Engineering Science and Technology, an International Journal 79 (2026) 102399

Contents lists available at ScienceDirect

Engineering Science and Technology, an International Journal

journal homepage: www.elsevier.com/locate/jestch

Residual deep neural network and response features for inverse modeling of microwave passives

Slawomir Koziel [a,b], Kaustab C. Sahu [a], Anna Pietrenko-Dabrowska [b,*]

[a] Engineering Optimization & Modeling Center, Reykjavik University, 102 Reykjavik, Iceland
[b] Faculty of Electronics, Telecommunications and Informatics, Gdansk University of Technology, 80-233 Gdansk, Poland

ARTICLE INFO

Keywords:
Microwave engineering
Inverse modeling
Deep learning
Residual neural networks
Feature extraction
Design-ready modeling

ABSTRACT

Electromagnetic (EM) solvers are widely used in microwave design for their accuracy, but are computationally expensive. Surrogate models offer a remedy, yet face challenges such as the curse of dimensionality. This work introduces a novel inverse modeling approach using deep learning and carefully selected response features to predict circuit geometry from desired operating parameters such as center frequency and power split ratio. Data samples for model training are collected through a performance-driven sequential process. Setting the model at the feature level implicitly reduces dimensionality, as the feature set is smaller than the geometry parameter set. This allows reliable regressors to be built with far fewer training samples, greatly lowering computational cost. Additional savings are achieved through explicit dimensionality reduction, carried out using global sensitivity analysis. To set up an inverse surrogate, we employ a customized residual deep residual neural network (ResNet), which is particularly well-suited for capturing dependencies between features and geometry parameters. This is due to ResNet's resilience to a certain degree of parametric redundancy in the circuit structure. The inverse model produces parameter vectors sufficiently close to the optimal ones, thus only minor refinement is necessary. Also, our model is shown to be more accurate than state-of-the-art forward and inverse metamodels. The proposed ResNet embedded in the inverse modeling regime and combined with dimensionality-reduced sequential sampling are the main artificial intelligence-related contributions of this study. Engineering applications encapsulate the successful employment of our framework for rapid global circuit optimization, as exemplified using three microwave passives.

1. Introduction

The design of modern microwave devices would be inconceivable without the use of simulation tools. It is particularly the initial phase of device development that heavily utilizes full-wave electromagnetic (EM) simulations [1–4]. The primary advantage of EM analysis is its ability to capture a wide range of phenomena affecting device performance, among others, cross-coupling [5], losses of various kind e.g., dielectric or radiation) [6], anisotropy [7], but also the influence of external appliances (e.g., connectors). These effects must be precisely quantified to ensure trustworthy analysis of circuit behavior. This is particularly important when designing more and more complex contemporary components (miniaturized structures, or those involving metamaterials or substrate-integrated waveguides [8–13]). The growing geometrical complexity of microwave components is, in turn, a byproduct of the increasing requirements of modern application areas (mobile communications [14], automotive radars [15], microwave imaging [16], energy harvesting [17], etc.). For these reasons, the use of EM analysis becomes indispensable. Still, its unparalleled accuracy comes at a considerable cost, often rendering simulation-intensive procedures uneconomical due to excessive CPU demands. This holds true for virtually all design procedures, including numerical optimization [18–20], and also robust [21–23] or multi-criterial design [24–26]. The excessive cost of simulation-driven design persists regardless of the specific type of numerical optimization algorithm employed. For example, nature-inspired routines [27,28]—although very popular nowadays—are typically cost-prohibitive. While the computational burden associated with gradient-based algorithms [29,30] is generally less severe, though still significant.

These obstacles led the researchers to resort to alternative and more cost-efficient approaches. The first group of the developed methodologies focuses on expediting direct simulation-driven design procedures

* Corresponding author.
E-mail addresses: koziel@ru.is (S. Koziel), kaustab23@ru.is (K.C. Sahu), anna.dabrowska@pg.edu.pl (A. Pietrenko-Dabrowska).

https://doi.org/10.1016/j.jestch.2026.102399
Received 3 October 2025; Received in revised form 9 April 2026; Accepted 29 April 2026
Available online 14 May 2026

using various mechanisms, e.g., by lowering the cost of sensitivity evaluation for gradient-based algorithms (sparse sensitivity updates [31,32], adjoint sensitivities [33], or mesh deformation [34]). The acceleration of gradient-based optimization can also be achieved through feature-based techniques (leveraging regular relationships between the design variables and feature points situated on the device response) [35,36]. Another option are variable-resolution methods such as space mapping [37], or dimensionality reduction approaches, e.g., model-order reduction [38]. The second important group of techniques aimed at cost-efficient microwave design relies on surrogate modeling [39–42]. Two types of models can be distinguished: physics-based [43–46] and behavioral [47–50]. The former is primarily applied in local search, whereas the latter is typically used for global optimization. The most common use of surrogate models is in machine learning (ML) procedures [51–54], where the model is typically refined iteratively throughout the course of the algorithm. In surrogate modeling of high-frequency components, the aforementioned response feature technique is also applied as it enables a remarkable reduction in the cost of acquiring training data [55,56].

The appeal of behavioral (data-driven) models in microwave engineering lies in their widespread accessibility and straightforward deployment [53,57–61]. Lately, artificial neural networks (ANNs) have experienced a rising interest within this group of surrogates and are frequently the modeling method of choice in ML-based frameworks. A range of ANN architectures with varying levels of complexity is employed. Simpler models include multi-layer perceptrons (MLPs) [62–68], deep neural networks (DNNs) [69–71], and convolutional neural networks (CNNs) [72–74]. Among the more advanced architectures in use are long short-term memory (LSTM) networks [75], evolutionary networks [76], autoencoders [77], and variable-fidelity NNs [78]. Nevertheless, the application of neural networks for standalone or broad-purpose modeling remains limited, although several recent attempts have been made [79–84]. Unfortunately, in the aforementioned works, ANNs serve merely as simple regressors, incapable of representing non-trivial relationships between design parameters and system responses. Thus, they exhibit limitations akin to those of other data-driven modeling techniques.

This study proposes a novel technique for the inverse modeling of microwave devices that exploits regression models and response features. In the developed framework, the model predicts geometry parameters of the component based on the problem-specific knowledge in the form of key points identified in its responses. These points (features) are selected to enable the evaluation of important performance indicators (e.g., center frequency, power split). The data acquisition procedure operates sequentially, with sample acceptance based on its quality. This allows for assessing the extent of the model domain. As the number of feature points is smaller than that of design variables, implicit dimensionality reduction occurs, lowering the amount of data needed for an accurate model setup. In addition, explicit dimensionality reduction is performed through global sensitivity analysis, further enhancing the cost efficiency of the entire process. For model construction, a customized residual neural network (ResNet) is employed, capable of effectively learning the interrelations between features and geometry parameters. The key advantage of ResNet lies in its enhanced generalization capability compared to other neural network-based modeling approaches, arising from the unique characteristics of the relationships between the feature space and circuit's geometrical parameters, and also the inherent parametric redundancy within the design space. The ultimate purpose of the developed inverse model is practical design applications. The surrogate enables the identification of globally optimal designs that meet the intended operating conditions over extended parameter ranges. Specifically, it facilitates the generation of near-optimal designs based on target performance specifications, which are subsequently refined at low cost using an accelerated gradient-based algorithm. The developed modeling framework has been demonstrated using three microwave devices and has been shown to outperform several kernel-based and neural network modeling approaches with regard to reliability. The proposed model was also demonstrated to be suitable for design applications, specifically, global circuit optimization under diverse performance requirements.

This study delivers several technical contributions, which include (i) development of a novel inverse modeling methodology operating at the level of response features to be used in microwave design, (ii) deployment of sequential procedure for training data generation, with embedded feature extraction routine, to maintain model validity across diverse operating scenarios, (iii) the development of a customized deep residual neural network for accurate representation of the feature space and reliable prediction of the circuit geometry parameters corresponding to the target values of the operating conditions, (iv) demonstrating the model's superiority over extensive set of state-of-the-art forward and inverse data-driven models, (v) demonstrating the model's usefulness for design purposes through extensive application case studies (global circuit optimization).

2. Regression-Based inverse modeling with response features

This section details the proposed modeling methodology. Section 2.1 introduces the core concept and key prerequisites of the approach. Section 2.2 describes the response features and the procedure used for their extraction. The dimensionality reduction technique employed to improve modeling efficiency is presented in Section 2.3. Section 2.4 depicts the introduced deep residual neural network (ResNet), developed to implement the inverse surrogate. A step-by-step summary of the entire modeling workflow is given in Section 2.5, followed by Section 2.6, which discusses the application of the developed model to circuit design tasks (global optimization).

2.1. Inverse versus forward modeling: prerequisites and benefits

Conventional surrogate models follow a forward-oriented framework, where circuit frequency responses (mainly *S*-parameters) are expressed as functions of design parameters, typically device dimensions. However, as highlighted in Section 1, constructing such models is difficult due to the high dimensionality of the design space, nonlinear system behavior, and wide parameter limits. Additionally, forward models do not directly support design tasks and require a separate optimization step. In this research, we adopt an inverse modeling approach, wherein the model inputs are meticulously selected characteristic (feature) points of the circuit responses, while the outputs are the design parameters. The features are chosen to represent essential performance indicators, such as center frequency, bandwidth, power split ratio, or even material properties (e.g., substrate permittivity), enabling the model to produce design vectors tailored to specific operating targets. To effectively capture the intricate mapping between response features and design variables, we utilize a customized deep neural network architecture (specifically, a residual neural network), described in detail in Section 2.3.

Some essential advantages of inverse surrogates compared to traditional forward models should be highlighted:

- Inverse models typically offer higher accuracy, as the mapping between response feature points and design variables is generally less nonlinear than that between geometric parameters and full frequency responses.
- Inverse models require considerably fewer training samples to achieve robust performance, reducing the overall modeling effort.
- Because the number of feature points is often smaller than the number of design variables, inverse modeling inherently introduces dimensionality reduction.
- Training is confined to regions of the design space that yield acceptable or high-performance solutions (i.e., where valid feature

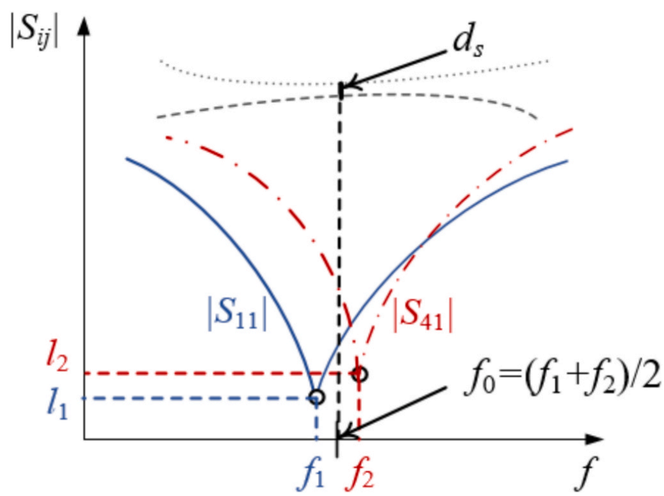

Fig. 1. Example of feature definitions for S-parameters of a microwave coupler: $|S_{11}|$ (—), $|S_{21}|$ (⋯·), $|S_{31}|$ |(– – –), $|S_{41}|$ (–.). The features include: two frequencies f_1 (blue) and f_2 (red) corresponding to the minima of $|S_{11}|$ (blue) and $|S_{41}|$ (red) characteristics, and the corresponding levels l_1 (blue) and l_2 (red). The operating parameters of the coupler can be estimated based on the features as: coupler's center frequency $f_0 = (f_1 + f_2)/2$ (black), and its power split $d_s = |S_{21}(f_0)| - |S_{31}(f_0)|$ (black).

points exist), improving model reliability and avoiding wasteful sampling in suboptimal areas.

