Bayesian Optimization of Microwave Filters: A Physics-Informed Approach Using the Szegö Kernel

1st Yens Lindemans   2nd Thijs Ullrick   3rd Ivo Couckuyt   4rd Dirk Deschrijver   5th Tom Dhaene
IDLAB
Ghent University - imec
Ghent, Belgium
Yens.Lindemans@UGent.be

***Abstract*—This work presents a data-efficient approach for optimizing microwave devices by combining Bayesian optimization with physics-informed Gaussian process models. The proposed approach models complex-valued frequency responses as a function of design parameters and frequency instead of modeling an objective function, enabling dynamic optimization of devices with intricate performance criteria. The method is validated through the optimization of a zig-zag microstrip bandpass filter to achieve the desired transmission characteristics. This work highlights the potential of Bayesian optimization to reduce computational costs in microwave device design.**

***Index Terms*—Bayesian optimization, Gaussian process, Microwave filters, Machine learning (ML), S-parameters**

I. Introduction

EFFICIENT optimization of microwave devices poses a significant challenge in modern engineering, particularly as these devices become increasingly complex. Typically, a device is designed with a specific frequency response in mind, but determining the optimal design parameters (e.g., the width or length of metallic traces) to achieve the desired response remains a nontrivial task. Traditional optimization algorithms iterate through design parameters until a predefined performance criterion is met. However, these methods require electromagnetic (EM) simulations at each iteration to evaluate the S-parameters of the new designs, resulting in high computational costs. To address this challenge, physics informed surrogate models can be employed as substitutes for expensive EM simulations during the optimization process [1], [2].

Bayesian optimization (BO) is a widely used algorithm for optimizing expensive black-box functions [3], [4]. BO uses a surrogate model with uncertainty quantification to iteratively sample new data points and refine its predictions of the objective function. The stochastic nature makes it data efficient and allows for an inherent balance between exploration (investigating uncharted design regions) and exploitation (refining promising designs). A critical component of any optimization algorithm is the formulation of the objective function. Simple objectives expressed directly in terms of design parameters may lead to suboptimal solutions if they fail to capture the nuances of the frequency response. Ideally, the surrogate model should directly model the S-parameters, enabling dynamic adjustments to the objective function *during* the optimization process. For instance, this flexibility allows the design expert to alter the objective in response to convergence toward infeasible designs or to incorporate trade-offs, all without requiring the surrogate model to be retrained.

Common stochastic surrogate models for BO are Gaussian process (GP) models [5]. These kernel-based machine learning (ML) models provide analytical uncertainty estimates for their predictions, enabling closed-form expressions for output probability distributions and making them well-suited for data modeling tasks. Standard GPs, however, are primarily designed for stationary functions, which can be characterized by a fixed *lengthscale*. In contrast, S-parameters are typically non-stationary functions of frequency, presenting unique challenges for GP modeling. Efforts to address these challenges include Deep GPs [6] and non-stationary kernels [7]–[9]. Notably, Bect et al. [9] introduced a kernel specifically tailored for GP regression of complex-valued S-parameters, and Ullrick et al. [10] further extended this approach to incorporate dependencies on design parameters.

In this work, we leverage a specialized kernel, initially proposed in [9] and further refined in [10], to enable data-efficient Bayesian optimization of a microstrip bandpass filter. This physics-informed GP model, designed to handle complex-valued functions, directly models the spectral S-parameters of the device. To guide the optimization, we define an objective function based on the magnitude of the S-parameters, capturing the desired filter performance. We employ Thompson sampling, drawing samples from the GP's posterior distribution. To address the non-linearity introduced by the magnitude operation, which distorts the Gaussianity of the posterior distribution, we incorporate an additional Monte Carlo sampling step within the optimization loop.

II. Methodology

A. Gaussian Process Model

GPs perform regression by sampling from a *posterior* Gaussian distribution over functions. This distribution is fully characterized by a mean, representing the model's prediction, and a variance, representing its uncertainty. Initially, the model assumes a multivariate normal distribution with zero mean and

This work is supported by Flemish Research Foundation (FWO-Vlaanderen) under grant G095224N and the Flanders AI Research Program.

a covariance matrix Σ, constructed using a *kernel* function. The kernel is central to the GP, as it defines the correlations between input points and, consequently, the covariance between outputs. Input points that are close in the input space are more likely to produce similar outputs. Moreover, the functional properties of the kernel—such as smoothness or periodicity—are directly reflected in the GP's output.

