Chance-Constrained and Yield-Aware Optimization of Photonic ICs With Non-Gaussian Correlated Process Variations

Chunfeng Cui, Kaikai Liu, and Zheng Zhang, *Member, IEEE*

***Abstract*—Uncertainty quantification has become an efficient tool for uncertainty-aware prediction, but its power in yield-aware optimization has not been well explored from either theoretical or application perspectives. Yield optimization is a much more challenging task. On the one side, optimizing the generally nonconvex probability measure of performance metrics is difficult. On the other side, evaluating the probability measure in each optimization iteration requires massive simulation data, especially, when the process variations are non-Gaussian correlated. This article proposes a data-efficient framework for the yield-aware optimization of photonic ICs. This framework optimizes the design performance with a yield guarantee, and it consists of two modules: 1) a modeling module that builds stochastic surrogate models for design objectives and chance constraints with a few simulation samples and 2) a novel yield optimization module that handles probabilistic objectives and chance constraints in an efficient deterministic way. This deterministic treatment avoids repeatedly evaluating probability measures at each iteration, thus it only requires a few simulations in the whole optimization flow. We validate the accuracy and efficiency of the whole framework by a synthetic example and two photonic ICs. Our optimization method can achieve more than 30× reduction of simulation cost and better design performance on the test cases compared with a Bayesian yield optimization approach developed recently.**

***Index Terms*—Chance constraints, integrated photonics, non-Gaussian correlations, photonic design automation, uncertainty quantification, yield optimization.**

I. Introduction

THE DEMAND for low-power, high-speed communications, and computations have boosted the advances in photonic integrated circuits. Based on the modern nanofabrication technology, hundreds to thousands of photonic components can be integrated on a single chip [1], [2]. However, process variations persist during all the fabrication processes and can cause a significant yield degradation in large-scale design and manufacturing [3]–[6]. Photonic ICs are more sensitive to process variations (e.g., geometric uncertainties) due to their large device dimensions compared with the small wavelength. To achieve an acceptable yield, uncertainty-aware design optimization algorithms are highly desirable [7].

Yield optimization algorithms try to increase the success ratio of a chip under random process variations, and they have been studied for a long time in the electronic circuit design [8]–[11]. However, it is still expensive to reuse existing yield optimization solvers for photonic ICs. The major difficulties include: 1) the quantity of interest (e.g., the probability distribution of a bandwidth) does not admit an explicit expression. Instead, we only know the simulation values at parameter sample points; 2) the design objectives and constraints are defined in a stochastic way. They are hard to compute directly and require massive numerical simulations to estimate their statistical distributions; and 3) practical photonic IC designs often involve non-Gaussian correlated process variations, which are more difficult to capture. To estimate the design yield efficiently, one alternative is to build a surrogate model. In [12]–[14], posynomials were used to model statistical performance, and geometric programming was employed to optimize the worst-case performance. Li *et al.* [15] proposed a Chebyshev affine arithmetic method to predict the cumulative distribution function. The recent Bayesian yield optimization (BYO) [10] approximated the probability density of the design variable under the condition of "pass" by the kernel density estimation. The work [11] further approximated the yield over the design variables directly by a Gaussian process regression. However, these machine learning techniques may still require many simulation samples. Furthermore, worst-case optimization or only optimizing the yield can lead to nonoptimal (and even poor) chip performance.

Recently, uncertainty quantification methods based on generalized polynomial chaos have achieved great success in modeling the impact caused by process variations in electronic and photonic ICs [16]–[27]. A novel stochastic collocation approach was further proposed in [28] and [29] to handle non-Gaussian correlated process variations, which shows significantly better accuracy and efficiency than [30] due to an optimization-based quadrature rule. These techniques construct stochastic surrogate models with a small number of simulation samples, but their power in yield optimization has not been well explored despite the recent robust optimization methods [7] based on generalized polynomial chaos.

Manuscript received August 20, 2019; revised November 13, 2019; accepted January 10, 2020. Date of publication January 22, 2020; date of current version November 20, 2020. This work was supported in part by NSF under Grant 1763699, in part by NSF CAREER Award under Grant 1846476, and in part by UCSB Start-Up Grant. This article was recommended by Associate Editor S. Mohanty. *(Chunfeng Cui and Kaikai Liu contributed equally to this work.)*

The authors are with the Department of Electrical and Computer Engineering, University of California at Santa Barbara, Santa Barbara, CA 93106 USA (e-mail: chunfengcui@ucsb.edu; kaikailiu@ucsb.edu; zhengzhang@ece.ucsb.edu).

Digital Object Identifier 10.1109/TCAD.2020.2968582

Leveraging the chance-constrained optimization [31] and our recently proposed uncertainty quantification solvers [28], [29], this article presents a data-efficient technique to optimize photonic ICs with non-Gaussian correlated process variations. Instead of just optimizing the yield, we optimize a target performance metric while enforcing the probability of violating design rules to be smaller than a user-defined threshold. Doing so can avoid performance degradation in yield optimization. Chance-constrained optimization [31] has been widely used in system control [32], autonomous vehicles [33], and reliable power generation [34], [35], but it has not been investigated for yield optimization of electronic or photonic ICs. Our specific contributions include.

1) A chance-constraint optimization framework that can achieve high chip performance and high yield simultaneously under non-Gaussian correlated process variations.
2) A surrogate model that approximates the stochastic objective and constraint functions with a few simulations. Since both the objective function and constraints are only available through a black-box simulator, we build a surrogate model based on the recent uncertainty quantification solver [29]. The main step is to compute a quadrature rule in the joint space of design variables and stochastic parameters by a new three-stage optimization process.
3) A deterministic reformulation. A major challenge of chance-constrained optimization is to reformulate the stochastic constraints into deterministic ones [36]. We reformulate the probabilistic objective function and constraints as nonsmooth deterministic functions. Afterward, we transform them into an equivalent polynomial optimization, which can be solved efficiently.
4) Validations on benchmarks. Finally, we validate the efficiency of our proposed framework on a synthetic example, a microring add-drop filter, and a Mach–Zehnder filter. Preliminary numerical experiments show that our proposed framework can find the optimal design variable efficiently. Compared with the BYO method [10], our proposed method can reduce the number of simulations by 30×, achieve better performance, and produce a similar yield on the test cases.

This article should be regarded as a preliminary result in this direction, and many topics can be investigated in the future.

II. Preliminaries

A. Yield Optimization

The yield is defined as the percentage of qualified products overall. For a photonic IC, denote the design variables by $\boldsymbol{x} = [x_1, x_2, \ldots, x_{d_1}]^T \in \mathcal{X}$ and the process variations by random parameters $\boldsymbol{\xi} = [\xi_1, \xi_2, \ldots, \xi_{d_2}]^T \in \Omega$. Suppose $\boldsymbol{x}$ is uniformly distributed in a bound domain and $\boldsymbol{\xi}$ follows a probability distribution $\rho(\boldsymbol{\xi})$. Let $\{y_i(\boldsymbol{x}, \boldsymbol{\xi})\}_{i=1}^n$ denote a set of performance metrics of interest, $\{u_i\}_{i=1}^n$ denote its required upper bound, and $I(\boldsymbol{x}, \boldsymbol{\xi})$ denote the indicator function

$$I(\boldsymbol{x}, \boldsymbol{\xi}) = \begin{cases} 1, & \text{if } y_i(\boldsymbol{x}, \boldsymbol{\xi}) \leq u_i \quad \forall i = 1, \ldots, n \\ 0, & \text{otherwise.} \end{cases} \tag{1}$$

The yield at a certain design choice $\boldsymbol{x}$ is defined as

$$Y(\boldsymbol{x}) = \text{Prob}_{\boldsymbol{\xi}}(\mathbf{y}(\boldsymbol{x}, \boldsymbol{\xi}) \leq \boldsymbol{u} | \boldsymbol{x}) = \mathbb{E}_{\boldsymbol{\xi}}\big[I(\boldsymbol{x}, \boldsymbol{\xi})\big]. \tag{2}$$

The yield optimization problem aims to find an optimal design variable $\boldsymbol{x}^*$ such that

$$\boldsymbol{x}^* = \underset{\boldsymbol{x} \in \mathcal{X}}{\text{argmax}} \quad Y(\boldsymbol{x}). \tag{3}$$

There are three major difficulties in solving the above yield optimization problem: 1) the indicator function $I(\boldsymbol{x}, \boldsymbol{\xi})$ does not always admit an explicit formulation; 2) computing the yield $Y(\boldsymbol{x})$ involves a nontrivial numerical integration, which requires numerous simulations at each design variable $\boldsymbol{x}$; and 3) $Y(\boldsymbol{x})$ is an implicit nonconvex function and it is difficult to compute its optimal solution.

B. Chance Constraints

The chance constraint is a powerful technique in uncertainty-aware optimization [31]. In comparison with the deterministic constraints or the worst-case constraints where the risk level ϵ is zero, a chance constraint enforces the probability of satisfying a stochastic constraint to be above a certain confidence level $1 - \epsilon$ (ϵ is usually not zero)

$$\text{Prob}_{\boldsymbol{\xi}}(y(\boldsymbol{x}, \boldsymbol{\xi}) \leq \boldsymbol{u}) \geq 1 - \epsilon \tag{4}$$

or equivalently, the probability of violating the constraint to be smaller than the risk level ϵ

$$\text{Prob}_{\boldsymbol{\xi}}(y(\boldsymbol{x}, \boldsymbol{\xi}) \geq \boldsymbol{u}) \leq \epsilon. \tag{5}$$

Under strict conditions, such as the parameters being independent and $y(\boldsymbol{x}, \boldsymbol{\xi})$ being a linear function, (4) can be reformulated into equivalent deterministic constraints [37]. In other words, one can reformulate the left-hand side of (4) by its probability density function (PDF) and substitute the right-hand side by a constant related to the cumulative density function. However, these conditions rarely hold in practice. Even if the conditions hold, computing the PDF or cumulative density function of an uncertain variable can be intractable [17], [36]. In these cases, we seek for deterministic reformulations that can well approximate the chance constraints. There is a tradeoff in choosing the reformulation: if the reformulation is aggressive (the feasible domain is enlarged), it may result in an infeasible solution; otherwise, if the reformulation is conservative (the feasible domain is decreased), the solution may be degraded.

One may convert the chance constraint (4) to a deterministic constraint via the mean and variance of $y(\boldsymbol{x}, \boldsymbol{\xi})$ [36], [37]

$$\mathbb{E}_{\boldsymbol{\xi}}\big[y(\boldsymbol{x}, \boldsymbol{\xi})\big] + \kappa_\epsilon \sqrt{\text{var}_{\boldsymbol{\xi}}\big[y(\boldsymbol{x}, \boldsymbol{\xi})\big]} \geq \boldsymbol{u}. \tag{6}$$

Here, $\mathbb{E}_{\boldsymbol{\xi}}[\cdot]$ denotes the mean value and $\text{var}_{\boldsymbol{\xi}}[\cdot]$ denotes the variance. The constant κ_ϵ is chosen as $\kappa_\epsilon = \sqrt{(1-\epsilon)/\epsilon}$. The detailed proof is shown in Appendix A. It is worth noting that (6) is a stronger constraint than (4): every feasible point of (6) is also a feasible point of the original chance constraint (4).

C. Stochastic Spectral Methods

Assume that $y(\boldsymbol{\xi})$ is a smooth function satisfying $\mathbb{E}[y^2(\boldsymbol{\xi})] \leq \infty$. The stochastic spectral methods can approximate $y(\boldsymbol{\xi})$ by orthonormal polynomial basis functions

$$y(\boldsymbol{\xi}) \approx \sum_{|\boldsymbol{\alpha}|=0}^{p} c_{\boldsymbol{\alpha}} \Psi_{\boldsymbol{\alpha}}(\boldsymbol{\xi}), \text{ with } \mathbb{E}\big[\Psi_{\boldsymbol{\alpha}}(\boldsymbol{\xi})\Psi_{\boldsymbol{\beta}}(\boldsymbol{\xi})\big] = \delta_{\boldsymbol{\alpha},\boldsymbol{\beta}}. \quad (7)$$

Here $|\boldsymbol{\alpha}| = \alpha_1 + \cdots + \alpha_{d_2}$, $\Psi_{\boldsymbol{\alpha}}(\boldsymbol{\xi})$ is an orthonormal basis function indexed by $\boldsymbol{\alpha}$, and $c_{\boldsymbol{\alpha}}$ is its corresponding coefficient.

