The 9th European Congress on Computational Methods in Applied Sciences and Engineering
ECCOMAS Congress 2024
3–7 June 2024, Lisboa, Portugal

GLOBALIZED TRUST-REGION OPTIMIZATION: AN UNASSUMING APPROACH

Yens Lindemans[1], Adrian Bekasiewicz[2], Ivo Couckuyt[1] and Tom Dhaene[1]

[1] Internet and Data Lab (IDLab)
Universiteit Gent
Ghent, Belgium
e-mail: yens.lindemans@ugent.be

[2] Faculty of Electronics, Telecommunications and Informatics
Gdansk University of Technology
Gdansk, Poland
email: bekasiewicz@pg.edu.pl

Key words: Gaussian Process, Bayesian Optimization, Antenna Optimization

Summary. Bayesian optimization (BO) with Gaussian process (GP) models form a powerful tool for the optimization of black-box functions. For this reason, a bi-stage multi-fidelity Bayesian framework is introduced for the optimization of a planar quasi-patch antenna for ISM band applications. The first stage of performance-oriented tuning involves exploration of the search space using a set of independent trust region (TR) based optimization runs executed on separate GP models. The evaluation of data-points in this step is performed using a numerically cheap low-fidelity model. In the second stage, the best identified regions are exploited using a multi-fidelity method that blends the information from the low- and high-fidelity simulations of the antenna into another GP. Data-fusion facilitates identification of high-quality design solutions using only a handful of expensive model simulations. Finally, the proposed method is benchmarked against a conventional BO where only high-fidelity data are used.

1 INTRODUCTION

The adoption of Internet of Things (IoT) devices has shown a continuous growth over the last few decades and it is expected to keep following this trend for the foreseeable future [1]. It is used daily in numerous devices such as smartphones, identification tags, or sensor networks and it has become indispensable for economic growth [2], [3]. Increased utilization of IoT devices is associated with a demand for more efficient and more robust wireless components that can be efficiently integrated within the IoT architecture. The development of such hardware is a tedious effort due to complex requirements and a large numbers of design variables. Circuit design constitutes only one step of the whole hardware development flowchart, but it is of significant importance regarding the efficiency of final product [4]. In this step, computer-aided-design (CAD) software is used to implement the device. With the help of these tools, numerical optimization techniques are adopted to further optimize the design and to ensure that the final device complies with all the necessary requirements. However, with the increasing complexity of wireless devices, CAD tools such as full-wave electromagnetic (EM) simulators suffer from high computational costs and lengthy simulation times [5], [6]. This challenge drives the search for faster optimization

algorithms capable of handling more parameters used for representation of contemporary structures, and hence, a more complex search space, since conventional population-based methods such as evolutionary algorithms or particle swarm optimization require numerous simulation iterations [7].

One research path that has gained significant attention of the community over the last years is optimization via machine learning (ML). Within this topic, several different strategies can be adopted in combination with various ML models. Examples include, but are not restricted to, design generation, dimensionality reduction, and surrogate modeling [6], [8]–[11]. The first considered concept involves application of generative ML models such as generative adversarial networks or diffusion models. These are trained on existing optimal device topologies and are subsequently used to generate more designs with, hopefully, similar or better performance characteristics [9], [12]. ML can also be used for dimensionality reduction through transformation of the design parameters into a low-dimensional latent space. Following this, any optimization algorithm can be deployed within the reduced space, thereby limiting overall computational cost [13]. Finally, surrogate modeling refers to building a cheap-to-evaluate representation of the structure at hand that tries to mimic the objective function. The surrogate model is then used to perform optimization in order to replace the costly, yet accurate, simulations [14], [15].

