Probabilistic Uncertainty Quantification of Microwave Circuits Using Gaussian Processes

Probabilistic Uncertainty Quantification of Microwave Circuits Using Gaussian Processes

Paolo Manfredi, *Senior Member, IEEE*

***Abstract*— In this article, a probabilistic machine learning framework based on Gaussian process regression (GPR) and principal component analysis (PCA) is proposed for the uncertainty quantification (UQ) of microwave circuits. As opposed to most surrogate modeling techniques, GPR models inherently carry information on the model prediction uncertainty due to unseen data. This article shows how the inherent uncertainty of GPR pointwise predictions can be combined with the uncertainty of the design parameters to provide global statistical information on the device performance with the inclusion of confidence bounds. The model confidence is possibly improved by increasing the amount of training data. In addition, PCA is employed to effectively deal with problems with multiple and possibly complex-valued output components, such as those involving the UQ of time-domain responses or multiport scattering parameters. The proposed technique is successfully applied to two low-noise amplifier designs subject to the process variation of up to 25 parameters. Comparisons against the state-of-the-art polynomial chaos expansion method demonstrates that GPR achieves superior accuracy, while additionally providing information on the prediction confidence.**

***Index Terms*— Gaussian process regression (GPR), kriging, machine learning, principal component analysis (PCA), surrogate modeling, uncertainty quantification (UQ).**

I. Introduction

PROCESS variations are an unavoidable factor in modern mass-production electronics, which makes uncertainty quantification (UQ) an essential task in the early stage design of microwave components and recently prompted an ever-growing interest in this field [1], [2], [3]. For example, threshold voltage, gate length, and oxide thickness are recognized as the leading variation sources in the CMOS technology [4], [5] appearing in many microwave circuits, such as power amplifiers.

Many surrogate modeling techniques were recently proposed for UQ as more effective alternatives to direct Monte Carlo (MC) sampling [6], [7], [8], [9], [10], [11]. Surrogate models provide a computationally inexpensive emulator of the true underlying model and can be classified as parametric or non-parametric. The former requires to define a priori the form of the model, and they exhibit a complexity that increases with the number of input parameters, possibly leading to the so-called "curse of dimensionality." This means that the model may break down even before attempting to train it, because the basis functions cannot be allocated in the available memory.

A typical example are the methods based on the framework of generalized polynomial chaos expansion (PCE) [12], [13], [14], [15], [16], [17], [18], [19], [20], [21], [22], which expand the stochastic quantities of interest in terms of orthogonal basis functions of the uncertain parameters. The number of basis functions grows exponentially with the input dimensionality and/or the expansion order, thus becoming intractable for problems with several parameters that require high-order expansions. This applies also to sparse PCEs [23], although they can be more effectively trained with a number of data samples that is smaller than the number of basis functions. Moreover, their accuracy cannot be improved beyond the maximum theoretical accuracy determined by the predefined form of the model. The popularity of PCE-based techniques resides in the fact that they exhibit optimal exponential convergence rate, provided that the correct polynomial basis is used according to the probability distribution of the random input parameters. Moreover, the first two statistical moments (mean and variance) are analytically derived from the model coefficients [24].

On the other hand, the complexity of nonparametric, kernel-based models mainly depends on the available training data, while in fact being almost transparent to the input dimensionality. Since in fact they do not assume any specific functional form, their accuracy is not limited a priori. Furthermore, the number of input parameters does not affect the model complexity directly, although it does have an impact on the number of observations that are required to train the model, which in turn affects the training efficiency. Examples are support-vector machines [25], least-square support-vector machines [26], and Kriging, also known as Gaussian process regression (GPR) [27], which were recently applied also to UQ [28], [29], [30].

A common feature of most surrogate modeling approaches is that, once the model coefficients are determined, the surrogate is in fact *deterministic*. Hence, it does not provide any indication on the confidence of the predictions, as if the surrogate were implicitly considered to be arbitrarily accurate. Confidence information is an important added value, since it allows assessing the accuracy of predictions in the absence of

Manuscript received 9 August 2022; revised 28 September 2022; accepted 30 November 2022. Date of publication 21 December 2022; date of current version 5 June 2023.

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

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TMTT.2022.3228953.

Digital Object Identifier 10.1109/TMTT.2022.3228953

reference results, and it is possibly used to drive a refinement of the model. In a surrogate-assisted MC analysis, the model confidence can be used to assess the accuracy of predicted statistical estimates in absolute terms, as well as to compare it with the inherent confidence of MC estimates and determine, e.g., whether the additional uncertainty introduced by the surrogate model can be deemed to be negligible [31].

In this regard, an attractive feature of GPR is that it provides a *probabilistic*, rather than a deterministic model. Indeed, the GPR model is given in terms of a Gaussian process with a mean function, or "trend," and a covariance function. The latter allows assigning confidence bounds to pointwise model predictions. The confidence is representative of the limited information that is used to train the model and is typically reduced by adding further data. In the context of microwave engineering, GPR has been widely applied to assist the optimization of computationally expensive systems [32], [33], [34], [35] (see also [36] and [37] for an overview and comparisons with other approaches), as well as for the generic surrogate modeling of microwave structures [38], [39]. An application to UQ, in which a GPR model is used in conjunction with least-square support-vector machines as a mere deterministic surrogate of an expensive underlying simulator, was proposed in [30].

Nevertheless, the use of GPR in a fully probabilistic framework for UQ requires to correctly propagate the inherent model uncertainty on local, pointwise predictions to "global" statistical estimates. To our best knowledge, this aspect received little attention in the literature. Some pioneering works in this regard [31], [40] consider a fully Bayesian approach, in which the uncertainty associated with the Bayesian inference of (some of) the GPR model parameters is propagated to UQ measures. Albeit very rigorous, the complexity of this approach considerably increases for high-dimensional problems as well as when kernel parameters are assumed to be unknown and must be inferred from the available data, which is the case of virtually any practical application scenario. Another approach based on local multioutput GPR was proposed in [41]. The method has some computational burden since it divides the input parameter space into several subdomains, for which it trains separate GPR models. Moreover, it has some limitations in that it assumes the same covariance function for each output component, it fits an individual GPR model for each statistical moment, thus possibly resulting in nonphysical estimates (e.g., negative variance predictions), and it neglects correlation between the predictions at different points in the input space. A later extension [42] assumes that the covariance matrix describing the correlation between distinct output variables can be factorized as the product of smaller matrices, which is not necessarily the case.

A simpler, though rigorous probabilistic framework based on GPR for the UQ of electronic circuits with the inclusion of confidence bounds for the statistical estimates, such as moments and probability distributions of the outputs of interest, has been recently developed [43]. The idea is to derive statistical estimates from pointwise predictions by suitably taking into account their uncertainty. This article extends and improves the above method for its application to the UQ of microwave circuits. Specifically, more accurate closed-form estimates are reported for the probabilistic prediction of the variance of the quantities of interest, when the GPR model is used as a surrogate in a MC-like analysis. The extension to complex-valued outputs, which is essential for the frequency-domain characterization of amplifiers as well as passive structures, is also discussed. The main difficulty in this regard is the suitable handling of the real and imaginary parts of the complex variable, since the standard GPR framework only applies to real-valued quantities. In this scenario, analytical estimates are also provided for the mean of the magnitude of the complex output.

Moreover, for problems with multiple outputs of interest, PCA is leveraged to compress data and make the model training more efficient [43], [44], [45], [46]. In this context, the use of PCA compression is not per se new, yet it fits particularly well in the context of GPR, because its linearity allows retaining the convenient properties of Gaussian random variables. It should be noted that PCA is already widely adopted in microwave engineering to identify digital predistortion parameters in power amplifiers [47], [48], [49], [50], [51], [52], [53], for model-order reduction [54], [55], [56], [57], or to reduce the dimensionality of correlated input parameters in statistical and yield analyses [58], [59]. In [60], PCA was used to reduce input space dimensionality in the GPR modeling of stochastic electromagnetic field exposure.

The findings presented in this article can be applied in conjunction with the state-of-the-art toolboxes for GPR modeling. The method is applied to the UQ of the performance of two low-noise amplifier (LNA) designs subject to process variations, for which both steady-state harmonic balance and small-signal ac simulations are considered. Comparisons against PCE show that GPR always outperforms this state-of-the-art method in terms of accuracy for a given training set size, while additionally providing confidence information.

The remainder of this article is organized as follows. Section II introduces the main notions on GPR modeling, as needed for the subsequent discussion. Section III discusses the application of GPR surrogates to UQ, and provides analytical probabilistic estimates for the predicted mean and variance. The application to multiple outputs by means of PCA is outlined in Section IV, whereas Section V extends the discussion to complex-valued outputs. Two application examples are reported in Section VI, with comparisons against the PCE technique. Finally, conclusions are drawn in Section VII.

II. The GPR Model

We consider the generic system

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

where $\boldsymbol{x} \in \mathbb{R}^d$ denotes a vector of selected circuit parameters, which we will later assume to be uncertain, and $\mathcal{M}$ represents the "full-computational" model that maps a given configuration of $\boldsymbol{x}$ to the corresponding output y. The computational model $\mathcal{M}$ can be anything from a simple analytical relationship to a complex circuit or full-wave simulator. In the

context of this article, it will be the SPICE simulator used to analyze the behavior of the considered microwave circuits. Hence, $\boldsymbol{x}$ is a set of tunable circuit parameters, and y the output of interest. For the sake of simplicity, we first base the discussion on a scalar and real-valued output y, like for example an amplifier gain. We discuss how to effectively deal with multiple outputs in Section IV, and we further extend the framework to complex-valued outputs in Section V.

