polynomial chaos expansion [26]:

$$y(\boldsymbol{\xi}) \approx \sum_{|\boldsymbol{\alpha}|=0}^{p} c_{\boldsymbol{\alpha}} \Psi_{\boldsymbol{\alpha}}(\boldsymbol{\xi}). \tag{1}$$

Here, $c_{\boldsymbol{\alpha}}$ is the coefficient and $\Psi_{\boldsymbol{\alpha}}(\boldsymbol{\xi})$ is an orthonormal polynomial satisfying

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

The operator $\mathbb{E}$ denotes expectation; $\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_d] \in \mathbb{N}^d$ is a vector with α_i being the highest polynomial order in terms of ξ_i. The total polynomial order $|\boldsymbol{\alpha}| = |\alpha_1| + \ldots + |\alpha_d|$ is

bounded by p, and thus there are $n = (p+d)!/(p!d!)$ basis functions in total.

The generalized polynomial chaos theory [26] assumes that all random parameters are mutually independent. In other words, if $\rho_k(\xi_k)$ denotes the marginal density of ξ_k, the joint density is $\rho(\boldsymbol{\xi}) = \prod_{k=1}^{d} \rho_k(\xi_k)$. Under this assumption, a multivariate basis function has a product form:

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

Here $\phi_{k,\alpha_k}(\xi_k)$ is a univariate degree-α_k orthonormal polynomial of parameter ξ_k, and it is calculated based on $\rho_k(\xi_k)$ via the three-term recurrence relation [41]. The unknown coefficients $c_{\boldsymbol{\alpha}}$ can be computed via various solvers such as stochastic Galerkin [6], stochastic testing [7], and stochastic collocation [8]. Once $c_{\boldsymbol{\alpha}}$ are computed, the mean value, variance and density function of $y(\boldsymbol{\xi})$ can be easily obtained.

However, if the domain is not exactly a tensor product or the parameters are not independent, the above theory cannot be applied directly. This is now an active research topic in both theoretical and application domains.

B. Existing Solutions for Correlated Cases

For general non-Gaussian correlated parameters, the reference [29] suggested the following basis functions:

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

However, the above basis functions have two limitations as shown by the numerical results in [22]:
- The basis functions are highly non-smooth and numerically unstable due to the first term on the right-hand side of (4). This is demonstrated in Fig. 1 (a).
- The basis functions do not allow an explicit expression for the expectation and variance of $y(\boldsymbol{\xi})$. This is because the basis function indexed by $\boldsymbol{\alpha} = 0$ is not a constant.

Recently, we proposed to build a new set of basis functions via a Gram-Schmidt approach [30], [31]. We also suggested to compute the coefficients via a stochastic collocation approach $c_{\boldsymbol{\alpha}} = \mathbb{E}[\Psi_{\boldsymbol{\alpha}}(\boldsymbol{\xi})y(\boldsymbol{\xi})] \approx \sum_k \Psi_{\boldsymbol{\alpha}}(\boldsymbol{\xi}_k)y(\boldsymbol{\xi}_k)w_k$. A quadrature rule based on an optimization model was developed to compute the quadrature points $\boldsymbol{\xi}_k$ and weights w_k, and the number of quadrature points was determined automatically. Our recent technique [30], [31] is highly accurate and efficient for low-dimensional problems, but it suffers from the curse of dimensionality: a huge number of simulation samples will be required if the number of random parameters is large.

C. Background: Tensor Train Decomposition

A tensor $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{n_1 \times n_2 \times \cdots n_d}$ is a d-way data array, which is a high-dimensional generalization of a vector and a matrix. A tensor has $n_1 n_2 \ldots n_d$ elements, leading to a prohibitive computation and storage cost. Fortunately, this challenge may be addressed by tensor decomposition [42]. Among various tensor decomposition approaches, tensor train decomposition [43] is highly suitable for factorizing high-dimensional tensors. Specifically, given a d-way tensor $\boldsymbol{\mathcal{A}}$, the tensor-train decomposition admits a decomposition as

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

where $\mathbf{A}_k(i_k)$ is an $r_{k-1} \times r_k$ matrix, and $r_0 = r_d = 1$.

Given two d-way tensors $\mathcal{A}, \mathcal{B}$ and their corresponding tensor train decomposition factors, the tensor train decomposition of their Hadamard (element-wise) product $\mathcal{C} = \mathcal{A} \circ \mathcal{B}$ has a closed form

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

Here $\otimes$ denotes a matrix Kronecker product.

III. BASIS FUNCTIONS WITH NON-GAUSSIAN CORRELATED UNCERTAINTIES

Assume that the elements of $\boldsymbol{\xi}$ are non-Gaussian correlated. A broad class of non-Gaussian correlated parameters can be described by fitting a Gaussian mixture model based on device/circuit testing data. In this general setting, the basis functions in (3) cannot be employed. Therefore, we will derive a set of multivariate polynomial basis functions, which can be obtained if a multivariate moment computation framework is available. We also show the theoretical completeness and expressive power of our basis functions.

A. Multivariate Basis Functions

Several orthogonal polynomials exist for specific density functions [44]. In general, one may construct multivariate orthogonal polynomials via the three-term recurrence in [45] or [46]. However, their theories either are hard to implement or can only guarantee weak orthogonality [45], [46]. Inspired by [47], we present a simple yet efficient method for computing a set of multivariate orthonormal polynomial basis functions.

Let $\boldsymbol{\xi}^{\boldsymbol{\alpha}} = \xi_1^{\alpha_1} \xi_2^{\alpha_2} \ldots \xi_d^{\alpha_d}$ be a *monomial* indexed by $\boldsymbol{\alpha}$ and the corresponding *moment* be

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

We resort all monomials bounded by order p in the graded lexicographic order, and denote them as

$$\mathbf{b}(\boldsymbol{\xi}) = [b_1(\boldsymbol{\xi}), \ldots, b_n(\boldsymbol{\xi})]^T \in \mathbb{R}^n. \tag{8}$$

Here, $n = \binom{d+p}{d}$. Further, we denote the *multivariate moment matrix* as $\mathbf{M} \in \mathbb{R}^{n \times n}$, where

$$\mathbf{M} = \mathbb{E}\left[\mathbf{b}(\boldsymbol{\xi})\mathbf{b}^T(\boldsymbol{\xi})\right], \text{ with } m_{ij} = \mathbb{E}[b_i(\boldsymbol{\xi})b_j(\boldsymbol{\xi})]. \tag{9}$$

Here, $\mathbf{M}$ is also the Gram matrix of $\mathbf{b}(\boldsymbol{\xi})$. Because the monomials $\mathbf{b}(\boldsymbol{\xi})$ are linearly independent, it hold that $\mathbf{M}$ is positive definite according to [48]. For instance, if $d = 2$ and $p = 1$, the monomials and the multivariate moment matrix are

$$\mathbf{b}(\boldsymbol{\xi}) = [1, \xi_1, \xi_2]^T \text{ and } \mathbf{M} = \begin{bmatrix} 1 & \mathbb{E}[\xi_1] & \mathbb{E}[\xi_2] \\ \mathbb{E}[\xi_1] & \mathbb{E}[\xi_1^2] & \mathbb{E}[\xi_1\xi_2] \\ \mathbb{E}[\xi_2] & \mathbb{E}[\xi_1\xi_2] & \mathbb{E}[\xi_2^2] \end{bmatrix}.$$

We intend to construct n multivariate orthonormal polynomials $\Psi_{\boldsymbol{\alpha}}(\boldsymbol{\xi})$ with their total degrees $|\boldsymbol{\alpha}|$ bounded by p. The monomials $\mathbf{b}(\boldsymbol{\xi})$ already contain n different polynomials, even though they are not orthogonal. The key idea of our method is to orthogonalize $\mathbf{b}(\boldsymbol{\xi})$ via a linear projection. This is fulfilled in three steps. Firstly, we compute all elements in the multivariate moment matrix $\mathbf{M}$. This step involves computing the moments up to order $2p$ because $\mathbf{b}(\boldsymbol{\xi})$ is up to order p. Secondly, we decompose $\mathbf{M}$ via the Cholesky factorization $\mathbf{M} = \mathbf{L}\mathbf{L}^T$, where $\mathbf{L}$ is a lower-triangular matrix. The process is stable when the diagonal elements $\mathbb{E}[\mathbf{b}_i^2(\boldsymbol{\xi})]$ are strictly positive. If $\mathbf{M}$ is close to singular, the Cholesky process maybe unstable. In such situations, we add a small permutation term

by $\mathbf{M} = \mathbf{M} + \varepsilon\mathbf{I}$. Thirdly, we define the orthogonal basis functions as

$$\boldsymbol{\Psi}(\boldsymbol{\xi}) = \mathbf{L}^{-1}\mathbf{b}(\boldsymbol{\xi}). \tag{10}$$

Here the n-by-1 functional vector $\boldsymbol{\Psi}(\boldsymbol{\xi})$ stores all basis functions $\{\Psi_{\boldsymbol{\alpha}}(\boldsymbol{\xi})\}$ in the graded lexicographic order. This factorization approach can reduce the high-dimensional functional inner-product and moment computations in [30], [31].

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

1) The basis functions are smooth (c.f. Fig. 1 (b)). In fact, the basis function $\Psi_{\boldsymbol{\alpha}}(\boldsymbol{\xi})$ is a multivariate polynomial. It differs from the standard generalized polynomial chaos [26] in the sense that our basis functions are not the products of univariate polynomials.
2) All basis functions are orthonormal to each other:

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

This property is important for the sparse approximation and for extracting the statistical information of $y(\boldsymbol{\xi})$.
3) Since $\Psi_{\mathbf{0}}(\boldsymbol{\xi}) = 1$ is a constant, the expectation and variance of $y(\boldsymbol{\xi})$ has a closed-form formula in (11).

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

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

B. Theoretical Analysis

Now we consider the expressive power of our basis functions. This can be described via *completeness*: a family of basis functions is complete in a space if any function in this space can be uniquely expressed as a linear combination of these basis functions. Denote $\mathcal{S}_p$ as the space including all polynomials bounded by order p. We have the following result.

Lemma 1 (Completeness): The basis functions defined via (10) are complete in the space $\mathcal{S}_p$.

