

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

Chunfeng Cui*, Kaikai Liu*, 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 one side, optimizing the generally non-convex 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 paper 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: a modeling module that builds stochastic surrogate models for design objectives and chance constraints with a few simulation samples, and 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—Photonic integrated circuits, photonic design automation, uncertainty quantification, yield optimization, chance constraints, non-Gaussian correlations.

I. INTRODUCTION

THE demand for low-power, high-speed communications and computing have boosted the advances in photonic integrated circuits. Based on the modern nano-fabrication 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 desired [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; 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. The reference [15] proposed a Chebyshev affine arithmetic method to predict the cumulative distribution function. The recent Bayesian yield optimization [10] approximated the probability density of the design variable under the condition of “pass” by a 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 non-optimal (and even poor) chip performance.

Recently, uncertainty quantification methods based on generalized polynomial chaos have gained great success in modeling the uncertainty caused by various process variations in electronic and photonic ICs [16]–[27]. A novel stochastic collocation approach was further proposed in [28], [29] to handle non-Gaussian correlated process variations, which shows significantly better accuracy and efficiency than [30] due to the smooth basis functions and an optimization-based quadrature rule. These techniques can construct stochastic surrogate models with a small number of simulation samples, but their power in yield optimization has not been well explored or exploited despite recent robust optimization methods [31] based on generalized polynomial chaos.

Paper Contributions. Leveraging the chance-constrained optimization [32] and our recent uncertainty quantification solvers [28], [29], this paper 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

* C. Cui and K. Liu contributed equally to this work.

This work was partly supported by NSF Grant No. 1763699, NSF CAREER Award No. 1846476, a UCSB start-up grant and a Samsung Gift Funding.

Chunfeng Cui, Kaikai Liu, and Zheng Zhang are in the Department of Electrical and Computer Engineering, University of California, Santa Barbara, CA 93106, USA (e-mail: chunfengcui@ucsb.edu, kaikailiu@ucsb.edu, and zhengzhang@ece.ucsb.edu).

threshold. Doing so can avoid performance degradation in yield optimization. Chance-constrained optimization [32] has been widely used in system control [33], autonomous vehicles [34], and reliable power generation [35], [36], but it has not been investigated for yield optimization of electronic or photonic ICs. Our specific contributions include:

- A chance-constraint optimization framework that can achieve high chip performance and high yield simultaneously under non-Gaussian correlated process variations.
- 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.
- A deterministic reformulation. A major challenge of chance-constrained problems is to reformulate the stochastic constraints into deterministic ones [37]. We reformulate the probabilistic objective function and constraints as non-smooth deterministic functions. Afterward, we transform them into an equivalent polynomial optimization, which can be solved efficiently.
- 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 Bayesian yield optimization 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 work should be regarded as a preliminary result in this direction, and many topics can be investigated in the future.

II. PRELIMINARIES

A. The Yield Optimization

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

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

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

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

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

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

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

B. Chance Constraints

The chance constraint is a powerful technique in uncertainty-aware optimization [32]. 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):

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

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

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

Under strict conditions, such as the parameters being independent and $y(\mathbf{x}, \boldsymbol{\xi})$ being a linear function, (4) can be reformulated into equivalent deterministic constraints [38]. In other words, one can reformulate the left-hand side of (4) by its probability density function 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 probability density function or probability density function of an uncertain variable can be intractable [17], [37]. In these cases, we seek for deterministic reformulations that can well approximate the chance constraints. There is a trade-off 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 popular method converts (4) to a deterministic constraint via the mean and variance of $y(\mathbf{x}, \boldsymbol{\xi})$ [37], [38]:

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

Here $\mathbb{E}_{\boldsymbol{\xi}}[\cdot]$ denotes the mean value, $\text{var}_{\boldsymbol{\xi}}[\cdot]$ 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:

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

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

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

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

These one-dimensional basis functions $\psi_i(\xi_i)$ can be constructed by the three term recursion [39]. Various stochastic spectral approaches have been proposed to compute the coefficients c_α , including the intrusive (i.e., non-sampling) solvers (e.g., stochastic Galerkin [40], the stochastic testing [16]) and the non-intrusive (i.e., sampling) solvers (e.g., stochastic collocation [41]). 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 [42], ANOVA (analysis of variance) or HDMR (the high-dimensional model representation) [43], and the hierarchical uncertainty quantification [18].

In practice, the random parameters may be correlated. If the parameters ξ are non-Gaussian correlated, the computation is more difficult. In such cases, $\Psi_\alpha(\xi)$ can be constructed by the Gram-Schmidt approach in [28], [29] or the Cholesky factorization in [44], [45]. The main difficulty lies in computing high order moments of ξ , which can be well resolved by the functional tensor train approach proposed in [45].

III. OUR YIELD-AWARE OPTIMIZATION MODEL

In this section, we will show our yield optimization model defined by a stochastic measure in the probability space, and will illustrate how to convert the stochastic formulation to a deterministic one. We first present the basic assumptions in this paper.

Assumption 1. *We made the following assumptions:*

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

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

Here $[n] = 1, \dots, n$ and $\mathbb{E}[y_i(\mathbf{x}, \xi)] \leq u_i$. Each individual quantity $\{y_i(\mathbf{x}, \xi)\}_{i=1}^n$ is a black-box function, and we can obtain its function values at given samples.

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

A. The Probabilistic Yield Optimization Model

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

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

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

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

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 is often non-unique, and we hope to chose a design that achieves good performance and high yield simultaneously. Therefore, we ensure the yield with a chance constraint

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

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

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

Here ϵ is a 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 $f(\mathbf{x}, \xi)$ 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.

When the yield function $Y(\mathbf{x})$ and the objective function $f(\mathbf{x}, \xi)$ are available, we may solve the above optimization problem directly. However, this is rarely true. Normally, one has to estimate the yield and objective at a given \mathbf{x} by the Monte Carlo method [8], [9] which requires a huge number of simulation samples at each design variable \mathbf{x} . This 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:

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

In this formulation, ϵ_i means the risk tolerance of violating the i -th design specification. Since $\text{Prob}_\xi(\mathbf{y}(\mathbf{x}, \xi) \leq \mathbf{u}) = \text{Prob}_\xi(\bigcap_{i=1}^n y_i(\mathbf{x}, \xi) \leq u_i) = 1 - \text{Prob}_\xi(\bigcup_{i=1}^n y_i(\mathbf{x}, \xi) \geq u_i)$, the probability of the joint chance constraint can be upper and lower bounded by the individual chance constraints:

$$\begin{aligned} & \max_{i=1, \dots, n} \text{Prob}_\xi(y_i(\mathbf{x}, \xi) \geq u_i) \\ & \leq \text{Prob}_\xi(\bigcup_{i=1}^n y_i(\mathbf{x}, \xi) \geq u_i) \leq \sum_{i=1}^n \text{Prob}_\xi(y_i(\mathbf{x}, \xi) \geq u_i). \end{aligned}$$

