A Hybrid Polynomial Chaos Expansion and Gaussian Process Regression Method for Forward Uncertainty Quantification of Integrated Circuits

A Hybrid Polynomial Chaos Expansion and Gaussian Process Regression Method for Forward Uncertainty Quantification of Integrated Circuits

Paolo Manfredi and Riccardo Trinchero
EMC Group, Department of Electronics and Telecommunications, Politecnico di Torino
Corso Duca degli Abruzzi 24, 10129 Torino, Italy
E-mail: {paolo.manfredi,riccardo.trinchero}@polito.it

***Abstract*—This paper introduces a novel kernel-based formulation for the efficient uncertainty quantification of integrated circuits. The method combines the polynomial chaos expansion (PCE) and Gaussian process regression (GPR) frameworks, the former to provide closed-form statistical information and the latter for an efficient training. In essence, the PCE coefficients are computed using a suitable Bayesian formulation that involves the definition of a special implicit kernel based on an infinite sequence of Hermite polynomials. The proposed method is illustrated based on a network with microstrip lines.**

***Index Terms*—Gaussian process regression, kernel method, least-square support-vector machine, machine learning, polynomial chaos, surrogate modeling, uncertainty quantification.**

I. Introduction

In recent years, polynomial chaos expansion (PCE) emerged as a primary tool for uncertainty quantification (UQ) in signal and power integrity [1], [2]. The main advantage of PCE lies in the closed-form statistical and sensitivity information, which is analytically derived from the model coefficients, and in its optimal convergence rate, making it more efficient than Monte Carlo (MC) for a moderate number of uncertain parameters.

Nevertheless, the burgeoning field of artificial intelligence led researches and engineers to start exploring alternative approaches based on machine learning methods due to their inherent data-driven nature. In particular, kernel-based methods such as least-square support-vector machine (LSSVM) and Gaussian process regression (GPR) exhibit attractive features due to their relatively simple form and flexible nature, which allows them to naturally scale well to a large number of uncertain parameters [1], [2], [3].

In this paper, we introduce a hybrid PCE-GPR approach that combines the advantages of both methods. With a suitable formulation that leverages a special kernel based on an infinite sequence of Hermite polynomials, the PCE coefficients are analytically derived from the GPR model via a simple and quick post-processing. Moreover, thanks to the inherent Bayesian setting of GPR, a prediction confidence is associated to the PCE, which accounts for the limited accuracy due to the finite number of training samples that is used.

II. Review of Baseline Surrogate Models

Let us consider a generic simulation problem expressed as

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

where $\mathcal{M}$ generically denotes the algorithm or simulator that allows obtaining the output of interest y for given a configuration of the input parameters $\boldsymbol{x}$. The goal of forward UQ is to statistically characterize the output y resulting from uncertainty on the inputs $\boldsymbol{x}$. We assume y to be a scalar and the inputs $\boldsymbol{x}$ to be independent and Gaussian-distributed.

A. PCE

The PCE seeks to approximate (1) using an expansion of orthogonal Hermite polynomials, i.e.,

$$y \approx \mathcal{M}_{\text{PCE}}(\boldsymbol{x}) = c_0 + \sum_{k=1}^{K} c_k H_k(\boldsymbol{x}) \tag{2}$$

where H_k are multivariate orthonormal polynomials that are built as the product of univariate Hermite polynomials.

A peculiarity of the PCE is that the first two statistical moments, the mean and the variance, are readily obtained from the coefficients as $\mu_y \approx c_0$ and $\sigma_y^2 \approx \sum_{k=1}^{K} c_k^2$, respectively. Several approaches exist for the calculation of the PCE coefficients, ranging from the accurate yet intrusive stochastic Galerkin method to non-intrusive techniques based on projection, interpolation, or regression [4] (Ch. 2).

