

# Bayesian active learning for multi-objective feasible region identification in microwave devices

Federico Garbuglia,<sup>✉</sup> Jixiang Qing, Nicolas Knudde, Domenico Spina, Ivo Couckuyt, Dirk Deschrijver, and Tom Dhaene

IDLab, Department of Information Technology, Ghent University-imec, Gent, Belgium

<sup>✉</sup>Email: federico.garbuglia@ugent.be

In microwave device and circuit design, many simulations are often needed to find a set of designs that satisfy one or multiple specifications chosen by the designer upfront: the feasible region. A novel Bayesian active learning framework is presented to accurately identify the feasible region with a low number of simulations. The technique leverages on a stochastic model to obtain an efficient and automated procedure. A suitable application example validates the proposed technique and shows its effectiveness to rapidly obtain many suitable designs.

**Introduction:** Over the last decades, the increase of available computing power has moved electronic designers away from hardware-based prototyping, towards computer aided design (CAD) simulations. However, simulations of modern microwave devices and circuits are expensive, both in terms of computational time as well as resources, due to the bandwidth requirements coupled with the complexity of modern microwave systems. Hence, several data-efficient techniques have been developed in recent years to reduce the number of expensive simulations required during the design process [1–4]. In this work, a novel Bayesian active learning framework for microwave applications is proposed. Contrary to optimization problems, where the goal is to find the best solution, this method identifies the set of all design configurations that satisfies chosen design specifications. This set of design solutions is called the *feasible region* in the rest of the contribution, while the chosen design specifications are referred to as *feasibility conditions*. The adopted active learning framework requires only a limited number of expensive CAD simulations (i.e. full-wave simulations) to efficiently identify the feasible region. This novel methodology is particularly well suited for design space exploration and reduction.

**Goal:** The main objective is to define a general methodology for feasible region discovery, which can be applied to a large range of microwave design problems. Hence, designers should be able to define multiple feasibility conditions, and the corresponding feasible region must be identified by an efficient and automated procedure. The proposed solution is a Bayesian active learning approach that relies on a stochastic model, coupled with a suitable sequential sampling strategy to minimize the number of expensive simulations needed to build and update such a model.

**Bayesian active learning framework:** Let us assume that the behaviour of the microwave system under study depends on a set of geometrical or electrical parameters, collected in the vector  $\mathbf{x}$ . Typical examples include the width and length of a metallic trace or the relative permittivity of a dielectric. Furthermore, the range of admissible values for each parameter is specified by the designers, and this defines the design space for the problem at hand. Finding a solution to the design problem corresponds to finding a point in the design space that satisfies all feasibility conditions. In general, the larger the design space (either by increasing the number of design parameters or their range), the higher the complexity of this problem. Our goal is to reduce this complexity by identifying the feasible region, that is, the area(s) in the design space that contains only feasible designs. Feasibility conditions are to be defined upon performance metrics that can be computed via CAD simulations. These performance metrics, which depend on the values assumed by design parameters, are called *objective functions* in this contribution. For example, in a filter design, the bandwidth can be specified as an objective function, and the corresponding feasible region is given by all the values of the design parameters leading to the desired bandwidth. In or-

der to reach the stated goal, a new Bayesian active learning approach is proposed, which is summarized in Figure 1. The first step is to evaluate via CAD simulations the objective function(s) of interest for a small set of initial samples in the design space, which are chosen via Latin hypercube design (LHD) [5]. Based on the initial simulations, a regression model of each objective function is built. Next, a new sample in the design space is selected according to a suitable sampling strategy. It is important to note that the model employed here is *stochastic*: The sampling strategy chooses the new design point to be evaluated based on the model prediction at the location of the feasible region. Subsequently, a new simulation is performed and the model can be updated accordingly. This iterative process is repeated until a stopping criterion is met. Finally, the location of the feasible region in the design space can be estimated from the stochastic model.

