Performance Variability Analysis of Photonic Circuits With Many Correlated Parameters

Abi Waqas, Paolo Manfredi, Daniele Melati

Performance Variability Analysis of Photonic Circuits with Many Correlated Parameters

Abi Waqas, *Member, IEEE.* Paolo Manfredi, *Senior Member, IEEE,* Daniele Melati, *Member, IEEE.*

***Abstract*—We propose a method to analyze the performance variability caused by fabrication uncertainty in photonic circuits with a large number of correlated parameters. By combining a sparse polynomial chaos expansion model with dimensionality reduction in the form of Karhunen-Loève transform and principal component analysis, we demonstrate the stochastic analysis of the transfer function of cascaded Mach-Zehnder interferometers with up to 38 correlated uncertain parameters.**

***Index Terms*—Correlated manufacturing variability, Karhunen-Loève transform, performance prediction, photonic devices, polynomial chaos, principal component analysis, process variations, silicon photonics, uncertainty quantification.**

I. Introduction

TECHNOLOGY progress in integrated photonics has resulted in complex photonic circuits combining many functions on a single chip, the possibility for significant production volumes and reduced fabrication costs [1], [2]. However, the effect of manufacturing variability remains a fundamental challenge. Photonic devices tend to be very sensitive to phase errors due to their inherently analog nature and their size, typically larger than the wavelength [3]. This means that small fabrication errors can significantly degrade their performance. Variability in the physical parameters of photonic components, e.g., waveguide width, thickness, gap aperture, or coupling length, results into deviations from the desired optical properties, such as the effective and group indices of waveguide modes, or coupling coefficients in directional couplers. As circuits become larger, deviations in the component performance propagate and accumulate hampering circuits fabrication yield, especially in high-index-contrast platforms [4], [5].

The simplest approach to study variability in photonic circuits is to exploit Monte Carlo sampling. Once the statistical information of the uncertain parameters is known, multiple samples can be generated based on these probability density functions (PDF) and the corresponding circuit behaviour can be obtained by means of a deterministic simulator. Although this approach is flexible and easy to implement, it is known for its slow convergence which requires a large number of samples to derive information on the stochastic behaviour of the circuit, thus imposing a great computational effort [6]. Alternatively, stochastic spectral methods based on polynomial chaos expansion (PCE) have emerged as a promising alternative, showing significant speedup over Monte Carlo in many engineering problems [7]. PCE has been also proposed for the variability analysis of photonic circuits [8], [9], [10], [11], [12], [13]. While classical PCE assumes random parameters to be mutually independent, photonic devices exhibit spatial correlation over rather long distances of few millimeters that must hence be taken into account to properly analyse the impact of fabrication uncertainty [3], [14]. Some PCE implementations proposed for photonics problems take into account parameters correlation [10], [13], but they either consider only a small set of random parameters or require the complex calculation of specialized polynomial basis. Indeed, one of the major drawbacks of PCE is that its computational cost grows exponentially with the number of stochastic variables, thus eventually becoming less efficient than Monte Carlo.

In this work, we tackle these issues by proposing an effective PCE implementation suitable for photonic circuits that can handle a large number of correlated uncertain parameters. The proposed method leverages the combination of several elements. Specifically, Karhunen-Loève transform (KLT) [15] is applied to decorrelate the physical parameters and reduce their number, thereby also reducing the complexity of the model to be trained. Principal component analysis (PCA) is then used to compress the training data and reduce the output quantities to be modeled. Therefore, the entire frequency-dependent transfer function is reduced to a small number of "principal components", for which an individual model is trained. An adaptive sparse PCE model [16] is built for the principal components, from which the complete transfer function is recovered via the inverse PCA transformation. The sparse implementation is particularly suited for training high-dimensional and/or high-order PCEs with a small amount of training data. Two test cases, i.e., a fifth-order and a seventh-order photonic filter realized with cascaded Mach-Zehnder interferometers, are used to demonstrate the proposed method. Filters have up to 38 correlated uncertain parameters and the variability analysis of their entire frequency-domain transfer function is performed more than 30 times faster than classical Monte Carlo analysis.

II. Proposed Method

We consider a generic system that depends on a set of d uncertain input parameters $\boldsymbol{x} = (x_1, \ldots, x_d)^\mathsf{T} \in \mathbb{R}^d$. We represent the system as

$$y = \mathcal{M}(\boldsymbol{x}), \tag{1}$$

A. Waqas is with the Department of Telecommunication, Mehran University of Engineering and Technology, Jamshoro, Pakistan (e-mail: abi.waqas@faculty.muet.edu.pk)

P. Manfredi is with the Department of Electronics and Telecommunications, Politecnico di Torino, Turin 10129, Italy (e-mail: paolo.manfredi@polito.it).

D. Melati is with the Centre de Nanosciences et de Nanotechnologies, CNRS, Université Paris-Saclay, 10 Bv. Thomas Gobert, 91120 Palaiseau, France (e-mail: daniele.melati@universite-paris-saclay.fr)

where $\mathcal{M}: \mathbb{R}^d \to \mathbb{R}$ is an implicit function that maps a given configuration of the input parameters $\boldsymbol{x}$ to the corresponding output y of the system. For the moment, we assume that the output of (1) is a scalar. We will later relax this assumption. In the context of this paper, $\boldsymbol{x}$ could be for example a collection of critical physical parameters of an optical device, e.g., the waveguide width or the gap of a directional coupler, and y an output of interest such as the transfer function at a given wavelength, the splitting ratio of a power splitter, or the bandwidth of a filter. The model can be equally applied to isolated components or entire photonic circuits.

