Modeling S-parameters of Interconnects using Periodic Gaussian Process Kernels

Federico Garbuglia
IDLab, Ghent University-imec
Ghent, Belgium
federico.garbuglia@ugent.be

Domenico Spina
IDLab,
Ghent University-imec
Ghent, Belgium

Torsten Reuschel
University of New Brunswick
Fredericton, NB, Canada

Christian Schuster
Hamburg University of Technology
Hamburg, Germany

Dirk Deschrijver
IDLab, Ghent University-imec
Ghent, Belgium

Tom Dhaene
IDLab, Ghent University-imec
Ghent, Belgium

2023 IEEE 27th Workshop on Signal and Power Integrity (SPI) | 979-8-3503-3282-7/23/$31.00 ©2023 IEEE | DOI: 10.1109/SPI57109.2023.10145548

***Abstract*—In this paper, we present a novel technique to model wide-band scattering parameter (S-parameter) curves of high-speed digital interconnects. The proposed technique utilizes a new kernel function with periodic components for Gaussian process (GP) models. After proper training, the GP models are able to predict the S-parameter values at arbitrary frequency points inside the trained interval. The performance of the proposed technique is reviewed by means of correlation with standard Gaussian Processes with squared exponential kernel and Matern kernel. Results for the proposed technique show an increased prediction accuracy when applied to interconnects.**

***Index Terms*—Interconnects, S-parameters, machine learning (ML), Gaussian processes (GP), kernels.**

I. Introduction

Macromodeling is a fundamental tool for the characterization of high frequency interconnects and has been used for several applications, including the efficiency increase of general purpose electromagnetic full-wave simulations, modeling lossy transmission lines, and design optimization [1]. Many popular macromodeling techniques exploit Vector Fitting (VF) [1] to build a rational function approximation for interconnect transfer functions. Recently, several machine learning methods have been proposed in literature to try to overcome the limitations of standard macromodeling approaches, when describing complex high frequency systems with a large number of ports and design parameters [2].

However, powerful ML models, such as neural networks or support vector machines [3], can be prone to overfitting. They may predict complicated patterns that are absent in the function (or stochastic process) underlying the data. Conversely, stochastic ML models such as Gaussian Processes (GP) [4] are less inclined to overfitting, thanks to their self-regularizing capability, while they present relatively high data-efficiency.

Furthermore, GPs have been successfully employed to optimize performance metrics of microwave devices, using Bayesian Learning [5], [6]. For a given DUT, they can rapidly identify the optimal values of performance metrics, which are modelled as a function of design variables. Hence, GPs appear to be generally suitable for building S-parameter models.

Unfortunately, deriving models for frequency-dependent data, such as S-parameters, over wide ranges is particularly challenging using GPs. The samples may resemble highly non-smooth and oscillating behavior, while standard GP configurations usually assume high smoothness of curves. In fact, this property of the GP derives from the typical covariance function used among the data points, also known as *kernels* [4].

In this paper, a new kernel is proposed to improve the GP accuracy on S-parameter curves across a wide frequency range. The new kernel is referred to as "periodic" for simplicity. It allows the GP to fit S-parameter curves even in the presence of ample oscillations, which may be a result of resonances or crosstalk. The usefulness of the obtained GP is demonstrated by the use of an example interconnect, and is compared to a GP model that uses a common, general-purpose kernel.

II. Methodology

A. Gaussian process modeling

In this study, the GP is applied to model interconnect S-parameter data across frequency. The GP is a non-parametric model: its computational complexity does not depend on a fixed number of trainable parameters, but it grows polynomially with the amount of training data samples. Therefore, an accurate GP can be built using a relatively low amount of data points and is suitable to represent low-dimensional functions or stochastic processes.

Moreover, the GP is a stochastic model: it represents each available data point as a realization of Gaussian probability density (prior distribution) with specified mean and variance, while the correlation between pairs of points is given by the kernel function. Most notably, the kernel has to be manually specified. This may include assumptions or prior knowledge of the function or stochastic process to be modeled, such as

This research received funding from the Flemish Government via the AI Research Programme and the Fonds Wetenschappelijk Onderzoek (FWO) programme.

smoothness, discontinuities, stationarity, or noise distribution on the data samples. If the kernel is properly designed to incorporate correct assumptions, the GP will achieve high prediction accuracy.

After specifying the kernel, the value of new data samples can be predicted via Gaussian Process regression (GPR) [4], also known as Kriging. For any new sample, the GPR provides the posterior distribution, i.e. the probability density for each new data sample. The mean of the obtained posterior represents the expected value, while the variance indicates the epistemic uncertainty on the value. Hence, the GP is able to provide a predicted value and a confidence interval for any new data sample. A comprehensive explanation on the functioning of the GPR is available in [4].