In particular, a Gaussian process (GP) is chosen as a stochastic model due to its flexibility and modelling power [4]. A GP is a distribution over functions, which can approximate a selected objective function  $f$ . In this framework, each point can be represented with a random Gaussian variable, with associated mean  $m$  and covariance  $k$ :

$$m(\mathbf{x}) = \mathbb{E}[f(\mathbf{x})] \quad (1a)$$

$$k(\mathbf{x}, \mathbf{x}') = \mathbb{E}[(f(\mathbf{x}) - m(\mathbf{x}))(f(\mathbf{x}') - m(\mathbf{x}'))], \quad (1b)$$

where  $\mathbb{E}$  designates the expectation operator. The radial basis kernel function with automatic relevance detection (ARD) is used as covariance of the GP model (also called kernel function) [6]:

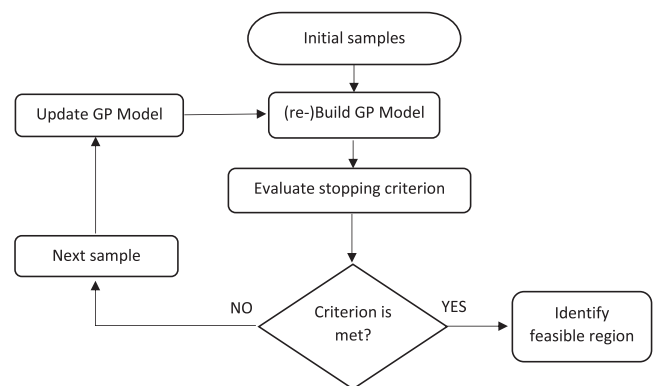
$$k(\mathbf{x}, \mathbf{x}') = \sigma_k^2 \exp\left(-\sum_{d=1}^D \frac{(\mathbf{x}_d - \mathbf{x}'_d)^2}{2\ell_d^2}\right), \quad (2)$$

where  $\mathbf{x}, \mathbf{x}'$  are vectors of data points, while  $\sigma_k$  and  $\ell_d$  are tunable hyper-parameters. GP regression [6] computes the predictive distribution for new test points  $\mathbf{x}_*$ , conditioned to the available training data  $D_n = \{\mathbf{x}_i, f(\mathbf{x}_i)\}_{i=1}^n$ . Under the previous assumptions, the predictive distribution  $p(f|\mathbf{x}_*, D_n)$  is also Gaussian, and is determined by the mean  $\mu$  and variance  $\sigma^2$  defined as:

$$\mu(\mathbf{x}^*) = \mathbb{E}[f_*|\mathbf{x}_*, D_n] = K_{\mathbf{x}\mathbf{x}^*}^T (K_{\mathbf{x}\mathbf{x}} + \sigma_r^2 I)^{-1} f(\mathbf{x}), \quad (3)$$

$$\sigma^2(\mathbf{x}^*) = \mathbb{V}[f_*|\mathbf{x}_*, D_n] = K_{\mathbf{x}^*\mathbf{x}^*} - K_{\mathbf{x}\mathbf{x}^*}^T (K_{\mathbf{x}\mathbf{x}} + \sigma_r^2 I)^{-1} K_{\mathbf{x}\mathbf{x}^*}, \quad (4)$$

where  $K_{\mathbf{x}\mathbf{x}}, K_{\mathbf{x}^*\mathbf{x}^*}, K_{\mathbf{x}\mathbf{x}^*}$  matrices consist of kernel functions (2) evaluated on pairs of data points, pairs of test points, pairs of a data and a test point, respectively. Also,  $\mathbb{V}$  represents the variance operator.  $\sigma_r^2$  is a Gaussian noise variance, which accounts for an eventual additive noise on data points. Thus, for each test point, the regression provides an estimate of the function value and its uncertainty, indicated by  $\mu(\mathbf{x}_*)$  and  $\sigma^2(\mathbf{x}_*)$ , respectively.



**Fig. 1** Flowchart of Bayesian active learning for feasible region identification. A Gaussian process stochastic model is built on initial design samples and updated after each new sample evaluation until the stopping criterion is met

When multiple feasibility conditions are defined, the corresponding objective function is multi-dimensional:  $\mathbf{f} = \{f_1, f_2, \dots, f_k\}$ . In this case, the model consists of one GP for each component of  $\mathbf{f}$ .

