Spectral Bayesian Optimization Using a Physics-Informed Rational Szegö Kernel for Microwave Design

Yens Lindemans, *Student Member, IEEE,* Thijs Ullrick, Ivo Couckuyt, *Member, IEEE,* Dirk Deschrijver, *Senior Member, IEEE,* Dries Vande Ginste, *Senior Member, IEEE,* Tom Dhaene, *Senior Member, IEEE*

***Abstract*—Microwave device design increasingly relies on surrogate modeling to accelerate optimization and reduce costly electromagnetic (EM) simulations. This paper presents a spectral Bayesian optimization (SBO) framework leveraging a physics-informed Gaussian process (GP) with a rational complex-valued Szegö kernel and input warping to enhance surrogate accuracy and data efficiency. Unlike conventional methods that model scalar objectives, our approach directly learns the complex-valued frequency response, enforcing causality and Hermitian symmetry. Effectiveness is demonstrated in two cases: a zig-zag microstrip bandpass filter optimized for magnitude response, and a passive differential equalizer optimized for both transmission magnitude and group delay. By embedding prior physics and modeling directly in the frequency domain, the method enables accurate, sample-efficient optimization of frequency-dependent behavior. This work shows how physics-informed Bayesian optimization can significantly improve microwave device design efficiency.**

***Index Terms*—Bayesian optimization, Gaussian process, Microwave devices, Machine learning (ML), Physics-informed, Szegö Kernel**

I. Introduction

THE optimization of microwave components is a persistent challenge in modern engineering, especially as circuit architectures grow in complexity. Engineers often aim to match a device's frequency response to a specified target, yet identifying the optimal design parameters—such as the dimensions of conductive elements or the dielectric properties of substrates—remains a difficult and computationally intensive task. Conventional optimization strategies explore the design space iteratively, requiring a full electromagnetic (EM) simulation at each step to evaluate the resulting S-parameters. This repeated simulation task leads to substantial computational costs. To mitigate this, physics-informed surrogate models offer a promising alternative by approximating EM behavior with far lower evaluation times, thereby accelerating the optimization workflow [1]–[3].

Bayesian optimization (BO) has become a common approach for optimizing costly black-box functions due to its sample efficiency and principled use of uncertainty [4]–[6]. It leverages a probabilistic surrogate model to guide the selection of new sample points, balancing the need to explore unknown regions of the design space with the desire to exploit areas that already show promise. This balance, rooted in the stochastic nature of BO, enables efficient navigation of complex design landscapes. A key factor in any optimization routine is the definition of the objective function(s). When objectives are derived directly from design specifications, they may overlook subtleties in the frequency response, potentially leading to suboptimal outcomes. For instance, objective values computed over a predefined frequency band can be influenced by both the density at which S-parameters are sampled and their sensitivity to the design variables. Consequently, different sampling strategies may favor different designs. A more robust alternative is to model the S-parameters directly with the surrogate and calculating the objective function dynamically throughout the optimization. This approach offers greater flexibility: objectives can be adjusted on-the-fly in response to observed behavior—such as convergence toward undesirable regions or newly identified trade-offs—without the need to retrain the surrogate model.

Gaussian process (GP) models are commonly used stochastic surrogate models for BO [7]. These kernel-based machine learning (ML) methods offer analytical uncertainty estimates alongside their predictions, allowing for closed-form representations of the predicted output distribution. This makes them particularly effective for data-efficient modeling and optimization tasks. However, traditional GP models are typically built under the assumption of stationarity, meaning they assume the function exhibits uniform behavior across its domain. This assumption does not hold in cases where the function contains both slowly varying and highly dynamic regions. This is often the case for S-parameters, which are typically non-stationary functions of frequency and frequently include features like sharp resonances and localized variations that challenge standard GP formulations. To overcome this limitation, several strategies have been proposed, including Deep Gaussian Processes [8] and a range of non-stationary kernels [9]–[11]. A particularly relevant advancement is the kernel developed by Bect et al. [11], specifically tailored for complex-valued S-parameter regression, and further extended by Ullrick et al. [12] to account for dependencies on design parameters.

In this paper, we employ this complex-valued, physics-informed kernel—originally introduced in [11] and enhanced in [12]—as part of a spectral GP model designed to handle frequency responses. This tailored GP surrogate enables efficient modeling of the S-parameters directly in the complex domain. For the purpose of optimization, physical observables such as the magnitude or phase response can be derived from the model and used to formulate appropriate objective functions aligned with design targets.

The remainder of the paper is organized as follows. Section II provides a brief introduction to Gaussian Processes and

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

Yens Lindemans, Thijs Ullrick, Ivo Couckuyt, Dirk Deschrijver, Dries Vande Ginste, and Tom Dhaene are with the Department of Information Technology, Ghent University - imec, Technologiepark-Zwijnaarde 126, 9052 Ghent, Belgium (Corresponding author email: Yens.Lindemans@UGent.be)

Bayesian Optimization. Section III outlines the methodology of the proposed approach. In Sections IV and V, we demonstrate the method on two passive microwave devices: a microstrip bandpass filter and a passive differential equalizer. The results are compared with those obtained using a standard BO implementation. Finally, Section VI presents the conclusions.

II. Theoretical Background

Consider a blackbox function $\mathcal{F}(x) : \mathbb{R}^d \to \mathbb{K}$, ($\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$), which we aim to model based on observed data. Let $\mathcal{D} = (X, y)$ denote a set of N observations, where $X \in \mathbb{R}^{N\times d}$ represents the inputs and $y \in \mathbb{K}^N$ the corresponding outputs. In many real-world applications, measurements are subject to noise, so the observations are modeled as

$$y_i = \mathcal{F}(x_i) + \epsilon, \tag{1}$$

where $\epsilon \sim \mathcal{N}(0, \sigma_n^2)$ is Gaussian noise with zero mean and variance σ_n^2.

A. Real-Valued Gaussian Process Regression

A Gaussian Process (GP) defines a *prior* over functions, meaning that any finite collection of function values follows a multivariate Gaussian distribution. This distribution is fully characterized by a mean function $m(x)$ and a covariance function $k(x,\, x')$ defined as

$$\begin{aligned} m(x) &= \mathbb{E}\left[\mathcal{F}(x)\right], \\ k(x,\, x') &= \mathbb{E}\left[\left(\mathcal{F}(x) - m(x)\right)\left(\mathcal{F}(x') - m(x')\right)\right]. \end{aligned} \tag{2}$$

We then denote the GP as

$$\mathcal{F}(x) \sim \mathcal{GP}\left(m(x),\, k(x,\, x')\right). \tag{3}$$

The prior mean function is usually set to zero, which can be justified by standardizing the observations. In contrast, the covariance function, or *kernel*, should be carefully chosen as it plays a central role in GP regression. Its properties characterize the general behavior of the predicted function, such as smoothness or periodicity. For example, if the kernel function is continuously differentiable and periodic, the resulting output of the GP will have the same characteristics.

The kernel effectively correlates similarities in input points with the covariance between outputs. Generally, the similarity of two inputs is measured as the distance between these points relative to a *lengthscale*, a hyperparameter of the GP that has to be trained by the model. A popular choice is the Matérn kernel:

$$k_\nu(x, x') = \frac{2^{1-\nu}}{\Gamma(\nu)}\left(\sqrt{2\nu\delta}\right)^{\nu} K_\nu\left(\sqrt{2\nu\delta}\right), \tag{4}$$

where

$$\delta = \sum_{k=1}^{d} \frac{(x^{(k)} - x'^{(k)})^2}{\lambda_k}, \tag{5}$$

with λ_k denoting the lengthscale parameters, $\Gamma(\cdot)$ the Gamma function, the modified Bessel function of the second kind, and $x^{(k)}$ the k-th dimension of input x. The parameter ν is a half-integer and controls the smoothness: a larger ν yields smoother functions.