Proof: See Appendix A. ∎

Denote $L^2(\boldsymbol{\xi}, \rho(\boldsymbol{\xi})) = \{y(\boldsymbol{\xi}) : \mathbb{E}_{\boldsymbol{\xi}}[y^2(\boldsymbol{\xi})] < \infty\}$ as the space of square-integrable functions. In the following lemma, we show that our basis function can approximate any function in $L^2(\boldsymbol{\xi}, \rho(\boldsymbol{\xi}))$ when the polynomial order p is large enough.

Lemma 2 (Expressive power): Assume that $\boldsymbol{\xi}$ is defined on a compact bounded domain or there exists a constant $a > 0$ such that $\mathbb{E}[\exp(a|\xi_i|)] < \infty$ for $i = 1, \ldots, d$, then

(i) the multivariate polynomials are dense in $L^2(\boldsymbol{\xi}, \rho(\boldsymbol{\xi}))$;
(ii) for any $y(\boldsymbol{\xi}) \in L^2(\boldsymbol{\xi}, \rho(\boldsymbol{\xi}))$, there exists an order-$p$ approximation $y_p(\boldsymbol{\xi}) = \sum_{|\boldsymbol{\alpha}|=0}^{p} c_{\boldsymbol{\alpha}}\Psi_{\boldsymbol{\alpha}}(\boldsymbol{\xi})$, such that $\mathbb{E}[(y(\boldsymbol{\xi}) - y_p(\boldsymbol{\xi}))^2] \to 0$ as $p \to \infty$.

Proof: The detailed proof is given in Appendix B. ∎

IV. HIGHER-ORDER MOMENT COMPUTATION

We further calculate the $2p$-order moments in order to build the basis functions in a high-dimensional setting.
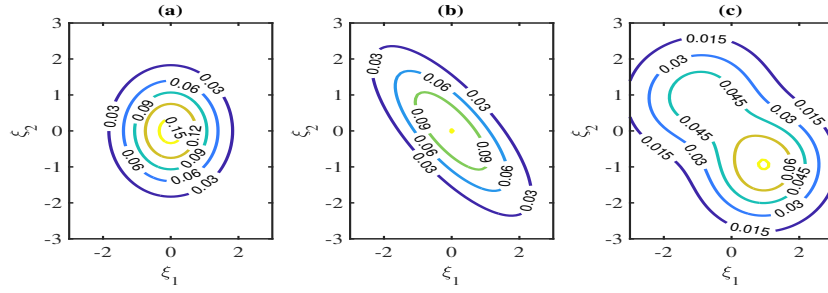
Fig. 2.　Joint density functions for (a): independent Gaussian, (b): correlated Gaussian, (c): correlated non-Gaussian (e.g., a Gaussian mixture distribution).

A. Gaussian Mixture Model

An excellent choice for the data-driven modeling of $\rho(\boldsymbol{\xi})$ is the Gaussian mixture model:

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

$\mathcal{N}(\boldsymbol{\xi}|\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$ denotes the multivariate Gaussian density function with mean $\boldsymbol{\mu}_i \in \mathbb{R}^d$ and a positive definite covariance matrix $\boldsymbol{\Sigma}_i \in \mathbb{R}^{d \times d}$. Fig. 2 has shown the difference of a Gaussian mixture model (e.g., $r > 1$) with independent Gaussian (e.g., $r = 1$ and $\boldsymbol{\Sigma}$ is diagonal) and correlated Gaussian distributions (e.g., $r = 1$ and $\boldsymbol{\Sigma}$ is not diagonal). The corresponding moment is

$$m_{\boldsymbol{\alpha}} = \sum_{i=1}^{r} w_i q_{\boldsymbol{\alpha},i}, \text{ with } q_{\boldsymbol{\alpha},i} = \int \boldsymbol{\xi}^{\boldsymbol{\alpha}} \mathcal{N}(\boldsymbol{\xi}|\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i) d\boldsymbol{\xi}.$$

Existing methods for calculating the higher-order moments for normal distributions rely on the characteristic function [49], [50]. The main bottleneck of these methods is enumerating an integers matrix. In this paper, we propose a functional tensor train approach to compute the higher-order moments.

For simplicity, we ignore the subscript index i in $\boldsymbol{\mu}_i$, $\boldsymbol{\Sigma}_i$ and $q_{\boldsymbol{\alpha},i}$. Denote $\mathbf{A}$ as the lower triangular matrix from the Cholesky decomposition of $\boldsymbol{\Sigma} = \mathbf{A}\mathbf{A}^T$. Then $\boldsymbol{\eta}$ from $\boldsymbol{\xi} = \mathbf{A}\boldsymbol{\eta} + \boldsymbol{\mu}$ satisfies $\boldsymbol{\eta} \sim \mathcal{N}(\boldsymbol{\eta}|\mathbf{0}, \mathbf{I})$. Consequently, $q_{\boldsymbol{\alpha}}$ can be calculated via

$$\begin{aligned} q_{\boldsymbol{\alpha}} &= \int \boldsymbol{\xi}^{\boldsymbol{\alpha}} \mathcal{N}(\boldsymbol{\xi}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) d\boldsymbol{\xi} \\ &= \int (\mathbf{A}\boldsymbol{\eta} + \boldsymbol{\mu})^{\boldsymbol{\alpha}} \frac{\exp(-\boldsymbol{\eta}^T\boldsymbol{\eta})}{\sqrt{(2\pi)^d}} d\boldsymbol{\eta}. \end{aligned} \quad (13)$$

The difficulty in computing (13) lies in $(\mathbf{A}\boldsymbol{\eta} + \boldsymbol{\mu})^{\boldsymbol{\alpha}}$, which is not the product of univariate functions of each η_i.

B. Functional Tensor Train Formula

Fortunately, $q_{\boldsymbol{\alpha}}$ can be computed exactly with an efficient functional tensor-train method. Specifically, we are seeking for $\mathbf{G}_0 \in \mathbb{R}^{1 \times r_0}$ and a set of univariate functional matrices $\mathbf{G}_i(\eta_i) \in \mathbb{R}^{r_{i-1} \times r_i}$ for $i = 1, \cdots d$ with $r_d = 1$, such that

$$(\mathbf{A}\boldsymbol{\eta} + \boldsymbol{\mu})^{\boldsymbol{\alpha}} = \mathbf{G}_0 \mathbf{G}_1(\eta_1) \mathbf{G}_2(\eta_2) \dots \mathbf{G}_d(\eta_d). \quad (14)$$

Afterwards, we can obtain $q_{\boldsymbol{\alpha}}$ via

$$q_{\boldsymbol{\alpha}} = \mathbf{G}_0 \mathbb{E}[\mathbf{G}_1(\eta_1)] \mathbb{E}[\mathbf{G}_2(\eta_2)] \dots \mathbb{E}[\mathbf{G}_d(\eta_d)]. \quad (15)$$

The detailed derivations of $\mathbf{G}_i(\eta_i)$ are as follow.

1) Derivation of (14) *with* $|\boldsymbol{\alpha}| = 1$: The j-th element in $\boldsymbol{\xi} = \mathbf{A}\boldsymbol{\eta} + \boldsymbol{\mu}$ satisfies

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

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

Theorem 1 (Theorem 2, [51]): Any function satisfies

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

can be written as a functional tensor train as in Eq. (17), which equals to the product of some univariate matrices and vectors.

Applying Theorem 1 to (16), we can derive a functional tensor train decomposition for ξ_j:

$$\xi_j = (\mu_j \ 1) \begin{pmatrix} 1 & 0 \\ a_{j1}\eta_1 & 1 \end{pmatrix} \cdots \begin{pmatrix} 1 & 0 \\ a_{j(d-1)}\eta_{d-1} & 1 \end{pmatrix} \begin{pmatrix} 1 \\ a_{jd}\eta_d \end{pmatrix}. \quad (18)$$

Then the expectation is

$$\mathbb{E}[\xi_j] = (\mu_j \ 1) \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \cdots \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \mu_j. \quad (19)$$

The obtained functional tensor trains (18) can be reused to compute higher-order moments.

2) Recurrence Formula for $1 < |\boldsymbol{\alpha}| \leq 2p$: For each $\boldsymbol{\alpha}$ with $1 < |\boldsymbol{\alpha}| \leq 2p$, there exist $\boldsymbol{\alpha}_1$ and $\boldsymbol{\alpha}_2$ with $|\boldsymbol{\alpha}_1|, |\boldsymbol{\alpha}_2| \leq p$, such that

$$\boldsymbol{\xi}^{\boldsymbol{\alpha}} = \boldsymbol{\xi}^{\boldsymbol{\alpha}_1} \cdot \boldsymbol{\xi}^{\boldsymbol{\alpha}_2}, \text{ where } \boldsymbol{\alpha} = \boldsymbol{\alpha}_1 + \boldsymbol{\alpha}_2.$$

According to (6), the tensor-train representation of $\boldsymbol{\xi}^{\boldsymbol{\alpha}}$ can be obtained as the Kronecker product of the tensor trains of $\boldsymbol{\xi}^{\boldsymbol{\alpha}_1}$ and $\boldsymbol{\xi}^{\boldsymbol{\alpha}_2}$. Suppose that $\boldsymbol{\xi}^{\boldsymbol{\alpha}_1} = \mathbf{E}_0(\boldsymbol{\mu})\mathbf{E}_1(\eta_1)\cdots\mathbf{E}_d(\eta_d)$ and $\boldsymbol{\xi}^{\boldsymbol{\alpha}_2} = \mathbf{F}_0(\boldsymbol{\mu})\mathbf{F}_1(\eta_1)\cdots\mathbf{F}_d(\eta_d)$, then

$$\boldsymbol{\xi}^{\boldsymbol{\alpha}} = \mathbf{G}_0(\boldsymbol{\mu})\mathbf{G}_1(\eta_1)\cdots\mathbf{G}_d(\eta_d), \text{ with } \mathbf{G}_i(\eta_i) = \mathbf{E}_i(\eta_i) \otimes \mathbf{F}_i(\eta_i). \quad (20)$$

Here $\mathbf{G}_i(\eta_i) \in \mathbb{R}^{2^{|\boldsymbol{\alpha}|} \times 2^{|\boldsymbol{\alpha}|}}$ for all $i = 1, \dots, d-1$. Because η_i's are mutually independent, finally we have

$$\mathbb{E}[\boldsymbol{\xi}^{\boldsymbol{\alpha}}] = \mathbf{G}_0(\boldsymbol{\mu})\mathbb{E}[\mathbf{G}_1(\eta_1)]\cdots\mathbb{E}[\mathbf{G}_d(\eta_d)]. \quad (21)$$

In other words, the moments can be easily computed via small-size matrix-vector products.

