Bayesian Optimization for Microwave Devices Using Deep GP Spectral Surrogate Models

Federico Garbuglia, Domenico Spina, *Member, IEEE*, Dirk Deschrijver, *Senior Member, IEEE*, Ivo Couckuyt, and Tom Dhaene, *Senior Member, IEEE*

***Abstract*— In microwave design, Bayesian optimization (BO) techniques have been widely applied to the optimization of the frequency response of components and devices. The common approach in BO is to model and maximize an objective function over the design parameters, in order to find the optimal spectral response. Such an approach avoids the direct modeling of spectral responses, which is a challenging task for the typical data-efficient surrogate models used in BO. Simple objective functions may lead to a suboptimal solutions, while complicated objectives require more powerful and less data-efficient surrogate models. To resolve this issue, this article proposes to adopt a deep Gaussian process (DGP) to directly model all relevant *S* coefficients over the frequency and the design parameter ranges of interest. Subsequently, an objective probability distribution is retrieved from the DGP model and maximized using a BO scheme. The proposed approach is tested on two suitable microwave examples and compared to the standard BO approach. Results show increased accuracy in identifying the optimal frequency response for the given design parameters and the desired objective, while maintaining high data efficiency.**

***Index Terms*— Bayesian optimization (BO), deep Gaussian processes (DGPs), electronic design automation (EDA), *S* coefficients.**

I. Introduction

A COMMON problem in electronic design automation (EDA) is to find the value of the design parameters (i.e., width and length of metallic traces, the relative permittivity of a dielectric) giving the desired response for a simulated device under test (DUT). This is typically done via optimization algorithms, that are able to adapt the value of the design parameters in an automated framework, until the desired performances are reached. However, any change to the design parameters requires to execute a new simulation, which can be computationally expensive. This is particularly relevant for modern microwave devices, considering that bandwidth and circuit complexity are increasing, and full-wave simulations are usually required to correctly characterize the DUT. In order to reduce the number of required simulations and thus increasing the efficiency of modern optimization strategies, several techniques based on *surrogate models* have been proposed in the last decade [1]. A surrogate model is a suitable mathematical model that, once trained, can effectively replace the simulator by representing the DUT's response over the design parameters space. Given that evaluating a surrogate model is computationally cheap compared to physics-based simulations, many different design configurations can be rapidly tested by querying the surrogate model.

When plenty of simulated data is readily available, a high-complexity model can be trained, such that the optimal parameter combination is accurately predicted. In this case, optimization strategies can be built upon powerful machine learning (ML) models such as artificial neural network (ANN), which have been widely employed for modeling and optimization of microwave devices [2], [3]. Less computationally intensive surrogates include Gaussian process (GP) regression [4], [5] and Support Vector Machines [6].

On the other hand, when the amount of simulated data is limited, it is beneficial to update a surrogate model in a sequential fashion, as soon as new simulation results become available. Then, at each iteration, new simulated data is acquired based on an adequate sampling strategy. Several optimization techniques have been presented in literature based on such iterative model-building approach [7], [8], also known as *active learning* or *adaptive sampling*. In this contribution, we focus on Bayesian optimization (BO) [9], [10], [11], where a *stochastic* surrogate model is used to select consecutive simulations in order to find the optimal performance.

For linear and passive microwave devices, design specifications are typically expressed in terms of requirements on the DUT transfer function (i.e., impedance, admittance or scattering parameters). For example, consider a bandpass filter: a typical design goal is to obtain the desired attenuation in a suitable frequency range. Hence, in recent years, active learning has been applied to the optimization of scattering parameters (*S* responses). In this framework, optimization strategies often rely on the definition of an *objective function* assigning an aggregate score to any *S* response, based on the design requirements. Then, a surrogate model can be trained to predict the objective function for each set of design parameters in order to be used in an optimization framework.

Note that, if the objective function is defined as an aggregated measure over the entire frequency range of interest,

Manuscript received 14 April 2022; revised 30 September 2022; accepted 5 December 2022. Date of publication 22 December 2022; date of current version 5 June 2023. This work was supported in part by the Flemish Government through the AI Research Program and in part by the Fonds Wetenschappelijk Onderzoek (FWO) Program. *(Corresponding author: Federico Garbuglia.)*

The authors are with the Department of Information Technology, Ghent University–imec, 9052 Ghent, Belgium (e-mail: federico.garbuglia@ugent.be).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TMTT.2022.3228951.

Digital Object Identifier 10.1109/TMTT.2022.3228951

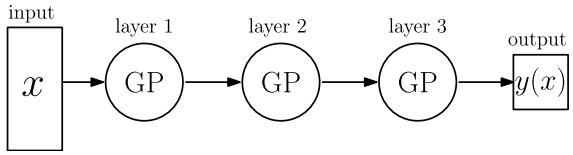

Fig. 1. Example of a DGP architecture.

it is not necessary to model *directly* the complex-valued and frequency-dependent scattering parameters of the DUT. This simplifies significantly the complexity of the function to be approximated via surrogate mode. An example of such an objective function is the maximum absolute error over the chosen frequency range between the desired S response and the response, obtained for the chosen design parameters. In this framework, relatively simple and efficient mathematical techniques can be adopted to accurately model the objective function. A popular choice in BO is the GP [12] model, thanks to its efficiency and flexibility. However, the definition of the objective function becomes crucial to obtain satisfactory optimization results: expressing the design requirements via simple objectives may lead to suboptimal performance of the DUT, while an objective function based on complex design specifications may require more powerful surrogate models than GPs in order to be modeled accurately.

In this article, a new strategy named spectral model BO (SMBO) is introduced for data-efficient optimization of S responses. The proposed approach exploits the increased modeling power of deep GPs (DGPs) [13] in order to directly learn the magnitude (or phase) of the scattering parameters, rather than modeling an objective function. Subsequently, S responses can be predicted by a DGP and ranked based on complex design specifications. The response ranking is executed by an *objective probability distribution* on the DGP's predictions. Finally, the objective distribution can be maximized following a Bayesian active learning strategy [15], in order to improve the model accuracy and to find the optimal design parameters. The rest of the contribution is organized as follows. In Section II, an overview of DGP characteristics and advantages over standard GP is available. Next, Section III presents a detailed description of the standard BO approach via objective function and the proposed SMBO technique, highlighting their differences. Two suitable application examples are presented in Sections IV and V to validate the new SMBO method, including a comparison between DGPs and standard GPs for the modeling of S responses over the frequency. Lastly, conclusions are drawn in Section VI.