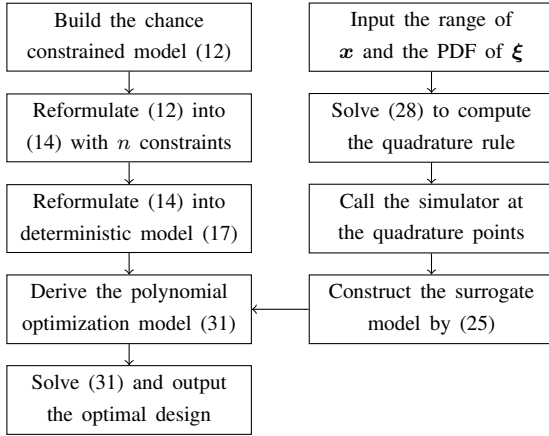


Fig. 1. The flowchart of our proposed framework for solving the chance constrained yield-aware optimization.

When $\epsilon_i = \epsilon$ for all i , (13) is a relaxation of (11) (e.g., the feasible domain is enlarged); when $\sum_{i=1}^n \epsilon_i \leq \epsilon$, (13) becomes more conservative than (11) (e.g., the feasible domain is decreased). In this paper, 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

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

B. Reformulate the Stochastic into Deterministic Model

The chance-constraint optimization in problem (14) is difficult to solve directly. This problem is more challenging when $y_i(\mathbf{x}, \xi)$ is nonlinear. In this case, 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:

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

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 [37], [38] and replace (13) by

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

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

- Constraint (16) is a stronger condition than (13). In other words, each feasible point of (16) is also a feasible solution of the chance constraint (13);

- 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;
- When the variance $\text{var}_{\xi}[y_i(\mathbf{x}, \xi)]$ is small enough, the feasible set of (16) is close to the deterministic constraint $\mathbb{E}_{\xi}[y_i(\mathbf{x}, \xi)] \leq u_i$.

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

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

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 $\{y_i(\mathbf{x}, \xi)\}_{i=1}^n$ and $f(\mathbf{x}, \xi)$. A direct approach is to apply a Monte Carlo method to estimate the mean values and variances for every iterate \mathbf{x} . However, this is not affordable because of the large number of numerical simulations.

In this section, we build the surrogate model for $f(\mathbf{x}, \xi)$ and $\{y_i(\mathbf{x}, \xi)\}_{i=1}^n$ by using generalized polynomial chaos [46] 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 $\Phi_{\alpha}(\mathbf{x})$, $\Psi_{\beta}(\xi)$ and compute the coefficients $c_{\alpha,\beta}^i$ and $h_{\alpha,\beta}$ such that

$$y_i(\mathbf{x}, \xi) \approx \sum_{|\alpha|+|\beta|=0}^p c_{\alpha,\beta}^i \Phi_{\alpha}(\mathbf{x}) \Psi_{\beta}(\xi), \quad (18)$$

and

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

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

$$\mathbb{E}_{\xi}[y_i(\mathbf{x}, \xi)] \approx \sum_{|\alpha|=0}^p c_{\alpha,0}^i \Phi_{\alpha}(\mathbf{x}), \quad (20)$$

and the variance is approximated by

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

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

For the uniform-distributed design variables \mathbf{x} , their basis functions $\Phi_\alpha(\mathbf{x})$ can be decoupled into the products of one-dimensional basis functions:

$$\Phi_\alpha(\mathbf{x}) = \phi_{\alpha_1}^1(x_1) \dots \phi_{\alpha_{d_1}}^{d_1}(x_{d_1}). \quad (22)$$

Here, $\phi_{\alpha_i}^i(x_i)$ is a Legendre polynomial [46] and can be constructed by the three-term recurrence relation [39].

For the random vector $\boldsymbol{\xi}$ describing non-Gaussian correlated process variations, we construct its basis functions $\Psi_\beta(\boldsymbol{\xi})$ by the Gram-Schmidt approach proposed in [28], [29]. Specifically, we first reorder the monomials $\boldsymbol{\xi}^\beta = \xi_1^{\beta_1} \dots \xi_{d_2}^{\beta_{d_2}}$ in the graded lexicographic order, and denote them as $\{p_j(\boldsymbol{\xi})\}_{j=1}^{N_p}$. Here, $N_p = \binom{d_2+p}{p}$ 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 $\{\Psi_j(\boldsymbol{\xi})\}_{j=2}^{N_p}$ in the correlated parameter space recursively by

$$\begin{aligned} \hat{\Psi}_j(\boldsymbol{\xi}) &= p_j(\boldsymbol{\xi}) - \sum_{i=1}^{j-1} \mathbb{E}[p_j(\boldsymbol{\xi})\Psi_i(\boldsymbol{\xi})]\Psi_i(\boldsymbol{\xi}), \\ \Psi_j(\boldsymbol{\xi}) &= \frac{\hat{\Psi}_j(\boldsymbol{\xi})}{\sqrt{\mathbb{E}[\hat{\Psi}_j^2(\boldsymbol{\xi})]}}, \quad j = 2, \dots, N_p. \end{aligned} \quad (23)$$

These basis functions $\{\Psi_j(\boldsymbol{\xi})\}_{j=1}^{N_p}$ can be re-ordered into the graded lexicographic order $\{\Psi_\beta(\boldsymbol{\xi})\}_{|\beta|=0}^p$.

B. How to Build the Surrogate Models?

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

$$c_{\alpha,\beta}^i = \mathbb{E}_{\mathbf{x},\boldsymbol{\xi}}[y_i(\mathbf{x},\boldsymbol{\xi})\Phi_\alpha(\mathbf{x})\Psi_\beta(\boldsymbol{\xi})]. \quad (24)$$

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

$$c_{\alpha,\beta}^i \approx \sum_{k=1}^M y_i(\mathbf{x}_k, \boldsymbol{\xi}_k) \Phi_\alpha(\mathbf{x}_k) \Psi_\beta(\boldsymbol{\xi}_k) w_k. \quad (25)$$

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

In this paper, we propose a three-stage optimization method to compute the quadrature points and weights:

- Firstly, we compute the quadrature rule $\{\mathbf{x}_l, v_l\}_{l=1}^{M_1}$ for the independent design variables \mathbf{x} .
- Secondly, we employ the optimization approach proposed in [28], [29] to calculate the quadrature points and weights $\{\boldsymbol{\xi}_l, u_l\}_{l=1}^{M_2}$ for the non-Gaussian correlated parameters $\boldsymbol{\xi}$.
- Finally, we use their tensor products ($M_1 M_2$ points) as an initialization and recall the optimization approach proposed in [28], [29] for the coupled space of \mathbf{x} and $\boldsymbol{\xi}$ to compute $M \leq M_1 M_2$ joint quadrature points $\{\mathbf{x}_k, \boldsymbol{\xi}_k, w_k\}_{k=1}^M$.