B. Periodic Gaussian process

As stated in the Introduction, one main limitation of characterizing high-speed digital interconnects with GPs is that GPs assume a relatively smooth behavior of the quantity to be modeled with respect to the considered parameters. However, this does not necessarily correspond to the reality of state-of-the-art interconnects. The wide-band S-parameters of such structures can exhibit a dynamic and oscillatory behavior. A potential solution for coping with this, is to adopt a type of GP with higher modeling power, like the Deep Gaussian Process (DGP) [6]. However, this significantly increases computational complexity and modeling time. Instead, here we propose a novel kernel to model the dynamic behavior of wide-band interconnects. The proposed kernel greatly improves GP's modeling accuracy at the cost of a limited increase in the number of hyperparameters to be learned during the training phase with regard to standard kernels.

In the following, it is applied to the real or imaginary part of one S-parameter curve, denoted as $S(f)$, across the frequency variable f. However, without loss of generality, the proposed kernel can also be used to model the magnitude and the phase of $S(f)$. The analytical expression for the new kernel function is k_{per} defined as

$$k_{\text{per}} = k_{\text{mean}} + k_{\text{envelope}} \cdot \prod_{m=1}^{N_p} k_{p,m} \tag{1}$$

where N_p is the number of periods. The three terms of the new kernel are defined as follows:

$$k_{\text{mean}}(f, f') = \exp\left(-\frac{(f-f')^2}{2l_1^2}\right) \tag{2}$$

$$k_{\text{envelope}}(f, f') = \exp\left(-\frac{(f-f')^2}{2l_2^2}\right) \tag{3}$$

$$k_{p,m}(f, f') = \exp\left(-2\frac{\sin^2\left(\pi\frac{1}{p_m}(f-f')\right)}{l_3^2}\right) \tag{4}$$

where the terms k_{mean} and k_{envelope} are standard squared exponential (SE) kernel functions, while $k_{p,m}$ is a periodic kernel of period p_m [7]. In other research areas, periodic kernels have been used to model functions that present visible oscillatory behaviours, such as seasonal time-series [7], or that are defined on periodic domains [8]; here we investigate applicability of periodic kernels to characterize high frequency interconnects. The quantities l_1, l_2 and l_3 are trainable length-scale hyperparameters: intuitively, they represent the maximum distance for which the correlation between two frequency samples is non-negligible. Note that all the kernel terms are stationary, since they depend only on the distance between two frequency samples f and f'. Furthermore, k_{per} is also stationary, since it is a linear combination of stationary functions.

The role of each term of the proposed kernel can be observed on a typical S-parameter curve of a microwave device. For example, Fig. 1 illustrates the near-end crosstalk between two adjacent ports of a 44-ports high-speed interconnect [9]. Considering the high oscillatory behaviour, the following assumptions can be made, either on the real or imaginary part of the S-parameters.

Assumption 1: The mean values of the oscillations are continuous and highly correlated between adjacent frequency samples, and weakly correlated otherwise. Therefore, the correlation among the mean values is represented by the decaying exponential in k_{mean}.

Assumption 2: Frequency samples separated by integer multiples of the oscillations' semi-period are highly correlated. This information is encoded in the kernel by the $\sin^2(\cdot)$ operator in $k_{p,m}$.

Assumption 3: The oscillations may have a non-constant amplitude. Hence, the term k_{envelope} is included to represent the envelope of the oscillations: the amplitude values are highly correlated between two adjacent frequency samples.

Furthermore, if oscillations of different periods can be observed in the S-parameter curve, multiple periodic terms $k_{p,m}$ can be combined into a single product, as shown in equation (1). The main advantage of the new kernel k_{per} is that it does not require the exact knowledge of the oscillation periods p_m. In fact, p_m values can be set as trainable hyperparameters for the GPR. Nonetheless, if prior knowledge on the