A. Classical PCE

At this stage, we further assume that the parameters $\boldsymbol{x}$ are statistically independent. The classical PCE framework seeks an approximation of (1) in the form of

$$y \approx \hat{\mathcal{M}}(\boldsymbol{x}) = \sum_{\boldsymbol{\alpha}\in\mathcal{A}} c_{\boldsymbol{\alpha}} \varphi_{\boldsymbol{\alpha}}(\boldsymbol{x}) \tag{2}$$

where $\mathcal{A}$ is a set of multi-indices defined as [16]

$$\mathcal{A} = \mathcal{A}_q^{d,p} = \left\{ \boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_d) \in \mathbb{N}^d : \|\boldsymbol{\alpha}\|_q \leq p \right\}, \tag{3}$$

$\varphi_{\boldsymbol{\alpha}}$ are multivariate orthonormal polynomials, and $c_{\boldsymbol{\alpha}}$ are the pertinent coefficients. The polynomial basis functions are constructed as the product

$$\varphi_{\boldsymbol{\alpha}}(\boldsymbol{x}) = \prod_{j=1}^{d} \phi_{\alpha_j}(x_j), \tag{4}$$

where ϕ_{α_j} is a univariate polynomial in the variable x_j, with $\deg(\phi_{\alpha_j}) = \alpha_j$. Therefore, the multi-indices define the order of the basis function in each variable. For example, the polynomial $\varphi_{(2,0,1)}$ is quadratic in x_1, constant in x_2, and linear in x_3.

The univariate polynomials ϕ_{α_j} in (4) are orthonormal based on the inner product

$$\left\langle \phi_{\alpha_j}, \phi_{\alpha_m} \right\rangle = \int_{\mathbb{R}} \phi_{\alpha_j}(x_j)\phi_{\alpha_m}(x_j) w(x_j) dx_j = \delta_{\alpha_j,\alpha_m}, \tag{5}$$

where $w(x_j)$ is the probability density function of x_j. This ensures also the orthogonality of the multivariate polynomials $\varphi_{\boldsymbol{\alpha}}$, as well as an exponential decay of the error norm

$$\|\mathcal{M} - \hat{\mathcal{M}}\|_2^2 = \int_{\mathbb{R}^d} \left(\mathcal{M}(\boldsymbol{x}) - \hat{\mathcal{M}}(\boldsymbol{x}) \right)^2 w(\boldsymbol{x}) d\boldsymbol{x} \tag{6}$$

for functions with finite variance [17], where $w(\boldsymbol{x})$ is the joint probability density function of the variables $\boldsymbol{x}$, as the PCE order p is increased. It should be noted that the error (6) is defined in statistical terms: a larger deviation is tolerated for values of $\boldsymbol{x}$ with lower probability.

A useful property of PCEs is that the first two statistical moments (i.e., the mean and the variance) are available in closed form as a function of the coefficients, i.e.,

$$\mathrm{E}\{y\} \approx \mathrm{E}\{\hat{\mathcal{M}}\} = \int_{\mathbb{R}^d} \hat{\mathcal{M}}(\boldsymbol{x}) w(\boldsymbol{x}) d\boldsymbol{x} = c_{\mathbf{0}}, \tag{7}$$

with $\mathbf{0}$ denoting the null element of $\mathbb{N}^d$, and

$$\begin{aligned} \mathrm{Var}\{y\} \approx \mathrm{Var}\{\hat{\mathcal{M}}\} &= \int_{\mathbb{R}^d} \left(\hat{\mathcal{M}}(\boldsymbol{x}) - c_{\mathbf{0}} \right)^2 w(\boldsymbol{x}) d\boldsymbol{x} \\ &= \sum_{\boldsymbol{\alpha}\in\mathcal{A}\setminus\mathbf{0}} |c_{\boldsymbol{\alpha}}|^2 \end{aligned} \tag{8}$$

Several choices are commonly used for choosing the indices in (3) based on the parameter q. The most popular is certainly $q = 1$, leading to the so-called "total-degree" truncation, with a number of terms in (2) (i.e., the cardinality of $\mathcal{A}$) given by

$$|\mathcal{A}| = |\mathcal{A}_1^{d,p}| = \frac{(p+d)!}{p!d!}. \tag{9}$$

Another common choice is $q = \infty$, which however leads to excessively large expansions. Instead, setting $q < 1$ provides increasingly sparser expansions that neglect higher-order interactions between input parameters, which is often acceptable [16].

B. Calculation of PCE Coefficients

There exist several strategies for the calculation of the PCE coefficients [18]. The most popular are the stochastic Galerkin method, which is intrusive and requires to modify the system of governing equations, numerical integration techniques that approximate projection integrals, and least-square regression or interpolation approaches. Here, we focus on the third ones, which are the most straightforward, versatile, and scalable.

Consider a set $\{\boldsymbol{x}_\ell\}_{\ell=1}^L$ of "training samples" of the input parameters, and a vector of corresponding observations $\boldsymbol{y} = (y_1, \ldots, y_L)^\mathsf{T} \in \mathbb{R}^L$, with $y_\ell = \mathcal{M}(\boldsymbol{x}_\ell)$, $\forall \ell = 1, \ldots, L$. The least-square regression minimizes the norm of the residual between the model prediction at the training samples and the observations. The solution is cast as

$$\boldsymbol{c}^* = \arg\min_{\boldsymbol{c}} \|\boldsymbol{\Psi}\boldsymbol{c} - \boldsymbol{y}\| = \boldsymbol{\Psi}^+ \boldsymbol{y}, \tag{10}$$

where $\boldsymbol{c}^* \in \mathbb{R}^K$ is a vector collecting all PCE coefficients, $\boldsymbol{\Psi} \in \mathbb{R}^{L\times K}$ is a matrix with entries $\Psi_{\ell,\boldsymbol{\alpha}} = \varphi_{\boldsymbol{\alpha}}(\boldsymbol{x}_\ell)$, and $\boldsymbol{\Psi}^+ = (\boldsymbol{\Psi}^\mathsf{T}\boldsymbol{\Psi})^{-1}\boldsymbol{\Psi}^\mathsf{T}$ is its Moore-Penrose pseudo-inverse, which generalizes the matrix inverse for non-square matrices.

For an acceptable solution, the regression problem needs to be overdetermined, i.e., $L > K$ (typically, at least two to three times larger). Plain random sampling, Latin hypercube sampling (LHS), or quasi-MC low-discrepancy sequences [19] can be used to generate the training samples, with an impact on the convergence performance.