Apart from the regression model, the other pillar of active learning is the definition of a suitable sequential sampling strategy. Here, the objective is to individuate the design point that maximizes the information gain on the location of the feasible region. Such point is chosen as new sample to update the regression model, as shown in Figure 1. In particular, the design point maximizing the information-gain function [7] is chosen as new sample point:

$$\alpha_g(\mathbf{x}) = \mathbb{H}(p(\mathbf{f}|D_n, \mathbf{x})) - \mathbb{E}_{p(g|D_n)}[\mathbb{H}(p(\mathbf{f}|D_n, \mathbf{x}, g))], \quad (5)$$

where  $\alpha_g$  is called the *acquisition function*,  $\mathbb{H}$  is the entropy operator and  $g$  is one of the intervals defined as follows. Let us suppose that the feasible region for the problem at hand is the area of the design space for which  $f \in [a, b]$ , where  $a$  and  $b$  are suitable scalar values. Then, the feasible region and its complement are defined by three intervals:

$$g_1 = \{f < a\}, \quad g_2 = \{a < f < b\}, \quad g_3 = \{f > b\}. \quad (6)$$

In this framework, the probability densities  $p(\mathbf{f}|D_n, \mathbf{x})$  and  $p(\mathbf{f}|D_n, \mathbf{x}, g)$  correspond to a normal distribution and a truncated normal distribution, respectively, both with mean  $\mu(\mathbf{x})$  and variance  $\sigma^2(\mathbf{x})$ ; it follows that their entropy can be expressed analytically. For instance, on the finite  $g_2$  interval the information gain becomes [7]:

$$\alpha_{g_2}(\mathbf{x}) = \frac{1}{2} \ln(2\pi e\sigma^2) - \left[ \frac{1}{2} \ln(2\pi e\sigma^2 Z^2) + \frac{1}{2Z} \left[ (a - \mu(\mathbf{x}))\mathcal{N}(\mu(\mathbf{x})|a, \sigma^2) - (b - \mu(\mathbf{x}))\mathcal{N}(\mu(\mathbf{x})|b, \sigma^2) \right] \right], \quad (7)$$

where  $Z = \Phi\left[\frac{b-\mu(\mathbf{x})}{\sigma(\mathbf{x})}\right] - \Phi\left[\frac{a-\mu(\mathbf{x})}{\sigma(\mathbf{x})}\right]$  is a normalization constant, and  $\Phi$  is the cumulative density function of the standard normal distribution  $\mathcal{N}$ . Then, the acquisition function on the entire domain is the sum of  $\alpha_{g_i}(\mathbf{x})$  for each interval:

$$\alpha_f(\mathbf{x}) = \sum_{i=1}^3 \alpha_{g_i}(\mathbf{x}). \quad (8)$$

Now, an optimization problem must be solved to identify the sample maximizing (5). In this work, an L-BFGS-B gradient-based optimizer [8] is used for this purpose. Note that the optimization process can be performed very efficiently: The acquisition function depends only on the GP model, which is very cheap to evaluate, and the training data  $D_n$ , which has already been computed.

However, the acquisition function in (5) can be adopted only for problems defined by a single feasibility constraint [7]. In the following, the information-gain metric defined above is generalized for multiple feasibility constraints. Indeed, (5) can be written as:

$$\alpha_g(\mathbf{x}) = \mathbb{H}(p(\mathbf{f}|D_n, \mathbf{x})) - \mathbb{E}_{p(g|D_n)}[\mathbb{H}(p(\mathbf{f}|D_n, \mathbf{x}, g))] = \frac{1}{2} \ln[\det(2\pi e\Sigma(\mathbf{x}))] - \sum_{i=1}^k \mathbb{H}(p(f_i|D_n, \mathbf{x}, g_i)), \quad (9)$$

where  $\Sigma$  is the covariance matrix of  $\mathbf{f}$ , while  $\mathbf{g}$  is the multi-dimensional interval to be discriminated. Assuming that the components of  $\mathbf{f}$  are independent, then  $\Sigma = \text{diag}\{\sigma_1^2, \dots, \sigma_k^2\}$  and (9) for interval  $g_2$  becomes:

$$\alpha_{g_2}(\mathbf{x}) = \sum_{i=1}^k \frac{1}{2} \ln(2\pi e\sigma_i^2) - \sum_{i=1}^k \left[ \frac{1}{2} \ln(2\pi e\sigma_i^2 Z_i^2) + \frac{1}{2Z_i} \left[ (a_i - \mu_i)\mathcal{N}(\mu_i|a_i, \sigma_i^2) - (b_i - \mu_i)\mathcal{N}(\mu_i|b_i, \sigma_i^2) \right] \right] \quad (10)$$

