

Conservative Gaussian Process Models for Uncertainty Quantification and Bayesian Optimization in Signal Integrity Applications

Original

Conservative Gaussian Process Models for Uncertainty Quantification and Bayesian Optimization in Signal Integrity Applications / Manfredi, Paolo. - In: IEEE TRANSACTIONS ON COMPONENTS, PACKAGING, AND MANUFACTURING TECHNOLOGY. - ISSN 2156-3950. - STAMPA. - 14:7(2024), pp. 1-12.
[10.1109/tcpmt.2024.3390402]

Availability:

This version is available at: 11583/2990109 since: 2024-07-01T15:03:15Z

Publisher:

IEEE

Published

DOI:10.1109/tcpmt.2024.3390402

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)

Conservative Gaussian Process Models for Uncertainty Quantification and Bayesian Optimization in Signal Integrity Applications

Paolo Manfredi, *Senior Member, IEEE*

Abstract—Surrogate modeling is being increasingly adopted in signal and power integrity analysis to assist design exploration, optimization, and uncertainty quantification tasks. In this scenario, machine learning methods are attracting an ever-growing interest over alternative and well-consolidated techniques due to their data-driven nature. However, an open issue is to properly assess the trustworthiness of predictions when generalizing beyond training data. Among various machine learning tools, Gaussian process regression (GPR) has the notable feature of providing an estimate of the prediction uncertainty (or confidence) due to the lack of data. Nevertheless, the uncertainty introduced by the estimation of the Gaussian process parameters, which is part of the training process, is typically not accounted for. In this paper, we introduce improved GPR formulations that take into account the additional uncertainty related to the estimation of (some of) the Gaussian process parameters, thereby providing a more accurate estimate of the actual prediction confidence. Furthermore, the advocated framework is extended to uncertainty quantification and Bayesian optimization settings. The technique is applied to two test cases concerning the analysis of crosstalk in a transmission line network and of the frequency-domain response of a microstrip line with a ground plane discontinuity.

Index Terms—Bayesian optimization, Gaussian processes, Kriging, machine learning, signal integrity, surrogate modeling, uncertainty quantification.

I. INTRODUCTION

The computer-aided design of modern mass production electronics is increasingly relying on surrogate models to assist expensive exploration tasks as required, e.g., by optimization and uncertainty quantification (UQ) [1], [2]. In particular, manufacturing tolerances play a key limiting factor to the miniaturization and downsizing of modern integrated circuits. The burgeoning field of machine learning led to the rapid availability of new surrogate modeling paradigms in the integrated circuit design and signal integrity analysis [3], [4], [5], including for example support-vector machines [6], [7], [8], [9], Gaussian processes [10], [11], [12], neural networks [13], [14], and reinforcement learning [15], [16]. Alternatively, the method of polynomial chaos expansion also became widely popular for UQ [17], [18], [19], [20], [21].

Among the abovementioned methods, the polynomial chaos expansion is mainly limited by the model complexity, which does not scale favorably with the number of independent design parameters. Moreover, the number of required training

samples is also somewhat related to the input dimensionality [17], and may become therefore intractable. Neural networks are also very data hungry, and the related approaches require a massive amount of simulations of the actual model to achieve a reasonable training accuracy. This issue can be mitigated by leveraging suitable knowledge-based networks [22] or multiple neural networks trained at different levels of fidelity [23], when applicable. Not surprisingly, neural networks turn out to be effective mainly for those tasks for which huge datasets are available, e.g., image processing. Furthermore, the selection of an optimal neural network architecture, i.e., the number of layers and hidden neurons, the activation functions, etc., is also rather arbitrary and case-dependent, and it mostly relies on experience or trial and error. As such, this approach is more suitable for models that can be reused many times, rather than for one-off simulations. Indeed, a lot of manual tuning is typically involved to obtain an acceptable model, which makes neural-network-based methods rather cumbersome and with scarce flexibility.

On the other hand, kernel-based machine learning methods, such as support-vector machines [24] and Gaussian process regression (GPR) [25], are very effective in this regard. They are characterized by a much simpler structure, in which the main key ingredient is the underlying kernel. This is usually chosen upfront, as there exist some standard kernels with excellent generalization performance. Alternatively, an optimal kernel can be selected from a pool of candidates during the training phase. Kernels are typically characterized by some free “hyperparameters”, which are optimized during the training phase (similarly to the weights in neural networks). The simpler structure of kernel-based methods does not imply lower modeling capabilities. Quite the opposite, it has been shown that GPR models are equivalent to infinitely-wide fully-connected neural networks [26]. Support-vector machine and GPR were successfully used for high-dimensional UQ, e.g., in [7], [9], [10], [11].

One of the main open questions related to surrogate models is their prediction accuracy. This is usually assessed based on the available training data or, more appropriately, on a (small) additional set of validation data that has not been considered for training. Nevertheless, very little information is available as to the generalization beyond the observed data. In this regard, one of the attractive features of GPR is that the model inherently embeds an estimate of the prediction uncertainty outside the training dataset. In the context of Bayesian optimization (BO) [27], this information is used to

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

drive the acquisition of additional observations in an iterative scheme [28], [29]. GPR was also applied to UQ in [10] and [11], in which the single-point prediction uncertainty was propagated to aggregate UQ estimates (e.g., statistical moments and distribution functions).

However, the estimated prediction confidence is rigorous only as long as prior information on the Gaussian process is fully available, which is virtually never the case in practice. In particular, the additional uncertainty introduced by the estimation of the GPR model parameters is not accounted for [30]. As a result, classical and naive implementations of GPR only provide a rough (typically, overconfident) estimation of the actual prediction uncertainty.

In this paper, the GPR framework is extended by improving the accuracy of the prediction confidence, with targeted application to signal integrity analysis. With some relatively simple corrections of the posterior covariance [31], the uncertainty in the estimation of (some of) the prior parameters is accounted for, which naturally leads to more robust surrogates that embed a more accurate (i.e., conservative) estimation of the prediction confidence. The enhanced GPR models are then used in the context of UQ and BO.

The present paper builds upon the preliminary work in [32], which presented the general idea for increasing the conservativity of GPR models. In [32], a single test case with a single output was considered. In the present paper, the theory and analysis are complemented and extended with:

- detailed information on the necessary corrections to increase the conservativity of the model prediction uncertainty;
- extension of the framework to BO;
- investigation of additional test cases with a larger number of independent parameters.

Moreover, the combination with principal component analysis (PCA) is adopted to address problems with multiple and/or time- or frequency-dependent outputs.

The remainder of the paper is organized as follows. Section II reviews the classical GPR framework. Section III introduces the necessary corrections to make the GPR models more conservative and account for the estimation of prior parameters. The improved GPR models are applied in the context of BO and UQ in Section IV. Sections V and VI present two application examples, whereas conclusions are drawn in Section VII.

II. STANDARD GPR MODELING

We start by considering a generic system in the form of

$$y = \mathcal{M}(\mathbf{x}), \quad (1)$$

with $\mathcal{M} : \mathcal{X} \subseteq \mathbb{R}^d \rightarrow \mathbb{R}$. This is a shorthand notation to describe an arbitrary computational model that, given a configuration of design parameters $\mathbf{x} \in \mathcal{X}$, provides a given output quantity of interest (QoI) y . For example, \mathbf{x} could be a set of optimizable or uncertain design variables (geometrical or material parameters, circuit elements, etc.), and y the result of a circuit or electromagnetic simulation (e.g., a crosstalk voltage, an absorbed power, a scattering parameter, etc.).

In surrogate modeling, the goal is to approximate (1) with an accurate though inexpensive mathematical representation that mimics the actual computational model. To begin with, we assume the QoI to be scalar. Clearly, especially in the context of UQ, the QoI may be an entire transient- or frequency-domain response. Hence, we later relax this restriction.

The fundamental assumption of GPR (also called Kriging, especially in geostatistics where the method originated [33]) is that the functional relationship (1) can be assimilated to a specific realization of a Gaussian process called *prior*, i.e.,

$$\mathcal{M}(\mathbf{x}) \sim \mathcal{GP}(\mu(\mathbf{x}), \sigma^2 r(\mathbf{x}, \mathbf{x}')), \quad (2)$$

where $\mu(\mathbf{x})$ is the prior mean function, or *trend*, $\sigma^2 r(\mathbf{x}, \mathbf{x}')$ is the covariance function, or *kernel*, and σ^2 is its variance [25].

A. Definition of the Prior

Usually, the prior trend is expressed as a linear combination of basis functions, i.e.

$$\mu(\mathbf{x}) = \sum_{k=1}^K \beta_k h_k(\mathbf{x}). \quad (3)$$

The basis functions h_k are often polynomials of increasing order (“universal Kriging”), with a constant trend and a polynomial chaos expansion being notable special cases in the so-called “ordinary Kriging” and “polynomial-chaos-based Kriging” [34], respectively. For a given amount of training data, the latter was shown to perform better or at least as good as plain polynomial chaos expansion and traditional GPR.

Furthermore, the kernel is often described by a stationary correlation function $r(\mathbf{x}, \mathbf{x}') = r(\mathbf{x} - \mathbf{x}' | \boldsymbol{\theta})$, which depends only on the difference between \mathbf{x} and \mathbf{x}' and on a set of lengthscales $\boldsymbol{\theta} = (\theta_1, \dots, \theta_d)$. The lengthscales describe the degree of correlation along each input dimension, which in turn defines the smoothness of the function in that dimension. These are the so-called *hyperparameters* of the GPR model. The squared-exponential and Matérn 5/2 covariaces are popular kernel functions (see, e.g., [25], [35]). The kernel is said to be *isotropic* if the same lengthscale is considered for each dimension, or *anisotropic* otherwise.