C. Sparse PCE

The standard approach rapidly becomes computationally intractable when the number of input parameters increases. For example, with an expansion order as low as $p = 2$, $d = 30$ variables, and $q = 1$, the number of basis functions is $K = 496$. Therefore, a number of training samples in the order of 10^3 is needed, with a computational effort that becomes comparable to a MC analysis. If order $p = 4$ is required, the number of training samples becomes of the order of 10^5, while remaining of the order of a few thousands even if a sparser expansion with $q = 0.7$ is adopted.

Regardless of the initial sparsity of set $\mathcal{A}_q^{d,p}$, typically most of the PCE coefficients further turns out to be negligible as a consequence of the "sparsity-of-effects principle" [16]. Therefore, sparse approaches were developed to identify only non-negligible coefficients, with a reduced requirement in terms of training samples. This allows extending the PCE framework to problems characterized by a large number of uncertain input parameters and/or requiring high-order expansions.

In this paper, we make use of the UQLab toolbox [20] to calculate sparse PCEs. The toolbox implements the adaptive method in [16], which is based on least-angle regression. Given the dimensionality d, the truncation parameter q, and a vector of possible p-values, the algorithm identifies the optimal expansion order p as well as the coefficients of a sparser subset of basis functions $\tilde{\mathcal{A}} \subset \mathcal{A}_q^{d,p}$, with $|\tilde{\mathcal{A}}| \ll |\mathcal{A}_q^{d,p}|$. We refer the Reader to the toolbox documentation for further information.

D. Correlated Variables: Karhunen-Loève Transform

The method so far assumes that the uncertain input parameters are mutually independent. In the most general case of parameters with arbitrary correlation, customized basis functions can be numerically computed [21]. However, for the special case of Gaussian correlated variables, KLT can be effectively used to decorrelate the variables [15].

Assume that $\boldsymbol{x}$ are correlated Gaussian variables with mean vector $\boldsymbol{\mu} \in \mathbb{R}^d$ and covariance matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{d\times d}$. The variables $\boldsymbol{x}$ can be expressed in terms of a set of independent and standard (i.e., with zero mean and unit variance) Gaussian variables $\boldsymbol{\xi} = (\xi_1, \dots, \xi_d)^\mathsf{T} \in \mathbb{R}^d$ as

$$\boldsymbol{x} = \boldsymbol{\mu} + \sum_{k=1}^{d} \sqrt{\lambda_k}\xi_k \boldsymbol{p}_k \tag{11}$$

where λ_k and $\boldsymbol{p}_k$ are the eigenvalues and corresponding eigenvectors of the covariance matrix $\boldsymbol{\Sigma}$. It should be noted that a trivial application of (11) is normally used to renormalize independent uncertain parameters to a common support, so that a standard set of univariate basis functions can be used for each variable. With the transformation 11, the PCE framework introduced so far can be seamlessly applied in terms of the variables $\boldsymbol{\xi}$.

Furthermore, for highly-correlated variables, the eigenvalues decay very fast. Hence, the KLT (11) can be truncated in such a way that the d variables $\boldsymbol{x}$ are expressed in terms of a *reduced* set of $\tilde{d}$ independent random variables [22], i.e.,

$$\boldsymbol{x} \approx \boldsymbol{\mu} + \sum_{k=1}^{\tilde{d}} \sqrt{\lambda_k}\xi_k \boldsymbol{p}_k \tag{12}$$

where the eigenvalues are now assumed to be sorted in decreasing order and $\tilde{d}$ is such that

$$\frac{\lambda_k}{\lambda_1} \leq \varepsilon_{\mathrm{KLT}}, \tag{13}$$

for $k > \tilde{d}$ and a prescribed threshold $\varepsilon_{\mathrm{KLT}}$. This dimensionality reduction allows alleviating the computational complexity of the PCE model.

E. Multiple Outputs: Principal Component Analysis

Another difficulty arises when the output of system (1) is not a scalar. This is the case, for instance, when a transfer functions is computed at different wavelengths within a given frequency band. The solution of the standard regression approach (10) is readily applied in a vectorized manner to the output vector thanks to the fact that the basis functions, and hence the regression matrix $\boldsymbol{\Psi}$, are the same for each component. On the contrary, sparse methods, such as that considered in this work, identify the basis functions based on the output data, and therefore the subset $\tilde{\mathcal{A}}$ potentially differs for each component of the vector of the outputs. One trivial solution is to apply the method individually to each output, which however is computationally inefficient when their number is large.

An effective strategy is to compress, using PCA, the output variables into a reduced set of components, for which the training of individual PCE models becomes feasible [23]. Let us assume that a set of L training observations are available for M outputs of interest, and stored into a matrix $\boldsymbol{Y} \in \mathbb{R}^{M\times L}$. Using an "economy-size" singular value decomposition (SVD), the matrix dataset is expressed as [24]

$$\boldsymbol{Y} = \bar{\boldsymbol{y}} + \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^\mathsf{T} = \bar{\boldsymbol{y}} + \boldsymbol{U}\boldsymbol{Z} \tag{14}$$

where $\bar{\boldsymbol{y}} \in \mathbb{R}^M$ is the columnwise mean of $\boldsymbol{Y}$ over the training observations, $\boldsymbol{U} \in \mathbb{R}^{M\times L}$, $\boldsymbol{V} \in \mathbb{R}^{L\times L}$, and $\boldsymbol{S} \in \mathbb{R}^{L\times L}$ is a diagonal matrix collecting the singular values $\{\sigma_i\}_{i=1}^L$ in decreasing order.

