

Data-Efficient Bayesian Optimization with Constraints for Power Amplifier Design

Nicolas Knudde*, Ivo Couckuyt*, Domenico Spina*, Konstanty Łukasik^{†‡},
Paweł Barmuta^{†‡}, Dominique Schreurs[†] and Tom Dhaene*

*Ghent University - imec, Department of Information Technology, 9052 Ghent, Belgium

[†]KU Leuven, Department of Electrical Engineering, 3000 Leuven, Belgium

[‡]Warsaw University of Technology, Faculty of Electronics and Information Technology, 00-665 Warsaw, Poland

Email: nicolas.knudde@ugent.be

Abstract—Finding the optimal working conditions for non-linear electrical components under large signal stimuli can be challenging, mainly due to the high number of input dimensions and multiple local minima of the goal function. In this paper a Bayesian optimization method is applied in order to limit the number of evaluations by a commercial harmonic balance simulator. The method is applied to amplifier optimization utilizing Wolfspeed CGH40010F GaN HEMT, for which input power, bias voltages and load at fundamental harmonic frequencies are changed in order to maximize for combined efficiency, gain, and output power. The optimum is found already after 80 iterations.

Index Terms—Bayesian optimization, Gaussian process, harmonic balance, load-pull, transistor

I. INTRODUCTION

Modern microwave power amplifiers have to fulfill multiple very stringent requirements, e.g., high output power, high Power Added Efficiency (PAE), etc. This makes the design process very cumbersome, and often leads to extensive use of optimization tools in the design process. Unfortunately, the commercial CAD environments lack efficient optimizers, especially when the objective function has a highly-dimensional parameter space and multiple local minima, which is the typical case for power amplifier design. This problem has been addressed in [1], [2], in which Bayesian Optimization (BO) is applied to the design of high-efficiency power amplifiers [3]. The optimal Large Signal Operating Point (LSOP), i.e., set of input variables' values, was found already within only a few tens of simulations.

Due to its efficiency, BO is also used in this work. However, the efficiency is further increased by employing a recently introduced sampling strategy, called Max-Value Entropy search [4]. We also incorporate the constraints, e.g., power dissipation or breakdown voltage, in the optimization process through so-called Probability of Feasibility (PoF) [5]. The work is organized as follows. In Sec. II, Bayesian optimization is explained with a particular focus on sampling strategies. Next, in Sec. III, the optimization results of the transistor's large signal operating point are shown and discussed, which leads to the conclusions drawn in Sec. IV.

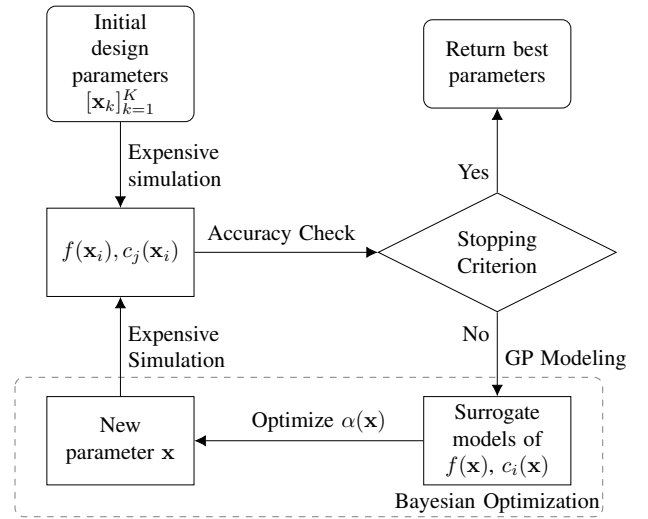


Fig. 1. Flowchart of the Bayesian optimization algorithm.

II. METHODOLOGY

The idea behind BO is to perform optimization on surrogate models, which mimic the real optimization problem, but contrary to the latter, are very cheap to evaluate [6]. The flowchart of BO algorithm is shown in Fig. 1. First, the objective function is evaluated over an initial set of values of the design parameters $[x_k]_{k=1}^K$ via simulation of the device under test. This allows to extract the first surrogate model of the objective function $f(x)$. Since the surrogate model is very cheap to evaluate, it is used by the optimizer to find the global minimum. This optimum is then evaluated via a new simulation, and when none of the stop criteria is met, the surrogate model is updated. Therefore, each additional simulation refines the surrogate model, which increases the chance of finding the global optimum of the true problem by using surrogate models.