B. Classical Prediction

If the prior trend and kernel are fully known a priori, the realization that best represents the function (1) is inferred by collecting a certain number of observations $\{(\mathbf{x}_l, y_l)\}_{l=1}^L$, with $y_l = \mathcal{M}(\mathbf{x}_l)$ (the “training samples”), and applying Bayesian inference. This leads to the best linear unbiased prediction of y at unseen points [31], which is provided by the *posterior* trend [35], [36]:

$$m(\mathbf{x}) = \mathbf{h}(\mathbf{x})^\top \hat{\boldsymbol{\beta}} + \mathbf{r}(\mathbf{x})^\top \mathbf{R}^{-1}(\mathbf{y} - \mathbf{H}\hat{\boldsymbol{\beta}}), \quad (4)$$

where:

- $\mathbf{y} = (y_1, \dots, y_L)^\top \in \mathbb{R}^L$ is the vector of training observations;
- $\mathbf{R} \in \mathbb{R}^{L \times L}$ is the correlation matrix of the training samples, with entries $R_{lm} = r(\mathbf{x}_l, \mathbf{x}_m)$;

- $\mathbf{r}(\mathbf{x}) = (r(\mathbf{x}_1, \mathbf{x}), \dots, r(\mathbf{x}_L, \mathbf{x}))^\top \in \mathbb{R}^L$ is the vector of cross-correlations between the training samples and the prediction point;
- $\mathbf{H} \in \mathbb{R}^{L \times K}$ is the matrix of the trend basis functions evaluated at the training samples, with entries $H_{lk} = h_k(\mathbf{x}_l)$;
- $\mathbf{h}(\mathbf{x}) = (h_1(\mathbf{x}), \dots, h_K(\mathbf{x}))^\top \in \mathbb{R}^K$ is the vector with the trend basis functions evaluated at the prediction point.

Without loss of generality, we assume the training observations to be noiseless. The framework is readily extended, with minimal modifications, to the case of noisy data [25], [35].

One important feature of this Bayesian approach is that it allows assigning *quantitative* information on the prediction uncertainty due to the limited amount of information that is used for the inference. This is expressed by the function

$$\varsigma_0(\mathbf{x}, \mathbf{x}') = r(\mathbf{x}, \mathbf{x}') - \mathbf{r}(\mathbf{x}')^\top \mathbf{R}^{-1} \mathbf{r}(\mathbf{x}), \quad (5)$$

which describes the posterior correlation between predictions at two different points. Hence, the GPR model (posterior) is a Gaussian process with mean function (4) and covariance function $c_0(\mathbf{x}, \mathbf{x}') = \sigma^2 \varsigma_0(\mathbf{x}, \mathbf{x}')$, i.e.,

$$\mathcal{M}_{\text{GPR-0}}(\mathbf{x}) \sim \mathcal{GP}(m(\mathbf{x}), c_0(\mathbf{x}, \mathbf{x}')). \quad (6)$$

In the above equations, the subscript “0” is used to differentiate this “primitive” covariance model from the modifications that will be introduced later. We shall label this model as “GPR-0”, indicating a model with the lowest confidence level. This corresponds to the most standard and popular model considered in the literature and available in commercial tools (e.g., MATLAB[®]).

The quantity $c_0(\mathbf{x}, \mathbf{x})$ provides the variance of the prediction. The variance collapses to zero at the training samples, as (4) provides an exact interpolation of the observations [31]. Since the posterior distribution remains Gaussian, knowing the prediction variance allows obtaining rigorous confidence bounds. For instance, we can expect the true value of y to lie with 95% probability within the interval $[m(\mathbf{x}) - 2\sqrt{c_0(\mathbf{x}, \mathbf{x})}, m(\mathbf{x}) + 2\sqrt{c_0(\mathbf{x}, \mathbf{x})}]$.

C. Estimation of the Prior Parameters

Unfortunately, the trend coefficients $\boldsymbol{\beta} = (\beta_1, \dots, \beta_K)^\top$ and the kernel hyperparameters $(\sigma^2, \boldsymbol{\theta})$ are in practice never known a priori, and they are typically left as free parameters to be estimated based on the training data. This allows better adapting the GPR model to the problem at hand.

The trend coefficients are estimated using a generalized least-square estimate, leading to [35], [36]

$$\hat{\boldsymbol{\beta}} = (\mathbf{H}^\top \mathbf{R}^{-1} \mathbf{H})^{-1} \mathbf{H}^\top \mathbf{R}^{-1} \mathbf{y}. \quad (7)$$

The kernel parameters are instead estimated by minimizing a suitable objective function $\mathcal{F}(\sigma^2, \boldsymbol{\theta})$, i.e.,

$$\hat{\sigma}^2, \hat{\boldsymbol{\theta}} = \arg \min_{\sigma^2, \boldsymbol{\theta}} \mathcal{F}(\sigma^2, \boldsymbol{\theta}). \quad (8)$$

Typical objective functions are the negative log-likelihood or a suitable (e.g., leave-one-out) cross-validation error [25], [35].

Once an estimate of the prior parameters $(\hat{\boldsymbol{\beta}}, \hat{\sigma}^2, \hat{\boldsymbol{\theta}})$ is available, they are plugged into (4) and (5) to provide the prediction and the corresponding uncertainty. However, this approach neglects the uncertainty that is introduced by the estimation of the prior parameters themselves, thereby leading to a likely overestimation of prediction confidence [30].

In this paper, we show that, by means of relatively simple modifications to the posterior covariance, we can account for the uncertainty in the estimate of – at least – $\boldsymbol{\beta}$ and σ^2 . Accounting also for the uncertainty in the estimation of $\boldsymbol{\theta}$ would make the framework much more complex instead [31].

D. Multi-Output Systems

The basic framework works with scalar QoIs, as GPR inherently provides a single-output model. On the other hand, the QoI is oftentimes a function of some sweep variable, e.g., time or frequency. Moreover, several concurrent outputs may be of interest, e.g., in multi-port systems. A naive solution is to train a separate GPR model for each time point, frequency, and/or port variable of interest. However, this would quickly become inefficient because of the massive amount of models to be trained.

To circumvent this problem, multi-output formulations of GPR were proposed in the literature [37], [38], [39], [40]. Another viable solution, which we will adopt in this paper, is to compress the multi-output data using PCA [9] or, similarly, proper orthogonal decomposition [41], [42]. This limits the separate models to be trained to the principal components only, thereby effectively reducing their number by orders of magnitude. The interested Reader is referred to [9], [10] for further details on the use of PCA in conjunction with surrogate models.

III. CONSERVATIVE GP MODELS

In this section, we outline the modifications of the posterior covariance that are required to account for the uncertainty in the estimation of $\boldsymbol{\beta}$ first and, in a second step, also of σ^2 . An important fact is that the posterior prediction remains the one provided by (4).

To account for the uncertainty in the prediction of $\boldsymbol{\beta}$, the posterior correlation (5) requires to be modified with an additional term, leading to [31], [36]

$$\varsigma_1(\mathbf{x}, \mathbf{x}') = \varsigma_0(\mathbf{x}, \mathbf{x}') + \mathbf{u}(\mathbf{x})^\top (\mathbf{H}^\top \mathbf{R}^{-1} \mathbf{H})^{-1} \mathbf{u}(\mathbf{x}'), \quad (9)$$

where

$$\mathbf{u}(\mathbf{x}) = \mathbf{H}^\top \mathbf{R}^{-1} \mathbf{r}(\mathbf{x}) - \mathbf{h}(\mathbf{x}). \quad (10)$$

The posterior distributions remaining Gaussian, with covariance function $c_1(\mathbf{x}, \mathbf{x}') = \sigma^2 \varsigma_1(\mathbf{x}, \mathbf{x}')$. We shall label this model “GPR-1”, with

$$\mathcal{M}_{\text{GPR-1}}(\mathbf{x}) \sim \mathcal{GP}(m(\mathbf{x}), c_1(\mathbf{x}, \mathbf{x}')). \quad (11)$$

To account also for the uncertainty in the prediction of σ^2 , we must use a Bayesian setting in which the kernel variance is inferred from the available data. In this case, the posterior distribution is no longer Gaussian, but it rather

becomes a Student's t -distribution with $\nu = L - K$ degrees of freedom [43] and scale function [31]

$$\varsigma_2(\mathbf{x}, \mathbf{x}') = \frac{\gamma}{\nu} \varsigma_1(\mathbf{x}, \mathbf{x}'), \quad (12)$$

where

$$\gamma = (\mathbf{y} - \mathbf{H}\hat{\boldsymbol{\beta}})^\top \mathbf{R}^{-1}(\mathbf{y} - \mathbf{H}\hat{\boldsymbol{\beta}}). \quad (13)$$

Recalling that the t -distribution is a sort of "wider" normal distribution, which approaches the latter for $\nu \rightarrow \infty$, we can observe that the prediction uncertainty is higher if the number of training samples L is low and/or the number of trend basis functions K is large, which is reasonable.¹ The posterior covariance is in this case given by $c_2(\mathbf{x}, \mathbf{x}') = \frac{\nu}{\nu-2} \varsigma_2(\mathbf{x}, \mathbf{x}')$, which is defined only for $\nu > 2$. We shall label this model "GPR-2", with

$$\mathcal{M}_{\text{GPR-2}}(\mathbf{x}) \sim \mathcal{TP}(m(\mathbf{x}), \varsigma_2(\mathbf{x}, \mathbf{x}'), \nu), \quad (14)$$

where the above notation indicates a Student's t -process with ν degrees of freedom.