The underlying assumption of GPR is that the target function (1) is a realization of some prior Gaussian process with a given mean function $\mu(\boldsymbol{x})$ and covariance or *kernel* function $k(\boldsymbol{x}, \boldsymbol{x}')$, that is,

$$\mathcal{M}_{\text{prior}}(\boldsymbol{x}) \sim \mathcal{GP}\big(\mu(\boldsymbol{x}), k(\boldsymbol{x}, \boldsymbol{x}')\big). \tag{2}$$

The "correct" realization is identified by conditioning the prior process (2) based on a certain amount of training observations $\{(\boldsymbol{x}_l^\dagger, y_l^\dagger)\}_{l=1}^L$, where $y_l^\dagger = \mathcal{M}(\boldsymbol{x}_l^\dagger)$, for $l = 1, \ldots, L$, are computed using the full-computational model (1). This results in the posterior Gaussian process

$$y \approx \mathcal{M}_{\text{GPR}}(\boldsymbol{x}) \sim \mathcal{GP}(m(\boldsymbol{x}), c(\boldsymbol{x}, \boldsymbol{x}')) \tag{3}$$

where the posterior trend $m(\boldsymbol{x})$ and covariance function $c(\boldsymbol{x}, \boldsymbol{x}')$ are readily obtained based on the prior parameters and the training observations [27]. If the training observations are assumed to be noiseless, which is reasonable if they come from a computer simulation, they are interpolated by the realizations, or "trajectories," of the process (3). Without loss of generality, in this article we will always assume noise-free observations, although the proposed framework applies, with minimal modifications, also to the case of noisy data. In that case, however, the posterior Gaussian process is no longer constrained to interpolate the observations.

It should be noted that, for a given value of $\boldsymbol{x}$, the model output is not a deterministic value, but rather a Gaussian random variable with mean $m(\boldsymbol{x})$ and standard deviation $(c(\boldsymbol{x}, \boldsymbol{x}))^{1/2}$, which allows assigning (local) confidence bounds to pointwise predictions. Therefore, the posterior trend $m(\boldsymbol{x})$ represents the most likely model, whereas the standard deviation is a measure of the residual uncertainty due to unseen data, and it is typically reduced by increasing the number of training observations.

There exist several common types of prior trends and covariance functions. Examples of the former include (unknown) constant values, with a fixed zero trend representing a particular, yet popular choice, as well as polynomials, with polynomial chaos expansions being a special case in the polynomial-chaos-based Kriging [61]. A plethora of kernel functions exist as well, with the anisotropic squared-exponential kernel

$$k(\boldsymbol{x}, \boldsymbol{x}') = \sigma^2 \exp\left(-\frac{1}{2} r^2\right) \tag{4}$$

and the anisotropic Matérn 5/2 kernel

$$k(\boldsymbol{x}, \boldsymbol{x}') = \sigma^2 \left(1 + \sqrt{5} r + \frac{5}{3} r^2\right) \exp\left(-\sqrt{5} r\right) \tag{5}$$

where

$$r = \sqrt{\sum_{j=1}^{d} \frac{(x_j - x_j')^2}{\theta_j^2}} \tag{6}$$

being among the most popular ones. The trend coefficients, as well as the kernel variance σ^2 and length scales $\{\theta_j\}_{j=1}^d$ (commonly referred to as the "hyperparameters"), are typically unknown and they must be computed as part of the training process. This is typically done by solving an optimization problem that either maximizes the likelihood function or minimizes the cross-validation error over the training samples [27], [62]. Bayesian inference is sometimes used as an alternative for the prior parameters, but it results in a much more cumbersome posterior model, which is no longer Gaussian distributed except for the case in which the kernel parameters are known and only the trend coefficients are estimated [40], [63], [64].

It should be noted that the choice of the prior trend and kernel is not necessarily critical, as GPR has excellent learning capabilities and generalization properties. The formulation scales favorably with the number of input parameters, since only the inner products in (6) needs to be computed. However, increasing the number of input parameters does have an impact of the training efficiency, since it increases the number of required observations as well as the size of the optimization problem to be solved for the estimation of the hyperparameters. Indeed, as the initial value for the number of training samples, it is reasonable to consider a multiple of the input dimensionality, i.e., $L = k \cdot d$, with $k \approx 3 - 10$. This is suggested both by some literature (e.g., [64]) and by similar applications of other kernel-based methods (e.g., [44]). Oftentimes, the length scale is assumed to be the same for each input dimension, which simplifies the form of the kernel, making it isotropic, and speeds-up the training phase. However, this choice may reduce the accuracy for high-dimensional parameter spaces that exhibit large variations.

III. Probabilistic UQ via GPR Surrogates

In an UQ scenario, the input parameters $\boldsymbol{x}$ are treated as random variables. Typically, the constructed model is used as a computationally cheap surrogate of the full-computational model (1) for a MC-like analysis. Therefore, the surrogate is sampled with a finite set $\{\boldsymbol{x}_i^*\}_{i=1}^N$ of values of the uncertain circuit parameters, drawn according to their distribution. For a GPR model, this results in a vector of N correlated Gaussian random variables $\boldsymbol{y}^* \sim \mathcal{N}(\boldsymbol{m}, \boldsymbol{C})$, whose mean vector and covariance matrix are found as

$$\boldsymbol{m} = \boldsymbol{\mu}^* + \boldsymbol{K}_* \boldsymbol{K}^{-1} (\boldsymbol{y}^\dagger - \boldsymbol{\mu}^\dagger) \tag{7}$$

and

$$\boldsymbol{C} = \boldsymbol{K}_{**} - \boldsymbol{K}_* \boldsymbol{K}^{-1} \boldsymbol{K}_*^{\mathsf{T}} \tag{8}$$

respectively, where

1) $\boldsymbol{\mu}^\dagger \in \mathbb{R}^L$ is a column vector with entries $\mu_l^\dagger = \mu(\boldsymbol{x}_l^\dagger)$, i.e., the prior trend evaluated at the training samples;
2) $\boldsymbol{\mu}^* \in \mathbb{R}^N$ is a column vector with entries $\mu_i^* = \mu(\boldsymbol{x}_i^*)$, i.e., the prior trend evaluated at the MC samples;

3) $\boldsymbol{K} \in \mathbb{R}^{L\times L}$ is a matrix with entries $K_{lm} = k(\boldsymbol{x}_l^\dagger, \boldsymbol{x}_m^\dagger)$, i.e., the covariance matrix of the training samples;
4) $\boldsymbol{K}_* \in \mathbb{R}^{N\times L}$ is a matrix with entries $K_{il} = k(\boldsymbol{x}_i^*, \boldsymbol{x}_l^\dagger)$, i.e., the cross-covariance matrix between the training and MC samples;
5) $\boldsymbol{K}_{**} \in \mathbb{R}^{N\times N}$ is a matrix with entries $K_{ij} = k(\boldsymbol{x}_i^*, \boldsymbol{x}_j^*)$, i.e., the covariance matrix of the MC samples;

It should be noted that the abovementioned Gaussian distribution refers to the posterior prediction, whereas there is no assumption on the probability distribution of the input parameters $\boldsymbol{x}$, which can be arbitrary.

The diagonal of matrix $\boldsymbol{C}$ in (8) provides the pointwise variance of the MC predictions. The variance vanishes at the training samples, and progressively reduces everywhere by increasing the density of the training samples [63]. The rate of variance reduction also depends on the kernel length scale, and therefore on the specific problem at hand. Intuitively, smoother functions require less dense sampling. Nevertheless, the information of the prediction confidence could drive the acquisition of additional training samples.

Drawing one random sample of $\boldsymbol{y}^*$ corresponds to generating a prediction of the MC samples along a specific posterior trajectory, from which an estimate of the moments or the probability distribution of y can be obtained. By considering a large number of samples, many estimates are produced, which allows assessing not only their expectation, but also their dispersion and confidence bounds. It is important to point out that, in the above process, it is essential to take into account the existing correlation between the samples $\boldsymbol{y}^*$, which is given by the full posterior covariance matrix $\boldsymbol{C}$ in (8). Assigning only a "local" uncertainty to each sample would lead to erroneous results when propagating uncertainty. In the state-of-the-art GPR-based approaches to UQ, the information on the prediction uncertainty is neglected, and deterministic statistical estimates are obtained by considering the mean prediction only, provided by (7) (cfr., e.g., [61]).

The abovementioned post-processing can be time consuming, especially if the number N of MC samples is large. Fortunately, closed-form expressions can be derived at least for the predicted mean and variance of y and, specifically, for their expectation and standard deviation. Some preliminary results in this regard were presented in [43]. This article provides more accurate results for the prediction of the variance, and it extends the framework to complex-valued outputs.

A. Probabilistic GPR Prediction of the Mean

In a MC setting, the mean of y is estimated by calculating the sample mean of the surrogate model predictions, that is,

$$\hat{\mu}_y = \frac{1}{N}\sum_{i=1}^{N} y_i^*. \tag{9}$$

By considering that each sample y_i^* is drawn from a multivariate normal distribution, with mean and covariance given by (7) and (8), we can calculate the expectation and the standard deviation of $\hat{\mu}_y$ to be

$$\mathrm{E}\{\hat{\mu}_y\} = \frac{1}{N}\sum_{i=1}^{N} m_i \tag{10}$$

and

$$\mathrm{Std}\{\hat{\mu}_y\} = \frac{1}{N}\sqrt{\sum_{i=1}^{N}\sum_{j=1}^{N} C_{ij}} \tag{11}$$

respectively.

It is worth noting the following.

1) The mean (10) is the average of the mean vector (7), and corresponds to the average computed by sampling along the posterior trend.
2) The standard deviation (11) is a measure of the prediction uncertainty due to the fact that the GPR model is trained with a limited amount of data, and it reflects the additional uncertainty introduced by the surrogate model in predicting the MC samples. By way of example, the standard deviation would vanish if the training samples and the MC samples coincided, as the GPR model would interpolate them without introducing errors (cfr. [63]). The GPR uncertainty in the prediction of the MC samples can be eventually combined with the inherent uncertainty of MC estimates due to finite sampling.
3) Since the sample mean (9) is a sum of Gaussian random variables, its distribution is still Gaussian. Hence, its mean and standard deviation are readily used to estimate quantitative confidence bounds.

