An Efficient Method for Antenna Design Based on a Self-Adaptive Bayesian Neural Network Assisted Global Optimization Technique

Yushi Liu, Bo Liu *Senior Member, IEEE*, Masood Ur-Rehman *Senior Member, IEEE*, Muhammad Imran *Senior Member, IEEE*, Mobayode O. Akinsolu *Senior Member, IEEE*, Peter Excell *Senior Member, IEEE*, Qiang Hua

***Abstract*—Gaussian process (GP) is arguably the most widely used machine learning method in the surrogate model-assisted antenna design optimization area. Despite many successes, two improvements are important for GP-based antenna global optimization methods, including (1) the convergence speed (i.e., the number of necessary electromagnetic simulations to obtain a high-performance design), and (2) the GP model training cost when there are several tens of design variables and/or specifications. In both aspects, state-of-the-art GP-based methods show practical but not desirable performance. Therefore, a new method, called self-adaptive Bayesian neural network surrogate model-assisted differential evolution for antenna optimization (SB-SADEA), is presented in this paper. The key innovations include: (1) The introduction of Bayesian neural network (BNN)-based antenna surrogate modeling method into this research area, replacing GP modeling, and (2) a bespoke self-adaptive lower confidence bound method for antenna design landscape making use of the BNN-based antenna surrogate model. The performance of SB-SADEA is demonstrated by two challenging design cases, showing considerable improvement in terms of both convergence speed and machine learning cost compared to state-of-the-art GP-based antenna global optimization methods.**

***Index Terms*—Antenna design, Antenna optimization, Bayesian neural network, Computationally expensive optimization, Differential evolution, Lower confidence bound, Surrogate modeling**

I. Introduction

Evolutionary algorithms (EAs) are widely used in antenna design [1], [2]. Due to their ability to jump out of local optima, without the need of an initial design, and generality, they are showing advantages for many design cases. Differential evolution (DE) [3] and particle swarm optimization (PSO) [4] algorithms are arguably playing the leading role in EA-driven antenna design [2], [5], [6]. However, considering full-wave electromagnetic (EM) simulations are often needed to obtain accurate performance of a candidate design, and EAs often need tens of thousands of such EM simulations to obtain the optimal design, the optimization time can be prohibitive.

Y. Liu, B. Liu, M. Ur-Rehman and M. Imran are with James Watt School of Engineering, University of Glasgow, Scotland. (e-mail: {Yushi.Liu, Bo.Liu, Masood.UrRehman, Muhammad.Imran}@glasgow.ac.uk)

M. O. Akinsolu and P. Excell are with Faculty of Computing and Engineering, Wrexham Glyndwr University, U.K. (e-mail: {m.o.akinsolu,p.excell}@glyndwr.ac.uk).

Q. Hua is with Faculty of Engineering and Technology, Liverpool John Moores University, U.K. e-mail: q.hua@ljmu.ac.uk

Corresponding author: Bo Liu

Therefore, surrogate models, constructed by machine learning techniques, are trained to approximate the antenna performance obtained by EM simulations [7]. Hence, in the optimization process, many computationally expensive EM simulations can be replaced by computationally cheap surrogate model predictions. The optimization time can, therefore, be largely reduced. Note that this research focuses on *online* surrogate model-assisted antenna *global* optimization, where the surrogate model keeps being updated in each iteration.

Three critical factors for online surrogate model-assisted antenna global optimization are the surrogate modeling method, the search operators, and the model management method. The surrogate modeling method refers to the machine learning core and its associated bespoke operators for antennas. Search operators refer to the optimization engine. Model management method refers to the framework making surrogate modeling and optimization work harmoniously. Since prediction uncertainty is unavoidable, which may lead to wrong convergence, identifying high potential candidate designs under uncertainty to maintain correct convergence and optimal update of the surrogate model are the goals of model management. The three factors are strongly interconnected.

Among online antenna global optimization methods, the surrogate model-assisted differential evolution for antenna optimization (SADEA) series [8], [9], [10], [11] is one of the state-of-the-art approaches. For design cases with fewer than 20 design variables and a few specifications, the latest P-SADEA method [10], [12] obtains better design quality than DE and PSO, while improving the speed by up to 30 times without parallel computing. P-SADEA shows success in challenging antenna cases where DE and PSO fail to obtain feasible designs, e.g., [13]. For antenna cases with several tens of design variables and/or specifications, machine learning cost becomes a new challenge, which may be even higher than EM simulation cost [11]. To the best of our knowledge, the TR-SADEA method firstly addresses this challenge and shows success for complex base station antennas [11].

In the evolution of the SADEA series, the main innovations locate in the model management method and the search operators. Considering the surrogate modeling method, to the best of our knowledge, not only the SADEA series but Gaussian process (GP) machine learning is also the routine of most state-of-the-art surrogate model-assisted antenna optimization methods [14], [15], [16], [17]. For the often highly multimodal antenna design landscape [10], a strong machine learning

method is essential. GP stands out due to its mathematical soundness, leading to stronger learning ability, with very few empirical parameters, and can provide a rigorous prediction uncertainty for each candidate design in optimization.

However, GP has its drawbacks, i.e., its training cost [18]. GP's training time grows cubically with the number of training data points, which is highly related to the number of design variables and specifications. Considering normal desktop computers (e.g., Intel i7 3.0 GHz CPU), for antennas with a few design variables, GP training time is short. However, for antennas with more than 20 and even around 50 design variables with more than 10 specifications, TR-SADEA uses 1-2 days for GP model training [11] (note that 90% of the expected training time compared to standard SADEA is reduced by a new GP model sharing method). Clearly, this training cost is practical but not desirable.

For design cases with very challenging specifications, e.g., [13], most surrogate model-assisted antenna global optimization methods still need a few thousand EM simulations, sometimes costing more than a week. Hence, much space is left for improving the convergence speed (i.e., the necessary number of EM simulations). The long GP training time restricts the number of surrogate models to be built, which also affects the convergence speed. For example, an idea like deep supervision [19], [20] in image recognition is suitable to many antenna problems to improve the convergence speed at the cost of training more surrogate models. However, considering GP modeling time, it is difficult to be used even for design cases with a moderate but not small number of design variables/specifications.

Owing to this, this research aims to seek a new machine learning method to replace GP-based antenna modeling and propose a method employing the new surrogate model. The goal is to considerably improve both the convergence speed and training cost and provide a universal method for antennas with various numbers of design variables and specifications. Hence, a new method, called self-adaptive Bayesian neural network surrogate model-assisted differential evolution for antenna optimization (SB-SADEA), is proposed. The key innovations include: (1) The introduction of Bayesian neural network (BNN)-based antenna surrogate modeling method into this area to replace GP, and (2) a bespoke self-adaptive lower confidence bound method for antenna design landscape making use of the BNN results as a part of model management.

The remainder of the paper is organized as follows. Section II presents the background knowledge. Section III elaborates on the SB-SADEA method. Section IV presents the advantages of SB-SADEA using a compact ultrawideband (UWB) monopole antenna (10 design variables, 2 specifications) and a compact multiband 5G mm-wave antenna (20 design variables, 12 specifications), both with challenging specifications. The concluding remarks are provided in Section V.

II. Background Knowledge

A. Online Surrogate Model-assisted Antenna Design Optimization

Machine learning is attracting much attention in electromagnetics recently [21], and surrogate model-assisted antenna optimization is an essential part, which can be classified into offline and online. In offline methods, a high-quality surrogate model is firstly built and there are no or few updates of the surrogate model in the optimization process [22]. The advantage is that the resulting effective antenna surrogate model is useful on many occasions, e.g., antenna circuit co-design, multiobjective Pareto optimization, showing excellent results [14]. The limitation is the "curse of dimensionality" [23]. When there are more than a few design variables and/or the modeling range is not narrow, the necessary number of EM simulations needed for building a high-quality surrogate model could be many, which grows exponentially with the number of design variables. Carrying out those simulations could cost tremendous time, canceling out the time saved by using surrogate models. Hence, design optimization is often a by-product for state-of-the-art antenna surrogate modeling methods.

Online methods, in contrast, keep improving the surrogate model quality in the optimization process. In each iteration, (a) surrogate model(s) is/are built using available simulated candidate designs. New candidate designs are generated by search operators, which are then predicted by the surrogate model. Using the prediction result, candidate designs with high potential are simulated and used to update the surrogate model for the next iteration. Hence, the surrogate model quality is not always high but is gradually improved. Particularly, at the beginning stage, the surrogate model quality may be poor due to the lack of training data points.