Since the singular values typically decay very fast due to the interdependency between the outputs, the SVD is truncated to retain only $\tilde{n}$ "principal components", leading to

$$\boldsymbol{Y} \approx \bar{\boldsymbol{y}} + \sum_{i=1}^{\tilde{n}} \boldsymbol{u}_i \boldsymbol{z}_i \tag{15}$$

where $\{\boldsymbol{u}_i\}_{i=1}^M \in \mathbb{R}^M$ are the first $\tilde{n}$ left-singular vectors of $\boldsymbol{Y} - \bar{\boldsymbol{y}}$, whereas $\{\boldsymbol{z}_i\}_{i=1}^{\tilde{n}} \in \mathbb{R}^L$ are in fact observations of the principal components. Similarly to the truncated KLT (12), $\tilde{n}$ is selected such that

$$\frac{\sigma_i}{\sigma_1} \leq \varepsilon_{\mathrm{PCA}} \tag{16}$$

for $i > \tilde{n}$. The KLT and the PCA are in fact equivalent, since the latter can be cast as the eigenvalue decomposition of the experimental covariance matrix of the dataset $\boldsymbol{Y}$ [24]. However, the purpose of KLT is to decorrelate and reduce the number of input parameters, whereas PCA is used to compress output data. In all simulations, we set $\varepsilon_{\mathrm{KLT}} = \varepsilon_{\mathrm{PCA}} = 10^{-3}$.

The observations $\{\boldsymbol{z}_i\}_{i=1}^{\tilde{n}}$ are used to train individual PCE models for the principal components, which is feasible because typically $\tilde{n} \ll M$. Once a model is available for the principal components, it is used to generate predictions thereof, while predictions of the original outputs are recovered via (15).

III. Application and Numerical Results

The method proposed in Sec. II can be applied to any photonic circuit design, including for example resonant structures, since no assumption is done in its formulation on the type

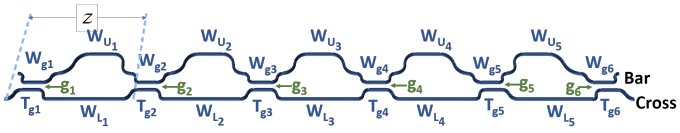

Fig. 1. Schematic of the fifth-order unbalanced Mach-Zehnder filter.

of system response (1). Here, a fifth-order and a seventh-order photonic filter realized with cascaded Mach-Zehnder interferometers are chosen as examples to demonstrate the performance of the proposed approach for sets of random variables of different sizes. In the following, we consider as a reference the typical parameter values for the silicon-on-insulator technology. The complex frequency response (transfer function) of the photonic filters is the stochastic process under investigation. The fabrication uncertainty is represented as correlated and Gaussian-distributed random variables.

A. Fifth-Order Unbalanced Mach-Zehnder Filter

The fifth-order Mach-Zehnder filter is assembled using two standard elementary building blocks, i.e., waveguides and directional couplers, as schematically shown in Fig. 1. The nominal design of the filter is obtained with the synthesis technique described in [25]. The nominal 3-dB bandwidth is $BW_0 = 39$ GHz and the coupling coefficients for the six directional couplers are $K_1 = K_6 = 0.993$, $K_2 = K_5 = 0.03$, and $K_3 = K_4 = 0.421$, with an in-band isolation larger than 30 dB. The dependence of the coupling constant of the directional couplers on waveguide gap, width, and thickness is first determined using coupled mode theory and preliminary electromagnetic simulations. For the nominal case, a waveguide width of 500 nm, thickness of 220 nm, and gap of 300 nm allow obtaining the required coupling coefficients K_1-K_6 with the coupling lengths $Lc_1 = Lc_6 = 26.38$ μm, $Lc_2 = Lc_5 = 3.09$ μm, and $Lc_3 = Lc_4 = 12.53$ μm. The effect of the transition regions in the couplers is neglected. For the waveguide building block, nominal width and thickness are again 500 nm and 220 nm, respectively, corresponding to an effective index and group index of 2.44 and 4.18. All the five stages of the filter have the same unbalance length of 681 μm, corresponding to a free spectral range of 100 GHz. Figure 2 shows, with bold lines, the ideal transfer function of the Mach-Zehnder filter at the bar and the cross ports, respectively.

Next, different sources of fabrication uncertainties are considered. For the waveguides in the two branches of each filter stage we assume an uncertainty only in the waveguide width, resulting in variability of the effective and group indices of the waveguide mode. The main effect of this variability will be a wavelength shift of the response of each stage. For the directional couplers, we consider a larger number of uncertainties, i.e., on the width and thickness of the waveguides and on the separating gap. Their collective effect is expected to mainly influence the resulting power coupling coefficients. [4]. The vector $\boldsymbol{x}$ of the uncertain input parameters is then split into two components, $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$. The former collects parameters W_{g_i} (waveguide width of the i-th coupler), g_i (coupling gap of the i-th coupler), W_{U_j} (width of the upper arm of the j-th stage), and W_{L_j} (width of the lower arm of the j-th stage), with $i = 1, \ldots, 5$ and $j = 1, \ldots, 6$. The vector $\boldsymbol{x}_2$ collects the thicknesses T_{g_j} of the coupler waveguides. The total number of uncertain parameter is thus $d = 28$. The aforementioned parameters are assumed to be Gaussian distributed, and specifically $W_{g_i} \sim \mathcal{N}(500\text{ nm}, 1\text{ nm})$, $g_i \sim \mathcal{N}(300\text{ nm}, 3\text{ nm})$, $W_{U_j}, W_{L_j} \sim \mathcal{N}(500\text{ nm}, 10\text{ pm})$, $T_{g_j} \sim \mathcal{N}(220\text{ nm}, 1\text{ nm})$. Organizing the parameters in two different groups allows defining a rather complex and realistic correlation scenario. The parameters in either of the two vectors $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are assumed to be mutually correlated among themselves, while no correlation is introduced between parameters belonging to different sets. This reflects the fact that the source of uncertainty is different for the two sets of parameters, mainly originating from lithography and etching for the first group, and from wafer fabrication for the second.