The details are described below.

1) *Initial Quadrature Points for \mathbf{x}* : One could employ the sparse grid approach [47], [48] to compute the quadrature samples and weights for the independent uniform-distribution variables $\mathbf{x} \in \mathbb{R}^{d_1}$. 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 propose refining the quadrature rule by the least square optimization solver

$$\min_{\mathbf{a} \leq \mathbf{x}_l \leq \mathbf{b}, v_l \geq 0} \sum_{j=1}^{N_{2p}} \left(\mathbb{E}[\Phi_j(\mathbf{x})] - \sum_{l=1}^{M_1} \Phi_j(\mathbf{x}_l) v_l \right)^2. \quad (26)$$

Here, the expectations $\mathbb{E}[\Phi_j(\mathbf{x})] = \delta_{1j}$ are already known from the orthogonality of basis functions. This model is similar to that of [28], [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 p -th 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], [29]. Specifically, we compute M_2 quadrature points $\boldsymbol{\xi}_l$ and weights w_l via solving the following optimization problem

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

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

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

Here $\delta_{1j_1} \delta_{1j_2} = 1$ if $j_1 = j_2 = 1$ and zero otherwise. Our numerical experiments show that the total number of optimized quadrature points is M is significantly smaller than $M_1 M_2$.

Remark: We can also solve (28) directly to obtain the optimized quadrature points. However, (28) is a non-convex optimization problem and is hard to optimize in general. The subproblems (26) and (27) help to provide a good initial guess for the joint optimization.

For all optimization subproblems (26), (27), and (28), we use the block coordinate-descent optimization method described in [29] to compute the quadrature points and weights alternatively. The following theorem ensures high accuracy for our surrogate model considering the unavoidable numerical optimization error and function approximation error.

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

Theorem 1. [29] Assume that $\{\mathbf{x}_k, \boldsymbol{\xi}_k, w_k\}_{k=1}^M$ 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

$$N_p = \frac{(d+p)!}{p!d!} \leq M \leq N_{2p} = \frac{(d+2p)!}{2p!d!}; \quad (29)$$

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

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

Here, $\tilde{y}(\mathbf{x}, \boldsymbol{\xi}) = \sum_{|\alpha|+|\beta|=0}^p c_{\alpha,\beta} \Phi_{\alpha}(\mathbf{x}) \Psi_{\beta}(\boldsymbol{\xi})$, 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 $y(\mathbf{x}, \boldsymbol{\xi})$ to the p -th order polynomial space, $\alpha_1 = N_p L T$, $\alpha_2 = 1 + N_p W$, $L = \max \|y(\mathbf{x}, \boldsymbol{\xi})\|_2$, $T = \max_{j_1+j_2, l_1+l_2=1, \dots, N_{2p}} \|\Phi_{j_1}(\mathbf{x}) \Psi_{j_2}(\boldsymbol{\xi}) \Psi_{l_1}(\mathbf{x}) \Psi_{l_2}(\boldsymbol{\xi})\|_2$, and $W = \sup \frac{\|y(\boldsymbol{\xi})\|}{\mathbb{E}[\|y(\boldsymbol{\xi})\|]}$ are constants.

Remark: This subsection 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 surrogate modeling and optimization, we may apply a high-dimensional solver such as the compressive sensing method [44] to build the surrogate model. Our framework shown in Fig. 1 is still applicable.

C. The 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:

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

However, the constraints are non-smooth because of the square-root terms, and may not admit a gradient at some points [49]. Instead, we use the equivalent smooth polynomial formula:

$$\kappa_{\epsilon_i}^2 \text{var}_{\boldsymbol{\xi}}[y_i(\mathbf{x}, \boldsymbol{\xi})] \leq (u_i - \mathbb{E}_{\boldsymbol{\xi}}[y_i(\mathbf{x}, \boldsymbol{\xi})])^2. \quad (33)$$

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

- Input:** The range of the design variable \mathbf{x} , probability density function 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 $\Phi_{\alpha}(\mathbf{x})$ and $\Psi_{\beta}(\boldsymbol{\xi})$ based on (22) and (23) independently.
 2. Initialize the quadrature points for design variables $\{\mathbf{x}_l, v_l\}_{l=1}^{M_1}$ 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 $\{\mathbf{x}_k, \boldsymbol{\xi}_k, w_k\}_{k=1}^M$ by (28).
 3. Call the simulator to compute $f(\mathbf{x}_k, \boldsymbol{\xi}_k)$, $y_i(\mathbf{x}_k, \boldsymbol{\xi}_k)$ for all $i = 1, \dots, n$ and $k = 1, \dots, M$.
 4. Build the coefficients $h_{\alpha,\beta}$ and $c_{\alpha,\beta}^i$ by equation (25).
 5. Set up the optimization problem (31), and then solve it via a global polynomial optimization solver, e.g., [52].

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

Consequently, (17) can be reduced to a deterministic and smooth optimization problem of \mathbf{x} in (31).

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 paper, we use the semi-definite relaxation based approaches [50], [51] 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 polynomial optimization sub-problem (31) is solved by the global optimization solver GloptiPoly 3 [52]. The yield is defined as

$$\text{yield}(\mathbf{x}) = \frac{\text{the number of } \boldsymbol{\xi}_j \text{ s.t. } y_i(\mathbf{x}, \boldsymbol{\xi}_j) \leq u_i, \forall i \in [n]}{\text{the total number of random parameters } \boldsymbol{\xi}_j}. \quad (34)$$

We set all risk thresholds to ϵ , i.e., $\epsilon = \epsilon_i, \forall i \in [n]$. 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 state-of-the-art Bayesian yield optimization method [10]. We summarize the key idea of the Bayesian yield optimization in Appendix C. The MATLAB codes and a demo example can be downloaded online ¹.