In summary, while GPR-0 only accounts for the prediction uncertainty resulting from the limited amount of data that is used to train the model, GPR-1 also accounts for the uncertainty introduced by the estimation of the trend coefficients $\hat{\boldsymbol{\beta}}$ and GPR-2 further accounts for the uncertainty due to the estimate of the kernel variance $\hat{\sigma}^2$. As we will show, this leads to an increasing level of conservativity and to a more accurate confidence information. It is important to remark that the proposed corrections to the posterior covariance are readily applied to any GPR formulation for which the prior trend can be cast as in (3), i.e., it is linear in its unknown coefficients $\boldsymbol{\beta}$, and the kernel variance σ^2 can be factored out from the prior correlation function. This covers the vast majority of the formulations available in the literature.

IV. SURROGATE MODELING FOR BO AND UQ

Using the GPR model to surrogate (1) implies using (4) to generate model predictions, possibly combined with (5), (9), or (12) to assess the confidence of such predictions with an increasing level of conservativity. For a significance level of α (i.e., a confidence level of $1 - \alpha$), the upper and lower confidence bounds are given by

$$\text{UCB}(\mathbf{x}) = m(\mathbf{x}) + F^{-1}(\alpha/2) \cdot s(\mathbf{x}) \quad (15)$$

and

$$\text{LCB}(\mathbf{x}) = m(\mathbf{x}) + F^{-1}(1 - \alpha/2) \cdot s(\mathbf{x}) \quad (16)$$

respectively, where $s(\cdot)$ is a scale parameter that is

$$s(\mathbf{x}) = \sqrt{c_i(\mathbf{x}, \mathbf{x})} = \hat{\sigma} \sqrt{\varsigma_i(\mathbf{x}, \mathbf{x})} \quad (17)$$

for GPR-0 and GPR-1, respectively with $i = 0$ and $i = 1$, or

$$s(\mathbf{x}) = \sqrt{\varsigma_2(\mathbf{x}, \mathbf{x})} \quad (18)$$

for GPR-2. Moreover, $F^{-1}(\cdot)$ is the inverse cumulative distribution function of the (normalized) posterior distribution, i.e., the standard normal inverse cumulative distribution function

¹The latter suggests that more degrees of freedom are "spent" to learn the trend coefficients.

for GPR-0 and GPR-1, or the Student's t inverse cumulative distribution function with ν degrees of freedom for GPR-2. For example, for a 95% confidence level, $\alpha = 0.05$ and $F^{-1}(0.975) = -F^{-1}(0.025) \approx 2$ for GPR-0 and GPR-1 (recall the "2-sigma rule") or ≈ 3.18 for a GPR-2 model with $\nu = 3$ degrees of freedom.

A. Application to BO

BO is a global optimization method that is suitable for objective functions that are expensive to evaluate [27]. Starting from a minimal set of observations, a GPR model is fit to the objective function and is iteratively refined by querying additional data samples. The acquisition is driven by the information on the prediction uncertainty [29]. Hence, having a more accurate estimate of the prediction uncertainty improves the optimization performance.

All selection strategies involve, at each iteration, the maximization (or minimization) or some "acquisition function". Hence, the next sampling point is selected as

$$\mathbf{x}_* = \arg \min_{\mathbf{x}} [-\text{AF}(\mathbf{x})] \quad (19)$$

which, for the sake of notation consistence, we cast as a minimization problem. Some popular acquisition functions are discussed in the following, all providing a trade-off between exploration and exploitation.

1) *Lower Confidence Bound (LCB)*: This strategy selects the next sampling point as the location where the LCB (16) is minimum. For a point with a similar prediction uncertainty, the lower confidence bound is more likely to be minimum in the neighborhood of the actual function minimum.

2) *Probability of Improvement (PoI)*: The next sampling point is selected as the one yielding the largest probability of improving the current estimate. The corresponding acquisition function is mathematically defined as

$$\text{PoI}(\mathbf{x}) = F\left(\frac{(y_{\min} - \epsilon) - m(\mathbf{x})}{s(\mathbf{x})}\right), \quad (20)$$

where

$$y_{\min} = \min_{\mathbf{x}} m(\mathbf{x}) \quad (21)$$

denotes the current estimate of the function minimum, i.e., the minimum of the posterior prediction, ϵ is a margin parameter (usually relative, e.g., $\epsilon = 0.001 \cdot |y_{\min}|$), and $F(\cdot)$ is the cumulative distribution function of the (normalized) posterior distribution. In essence, (20) defines the probability of finding a value of y that is smaller than $y_{\min} - \epsilon$.

3) *Expected Improvement (EI)*: This acquisition function is designed to take into account not only the PoI, but also its magnitude. The EI is defined as [44]

$$\text{EI}(\mathbf{x}) = (y_{\min} - m(\mathbf{x})) F(z(\mathbf{x})) + s(\mathbf{x}) \cdot \phi(z(\mathbf{x})) \quad (22)$$

for a Gaussian posterior distribution (GPR-0 and GPR-1), and

$$\begin{aligned} \text{EI}(\mathbf{x}) &= (y_{\min} - m(\mathbf{x})) F(z(\mathbf{x})) \\ &+ \frac{\nu}{\nu - 1} \left(1 + \frac{z^2(\mathbf{x})}{\nu}\right) s(\mathbf{x}) \cdot \phi(z(\mathbf{x})) \end{aligned} \quad (23)$$

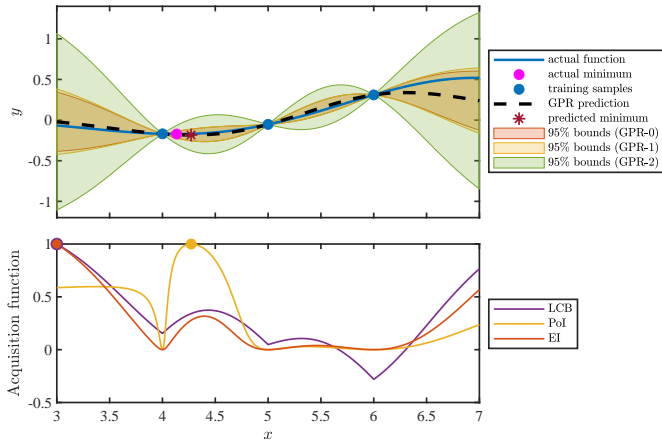


Fig. 1. BO of function (25). The top panel shows the actual function (solid blue line), the training samples (blue circles), and the GPR prediction (dashed black line). The magenta circle and the red asterisk are the actual and predicted minimum, respectively. The shaded areas represent the 95% confidence bounds of GPR-0 (red), GPR-1 (yellow), and GPR-2 (green). The bottom panel shows the LCB (purple), PoI (yellow), and EI (red) acquisition functions.

for a Student's t posterior distribution (GPR-2) [45], where $\phi(\cdot)$ is the probability density function (PDF) of the (normalized) posterior distribution and

$$z(\mathbf{x}) = \frac{y_{\min} - m(\mathbf{x})}{s(\mathbf{x})} \quad (24)$$

is the standardized output minimum.

It should be noted that the aforementioned acquisition functions are all analytical and rather well-behaved. Therefore, the minimization problem (19) is readily solved using some other standard optimization method, such as genetic algorithm.

In Fig. 1, the concept of BO is illustrated with the aim of finding the minimum of the analytical function

$$\mathcal{M}(x) = e^{-\frac{10}{x}} (2 \cos(x) + \sin(x)) \quad (25)$$

in the interval $x \in [3, 7]$. The actual function is shown in the top panel (solid blue line) together with the corresponding prediction (dashed black line) of a GPR model with a constant prior trend and squared-exponential kernel, trained with $L = 3$ samples at points $x = \{4, 5, 6\}$ (blue circles). The red, yellow, and green lines are the 95% confidence bounds of GPR-0, GPR-1, and GPR-2, respectively. The magenta circle indicates the true function minimum (i.e., $y_{\min} = -0.1719$), whereas the red asterisk is the current estimated minimum (i.e., the minimum of the GPR prediction).

The bottom panel of Fig. 1 shows the LCB, PoI, and EI acquisition functions computed based on the GPR-2 posterior, normalized w.r.t. their maximum value for the ease of comparison. The next sampling point, selected as the location where the acquisition function is maximum, is denoted with a circle. It is noted that the LCB and EI prescribe to sample in the corner of the interval, where the prediction uncertainty is the highest, whereas the PoI suggests sampling a point that is much closer to the actual minimum. Nevertheless, any acquisition function leads to the correct result within a few iterations, as indicated by the results summarized in Table I, with PoI providing indeed a slightly faster convergence.

TABLE I
CONVERGENCE OF BO APPLIED TO FUNCTION (25) AND BASED ON GPR-2 WITH DIFFERENT ACQUISITION FUNCTIONS.