- Inverse models enable direct prediction of design parameters for given target specifications, eliminating the need for a separate optimization phase as required in traditional forward-model-based design workflows.

2.2. Response features and feature extraction. training data acquisition

The core of the introduced inverse modeling methodology lies in the use of response features, whose definition corresponds to the considered design specifications. The two following vectors are of interest: a vector of design parameters $\boldsymbol{x} = [x_1 \ldots x_n]^T$ (typically, comprising the device's dimensions), and a vector of response features $\boldsymbol{v} = [v_1 \ldots v_K]^T$. For

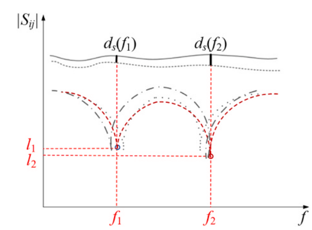

Fig. 3. Example of feature definitions for a dual-band power divider: $|S_{11}|$(– – –) (red), $|S_{21}|$ (—), $|S_{31}|$ (–·–), $|S_{22}|$ (–.), and $|S_{33}|$ (⋯·). For simplicity, the features are extracted solely from one of the component's characteristics, i.e., S_{11} (red). The feature set comprises four elements: frequencies f_1 and f_2 (red), corresponding to the minima of the S_{11} characteristics, and the associated levels l_1 and l_2 (red).

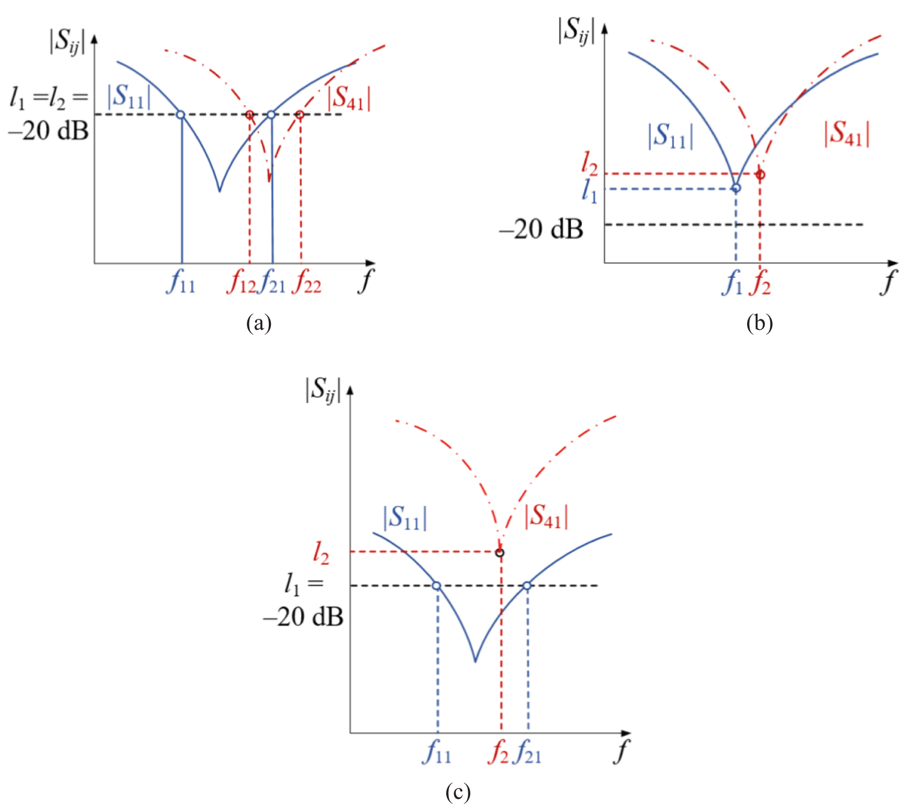

Fig. 2. Example of feature definitions for band enhancement of a microwave coupler (only $|S_{11}|$ and $|S_{41}|$ responses are shown for clarity): (a) Case I (both minima below –20 dB), the features are: the frequencies delimiting –20-dB band of $|S_{11}|$ (blue) response, i.e., f_{11} and f_{21} (blue), along with the frequencies delimiting –20-dB band of $|S_{41}|$ (red) response, i.e., f_{12} and f_{22} (red); (b) Case II (both minima above –20 dB), the features are: the frequencies f_1 (blue) and f_2 (red) indicating the minima of $|S_{11}|$ (blue) and $|S_{41}|$ (red) responses, respectively, and the corresponding levels l_1 (blue) and l_2 (red); (c) Case III (only one minimum above –20 dB, here, for $|S_{41}|$ (red) characteristics, similar definitions hold for the opposite case), the features are: the frequencies corresponding to –20-dB band of $|S_{11}|$ (blue) response, i.e., f_{11} and f_{21} (blue), and the level $l_1 = -20$ dB (black), and also frequency f_2 (red) corresponding to the minimum of $|S_{41}|$ (red) response and the respective level l_2 (red).

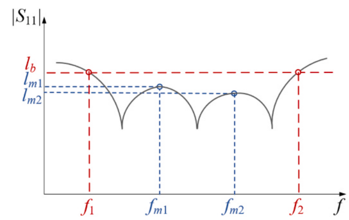

Fig. 4. Exemplary feature definitions for a microwave filter. The feature set comprises six elements: frequencies f_1 and f_2 (red), which define the filter's passband, the corresponding value l_b (red) of the S_{11} characteristics at these frequencies, and the levels l_{m1} and l_{m2} (blue) corresponding to the maxima of the component's response within the band, as well as the respective frequencies f_{m1} and f_{m2} (blue). For higher-order filter structures, the feature set can be expanded to include additional minima.

illustration, Figs. 1 through 4 present examples of how response features can be defined for various microwave components and various design problems. Fig. 1 shows S-parameters of a microwave coupler, for which the performance indicators include the operating frequency and power split. For this type of the device, the feature set contains two frequencies f_1 and f_2 matching the minima of $|S_{11}|$ and $|S_{41}|$ characteristics, and the respective levels l_1 and l_2. Using these, the operating parameters of the coupler can be evaluated as follows. The center frequency can be assessed as f_0 as the average value of the frequencies f_1 and f_2 (i.e., $f_0 = (f_1 + f_2)/2$). While power split is evaluatd at the estimated operating frequency f_0 as $d_s = |S_{21}(f_0)| - |S_{31}(f_0)|$. Consequently, the feature vector would be $\boldsymbol{v} = [f_1\ f_2\ l_1\ l_2\ d_s]^T$.

If, however, the enhancement of the coupler's bandwidth is of interest, a different choice of response features is required (see Fig. 2). Here, the feature vector would comprise the frequencies defining the –20 dB bandwidth: f_{11} and f_{21} (assessed based on $|S_{11}|$ characteristics), as well as the frequencies f_{21} and f_{22} (assessed based on S_{41} characteristics). The remaining feature would be the level l_1 = –20 dB. Consequently, the feature set would consist be $\boldsymbol{v} = [f_{11}\ f_{21}\ f_{12}\ f_{22}\ l_1\ l_2]^T$. The aforementioned feature arrangement pertains to Case I shown in Fig. 2 (a), in which the frequencies corresponding to –20 dB band can be discerned for both responses. Whereas Fig. 2(b) illustrates Case II, where the minima of both considered responses exceed –20 dB. In this case, we assign $f_{11} = f_{21} = f_1$ (the frequency at the minimum of $|S_{11}|$) and similarly $f_{12} = f_{22} = f_2$ (the frequency at the minimum of $|S_{41}|$). As for the levels, we set l_1 and l_2 to the values of the minima of $|S_{11}|$ and $|S_{41}|$ characteristics, respectively. This approach allows for maintaining a consistent number of feature points regardless of the design quality. Furthermore, Case III shown in Fig. 2(c), refers to a situation, in which only one of the minima exceeds the –20 dB level. This case is illustrated in Fig. 2(c) for the $|S_{11}|$ characteristics; a similar definition applies when the minimum of the S_{41} characteristics exceeds –20 dB instead. In this case, the features comprise the frequencies delimiting the –20-dB band of the S_{11} response, i.e., f_{11} and f_{21}, while $f_{12} = f_{22} = f_2$ (i.e., the frequency corresponding to the minimum of $|S_{41}|$ response). In addition, the level l_1 = –20 dB is included in the response set, along with the level l_2 at the minimum of the S_{41} characteristics. Consequently, the feature vector becomes $\boldsymbol{v} = [f_{11}\ f_{21}\ f_{12}\ f_{22}\ l_1\ l_2]^T$.

Another example is presented in Fig. 3, which shows the S-parameters of a dual-band power divider. To simplify feature extraction, despite existing multiple responses, the features are assessed solely from one of the component's characteristics, i.e., $|S_{11}|$. Consequently, the feature vector contains four entries: frequencies f_1 and f_2, corresponding to the minima of the $|S_{11}|$ characteristics, and the associated levels l_1 and l_2. In this case, the feature vector takes the form of $\boldsymbol{v} = [f_1\ f_2\ l_1\ l_2]^T$.

The last example is shown in Fig. 4, for a microwave filter (here, a third-order one). For this component, the feature vector is as follows $\boldsymbol{v} = [f_1\ f_2\ l_1\ l_2\ l_{m1}\ l_{m2}]^T$. The first two entries refer to the frequencies f_1 and f_2 delimiting the filter's passband. The next two entries are the values of the $|S_{11}|$ characteristics at these frequencies. Finally, l_{m1} and l_{m2} correspond to the maxima of the component's response within the band.

This example can be generalized to higher-order filter structures by including additional features corresponding to other minima. It should also be noted that, regardless of the component under consideration, the feature vector may also include additional parameters. An example of such a parametr is the substrate permittivity ε_r, if the model is intended to include the possibility of implementing the component of interest on different substrate materials.

Once the response features are defined, the next step is to acquire the training data. In this work, we adopt a randomized sequential sampling approach. A series of candidate design vectors $\boldsymbol{x}_{tmp}^{(j)}$ is generated, and for each candidate, the corresponding feature vector $\boldsymbol{v}(\boldsymbol{x}_{tmp}^{(j)})$ is extracted from the EM-evaluated response $\boldsymbol{R}(\boldsymbol{x}_{tmp}^{(j)})$. A sample is accepted into the training dataset only if it meets predefined threshold criteria for all feature components. Fig. 5 outlines the pseudocode of this sampling procedure, along with relevant commentary. Importantly, this process not only builds the training dataset but also implicitly defines the region of validity for the inverse model, based on the spatial

1. Input arguments:
   - Maximum number of random samples $N_{\max}$;
   - Acceptance thresholds $v_{k.\min}$ and $v_{k.\max}$, $k = 1, \ldots, K$;
   - Design variable space X (delimited by lower and upper parameter bounds);
2. Initialize training dataset $X_t = \varnothing$ (size $q = 0$);
3. Initialize test sample counter $j = 1$;
4. Generate sample $\boldsymbol{x}_{tmp}^{(j)} \in X$;
5. Perform EM simulation and obtain circuit responses $\boldsymbol{R}(\boldsymbol{x}_{tmp})$;
6. Extract response features $\boldsymbol{v}(\boldsymbol{x}_{tmp}^{(j)})$;
7. If $\boldsymbol{v}(\boldsymbol{x}_{tmp}^{(j)}) = [v_1(\boldsymbol{x}_{tmp}^{(j)}) \ldots v_K(\boldsymbol{x}_{tmp}^{(j)})]^T$ fulfill acceptance conditions, i.e., $v_{k.\min} \le v_k(\boldsymbol{x}_{tmp}^{(j)}) \le v_{k.\max}$, for $k = 1, \ldots, K$, set $q = q + 1$ and $\boldsymbol{x}_t^{(q)} = \boldsymbol{x}_{tmp}^{(j)}$; update $X_t = X_t \cup \{\boldsymbol{x}_t^{(q)}\}$;
8. If $j \le N_{\max}$, go to 4;
9. Return training set $X_t = \{\boldsymbol{x}_t^{(q)}\}_{q = 1, \ldots, Nt}$.