B. LSSVM

The model can be expressed using two equivalent formulations, i.e.,

$$y \approx \mathcal{M}(\boldsymbol{x}) = b + \sum_{k=1}^{K} w_k \varphi_k(\boldsymbol{x}) = b + \sum_{l=1}^{L} \alpha_l k(\boldsymbol{x}, \boldsymbol{x}_l), \tag{3}$$

where φ_k are generic "feature-space functions" in the *primal space*, whereas $k(\cdot,\cdot)$ is a kernel function in the *dual space*. Moreover, $\{\boldsymbol{x}_l\}_{l=1}^{L}$ are a set of training points at which observations $y_l = \mathcal{M}(\boldsymbol{x}_l)$ of the actual system are collected. The two formulations are readily shown to be equivalent if the kernel function is defined explicitly as

$$k(\boldsymbol{x}, \boldsymbol{x}') = \sum_{k=1}^{K} \varphi_k(\boldsymbol{x}) \varphi_k(\boldsymbol{x}'). \tag{4}$$

The bias term b and the dual-space coefficients $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_L)^{\mathsf{T}}$ are obtained by solving the linear system

$$\begin{pmatrix} \boldsymbol{K} + \gamma^{-1}\boldsymbol{I} & \boldsymbol{1} \\ \boldsymbol{1}^{\mathsf{T}} & 0 \end{pmatrix} \begin{pmatrix} \boldsymbol{\alpha} \\ b \end{pmatrix} = \begin{pmatrix} \boldsymbol{y} \\ 0 \end{pmatrix}, \tag{5}$$

979-8-3503-5123-1/24/$31.00 ©2024 IEEE

where $\boldsymbol{K}$ is the $L \times L$ kernel matrix of the training points, with entries $K_{lm} = k(\boldsymbol{x}_l, \boldsymbol{x}_m)$ for $l, m = 1, \ldots, L$, $\boldsymbol{I}$ is the identity matrix of the same size, $\mathbf{1}$ is a column vector of ones, $\boldsymbol{y} = (y_1, \ldots, y_L)^\mathsf{T}$ is the vector of the training observations, and γ is a regularization hyperparameter.

The advantage of the LSSVM formulation is that the coefficients are computed by means of L observations only, regardless of the dimensionality of $\boldsymbol{x}$ and the number of primal-space functions K. Furthermore, the dual-space formulation is also valid for implicit (e.g., radial basis function or squared-exponential) kernels, which are more efficient to evaluate even in high-dimensional settings. However, in that case, an explicit primal-space formulation is lost due to the unavailability of the corresponding feature-space functions φ_k.

It was recently shown that the PCE model (2) can be cast as a LSSVM problem (3) if Mehler kernel is used [5]. The univariate kernel reads

$$k(x, x') = \sum_{k=0}^{\infty} \rho^k H_k(x) H_k(x') = \frac{e^{\frac{-\rho^2(x^2+x'^2)-2xx'}{2(1-\rho^2)}}}{\sqrt{1-\rho^2}} \tag{6}$$

and corresponds to the sum of an (infinite!) sequence of (univariate) Hermite polynomials. Comparing with (4) shows that the Hermite polynomials, rescaled by a factor $\sqrt{\rho^k}$, form the feature-space functions for this kernel.

The multivariate kernel is build as the product of univariate kernels and is further equipped with a variance σ^2, i.e.,

$$k(\boldsymbol{x}, \boldsymbol{x}') = \sigma^2 \prod_{j=1}^{d} k(x_j, x'_j). \tag{7}$$

Apart from trivial rescaling factors, this makes the corresponding feature-space functions generalize to the multivariate Hermite polynomials as in (2). Since the feature-space functions for this kernel are known to be the Hermite polynomials, the primal-space LSSVM coefficients w_k are readily obtained from the dual-space coefficients α_l and correspond to the PCE coefficients c_k (virtually, up to an infinite order!), whereas the bias term b is equivalent to c_0.

C. GPR

The model assumes that (1) be one specific realization of a prior Gaussian process $\mathcal{GP}\left(\mu(\boldsymbol{x}), k(\boldsymbol{x}, \boldsymbol{x}')\right)$, where the function $k(\cdot, \cdot)$ now plays the role of prior covariance.

Starting from a collection of training observations $\{(\boldsymbol{x}_l, y_l)\}_{l=1}^{L}$, similarly as for the LSSVM, posterior predictions are computed as

$$y \approx \mathcal{M}(\boldsymbol{x}) = \mu(\boldsymbol{x}) + \sum_{l,m=1}^{L} B_{lm} \left(y_m - \mu(\boldsymbol{x}_m)\right) k(\boldsymbol{x}, \boldsymbol{x}_l) \tag{8}$$

where B_{lm} are the elements of $\boldsymbol{B} = \boldsymbol{K}^{-1}$, the inverse of the kernel matrix with the same definition as in (5).