Within the domain of surrogate-based optimization, Gaussian processes (GPs) have become highly popular due to their analytic description of uncertainty [16], [17]. Owing to this capability, GPs are well-suited for use within a Bayesian optimization (BO) framework, which is known for its data efficiency in black-box optimization [18], [19]. Given the usefulness of BO, we employ this tool in our work. Specifically, we introduce a bi-stage Bayesian method for the multi-fidelity optimization of a planar quasi-patch antenna, depicted in Figure 1. This antenna comprises six geometrical parameters, highlighted by red boxes in Figure 1b, which can be adjusted to minimize the antenna's reflection. The motivation for exploring variations of conventional BO arises from the challenge that optimizing over six features is already considered *high-dimensional* [20]. By implementing our bi-stage Bayesian approach, we aim to enhance the optimization process for such high-dimensional spaces.

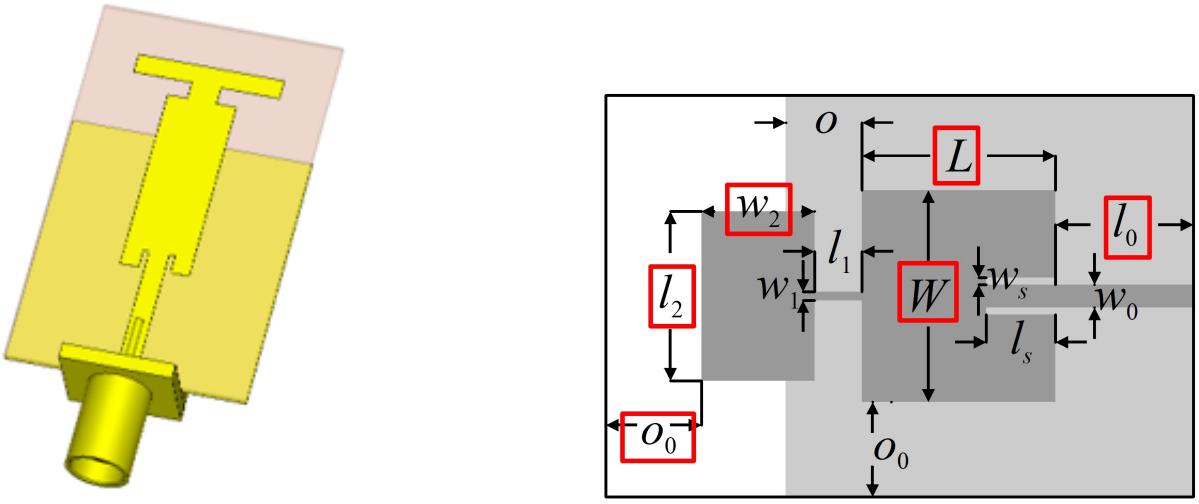
The subsequent part of this article is organized as follows: Section 2 introduces the proposed method in the BO framework and explains the choices that were made during the development of the method. Section 3 covers the results and compares them with standard BO methods. Finally, Section 4 concludes the article.

2 METHODOLOGY

The main challenge associated with design of antenna structures represented using EM simulation models is high computational cost as well as a large number of design parameters which hinders construction of accurate models (due to the curse of dimensionality). To try and mitigate these factors, different adaptations to the conventional BO method employing a GP surrogate model are proposed. In what follows, a brief description of GPs and BO will be given, followed by the new techniques that are adopted in this work.

2.1 Gaussian Processes

A Gaussian process can be seen as a distribution of functions, where the mean of that distribution represents the prediction of the model and the variance represents the uncertainty.



(a) (b)

Figure 1: A planar quasi-patch antenna for ISM band applications. (a) 3D schematic of the antenna. (b) top view of the antenna, with optimizable parameters indicated by red squares. The ISM band is an internationally reserved frequency-band for industrial, scientific, and medical applications. [7]

The initial distribution, i.e., the one provided before training data has been introduced, is called the *prior* and is defined as a multivariate normal with mean vector $\boldsymbol{\mu}$ and covariance matrix Σ. The model is trained by making use of Bayes' rule

$$\text{posterior} = \frac{\text{likelihood} \times \text{prior}}{\text{normalization constant}}\,, \tag{1}$$

which defines the *posterior* distribution, given a prior and likelihood. The likelihood represents the probability of obtaining the observed data, given the parameters of the model. Since Gaussian distributions are closed under marginalization and conditioning, updating them using Bayes' rule also yields a Gaussian distribution. This analytical tractability is very useful when determining the posterior distribution.