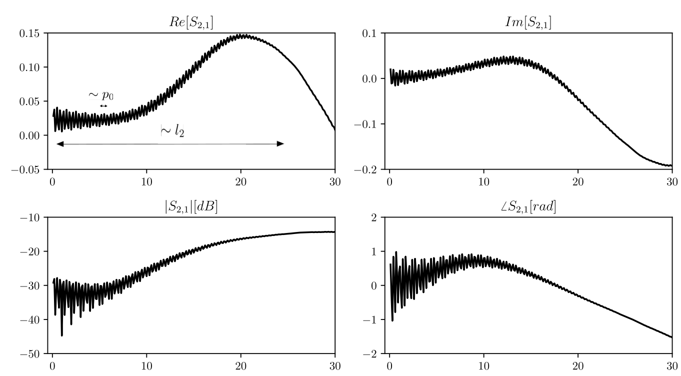

Fig. 1. Real/imaginary (top) and magnitude/phase (bottom) representation of the near-end crosstalk (NEXT), $S_{2,1}$, between interconnects available from [9]. The depicted S-parameters are analyzed in Section III.

period(s) is available, it can be integrated in the tuning process. In fact, it is possible to select the optimal period from a specific interval $p_m \in [p_L, p_H]$, where p_L, p_H are arbitrary bounds. Alternatively, if a prior probability distribution of period values is available, the GP regression provides an estimated posterior distribution from which the optimal period can be sampled. For example, by observing the data in Fig. 1, the period length p_0 can be constrained to values in the range of $[0.2, 0.4]$ GHz. In addition, the envelope length-scale l_2 can be assumed to be larger than a 20 GHz, since oscillations seems to attenuate in a larger frequency window. Thus, l_2 can be drawn from a prior gamma distribution with rate and scale parameter equal to 5: $l_2 \sim \Gamma(5, 5)$. Subsequently, the optimal value for each hyperparameter is extracted during the GP training, using a gradient ascend algorithm for likelihood maximization [10]. It is worth noting that the optimal solution for hyperparameter values may not be unique and can be sensitive to the available training data.

The main disadvantage of the proposed kernel k_{per} is that the number of periods N_p can heavily affect the performance of the GP. In fact, if the number of periods is under-estimated, the GP is more likely to underfit the available data samples, leading to reduced accuracy. Conversely, if the number of periods is over-estimated, GP may exhibit numerical instabilities due to the excessive amount of hyperparameters. However, the latter problem is minor in practice, because of the low computational cost of the GP regression. If in doubt, the GP training may be repeated for different numbers of periods until a sufficient accuracy is reached with regard to the validation data set.

III. Application Example

The GP model is tested on previously validated interconnect S-parameters that were extracted using physics-based modeling [9]. While several routing lengths, topologies, and dielectric materials have been investigated, the following discussion focuses on a single-board interconnect with a routing length of 10 in. The interconnect contains 22 differential signal pairs and occupies the outer nine layers of a board with 40 layers and an effective dielectric $\varepsilon_r = 3.7$, $\tan\delta = 0.01$. The signals are routed through via pin fields with a pitch of 60 mil, which are the main cause of crosstalk between lines of the interconnect. The single-ended S-parameter matrix is available in the frequency range from 0.05 GHz to 50 GHz in steps of 0.05 GHz, for a total of 1000 frequency samples. The samples are equally split into two sets: the first set is used to train the GPs (training set), while the second set is used to validate the performance of the trained models (validation set). Two GPs are built respectively for the real and imaginary part of each single-ended S-parameter, between any ports a and b of the interconnect: $Re[S_{a,b}(f)] \sim GP_{re}|_{a,b}(f)$ and $Im[s_{a,b}(f)] \sim GP_{im}|_{a,b}(f)$. Therefore, the predicted complex-valued S-parameters are

$$S_{a,b} = GP_{re}|_{a,b} + j \cdot GP_{im}|_{a,b} \ . \quad (5)$$

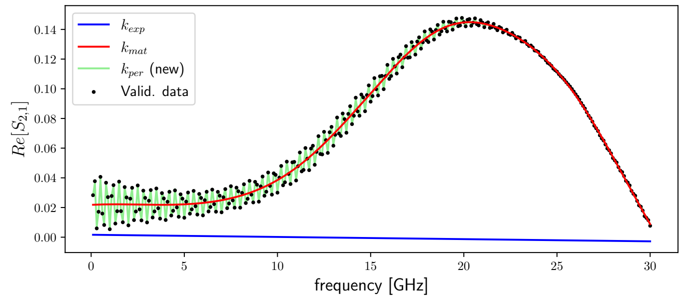

Fig. 2. Comparison of real part of near-end crosstalk, $S_{2,1}$, as predicted by GP_{re} with different kernels, and validation data. Only the newly proposed, periodic kernel readily captures all features of the original S-parameters.

TABLE I
Prediction error of GP models for entire frequency range.