Iteration	Acquisition function					
	LCB		PoI		EI	
	x_{\min}	y_{\min}	x_{\min}	y_{\min}	x_{\min}	y_{\min}
0	4.2706	-0.1834	4.2706	-0.1834	4.2706	-0.1834
1	4.1804	-0.1740	4.1803	-0.1740	4.1804	-0.1740
2	4.1344	-0.1718	4.1349	-0.1719	4.1453	-0.1722
3	4.1346	-0.1719	4.1344	-0.1719	4.1358	-0.1718
4	4.1343	-0.1719	4.1352	-0.1719	4.1349	-0.1719
5	4.1349	-0.1719	4.1350	-0.1719	4.1348	-0.1719

B. Application to UQ

GPR was also applied in the context of UQ, in which the prediction confidence was propagated from point-wise predictions to UQ measures, such as moments and distribution functions [10], [11]. Although some semi-analytical results are available that require the integration of the posterior [10], for high-dimensional problems the most viable and convenient approach is to use the GPR model to surrogate the actual system (1) in a Monte Carlo (MC) analysis.

In the above scenario, the GPR prediction is calculated for a (typically large) ensemble of samples $\{\mathbf{x}_i^*\}_{i=1}^N$. This leads to a set of N correlated random variables defining the model predictions at the above ensemble of points, which are described by the mean vector

$$\bar{\mathbf{m}} = \mathbf{h}_*^T \hat{\boldsymbol{\beta}} + \mathbf{r}_*^T \mathbf{R}^{-1} (\mathbf{y} - \mathbf{H} \hat{\boldsymbol{\beta}}) \quad (26)$$

and by the pertinent covariance matrix computed according to the posterior distribution. In (26), $\mathbf{h}_* \in \mathbb{R}^{K \times N}$ and $\mathbf{r}_* \in \mathbb{R}^{L \times N}$ are matrices with entries $[h_*]_{ki} = h_k(\mathbf{x}_i^*)$ and $[r_*]_{li} = r(\mathbf{x}_l, \mathbf{x}_i^*)$, respectively.

To obtain the covariance matrices, we first define the correlation matrices stemming from (5), (9), and (12), i.e.,

$$\boldsymbol{\Sigma}_0 = \mathbf{R}_{**} - \mathbf{r}_*^T \mathbf{R}^{-1} \mathbf{r}_* \quad (27)$$

for GPR-0,

$$\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_0 + \mathbf{u}_*^T (\mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})^{-1} \mathbf{u}_* \quad (28)$$

for GPR-1, and

$$\boldsymbol{\Sigma}_2 = \frac{\gamma}{\nu} \boldsymbol{\Sigma}_1 \quad (29)$$

for GPR-2, where $\mathbf{R}_{**} \in \mathbb{R}^{N \times N}$ is a matrix with entries $[R_{**}]_{ij} = r(\mathbf{x}_i^*, \mathbf{x}_j^*)$ and

$$\mathbf{u}_* = \mathbf{H}^T \mathbf{R}^{-1} \mathbf{r}_* - \mathbf{h}_*. \quad (30)$$

Then, the covariance matrix is $\mathbf{C}_i = \hat{\sigma}^2 \boldsymbol{\Sigma}_i$ for GPR-0 and GPR-1, respectively with $i = 0$ and $i = 1$, and $\mathbf{C}_2 = \frac{\nu}{\nu-2} \boldsymbol{\Sigma}_2$ for GPR-2. The covariance matrix for GPR-2 is well defined only for $\nu > 2$, which requires to have a number of training samples

$$L > K + 2. \quad (31)$$

Based on the above definitions, the mean of the MC predictions (an estimate of the mean of the QoI) is a random variable whose expected value and variance are analytically computed from the mean vector (26) and the pertinent covariance matrix, as reported in [11]. Indeed, the results for the output mean in [11] readily apply regardless of the posterior distribution.

Furthermore, the variance of the MC predictions (an estimate of the variance of the QoI) is also a random variable. Closed-form results for its expected value and variance were also derived in [11], yet the latter is valid only for a Gaussian posterior distribution, i.e., for GPR-0 and GPR-1 models. Nevertheless, empirical results suggest that assuming a Gaussian approximation for the distribution of both the posterior and the variance incurs a small error in the calculation of confidence bounds, while providing a useful assessment of the dispersion due to the limited amount of data used to train the GPR model.

In summary, the outlined procedure allows assigning confidence bounds to the MC prediction of the output mean and output variance, respectively, obtained when the GPR model is used to surrogate the actual model (1). When the posterior covariance of GPR-0 is considered, the confidence bounds only account for the uncertainty due to the limited availability of training data. Using the covariance of GPR-1 and GPR-2 allows accounting also for the uncertainty introduced by the estimation of the trend coefficients $\hat{\beta}$ and kernel variance $\hat{\sigma}^2$, respectively. Confidence bounds for higher-order moments and probability distributions are computed numerically by generating a large number of posterior realizations instead [32].

C. Implementation

The outlined framework is implemented in MATLAB[®] by leveraging the Statistics and Machine Learning Toolbox[™] [46] for training the GPR models. The toolbox estimates the trend coefficients via (7) and the kernel hyperparameters using maximum likelihood. The provided predictive covariance is the one of GPR-0, i.e., (5). For GPR-1 and GPR-2 models, we correct the posterior covariance as outlined in Section III. For the latter, only the estimated lengthscales are retained, whereas the estimate of the kernel variance is already embedded in the posterior distribution.

V. APPLICATION EXAMPLE #1: MULTICONDUCTOR TRANSMISSION LINE NETWORK

The first application test case considers the interconnect of Fig. 2, consisting of three sections of coupled embedded transmission lines, some of which are terminated by nonlinear inverters at the far end [47]. The cross-section of the multiconductor transmission line sections is depicted at the bottom of the figure. The nominal values of the geometrical parameters and element values are provided in the second column of Table II. The central line of the first section is driven by a trapezoidal voltage pulse with an amplitude of 5 V, a duration of 1 ns, and rise/fall times of 100 ps.

The QoI is taken as the maximum absolute value of the far-end crosstalk occurring over time at the 0.5-pF termination (i.e., C_{c3}). We perform an UQ task first, and then we optimize the parameters to minimize crosstalk. The simulations are

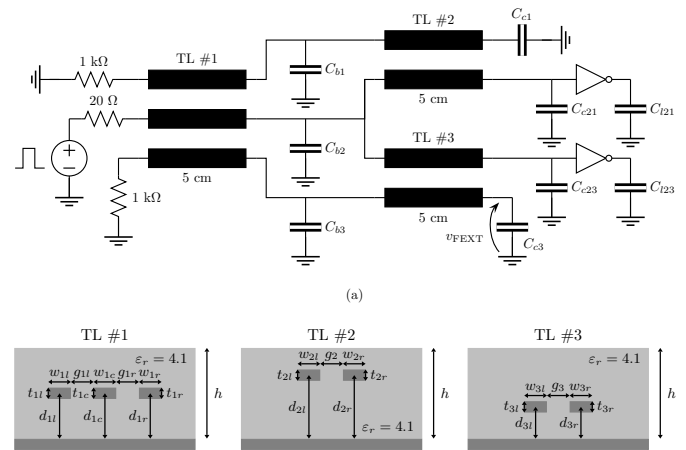


Fig. 2. Schematic of the transmission line network.

carried out using HSPICE [48] which, for this example, implements the map (1) between the circuit parameters and the QoI.

TABLE II
NOMINAL VALUES OF THE INTERCONNECT PARAMETERS, TOGETHER WITH TOLERANCE AND RANGE OF VARIATION FOR UQ AND BO TASKS.

Parameter(s)	Nominal value	Tolerance (UQ)	Range (BO)
w_{1l}, w_{1c}, w_{1r}	150 μm	10%	$\pm 30 \mu\text{m}$
w_{2l}, w_{2r}	130 μm	10%	$\pm 30 \mu\text{m}$
w_{3l}, w_{3r}	170 μm	10%	$\pm 30 \mu\text{m}$
d_{1l}, d_{1c}, d_{1r}	100 μm	10%	$\pm 20 \mu\text{m}$
d_{2l}, d_{2r}	140 μm	10%	$\pm 20 \mu\text{m}$
d_{3l}, d_{3r}	70 μm	10%	$\pm 20 \mu\text{m}$
g_{1l}, g_{1r}	150 μm	10%	$\pm 30 \mu\text{m}$
g_2	150 μm	10%	$\pm 30 \mu\text{m}$
g_3	150 μm	10%	$\pm 30 \mu\text{m}$
h	200 μm	-	$\pm 20 \mu\text{m}$
t_{1l}, t_{1c}, t_{1r}	30 μm	-	$\pm 10 \mu\text{m}$
t_{2l}, t_{2r}	20 μm	-	$\pm 10 \mu\text{m}$
t_{3l}, t_{3r}	40 μm	-	$\pm 10 \mu\text{m}$
C_{b1}, C_{b2}, C_{b3}	1 pF	-	$\pm 0.25 \text{ pF}$
C_{c1}, C_{c21}, C_{c23}	1 pF	-	$\pm 0.25 \text{ pF}$
C_{c3}	0.5 pF	-	$\pm 0.25 \text{ pF}$
C_{l21}, C_{l23}	1 pF	-	$\pm 0.25 \text{ pF}$

A. UQ of Maximum Crosstalk

For this analysis, the uncertainty is provided by $d = 9$ parameters, namely the trace widths, gaps, and distances from the ground plane in each of the three sections, whose nominal values are indicated in the first nine rows of Table II. These parameters are assumed to have a Gaussian distribution with a 10% standard deviation from the nominal value, and the same variation is considered for all traces within the same section (i.e., for all parameters in the same row of Table II).