To perform regression, the GP is conditioned on the data using Bayes' theorem to form a *posterior*. The posterior mean and variance at a test point x_t are given by:

$$\begin{aligned} \hat{m}(x_t) &= K(x_t,\, X) A\, y\,, \\ \hat{k}(x_t,\, x_t) &= K(x_t,\, x_t) - K(x_t,\, X) A\, K(X,\, x_t)\,, \end{aligned} \tag{6}$$

with

$$A = \left(K(X,\, X) + \sigma_n^2 I\right)^{-1}. \tag{7}$$

Here, $K(X, X)$ is the matrix constructed by evaluating the kernel k on all pairs of training inputs, and $k(x_t, X)$ is the vector of covariances between the test point and training inputs. The noise on the observations is captured by the trainable hyperparameter σ_n, which is multiplied with the identity matrix I in (7). This hyperparameter allows the model to distinguish real system fluctuations with random Gaussian noise which makes it more robust in real-life scenarios compared to e.g., vector fitting [13]. The posterior mean serves as the regression prediction, and the posterior variance quantifies the uncertainty.

This uncertainty-aware modeling approach makes GPs especially useful in settings where data is expensive or limited, such as engineering design and simulation-based optimization. The flexibility in kernel design also allows tailoring the model to suit the characteristics of the application domain, which will be explored in the remainder of this work.

B. Complex-Valued Gaussian Process Regression

Frequency-domain responses are complex-valued, so modeling them requires extending the real-valued GP to complex-valued functions. The core concept of complex-valued GP (CGP) models remains fairly similar. Let now $\mathcal{F}(x) : \mathbb{R}^d \to \mathbb{C}$. The definition of the mean function remains

$$\tilde{m}(x) = \mathbb{E}[\mathcal{F}(x)]\,, \tag{8}$$

but the covariance must now capture correlations between real and imaginary parts. This leads to the definition of the covariance $\tilde{k}$ and pseudo-covariance $\tilde{c}$ functions:

$$\begin{aligned} \tilde{k}(x,\, x') &= \mathbb{E}\left[\left(\mathcal{F}(x) - \tilde{m}(x)\right)\left(\mathcal{F}(x') - \tilde{m}(x')\right)^*\right], \\ \tilde{c}(x,\, x') &= \mathbb{E}\left[\left(\mathcal{F}(x) - \tilde{m}(x)\right)\left(\mathcal{F}(x') - \tilde{m}(x')\right)\right], \end{aligned} \tag{9}$$

where $(\cdot)^*$ represents the complex conjugation operator [14]. The mean, together with these two functions then fully specify the CGP model:

$$\mathcal{F}(x) \sim \mathcal{GP}\left(\tilde{m}(x),\, \tilde{k}(x,\, x'),\, \tilde{c}(x,\, x')\right). \tag{10}$$

In the remainder of this paper, the tilde symbol will be used to specify complex-valued variables or functions.

The mathematical framework for CGPs already exists for quite some time [15], [16]. However, direct implementations are rare in standard libraries. Instead, we adopt a real-valued multi-output GP (MOGP) representation [11], [12], where the real and imaginary parts of $\mathcal{F}$ are treated as two correlated outputs. A MOGP is similar to a GP but models two outputs simultaneously. This can be more efficient than modeling them separately when the two outputs are correlated. To convert (10)

into a MOGP, the covariance and pseudo-covariance functions are first rewritten as

$$\begin{aligned} \tilde{k} &= k_{\Re\Re} + k_{\Im\Im} + j(k_{\Im\Re} - k_{\Re\Im})\,, \qquad \text{and} \\ \tilde{c} &= k_{\Re\Re} - k_{\Im\Im} + j(k_{\Im\Re} + k_{\Re\Im})\,, \end{aligned} \tag{11}$$

where $k_{\Re\Re}$ and $k_{\Im\Im}$ are the covariances, and $k_{\Im\Re}$ and $k_{\Re\Im}$ are the cross-covariances of the real and imaginary parts of $\mathcal{F}$, respectively. After splitting the observations into their real and imaginary part, we can construct the MOGP mean,

$$\mathcal{M}(x) = \begin{pmatrix} m_{\Re}(x) \\ m_{\Im}(x) \end{pmatrix} \tag{12}$$

and covariance matrix,

$$\mathcal{K}(x,\, x') = \begin{pmatrix} k_{\Re\Re}(x,\, x') & k_{\Re\Im}(x,\, x') \\ k_{\Im\Re}(x,\, x') & k_{\Im\Im}(x,\, x') \end{pmatrix}, \tag{13}$$

with

$$\begin{aligned} k_{\Re\Re} &= \frac{\mathrm{Re}\{\tilde{k} + \tilde{c}\}}{2}\,, \qquad k_{\Re\Im} = \frac{\mathrm{Im}\{-\tilde{k} + \tilde{c}\}}{2}\,, \\ k_{\Im\Re} &= \frac{\mathrm{Im}\{\tilde{k} + \tilde{c}\}}{2}\,, \qquad k_{\Im\Im} = \frac{\mathrm{Re}\{\tilde{k} - \tilde{c}\}}{2}\,. \end{aligned} \tag{14}$$

Conditioning and prediction follow the same equations as in the real-valued case, applied to the two-output GP. This approach enables efficient and accurate modeling of complex-valued S-parameters while leveraging existing GP software.

III. Methodology

A. Rational Szegö Kernel

In this work, we use the physics-informed rational *Szegö kernel* to model complex-valued S-parameters. The choice of kernel (and pseudo-kernel) is based on the paper by Bect et al. [11]. It was shown there that the Szegö kernel is a reproducing kernel for the Hilbert space of complex holomorphic functions on a right half-plane Γ_α. Consequently, employing this kernel ensures that the GP output satisfies the Kramers-Kronig relations, which are inherent to frequency response functions. Additionally, the Szegö pseudo-kernel was constructed to incorporate Hermitian symmetry in the GP output, ensuring it represents a real-valued signal in the time domain. The Szegö kernel and pseudo-kernel are given by

$$\begin{aligned} \tilde{k}(s,\, s') &= \frac{1}{2\alpha + s + (s')^*}\,, \\ \tilde{c}(s,\, s') &= \frac{1}{2\alpha + s + s'}\,, \end{aligned} \tag{15}$$

where $s = j2\pi f$ is the Laplace variable (with f the frequency), and α is a trainable hyperparameter of the kernel, similar to the lengthscales mentioned in the previous section. However, in this context, α has a more physical interpretation, as it corresponds to the real part of a pole in a rational function.

B. Parametrized Rational Szegö Kernel with Input Warping

Optimizing devices based on their frequency response requires parametric modeling. The change in response caused by modifying design parameters $p \in \mathbb{R}^d$ can be captured by the Matérn kernel $k_\nu(p,\, p')$ described in Section II-A. To include this parametric dependence in the GP model, a *composite* covariance matrix is created by taking the Kronecker product of (13) and the covariance matrix generated by the Matérn kernel (4):

$$\mathcal{K}_c(X,\, X') = \mathcal{K}(s,\, s') \otimes k_\nu(p, p')\,, \tag{16}$$

where $X = \{x_{ij} = (f_i,\, p_j)\}$ is the extended vector containing all combinations of frequency and design samples.

Although this method of introducing design parameters is straightforward to implement, in practice it typically does not yield significantly better performance. The main reason for this is twofold. First, the data collection is highly *asymmetric*: for each design parameter p, a full frequency sweep is performed, resulting in many more frequency points than parameter points. This imbalance indirectly places a bias towards optimizing the hyperparameters of the Szegö kernel rather than those of the Matérn kernel. Second, the frequency response can change in a non-stationary manner with respect to the design parameters. Since the Matérn kernel is stationary, it performs poorly in capturing these variations.

