Modeling Microwave S-parameters using Frequency-scaled Rational Gaussian Process Kernels

Thijs Ullrick
IDLab
Ghent University - imec
Ghent, Belgium
Thijs.Ullrick@UGent.be

Dirk Deschrijver
IDLab
Ghent University - imec
Ghent, Belgium
Dirk.Deschrijver@UGent.be

Wim Bogaerts
Photonics Research Group
Ghent University - imec
Ghent, Belgium
Wim.Bogaerts@UGent.be

Tom Dhaene
IDLab
Ghent University - imec
Ghent, Belgium
Tom.Dhaene@UGent.be

***Abstract*—This work presents a machine learning technique to model the complex-valued scattering parameters (S-parameters) of passive microwave devices as a function of frequency and a set of design variables. The proposed Gaussian process (GP) model intricately models the real and imaginary parts of the S-parameters by employing a physics-informed kernel, adept at representing complex holomorphic functions and incorporating the Hermitian symmetry inherent in scattering parameters. Additionally, to extend the kernel's capabilities to higher dimensions beyond standard GP techniques, it is extended with a frequency scaling, enhancing the modeling capacity. The resulting physics-informed frequency-scaled GP model accurately predicts the S-parameter values at desired parameter configurations in the design space. One application example demonstrates the superiority of the new kernel, compared to standard GP kernels.**

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

I. Introduction

PARAMETRIC macromodeling is indispensable for the characterization of high-frequency electromagnetic (EM) systems, and plays a crucial role in design space exploration, optimization, and sensitivity analysis. Several widely used macromodeling techniques rely on vector fitting (VF) to build a rational function approximation [1]. A major advantage of these rational models is their seamless conversion into state-space form, facilitating integration in SPICE-like solvers for time-domain simulations [2]. These rational models, however, do not provide any uncertainty estimation, limiting their applicability for design optimization purposes.

Recent advancements in macromodeling utilize machine learning techniques such as artificial neural networks (ANN) [3], [4] and support vector machines (SVM) [5] to address the limitations of standard approaches like VF [1] and AAA [6]. However, ANNs, effective for high-dimensional and nonlinear functions, require substantial amounts of training data and are prone to overfitting. Conversely, SVMs offer robust regularization but lack probabilistic interpretability.

Stochastic models, such as Gaussian processes (GP), offer a promising alternative due to their data efficiency and posterior variance estimation, particularly advantageous in computationally expensive optimization scenarios [7]. Standard GPs, however, are typically adopted for modeling smooth functions, while microwave S-parameters often exhibit a dynamic behavior. In fact, this property of the GP derives from the typical covariance function used among the data points, also known as kernels.

While S-parameters may exhibit dynamic behavior, changes in the overall frequency response due to adjustments in the device's geometrical dimensions or dielectric properties are usually smooth or can be represented as a compression/expansion transformation along the frequency axis. In this regard, the present work introduces a novel kernel for modeling parametric S-parameters. It combines the rational Szegö kernel, originally proposed by Bect et al. [8], with standard GP kernels to extend its applicability to a multidimensional settings. Moreover, a parametric scaling in the frequency dimension is incorporated in the kernel, significantly enhancing its performance in modeling microwave S-parameters.

II. Methodology

A. Gaussian Process Modeling

GPs are probabilistic models that define distributions over functions, where the joint distribution of any collection of points on that function follows a multivariate normal distribution. In particular, the GP is data-driven: it doesn't depend on a fixed number of parameters, such as the poles and residues of a rational VF macromodel. Instead, its complexity and capacity to express patterns increase with the volume of training data. This adaptability makes GPs highly data-efficient, facilitating accurate predictions even when the dataset is limited in size.

GPs are defined by a mean function, representing the expected value of the function at each point, and a covariance function, also referred to as the kernel, which captures the correlation between pairs of points. In many cases, prior information on the underlying stochastic process or function, such as the periodicity or smoothness, can be encoded in either the mean function or kernel. By properly incorporating these assumptions, GPs can achieve high accuracy, even when trained on small datasets.

This work was supported by the 'Flemish Research Foundation (FWO-Vlaanderen) under grant G031421N' and 'Onderzoeksprogramma Artificiële Intelligentie (AI) Vlaanderen' programs.