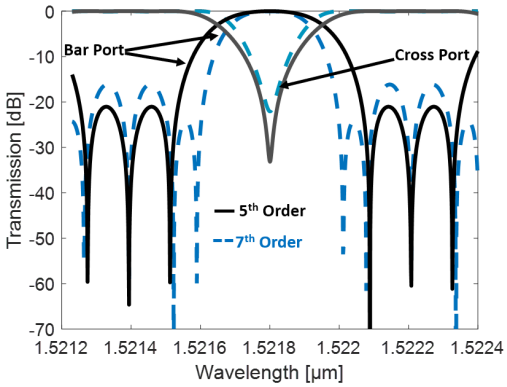

Fig. 2. Transfer function of the fifth-order (bold balck lines) and seventh-order (bold blue lines) Mach-Zehnder filter at the bar and the cross ports for the nominal design

The pertinent correlation coefficients are calculated by considering the correlation function [6] $g(z) = e^{-z^2/2l^2}$, where l is the correlation length and z is the distance between the circuit components along the horizontal direction. As an example, the vertical dashed lines in Fig. 1 highlight the distance between the first and the second directional coupler, which is used for the definition of z in the calculation of the correlation between the parameters of these two components. A realistic correlation length of $l = 4.5$ mm is considered for both parameter groups [14]. Since the distance between the upper and lower waveguide arms is relatively small in an unbalanced Mach-Zehnder filter, a correlation coefficient of one is considered between their parameters. This means that, in the simulation, the same variation is assigned to the upper and lower waveguides.

B. Seventh-Order Unbalanced Mach-Zehnder Filter

The seventh-order Mach-Zehnder filter is assembled in the same way as described above, with seven delay sections and eight directional couplers. The nominal design exhibits a nominal 3-dB bandwidth of $BW_0 = 29$ GHz. The coupling coefficients of the eight directional couplers are $K_1 = K_8 = 0.9914$, $K_2 = K_7 = 0.04$, $K_3 = K_6 = 0.0690$, and $K_4 = K_5 = 0.3489$, with an in-band isolation larger

than 20 dB. Width and thickness of the waveguides are the same as for the fifth-order filter. The nominal gap of the directional couplers is again 300 nm, leading to coupling lengths $Lc_1 = Lc_8 = 26.22$ μm, $Lc_2 = Lc_7 = 0.35$ μm, $Lc_3 = Lc_6 = 4.71$ μm, and $Lc_4 = Lc_5 = 11.21$ μm. All the seven stages have again the same unbalance length of 681 μm, corresponding to a free spectral range of 100 GHz. For each building block, the same geometrical parameters as for the previous test case are considered to be uncertain, with the same probability distribution and correlation. Owing to the increased number of stages, the total number of uncertain parameters is now $d = 38$.

C. Simulation Results

Circuit analysis is performed using circuit-level simulations. In each filter stage, waveguide sections (including both upper and lower waveguide arms) and directional couplers are building blocks with 2 input and 2 output ports that are represented by means of 2×2 transfer matrices with complex coefficients [25]. The covariance matrix of the input parameters is first computed as described in Section III-A, taking into account the relative distance of each building block within the circuit. For each simulation run, the value of the parameters is then determined according to their mean and the computed covariance matrix, and used to calculate the coefficients of the transfer matrix of each building block. Lastly, transfer matrices are multiplied according to the building block connections to obtain the complex transfer function of the circuit in the spectral domain. In each simulation, 1000 wavelength samples are used.

For both test circuits, a Monte Carlo analysis is performed running 10000 simulations and it is used as a reference. As an example, Fig. 3 (a) and (c) report, with thin gray lines, the absolute value of some of the computed transfer functions (transmission) of the fifth-order filter at the bar and cross ports, respectively. The black dashed line marks $\lambda = 1521.8$ nm, which is the designed central wavelength of the filter. As can be seen, the transmission is severely affected, especially with large fluctuations of the 3-dB pass-band as well as a reduced isolation both in band (cross port) and out of band (bar port). Remarkably, these large fluctuations results from relatively small variations of the geometrical parameters. Indeeed, a variation of 10 pm in the waveguide width corresponds to an effective index variation of $8 \cdot 10^{-5}$, whereas a coupling gap variation of 1 nm corresponds to a power coupling coefficient variation of $5 \cdot 10^{-3}$. In the two figures, the mean (blue solid line, computed at each wavelength as the absolute value of the mean of the complex transfer function) and standard deviation (blue dashed line, again computed using the complex transfer function) of the Monte Carlo samples are reported as well. To further investigate the effect of variability, Fig. 3(b) and (d) show the probability density function of the transmission at the bar and cross port, respectively, obtained from the Monte Carlo samples at $\lambda = 1521.8$ nm (blue bars). While the transmission at the bar port is confirmed to be almost unaffected by the variability, the isolation at the cross port varies across a range of more than 30 dB, with -20 dB as the most probable value compared to the -32 dB of the ideal design without variability.