Ulrick et al. [12] showed that an additional step is required to significantly improve modeling performance. They introduced input warping of the frequency variable. The idea behind this step is to allow the design parameters to directly affect the part of the kernel that models the frequency response. To capture a general compression, expansion, or shift of the frequency response function, a simple linear warping function is sufficient:

$$\varphi(f,\, p) = \nu^\top \cdot p + (1 + \gamma^\top \cdot p) f\,, \tag{17}$$

where the frequency is shifted by a parameter-dependent term $\nu^\top \cdot p$ and scaled by $1 + \gamma^\top \cdot p$. Here, $\nu,\, \gamma \in \mathbb{R}^d$ are vectors of hyperparameters that are trained by maximizing the log marginal likelihood of the model, similar to the lengthscales in (4).

This warping function provides a practical trade-off between modeling flexibility and computational complexity. It allows the GP model to express linear shifts and scaling (or dilation) of the frequency axis, which are common in microwave circuits as their design parameters vary. The approach is inspired by techniques in parametric macromodeling, particularly in complex vector fitting, where such transformations have been shown to effectively align resonant features across different parameter configurations [17].

While the proposed input warping significantly improves the problem of asymmetric data collection between frequency and design parameters, the transformation is still linear in p. This implies that more complex non-linear interactions or higher-order distortions in the frequency response induced by parameter changes are not captured. However, we find this choice sufficiently general for a wide range of practical devices and advantageous due to its low number of hyperparameters.

The final, warped composite covariance matrix is then given by

$$\mathcal{K}_{wc}(X,\, X') = \mathcal{K}\left(j2\pi\varphi(X),\, j2\pi\varphi(X')\right) \odot k_\nu(P, P')\,, \tag{18}$$

where X is the same vector as in (16), and

$$P = (\underbrace{p_1,\, p_1,\, \ldots,\, p_1}_{n_f},\, \ldots,\, \underbrace{p_N,\, p_N,\, \ldots,\, p_N}_{n_f})^\top \tag{19}$$

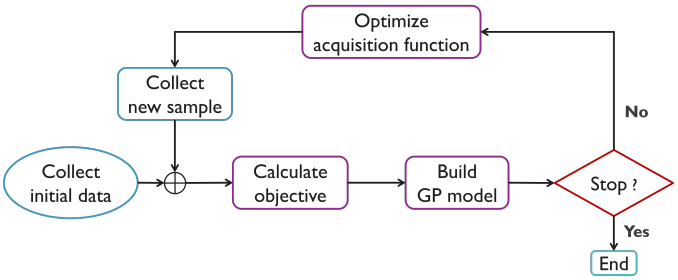


Fig. 1. Standard Bayesian optimization (BO) workflow: first some initial data is gathered. Next, the objective is calculated on that data, eliminating dependence on frequency. Then the Gaussian process (GP) is trained to model the objective function. If the stop condition is not met, a new sampling point is chosen by optimizing an acquisition function (AF).

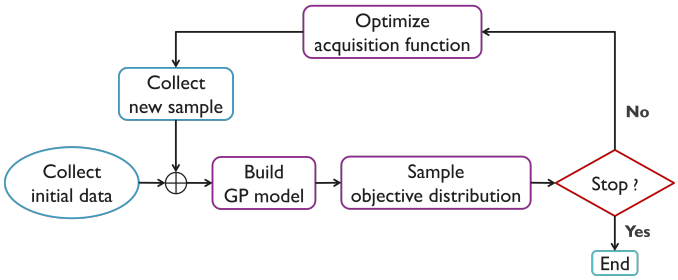


Fig. 2. Spectral Bayesian optimization (SBO) workflow. A Gaussian process (GP) model is trained on the complex-valued frequency response, taking both frequency and design parameters as inputs. An acquisition function (AF) selects the next design configuration based on the GP posterior. Unlike standard BO, the objective function is defined on the output of the GP and can be updated dynamically without retraining, enabling flexible and efficient design optimization.

is the vector of N design samples extended to the number of frequency samples n_f. Note that (18) does not contain the Kronecker product anymore, but simply uses a Hadamard product due to the extended vectors.

C. Standard Bayesian Optimization

In our current setup, we want to optimize the frequency response $S(f, p) \in \mathbb{C}$ which depends both on frequency f and design parameters p. When applying BO, we first start with gathering a small, random initial set of data $\mathcal{D} = \{(\boldsymbol{f}, p_i), \boldsymbol{S}_i\}_{i=0}^{N}$. Here, $\boldsymbol{f}$ and $\boldsymbol{S}_i$ are indicated in bold to denote that they are vectors, since each design configuration p_i yields a full sweep of the S-parameters at n_f discrete frequency points.

A standard BO scheme for MW/RF device optimization follows the workflow in Fig. 1. After gathering the initial dataset, an objective function $q(p)$ is defined to eliminate the frequency variable, since the goal is to identify an optimal design over a specified frequency range. This is typically done using a scoring function $g(S(f, p))$ that quantifies the deviation from a desired response at each frequency point. The overall objective is then computed as

$$q(p) = \sum_{k=1}^{n_f} g\big(S(f_k, p)\big)\,, \tag{20}$$

such that the optimal design parameters are given by

$$p_{\text{opt}} = \arg\max_{p} \{q(p)\} \tag{21}$$

The GP is then trained to model this objective as a function of the design parameters. Next, an *acquisition function* (AF) is optimized to select the next design to sample. The AF acts as an optimization policy to effectively balance exploration and exploitation. A fully exploitative algorithm would sample only near points that the model beliefs maximize the objective, whilst a purely exploratory algorithm would spread all samples evenly over the search domain. The AF balances this by taking the uncertainty of the GP predictions into account. The choice of AF depends on the type of optimization and the number of objectives. Common choices for single-objective optimization are the upper confidence bound (UCB), expected improvement (EI), and Thompson sampling, while the hypervolume expected improvement (HVEI) is often used in multi-objective settings [18]–[20]. Details on the selected AF are provided in Sections IV and V. Once the next design candidate p_{i+1} is identified, the S-parameters are simulated for this configuration and added to the dataset $\mathcal{D}$. This process continues until a predefined stopping criterion—such as a maximum number of simulations—is met.

D. Spectral Bayesian Optimization

In this work, however, we adopt the spectral BO (SBO) workflow introduced in [8] and illustrated in Fig. 2. In this approach, the GP model is trained directly on the complex-valued S-parameters instead of the objective, taking both frequency and design parameters as inputs—hence the name *spectral*. This formulation enables a more flexible and interpretable optimization process: since the GP captures the full frequency response, the objective function can be defined or adjusted dynamically without retraining the model, allowing for efficient human-in-the-loop (HITL) design iterations.

Moreover, by modeling the complex-valued response directly rather than separating magnitude and phase into distinct objectives, the GP can leverage known physical constraints such as the Kramers-Kronig relations, causality, and Hermitian symmetry. These relations intrinsically link the real and imaginary components of the response, and are naturally enforced through the use of the Szegö kernel. This physics-informed modeling approach improves data efficiency and generalization, particularly when the objective depends on both amplitude and phase characteristics of the device.

However, special care is required when applying SBO in practice. As discussed earlier, the AF selects new samples based on the predictive uncertainty of the GP model, which is characterized by its posterior distribution. While the GP outputs are, by construction, jointly Gaussian in the complex domain, this Gaussianity does not hold after nonlinear transformations. In particular, objective functions are often defined on measurable quantities such as magnitude, $|S| = \sqrt{\Re(S)^2 + \Im(S)^2}$, and phase, $\arg(S) = \arctan(\Im(S)/\Re(S))$. These nonlinear operations break the Gaussian structure, meaning the resulting objective no longer has a tractable closed-form distribution.