Remark: We use relaxed acceptance criteria (e.g., maximum $|S_{11}|$, $|S_{41}|$ around –10 dB) to incorporate a larger number of samples and construct a reliable representation of the dependencies between the feature points and design variables. Nonetheless, the number of accepted samples is typically $N_t = N_{\max}/3$ to $N_{\max}/2$.

Fig. 5. Training data generation for inverse modeling of microwave circuits.

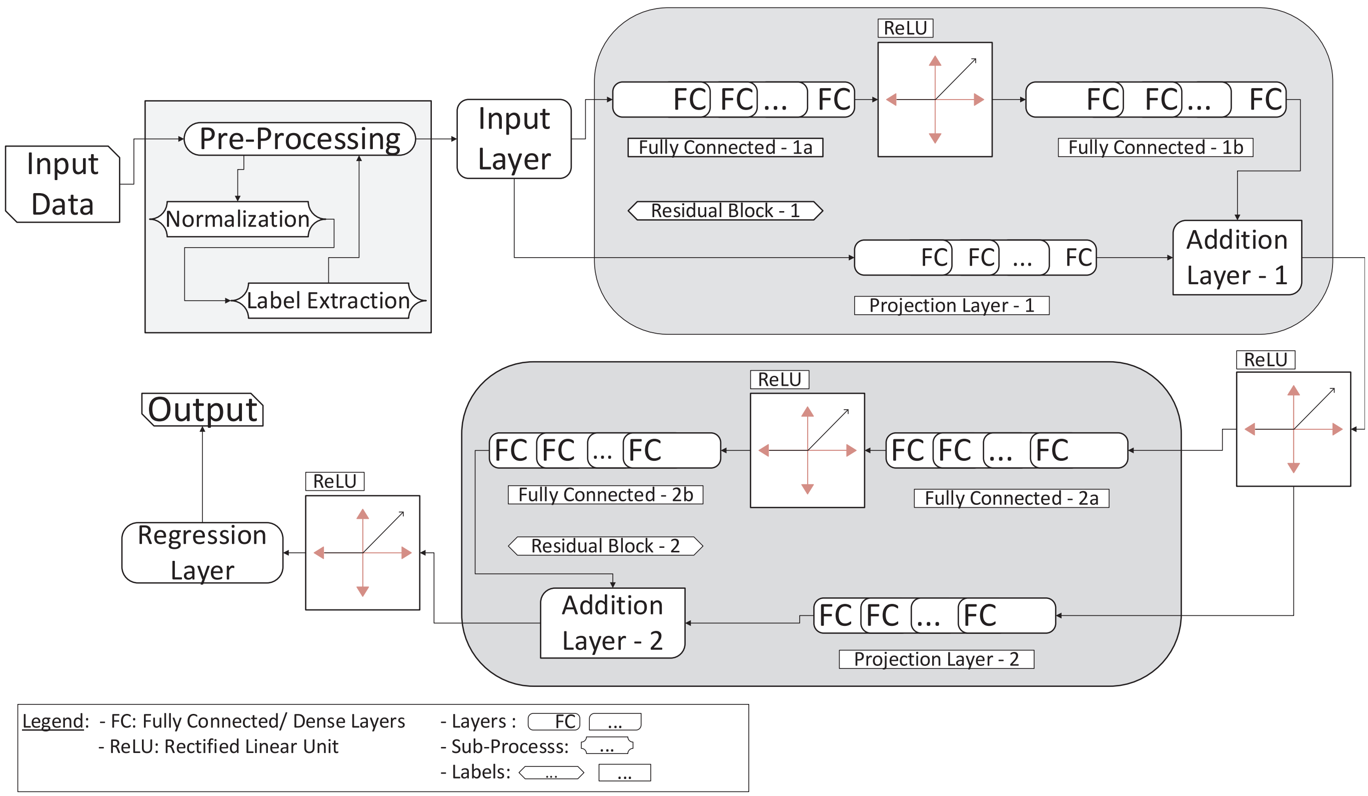

Fig. 6. The proposed ResNet architecture: block diagram. The network processes feature vectors derived from frequency responses through two residual blocks, each combining a main path with a projection shortcut. The final dense layer outputs the predicted geometric parameters. All hidden layers use 256 neurons, and layer types and dimensions are indicated.

distribution of accepted samples. Conversely, rejected candidates correspond to regions associated with poor-performance designs and are therefore excluded from the modeling process.

After completing the data acquisition process, we obtain a set of sample pairs $\{\boldsymbol{v}(\boldsymbol{x}_t^{(j)}), \boldsymbol{x}_t^{(j)}\}$, $j = 1, \ldots, N_t$, which serve as the basis for constructing the inverse model $s(\boldsymbol{v})$: $F \rightarrow X$, where F represents the feature space. During the model identification phase (see Section 2.4), the feature vectors $\boldsymbol{v}(\boldsymbol{x}_t^{(j)})$ are used as inputs to the surrogate model, and the corresponding design vectors $\boldsymbol{x}_t^{(j)}$ as outputs. The chosen modeling framework—a deep residual neural network—is described in detail in Section 2.4.

2.3. Dimensionality reduction using global sensitivity analysis

As previously noted, the curse of dimensionality poses a significant obstacle in building accurate data-driven models. To address this, the proposed modeling framework incorporates explicit dimensionality reduction. Rather than operating within the full design space X, training data is generated within a reduced affine subspace X_d, identified through fast global sensitivity analysis (FGSA) [85]. FGSA efficiently determines the most influential directions—those that have a notable impact on the circuit's electromagnetic (EM) responses. The process involves evaluating the differences in EM responses between randomly generated sample pairs, from which a relocation matrix $\boldsymbol{S}$ is constructed. Spectral decomposition of $\boldsymbol{S}$ yields eigenvectors $\boldsymbol{e}_j$ that represent principal directions of variation in the circuit output. The importance of each direction is quantified by its corresponding eigenvalue λ_j, $j = 1,\ldots,n$. The reduced domain X_d is then spanned by the $N_d < n$ most significant eigenvectors. The latter are chosen such that the "energy" captured by these directions satisfies the condition $|| [\lambda_1 \ldots \lambda_{Nd}]^T || \, / \, || [\lambda_1 \ldots \lambda_n]^T || \geq C_{\min}$ (here $C_{\min} = 0.9$, see [85] for details).

$$X_d = \left\{ \boldsymbol{x} \in X : \boldsymbol{x} = \boldsymbol{x}_c + \sum_{j=1}^{N_d} a_j \boldsymbol{e}_j \right\} \cap X \qquad (1)$$

where a_j are real numbers, and $\boldsymbol{x}_c$ is the geometric center of X. One of the practical advantages of FGSA is its low computational cost: in this study, it is carried out using only fifty random samples.

2.4. Residual deep neural network for inverse modeling

Predicting geometric design parameters from microwave frequency response features presents two principal challenges. First, the input space is redundant—distinct geometries may yield similar frequency responses (e.g., overlapping resonance bands). Second, the number of valid training samples is limited, particularly after discarding low-quality data based on frequency-domain acceptance criteria (cf. Section 2.3). These factors make the inverse mapping task highly sensitive to both underfitting and overfitting: shallow models often lack sufficient capacity, while deeper ones tend to overfit even with careful regularization.

Given this trade-off, our choice of model architecture was driven by the need for sufficient depth to capture non-linear structure, while preserving generalization. Preliminary experiments showed that standard deep neural networks (DNNs), convolutional neural networks (CNNs), and radial basis function (RBF) models overfit rapidly. While simpler alternatives such as support vector regressors (SVRs) and ridge regression consistently underfit, even with extensive tuning (see Section 3 for comparative results).

2.4.1. ResNets: motivation and background

To address the trade-off between model complexity and generalization in the inverse modeling task, we adopt a compact Residual Neural

Network (ResNet) architecture [86]. ResNets, originally introduced for image classification, incorporate skip connections that allow neural layers to learn residual functions—i.e., corrections over the identity—rather than full input–output mappings. This structure improves gradient flow and reduces convergence issues, particularly in deep networks [87].

Although ResNets were designed for convolutional image models, their core mechanism of residual learning extends naturally to regression tasks with redundant or correlated inputs [88]—such as ours. We adapt this architecture for 1D feature vectors derived from frequency responses, enabling stable training even with limited and noisy data. Unlike typical ResNet variants used in computer vision, our version omits convolutional layers entirely and employs a shallow structure with just two residual blocks. The design choice was made after repeated trials with increasingly deeper variants under constrained data conditions. The chosen configuration represents the smallest network that reliably generalizes across all design cases considered.

By using residual connections in a fully connected setting, the model avoids the optimization instability seen in standard multilayer perceptrons (MLPs), while maintaining enough capacity to capture non-linear dependencies in the inverse mapping. Further, the ability to reuse low-level features across blocks is particularly useful in our task, where key response traits (e.g., resonance depth, bandwidth) recur across parameter combinations.

It should be emphasized that the use of residual learning is particularly important in the presence of parametric redundancy, as discussed in Section 3.1. In such cases, the inverse mapping from response features to design parameters is inherently non-unique, i.e., multiple distinct parameter vectors may correspond to similar operating characteristics. This leads to an ill-posed regression problem, where conventional feedforward models tend to either average across feasible solutions or overfit to specific samples. The residual architecture alleviates these issues by learning incremental corrections over intermediate representations, thereby improving stability and generalization performance in the presence of input–output ambiguity. This property is particularly relevant for structures such as Circuit II.

This compact ResNet architecture thus offers a practical balance: it leverages the known strengths of residual learning while being tailored for 1D, regression-based, data-limited microwave circuit modeling. In this study, we extend ResNet architectures to a regression setting, focusing on predicting geometric design parameters of microwave components from frequency response features (as discussed in Section 2.2). Comparative results against alternative architectures are presented in Section 3.

2.4.2. Model architecture

Building on the requirements outlined previously, we design a customized Residual Network (ResNet) architecture tailored explicitly for the inverse modeling of microwave geometries from frequency-domain features. The final architecture, shown in Fig. 6, was the smallest configuration that consistently avoided overfitting across validation splits. It retains sufficient non-linear capacity through residual learning while remaining stable and tractable during training. Each component is designed to balance representational flexibility with robustness in a low-data regime.

The input layer receives the preprocessed 1D feature vector (see Section 2.2), which encodes resonance frequencies, bandwidths, reflection levels, and other spectral features. No further normalization is applied at this stage, as all inputs are pre-scaled during preprocessing.

Residual block 1:

- Main path: Two fully connected (dense) layers, each with 256 neurons and ReLU activation (FC1 → ReLU1 → FC2).
- A projection layer (PROJ1) aligns the input dimensions for residual addition.
- Output: The main and shortcut paths are summed (ADD1), followed by another ReLU (ReLU_ADD1).

This structure allows the network to learn residual mappings $F(\boldsymbol{x}) = H(\boldsymbol{x}) - \boldsymbol{x}$, enabling more stable training and mitigating vanishing gradients.