As one can see, BO is very general and several mathematical techniques can be adopted to build the desired surrogate model, such as Random Forest [7] or Bayesian neural network [8]. In this paper, we have chosen Gaussian Processes (GPs) [9], thanks to their accuracy and modeling power. A GP represents a distribution over functions $f : \mathcal{X} \rightarrow \mathbb{R}$, that is completely defined by a suitable mean function $m : \mathcal{X} \rightarrow \mathbb{R}$

and covariance function $k : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$, where the type of mean and covariance functions depends on the particular problem under study.

A sampling strategy is defined based on the surrogate model that is computed so far, which is called acquisition function. Typical acquisition functions include Expected Improvement (EI) [3] and Probability of Improvement (POI) [10]. In this work, the recently introduced Max-Value Entropy Search (MES) is adopted [4], which automatically balances exploration and exploitation. In particular, the MES aims to maximize the information gain about the minimum value f_{\min} , which is quantified by the differential entropy:

$$\alpha_{\text{MES}}(\mathbf{x}) = H(p(f^{GP}|\mathcal{D}, \mathbf{x})) - \mathbb{E}(H(p(f^{GP}|\mathcal{D}, \mathbf{x}, f_{\min}))), \quad (1)$$

where \mathcal{D} represents the already observed data $\{\mathbf{x}_i, y_i\}_{i=1}^N$, H denotes the entropy and the expectation \mathbb{E} is computed over the random variable f_{\min} . Note that the distribution of f_{\min} cannot be calculated analytically, but can be sampled via Gumbel sampling [11]. Furthermore, $p(f^{GP}|\mathcal{D}, \mathbf{x}, f_{\min})$ is a truncated Gaussian, since all y values have to be greater than f_{\min} .

However, in many engineering problems it is not sufficient to find a global maximum of the objective, but some physical constraints must be satisfied as well. Hence, the acquisition function has to be modified accordingly. In practice, this is done through the Probability of Feasibility (PoF) [5], where each constraint is often formulated as a suitable function $c : \mathcal{X} \rightarrow \mathbb{R}$ that has to be negative.

In our problem setting, the constraints depend on the physical characteristics of the transistor considered and, typically, include limitations on the power balance and the maximum DC currents of the device. Unfortunately, it is not possible to know upfront (before simulating the device) if a particular combination of design parameters will lead to a constraints violation. Hence, first a suitable GP-based surrogate model is built for each function c_i for $i = 1, \dots, N$, where N is the total number of constraints, then the corresponding PoF is calculated:

$$\alpha_{\text{PoF},i}(\mathbf{x}) = p(c_i^{GP}(\mathbf{x}) < 0). \quad (2)$$

Note that $\alpha_{\text{PoF},i}$ is equal to the probability that the i -th constraint is satisfied and such probability is estimated with increasing precision as the number of iterations of the optimization algorithm increases.

Finally, in order to combine the ability of selecting feasible design points via PoF and the capability of the MES to find the optimum, the overall acquisition function is given by

$$\alpha(\mathbf{x}) = \alpha_{\text{MES}}(\mathbf{x}) \prod_i^N \alpha_{\text{PoF},i}(\mathbf{x}) \quad (3)$$

III. RESULTS AND DISCUSSION

The device under test is the GaN HEMT transistor CGH40010F 10 W from Wolfspeed, which is simulated in ADS¹. The gate $V_G \in [-2, 0]$ V and drain voltage $V_D \in [16, 42]$ V, the input power $P_{\text{in}} \in [0.01, 1]$ W and the complex load impedance Z_L constitute LSOP \mathbf{x} . Furthermore, the load

reflection coefficient Γ_L must have a magnitude less than one and it is parameterized in polar coordinates to efficiently enforce this constraint. A typical design objective is to have the gain G , PAE, and output power P_{out} as high as possible. Since these quantities are highly correlated, in order to reach this goal it is sufficient to minimize a suitable objective function f , defined as:

$$f(\mathbf{x}) = -G(\mathbf{x})\text{PAE}(\mathbf{x})P_{\text{out}}(\mathbf{x}), \quad (4)$$

where the vector \mathbf{x} is LSOP. Additionally, the following constraints must hold:

$$P_{\text{dc}} - P_{\text{out}} + P_{\text{in}} - 14 \text{ W} \leq 0, \quad (5)$$

$$I_{\text{dsi}} - 4.5 \text{ A} \leq 0, \quad (6)$$

$$I_{\text{gsi}} - 4 \text{ mA} \leq 0, \quad (7)$$

where P_{dc} and P_{in} are the DC and input power, respectively, and I_{dsi} , I_{gsi} are the maximum instantaneous drain and gate currents, respectively. Note that all these conditions can be expressed as $c_i < 0$.