B. Probabilistic GPR Estimate of the Variance

The variance of y is estimated by calculating the sample variance of the surrogate model predictions, that is,

$$\begin{aligned}\hat{\sigma}_y^2 &= \frac{1}{N-1}\sum_{i=1}^{N}\left(y_i^* - \hat{\mu}_y\right)^2 \\ &= \frac{1}{N-1}\sum_{i=1}^{N}\left(y_i^* - \frac{1}{N}\sum_{j=1}^{N} y_j^*\right)^2.\end{aligned} \tag{12}$$

In [43], the expected value and standard deviation of (12) were approximatively derived under the simplifying assumption that $\hat{\mu}_y \approx \mathrm{E}\{\hat{\mu}_y\}$. For a more rigorous calculation, it is useful to rewrite (12) as the quadratic form [65]

$$\hat{\sigma}_y^2 = \frac{1}{N-1}\boldsymbol{y}^{*\mathsf{T}}\boldsymbol{\Lambda}\boldsymbol{y}^* \tag{13}$$

where $\boldsymbol{\Lambda}$ is a $N \times N$ matrix with entries

$$\Lambda_{ij} = \begin{cases} \dfrac{N-1}{N} & i = j \\ -\dfrac{1}{N} & i \neq j. \end{cases} \tag{14}$$

The properties of quadratic forms [65] allow obtaining the expected value and standard deviation of (13) as

$$\mathrm{E}\{\hat{\sigma}_y^2\} = \frac{1}{N-1}(\boldsymbol{m}^{\mathsf{T}}\boldsymbol{\Lambda}\boldsymbol{m} + \mathrm{tr}(\boldsymbol{\Lambda}\boldsymbol{C})) \tag{15}$$

and

$$\mathrm{Std}\{\hat{\sigma}_y^2\} = \frac{2}{(N-1)}\sqrt{\boldsymbol{m}^\top \boldsymbol{\Lambda C \Lambda m} + \frac{1}{2}\mathrm{tr}(\boldsymbol{\Lambda C \Lambda C})} \quad (16)$$

respectively. It should be noted that the second result applies only because the random variables $\boldsymbol{y}^*$ are normally distributed. However, contrary to the prediction of the mean, the distribution of (13) is no longer Gaussian, but rather a generalized chi-square [66]. Hence, the knowledge of the standard deviation only provides a qualitative indication of the prediction confidence.

It is important to remark that the outlined estimates seamlessly apply with any definition of the prior trend and kernel, including the PCE-based Kriging [61], and for any distribution of the input parameters. The prediction uncertainty is reduced by increasing the training samples since, as already discussed, this reduces the entries of the covariance matrix (8) appearing in the standard deviation of the estimates.

IV. Multioutput Problems: PCA Compression

In practical applications, the output of interest is rarely scalar, but most likely vectorial. Typical examples are the steady-state response of an oscillator, or the (small-signal) multiport scattering parameters of an amplifier. In the latter case, the output is also complex-valued, which requires extra care, since the standard GPR framework only applies to real-valued quantities.

A common, albeit naive approach is to train individual GPR models for each component, which however becomes unfeasible if the output size is large. Moreover, this would cause the time/frequency resolution to largely affect the training time. Indeed, the general inability of dealing with multiple outputs is one of the main shortcomings of most of the standard implementations of kernel-based machine learning techniques, which jeopardizes their application to real-life scenarios and motivates the adoption of more flexible, but data-hungry, multioutput neural network architectures. An effective approach to compress the training data, thereby reducing to a feasible amount the number of individual models to be trained, has been proposed in [44]. The method is based on PCA, and fortunately it fits particularly well also to GPR models.

Let us assume that a training dataset $\boldsymbol{Y}^\dagger \in \mathbb{C}^{S\times L}$ be available, collecting training samples for S output components, which we individually denote as y_s. An output component can be any voltage, current, or scattering parameter at a particular time or frequency point. We further denote the vector of training samples for each component, i.e., the rows of $\boldsymbol{Y}^\dagger$, as $\boldsymbol{y}_s^\dagger = (y_{s,1}^\dagger, \ldots, y_{s,L}^\dagger)$, for $s = 1, \ldots, S$.

The training dataset $\boldsymbol{Y}^\dagger$ is compressed using PCA, leading to a reduced set of training observations [44]

$$\tilde{z}_n^\dagger = \sum_{s=1}^{S} U_{sn}\left(\boldsymbol{y}_s^\dagger - \bar{y}_s\right) \quad (17)$$

with $n = 1, \ldots, \tilde{n}$, and typically $\tilde{n} \ll S$. In (17), $\bar{y}_s$ is the sth component of the dataset mean, that is,

$$\bar{y}_s = \frac{1}{L}\sum_{l=1}^{L} y_{s,l}^\dagger \quad (18)$$

whereas U_{sn} are the elements of the first $\tilde{n}$ left-singular vectors obtained from the singular value decomposition (SVD) of the zero-mean dataset in which the entries of $\boldsymbol{Y}^\dagger$ are replaced with $y_{s,l}^\dagger - \bar{y}_s$. The number of principal components $\tilde{n}$ is determined by truncating the SVD based on the relative magnitude of the singular values. Setting a relative truncation threshold ϵ determines an analogous maximum error on the norm of the training dataset. A threshold of $\epsilon = 1\%$ usually leads to satisfactory accuracy, with good generalization also beyond training data [45], and shall be therefore used in this article. In the following, we momentarily restrict the discussion to real-valued outputs, which makes all the quantities in (17) to be also real-valued. We will relax this assumption in the next section.

The compressed datasets (17) are used to train a small set of $\tilde{n}$ GPR models of the principal components in the form of (3). When sampled in the context of a MC analysis, these models yield $\tilde{n}$ independent sets of multivariate Gaussian random variables $\tilde{\boldsymbol{z}}_n^* \sim \mathcal{N}(\tilde{\boldsymbol{m}}_n, \tilde{\boldsymbol{C}}_n)$, where the mean vectors $\tilde{\boldsymbol{m}}_n$ and covariance matrices $\tilde{\boldsymbol{C}}_n$ are computed from the respective GPR models as in (7) and (8). It is worth noting that the MC predictions within each set are still correlated, and hence the covariance matrices $\tilde{\boldsymbol{C}}_n$ are also full.

Samples of the original outputs are then recovered by applying the inverse PCA transformation, leading to

$$\boldsymbol{y}_s^* = \bar{y}_s + \sum_{n=1}^{\tilde{n}} U_{sn}\tilde{\boldsymbol{z}}_n^*. \quad (19)$$

Given the linearity of (19), the samples $\boldsymbol{y}_s^*$ are still Gaussian-distributed, with mean vector

$$\boldsymbol{m}_s = \bar{y}_s + \sum_{n=1}^{\tilde{n}} U_{sn}\tilde{\boldsymbol{m}}_n \quad (20)$$

and covariance matrix [65] gpr-pca-dosimetry

$$\boldsymbol{C}_s = \sum_{n=1}^{\tilde{n}} U_{sn}^2\tilde{\boldsymbol{C}}_n. \quad (21)$$

It should be noted that, to avoid confusion, a tilde has been used to discriminate between the quantities associated with the principal components and the ones related to the original outputs.

Based on (20) and (21), probabilistic estimates for each output component are obtained as discussed in Section II. For the closed-form probabilistic estimates of the predicted mean and variance, (20) and (21) are plugged directly into (10), (11), (15), and (16) to obtain more efficient vectorized implementations, whose detailed discussion is however outside the scope of this article.

V. Complex-Valued Outputs

We now consider the case of complex-valued quantities of interest. We start again from the scalar case, and we recall that the mean and variance of a complex random variable $y = u + jv$ are computed as

$$\mathrm{E}\{y\} = \mathrm{E}\{u\} + j\mathrm{E}\{v\} \in \mathbb{C} \quad (22)$$

and

$$\mathrm{Var}\{y\} = \mathrm{Var}\{u\} + \mathrm{Var}\{v\} \in \mathbb{R} \tag{23}$$

respectively. Hence, sample estimates (9) and (12) are readily combined with (22) and (23) to predict the mean and variance of a complex output.

The classical GPR framework utilizes real-valued kernels and therefore it only applies to real-valued data. There exist some extensions of kernel-based techniques, including GPR, to complex-valued quantities [67], [68]. However, their effectiveness strongly relies on the specific features of ad hoc complex kernel functions, which must be carefully chosen depending on the problem at hand. A viable solution is the so-called dual-channel formulation, which consists in training a separate GPR model for the real and imaginary parts.

Let us denote with $\boldsymbol{u}^* \sim \mathcal{N}(\boldsymbol{m}', \boldsymbol{C}')$ and $\boldsymbol{v}^* \sim \mathcal{N}(\boldsymbol{m}'', \boldsymbol{C}'')$ the vectors of MC predictions of the real and imaginary parts of y, respectively. From the mean vectors and covariance matrices of the two GPR models, the closed-form expressions (10), (11), (15), and (16) are readily used to obtain the expected value and standard deviation of the complex mean and of the variance, thanks to the additive properties of the expectation and variance operators. It is important to point out that the standard deviation of the prediction of the variance can be computed from the sum of the variances of the real and imaginary parts because they are modeled as two separate (and hence, independent) GPR models, and therefore their covariance is zero.