If the parameters $\boldsymbol{\xi}$ are independent, $\rho(\boldsymbol{\xi})$ equals the products of its 1-D marginal density function $\rho_i(\xi_i)$. In this case, the basis function $\Psi_{\boldsymbol{\alpha}}(\boldsymbol{\xi})$ is the product of multiple 1-D orthogonal basis functions

$$\Psi_{\boldsymbol{\alpha}}(\boldsymbol{\xi}) = \psi_1(\xi_1) \ldots \psi_{d_2}\big(\xi_{d_2}\big). \quad (8)$$

These 1-D basis functions $\psi_i(\xi_i)$ can be constructed by the three term recursion [38]. Various stochastic spectral approaches have been proposed to compute the coefficients $c_{\boldsymbol{\alpha}}$, including the intrusive (i.e., nonsampling) solvers (e.g., stochastic Galerkin [39], the stochastic testing [16]) and the nonintrusive (i.e., sampling) solvers (e.g., stochastic collocation [40]). In the past few years, there has also been a rapid progress in handling high-dimensional parameters, such as the tensor recovery method [19], the compressive sensing technique [41], analysis of variance (ANOVA) or the high-dimensional model representation (HDMR) [42], and the hierarchical uncertainty quantification [18].

In practice, the random parameters may be correlated. If the parameters $\boldsymbol{\xi}$ are non-Gaussian correlated, the computation is more difficult. In such cases, $\Psi_{\boldsymbol{\alpha}}(\boldsymbol{\xi})$ can be constructed by the Gram–Schmidt approach [28], [29] or the Cholesky factorization [43], [44]. The main difficulty lies in computing high order moments of $\boldsymbol{\xi}$, which can be well resolved by the functional tensor train approach [44].

III. Our Yield-Aware Optimization Model

In this section, we show our yield-aware chance-constrained optimization model, and illustrate how to convert the stochastic formulation to a deterministic one. The basic assumptions are listed as follows.

Assumption 1: We made the following assumptions.

1) The design variable $\boldsymbol{x}$ is upper and lower bounded, i.e., $\boldsymbol{x} \in \mathcal{X} = [\mathbf{a}, \mathbf{b}]^{d_1}$.
2) The stochastic parameter $\boldsymbol{\xi} \in \Omega \in \mathbb{R}^{d_2}$ admits a non-Gaussian correlated density function $\rho(\boldsymbol{\xi})$.
3) The yield is qualified by the following constraints:

$$y_i(\boldsymbol{x}, \boldsymbol{\xi}) \leq u_i \quad \forall\, i \in [n]. \quad (9)$$

Here, $[n] = 1, \ldots, n$ and $\mathbb{E}[y_i(\boldsymbol{x}, \boldsymbol{\xi})] \leq u_i$. Each individual quantity $y_i(\boldsymbol{x}, \boldsymbol{\xi})$ is a black-box function, and we can obtain its function values at given sample points.

The design variables $\boldsymbol{x}$ are deterministic without any probability measures, and all samples of $\boldsymbol{x}$ are equally important in the optimization process. Therefore, we treat $\boldsymbol{x}$ as mutually independent random variables with a uniform distribution and use Legendre polynomials as their basis functions. The process variations $\boldsymbol{\xi}$ are non-Gaussian correlated, which enables our model to handle generic cases.

A. Probabilistic Yield Optimization Model

The yield at a given design variable $\boldsymbol{x}$ can be defined as the probability that the yield conditions (9) are satisfied, i.e.,

$$Y(\boldsymbol{x}) = \text{Prob}_{\boldsymbol{\xi}}(\boldsymbol{y}(\boldsymbol{x}, \boldsymbol{\xi}) \leq \boldsymbol{u}).$$

Here, $\boldsymbol{y}(\boldsymbol{x}, \boldsymbol{\xi}) = [y_1(\boldsymbol{x}, \boldsymbol{\xi}), \ldots, y_n(\boldsymbol{x}, \boldsymbol{\xi})]^T$ and $\boldsymbol{u} = [u_1, \ldots, u_n]^T$. Consequently, the yield optimization problem can be described as

$$\max_{\boldsymbol{x} \in \mathcal{X}} \quad \text{Prob}_{\boldsymbol{\xi}}(\boldsymbol{y}(\boldsymbol{x}, \boldsymbol{\xi}) \leq \boldsymbol{u}). \quad (10)$$

However, the above yield maximization often contradicts with our performance goals. For instance, one may have to reduce the clock rate of a processor significantly in order to achieve a high yield. As a result, directly optimizing the yield may lead to an over-conservative design. In practice, the design variables that provide the best yield may be nonunique, and we hope to chose a design that achieves good performance and high yield simultaneously. Therefore, we ensure the yield with a chance constraint

$$\text{Prob}_{\boldsymbol{\xi}}(\boldsymbol{y}(\boldsymbol{x}, \boldsymbol{\xi}) \leq \boldsymbol{u}) \geq 1 - \epsilon \quad (11)$$

and optimize the expected value of an uncertain performance metric $f(\boldsymbol{x}, \boldsymbol{\xi})$ by the following yield-aware optimization:

$$\begin{aligned} \min_{\boldsymbol{x} \in \mathcal{X}} \ & \mathbb{E}_{\boldsymbol{\xi}}\big[f(\boldsymbol{x}, \boldsymbol{\xi})\big] \\ \text{s.t.} \ & \text{Prob}_{\boldsymbol{\xi}}(\boldsymbol{y}(\boldsymbol{x}, \boldsymbol{\xi}) \leq \boldsymbol{u}) \geq 1 - \epsilon. \end{aligned} \quad (12)$$

Here, ϵ is a given risk level to control the yield. The above formulation is not equivalent to (10). It can describe, for instance, the following design optimization problem: minimize the average power consumption of a photonic IC while ensuring at least 95% yield (i.e., with 5% probability of violating timing and bandwidth constraints) under process variations. Note that $f(\boldsymbol{x}, \boldsymbol{\xi})$ may also be the function (e.g., weighted sum) of several performance metrics that we intend to optimize simultaneously. The parameter ϵ can help designers balance between the yield and a target performance goal (i.e., power consumption). A small ϵ results in a higher yield but possibly a worse performance metric. Therefore, the value of ϵ can be chosen adaptively and case-dependently by the users based on their specific requirements on the performance and yield.

Because the yield function $Y(\boldsymbol{x})$ and the objective function $f(\boldsymbol{x}, \boldsymbol{\xi})$ are not available, we have to estimate the yield and objective at a certain design variable $\boldsymbol{x}$ by the Monte Carlo (MC) method [8], [9]. This requires a huge number of simulation samples at each design variable $\boldsymbol{x}$, which is infeasible for many simulation-expensive photonic IC design problems.

Due to the ease of implementation, we reformulate the joint chance constraint in (11) into individual chance constraints

$$\text{Prob}_{\boldsymbol{\xi}}(y_i(\boldsymbol{x}, \boldsymbol{\xi}) \leq u_i) \geq 1 - \epsilon_i \quad \forall\, i \in [n]. \quad (13)$$

In this formulation, ϵ_i means the risk tolerance of violating the ith design specification. Since $\text{Prob}_{\boldsymbol{\xi}}(\mathbf{y}(\boldsymbol{x}, \boldsymbol{\xi}) \leq \mathbf{u}) =$

$\text{Prob}_{\boldsymbol{\xi}}(\cap_{i=1}^{n}(y_i(\boldsymbol{x},\boldsymbol{\xi}) \le u_i)) = 1 - \text{Prob}_{\boldsymbol{\xi}}(\cup_{i=1}^{n}(y_i(\boldsymbol{x},\boldsymbol{\xi}) \ge u_i))$, the probability of the joint chance constraint can be upper and lower bounded by the individual chance constraints

$$\max_{i=1,\ldots,n} \text{Prob}_{\boldsymbol{\xi}}(y_i(\boldsymbol{x},\boldsymbol{\xi}) \ge u_i) \le \text{Prob}_{\boldsymbol{\xi}}\left(\cup_{i=1}^{n}(y_i(\boldsymbol{x},\boldsymbol{\xi}) \ge u_i)\right) \le \sum_{i=1}^{n} \text{Prob}_{\boldsymbol{\xi}}(y_i(\boldsymbol{x},\boldsymbol{\xi}) \ge u_i).$$

When $\epsilon_i = \epsilon$ for all i, (13) is a relaxation of (11) (e.g., the feasible domain is enlarged); when $\sum_{i=1}^{n} \epsilon_i \le \epsilon$, (13) becomes more conservative than (11) (e.g., the feasible domain is reduced). In this article, we do not give the universal best choice of ϵ_i. Instead, the users can tune the parameters adaptively based on their requirements.

Consequently, we have the following chance-constrained yield-aware optimization model:

$$\begin{aligned} \min_{\boldsymbol{x}\in\mathcal{X}} \ & \mathbb{E}_{\boldsymbol{\xi}}\left[f(\boldsymbol{x},\boldsymbol{\xi})\right] \\ \text{s.t.} \ & \text{Prob}_{\boldsymbol{\xi}}(y_i(\boldsymbol{x},\boldsymbol{\xi}) \le u_i) \ge 1-\epsilon_i \quad \forall\, i \in [n]. \end{aligned} \tag{14}$$

B. Deterministic Reformulation

The chance-constrained optimization problem (14) is difficult to solve directly. This problem is more challenging when $y_i(\boldsymbol{x},\boldsymbol{\xi})$ is nonlinear because it is almost impossible to formulate the chance constraints in (14) to equivalent deterministic formulations. A naive approach is to replace the stochastic constraints by inequality constraints over the expected constraints

$$\begin{aligned} \min_{\boldsymbol{x}\in\mathcal{X}} \ & \mathbb{E}_{\boldsymbol{\xi}}\left[f(\boldsymbol{x},\boldsymbol{\xi})\right] \\ \text{s.t.} \ & \mathbb{E}_{\boldsymbol{\xi}}\left[y_i(\boldsymbol{x},\boldsymbol{\xi})\right] \le u_i \quad \forall\, i \in [n]. \end{aligned} \tag{15}$$

However, this treatment will lose the probability density information and may not provide a high-quality solution, although it can help improve the yield in practice. We will illustrate this phenomenon in numerical experiments in Section V-A.

Therefore, we do not use the formulation in (15). Instead, we adopt the second-order moment approach in [36] and [37] and replace (13) by

$$\mathbb{E}_{\boldsymbol{\xi}}\left[y_i(\boldsymbol{x},\boldsymbol{\xi})\right] + \kappa_{\epsilon_i}\sqrt{\text{var}_{\boldsymbol{\xi}}\left[y_i(\boldsymbol{x},\boldsymbol{\xi})\right]} \le u_i \quad \forall\, i \in [n]. \tag{16}$$

Here, $\kappa_{\epsilon_i} = \sqrt{(1-\epsilon_i/\epsilon_i)}$ is a scaling parameter. We present the detailed proof in Appendix A and point out the following.

1) Constraint (16) is a stronger condition than (13). In other words, each feasible point of (16) is also feasible for the chance constraint (13).
2) The parameter ϵ_i is a user-defined risk tolerance. When ϵ_i decreases, the feasible set will become smaller. However, the optimal solution may result in a higher yield.
3) When the variance $\text{var}_{\boldsymbol{\xi}}[y_i(\boldsymbol{x},\boldsymbol{\xi})]$ is small enough, the feasible set of (16) is close to the deterministic constraint $\mathbb{E}_{\boldsymbol{\xi}}[y_i(\boldsymbol{x},\boldsymbol{\xi})] \le u_i$.

Consequently, the probabilistic optimization model (14) is reformulated into a deterministic optimization problem

$$\begin{aligned} \min_{\boldsymbol{x}\in\mathcal{X}} \ & \mathbb{E}_{\boldsymbol{\xi}}\left[f(\boldsymbol{x},\boldsymbol{\xi})\right] \\ \text{s.t.} \ & \mathbb{E}_{\boldsymbol{\xi}}\left[y_i(\boldsymbol{x},\boldsymbol{\xi})\right] + \kappa_{\epsilon_i}\sqrt{\text{var}_{\boldsymbol{\xi}}\left[y_i(\boldsymbol{x},\boldsymbol{\xi})\right]} \le u_i \quad \forall\, i \in [n]. \end{aligned} \tag{17}$$

IV. Algorithm and Implementation Details

We cannot solve problem (17) directly because we do not know the mean values and variances for the black-box functions $\{y_i(\boldsymbol{x},\boldsymbol{\xi})\}_{i=1}^{n}$ and $f(\boldsymbol{x},\boldsymbol{\xi})$. A direct approach is to apply an MC method to estimate the mean values and variances for every iterate $\boldsymbol{x}$. However, this is not affordable because of the large number of numerical simulations.