Residual block 2: Identical in structure to the first block: a pair of dense layers (FC3, FC4), projection path (PROJ2), addition (ADD2), and a final ReLU (ReLU_ADD2).

Output Layer:

- A final fully connected layer reduces the 256-dimensional latent representation to the output size — equal to the number of target geometric parameters.
- A regression output computes the mean squared error (MSE) loss between predictions and true parameter values.

The final model comprises approximately 270,000 trainable parameters, with each dense layer set to 256 neurons. Two residual blocks (four dense layers in total) were found sufficient to capture non-linear relationships without overfitting. ReLU was selected as the activation function due to consistently superior performance over alternatives like tanh and leaky ReLU. This configuration offers a practical trade-off between depth, capacity, and generalization in a data-constrained regression setting.

As mentioned earlier, while alternative architectures such as Multilayer Perceptrons (MLPs), Convolutional Neural Networks (CNNs), and Radial Basis Function (RBF) networks are commonly used in regression tasks, they failed to offer reliable generalization in our case.

MLPs, even with extensive tuning, were prone to either underfitting or overfitting, with unstable optimization in deeper configurations. CNNs, better suited for structured spatial data, performed poorly when applied to flat, unstructured input vectors. RBF networks and Support Vector Regressors (SVRs) struggled with scalability and generalization, often collapsing to trivial mappings due to input redundancy and lack of inductive structure.

Furthermore, we also considered Graph Neural Networks (GNNs) and Transformer-based models. GNNs were ruled out due to the lack of inherent topological structure in the input features, which would require significant preprocessing and task-specific graph construction. Transformers, while powerful, introduced excessive architectural complexity and computational overhead, making them especially prone to overfitting in our data-constrained setting. Residual architectures, in contrast, offer a strong balance between depth, expressiveness, and training stability, a desirable property when the mapping is non-linear but partially invertible.

2.4.3. Training setup and hyperparameters

The ResNet model is trained using the Adam optimizer [89], which adaptively adjusts learning rates using first and second moment estimates. This introduces between momentum and RMSProp [90] and resulting in efficient, stable convergence across trials. The model is initialized with a relatively low learning rate of 5×10^{-4}, where higher rates often led to erratic updates and unstable loss behavior. This base rate is decayed by a factor of 0.96 every 128 epochs using a piecewise schedule, which allows for finer updates as training progresses. This gradual decay proved critical in ensuring convergence to smooth, generalizable solutions.

The model is trained for up to 1024 epochs, with early stopping typically triggered before then. Early stopping is based on validation error with a patience of 128 epochs, i.e., training is terminated if no improvement in validation error is observed over 128 consecutive epochs. In addition, convergence behavior is monitored, and training is terminated once the validation loss stabilizes and exhibits negligible changes. A mini-batch size of 16 was selected to balance gradient stability and regularization via batch noise—especially important given the

Table 1
Proposed ResNet: training setup and key parameters.

Parameter	Value	Description
Optimizer	Adam	Combines momentum and RMSProp
Initial learning rate	5×10^{-4}	Starting point for weight updates
Learning rate schedule	Piecewise	Multiplied by 0.96 every 128 epochs
Max epochs	1024	Upper limit for training iterations
Mini-batch size	16	Number of samples per training update
Validation frequency	32 iterations	Intervals at which validation loss is evaluated
Residual blocks	2	Each with 2 dense layers and a projection shortcut
Hidden layer width	256 neurons	Used in all dense layers within residual blocks
Activation function	ReLU	Applied after each dense layer
Output layer	Fully connected + Regression layer	Maps to target geometric parameters

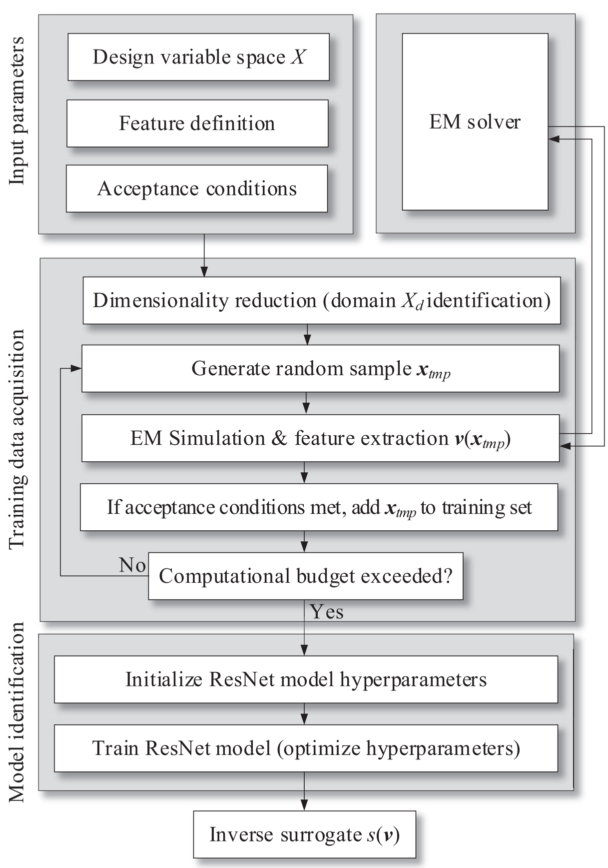

Fig. 7. Proposed modeling framework: flow diagram.

small dataset and the loss landscape's sensitivity. These hyperparameters were informed by prior modeling efforts in related microwave inverse problems. Next, they were refined through iterative experimentation, progressively lowering model complexity and learning rates until stable convergence was achieved.

We also observed that under this training regime, model performance became largely insensitive to initialization—i.e., random weight initialization resulted in similar convergence behavior across seeds. This reinforces the stability of the optimization process under the chosen setup.

Validation loss is evaluated every 32 iterations on a held-out set and monitored throughout training for both hyperparameter tuning and early stopping criteria. The held-out subset comprised approximately 5 % of the available samples, selected at random prior to training. A summary of all key training parameters is provided in Table 1.

2.5. Modeling procedure

This section outlines the complete modeling framework, with a flow diagram provided in Fig. 7. The process is initiated with specifying the input parameters: the design variable space X, the definition of response feature points, and the acceptance criteria for training samples (see Section 2.2). The modeling procedure consists of two primary stages:

- Training data acquisition, carried out through performance-driven sequential sampling.
- Inverse model construction, which involves initializing the hyperparameters of the ResNet architecture and subsequently optimizing them. Further implementation details are given in Section 2.4.

It is significant to highlight that the only user-defined control parameter is the computational budget for data acquisition, specified as the maximum number of training candidates $N_{\max}$. The selection of feature point definitions and acceptance conditions is problem-specific and will be addressed individually for each test case in Section 3.

2.6. Inverse modeling for design optimization

As mentioned earlier, the fundamental role of the inverse surrogate is to expedite circuit optimization. There are several simple steps involved in this process, as explained in Fig. 8. The essential ones are to reformulate the design specifications in relation to the target feature vector $\boldsymbol{v}_t$, and to evaluate the inverse model to find the candidate design $\boldsymbol{x}^{(0)} = s(\boldsymbol{v}_t)$. Owing to the good accuracy of the surrogate, the actual feature vector $\boldsymbol{v}(\boldsymbol{x}^{(0)})$ is likely to be close to $\boldsymbol{v}_t$, so that only a minor tuning (here, realized using an accelerated gradient-based algorithm) is necessary.

Some additional remarks should be made at this point. Recall that the training dataset $\{\boldsymbol{v}(\boldsymbol{x}_t^{(j)})\}$ defined the surrogate's region of validity. Thus, for the feature vector $\boldsymbol{v}_t$ to be attainable, it must be allocated in the neighborhood of $\{\boldsymbol{v}(\boldsymbol{x}_t^{(j)})\}$. Verification of this condition is straightforward: it is sufficient to check whether $\boldsymbol{v}_t \in H_t$ (the convex hull of $\{\boldsymbol{v}(\boldsymbol{x}_t^{(j)})\}$). The latter is the smallest convex set containing $\{\boldsymbol{v}(\boldsymbol{x}_t^{(j)})\}$. In practice, H_t is determined through triangulation of $\{\boldsymbol{v}(\boldsymbol{x}_t^{(j)})\}$, and finding if $\boldsymbol{v}_t$ belongs to any of the simplexes obtained this way. If not, the circuit is unlikely to meet the design requirements.

Subsequently, one needs to discuss the function $U(\boldsymbol{x})$. In this work, we employ the minimax formulation. Let us consider two examples, pertinent to the test cases of Section 3. The first one is a microwave coupler designed to improve its impedance matching $|S_{11}|$ and port isolation $|S_{41}|$ at the target center frequency f_0, and to realize the target power division ratio $K_P = |S_{21}| - |S_{31}|$ (also at f_0). In this case, we set

$$U(\boldsymbol{x}) = \max\{|S_{11}(\boldsymbol{x},f_0)|, |S_{41}(\boldsymbol{x},f_0)|\} + \beta\left[|S_{21}(\boldsymbol{x},f_0)| - |S_{31}(\boldsymbol{x},f_0)| - K_P\right]^2 \quad (3)$$

where the second term enforces the power split condition with $\beta > 0$ being a penalty coefficient.

The second example is a dual-band power divider, designed to ensure equal power split at the operating frequencies f_{01} and f_{02}, and minimization of the input matching $|S_{11}|$, output matching $|S_{22}|$, $|S_{33}|$, and port isolation $|S_{23}|$ at f_{01} and f_{02}simultaneously. In this case, we set

Inverse model for (global) circuit optimization:

1. Represent design specifications in the form of the target feature vector $\boldsymbol{v}_t$;
2. Check if $\boldsymbol{v}_t$ is in the convex hull H of the training feature vector set $\{\boldsymbol{v}(\boldsymbol{x}_t^{(j)})\}_{j=1,\ldots,Nt}$. If not, relax the specifications so that $\boldsymbol{v}_t \in H$.
3. Evaluate the inverse model to generate the initial design $\boldsymbol{x}^{(0)} = s(\boldsymbol{v}_t)$;
4. Fine tune $\boldsymbol{x}^{(0)}$ to obtain the final design

$$\boldsymbol{x}^* = \arg\min_{\boldsymbol{x}} U(\boldsymbol{x}) \tag{2}$$

where $U(\boldsymbol{x})$ is the objective function that represents design quality.

Fig. 8. Application of the inverse surrogate for circuit optimization.

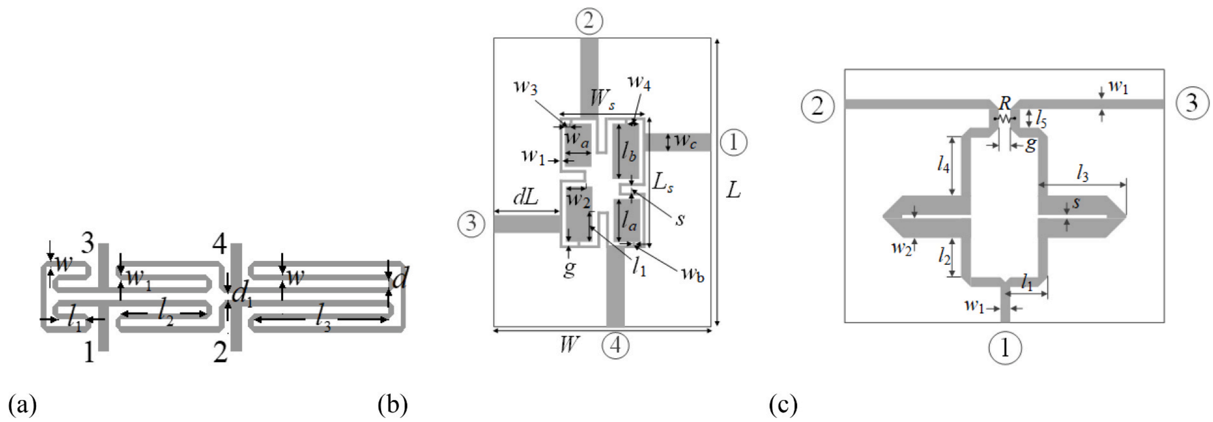

Fig. 9. Test circuits: (a) compact rat-race coupler (Circuit I) [95], compact branch-line coupler (Circuit II) [96], (c) dual-band power divider (Circuit III) [97].