¹<https://www.dropbox.com/sh/stvpiz0aa14pmpm/AABKrCGKjWzQY8OrrD-GNLJAa?dl=0>

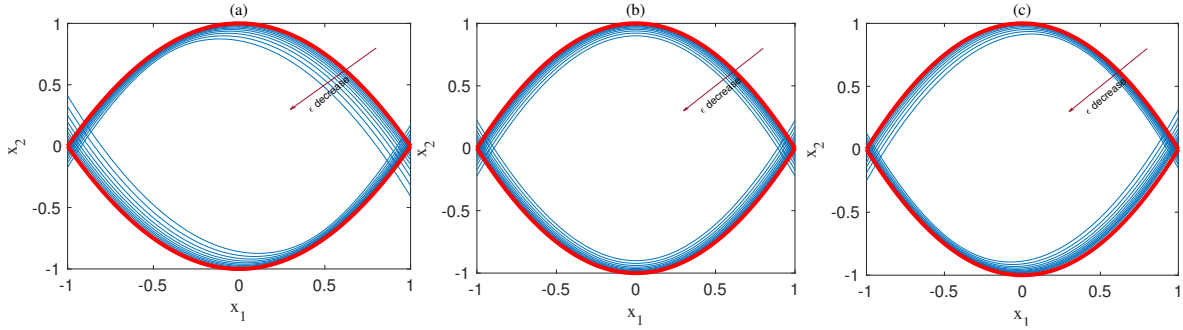


Fig. 2. The feasible set of the synthetic example with risk tolerance levels $\epsilon \in [10^{-2}, 10^{-0.1}]$ under different uncertainty distributions. (a): a positive-correlated non-Gaussian distribution; (b): a Gaussian independent distribution; (c): a negative correlated non-Gaussian distribution. 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.

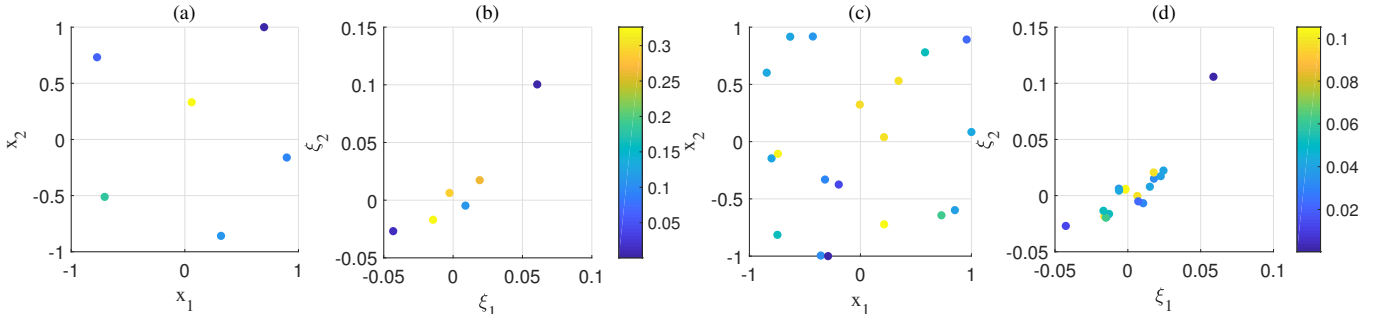


Fig. 3. The quadrature points and weights in the synthetic experiment. (a) and (b): The initial 2-D quadrature points for the design variable \mathbf{x} and uncertain parameters $\boldsymbol{\xi}$ by solving (26) and (27), respectively. (c) and (d): The optimized quadrature points for the joint 4-D space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

A. Synthetic Example

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

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

To illustrate the effects of different parameter distributions, we study three probability density functions: the independent distribution $\mathcal{N}(\mathbf{0}, 10^{-4}\mathbf{I})$, the non-Gaussian positive correlations $\frac{1}{2}\mathcal{N}(\mathbf{0.01}, 10^{-4}\boldsymbol{\Sigma}) + \frac{1}{2}\mathcal{N}(-\mathbf{0.01}, 10^{-4}\boldsymbol{\Sigma})$ with $\boldsymbol{\Sigma} = \begin{pmatrix} 1 & 0.75 \\ 0.75 & 1 \end{pmatrix}$, and the non-Gaussian negative correlations $\frac{1}{2}\mathcal{N}([0.01, -0.01]^T, 10^{-4}\boldsymbol{\Sigma}) + \frac{1}{2}\mathcal{N}([-0.01, 0.01]^T, 10^{-4}\boldsymbol{\Sigma})$ 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.

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

Algorithm	\mathbf{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 ($\epsilon = 0.20$)	0.9751	-0.0293	2.8959	87.49
(15)	0.9999	0	2.9997	41.66

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 $\{\mathbf{x}_l, v_l\}_{l=1}^6$ 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.

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 trade-off between the performance and yield

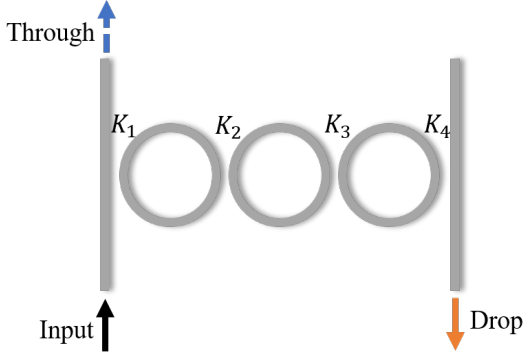


Fig. 4. An optical add-drop filter with three microrings coupled in series.

requirements. Compared with the solution $\tilde{x} = [0.9999, 0]^T$ 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 [53], [54]. A broad and flat passband with a high extinction ratio can be achieved by optimizing the coupling strengths between the microrings [53]. 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 $1.55 \mu\text{m}$. The design variables are the coupling coefficients $\mathbf{x} = [K_1, K_2, K_3, K_4]$ 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 ξ follows a non-Gaussian correlated distribution

$$p(\xi) = \frac{1}{2}\mathcal{N}(\mu_1, \Sigma) + \frac{1}{2}\mathcal{N}(\mu_2, \Sigma), \quad (36)$$

where $\mu_1 = -\mu_2 = 0.006[1, 1, 1, 1]^T$, and the variance is

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

We mainly focus on three metrics of the microring filter: the 3dB bandwidth (BW, in GHz), the extinction ratio (RE, in dB) of the transmission at the drop port, and 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:

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

TABLE II
OPTIMIZATION RESULTS FOR THE MICRORING ADD-DROP FILTER.

