Statistical crosstalk analysis via probabilistic machine learning surrogates

Statistical Crosstalk Analysis via Probabilistic Machine Learning Surrogates

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 discusses the application of a probabilistic surrogate modeling technique, based on Gaussian process regression (GPR), to the uncertainty quantification (UQ) of crosstalk. Compared to traditional deterministic surrogate models, the GPR provides a stochastic process that carries an estimate of the model uncertainty. This allows assigning confidence bounds to the model prediction and, in an UQ scenario, to statistical estimates. The advocated method is illustrated through its application to a literature test case.**

***Index Terms*—Crosstalk, Gaussian process regression, machine learning, surrogate modeling, uncertainty quantification.**

I. Introduction

In the past few years, the increasing impact of manufacturing variability on the performance of large-scale integration circuits has drawn an enormous attention on techniques for uncertainty quantification (UQ). While polynomial chaos expansion emerged as a robust and powerful method [1]–[3], more recent works investigated the use of machine learning methods for UQ tasks [4], [5].

An attractive feature of machine learning techniques is that they are nonparametric. This means that the model complexity is mainly determined by the available training data, rather than by the problem dimensionality. Therefore, compared to conventional parametric approaches like the ones based on polynomial chaos expansion [1], it scales more favorably with the number of random variables. While most methods, such as those based on plain neural networks or support vector machine regression, provide deterministic surrogates, Gaussian process regression (GPR) models inherently carry an estimate of the model uncertainty [6]. In an UQ scenario, this information can be suitably propagated to obtain a probabilistic estimate of the statistics of interest, such as mean value, variance, and probability density function (PDF), with the inclusion of confidence levels. This paper discusses the application of probabilistic GPR models to the UQ of crosstalk. Moreover, analytical estimates are provided for the prediction of the mean and the variance, and for their confidence bounds.

II. Problem Statement

Let us consider a generic system in the form of

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

where $\boldsymbol{x} \in \mathbb{R}^d$ is a set of uncertain design parameters, y is a scalar output quantity of interest, and $\mathcal{M} : \mathbb{R}^d \to \mathbb{R}$ denotes a function (the "full computational model") that allows computing y for a given configuration of the parameters $\boldsymbol{x}$. Extension to multiple outputs will be discussed in an expanded paper.

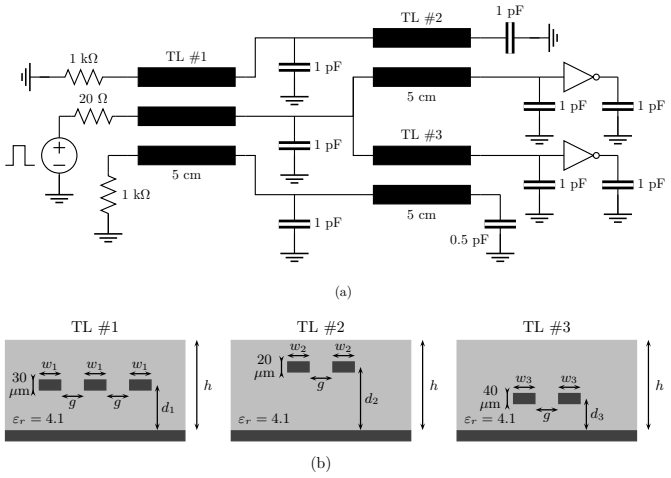

Fig. 1. Interconnect network for the considered application example, reproduced from [3]: (a) network topology; (b) cross-section of the transmission line sections.

For illustration purposes, in this paper we analyze the high-speed interconnect depicted in Fig. 1, which was investigated in [3]. We consider two test cases, summarized in Table I. In Test Case #1, there are $d = 2$ uncertain parameters, namely the substrate thickness h and the gap g between the traces. In Test Case #2, there are $d = 6$ uncertain parameters, i.e., the widths w_n and the vertical positions d_n (with $n = 1, 2, 3$) of the traces in each of the three subnetworks. All parameters are assumed to be Gaussian distributed with a 10% relative standard deviation from the nominal value. The output of interest is the maximum crosstalk over time produced at the 0.5-pF termination by the transmission of a voltage pulse with an amplitude of 5 V, rise and fall times of 100 ps, and duration of 1 ns. The full computational model (1) is in this case an HSPICE transient simulation of the network, performed with a maximum step size of 1 ps.