Besides performing UQ of a complex output, it is often of interest to perform UQ also of its magnitude $|y| = (u^2 + v^2)^{1/2}$, especially when the information on the phase is of minor importance. Unfortunately, deriving closed-form predictions of the magnitude from the model of a complex variable is hindered by the square-root operator. Of course, one can train a GPR model directly for the magnitude of the output, instead of considering it as a complex number. However, from a complex model, some closed-form estimates can be derived, with no additional effort, for the squared magnitude $|y|^2 = u^2 + v^2$, which is still of interest in microwave circuits because it is associated with power/gain.[1] In particular, the sample mean of the squared magnitude is computed as

$$\hat{\mu}_{|y|^2} = \frac{1}{N}\sum_{i=1}^{N}\left|y_i^*\right|^2 = \frac{1}{N}\sum_{i=1}^{N}\left(u_i^*\right)^2 + \frac{1}{N}\sum_{i=1}^{N}\left(v_i^*\right)^2 \tag{24}$$

which can be cast as the sum of two quadratic forms

$$\hat{\mu}_{|y|^2} = \frac{1}{N}\boldsymbol{u}^{*\mathsf{T}}\boldsymbol{u}^* + \frac{1}{N}\boldsymbol{v}^{*\mathsf{T}}\boldsymbol{v}^*. \tag{25}$$

With the above definition, and using again the properties of quadratic forms, the expected value and the standard deviation of $\hat{\mu}_{|y|^2}$ are computed as

$$\mathrm{E}\left\{\hat{\mu}_{|y|^2}\right\} = \frac{1}{N}\left(\boldsymbol{m}'^{\mathsf{T}}\boldsymbol{m}' + \boldsymbol{m}''^{\mathsf{T}}\boldsymbol{m}'' + \mathrm{tr}\left(\boldsymbol{C}' + \boldsymbol{C}''\right)\right) \tag{26}$$

[1]Besides, it practically makes no difference when working with logarithmic scales.

and

$$\mathrm{Std}\left\{\hat{\mu}_{|y|^2}\right\} = \frac{2}{N}\sqrt{\boldsymbol{m}'^{\mathsf{T}}\boldsymbol{C}'\boldsymbol{m}' + \boldsymbol{m}''^{\mathsf{T}}\boldsymbol{C}''\boldsymbol{m}'' + \frac{1}{2}\mathrm{tr}\left(\boldsymbol{C}'\boldsymbol{C}' + \boldsymbol{C}''\boldsymbol{C}''\right)} \tag{27}$$

where we used the additivity of the trace operator. Like in the case of (13), the distribution of (25) is not Gaussian, which does not allow obtaining quantitative confidence bounds from the standard deviation.

A similar reasoning can be applied to the prediction of the variance of $|y|^2$. However, the calculation involves a quadratic form of the squares of vectors $\boldsymbol{u}^*$ and $\boldsymbol{v}^*$, which are no longer Gaussian-distributed. Hence, closed-form results can be obtained only for the expectation of the variance prediction, but not for its standard deviation.

We now come to the case of multiple complex-valued outputs, which requires some further elaboration. We consider a suitable complex-valued training dataset $\boldsymbol{Y}^\dagger$. The PCA-based framework outlined in Section IV can be applied, with some care, also in this scenario. There are basically three possibilities.

1) To perform the SVD of the complex dataset as is. This leads to complex-valued PCA coefficients U_{sn} and principal components. The latter are then modeled using separate GPR models for their real and imaginary parts, leading to a total of $2\tilde{n}$ models to be trained.
2) To split the real and imaginary parts of the training dataset and to stack them into a new, real-valued dataset of double size. The SVD thereof leads to real-valued PCA coefficients and principal components, which can be therefore modeled using the standard GPR framework. The real and imaginary parts of each original output are then suitably recombined after the inverse PCA transformation.
3) To split the real and imaginary parts of the training dataset, and to model them completely independently. This implies performing two separate, real-valued SVDs, which potentially lead to a different number of principal components.

Approaches 1) and 2) are expected to perform very similarly. However, since the real and imaginary parts of the reconstructed output are in that case a linear combination of the *same* set of principal components, and hence of GPR models, they become statistically correlated. This hinders in particular the calculation of the standard deviation of the predicted variance using the analytical relationships outlined in Section III, since the covariance between the real and imaginary parts should be also accounted for. Therefore, approach 3) is preferable to take advantage of the closed-form predictions, whereas 1) or 2) can be interchangeably used if the predictions and confidence information are calculated numerically.

At this point, it is important to mention that the separate and PCA-compressed modeling of the real and imaginary parts may destroy some physical properties of the data, such as causality. Although this is not crucial in UQ, where the focus is on estimating the statistical properties of the quantities under study, a strategy that can simultaneously handle the real and

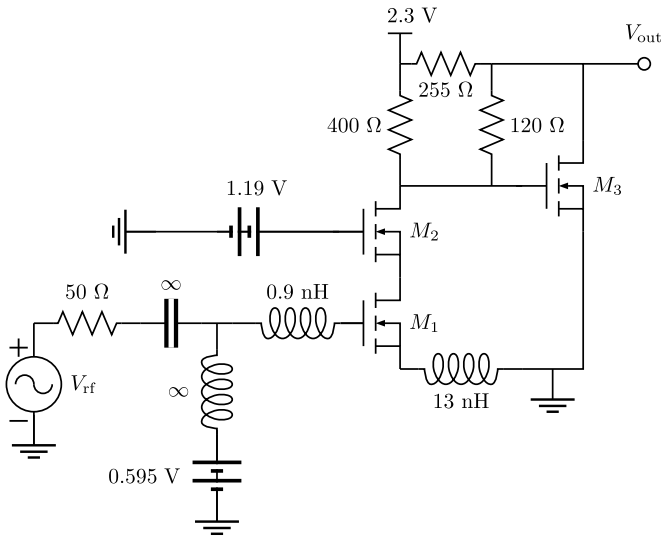

Fig. 1. Schematic of the 900-MHz CMOS cascode amplifier.

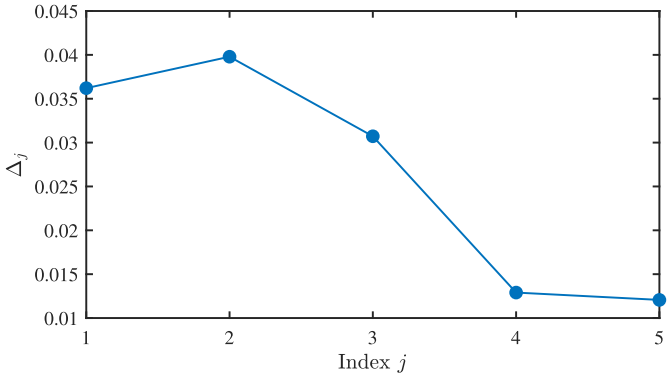

Fig. 2. Convergence of the MC estimate of the variance for increasing sample size.

imaginary parts would yield physically sounder models and is the subject of ongoing research.

VI. Numerical Results

In this section, the proposed UQ framework is applied to two LNA designs, namely a 900-MHz CMOS cascode amplifier and a 2-GHz BJT amplifier. The circuits are simulated with HSPICE, which represents the full computational model (1). For the GPR prior, we consider a constant unknown trend, i.e., $\mu(\boldsymbol{x}) = \beta_0$ in (2), and an anisotropic Matérn 5/2 kernel (5). The models are trained and evaluated by means the MATLAB® Statistics and Machine Learning Toolbox™ [69]. However, any other toolbox for GPR modeling, such as the MATLAB-based UQLab [62], or Scikit-learn in Python [70], can be alternatively used for training the models as needed for the following simulations. All simulations are performed on a Lenovo Thinkpad X13 Yoga laptop with an Intel(R) Core(TM) i7-10510U, CPU running at 1.8 GHz, and 16 GB of RAM.

A. 900-MHz CMOS LNA

The first test case is a 900-MHz cascode power amplifier realized in the 250-nm process technology, whose schematic is illustrated in Fig. 1. The device is taken from the HSPICE 2008-09 benchmark library [71]. The threshold voltage, gate length, and oxide thickness of the CMOS transistors are assumed to be affected by uncertainty. To stress the fact that the probability distribution of the input parameters can be arbitrary, three different distributions are deliberately ascribed to the aforementioned parameters, namely a Gaussian distribution $\mathcal{N}(0.382, 0.038)$ V to the threshold voltage, a uniform distribution $\mathcal{U}(200, 300)$ nm to the gate length, and a beta distribution $\mathcal{B}(2, 5)$ in the interval $[4.9, 6.7]$ nm to the oxide thickness. For the simulation, two input tones at 900 and 910 MHz are considered, with a power of 0 dBm. A harmonic balance simulation is performed to obtain the time-domain steady-state output power at 8009 time points over a window of 200 ns.

Hence, the input of (1) is here a vector $\boldsymbol{x} = (x_1, x_2, x_3)$, collecting the three aforementioned uncertain parameters, which is mapped by the SPICE simulator to a vector of $S = 8009$ time-domain values of the output power. To generate reference results, we run a MC analysis by progressively increasing the number of samples, drawn according to a Latin hypercube design. Specifically, we start from $N = 125$ and we double the number up to $N = 4000$. We assess the convergence by calculating the integral over time of the absolute deviation of subsequent variance estimates, that is,

$$\Delta_j = \frac{1}{T}\int_0^T \left|\sigma^2_{\text{MC},j}(t) - \sigma^2_{\text{MC},j-1}(t)\right| dt \tag{28}$$

where $\sigma^2_{\text{MC},j}$ denotes the variance obtained from $N = 2^j \cdot 125$ samples, and $j = 1, \ldots, 5$. Briefly speaking, Δ_j is a measure on the improvement achieved by doubling the number of samples.

Fig. 2 shows the behavior of Δ_j, highlighting that there is no substantial improvement from $j = 4$ on, which corresponds to increasing the number of samples from $N = 1000$ to $N = 2000$. Therefore, we shall use $N = 1000$ samples as a reference, and we use the same set of MC samples to calculate the GPR predictions. We train two GPR models, with $L = 20$ and $L = 40$ samples, also drawn according to a Latin hypercube scheme. The corresponding datasets are compressed with PCA from $S = 8009$ components to $\tilde{n} = 11$ and $\tilde{n} = 12$ principal components, respectively.