When employing a GP, the prior mean and covariance matrix can be seen as the hyperparameters of the model. For simplicity, the mean is set to zero. This approach is justified because the data can always be normalized to have a zero mean and then rescaled back to its original distribution after inference. The covariance matrix on the other hand has to be determined more carefully. It is derived from a covariance function or *kernel* $k(\boldsymbol{x}, \boldsymbol{x}')$ that ultimately defines the properties of the regression model. For example, if $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ are elements of the training set X, then the ij^{th} element of the covariance matrix is given by

$$[\Sigma]_{i,j} = k(\boldsymbol{x}_i, \boldsymbol{x}_j)\,. \tag{2}$$

If the kernel is a smooth function, the functions sampled from the GP will be smooth. If it is periodic, the functions sampled from the GP will be periodic. In essence, the kernel comprises

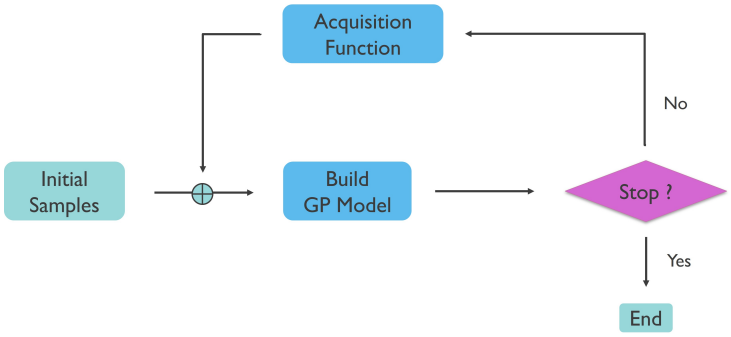


Figure 2: A flowchart showing the different steps in BO.

the properties of the GP model. A popular choice is the *Matérn* kernel

$$k(\boldsymbol{x}, \boldsymbol{x}') = \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\sqrt{2\nu \sum_{d=1}^{D} \frac{(x_d - x'_d)^2}{l_d^2}} \right)^{\nu} K_\nu \left(\sqrt{2\nu \sum_{d=1}^{D} \frac{(x_d - x'_d)^2}{l_d^2}} \right) , \tag{3}$$

where D is the dimensionality of the data, ν is a hyperparameter defining the smoothness, Γ and K_ν are the Gamma and Bessel function respectively, and l_d are length scales defined separately for each dimension. This kernel is used since the smoothness can be specified manually, making it more likely to coincide with the properties of the objective function. The value of this order parameter is chosen as $\nu = 5/2$. The length scales are part of the hyperparameters of the GP model and trained by maximizing the marginal log-likelihood of the posterior distribution. After training the length scales, the mean and covariance function of the posterior are then calculated through conditioning the GP. Specifically, if X represents the training set with labels Y and X^* represents a test set, the mean and covariance matrix evaluated on the test set given the training set are

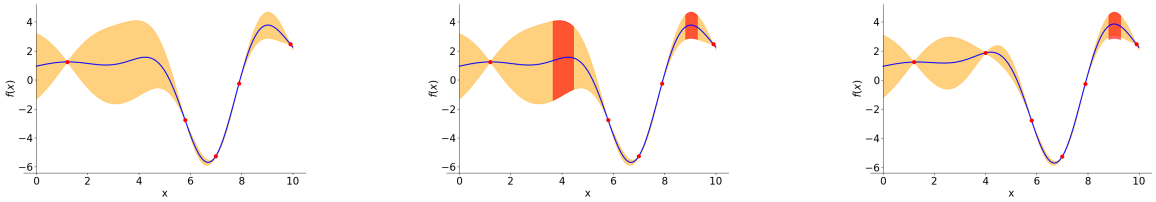
$$\mu_{X^*|X} = \Sigma_{X^*X} \Sigma_{XX}^{-1} \cdot Y \qquad \text{and} \tag{4}$$

$$\Sigma_{X^*|X} = \Sigma_{X^*X^*} - \Sigma_{X^*X} \Sigma_{XX}^{-1} \Sigma_{XX^*} , \tag{5}$$

respectively In these equations, the covariance matrices are built using Eq. (2).