III. Probabilistic GPR Model

As opposed to most surrogates, a GPR model is a stochastic process rather than a deterministic function [6]. Specifically,

978-1-6654-4269-5/21/$31.00 ©2021 IEEE

TABLE I
DESCRIPTION OF THE TWO TEST CASES.

Test case	d	Parameter	Nominal value	Standard deviation
#1	2	h	200 μm	10% Gaussian
		g	150 μm	
#2	6	w_1	150 μm	10% Gaussian
		w_2	130 μm	
		w_3	170 μm	
		d_1	100 μm	
		d_2	140 μm	
		d_3	70 μm	

GPR seeks for y a *probabilistic* model in the form of

$$y \approx \hat{\mathcal{M}}(\boldsymbol{x}) \sim \mathcal{GP}(m(\boldsymbol{x}), c(\boldsymbol{x}, \boldsymbol{x}')), \tag{2}$$

meaning that the surrogate $\hat{\mathcal{M}}$ is a Gaussian process with mean function $m(\boldsymbol{x})$ and covariance function $c(\boldsymbol{x}, \boldsymbol{x}')$. Hence, for a given value of the input parameters $\boldsymbol{x}$, the output prediction is a Gaussian random variable with mean $m(\boldsymbol{x})$ and standard deviation $\sqrt{c(\boldsymbol{x}, \boldsymbol{x})}$. This fact can be exploited to obtain probabilistic predictions of statistical moments and PDFs. In order to do so, however, it is important to correctly take into account the correlation in the output distribution described by the covariance function.

The process mean and covariance are found by choosing a prior Gaussian process with a given mean $\mu(\boldsymbol{x})$ and covariance $k(\boldsymbol{x}, \boldsymbol{x}')$, also called "kernel", and by conditioning it to interpolate a set of available training observations $\{(\boldsymbol{x}_l^\dagger, y_l^\dagger)\}_{l=1}^L$, with $y_l^\dagger = \mathcal{M}(\boldsymbol{x}_l^\dagger)$, calculated using the full computational model (1). The prior mean could be any function, including another surrogate model. However, taking $\mu(\boldsymbol{x}) = 0$ is a common option that we shall adopt in this paper. Exponential, squared exponential, and Matérn functions are popular choices for the kernel. Without loss of generality, we shall use a Matérn 5/2 kernel, i.e.,

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

where $r = \sqrt{(\boldsymbol{x} - \boldsymbol{x}')^\mathsf{T}(\boldsymbol{x} - \boldsymbol{x}')}$ is the Euclidean distance between $\boldsymbol{x}$ and $\boldsymbol{x}'$, whereas σ_f and σ_l are two *hyperparameters* that are estimated when training the model.

With the above definitions, it is possible to sample (2) with a finite number of samples $\{\boldsymbol{x}_i^*\}_{i=1}^N$ of the uncertain input parameters, as is typically done when performing UQ using Monte Carlo (MC) analysis in conjunction with surrogate models. However, the substantial difference here is that the result is not an ensemble of deterministic predictions, but rather a set of N correlated Gaussian random variables $\boldsymbol{y}^* \sim \mathcal{N}(\boldsymbol{\mu}_y, \boldsymbol{\Sigma}_y)$. The random variables are characterized by the mean vector

$$\boldsymbol{\mu}_y = \boldsymbol{K}_* \boldsymbol{K}^{-1} \boldsymbol{y}^\dagger \tag{4}$$

and covariance matrix

$$\boldsymbol{\Sigma}_y = \boldsymbol{K}_{**} - \boldsymbol{K}_* \boldsymbol{K}^{-1} \boldsymbol{K}_*^\mathsf{T}, \tag{5}$$

where

- $\boldsymbol{y}^\dagger \in \mathbb{R}^L$ is a column vector with entries $y_l^\dagger = \mathcal{M}(\boldsymbol{x}_l^\dagger)$;
- $\boldsymbol{K} \in \mathbb{R}^{L \times L}$ is a matrix with entries $K_{lm} = k(\boldsymbol{x}_l^\dagger, \boldsymbol{x}_m^\dagger)$;
- $\boldsymbol{K}_* \in \mathbb{R}^{N \times L}$ is a matrix with entries $K_{il} = k(\boldsymbol{x}_i^*, \boldsymbol{x}_l^\dagger)$;
- $\boldsymbol{K}_{**} \in \mathbb{R}^{N \times N}$ is a matrix with entries $K_{ij} = k(\boldsymbol{x}_i^*, \boldsymbol{x}_j^*)$;