II. Background on DGPs

Let $\boldsymbol{x}$ be a sample vector of variables and $y(\boldsymbol{x})$ a scalar function to be modeled. A GP represents any observed sample as a random Gaussian variable with specified mean $\boldsymbol{\mu}(\boldsymbol{x})$ and covariance matrix $\boldsymbol{K}(\boldsymbol{x}, \boldsymbol{x}')$, which contains the covariance between any two samples $\boldsymbol{x}$ and $\boldsymbol{x}'$. Consequently, given a new test sample $\boldsymbol{x}^*$, the probability distribution of $y(\boldsymbol{x}^*)$ conditioned to the available samples is Gaussian as well (also known as *posterior*). This property allows one to analytically compute the mean and the variance of the posterior, by applying GP regression [12], which exploits the principle of Bayesian inference. The obtained mean and variance represent the predicted value and the confidence interval of the GP, respectively. In ML, many popular models are parametric: they represent the observed data as a function of trainable parameters of high, prefixed, cardinality. ANNs are a typical example. On the contrary, GP is a non-parametric model: the observed data is fully expressed as realizations of random Gaussian variables by the mean and the covariance matrix. Thus, the GP prediction can be efficiently computed from relatively few data samples. However, GP modeling may become intractable for a high amount of data, since it requires the inversion of an initial covariance matrix $\boldsymbol{K}$ whose size grows as the square of the number of the data samples. This problem is exacerbated by the curse of dimensionality: functions y that are defined over more input dimensions require increasingly more data to be represented accurately. In order to mitigate this intractability, the sparse GP (SGP) has been introduced [17].

The main disadvantage of standard GP and SGP models is their poor accuracy in modeling discontinuous functions or functions that are generated by a nonstationary stochastic process. Indeed, these properties can only be enforced by the user via a specific covariance function design. Therefore, high inaccuracy may occur when no prior knowledge is available on the stationarity or the discontinuity in y. This is often the case for response functions of engineering systems that operate in different regimes, depending on the input variables [16].

To solve this problem, the DGP model has been introduced [13], [14], which is able to capture local properties of the function by using a composition of multiple GPs. In fact, in its simplest form, the DGP is composed by a stack of GPs, from the input variables to the output. Each intermediate GP, except the last, constitutes a hidden layer that produces a latent encoding of the input variables. Fig. 1 shows an example of a DGP architecture. Similar to SGP, the variational inference is applied in the DGP, allowing the training with a high amount of data. In fact, the output distribution at each intermediate layer is approximated via inducing points.

Being non-parametric, the DGP model preserves high data efficiency. Furthermore, its composite structure allows for higher modeling accuracy for complicated functions. However, since the DGP posterior is not analytical, its mean and variance can only be approximated statistically, by recording multiple model predictions for each input value. Therefore, the DGP uncertainty on the prediction is more computationally expensive than a simple GP or SGP. The interested reader is referred to [13] and [14] for a detailed description of the DPG properties.

To the best of our knowledge, in this contribution, DGPs are employed for the first time in microwave device optimization as surrogate models for S responses over the design parameter and the frequency. In fact, the previously described properties enable an excellent trade-off between model complexity and data efficiency.

III. Methodology

As mentioned in the Introduction, surrogate models have been developed to speed up microwave devices optimization by replacing expensive simulations. In particular, stochastic

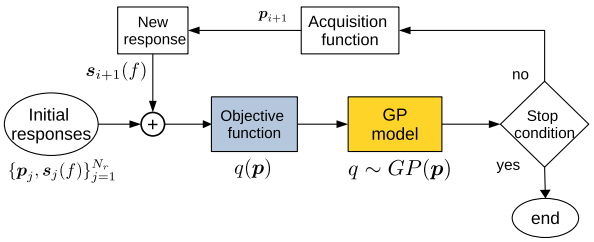

Fig. 2. Standard BO scheme using an objective function model over the design parameters.

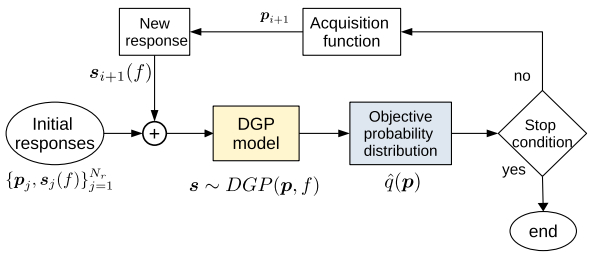

Fig. 3. Proposed SMBO scheme based on a DGP model of the S response over the frequency.

models such as GP and DGP offer an additional advantage: they both provide an expected value (mean) and a confidence interval (variance) for the modeled quantity. This information is crucial in the execution of optimization schemes based on Bayesian inference, as described in detail in the following paragraph.

A. BO via Objective Function Modeling

Let $\boldsymbol{p} = [p_0, \ldots, p_{n_p}]$ be the vector containing specific values of n_p design parameters for the microwave device under study. Additionally, let f be the discrete frequency value at which the scattering parameters of the device are simulated. Each S response, i.e., the magnitude or phase of an element of the scattering matrix, can be considered a function s over the design parameters and the frequency: $s(\boldsymbol{p}, f)$. Then, an S response is constituted by all the samples of s acquired at different frequencies for the same values of $\boldsymbol{p}$. Using this notation, an objective function q can be defined over the space of all the possible $\boldsymbol{p}$ values

$$q(\boldsymbol{p}) = q(s(\boldsymbol{p}, f)). \tag{1}$$

The purpose of the objective function is to assign a single score to any S response. In this way, the design parameters $\boldsymbol{p}_{\text{opt}}$ corresponding to the optimal response can be identified by solving a simple minimization or maximization problem. For example, minimization can be adopted when the objective function is defined as an error measure (i.e., optimal performance is characterized by minimum error), while a maximization problem can be solved when the objective function represents the desired performance measure (i.e., transmissivity of a filter in the passband). Without loss of generality, let us focus on the formulation of the optimization as a maximization problem

$$\boldsymbol{p}_{\text{opt}} = \arg\max_{\boldsymbol{p}} q(\boldsymbol{p}). \tag{2}$$