2.2 Bayesian optimization

When working with surrogate models capable of expressing uncertainty and performing Bayesian inference, we can speak of BO when we use them in an optimization algorithm [18]. The general workflow of BO is illustrated in Figure 2: After obtaining some initial samples, a GP model is built and a stop condition is evaluated. If the stop condition is not met, an *acquisition function* (AF) has to be optimized in order to obtain new sampling points and restart the loop. The AF can be seen as an evaluation metric that gives a score to each point in the search space. The score is given based on how likely these points are to yield the optimum of the objective function. Different types of AFs exist, but the most common ones are probability of improvement, expected improvement, and upper confidence bound [21]. The manner in which these AFs differ is how much value they give to uncertainty or to expected optima. Essentially, choosing an AF is choosing how much the balance between exploration and exploitation.



(a) (b) (c)

Figure 3: Different steps of BO for maximizing a 1D example function. The blue line indicates the mean of the GP and hence the prediction of the model. The yellow region indicates the uncertainty. (a) Initial GP model built on initial samples (red dots). (b) Regions of interest (red zones) evaluated by the acquisition function. (c) GP updated with new sampling point.

An illustration of BO using the expected improvement AF is shown in Figure 3. Note that the regions of interest can either be interesting because the model predicts an optimum value at the given point, or because the model is very uncertain at that point.

As mentioned in the introduction, BO comes with some limitations. In particular, if the objective function consist of many local optima, the optimization scheme can take a long time to explore and find a global optimum. This problem is exacerbated by the *curse of dimensionality*, which refers to the phenomenon where points in a higher-dimensional space are relatively farther apart from each other than points in a lower-dimensional space. Since GPs work by measuring distances between points, as shown in Eq. (3), this effect can cause the model to perform poorly in high dimensions. It should be noted that six dimensions is not too high for GPs to handle properly, but as mentioned in [20], it can already cause some problems in BO.

To address these issues, two strategies are employed: trust region BO and multi-fidelity BO. The two methods are beneficial on different aspects of the optimization problem and hence can be used in conjunction with each other.

2.3 Trust-region Bayesian optimization

In trust region BO, different independent local GP models are trained on several trusted regions (TRs) spread throughout the search space to guide the process quicker towards a global optimum instead of a local one. A popular algorithm is the TuRBO algorithm, which stands for trust region Bayesian optimization and is developed in [22]. In TuRBO, a predefined number of equal-sized TRs are initialized randomly using Latin hypercube sampling and uses Thompson sampling to find promising points within these regions. Subsequently, the TRs are shifted towards the best solution found so-far within each region and the algorithm continues. Additionally, if the optimizer keeps finding better objective values, the TR will expand. On the other hand, if it fails to find a better optimum after a specified number of iterations, the TR will shrink. These expansions or shrinkages are proportional to the kernel's characteristic length scale of that particular dimension.

2.4 Multi-fidelity Bayesian optimization

As the EM simulations are performed using the CST Microwave Studio software package, a multi-fidelity BO (MFBO) strategy can be deployed by specifying different models to speed up

the optimization process. At the end of the process, it is important that the optimizer identifies a global solution using simulation data that closely resembles the real-world situation. However, it would be beneficial if part of the process can be done using low-fidelity simulations that are less accurate, but faster to generate. One way of doing this, is utilizing the method described in [23]. Here, multi-fidelity data can be used to train GPs at different fidelity levels and allows for all types of correlation between these different levels, wether they are linear, non-linear, or even non-existent.