A random draw of $\boldsymbol{y}^*$ yields one possible prediction (realization) of the MC samples. An ensemble of predictions allows one to estimate statistical moments and PDFs in a probabilistic way, e.g., in terms of expectation and confidence bounds. Moreover, lengthy derivations lead to analytical expressions for the expectation of the mean

$$\bar{\mu}_y = \tfrac{1}{N} \textstyle\sum_{i=1}^N \mu_{y,i} \tag{6}$$

and the variance

$$\bar{\sigma}_y^2 = \tfrac{1}{N-1} \left[\textstyle\sum_{i=1}^N (\mu_{y,i} - \bar{\mu}_y)^2 + \mathrm{tr}(\boldsymbol{\Sigma}_{\boldsymbol{y}})\right] \tag{7}$$

of y, as well as for their standard deviation:

$$\mathrm{Std}\{\bar{\mu}_y\} = \tfrac{1}{N} \sqrt{\textstyle\sum_{i=1}^N \sum_{j=1}^N (\boldsymbol{\Sigma}_y)_{ij}} \tag{8}$$

$$\mathrm{Std}\{\bar{\sigma}_y^2\} = \tfrac{2}{N-1} \sqrt{(\boldsymbol{\mu}_y - \bar{\mu}_y)^\mathsf{T} \boldsymbol{\Sigma}_y (\boldsymbol{\mu}_y - \bar{\mu}_y) + \tfrac{1}{2}\mathrm{tr}(\boldsymbol{\Sigma}_y^2)}. \tag{9}$$

The derivation of the above expressions is deferred to an expanded paper.

IV. NUMERICAL RESULTS

The proposed probabilistic method is applied to the interconnect of Fig. 1 and the two test cases described in Table I. Reference results for the maximum crosstalk are generated with a MC analysis using 5000 samples of the uncertain parameters, drawn according to a Latin hypercube sampling (LHS) scheme. With HSPICE, this analysis takes about 1 h 25 min for each test case.

TABLE II
GPR PREDICTION OF THE MEAN AND VARIANCE OF THE MAXIMUM CROSSTALK.

	Moment	MC	GPR
Test Case #1	Mean (V)	0.9537	$[\underline{0.9449}, \mathbf{0.9533}, \overline{0.9617}]$
	Variance (V^2)	0.0223	$[\underline{0.0185}, \mathbf{0.0227}, \overline{0.0269}]$
Test Case #2	Mean (V)	0.9413	$[\underline{0.9368}, \mathbf{0.9402}, \overline{0.9437}]$
	Variance (V^2)	0.0202	$[\underline{0.0193}, \mathbf{0.0205}, \overline{0.0218}]$

GPR models are trained by considering $L = 10 \cdot d$ samples, also drawn by means of LHS, leading to $L = 20$ for Test Case #1 and $L = 60$ for Test Case #2. Table II reports the mean and the variance of the maximum crosstalk computed from the MC samples, as well as the ones predicted with the GPR models based on the same set of parameter samples. The expected value of the GPR prediction, obtained with (6) and (7), is highlighted with bold font, whereas the upper and lower 95% confidence bounds, corresponding to ± 1.96 times the standard deviation computed with (8) and (9), are denoted with overlines and underlines, respectively. It is noted

that the expected value of the prediction is always very clos to the reference MC value, with errors below 0.2% for tl mean and 1.8% for the variance. The latter corresponds an error below 0.9% for the standard deviation. The accurac of the prediction is further confirmed by the tightness of tl confidence intervals, which include the reference result ar amount to less than 0.9% of the expected value for the mea and 19% for the variance. The information on the mod accuracy is summarized in Table III. Experiments demonstra that the confidence interval can be reduced by increasing tl number of training samples and/or choosing a more compl prior kernel (results not shown here due to the lack of space

TABLE III
ACCURACY OF THE GPR PREDICTIONS.

	Moment	Prediction error	Confidence tightness
Test Case #1	Mean	$< 0.1\%$	0.9%
	Variance	1.8%	18.5%
Test Case #2	Mean	0.1%	0.4%
	Variance	1.5%	6.3%

The GPR models are further used to provide a probabilistic estimate of the PDF of the maximum crosstalk, based on 5000 realizations of the posterior random variable $\boldsymbol{y}^*$. From each realization, a different prediction of the PDF is obtained. Figure 2 shows, with a blue line, the distribution of the reference MC samples. The 67%, 95%, and 99% confidence intervals of the GPR prediction are shown instead with different shades of red. It is noted that the reference MC distribution is well within the confidence bounds, and mostly within the tightest 67% interval. Moreover, the confidence bounds for Test Case #2 are tighter, consistently with the higher accuracy observed in Table III.