The kernel determines the similarity between inputs based on a *lengthscale*, a key hyperparameter that defines the range over which inputs are considered correlated. This lengthscale is optimized by fitting the model to data via Bayes' theorem. Because the lengthscale assumes a fixed variation across the input space, GPs are typically suited for modeling stationary functions.

B. Parametrized Rational Szegö Kernel with Input Warping

To accurately model S-parameters, the physics-informed Szegö kernel is used. It is specifically designed to represent the space of complex holomorphic functions and to incorporate Hermitian symmetry inherent to frequency response functions of passive MW/RF devices [9]. Bect et al. [9] achieve this by constructing a multi-output GP (MOGP) which models the real and imaginary part of the S-parameters as separate correlated outputs. This method allows us to use real-valued GPs, of which there are many libraries for training and optimization. The Szegö kernel is defined as

$$\tilde{k}(s_0,\, s_1) = \begin{pmatrix} \Re\left(\frac{k+c}{2}\right) & \Im\left(\frac{-k+c}{2}\right) \\ \Im\left(\frac{k+c}{2}\right) & \Re\left(\frac{k-c}{2}\right) \end{pmatrix}, \tag{1}$$

with

$$k(s_0,\, s_1) = \frac{\sigma^2}{2\alpha + s_0 + s_1^*}, \tag{2}$$

$$c(s_0,\, s_1) = \frac{\sigma^2}{2\alpha + s_0 + s_1}, \tag{3}$$

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

To capture the dependence on design parameters $\boldsymbol{x}$, a Matérn kernel can be multiplied with the Szegö kernel. However, Ullrick et al. [10] showed that simply incorporating a Matérn kernel does not significantly improve modeling performance. Instead, they included an extra input warping step to the Szegö kernel, which they called *frequency scaling*. The reasoning behind this step is that S-parameters are often compressed, expanded, or shifted along the frequency axis when changing design parameters. Therefore, the linear parameter-dependent transformation $s \mapsto s(1 + \boldsymbol{\gamma} \cdot \boldsymbol{x})$ is added, which gives

$$K = \tilde{k}(s_0(1 + \boldsymbol{\gamma} \cdot \boldsymbol{x}_0),\, s_1(1 + \boldsymbol{\gamma} \cdot \boldsymbol{x}_1)) \cdot k_{\text{Mat}}(\boldsymbol{x}_0,\, \boldsymbol{x}_1), \tag{4}$$

where $\boldsymbol{\gamma}$ is a vector of hyperparameters.

C. Bayesian Optimization

BO is used to optimize a device with d design parameters by starting with a small initial dataset $\mathcal{D}$, which contains parameter configurations $\boldsymbol{x}_i \in \mathbb{R}^d$ and their corresponding frequency responses $S_i(\boldsymbol{x}, f) \in \mathbb{C}$. For each configuration $\boldsymbol{x}_i$, the S-parameters are sampled at n_f discrete frequency points. Both the design parameters and frequency points serve as inputs to the GP model, while the frequency response is treated as the target output.

To optimize the design, an objective function $q(\boldsymbol{x})$ is defined over the parameter space by assigning a *score* to each frequency response $S(\boldsymbol{x}, f)$. Specifically, a scoring function $g(\boldsymbol{x}, f)$ evaluates the difference between the desired and current responses at each frequency point. The objective is then computed as

$$q(\boldsymbol{x}) = q\big(S(\boldsymbol{x}, f)\big) = \sum_{k=1}^{n_f} g(\boldsymbol{x}, f_k)\,. \tag{5}$$

In such manner, the optimal design parameters $\boldsymbol{x}_{\text{opt}}$ are obtained by maximizing the objective:

$$\boldsymbol{x}_{\text{opt}} = \arg\max_{\boldsymbol{x}} \{q(\boldsymbol{x})\} \tag{6}$$

This is different than standard BO, where typically the GP models the objective function directly. As stated in Section I and elaborated in [6], this allows the objective to be dynamically redefined during optimization.

The optimization proceeds iteratively. A stop condition, such as a maximum number of simulations, determines whether the algorithm ends or continues. If not met, an acquisition function is optimized to select the next design to evaluate. This function scores potential designs based on the GP posterior, balancing exploration and exploitation. Common acquisition functions, such as the upper confidence bound (UCB) and expected improvement (EI) [11], are computationally efficient and rely on closed-form expressions of the Gaussian posterior.

However, our GP model outputs a posterior distribution on the real and imaginary parts of $S(\boldsymbol{x}, f)$. When transforming to the magnitude $|S| = \sqrt{\Re(S)^2 + \Im(S)^2}$, the Gaussian posterior is no longer analytically tractable. To address this, Monte Carlo (MC) sampling is used to approximate the posterior on the objective function. This process can be computationally demanding, as the GP must evaluate $n_{\text{MC}} \times n_s \times n_f$ inputs, where n_{MC} is the number of MC samples and n_s is the number of design configurations in the dataset.