2024 IEEE 33rd Conference on Electrical Performance of Electronic Packaging and Systems (EPEPS) | 979-8-3503-5123-1/24/$31.00 ©2024 IEEE | DOI: 10.1109/EPEPS61853.2024.10754263

Once the kernel is defined, new function values are predicted via Gaussian Process Regression (GPR), also known as Kriging. In essence, GPR entails two primary steps: specifying a prior distribution based on assumptions about the underlying data-generating process, typically achieved through the design of appropriate mean and kernel functions, followed by updating this prior using Bayes' theorem to derive the posterior distribution. The posterior distribution in GPR captures both the predictive mean and the uncertainty associated with each prediction, making it a powerful tool for regression tasks.

B. Rational Szegö Kernel

Few physics-informed kernels have been introduced in the literature for the modeling of microwave S-parameters [8], [9]. For instance, the delayed GP method introduced by Garbuglia et al. [9] demonstrates comparable accuracy to delayed vector fitting for modeling elecrically long interconnects prone to significant cross-talk. However, it fails to leverage the intricate relation of the real and imaginary part of the scattering parameters, and instead, models them as independent variables using separate GPs. To use the data more effectively, a Multi-Output Gaussian Process (MOGP) is adopted in this work. An MOGP is an extension of the standard GP that can simultaneously model multiple related outputs. This is particularly useful in scenarios where outputs are correlated or share common characteristics, allowing for more efficient and accurate predictions compared to modeling each output independently.

Bect et al. [8] recently introduced a novel covariance function, referred to as the rational Szegö kernel, for modeling complex-valued functions. This kernel has been designed to represent a space of complex holomorphic functions and incorporates the Hermitian symmetry inherent in the frequency response of dynamical systems. It effectively captures the correlation between the real and imaginary parts, leading to a fitting procedure that converges significantly faster compared to standard GP kernels. For a detailed discussion of its derivation and properties, readers are referred to [8]. The Szegö kernel is adopted in present work as the covariance function of the MOGP and can be expressed as follows

$$K_{sz}(s_0, s_1) = \begin{bmatrix} \Re(\frac{k+c}{2}) & \Im(\frac{-k+c}{2}) \\ \Im(\frac{k+c}{2}) & \Re(\frac{k-c}{2}) \end{bmatrix} \quad (1)$$

with

$$\begin{aligned} k(s_0, s_1) &= \frac{\sigma^2}{2\alpha + s_0 + s_1^*} \\ c(s_0, s_1) &= \frac{\sigma^2}{2\alpha + s_0 + s_1} \end{aligned} \quad (2)$$

where α and σ^2 are the hyperparameters of the rational kernel and $s = j2\pi f$ is the Laplace variable.

C. Rational Kernel Extension to Higher Dimensions

While the Szegö kernel excels in representing complex-valued functions across frequency, the dynamics associated with the parameterization of these functions can be effectively captured by standard GP kernels. In this work, the Szegö kernel is extended to higher dimensions by combining it with a standard Matérn 5/2 kernel [9], for modeling S-parameter variations with respect to design variables. Leveraging the distinct strengths of each kernel, the covariance function of the MOGP can be expressed as

$$K_{cm}(s_0, \mathbf{x}_0, s_1, \mathbf{x}_1) = K_{sz}(s_0, s_1) \circ K_{mat}(\mathbf{x}_0, \mathbf{x}_1) \quad (3)$$

where $\mathbf{x}_0$ and $\mathbf{x}_1$ are vectors containing the design variables.

D. Frequency-scaled Kernels

Microwave S-parameters often exhibit compression or expansion along the frequency axis. Assuming the frequency response remains unaffected by other transformations or changes, this implies that we can express $S(s, \hat{\mathbf{x}}_0) = S(\gamma s, \hat{\mathbf{x}}_1)$. Here, γ represents the scaling of the frequency response relative to the Laplace variable s as the design variables are tuned from $\hat{\mathbf{x}}_0$ to $\hat{\mathbf{x}}_1$. Consequently, the parameter configurations $(s,\ \mathbf{x}_0)$ and $(\gamma s,\ \mathbf{x}_1)$ are highly correlated. However, stationary kernels like the SE kernel, which rely solely on the distance between points, struggle to capture such correlations effectively. To address this, the covariance function (3) is enhanced by incorporating a linear frequency scaling with respect to the design variables, leading to

$$K_{fs} = K_{cm}(s_0(1 + \boldsymbol{\gamma} \cdot \mathbf{x}_0), \mathbf{x}_0, s_1(1 + \boldsymbol{\gamma} \cdot \mathbf{x}_1), \mathbf{x}_1) \quad (4)$$

where $\boldsymbol{\gamma}$ serves as an additional hyperparameter describing the linear frequency scaling. It is noteworthy that the use of a frequency-scaling coefficient to improve modeling accuracy is inspired by techniques employed in rational modeling [10].