Fig. 3 shows the steady-state output power. The gray lines are a subset of responses from the MC simulation, whereas the green line is the mean power predicted with the GPR model trained with $L = 40$ samples. Fig. 4 provides enlargements around 50 and 100 ns. The top figures show, in addition to the MC samples, the mean μ_{MC} of the MC samples (blue solid line), its 95% confidence interval (dash-dotted blue lines), as well as the 95% confidence bounds of the GPR prediction of the mean (shaded areas). Red and green areas refer to the models trained with $L = 20$ and $L = 40$ samples, respectively. The mean predicted with an adaptive sparse PCE, again in conjunction with PCA compression [44] and with a maximum order of five, is also shown (dashed black lines). The PCE is trained by means of the pertinent module in UQLab [72] using the dataset with $L = 40$ samples, and it

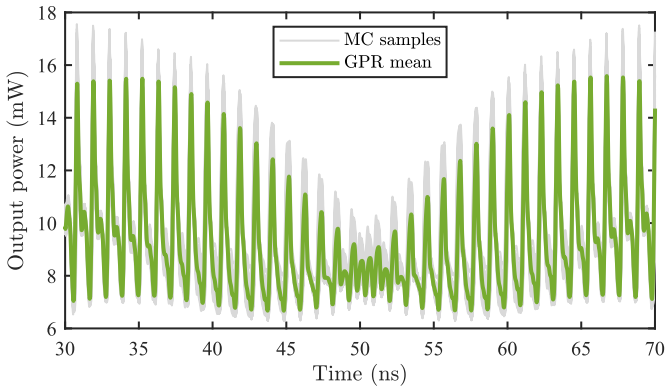

Fig. 3. Steady-state output power. Gray lines: subset of responses from the MC simulation. Green line: mean predicted with the GPR model.

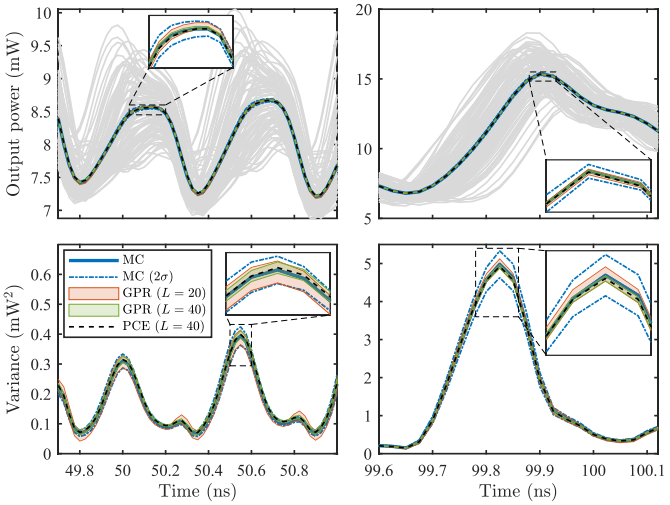

Fig. 4. Steady-state output power around 50 and 100 ns. Top: MC samples (gray lines), MC mean (blue solid line), 95% confidence interval of the MC mean (blue dash-dotted lines), 95% confidence bounds of the mean predicted with the GPR models trained with $L = 20$ (red area) and $L = 40$ samples (green area), and mean predicted with the PCE trained with $L = 40$ samples (black dashed line). Bottom: variance, 2σ bounds of the MC variance and of the corresponding GPR predictions, and PCE estimate.

has no confidence associated. The confidence of the MC mean is instead computed based on the well-known estimate of its standard deviation, i.e., [73]

$$\mathrm{Std}\{\mu_{\mathrm{MC}}\} \approx \sqrt{\frac{\sigma^2_{\mathrm{MC}}}{N}}. \tag{29}$$

Similar results are reported for the variance in the bottom of Fig. 4. In this case, however, the MC and GPR confidence intervals correspond to the 2σ bounds, which in both cases do not indicate exactly a 95% confidence due to the non-Gaussian distribution of the variance predictions. The standard deviation of the MC variance is computed as [73]

$$\mathrm{Std}\{\sigma^2_{\mathrm{MC}}\} \approx \sqrt{\frac{\left(\sigma^2_{\mathrm{MC}}\right)^2}{N}\left(\kappa_{\mathrm{MC}} - 1 + \frac{2}{N-1}\right)} \tag{30}$$

where κ_{MC} denotes the kurtosis of the MC samples.

All the plots show that the confidence interval of the GPR predictions shrinks by increasing the number of training

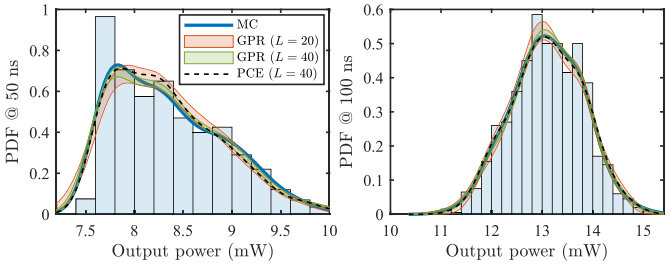

Fig. 5. PDF of the steady-state output power at 50 ns (left) and 100 ns (right). Histogram and blue solid line: distribution of the MC samples; red and green areas: 95% confidence bounds of the GPR predictions; black dashed line: prediction of the PCE.

samples and tightens around the MC prediction, which is mostly enclosed by it. This is further appreciated in the enlargements in the insets. The prediction is more accurate for the mean, for which both models provide extremely thin, and hence nearly indistinguishable bounds. It is also noted that the GPR confidence bounds are narrower than the ones of the MC estimates, indicating that the surrogate model introduces marginal additional uncertainty. The PCE results also agree well with the reference MC curves.

Furthermore, Fig. 5 shows the probability density function (PDF) of the output power at 50 and 100 ns. The histogram is the distribution of the MC samples, whereas the blue line is its kernel density estimate. As in Fig. 4, the red and green areas indicate the 95% confidence bounds of the distributions obtained with the GPR models trained with $L = 20$ and $L = 40$ samples, whereas the dashed black line is the distribution predicted by the PCE. The GPR confidence bounds are obtained by calculating the kernel density estimates of the distributions along 1000 posterior trajectories. Like for the moments in Fig. 4, it is observed that the bounds mostly enclose the MC result, and they narrow as the number of training samples is increased. Moreover, the agreement is similar or better compared to the PCE prediction, especially at 50 ns.

To better assess the accuracy of the GPR and PCE models, we introduce the root-mean-square error (RMSE) and the coefficient of determination (R^2), respectively defined as

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} |y_i - y_i^*|^2} \tag{31}$$

and

$$R^2 = 1 - \frac{\sum_{i=1}^{N} |y_i - y_i^*|^2}{\sum_{i=1}^{N} |y_i - \bar{y}|^2} \tag{32}$$

where y_i denotes the reference MC samples, y_i^* are the corresponding predictions, and $\bar{y}$ is the mean of the MC samples. Since the output is a function of time, the RMSE and R^2 are also computed as a function of time. Table I reports the average and the worst RMSE e R^2 values over time achieved by the GPR and PCE models, trained with both $L = 20$ and $L = 40$ samples. It is observed that for both training set sizes, the GPR model outperforms the PCE one. In particular, the accuracy of the PCE model trained with

TABLE I
ACCURACY OF THE GPR AND PCE MODELS FOR THE 900-MHZ CMOS LNA EXAMPLE

	RMSE		R^2	
Model	avg	max	avg	min
GPR ($L = 20$)	0.0633	0.1725	0.9777	0.6743
GPR ($L = 40$)	0.0300	0.0747	0.9947	0.8062
PCE ($L = 20$)	0.1342	0.3599	0.9141	-0.4018
PCE ($L = 40$)	0.0661	0.1861	0.9767	0.6457

TABLE II
COMPUTATIONAL TIMES REQUIRED BY THE GPR- AND PCE-BASED UQ OF THE 900-MHZ CMOS LNA

	$L = 20$		$L = 40$	
	GPR	PCE	GPR	PCE
SPICE simulation	1.01 s		3.01 s	
PCA compression	0.02 s		0.03 s	
Model training	1.22 s	1.34 s	1.30 s	1.59 s
Model evaluation	3.60 s	0.29 s	4.07 s	0.43 s
Confidence of moments	6.18 s	—	6.80 s	—
Confidence of PDF	5.24 s	—	5.52 s	—

40 samples is comparable to the one of the GPR model trained with 20 samples only.

Finally, Table II collects the main figures concerning the simulation times. It is observed that the time required by the PCA compression is negligible. For the GPR, the training time (i.e., the calculation of the prior trend and kernel hyperparameters) is smaller than the evaluation time (i.e., the calculation of the posterior mean vector and covariance matrix for the set of MC samples). Most of the time is required by the calculation of the confidence bounds for the moments (mean and variance) and the distributions. The former are computed for all the 8009 times points, whereas the latter only for the two time points shown in Fig. 5. Hence, the calculation is much more efficient for the moments because it benefits from the closed-form formulas. On the other hand, the PDF is usually computed for a smaller set of outputs. The PCE has a comparable training time and a negligible evaluation time, since only a deterministic prediction is provided. It is also important to mention that, for more complex circuits, the post-processing times become negligible compared to the SPICE simulations. For this simple design, the SPICE simulation is rather fast, and the MC analysis itself took only 95.3 s.

B. 2-GHz BJT LNA

The second application example considers the 2-GHz BJT LNA described in [74], and already investigated in [16] by assuming uncertainty in $d = 25$ circuit parameters, including the transistor's forward current gain, some of its parasitics, the external lumped components, and the widths of the microstrip line sections. All these parameters are assumed to be Gaussian-distributed and independent, with a 10% relative standard deviation.

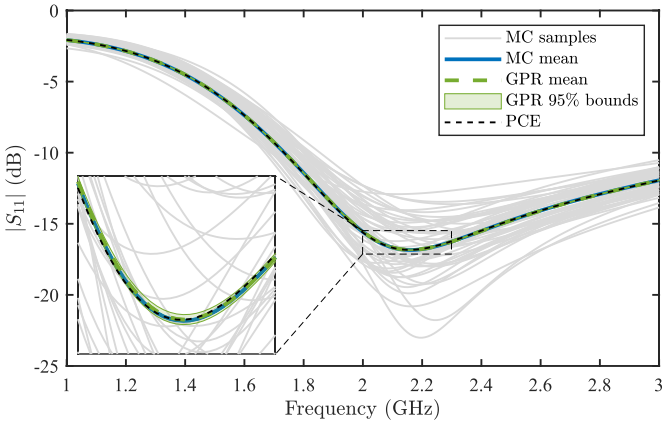

Fig. 6. S_{11} of the BJT LNA. The gray lines are a subset of samples from the MC simulation. The blue solid and the green dashed lines are the magnitude of the average S_{11} obtained from the MC samples and with the GPR model, respectively. The green shaded area is the 95% confidence interval of the GPR prediction. The black dashed line is the average obtained with the PCE model.