In this section, we build the surrogate model for $f(\boldsymbol{x},\boldsymbol{\xi})$ and $\{y_i(\boldsymbol{x},\boldsymbol{\xi})\}_{i=1}^{n}$ by using generalized polynomial chaos [45] and our recent developed uncertainty quantification solver [28], [29]. Once the surrogate models are constructed, we can perform deterministic optimization. The main task is to build the orthogonal basis functions $\Phi_{\boldsymbol{\alpha}}(\boldsymbol{x})$ and $\Psi_{\boldsymbol{\beta}}(\boldsymbol{\xi})$, and compute the coefficients $c^{i}_{\boldsymbol{\alpha},\boldsymbol{\beta}}$ and $h_{\boldsymbol{\alpha},\boldsymbol{\beta}}$ such that

$$y_i(\boldsymbol{x},\boldsymbol{\xi}) \approx \sum_{|\boldsymbol{\alpha}|+|\boldsymbol{\beta}|=0}^{p} c^{i}_{\boldsymbol{\alpha},\boldsymbol{\beta}}\Phi_{\boldsymbol{\alpha}}(\boldsymbol{x})\Psi_{\boldsymbol{\beta}}(\boldsymbol{\xi}) \tag{18}$$

and

$$f(\boldsymbol{x},\boldsymbol{\xi}) \approx \sum_{|\boldsymbol{\alpha}|+|\boldsymbol{\beta}|=0}^{p} h_{\boldsymbol{\alpha},\boldsymbol{\beta}}\Phi_{\boldsymbol{\alpha}}(\boldsymbol{x})\Psi_{\boldsymbol{\beta}}(\boldsymbol{\xi}). \tag{19}$$

Once the above surrogate models are obtained, the mean value of $y_i(\boldsymbol{x},\boldsymbol{\xi})$ can be approximated by

$$\mathbb{E}_{\boldsymbol{\xi}}\left[y_i(\boldsymbol{x},\boldsymbol{\xi})\right] \approx \sum_{|\boldsymbol{\alpha}|=0}^{p} c^{i}_{\boldsymbol{\alpha},\mathbf{0}}\Phi_{\boldsymbol{\alpha}}(\boldsymbol{x}) \tag{20}$$

and the variance is approximated by

$$\text{var}_{\boldsymbol{\xi}}\left[y_i(\boldsymbol{x},\boldsymbol{\xi})\right] \approx \sum_{|\boldsymbol{\beta}|=1}^{p}\left(\sum_{|\boldsymbol{\alpha}|=0}^{p-|\boldsymbol{\beta}|} c^{i}_{\boldsymbol{\alpha},\boldsymbol{\beta}}\Phi_{\boldsymbol{\alpha}}(\boldsymbol{x})\right)^2. \tag{21}$$

Equation (21) is obtained based on the orthonormal property of the basis functions. The detailed proof is shown in Appendix B. The mean value of the objective function $f(\boldsymbol{x},\boldsymbol{\xi})$ can be evaluated in the same way. Finally, the deterministic yield optimization model (17) has an explicit expression and can be solved.

The overall framework is summarized in Algorithm 1. In the following, we explain the implementation details.

A. Basis Functions for Design and Uncertainty Variables

For the mutually independent uniform-distributed design variable $\boldsymbol{x}$, their basis functions $\Phi_{\boldsymbol{\alpha}}(\boldsymbol{x})$ can be decoupled into the products of 1-D basis functions

$$\Phi_{\boldsymbol{\alpha}}(\boldsymbol{x}) = \phi^{1}_{\alpha_1}(x_1)\ldots\phi^{d_1}_{\alpha_{d_1}}\left(x_{d_1}\right). \tag{22}$$

Here, $\phi^{i}_{\alpha_i}(x_i)$ is a Legendre polynomial [45] and can be constructed by the three-term recurrence relation [38].

For the random vector $\boldsymbol{\xi}$ describing non-Gaussian correlated process variations, we construct its basis functions $\Psi_{\boldsymbol{\beta}}(\boldsymbol{\xi})$ by the Gram–Schmidt approach proposed in [28] and [29]. Specifically, we first reorder the monomials $\boldsymbol{\xi}^{\boldsymbol{\beta}} = \xi_1^{\beta_1}\ldots\xi_{d_2}^{\beta_{d_2}}$ in the graded lexicographic order, and denote them as $\{p_j(\boldsymbol{\xi})\}_{j=1}^{N_p}$. Here, $N_p^{d_1} = \binom{d_2+p}{p}$ is the total number of basis functions for $\boldsymbol{\xi} \in \mathbb{R}^{d_2}$ bounded by order p. Then we set $\Psi_1(\boldsymbol{\xi}) = 1$ and generate the orthonormal polynomials

Algorithm 1: Our Proposed Chance-Constrained Yield-Aware Optimization Solver

Input: The range of the design variable $\boldsymbol{x}$, PDF of the non-Gaussian correlated random parameters $\rho(\boldsymbol{\xi})$, the polynomial order p, the upper bounds of performance metrics $\{u_i\}_{i=1}^n$, and the chance constraint thresholds $\{\epsilon_i\}_{i=1}^n$.

1. Construct the basis functions $\Phi_{\boldsymbol{\alpha}}(\boldsymbol{x})$ and $\Psi_{\boldsymbol{\beta}}(\boldsymbol{\xi})$ based on (22) and (23) independently.
2. Initialize the quadrature points for design variables $\{\boldsymbol{x}_l, v_l\}_{l=1}^{M_1}$ by (26), and quadrature points for stochastic parameters $\{\boldsymbol{\xi}_l, u_l\}_{l=1}^{M_2}$ by the optimization problem (27), respectively. Then co-optimize the quadrature rule to obtain $\{\boldsymbol{x}_k, \boldsymbol{\xi}_k, w_k\}_{k=1}^{M}$ by (28).
3. Call the simulator to compute $f(\boldsymbol{x}_k, \boldsymbol{\xi}_k)$, $y_i(\boldsymbol{x}_k, \boldsymbol{\xi}_k)$ for all $i = 1, \ldots, n$ and $k = 1, \ldots, M$.
4. Build the coefficients $h_{\boldsymbol{\alpha},\boldsymbol{\beta}}$ and $c^i_{\boldsymbol{\alpha},\boldsymbol{\beta}}$ by equation (25).
5. Set up the optimization problem (31), and then solve it via a global polynomial optimization solver, e.g., [51].

Output: The optimized design variable $\boldsymbol{x}^*$

$\{\Psi_j(\boldsymbol{\xi})\}_{j=2}^{N_p}$ in the correlated parameter space recursively by

$$\hat{\Psi}_j(\boldsymbol{\xi}) = p_j(\boldsymbol{\xi}) - \sum_{i=1}^{j-1} \mathbb{E}\big[p_j(\boldsymbol{\xi})\Psi_i(\boldsymbol{\xi})\big]\Psi_i(\boldsymbol{\xi})$$

$$\Psi_j(\boldsymbol{\xi}) = \frac{\hat{\Psi}_j(\boldsymbol{\xi})}{\sqrt{\mathbb{E}\Big[\hat{\Psi}_j^2(\boldsymbol{\xi})\Big]}}, \quad j = 2, \ldots, N_p. \tag{23}$$

These basis functions $\{\Psi_j(\boldsymbol{\xi})\}_{j=1}^{N_p}$ can be reordered into $\{\Psi_{\boldsymbol{\beta}}(\boldsymbol{\xi})\}_{|\boldsymbol{\beta}|=0}^{p}$.

B. How to Compute the Coefficients?

By a projection approach, the coefficient $c^i_{\boldsymbol{\alpha},\boldsymbol{\beta}}$ for the basis function can be computed by

$$c^i_{\boldsymbol{\alpha},\boldsymbol{\beta}} = \mathbb{E}_{\boldsymbol{x},\boldsymbol{\xi}}\big[y_i(\boldsymbol{x},\boldsymbol{\xi})\Phi_{\boldsymbol{\alpha}}(\boldsymbol{x})\Psi_{\boldsymbol{\beta}}(\boldsymbol{\xi})\big]. \tag{24}$$

The above integration can be well computed given a suitable set of quadrature points $\{\boldsymbol{x}_k, \boldsymbol{\xi}_k\}_{k=1}^{M}$ and weights $\{w_k\}_{k=1}^{M}$

$$c^i_{\boldsymbol{\alpha},\boldsymbol{\beta}} \approx \sum_{k=1}^{M} y_i\big(\boldsymbol{x}_k, \boldsymbol{\xi}_k\big)\Phi_{\boldsymbol{\alpha}}(\boldsymbol{x}_k)\Psi_{\boldsymbol{\beta}}(\boldsymbol{\xi}_k)w_k. \tag{25}$$

We need to design a proper quadrature rule. The main challenge here is that $\boldsymbol{x}$ is an independent vector but $\boldsymbol{\xi}$ describes non-Gaussian correlated uncertainties.

In this article, we propose a three-stage optimization method to compute the quadrature rule.

1) We compute the quadrature rule $\{\boldsymbol{x}_l, v_l\}_{l=1}^{M_1}$ for the independent design variable $\boldsymbol{x}$.
2) We employ the optimization approach proposed in [28] and [29] to calculate the quadrature points and weights $\{\boldsymbol{\xi}_l, u_l\}_{l=1}^{M_2}$ for the non-Gaussian correlated parameter $\boldsymbol{\xi}$.
3) We use their tensor products (M_1M_2 points) as an initialization and call the optimization approach proposed in [28] and [29] for the coupled space of $\boldsymbol{x}$ and $\boldsymbol{\xi}$ to compute $M \le M_1M_2$ joint quadrature points and weights $\{\boldsymbol{x}_k, \boldsymbol{\xi}_k, w_k\}_{k=1}^{M}$.

The details are described as follows.

1) Initial Quadrature Rule for $\boldsymbol{x}$: One could employ the sparse grid approach [46], [47] to compute the quadrature samples and weights for the independent uniform random variables $\boldsymbol{x} \in \mathbb{R}^{d_1}$. However, the quadrature weights from a sparse grid method can be negative, and the number of quadrature points is not small enough. Therefore, after obtaining the sparse-grid quadrature rule, we refine the quadrature rule by the least square optimization solver

$$\min_{\mathbf{a}\le\boldsymbol{x}_l\le\mathbf{b}, v_l\ge 0} \sum_{j=1}^{N_{2p}^{d_1}} \left(\mathbb{E}\big[\Phi_j(\boldsymbol{x})\big] - \sum_{l=1}^{M_1}\Phi_j(\boldsymbol{x}_l)v_l\right)^2. \tag{26}$$

Here, the expectations $\mathbb{E}[\Phi_j(\boldsymbol{x})] = \delta_{1j}$ are already known from the orthogonality of basis functions, and $N_{2p}^{d_1} = \binom{d_1+2p}{2p}$. This model is similar to that of [28] and [29], which provides the quadrature points and weights to compute the numerical integral of all basis functions upper bounded by order $2p$. If the optimized objective in (26) is small, the numerical integral of any functions in the pth order polynomial space will also be accurate. Further, the number of points M_1 can also be updated adaptively. The theoretical proofs for the number of quadrature points and the numerical approximation error are provided in [29].

2) Initial Quadrature Points for $\boldsymbol{\xi}$: For the non-Gaussian correlated parameters $\boldsymbol{\xi}$, we adopt the optimization-based quadrature rule in [28] and [29]. Specifically, we compute M_2 quadrature points $\boldsymbol{\xi}_l$ and weights w_l via solving the following optimization problem:

$$\min_{\boldsymbol{\xi}_l, u_l\ge 0} \sum_{j=1}^{N_{2p}^{d_2}} \left(\mathbb{E}\big[\Psi_j(\boldsymbol{\xi})\big] - \sum_{l=1}^{M_2}\Psi_j\big(\boldsymbol{\xi}_l\big)u_l\right)^2. \tag{27}$$

3) Optimized Joint Quadrature Points for $\boldsymbol{x}$ and $\boldsymbol{\xi}$: The tensor product of the two sets of quadrature points $\{\boldsymbol{x}_l, v_l\}_{l=1}^{M_1}$ and $\{\boldsymbol{\xi}_l, u_l\}_{l=1}^{M_2}$ result in M_1M_2 simulation points in total, which may be still unaffordable for large-scale photonic design problems. We propose an optimization model to compute the joint quadrature rule for both the design variables $\boldsymbol{x}$ and the uncertain parameters $\boldsymbol{\xi}$ to further reduce the simulation cost of building surrogate models

$$\min_{\substack{\mathbf{a}\le\boldsymbol{x}_k\le\mathbf{b}\\ \boldsymbol{\xi}_k, w_k\ge 0}} \sum_{j_1=1}^{N_{2p}^{d}} \sum_{j_2=1}^{N_{2p}^{d}-j_1} \left(\delta_{1j_1}\delta_{1j_2} - \sum_{k=1}^{M}\Phi_{j_1}(\boldsymbol{x}_k)\Psi_{j_2}\big(\boldsymbol{\xi}_k\big)w_k\right)^2. \tag{28}$$

Here $\delta_{1j_1}\delta_{1j_2} = 1$ if $j_1 = j_2 = 1$ and zero otherwise, and $d = d_1 + d_2$. Our numerical experiments show that the total number of optimized quadrature points is M is significantly smaller than M_1M_2.