III. Application Example

In this application example, the proposed GPR framework is evaluated for modeling the transmission of a double-folded microstrip band-stop filter, formed by a dielectric substrate between a top metallization and a bottom ground plane. As illustrated in Fig. 1, the geometry of the upper layer consists of two stubs, with identical length and spacing, folded onto the two sides of a transmission line. The S-parameters of the device are simulated for a set of 15 equispaced frequencies within the $[5, 25]$ GHz range using ADS Momentum. The chosen design parameters are the stub length $L \in [1.5, 3.0]$ μm and line-stub spacing $S \in [0.0, 0.3]$ μm.

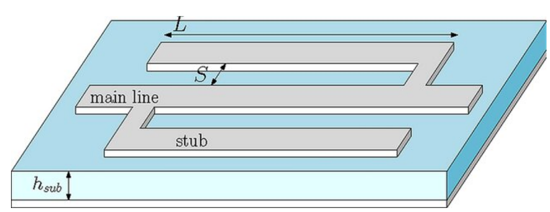

Fig. 1: Double-folded microstrip band-stop filter geometry.

In the following analysis, three kernel functions are compared: two separate GPs for independent modeling of the real and imaginary part using a Matérn kernel K_{mat}, an

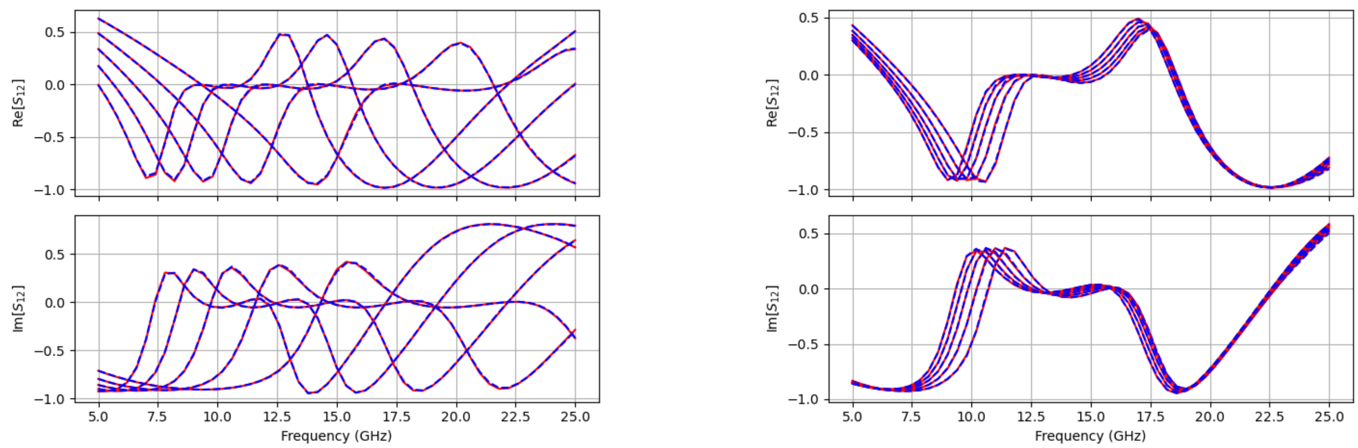

Fig. 2: Comparison of real and imaginary part of S_{12}, as predicted with K_{fs} (blue), and the original S-parameter (red), for varying L (left) and S (right).