Comparing (7) and (10), the information gain on multiple dimensions is the sum of the gains of each component. Under the same considerations, this result is also valid for the other non-finite  $\mathbf{g}$  intervals. Finally,

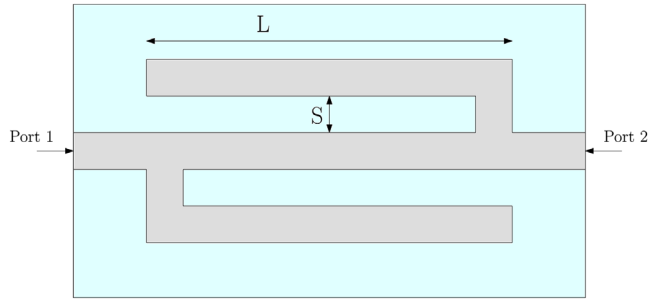


Fig. 2 Geometry of the microstrip band-stop filter under study (top view)

Table 1. Design parameters for the microstrip band-stop filter

Parameter	Description	Range
L	Stub length	[2, 2.5] mm
S	Line-stub spacing	[0.1, 0.2] mm
h	Dielectric height	[0.10, 0.15] mm
$\epsilon$	Substrate permittivity	[9.5, 11]

by applying (8), the acquisition function for multiple constraints can be expressed as:

$$\alpha_f(\mathbf{x}) = \sum_{i=1}^k \alpha_{f_i}(\mathbf{x}). \quad (11)$$

It follows that the overall acquisition function is simply the sum of  $\alpha_f(\mathbf{x})$  for each dimension of  $\mathbf{f}$ .

A suitable criterion must be defined to stop the iterative refinement of the model, as shown in Figure 1. In this work, an upper bound on the number of total samples is adopted for simplicity. Note that alternative criteria can also be employed, based, for example, on the variance of the regression model [9].

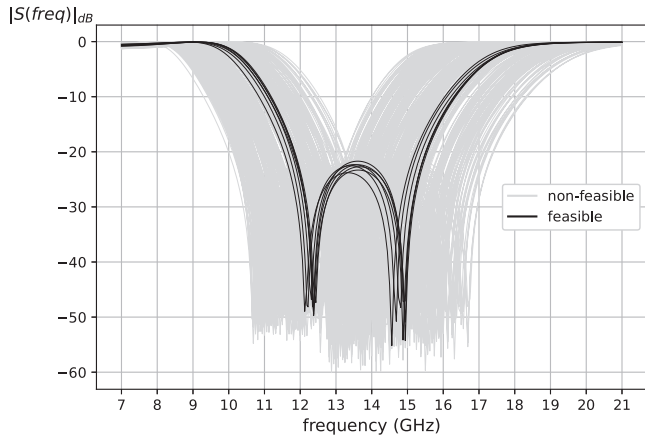
In this application, the main task of the regression model is to evaluate whether a point is inside or outside the feasible region. Hence, the accuracy of the model predictions is evaluated by using a binary classification metric: the F1 score [10]. Typically the F1 score is evaluated on a test set, an independent set of samples. Computing the test set requires to perform CAD simulations over a large number of points in the design space and to compare the simulation results with the model predictions. In fact, the F1 score is defined as the harmonic mean between precision and recall:

$$\text{precision} = \frac{\text{true positives}}{\text{true positives} + \text{false positives}}, \quad (12)$$

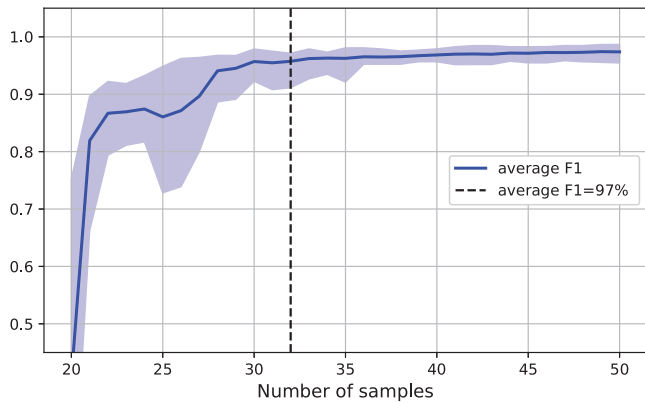
$$\text{recall} = \frac{\text{true positives}}{\text{true positives} + \text{false negatives}}, \quad (13)$$

where ‘*true positives*’ are the points correctly identified as feasible, while ‘*false*’ are incorrectly labelled points, which can be either ‘*positives*’ (if predicted feasible) or ‘*negatives*’ (if predicted unfeasible). Note that the F1 score cannot be integrated in the algorithm as stopping criterion, due to the large number of simulations required, leading to a high computational cost.