In order to solve (2), a Bayesian strategy typically follows the scheme in Fig. 2: a stochastic surrogate model reproduces the objective function over the design parameters, then the maximum of the objective is identified using the surrogate. The first step of the strategy is to collect a small initial set of S response samples (N_r). Second, an objective function q is computed for each response: a set of samples of the objective function is obtained. Third, a surrogate model is trained to reproduce q. By employing a GP model: $q(\boldsymbol{p}) \sim \text{GP}(\boldsymbol{p})$. At this point, the model can predict the value of the objective function for any combination of design parameters. Next, the optimization process is halted if a stop condition is verified. Otherwise, an acquisition function is computed. In a BO framework, the acquisition function α is defined in order to select the next sample of the design parameters $\boldsymbol{p}_{i+1}$ to be evaluated, which corresponds to the value of the design parameters that is most likely to lead to the optimal solution. In particular, this is performed by solving the following problem:

$$\boldsymbol{p}_{i+1} = \arg\max_{\boldsymbol{p}}[\alpha(q(\boldsymbol{p}))]. \tag{3}$$

Then, a new S response $s(\boldsymbol{p}_{i+1}, f)$ is simulated for selected design parameters and it is added to the initial set of samples. These steps are iterated until the stop condition is verified. At each iteration, the $\boldsymbol{p}_{\text{opt}}$ can be easily estimated by applying a strategy based on Monte Carlo [18] sampling of the GP model, which is computationally efficient to evaluate. During this process, the accuracy of the surrogate model is likely to improve for parameters values that are close to the maximum of the objective function.

Note that in the described optimization strategy, the performance of the surrogate model is strongly dependent on the definition of the objective function. In fact, the objective function should be sufficiently simple to model, allowing for an accurate prediction and a fast convergence to the maximum. However, in practice, the objective function must account for several design specifications on the S response. These can be imposed by defining multiple objective functions and then executing a multiobjective optimization algorithm [19], [20]. However, since it exploits multioutput models, this technique does not return a unique solution, but a Pareto set of possible solutions. Alternatively, multiple objectives can be formulated into a single function before modeling, using scalarization [21]. For example, multiple design objectives can be combined into a single objective function via a user-defined weighted sum. Nonetheless, scalarized objective functions are typically more complex than their single objective constituents. Thus, their maximization may require more complicated surrogate models and more sampling iterations.

B. Spectral Model BO

In this section, the new SMBO strategy is developed. The proposed approach overcomes the previously described trade-off between the multiple S response requirements and the objective function's complexity. In fact, the surrogate modeling of the objective function is avoided altogether, by following the scheme in Fig. 3. In this strategy, like in standard BO,

a surrogate model is sequentially updated and the next value of the design parameters to be evaluated $\boldsymbol{p}_{i+1}$ is selected by an acquisition function. However, in this case, a DGP model is used to directly learn the S response values as a function of the design parameters and the frequency: $s(\boldsymbol{p}, f) \sim \mathrm{DGP}(\boldsymbol{p}, f)$. Next, an objective value is assigned to each possible S response predicted by the DGP. In standard BO, this task is performed by the objective function q. However, using q as a deterministic function on the model prediction is not sufficient in this case. In fact, the subsequent acquisition function requires a confidence interval on every objective value. For this purpose, a new objective probability distribution $\hat{q}$ is hereby derived.

First, let g be a goal function over the S response, which is linear for any test frequency in a user-defined set T_f

$$g(s, f) = a(f)s + b(f); \quad a(f), b(f) \in \mathbb{R} \quad \forall f \in T_f \tag{4}$$

where $a(f), b(f)$ are frequency-dependent linear coefficients. These coefficients can be selected according to desired device specifications, such that higher values of g correspond to better values of s. For example, g can represent the distance between the S response and a lower (or upper) limit given by the specifications. Then, by computing g on the DGP model

$$g(\boldsymbol{p}, f) = g(\mathrm{DGP}(\boldsymbol{p}, f), f); \quad f \in T_f. \tag{5}$$

Since the values predicted by the DGP obey to a Normal distribution and g is a linear function for any test frequency, then the g values are also normally distributed

$$\mathbb{E}[g(\boldsymbol{p}, f)] = g(\mathbb{E}[\mathrm{DGP}(\boldsymbol{p}, f)], f); \quad f \in T_f \tag{6a}$$

$$\mathrm{Var}[g(\boldsymbol{p}, f)] \propto \mathrm{Var}(\mathrm{DGP}(\boldsymbol{p}, f), f); \quad f \in T_f \tag{6b}$$

where Var is the variance operator. Next, by summing up all the values of g across all the test frequencies, the new objective probability distribution is obtained

$$\hat{q}(\boldsymbol{p}) = \sum_{f \in T_f} g(\boldsymbol{p}, f). \tag{7}$$

Since the $\hat{q}$ is a sum of random Gaussian variables, its mean and variance can be computed analytically

$$\mathbb{E}[\hat{q}(\boldsymbol{p})] = \sum_{f \in T_f} (\mathbb{E}[g(\boldsymbol{p}, f)]) \tag{8a}$$

$$\mathrm{Var}[\hat{q}(\boldsymbol{p})] = \sum_{f \in T_f} \sum_{f' \in T_f} \mathrm{Var}[g(\boldsymbol{p}, f), g(\boldsymbol{p}, f')] \tag{8b}$$

where Cov is the covariance operator for each pair of test frequencies $f, f' \in T_f$. The mean and variance of the objective distribution provide an objective estimation and a confidence interval for each combination of design parameters $\boldsymbol{p}$, respectively.

Finally, similar to (3), the acquisition function can be computed on $\hat{q}$, in order to select the value of the design parameters $\boldsymbol{p}_{i+1}$ for next simulation. As a result, the SMBO aims at maximizing the expectation of $\hat{q}$

$$\boldsymbol{p}_{\mathrm{opt}} = \arg\max_{\boldsymbol{p}} \mathbb{E}[\hat{q}(\boldsymbol{p})]. \tag{9}$$

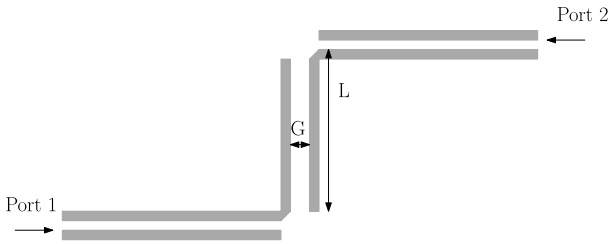

Fig. 4. Application I: Geometry of the microstrip bandpass filter (top layer).