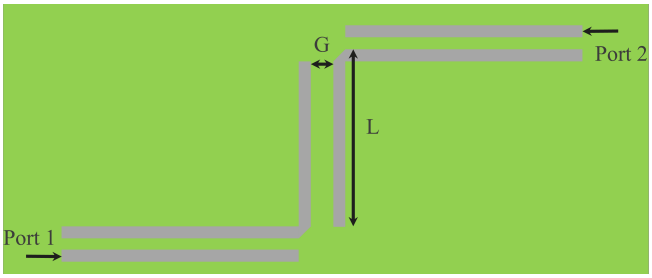


Fig. 3. Top view of the 2-port zig-zag microstrip bandpass filter. Gray lines indicate the metallic conductors, while the green background represents the dielectric substrate. The design parameters subject to optimization are the gap width G and the vertical conductor length L.

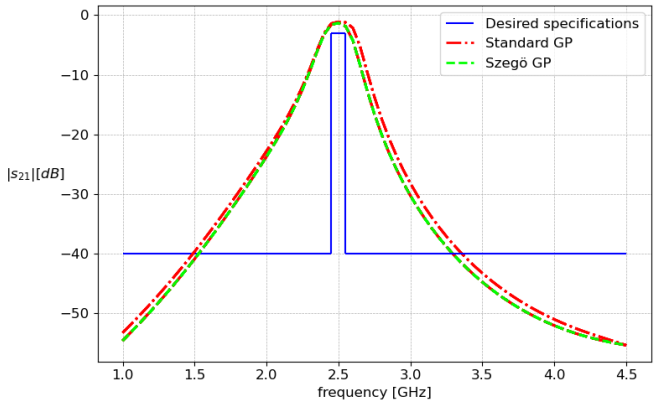


Fig. 4. Simulated best responses predicted by the standard (red) and Szegö (green) GP models after 15 Bayesian optimization (BO) iterations. Each BO run was repeated 10 times to ensure statistical robustness, yielding ten distinct optimized responses per model.

To address this, we employ Monte Carlo (MC) or Thompson sampling to approximate the distribution of the objective and evaluate the AF accordingly.

While the SBO workflow described above provides a general framework, specific implementation details—such as the number of initial samples, the definition of the objective function, and the choice and computation of the acquisition function—depend on the particular optimization task. Therefore, these aspects are discussed separately within each application section (Sections IV and V) to reflect their context-specific configurations and to ensure reproducibility. In both examples, the complete algorithm is implemented in Python using the BoTorch [21] library to optimize the AFs, and GPyTorch [22] to implement the warped composite kernel (18) and train the GP models.

IV. Application I: Zig-Zag Bandpass Filter

We first evaluate the proposed method on a 2-port zig-zag microstrip bandpass filter, previously studied in [8]. A top view of the layout is shown in Fig. 3. The device is fabricated on a dielectric substrate with a thickness of 0.5 mm, relative permittivity $\epsilon_r = 2.2$, and loss tangent $\tan\delta = 9 \times 10^{-4}$. The horizontal conductor sections have a fixed width of 0.4 mm, and the gaps between them are fixed at 0.3 mm.

The filter's geometry is parametrized by two design variables: the width of the vertical conductor gaps $G \in [0.3,\ 1.2]$ mm, and the length of the vertical conductors $L \in [5.0,\ 25.0]$ mm. The frequency response for each design is simulated using the ADS Momentum software [23] over $n_f = 71$ uniformly spaced frequency points in the range $[1.0,\ 4.5]$ GHz. To evaluate the model performance, a separate validation set of 500 randomly selected designs is generated using Latin hypercube sampling (LHS) and simulated over 141 frequency points.

The objective function is designed to obtain a bandpass response centered around 2.5 GHz. Specifically, the scoring function $g(S(f,\,p))$ is defined as

$$g(f,\,p) = \begin{cases} 10(|S_{21}(f,\,p)| - 0.708) & \text{if } f \in \Omega \\ 0.01 - |S_{21}(f,\,p)| & \text{else} \end{cases} \tag{22}$$

where the passband $\Omega = [2.45,\ 2.55]$ GHz. The values 0.708 and 0.01 correspond to transmission levels of -3 dB and -40 dB, respectively, thereby encouraging high transmission within the passband and strong attenuation outside of it. Note that in this example, the design parameters are given by $p = (G,\,L)$, and the objective is based solely on the magnitude of the transmission coefficient $|S_{21}|$. No constraints are imposed on the phase response.

Optimization is initialized with eight design samples selected to provide broad coverage of the design space: four correspond to the corners of the $(G,\,L)$ domain, while the remaining four are selected via LHS to maximize global exploration. The choice of optimization policy goes to Thompson sampling (TS). In this approach, a single sample S response is drawn from the GP posterior, and the objective is maximized based on this sample:

$$\tilde{S}(\boldsymbol{x}, f) \sim GP(f,\,p)\,, \tag{23}$$

$$\boldsymbol{x}_{i+1} = \arg\max_{\boldsymbol{x}} \{q(\tilde{S}(f,\,p))\}\,. \tag{24}$$

This method effectively optimizes a random AF at each BO iteration [4], significantly reducing computational complexity as there is no need for MC sampling. 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.

To benchmark our approach, we compare it against a standard GP model equipped with a Matérn $5/2$ kernel. This baseline model is trained only on the magnitude response, which is sufficient for the current objective. In contrast, the proposed Szegö GP models the complex-valued S-parameters directly and captures both the real and imaginary components, though only the magnitude is used in the scoring function. Both methods are run for 15 optimization iterations under identical starting conditions.

Performance is assessed using a regret metric that quantifies the discrepancy between the design predicted to be optimal by the surrogate model and the true optimal design identified from evaluation of the reference test set. Specifically, the regret is defined as

$$\text{regret} = \frac{\|p_{\text{opt}} - \tilde{p}_{\text{opt}}\|}{\|p_H - p_L\|}\,, \tag{25}$$

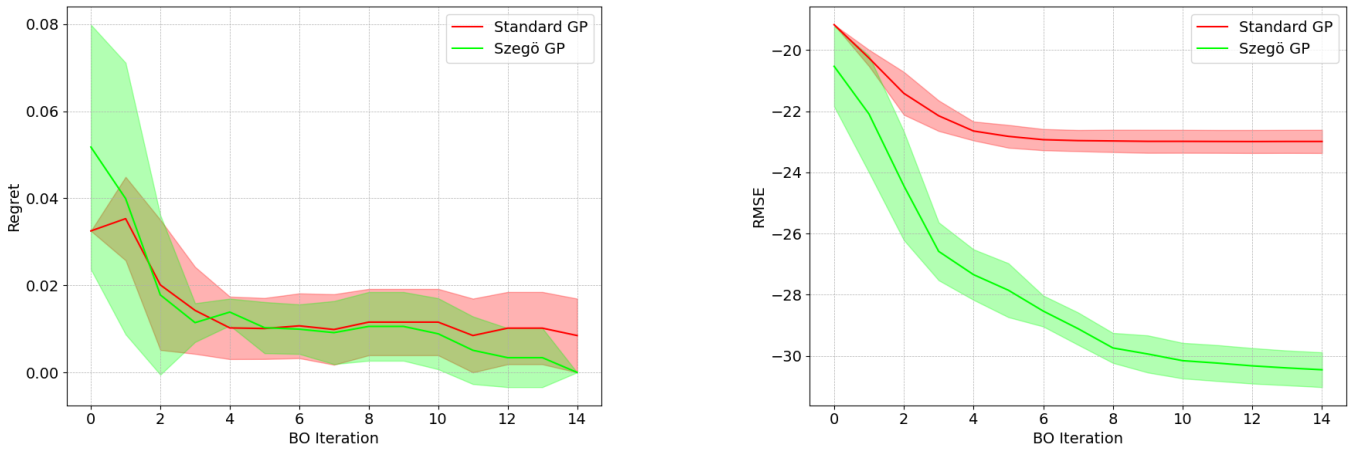


Fig. 5. Comparison of standard (red) and Szegö (green) GP models over 10 Bayesian optimization runs. (a) Regret metric as a function of the BO iteration. (b) Root-mean-square error (RMSE) as a function of the BO iteration. The shaded regions indicate the ± 1 standard deviation across the runs.

where $\|\cdot\|$ denotes the Euclidean norm, p_H and p_L are the upper and lower bounds of the design space, and $\tilde{p}_{\text{opt}}$ is the design predicted to yield the best objective according to the surrogate model. The true optimal design, found by evaluating all 500 test points, is given by $p_{\text{opt}} = (G_{\text{opt}}, L_{\text{opt}}) = (0.808, 18.325)$ mm. The regret thus provides a normalized measure of optimization accuracy, scaled to the size of the search space.