Unlike assuming an accurate surrogate model like offline methods, prediction quality and uncertainty largely affect the optimization. When the prediction quality is insufficient and the prediction uncertainty is not appropriately considered, the optimization is highly likely to converge to a local optimum far away from the design specifications [23], [24]. Therefore, appropriate collaboration of the surrogate modeling method, the search operators, and the model management method (i.e., the three key factors in Section I) is essential.

An important task in model management is to co-use the prediction uncertainty together with the predicted value to judge the potential of a candidate antenna design. This is also called prescreening or acquisition function [25], which will be discussed in the next subsection. Explicitly providing the prediction uncertainty for each candidate design (instead of the overall uncertainty of the surrogate model) is necessary for most prescreening methods.

B. GP Modeling and Prescreening, Advantages and Drawbacks

The basic principle of GP is as follows [18]. Giving n observations of $y(x)$ ($x = (x^1, \ldots, x^n)$ and $y = (y^1, \ldots, y^n)$), GP assumes that $y(x)$ is a sample of a Gaussian distributed stochastic process with mean μ and variance σ. GP then predicts the value of $y(x)$ for a new x using its relation with the n observations. For example, a correlation function can be described as (1).

$$\begin{aligned} &\mathrm{Corr}(x^i, x^j) = \exp(-\textstyle\sum_{l=1}^{d} \theta_l |x_l^i - x_l^j|^{p_l}) \\ &\theta_l > 0, 1 \le p_l \le 2 \end{aligned} \tag{1}$$

where d is the dimension of x. θ_l and p_l are hyper-parameters, which are determined by maximizing the likelihood function in (2).

$$\frac{1}{(2\pi\sigma^2)^{n/2}\sqrt{\det(R)}}\exp\left[-\frac{(y-\mu I)^T R^{-1}(y-\mu I)}{2\sigma^2}\right] \quad (2)$$

where R is a $n \times n$ covariance matrix and I is a $n \times 1$ vector having all its elements as unity. By maximizing the likelihood function that $y = y^i$ at $x = x^i (i = 1, \ldots, n)$ and handling the prediction uncertainty based on the best linear unbiased prediction, for a new point x^*, the predicted value and prediction uncertainty are $\hat{y}(x^*)$ and $\hat{s}^2(x^*)$, which are as follows.

$$\hat{y}(x^*) = \mu + r^T R^{-1}(y - I\mu) \quad (3)$$

where

$$\begin{aligned} &R_{i,j} = \text{Corr}(x^i, x^j), i, j = 1, 2, \ldots, n \\ &r = [\text{Corr}(x^*, x^1), \text{Corr}(x^*, x^2), \ldots, \text{Corr}(x^*, x^n)] \\ &\hat{\mu} = (I^T R^{-1} y)(I^T R^{-1} I)^{-1} \end{aligned} \quad (4)$$

$$\hat{s}^2(x^*) = \hat{\sigma}^2[I - r^T R^{-1} r + (I - r^T R^{-1} r)^2 (I^T R^{-1} I)^{-1}] \quad (5)$$

where

$$\hat{\sigma}^2 = (y - I\widehat{\mu})^T R^{-1}(y - I\widehat{\mu}) n^{-1} \quad (6)$$

Based on the above, two advantages of GP include: (1) There are almost no empirical parameters in GP modeling except deciding the type of correlation function; A few appropriate correlation functions are already found by antenna surrogate modeling researchers [14], [15], [16], [8]. Hence, overfitting or underfitting like artificial neural networks (ANNs) is less likely to happen, which improves the prediction quality. (2) The prediction uncertainty (5) is rigorous, which can play an important role when judging the full potential of a candidate antenna design in prescreening or acquisition function.

In contrast, many fitting-based machine learning methods, such as most kinds of ANN, radial basis function [26], response surface models either cannot provide a prediction uncertainty for each candidate design or the prediction uncertainty is not rigorous. They are, therefore, less suitable for online surrogate model-assisted antenna global optimization because this largely affects the convergence speed and the ability to jump out of local optima. Although researchers suggest using the dropout method in ANN training to provide a prediction uncertainty [27], [28], [29], our experiments using antenna problems show much worse results than GP.

With the prediction uncertainty, widely used prescreening methods include expected improvement [25], probability of improvement [30], and lower confidence bound (LCB) [31]. LCB is the fundamental of the new prescreening method in this paper and is introduced as follows. Given the objective function $y(x)$ has a predictive distribution of $N(\hat{y}(x), \hat{s}^2(x))$, an LCB prescreening of $y(x)$ is:

$$\begin{aligned} &y_{\text{lcb}}(x) = \hat{y}(x) - \omega\hat{s}(x) \\ &\omega \in [0, 3] \end{aligned} \quad (7)$$

where ω is a constant, and is often set to 2 in the AI domain [23], and is applicable to antenna problems [8].

However, the main drawback of GP is its training cost. In online surrogate model-assisted antenna global optimization, the total training time of GP models in the optimization process can be estimated as $T_{GP} \times N_{specs} \times N_{pop} \times N_{it}$ [11], where T_{GP} is the training time of each GP model, N_{specs} is the number of specifications, N_{pop} is the number of candidate designs in a population, and N_{it} is the number of iterations in antenna design optimization.

For a GP model, the computational complexity is $O(N_o n^3 d)$ [23], where N_o is the number of iterations spent in hyper-parameter optimization (i.e., (2)) and n is the number of training data points. n is highly affected by d in order to construct a reliable surrogate model. [8], [11] shows that at least $4 \times d$ training data points are needed for antenna problems. Often, when d reaches 20, T_{GP} could be in minutes for a normal computer and then grows cubically [11]. Also, to maintain the exploration ability, N_{pop} is also highly affected by d (e.g., often at least $4 \times d$ when using DE operators). This makes the GP modeling time in antenna optimization become long when d is large and could be even longer than EM simulation time.

To the best of our knowledge, TR-SADEA [11] is the first method addressing this challenge. A self-adaptive GP model sharing method is proposed aiming to highly decrease the number of GP modeling while maintaining correct convergence. Although this method decreases the GP model training time to be practical, the time consumption (e.g., 1-2 days for complex base station antennas with several tens of design variables and specifications) is not desirable.

C. Antenna Design Optimization and the DE Algorithm

Antenna design optimization can be classified into local optimization [32] and global optimization [1], [6]. When an initial design is available, the former is more appropriate; otherwise, the latter is more appropriate, on which, this paper focuses. In antenna global optimization, EAs were introduced into this area two decades ago and are widely used. Multi-start local search and surrogate model-assisted multi-start local search for antenna global optimization are introduced recently [17], [33], showing successful results. This paper focuses on using EAs. As said in Section I, DE and PSO are arguably the most widely used EAs in antenna global optimization [2], [5], [34], [35], and DE operators are used in this work. A brief introduction to DE is as follows [3].

P is a population composed of N_{pop} individual solutions $x = (x_1, \ldots, x_d) \in R^d$. To create a child solution $u = (u_1, \ldots, u_d)$, firstly, mutation happens to generate a donor vector:

$$v^i = x^i + F \cdot (x^{best} - x^i) + F \cdot (x^{r_1} - x^{r_2}) \quad (8)$$

where x^i is the i^{th} vector in the current population; x^{best} is the best candidate solution in the current population P; x^{r_1} and x^{r_2} are two mutually exclusive solutions randomly selected from P; v^i is the i^{th} mutant vector; $F \in (0, 2]$ is the scaling factor. The mutation strategy in (8) is called DE/current-to-best/1, which is used in SB-SADEA.

The crossover operator then happens to produce the child solution u:

1 Randomly select a variable index $j_{rand} \in \{1, \ldots, d\}$,
2 For each $j = 1$ to d, generate a uniformly distributed random number $rand$ from $(0, 1)$ and set:

$$u_j = \begin{cases} v_j, & \text{if } (rand \leq CR) | j = j_{rand} \\ x_j, & \text{otherwise} \end{cases} \quad (9)$$

where $CR \in [0, 1]$ is the crossover rate.