The method is implemented by stacking multiple posteriors. Specifically, if X represents the input data and Y_i represents output data of fidelity level $i \in \{1, \ldots, N\}$, then the first step involves training a GP on the lowest fidelity level (X, Y_1). After training, the obtained posterior is used to predict output values $Y_1^* = \mathcal{GP}_1(X)$ for X, which are then collocated to X and given as new input to the GP representing the next level of fidelity. In other words, given two levels of fidelity, the input of the second fidelity GP is represented by $[X, Y_1^*]$ with output Y_2. It is important to mention that a posterior is used to produce the prediction values such that this method is not confused with a deep GP, i.e., a stacking of GPs that are trained simultaneously [24]. To make sure that the GP models are able to differentiate between the search space input X and the posterior predictions Y_i^*, the kernel functions are adapted as well. If a conventional kernel, such as Eq. (3), is given by $k(\boldsymbol{x}, \boldsymbol{x}')$, then the kernel for the i^{th} fidelity level is defined as

$$k_i = k_x(\boldsymbol{x}, \boldsymbol{x}')k_y(Y_{i-1}^*, Y_{i-1}^{*'}) + k_\delta(\boldsymbol{x}, \boldsymbol{x}')\,. \tag{6}$$

Here, three kernels all have independent length scales that are trained by maximizing the marginal log-likelihood. This particular combination of three kernels allows for identification of non-linear correlations between different fidelity levels [23].

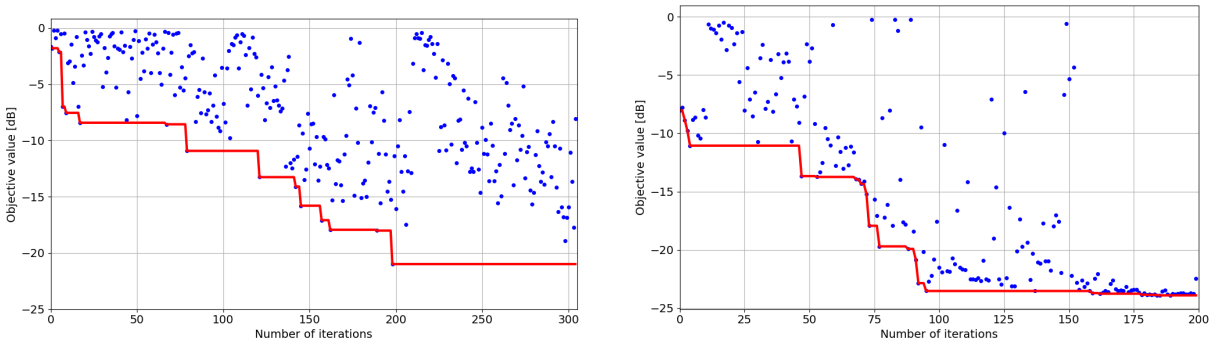
In this work, only two fidelity levels are used. They are distinguished by the fact that the low-fidelity data is generated using a coarser grid and excludes the feeding tube that goes into the antenna. The difference in simulation time is generally a factor 2: 60 s for a low-fidelity simulation and 120 s for a high-fidelity simulation.

3 RESULTS

The goal of this work is to optimize the planar quasi-patch antenna shown in Figure 1. Optimizing in this context is defined as minimizing the reflection in frequency range [5, 6] GHz. Practically, this can be achieved by tuning the aforementioned six geometrical parameters and running a CST simulation. The simulator then outputs the S-parameters of the antenna, from which the reflection coefficient is taken and post-processed into the objective function

$$f_{\text{obj}}(\boldsymbol{x}) = \min_{f \in [5,6]\,\text{GHz}} \Big\{ 20 \log(|R(f;\, \boldsymbol{x})|) \Big\}\,. \tag{7}$$