Table 2
Essential parameters of Circuits I through III.

Parameter	Circuit		
	I	II	III
Substrate	RO4003 substrate ($\varepsilon_r = 3.38$, $h = 0.762$ mm)	ε_r – operating parameter, $h = 0.76$ mm	AD250 substrate ($\varepsilon_r = 2.5$, $h = 0.81$ mm)
Design variables	$\boldsymbol{x} = [l_1\ l_2\ l_3\ d\ w\ w_1]^T$	$\boldsymbol{x} = [g\ l_{1r}\ l_a\ l_b\ w_1\ w_{2r}\ w_{3r}\ w_{4r}\ w_a\ w_b]^T$	$\boldsymbol{x} = [l_1\ l_2\ l_3\ l_4\ l_5\ s\ w_2]^T$
Other parameters	$d_1 = d + \lvert w - w_1\rvert$, $d = 1.0$, $w_0 = 1.7$, $l_0 = 15$	$L = 2dL + L_s$, $L_s = 4w_1 + 4g + s + l_a + l_b$, $W = 2dL + W_s$, $W_s = 4w_1 + 4g + s + 2w_a$, $l_1 = l_b l_{1r}$, $w_2 = w_a w_{2r}$, $w_3 = w_{3r} w_a$, $w_4 = w_{4r} w_a$	$w_1 = 2.2$ (to ensure 50-ohm line impedance), $g = 1$
Parameter space X	$\boldsymbol{l} = [2.0\ 7.0\ 12.5\ 0.2\ 0.7\ 0.2]^T$, $\boldsymbol{u} = [4.5\ 12.5\ 22.0\ 0.65\ 1.5\ 0.9]^T$	$\boldsymbol{l} = [0.4\ 0.43\ 5.9\ 7.7\ 0.68\ 0.28\ 0.1\ 0.1\ 2.0\ 0.2]^T$, $\boldsymbol{u} = [1.0\ 0.86\ 14.0\ 16.5\ 1.5\ 0.99\ 0.65\ 0.25\ 5.5\ 0.8]^T$	$\boldsymbol{l} = [14.5\ 1.1\ 13.0\ 0.5\ 1.6\ 0.19\ 3.9]^T$, $\boldsymbol{u} = [37.0\ 16.6\ 35.0\ 15.0\ 5.6\ 1.5\ 5.8]^T$
Frequency range	0.5 GHz to 2.5 GHz	0.1 GHz to 2.5 GHz	0.5 GHz to 7.0 GHz

$$U(\boldsymbol{x}) = \max\left\{ \max_{k\in\{1,2,3\},l\in\{1,2\}} \left|S_{kk}(\boldsymbol{x}, f_{0l})\right|, \max_{l\in\{1,2\}} \left|S_{23}(\boldsymbol{x}, f_{0l})\right| \right\} + \\ +\beta \sum_{l=1}^{2} \left(\left|S_{21}(\boldsymbol{x}_{0l})\right| - \left|S_{31}(\boldsymbol{x}, f_{0l})\right|\right)^2 \tag{4}$$

Similarly to (2), the second term with β > 0 is to enforce the power split condition.

A final note concerns the algorithm used for fine-tuning the circuit designs. In this study, we employ a trust-region gradient-based optimization method [91]. The response Jacobian is initially assessed using finite-difference (FD) approximation [92] during the first iteration. For subsequent steps, a Broyden update formula is applied [93,94], which is adequate given the typically small adjustments required. Thanks to the high predictive accuracy of the inverse surrogate, the fine-tuning process generally completes within fewer than thirty EM simulations.

Table 3
Design objectives for Circuits I through III.

Design goals	Circuit		
	I	II	III
S-parameters	Minimize $\lvert S_{11}\rvert$ at frequency f_0 Minimize $\lvert S_{41}\rvert$ at frequency f_0 Ensure the target power split $\lvert S_{31}\rvert - \lvert S_{21}\rvert = d_s$ at frequency f_0	Minimize $\lvert S_{11}\rvert$ at frequency f_0 Minimize $\lvert S_{41}\rvert$ at frequency f_0 Ensure an equal power split $\lvert S_{31}\rvert - \lvert S_{21}\rvert = 0$ at frequency f_0	Minimize $\lvert S_{11}\rvert$, $\lvert S_{22}\rvert$, and $\lvert S_{32}\rvert$, at target frequencies f_{01} and f_{02} Ensure target power split $\lvert S_{31}\rvert - \lvert S_{21}\rvert = 0$ at f_{01} and f_{02}
Substrate	–	Implement circuit on substrate of target permittivity ε_r	–

Table 4
Response feature definitions for Circuits I through III.

Circuit	Feature vector	Response feature definitions	Acceptance criteria
I	$\boldsymbol{v} = [f_1\ f_2\ l_1\ l_2\ d_s]^T$	• Frequencies f_1 and f_2 marking minima of $\lvert S_{11}\rvert$ and $\lvert S_{41}\rvert$ • $\lvert S_{11}\rvert$ level l_1 at frequency f_1 • $\lvert S_{41}\rvert$ level l_2 at frequency f_2 • Power division $d_s = \lvert S_{31}\rvert - \lvert S_{21}\rvert$ at frequency $(f_1 + f_2)/2$	• $0.5 \le f_1 \le 2.5$ • $0.5 \le f_2 \le 2.5$ • $l_1 \le -10$ dB • $l_2 \le -10$ dB • $-6 \le d_s \le 1$
II	$\boldsymbol{v} = [f_1\ f_2\ l_1\ l_2\ d_s\ \varepsilon_r]^T$	• Frequencies f_1 and f_2 marking minima of $\lvert S_{11}\rvert$ and $\lvert S_{41}\rvert$ • $\lvert S_{11}\rvert$ level l_1 at frequency f_1 • $\lvert S_{41}\rvert$ level l_2 at frequency f_2 • Power division $d_s = \lvert S_{31}\rvert - \lvert S_{21}\rvert$ at frequency $(f_1 + f_2)/2$	• $0.5 \le f_1, f_2 \le 2.5$ • $l_1, l_2 \le -10$ dB • $-1 \le d_s \le 1$ • Substrate permittivity $2.0 \le \varepsilon_r \le 5.0$ is user provided
III	$\boldsymbol{v} = [f_1\ f_2\ l_1\ l_2]^T$	• Frequencies f_1 and f_2 marking minima of $\lvert S_{11}\rvert$ at two operating bands • $\lvert S_{11}\rvert$ level l_1 at frequency f_1 • $\lvert S_{11}\rvert$ level l_2 at frequency f_2	• $1.0 \le f_1 \le 4.0$ • $1.4 \le f_2/f_1 \le 2.0$ • $l_1 \le -10$ dB • $l_2 \le -10$ dB

3. Verification and benchmarking

Verification of the inverse modeling approach of Section 2 is carried out using three microstrip circuits introduced in Section 3.1. The proposed feature-based ResNet surrogate is compared to several forward and inverse metamodels, including diverse regression- and neural-network-based strategies, all outlined in Section 3.2. The proposed model setup and the findings are summarized in Sections 3.3 and 3.4, respectively. Design applications, specifically circuit optimization for various sets of target operating parameters and substrate materials, are elaborated on in Section 3.5.

3.1. Verification circuits

The test circuits are shown in Fig. 9. Table 2 puts together their essential parameters. The characteristics of interest are *S*-parameters versus frequency: S_{11}, S_{21}, S_{31}, and S_{41} for Circuits I and II, as well as S_{11}, S_{22}, S_{32}, and S_{21} for Circuit III. The design specifications are imposed upon these responses as explained in Table 3. For Circuits I and II, the specs are: improvement of the impedance matching and port isolation, as well as ensuring required the power division ratio $d_s = |S_{31}| - |S_{21}|$). Moreover, in the case of Circuit II, the relative permittivity is an additional parameter assuming values from the interval from 2.0 to 5.0. This enables using the surrogate model to optimize the structure for various substrate materials. Finally, for Circuit III, the design specifications include: minimization of $|S_{11}|$, $|S_{22}|$, and $|S_{32}|$ characteristics at the target frequencies f_{01} and f_{02} and to ensure target power split. Model implementation is carried out in CST Microwave Studio, employing the time-domain solver for evaluation. Among the circuits considered, Circuit II is the best example to illustrate parametric redundancy mentioned earlier. Circuit II is a branch-line coupler using compact microstrip resonant cells (CMRCs). While it produces a single operating band, it is described by ten geometric parameters.

This number of parameters is much larger than what is necessary to control the center frequency, impedance matching and port isolation, as well as the power division ratio, meaning that there are multiple parameter vectors that result in similar operating conditions. On the other hand, the dependence between frequency responses and design variable vectors is still unique because the amount of data contained in the circuit characteristics is significant (four *S*-parameters evaluated as hundreds of frequency points).

Table 3 summarizes the design objectives for Circuits I through III.

Table 5
Benchmark modeling procedures.

Model type	Method	Setup
Forward	Kriging [99]	• Second-order polynomial as a trend function • Gaussian correlation function • Hyperparameters found through maximum likelihood optimization
	Radial basis functions (RBF) [99]	• Gaussian basis functions • Scaling coefficient adjusted through cross-validation
	Gaussian process regression (GPR) [100]	• Rational Quadratic kernel functions • Separate GPR models for real and imaginary parts of the response • Hyperparameters optimized through maximum likelihood estimation
	Support vector regression (SVR) [101]	• Gaussian (RBF) kernel function • Separate SVM models for real and imaginary parts of the response • Kernel scale optimized automatically for each frequency
	Artificial neural network (ANN 1) [102]	• Fully connected architecture with ReLU activation • Layers: 512 → 256 → 128 → 64 → output (real + imaginary) • Trained using the Adam optimizer with 400 epochs
	ANN 2 [102]	• Fully connected architecture with ReLU activation • Layers: 128 → 128 → 128 → 128 → 128 → 128 → output (real + imaginary) • Trained using the Adam optimizer with 400 epochs
	ANN 3 [102]	• Fully connected architecture with ReLU activation • Layers: 256 → 256 → 256 → 256 → 256 → output (real + imaginary) • Trained using the Adam optimizer with 400 epochs
Inverse	Ridge Regression (RR) [103]	• Least-squares regression with ridge (L2) regularization • Regularization parameter selected via cross-validation
	RBF [99]	• Gaussian basis functions • Scaling coefficient adjusted through cross-validation
	GPR [100]	• Rational Quadratic kernel functions • Hyperparameters optimized through maximum likelihood estimation
	SVR [101]	• Gaussian(RBF) kernel function • Kernel scale optimized automatically for each frequency
	Random Forest (RF) [104]	• Ensemble of 100 regression trees • Bootstrap aggregation • No explicit hyperparameter tuning
	ANN 1 [102]	• Feedforward neural network • 3 hidden layers (256 → 512 → 256 units), ReLU activations • Trained with Adam optimizer, 512 epochs, with a batch size of 16
	ANN 2 [102]	• Feedforward neural network with attention layer; • 64 → 128 → 64 architecture; • Adam optimizer, for 512 epochs, with a batch size of 16
	ResNet (This work)	• Two residual blocks with fully connected layers (256 units), ReLU activations(Ref Fig. 1) • Adam optimizer, with piecewise learning rate decay. • 1024 epochs, batch size of 16

Table 4 summarizes the definitions of the response features (cf. Section 2.2 for graphical illustrations) and gathers the acceptance criteria for the sampling algorithm outlined in Section 2.2.