From another perspective, the combination of DGP and g is analogous to a multiobjective model, where $g(\boldsymbol{p}, f)$ is a collection of many objective functions over $\boldsymbol{p}$. Then, the definition of $\hat{q}$ in (7) can be interpreted as a scalarization of g objective functions over the frequency. However, in contrast with the modeling of scalarized objectives, the mutual information between different frequencies is not lost, but it is learned by the spectral surrogate model and combined in the objective distribution.

Compared to the standard BO described in Section III-A, the new strategy adds another input dimension to the surrogate model: the frequency. Consequently, higher training and prediction times are necessary in each iteration for the computation of the objective distribution. However, the advantage of SMBO is that the surrogate model is independent of the choice of the objective function since the S response is modeled instead. Therefore, the objective definition can be changed without having to build again the spectral model from scratch. This is beneficial in the human-in-the-loop (HITL) optimization scenario, or when the chosen objectives need some fine-tuning.

Furthermore, no assumption is made on the test frequencies T_f, since they should be selected according to the design specifications. Thus, g can be calculated at frequencies that are different from the simulated data samples, thanks to the interpolation operated by the surrogate. Then, s values can also be acquired at different frequencies for each parameter combination p, without modifying the objective definition. This enables the usage of adaptive frequency sampling schemes [22] in the simulation software, thus further accelerating data acquisition during the model training.

IV. Application I: Zig-Zag Bandpass Filter

The proposed SMBO methodology is tested on a microstrip bandpass filter [23]. The material chosen as substrate is 0.5 mm thick, with relative permittivity $\epsilon = 2.2$ and loss tangent $\tan\delta = 0.0009$. The top layer geometry of the DUT is shown in Fig. 4. The gap among the horizontal conductors is 0.3 mm, while all the conductors have a width of 0.4 mm. This structure results in a very narrowband response. The S matrix of the DUT is computed via the advanced design system (ADS) Momentum simulator [24] for a set of 71 equispaced frequencies $X_f \in [1, 4.5]$ GHz. In this example, the chosen design parameters are the length of the vertical conductors $L \in [5, 25]$ mm and gap between them $G \in [0.3, 1.2]$ mm. Before executing the optimization, the desired filter response needs to be specified. For this example, a bandpass behavior in the interval $[2.45, 2.55]$ GHz is imposed by the following

specifications on the magnitude of the s_{21} coefficient:

$$|s_{21}(\boldsymbol{p}, f)| \leq 0.01, \quad \text{for } f < 2.45 \text{ GHz} \tag{10a}$$
$$|s_{21}(\boldsymbol{p}, f)| \geq 0.708, \quad \text{for } f \in [2.45, 2.55] \text{ GHz} \tag{10b}$$
$$|s_{21}(\boldsymbol{p}, f)| \leq 0.01, \quad \text{for } f > 2.55 \text{ GHz} \tag{10c}$$

where $\boldsymbol{p} = [G, L]$ is the vector of parameters, while f is any frequency in X_f. Note that 0.01 and 0.708 correspond to -40 and -3 dB, respectively. Then, the goal function $g(\boldsymbol{p}, f)$ can be defined as the distance between $|s_{21}|$ and the limit values of the specifications

$$g(\boldsymbol{p}, f) = 0.01 - |s_{21}(\boldsymbol{p}, f)|, \quad \text{for } f < 2.45 \text{ GHz} \tag{11a}$$
$$g(\boldsymbol{p}, f) = 10(|s_{21}(\boldsymbol{p}, f)| - 0.708) \quad \text{for } f \in [2.45, 2.55] \text{ GHz} \tag{11b}$$
$$g(\boldsymbol{p}, f) = 0.01 - |s_{21}(\boldsymbol{p}, f)|, \quad \text{for } f > 2.55 \text{ GHz}. \tag{11c}$$

Using this definition, the value of $g(\boldsymbol{p}, f)$ increases when the magnitude of s_{21} is closer to one in the passband [2.45, 2.55] GHz, and also when the magnitude of s_{21} is closer to zero elsewhere. In particular, the weighting factor 10 in (11b) is introduced to assign higher importance to the in-band response of the filter. Therefore, maximizing $g(\boldsymbol{p}, f)$ at any simulated frequency means finding the values of the magnitude of s_{21} that are closest to an ideal bandpass response. While reaching an ideal response is not feasible, in practice, this choice allows us to clearly illustrate the performance of the proposed methodology. In practical usages, g can be modified by the designer to individuate more realistic responses, as shown in Section V. Indeed, the definition of g can be changed according to the design specifications of the problem at hand, as long as it satisfies (4).

Then, both the standard BO (see Section III-A) and the new SMBO strategy (see Section III-B) are executed to find the optimal s_{21} response using a GP and a DGP model, respectively. In the former, the chosen objective function is

$$q(\boldsymbol{p}) = \sum_{f \in X_f} g(|s_{21}(\boldsymbol{p}, f)|, f), \quad \text{for } f \in X_f. \tag{12}$$

Instead, the objective probability distribution in the SMBO is

$$\hat{q}(\boldsymbol{p}) = \sum_{f \in T_f} g(\text{DGP}(\boldsymbol{p}, f), f), \quad \text{for } f \in T_f \tag{13}$$

where T_f is a set of 36 equispaced test frequencies in the range [1, 4.5] GHz. Note that, if both models are accurate, the objective function prediction in BO should correspond to the objective distribution prediction in SMBO

$$\mathbb{E}[\text{GP}(\boldsymbol{p})] \sim \mathbb{E}[\hat{q}(\text{DGP}(\boldsymbol{p}, f))]. \tag{14}$$

This choice of q, g, and $\hat{q}$ is not unique, but it enables a consistent comparison of the two techniques. It is worth noticing that these objectives do not impose constraints on the optimization: the corresponding optimal S response may still violate the initial design constraints.