The constraints and the objective f are modeled separately. In particular, a periodic kernel is used to model the phase of Γ_L , indicated as ϕ , which is a periodic function. For the remaining design parameters $\tilde{\mathbf{x}}$, a Matérn 3/2 kernel is used to compute the GP model:

$$k(\mathbf{x}, \mathbf{x}') = k_M(\tilde{\mathbf{x}}, \tilde{\mathbf{x}}')k_P(\phi, \phi') \quad (8)$$

$$k_M(\tilde{\mathbf{x}}, \tilde{\mathbf{x}}') = \sigma^2 \left(1 + \frac{\sqrt{3}\|\tilde{\mathbf{x}} - \tilde{\mathbf{x}}'\|}{\ell} \right) \exp \left(-\frac{\sqrt{3}\|\tilde{\mathbf{x}} - \tilde{\mathbf{x}}'\|}{\ell} \right) \quad (9)$$

$$k_P(\phi, \phi') = \exp \left(-\frac{\sin(\phi - \phi')^2}{\lambda} \right) \quad (10)$$

where the mean function is considered equal to zero. The initial set of values of the design parameters (needed to compute the GP models) are generated by a minimax Latin hypercube of 10 points. The kernel (9) is adopted to model the constraints as well. Next, the acquisition function is optimized by evaluating the GP models for 500 randomly sampled points and by considering the best result as starting point of the Limited-memory Broyden–Fletcher–Goldfarb–Shannon optimizer (L-BFGS). Finally, the Bayesian optimization was performed by *GPflowOpt*, a Bayesian optimization package in Python [12].

Furthermore, the proposed method is compared with a regular gradient optimizer (the Sequential Least Squares Programming (SLSQP)) and two Bayesian optimization approaches based on acquisition functions computed via EI and POI, used in [1]. Note that, in order to show statistically relevant results, each optimization algorithm was executed ten times and the mean results are shown in Figs. 2 and 3.

In particular, Fig. 2 shows that the proposed MES-based technique has superior performance with respect to similar Bayesian optimization approaches, thanks to its trade-off between exploration and exploitation; while the SLSQP gradient optimizer is unable to find a feasible minimum of the objective function. For the Bayesian optimization methods considered,

¹Advanced Design System (ADS), Keysight Technologies, Santa Rosa, CA.

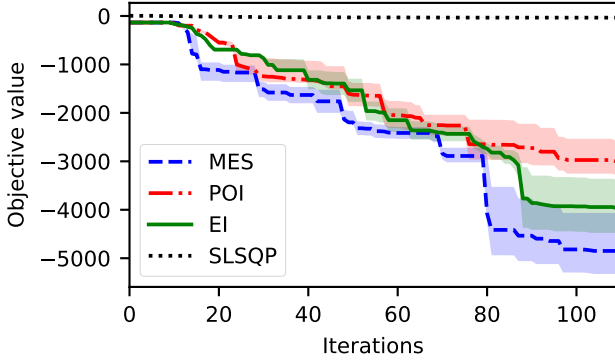


Fig. 2. Feasible minimum of the objective function with respect to the number of iterations. The thick lines are the means of ten runs of the methods surrounded by their 98% confidence interval, calculated by bootstrapping.

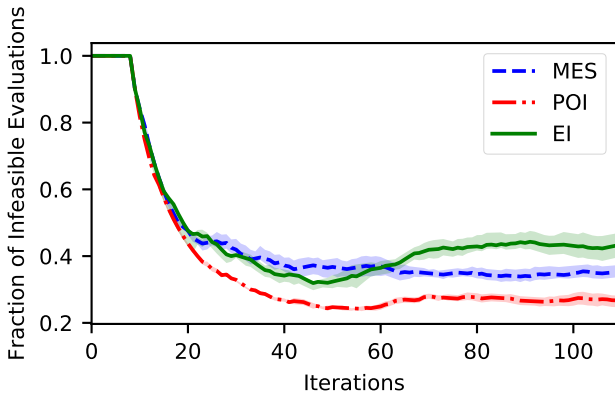


Fig. 3. Fraction of infeasible samples with respect to the number of iterations. The thick lines are the means of ten runs of the methods surrounded by their 98% confidence interval, calculated by bootstrapping.

the number of evaluations on design points resulting in constraint violations is shown in Fig. 3: as expected, the number of infeasible points decreases with respect to the number of iterations and drops to a fraction of approximately 0.4 fairly quick after 40 samples. As can be seen EI and MES have overlapping confidence intervals, but on average, MES still performs better.