III. The SB-SADEA Algorithm

A. The Algorithm Framework

As discussed in Section I, this research aims to seek a new machine learning core to replace GP and propose a new prescreening method for it. The resulted SB-SADEA algorithm is expected to significantly improve both the convergence speed (i.e., the number of EM simulations used to obtain the optimal design) and the surrogate model training cost compared to GP-based methods. Also, it should be universal for antenna cases with various numbers of design variables and specifications. In the following, the general framework of SB-SADEA is first provided and details of the two innovations are then described in Section III (B) and (C).

The SB-SADEA framework is shown in Fig. 1, and the algorithm works as follows.

Step 1: Sample α (often a small number of) candidate designs from the design space $[LB, UB]^d$ (LB and UB are the lower and upper bounds of design variables, respectively) using Latin Hypercube Sampling (LHS) [36]. Carry out EM simulations to obtain their performance values and form the initial database.

Step 2: If a preset stopping criterion is met (e.g., the computing budget is exhausted, satisfy the specifications), output the best candidate design from the database; otherwise go to Step 3.

Step 3: Select the λ best candidate designs from the database to form a population P.

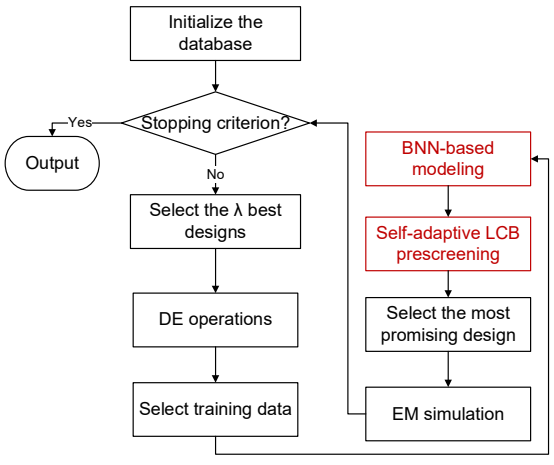


Figure 1. Flow diagram of SB-SADEA

Step 4: Apply the DE/current-to-best/1 operator (8) to P to create λ new child solutions.

Step 5: For each child solution, select τ nearest samples (based on Euclidean distance) as the training data points and construct a Bayesian neural network (BNN)-based surrogate model (Section III (B)).

Step 6: Prescreen the child solutions generated in Step 4 using the BNN model and the self-adaptive LCB method (Section III (C)).

Step 7: Carry out EM simulation to the estimated best child solution from Step 6. Add this evaluated candidate design and its performance values to the database. Go back to Step 2.

It can be seen that some model management operators are borrowed from standard SADEA [8]. This model management method is attracting much attention in the AI domain [22] and its advantages are detailed in [8], [10], [22]. The two novel methods, including the BNN-based antenna surrogate modeling (Step 5) and the self-adaptive LCB method (Step 6), which are red blocks in Fig. 1 are introduced in the following subsections. Note that they are compatible with model management frameworks in other SADEA versions as well as some other online antenna global optimization methods.

B. The Bayesian Neural Network-based Antenna Surrogate Modeling Method

To replace GP modeling, the machine learning core must satisfy the following requirements: (1) Can provide a high-quality predicted value comparable to GP and has less risk to be overfitted; (2) The prediction uncertainty of each candidate design is statistically grounded; and (3) The training cost is much lower than GP when the targeted antenna has many design variables/specifications.

To meet the above requirements, an alternative is a stochastic neural network, in particular, a Bayesian neural network. To the best of our knowledge, BNN has not been used in antenna design optimization and is seldom considered for surrogate model-assisted optimization even in the AI domain. In the following, BNN concepts are introduced.

Considering the antenna design variables as x, and the performance as y, for an ANN, the model parameters are $\theta = [w_1, \ldots, w_j, b_1, \ldots, b_k]$, where w are the weights and b are the biases. In a multi-layer ANN, each layer is a linear transformation, followed by a nonlinear activation function. The training optimizes the cost function, which is often the log likelihood of the training data points, i.e., maximize $\sum_{i=1}^{n} \log(p(x^i; \theta))$ with a regularization term. The optimized θ, which are fixed values, are then used for prediction.

For BNN, the network structure does not change compared to a standard ANN, but θ become stochastic variables with their probability distribution $p(\theta)$. Fig. 2 shows an illustrative figure of BNN. By applying Bayes' theorem, the Bayesian posterior can be expressed as:

$$\begin{aligned} p(\theta|x, y) &= \frac{p(y|x, \theta)p(\theta)}{\int_\theta p(y|x, \theta')p(\theta')\mathrm{d}\theta'} \\ &\propto p(y|x, \theta)p(\theta), \end{aligned} \quad (10)$$

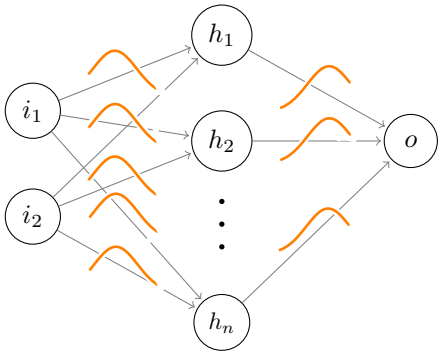


Figure 2. An illustrative figure of a basic BNN.

where $p(y|x,\theta)$ is the likelihood, $p(\theta)$ is the prior, the denominator is the evidence, and $p(\theta|x,y)$ is the posterior. The posterior is what we acquire, which is used in obtaining the predicted value and prediction uncertainty. Obtaining $p(\theta|x,y)$ by standard sampling method is intractable. Hence, the variational inference method [37] is proposed.

In variational inference, a new distribution $q(\phi)$ (ϕ are the model parameters), called a variational distribution, is proposed to approximate $p(\theta|x,y)$. By minimizing the Kullback-Leibler (KL) divergence between $q(\phi)$ and $p(\theta|x,y)$, the closest distribution can be found to replace the posterior. Compared to the posterior, $q(\phi)$ has a smaller set of parameters, which are often considered as means and variances of a multivariate Gaussian distribution, and are optimized in training.

The cost function to be maximized is:

$$E_{\phi\sim q}(\log p(y|x;\phi)) - D_{KL}(q(\phi)||p(\theta)) \tag{11}$$

(11) is called the evidence lower bound. The first term is the standard maximum likelihood loss and the second term is the regularization loss, which is a closed form for Gaussian distribution. The first item can be obtained by sampling. After the optimization or training, the posterior, $p(\theta|x,y)$, is approximated, and the BNN is ready to be used.

When performing prediction by BNN, given the posterior, $p(\theta|x,y)$, the model's prediction uncertainty can be derived from $p(y|x,x,y)$. Mathematically, it can be written as

$$p(y|x,x,y) = \sum_{\theta} p(y|x,\theta')p(\theta'|x,y)\mathrm{d}\theta', \tag{12}$$

In practice, this is done by sampling [38].

$$\theta \sim p(\theta|x,y). \tag{13}$$

The predicted value is the average of BNN model output samples.

$$\hat{y} = \frac{1}{|\Theta|}\sum_{\theta_i\in\Theta}\Phi_{\theta_i}(x), \tag{14}$$

where $\Phi_\theta(x)$ is the BNN model and $\hat{y}$ is the estimated output. The uncertainty quantification is given by the covariance matrix $\Sigma_{y|x,x,y}$, which is:

$$\Sigma_{y|x,x,y} = \frac{1}{|\Theta|-1}\sum_{\theta_i\in\Theta}(\Phi_{\theta_i}(x)-\hat{y})(\Phi_{\theta_i}(x)-\hat{y})^T. \tag{15}$$

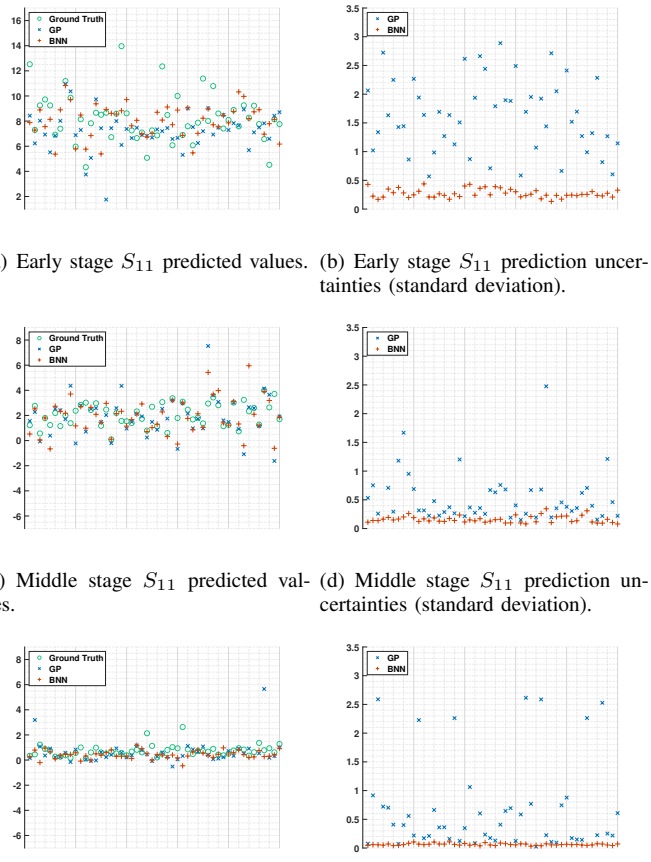
Some clarifications in terms of the requirements at the beginning of this subsection are as follows.