Notably, a (co)variance

$$c(\boldsymbol{x}, \boldsymbol{x}') = k(\boldsymbol{x}, \boldsymbol{x}') - \sum_{l,m=1}^{K} k(\boldsymbol{x}, \boldsymbol{x}_l) B_{lm} k(\boldsymbol{x}, \boldsymbol{x}_m), \tag{9}$$

is also associated to the model predictions, which reflects their uncertainty due to the limited amount of observations. This information can be used to obtain confidence bounds on pertinent statistical information when the GPR model is used as a surrogate in a MC analysis [3]. However, the calculation becomes prohibitive if the number of MC trials is large, due to the large size of the resulting covariance matrix.

III. Proposed Hybrid PCE-GPR Method

It is readily shown that the GPR and the dual-space LSSVM formulations are in fact equivalent if a constant prior mean function $\mu(\boldsymbol{x}) = b$ is assumed and the same LSSVM kernel is taken as the GPR prior covariance function. With the above settings, the proof is found by comparing (8) and the dual-space formulation in (3), and noting that the coefficients

$$\alpha_l = \sum_{m=1}^{L} B_{lm}(y_m - b) \tag{10}$$

are equivalent to the solution of (5) if the term $\gamma^{-1}\boldsymbol{I}$ is embedded in the kernel matrix $\boldsymbol{K}$. In the GPR settings, this is equivalent to assuming a Gaussian noise on the training data with standard deviation $\sigma_n = \sqrt{\gamma^{-1}}$ [3]. For the sake of simplicity, in the following we assume a noise-free interpolation, which is a standard scenario if no noise is expected on the observations and is equivalent to letting $\gamma \to \infty$ in (5).

In summary, based on the above considerations, a GPR model with constant trend that uses the Mehler kernel (6) as prior covariance is equivalent to a dual-space LSSVM formulation, whose primal-space formulation is in turn equivalent to a Hermite PCE. This Bayesian interpretation allows associating a predictive covariance in the form of (9), from which the covariance matrix of the PCE coefficients is analytically derived, expressing the estimation uncertainty resulting from using a limited amount of observations. Owing to the lack of space, the theoretical details are deferred to a future report.

At present, an analogous kernel is available for Legendre polynomials and uniform variability [5]. As in standard PCE, dealing with Gaussian correlation is straightforward and is readily achieved by preliminarily decorrelating the input variables. Suitable kernels should be instead identified for other distributions, including non-Gaussian correlated ones.

IV. Application Example and Numerical Results

The proposed hybrid PCE-GPR method is applied to the UQ of the transmitted voltage in the network of Fig. 1, which was one of the test cases in [4] (Ch. 3). The uncertainty is in the thickness, relative permittivity, and loss tangent of the substrate. These parameters are Gaussian-distributed with nominal values of $\varepsilon_r = 4.1$, $\tan\delta = 0.02$, and $h = 100\,\mu\text{m}$ and a standard deviation of 10%. We refer to [4] for additional details. For the sake of simplicity, we restrict the analysis to the transmitted voltage at a specific time point, i.e., $t = 1.46$ ns, corresponding to the location of the maximum overshoot.

We compare the advocated method against the following state-of-the-art techniques [4]: a) stochastic Galerkin method

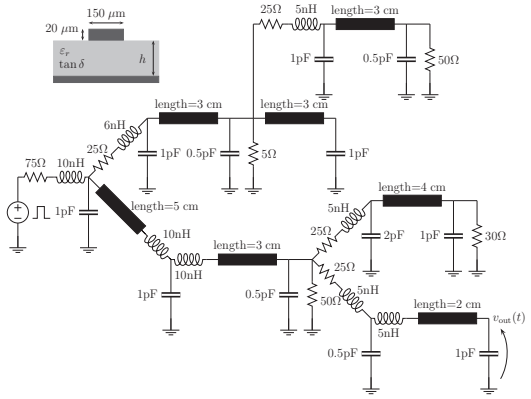

Fig. 1. Schematic of the considered test case.

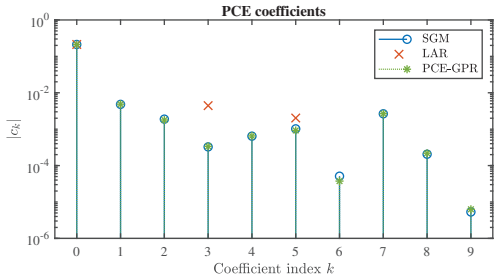

Fig. 2. PCE coefficients of the transmitted voltage at 1.46 ns (log scale). Circles (○): SGM; crosses (×): LAR; stars (∗): proposed PCE-GPR.