TABLE I
OPTIMAL LARGE-SIGNAL OPERATING POINT WITH CORRESPONDING PERFORMANCE AND CONSTRAINT METRICS.

Transistor Characteristics	
G	1810.55
PAE	0.43
P_{out}	0.053 W
Design parameters	
V_G	0.16 V
V_D	0.22 V
P_{in}	0.16 W
Γ_L	0.053
Constraints	
c_1	-0.056 W
c_2	-2.25 A
c_3	-0 mA

Finally, the overall minimum value of the objective f is summarized in Table I, where c_1 , c_2 and c_3 correspond to equations (5), (6) and (7), respectively.

IV. CONCLUSIONS

In this work, a Bayesian optimization strategy was used for the optimization of a simulated device. The Max-Value Entropy search shows good results in combination with the Probability of Feasibility to take into account constraints. Using this combination allowed to achieve lower values of the objective function despite the number of algorithm iterations in comparison to Probability of Improvement and Expected Improvement. It also has to be noted that the ratio of infeasible samples with respect to the number of iterations drops very quickly to around 0.4 for already 40 samples, and is expected to decrease further as more LSOPs are evaluated.

ACKNOWLEDGEMENTS

This work was supported by FWO-Vlaanderen. P. Barmuta is supported by the C2 project (C24/15/015). I. Couckuyt is a postdoctoral research fellow of FWO-Vlaanderen.

REFERENCES

- [1] P. Chen, B. M. Merrick, and T. J. Brazil, "Bayesian optimization for broadband high-efficiency power amplifier designs," *IEEE Transactions on Microwave Theory and Techniques*, vol. 63, no. 12, pp. 4263–4272, 2015.
- [2] P. Chen, J. Xia, B. M. Merrick, and T. J. Brazil, "Multiobjective bayesian optimization for active load modulation in a broadband 20-w gan doherty power amplifier design," *IEEE Transactions on Microwave Theory and Techniques*, vol. 65, no. 3, pp. 860–871, 2017.
- [3] D. R. Jones, M. Schonlau, and W. J. Welch, "Efficient global optimization of expensive black-box functions," *Journal of Global optimization*, vol. 13, no. 4, pp. 455–492, 1998.
- [4] Z. Wang and S. Jegelka, "Max-value entropy search for efficient bayesian optimization," in *International Conference on Machine Learning (ICML)*, 2017.
- [5] M. Schonlau, "Computer experiments and global optimization," Ph.D. dissertation, Waterloo, Ont., Canada, Canada, 1997, aAINQ22234.
- [6] I. Couckuyt, T. Dhaene, and P. Demeester, "oodace toolbox: A flexible object-oriented kriging implementation," *Journal of Machine Learning Research*, vol. 15, pp. 3183–3186, 2014.
- [7] F. Hutter, H. H. Hoos, and K. Leyton-Brown, "Sequential model-based optimization for general algorithm configuration," in *International Conference on Learning and Intelligent Optimization*. Springer, 2011, pp. 507–523.
- [8] J. T. Springenberg, A. Klein, S. Falkner, and F. Hutter, "Bayesian optimization with robust bayesian neural networks," in *Advances in Neural Information Processing Systems 29*. Curran Associates, Inc., 2016, pp. 4134–4142.
- [9] C. E. Rasmussen, "Gaussian processes in machine learning," in *Advanced lectures on machine learning*. Springer, 2004, pp. 63–71.
- [10] H. J. Kushner, "A new method of locating the maximum point of an arbitrary multipeak curve in the presence of noise," *Journal of Basic Engineering*, vol. 86, no. 1, pp. 97–106, 1964.
- [11] R. A. Fisher, *The genetical theory of natural selection: a complete variorum edition*. Oxford University Press, 1999.
- [12] N. Knudde, J. van der Herten, T. Dhaene, and I. Couckuyt, "GPflowOpt: A Bayesian Optimization Library using TensorFlow," *arXiv preprint - arXiv:1711.03845*, 2017. [Online]. Available: <https://arxiv.org/abs/1711.03845>