A. Results

BO is repeated ten times for each model to ensure statistical robustness. Fig. 4 presents the best responses identified by the standard (red) and Szegö (green) GP models after 15 BO iterations across the ten independent runs. The responses obtained using the Szegö GP consistently cluster around the desired bandpass response (blue), while those from the standard GP exhibit significantly more variation. This suggests that the Szegö GP more reliably identifies accurate optima.

This observation is further supported by Fig. 5a, which shows the average regret over BO iterations, with shaded regions indicating ± 1 standard deviation across the ten runs. While both models are able to converge towards the optimum—reflected by a zero regret value—the Szegö GP demonstrates consistently faster and closer convergence.

A more pronounced distinction between the models is visible in Fig. 5b, which shows the RMSE over the course of optimization. The Szegö GP achieves an RMSE below -30 dB after approximately ten iterations, whereas the standard GP does not improve beyond -24 dB. This result highlights the superior modeling capability of the Szegö GP. Such improvement is particularly valuable in applications requiring surrogate model flexibility—such as re-optimization under modified objectives or integration into human-in-the-loop frameworks.

To further illustrate this modeling advantage, Fig. 6 shows several predicted frequency responses for different design parameters. The GP models in these examples are trained using 71 equidistant frequency samples and 20 design parameter samples generated via LHS. Due to its lack of physical awareness, the standard GP is prone to producing unphysical predictions. For instance, Fig. 6b shows frequency response predictions from the standard model for various values of L at a fixed gap $G = 1$ mm, where it incorrectly predicts negative values for the magnitude. Such behavior does not occur with the Szegö GP due to its physics-informed construction.

Table I compares the computation times associated with each model at the beginning and end of the BO process. As expected, the Szegö GP exhibits significantly longer training times, primarily due to the increased kernel complexity and the absence of acceleration features such as lazy tensor evaluation or GPU training. However, this added cost yields clear gains in modeling accuracy, as discussed above. Moreover, the Szegö GP provides physically consistent and accurate surrogates near the optima, which is particularly valuable in downstream tasks such as sensitivity analysis, tolerance design, or interactive design frameworks.

V. Application II: Passive Differential Equalizer

In this example, we demonstrate the full potential of SBO by targeting not only the magnitude response but also the phase response of the device. This task showcases the advantage of directly modeling the complex-valued frequency response using a physics-informed GP. The second device under test is a passive equalizer embedded in a high-speed, differential interconnect, designed to enhance signal integrity (SI) in high-speed data transmission systems. At these high data rates, PCB

TABLE I
Computation times in seconds for standard GP and Szegö GP at iterations 1 and 15.

	Standard GP		Szegö GP	
	Iter 1	Iter 15	Iter 1	Iter 15
Training time	0.812	10.770	137.745	2136.503
Acquisition opt. time	2.177	5.716	14.595	44.060
Simulation time	25.671			

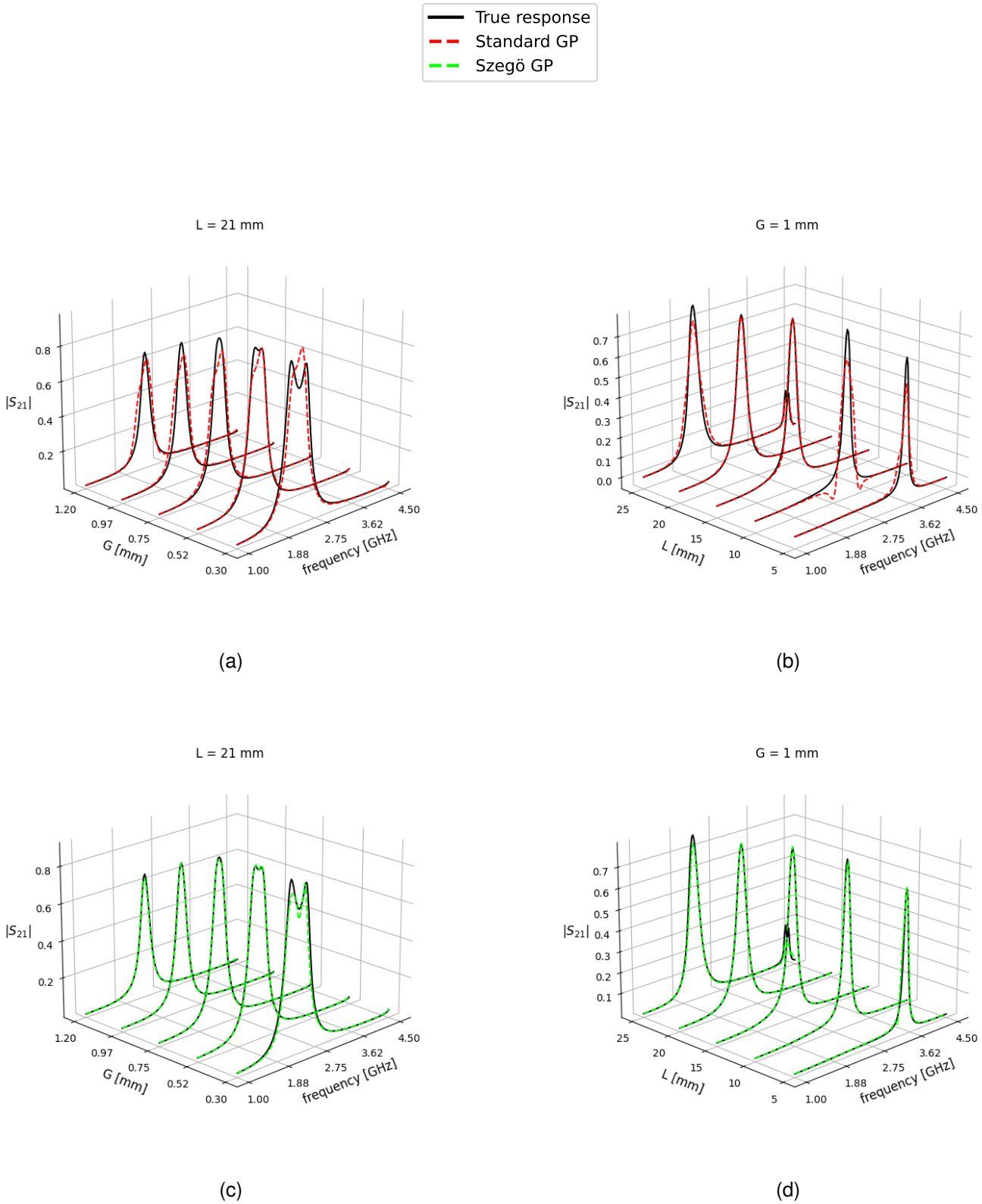


Fig. 6. Comparison of predicted transmission magnitude $|S_{21}(f, p)|$ from the standard and Szegö GP models against the true response (black) for select design parameter sweeps. (a) Standard GP predictions (red) as G varies over $[0.3, 1.2]$ mm with $L = 21$ mm fixed. (b) Standard GP predictions (red) as L varies over $[5.0, 25.0]$ mm with $G = 1.0$ mm fixed. (c) Szegö GP predictions (green) under the same G-sweep at $L = 21$ mm. (d) Szegö GP predictions (green) under the same L-sweep at $G = 1.0$ mm.