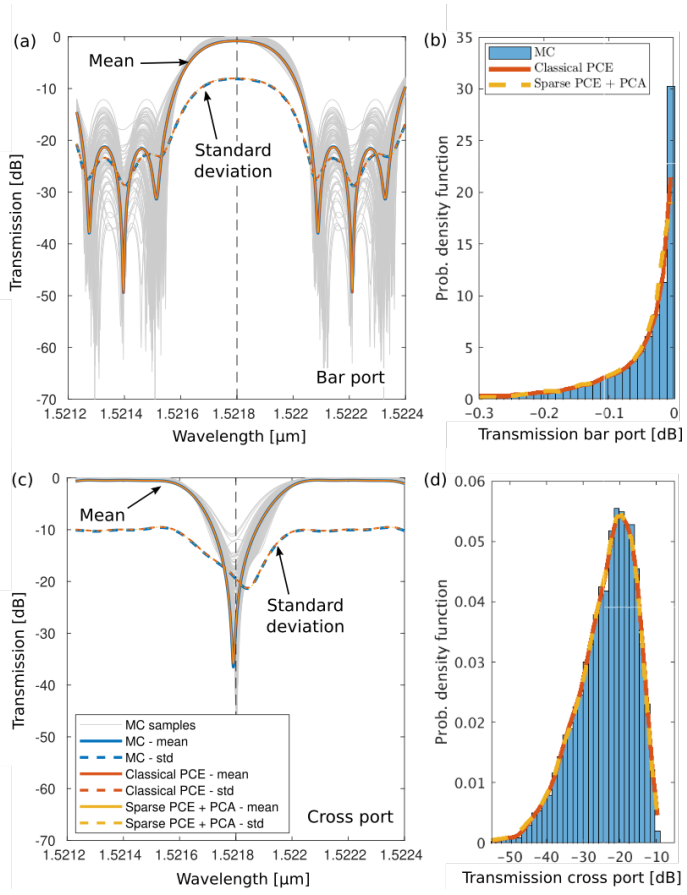

Fig. 3. Stochastic analysis of the transfer function for the fifth-order Mach-Zehnder filter (28 random variables). Magnitude of the transfer function at the bar (a) and cross (c) port is considered. Solid gray lines represents a subset of the Monte Carlo samples. Other solid lines show the mean computed with Monte Carlo (blue), classical PCE (red), and sparse PCE with PCA (yellow). Dashed lines are the standard deviations computed with the same three methods. The probability density functions computed at $\lambda = 1521.8$ nm, dashed black line in (a) and (c), are reported as well for the bar port (b) and the cross port (d). The reference histogram (blue bars) is obtained from the Monte Carlo samples while solid red and dashed yellow lines report the results of classical PCE and sparse PCE with PCA, respectively.

In order to test the PCE-based methods, we first apply KLT to the 28 input correlated variables and we train a classical PCE model to compare its performance with our proposed method. By considering a threshold of $\varepsilon_{\text{KLT}} = 10^{-3}$ for the truncation, the original correlated variables are reduced to a set of $\tilde{d} = 5$ independent variables. A different PCE model needs to be trained for each of the 1000 wavelength samples of the complex transfer function of the filter at the bar and cross port. For each model, we select an order $p = 6$ and $q = 1$, resulting in $K = 462$ coefficients, and we use $L = 3K = 1386$ samples for the training, drawn according to a LHS design. It should be noted that, without KLT truncation, the number of PCE coefficients would have been $\sim 1.3 \cdot 10^6$, thus making the model training unfeasible.

Mean and standard deviation of the transfer functions are computed using (7) and (8) and the absolute values are reported in Fig. 3(a) and (b) for the two ports using red solid and dashed lines, respectively. PDFs of the transmission at the two ports at the central wavelength, efficiently computed by sampling the model with the same samples used for the

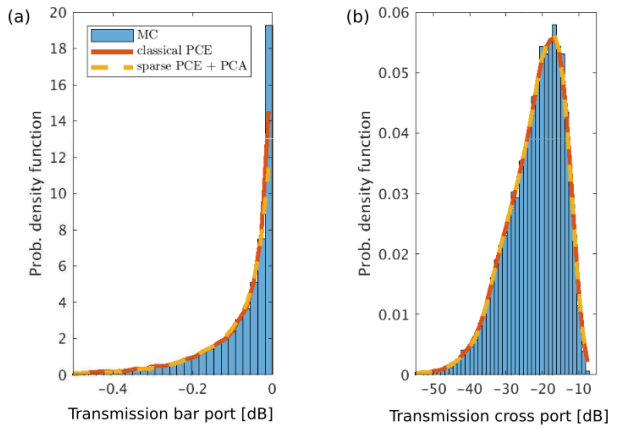

Fig. 4. Stochastic analysis of the seventh-order Mach-Zehnder filter (38 random variables). Probability density functions of the transmission at $\lambda = 1521.8$ nm for the bar port (a) and cross port (b) are computed with Monte Carlo (blue bars), classical PCE (red solid lines) and the proposed sparse PCE with PCA (yellow dashed lines).

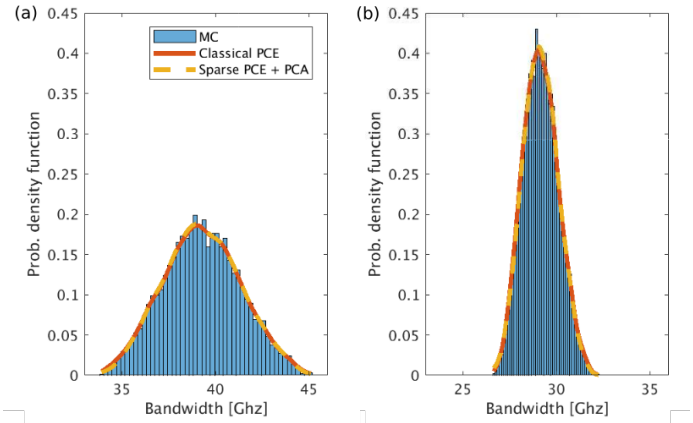

Fig. 5. Probability density function of the 3-dB bandwidth of the cross port at $\lambda = 1521.8$ nm. (a) Fifth-order filter and (b) seventh-order filter. The reference histogram is computed with Monte Carlo. Results for classical and sparse PCE are reported with solid red and dashed yellow lines, respectively.

reference Monte Carlo analysis, are shown with red solid lines in Fig. 3(c) and (d). Results are in excellent agreement.

Lastly, we train the sparse PCE model described in Section II-C for the same set of 5 independent variables resulting from the truncation of the KLT. For the adaptive training, we set the expansion order to be in the range $[2, 6]$, and we use only $L = 100$ samples, again drawn from a LHS design. The use of PCA with a truncation threshold of $\varepsilon_{\text{PCA}} = 10^{-3}$ allows to vastly compress the 2000 wavelength samples (1000 for each of the two ports, bar and cross) to a mere $\tilde{n} = 9$ principal components. Individual sparse PCE models needs hence to be trained only for these principal components. It is worth noting that the UQLab toolbox only handles real data, therefore a separate model is trained for the real and imaginary part of the principal components, leading to a total of 18 models to be computed. The number of non-zero coefficients identified for these models by the adaptive algorithm ranges from 13 to 38. Results for the mean and standard deviation of the transmission are reported in Fig. 3(a) and (b) for the two ports using yellow solid and dashed lines, respectively, while probability density functions at the central wavelength are shown in Fig. 3(c) and (d) with yellow dashed lines. Results are once again in excellent agreement with the Monte Carlo analysis. However, compared to the classical PCE, a much smaller number of training samples is needed.