Algorithm	Simulations	$\mathbb{E}_{\xi}[\text{BW}]$ (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 25dB and 0.5dB, 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 \mathbf{x} by solving (26), 16 quadrature points for the parameters ξ by solving (27), and 64 quadrature points for the joint optimization of \mathbf{x} and ξ 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 Monte Carlo 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 Bayesian yield optimization (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 $\mathbb{E}_{\xi}[\text{BW}] = 115.6$ GHz with 99.8% yield at the optimal solution $\mathbf{x}^* = [0.5582, 0.4208, 0.3000, 0.6000]$, while BYO takes 2020 simulations to achieve the result of $\mathbb{E}_{\xi}[\text{BW}] = 112.3$ 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

$$\kappa = \exp\left(-\frac{g}{260}\right). \quad (38)$$

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

$$p(\xi) = \frac{1}{2}\mathcal{N}(\mu_1, \Sigma) + \frac{1}{2}\mathcal{N}(\mu_2, \Sigma), \quad (39)$$

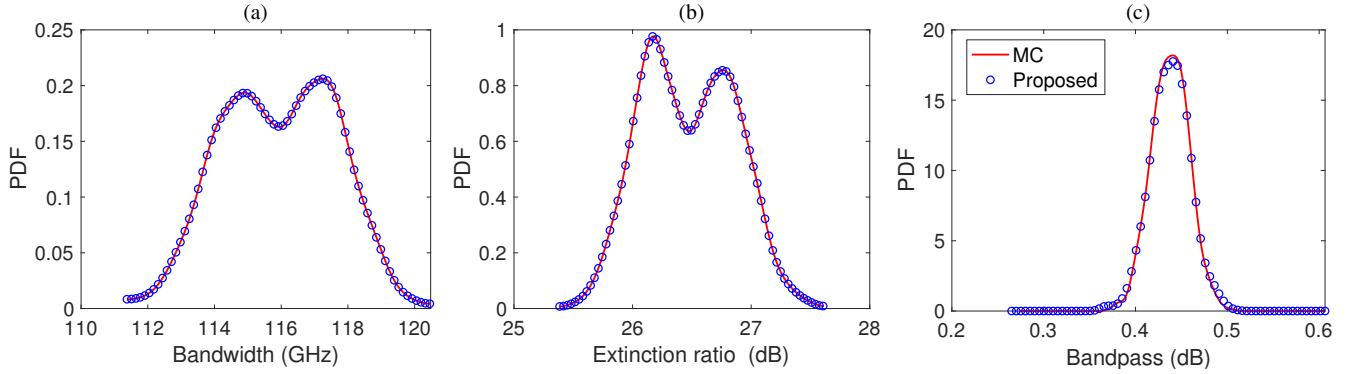


Fig. 5. The probability density functions (PDF) of the bandwidth, extinction ratio and roughness for the microring add-drop filter at the optimal solution $\mathbf{x}^* = [0.5582, 0.4208, 0.3000, 0.6000]$ by our proposed optimization with $\epsilon = 0.05$. Our surrogate model uses only 64 simulations,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

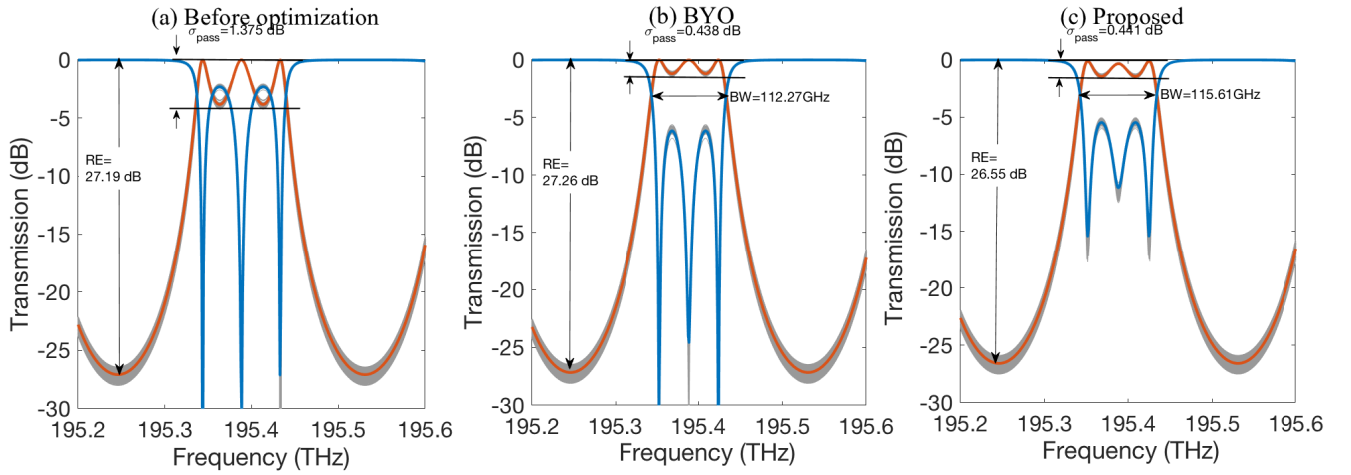


Fig. 6. The transmission curves of the microring add-drop filter at different design choices. The grey lines show the uncertainties caused by the process variations. The orange and blue curves show the mean transmission rates at the drop port and the through port, respectively. Here RE, BW and σ_{pass}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

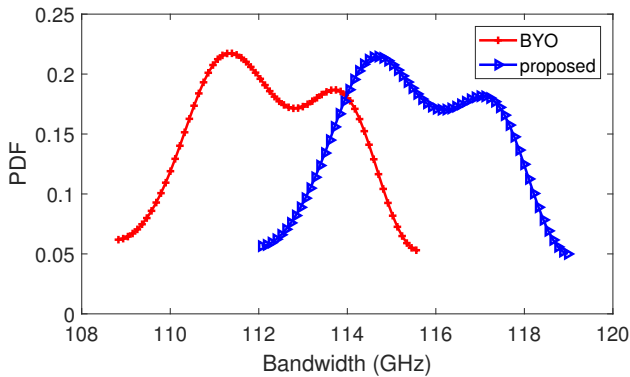


Fig. 7. The optimized bandwidth probability density distribution of the microring filter. Our chance-constrained optimization obtain an expected value of 115.6 GHz while the Bayesian yield optimization (BYO) only produces an expected value of 112.3 GHz.