Remark: Problem (28) is a nonconvex optimization and is hard to optimize in general. The subproblems (26) and (27) help to provide a good initial guess for the joint optimization.

We use the block coordinate descent optimization method described in [29] to solve all optimization subproblems (26)–(28). The following theorem ensures high accuracy for our surrogate model considering the unavoidable numerical optimization error and function approximation error.

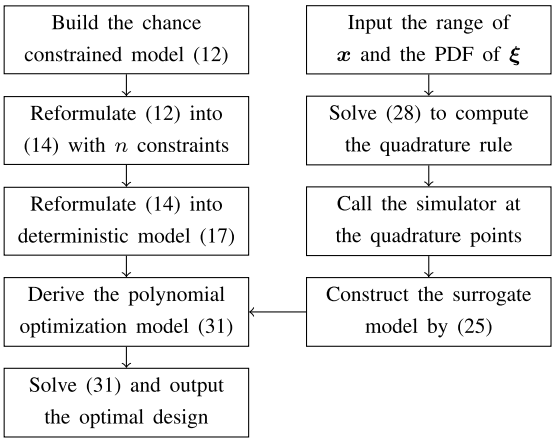


Fig. 1. Flowchart of our proposed framework for solving the chance constrained yield-aware optimization.

Theorem 1 [29]: Assume that $\{\boldsymbol{x}_k, \boldsymbol{\xi}_k, w_k\}_{k=1}^M$ are the numerical solution to (28).

1) Suppose that the objective function of (28) decays to zero. The required number of quadrature points is upper and lower bounded by

$$N_p^d = \frac{(d+p)!}{p!d!} \le M \le N_{2p}^d = \frac{(d+2p)!}{(2p)!d!}. \tag{29}$$

2) For any smooth and square-integrable function $y(\boldsymbol{\xi})$, the approximation error of its pth order stochastic approximation $\tilde{y}(\boldsymbol{\xi})$ satisfies

$$\|y(\boldsymbol{x},\boldsymbol{\xi}) - \tilde{y}(\boldsymbol{x},\boldsymbol{\xi})\|_2 \le \alpha_1\delta_1 + \alpha_2\delta_2. \tag{30}$$

Here, $\tilde{y}(\boldsymbol{x},\boldsymbol{\xi}) = \sum_{|\boldsymbol{\alpha}|+|\boldsymbol{\beta}|=0}^{p} c_{\boldsymbol{\alpha},\boldsymbol{\beta}}\Phi_{\boldsymbol{\alpha}}(\boldsymbol{x})\Psi_{\boldsymbol{\beta}}(\boldsymbol{\xi})$, δ_1 is the ℓ_1-norm of the objective function of (28) evaluated at its final numerical solution, δ_2 is the distance of $y(\boldsymbol{x},\boldsymbol{\xi})$ to the pth order polynomial space, $\alpha_1 = N_pLT$, $\alpha_2 = 1 + N_pW$, $L = \max\|y(\boldsymbol{x},\boldsymbol{\xi})\|_2$, $T = \max_{j_1+j_2,l_1+l_2=1,\ldots,N_{2p}}\|\Phi_{j_1}(\boldsymbol{x})\Psi_{j_2}(\boldsymbol{\xi})\Psi_{l_1}(\boldsymbol{x})\Psi_{l_2}(\boldsymbol{\xi})\|_2$, and $W = \sup[(|\mathbb{I}[y(\boldsymbol{\xi})]|)/(\mathbb{E}[|y(\boldsymbol{\xi})|])]$ are constants.

Remark: This section focuses on the theory and implementation for building a surrogate model for low-dimensional problems. For high-dimensional problems that are more costly in both computing the quadrature rule and difficult in reducing the number of samples, we may apply a high-dimensional solver such as the compressive sensing [43] to build the surrogate model. Our framework in Fig. 1 is still applicable.

C. Proposed Polynomial Optimization

With the formula for the mean value (20) and the variance (21), we obtain the following deterministic formula for the chance-constrained optimization:

$$\begin{aligned}\min_{\boldsymbol{x}\in\mathcal{X}} \ & \sum_{|\boldsymbol{\alpha}|=0}^{p} h_{\boldsymbol{\alpha},\mathbf{0}}\Phi_{\boldsymbol{\alpha}}(\boldsymbol{x}) \\ \text{s.t.} \ & \kappa_{\epsilon_i}\sqrt{\sum_{|\boldsymbol{\beta}|=1}^{p}\left(\sum_{|\boldsymbol{\alpha}|=0}^{p-|\boldsymbol{\beta}|} c_{\boldsymbol{\alpha},\boldsymbol{\beta}}^i\Phi_{\boldsymbol{\alpha}}(\boldsymbol{x})\right)^2} \\ & + \sum_{|\boldsymbol{\alpha}|=0}^{p} c_{\boldsymbol{\alpha},\mathbf{0}}^i\Phi_{\boldsymbol{\alpha}}(\boldsymbol{x}) \le u_i \quad \forall\, i\in[n].\end{aligned} \tag{32}$$

However, the constraints are nonsmooth because of the square-root terms, and may not admit a gradient at some points [48]. Instead, we use the equivalent smooth polynomial formula

$$\kappa_{\epsilon_i}^2\mathrm{var}_{\boldsymbol{\xi}}\big[y_i(\boldsymbol{x},\boldsymbol{\xi})\big] \le \big(u_i - \mathbb{E}_{\boldsymbol{\xi}}\big[y_i(\boldsymbol{x},\boldsymbol{\xi})\big]\big)^2. \tag{33}$$

Consequently, (17) can be reduced to a deterministic and smooth optimization problem of $\boldsymbol{x}$ in (31), as shown at the bottom of this page.

Noting that both the objective function and the constraints of (31) are polynomials, we can obtain the optimal solution by using any polynomial solvers. In this article, we use the semi-definite relaxation-based approaches [49], [50] because they can find the global optimal solution.

V. Numerical Experiments

In this section, we verify our proposed approach by a synthetic example and two photonic IC examples. The p subproblem (31) is solved by the global optimization solver GloptiPoly 3 [51]. For a design variable $\boldsymbol{x}$, we generate M parameters $\boldsymbol{\xi}_j$ and approximate the yield by

$$\text{yield}(\boldsymbol{x}) = \frac{\text{the number of } \boldsymbol{\xi}_j \text{ such that } y_i(\boldsymbol{x},\boldsymbol{\xi}_j) \le u_i}{M}. \tag{34}$$

We set all risk thresholds to ϵ. For the synthetic example, we will compare our method with the deterministic formulation (15). For the photonic IC examples, we will compare our method with the BYO method [10]. We summarize the key idea of the BYO in Appendix C. The MATLAB codes and a demo example can be downloaded online.[1]

A. Synthetic Example

First, we consider a synthetic example with two design variables and two non-Gaussian correlated random parameters. The design variable $\boldsymbol{x}$ admits a uniform distribution $\mathcal{U}[-1,1]^2$

[1] https://web.ece.ucsb.edu/∼zhengzhang/codes_dataFiles/ccyopt/

$$\begin{aligned}\min_{\boldsymbol{x}\in\mathcal{X}} \ & \sum_{|\boldsymbol{\alpha}|=0}^{p} h_{\boldsymbol{\alpha},\mathbf{0}}\Phi_{\boldsymbol{\alpha}}(\boldsymbol{x}) \\ \text{s.t.} \ & \kappa_{\epsilon_i}^2\sum_{|\boldsymbol{\beta}|=1}^{p}\left(\sum_{|\boldsymbol{\alpha}|=0}^{p-|\boldsymbol{\beta}|} c_{\boldsymbol{\alpha},\boldsymbol{\beta}}^i\Phi_{\boldsymbol{\alpha}}(\boldsymbol{x})\right)^2 \le \left(u_i - \sum_{|\boldsymbol{\alpha}|=0}^{p} c_{\boldsymbol{\alpha},\mathbf{0}}^i\Phi_{\boldsymbol{\alpha}}(\boldsymbol{x})\right)^2, \quad \sum_{|\boldsymbol{\alpha}|=0}^{p} c_{\boldsymbol{\alpha},\mathbf{0}}^i\Phi_{\boldsymbol{\alpha}}(\boldsymbol{x}) \le u_i \quad \forall\, i\in[n]\end{aligned} \tag{31}$$

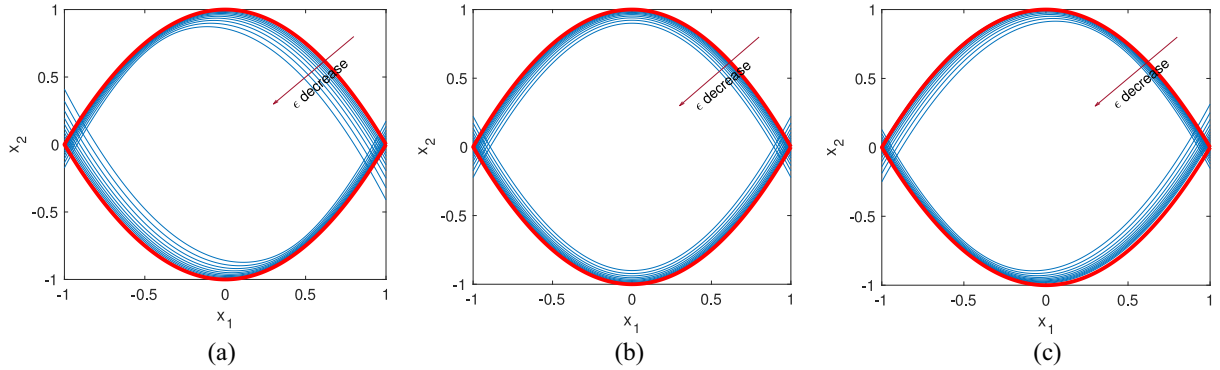


Fig. 2. Feasible set of the synthetic example with risk tolerance levels $\epsilon \in [10^{-2}, 10^{-0.1}]$ under different uncertainty distributions. (a) Positive-correlated non-Gaussian distribution. (b) Gaussian independent distribution. (c) Negative correlated non-Gaussian distribution. The domain between the red lines are the deterministic feasible set $x_1^2 \pm x_2 \leq 1$, and the blue lines demonstrate the effects of chance constraints.