Then, for both strategies, the upper confidence bound (UCB) [25] is selected as the acquisition function, due to its simplicity in interpretation. In addition, a squared exponential (SE) covariance matrix, also called kernel, Wilson and Adams [26] is used for both GP and DGP models. This covariance assumes high local smoothness in the modeled function, which is typical of S responses. In particular, the DGP architecture used in this example for the SMBO is as follows: three stacked GP layers, 50 inducing points at each layer, and the training is executed with an Adam optimizer for 300 iterations and a learning rate of 0.01. Similar to the ANN hyperparameters, these settings are identified with a trial and error approach, by testing the model on the initial samples.

A. Results and Discussion

BO and SMBO are executed for 30 iterations, starting from five different sets of initial S responses, collected by running ADS Momentum simulations on ten combinations of the design parameters. These parameter values are chosen randomly, according to a Latin hypercube design (LHD) [27]. This allows one to evaluate the robustness of the optimization strategies to the choice of the initial samples. Based on a validation set of 1000 (G, L) samples, which are also chosen according to an LHD, it is found that the design parameters corresponding to the highest objective value are $\boldsymbol{p}_* = [0.790, 18.678]$ mm.

To validate the performance of BO and SMBO, a large set of simulations is performed in order to estimate the value of the design parameters giving the best performance for the problem at hand, indicated as $\boldsymbol{p}_*$. In particular, a set of 1000 (G, L) samples chosen according to an LHD is evaluated (referred to as validation set in the following), leading to $\boldsymbol{p}_* = [0.790, 18.678]$ mm.

Then, after each iteration of BO and SMBO, the normalized regret is computed

$$\text{regret}_i = \frac{||\boldsymbol{p}_{\text{opt},i} - \boldsymbol{p}_*||}{||\boldsymbol{p}_H - \boldsymbol{p}_L||} \tag{15}$$

where i is the iteration index, $\boldsymbol{p}_{\text{opt},i}$ are optimal parameters predicted by the models, while $\boldsymbol{p}_H$ and $\boldsymbol{p}_L$ are vectors containing the higher and the lower bounds of the parameter ranges, respectively; $||\cdot||$ is the Euclidean norm. As a result, the normalized regret in (15) ranges from 0 to 1. The previous metric is averaged across five different runs of the strategy, one for each set of initial samples. The resulting regret at each iteration is reported in Fig. 5. In addition, Fig. 6 reports the best among the simulated S responses, for each run of the two strategies. These correspond to the highest objective value found during each run. It is apparent that the two strategies identify similar optimal responses, which are correctly centered around the requested passband [2.45, 2.55] GHz. Thus, the optimization results of SMBO are equivalent to the standard BO for sufficiently simple objectives.

The previous observations are confirmed by Fig. 7, which compares the prediction of the BO model of the objective function with the prediction of the objective distribution in SMBO, after 30 iterations. In this figure, the black dots represent all objective samples gathered during the optimization. Both BO and SMBO focus the simulated samples near the maximum (green dot), where they produce a similar prediction for the objective values.

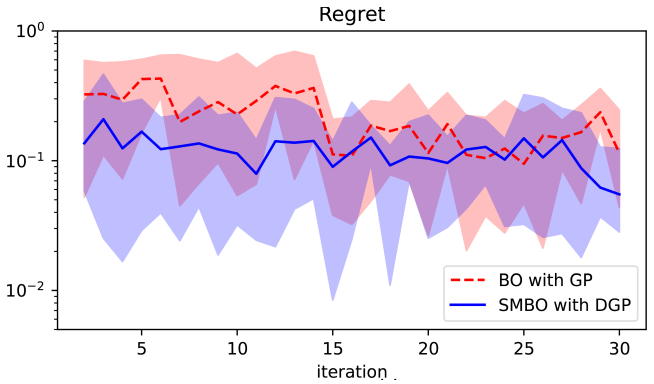

Fig. 5. Application I: Regret for different optimization strategies over increasing optimization iterations. Regret is computed for a validation set of 1000 $|s_{21}|$ responses and averaged across five complete runs, with different initial samples. The shaded areas represent the range between the minimum and maximum values recorded across the runs.

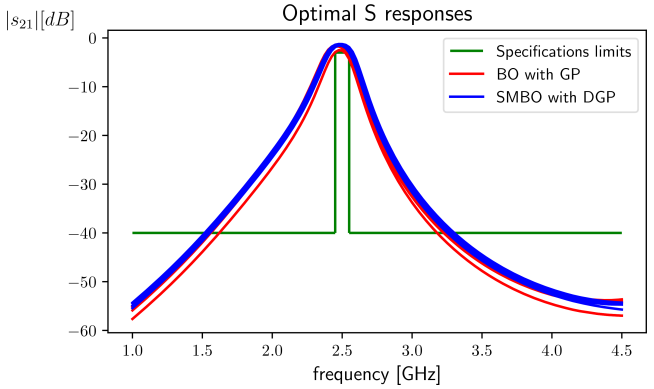

Fig. 6. Application I: Optimal S-responses identified with five different runs of both optimization strategies, after 30 iterations.

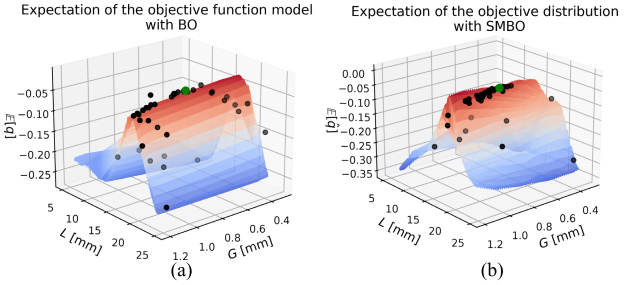

Fig. 7. Application I. (a) Expectation of the objective function model in the standard BO strategy using a GP. (b) The expectation of the objective distribution in the proposed SMBO strategy using a DGP model. Black dots represent all the simulated samples of design parameters during the optimization, while the green dot corresponds to the true optimum $\boldsymbol{p}_* = [0.790, 18.678]$.

Further tests on the BO strategy show that substituting the standard GP with SGP does not improve the identification of the optimum in this application example. Hence, the results obtained via SGP are not reported in Figs. 5–7, for simplicity.