Table 6
Modeling results for the circuits of Fig. 9.

Modeling approach		Average RRMSE#		
Model type	Modeling technique	Circuit I	Circuit II	Circuit III
Forward	Kriging	8.0 %	23.3 %	32.3 %
	RBF	8.9 %	27.2 %	34.7 %
	GPR	10.4 %	27.2 %	70.2 %
	SVR	16.1 %	31.6 %	59.6 %
	ANN1	5.9 %	9.0 %	33.3 %
	ANN2	9.8 %	13.9 %	42.3 %
	ANN3	11.2 %	14.5 %	44.1 %
Inverse	LR	3.0 %	7.5 %	12.0 %
	RBF	1.0 %	4.6 %	12.0 %
	SVR	1.1 %	5.6 %	11.4 %
	GPR	0.7 %	4.2 %	11.3 %
	RF	1.5 %	6.9 %	11.9 %
	ANN1	0.7 %	4.5 %	11.2 %
	ANN2	1.0 %	5.5 %	11.5 %
	ResNet (this work)	0.4 %	3.5 %	9.2 %

In the case of inverse models, the numbers in brackets represent the correlation coefficient R^2, which is a suitable metric for model outputs being design variable vectors.

3.2. Benchmark procedures

The proposed modeling technique is evaluated against several forward and inverse surrogate models, as summarized in Table 5. The forward benchmarks include kernel-based regression methods, such as kriging, Gaussian Process Regression (GPR), and Support Vector Regression (SVR), along with three variants of deep neural networks. The inverse benchmark models extend these approaches with additional techniques, including ridge regression and random forest. Forward surrogates are trained using 800 samples distributed across the original parameter space via Latin Hypercube Sampling (LHS) [98]. In contrast, the inverse models—including the proposed approach—are developed in a reduced-dimensional parameter space (see Section 2.3), using training data generated according to the procedure outlined in Section 2.2. The total computational budget for data acquisition is set to $N_{\max} = 1000$. Model performance is assessed using the relative root-mean-squared error (RRMSE), calculated on a separate hold-out set of 100 testing samples.

3.3. Inverse model setup

The setup of the proposed ResNet-based model architecture has been detailed in Section 2.4. The training data was generated as described in Section 2.2 with the computational budget of $N_{\max} = 1000$. The actual number of accepted samples is between 600 and 750 (circuit dependent). The dimensionality of the reduced domain X_d of the surrogate is determined using global sensitivity analysis as explained in Section 2.3. The sensitivity analysis was carried out using fifty random parameter vectors and $C_{\min} = 0.9$. The normalized eigenvalues for the test circuits are as follows:

- Circuit I: $\lambda_1 = 1.00$, $\lambda_2 = 0.65$, $\lambda_3 = 0.51$, $\lambda_4 = 0.46$, $\lambda_5 = 0.37$, $\lambda_6 = 0.28$ (reduced dimensionality $N_d = 3$ with $C = || [\lambda_1 \ldots \lambda_{Nd}]^T || / || [\lambda_1 \ldots \lambda_n]^T || = 0.9$),
- Circuit II: $\lambda_1 = 1.00$, $\lambda_2 = 0.66$, $\lambda_3 = 0.54$, $\lambda_4 = 0.48$, $\lambda_5 = 0.41$, $\lambda_6 = 0.39$, $\lambda_7 = 0.30$,$\lambda_8 = 0.25$,$\lambda_9 = 0.22$,$\lambda_{10} = 0.16$,$\lambda_{11} = 0.13$ (reduced dimensionality $N_d = 4$, $C = 0.9$),
- Circuit III: $\lambda_1 = 1.00$, $\lambda_2 = 0.77$, $\lambda_3 = 0.66$, $\lambda_4 = 0.64$, $\lambda_5 = 0.50$, $\lambda_6 = 0.48$, $\lambda_7 = 0.45$ (reduced dimensionality $N_d = 4$, $C = 0.9$).

3.4. Results and Discussion

Table 6 summarizes the results found by the proposed inverse modeling methodology and the benchmark techniques. The modeling error, RRMSE, is presented in percent. The scatter plots of the EM-simulated and inverse-model-predicted circuit parameters for selected design variables can be found in Figs. 10–12, for Circuits I through III.

We begin our analysis by assessing the performance of the forward surrogate models, which consistently underperform across all test cases, most notably for Circuit III. Both kernel-based methods (GPR, SVR, RBF) and artificial neural networks (ANNs) exhibit poor predictive accuracy. The only partial exception is Circuit I, the simplest of the evaluated structures; however, even in this case, the RRMSE ranges from approximately 10 % to 15 %. For a model to be considered viable for design applications, the RRMSE should ideally remain below 5–7 %. This level of underperformance occurs despite the use of relatively large training datasets (800 samples) and highlights common challenges in data-driven modeling, such as the curse of dimensionality, expansive parameter spaces, and strong nonlinearities in the system response.

In contrast, the inverse metamodels (see the lower portion of Table 6) exhibit markedly better performance. This improvement is primarily due to the use of explicit dimensionality reduction and the relatively weak nonlinearity between the reduced feature space and the design variables. Additionally, the number of characteristic features is lower than the number of design parameters (typically five versus more than seven), which further simplifies the modeling task. These factors contribute to significantly lower RRMSE values compared to the forward models. Among the inverse approaches, the proposed ResNet-based model consistently outperforms the benchmark techniques—including ridge regression, RBF, SVR, GPR, and ANNs—delivering higher predictive accuracy across all test cases.

The proposed modeling technique is superior to all benchmark methods regarding RRMSE. Only for specific test cases, individual benchmark approaches achieve comparable error values, e.g., ANN1 for Circuit I (0.7 % versus 0.4 % for our method). The main reason for this performance is a particular choice of the model architecture, as elucidated in Section 2.4, which prevents overfitting and allows handling parameter redundancy (i.e., the existence of different designs corresponding to similar values of response features). The mentioned redundancy generally makes simpler regression techniques (e.g., GPR or SVR) prevail over deep neural network surrogates, as reflected in the

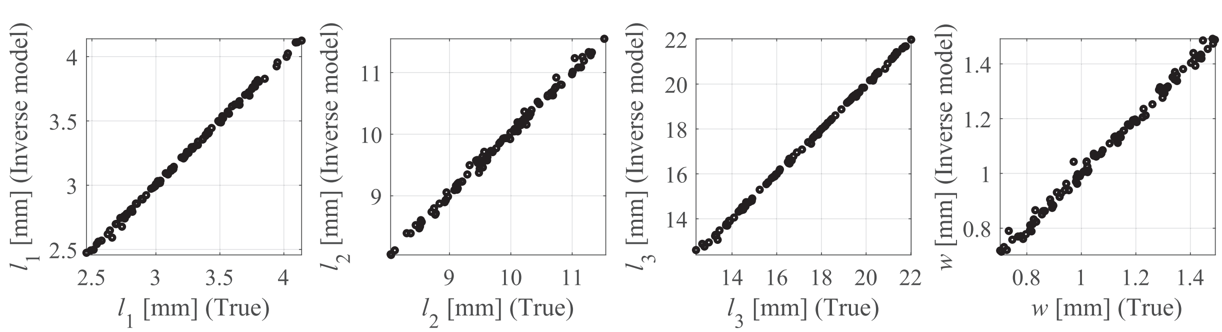

Fig. 10. Scatter plots of selected design variables (inverse model outputs versus true feature values) for Circuit I.

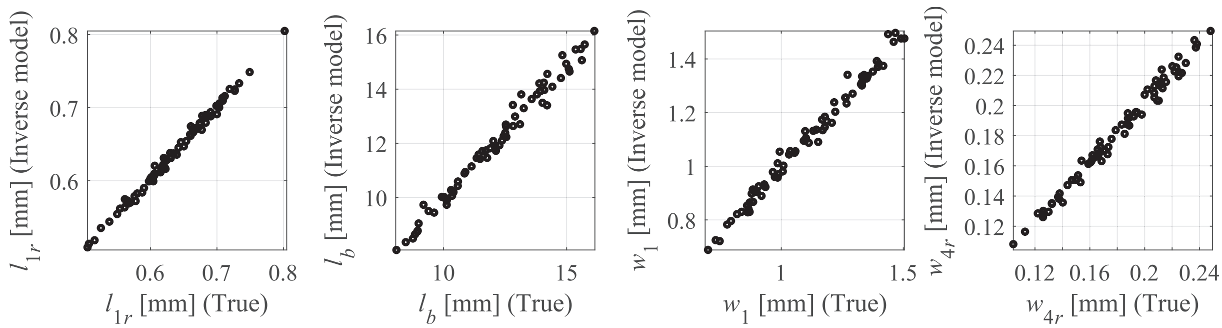

Fig. 11. Scatter plots of selected design variables (inverse model outputs versus true feature values) for Circuit II.

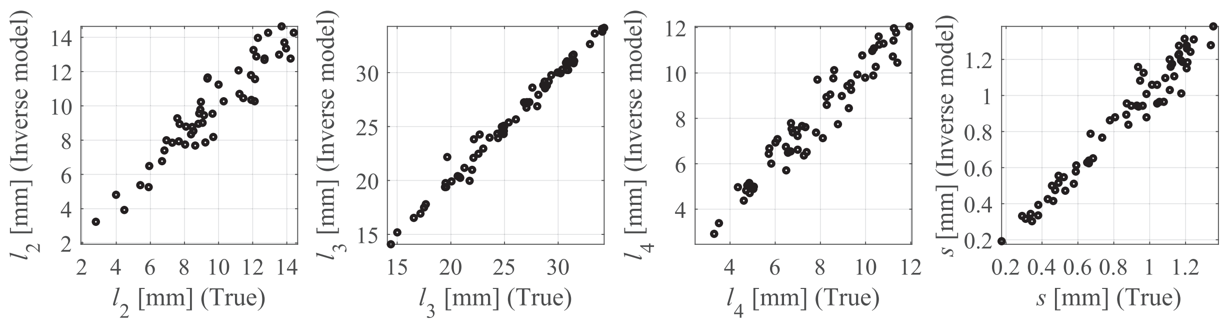

Fig. 12. Scatter plots of selected design variables (inverse model outputs versus true feature values) for Circuit III.

Table 7
Design optimization scenarios for Circuits I through III.