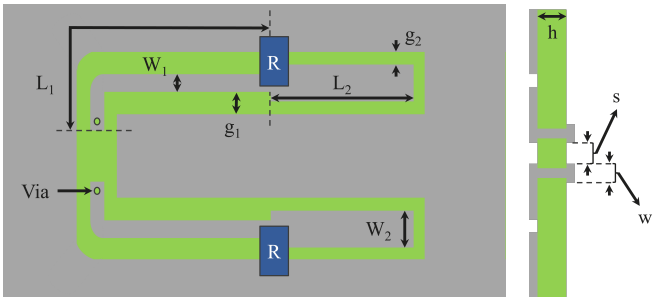


Fig. 7. Geometry of the passive differential equalizer: (a) Bottom view of the slot layout showing metallic layers (gray) and the dielectric substrate (green). Key features include the open-circuited stubs of length L_2 (the design parameter) and width W_2, transmission-line segments of length L_1 and width W_1, shunt resistors R, and the two via connections. (b) Cross-sectional view of the structure, illustrating the conductor width w, substrate thickness h, and spacing s. Component dimensions and electrical parameters are listed in Table II.

traces become electrically long, resulting in a detrimental low-pass effect caused by several types of frequency-dependent losses, e.g., the skin-effect, substrate, and radiation losses. This low-pass effect attenuates higher frequencies more than low frequencies, causing pulse distortion which limits the maximum achievable data rate. The equalizer mitigates this behavior by shaping the differential-mode S-parameters to achieve a flat magnitude response and a linear phase profile across the useful signal bandwidth.

Originally introduced by Pattyn et al. [24], the considered topology offers a passive implementation based on open-circuited stubs and shunt resistors. It attenuates low-frequency differential signals while allowing higher-frequency components to pass with relatively less attenuation, effectively flattening the overall frequency response. In addition, the equalizer integrates a common-mode filtering mechanism that reduces susceptibility to common-mode noise through resistive attenuation and disruption of the return path via a slot in the ground plane.

For this study, we optimize the equalizer's differential-mode performance over the frequency band $\Omega = [1.0,\ 5.0]$ GHz. Our dual objectives are to flatten both the magnitude and the group delay of the differential transmission coefficient $S_{dd21}(f,\ p)$. A flat magnitude response compensates for the channel's low-pass characteristics, while a constant group delay—equivalent to linear phase—minimizes inter-symbolic interference, thereby preserving high-speed pulse integrity at data rates up to 20 Gbps using a PAM-4 modulation scheme. A schematic of the equalizer slot layout is shown in Fig. 7. The equalizer is realized on an FR-4 substrate with a relative permittivity of $\epsilon_r = 3.9$, a loss tangent of $\tan\delta = 0.022$, and the microstrip lines at the top of the geometry are of length 220 mm and width 0.9 mm.

We formalize these goals via the multi-objective function

$$G\big(S_{dd21}\big) = \begin{bmatrix} g_1\big(S_{dd21}\big) \\ g_2\big(S_{dd21}\big) \end{bmatrix} = \begin{bmatrix} -\mathrm{Var}_{f\in\Omega}\Big(\big|S_{dd21}(f,\ p)\big|\Big) \\ -\mathrm{Var}_{f\in\Omega}\Big(\tau(f,\ p)\Big) \end{bmatrix}, \tag{26}$$

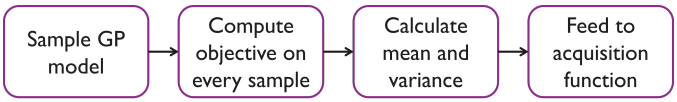


Fig. 8. Monte Carlo sampling workflow: first, 100 samples are drawn from the Gaussian process (GP) posterior. The multi-objective function is then evaluated for each sample, after which the mean and variance of the resulting distribution are computed. Finally, the distribution is approximated as Gaussian and passed to the acquisition function (AF).

where

$$\tau(f,\ p) = -\frac{\partial}{\partial\omega}\arg\big(S_{dd21}(f,\ p)\big)\,, \tag{27}$$

is the group delay, and $\mathrm{Var}_{f\in\Omega}(\cdot)$ denotes the empirical variance over the frequency range Ω. Here, $g_1\big(S_{dd21}\big)$ encourages a uniform attenuation profile, and $g_2\big(S_{dd21}\big)$ penalizes delay fluctuations. The sole design variable is the open-circuited stub length $p = L_2 \in [1.0,\ 20.0]$ mm; all other geometric parameters are fixed and listed in Table II.

While this example involves only a single design parameter, it serves as a meaningful test case to validate the proposed methodology in a multi-objective setting. Despite the low-dimensional input space, the simultaneous optimization of magnitude and phase characteristics presents a non-trivial challenge representative of real-world trade-offs. Extending the approach to multi-parameter scenarios is a natural next step, but this 1D case is sufficient to demonstrate the core strengths of our method, including the ability to model complex-valued responses accurately inside a design region of interest and handle multiple objectives effectively.

The initial training dataset comprises five design samples evaluated at 30 equidistant frequency points. These include the two boundary values of L_2 and three additional samples generated via LHS. The test dataset consists of 200 design samples, also generated via LHS, evaluated at 251 equidistant frequency points.

Since this example involves multi-objective optimization, a different acquisition function is employed. Specifically, the expected hypervolume improvement (EHVI) is used. At each iteration, the EHVI quantifies the *expected* gain in hypervolume if a new candidate design were to be evaluated. The hypervolume is a scalar metric representing the portion of the objective space dominated by the current Pareto front. The Pareto front comprises all non-dominated objectives—those for which no other designs exist that improve both objectives simultaneously. Unlike standard hypervolume improvement, EHVI accounts for the uncertainty in the surrogate model, enabling a principled trade-off between exploration and exploitation. Because the distribution of the predicted objectives is generally not available in closed form (as explained in Section III-D), MC sampling

TABLE II
DESIGN PARAMETER VALUES FOR PASSIVE EQUALIZER.

Parameter	Value	Parameter	Value	Parameter	Value
W_1	0.35 mm	g_1	1.0 mm	s	0.32 mm
W_2	1.0 mm	g_2	0.675 mm	h	1.0 mm
L_1	18 mm	R	45 Ω	w	0.9 mm

is required to approximate the EHVI. To this end, 100 samples are drawn from the GP posterior at each iteration. For each sample, the multi-objective function (26) is evaluated, and the resulting distribution is summarized by computing the sample mean and variance. These are then used to construct a Gaussian approximation of the objective distribution, which is fed to the EHVI AF. A schematic of this workflow is provided in Fig. 8. A detailed description of the BoTorch implementation of EHVI is provided in [25].

To evaluate optimization performance, the log hypervolume difference (LHVD) is used, analogous to the regret metric in single-objective settings. LHVD is defined as the common logarithm of the difference in hypervolume between the region dominated by the true Pareto front and that dominated by the approximate front identified during optimization. The true Pareto front is in this case determined by evaluating all test samples on (26).

A. Results

Fig. 9 plots the LHVD over 30 BO iterations, averaged across ten independent runs. The shaded band denotes the minimum and maximum LHVD observed at each iteration. When the approximate hypervolume matches the true hypervolume (evaluated on the test set), the LHVD is clipped to -10 to avoid undefined logarithms. The rapid decline of LHVD in most runs indicates that the BO procedure consistently approaches the true Pareto front.