It is important to remark that standard GPs could not be used reliably as spectral models in the proposed SMBO. In fact, GPs are not sufficiently powerful to reproduce the S responses over the frequency. This often causes numerical errors in the kernel matrix inversion, especially for high amount of training data. As an illustration, Fig. 8 compares projections of a GP and a DGP spectral model, trained on 50 $\boldsymbol{p}$ samples and 71 f samples, with $\boldsymbol{p}$ chosen from an LHD design and $f \in X_f$.

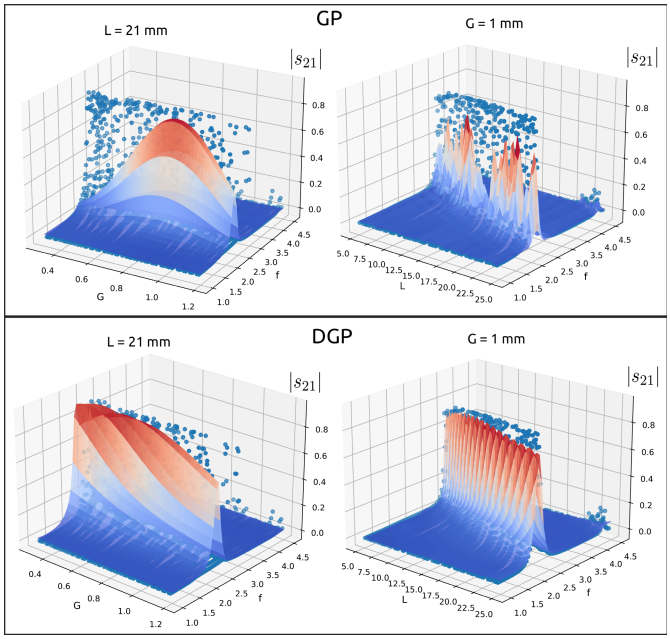

Fig. 8. Application I: Comparison of $|s_{21}|$ values predicted by GP (top) and DGP (bottom) spectral models, in function of frequency and one design parameter. The models are trained on 50 $\boldsymbol{p} \in X_{\boldsymbol{p}}$ samples and 71 $f \in X_f$ frequencies. Blue dots represent the training samples (50×71).

Here, the GP is noticeably less accurate than the DGP in fitting the data samples.

Finally, Table I reports the time needed for model training and the maximization of the UCB acquisition function, recorded at the first and at the last iteration. The UCB maximization also includes the cost of inferring new values from the surrogate models and computing the objective distribution in the SMBO. Each value in the table is averaged across five runs and does not include the simulation cost of the filter response in ADS. Since all optimization methods are evaluated for the same number of iterations and with the same simulation settings (i.e., number of chosen frequency samples), the computational cost of ADS Momentum simulations does not show meaningful differences across all optimization strategies. This table shows that both the training and the inference time for the DGP is significantly higher than a simple GP. Moreover, by comparing the BO with DGP and SMBO, it can be observed that modeling the S coefficients over the frequency further increases the computational cost of the iterations. Therefore, the new SMBO strategy is more advantageous for the optimization of devices with high simulation costs.

V. Application II: Dual-Band Slot Antenna

In this section, the two optimization strategies are tested on a dual-band slot antenna [28]. The top conductive layer of this device (see Fig. 9) is deposited on a dielectric material of permittivity $\epsilon = 3.5$, loss tangent $\tan\delta = 0.0009$ and thickness of 0.76 mm. For this DUT, three design parameters are chosen: $L_1 \in [29.5, 40]$ mm, $L_2 \in [5, 12]$ mm, $L_3 \in [17, 25]$ mm. The reflection scattering coefficient Γ of the DUT is computed via ADS Momentum [24] for 50 equispaced frequencies $X_f = [1.5, 6]$ GHz. Then, the following goal function g is applied

TABLE I
COMPUTATIONAL TIMES IN SECONDS FOR APPLICATION I, RECORDED FOR EACH OPTIMIZATION STRATEGY

	1-st iteration		30-th iteration	
Strategy	Model training	UCB maximization	Model training	UCB maximization
BO, GP	1.112	0.024	1.312	0.014
SMBO, DGP	13.943	6.571	31.945	7.117

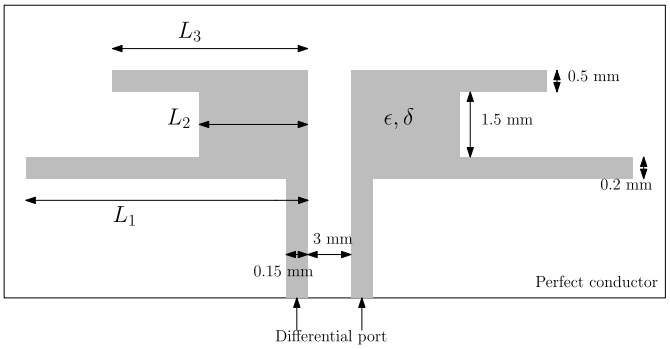

Fig. 9. Application II: Dual-band slot antenna (top view); the geometry is symmetric with respect to the vertical axis.

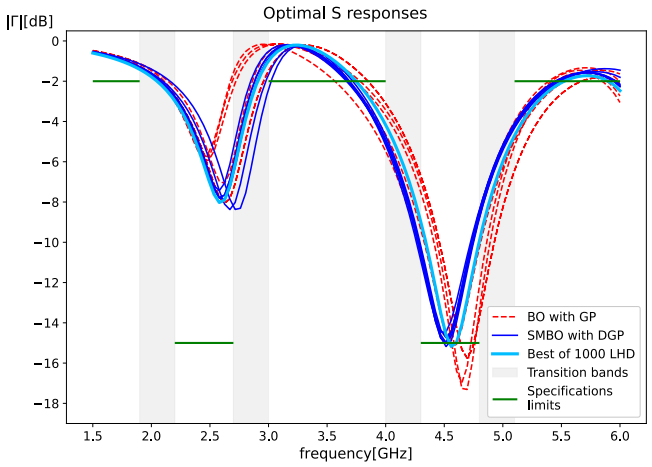

Fig. 10. Application II: Optimal S-responses identified with five different runs of both optimization strategies, after 30 iterations.