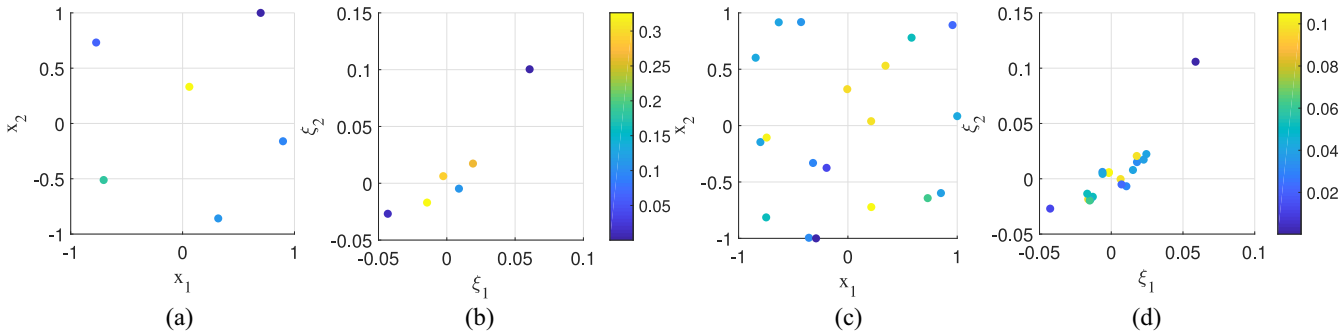


Fig. 3. Quadrature points and weights in the synthetic experiment. (a) and (b) Initial 2-D quadrature points for the design variable $\boldsymbol{x}$ and uncertain parameters $\boldsymbol{\xi}$ by solving (26) and (27), respectively. (c) and (d) Optimized quadrature points for the joint 4-D space of $\boldsymbol{x}$ and $\boldsymbol{\xi}$ by solving (28). Here, we project the optimized 4-D quadrature points to the 2-D subspace of $\boldsymbol{x}$ and $\boldsymbol{\xi}$, respectively. The quadrature weights are shown in colors.

and the uncertain parameter $\boldsymbol{\xi}$ follows a Gaussian mixture distribution. We define the yield criterion as $(x_1+\xi_1)^2 \pm (x_2+\xi_2) \leq 1$ and our goal is to maximize $\mathbb{E}_{\boldsymbol{\xi}}[3(x_1+\xi_1)+(x_2+\xi_2)]$. We formulate the yield into chance constraints and derive the following problem:

$$\begin{aligned} \max_{\boldsymbol{x}} \quad & \mathbb{E}_{\boldsymbol{\xi}}[3(x_1+\xi_1)-(x_2+\xi_2)] \\ \text{s.t.} \quad & \text{Prob}_{\boldsymbol{\xi}}\Big((x_1+\xi_1)^2-(x_2+\xi_2)\leq 1\Big) \geq 1-\epsilon \\ & \text{Prob}_{\boldsymbol{\xi}}\Big((x_1+\xi_1)^2+(x_2+\xi_2)\leq 1\Big) \geq 1-\epsilon. \quad (35) \end{aligned}$$

To illustrate the effects of different parameter distributions, we study three PDFs: 1) the independent distribution $\mathcal{N}(\mathbf{0}, 10^{-4}\mathbf{I})$; 2) the non-Gaussian positive correlations $(1/2)\mathcal{N}(\mathbf{0.01}, 10^{-4}\boldsymbol{\Sigma}) + (1/2)\mathcal{N}(-\mathbf{0.01}, 10^{-4}\boldsymbol{\Sigma})$ with $\boldsymbol{\Sigma} = \begin{pmatrix} 1 & 0.75 \\ 0.75 & 1 \end{pmatrix}$; and 3) the non-Gaussian negative correlations $(1/2)\mathcal{N}([0.01, -0.01]^T, 10^{-4}\boldsymbol{\Sigma}) + (1/2)\mathcal{N}([-0.01, 0.01]^T, 10^{-4}\boldsymbol{\Sigma})$ with $\boldsymbol{\Sigma} = \begin{pmatrix} 1 & -0.75 \\ -0.75 & 1 \end{pmatrix}$. The feasible sets under three probability density distributions are shown in Fig. 2. The comparison clearly shows that the effects of different uncertainties. For all three density functions, the feasible regions are reduced when the risk level ϵ decreases.

Next, we take the non-Gaussian positive correlated distribution as an example to compute the optimal solution of (35). We first build the surrogate models for both the objective and constraints by the second-order polynomial basis functions. The optimized quadrature points $\{\boldsymbol{x}_l, v_l\}_{l=1}^6$ for the design variables by (26) and $\{\boldsymbol{\xi}_l, u_l\}_{l=1}^6$ for the random parameter by (27) are shown in Fig. 3(a) and (b), respectively. Directly tensorizing the two sets of quadrature points generates 36 samples. We further solve (28) to reduce them to $M = 19$ optimized samples and weights. According to Theorem 1, the number of quadrature samples for $d = 4, p = 2$ should be in the range $[15, 70]$. Our optimization algorithm obtains $M = 19$, which is close to the theoretical lower bound.

TABLE I
OPTIMAL SOLUTION FOR THE SYNTHETIC EXPERIMENT UNDER DIFFERENCE RISK THRESHOLD ϵ

Algorithm	$\boldsymbol{x}^*$		Objective	Yield (%)
Proposed ($\epsilon = 0.01$)	0.8630	-0.1172	2.4717	100
Proposed ($\epsilon = 0.05$)	0.9379	-0.0522	2.7616	100
Proposed ($\epsilon = 0.10$)	0.9587	-0.0402	2.8360	99.42
Proposed ($\epsilon = 0.15$)	0.9689	-0.0351	2.8717	93.84
Proposed (ϵ =0.20)	0.9751	-0.0293	2.8959	87.49
(15)	0.9999	0	2.9997	41.66

We further show the results for different risk tolerance levels ϵ in Table I. A smaller ϵ results in a smaller feasible domain (as shown in Fig. 2), and generates a higher yield but a smaller objective value. In practice, ϵ can be chosen case-by-case based on the tradeoff between the performance

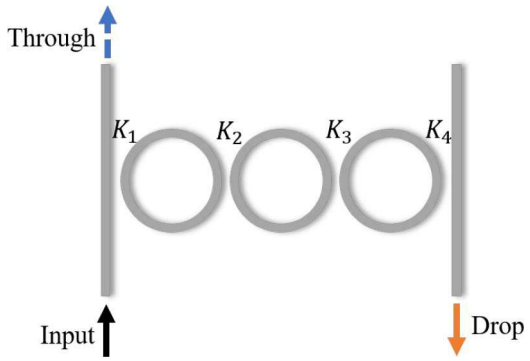


Fig. 4. Optical add-drop filter with three microrings coupled in series.

and yield requirements. Compared with the solution $\tilde{\boldsymbol{x}} = [0.9999, 0]^T$ from solving (15), our method can achieve a significantly higher yield: our optimized yield is above 87% while solving (15) only leads to a yield of 41.66%.

B. Microring Add-Drop Filter

We continue to consider the design of an optical add-drop filter consisting of three identical silicon microrings coupled in series, as shown in Fig. 4. In designing such a broadband optical filter, the coupling coefficients play an important role in determining key performance metrics, such as the bandwidth and extinction ratio [52], [53]. A broad and flat passband with a high extinction ratio can be achieved by optimizing the coupling strengths between the microrings [52]. In this example, we employ silicon as the waveguide material and assume the effective refractive index to be $n_{\text{eff}} = 2.44$ and the effective group index to be $n_g = 4.19$ near the wavelength of 1.55 μm. The design variables are the coupling coefficients $\boldsymbol{x} = [K_1, K_2, K_3, K_4]$ that are to be optimized within the interval of $[0.3, 0.6]$. The random variables are set as small deviations of the coupling coefficients. We assume that $\boldsymbol{\xi}$ follows a non-Gaussian correlated distribution:

$$p(\boldsymbol{\xi}) = \frac{1}{2}\mathcal{N}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}) + \frac{1}{2}\mathcal{N}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}) \tag{36}$$

where $\boldsymbol{\mu}_1 = -\boldsymbol{\mu}_2 = 0.006[1, 1, 1, 1]^T$, and the variance is defined as

$$\boldsymbol{\Sigma} = 0.006^2 \begin{bmatrix} 1 & 0.4 & 0.1 & 0.4 \\ 0.4 & 1 & 0.4 & 0.1 \\ 0.1 & 0.4 & 1 & 0.4 \\ 0.4 & 0.1 & 0.4 & 1 \end{bmatrix}.$$

We mainly focus on three metrics of the microring filter: 1) the 3dB bandwidth (BW, in GHz); 2) the extinction ratio (RE, in dB) of the transmission at the drop port; and 3) the roughness (σ_{pass}, in dB) of the passband that takes a standard deviation of the passband. The yield-aware optimization problem of the microring filter design can be formulated as

$$\begin{aligned} \max_{\boldsymbol{x}\in\mathcal{X}} \quad & \mathbb{E}_{\boldsymbol{\xi}}\big[\text{BW}(\boldsymbol{x}, \boldsymbol{\xi})\big] \\ \text{s.t.} \quad & \text{Prob}_{\boldsymbol{\xi}}(\text{RE}(\boldsymbol{x}, \boldsymbol{\xi}) \geq \text{RE}_0) \geq 1 - \epsilon \\ & \text{Prob}_{\boldsymbol{\xi}}\big(\sigma_{\text{pass}}(\boldsymbol{x}, \boldsymbol{\xi}) \leq \sigma_0\big) \geq 1 - \epsilon \end{aligned} \tag{37}$$

TABLE II
OPTIMIZATION RESULTS FOR THE MICRORING ADD-DROP FILTER

Algorithm	Simulations	$\mathbb{E}_{\boldsymbol{\xi}}$[BW] (GHz)	Yield (%)
Proposed ($\epsilon = 0.03$)	64	113.4	100
Proposed ($\epsilon = 0.05$)	64	115.6	99.8
Proposed ($\epsilon = 0.07$)	64	117.2	99.5
Proposed ($\epsilon = 0.10$)	64	118.4	98.1
BYO [10]	2020	112.3	99.8

where the yield is defined via some chance constraints on the extinction ratio and the roughness of the passband. In our simulation, the threshold extinction ratio (RE_0) and the roughness of the passband (σ_0) are 25 dB and 0.5 dB, respectively.

We first build the second-order polynomial surrogate model by our proposed Algorithm 1. We only need 17 initial quadrature points for the variable $\boldsymbol{x}$ by solving (26), 16 quadrature points for the parameters $\boldsymbol{\xi}$ by solving (27), and 64 quadrature points for the joint optimization of $\boldsymbol{x}$ and $\boldsymbol{\xi}$ by solving (28). Fig. 5 shows that our surrogate model can well approximate the probabilistic distributions of the performance metrics with the comparison of 10^3 MC simulations, although our method only needs 64 simulation samples for this example.

We summarize the results of our proposed method with different choices of ϵ and the results obtained by the BYO in Table II. It shows that when risk tolerance level ϵ decreases, our proposed method can achieve higher yield and lower bandwidth. This is corresponding to our theory that a lower risk level ϵ results in a smaller feasible region. Our proposed method can always achieve a large bandwidth because it computes the global optimal solution of the polynomial optimization problem. When $\epsilon = 0.05$, we get a bandwidth $\mathbb{E}_{\boldsymbol{\xi}}[\text{BW}] = 115.6$ GHz with 99.8% yield at the optimal solution $\boldsymbol{x}^* = [0.5582, 0.4208, 0.3000, 0.6000]$, while BYO takes 2020 simulations to achieve the result of $\mathbb{E}_{\boldsymbol{\xi}}[\text{BW}] = 112.3$ GHz with the yield 99.8%. Fig. 6 compares the frequency response before and after the yield-aware optimization. Both our proposed method and BYO can achieve a higher bandwidth with a smoother passband compared to the design before optimization. In Fig. 7, we plot the probability density of the bandwidth at the optimal design by our chance-constrained optimization with $\epsilon = 0.05$ and by the BYO, respectively. It clearly shows that our proposed method can increase the bandwidth while achieving the same yield.

C. Mach–Zehnder Interferometer

We apply the same framework to optimize a third-order Mach–Zehnder interferometer (MZI) which consists of three port coupling and two arms, as shown in Fig. 8. The coupling coefficients between the MZ arms play the most important role in the design. The relationship between the coupling coefficient κ and the gap g (nm) is

$$\kappa = \exp\Big(-\frac{g}{260}\Big). \tag{38}$$

In this experiment, the design variables $\boldsymbol{x} = [g_1, g_2, g_3]$ are optimized in the interval of [100 nm, 300 nm]^3. The random variable $\boldsymbol{\xi}$ follows the Gaussian mixture distribution:

$$p(\boldsymbol{\xi}) = \frac{1}{2}\mathcal{N}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}) + \frac{1}{2}\mathcal{N}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}) \tag{39}$$

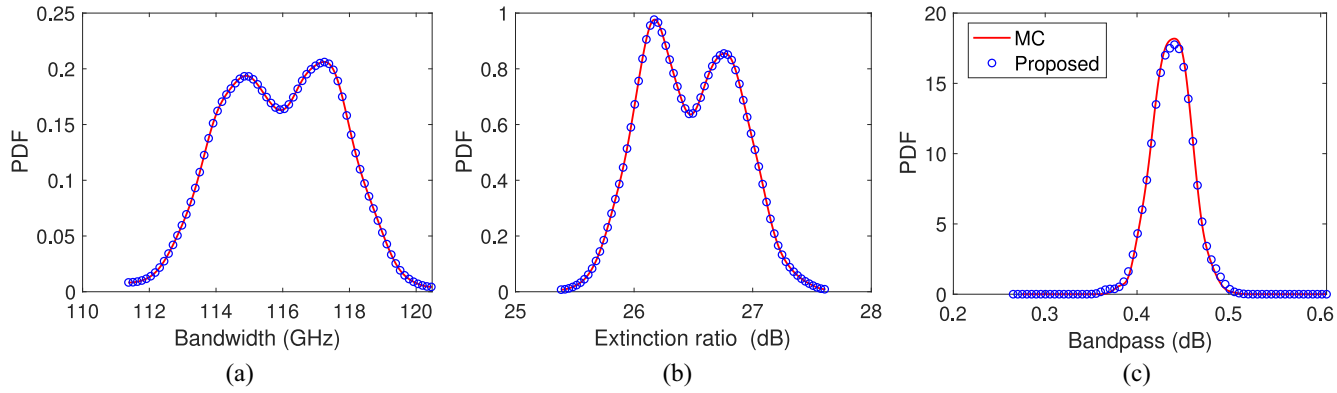


Fig. 5. PDFs of the bandwidth, extinction ratio, and roughness for the microring add-drop filter at the optimal solution $\boldsymbol{x}^* = [0.5582, 0.4208, 0.3000, 0.6000]$ by our proposed optimization with $\epsilon = 0.05$. Our surrogate model uses only 64 simulations, and MC uses 10^3 simulations.