- BNN has a good potential to provide high-quality prediction results. BNN can be interpreted as a special case of ensemble methods [39]. Ensemble methods are well known for taking advantage of the fact that the aggregation of multiple averaged and independent predictors may outperform a single expert predictor, given the same training information [40]. BNN's stochastic components similarly improve a normal ANN. Also, BNN can avoid overfitting when learning from a small dataset (i.e., available training data points via EM simulations) by considering both aleatoric uncertainty and epistemic uncertainty, as evidenced in [41].
- The prediction uncertainty of BNN is statistically grounded [42], [43]. Intuitively, as in (15), for any input x, low uncertainty is given when multiple sample models yield close estimated outputs $\hat{y}$; high, otherwise.
- BNN has much less training complexity compared to GP. As discussed earlier, the computational complexity of a single GP model is $O(N_o n^3 d)$. In the SADEA series, n is linearly increased with d and there are m specifications. Hence, the complexity is $O(N_o d^4 m)$. For the BNN used in SB-SADEA, which uses two hidden layers, the computational complexity is $O(N_b d(d+m)^2 s)$, where s is the sampling cost, N_b is the number of iterations in training. More verifications are in Section IV.

Due to the considerably reduced training cost of BNN, an idea inspired by deeply supervised nets [19] for image recognition is proposed, which we call it "fine supervision". Often, the antenna response over the operating band is considered as a whole and a maximum or minimum is obtained as the performance (e.g., $max(|S_{11}|)$). By using a point to represent a whole curve, much information is lost.

In the proposed fine supervision, the response curve is divided by resonances, and for each part of the response curve, its maximum or minimum value is used. In this way, much more information participates in the learning with the cost of increasing the number of specifications (i.e., number of surrogate models). This is a significant burden to GP modeling when the number of design variables is large because several times more GP models need to be trained [11]. However, for BNN, this is affordable because only the number of neurons in the output layer increases. This conclusion is verified by case study 2 (4-band 5G mm-wave antenna) in Section IV.

The parameter setting of BNN is as follows. The BNN structure has 2 hidden layers and the number of neurons are d (input layer), $2d$ (the first hidden layer), $max([d, 2m])$ (the second hidden layer) and m (the output layer), respectively. The prior standard deviation is defaulted to be 0.1. The Adam optimizer is used for training with the initial learning rate of 0.05, and a decay rate of 0.999 in every step of the model parameter optimization. An early stop criterion is set within the training to stop any training proceeding with an insignificant loss decrease. All the above are based on rules of thumbs or empirical settings and are verified by antennas with various characteristics.



(a) Early stage S_{11} predicted values. (b) Early stage S_{11} prediction uncertainties (standard deviation).

(c) Middle stage S_{11} predicted values. (d) Middle stage S_{11} prediction uncertainties (standard deviation).

(e) Late stage S_{11} predicted values. (f) Late stage S_{11} prediction uncertainties (standard deviation).

Figure 3. GP and BNN predicted values and prediction uncertainty during early, middle and late stage of the optimization (ground truth is from EM simulations).

C. The Self-adaptive Lower Confidence Bound Method

For a machine learning core (i.e., BNN-based model in this case), a corresponding prescreening method (Step 6 in SB-SADEA) is often needed considering its own data characteristics. Most existing prescreening methods consider the data characteristics of the GP model. In our pilot experiments, the widely used expected improvement [25] and probability of improvement [30] prescreening are employed together with the BNN-based model for multiple antenna cases and more than 50% of the runs are stuck in local optima. This invites us to study the difference in data characteristics between BNN and GP in terms of predicted values and prediction uncertainty, so as to propose a prescreening method that can jump out of local optima and also improve the convergence speed for the BNN-based model.

Using case study 1 in Section IV (i.e., UWB monopole antenna), Fig. 3 shows the GP's and BNN's predicted values and prediction uncertainty for three sample populations of candidate designs during the early, middle, and late stages in the optimization process. $max(|S_{11}|)$ is used. It can be observed that: (1) In terms of the predicted values, the BNN-based model and GP-based model are comparable, and both show reasonably low prediction error compared to the simulated values (i.e., ground truth) considering all three sample populations. (2) In terms of the prediction uncertainty, the BNN-based model shows much smaller values than that of GP, and the gap between them is much clearer in the later iterations. For example, when the optimization is at its late stage and nearly converges, the BNN prediction uncertainty is at the level of 0.05, while GP prediction uncertainty is at the level of 0.5.

The reason why a surrogate model-assisted antenna global optimization method falls into local optima is the lack of exploration ability. In the optimization theory, exploration refers to exploring the search region that currently lacks knowledge, while exploitation refers to finding the optimum in the search region with sufficient knowledge. Antenna design landscapes are often highly multimodal, and strong exploration ability is required [10]. Fully considering prediction uncertainty is important for exploration, which is the reason for prescreening methods. For the popular expected improvement and potential of improvement prescreening methods, there are no hyperparameters controlling the extent of exploration, and the prediction uncertainty obtained by the BNN-based model is small. Hence, it is not a surprise that using a BNN-based model often leads to falling into a local optimum for antenna cases compared to using the GP model.

A solution is using the LCB prescreening method (7) [31], which has a hyperparameter ω to control the extent of exploration. The value of ω can be set empirically using experiments with various antenna design cases, and the recommended value is 14. Clearly, using a large value for ω can promote the exploration ability, but high exploration ability inevitably slows down the convergence (i.e., more EM simulations) due to no-free-lunch. Hence, a novel method to obtain the appropriate trade-off, called self-adaptive LCB, is proposed. Given the λ current best candidate designs, called P_b, and a vector called S, where $S_i(i = 1, 2, \ldots, k)$ saves the smallest distance between the current predicted best candidate design to all candidate designs in P_b in each iteration, self-adaptive LCB (Step 6 in SB-SADEA) works as follows.

Step 1: Select the best candidate design x_b in the child population in Step 4 of SB-SADEA using the BNN-based model predicted values.

Step 2: Calculate the distance between x_b and each individual in P_b and obtain the smallest distance, S_{k+1}.

Step 3: Taking the last 10 elements of S, check if S_{k+1} from Step 2 is smaller than $\hat{S}(k-9:k) - 0.5 \times \sigma(S(k-9:k))$. If yes, go to Step 4; Otherwise, output x_b.

Step 4: Prescreen the child population using the LCB method (7) with the recommended ω value.

Step 5: Output the best candidate design according to LCB values.

Some clarifications are as follows.

- The self-adaptive LCB method alternatively uses the BNN model predicted value and the LCB prescreened value for selecting the estimated best candidate design from the child population. The former is for promoting exploitation so as to improve the convergence speed,

while the latter is for promoting exploration for jumping out of local optima.
- Whether the algorithm has sufficient exploration ability or not highly depends on the diversity of P_b (Step 3 of SB-SADEA). Hence, the predicted values are used when the diversity is reasonable, while LCB values are imposed when the diversity is small.
- The method to judge the extent of introduced diversity is to compare with the smallest distance to any individual in P_b with those in the last 10 iterations. Assuming $S_i(i = 1, 2, \ldots, k)$ is Gaussian distributed, the $0.5 \times \sigma$ value is used as the threshold to find those introducing low diversity to P_b when using them.