TABLE IV
EFFICIENCY OF THE GPR MODELS.

	Model training		Model	Total	Speed-up
	Samples	Building	evaluation	time	w.r.t. MC
Test Case #1	27.3 s	0.5 s	52.3 s	80.1 s	63.7×
Test Case #2	80.3 s	0.1 s	52.7 s	133.1 s	38.3×

Finally, Table IV provides the relevant figures concerning the efficiency of the proposed method. The computational time consists of two components: the model training and the model evaluation. The former can be further broken down into the time required by the simulation of the training samples and by the model building. Of these three contributions, the first one is typically dominant, especially when the simulation of the full computational model is expensive. As opposed to parametric models, the second contribution is only determined by the number of available training samples, regardless of the model complexity in terms of number of uncertain parameters, and it is therefore usually negligible. The third contribution is related

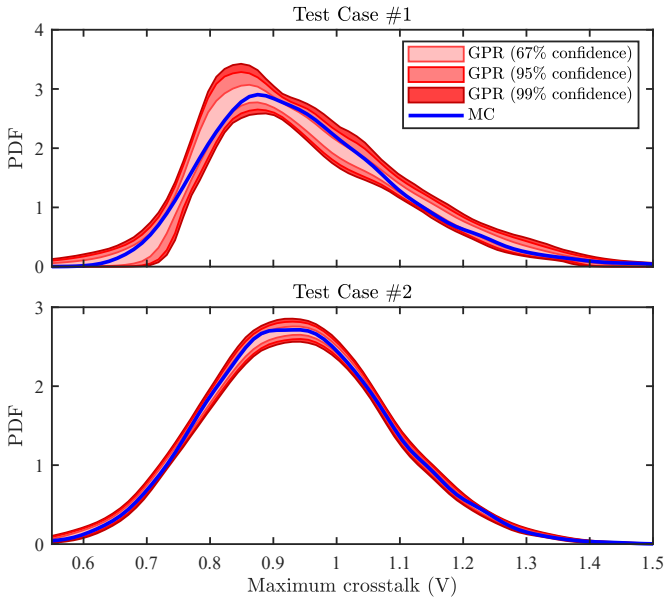

Fig. 2. PDF of the maximum crosstalk for the two test cases. Blue line: reference distribution of the MC samples; shades of red: confidence levels of the GPR prediction.

to the post-processing for the extraction of the statistical information. In the considered application, it is dominated by the calculation of the posterior covariance matrix and of the kernel density estimates of the PDF for the probabilistic assessment of the crosstalk distribution reported in Fig. 2. All in all, the speed-up w.r.t. the MC analysis is remarkable.

V. CONCLUSIONS

This paper introduced a probabilistic surrogate model for UQ based on GPR. The proposed method allows calculating not only predictions of relevant statistical information, but also an estimate of their confidence bounds. The technique was illustrated via the UQ of the maximum crosstalk in an interconnect network.

REFERENCES

[1] M. R. Rufuie, E. Gad, M. Nakhla, and R. Achar, "Generalized Hermite polynomial chaos for variability analysis of macromodels embedded in nonlinear circuits," *IEEE Trans. Compon. Packag. Manuf. Technol.*, vol. 4, no. 4, pp. 673–684, Apr. 2014.

[2] J. B. Preibisch, T. Reuschel, K. Scharff, and C. Schuster, "Impact of continuous time linear equalizer variability on eye opening of high-speed links," in *Proc. 2016 IEEE 20th Workshop on Signal and Power Integrity (SPI)*, Turin, Italy, May 2016, pp. 1-4.

[3] M. Ahadi and S. Roy, "Sparse linear regression (SPLINER) approach for efficient multidimensional uncertainty quantification of high-speed circuits," *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst*, vol. 35, no. 10, pp. 1640–1652, Oct. 2016.

[4] 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, 2019.

[5] X. Ma, M. Raginsky, and A. C. Cangellaris, "A machine learning methodology for inferring network S-parameters in the presence of variability," in *Proc. 2018 IEEE 22nd Workshop on Signal and Power Integrity (SPI)*, Brest, France, May 2018, pp. 1-4.

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