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: the 3dB

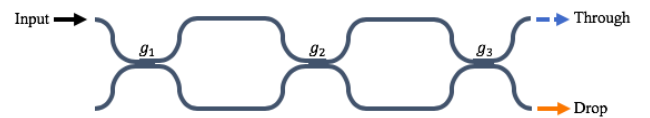


Fig. 8. The schematic of a third-order Mach-Zehnder Interferometer.

bandwidth (BW, in GHz), the crosstalk (XT, in dB), and the attenuation (α , in dB) of the peak transmission. The yield is defined through the crosstalk and the attenuation. The yield-aware optimization is formulated as:

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

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

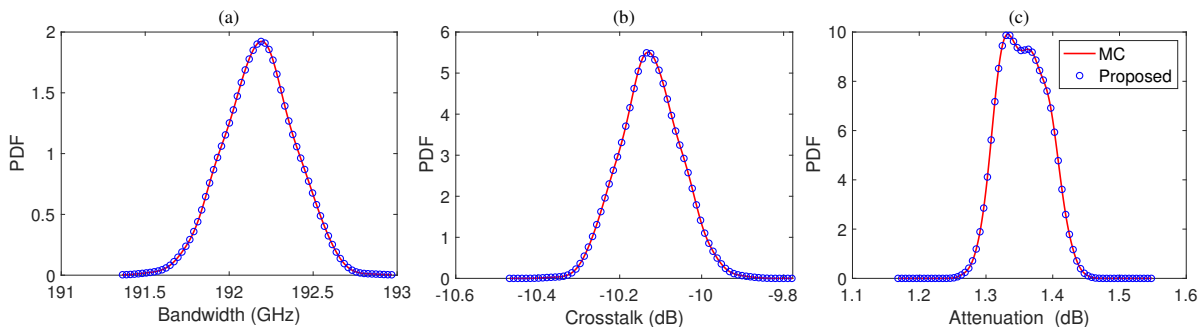


Fig. 9. The probability density functions (PDF) for the bandwidth, crosstalk, and attenuation of the MZI at our optimized design parameters $\mathbf{x}^* = [0.300, 0.5036, 0.300]$. Our surrogate model uses only 36 simulations and Monte Carlo (MC) uses 1000 simulations.

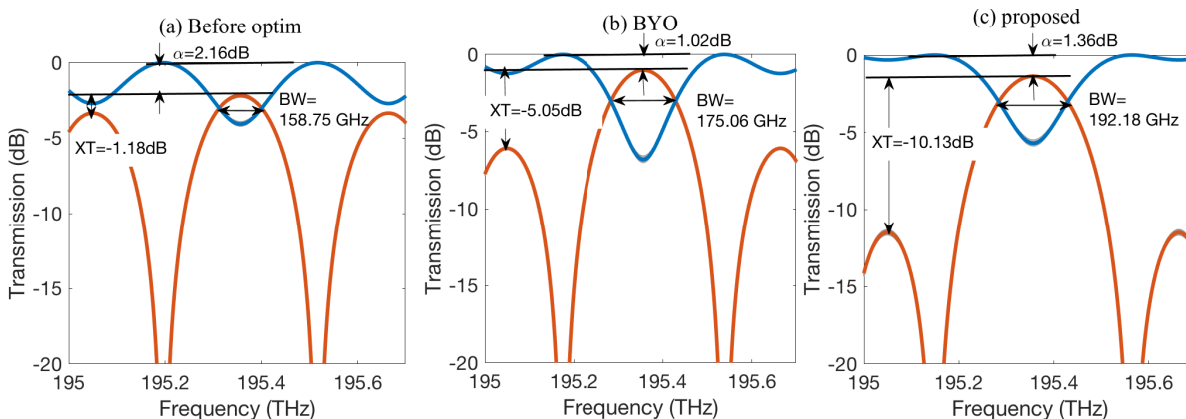


Fig. 10. The transmission curves of the MZI. The grey lines show the performance uncertainties. The orange and blue curves show the transmission rates at the drop and through ports, respectively. The mean values of the bandwidth, crosstalk and attenuation are denoted as BW, XT and α , respectively. (a)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

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 \mathbf{x} , 10 initial quadrature points for the uncertainty parameter ξ . 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 Monte Carlo with 10^3 samples.

We also compare our proposed method and BYO in Table III. Similar to the result in Table II, lower risk tolerance results in higher yield and lower expected value of bandwidth. Our method requires $56\times$ 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 $\mathbf{x}^* = [300, 111.2, 300]$ and its expected bandwidth is 192.2GHz. 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 Bayesian yield optimization and the initial design. Fig. 11 further shows the probability density of the optimized bandwidth by our chance-constrained optimization and the Bayesian yield optimization, respectively. It clearly shows that our proposed method produces higher bandwidth.

TABLE III
OPTIMIZATION RESULT FOR THE MZI.

Algorithm	Simulations	$\mathbb{E}_\epsilon[\text{BW}]$ (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

VI. CONCLUSIONS AND REMARKS

This paper 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 Bayesian yield optimization, our method has consumed much fewer simulation samples and produced better design performance while achieving the same yield.

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

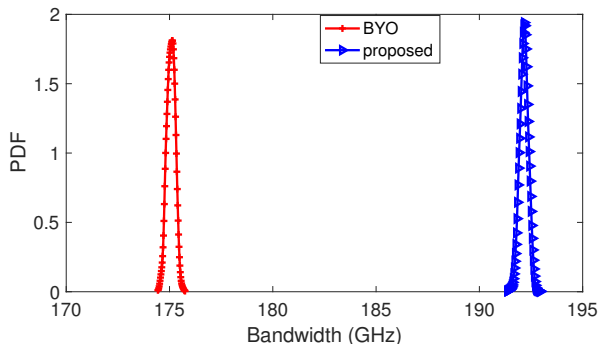


Fig. 11. The optimized bandwidth of the MZI by the Bayesian yield optimization and our proposed method, respectively. The expectation bandwidth of the Bayesian yield optimization is 175.4 GHz while our proposed method with $\epsilon = 0.05$ can get 186.4 GHz.

- **Non-Smoothness.** Similar to generalized polynomial chaos [46], the surrogate modeling techniques in [28], [29] require the stochastic functions to be smooth. However, performance metrics of a photonic IC may be non-smooth with respect to the design variables and process variations. How to handle non-smoothness in this optimization framework is a critical issue.
- **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.

ACKNOWLEDGMENT

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

APPENDIX A

DETAILED DERIVATION OF EQUATION (6)

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

$$\mathbb{E}_{\xi}[y(\mathbf{x}, \xi)] + \sqrt{(1-\epsilon)/\epsilon} \sqrt{\text{var}_{\xi}[y(\mathbf{x}, \xi)]} \leq u$$

is a sufficient but not necessary condition for the following probability constraint:

$$\text{Prob}_{\xi}(y(\mathbf{x}, \xi) \leq u) \geq 1 - \epsilon.$$

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 $\mathbf{X} = y(\mathbf{x}, \xi)$. Cantelli's inequality [55] states that for any random variable \mathbf{X} with a mean value $\mathbb{E}[\mathbf{X}] = \mathbb{E}_{\xi}[y(\mathbf{x}, \xi)]$ and variance $\sigma^2 = \text{var}_{\xi}[y(\mathbf{x}, \xi)]$, it holds that the probability of a single tail can be bounded as follows:

$$\text{Prob}(\mathbf{X} - \mathbb{E}[\mathbf{X}] \leq \lambda) \geq 1 - \frac{\sigma^2}{\sigma^2 + \lambda^2} \quad \text{if } \lambda > 0. \quad (41)$$