The proposed MFBO method is initiated by employing the TuRBO algorithm with low-fidelity data. The results of this step are shown in Figure 4a. Here, the algorithm uses five TR instances and a maximum number of 300 iterations. The reason for using TuRBO to gather low-fidelity data, instead of alternative techniques such as uniform or Latin hypercube sampling, is to concentrate training data in the more interesting regions of the search space. The approach is based on the assumption that regions of interest in low-fidelity data generally align with those in high-fidelity data. The algorithm clearly converges to a minimum, but as this is low-fidelity



(a) (b)

Figure 4: Convergence plot for (a) TuRBO algorithm and (b) the MFBO scheme proposed in this work. The blue dots represent the objective value found in each iteration and the red line shows the best solution found so far.

data, no conclusions can be drawn on the quality of the antenna at this point. In the next step, the multi-fidelity technique is employed and the high-fidelity data are generated along with the low-fidelity model predictions. Since TuRBO trains several local GPs, a global low-fidelity GP model is built with the data gathered in the first step. The results are depicted in Figure 4b for 200 iterations and show a rapid convergence towards optimal solutions characterized by objective values below -10 dB.

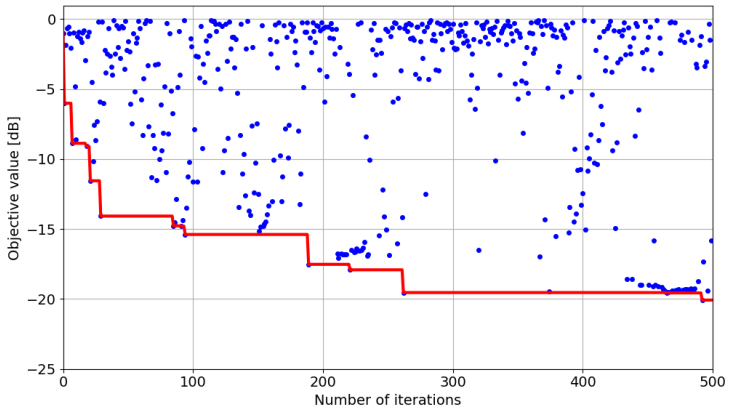
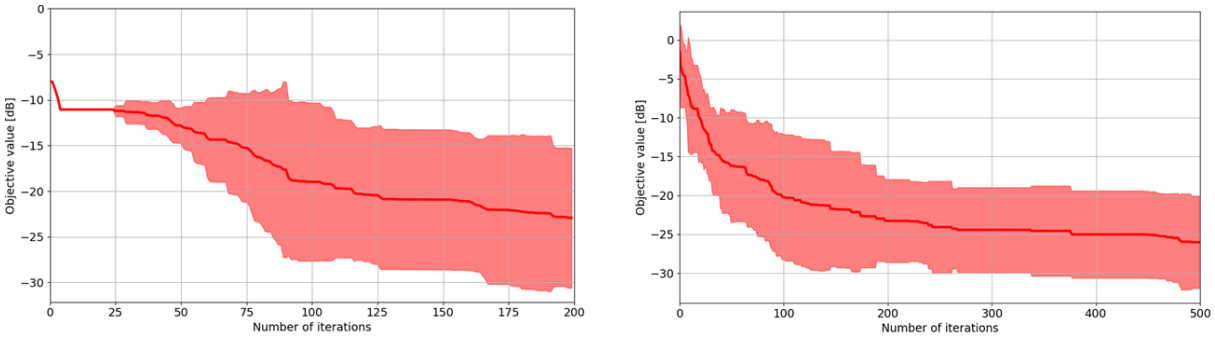


Figure 5: Convergence plot for the conventional BO on high-fidelity data only.