(SGM): requires a single deterministic simulation of a $(K+1)$-times augmented counterpart of the original stochastic circuit; b) stochastic testing (ST): solves the stochastic circuit at $K+1$ specific configurations of the uncertain parameters and interpolates the resulting responses; c) least-angle regression (LAR): collects observations for some configurations of the uncertain parameters and performs a sparse regression, here implemented using UQLab [6]; d) plain GPR: baseline formulation considering a squared-exponential kernel. Furthermore, reference results are generated based on a MC simulation with 10'000 runs. For the PCE-based techniques, including the proposed hybrid method, we consider a second-order expansion as in [4], which features $K+1=10$ terms. However, it is important to note that with PCE-GPR the expansion order does not need to be predetermined, as opposed to classical PCE-based methods. The 10 simulations of the ST method are also used to train the data-driven methods, i.e., LAR and the two GPR-based techniques.

Figure 2 compares the magnitude of the PCE coefficients obtained with the SGM method against the coefficients estimated using LAR and the proposed PCE-GPR method. Despite using exactly the same training data, LAR is able to identify only three coefficients (i.e., c_0, c_3, and c_6), two of which with rather low accuracy. Conversely, the proposed method obtains a very accurate estimate.

Table I collects the relevant figures related to the mean and the variance, as well as the root-mean-square error (RMSE) obtained with the various methods. For the GPR-based techniques, the moments are expressed in terms of expected value

TABLE I
MEAN, VARIANCE, AND RMSE OBTAINED WITH THE VARIOUS METHODS.

	Mean (V)	Variance ($V^2 \times 10^5$)	RMSE (V)
MC	0.2120	3.5628	–
SGM	0.2120	3.5234	4.4809×10^{-4}
ST	0.2120	3.5546	4.7095×10^{-4}
LAR	0.2122	2.3751	2.4443×10^{-3}
Plain GPR	0.2122	2.7771	1.7566×10^{-3}
2-sigma interval	$[\underline{0.2120}, \overline{0.2124}]$	$[\underline{2.2139}, \overline{3.3402}]$	–
Hybrid PCE-GPR	0.2120	3.5377	4.7845×10^{-4}
2-sigma interval	$[\underline{0.2120}, \overline{0.2120}]$	$[\underline{3.2843}, \overline{3.8181}]$	–

and 2-sigma confidence interval. First of all, it is observed that the SGM and ST provide very accurate results, whereas the LAR exhibits a larger error. The performance of the standard GPR is similar to LAR, whereas the proposed hybrid method achieves a better accuracy that is comparable to the one of the SGM and ST. However, PCE-GPR is expected to scale better to high-dimensional problems. Indeed, GPR methods were demonstrated to achieve accurate results with a moderate number of training samples (generally, in the order of a few hundreds) even in high-dimensional settings [3], which also makes the required factorization of matrix $\boldsymbol{K}$ computationally tractable. The confidence intervals of PCE-GPR are also narrower and enclose the reference MC result.

V. CONCLUSIONS

This paper introduced a hybrid approach combining the advantages of both PCE and kernel methods. The model uses a GPR formulation and a special kernel corresponding to an infinite sequence of Hermite polynomials. Thanks to the inherent Bayesian formulation, the PCE coefficients are analytically calculated with the inclusion of confidence bounds, which allow assigning a prediction confidence to statistical metrics. This information could be used to drive the acquisition of additional training samples in an active learning scenario. The method was illustrated based on the simulation of a transmission-line network, for which a very good accuracy, comparable to the state-of-the-art SGM, was obtained.

REFERENCES

[1] R. Trinchero *et al.*, "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.

[2] T. Nguyen *et al.*, "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] 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. 2021.

[4] S. Roy, *Uncertainty Quantification of Electromagnetic Devices, Circuits, and Systems*. Stevenage, United Kingdom: IET, 2021.

[5] P. Manfredi and R. Trinchero, "Nonparametric formulation of polynomial chaos expansion based on least-square support-vector machines," *Eng. Appl. Artif. Intell.*, vol. 133, p. 108182, 2024.

[6] 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.