For the GPR models, we consider a polynomial-chaos-Kriging model [34] with a second-order Hermite expansion for the trend and an isotropic squared-exponential kernel. As a rule of thumb, it is often recommended to use a number of training samples from 5 to 10 times the number of uncertain

parameters (see, e.g., [11], [31]). Since the trend features $K = 55$ expansion terms, we consider $L = \{60, 70, 70\}$ training samples so that condition (31) is also fulfilled. The training points are generated randomly using a uniform Latin hypercube sampling scheme in the hyperinterval spanning three times the standard deviation (i.e., $\pm 30\%$) around the nominal value of the parameters.

TABLE III
MEAN AND VARIANCE OF THE MAXIMUM CROSSTALK VOLTAGE.

Method	Observations	Mean (V)	Variance (V ²)
MC	$N = 1000$	0.4795	0.0043
GPR-0		[0.4700, 0.4705]	[0.0036, 0.0037]
GPR-1	$L = 60$	[0.4588, 0.4817]	[0.0035, 0.0040]
GPR-2		[0.4191, 0.5215]	[0.0025, 0.0086]
GPR-0		[0.4921, 0.4931]	[0.0039, 0.0040]
GPR-1	$L = 70$	[0.4856, 0.4996]	[0.0037, 0.0043]
GPR-2		[0.4763, 0.5089]	[0.0037, 0.0050]
GPR-0		[0.4813, 0.4818]	[0.0040, 0.0041]
GPR-1	$L = 80$	[0.4780, 0.4851]	[0.0040, 0.0042]
GPR-2		[0.4749, 0.4882]	[0.0039, 0.0043]

Table III provides the results for the first two statistical moments, i.e., mean and variance, of the maximum crosstalk voltage. The first row reports the result obtained with a MC simulation performed in HSPICE for $N = 1000$ random configurations of the geometrical parameters. The remaining rows provide the 2-sigma intervals of the moments predicted with the GPR model based on the *same* ensemble of MC samples², computed for different conservativity levels of the prediction uncertainty and for increasing training set size. It is observed that the confidence intervals of GPR-0 are very narrow and do not include the reference values, even for the largest training dataset with $L = 80$ samples. On the contrary, the estimates of GPR-1 and GPR-2 models are more conservative, yet only the latter always include the correct result. As the number of training samples is increased, the confidence interval of GPR-2 shrinks around the correct result.

This example shows that the classical model of the posterior covariance is prone to an underestimation of the prediction uncertainty, while the introduced corrections lead to a more accurate estimation. As mentioned before, confidence bounds can be obtained numerically also for the PDF of the QoI. Results in this regard were provided in [32] are therefore they are not repeated here.

So far, the results were obtained with a single configuration of the training dataset for a given size. To evaluate the dispersion of the performance across different training datasets, the analysis is repeated for 100 independent runs for each training set size (i.e., $L = \{60, 70, 80\}$). The aim of this analysis is to assess whether the better conservativity holds for many random configurations of the training samples.

Figure 3 shows the distribution of the upper and lower 2-sigma bounds of the predicted mean (top panels) and predicted variance (bottom panels) obtained with the smallest

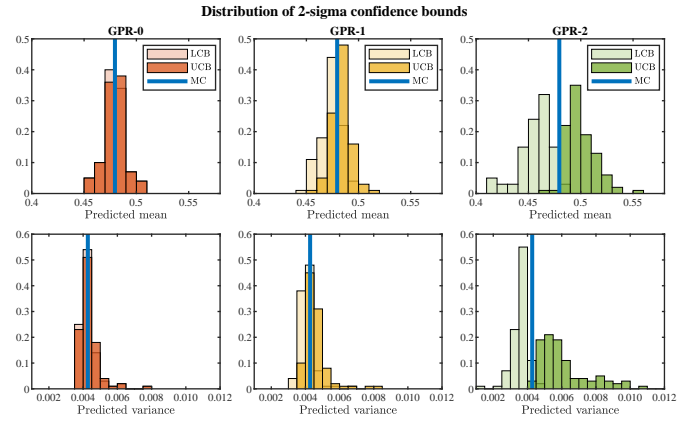


Fig. 3. Distribution of the upper and lower bounds of the 2-sigma confidence interval of the maximum crosstalk mean (top panels) and variance (bottom panels) computed over 100 independent runs for the GPR-0, GPR-1, and GPR-2 models (red, yellow, and green histograms, respectively) trained with $L = 60$ samples, and comparison with the reference MC values (blue bars).

training set size (i.e., $L = 60$) and the three different conservativity levels (left, central, and right panels). In particular, the lighter and darker histograms show the distribution of the LCB and UCB, respectively. The vertical blue line indicates the reference value from the MC analysis (cfr. Table III).

It is interesting to note that for GPR-0 (left plot, red histograms) the two distributions overlap completely, indicating that it is very likely that the reference MC value falls outside the confidence interval. The two distributions become increasingly separated for the more conservative GPR-1 and GPR-2 models (central and right plots, respectively). In particular, for GPR-2 (green histograms) the two distributions are well apart, with the reference value lying in the middle. This signifies that the confidence bounds almost always enclose the correct result, as their distributions rarely cross its value.

TABLE IV
A POSTERIORI VERIFICATION OF THE ACTUAL PREDICTION CONFIDENCE.

Model	Mean	Variance	Mean	Variance	Mean	Variance
	$L = 60$		$L = 70$		$L = 80$	
GPR-0	5%	5%	5%	9%	11%	8%
GPR-1	41%	45%	64%	47%	82%	32%
GPR-2	95%	94%	99%	96%	97%	86%

The above conclusion is further corroborated by counting the rate at which the correct value falls within the 2-sigma confidence interval over the 100 runs. As indicated in Table IV, the rate is remarkably close to 95% for GPR-2, whereas it is much lower for GPR-0 and GPR-1.³ This indicates a severe overestimation of the prediction confidence, which turns out to be actually much lower than 95% for those models.

Finally, Fig. 4 compares the performance of the GPR-2 model against the popular polynomial chaos method based on least-angle regression [49], implemented via the UQLab toolbox [50] and using the same second-order Hermite expansion as the GPR trend. In particular, the boxplot assesses the

³The 2-sigma bounds theoretically correspond to a 95% confidence only for the mean computed with GPR-0 and GPR-1 models.

²This ensures the results are comparable regardless of the MC sample size.

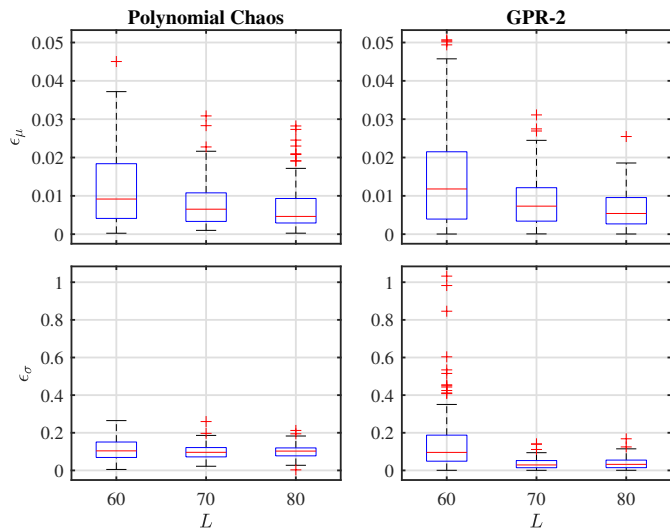


Fig. 4. Relative error on the mean (top) and variance (bottom) obtained with the plain polynomial chaos method (left) and the GPR-2 model with the same polynomial chaos expansion as trend (right).

dispersion across the 100 training datasets of the relative error on the mean and variance w.r.t. the MC result, respectively denoted with ϵ_μ and ϵ_σ . The red lines indicate the median error, whereas the bottom and top edges of the boxes are the 25th and 75th percentiles, respectively. The whiskers indicate the minimum and maximum values, while the red crosses are outliers. For the polynomial chaos method, the mean and variance are analytically derived from the model coefficients. For the GPR model, they are obtained via the analytical relations in [11].

It is noted that the two methods achieves a similar error for the mean, which is usually estimated with very good accuracy by most methods. For the variance instead, the GPR model overall provides a better accuracy, with a median error of about 3% for the datasets with 70 and 80 training samples, compared to about 10% obtained with polynomial chaos. The higher dispersion of the GPR-2 model trained with 60 samples reflects its higher and more conservative prediction uncertainty, while the median error is still slightly lower compared to polynomial chaos. These results confirm that GPR models provide better or comparable accuracy as polynomial chaos, an outcome that was also highlighted in [11] and [34].

B. BO of Maximum Crosstalk

In this second scenario, we aim to minimize the far-end crosstalk w.r.t. all the 35 design variables indicated in Table II, which are now allowed to vary independently in the range indicated in the last column. To this end, we implement the BO scheme outlined in Section IV-A in conjunction with both GPR-0 and GPR-2 models.