The basis construction framework is summarized in Alg. 1

Remark. If the parameters have a block-wise correlation structure, we can divide $\boldsymbol{\xi} = [\xi_1, \dots, \xi_d]^T$ into several disjoint groups $\boldsymbol{\xi}_{g_1}, \dots, \boldsymbol{\xi}_{g_r}$ in the following way: the random parameters inside each group are correlated, while the parameters among different groups are mutually independent. Under this assumption, the basis functions can be constructed by

$$\Psi_{\boldsymbol{\alpha}}(\boldsymbol{\xi}) = \Psi_{\boldsymbol{\alpha}_1}(\boldsymbol{\xi}_{g_1}) \dots \Psi_{\boldsymbol{\alpha}_r}(\boldsymbol{\xi}_{g_r}).$$

$$f(x_0, x_1, \ldots, x_d) = \begin{pmatrix} \omega_0(x_0) & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ \omega_1(x_1) & 1 \end{pmatrix} \cdots \begin{pmatrix} 1 & 0 \\ \omega_{d-1}(x_{d-1}) & 1 \end{pmatrix} \begin{pmatrix} 1 \\ \omega_d(x_d) \end{pmatrix}. \tag{17}$$

Algorithm 1: A Functional Tensor-Train Method for Computing Basis Functions of Gaussian Mixture Distributions

Input: The mean value, covariance, and weight for Gaussian mixtures $\{\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i, w_i\}_{i=1}^r$, and order p.
for $i = 1, \ldots, r$ **do**
    Compute the Cholesky factor $\mathbf{A}$ via $\boldsymbol{\Sigma}_i = \mathbf{A}\mathbf{A}^T$;
    Calculate the functional tensor trains for the first-order and higher-order monomials via (18) and (20), respectively;
    Obtain the moments via (19) and (21).
Assemble the multivariate moment matrix $\mathbf{M}$ in (9);
Compute the basis functions via (10).
Output: The multivariate basis functions $\Psi_{\boldsymbol{\alpha}}(\boldsymbol{\xi})$.

If there are multiple parameters inside a group, we can construct the basis function $\Psi_{\boldsymbol{\alpha}_i}(\boldsymbol{\xi}_{g_i})$ by the proposed formula (10). Otherwise, the univariate orthogonal basis functions can be calculated via the three-term recurrence relation [41].

V. A SPARSE SOLVER: WHY AND HOW DOES IT WORK?

After constructing the basis functions $\{\Psi_{\boldsymbol{\alpha}}(\boldsymbol{\xi})\}_{|\boldsymbol{\alpha}|=0}^p$, we need to compute the weights (or coefficients) $\{c_{\boldsymbol{\alpha}}\}$. For the independent case, many high-dimensional solvers have been developed, such as compressed sensing [33], [34], analysis of variance [21], [35], model order reduction [36], hierarchical methods [21], [37], [38], and tensor computation [38]–[40]. For the non-Gaussian correlated case discussed in this paper, we employ a sparse solver to obtain the coefficients.

For convenience, we resort all basis functions and their weights $\{\Psi_{\boldsymbol{\alpha}}(\boldsymbol{\xi}), c_{\boldsymbol{\alpha}}\}_{|\boldsymbol{\alpha}|=0}^p$ into $\{\Psi_j(\boldsymbol{\xi}), c_j\}_{j=1}^n$. Given m pairs of parameter samples and simulation values $\{\boldsymbol{\xi}_k, y(\boldsymbol{\xi}_k)\}_{k=1}^m$, our task is to find the coefficient $\mathbf{c}$ such that

$$\boldsymbol{\Phi}\mathbf{c} = \mathbf{y}, \quad \text{with } \Phi_{kj} = \Psi_j(\boldsymbol{\xi}_k), \ y_k = y(\boldsymbol{\xi}_k), \tag{22}$$

where $\boldsymbol{\Phi} \in \mathbb{R}^{m \times n}$ stores the values of n basis functions at m samples and $\mathbf{y} \in \mathbb{R}^m$ stores the m simulation values. In practice, computing each sample $y(\boldsymbol{\xi}_k)$ requires calling a time-consuming device- or circuit-level simulator. Therefore, it is desired to use as few simulation samples as possible.

We consider the compressed sensing technique [33], [34] with $m \ll n$. We seek for the sparsest solution by solving the ℓ_0-minimization problem

$$\min_{\mathbf{c} \in \mathbb{R}^n} \|\mathbf{c}\|_0 \quad \text{s.t.} \quad \boldsymbol{\Phi}\mathbf{c} = \mathbf{y}. \tag{23}$$

Here $\|\mathbf{c}\|_0$ denotes the number of nonzero elements. The compressed sensing technique is subject to some assumptions. Firstly, the solution $\mathbf{c}$ should be sparse in nature, which is generally true in high-dimensional uncertainty quantification. Secondly, the matrix $\frac{1}{\sqrt{m}}\boldsymbol{\Phi}$ should satisfy the restricted isometry property (RIP) [52]: there exists $0 < \kappa_s < 1$ such that

$$(1 - \kappa_s)\|\mathbf{c}\|_2^2 \leq \frac{1}{m}\|\boldsymbol{\Phi}\mathbf{c}\|_2^2 \leq (1 + \kappa_s)\|\mathbf{c}\|_2^2 \tag{24}$$

holds for any $\|\mathbf{c}\|_0 \leq s$. Here, $\|\cdot\|_2$ is the Euclidean norm. Intuitively, this requires that all columns of $\boldsymbol{\Phi}$ are nearly orthogonal to each other.

Compressed sensing techniques have been extensively studied in signal processing. Now we investigate its theoretical condition and accuracy guarantees in our specific setting: high-dimensional uncertainty quantification with non-Gaussian correlated process variations.

A. Conditions to Achieve RIP

In general, it is NP-hard to check whether $\boldsymbol{\Phi}$ satisfies the RIP condition [53]. When the number of samples m is large enough, our matrix $\boldsymbol{\Phi}$ satisfies $\frac{1}{m}\boldsymbol{\Phi}^T\boldsymbol{\Phi} \approx \mathbf{I}$, i.e.,

$$\frac{1}{m}\sum_{k=1}^m (\Psi_i(\boldsymbol{\xi}_k)\Psi_j(\boldsymbol{\xi}_k)) \approx \mathbb{E}[\Psi_i(\boldsymbol{\xi})\Psi_j(\boldsymbol{\xi})] = \delta_{ij}. \tag{25}$$

Here δ_{ij} is the delta function. Hence the RIP condition (24) will be satisfied with a high probability. The following theorem provides a rigorous guarantee.

Theorem 2 (Conditions for RIP): Denote the random variable $X_k^{ij} = \Psi_i(\boldsymbol{\xi}_k)\Psi_j(\boldsymbol{\xi}_k)$. Assume that X_k^{ij} is sub-Gaussian [54] with variance proxy σ for any i, j, k, i.e.,

$$\mathbb{E}[\exp(\lambda(X_k^{ij} - \delta_{ij}))] \leq \exp(\frac{\sigma^2\lambda^2}{2}), \quad \forall \ \lambda \in \mathbb{R}, \tag{26}$$

and the random samples $\{\boldsymbol{\xi}_k\}_{k=1}^m$ are generated independently. Then the RIP condition (24) holds with a probability at least $1 - \eta$ provided that

$$m \geq 2\log(2/\eta)\frac{s^2\sigma^2}{\kappa_s^2}. \tag{27}$$

Proof: See Appendix C. ∎

B. Error Bounds in Uncertainty Quantification

Under the RIP condition, we are able to approximate the solution with good accuracy. Now we provide the error bounds for $\mathbf{c}$ and for the stochastic solution $\mathbf{y}(\boldsymbol{\xi})$. In our implementation, we solve the following constrained optimization

$$\mathbf{c}^* = \arg\min_{\mathbf{c}} \|\boldsymbol{\Phi}\mathbf{c} - \mathbf{y}\|_2 \quad \text{s.t.} \quad \|\mathbf{c}\|_0 \leq s. \tag{28}$$

The above problem can be solved by any ℓ_0-minimization solver, such as COSAMP [55], difference-of-convex [56], and penalty decomposition [57]. The error bound is presented in the following theorem.

Theorem 3 (Coefficient error): Suppose $\mathbf{y} = \boldsymbol{\Phi}\mathbf{c} + \mathbf{e}$, where $\mathbf{e}$ is some random noise and $\mathbf{c}$ is the exact solution. Let $\mathbf{c}_s$ be a sparse vector that remains the s largest-magnitude components of $\mathbf{c}$ and keeps all other components to be zero, and $\epsilon = \|\boldsymbol{\Phi}\mathbf{c}^* - \mathbf{y}\|_2$ is the residue in (28). If $\boldsymbol{\Phi}$ satisfies the $(2s, \kappa_{2s})$-RIP condition, then any solution $\mathbf{c}^*$ of (28) satisfies

$$\|\mathbf{c} - \mathbf{c}^*\|_2 \leq \alpha_0\|\mathbf{c}_s - \mathbf{c}\|_1 + \alpha_1\|\mathbf{e}\|_2 + \alpha_1\epsilon.$$

Here, $\alpha_0 = 1 + \frac{1.7071\sqrt{1+\kappa_{2s}}}{m(1-\kappa_{2s})\sqrt{s}}$ and $\alpha_1 = \frac{1}{m(1-\kappa_{2s})}$ are constants.

Proof: See Appendix D. ∎

Theorem 3 shows that numerical error of computing $\mathbf{c}$ consists of three parts. The first part exists because the exact solution may not be exactly s-sparse. The second part is caused

---

Algorithm 2: An Adaptive Sparse Solver

Input: Input a set of candidate samples Ω^0 and basis functions $\{\Psi_j(\boldsymbol{\xi})\}_{j=1}^n$.