This convergence is further illustrated in Fig.10, which overlays the sampled objective values (red markers) on the full test-set Pareto plot (blue markers). The Pareto front calculated on the test set is indicated with a green line. The proximity of the red points to the green line confirms that the algorithm effectively identifies high-quality trade-off solutions. A zoomed-in view in Fig.10b focuses on the Pareto set, clearly showing that the estimated Pareto front closely tracks the true front with minimal extraneous exploration of the objective space. Together, these results demonstrate that EHVI-guided sampling yields an accurate and efficient approximation of the Pareto boundary.

To illustrate the effectiveness of the optimization, Fig.11 compares the magnitude and group delay of the differential transmission coefficient S_{dd21} for an unoptimized design ($L_2 = 11.216$ mm) and an optimized design on the Pareto front ($L_2 = 18.189$ mm). The optimized sample demonstrates significantly reduced variation in both magnitude and group delay across the frequency band of interest, confirming the success of the SBO procedure in jointly flattening the transmission response and improving phase linearity.

Table III provides the training and acquisition optimization times for the Szegö GP model at the first and final optimization iterations. The average simulation time, approximately 355 seconds per evaluation, reflects the cost of full-wave EM simulations. Notably, even at iteration 30, the combined cost of model training and acquisition optimization remains on the same order of magnitude as the simulation time, underscoring the practical efficiency of the surrogate model relative to the high cost of direct evaluations. These values offer a realistic indication of the computational footprint of the proposed method when applied to multi-objective design tasks involving frequency-domain simulations.

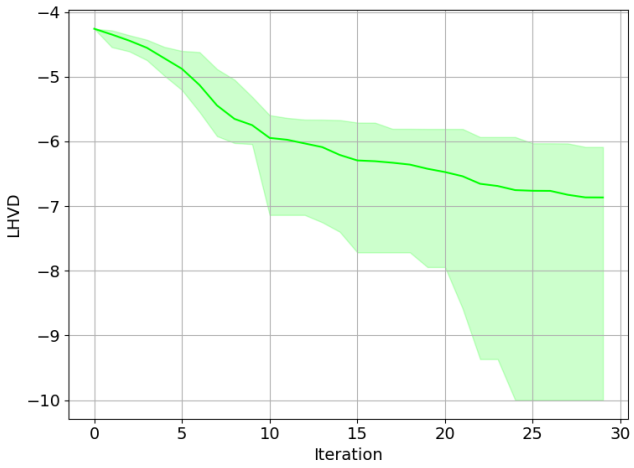


Fig. 9. Log hypervolume difference (LHVD) of the Szegö GP over 30 Bayesian optimization iterations. The solid line denotes the mean LHVD across ten independent runs, and the shaded band spans the minimum to maximum LHVD values observed. All zero hypervolume improvement values (i.e., when the approximate front coincides with the true front) are clipped to $1e-10$ to avoid undefined logarithms, corresponding to a lower bound of -10.

VI. Conclusions

The paper has presented a novel spectral Bayesian optimization (SBO) framework for microwave device design, leveraging a physics-informed Gaussian process (GP) with the rational complex-valued Szegö kernel and input warping. By modeling the full complex S-parameter response, the approach naturally enforces stability, causality, and Hermitian symmetry, yielding markedly improved accuracy and data efficiency of the surrogate model over standard (real-valued) GPs.

In the single-objective case of the 2-port zig-zag microstrip bandpass filter, the Szegö GP consistently converged to the true optimum in fewer iterations, achieving lower regret and RMSE than the Matérn $5/2$ baseline. In the multi-objective optimization of a passive differential equalizer, our approach efficiently approximated the true Pareto front—demonstrated by rapidly decreasing LHVD trajectories and close alignment of sampled points with the test-set front.

Together, these results demonstrate that the proposed SBO framework achieves both faster convergence and higher predictive accuracy across diverse optimization scenarios.

TABLE III
Computation times in seconds for Szegö GP at iterations 1 and 30.

	Iter 1	Iter 30
Training time	9.187	631.806
Acquisition opt. time	6.595	19.003
Simulation time	355.280	

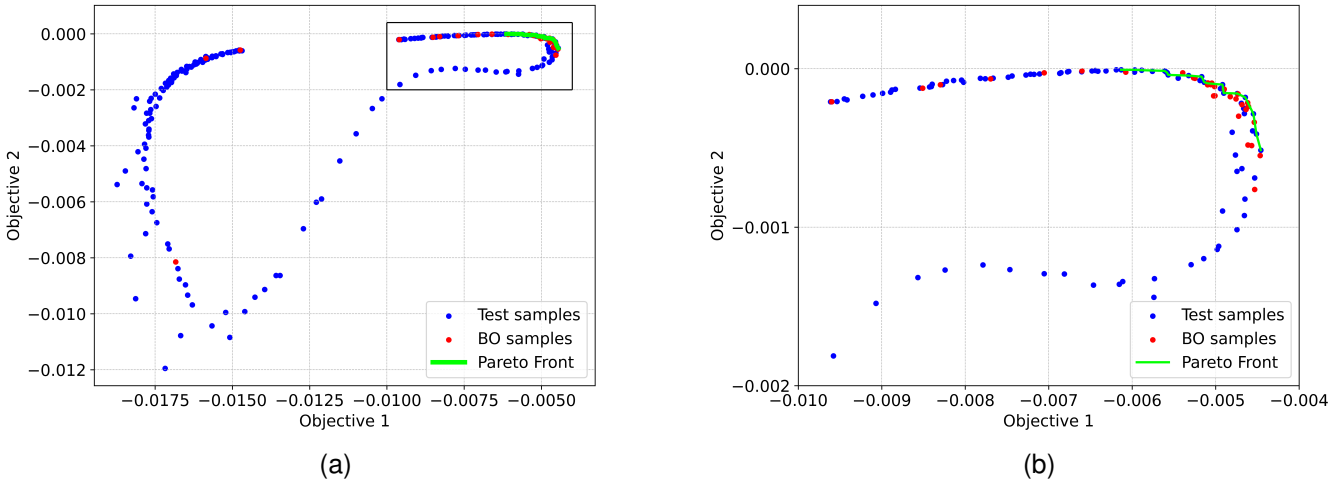


Fig. 10. Visualization of all sampled designs in the two-objective space during Bayesian optimization. (a) Full overview of all evaluated samples across the design space. (b) Zoomed-in view near the Pareto front. Blue dots denote the objective values of the 200 test designs, red dots indicate the samples selected during the optimization process, and the green line represents the Pareto front.

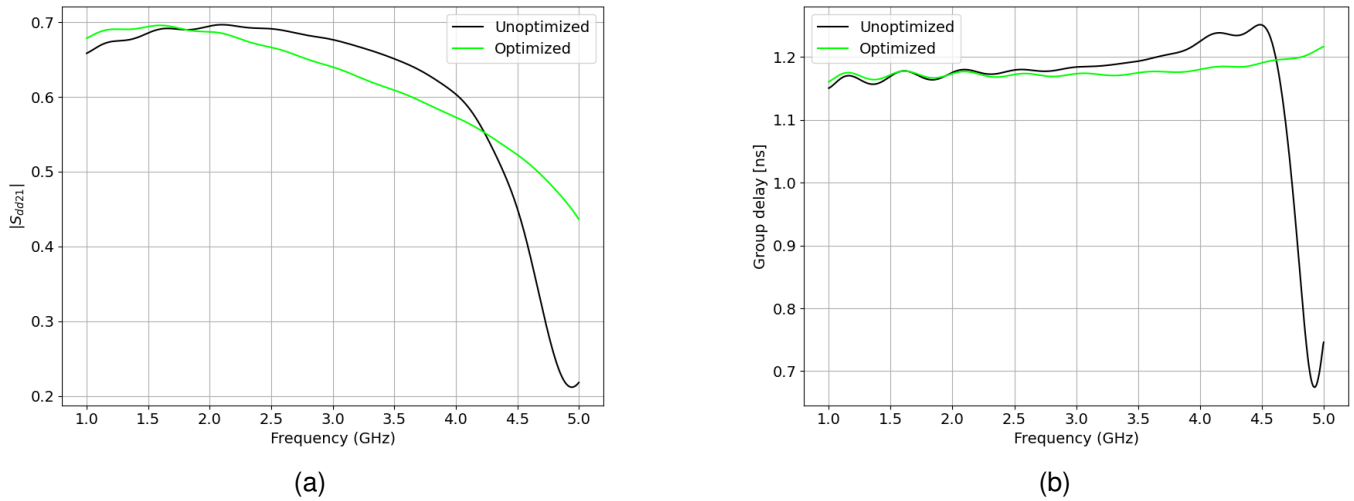


Fig. 11. Frequency response of the passive equalizer before and after optimization. (a) Magnitude of the differential transmission coefficient for an unoptimized design (black, $L_2 = 11.216$ mm) and an optimized design located on the Pareto front (green, $L_2 = 18.189$ mm). (b) Corresponding group delay profiles.