To reduce computational cost, we employ Thompson sampling as the optimization policy. In this approach, a single sample S response is drawn from the GP posterior, and the objective is maximized based on this sample:

$$\tilde{S}(\boldsymbol{x}, f) \sim GP(\boldsymbol{x}, f)\,, \tag{7}$$

$$\boldsymbol{x}_{i+1} = \arg\max_{\boldsymbol{x}} \{q\big(\tilde{S}(\boldsymbol{x}, f)\big)\}\,. \tag{8}$$

This method effectively optimizes a random acquisition function at each BO iteration [3], significantly reducing computational complexity. Exploration remains present, as each sampled output can differ significantly in regions where the model is uncertain. Consequently, the algorithm explores the parameter space until it develops a sufficiently global understanding of the objective behavior.

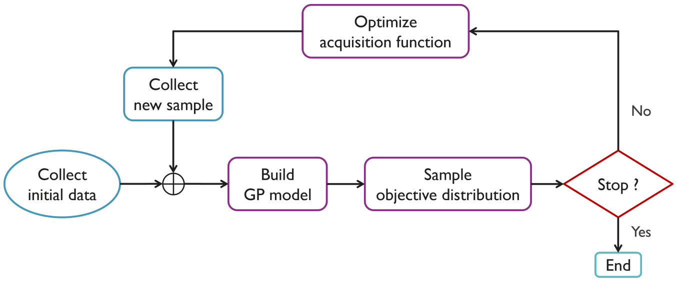

Figure 1. Bayesian optimization workflow

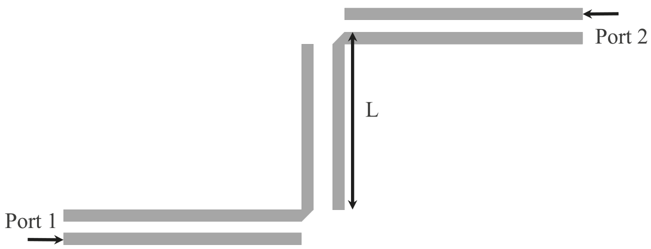

Figure 2. Microstrip bandpass filter geometry.

Once the next design $\boldsymbol{x}_{i+1}$ is determined, the S-parameters are simulated for this configuration and added to the dataset $\mathcal{D}$. The process repeats until the stopping criterion is met. The general workflow of the BO algorithm is illustrated in Fig. 1.

III. Application: Zig-Zag Bandpass Filter

The proposed method is tested by optimizing the 2-port zig-zag microstrip bandpass filter depicted in Fig. 2, also studied in [6]. The device consists of metallic conductors on a 0.5 mm thick dielectric substrate with relative permittivity $\epsilon_r = 2.2$ and loss tangent $\tan\delta = 9 \times 10^{-4}$. The gap between horizontal conductors is fixed at 0.3 mm, between the vertical conductors at 1.0 mm, and the conductor width is fixed at 0.4 mm. The optimization parameter is the length of the vertical conductors, $L \in [5.0, 25.0]$ mm. The frequency responses are simulated using the advanced design system (ADS) Momentum software [12] at $n_f = 71$ equidistant frequency points in $[1.0, 4.5]$ GHz. A validation set is generated by evaluating 50 designs at 125 equidistant frequency points.

The scoring function $g(\boldsymbol{x}, f)$ is designed to reflect the desired bandpass behavior in frequency range $f \in [2.45, 2.55]$ GHz:

$$g(\boldsymbol{x}, f) = \begin{cases} 10(|S_{21}(\boldsymbol{x}, f)| - 0.708) & \text{if } f \in [2.45, 2.55] \text{ GHz} \\ 0.01 - |S_{21}(\boldsymbol{x}, f)| & \text{else} \end{cases} \tag{9}$$

Here, the values 0.01 and 0.708 correspond to -40 and -3 dB, respectively, emphasizing high transmission in the passband and low transmission elsewhere. Note that for this application, the design parameter is $\boldsymbol{x} = L$.