Choose an initial sample set $\Omega \subset \Omega^0$ via the rank-revealing QR factorization, with $m = |\Omega| \ll n$.

Call the simulator to calculate $y(\boldsymbol{\xi})$ for all $\boldsymbol{\xi} \in \Omega$.

for *Iteration* $t = 1, 2, \ldots$ **do**

    Solve the ℓ_0 minimization problem (28) to obtain $\mathbf{c}$.

    Chose the next optimal sample point $\boldsymbol{\xi}_{m+1}$ via the D-optimal, R-optimal, or E-optimal criteria.

    Call the simulator to calculate $y(\boldsymbol{\xi}_{m+1})$.

    Let $m \leftarrow m + 1$.

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

        Stop

Output: The coefficient $\mathbf{c}$ and the surrogate model $y(\boldsymbol{\xi}) = \sum_{|\boldsymbol{\alpha}|=0}^p c_{\boldsymbol{\alpha}} \Psi_{\boldsymbol{\alpha}}(\boldsymbol{\xi})$.

---

by the numerical errors in device/circuit simulation. The third part is caused by the numerical error in an optimization solver.

Now we consider the error of approximating $y(\boldsymbol{\xi})$. For any square-integrable $y(\boldsymbol{\xi})$, denote its ℓ_2 norm as

$$\|y(\boldsymbol{\xi})\|_2 = \sqrt{\mathbb{E}_{\boldsymbol{\xi}}[y^2(\boldsymbol{\xi})]}. \quad (29)$$

Further, let $y_p(\boldsymbol{\xi})$ be the projection of $y(\boldsymbol{\xi})$ to the space $\boldsymbol{\mathcal{S}}_p$. In other words, $y_p(\boldsymbol{\xi}) = \sum_{|\boldsymbol{\alpha}|=0}^p c_{\boldsymbol{\alpha}} \Psi_{\boldsymbol{\alpha}}(\boldsymbol{\xi})$ is the pth-order approximation that we are seeking for. The approximation error is shown as follows.

Theorem 4 (Approximation error): For any square integrable function $y(\boldsymbol{\xi})$, the approximation error is bounded by

$$\|y(\boldsymbol{\xi}) - y^*(\boldsymbol{\xi})\|_2 \leq \|\mathbf{c} - \mathbf{c}^*\|_2 + \|y(\boldsymbol{\xi}) - y_p(\boldsymbol{\xi})\|_2, \quad (30)$$

where $y^*(\boldsymbol{\xi}) = \sum_{|\boldsymbol{\alpha}|=0}^p c_{\boldsymbol{\alpha}}^* \Psi_{\boldsymbol{\alpha}}(\boldsymbol{\xi})$ is our constructed surrogate function, and $\mathbf{c}^* = [\cdots, c_{\boldsymbol{\alpha}}^*, \cdots]$ is the solution of (28).

Proof: The detailed proof is shown in Appendix E. ∎

For the bound shown in (30), the first term is caused by the numerical error in computing $\mathbf{c}$; The second term arises from the distance from $y(\boldsymbol{\xi})$ to the pth-order polynomial space $\boldsymbol{\mathcal{S}}_p$, which will be sufficiently small if p is large enough.

VI. ADAPTIVE SAMPLE SELECTION

The system equation (22) is normally set up by using some random samples [33]. In practice, some samples are informative, yet others are not. Therefore, we present some adaptive sampling methods to improve the performance of the sparse solver. Our method uses a rank-revealing QR decomposition to pick initial samples, then it selects subsequent samples using a D-optimal, R-optimal, or E-optimal criterion. The whole framework is summarized in Alg. 2.

A. Initial Sample Selection

We first select a small number of initial samples from a pool of m_0 random candidate samples Ω^0. By evaluating the basis functions at all candidate samples, we form a matrix $\boldsymbol{\Phi}^0 \in \mathbb{R}^{m_0 \times n}$ whose j-th rows stores the values of n basis functions at the j-th candidate sample. Here, we use the determinant to measure the importance. Specifically, suppose that the rank of $\boldsymbol{\Phi}^0$ is greater than m, we want to separate the rows of $\boldsymbol{\Phi}$ into two parts via a maximization problem

$$\max_{\mathbf{P}} \ \det(\boldsymbol{\Phi}_a^T \boldsymbol{\Phi}_a), \ \text{where} \ (\boldsymbol{\Phi}^0)^T \mathbf{P} = [\boldsymbol{\Phi}_a^T \quad \boldsymbol{\Phi}_b^T]. \quad (31)$$

We achieve this via a rank-revealing QR factorization [58]:

$$(\boldsymbol{\Phi}^0)^T \mathbf{P} = \mathbf{Q}\mathbf{R}, \ \text{where} \ \mathbf{R} = \begin{bmatrix} \mathbf{R}_{11} & \mathbf{R}_{12} \\ \mathbf{0} & \mathbf{R}_{22} \end{bmatrix}. \quad (32)$$

Here $\mathbf{P}$ is a permutation matrix, $\mathbf{Q}$ is an orthogonal matrix, and $\mathbf{R}_{11} \in \mathbb{R}^{m \times m}$ is an upper triangular matrix. We compute the permutation matrix $\mathbf{P}$ such that the absolute values of diagonal elements in $\mathbf{R}$ are in a descending order. In other words, the first m columns of $\mathbf{P}$ indicate the most informative m rows of $\boldsymbol{\Phi}^0$. We keep these rows and the associated parameter samples Ω. The above initial samples have generated a matrix $\boldsymbol{\Phi} \in \mathbb{R}^{m \times n}$, then we further add new informative samples based on a D-optimal, R-optimal, or E-optimal criterion.

B. D-optimal Adaptive Sampling

Our first method is motivated by [59], [60] but differs in the numerical implementation: the method in [59] defined the candidate sample set as quadrature nodes, which are unavailable for non-Gaussian correlated case; the method in [60] used rank-revealing QR at each iteration, whereas our method selects new samples via an optimization method.

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 solutions sequentially. We denote the locations of nonzero coefficients as $\mathcal{C} = \{i_1, \ldots, i_s\}$ with $m > s$, and denote $\boldsymbol{\Phi}_s \in \mathbb{R}^{m \times s}$ as the submatrix of $\boldsymbol{\Phi}$ generated by extracting the columns associated with $\mathcal{C}$. The next most informative sample $\boldsymbol{\xi}_k$ associated with the row vector $\mathbf{x}(\boldsymbol{\xi}_k) = [\Psi_{i_1}(\boldsymbol{\xi}_k) \cdots, \Psi_{i_s}(\boldsymbol{\xi}_k)] \in \mathbb{R}^{1 \times s}$ can be decided via solving the following problem:

$$\max_{\boldsymbol{\xi}_k \in \Omega^0 \setminus \Omega} \ \det(\boldsymbol{\Phi}_s^T \boldsymbol{\Phi}_s + \mathbf{x}(\boldsymbol{\xi}_k)^T \mathbf{x}(\boldsymbol{\xi}_k)), \quad (33)$$

where $\Omega^0 \setminus \Omega$ includes the sample points in Ω^0 but not in Ω. It is unnecessary to compute the above determinant for every sample. The matrix determinant lemma [61] shows $\det(\boldsymbol{\Phi}_s^T \boldsymbol{\Phi}_s + \mathbf{x}(\boldsymbol{\xi}_k)^T \mathbf{x}(\boldsymbol{\xi}_k)) = \det(\boldsymbol{\Phi}_s^T \boldsymbol{\Phi}_s)(1 + \mathbf{x}(\boldsymbol{\xi}_k)(\boldsymbol{\Phi}_s^T \boldsymbol{\Phi}_s)^{-1} \mathbf{x}(\boldsymbol{\xi}_k)^T)$. Therefore, (33) can be solved via

$$\max_{\boldsymbol{\xi}_k \in \Omega^0 \setminus \Omega} \ \mathbf{x}(\boldsymbol{\xi}_k)(\boldsymbol{\Phi}_s^T \boldsymbol{\Phi}_s)^{-1} \mathbf{x}(\boldsymbol{\xi}_k)^T. \quad (34)$$

In our experiments, we obtain the optimal solution of (34) by comparing the objective values for all sample points in $\Omega^0 \setminus \Omega$.

After getting the new sample, we update the matrix $\boldsymbol{\Phi}_s := \begin{bmatrix} \boldsymbol{\Phi}_s \\ \mathbf{x}(\boldsymbol{\xi}_k) \end{bmatrix}$, update $(\boldsymbol{\Phi}_s^T \boldsymbol{\Phi}_s)^{-1}$ via the Sherman-Morrison formula [62], and recompute the s nonzero elements of $\mathbf{c}$ by

$$\mathbf{c}_1 = (\boldsymbol{\Phi}_s^T \boldsymbol{\Phi}_s)^{-1} \boldsymbol{\Phi}_s^T \mathbf{y}. \quad (35)$$

Inspired by [63], we stop the iteration if $\mathbf{c}_1$ is close to its previous step or if the maximal iteration number is reached.

C. R-optimal Adaptive Sampling

The RIP condition (24) is equivalent to

$$(1 - \kappa_s) \leq \lambda_{\min}(\frac{1}{m} \boldsymbol{\Phi}_s^T \boldsymbol{\Phi}_s) \leq \lambda_{\max}(\frac{1}{m} \boldsymbol{\Phi}_s^T \boldsymbol{\Phi}_s) \leq (1 + \kappa_s). \quad (36)$$

Here $\boldsymbol{\Phi}_s$ contains arbitrary s columns of $\boldsymbol{\Phi}$. The constraint in (36) is equivalent to $\|\frac{1}{m} \boldsymbol{\Phi}_s^T \boldsymbol{\Phi}_s - \mathbf{I}\|_2 \leq \kappa_s$. Therefore, we can select the next sample by minimizing $\|\frac{1}{m} \boldsymbol{\Phi}_s^T \boldsymbol{\Phi}_s - \mathbf{I}\|_2$. We refer this method as R-optimal because it optimizes the RIP condition.
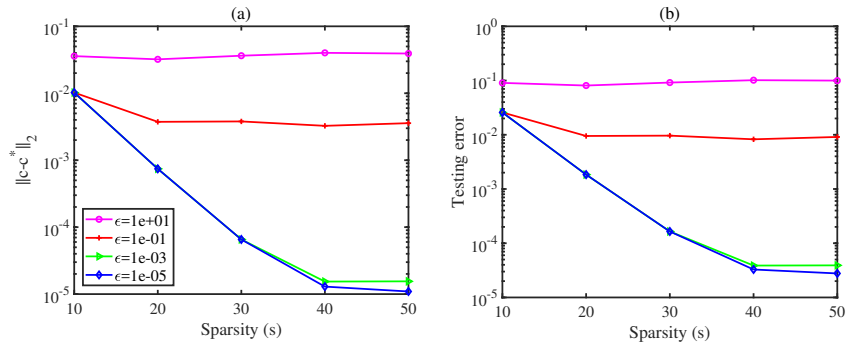
Fig. 3. Numerical error for the synthetic experiment. (a) Errors in the computed coefficients of our basis functions; (b) Testing errors of approximating $y(\boldsymbol{\xi})$.