To compare the proposed technique with standard BO, several optimization runs are performed using a GP model with a Matérn kernel and trained using the high-fidelity input only. An example of a single run is shown in Figure 5, to ensure there is a fair comparison, ten optimization runs are performed using both techniques and their results are summarized in Figure 6. Note that the stop condition for the standard BO technique is set to 500 iterations, aligning with the MFBO method's total of 500 iterations (300 low-fidelity and 200 high-fidelity). However, due to a significant runtime difference between both methods, a fair comparison based on equal runtime requires setting the stop condition for standard BO at an average of 350 iter-



(a) (b)

Figure 6: Mean convergence plot for (a) the MFBO algorithm and (b) a standard BO scheme. The red line represents the mean of the different runs and the red colored region shows the standard deviation.

ations. Notably, the standard BO technique performs equally well in finding an optimum for the considered optimization problem. Nevertheless, there is a significant difference in how both methods explore the search space. Figure 4b shows a more targeted search with little samples taken far from an optimum, whilst the samples in Figure 5 are more scattered with only a few being located close to an optimum.

4 CONCLUSION

In this work, we introduced a bi-stage Bayesian framework for multi-fidelity optimization of a quasi-patch antenna for ISM band applications. The method is designed to address challenges such as the high computational costs of gathering simulation data and the presence of a multi-modal objective function. It is based on the principles of TuRBO and non-linear multi-fidelity modeling and is compared with conventional BO, which utilizes high-fidelity data only. The results show that both methods perform equally well in terms of finding a qualitative optimum. However, the proposed method demonstrates a more targeted exploitative search, with less exploration, whereas the conventional method focuses more on exploration.

Future work will focus on adaptation of the proposed methodology in a higher-dimensional setting. For instance, extending to 20 dimensions would push the boundaries of standard BO, potentially highlighting performance distinctions between the proposed method and the standard method [20]. Additionally, further improvements can be made by carefully tuning the number of TRs used by the TuRBO algorithm and by examining the performance when varying the ratio of low-fidelity samples to high-fidelity samples.

References

[1] M. Capra, R. Peloso, G. Masera, M. Ruo Roch, and M. Martina, "Edge computing: A survey on the hardware requirements in the internet of things world," *Future Internet*, vol. 11, no. 4, 2019, ISSN: 1999-5903. DOI: 10.3390/fi11040100.

[2] M. Mansour, A. Gamal, A. I. Ahmed, *et al.*, "Internet of things: A comprehensive overview on protocols, architectures, technologies, simulation tools, and future directions," *Energies*, vol. 16, no. 8, 2023, ISSN: 1996-1073. DOI: 10.3390/en16083465.

[3] D. Singh, "Internet of things," in *Factories of the Future*. John Wiley & Sons, Ltd, 2023, ch. 9, pp. 195–227, ISBN: 9781119865216. DOI: doi.org/10.1002/9781119865216.ch9. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/9781119865216.ch9.

[4] H. Evans. "Iot hardware development: A guide to all essential considerations." (2024), [Online]. Available: https://www.velvetech.com/blog/iot-hardware-development/ (visited on 06/24/2024).

[5] K. Sankaran, "Are you using the right tools in computational electromagnetics?" *Engineering Reports*, vol. 1, no. 3, e12041, 2019, e12041 eng2.12041. DOI: doi.org/10.1002/eng2.12041. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/eng2.12041.

[6] F. Lucchini, R. Torchio, V. Cirimele, P. Alotto, and P. Bettini, "Topology optimization for electromagnetics: A survey," *IEEE Access*, vol. 10, pp. 98 593–98 611, 2022. DOI: 10.1109/ACCESS.2022.3206368.

[7] A. Bekasiewicz, "Optimization of the hardware layer for iot systems using a trust-region method with adaptive forward finite differences," *IEEE Internet of Things Journal*, vol. 10, no. 11, pp. 9498–9512, 2023. DOI: 10.1109/JIOT.2023.3234107.

[8] M. M. Khan, S. Hossain, P. Mozumdar, S. Akter, and R. H. Ashique, "A review on machine learning and deep learning for various antenna design applications," *Heliyon*, vol. 8, no. 4, 2022.

[9] S. Shin, D. Shin, and N. Kang, "Topology optimization via machine learning and deep learning: a review," *Journal of Computational Design and Engineering*, vol. 10, no. 4, pp. 1736–1766, Jul. 2023, ISSN: 2288-5048. DOI: 10.1093/jcde/qwad072. eprint: https://academic.oup.com/jcde/article-pdf/10/4/1736/52600359/qwad072.pdf.