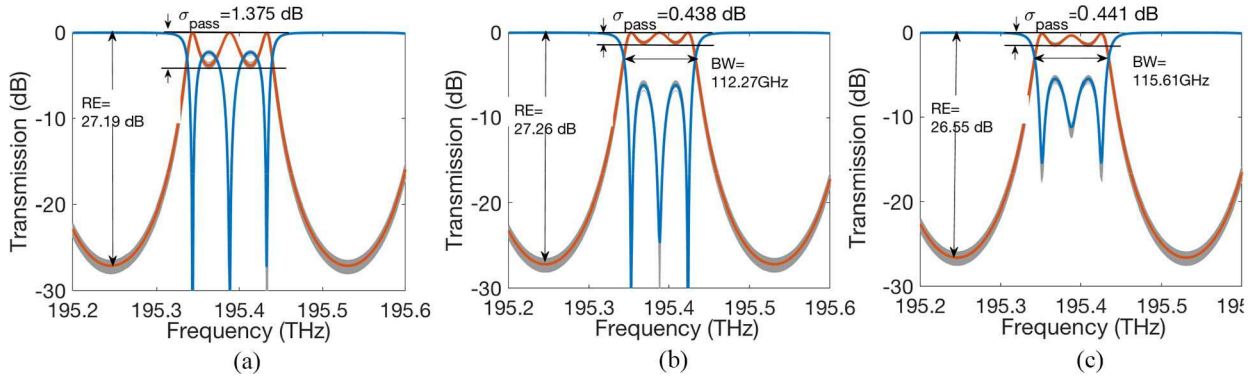


Fig. 6. Transmission curves of the microring add-drop filter at different design choices. The gray lines show the uncertainties caused by the process variations. The orange and blue curves show the mean transmission rates at the drop port and the through port, respectively. Here, RE, BW, and σ_{pass} denote the mean values of extinction ratio, bandwidth, and roughness, respectively. (a) Transmission at $\boldsymbol{x}^0 = [0.45, 0.45, 0.45, 0.45]$ without any optimization. It does not have a clear passband because σ_{pass} is too large. (b) Results after the BYO. (c) Results obtained from our chance-constrained optimization with $\epsilon = 0.05$.

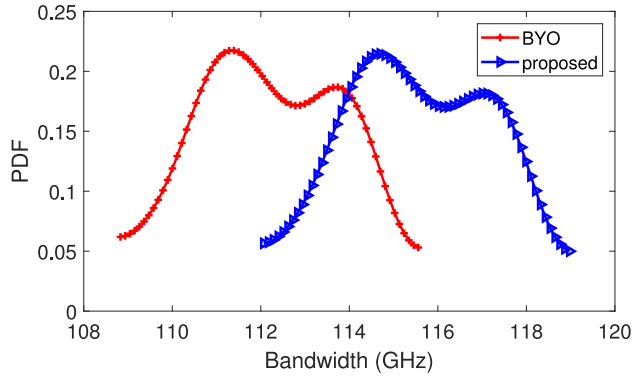


Fig. 7. Optimized bandwidth probability density distribution of the microring filter. Our chance-constrained optimization obtain an expected value of 115.6 GHz while the BYO only produces an expected value of 112.3 GHz.

where $\boldsymbol{\mu}_1 = -\boldsymbol{\mu}_2 = [1, 1, 1]^T$, and

$$\boldsymbol{\Sigma} = \begin{bmatrix} 1 & 0.4 & 0.1 \\ 0.4 & 1 & 0.4 \\ 0.1 & 0.4 & 1 \end{bmatrix}.$$

We consider three performance metrics of the MZI: 1) the 3 dB bandwidth (BW, in GHz); 2) the crosstalk (XT, in dB); and 3) the attenuation (α, in dB) of the peak transmission. The yield is defined through the crosstalk and the attenuation. The

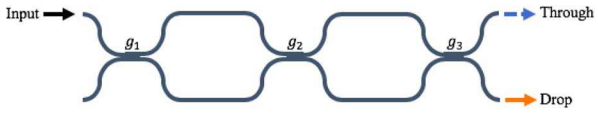


Fig. 8. Schematic of a third-order MZI.

yield-aware optimization is formulated as

$$\begin{aligned} \max_{\boldsymbol{x}} \quad & \mathbb{E}_{\boldsymbol{\xi}}\big[\text{BW}(\boldsymbol{x}, \boldsymbol{\xi})\big] \\ \text{s.t.} \quad & \text{Prob}_{\boldsymbol{\xi}}\big(\text{XT}(\boldsymbol{K}, \boldsymbol{\xi}) \le \text{XT}_0\big) \ge 1 - \epsilon \\ & \text{Prob}_{\boldsymbol{\xi}}\big(\alpha(\boldsymbol{x}, \boldsymbol{\xi}) \le \alpha_0\big) \ge 1 - \epsilon \end{aligned} \tag{40}$$

where the yield risk level is ϵ. The threshold crosstalk (XT_0) and attenuation (α_0) are -4 dB and 2 dB, respectively.

We first build three second-order polynomial surrogate models for BW, XT, and α by our proposed Algorithm 1. We generate 11 initial quadrature points for the design variable $\boldsymbol{x}$, 10 initial quadrature points for the uncertainty parameter $\boldsymbol{\xi}$. Then we apply the tensor product of those 110 points to problem (28) and eventually get 36 quadrature points for the joint space after co-optimization. Fig. 9 shows that our surrogate models constructed with 36 quadrature points can well approximate the density functions of all three performance metrics compared with MC with 10^3 samples.

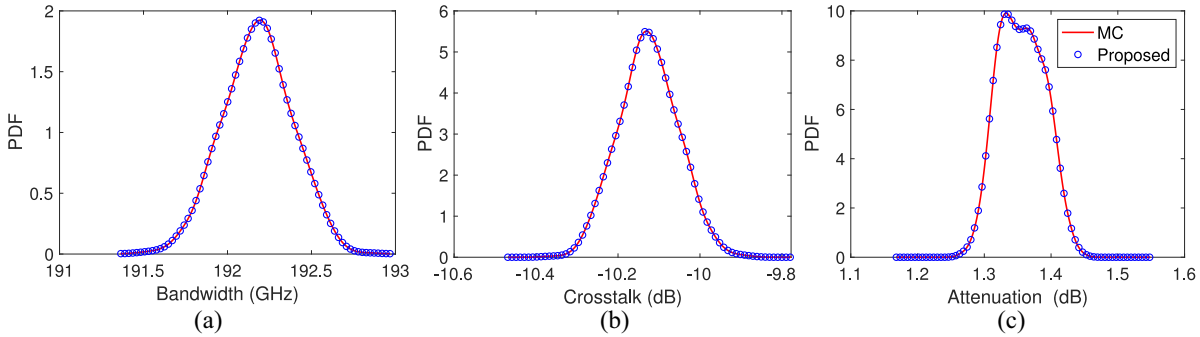


Fig. 9. PDFs for the bandwidth, crosstalk, and attenuation of the MZI at our optimized design parameters $\boldsymbol{x}^* = [0.300, 0.5036, 0.300]$. Our surrogate model uses only 36 simulations and MC uses 1000 simulations.

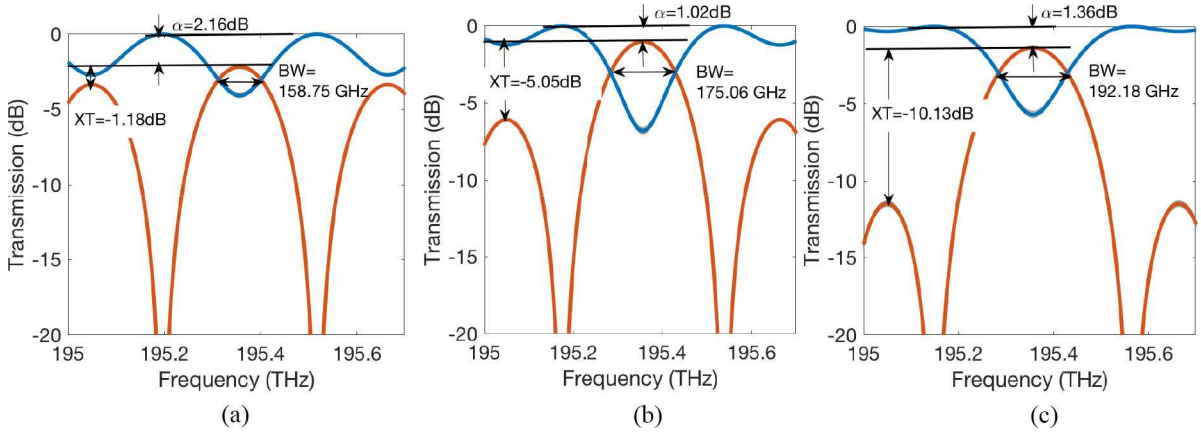


Fig. 10. Transmission curves of the MZI. The gray lines show the performance uncertainties. The orange and blue curves show the transmission rates at the drop and through ports, respectively. The mean values of the bandwidth, crosstalk, and attenuation are denoted as BW, XT, and α, respectively. (a) Initial design $\boldsymbol{x}^0 = [150, 150, 150]$. (b) Design after BYO. (c) Design with the proposed chance-constrained yield optimization.

TABLE III
OPTIMIZATION RESULT FOR THE MZI

Algorithm	Simulations	$\mathbb{E}_{\boldsymbol{\xi}}$[BW] (GHz)	Yield (%)
Proposed ($\epsilon = 0.03$)	36	188.8	100
Proposed ($\epsilon = 0.05$)	36	192.2	100
Proposed ($\epsilon = 0.07$)	36	194.5	100
Proposed ($\epsilon = 0.10$)	36	195.0	87.7
BYO [10]	2020	175.0	100

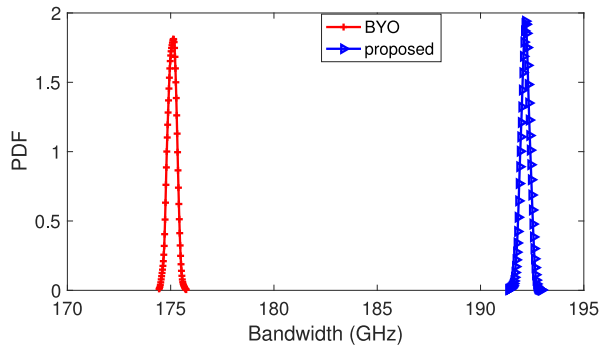


Fig. 11. Optimized bandwidth of the MZI by the BYO and our proposed method, respectively. The expectation bandwidth of the BYO is 175.4 GHz while our proposed method with $\epsilon = 0.05$ can get 186.4 GHz.

We also compare our proposed method and BYO in Table III. Similar to the result in Table II, a lower risk tolerance results in higher yield and a lower expected value of bandwidth. Our method requires 56× fewer simulation points than BYO, which is a great advantage for design cases with the time-consuming simulations. For $\epsilon = 0.05$, the optimized nominal design is $\boldsymbol{x}^* = [300, 111.2, 300]$ and its expected bandwidth is 192.2 GHz. In Fig. 10, we compare the frequency response before and after the yield-aware optimization. Our proposed method can have a higher bandwidth and a smaller crosstalk compared to BYO and the initial design. Fig. 11 further shows the probability density of the optimized bandwidth by our chance-constrained optimization and the BYO, respectively. It clearly shows that our proposed method produces higher bandwidth.