The two-port scattering parameters S_{11} and S_{21} of the LNA are simulated at 201 frequency points with a small-signal ac analysis. A GPR model is trained with $L = 100$ response samples. As described in Section V, the real and imaginary parts of the scattering parameters are modeled separately, leading to two datasets with $S = 402$ output components each. The PCA compression of these datasets leads to 11 principal components for the real part, and 10 for the imaginary part. Reference results are generated based on $N = 1000$ MC simulations.

Fig. 6 shows the results obtained for the magnitude of S_{11}. The gray lines are a subset of MC responses, highlighting the large variability of S_{11} resulting from the uncertainty in the circuit parameters. The magnitude of the average S_{11} obtained from the MC samples (blue solid line) is compared with the GPR (dashed green line) and PCE (dashed black line) predictions, showing excellent agreement. The 95% confidence (2σ) bounds of the GPR prediction are also shown by the shaded green area. The enlargement around 2 GHz, displayed in the inset, allows for a better assessment of the accuracy of the predictions and, in particular, of the tightness of the GPR confidence bounds.

Fig. 7 illustrates the results for the variance of S_{11}. Similar to Fig. 6, the blue line indicates the value obtained from the MC samples, whereas the dashed green and dashed black lines are the predictions of the GPR and PCE models, respectively. It is noted that the PCE estimate is in this case way less accurate compared to the GPR prediction. However, the lack of confidence information would not allow recognizing it in the absence of references results. On the contrary, the shaded green area indicates again the 2σ interval of the GPR prediction. Even though it no longer corresponds to a 95% confidence since, as already noted for the previous example, the distribution of the variance prediction is not Gaussian, the interval still provides a good indication of the prediction confidence, with the reference MC result being well enclosed, as shown in the inset.

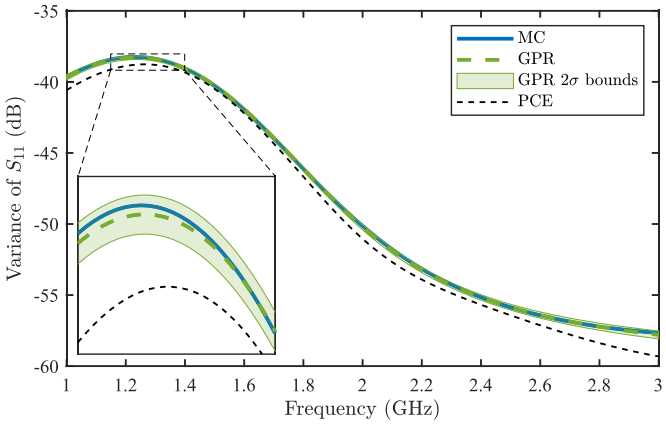

Fig. 7. Variance of S_{11} obtained from the MC samples (blue solid line) and with the GPR (green dashed line) and PCE (black dashed line) models. The green shaded area indicates the 2σ interval of the GPR prediction.

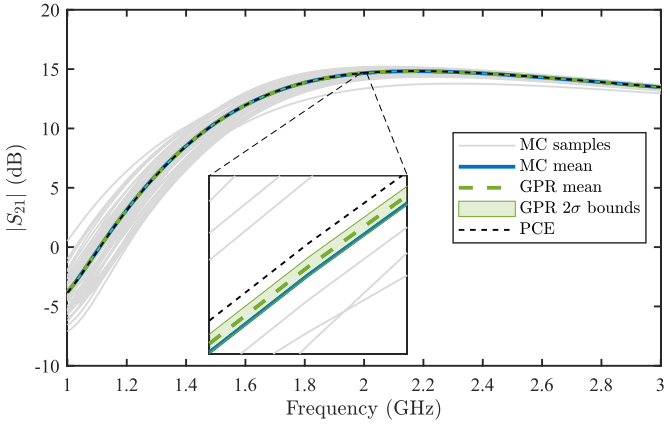

Fig. 8. S_{21} of the BJT LNA. The gray lines are a subset of samples from the MC simulation. The blue solid, green dashed, and black dashed lines are the average of $|S_{21}|$ obtained from the MC samples and with the GPR and PCE models, respectively. The green shaded area is the 2σ interval of the GPR prediction.

Fig. 8 provides similar results for S_{21}, which also corresponds to the small-signal gain. As opposed to Fig. 6, the solid blue, dashed green, and dashed black lines now indicate the average of the magnitude, rather than the magnitude of the average of the scattering parameter. The GPR prediction thereof is obtained by means of (26) and (27), in conjunction with PCA compression. As for the variance, the distribution of this prediction is not Gaussian. Therefore, the reported 2σ bounds do not correspond to the 95% confidence interval. Nonetheless, the interval provides again a good indication of the model confidence, and it does enclose the MC result. On the contrary, the PCE prediction is again less accurate, but this cannot be easily assessed without the information on its confidence.

Finally, Fig. 9 shows the probability distribution of $|S_{21}|$ (gain) at the target operating frequency of 2 GHz. The histogram is the distribution of the MC samples, whereas the blue solid line is its kernel density estimate. The shaded green area indicates the 95% confidence interval of the GPR prediction, obtained by sampling 1000 posterior trajectories with the same

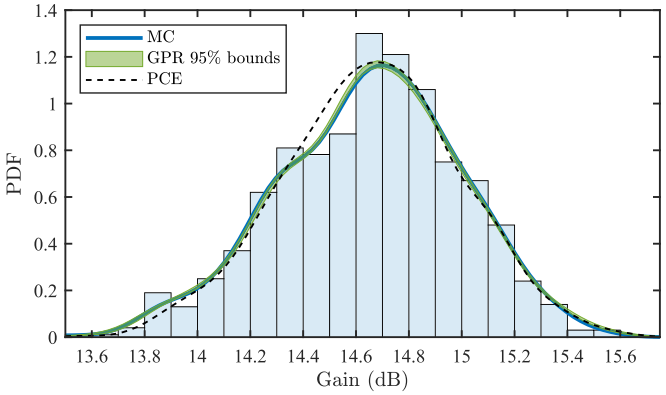

Fig. 9. Probability distribution of the LNA gain at 2 GHz. Histogram and blue solid line: distribution of the MC samples and its kernel density estimate, respectively. Green shaded area: 95% confidence interval of the GPR prediction. Black dashed line: PCE prediction.

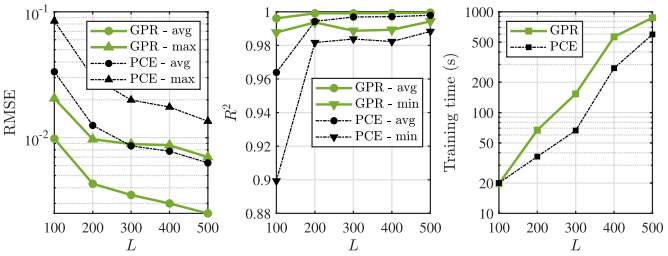

Fig. 10. Accuracy and training cost of the GPR and PCE models. Left: average and maximum RMSE; central: average and minimum R^2; right: CPU time required to train the models.

set of MC samples, whereas the dashed black line is the PCE estimate. A superior accuracy of the GPR prediction over the PCE one is again established, and it is further confirmed by the tightness of the confidence bounds, which wrap the MC result.

To further investigate the accuracy and the training efficiency of the GPR in comparison with the PCE method, we consider models trained with different sample sizes, and we calculate the RMSE and R^2 values over frequency. Fig. 10 shows the average and maximum RMSE (left), the average and minimum R^2 (central), and the training time (right) for the two methods, when the training set size L is increased from 100 to 500 samples in steps of 100 samples. From the comparison, we can draw some interesting conclusions. First of all, the GPR always outperforms the PCE in both metrics, and for all sample sizes. For the RMSE, the maximum error obtained with the GPR is comparable with the average error achieved by the PCE. On the contrary, the training of the GPR models is slightly more costly, yet they scale similarly. This is due to the larger number of uncertain parameters, which on the one hand motives the use of a larger number of training samples, and on the other hand requires the optimization of a larger number of hyperparameters, since an anisotropic kernel is used. In this regard, the plot considers only the actual training time. The time required by the PCA compression, which is common to both methods, is always well below one second, and hence negligible. The evaluation time is instead roughly

constant with the training set size, even for the GPR model, for which it amounts to up to 17 s to evaluate the posterior mean vector and covariance matrices, and to about 8 s to extract the expected value and standard deviation of the mean and variance estimates. As to the PCE instead, the evaluation time is within one second, as in the previous application example. For reference, the MC simulation took 1249 s with HSPICE. We can therefore conclude that, compared to the PCE and for a given training set size, the GPR method is more accurate, at the price of a slightly higher training cost. Moreover, as we show in this article, it additionally provides information on the confidence of the predicted statistics.

VII. Conclusion

This article presented a probabilistic machine learning framework for the UQ of microwave circuits. The method is based on GPR and allows estimating statistical information, like moments and PDFs, with the inclusion of confidence information.

Closed-form results are reported for the expected value and standard deviation of the predictions of the mean and of the variance of a stochastic output, and of the mean of the magnitude of a complex one. Confidence bounds can be obtained numerically for higher order moments and distribution functions, as well as for any other statistical metric. Such confidence information reflects the uncertainty introduced by the GPR surrogate in the prediction of the response samples in a MC-like analysis. In future research, it could be integrated into an adaptive learning strategy to optimize the surrogate-based UQ. Moreover, it could be combined with the inherent uncertainty of MC estimates to simultaneously account for finite sampling and limited training set size.