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

$$\begin{aligned} \text{Prob}(\mathbf{X} \leq u) &= \text{Prob}(\mathbf{X} - \mathbb{E}[\mathbf{X}] \leq u - \mathbb{E}[\mathbf{X}]) \\ &\geq 1 - \frac{\sigma^2}{\sigma^2 + (u - \mathbb{E}[\mathbf{X}])^2}. \end{aligned}$$

For any ϵ , a sufficient condition for $\text{Prob}(\mathbf{X} \leq u) \geq 1 - \epsilon$ is $1 - \frac{\sigma^2}{\sigma^2 + (u - \mathbb{E}[\mathbf{x}])^2} \geq 1 - \epsilon$, i.e.,

$$\mathbb{E}[\mathbf{X}] + \sqrt{(1-\epsilon)/\epsilon} \sigma \leq u. \quad (42)$$

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

APPENDIX B

DETAILED DERIVATION OF EQUATIONS (20) AND (21)

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

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

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

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

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

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

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

APPENDIX C

BAYESIAN YIELD OPTIMIZATION (BYO)

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

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

With the Bayes' theorem, it holds that $\text{Prob}(S|\mathbf{x}) = \frac{\text{Prob}(S)}{\text{Prob}(\mathbf{x})} \text{Prob}(\mathbf{x}|S)$. In our problem setting, $\text{Prob}(\mathbf{x})$ is a constant because we assume that \mathbf{x} follows a uniform distribution and $\text{Prob}(S)$ should also be a constant without the dependence

on the variable \mathbf{x} . Therefore, $\text{Prob}(S|\mathbf{x}) \propto \text{Prob}(\mathbf{x}|S)$ and the original yield optimization problem (3) is equivalent to

$$\mathbf{x}^{BYO} = \underset{\mathbf{x} \in \mathcal{X}}{\text{argmax}} \text{Prob}(\mathbf{x}|S). \quad (44)$$

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, $N = 100$ samples are generated randomly and a simulator is called to compute the quantity of interests at those samples. Then choose $M \leq N$ “pass” samples to perform the kernel density estimation

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

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 $\mathbf{x}^{BYO,t}$. 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 highest bandwidth

$$\mathbf{x}^{BYO} = \underset{\mathbf{x} \in \mathbf{x}^{BYO,t}}{\text{arg max}} \text{BW}(\mathbf{x}) \text{ s.t. pass}(\mathbf{x}^{BYO,t}) = 1.$$

REFERENCES

- [1] S. C. Nicholes, M. L. Masanovic, B. Jevremovic, E. Lively, L. A. Coldren, and D. J. Blumenthal, “The world’s first InP 8×8 monolithic tunable optical router (MOTOR) operating at 40 Gbps line rate per port,” in *Proc. Optical Fiber Communication*, 2009, pp. 1–3.
- [2] M. Kato, R. Nagarajan, J. Pleumeekers, P. Evans, A. Chen, A. Mathur, A. Dentai, S. Hurtt, D. Lambert, P. Chavarkar *et al.*, “40-channel transmitter and receiver photonic integrated circuits operating at a per channel data rate 12.5 Gbit/s,” in *National Fiber Optic Engineers Conference*. Optical Society of America, 2007, p. JThA89.
- [3] X. Chen, M. Mohamed, Z. Li, L. Shang, and A. R. Mickelson, “Process variation in silicon photonic devices,” *Applied optics*, 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,” *Journal of Lightwave Technology*, vol. 34, no. 13, pp. 3163–3170, 2016.
- [5] Z. Lu, J. Jhoja, J. Klein, X. Wang, A. Liu, J. Flueckiger, J. Pond, and L. Chrostowski, “Performance prediction for silicon photonics integrated circuits with layout-dependent correlated manufacturing variability,” *Optics express*, vol. 25, no. 9, pp. 9712–9733, 2017.
- [6] J. Pond, J. Klein, J. Flückiger, X. Wang, Z. Lu, J. Jhoja, and L. Chrostowski, “Predicting the yield of photonic integrated circuits using statistical compact modeling,” in *Integrated Optics: Physics and Simulations III*, vol. 10242, 2017, p. 102420S.
- [7] T. W. Weng, D. Melati, A. I. Melloni, L. Daniel *et al.*, “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. Intl. Conf. on Computer-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 Transactions on Circuits and Systems I: Regular Papers*, vol. 58, no. 12, pp. 2906–2918, 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 Automation Conference*, 2017, pp. 1–6.
- [11] M. Wang, W. Lv, F. Yang, C. Yan, W. Cai, D. Zhou, and X. Zeng, “Efficient yield optimization for analog and SRAM circuits via Gaussian process regression and adaptive yield estimation,” *IEEE Trans. Computer-Aided Design of Integrated Circuits and Systems*, vol. 37, no. 10, pp. 1929–1942, 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 Automation Conference*, 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. CAD of Integr. Circ. Syst.*, vol. 26, no. 1, pp. 2–15, 2006.
- [14] X. Li, J. Le, L. T. Pileggi *et al.*, “Statistical performance modeling and optimization,” *Foundations and Trends® in Electronic Design Automation*, 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,” *Frontiers of Information Technology & Electronic Engineering*, 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. Computer-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. CAD of Integrated Circuits and Systems*, vol. 33, no. 5, pp. 728–740, 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. CAD of Integrated Circuits and Systems*, 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. Components, Packaging and Manufacturing Technology*, vol. 7, no. 5, pp. 687–697, 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. Components, Packaging and Manufacturing Technology*, vol. 3, no. 7, pp. 1252–1258, 2013.
- [21] S. Vruthula, J. M. Wang, and P. Ghanta, “Hermite polynomial based interconnect analysis in the presence of process variations,” *IEEE Trans. Computer-Aided Design of Integrated Circuits and Systems*, vol. 25, no. 10, pp. 2001–2011, 2006.
- [22] 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, p. 15, 2008.
- [23] M. R. Ruffe, E. Gad, M. Nakhla, and R. Achar, “Generalized Hermite polynomial chaos for variability analysis of macromodels embedded in nonlinear circuits,” *IEEE Trans. Components, Packaging and Manufacturing Technology*, vol. 4, no. 4, pp. 673–684, 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 and South Pacific Design Automation Conference*, 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 transactions on very large scale integration (VLSI) systems*, vol. 18, no. 11, pp. 1556–1566, 2009.
- [26] A. Waqas, D. Melati, Z. Mushtaq, and A. Melloni, “Uncertainty quantification and stochastic modelling of photonic device from experimental data through polynomial chaos expansion,” in *Integrated Optics: Devices, Materials, and Technologies XXII*, vol. 10535. International Society for Optics and Photonics, 2018, p. 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,” *Optics express*, 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 Conference on Electrical Performance of Electronic Packaging and Systems*, 2018, pp. 57–59.
- [29] C. Cui and Z. Zhang, “Stochastic collocation with non-Gaussian correlated process variations: Theory, algorithms and applications,” *IEEE Trans. Components, Packaging and Manufacturing Technology*, vol. 9, no. 7, pp. 1362–1375, July 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,” *Optics Express*, vol. 23, no. 4, pp. 4242–4254, 2015.