Suppose that the locations of nonzero coefficients are fixed as $\mathcal{C} = \{i_1, \ldots, i_s\}$. The next sample $\boldsymbol{\xi}_k$ associated with the row vector $\mathbf{x}(\boldsymbol{\xi}_k) = [\Psi_{i_1}(\boldsymbol{\xi}_k) \cdots, \Psi_{i_s}(\boldsymbol{\xi}_k)]$ is found via the following optimization problem

$$\min_{\boldsymbol{\xi}_k \in \Omega^0 \backslash \Omega} \left\| \frac{1}{m+1} \left(\boldsymbol{\Phi}_s^T \boldsymbol{\Phi}_s + \mathbf{x}(\boldsymbol{\xi}_k)^T \mathbf{x}(\boldsymbol{\xi}_k) \right) - \mathbf{I} \right\|_2. \quad (37)$$

In our implementation, we solve (37) by comparing the objective function values of all available sample points.

D. E-optimal Adaptive Sampling

Both D-optimal and R-optimal methods solve an optimization problem and have to explore the entire sample sets. In contrast, we can exploit local information to pick the next sample. Inspired by the exploitation in Bayesian optimization, we propose to find the next sample in a neighbourhood where the approximation error is large. Specifically, we group the existing samples into k clusters $\mathcal{U}^1, \ldots, \mathcal{U}^k$, and compute the average approximation error as

$$\text{res}(\mathcal{U}^i) = \text{mean}(\boldsymbol{\Phi}^i \mathbf{c} - \mathbf{y}^i), \quad (38)$$

where $\boldsymbol{\Phi}^i \in \mathbb{R}^{|\mathcal{U}^i| \times n}$ contains the $|\mathcal{U}^i|$ rows of $\boldsymbol{\Phi}$ that are associated with all samples in $\mathcal{U}^i$. Afterwards, we choose the next sample nearest to the i-th cluster center, where i is the index of cluster with the maximal residue, i.e., $i = \arg \max \text{res}(\mathcal{U}^i)$. This approach is called E-optimal because it *exploits* the samples in a neighbourhood.

VII. NUMERICAL RESULTS

We test our algorithms by a synthetic example and three real-world benchmarks, including a photonic band-pass filter, a 7-stage CMOS ring oscillator, and an array waveguide grating (AWG) with 41 waveguides. 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. We employ COSAMP [55] to solve (28), and stop it when the residue in the objective function satisfies $\|\boldsymbol{\Phi}\mathbf{c} - \mathbf{y}\|_2 \leq \epsilon$. We define the relative error as:

$$\epsilon_r = \|\boldsymbol{\Phi}\mathbf{c} - \mathbf{y}\|_2 / \|\mathbf{y}\|_2. \quad (39)$$

We refer ϵ_r as a training error if the samples are those used in our sparse solver, and as a testing (or prediction) error if an entirely new set of samples are used. The stopping criterion for Alg. 2 is either the maximal number of samples is attained or the training error is small enough. We refer our methods as "D-optimal", "E-optimal", "R-optimal", and "hybrid" (combinations of all three methods), dependent on

different sample selection criterion. We compare our methods with "rand" approach that chooses all samples by Monte Carlo. For the "rand" approach, we run the experiment 10 times using 10 sets of different samples, and report the mean values and variances of ϵ_r. The CPU time of obtaining each simulation sample highly depends on the specific design problem and on the hardware platform. In most cases, the simulation cost dominates the total cost of uncertainty quantification. Therefore, we compare the costs of different methods mainly based on their total numbers of simulation samples.

A. A Synthetic Example

We firstly use a synthetic example to verify our theoretical results in Section V. This example contains $d = 8$ non-Gaussian correlated random parameters $\boldsymbol{\xi}$ and we approximate the stochastic solution $y(\boldsymbol{\xi})$ by our basis functions with a total order bounded by $p = 3$. The sparse coefficient $\mathbf{c}$ is given *a priori*, and the output $y(\boldsymbol{\xi})$ has a closed-form as

$$y(\boldsymbol{\xi}) = \boldsymbol{\Phi}(\boldsymbol{\xi})\mathbf{c} + \mathbf{e}, \quad (40)$$

where $\mathbf{e}$ is a random simulation noise satisfying $\|\mathbf{e}\|_2 = 10^{-6}$.

In order to verify Theorems 3 and 4, we generate $m = 200$ random samples and approximate the sparse coefficients via an ℓ_0-minimization (28). We stop the algorithm when $\|\boldsymbol{\Phi}\mathbf{c} - \mathbf{y}\|_2 \leq \epsilon$. Fig. 3 (a) shows that when the numerical error ϵ is too large, the error will always be dominated by ϵ. Otherwise, when ϵ is small enough, the prescribed sparsity s will dominate the error. This is consistent with Theorem 3. Fig. 3 (b) confirms Theorem 4: when $y(\boldsymbol{\xi})$ and the polynomial order p are fixed (hence $\|y(\boldsymbol{\xi}) - y_p(\boldsymbol{\xi})\|_2$ is fixed), the overall error is entirely dependent on the coefficient error.

Remark. Fig. 3 shows that a large s leads to smaller errors when ϵ is small enough. However, we cannot set the sparsity s to be too large, because a larger s requires more samples to achieve the RIP condition. Therefore, in the following experiments, we set s as the largest integer below $\frac{m}{3}$.

B. Photonic Band-pass Filter (19 Parameters)

Now we consider the photonic band-pass filter in Fig. 5. This photonic IC has 9 micro-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-μm operation wavelength. A total of 19 random parameters are used to describe the variations of the effective phase index (n_{neff}) 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 with three components.
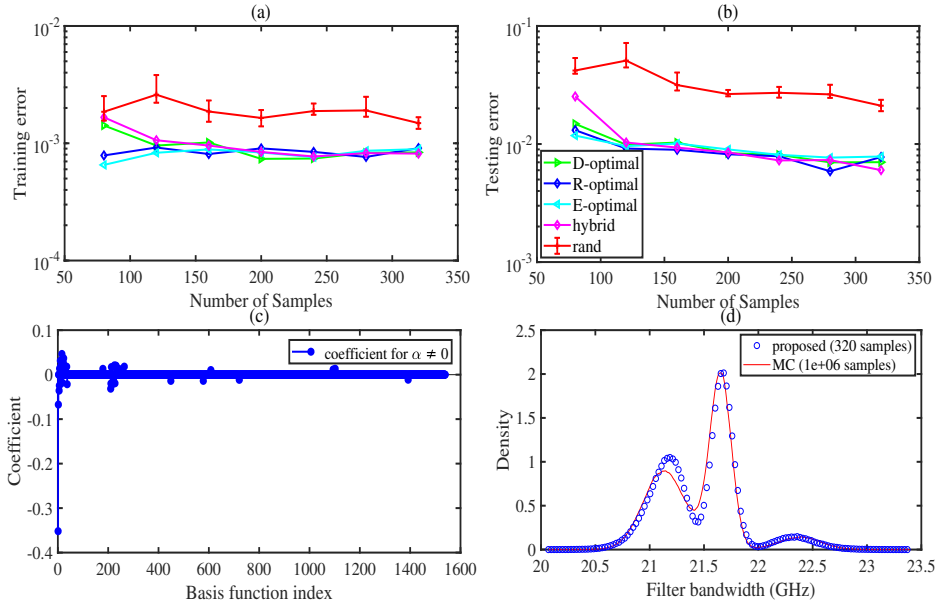
Fig. 4. Results for the photonic band-pass filter. (a) training error; (b) testing error on 9000 new samples; (c) calculated coefficients/weights of our proposed basis functions; (d) probability density functions of the 3-dB bandwidth obtained with our proposed method and with Monte Carlo (MC), respectively.
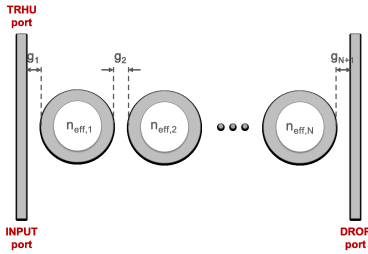

Fig. 5. A band-pass filter with 9 micro-ring resonators.

TABLE I
ACCURACY COMPARISON ON THE PHOTONIC BAND-PASS FILTER. THE
UNDERSCORES INDICATE PRECISION.

method	Proposed	Monte Carlo		
# samples	320	10^2	10^4	10^6
mean (GHz)	21.4773	21.5297	21. 4867	21.4782
std (GHz)	0.3884	0.4131	0.3767	0.3808

We approximate the 3-dB bandwidth f_{3dB} at the DROP port using our basis functions with the total order bounded by $p = 3$. It takes 350.59 seconds to generate the 1540 basis functions. We verify D-optimal, R-optimal, E-optimal methods, and their combinations (denoted as "hybrid"). Fig. 4 (b) clearly shows that all four adaptive sampling methods lead to significantly lower testing (i.e., prediction) errors because they choose more informative samples. Finally, we use 320 samples to assemble a linear system and solve it by an ℓ_0 minimization, and obtain the sparse coefficients of our basis functions in Fig. 4 (c). Although a third-order expansion involves more than 1000 basis functions, only a few dozens are important. Fig. 4 (d) shows the predicted probability density function of the filter's 3-dB bandwidth, and it matches the result from Monte Carlo very well. 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 stochastic

TABLE II
ACCURACY COMPARISON ON THE CMOS RING OSCILLATOR. THE
UNDERSCORES INDICATE PRECISION.

method	Proposed	Monte Carlo		
# samples	300	10^2	10^4	10^6
mean (MHz)	90.5797	89.7795	90.4945	90.5253
std (MHz)	14.6068	14.4512	14.6975	14.7400

model, we compare the computed mean values and standard variations of f_{3dB} from our methods with that of Monte Carlo in Table I. Our method provides a closed-form expression for the mean value. Monte Carlo method converges very slowly and requires $3125\times$ more simulation samples to achieve the similar level of accuracy (with 2 accurate fractional digits).