to search the desired dual-band behavior:

$$g(\boldsymbol{p}, f) = |\Gamma(\boldsymbol{p}, f)| - 0.79, \quad \text{for } f < 1.9 \text{ GHz} \tag{16a}$$
$$g(\boldsymbol{p}, f) = 0.18 - |\Gamma(\boldsymbol{p}, f)|, \quad \text{for } f \in [2.2, 2.7] \text{ GHz} \tag{16b}$$
$$g(\boldsymbol{p}, f) = |\Gamma(\boldsymbol{p}, f)| - 0.79, \quad \text{for } f \in [3.0, 4.0] \text{ GHz} \tag{16c}$$
$$g(\boldsymbol{p}, f) = 0.18 - |\Gamma(\boldsymbol{p}, f)|, \quad \text{for } f \in [4.3, 4.8] \text{ GHz} \tag{16d}$$
$$g(\boldsymbol{p}, f) = |\Gamma(\boldsymbol{p}, f)| - 0.79, \quad \text{for } f > 5.1 \text{ GHz} \tag{16e}$$

where 0.79 and 0.18 correspond to -2 and -15 dB, respectively. This goal function returns higher value either when $|\Gamma|$ increases in the stop bands $[1, 1, 9]$, $[3, 4]$, and $[4.3, 4.8]$ GHz, and when $|\Gamma|$ decreases in the pass bands $[2.2, 2.7]$ and $[4.3, 4.8]$ GHz. Thus, higher values of g correspond to better dual-band responses in the considered frequency range. Compared to the previous application, this goal function represents more realistic specifications for the DUT, thus resulting in a more complex objective function q and objective distribution $\hat{q}$.

Then, similar to the previous example, the BO and SMBO strategies are executed for 30 iterations, for five different sets of initial samples. The same settings of Application I are used for both GP and DGP models. The best responses found in each run of the strategies are reported in Fig. 10. For reference, the $|\Gamma|$ response corresponding to the best parameters values $\boldsymbol{p}_*$ is included (green curve). The best $\boldsymbol{p}_* = [34.25, 11.42, 22.41]$ is obtained from 1000 simulations of the design parameters chosen according to LHD sampling. This graph shows that the DGP solutions are closer to the actual optimum. In addition, the average normalized regret 15 after 30 iterations is 0.34 for BO and 0.04 for SMBO. These results indicate that, for this example, the BO using GP converges to a suboptimal solution. On the contrary, the SMBO with DGP is able to identify responses that are close to the best response $|\Gamma(\boldsymbol{p}_*)|$. Thus, this application illustrates the advantage of SMBO when more complicated objectives are imposed on the DUT.

VI. CONCLUSION

A new BO strategy based on DGP spectral models has been presented for the optimization of microwave devices. This SMBO strategy avoids the standard direct modeling of objective functions, thus overcoming the trade-off between good optimization results and low objective function complexity.

First, a new spectral surrogate model is employed to model the S coefficients as a function of the frequency and design parameters. Second, an objective probability distribution is derived from the model prediction and maximized using a conventional BO scheme. The new SMBO strategy better identifies the optimal parameter values, compared to the standard BO (based on the objective function modeling with GPs).

Note however that the increased modeling accuracy, flexibility, and optimization performance come at a cost. In fact, it requires a longer iteration time, which makes it more beneficial for devices with a high simulation cost, or, in the HITL optimization scenario, for problems where the initial objectives might need some fine-tuning. Further studies are necessary to extend the proposed strategy, first toward a joint model of the S coefficient magnitude and phase, and later toward the optimization of full S matrices.

REFERENCES

[1] J. E. Rayas-Sanchez, S. Koziel, and J. W. Bandler, "Advanced RF and microwave design optimization: A journey and a vision of future trends," *IEEE J. Microw.*, vol. 1, no. 1, pp. 481–493, Jan. 2021.

[2] J. Jin, F. Feng, W. Zhang, J. Zhang, Z. Zhao, and Q.-J. Zhang, "Recent advances in deep neural network technique for high-dimensional microwave modeling," in *IEEE MTT-S Int. Microw. Symp. Dig.*, Hangzhou, China, Dec. 2020, pp. 1–3.

[3] H. Yu, H. M. Torun, M. U. Rehman, and M. Swaminathan, "Design of SIW filters in D-band using invertible neural nets," in *IEEE MTT-S Int. Microw. Symp. Dig.*, Los Angeles, CA, USA, Aug. 2020, pp. 72–75.