Multioutput problems are dealt with by leveraging PCA compression, which allows reducing the number of output components to be modeled by one or two orders of magnitude. In case of complex-valued outputs, the real and imaginary parts are split and modeled individually to avoid introducing statistical correlation between them. In this regard, more rigorous approaches to simultaneously handle complex-valued data, and preserve physical properties like causality, will be considered in future research.

The application to two LNA designs, one with 25 uncertain parameters, illustrated and validated the advocated approach. Accurate results were obtained with a very limited number of training samples. Moreover, it was shown that the predicted confidence provides a good indication of the model accuracy, and it is improved by increasing the number of training samples. Comparisons with one of the prime and state-of-the-art tools for UQ, i.e., the sparse PCE, show that GPR achieves higher accuracy at the price of a slightly higher training cost, while providing in addition information of the prediction confidence.

Acknowledgment

The author thank Prof. Riccardo Trinchero for the many fruitful discussions about machine learning methods.

References

[1] M. Gossye, G.-J. Gordebeke, K. Y. Kapusuz, D. V. Ginste, and H. Rogier, "Uncertainty quantification of waveguide dispersion using sparse grid stochastic testing," *IEEE Trans. Microw. Theory Techn.*, vol. 68, no. 7, pp. 2485–2494, Jul. 2020.

[2] Y. Tao, F. Ferranti, and M. S. Nakhla, "Uncertainty quantification using parameter space partitioning," *IEEE Trans. Microw. Theory Techn.*, vol. 69, no. 4, pp. 2110–2119, Apr. 2021.

[3] M. Yusuf and S. Roy, "A polymorphic polynomial chaos formulation for mixed epistemic-aleatory uncertainty quantification of RF/microwave circuits," *IEEE Trans. Microw. Theory Techn.*, vol. 70, no. 1, pp. 926–937, Jan. 2022.

[4] W. Zhao et al., "Rigorous extraction of process variations for 65-nm CMOS design," *IEEE Trans. Semicond. Manuf.*, vol. 22, no. 1, pp. 196–203, Feb. 2009.

[5] S. Mittal, "A survey of architectural techniques for managing process variation," *ACM Comput. Surv.*, vol. 48, no. 4, pp. 1–29, May 2016.

[6] S. Koziel and A. Bekasiewicz, "Variable-fidelity response feature surrogates for accelerated statistical analysis and yield estimation of compact microwave components," *IET Microw., Antennas Propag.*, vol. 13, no. 14, pp. 2539–2543, Nov. 2019.

[7] J. Zhang, F. Feng, J. Jin, W. Zhang, Z. Zhao, and Q.-J. Zhang, "Adaptively weighted yield-driven EM optimization incorporating neurotransfer function surrogate with applications to microwave filters," *IEEE Trans. Microw. Theory Techn.*, vol. 69, no. 1, pp. 518–528, Jan. 2021.

[8] A. Pietrenko-Dabrowska and S. Koziel, "Design centering of compact microwave components using response features and trust regions," *Energies*, vol. 14, no. 24, p. 8550, Dec. 2021.

[9] A. Pietrenko-Dabrowska and S. Koziel, "Optimization-based robustness enhancement of compact microwave component designs with response feature regression surrogates," *Knowl.-Based Syst.*, vol. 240, Mar. 2022, Art. no. 108161.

[10] S. Koziel and A. Pietrenko-Dabrowska, "Performance-driven yield optimization of high-frequency structures by Kriging surrogates," *Appl. Sci.*, vol. 12, no. 7, p. 3697, Apr. 2022.

[11] S. Koziel and A. Pietrenko-Dabrowska, "Tolerance-aware multi-objective optimization of antennas by means of feature-based regression surrogates," *IEEE Trans. Antennas Propag.*, vol. 70, no. 7, pp. 5636–5646, Jul. 2022.

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

[13] A. C. M. Austin and C. D. Sarris, "Efficient analysis of geometrical uncertainty in the FDTD method using polynomial chaos with application to microwave circuits," *IEEE Trans. Microw. Theory Techn.*, vol. 61, no. 12, pp. 4293–4301, Dec. 2013.

[14] D. Spina, D. De Jonghe, D. Deschrijver, G. Gielen, L. Knockaert, and T. Dhaene, "Stochastic macromodeling of nonlinear systems via polynomial chaos expansion and transfer function trajectories," *IEEE Trans. Microw. Theory Techn.*, vol. 62, no. 7, pp. 1454–1460, Jul. 2014.

[15] P. Manfredi and F. Canavero, "Efficient statistical simulation of microwave devices via stochastic testing-based circuit equivalents of nonlinear components," *IEEE Trans. Microw. Theory Techn.*, vol. 63, no. 5, pp. 1502–1511, May 2015.

[16] P. Manfredi, D. V. Ginste, D. De Zutter, and F. G. Canavero, "Generalized decoupled polynomial chaos for nonlinear circuits with many random parameters," *IEEE Microw. Wireless Compon. Lett.*, vol. 25, no. 8, pp. 505–507, Aug. 2015.

[17] A. K. Prasad, M. Ahadi, and S. Roy, "Multidimensional uncertainty quantification of microwave/RF networks using linear regression and optimal design of experiments," *IEEE Trans. Microw. Theory Techn.*, vol. 64, no. 8, pp. 2433–2446, Aug. 2016.

[18] D. Spina, T. Dhaene, L. Knockaert, and G. Antonini, "Polynomial chaos-based macromodeling of general linear multiport systems for time-domain analysis," *IEEE Trans. Microw. Theory Techn.*, vol. 65, no. 5, pp. 1422–1433, May 2017.

[19] A. K. Prasad and S. Roy, "Accurate reduced dimensional polynomial chaos for efficient uncertainty quantification of microwave/RF networks," *IEEE Trans. Microw. Theory Tech.*, vol. 65, no. 10, pp. 3697–3708, Oct. 2017.

[20] J. Zhang, C. Zhang, F. Feng, W. Zhang, J. Ma, and Q.-J. Zhang, "Polynomial chaos-based approach to yield-driven EM optimization," *IEEE Trans. Microw. Theory Techn.*, vol. 66, no. 7, pp. 3186–3199, May 2018.

[21] S. Bakirtzis, X. Zhang, and C. D. Sarris, "Stochastic modeling of wave propagation in waveguides with rough surface walls," *IEEE Trans. Microw. Theory Techn.*, vol. 69, no. 1, pp. 500–508, Jan. 2021.

[22] A. B. Narendranath and K. J. Vinoy, "A novel method for intrusive stochastic estimation of geometric tolerance effects in finite element electromagnetic analysis," *IEEE Trans. Microw. Theory Techn.*, vol. 69, no. 10, pp. 4329–4340, Oct. 2021.

[23] 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, 2011.

[24] P. Manfredi and D. V. Ginste, "Polynomial chaos based uncertainty quantification in electrical engineering: Theory," in *Uncertainty Quantification of Electromagnetic Devices, Circuits, and Systems*, S. Roy, Ed. Stevenage, U.K.: IET, 2021, ch. 2, pp. 21–48.

[25] V. Vapnik, *The Nature of Statistical Learning Theory*, 2nd ed. New York, NY, USA: Springer, 1999.

[26] J. A. K. Suykens, *Least Squares Support Vector Machines*. Singapore: World Scientific, 2002.

[27] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. Cambridge, MA, USA: MIT Press, 2006.

[28] R. Trinchero, P. Manfredi, I. S. Stievano, and F. G. Canavero, "Machine learning for the performance assessment of high-speed links," *IEEE Trans. Electromagn. Compat.*, vol. 60, no. 6, pp. 1627–1634, Dec. 2018.

[29] R. Trinchero, M. Larbi, H. Torun, F. G. Canavero, and M. Swaminathan, "Machine learning and uncertainty quantification for surrogate models of integrated devices with a large number of parameters," *IEEE Access*, vol. 7, pp. 4056–4066, 2019.

[30] R. Trinchero and F. G. Canavero, "Combining LS-SVM and GP regression for the uncertainty quantification of the EMI of power converters affected by several uncertain parameters," *IEEE Trans. Electromagn. Compat.*, vol. 62, no. 5, pp. 1755–1762, Oct. 2020.

[31] A. O'Hagan, "Bayesian analysis of computer code outputs: A tutorial," *Rel. Eng. Syst. Saf.*, vol. 91, nos. 10–11, pp. 1290–1300, Oct. 2006.

[32] E. S. Siah, M. Sasena, J. L. Volakis, P. Y. Papalambros, and R. W. Wiese, "Fast parameter optimization of large-scale electromagnetic objects using DIRECT with Kriging metamodeling," *IEEE Trans. Microw. Theory Tech.*, vol. 52, no. 1, pp. 276–285, Jan. 2004.

[33] P. Chen, B. M. Merrick, and T. J. Brazil, "Bayesian optimization for broadband high-efficiency power amplifier designs," *IEEE Trans. Microw. Theory Techn.*, vol. 63, no. 12, pp. 4263–4272, Dec. 2015.

[34] B. Liu, H. Yang, and M. J. Lancaster, "Global optimization of microwave filters based on a surrogate model-assisted evolutionary algorithm," *IEEE Trans. Microw. Theory Techn.*, vol. 65, no. 6, pp. 1976–1985, Jun. 2017.

[35] H. M. Torun and M. Swaminathan, "High-dimensional global optimization method for high-frequency electronic design," *IEEE Trans. Microw. Theory Techn.*, vol. 67, no. 6, pp. 2128–2142, Jun. 2019.

[36] F. E. Rangel-Patino, J. L. Chavez-Hurtado, A. Viveros-Wacher, J. E. Rayas-Sanchez, and N. Hakim, "System margining surrogate-based optimization in post-silicon validation," *IEEE Trans. Microw. Theory Techn.*, vol. 65, no. 9, pp. 3109–3115, Sep. 2017.

[37] J. E. Rayas-Sanchez, S. Koziel, and J. W. Bandler, "Advanced RF and microwave design optimization: A journey and a vision of future trends," *IEEE J. Microw.*, vol. 1, no. 1, pp. 481–493, Jan. 2021.