The BO algorithm is implemented in Python using the BoTorch [13] and GPyTorch libraries. Optimization begins with three initial design samples selected to maximize global information: $L \in \{5.0, 25.0\}$ mm to cover the design limits,

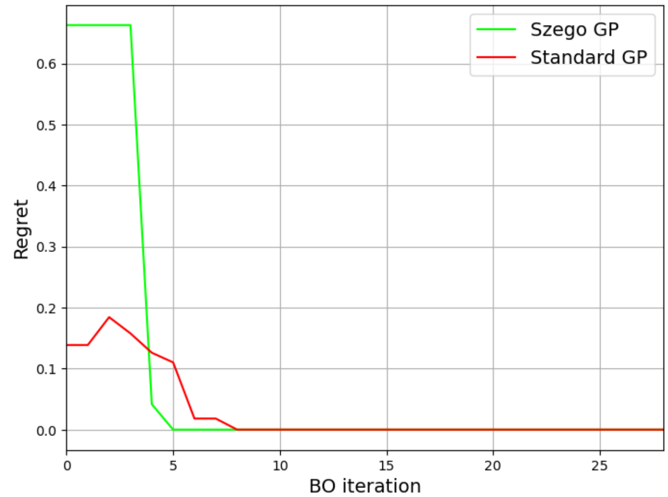

Figure 3. Regret as function of BO iterations. Regret is calculated with respect to the optimal design found by evaluating a test set of 50 L parameters. BO starts with an initial trainig set consisting of 3 L parameters.

with the third sample determined via Latin hypercube design. For comparison, a standard GP with a Matérn 5/2 kernel is also used to optimize the filter. Unlike the proposed Szegö GP model, which models both the real and imaginary components of the frequency response, the standard GP only models the magnitude. This limitation reduces the overall modeling power of the standard GP compared to the Szegö GP. The BO algorithm is carried out for 29 iterations.

The performance of BO is evaluated using the regret metric, defined as:

$$\text{regret} = \frac{\|\boldsymbol{x}_{\text{opt}} - \tilde{\boldsymbol{x}}_{\text{opt}}\|}{\|\boldsymbol{x}_H - \boldsymbol{x}_L\|}. \tag{10}$$

In this expression, $\boldsymbol{x}_H$ and $\boldsymbol{x}_L$ represent the upper and lower bounds of the design space, respectively. The term $\tilde{\boldsymbol{x}}_{\text{opt}}$ refers to the optimal design predicted by the surrogate model, while $\boldsymbol{x}_{\text{opt}} = 18.257$ mm is the true optimal design, determined by evaluating the objective function on the test set. The regret quantifies the normalized difference between the predicted and true optimal designs, providing a measure of the accuracy of the optimization relative to the size of the parameter space.

Fig. 3 shows the regret of the proposed and standard method over each BO iteration. Both methods converge towards the optimal design, with the Szegö GP converging faster than the standard GP.

As stated earlier, the advantage of modeling the frequency response instead of the objective is that it enables a human-in-the-loop scenario, allowing dynamic adjustments to the objective during or after optimization. For this approach to be effective, the GP must maintain high accuracy at unseen values of L. Fig. 4 compares the predictions of the Szegö GP and the standard GP for several test samples. Both models were trained only on the data collected during optimization, resulting in training sets concentrated near the optimal value, $L = 18.257$ mm.

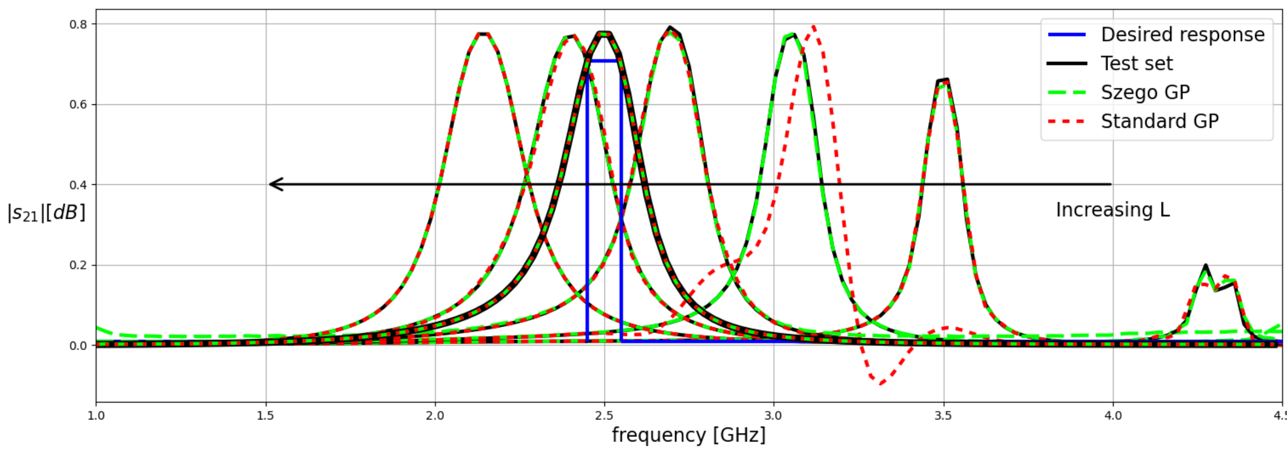

Figure 4. Comparison between predicted and simulated S responses for different L parameters: $L \in [5.0, 10.0, 15.0, \mathbf{18.2}, 20.0, 25.0]$. The bold black line represents the optimal frequency response.