D. Parameter Settings

Compared to standard SADEA [8], SB-SADEA only introduces one new parameter, ω, in the self-adaptive LCB method. Using various challenging antennas from fewer than 10 design variables to 45 design variables, from a few specifications to nearly 20 specifications, ω is suggested to be set to 14 for successfully jumping out of local optima. In BNN modeling, the network parameters are pre-decided by the rules of thumb and do not need the users to alter. For all other parameters, the setting rule in other SADEA versions is still applicable to SB-SADEA, which are: $\alpha = 4 \times d$, $\lambda = 4 \times d$, $\tau = 4$, $F = 0.8$, $CR = 0.8$. They are used in all the test cases in Section IV.

IV. Experimental Results and Verifications

SB-SADEA is tested by 7 challenging antennas with various characteristics and the comparisons show the same conclusion. In this section, two typical cases from them are used to demonstrate SB-SADEA's performance in different aspects.

The first case study is a slotted monopole antenna for UWB microwave imaging applications [44]. The antenna has 10 design variables and 3 specifications. The design optimization of this antenna is challenging due to its compact size to ensure proper physical placement and integration of its antenna structure with compact components on the same printed circuit board. For antennas with 10 design variables and 3 specifications, the machine learning cost using most methods is often small. Hence, the purpose of this case study is to test SB-SADEA's convergence speed (i.e., the number of EM simulations needed to obtain the optimal design) when facing stringent design specifications.

The second case study is a 4-band mm-wave antenna for wearable 5G and beyond applications [45]. It has 20 design variables and 12 specifications. The design optimization of high-performance 5G mm-wave antenna is often challenging [46] and this case study has particular challenges due to its compact size, lightweight, low profile, and low maintenance with a simple off-centered microstrip feeding structure. Moreover, maintaining a multi-band, high gain operation in wearable scenarios for body-centric wireless communications at mm-wave frequencies increases the design complexity and sensitivity. Considering the number of design variables and specifications, the machine learning cost can be considered computationally expensive for GP-based methods. Although antennas with more design variables and specifications can make the advantages of SB-SADEA even clearer, considering the time to draw statistical conclusions (i.e, using sufficient runs), this antenna is selected as a representative. Hence, the purpose of this case study is to test SB-SADEA's performance in terms of both convergence speed and machine learning cost.

For both antennas, because no reasonably good initial designs can be provided, the search ranges provided by the antenna designers are relatively wide, although restricted by the compact size. The antennas are optimized in a workstation with an AMD Ryzen™ 9 3900X 12-core processor (3.8GHz) and an NVIDIA® GeForce® GT 710 GPU. 40 MATLAB parallel workers are activated for GP/BNN-based surrogate model training.

The SADEA series are stochastic algorithms and 10 independent runs are carried out for case study 1 to draw statistical conclusions. 10 runs of design optimization are expected to be over a month for case study 2, and 5 independent runs are carried out. P-SADEA [10], [12] is selected as the reference method for case study 1 since it is one of the state-of-the-art methods for antenna design global optimization with fewer than 20 design variables with a few specifications. PSO, as one of the most popular evolutionary algorithms for antenna global optimization, is also used a reference.

TR-SADEA [11] is selected as the reference method for case study 2, since to the best of our knowledge, it is the only published method for antenna global optimization with many design variables and specifications, addressing the challenge in machine learning cost. DE, as one of the most popular evolutionary algorithms for antenna global optimization, is also used as a reference. In terms of parameter setting, SB-SADEA follows Section III (D), P-SADEA follows [10] and TR-SADEA follows [11]. The PSO optimizer with default parameters in Computer Simulation Technology - Microwave Studio (CST-MWS) is used. The setting of the DE optimizer follows [3].

A. Case Study 1: A Compact UWB Slotted Monopole Antenna

The layout of the slotted monopole antenna is shown in Fig. 4. The antenna is implemented on an FR-4 substrate with a thickness of 0.8 mm, a relative permittivity of 4.4, and a loss tangent of 0.02. It consists of a driven circular patch radiator and two uniform rectangular metal planes separated by the microstrip line. Two slots are fused at the center of the driven circular patch radiator to form a quasi-cross slot, and the geometry of the slot helps control the surface current distribution. Meanwhile, the rectangular planes act as a coplanar partial ground.

The slotted monopole antenna is modeled and discretized in CST-MWS with over 162,000 mesh cells in total. Each EM simulation costs about 1 minute on average. For the optimization of the slotted monopole antenna, the design variables shown in Fig. 4 and their search ranges in Table I are considered. The optimization goal is to minimize the fitness function, F_{monopole}, to satisfy the design specifications

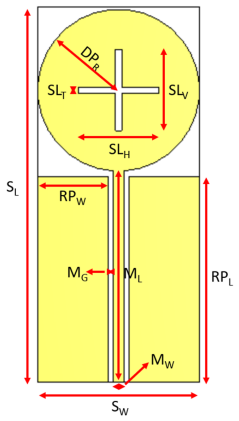


Figure 4. Layout of the compact UWB slotted monopole antenna.

Table I
SEARCH RANGES OF THE DESIGN VARIABLES AND THE OPTIMAL DESIGN BY SB-SADEA (ALL SIZES IN MM) (CASE STUDY 1)

Parameters	Lower bound	Upper bound	SB-SADEA Optimum
Circular patch radius (DP_R)	2	25	7.14
Substrate width (S_W)	$2\times DP_R$	$3\times DP_R$	14.40
Width of slot throat (SL_T)	0	$2\times DP_R$	8.21
Vertical slots' depth (SL_V)	0	$2\times DP_R$	0.75
Horizontal slots' depth (SL_H)	0	$2\times DP_R$	0.11
Microstrip width (M_L)	RP_L	50	26.21
Partial ground plane length (RP_L)	DP_R	M_L	8.21
Microstrip length (M_W)	0.50	7.50	1.20
Microstrip gap (M_G)	0	21.5	0.34
Feed guide width (P_W)	$6\times M_W$	$10\times M_W$	8.90
Substrate width (S_L) $= M_L + 2\times DP_R + 0.2$ (mm)			
Partial ground plane width (RP_W) $= (S_W - 2\times M_G - M_W) \div 2$ (mm)			

shown in Table II, mathematically,

$$F_{\text{monopole}} = w_1 \times \max(|S_{11}| + 10, 0) + w_2 \times \max(G_{\max} - 3, 0) + w_3 \times \max(1 - G_{\min}, 0) \quad (16)$$

where w_1, w_2 and w_3 are the penalty coefficients set to 1, 50 and 50, respectively. When all the design specifications in Table II are satisfied, F_{monopole} is equal to 0. 10 independent runs are carried out for SB-SADEA and all other reference methods except PSO. Three runs are carried out for PSO because more runs are not affordable.

In all the 10 runs, SB-SADEA satisfies the design specifications shown in Table II using an average of 924 EM simulations (15 hours). Fig. 5 shows the convergence trends. Fig. 6 shows the reflection coefficient and the realized gain of a typical optimal design mentioned in Table II. The size of the antenna shrinks about 60% compare with a state-of-the-art reference design [47].

Table II
DESIGN SPECIFICATIONS AND THE PERFORMANCE OF A TYPICAL OPTIMAL DESIGN OBTAINED BY SB-SADEA (CASE STUDY 1)

Item	Specification	SB-SADEA Optimum
Maximum Reflection Coefficient ($\lvert S_{11}\rvert$) (3.1 to 10.6 GHz)	$\leq$ -10 dB	-10.13 dB
Maximum Realized Gain ($G_{\max}$)	$\leq$ 3 dB	2.90 dB
Minimum Realized Gain ($G_{\min}$)	$\geq$ 1 dB	1.19 dB

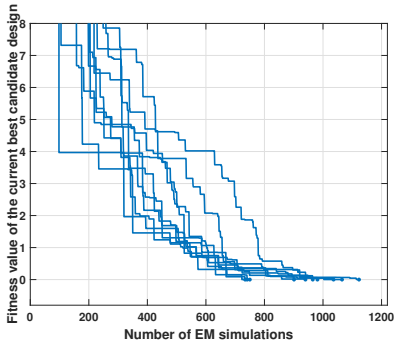
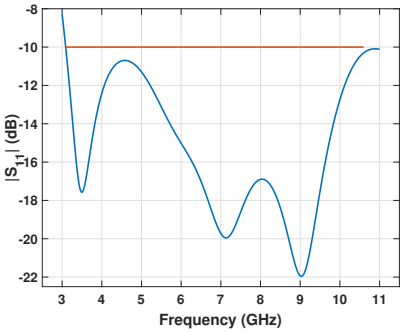
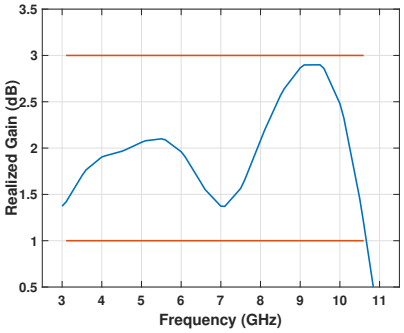


Figure 5. Convergence trends of SB-SADEA (Case Study 1, 10 runs).