VI. CONCLUSION

This article has presented a data-efficient framework for the yield-aware optimization of photonic ICs under non-Gaussian correlated process variations. We have proposed to reformulate the stochastic chance-constrained optimization into a deterministic polynomial optimization problem. Our framework only requires simulation at a small number of important points and admits a surrogate model for yield-aware optimization. In the experiments by the microring filter and the Mach–Zehnder filter, we have demonstrated that our optimization scheme can give high yield and high bandwidth. Compared with BYO, our method has consumed much fewer simulation samples and

produced better design performance while achieving the same yield.

This article should be regarded as a presentation of preliminary results in this direction. Many problems are worth further investigation in the future, for instance.

1) *Nonsmoothness:* Similar to generalized polynomial chaos [45], the surrogate modeling techniques in [28] and [29] require the stochastic functions to be smooth. However, the performance metrics of a photonic IC may be nonsmooth with respect to the design variables and process variations. How to handle nonsmoothness in this optimization framework is a critical issue.
2) *High Dimensionality:* Large-scale photonic ICs may have a huge number of design variables and process variation parameters. This brings new challenges to the surrogate modeling and the resulting polynomial optimization in our framework.

APPENDIX A
DETAILED DERIVATION OF (6)

Suppose $u > \mathbb{E}_{\boldsymbol{\xi}}[y(\boldsymbol{x},\boldsymbol{\xi})]$. We show that the following deterministic constraint

$$\mathbb{E}_{\boldsymbol{\xi}}\big[y(\boldsymbol{x},\boldsymbol{\xi})\big] + \sqrt{(1-\epsilon)/\epsilon}\sqrt{\mathrm{var}_{\boldsymbol{\xi}}\big[y(\boldsymbol{x},\boldsymbol{\xi})\big]} \le u$$

is a sufficient but not necessary condition for the chance constraint

$$\mathrm{Prob}_{\boldsymbol{\xi}}(y(\boldsymbol{x},\boldsymbol{\xi}) \le u) \ge 1-\epsilon.$$

In other words, we want to show that each feasible point of (16) is a feasible point of the chance constraint (13).

Denote the random variable as $\boldsymbol{X} = y(\boldsymbol{x},\boldsymbol{\xi})$. Cantelli's inequality [54] states that for any random variable $\boldsymbol{X}$ with a mean value $\mathbb{E}[\boldsymbol{X}] = \mathbb{E}_{\boldsymbol{\xi}}[y(\boldsymbol{x},\boldsymbol{\xi})]$ and variance $\sigma^2 = \mathrm{var}_{\boldsymbol{\xi}}[y(\boldsymbol{x},\boldsymbol{\xi})]$, it holds that the probability of a single tail can be bounded as follows:

$$\mathrm{Prob}(\boldsymbol{X} - \mathbb{E}[\boldsymbol{X}] \le \lambda) \ge 1 - \frac{\sigma^2}{\sigma^2+\lambda^2} \quad \text{if } \lambda > 0. \tag{41}$$

Therefore, for any constant $u \ge \mathbb{E}[\boldsymbol{x}]$ we have

$$\begin{aligned}\mathrm{Prob}(\boldsymbol{X} \le u) &= \mathrm{Prob}(\boldsymbol{X} - \mathbb{E}[\boldsymbol{X}] \le u - \mathbb{E}[\boldsymbol{X}]) \\ &\ge 1 - \frac{\sigma^2}{\sigma^2 + (u-\mathbb{E}[\boldsymbol{X}])^2}.\end{aligned}$$

For any ϵ, a sufficient condition for $\mathrm{Prob}(\boldsymbol{X} \le u) \ge 1-\epsilon$ is $1 - [\sigma^2/(\sigma^2 + (u-\mathbb{E}[\boldsymbol{x}])^2)] \ge 1-\epsilon$, i.e.,

$$\mathbb{E}[\boldsymbol{X}] + \sqrt{(1-\epsilon)/\epsilon}\sigma \le u. \tag{42}$$

Substituting $\boldsymbol{X} = y(\boldsymbol{x},\boldsymbol{\xi})$ into the above equation we get (6). The proof is completed.

APPENDIX B
DETAILED DERIVATION OF (20) AND (21)

Suppose that the smooth function $y(\boldsymbol{x},\boldsymbol{\xi})$ is already represented by a linear combination of some basis functions

$$y(\boldsymbol{x},\boldsymbol{\xi}) = \sum_{|\boldsymbol{\alpha}|+|\boldsymbol{\beta}|=0}^{p} c_{\boldsymbol{\alpha},\boldsymbol{\beta}}\Phi_{\boldsymbol{\alpha}}(\boldsymbol{x})\Psi_{\boldsymbol{\beta}}(\boldsymbol{\xi}) \tag{43}$$

where $\mathbb{E}[\Psi_{\boldsymbol{\beta}}(\boldsymbol{\xi})\Psi_{\boldsymbol{\gamma}}(\boldsymbol{\xi})] = \delta_{\boldsymbol{\beta},\boldsymbol{\gamma}}$. The mean value of $y(\boldsymbol{x},\boldsymbol{\xi})$ is

$$\begin{aligned}\mathbb{E}_{\boldsymbol{\xi}}\big[y(\boldsymbol{x},\boldsymbol{\xi})\big] &= \sum_{|\boldsymbol{\alpha}|=0}^{p}\sum_{|\boldsymbol{\beta}|=0}^{p-|\boldsymbol{\alpha}|} c_{\boldsymbol{\alpha},\boldsymbol{\beta}}\Phi_{\boldsymbol{\alpha}}(\boldsymbol{x})\mathbb{E}\big[\Psi_{\boldsymbol{\beta}}(\boldsymbol{\xi})\big] \\ &= \sum_{|\boldsymbol{\alpha}|=0}^{p} c_{\boldsymbol{\alpha},0}\Phi_{\boldsymbol{\alpha}}(\boldsymbol{x})\end{aligned}$$

where the last equality is due to $\Psi_0(\boldsymbol{\xi}) = 1$ and $\mathbb{E}[\Psi_{\boldsymbol{\beta}}(\boldsymbol{\xi})] = \mathbb{E}[\Psi_{\boldsymbol{\beta}}(\boldsymbol{\xi})\Psi_0(\boldsymbol{\xi})] = 0\ \forall \boldsymbol{\beta} \ne 0$. The variance is

$$\begin{aligned}\mathrm{var}_{\boldsymbol{\xi}}\big[y(\boldsymbol{x},\boldsymbol{\xi})\big] &= \mathbb{E}_{\boldsymbol{\xi}}\Big[\big(y(\boldsymbol{x},\boldsymbol{\xi}) - \mathbb{E}_{\boldsymbol{\xi}}\big[y(\boldsymbol{x},\boldsymbol{\xi})\big]\big)^2\Big] \\ &= \mathbb{E}_{\boldsymbol{\xi}}\left[\left(\sum_{|\boldsymbol{\beta}|=1}^{p}\left(\sum_{|\boldsymbol{\alpha}|=0}^{p-|\boldsymbol{\beta}|} c_{\boldsymbol{\alpha},\boldsymbol{\beta}}\Phi_{\boldsymbol{\alpha}}(\boldsymbol{x})\right)\Psi_{\boldsymbol{\beta}}(\boldsymbol{\xi})\right)^2\right] \\ &= \sum_{|\boldsymbol{\beta}|=1}^{p}\left(\sum_{|\boldsymbol{\alpha}|=0}^{p-|\boldsymbol{\beta}|} c_{\boldsymbol{\alpha},\boldsymbol{\beta}}\Phi_{\boldsymbol{\alpha}}(\boldsymbol{x})\right)^2\end{aligned}$$

where the last equality is due to the basis functions $\{\Psi_{\boldsymbol{\beta}}(\boldsymbol{\xi})\}$ are orthogonal in the stochastic parameter space.

APPENDIX C
BAYESIAN YIELD OPTIMIZATION

BYO is a state-of-the-art tool for the yield optimization of electronic devices and circuits [10]. This method approximates and optimizes the posterior distribution of design variable under the condition of pass events

$$S = \{(\boldsymbol{x},\boldsymbol{\xi}) : (\boldsymbol{x},\boldsymbol{\xi}) \text{ satisfies all performance constraints}\}.$$

With the Bayes' theorem, it holds that $\mathrm{Prob}(S|\boldsymbol{x}) = [(\mathrm{Prob}(S))/(\mathrm{Prob}(\boldsymbol{x}))]\mathrm{Prob}(\boldsymbol{x}|S)$. In our problem setting, $\mathrm{Prob}(\boldsymbol{x})$ is a constant because we assume that $\boldsymbol{x}$ follows a uniform distribution and $\mathrm{Prob}(S)$ should also be a constant without the dependence on the variable $\boldsymbol{x}$. Therefore, $\mathrm{Prob}(S|\boldsymbol{x}) \propto \mathrm{Prob}(\boldsymbol{x}|S)$ and the original yield optimization problem (3) is equivalent to

$$\boldsymbol{x}^{\mathrm{BYO}} = \underset{\boldsymbol{x}\in\mathcal{X}}{\mathrm{argmax}}\ \ \mathrm{Prob}(\boldsymbol{x}|S). \tag{44}$$

The paper [10] proposed an expectation-maximization framework to solve problem (44). At the t-th iteration, the expectation step approximates the probability by the kernel density estimation. Specifically, we generate $N = 100$ samples $(\boldsymbol{x}_i,\boldsymbol{\xi}_i)$ randomly and call a simulator to compute the quantity of interests at those samples. Then choose $M \le N$ pass samples to perform the kernel density estimation

$$\mathrm{Prob}(\boldsymbol{x}|S) \approx \frac{1}{M}\sum_{i=1}^{M}\frac{1}{\sqrt{2\pi h}}\exp\left(-\frac{1}{2h}(\boldsymbol{x}-\boldsymbol{\mu}_i)^T(\boldsymbol{x}-\boldsymbol{\mu}_i)\right)$$

where $\{\boldsymbol{\mu}_i\}_{i=1}^{M} \in S$ are design samples that satisfies the performance constraints and $h = 0.3$ is a bandwidth parameter. Afterward, the maximization step returns an updated design variable $\boldsymbol{x}^{\mathrm{BYO},t}$. We will call the simulator again at this design variable to record its objective value and pass status. We terminate the algorithm if the maximal iteration number 20 is reached, or the residue of two consecutive iterations is below

10^{-6}. After the whole optimization process, we return the design variable that can pass the yield constraints with the best objective value

$$\boldsymbol{x}^{\mathrm{BYO}} = \arg \max_{\boldsymbol{x} \in \boldsymbol{x}^{\mathrm{BYO},t}} \mathbb{E}_{\boldsymbol{\xi}}\big[f(\boldsymbol{x}, \boldsymbol{\xi})\big] \text{ s.t. } \mathrm{pass}(\boldsymbol{x}) = 1.$$

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their detailed comments. They also appreciate P. Pintus for his helpful discussions on the benchmarks.

REFERENCES

[1] S. C. Nicholes *et al.*, "The world's first InP 8× 8 monolithic tunable optical router (MOTOR) operating at 40 Gbps line rate per port," in *Proc. Opt. Fiber Commun.*, 2009, pp. 1–3.

[2] M. Kato *et al.*, "40-channel transmitter and receiver photonic integrated circuits operating at a per channel data rate 12.5 Gbit/s," in *Proc. Nat. Fiber Opt. Eng. Conf.*, 2007.

[3] X. Chen, M. Mohamed, Z. Li, L. Shang, and A. R. Mickelson, "Process variation in silicon photonic devices," *Appl. Opt.*, vol. 52, no. 31, pp. 7638–7647, 2013.

[4] T. Lipka, J. Müller, and H. K. Trieu, "Systematic nonuniformity analysis of amorphous silicon-on-insulator photonic microring resonators," *J. Lightw. Technol.*, vol. 34, no. 13, pp. 3163–3170, Jul. 1, 2016.

[5] Z. Lu *et al.*, "Performance prediction for silicon photonics integrated circuits with layout-dependent correlated manufacturing variability," *Opt. Exp.*, vol. 25, no. 9, pp. 9712–9733, 2017.

[6] J. Pond *et al.*, "Predicting the yield of photonic integrated circuits using statistical compact modeling," in *Proc. Integr. Opt. Phys. Simulat. III*, vol. 10242, 2017, Art. no. 102420S.