[4] S. Koziel, A. Pietrenko-Dabrowska, and U. Ullah, "Low-cost modeling of microwave components by means of two-stage inverse/forward surrogates and domain confinement," *IEEE Trans. Microw. Theory Techn.*, vol. 69, no. 12, pp. 5189–5202, Dec. 2021.
[5] S. Koziel and A. Pietrenko-Dabrowska, "Design-oriented computationally-efficient feature-based surrogate modelling of multi-band antennas with nested Kriging," *AEU-Int. J. Electron. Commun.*, vol. 120, Jun. 2020, Art. no. 153202.
[6] R. Trinchero, M. Larbi, H. 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.
[7] S. Koziel and A. Pietrenko-Dabrowska, *Performance-Driven Surrogate Modeling of High-Frequency Structures*, 1st ed. Cham, Switzerland: Springer, 2020.
[8] J. Qing, N. Knudde, I. Couckuyt, D. Spina, and T. Dhaene, "Bayesian active learning for electromagnetic structure design," in *Proc. 14th Eur. Conf. Antennas Propag. (EuCAP)*, Copenhagen, Denmark, Mar. 2020, pp. 1–5.
[9] H. M. Torun and M. Swaminathan, "High-dimensional global optimization method for high-frequency electronic design," *IEEE Trans. Microw. Theory Techn.*, vol. 67, no. 6, pp. 2128–2142, Jun. 2019.
[10] S. Koziel, S. Ogurtsov, I. Couckuyt, and T. Dhaene, "Variable-fidelity electromagnetic simulations and co-Kriging for accurate modeling of antennas," *IEEE Trans. Antennas Propag.*, vol. 61, no. 3, pp. 1301–1308, Mar. 2013.
[11] N. Knudde et al., "Data-efficient Bayesian optimization with constraints for power amplifier design," in *IEEE MTT-S Int. Microw. Symp. Dig.*, Reykjavik, Iceland, Aug. 2018, pp. 1–3.
[12] C. E. Rasmussen and C. K. Williams, *Gaussian Processes for Machine Learning*, 1st ed. Cambridge, MA, USA: MIT Press, 2008.
[13] A. C. Damianou and N. D. Lawrence, "Deep Gaussian processes," in *Proc. 16th Int. Conf. Artif. Intell. Statist.*, Scottsdale, AZ, USA, Apr. 2013, pp. 207–215.
[14] N. Knudde, V. Dutordoir, J. V. D. Herten, I. Couckuyt, and T. Dhaene, "Hierarchical Gaussian process models for improved metamodeling," *ACM Trans. Modeling Comput. Simulation*, vol. 30, no. 4, pp. 1–17, Oct. 2020.
[15] A. Hebbal, L. Brevault, M. Balesdent, E.-G. Talbi, and N. Melab, "Bayesian optimization using deep Gaussian processes," 2019, *arXiv:1905.03350*.
[16] A. Hebbal, "Deep Gaussian processes for the analysis and optimization of complex systems application to aerospace system design," Ph.D. dissertation, Centre de Recherche en Informatique, Signal et Automatique de Lille-UMR 9189, Université de Lille, Lille, France, Jul. 2021.
[17] M. Titsias, "Variational learning of inducing variables in sparse Gaussian processes," in *Proc. 12th Int. Conf. Artif. Intell. Statist.*, Clearwater Beach, FL, USA, Apr. 2009, pp. 567–574.
[18] J. Wilson, F. Hutter, and M. Deisenroth, "Maximizing acquisition functions for Bayesian optimization," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 31, 2018, pp. 1–15.
[19] N. Loka, I. Couckuyt, F. Garbuglia, D. Spina, I. Van Nieuwenhuyse, and T. Dhaene, "Bi-objective Bayesian optimization of engineering problems with cheap and expensive cost functions," *Eng. Comput.*, vol. 2022, pp. 1–11, Jan. 2022.
[20] S. Koziel and S. Ogurtsov, "Multi-objective design of antennas using variable-fidelity simulations and surrogate models," *IEEE Trans. Antennas Propag.*, vol. 61, no. 12, pp. 5931–5939, Dec. 2013.
[21] D. M. Roijers, P. Vamplew, S. Whiteson, and R. Dazeley, "A survey of multi-objective sequential decision-making," *J. Artif. Intell. Res.*, vol. 48, pp. 67–113, Oct. 2013.
[22] S. De Ridder, D. Deschrijver, D. Spina, T. Dhaene, and D. V. Ginste, "Adaptive frequency sampling using linear Bayesian vector fitting," *Electron. Lett.*, vol. 55, no. 2, pp. 74–76, Jan. 2019.
[23] D. Puttadilok, D. Eungdamrong, and W. Tanacharoenwat, "A study of narrow-band and compact size microstrip bandpass filters for wireless communications," in *Proc. SICE Annu. Conf.*, Takamatsu, Japan, Sep. 2007, pp. 1418–1421.
[24] (Jun. 2018). *Advanced Design System*. [Online]. Available: http://www.keysight.com/find/eesof-ads
[25] P. Auer, "Using upper confidence bounds for online learning," in *Proc. 41st Annu. Symp. Found. Comput. Sci.*, Washington, DC, USA, 2000, pp. 270–279.
[26] A. Wilson and R. Adams, "Gaussian process kernels for pattern discovery and extrapolation," in *Proc. 30th Int. Conf. Mach. Learn.*, Atlanta, GA, USA, Jun. 2013, pp. 1067–1075.
[27] F. A. C. Viana, G. Venter, and V. Balabanov, "An algorithm for fast optimal Latin hypercube design of experiments," *Int. J. Numer. Methods Eng.*, vol. 82, no. 2, pp. 135–156, Apr. 2010.
[28] S. Koziel, N. Çalık, P. Mahouti, and M. A. Belen, "Accurate modeling of antenna structures by means of domain confinement and pyramidal deep neural networks," *IEEE Trans. Antennas Propag.*, vol. 70, no. 3, pp. 2174–2188, Mar. 2022.

Federico Garbuglia received the M.Sc. degree in electronic engineering from the Polytechnic University of Marche, Ancona, Italy, in 2019. He is currently pursuing the Ph.D. degree in electrical engineering at the Internet Technology and Data Science Laboratory (IDLab), Ghent University, Ghent, Belgium, working with the Surrogate Modeling Laboratory (SUMO Lab).

His current research interests include machine learning for electronic design automation and optimization, and adaptive system identification.

Domenico Spina (Member, IEEE) received the M.S. degree (cum laude) in electronics engineering from the University of L'Aquila, L'Aquila, Italy, in 2010, and the joint Ph.D. degree in electrical engineering from Ghent University, Ghent, Belgium, and the University of L'Aquila, in 2014.

Since 2015, he has been a member with the Internet Technology and Data Science Laboratory (IDLab), Department of Information Technology (INTEC), Ghent University–imec, Ghent, first as a Post-Doctoral Researcher and currently as the Research and Development Technical Lead. His research interests include modeling and simulation, uncertainty quantification, and machine learning for analog electronic and photonic systems.

Dirk Deschrijver (Senior Member, IEEE) received the first Ph.D. degree from the Department of Mathematics and Computer Science, Ghent University, Ghent, Belgium, in 2007 and the second Ph.D. degree in engineering from Ghent University, in 2012.

From 2008 to 2014, he has worked as an FWO Post-Doctoral Research Fellow with the IBCN Research Group, Internet Technology and Data Science Laboratory (IDLab), Ghent University. Since 2016, he has been an Associate Professor with the IDLab Research Group, Ghent University, working on data analytics, machine learning, and surrogate modeling.

Ivo Couckuyt received the M.S. 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 Laboratory (IDLab), Ghent University, Ghent, Belgium, in 2013.

In October 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.

Tom Dhaene (Senior Member, IEEE) received the M.S. 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 Laboratory (IDLab), Ghent University, researching on distributed scientific computing, machine learning, bioinformatics, signal integrity, electromagnetic compatibility, optimal design, surrogate modeling, circuit, and EM modeling.