(a) Reflection coefficient of the optimal design



(b) Realized gain of the optimal design

Figure 6. Response of the optimal design obtained by SB-SADEA (Case Study 1)

As discussed earlier, as one of the state-of-the-art methods for antennas with stringent specifications but without many design variables and specifications, P-SADEA is considered as the reference method. P-SADEA also shows a 100% success rate but uses an average of 1574 EM simulations to satisfy all the specifications. Therefore, SB-SADEA saves 40% of the EM simulations compared to P-SADEA. Note that compared to standard SADEA [8], P-SADEA improves the convergence speed at the cost of more GP modeling [10], [12] by its new model management framework. SB-SADEA, on the other hand, only uses the model management framework of standard SADEA [8], and the comparison result shows the effectiveness of BNN-based modeling and self-adaptive LCB techniques.

Table III
NUMBER OF EM SIMULATIONS (AVERAGE NUMBER) USED TO SATISFY THE SPECIFICATIONS FOR DIFFERENT METHODS (CASE STUDY 1)

	SB-SADEA	GP-ALCB	FBN-LCB	BN-ALCB	GP-LCB
ML models	BNN	GP	BNN	BNN	GP
Fine-supervision	Yes	No	Yes	No	No
Prescreening	AdapLCB	AdapLCB	LCB	AdapLCB	LCB
number of EM simulations	**924**	1262	1329	1104	1991

Moreover, they are compatible with the model management framework of P-SADEA, forming an even faster method.

To verify the effectiveness of the BNN-based antenna modeling, including the fine supervision, and the self-adaptive LCB-based prescreening, more comparisons are shown in Table III. When the GP model is used, the ω value for LCB is set to 2 as other SADEA versions, instead of 14 for the BNN-based model.

The following conclusions can be drawn from Table III: (1) By comparing SB-SADEA with GP-ALCB, when both make use of the self-adaptive LCB-based prescreening, nearly 25% fewer EM simulations are saved by the BNN-based surrogate modeling compared to GP. (2) By comparing SB-SADEA with FBN-LCB, when both make use of the BNN-based surrogate modeling, nearly 30% fewer EM simulations are saved. This indicates the effectiveness of the self-adaptive LCB prescreening and its co-working with the BNN-based model. For the BNN-based model, the prediction uncertainty is smaller than that of GP (Section III), and a larger ω has to be used in LCB prescreening to guarantee the exploration ability, which inevitably slows down the convergence speed. Hence, the self-adaptive LCB technique is essential for the BNN-based model. (3) By comparing SB-SADEA with BN-ALCB, where the only difference is the use of fine supervision, about 15% fewer EM simulations are saved, showing the effect of fine supervision. (4) GP-LCB (i.e., standard SADEA) is the slowest and SB-SADEA decreases 53% of the necessary EM simulations to obtain the optimal design, showing the combined effect of the BNN-based antenna surrogate model and the self-adaptive LCB method.

In the three PSO runs, the specifications on realized gain are satisfied, but the specification on $max(|S_{11}|)$ is not, and the average value is -5.2 dB. This can be attributed to the compactness of the structure and the stringency of the design specifications. Considering all these comparisons, this case study verifies the advantages of SB-SADEA in terms of convergence speed.

B. Case Study 2: A 4-band mm-wave Antenna

This case is designed to exhibit a quad-band operation with significant band discrimination and high gain at mm-wave frequencies of 28 GHz, 38 GHz, 50 GHz, and 60 GHz. It aims to achieve a minimum realized gain of 4.5 dB and a total efficiency better than 80% for all four operating bands. This low-profile antenna employs a patch geometry combining a square patch with an L- and an F-shaped slot on a Rogers RT/Duroid 5880 substrate of 0.254 mm thickness, relative

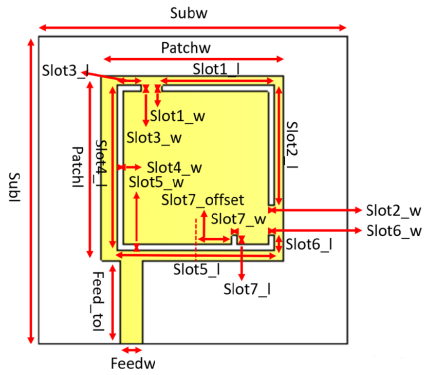


Figure 7. The layout of the 4-band 5G mm-wave antenna.

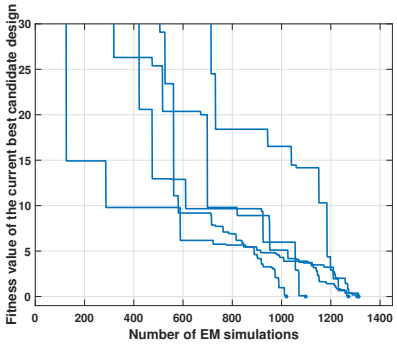


Figure 8. Convergence trends of SB-SADEA (Case Study 2, 5 runs).

permittivity of 2.2, and loss tangent of 0.0009. This single layer 5.1 mm $\times$ 5 mm $\times$ 0.254 mm antenna is excited by a 50Ω off-centered single-feed microstrip line. The slots positioned close to the edges of the patch make the current mostly concentric there and generate inductive and capacitive effects resulting in the multi-frequency operation.

The 4-band mm-wave antenna is modeled and discretized in CST-MWS with nearly 300,000 mesh cells in total. Each EM simulation costs about 2.5 minutes on average. For the optimization of the targeted antenna, the design variables shown in Fig. 7 and their search ranges in Table IV are considered. The optimization goal is to minimize the fitness function, F_{mmwave}, to satisfy the design specifications shown in Table V, mathematically,

$$\begin{aligned} F_{\text{mmwave}} = &\sum_{i=1}^{4} w_1 \times \max(|S_{11}^i| + 10, 0) \\ &+ \sum_{i=1}^{4} w_2 \times \max(4.5 - G_{\text{min}}^i, 0) \\ &+ \sum_{i=1}^{4} w_3 \times \max(0.8 - E_{\text{total}}^i, 0) \end{aligned} \tag{17}$$

where i is the index for the current frequency band out of the 4 frequency bands. w_1, w_2 and w_3 are the penalty coefficients set to 1, 50 and 50, respectively. When all the design specifications in Table V are satisfied, F_{mmwave} is equal to 0.

Five independent runs are carried out to test SB-SADEA. All of them satisfy the design specifications shown in Table

Table IV
SEARCH RANGES OF THE DESIGN VARIABLES AND A TYPICAL OPTIMAL DESIGN OBTAINED BY SB-SADEA (ALL SIZES IN MM) (CASE STUDY 2)

Variable	Lower bound	Upper bound	SB-SADEA Optimum
slot1_w	0	3	0.059
slot2_w	0	3	0.72
slot3_w	0	3	0.033
slot4_w	0	3	2.23
slot5_w	-3	0.2	-0.71
slot6_w	-3	0.2	-0.081
slot7_w	-2.2	0.9	-1.34
slot7_offset	0	2.5	2.24
feedw	0.1	0.45	0.18
feed_offset	0	3-feedw	0.017
slot1_l	0	3	0.18
slot2_l	0	3	1.98
slot3_l	0	3	1.95
slot4_l	0	3	0.60
slot5_l	0	3	1.40
slot6_l	0	3	2.15
slot7_l	0	3	0.87
feed_tol	0	3	2.02
patchl	0.5	4.3	4.26
patchw	0.5	5	4.55