It should be noted that the goal of this example is to assess the efficacy of the optimization scheme with a more conservative posterior covariance on a large input space, but some of the parameters considered may not be optimizable in practice (or, at least, not individually) due to other design or fabrication constraints. Moreover, even if the optimal design

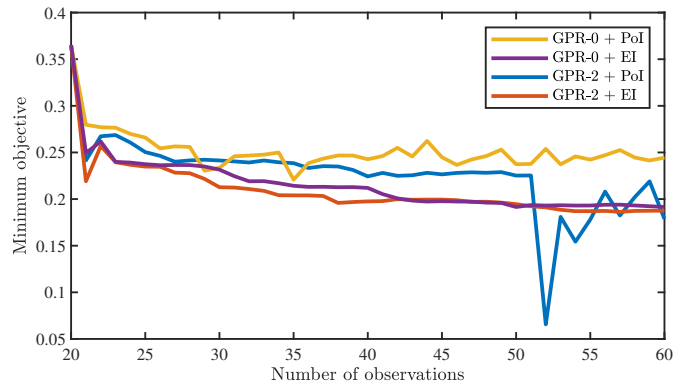


Fig. 5. Evolution of the minimum objective (maximum crosstalk) throughout the iterations of the BO scheme implemented with GPR-0 and GPR-2 models in conjunction with PoI and EI acquisition functions.

may lie in a corner of the design space, a full corner analysis would require to test $2^{35} \approx 3 \times 10^{10}$ configurations.

For the GPR models, we use an isotropic squared-exponential kernel and a constant trend, since a polynomial trend would feature an excessive number of terms in such a high-dimensional space. We start from a common initial dataset of $L = 20$ observations to initialize the GPR model, and we run the optimization scheme for 40 iterations using either the PoI or EI acquisition function.

Figure 5 compares the evolution of the minimum predicted objective (maximum crosstalk, in volts) throughout the BO iterations obtained with the four combinations of the two GPR models and the two acquisition functions. It is observed that the EI acquisition function (red and purple curves) provides a faster and smoother convergence compared to PoI (yellow and blue curves) for a given model type. Moreover, for both acquisition functions, a better convergence is achieved by using the GPR-2 model instead of the GPR-0 one, especially in the first iterations. This is reasonable, since increasing the number of observations in turn increases the number of degrees of freedom in the GPR-2 posterior distribution, thereby reducing the difference in the covariance. In general, since GPR-2 provides a more accurate estimation of the prediction confidence, on which BO is highly reliant, we expect GPR-2 to always provide better, or at least similar performance compared to the standard GPR-0 model. Hence, GPR-2 should be preferred over GPR-0, also in view of its limited additional computational complexity.

Furthermore, the top panel of Fig. 6 compares the transient crosstalk response of the optimal design achieved with GPR-2 and EI (red line) against the spread of 1000 MC configurations drawn randomly within the design space (gray area). The comparison highlights that the maximum crosstalk over time of the optimal design is lower than any of the random MC configurations.

The bottom panel further enhance the comparison, with the cyan circles indicating the maximum crosstalk level of all the random configurations, whereas the colored lines indicate the maximum crosstalk of the optimal designs achieved with the aforementioned four configurations of BO (i.e., GPR-0 and GPR-2 models with PoI and EI acquisition functions). It

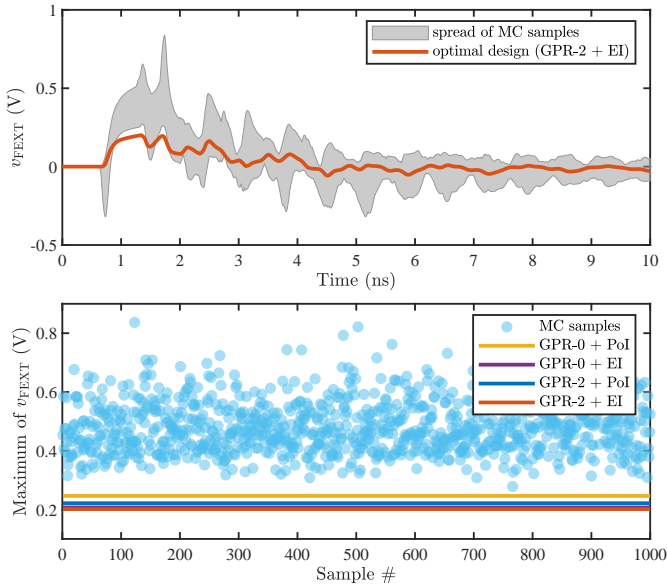


Fig. 6. Performance of the optimal design(s) in comparison with 1000 random configurations within the design space. The top panel compares the spread of the entire transient crosstalk of the random configurations (gray area) against the optimal design achieved with BO in conjunction with the GPR-2 model and EI the acquisition function (red curve). The bottom panel compares the maximum crosstalk over time of each random configuration (cyan circles) against the maximum crosstalk of the optimal designs achieved with BO and the four different configurations of model type and acquisition function (colored lines).

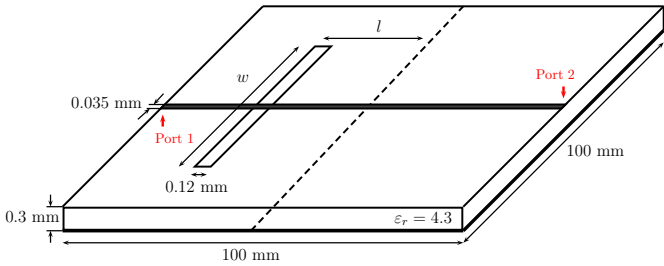


Fig. 7. Layout of the microstrip line with ground plane discontinuity.

should be noted that, as opposed to Fig. 5, the results in Fig. 6 now refer to the actual maximum crosstalk that is achieved in the SPICE simulation, and not to the one predicted by the GPR model. It is observed that the all optimal designs exhibit a lower crosstalk level than any of the random configurations, and the overall best result is achieved by using GPR-2 in conjunction with the EI acquisition function (red line).

VI. APPLICATION EXAMPLE #2: MICROSTRIP WITH GROUND PLANE DISCONTINUITY

The second application example concerns the analysis of the signal integrity in a microstrip line with a discontinuity in the ground plane, shown in Fig. 7 [20], [51]. The free design parameters are the location l and the width w of the slot in the ground plane, both with a nominal value of 15 mm. The scattering (S)-parameters are simulated with CST Studio Suite® software from Dassault Systemès [52] at 1859 equally-spaced frequency points from DC to 10 GHz.

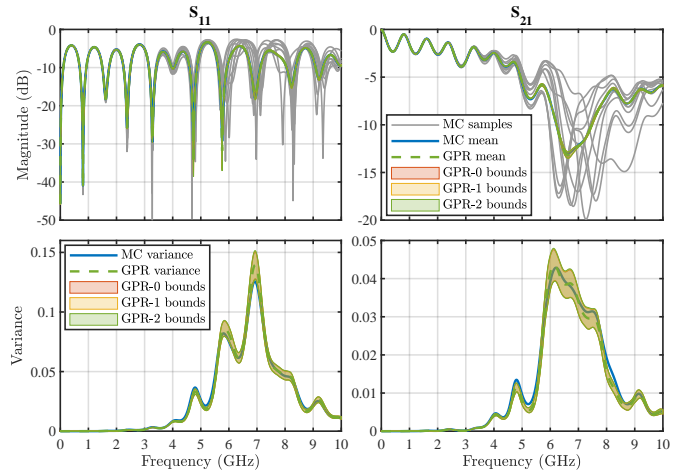


Fig. 8. UQ of S_{11} and S_{21} . The top panels show the magnitude of a subset of samples from the MC simulation (gray lines) together with the average value computed from the MC samples (solid blue line) and predicted by the GPR models (dashed green line). The shaded areas are the 2-sigma confidence intervals of the GPR-0 (red), GPR-1 (yellow), and GPR-2 (green) predictions. The bottom panels compare the variance obtained with MC and the one predicted by the GPR models (same color labeling as above).

A. UQ of the S-Parameters in the Frequency Domain

In this first scenario, the two design parameters are ascribed an independent Gaussian distribution with a 10% relative standard deviation. We aim to assess the variability of the S-parameters in the entire frequency range. To this end, we compute $L = 50$ training configurations of the two-port S-parameters and we compress the resulting dataset using a PCA with a 1% truncation threshold on the singular values. This allows compressing the dataset from $1859 \times 4 = 7436$ outputs to a mere 18 principal components. Since we build a separate model for the real and imaginary part of the scattering data [11], the total number of individual GPR models to be trained is 36. A constant trend and an anisotropic Matérn 5/2 kernel are considered for the prior.

The top panels of Fig. 8 show the magnitude of S_{11} (left) and S_{21} (right) for some random configurations of the design parameters (gray lines), as well as their mean computed based on 1000 MC samples (solid blue line) and the GPR models (dashed green line). The 2-sigma confidence intervals of the GPR models (shaded areas) are also provided. The bottom panels show a similar comparison between the variance of the MC samples and the corresponding GPR predictions. In this case, the prediction confidence of GPR-0, GPR-1, and GPR-2 models is similar due to the relatively large number of training samples that is required to properly capture the high-frequency variability of the scattering responses.