C. CMOS Ring Oscillator (57 Parameters)

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

We employ a 2nd-order expansion to model the oscillator frequency. Generating the 1711 basis functions takes 1657 seconds. The simulation samples are obtained by calling a periodic steady-state simulator repeatedly. The detailed results are shown in Fig. 6. Our adaptive sparse solver produces a sparse and highly accurate stochastic solution with better prediction behaviors than the standard compressed 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 II compares our method with Monte Carlo. Our method takes about $3333\times$ less samples than Monte Carlo to achieve a precision of one fractional digit for the mean value.
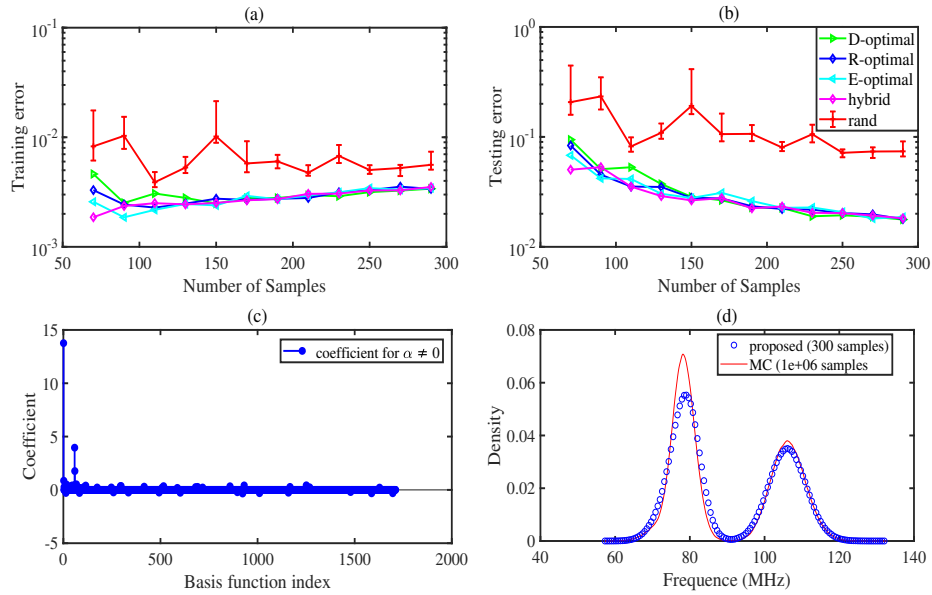
Fig. 6. Results for the CMOS ring oscillator. (a) training error; (b) testing error on 9000 new samples; (c) calculated coefficients/weights of our proposed basis functions; (d) probability density functions of the oscillator frequency obtained by our method with 300 training samples and by Monte Carlo, respectively.
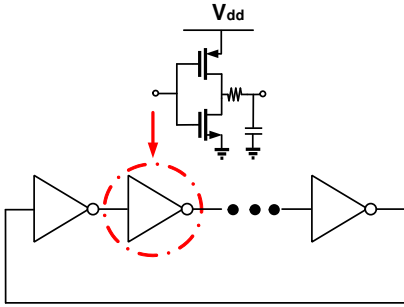

Fig. 7. Schematic of a CMOS ring oscillator.

D. Array Waveguide Grating (AWG, 43 Parameters)

Finally, we investigate an arrayed waveguide grating (AWG) [64]. The AWG is essential for wavelength division multiplexing in photonic systems. We use an AWG with 41 waveguide arrays and two-star couplers, as shown in Fig. 8 (a). In the nominal design, the radius of each star coupler is $R_1 = R_2 = 2.985$ mm, and the waveguide lengths $L_1, \ldots, L_{41}$ range from 46 μm to 1.9 mm. We use a Gaussian mixture distribution with two components to describe the uncertainties in the waveguide lengths, and a Gamma distribution to formulate the uncertainties in each star coupler. The resulting transmission with uncertainties is shown in Fig. 8 (b).

We approximate the transmission rate of peak 1 at the output 1 by our proposed basis functions with a total order $p = 2$. It takes 227.39 seconds to generate 990 basis functions. The numerical results are presented in Fig. 9. Similar to the previous examples, our adaptive sparse solver produces a sparse and highly accurate stochastic solution with better prediction accuracy than using random samples. Table III compares our method with Monte Carlo. Our method consumes about $1111\times$ less sample points than Monte Carlo to get two exact fractional digits for the mean value.

TABLE III
ACCURACY COMPARISON ON THE AWG BY ADAPTIVE SAMPLING. THE UNDERSCORES INDICATE PRECISION.

method	Proposed	Monte Carlo	
# samples	90	10^3	10^5
mean	0.8739	0.8802	0.8792
std	0.0120	0.0134	0.0135

VIII. CONCLUSION

This paper has presented a set of theoretical and numerical results for high-dimensional uncertainty quantification with non-Gaussian correlated process variations. We have proposed a set of basis functions for non-Gaussian correlated cases, and have provided a functional tensor-train method for their high-dimensional implementation. Theoretical results on the expressivity of our basis function are presented. In order to reduce the computational time of analyzing process variations, we have justified the theoretical foundations (i.e., theoretical conditions and numerical errors) of compressed sensing in our problem setting. We have also proposed several adaptive sampling techniques to improve the performance of compressed sensing. Our approach has been verified with a synthetic example and three electronic and photonic ICs with up to 57 random parameters. On these benchmarks, our method has achieved high accuracy in predicting the multi-peak output probability density functions and in estimating the output mean value. Our method has consumed $1111\times$ to $3333\times$ less samples than Monte Carlo to achieve a similar level of accuracy.

APPENDIX A
PROOF OF LEMMA 1

We show the completeness of our basis function in $\mathcal{S}_p$ via two steps. Firstly, it follows from the definition of polynomials that the monomials (8) are complete basis functions for $\mathcal{S}_p$: any $y(\boldsymbol{\xi}) \in \mathcal{S}_p$ can be written as $y(\boldsymbol{\xi}) = \mathbf{c}_0^T \mathbf{b}(\boldsymbol{\xi})$. Secondly, our basis function is an equivalent linear transformation from the monomials $\mathbf{b}(\boldsymbol{\xi})$. Consequently,

$$y(\boldsymbol{\xi}) = \mathbf{c}_0^T \mathbf{b}(\boldsymbol{\xi}) = \mathbf{c}_0^T \mathbf{L} \boldsymbol{\Phi}(\boldsymbol{\xi}) := \mathbf{c}^T \boldsymbol{\Phi}(\boldsymbol{\xi}).$$
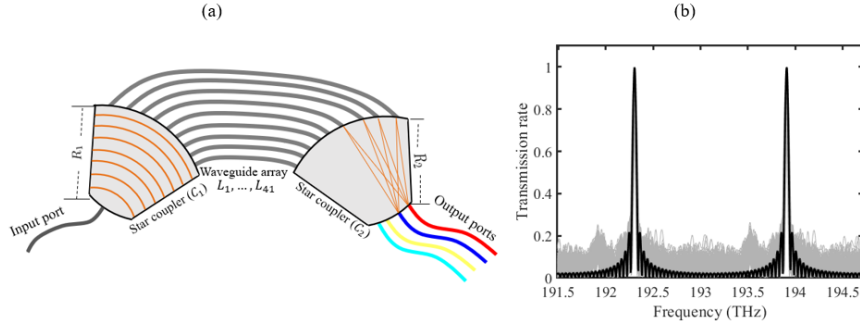
Fig. 8.   (a) An AWG with 41 waveguide arrays; (b) The transmission rates from the input to output Port 1. The black curve shows the nominal result without any uncertainties, and the grey curves show the effects caused by the fabrication uncertainties of radius R_1, R_2 and waveguide lengths $L_1, \ldots, L_{41}$.
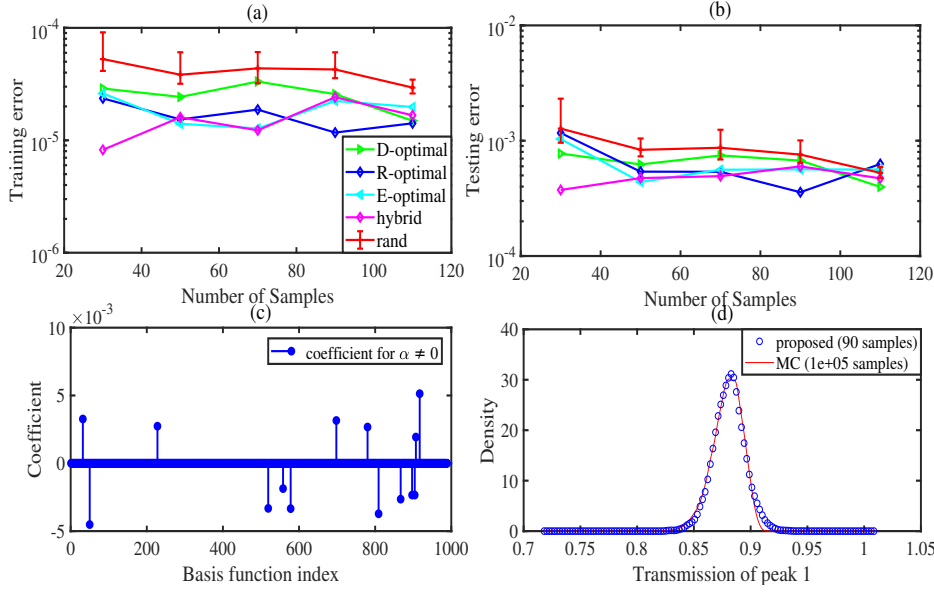

Fig. 9.   Numerical results of the AWG example. (a) Training error; (b) testing error on 10^5 new samples; (c) computed coefficients/weights of our basis functions; (d) probability density function of transmission of peak one with our proposed method and Monte Carlo, respectively.