A similar analysis is performed for the seventh-order Mach-Zehnder filter, comprising 38 random variables. The probability density functions of the bar and cross transmissions at $\lambda = 1521.8$ nm are reported in Fig. 4(a) and (b), respectively. Blue bars refer to the results of the Monte Carlo analysis, red solid lines to the classical PCE computed after the application of the truncated KLT, and yellow dashed lines to the proposed sparse method in conjunction with PCA. By using the same truncation thresholds for the KLT and PCA as in the previous test case, the initial set of correlated parameters is reduced to $\tilde{d} = 7$ independent variables, whereas the output wavelength samples is compressed once again to only $\tilde{n} = 9$ principal components. An order of $p = 6$ is again used for the classical PCE, leading to $K = 1716$ coefficients. Also in this case, the training of the classical PCE model would have been unfeasible without the preliminary KLT truncation. Such a reduced model is instead trained with $L = 3K = 5184$ samples. For the adaptive sparse scheme, the same (maximum) order is used, and the models of the principal components is trained with only $L = 200$ samples. In both cases, training samples are generated from a LHS design. The number of non-zero coefficients identified ranges from 15 to 120. As can be seen in Fig. 4, similarly to the fifth-order filter, the transmission at the bar port is only minimally impacted by the variability, while the rejection at the cross port varies across a wide range. Here the most probable rejection value is -17 dB, compared to a value of -22 dB for the ideal case without parameter uncertainty.

To further demonstrate the accuracy of the proposed method, Fig. 5 (a) and (b) report the probability density functions of the 3-dB bandwidth for the fifth- and seventh-order filters, respectively. The bandwidth is derived as a post-processing step starting from the full transfer functions described above, computed with a classical (solid red lines) and sparse PCE (yellow dashed lines) models after the application of the truncated KLT, and using PCA in the latter case. The bandwidth calculated from the Monte Carlo samples is used as reference (blue bars). Despite the fact that the analysis is conducted on a derived synthetic quantity (the bandwidth), the proposed method still retains its high accuracy, with an almost perfect agreement between the three techniques. It should also be noted that the two probability density functions are centered at different values because the designs of the two filters targets two different nominal bandwidths (39 GHz for the fifth-order filter and 29 GHz for the seventh-order one). Both density functions are symmetric, indicating an equal probability for the bandwidth to become either larger or narrower under the considered parameter uncertainty.

TABLE I
COMPARISON OF THE DIFFERENT METHODS USED FOR THE ANALYSIS OF THE MACH-ZEHNDER FILTERS WITH 28 AND 38 VARIABLES.

	Number of samples		**Computational time**	
	$d = 28$	$d = 38$	$d = 28$	$d = 38$
Monte Carlo	10000	10000	847 s	899 s
Classical PCE	4 mln	21 mln	n/a	n/a
Classical PCE + KLT	1386	5148	118 s	507 s
Sparse PCE + KLT + PCA	100	200	13.2 s	27 s

TABLE II
ACCURACY OF THE PROPOSED METHOD AS A FUNCTION OF THE NUMBER OF TRAINING SAMPLES.

	RMSE			
	Fifth-order filter ($d = 28$)		Seventh-order filter ($d = 38$)	
	average	maximum	average	maximum
$L = 25$	2.0999×10^{-2}	3.0723×10^{-2}	2.6623×10^{-2}	4.7079×10^{-2}
$L = 50$	6.0095×10^{-3}	1.0539×10^{-2}	2.7191×10^{-2}	5.1511×10^{-2}
$L = 100$	3.7550×10^{-3}	5.8772×10^{-3}	9.1346×10^{-3}	1.5714×10^{-2}
$L = 200$	3.8724×10^{-3}	5.7735×10^{-3}	5.6308×10^{-3}	9.6840×10^{-3}
$L = 400$	2.7967×10^{-3}	4.1644×10^{-3}	3.7206×10^{-3}	5.8509×10^{-3}

D. Efficiency and Accuracy

Table I summarizes the efficiency of the discussed methods in terms of number of required samples and computational time for the analysis of the full transfer functions of the two test cases. All simulations were performed on a Lenovo ThinkPad X1 Yoga laptop with an Intel(R) Core(TM) i7-7500U processor, CPU running at 2.7 GHz, and 16 GB of RAM. It is important to mention that most of the computational time required by the PCE methods (about 98% for the classical method, and 65% for the proposed sparse method) is due to the simulation of the training samples.

As already noted, for both the reported examples, the application of the classical PCE would have been unfeasible without the preliminary KLT reduction, since it would have required a number of training samples in the order of several millions. It is also important to mention that this is not necessarily the case for the sparse method, thanks to its capability of identifying only the non-negligible terms. However, its computational efficiency also benefits from this dimensionality reduction, since the algorithms needs to search through a largely reduced sets of terms. Because of this, the performance of the method is expected to reduce if a large number of independent or weakly correlated parameters is included, or if the amount of variability is increased up to the point of requiring a model of prohibitive complexity. The advocated method achieves a further speed-up of about one order of magnitude w.r.t. the classical PCE method ($8.9\times$ and $18.8\times$ for the first and the second test case, respectively), and an overall speed-up of $64.2\times$ and $33.3\times$ against Monte Carlo.

Table II investigates instead the impact of the number of training samples on the achieved accuracy. The table reports the average and maximum, over frequency and port, of the root-mean-square error (RMSE) computed w.r.t. the 10000 MC samples. It is observed that the RMSE steadily decreases by increasing the training set size L. In particular, the maximum RMSE drops below 10^{-2} for $L \geq 100$ in the first test case (fifth-order filter), and for $L \geq 200$ in the second test case (seventh-order filter). The improvement becomes marginal beyond those values. A generally good practice is to select a number of training samples proportional to the number of uncertain parameters, e.g., $L > 3 \cdot d$ or $L > 10 \cdot \tilde{d}$, where d is the original number of uncertain parameters, and $\tilde{d}$ is the number of reduced variables after KLT.