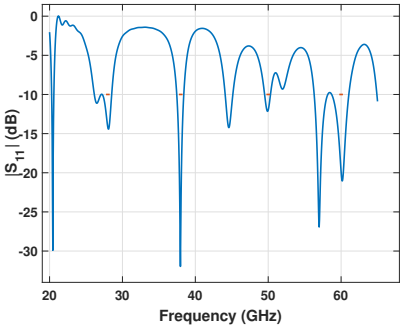
Table V
DESIGN SPECIFICATIONS AND THE PERFORMANCE OF A TYPICAL OPTIMAL DESIGN OBTAINED BY SB-SADEA (CASE STUDY 2)

Items	Specification	SB-SADEA Optimum
Maximum in-band reflection coefficients ($\|S11\|$) (27.75 to 28.25 GHz)	$\leq$ -10 dB	-12.28 dB
Maximum in-band reflection coefficients ($\|S11\|$) (37.75 to 38.25 GHz)	$\leq$ -10 dB	-13.04 dB
Maximum in-band reflection coefficients ($\|S11\|$) (49.75 to 50.25 GHz)	$\leq$ -10 dB	-10.54 dB
Maximum in-band reflection coefficients ($\|S11\|$) (59.75 to 60.25 GHz)	$\leq$ -10 dB	-16.18 dB
Minimum in-band realized gain ($G_{\min}$) (27.75 to 28.25 GHz)	$\geq$ 4.5 dB	5.67 dB
Minimum in-band realized gain ($G_{\min}$) (37.75 to 38.25 GHz)	$\geq$ 4.5 dB	4.88 dB
Minimum in-band realized gain ($G_{\min}$) (49.75 to 50.25 GHz)	$\geq$ 4.5 dB	6.75 dB
Minimum in-band realized gain ($G_{\min}$) (59.75 to 60.25 GHz)	$\geq$ 4.5 dB	7.01 dB
Minimum in-band total efficiency (E_{tot}) (27.75 to 28.25 GHz)	$\geq$ 80%	82.4%
Minimum in-band total efficiency (E_{tot}) (37.75 to 38.25 GHz)	$\geq$ 80%	86.2%
Minimum in-band total efficiency (E_{tot}) (49.75 to 50.25 GHz)	$\geq$ 80%	84.4%
Minimum in-band total efficiency (E_{tot}) (59.75 to 60.25 GHz)	$\geq$ 80%	89.3%

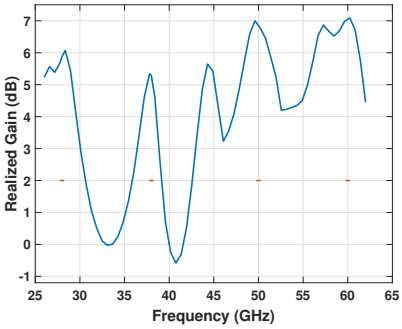
V using an average of 1202 EM simulations. Fig. 8 shows the convergence trends. Fig. 9 shows the reflection coefficient, realized gain and the total efficiency of a typical optimal design in Table IV.

As discussed earlier, TR-SADEA [11] is selected as the reference method. In all the 5 runs, it also has a 100% success rate but uses an average of 2426 EM simulations. Hence, SB-SADEA decreases the number of EM simulations by more than 50% compared to TR-SADEA in this case study, verifying the advantages in convergence speed again.

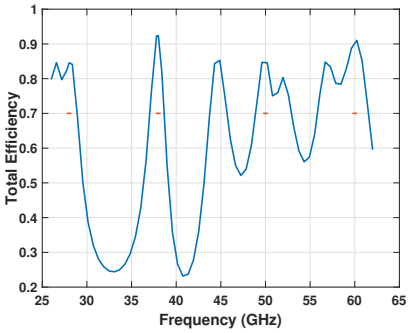
The other aim of this case study is to compare the machine learning cost. TR-SADEA is proposed for antennas with many design variables and specifications, where GP modeling time becomes a challenge. By its GP model sharing method, TR-SADEA often reduces the GP modeling time by 90% [11]. Still, for the targeted antenna, an average of over 426,000 GP surrogate models are built in the optimization using TR-



(a) Reflection coefficient of the optimal design



(b) Realized gain of the optimal design



(c) Total efficiency of the optimal design

Figure 9. Response of the optimal design obtained by SB-SADEA (Case Study 2)

Table VI
COMPARISON BETWEEN SB-SADEA AND TR-SADEA (CASE STUDY 2, AVERAGE VALUES)

	SB-SADEA	TR-SADEA
ML models	BNN	GP (model sharing)
Fine-supervision	Yes	Yes
Prescreening	AdapLCB	LCB
Number of surrogate models	**102,000**	426,000
Modeling time (hours)	**2.8**	15.5
Number of EM simulations	**1202**	2426
Total optimization time (hours)	**55.8**	128

SADEA, taking 15 hours on average. This time consumption is practical but not desirable. With BNN-based surrogate modeling in SB-SADEA, only 2.8 hours of surrogate model training time on average is used. Table VI demonstrates the number of EM simulations used, the number of surrogate models trained, and the total time used for the two methods. The average values of the 5 independent runs are used. The significant improvement in terms of the machine learning cost of SB-SADEA is also shown. The total optimization time decreased by more than a half compared to the reference method.

DE is carried out for the 5G mm-wave antenna. After two weeks' optimization, none of the reflection coefficient specifications is satisfied and only half of the gain and total efficiency specifications are met. Longer run may improve the performance, but the optimization time is too long for practical use. Considering all these comparisons, this case study verifies the advantages of SB-SADEA in terms of both convergence speed and machine learning cost.

V. Conclusions

In this paper, the SB-SADEA method has been proposed, which is a universal global optimization method for antennas with various numbers of design variables and specifications. Its effectiveness and efficiency are demonstrated by two real-world challenging antenna design cases. Thanks to the BNN-based surrogate model, which is introduced into antenna global optimization area by this research, and the new self-adaptive LCB method, which is essential for using the BNN-based surrogate model, significant advantages in terms of both convergence speed (i.e., the number of EM simulations needed to obtain the optimal design) and the machine learning cost are obtained. Future works will include behavioral analysis of SB-SADEA and its improvement.

References

[1] Y. Sato, F. Campelo, and H. Igarashi, "Meander line antenna design using an adaptive genetic algorithm," *IEEE Transactions on Magnetics*, vol. 49, no. 5, pp. 1889–1892, 2013.

[2] Z. D. Zaharis and T. V. Yioultsis, "A novel adaptive beamforming technique applied on linear antenna arrays using adaptive mutated boolean pso," *Progress In Electromagnetics Research*, vol. 117, pp. 165–179, 2011.

[3] R. Storn and K. Price, "Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces," *Journal of Global Optimization*, vol. 11, no. 4, pp. 341–359, 1997.

[4] J. Kennedy, "Particle swarm optimization," *Encyclopedia of machine learning*, pp. 760–766, 2010.

[5] P. Rocca, G. Oliveri, and A. Massa, "Differential evolution as applied to electromagnetics," *IEEE Antennas and Propagation Magazine*, vol. 53, no. 1, pp. 38–49, 2011.

[6] Z. Zhang, H. C. Chen, and Q. S. Cheng, "Surrogate-assisted quasi-newton enhanced global optimization of antennas based on a heuristic hypersphere sampling," *IEEE Transactions on Antennas and Propagation*, vol. 69, no. 5, pp. 2993–2998, 2020.

[7] S. Koziel and S. Ogurtsov, *Antenna design by simulation-driven optimization*. Springer, 2014.

[8] B. Liu *et al.*, "An efficient method for antenna design optimization based on evolutionary computation and machine learning techniques," *IEEE Trans. on Antennas and Propagation*, vol. 62, no. 1, pp. 7–18, 2014.

[9] B. Liu, S. Koziel, and N. Ali, "Sadea-ii: A generalized method for efficient global optimization of antenna design," *Journal of Computational Design and Engineering*, vol. 4, no. 2, pp. 86–97, 2017.

[10] M. O. Akinsolu, B. Liu, V. Grout, P. I. Lazaridis, M. E. Mognaschi, and P. Di Barba, "A parallel surrogate model assisted evolutionary algorithm for electromagnetic design optimization," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 3, no. 2, pp. 93–105, 2019.

[11] B. Liu, M. O. Akinsolu, C. Song, Q. Hua, P. Excell, Q. Xu, Y. Huang, and M. A. Imran, "An efficient method for complex antenna design based on a self adaptive surrogate model-assisted optimization technique," *IEEE Transactions on Antennas and Propagation*, vol. 69, no. 4, pp. 2302–2315, 2021.