This shows that our proposed basis function is complete (any function $y(\boldsymbol{\xi}) \in \boldsymbol{\mathcal{S}}_p$ can be expressed by a linear transformation of our proposed basis function).

APPENDIX B
PROOF OF LEMMA 2

The statements (i) and (ii) hold if $\boldsymbol{\mathcal{S}}_p$ is dense in $L^2(\boldsymbol{\xi}, \rho(\boldsymbol{\xi}))$. According to Theorem 3 of [65], a sufficient condition would be the following: there exists $q > 2$ such that the 1-D polynomials are dense in $L^q(\xi_i, \rho_i(\xi_i)), \forall i = 1, \ldots, d$. Here $\rho_i(\xi_i)$ is the marginal distribution of ξ_i.

Consider the following two cases. (i), the marginal distribution $\rho_i(\xi_i)$ is defined on a compact domain. Then the 1-D polynomials are dense in $L^q(\xi_i, \rho_i(\xi_i))$ under the Weierstrass Approximation Theorem [66]. (ii), $\rho_i(\xi_i)$ is defined on a non-compact domain. In this case, the 1-D polynomials are dense in $L^q(\xi_i, \rho_i(\xi_i))$ under the condition that the random variables are exponentially integrable [65]. Namely, there exits a constant $a \geq 0$ such that

$$\mathbb{E}[\exp(a|\xi_i|)] = \int_{\mathbb{R}} \exp(a|\xi_i|)\rho_i(\xi_i)d\xi_i < \infty, \forall i. \quad (41)$$

In both two cases, the 1-D polynomials are dense in $L^q(\xi_i, \rho_i(\xi_i)), \forall i = 1, \ldots, d$, hence the multidimensional polynomials are dense in $L^2(\boldsymbol{\xi}, \rho(\boldsymbol{\xi}))$ [65].

Remark. Inequality (41) holds for many well-known distributions, such as normal distribution, Gaussian mixture distribution, and Gamma distribution.

APPENDIX C
PROOF OF THEOREM 2

A sufficient condition to achieve the (s, κ_s)-RIP condition is if the following inequality holds

$$\|\frac{1}{m}\boldsymbol{\Phi}_s^T\boldsymbol{\Phi}_s - \mathbf{I}_s\|_F \leq \kappa_s \quad (42)$$

for any $\boldsymbol{\Phi}_s$ constructed by arbitrary s columns of $\boldsymbol{\Phi}$. Equation (42) can be derived if each element satisfies

$$|\frac{1}{m}\sum_k X_k^{ij} - \delta_{ij}| \leq \frac{\kappa_s}{s}. \quad (43)$$

Here $\delta_{ij} = 1$ if $i = j$, and $\delta_{ij} = 0$ otherwise. It follows from the concentration bounds of sub-Gaussian random variables [54] that for any $t \geq 0$ there is

$$P\left(\left|\frac{1}{m}\sum_{k=1}^m(X_k^{ij} - \delta_{ij})\right| \geq \frac{t}{m}\right) \leq 2\exp(-\frac{t^2}{2m\sigma^2}). \quad (44)$$

Substituting $t = m\frac{\kappa_s}{s}$ and $2\exp(-\frac{t^2}{2m\sigma^2}) \leq \eta$ into the above equation, we have that (43) holds with a probability $\geq 1 - \eta$ if $2\exp(-\frac{t^2}{2m\sigma^2}) = 2\exp(-\frac{m\kappa_s^2}{2s^2\sigma^2}) \leq \eta$ (i.e., $m \geq 2\log\left(\frac{2}{\eta}\right)\frac{s^2\sigma^2}{\kappa_s^2}$).

APPENDIX D
PROOF OF THEOREM 3

Denote vector $\mathbf{c}$ as the exact unknown coefficients, vector $\mathbf{c}_s$ as the s-sparsity approximation of $\mathbf{c}$, and vector $\mathbf{c}^*$ as the solution from our ℓ_0-minimization solver. The error of $\mathbf{c}^*$ satisfies

$$\|\mathbf{c} - \mathbf{c}^*\|_2 \leq \|\mathbf{c} - \mathbf{c}_s\|_2 + \|\mathbf{c}_s - \mathbf{c}^*\|_2$$
$$\leq \|\mathbf{c} - \mathbf{c}_s\|_2 + \frac{1}{m(1 - \kappa_{2s})}\|\mathbf{\Phi}\mathbf{c}_s - \mathbf{\Phi}\mathbf{c}^*\|_2,$$

where

$$\|\mathbf{\Phi}\mathbf{c}_s - \mathbf{\Phi}\mathbf{c}^*\|_2 \leq \|\mathbf{\Phi}\mathbf{c}_s - \mathbf{\Phi}\mathbf{c}\|_2 + \|\mathbf{\Phi}\mathbf{c} - \mathbf{\Phi}\mathbf{c}^*\|_2$$
$$\leq \|\mathbf{\Phi}\mathbf{c}_s - \mathbf{\Phi}\mathbf{c}\|_2 + \|\mathbf{y} - \mathbf{\Phi}\mathbf{c}^*\|_2 + \|\mathbf{e}\|_2$$
$$\leq \|\mathbf{\Phi}\mathbf{c}_s - \mathbf{\Phi}\mathbf{c}\|_2 + \epsilon + \|\mathbf{e}\|_2.$$

By Proposition 3.5 in [55], it holds that

$$\|\mathbf{\Phi}(\mathbf{c}_s - \mathbf{c})\|_2 \leq \sqrt{1 + \kappa_{2s}}(\|\mathbf{c}_s - \mathbf{c}\|_2 + \frac{1}{\sqrt{2s}}\|\mathbf{c}_s - \mathbf{c}\|_1)$$
$$\leq \frac{1.7071\sqrt{1 + \kappa_{2s}}}{\sqrt{s}}\|\mathbf{c}_s - \mathbf{c}\|_1.$$

Combing the above equations together, we have

$$\|\mathbf{c} - \mathbf{c}^*\|_2$$
$$\leq \|\mathbf{c} - \mathbf{c}_s\|_1 + \left(\frac{1.7071\sqrt{1 + \kappa_{2s}}}{m(1 - \kappa_{2s})\sqrt{s}}\|\mathbf{c}_s - \mathbf{c}\|_1 + \epsilon + \|\mathbf{e}\|_2\right)$$
$$= \alpha_0\|\mathbf{c}_s - \mathbf{c}\|_1 + \alpha_1(\epsilon + \|\mathbf{e}\|_2), \tag{45}$$

where $\alpha_0 = 1 + \frac{1.7071\sqrt{1 + \kappa_{2s}}}{m(1 - \kappa_{2s})\sqrt{s}}$, $\alpha_1 = \frac{1}{m(1 - \kappa_{2s})}$ are constants.

APPENDIX E
PROOF OF THEOREM 4

Denote $y(\boldsymbol{\xi})$ as the unknown quantity of interest, $y_p(\boldsymbol{\xi}) = \sum_{|\boldsymbol{\alpha}|=0}^{p} c_{\boldsymbol{\alpha}}\Psi_{\boldsymbol{\alpha}}(\boldsymbol{\xi})$ as the projection of $y(\boldsymbol{\xi})$ onto the p-th order polynomial space $\boldsymbol{S}_p$, and $y^*(\boldsymbol{\xi}) = \sum_{|\boldsymbol{\alpha}|=0}^{p} c_{\boldsymbol{\alpha}}^*\Psi_{\boldsymbol{\alpha}}(\boldsymbol{\xi})$ as the model from our numerical framework, then we have

$$\|y(\boldsymbol{\xi}) - y^*(\boldsymbol{\xi})\|_2 \leq \|y(\boldsymbol{\xi}) - y_p(\boldsymbol{\xi})\|_2 + \|y_p(\boldsymbol{\xi}) - y^*(\boldsymbol{\xi})\|_2. \tag{46}$$

The first term is the distance of $y(\boldsymbol{\xi})$ to $\boldsymbol{S}_p$, which can be very small if p is large enough. The second term is due to the error caused by a compressed sensing solver:

$$\|y_p(\boldsymbol{\xi}) - y^*(\boldsymbol{\xi})\|_2 = \sqrt{\mathbb{E}[(y_p(\boldsymbol{\xi}) - y^*(\boldsymbol{\xi}))^2]}$$
$$= \sqrt{\mathbb{E}[(\sum_{|\alpha|=0}^{p}(c_{\boldsymbol{\alpha}} - c_{\boldsymbol{\alpha}}^*)\Psi_{\boldsymbol{\alpha}}(\boldsymbol{\xi}))^2]} = \sqrt{\sum_{|\alpha|=0}^{p}(c_{\boldsymbol{\alpha}} - c_{\boldsymbol{\alpha}}^*)^2}$$
$$= \|\mathbf{c} - \mathbf{c}^*\|_2,$$

where the third equality is due to the orthonormal property of our basis functions and the last equality is derived from the definition of the ℓ_2-norm in the Euclidean space.
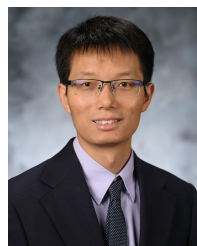
REFERENCES

[1] C. Cui and Z. Zhang, "Uncertainty quantification of electronic and photonic ICs with non-Gaussian correlated process variations," in *Proc. Intl. Conf. Computer-Aided Design*, 2018, pp. 1–8.

[2] D. S. Boning, K. Balakrishnan, H. Cai *et al.*, "Variation," *IEEE Trans. Semiconductor Manufacturing*, vol. 21, no. 1, pp. 63–71, 2008.

[3] 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.

[4] 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.

[5] S. Weinzierl, "Introduction to Monte Carlo methods," NIKHEF, Theory Group, The Netherlands, Tech. Rep. NIKHEF-00-012, 2000.

[6] 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.

[7] 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.

[8] 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.