The figure highlights the superior accuracy of the Szegö GP on unseen data. Its physics-informed kernel further reduces the likelihood of nonphysical predictions compared to the standard GP. For example, the standard GP erroneously predicts negative values for $|S_{21}|$ at some values of L, which is unrealistic. This occurs because the standard GP does not restrict its outputs to be positive. In contrast, the Szegö GP, by separately modeling the real and imaginary components, inherently avoids such nonphysical errors.

IV. Conclusion

This paper presents the physics-informed Szegö GP model with input-warping for efficient optimization of electromagnetic devices, demonstrated on a zig-zag microstrip bandpass filter. By modeling the real and imaginary parts of the frequency response, the Szegö GP outperforms a standard GP with a Matérn $5/2$ kernel in accuracy and avoids nonphysical predictions, such as negative transmission values.

The proposed method enables dynamic, human-in-the-loop optimization, offering flexibility to adjust objectives during or after optimization. Results show that the Bayesian Optimization framework effectively identifies the optimal design with fewer iterations, reducing simulation costs. These findings validate the Szegö GP as an efficient tool for physics-informed optimization of electromagnetic devices.

References

[1] R. Alizadeh, J. K. Allen, and F. Mistree, “Managing computational complexity using surrogate models: A critical review,” *Res. Eng. Des.*, vol. 31, no. 3, pp. 275–298, 2020.

[2] G. E. Karniadakis, I. G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, and L. Yang, “Physics-informed machine learning,” *Nat. Rev. Phys.*, vol. 3, no. 6, pp. 422–440, 2021.

[3] P. I. Frazier, “Bayesian optimization,” in *Recent advances in optimization and modeling of contemporary problems*, Informs, 2018, pp. 255–278.

[4] X. Wang, Y. Jin, S. Schmitt, and M. Olhofer, “Recent advances in bayesian optimization,” *ACM Comput. Surv.*, vol. 55, no. 13s, pp. 1–36, 2023.

[5] M. Seeger, “Gaussian processes for machine learning,” *Int. J. Neural Syst.*, vol. 14, no. 02, pp. 69–106, 2004.

[6] F. Garbuglia, D. Spina, D. Deschrijver, I. Couckuyt, and T. Dhaene, “Bayesian optimization for microwave devices using deep gp spectral surrogate models,” *IEEE Trans. Microwave Theory Tech.*, vol. 71, no. 6, pp. 2311–2318, 2022.

[7] C. Paciorek and M. Schervish, “Nonstationary covariance functions for gaussian process regression,” *Advances in neural information processing systems*, vol. 16, 2003.

[8] P. Pandita, P. Tsilifis, N. M. Awalgaonkar, I. Bilionis, and J. Panchal, “Surrogate-based sequential bayesian experimental design using non-stationary gaussian processes,” *Comput. Methods Appl. Mech. Eng.*, vol. 385, p. 114 007, 2021.

[9] J. Bect, N. Georg, U. Römer, and S. Schöps, “Rational kernel-based interpolation for complex-valued frequency response functions,” *SIAM J Sci. Comput.*, vol. 46, no. 6, A3727–A3755, 2024.

[10] T. Ullrick, D. Deschrijver, W. Bogaerts, and T. Dhaene, “Modeling microwave s-parameters using frequency-scaled rational gaussian process kernels,” in *2024 IEEE 33rd Conference on Electrical Performance of Electronic Packaging and Systems (EPEPS)*, IEEE, 2024, pp. 1–3.

[11] J. Wilson, F. Hutter, and M. Deisenroth, “Maximizing acquisition functions for bayesian optimization,” *Advances in neural information processing systems*, vol. 31, 2018.

[12] “Advanced design system.” (Dec. 2024), [Online]. Available: http://www.keysight.com/find/eesof-ads.

[13] M. Balandat, B. Karrer, D. R. Jiang, *et al.*, “BoTorch: A Framework for Efficient Monte-Carlo Bayesian Optimization,” in *Advances in Neural Information Processing Systems 33*, 2020.