[12] B. Liu, M. O. Akinsolu, N. Ali, and R. Abd-Alhameed, "Efficient global optimisation of microwave antennas based on a parallel surrogate model-assisted evolutionary algorithm," *IET Microwaves, Antennas & Propagation*, vol. 13, no. 2, pp. 149–155, 2018.

[13] M. Alibakhshikenari, B. S. Virdee, C. H. See, P. Shukla, S. M. Moghaddam, A. U. Zaman, S. Shafqaat, M. O. Akinsolu, B. Liu, J. Yang *et al.*, "Dual-polarized highly folded bowtie antenna with slotted self-grounded structure for sub-6 ghz 5g applications," *IEEE Transactions on Antennas and Propagation*, 2021.

[14] Q. Wu, H. Wang, and W. Hong, "Multistage collaborative machine learning and its application to antenna modeling and optimization," *IEEE Transactions on Antennas and Propagation*, vol. 68, no. 5, pp. 3397–3409, 2020.

[15] 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 Transactions on Microwave Theory and Techniques*, vol. 69, no. 12, pp. 5189–5202, 2021.

[16] S. Koziel, A. Bekasiewicz, I. Couckuyt, and T. Dhaene, "Efficient multi-objective simulation-driven antenna design using co-kriging," *IEEE Transactions on Antennas and Propagation*, vol. 62, no. 11, pp. 5900–5905, 2014.

[17] J. Zhou, Z. Yang, Y. Si, L. Kang, H. Li, M. Wang, and Z. Zhang, "A trust-region parallel bayesian optimization method for simulation-driven antenna design," *IEEE Transactions on Antennas and Propagation*, 2020.

[18] C. Rasmussen, "Gaussian processes in machine learning," *Advanced Lectures on Machine Learning*, pp. 63–71, 2004.

[19] C.-Y. Lee, S. Xie, P. Gallagher, Z. Zhang, and Z. Tu, "Deeply-supervised nets," in *Artificial intelligence and statistics*. PMLR, 2015, pp. 562–570.

[20] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "Unet++: A nested u-net architecture for medical image segmentation," in *Deep learning in medical image analysis and multimodal learning for clinical decision support*. Springer, 2018, pp. 3–11.

[21] D. Erricolo, P.-Y. Chen, A. Rozhkova, E. Torabi, H. Bagci, A. Shamim, and X. Zhang, "Machine learning in electromagnetics: A review and some perspectives for future research," in *2019 International Conference on Electromagnetics in Advanced Applications (ICEAA)*. IEEE, 2019, pp. 1377–1380.

[22] B. Liu, Q. Zhang, and G. G. E. Gielen, "A gaussian process surrogate model assisted evolutionary algorith for medium scale expensive optimization problems," *IEEE Trans. on Evolutionary Computation*, vol. 18, no. 2, pp. 180–192, 2014.

[23] M. Emmerich, K. Giannakoglou, and B. Naujoks, "Single-and multi-objective evolutionary optimization assisted by Gaussian random field metamodels," *IEEE Transactions on Evolutionary Computation*, vol. 10, no. 4, pp. 421–439, 2006.

[24] Y. Jin, "Surrogate-assisted evolutionary computation: Recent advances and future challenges," *Swarm and Evolutionary Computation*, vol. 1, no. 2, pp. 61–70, 2011.

[25] D. R. Jones *et al.*, "Efficient global optimization of expensive black-box functions," *Journal of Global Optimization*, vol. 13, no. 4, pp. 455–492, 1998.

[26] G. B. Wright, *Radial basis function interpolation: numerical and analytical developments*. University of Colorado at Boulder, 2003.

[27] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *The journal of machine learning research*, vol. 15, no. 1, pp. 1929–1958, 2014.

[28] Y. Gal and Z. Ghahramani, "Dropout as a bayesian approximation: Representing model uncertainty in deep learning," in *international conference on machine learning*. PMLR, 2016, pp. 1050–1059.

[29] ——, "A theoretically grounded application of dropout in recurrent neural networks," *Advances in neural information processing systems*, vol. 29, pp. 1019–1027, 2016.

[30] H. Ulmer, F. Streichert, and A. Zell, "Evolution strategies assisted by gaussian processes with improved preselection criterion," in *The 2003*

Congress on Evolutionary Computation, 2003. CEC'03., vol. 1. IEEE, 2003, pp. 692–699.

[31] J. Dennis and V. Torczon, "Managing approximation models in optimization," *Multidisciplinary design optimization: State-of-the-art*, pp. 330–347, 1997.

[32] A. Bekasiewicz, S. Koziel, and Q. S. Cheng, "Reduced-cost constrained miniaturization of wideband antennas using improved trust-region gradient search with repair step," *IEEE Antennas and Wireless Propagation Letters*, vol. 17, no. 4, pp. 559–562, 2018.

[33] A. Pietrenko-Dabrowska and S. Koziel, "Expedited gradient-based design closure of antennas using variable-resolution simulations and sparse sensitivity updates," *IEEE Transactions on Antennas and Propagation*, 2022.

[34] Y. Youn, J. Choi, D. Kim, A. A. Omar, J. Choi, I. Yoon, S.-T. Ko, J. Lee, Y. Lee, and W. Hong, "Pso-aided ila methodology for hemispherical beam coverage and scan loss mitigation," in *2020 International Symposium on Antennas and Propagation (ISAP)*. IEEE, 2021, pp. 153–154.

[35] K. Dutta, D. Guha, and C. Kumar, "Theory of controlled aperture field for advanced superstrate design of a resonance cavity antenna with improved radiations properties," *IEEE Transactions on Antennas and Propagation*, vol. 65, no. 3, pp. 1399–1403, 2017.

[36] M. Stein, "Large sample properties of simulations using latin hypercube sampling," *Technometrics*, pp. 143–151, 1987.

[37] D. M. Blei, A. Kucukelbir, and J. D. McAuliffe, "Variational inference: A review for statisticians," *Journal of the American statistical Association*, vol. 112, no. 518, pp. 859–877, 2017.

[38] Y. Gal and Z. Ghahramani, "Bayesian convolutional neural networks with bernoulli approximate variational inference," *arXiv preprint arXiv:1506.02158*, 2015.

[39] Z.-H. Zhou, *Ensemble methods: foundations and algorithms*. Chapman and Hall/CRC, 2019.

[40] L. Breiman, "Bagging predictors," *Machine learning*, vol. 24, no. 2, pp. 123–140, 1996.

[41] S. Depeweg, J.-M. Hernandez-Lobato, F. Doshi-Velez, and S. Udluft, "Decomposition of uncertainty in bayesian deep learning for efficient and risk-sensitive learning," in *International Conference on Machine Learning*. PMLR, 2018, pp. 1184–1193.

[42] P. Westermann and R. Evins, "Using bayesian deep learning approaches for uncertainty-aware building energy surrogate models," *Energy and AI*, vol. 3, p. 100039, 2021.

[43] L. V. Jospin, W. Buntine, F. Boussaid, H. Laga, and M. Bennamoun, "Hands-on bayesian neural networks–a tutorial for deep learning users," *arXiv preprint arXiv:2007.06823*, 2020.

[44] I. M. Danjuma, M. O. Akinsolu, C. H. See, R. A. Abd-Alhameed, and B. Liu, "Design and optimization of a slotted monopole antenna for ultra-wide band body centric imaging applications," *IEEE Journal of Electromagnetics, RF and Microwaves in Medicine and Biology*, vol. 4, no. 2, pp. 140–147, 2020.

[45] M. Ur-Rehman, M. Adekanye, and H. T. Chattha, "Tri-band millimetre-wave antenna for body-centric networks," *Nano communication networks*, vol. 18, pp. 72–81, 2018.

[46] Z. U. Khan, T. H. Loh, A. Belenguer, and A. Alomainy, "Empty substrate-integrated waveguide-fed patch antenna array for 5g millimeter-wave communication systems," *IEEE Antennas and Wireless Propagation Letters*, vol. 19, no. 5, pp. 776–780, 2020.

[47] B. Yeboah-Akowuah, P. Kosmas, and Y. Chen, "A q-slot monopole for uwb body-centric wireless communications," *IEEE Transactions on Antennas and Propagation*, vol. 65, no. 10, pp. 5069–5075, 2017.