Circuit	Design targets & optimization cost	Design case 1	2	3	4
I	Target operating parameters: f_0 [GHz], d_s [dB]	1, 0	1.2, –6	1.5, –1	1.8, –3
	Target feature vector $\boldsymbol{v} = [f_1\ f_2\ l_1\ l_2\ d_s]^T$	[1.0 1.0 0.1 0.1 0]^T	[1.2 1.2 0.1 0.1 –6]^T	[1.5 1.5 0.1 0.1 –1]^T	[1.8 1.8 0.1 0.1 –3]^T
	Fine tuning cost#	13	17	16	14
II	Target operating parameters: f_0 [GHz], ε_r	1.2, 2.5	1.0, 3.0	2.0, 3.5	1.5, 4.3
	Target feature vector $\boldsymbol{v} = [f_1\ f_2\ l_1\ l_2\ d_s\ \varepsilon_r]^T$	[1.2 1.2 0.1 0.1 0 2.5]^T	[1 1 0.1 0.1 0 3.0]^T	[2 2 0.1 0.1 0 3.5]^T	[1.5 1.5 0.1 0.1 0 4.3]^T
	Fine tuning cost#	19	20	19	21
III	Target operating parameters: f_{01} and f_{02} [GHz]	1.5, 2.3	2.5, 3.8	3.0, 4.5	4.0, 6.0
	Target feature vector $\boldsymbol{v} = [f_1\ f_2\ l_1\ l_2]^T$	[1.5 2.5 0.1 0.1]^T	[2.5 3.8 0.1 0.1]^T	[3.0 4.5 0.1 0.1]^T	[4.0 6.0 0.1 0.1]^T
	Fine tuning cost#	16	18	17	23

Cost expressed in numbers of EM simulations.

RRMSE data in Table 6.

It is also worth noting that the computational cost of training the surrogate model is negligible compared to that of electromagnetic (EM) simulations required for data generation. In the considered cases, EM simulations account for approximately 95–98 % of the total modeling cost, whereas network training constitutes only a minor fraction. This further supports the practical efficiency of the proposed framework.

3.5. Design applications

The main intended application of the proposed inverse surrogate is global optimization. The details of the surrogate-aided optimization procedure have been highlighted in Section 2.6. For demonstration, Circuits I, II, and III are optimized for different values of target operating parameters (including substrate permittivity for Circuit II). The specifications are incorporated into the target feature vector $\boldsymbol{v}_t$, which becomes the inverse surrogate model input. The surrogate generates the starting point $\boldsymbol{x}^{(0)} = s(\boldsymbol{v}_t)$, which is subject to fine-tuning, realized using the accelerated gradient-based algorithm outlined in Section 2.6.

For each circuit, four design scenarios are considered as listed in Table 7. The same table reports the overall cost of the optimization process. Furthermore, Figs. 13–15 illustrate the circuit frequency characteristics at the inverse-model-predicted initial designs and the final designs for Circuits I through III, respectively.

It should be emphasized that the design specifications (e.g., center frequencies) are first converted to the associated target feature vectors, which, in turn, become the inverse model inputs. The feature vector is subsequently used to obtain the initial design for further fine-tuning. For certain feature points, such as frequencies and power division ratio, the design specifications and feature vectors coincide. For the remaining parameters, specifically, the levels of $|S_{11}|$ or $|S_{41}|$, one needs to choose reasonable values, here, 0.1 (or –20 dB). Finally, the desired relative permittivity ε_r is directly applied as the model input in the case of Circuit II.

The results shown in Figs. 13 through 15 underscore the exquisite performance of the proposed inverse modeling methodology. For all tested circuits and design cases, the surrogate-produced starting designs demonstrate satisfactory quality. In particular, these designs either meet the assumed specifications or are close to fulfilling those. Consequently, the gradient-based fine-tuning is achieved at the cost of few EM analyses to identify the optimum parameter vectors. The average tuning cost is just 18 EM analyses, and it is weakly dependent on the problem dimensionality. The average cost for Circuit I (six parameters) is 15 simulations, the cost for Circuit III (seven parameters) is 18.5 analyses,

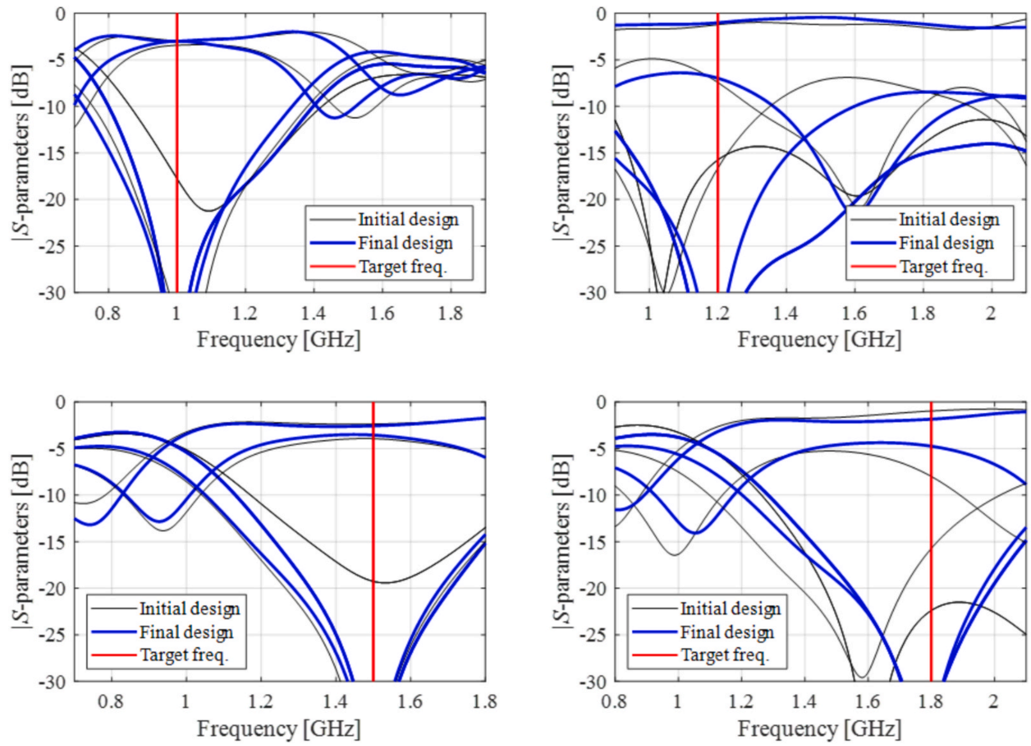

Fig. 13. Optimization results for Circuit I: the picture shows the responses at the initial design predicted by the proposed inverse surrogate, and EM-simulated characteristics at the fine-tuned design. Cases 1 through 4 are shown from top left to bottom right.

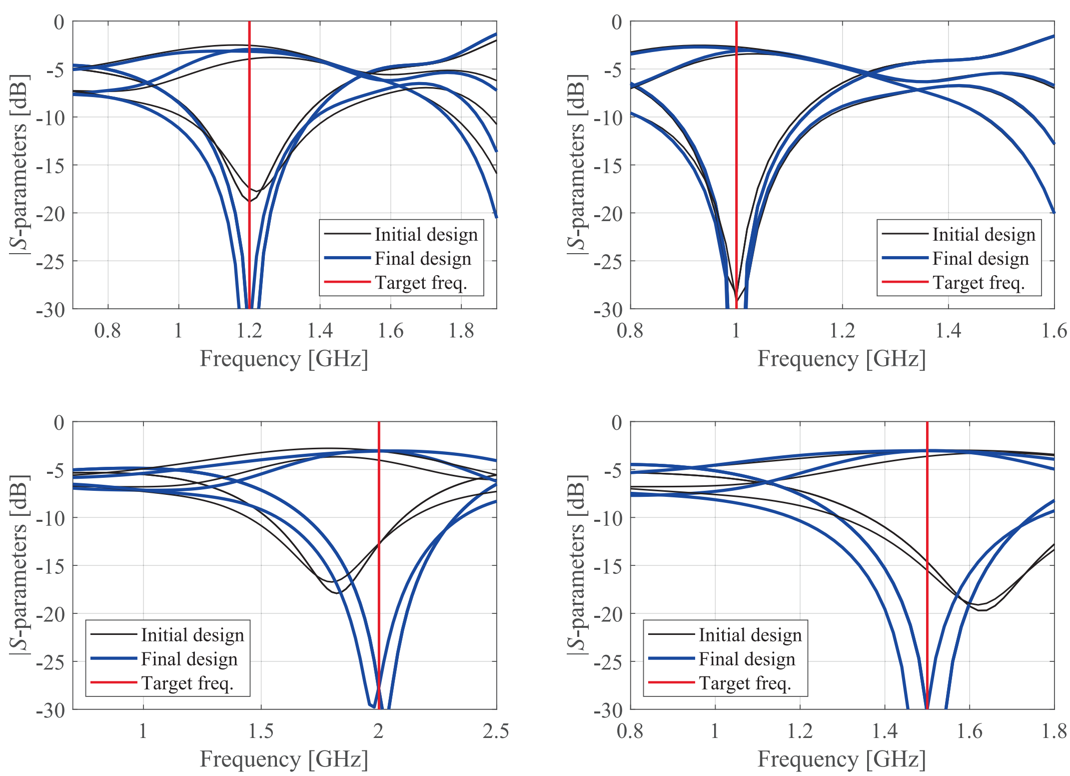

Fig. 14. Optimization results for Circuit II: the picture shows the responses at the initial design predicted by the proposed inverse surrogate, and EM-simulated characteristics at the fine-tuned design. Cases 1 through 4 are shown from top left to bottom right.

and it is 19.8 simulations for Circuit II (ten parameters).

The aforementioned results verify design applicability of the presented inverse modeling approach, especially regarding its ability to facilitate global optimization of circuits across extended ranges of operating conditions and material parameters.

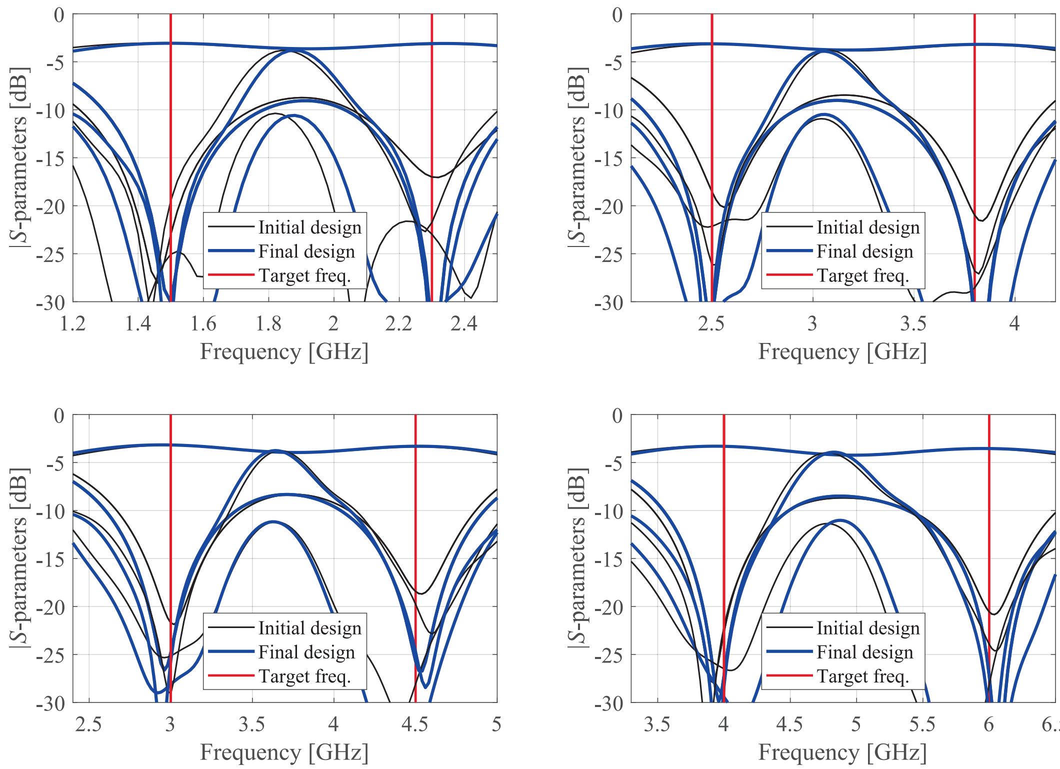

Fig. 15. Optimization results for Circuit III: the picture shows the responses at the initial design predicted by the proposed inverse surrogate, and EM-simulated characteristics at the fine-tuned design. Cases 1 through 4 are shown from top left to bottom right.