[10] F. Garbuglia, D. Deschrijver, and T. Dhaene, "On the role of bayesian learning for electronic design automation: A survey," *IEEE Electromagnetic Compatibility Magazine*, vol. 12, no. 4, pp. 77–84, 2023. DOI: 10.1109/MEMC.2023.10466425.

[11] F. Feng, W. Na, J. Jin, J. Zhang, W. Zhang, and Q.-J. Zhang, "Artificial neural networks for microwave computer-aided design: The state of the art," *IEEE Transactions on Microwave Theory and Techniques*, vol. 70, no. 11, pp. 4597–4619, 2022. DOI: 10.1109/TMTT.2022.3197751.

[12] Y. Zhong, P. Renner, W. Dou, G. Ye, J. Zhu, and Q. H. Liu, "A machine learning generative method for automating antenna design and optimization," *IEEE Journal on Multiscale and Multiphysics Computational Techniques*, vol. 7, pp. 285–295, 2022. DOI: 10.1109/JMMCT.2022.3211178.

[13] K. Peng and F. Xu, "Optimization of antenna performance based on vae-bpnn-pca," in *2022 International Conference on Microwave and Millimeter Wave Technology (ICMMT)*, 2022, pp. 1–3. DOI: 10.1109/ICMMT55580.2022.10023171.

[14] A. Belen, Ö. Tarı, P. Mahouti, M. A. Belen, and A. Caliskan, "Surrogate-based design optimization of multi-band antenna," *Applied Computational Electromagnetics Society Journal*, 2022.

[15] 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," *Engineering with Computers*, vol. 39, no. 3, pp. 1923–1933, 2023.

[16] 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, 2022, ISSN: 2079-9292. DOI: 10.3390/electronics11142267.

[17] P. Manfredi, "Probabilistic uncertainty quantification of microwave circuits using gaussian processes," *IEEE Transactions on Microwave Theory and Techniques*, vol. 71, no. 6, pp. 2360–2372, 2023. DOI: 10.1109/TMTT.2022.3228953.

[18] R. Garnett, *Bayesian optimization*. Cambridge University Press, 2023.

[19] X. Wang, Y. Jin, S. Schmitt, and M. Olhofer, "Recent advances in bayesian optimization," *ACM Comput. Surv.*, vol. 55, no. 13s, Jul. 2023, ISSN: 0360-0300. DOI: 10.1145/3582078.

[20] M. Binois and N. Wycoff, "A survey on high-dimensional gaussian process modeling with application to bayesian optimization," *ACM Trans. Evol. Learn. Optim.*, vol. 2, no. 2, Aug. 2022. DOI: 10.1145/3545611.

[21] J. Wilson, F. Hutter, and M. Deisenroth, "Maximizing acquisition functions for bayesian optimization," in *Advances in Neural Information Processing Systems*, S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, Eds., vol. 31, Curran Associates, Inc., 2018.

[22] D. Eriksson, M. Pearce, J. Gardner, R. D. Turner, and M. Poloczek, "Scalable global optimization via local bayesian optimization," in *Advances in Neural Information Processing Systems*, H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, Eds., vol. 32, Curran Associates, Inc., 2019.

[23] P. Perdikaris, M. Raissi, A. Damianou, N. D. Lawrence, and G. E. Karniadakis, "Nonlinear information fusion algorithms for data-efficient multi-fidelity modelling," *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 473, no. 2198, p. 20 160 751, 2017. DOI: 10.1098/rspa.2016.0751. eprint: https://royalsocietypublishing.org/doi/pdf/10.1098/rspa.2016.0751.

[24] M. M. Dunlop, M. A. Girolami, A. M. Stuart, and A. L. Teckentrup, "How deep are deep gaussian processes?" *Journal of Machine Learning Research*, vol. 19, no. 54, pp. 1–46, 2018.