[7] T.-W. Weng, D. Melati, A. Melloni, and L. Daniel, "Stochastic simulation and robust design optimization of integrated photonic filters," *Nanophotonics*, vol. 6, no. 1, pp. 299–308, 2017.

[8] T.-K. Yu, S.-M. Kang, J. Sacks, and W. J. Welch, "An efficient method for parametric yield optimization of MOS integrated circuits," in *Proc. Int. Conf. Comput.-Aided Design*, 1989, pp. 190–193.

[9] Y. Li, H. Schneider, F. Schnabel, R. Thewes, and D. Schmitt-Landsiedel, "DRAM yield analysis and optimization by a statistical design approach," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 58, no. 12, pp. 2906–2918, Dec. 2011.

[10] M. Wang, F. Yang, C. Yan, X. Zeng, and X. Hu, "Efficient Bayesian yield optimization approach for analog and SRAM circuits," in *Proc. Design Autom. Conf.*, 2017, pp. 1–6.

[11] M. Wang *et al.*, "Efficient yield optimization for analog and SRAM circuits via Gaussian process regression and adaptive yield estimation," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 37, no. 10, pp. 1929–1942, Oct. 2018.

[12] Y. Xu, K.-L. Hsiung, X. Li, I. Nausieda, S. Boyd, and L. Pileggi, "OPERA: Optimization with ellipsoidal uncertainty for robust analog IC design," in *Proc. Design Autom. Conf.*, 2005, pp. 632–637.

[13] X. Li, P. Gopalakrishnan, Y. Xu, and L. T. Pileggi, "Robust analog/RF circuit design with projection-based performance modeling," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 26, no. 1, pp. 2–15, Jan. 2007.

[14] X. Li, J. Le, and L. T. Pileggi, "Statistical performance modeling and optimization," *Found. Trends® Electron. Design Autom.*, vol. 1, no. 4, pp. 331–480, 2007.

[15] X. Li, J. Sun, F. Xiao, and J.-S. Tian, "An efficient bi-objective optimization framework for statistical chip-level yield analysis under parameter variations," *Front. Inf. Technol. Electron. Eng.*, vol. 17, no. 2, pp. 160–172, 2016.

[16] 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. Comput.-Aided Design Integr. Circuits Syst.*, vol. 32, no. 10, pp. 1533–1545, Oct. 2013.

[17] 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. Comput.-Aided Design Integr. Circuits Syst.*, vol. 33, no. 5, pp. 728–740, May 2014.

[18] Z. Zhang, I. Osledets, X. Yang, G. E. Karniadakis, and L. Daniel, "Enabling high-dimensional hierarchical uncertainty quantification by ANOVA and tensor-train decomposition," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 34, no. 1, pp. 63–76, Jan. 2015.

[19] Z. Zhang, T.-W. Weng, and L. Daniel, "Big-data tensor recovery for high-dimensional uncertainty quantification of process variations," *IEEE Trans. Compon. Packag. Manuf. Technol.*, vol. 7, no. 5, pp. 687–697, May 2017.

[20] P. Manfredi, D. V. Ginste, D. De Zutter, and F. G. Canavero, "Uncertainty assessment of lossy and dispersive lines in SPICE-type environments," *IEEE Trans. Compon. Packag. Manuf. Technol.*, vol. 3, no. 7, pp. 1252–1258, Jul. 2013.

[21] S. Vrudhula, J. M. Wang, and P. Ghanta, "Hermite polynomial based interconnect analysis in the presence of process variations," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 25, no. 10, pp. 2001–2011, Oct. 2006.

[22] K. Strunz and Q. Su, "Stochastic formulation of SPICE-type electronic circuit simulation with polynomial chaos," *ACM Trans. Model. Comput. Simulat.*, vol. 18, no. 4, p. 15, 2008.

[23] M. R. Rufuie, E. Gad, M. Nakhla, and R. Achar, "Generalized Hermite polynomial chaos for variability analysis of macromodels embedded in nonlinear circuits," *IEEE Trans. Compon. Packag. Manuf. Technol.*, vol. 4, no. 4, pp. 673–684, Apr. 2013.

[24] 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 South Pac. Design Autom. Conf.*, 2007, pp. 474–479.

[25] R. Shen, S. X.-D. Tan, J. Cui, W. Yu, Y. Cai, and G.-S. Chen, "Variational capacitance extraction and modeling based on orthogonal polynomial method," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 18, no. 11, pp. 1556–1566, Nov. 2010.

[26] A. Waqas, D. Melati, Z. Mushtaq, and A. Melloni, "Uncertainty quantification and stochastic modeling of photonic device from experimental data through polynomial chaos expansion," in *Proc. Integr. Opt. Devices Mater. Technol. XXII*, vol. 10535, 2018, Art. no. 105351A.

[27] A. Waqas, D. Melati, P. Manfredi, and A. Melloni, "Stochastic process design kits for photonic circuits based on polynomial chaos augmented macro-modelling," *Opt. Exp.*, vol. 26, no. 5, pp. 5894–5907, 2018.

[28] C. Cui, M. Gershman, and Z. Zhang, "Stochastic collocation with non-Gaussian correlated parameters via a new quadrature rule," in *Proc. 27th IEEE Conf. Elect. Perform. Electron. Packag. Syst.*, 2018, pp. 57–59.

[29] C. Cui and Z. Zhang, "Stochastic collocation with non-Gaussian correlated process variations: Theory, algorithms and applications," *IEEE Trans. Compon. Packag. Manuf. Technol.*, vol. 9, no. 7, pp. 1362–1375, Jul. 2019.

[30] 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," *Opt. Exp.*, vol. 23, no. 4, pp. 4242–4254, 2015.

[31] P. Li, H. Arellano-Garcia, and G. Wozny, "Chance constrained programming approach to process optimization under uncertainty," *Comput. Chem. Eng.*, vol. 32, nos. 1–2, pp. 25–45, 2008.

[32] A. Mesbah, S. Streif, R. Findeisen, and R. D. Braatz, "Stochastic nonlinear model predictive control with probabilistic constraints," in *Proc. Amer. Control Conf.*, 2014, pp. 2413–2419.

[33] L. Blackmore, M. Ono, A. Bektassov, and B. C. Williams, "A probabilistic particle-control approximation of chance-constrained stochastic predictive control," *IEEE Trans. Robot.*, vol. 26, no. 3, pp. 502–517, Jun. 2010.

[34] H. Akhavan-Hejazi and H. Mohsenian-Rad, "Energy storage planning in active distribution grids: A chance-constrained optimization with non-parametric probability functions," *IEEE Trans. Smart Grid*, vol. 9, no. 3, pp. 1972–1985, May 2018.

[35] Z. Wang, C. Shen, F. Liu, X. Wu, C.-C. Liu, and F. Gao, "Chance-constrained economic dispatch with non-Gaussian correlated wind power uncertainty," *IEEE Trans. Power Syst.*, vol. 32, no. 6, pp. 4880–4893, Nov. 2017.

[36] A. Nemirovski and A. Shapiro, "Convex approximations of chance constrained programs," *SIAM J. Optim.*, vol. 17, no. 4, pp. 969–996, 2006.

[37] G. Calafiorea and L. El Ghaoui, "Distributionally robust chance-constrained linear programs with applications," DAUIN, Politecnico di Torino, Torino, Italy, Rep., 2005.

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

[39] R. Ghanem and P. Spanos, *Stochastic Finite Elements: A Spectral Approach*. New York, NY, USA: Springer-Verlag, 1991.

[40] 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, Mar. 2005.
[41] X. Li, "Finding deterministic solution from under-determined equation: Large-scale performance variability modeling of analog/RF circuits," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 29, no. 11, pp. 1661–1668, Nov. 2010.
[42] Z. Zhang *et al.*, "Stochastic testing simulator for integrated circuits and MEMS: Hierarchical and sparse techniques," in *Proc. IEEE Custom Integr. Circuits Conf.*, San Jose, CA, USA, Sep. 2014, pp. 1–8.
[43] C. Cui and Z. Zhang, "High-dimensional uncertainty quantification of electronic and photonic IC with non-Gaussian correlated process variations," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, early access, doi: 10.1109/TCAD.2019.2925340.
[44] C. Cui and Z. Zhang, "Uncertainty quantification of electronic and photonic ICs with non-Gaussian correlated process variations," in *Proc. Int. Conf. Comput.-Aided Design*, 2018, pp. 1–8.
[45] D. Xiu and G. E. Karniadakis, "The Wiener–Askey polynomial chaos for stochastic differential equations," *SIAM J. Sci. Comput.*, vol. 24, no. 2, pp. 619–644, 2002.
[46] V. Barthelmann, E. Novak, and K. Ritter, "High dimensional polynomial interpolation on sparse grids," *Adv. Comput. Math.*, vol. 12, no. 4, pp. 273–288, Mar. 2000.
[47] T. Gerstner and M. Griebel, "Numerical integration using sparse grids," *Numer. Alg.*, vol. 18, pp. 209–232, Mar. 1998.
[48] F. H. Clarke, *Optimization and Nonsmooth Analysis*, vol. 5. Philadelphia, PA, USA: SIAM, 1990,
[49] H. Waki, S. Kim, M. Kojima, and M. Muramatsu, "Sums of squares and semidefinite program relaxations for polynomial optimization problems with structured sparsity," *SIAM J. Optim.*, vol. 17, no. 1, pp. 218–242, 2006.
[50] J. Nie, "An exact Jacobian SDP relaxation for polynomial optimization," *Math. Program.*, vol. 137, nos. 1–2, pp. 225–255, 2013.
[51] D. Henrion, J.-B. Lasserre, and J. Löfberg, "GloptiPoly 3: Moments, optimization and semidefinite programming," *Optim. Methods Softw.*, vol. 24, nos. 4–5, pp. 761–779, 2009.
[52] R. Orta, P. Savi, R. Tascone, and D. Trinchero, "Synthesis of multiple-ring-resonator filters for optical systems," *IEEE Photon. Technol. Lett.*, vol. 7, no. 12, pp. 1447–1449, Dec. 1995.
[53] P. Pintus, P. Contu, N. Andriolli, A. D'Errico, F. Di Pasquale, and F. Testa, "Analysis and design of microring-based switching elements in a silicon photonic integrated transponder aggregator," *J. Lightw. Technol.*, vol. 31, no. 24, pp. 3943–3955, Dec. 15, 2013.
[54] F. P. Cantelli, "Sui confini della probabilita," in *Proc. Atti del Congresso Internazionale dei Matematici: Bologna del 3 al 10 de settembre di 1928*, 1929, pp. 47–60.

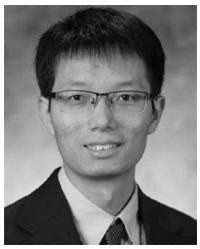
Chunfeng Cui received the Ph.D. degree in computational mathematics from the Chinese Academy of Sciences, Beijing, China, in 2016 with a specialization in numerical optimization.

From 2016 to 2017, she was a Postdoctoral Fellow with the City University of Hong Kong, Hong Kong. In 2017, she joined the Department of Electrical and Computer Engineering, University of California at Santa Barbara, Santa Barbara, CA, USA, as a Postdoctoral Scholar. Her research activities are mainly focused on the areas of tensor computing, uncertainty quantification, machine learning, and their interface.

Dr. Cui is a recipient of the 2019 Rising Stars in Computational and Data Sciences, the 2019 Rising Stars in EECS, the 2018 Best Paper Award of IEEE Electrical Performance of Electronic Packaging and Systems, and the Best Journal Paper Award of *Scientia Sinica Mathematica*.

Kaikai Liu received the B.S. degree in physics from the Huazhong University of Science and Technology, Wuhan, China, in 2018. He is currently pursuing the Ph.D. degree with the Department of Electrical and Computer Engineering, University of California at Santa Barbara, Santa Barbara, CA, USA.

He has been working on the tensorized Bayesian optimization and the chance-constraint optimization method. His research focuses on developing the novel data-driving algorithms for photonic integrated circuits design.



Zheng Zhang (Member, IEEE) received the Ph.D. degree in electrical engineering and computer science from the Massachusetts Institute of Technology (MIT), Cambridge, MA, USA, in 2015.

He is an Assistant Professor of electrical and computer engineering with the University of California at Santa Barbara, Santa Barbara, CA, USA. His research interests include uncertainty quantification for the design automation of multidomain systems, and tensor methods for high-dimensional data analytics.

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, and the Two Best Conference Paper Awards (IEEE EPEPS 2018 and IEEE SPI 2016). 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 received the NSF CAREER Award in 2019.