- [31] 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.
- [32] P. Li, H. Arellano-Garcia, and G. Wozny, "Chance constrained programming approach to process optimization under uncertainty," *Computers & chemical engineering*, vol. 32, no. 1-2, pp. 25–45, 2008.
- [33] A. Mesbah, S. Streif, R. Findeisen, and R. D. Braatz, "Stochastic nonlinear model predictive control with probabilistic constraints," in *American Control Conference*, 2014, pp. 2413–2419.
- [34] L. Blackmore, M. Ono, A. Bektassov, and B. C. Williams, "A probabilistic particle-control approximation of chance-constrained stochastic predictive control," *IEEE Trans. Robotics*, vol. 26, no. 3, pp. 502–517, 2010.
- [35] H. Akhavan-Hejazi and H. Mohsenian-Rad, "Energy storage planning in active distribution grids: A chance-constrained optimization with non-parametric probability functions," *IEEE Transactions on Smart Grid*, vol. 9, no. 3, pp. 1972–1985, 2018.
- [36] 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 Systems*, vol. 32, no. 6, pp. 4880–4893, 2017.
- [37] A. Nemirovski and A. Shapiro, "Convex approximations of chance constrained programs," *SIAM Journal on Optimization*, vol. 17, no. 4, pp. 969–996, 2006.
- [38] G. Calafiore, L. El Ghaoui *et al.*, "Distributionally robust chance-constrained linear programs with applications," *Technical Report, DAUIN, Politecnico di Torino, Torino, Italy*, 2005.
- [39] W. Gautschi, "On generating orthogonal polynomials," *SIAM J. Sci. Stat. Comput.*, vol. 3, no. 3, pp. 289–317, Sept. 1982.
- [40] R. Ghanem and P. Spanos, *Stochastic finite elements: a spectral approach*. Springer-Verlag, 1991.
- [41] 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.
- [42] X. Li, "Finding deterministic solution from under-determined equation: large-scale performance variability modeling of analog/RF circuits," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 29, no. 11, pp. 1661–1668, 2010.
- [43] 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. IEEE Custom Integrated Circuits Conf.* San Jose, CA, Sept. 2014, pp. 1–8.
- [44] C. Cui and Z. Zhang, "High-dimensional uncertainty quantification of electronic and photonic IC with non-Gaussian correlated process variations," *IEEE Trans. Computer Aided Design of Integrated Circuits and Systems*, June 2019, doi: 10.1109/TCAD.2019.2925340.
- [45] —, "Uncertainty quantification of electronic and photonic ICs with non-Gaussian correlated process variations," in *Proc. Intl. Conf. Computer-Aided Design*, 2018, pp. 1–8.
- [46] D. Xiu and G. E. Karniadakis, "The Wiener–Askey polynomial chaos for stochastic differential equations," *SIAM journal on scientific computing*, vol. 24, no. 2, pp. 619–644, 2002.
- [47] 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.
- [48] T. Gerstner and M. Griebel, "Numerical integration using sparse grids," *Numer. Algor.*, vol. 18, pp. 209–232, Mar. 1998.
- [49] F. H. Clarke, *Optimization and nonsmooth analysis*. Siam, 1990, vol. 5.
- [50] H. Waki, S. Kim, M. Kojima, and M. Muramatsu, "Sums of squares and semidefinite program relaxations for polynomial optimization problems with structured sparsity," *SIAM Journal on Optimization*, vol. 17, no. 1, pp. 218–242, 2006.
- [51] J. Nie, "An exact Jacobian SDP relaxation for polynomial optimization," *Mathematical Programming*, vol. 137, no. 1-2, pp. 225–255, 2013.
- [52] D. Henrion, J.-B. Lasserre, and J. Löfberg, "GloptiPoly 3: moments, optimization and semidefinite programming," *Optimization Methods & Software*, vol. 24, no. 4-5, pp. 761–779, 2009.
- [53] R. Orta, P. Savi, R. Tascone, and D. Trincherio, "Synthesis of multiplexer-resonator filters for optical systems," *IEEE Photonics Technology Letters*, vol. 7, no. 12, pp. 1447–1449, 1995.
- [54] 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," *Journal of Lightwave Technology*, vol. 31, no. 24, pp. 3943–3955, 2013.
- [55] F. P. Cantelli, "Sui confini della probabilita," in *Atti del Congresso Internazionale dei Matematici: Bologna del 3 al 10 de settembre di 1928*, 1929, pp. 47–60.



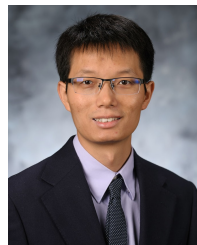
Chunfeng Cui received the Ph.D. degree in computational mathematics from Chinese Academy of Sciences, Beijing, China, in 2016 with a specialization in numerical optimization. From 2016 to 2017, she was a Postdoctoral Fellow at City University of Hong Kong, Hong Kong. In 2017 She joined the Department of Electrical and Computer Engineering at University of California Santa Barbara as a Postdoctoral Scholar.

Dr. Cui's research activities are mainly focused on the areas of tensor computing, uncertainty quantification, machine learning, and their interface. She is the recipient of the 2019 Rising Stars in Computational and Data Sciences, 2019 Rising Stars in EECSS, the 2018 Best Paper Award of IEEE Electrical Performance of Electronic Packaging and Systems (EPEPS), and the Best Journal Paper Award of Scientia Sinica Mathematica.



Kaikai Liu received the B.S. degree in physics in 2018 from Huazhong University of Science and Technology, Wuhan, China. In 2018 he joined the Department of Electrical and Computer Engineering at University of California Santa Barbara as a Ph.D. student.

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



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 (UCSB), CA. His research interests include uncertainty quantification with applications to the design automation of multi-domain systems (e.g., nano-scale electronics, integrated photonics, and autonomous systems), and tensor computational 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, 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, and the NSF CAREER Award in 2019.