*Application example:* The proposed methodology is tested on a microstrip stop-band filter [11]. This device consists of a dielectric substrate between a top metallization and a bottom ground plane. The geometry of the top layer consists of two stubs, with identical length and spacing, folded along the transmission line, as shown in Figure 2. On this device, a feasible region identification problem is set up as follows. Four design parameters  $\mathbf{x} = (L, S, h, \epsilon)$  are considered and described in Table 1, along with the corresponding range in the design space. Then, two feasibility conditions are defined on the  $-3$  dB bandwidth ( $BW$ ) and



**Fig. 3** Magnitude of the element  $S_{21}$  of the scattering matrix, computed for 300( $L, S, h, \epsilon$ ) combinations. Only seven designs are feasible



**Fig. 4** Average F1 score (blue line) for 10 different GP models, after 20 samples. The light blue color indicates the variability of the score

central frequency ( $freq_0$ ) of the filter stop-band:

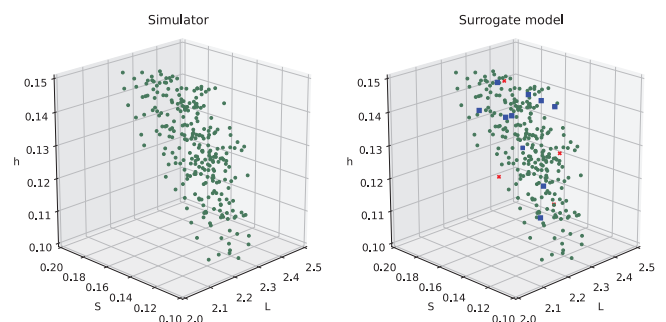
$$7.0 \text{ GHz} \leq BW \leq 7.1 \text{ GHz}, \quad (14a)$$

$$13.5 \text{ GHz} \leq freq_0 \leq 14 \text{ GHz}. \quad (14b)$$

In this way, these two quantities represent the objective function to be modelled:  $f = \{BW(x), freq_0(x)\}$ . In order to evaluate (14) for a particular set of parameters, the filter frequency response is simulated via the momentum electromagnetic field simulator of Advanced Design System [12], by adopting an adaptive frequency sampling in the range [7 – 21] GHz. Note that the filter response is very dynamic with respect to the chosen design parameters, as shown in Figure 3, thus making it a challenging feasible region identification problem. Indeed, Figure 3 shows that, by evaluating 300 ( $L, S, h, \epsilon$ ) samples randomly chosen in the design space, only 7 acceptable designs are found. In order to identify the feasible region, a GP regression model consisting of one GP for each component of the objective function is built. First, the GP model is trained on 10 samples chosen via LHD [5]. Then, the model is updated one sample at a time, by using the proposed information-based acquisition function: For each new samples, the value of  $f$  is computed via an Advanced Design System simulation, by setting design parameters  $x$  which maximizes the total information gain (11). Finally, this procedure is halted after 40 iterations, reaching the maximum computational budget chosen for this problem: 50 total samples. The model accuracy is validated by computing the F1 score with 10,000 ( $L, S, h, \epsilon$ ) validation samples, which are randomly chosen in the design space. In particular, Figure 4 shows the model F1 score as a function of the number of samples used to compute the model. Additionally, in order to verify the robustness of the proposed methodology with respect to the choice of initial samples, the model construction is repeated 10 times for different initial samples chosen via LHD. As demonstrated by Figure 4, the feasible region estimated by the regression model rapidly converges to the

**Table 2.** Number of classification errors (false positives + false negatives) for each condition (14), occurred out of 10,000 randomly chosen validation points, on different runs of the proposed algorithm