S-parameter	Mean Abs. Err. (dB)			Max. Abs. Err. (dB)		
	k_{exp}	k_{mat}	k_{per}	k_{exp}	k_{mat}	k_{per}
Avg. $S_{n,n}$	-16.03	-15.20	-59.07	-11.05	-10.68	-38.74
Avg. $S_{n,n+1}$	-17.36	-17.24	-60.82	-12.53	-12.52	-44.64

In the following, three kernel functions are compared: the general-purpose squared exponential (k_{exp}) and Matern (k_{mat}) kernels [4] and the new periodic kernel (k_{per}). For this application example, it is experimentally verified that the k_{per} kernel achieves the best performance when 2 periodic terms are used ($N_p = 2$). Furthermore, results indicate that the standard kernels do not accurately interpolate the training data for most of the S-parameter curves. For instance, Fig. 2 shows the real values of $S_{2,1}$ predicted at the validation frequencies by the GP. It can be observed that the GP with k_{per} correctly predicts the validation samples. On the other hand, the conventional kernels k_{exp} and k_{mat} fail to model the oscillations. For some S-parameters, k_{mat} manages to fit the average value of the oscillations as shown in Fig. 2. The kernel k_{exp} regularly yields significantly deviating results.

The predictions using the new kernel are highly accurate for all S-matrix elements. Indeed, Fig. 3 illustrates this by example of four distinct interconnect S-parameters: reflection, transmission, near-end crosstalk, and far-end crosstalk. The predicted real part (green) matches the validation samples (black) and the predicted magnitude (orange) matches the validation samples (black) in all randomly selected cases, respectively. In particular, the model captures the amplitude oscillations of the crosstalk. For completeness, Table I lists the mean and maximum prediction errors for reflections and near-end crosstalk S-parameter across the entire frequency range with regard to each kernel.

Moreover, the high accuracy obtained for both the real and imaginary part allows one to recover the S magnitude, phase, mixed-mode S parameters, or other figures of merit that are derived from the single-ended S-parameters. For example, the mixed-mode insertion loss (IL) and return loss (RL) are of

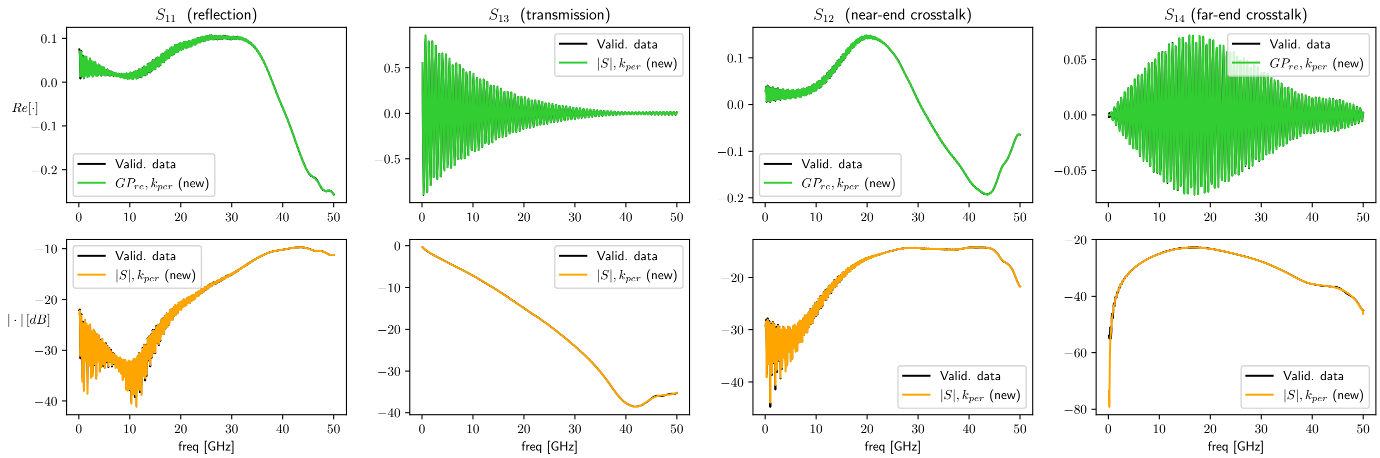

Fig. 3. Single-ended S-parameters of interest, between the first pair of the interconnect: comparison of validation set (black lines) with predicted values using the new kernel. First and second rows show the real part and the magnitude, respectively.