The GPR results are computed by recombining the 18 principal components to obtain the model of the real and imaginary part of the S-parameters at each port and frequency, as indicated in Section II-D. The posterior mean and covariance are then combined with the analytical estimates of the output mean and variance. In this process, the posterior distribution is assumed to be Gaussian, even though this is not the case for GPR-2. Nevertheless, a good accuracy of the GPR model is established, despite the very large variability of the S-

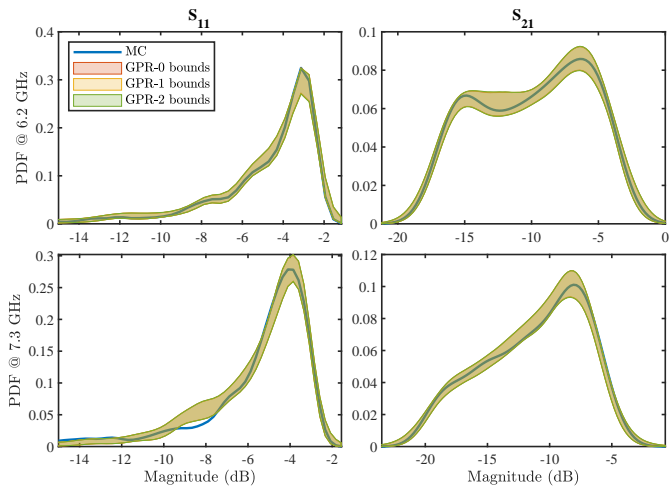


Fig. 9. PDF of $|S_{11}|$ (left) and $|S_{21}|$ (right) computed at 6.2 GHz (top) and 7.3 GHz (bottom). The distribution of the MC samples (solid blue line) is compared against the 95% confidence intervals of the GPR-0, GPR-1, and GPR-2 predictions (red, yellow, and green areas, respectively).

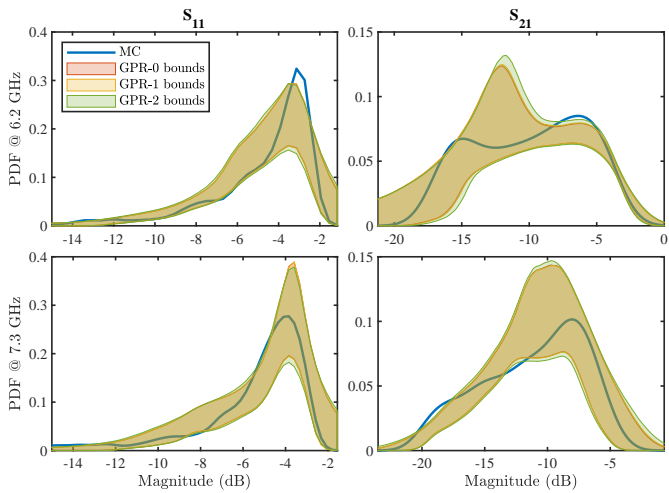


Fig. 10. Comparison between the confidence of GPR-0, GPR-1, and GPR-2 models trained with only 10 samples in predicting the distribution of $|S_{11}|$ and $|S_{21}|$ at 6.2 GHz and 7.3 GHz.

parameters that can be observed in Fig. 8. Indeed, as it was shown in [20], the accurate modeling of the high-frequency response is particularly challenging for this structure.

Furthermore, Fig. 9 provides a comparison on the PDF at the frequencies of 6.2 GHz and 7.3 GHz (top and bottom panels, respectively). The blue lines are the distribution of the MC samples of the S-parameter magnitude at those frequencies. The shaded areas are the 95% confidence bounds of the GPR-0, GPR-1, and GPR-2 predictions. These are obtained by generating a large number of posterior predictions and, as noted before, turn out to be similar in this case. Once again, they compare well with the reference MC distribution.

To better illustrate the difference between the prediction confidence of the various GPR models, Fig. 10 compares the GPR-0, GPR-1, and GPR-2 confidence intervals in predicting the same distributions based on 10 training samples only. The confidence intervals are now wider to reflect the larger

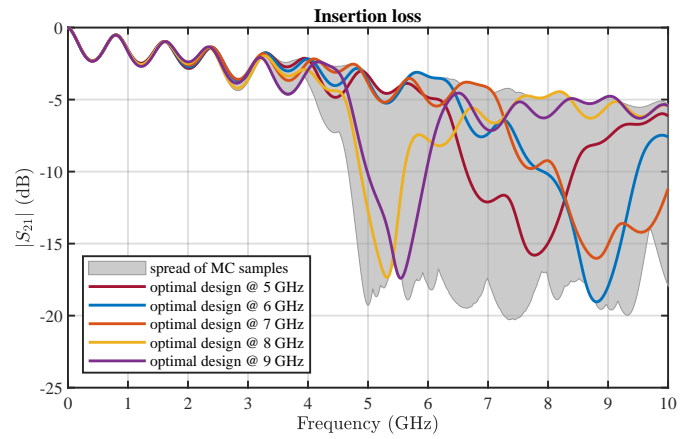


Fig. 11. Insertion loss of the optimal designs at various frequencies (colored lines) and comparison against the spread of responses of 1000 random configurations of the design parameters (gray area).

prediction uncertainty that results from considering a much lower number of training samples. In turn, the discrepancy between the various confidence intervals is also more visible, which reflects the higher uncertainty in the estimation of the trend coefficient and kernel variance.

B. BO of the Insertion Loss at Specific Frequencies

We now want to obtain the configurations that maximize the transmission (i.e., minimize the insertion loss) at some specific frequencies, namely $f_0 = (5, 6, 7, 8, 9)$ GHz. To this end, we run a BO scheme by considering the insertion loss at the operating frequency, i.e., $IL = -20 \log_{10} |S_{21}(f_0)|$, as the objective function. We consider a range of the design parameters within $\pm 35\%$ around their nominal value. We start from a set of $L = 10$ configurations to initialize the GPR-2 model of the insertion loss and we run 20 iterations of the optimizer. Since the optimal configuration will differ for each operating frequency, we run an individual optimization for each value of f_0 .

TABLE V
OPTIMAL VALUE OF THE DESIGN PARAMETERS MINIMIZING THE INSERTION LOSS AT VARIOUS OPERATING FREQUENCIES.

Frequency	f_0 (GHz)	5	6	7	8	9
Optimal value	l (mm)	11.5	13.7	17.6	15.5	19.0
	w (mm)	13.1	11.4	11.0	18.7	18.3

The optimal design values are indicated in Table V, whereas Fig. 11 illustrates the corresponding responses. The insertion loss of the optimal designs (colored lines) is compared against the spread of 1000 random responses from the MC analysis (gray area). It is observed that the optimal design provides indeed the smallest insertion loss at the corresponding operating frequency.

VII. CONCLUSIONS

This paper introduced conservative GPR models with application to UQ and BO in signal integrity analyses. Compared to

classical GPR implementations, whose prediction uncertainty accounts only for the lack of training data, the proposed models also take into account the estimation of the prior trend coefficients and kernel variance. As a result, the estimated prediction confidence turns out to be more accurate and much closer to the actual one.

The advocated framework was applied to two test cases involving the analysis of crosstalk in a multiconductor transmission line network and of the insertion loss in a microstrip with a ground plane discontinuity, for which both UQ and optimization tasks were performed. The results highlighted that the proposed conservative GPR models yield more accurate results compared to state-of-the-art implementations of GPR.