Run	$N_{BW}$	$N_{freq_0}$
1	8	1
2	15	3
3	19	2
4	12	0
5	9	0
6	13	0
7	9	0
8	18	2
9	9	2
10	14	1
Average	12.6	1.1



**Fig. 5** Example of feasible designs computed for 10,000 random combinations of all parameters, plotted for varying( $L, S, h$ ). True positives (green), false positives (blue) and false negatives (red) are highlighted for the GP model

one provided by the simulator: The average score reaches 97% after only 32 samples. One can also observe that the F1 scores of all different runs converges rapidly: The choice of the initial samples has a limited impact on the model performance. In Table 2, the classification errors (false positives and false negatives) are indicated for the 10 models computed. On average, 13.7 errors are recorded over 10,000 samples and most of them are due to the bandwidth constraint (14a). Among all 10 models computed, the maximum discrepancy between model and simulator results is 10.8 MHz for the bandwidth estimation and 6.93 MHz for the central frequency, meaning that errors occur near the edge of the feasible region. Finally, an illustration of the feasible region is given in Figure 5. The figure confirms the high modelling accuracy of the proposed technique, since only few errors are present at the border of the feasible region.

**Conclusion:** A novel feasible region identification methodology was presented, based on Bayesian active learning. In particular, this novel approach is able to classify designs according to one or multiple feasibility conditions, and it can be applied to a large range of microwave devices and systems, thanks to the flexibility and modelling power of GP models. A suitable information-based sampling strategy is adopted in the model-building phase, which allows for the feasible region identification with an efficient and automated framework. A suitable application example validates the proposed methodology.

**Acknowledgements:** This work has been supported by the Flemish Government under the ‘Onderzoeksprogramma Artificiële Intelligentie (AI) Vlaanderen’ and the ‘Fonds Wetenschappelijk Onderzoek (FWO)’ programmes.

© 2021 The Authors. *Electronics Letters* published by John Wiley & Sons Ltd on behalf of The Institution of Engineering and Technology

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Received: 3 July 2020 Accepted: 20 October 2020

doi: 10.1049/el12.12022

## References

- 1 Stievano, I.S., et al.: On-the-fly estimation of IC macromodels. *Electron. Lett.* **42**(14), 801–803 (2006)
- 2 Chemmangat, K., et al.: Parametric macromodelling of linear high frequency systems using multiple frequency scaling and sequential sampling. *Electron. Lett.* **50**(6), 475–476 (2014)
- 3 Passos, F., et al.: Parametric macromodeling of integrated inductors for RF circuit design. *Microw. Opt. Technol. Lett.* **59**, 1207–1212 (2017)
- 4 Koziel, S., Pietrenko-Dabrowska, A.: *Performance-Driven Surrogate Modeling of High-Frequency Structures*. 1st edition. Springer, Cham, Switzerland (2020)
- 5 Viana, F.A.C., et al.: An algorithm for fast optimal Latin hypercube design of experiments. *Int. J. Numer. Methods Eng.* **82**(2), 135–156 (2010)
- 6 Rasmussen, C.E., Williams, C.K.: *Gaussian Processes for Machine Learning*. 1st edition. MIT Press, Cambridge (2008)
- 7 Knudde, N., et al.: Active learning for feasible region discovery. In: Proceedings of IEEE ICMLA, Boca Raton, FL, USA, pp. 567–572 (2019)
- 8 Mokhtari, A., Ribeiro, A.: Global convergence of online limited memory BFGS. *J. Mach. Learn. Res.* **16**, 3151–3181 (2015)
- 9 Picheny, V., et al.: Adaptive designs of experiments for accurate approximation of a target region. *J. Mech. Des.* **132**(7), 567–572 (2010)
- 10 Lipton, Z.C., et al.: Thresholding classifiers to maximize F1 score. arXiv preprint (2014), arXiv:1402.1892v2
- 11 Spina, D., et al.: Variability analysis of multiport systems via polynomial-chaos expansion. *IEEE Trans. Microw. Theory Tech.* **60**(8), 2329–2338 (2012)
- 12 Advanced Design System. 471.update1.0 (2018), <https://www.keysight.com/be/en/lib/resources/software-releases/ads-201501.html>, April 19, 2020. Accessed 2 July 2020.