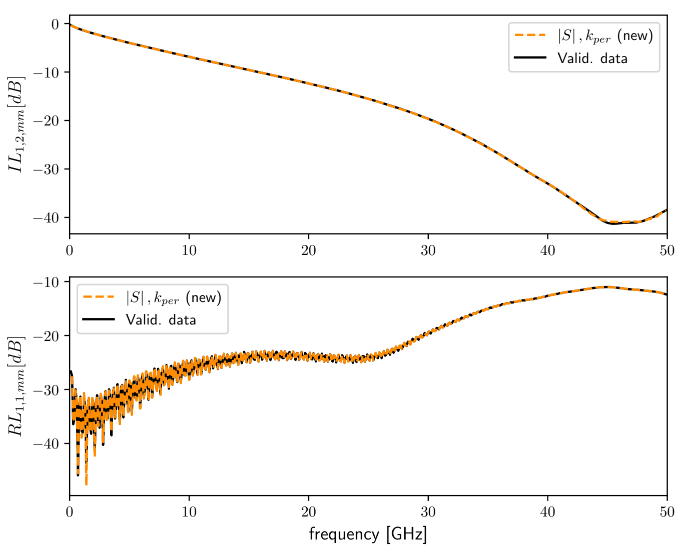

Fig. 4. Comparison of quantities, that are derived from the original S-parameter representation as real/imaginary part for the previously analyzed interconnect: differential insertion loss (IL) and differential return loss (RL). The validation data is plotted in solid black, the GP with proposed kernel is plotted in dashed orange.

particular interest in interconnect design. Figure III correlates the predicted IL and RL with the respective ones from the validation data set.

IV. Conclusion

The novel kernel uses periodic components that enable the Gaussian process modeling of S-parameters that show dynamic and oscillatory behavior, e.g. electrically long interconnects with considerable delays. It is superior to a standard kernel that does not achieve sufficient accuracy. In addition, the new kernel allows retrieval of frequency-dependent figures of merit thanks to the low prediction error on the real and imaginary part of the S-parameters. Further studies are necessary to guarantee physical properties, such as the S-matrix causality, and to extend the proposed kernel to the modeling of S-parameters to cover additional design variables.

References

[1] Grivet-Talocia and Gustavsen, "Black-box macromodeling and its EMC applications," *IEEE Electromagnetic Compatibility Magazine*, vol. 5, no. 3, pp. 71–78, 2016.

[2] J. Jin, F. Feng, W. Zhang, J. Zhang, Z. Zhao, and Q.-J. Zhang, "Recent advances in deep neural network technique for high-dimensional microwave modeling," in *IEEE MTT-S International Conference on Numerical Electromagnetic and Multiphysics Modeling and Optimization (NEMO)*, 2020.

[3] 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.

[4] C. Rasmussen and C. Williams, "Regression," in *Gaussian Processes for Machine Learning*. MIT Press, 2005, ch. 2, pp. 7–32.

[5] S. Koziel, S. Ogurtsov, I. Couckuyt, and T. Dhaene, "Variable-fidelity electromagnetic simulations and co-kriging for accurate modeling of antennas," *IEEE Transactions on Antennas and Propagation*, vol. 61, no. 3, pp. 1301–1308, 2013.

[6] F. Garbuglia, D. Spina, D. Deschrijver, I. Couckuyt, and T. Dhaene, "Bayesian optimization for microwave devices using deep GP spectral surrogate models," *IEEE Transactions on Microwave Theory and Techniques*, pp. 1–8, 2022.

[7] A. Wilson and R. Adams, "Gaussian process kernels for pattern discovery and extrapolation," in *Proceedings of the 30th International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, S. Dasgupta and D. McAllester, Eds., vol. 28, no. 3. Atlanta, Georgia, USA: PMLR, 17–19 Jun 2013, pp. 1067–1075. [Online]. Available: https://proceedings.mlr.press/v28/wilson13.html

[8] F. Garbuglia, T. Claeys, I. Couckuyt, D. Deschrijver, and T. Dhaene, "Bayesian active learning for radiation pattern sampling over cylindrical surfaces," *IEEE Transactions on Electromagnetic Compatibility*, vol. 64, no. 5, pp. 1391–1398, 2022.

[9] T. Reuschel, J. B. Preibisch, K. Scharff, R. Rimolo-Donadio, X. Duan, Y. H. Kwark, and C. Schuster, "Efficient prediction of equalization effort and channel performance for PCB-based data links," *IEEE Transactions on Components, Packaging and Manufacturing Technology*, vol. 7, no. 11, pp. 1842–1851, 2017.

[10] C. Rasmussen and C. Williams, "Model selection and adaptation of hyperparameters," in *Gaussian Processes for Machine Learning*. MIT Press, 2005, ch. 5, pp. 105–129.