[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] X. Chen, J. S. Ochoa, J. E. Schutt-Ainé, and A. C. Cangellaris, "Optimal relaxation of I/O electrical requirements under packaging uncertainty by stochastic methods," in *Electronic Components and Technology Conference*, 2014, pp. 717–722.

[14] T.-A. Pham, E. Gad, M. S. Nakhla, and R. Achar, "Decoupled polynomial chaos and its applications to statistical analysis of high-speed interconnects," *IEEE Trans. Components, Packaging and Manufacturing Technology*, vol. 4, no. 10, pp. 1634–1647, 2014.

[15] 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.

[16] 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 *Porc. Asia and South Pacific Design Automation Conference*, 2007, pp. 474–479.

[17] D. Spina, F. Ferranti, T. Dhaene, L. Knockaert, G. Antonini, and D. V. Ginste, "Variability analysis of multiport systems via polynomial-chaos expansion," *IEEE Trans. Microwave Theory and Techniques*, vol. 60, no. 8, pp. 2329–2338, 2012.

[18] P. Manfredi, D. V. Ginste, D. De Zutter, and F. G. Canavero, "Stochastic modeling of nonlinear circuits via spice-compatible spectral equivalents," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 61, no. 7, pp. 2057–2065, 2014.

[19] M. Rufuie, E. Gad, M. Nakhla, R. Achar, and M. Farhan, "Fast variability analysis of general nonlinear circuits using decoupled polynomial chaos," in *Workshop Signal and Power Integrity*, May 2014, pp. 1–4.

[20] M. Ahadi and S. Roy, "Sparse linear regression (spliner) approach for efficient multidimensional uncertainty quantification of high-speed circuits," *IEEE Trans. CAD of Integr. Circuits Syst.*, vol. 35, no. 10, pp. 1640–1652, 2016.

[21] 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.

[22] 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.

[23] A. Waqas, D. Melati, P. Manfredi, F. Grassi, and A. Melloni, "A polynomial-chaos-expansion-based building block approach for stochastic analysis of photonic circuits," in *Physics and Simulation of Optoelectronic Devices XXVI*, vol. 10526, 2018, p. 1052617.

[24] D. Melati, E. Lovati, and A. Melloni, "Statistical process design kits: analysis of fabrication tolerances in integrated photonic circuits," in *Integrated Photonics Research, Silicon and Nanophotonics*, 2015, pp. IT4A–5.

[25] Z. He, W. Cui, C. Cui, T. Sherwood, and Z. Zhang, "Efficient uncertainty modeling for system design via mixed integer programming," in *Proc. Intl. Conf. Computer-Aided Design*, Nov. 2019.

[26] 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.

[27] S. Wold, K. Esbensen, and P. Geladi, "Principal component analysis," *Chemometrics and intelligent laboratory systems*, vol. 2, no. 1-3, pp. 37–52, 1987.

[28] J. Singh and S. Sapatnekar, "Statistical timing analysis with correlated non-Gaussian parameters using independent component analysis," in *Proc. DAC*, 2006, pp. 155–160.

[29] 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.

[30] C. Cui, M. Gershman, and Z. Zhang, "Stochastic collocation with non-Gaussian correlated parameters via a new quadrature rule," in *Proc. IEEE Conf. Electrical Performance of Electronic Packaging and Syst.*, 2018, pp. 57–59.

[31] C. Cui and Z. Zhang, "Stochastic collocation with non-Gaussian correlated process variations: Theory, algorithms and applications," *IEEE Trans. Components, Packaging and Manufacturing Technology*, 2018.

[32] J. D. Jakeman, F. Franzelin, A. Narayan, M. Eldred, and D. Plfüger, "Polynomial chaos expansions for dependent random variables," *Computer Methods in Applied Mechanics and Engineering*, 2019.

[33] 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.

[34] 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.

[35] X. Ma and N. Zabaras, "An adaptive high-dimensional stochastic model representation technique for the solution of stochastic partial differential equations," *J. Comp. Physics*, vol. 229, no. 10, pp. 3884–3915, 2010.

[36] T. El-Moselhy and L. Daniel, "Variation-aware interconnect extraction using statistical moment preserving model order reduction," in *Proc. DATE*, 2010, pp. 453–458.

[37] Z. Zhang, T. A. El-Moselhy, I. M. Elfadel, and L. Daniel, "Calculation of generalized polynomial-chaos basis functions and *G*auss quadrature rules in hierarchical uncertainty quantification," *IEEE Trans. CAD of Integrated Circuits and Systems*, vol. 33, no. 5, pp. 728–740, 2014.

[38] Z. Zhang, X. Yang, I. 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.

[39] 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.

[40] 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.

[41] W. Gautschi, "On generating orthogonal polynomials," *SIAM J. Sci. Stat. Comput.*, vol. 3, no. 3, pp. 289–317, Sept. 1982.

[42] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM Rev.*, vol. 51, no. 3, pp. 455–500, 2009.

[43] I. Oseledets, "Tensor-train decomposition," *SIAM Journal Sci. Comp.*, vol. 33, no. 5, pp. 2295–2317, 2011.

[44] 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.

[45] Y. Xu, "On multivariate orthogonal polynomials," *SIAM Journal Math. Analysis*, vol. 24, no. 3, pp. 783–794, 1993.

[46] 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.

[47] G. H. Golub and J. H. Welsch, "Calculation of Gauss quadrature rules," *Math. Comp.*, vol. 23, pp. 221–230, 1969.

[48] R. A. Horn, R. A. Horn, and C. R. Johnson, *Matrix analysis (Second Edition), 7.2 Characterizations and Properties*. Cambridge university press, 2012.

[49] D. S. Tracy and S. A. Sultan, "Higher order moments of multivariate normal distribution using matrix derivatives," *Stochastic Analysis and Applications*, vol. 11, no. 3, pp. 337–348, 1993.

[50] K. Phillips *et al.*, "R functions to symbolically compute the central moments of the multivariate normal distribution," *Journal of Statistical Software, Code Snippets*, vol. 33, no. 1, pp. 1–14, 2010.

[51] I. Oseledets, "Constructive representation of functions in low-rank tensor formats," *Constructive Approximation*, vol. 37, no. 1, pp. 1–18, 2013.

[52] E. J. Candès, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Comm. Pure Appl. Math.*, vol. 59, no. 8, pp. 1207–1223, 2006.

[53] A. S. Bandeira, E. Dobriban, D. G. Mixon, and W. F. Sawin, "Certifying the restricted isometry property is hard," *IEEE transactions on information theory*, vol. 59, no. 6, pp. 3448–3450, 2013.

[54] V. V. Buldygin and Y. V. Kozachenko, "Sub-Gaussian random variables," *Ukrainian Mathematical Journal*, vol. 32, no. 6, pp. 483–489, 1980.

[55] 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.

[56] X. Zheng, X. Sun, D. Li, and J. Sun, "Successive convex approximations to cardinality-constrained convex programs: a piecewise-linear dc approach," *Computational Optimization and Applications*, vol. 59, no. 1-2, pp. 379–397, 2014.

[57] Z. Lu, Y. Zhang, and X. Li, "Penalty decomposition methods for rank minimization," *Optimization Methods and Software*, vol. 30, no. 3, pp. 531–558, 2015.

[58] 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.

[59] A. K. Prasad, M. Ahadi, and S. Roy, "Multidimensional uncertainty quantification of microwave/RF networks using linear regression and optimal design of experiments," *IEEE Transactions on Microwave Theory and Techniques*, vol. 64, no. 8, pp. 2433–2446, 2016.

[60] P. Diaz, A. Doostan, and J. Hampton, "Sparse polynomial chaos expansions via compressed sensing and *D*-optimal design," *arXiv preprint arXiv:1712.10131*, 2017.

[61] D. A. Harville, *Matrix algebra from a statistician's perspective*. Springer, 1997, vol. 1.

[62] 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.

[63] 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.

[64] K. Zhang, X. Xiao, Y. Zhang, and S. B. Yoo, "Verilog-A compact modeling and simulation of AWGR based all-to-all optical interconnects," in *CLEO: QELS_Fundamental Science*. OSA, 2018, pp. JW2A–49.

[65] L. Petersen, "On the relation between the multidimensional moment problem and the one-dimensional moment problem," *Mathematica Scandinavica*, pp. 361–366, 1983.

[66] D. Pérez and Y. Quintana, "A survey on the Weierstrass approximation theorem," *Divulgaciones Matemáticas*, vol. 16, no. 1, pp. 231–247, 2008.

Chunfeng Cui received the Ph.D. degree in computational mathematics from Chinese Academy of Sciences, Beijing, China, in 2016 with a specialization in numerical optimization. From 2016 to 2017, she was a Postdoctoral Fellow at City University of Hong Kong, Hong Kong. In 2017, she joined the Department of Electrical and Computer Engineering at University of California Santa Barbara as a Postdoctoral Scholar. From 2011 her research activity is mainly focused in the areas of tensor computations, uncertainty quantification, and machine learning.

She received the Best Paper Award of the IEEE EPEPS 2018.

Zheng Zhang (M'15) received his Ph.D degree in Electrical Engineering and Computer Science from the Massachusetts Institute of Technology (MIT), Cambridge, MA, in 2015. He is an Assistant Professor of Electrical and Computer Engineering with the University of California at Santa Barbara, CA. His research interests include uncertainty quantification and tensor computation, with applications to multi-domain design automation, data analysis and algorithm/hardware co-design of machine learning.

Dr. Zhang received the Best Paper Award of IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems in 2014, the Best Paper Award of IEEE Transactions on Components, Packaging and Manufacturing Technology in 2018, two Best Paper Awards (IEEE EPEPS 2018 and IEEE SPI 2016) and three additional Best Paper Nominations (CICC 2014, ICCAD 2011 and ASP-DAC 2011) at international conferences. His Ph.D. dissertation was recognized by the ACM SIGDA Outstanding Ph.D. Dissertation Award in Electronic Design Automation in 2016, and by the Doctoral Dissertation Seminar Award (i.e., Best Thesis Award) from the Microsystems Technology Laboratory of MIT in 2015. He was a recipient of the Li Ka-Shing Prize from the University of Hong Kong in 2011. He received the NSF CAREER Award in 2019.