TABLE I: Prediction accuracy of the GP models

$N_f \times N_L \times N_S$	RMSE (dB)			MAE (dB)		
	K_{mat}	K_{cm}	K_{fs}	K_{mat}	K_{cm}	K_{fs}
$15 \times 2 \times 2$	-3.7	-4.5	-23.7	2.8	2.0	-13.5
$15 \times 3 \times 3$	-8.9	-8.7	-36.6	-0.8	-0.5	-25.5
$15 \times 4 \times 4$	-11.0	-12.5	-40.9	-3.6	-3.0	-31.3
$15 \times 5 \times 5$	-12.3	-16.4	-41.4	-4.6	-6.0	-32.5

MOGP using the composite Szegö Matérn kernel K_{cm}, and an MOGP using the frequency-scaled enhanced version K_{fs}. The implementation of the GP is done in Python using the GPyTorch library. It is worth noting that the implementation employs real-valued kernels and inputs, where the outputs of the MOGP correspond to the real and imaginary parts, respectively.

The training data is generated by simulating the frequency response of the double folded stub filter on a uniform N×N grid within the 2D design space. The hyperparameters of each model are then selected by minimizing the marginal log-likelihood. Once the models are fit to the data, their accuracy is quantified in terms of the root-mean-squared error (RMSE) and maximum absolute error (MAE), which are evaluated on a set of 400 (f, L, s) samples chosen randomly according to an Latin hypercube design. Both metrics are reported in Table I for each kernel.

The new kernel yields highly accurate predictions despite the significant variability in the S-parameters. Indeed, this is demonstrated in Fig. 2, which plots S_{12} for varying L and S respectively. The predicted real and imaginary parts (blue) accurately match the validation samples (red) computed via EM simulation. In particular, the model effectively captures the compression and expansion observed in the S-parameters.

IV. Conclusion

The rational kernel used in this work enables detailed modeling of the real and imaginary parts of highly dynamic S-parameters. Additionally, extending the kernel to higher dimensions, particularly by incorporating a linear frequency scaling, has shown superior performance in modeling parametric microwave S-parameters compared to standard kernels, which lack sufficient accuracy.

References

[1] S. Grivet-Talocia and B. Gustavsen, *Passive macromodeling*. Wiley series in microwave and optical engineering, Hoboken, New Jersey: John Wiley & Sons, Inc, 2016.

[2] Y. Ye, T. Ullrick, W. Bogaerts, T. Dhaene, and D. Spina, "SPICE-Compatible Equivalent Circuit Models for Accurate Time-Domain Simulations of Passive Photonic Integrated Circuits," *Journal of Lightwave Technology*, vol. 40, pp. 7856–7868, Dec. 2022.

[3] F. Feng, W. Na, J. Jin, W. Zhang, and Q.-J. Zhang, "ANNs for Fast Parameterized EM Modeling: The State of the Art in Machine Learning for Design Automation of Passive Microwave Structures," *IEEE Microwave Magazine*, vol. 22, pp. 37–50, Oct. 2021.

[4] M. Schierholz, I. Erdin, J. Balachandran, C. Yang, and C. Schuster, "Parametric S-Parameters for PCB based Power Delivery Network Design Using Machine Learning," in *2022 IEEE 26th Workshop on Signal and Power Integrity (SPI)*, (Siegen, Germany), pp. 1–4, IEEE, May 2022.

[5] 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.

[6] Y. Nakatsukasa, O. Sète, and L. N. Trefethen, "The AAA Algorithm for Rational Approximation," *SIAM Journal on Scientific Computing*, vol. 40, pp. A1494–A1522, Jan. 2018. Publisher: Society for Industrial and Applied Mathematics.

[7] C. E. Rasmussen and C. K. I. Williams, *Gaussian processes for machine learning*. Adaptive computation and machine learning, Cambridge, Mass: MIT Press, 2006. OCLC: ocm61285753.

[8] J. Bect, N. Georg, U. Römer, and S. Schöps, "Rational kernel-based interpolation for complex-valued frequency response functions," Sept. 2023. arXiv:2307.13484 [cs, math].

[9] F. Garbuglia, T. Reuschel, C. Schuster, D. Deschrijver, T. Dhaene, and D. Spina, "Modeling Electrically Long Interconnects Using Physics-Informed Delayed Gaussian Processes," *IEEE Transactions on Electromagnetic Compatibility*, vol. 65, pp. 1715–1723, Dec. 2023.

[10] T. Ullrick, D. Spina, W. Bogaerts, and T. Dhaene, "Wideband parametric baseband macromodeling of linear and passive photonic circuits via complex vector fitting," *Scientific Reports*, vol. 13, p. 15407, Sept. 2023. Number: 1 Publisher: Nature Publishing Group.