Beyond performance gains, the integration of prior physical knowledge into the GP surrogate offers a principled way to improve generalization and reduce simulation overhead.

Future work will extend this physics-informed SBO to feasibility region identification, by which the GP is trained to map out high-performance regions within the design space with high confidence. This extension could enable adaptive design and rapid re-optimization under shifting objectives and support interactive, human-in-the-loop design workflows across a wide range of microwave and RF components.

ACKNOWLEDGMENTS

The authors would like to thank Tim Pattyn of the Quest research group in IDLab at Ghent University, Belgium, for providing the equalizer example and its corresponding simulation model.

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] Y. Lindemans, T. Ullrick, I. Couckuyt, D. Deschrijver, and T. Dhaene, "Bayesian optimization of microwave filters: A physics-informed approach using the szegö kernel," in *2025 IEEE 29th Workshop on Signal and Power Integrity (SPI)*, pp. 1–4, IEEE, 2025.

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

[5] 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.

[6] D. De Witte, J. Qing, I. Couckuyt, T. Dhaene, D. Vande Ginste, and D. Spina, "A robust bayesian optimization framework for microwave circuit design under uncertainty," *Electronics*, vol. 11, no. 14, p. 2267, 2022.

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

[8] 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.

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

[10] 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. 114007, 2021.

[11] 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, pp. A3727–A3755, 2024.

[12] 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)*, pp. 1–3, IEEE, 2024.

[13] 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, no. 6, pp. 1715–1723, 2023.

[14] R. Boloix-Tortosa, J. J. Murillo-Fuentes, F. J. Payán-Somet, and F. Pérez-Cruz, "Complex gaussian processes for regression," *IEEE transactions on neural networks and learning systems*, vol. 29, no. 11, pp. 5499–5511, 2018.

[15] R. A. Wooding, "The multivariate distribution of complex normal variables," *Biometrika*, vol. 43, no. 1/2, pp. 212–215, 1956.

[16] K. S. Miller, "Complex gaussian processes," *Siam Review*, vol. 11, no. 4, pp. 544–567, 1969.

[17] 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, no. 1, p. 15407, 2023.

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

[19] D. J. Russo, B. Van Roy, A. Kazerouni, I. Osband, Z. Wen, *et al.*, "A tutorial on thompson sampling," *Foundations and Trends® in Machine Learning*, vol. 11, no. 1, pp. 1–96, 2018.

[20] K. Yang, M. Emmerich, A. Deutz, and T. Bäck, "Multi-objective bayesian global optimization using expected hypervolume improvement gradient," *Swarm and evolutionary computation*, vol. 44, pp. 945–956, 2019.

[21] M. Balandat, B. Karrer, D. R. Jiang, S. Daulton, B. Letham, A. G. Wilson, and E. Bakshy, "BoTorch: A Framework for Efficient Monte-Carlo Bayesian Optimization," in *Advances in Neural Information Processing Systems 33*, 2020.

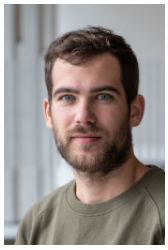
[22] J. R. Gardner, G. Pleiss, D. Bindel, K. Q. Weinberger, and A. G. Wilson, "Gpytorch: Blackbox matrix-matrix gaussian process inference with gpu acceleration," in *Advances in Neural Information Processing Systems*, 2018.

[23] *Advanced Design System*. (2022). Keysight. Accessed: December 5, 2024. [Online]. Available: http://www.keysight.com/find/eesof-ads.

[24] T. Pattyn, A. Moerman, M. Huynen, and D. Vande Ginste, "Differential interconnects with integrated equalization and common-mode filtering for broadband signal integrity enhancement in high-speed pam-4 signaling," in *2024 IEEE 28th Workshop on Signal and Power Integrity (SPI)*, pp. 1–4, IEEE, 2024.

[25] K. Yang, M. Emmerich, A. Deutz, and T. Bäck, "Efficient computation of expected hypervolume improvement using box decomposition algorithms," *Journal of Global Optimization*, vol. 75, pp. 3–34, 2019.

Yens Lindemans (Student Member, IEEE) received the M.Sc. degree in engineering physics from Ghent University, Ghent, Belgium, in 2023. He is currently pursuing a Ph.D. in electrical engineering at the Internet Technology and Data Science Lab (IDLab), Ghent University, Ghent, Belgium, working for the Surrogate Modeling Lab (SUMO Lab). His current research interests include machine learning for electronic design modeling and optimization.



Thijs Ullrick received the M.Sc. degree in electrical engineering from Ghent University, Ghent, Belgium, in 2018. He is currently pursuing a Ph.D. in electrical engineering at the Photonics Research Group and the Surrogate Modeling Lab (SUMO Lab), Ghent University, Ghent, Belgium. His current research interests include circuit theory, macromodeling, photonic integrated circuits, and machine learning for electronic and photonic design.

Ivo Couckuyt (Member, IEEE) received the M.Sc. degree in computer science from the University of Antwerp (UA), Antwerp, Belgium, in 2007, and the Ph.D. degree from the Internet Technology and Data Science Lab (IDLab), Ghent University, Ghent, Belgium, in 2013. In 2007, he joined the Computer Modeling and Simulation (COMS) Research Group. Since 2021, he has been an Associate Professor with the IDLab Research Group, Ghent University, working on automation in machine learning, data analytics, data-efficient machine learning, and surrogate modeling.

Dirk Deschrijver (Member, IEEE) obtained his Ph.D. degree in Computer Science from the Department of Mathematics and Computer Science at the University of Antwerp, Belgium, in 2007. From May to October 2005, he was a Marie Curie Fellow in the Scientific Computing group at the Eindhoven University of Technology, Eindhoven, The Netherlands. From 2008 to 2014, he worked as an FWO postdoctoral research fellow in the IDLab research group at Ghent University. In 2012, he obtained a second Ph.D. degree, in Engineering, from Ghent University. Since October 2014, he has been working as a senior researcher at iMinds/imec. Since October 2016, he has been an Associate Professor in the IDLab Research Group at Ghent University, focusing on data analytics, machine learning, and surrogate modeling algorithms.

Tom Dhaene (Senior Member, IEEE) obtained the M.Sc. and Ph.D. degrees in electrical engineering from Ghent University, Ghent, Belgium, in 1989 and 1993, respectively. Since 2000, he has been a Professor with the Computer Modeling and Simulation Research Group, University of Antwerp, Antwerp, Belgium. Since 2007, he has been a Full Professor with the Internet Technology and Data Science Lab (IDLab), Ghent University. His current research interests include data-efficient machine learning, surrogate modelling, Gaussian processes, Bayesian optimization, and system identification.