[38] S. Koziel and J. W. Bandler, "Reliable microwave modeling by means of variable-fidelity response features," *IEEE Trans. Microw. Theory Techn.*, vol. 63, no. 12, pp. 4247–4254, Dec. 2015.

[39] N. Leszczynska, I. Couckuyt, T. Dhaene, and M. Mrozowski, "Low-cost surrogate models for microwave filters," *IEEE Microw. Wireless Compon. Lett.*, vol. 26, no. 12, pp. 969–971, Dec. 2016.

[40] R. G. Haylock and A. O'Hagan, "On inference for outputs of computationally expensive algorithms with uncertainty on the inputs," in *Bayesian Statistics*. New York, NY, USA: Oxford Univ. Press, 1996, pp. 629–637.

[41] I. Bilionis and N. Zabaras, "Multi-output local Gaussian process regression: Applications to uncertainty quantification," *J. Comput. Phys.*, vol. 231, no. 17, pp. 5718–5746, Jul. 2012.

[42] I. Bilionis, N. Zabaras, B. A. Konomi, and G. Lin, "Multi-output separable Gaussian process: Towards an efficient, fully Bayesian paradigm for uncertainty quantification," *J. Comput. Phys.*, vol. 241, pp. 212–239, May 2013.

[43] P. Manfredi and R. Trinchero, "A probabilistic machine learning approach for the uncertainty quantification of electronic circuits based on Gaussian process regression," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 41, no. 8, pp. 2638–2651, Aug. 2022.

[44] P. Manfredi and R. Trinchero, "A data compression strategy for the efficient uncertainty quantification of time-domain circuit responses," *IEEE Access*, vol. 8, pp. 92019–92027, 2020.

[45] P. Manfredi and S. Grivet-Talocia, "Fast stochastic surrogate modeling via rational polynomial chaos expansions and principal component analysis," *IEEE Access*, vol. 9, pp. 102732–102745, 2021.

[46] M. Sedaghat, R. Trinchero, Z. H. Firouzeh, and F. G. Canavero, "Compressed machine learning-based inverse model for design optimization of microwave components," *IEEE Trans. Microw. Theory Techn.*, vol. 70, no. 7, pp. 3415–3427, Jul. 2022.

[47] P. L. Gilabert and G. Montoro, "3-D distributed memory polynomial behavioral model for concurrent dual-band envelope tracking power amplifier linearization," *IEEE Trans. Microw. Theory Techn.*, vol. 63, no. 2, pp. 638–648, Feb. 2015.

[48] Y. Lin, C. Quindroit, H. Jang, and P. Roblin, "3-D Fourier series based digital predistortion technique for concurrent dual-band envelope tracking with reduced envelope bandwidth," *IEEE Trans. Microw. Theory Techn.*, vol. 63, no. 9, pp. 2764–2775, Sep. 2015.

[49] D. López-Bueno, Q. A. Pham, G. Montoro, and P. L. Gilabert, "Independent digital predistortion parameters estimation using adaptive principal component analysis," *IEEE Trans. Microw. Theory Techn.*, vol. 66, no. 12, pp. 5771–5779, Dec. 2018.

[50] Y. Liu et al., "Multiband linearization technique for broadband signal with multiple closely spaced bands," *IEEE Trans. Microw. Theory Techn.*, vol. 67, no. 3, pp. 1115–1129, Mar. 2019.

[51] Q. A. Pham, G. Montoro, D. Lopez-Bueno, and P. L. Gilabert, "Dynamic selection and estimation of the digital predistorter parameters for power amplifier linearization," *IEEE Trans. Microw. Theory Techn.*, vol. 67, no. 10, pp. 3996–4004, Oct. 2019.

[52] Y. Li, X. Wang, and A. Zhu, "Complexity-reduced model adaptation for digital predistortion of RF power amplifiers with pretraining-based feature extraction," *IEEE Trans. Microw. Theory Techn.*, vol. 69, no. 3, pp. 1780–1790, Mar. 2021.

[53] H. Li et al., "Signed orthogonal regressor algorithm for digital predistortion of power amplifiers," *IEEE Microw. Wireless Compon. Lett.*, vol. 31, no. 7, pp. 869–872, Jul. 2021.

[54] S. Mijalkovic, "Using frequency response coherent structures for model-order reduction in microwave applications," *IEEE Trans. Microw. Theory Techn.*, vol. 52, no. 9, pp. 2292–2297, Sep. 2004.

[55] B. Nouri, M. S. Nakhla, and R. Achar, "Efficient simulation of nonlinear transmission lines via model-order reduction," *IEEE Trans. Microw. Theory Techn.*, vol. 65, no. 3, pp. 673–683, Mar. 2017.

[56] B. Nouri and M. S. Nakhla, "Model order reduction of nonlinear transmission lines using interpolatory proper orthogonal decomposition," *IEEE Trans. Microw. Theory Techn.*, vol. 66, no. 12, pp. 5429–5438, Dec. 2018.

[57] K. T. J. Gladwin and K. J. Vinoy, "Fast solution of high stochastic dimensional EM problems using proper orthogonal decomposition," *IEEE Microw. Wireless Compon. Lett.*, vol. 32, no. 6, pp. 483–486, Jun. 2022.

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

[59] Z. Chen, Y. Xu, B. Zhang, T. Chen, T. Gao, and R. Xu, "A GaN HEMTs nonlinear large-signal statistical model and its application in S-band power amplifier design," *IEEE Microw. Wireless Compon. Lett.*, vol. 26, no. 2, pp. 128–130, Feb. 2016.

[60] E. Chiaramello et al., "Innovative stochastic modeling of residential exposure to radio frequency electromagnetic field sources," *IEEE J. Electromagn., RF Microw. Med. Biol.*, vol. 5, no. 1, pp. 62–69, Mar. 2021.

[61] R. Schobi, B. Sudret, and J. Wiart, "Polynomial-chaos-based Kriging," *Int. J. Uncertainty Quantification*, vol. 5, no. 2, pp. 171–193, 2015.

[62] C. Lataniotis, D. Wicaksono, S. Marelli, and B. Sudret, "UQLab user manual: Kriging (Gaussian process modeling)," UQLab, Zürich, Switzerland, Tech. Rep., UQLab-V1.4-105, 2021.

[63] V. Dubourg, "Adaptive surrogate models for reliability analysis and reliability-based design optimization," Ph.D. thesis, Laboratoire de Mécanique et Ingénieries, Université Blaise Pascal Clermont-Ferrand II, Aubière, France, 2011.

[64] T. J. Santner, B. J. Williams, and W. I. Notz, *The Design and Analysis of Computer Experiments*. 2nd ed. New York, NY, USA: Springer, 2018.

[65] A. C. Rencher and G. B. Schaalje, *Linear Models in Statistics*, 2nd ed. Hoboken, NJ, USA: Wiley, 2008.

[66] A. Das and W. S. Geisler, "A method to integrate and classify normal distributions," 2020, *arXiv:2012.14331*.

[67] N. Soleimani and R. Trinchero, "Compressed complex-valued least squares support vector machine regression for modeling of the frequency-domain responses of electromagnetic structures," *Electronics*, vol. 11, no. 4, p. 551, Feb. 2022.

[68] R. Boloix-Tortosa, J. J. Murillo-Fuentes, F. J. Payán-Somet, and F. Pérez-Cruz, "Complex Gaussian processes for regression," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 11, pp. 5499–5511, Nov. 2018.

[69] *Statistics and Machine Learning Toolbox, Version 12.1*, The MathWorks, Inc., Natick, MA, USA, 2021.

[70] F. Pedregosa et al., "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, Feb. 2011.

[71] *HSPICE User Guide: RF Analysis, Version B-2008.09*, Synopsys, Inc., Mountain View, CA, USA, Sep. 2008.

[72] S. Marelli, N. Lüthen, and B. Sudret, "UQLab user manual: Polynomial chaos expansions," UQLab, Zürich, Switzerland, Tech. Rep., UQLab-V1.4-104, 2021.

[73] A. M. Mood, F. A. Graybill, and D. C. Boes, *Introduction to the Theory of Statistics*, 3rd ed. New York, NY, USA: McGraw-Hill, 1974.

[74] T. Buss, "2 GHz low noise amplifier with the BFG425 W Philips semiconductors," Philips Semiconductors BV, Nijmegen, The Netherlands, Appl. Note RNRT45-96-B-773, Nov. 1996.

Paolo Manfredi (Senior Member, IEEE) received the M.Sc. degree in electronic engineering and the Ph.D. degree in information and communication technology from the Politecnico di Torino, Turin, Italy, in 2009 and 2013, respectively.

From 2014 to 2017, he was a Post-Doctoral Research Fellow of the Research Foundation–Flanders (FWO) with the Electromagnetics Group, Department of Information Technology, Ghent University, Ghent, Belgium. He is currently an Associate Professor with the EMC Group, Department of Electronics and Telecommunications, Politecnico di Torino. His research interests comprise the several aspects of circuit and interconnect modeling and simulation, including uncertainty quantification, surrogate modeling, signal integrity, and electromagnetic compatibility.

Dr. Manfredi was a recipient of the Outstanding Young Scientist Award at the 2018 Joint IEEE International Symposium on Electromagnetic Compatibility and Asia-Pacific Symposium on Electromagnetic Compatibility, the Best Paper Award at the 2016 IEEE Electrical Design of Advanced Packaging and Systems Symposium, the Best Oral Paper Award and the Best Student Paper Award at the 22nd, and 19th IEEE Conference on Electrical Performance of Electronic Packaging and Systems, respectively, a Young Scientist Award at the XXX International Union of Radio Science General Assembly and Scientific Symposium, and an Honorable Mention at the 2011 IEEE Microwave Theory and Techniques Society International Microwave Symposium. He is currently serving as an Associate Editor for the IEEE JOURNAL ON MULTISCALE AND MULTIPHYSICS COMPUTATIONAL TECHNIQUES and the *International Journal of Circuit Theory and Applications*, and as an Academic Editor for *Mathematical Problems in Engineering*.

Open Access funding provided by 'Politecnico di Torino' within the CRUI CARE Agreement