REFERENCES

- [1] M. B. Yelten, T. Zhu, S. Koziel, P. D. Franzon, and M. B. Steer, "Demystifying surrogate modeling for circuits and systems," *IEEE Circuits Syst. Mag.*, vol. 12, no. 1, pp. 45–63, Jan.–Mar. 2012.
- [2] T. Nguyen, B. Shi, H. Ma, E.-P. Li, X. Chen, A. C. Cangellaris, and J. Schutt-Ainé, "Comparative study of surrogate modeling methods for signal integrity and microwave circuit applications," *IEEE Trans. Compon. Packag. Manuf. Technol.*, vol. 11, no. 9, pp. 1369–1379, Sep. 2021.
- [3] T. Lu, J. Sun, K. Wu, and Z. Yang, "High-speed channel modeling with machine learning methods for signal integrity analysis," *IEEE Trans. Electromagn. Compat.*, vol. 60, no. 6, pp. 1957–1964, Dic. 2018.
- [4] M. Swaminathan, H. M. Torun, H. Yu, J. A. Hejase, and W. D. Becker, "Demystifying machine learning for signal and power integrity problems in packaging," *IEEE Trans. Compon. Packag. Manuf. Technol.*, vol. 10, no. 8, pp. 1276–1295, Aug. 2020.
- [5] C. S. Geyik, Z. Zhang, K. Aygün, and J. T. Aberle, "Machine learning for evaluating the impact of manufacturing process variations in high-speed interconnects," in *Proc. 22nd Int. Symp. Quality Electron. Des. (ISQED)*, Santa Clara, CA, USA, Apr. 2021, pp. 160–163.
- [6] A. Ciccazzo, G. D. Pillo, and V. Latorre, "Support vector machines for surrogate modeling of electronic circuits," *Neural Comput. Appl.*, vol. 24, pp. 69–76, 2014.
- [7] R. Trinchero, M. Larbi, H. M. 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, 2018.
- [8] H. Ma, E.-P. Li, A. C. Cangellaris, and X. Chen, "Support vector regression-based active subspace (SVR-AS) modeling of high-speed links for fast and accurate sensitivity analysis," *IEEE Access*, vol. 8, pp. 74 339–74 348, 2020.
- [9] 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.
- [10] —, "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. 2021.
- [11] P. Manfredi, "Probabilistic uncertainty quantification of microwave circuits using Gaussian processes," *IEEE Trans. Microw. Theory Techn.*, vol. 71, no. 6, pp. 2360–2372, Jun. 2023.
- [12] F. Garbuglia, T. Reuschel, C. Schuster, D. Deschrijver, T. Dhaene, and D. Spina, "Modeling electrically long interconnects using physics-informed delayed Gaussian processes," *IEEE Trans. Electromagn. Compat.*, vol. 65, no. 6, Dec. 2023.
- [13] H. D. Kabir, A. Khosravi, M. A. Hosen, and S. Nahavandi, "Neural network-based uncertainty quantification: A survey of methodologies and applications," *IEEE Access*, vol. 6, pp. 218–36 234, 2018.
- [14] S. Chen, J. Chen, T. Zhang, and S. Wei, "Semi-supervised learning based on hybrid neural network for the signal integrity analysis," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 67, no. 10, pp. 1934–1938, Oct. 2019.
- [15] K. Son, M. Kim, H. Park, D. Lho, K. Son, K. Kim, S. Lee, S. Jeong, S. Park, S. Hong, G. Park, and J. Kim, "Reinforcement-learning-based signal integrity optimization and analysis of a scalable 3-D X-point array structure," *IEEE Trans. Compon. Packag. Manuf. Technol.*, vol. 12, no. 1, pp. 100–110, Jan. 2021.
- [16] K. Kim, H. Park, S. Kim, Y. Kim, K. Son, D. Lho, K. Son, T. Shin, B. Sim, J. Park, S. Park, and J. Kim, "Policy-based reinforcement learning for through silicon via array design in high-bandwidth memory considering signal integrity," *IEEE Trans. Electromagn. Compat.*, 2024.
- [17] A. Kaintura, T. Dhaene, and D. Spina, "Review of polynomial chaos-based methods for uncertainty quantification in modern integrated circuits," *Electronics*, vol. 7, no. 3, p. 30, 2018.
- [18] M. A. Dolatsara, J. A. Hejase, W. D. Becker, and M. Swaminathan, "A hybrid methodology for jitter and eye estimation in high-speed serial channels using polynomial chaos surrogate models," *IEEE Access*, vol. 7, pp. 53 629–53 640, 2019.
- [19] M. E. Duman and O. Suvak, "Uncertainty quantification of CMOS active filter circuits: A non-intrusive computational approach based on generalized polynomial chaos," *IEEE Access*, vol. 8, pp. 189 246–189 261, 2020.
- [20] P. Manfredi and S. Grivet-Talocia, "Fast stochastic surrogate modeling via rational polynomial chaos expansions and principal component analysis," *IEEE Access*, vol. 9, pp. 102 732–102 745, 2021.
- [21] S. Gugliani, K. Dimple, S. Roy, R. Sharma, and B. K. Kaushik, "A bilevel multi-fidelity polynomial chaos approach for the uncertainty quantification of mwnt interconnect networks with variable imperfect contact resistances," *IEEE Access*, vol. 10, pp. 109 925–109 936, 2022.
- [22] K. Dimple, S. Gugliani, A. Dasgupta, R. Sharma, S. Roy, and B. K. Kaushik, "Modified knowledge-based neural networks using control variates for the fast uncertainty quantification of on-chip MWCNT interconnects," *IEEE Trans. Electromagn. Compat.*, vol. 65, no. 4, pp. 1232–1246, Aug. 2023.
- [23] S. Gugliani, A. K. Jakhar, K. Dimple, A. Sukhija, A. Dasgupta, R. Sharma, B. K. Kaushik, and S. Roy, "Artificial neural networks with fast transfer learning for statistical signal integrity analysis of MWCNT and MLGMR interconnect networks," *IEEE Trans. Electromagn. Compat.*, (early access).
- [24] J. A. K. Suykens, T. Van Gestel, J. De Brabanter, B. De Moor, and J. Vandewalle, *Least Squares Support Vector Machines*. Singapore: World Scientific Pub. Co., 2002.
- [25] C. K. Williams and C. E. Rasmussen, *Gaussian Processes for Machine Learning*. Cambridge, MA: MIT Press, 2006.
- [26] J. Lee, Y. Bahri, R. Novak, S. S. Schoenholz, J. Pennington, and J. Sohl-Dickstein, "Deep neural networks as Gaussian processes," *arXiv preprint arXiv:1711.00165*, 2017.
- [27] R. Garnett, *Bayesian Optimization*. Cambridge, UK: Cambridge University Press, 2023.
- [28] F. Garbuglia, D. Spina, D. Deschrijver, I. Couckuyt, and T. Dhaene, "Bayesian optimization for microwave devices using deep GP spectral surrogate models," *IEEE Trans. Microw. Theory Techn.*, vol. 71, no. 6, Jun. 2022.
- [29] D. De Witte, J. Qing, I. Couckuyt, T. Dhaene, D. Vande Ginste, and D. Spina, "A robust Bayesian optimization framework for microwave circuit design under uncertainty," *Electronics*, vol. 11, no. 14, p. 2267, 2022.
- [30] M. S. Handcock and M. L. Stein, "A Bayesian analysis of kriging," *Technometrics*, vol. 35, no. 4, pp. 403–410, 1993.
- [31] T. J. Santner, B. J. Williams, W. I. Notz, and B. J. Williams, *The Design and Analysis of Computer Experiments*. Springer, 2003, vol. 1.
- [32] P. Manfredi, "Conservative surrogate modeling of crosstalk with application to uncertainty quantification," in *Proc. IEEE 27th Workshop Signal Power Integr. (SPI)*, Aveiro, Portugal, May 2023, pp. 1–4.
- [33] J.-P. Chiles and P. Delfiner, *Geostatistics: Modeling Spatial Uncertainty*. Hoboken, NJ, USA: John Wiley & Sons, 2012, vol. 713.
- [34] R. Schobi, B. Sudret, and J. Wiart, "Polynomial-chaos-based Kriging," *Int. J. Uncertainty Quantification*, vol. 5, no. 2, 2015.
- [35] C. Lataniotis, D. Wicaksono, S. Marelli, and B. Sudret, *UQLab user manual – Kriging (Gaussian process modeling)*, 2022, report UQLab-V2.0-105.
- [36] V. Dubourg, "Adaptive surrogate models for reliability analysis and reliability-based design optimization," Ph.D. dissertation, Université Blaise Pascal-Clermont-Ferrand II, 2011.
- [37] M. Alvarez and N. Lawrence, "Sparse convolved Gaussian processes for multi-output regression," *Adv. Neural Inf. Process. Syst.*, vol. 21, 2008.
- [38] I. Bilonis and N. Zabarar, "Multi-output local Gaussian process regression: Applications to uncertainty quantification," *J. Comput. Phys.*, vol. 231, no. 17, pp. 5718–5746, 2012.
- [39] P. Moreno-Muñoz, A. Artés, and M. Alvarez, "Heterogeneous multi-output Gaussian process prediction," *Adv. Neural Inf. Process. Syst.*, vol. 31, 2018.
- [40] H. Liu, J. Cai, and Y.-S. Ong, "Remarks on multi-output Gaussian process regression," *Knowl. Based Syst.*, vol. 144, pp. 102–121, 2018.

- [41] B. Bhattacharyya, "Uncertainty quantification of dynamical systems by a POD-Kriging surrogate model," *J. Comput. Sci.*, vol. 60, p. 101602, 2022.
- [42] Q. Lu, L. Wang, and L. Li, "Efficient uncertainty quantification of stochastic problems in CFD by combination of compressed sensing and POD-Kriging," *Comput. Methods Appl. Mech. Eng.*, vol. 396, p. 115118, 2022.
- [43] R. Haylock and A. O'Hagan, "On inference for outputs of computationally expensive algorithms with uncertainty on the inputs," in *Bayesian Statistics 5*, J. Bernardo, J. Berger, A. Dawid, and A. Smith, Eds. Oxford: Oxford University Press, 1996, pp. 629–637.
- [44] D. R. Jones, M. Schonlau, and W. J. Welch, "Efficient global optimization of expensive black-box functions," *J. Global Optim.*, vol. 13, pp. 455–492, 1998.
- [45] B. D. Tracey and D. Wolpert, "Upgrading from Gaussian processes to Student's-t processes," in *Proc. 2018 AIAA Non-Deterministic Approaches Conf.*, Kissimmee, FL, USA, Jan. 2018, p. 1659.
- [46] *Statistics and Machine Learning Toolbox, Version 12.5*. Natick, MA, USA: The MathWorks Inc., 2023. [Online]. Available: <https://www.mathworks.com>
- [47] M. Ahadi and S. Roy, "Sparse linear regression (SPLINER) approach for efficient multidimensional uncertainty quantification of high-speed circuits," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 35, no. 10, pp. 1640–1652, Oct. 2016.
- [48] *HSPICE User Guide: RF Analysis, Version B-2008.09*. Mountain View, CA, USA: Synopsys, Inc., 2008. [Online]. Available: <https://www.synopsys.com/implementation-and-signoff/ams-simulation/primesim-hspice.html>
- [49] G. Blatman and B. Sudret, "Adaptive sparse polynomial chaos expansion based on least angle regression," *Journal of computational Physics*, vol. 230, no. 6, pp. 2345–2367, 2011.
- [50] S. Marelli and B. Sudret, "Uqlab: A framework for uncertainty quantification in matlab," in *Vulnerability, uncertainty, and risk: quantification, mitigation, and management*, 2014, pp. 2554–2563.
- [51] S. Grivet-Talocia and R. Trinchero, "Behavioral, parameterized, and broadband modeling of wired interconnects with internal discontinuities," *IEEE Trans. Electromagn. Compat.*, vol. 60, no. 1, pp. 77–85, Feb. 2017.
- [52] *CST Studio Suite, Version 2022.00*. Vélizy-Villacoublay, France: Dassault Systemès, 2022. [Online]. Available: <https://www.3ds.com/products-services/simulia/products/cst-studio-suite/>