IV. CONCLUSIONS

In this paper, we have demonstrated an efficient and accurate method to analyze the performance variability induced by fabrication uncertainty in photonic circuits with a large number of Gaussian correlated design parameters and multiple outputs of interest. The proposed scheme combines a sparse PCE approach with KLT to both remove correlation in the parameter set and reduce its dimensionality, alleviating the computational cost of the PCE model. Similarly, the use of PCA compression allows reducing the number of output quantities to be modeled, thereby further improving efficiency.

As an example of the method potentialities, we have shown the stochastic analysis of a cascaded Mach-Zehnder filter with up to 38 design parameters with a complex correlation scenario. The filter transfer function at different wavelengths represented the multiple output of interest for the analysis. Results were in excellent agreement with the reference Monte Carlo simulation, while the computational time was reduced by more than 30 times.

With the widespread of photonic devices controlled by large sets of parameters and with multiple objective requirements, this method represents a promising and valuable approach for the stochastic analysis, yield estimation, and optimization of their performance.

REFERENCES

[1] M. Smit et al., "An introduction to InP-based generic integration technology," *Semicond. Sci. Technol.*, vol. 29, no. 8, p. 083001, Jun. 2014.

[2] W. Bogaerts, D. Pérez, J. Capmany, D. A. B. Miller, J. Poon, D. Englund, F. Morichetti, and A. Melloni, "Programmable photonic circuits," *Nature*, vol. 586, no. 7828, pp. 207–216, Oct. 2020.

[3] 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. International Society for Optics and Photonics, 2017, p. 102420S.

[4] A. Waqas, D. Melati, and A. Melloni, "Sensitivity analysis and uncertainty mitigation of photonic integrated circuits," *Journal of Lightwave Technology*, vol. 35, no. 17, pp. 3713–3721, 2017.

[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] R. L. Wagner, J. Song, and W. C. Chew, "Monte carlo simulation of electromagnetic scattering from two-dimensional random rough surfaces," *IEEE Transactions on Antennas and Propagation*, vol. 45, no. 2, pp. 235–245, 1997.

[7] M. S. Eldred, "Recent advances in non-intrusive polynomial chaos and stochastic collocation methods for uncertainty analysis and design," *AIAA Paper*, vol. 2274, no. 2009, p. 37, 2009.

[8] 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.

[9] 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.

[10] Y. Xing, D. Spina, A. Li, T. Dhaene, and W. Bogaerts, "Stochastic collocation for device-level variability analysis in integrated photonics," *Photonics Research*, vol. 4, no. 2, pp. 93–100, 2016.

[11] A. Waqas, D. Melati, B. S. Chowdhry, and A. Melloni, "Efficient variability analysis of photonic circuits by stochastic parametric building blocks," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 26, no. 2, pp. 1–8, 2019.

[12] X. Cao, S. Bhatnagar, M. Nikdast, and S. Roy, "Hierarchical polynomial chaos for variation analysis of silicon photonics microresonators," in *2019 International Applied Computational Electromagnetics Society Symposium (ACES)*, 2019.

[13] C. Cui, K. Liu, and Z. Zhang, "Chance-constrained and yield-aware optimization of photonic ICs with non-gaussian correlated process variations," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 39, no. 12, pp. 4958–4970, 2020.

[14] Y. Yang, Y. Ma, H. Guan, Y. Liu, S. Danziger, S. Ocheltree, K. Bergman, T. Baehr-Jones, and M. Hochberg, "Phase coherence length in silicon photonic platform," *Optics express*, vol. 23, no. 13, pp. 16 890–16 902, 2015.

[15] M. Loève, *Probability Theory*. New York, NY, USA: Springer-Verlag, 1977.

[16] G. Blatman and B. Sudret, "Adaptive sparse polynomial chaos expansion based on least angle regression," *J. Comput. Phys.*, vol. 230, no. 6, pp. 2345–2367, Mar. 2011.

[17] D. Xiu and G. E. Karniadakis, "The wiener-askey polynomial chaos for stochastic differential equations," *SIAM J. Sci. Computation*, vol. 24, no. 2, pp. 619–644, 2002.

[18] D. Xiu, "Fast numerical methods for stochastic computations: A review," *Commun. Computational Physics*, vol. 5, no. 2–4, pp. 242–272, Feb. 2009.

[19] A. Singhee and R. A. Rutenbar, "Why quasi-monte carlo is better than monte carlo or latin hypercube sampling for statistical circuit analysis," *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.*, vol. 29, no. 11, pp. 1763–1776, Nov. 2010.

[20] S. Marelli and B. Sudret, "Uqlab: A framework for uncertainty quantification in matlab," in *Proc. 2nd Int. Conf. on Vulnerability, Risk Analysis and Management, Liverpool, United Kingdom, Apr. 2014, pp. 2554–2563.*

[21] 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. 2018.

[22] J. S. Ochoa and A. C. Cangellaris, "Random-space dimensionality reduction for expedient yield estimation of passive microwave circuits," *IEEE Trans. Microw. Theory Techn.*, vol. 61, no. 12, pp. 4313–4321, Dec. 2013.

[23] P. Manfredi and R. Trinchero, "A data compression strategy for the efficient uncertainty quantification of time-domain circuit responses," *IEEE Access*, vol. 8, pp. 92 019–92 027, 2020.

[24] V. Yaghoubi, S. Marelli, B. Sudret, and T. Abrahamsson, "Sparse polynomial chaos expansions of frequency response functions using stochastic frequency transformation," *Probabilistic Eng. Mech.*, vol. 48, pp. 39–58, Apr. 2017.

[25] C. K. Madsen and J. H. Zhao, *Optical Filter Design and Analysis: A Signal Processing Approach Optical Filter Design and Analysis: A Signal Processing Approach*. Wiley Online Library, 1999.