4. Conclusion

This research introduced a novel inverse modeling approach for microwave passive components, tailored for design applications. The core of the method is a deep-learning surrogate that maps the circuit's operating parameters directly to its design variables. By exploiting the weak nonlinearity between these two domains, the model achieves high predictive accuracy across wide ranges of operating frequencies, bandwidths, and other relevant conditions, while requiring only a reduced number of training data. The modeling process is enhanced through explicit dimensionality reduction, enabled by fast global sensitivity analysis. The surrogate itself is implemented as a custom deep residual network (ResNet), designed to capture the relationship between key response features and the component's geometric parameters. To further improve efficiency, the model focuses only on a small set of critical performance indicators, rather than modeling the full frequency response.

The proposed technique has been comprehensively verified using three microstrip components and compared against a range of benchmark methods, including both forward and inverse metamodels. Given the complexity of the test cases, all forward modeling approaches yielded limited predictive accuracy, resulting in models of little or no practical design value. In contrast, the inverse models delivered significantly better performance, with the proposed ResNet-based surrogate consistently demonstrating the highest reliability across all circuits. To highlight the practical utility of the method, the circuits were optimized under various performance scenarios. In every case, the surrogate successfully generated high-quality initial designs that required only minimal refinement, achieved with an average cost of less than twenty electromagnetic (EM) simulations. The focus of the future work will be on further enhancing model robustness and computational efficiency. Planned improvements include incorporating spatial domain confinement and adopting variable-resolution EM analysis.

CRediT authorship contribution statement

Slawomir Koziel: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Kaustab C. Sahu:** Writing – original draft, Visualization, Methodology, Investigation. **Anna Pietrenko-Dabrowska:** Writing – review & editing, Writing – original draft, Project administration, Conceptualization.

Funding

This work was partly supported by the Icelandic Research Fund Grant 2410297 and by the National Science Centre of Poland Grant 2024/55/B/ST7/01413.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

The authors thank Dassault Systemes, France, for making CST Microwave Studio available.

Data availability

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

References

[1] S. Leng, S. Sui, Y. Cui, Microwave sensing aqueous solutions based on ring-assisted Mach–Zehnder interferometer, IEEE Microwave Wireless Techn. Lett. 33 (2) (2023) 216–219, https://doi.org/10.1109/LMWC.2022.3213898.

[2] P. Li, J. Ren, Y. Chen, X. Ren, K.D. Xu, Y.Z. Yin, M. Shen, Design of low-cost single-layer 2-bit reflective programmable metasurface based on folded ground, IEEE Trans. Microwave Theory Techn. 71 (8) (2023) 3455–3465, https://doi.org/10.1109/TMTT.2023.3259526.

[3] Z. Luo, L. Yang, T. Su, X. Zhu, R. Gómez-García, Single-ended and balanced flat-group-delay RF low-pass filters with input-quasi-reflectionless behavior for digital-communication systems, IEEE Trans. Microwave Theory Techn. 73 (1) (2025) 321–334, https://doi.org/10.1109/TMTT.2024.3430485.

[4] S. Lee, J. Lee, B. Lee, J. Lim, J. Han, J. Lee, K-band waveguide resonator bandpass filter structure with adjustable transmission zeros for producing various symmetric and asymmetric responses, IEEE Access 13 (2025) 8664–8675, https://doi.org/10.1109/ACCESS.2025.3527545.

[5] L. Liu, Q. Xiang, Q. Feng, A self-packaged tunable LC fourth-order filter with cross-coupling and bandwidth compensation, IEEE Microwave Wireless Techn. Lett. 35 (4) (2025) 420–423, https://doi.org/10.1109/LMWT.2025.3535706.

[6] C.-M. Liu, L.-P. Carignan, K. Wu, Hybrid metallo-dielectric waveguide architecture for compact low-loss THz applications, IEEE Microwave Wireless Techn. Lett. 34 (5) (2024) 486–489, https://doi.org/10.1109/LMWT.2024.3380445.

[7] L. Lum, Z.K. Ng, R.T. Jiang, C.W. Tan, E.H.T. Teo, B.K. Tay, Anisotropic microwave properties of vertically aligned carbon nanotube arrays, IEEE Trans. Microwave Theory Techn. 72 (2) (2024) 919–926, https://doi.org/10.1109/TMTT.2023.3305072.

[8] X. Tan, Y. Zhang, A compact rat-race coupler with widely tunable frequency- and power-dividing ratio, IEEE Microwave Wireless Techn. Lett. 34 (1) (2024) 17–20, https://doi.org/10.1109/LMWT.2023.3330491.

[9] S. Kumari, P.K. Jain, M. Mahto, Study of a modified complementary electric split ring resonator-based metamaterial slow wave structure, IEEE Trans. Plasma Sc. 51 (8) (2023) 2205–2211, https://doi.org/10.1109/TPS.2023.3299418.

[10] A. Noor, S. Koziel, Dual-polarized wideband bandpass metasurface-based filter, IEEE Ant. Wireless Propag. Lett. 24 (1) (2024) 404–408, https://doi.org/10.1109/LAWP.2023.3325930.

[11] X. Tang, Y. Zhang, S. Jian, J. Zhang, L. Zhang, Q. Wang, X. Li, Z. Duan, Metamaterial-inspired backward wave oscillator using a dual band coaxial coupler, IEEE Electron Dev. Lett. 44 (9) (2023) 1559–1562, https://doi.org/10.1109/LED.2023.3294759.

[12] X. He, D. Ding, X.Y. Zhang, An SIW-based wideband 5G mmWave bandpass filter by using multimode resonator, IEEE Microwave Wireless Techn. Lett. 35 (4) (2025) 412–415, https://doi.org/10.1109/LMWT.2025.3533482.

[13] P. Chu, M. Luo, W. Zhang, F. Zhu, L. Liu, K. Wu, Dual band dual mode substrate integrated waveguide filter with mixed coupling, IEEE Microwave Wireless Techn. Lett. 34 (11) (2024) 1239–1242, https://doi.org/10.1109/LMWT.2024.3462729.

[14] Y.-P. Hong, I.-J. Hwang, D.-J. Yun, D.-J. Lee, I.-H. Lee, Design of single-layer metasurface filter by conformational space annealing algorithm for 5G mm-wave communications, IEEE Access 9 (2021) 29764–29774, https://doi.org/10.1109/ACCESS.2021.3059019.

[15] K.Y. Kapusuz, A.V. Berghe, S. Lemey, H. Rogier, Partially filled half-mode substrate integrated waveguide leaky-wave antenna for 24 GHz automotive radar, IEEE Ant. Wireless Propag. Lett. 20 (1) (2021) 33–37, https://doi.org/10.1109/LAWP.2020.3038201.

[16] S. Paul, M.J. Akhtar, Novel metasurface lens-based RF sensor structure for SAR microwave imaging of layered media, IEEE Sens. J. 21 (16) (2021) 17827–17837, https://doi.org/10.1109/JSEN.2021.3084614.

[17] M.M. Mansour, H. Kanaya, High-efficient broadband CPW RF rectifier for wireless energy harvesting, IEEE Microwave Wireless Comp. Lett. 29 (4) (2019) 288–290, https://doi.org/10.1109/LMWC.2019.2902461.

[18] A. Pietrenko-Dabrowska, S. Koziel, Globalized parametric optimization of microwave passive components using simplex-based surrogates, IEEE Trans. Microwave Theory Techn. 72 (5) (2024) 2899–2913, https://doi.org/10.1109/TMTT.2023.3318702.

[19] J. Fu, J. Jin, J. Yang, Z. Liu, J. Qian, H. Lin, Feature-assisted neural network surrogate-based multiphysics optimization for microwave filters, IEEE Microwave Wireless Techn. Lett. 34 (5) (2024) 474–477, https://doi.org/10.1109/LMWT.2024.3374792.

[20] J.E. Rayas-Sánchez, et al., Microwave modeling and design optimization: the legacy of John Bandler, IEEE Trans. Microwave Theory Techn. 73 (1) (2025) 87–101, https://doi.org/10.1109/TMTT.2024.3437198.

[21] D. Klink, P. Meyer, W. Steyn, Efficient yield estimation of multiband patch antennas using NLPLS-based PCE, IEEE Trans. Ant. Propag. 70 (8) (2022) 7037–7045, https://doi.org/10.1109/TAP.2021.3138496.

[22] Z. Zhang, B. Liu, Y. Yu, M. Imran, Q.S. Cheng, M. Yu, A surrogate modeling space definition method for efficient filter yield optimization, IEEE Microwave Wireless Techn. Lett. 33 (6) (2023) 631–634, https://doi.org/10.1109/LMWT.2023.3243524.

[23] P. Manfredi, Probabilistic uncertainty quantification of microwave circuits using Gaussian processes, IEEE Trans. Microwave Theory Techn. 71 (6) (2023) 2360–2372, https://doi.org/10.1109/TMTT.2022.3228953.

[24] M. Mahmood, A. Koc, R. Morawski, T. Le-Ngoc, Achieving capacity gains in practical full-duplex massive MIMO systems: a multi-objective optimization approach using hybrid beamforming, IEEE Open J. Comm. Soc. 5 (2024) 2268–2286, https://doi.org/10.1109/OJCOMS.2024.3384988.

[25] V.H. Hasbestan, Y. Farhang, K. Majidzadeh, C. Ghobadi, Multi-objective hybrid optimization algorithm for design a printed MIMO antenna with n78–5G NR frequency band applications, IEEE Access 11 (2023) 68231–68242, https://doi.org/10.1109/ACCESS.2023.3292307.

[26] S. Koziel, A. Pietrenko-Dabrowska, Constrained multi-objective optimization of compact microwave circuits by design triangulation and Pareto front interpolation, Eur. J. Op. Res. 299 (1) (2022) 302–312, https://doi.org/10.1016/j.ejor.2021.08.021.

[27] E.H. Houssein, M.E. Hosney, D. Oliva, E.M.G. Younis, A.A. Ali, W.M. Mohamed, An efficient discrete rat swarm optimizer for global optimization and feature selection in chemoinformatics, Knowl.-Based Syst. 275 (2023) 110697, https://doi.org/10.1016/j.knosys.2023.110697.

[28] S. Kumar, B.S. Yildiz, P. Mehta, N. Panagant, S.M. Sait, S. Mirjalili, A.R. Yildiz, Chaotic marine predators algorithm for global optimization of real-world engineering problems, Knowl.-Based Syst. 261 (2023) 110192, https://doi.org/10.1016/j.knosys.2022.110192.

[29] A. Afsari, A. Abbosh, Y. Rahmat-Samii, Quasi-gradient nonlinear simplex optimization method in electromagnetics, IEEE Access 11 (2023) 59599–59611, https://doi.org/10.1109/ACCESS.2023.3285602.

[30] T. Brown, P. Mojabi, Cascaded metasurface design using electromagnetic inversion with gradient-based optimization, IEEE Trans. Ant. Propag. 70 (3) (2022) 2033–2045, https://doi.org/10.1109/TAP.2021.3119115.

[31] A. Pietrenko-Dabrowska, S. Koziel, Fast EM-driven parameter tuning of microwave circuits with sparse sensitivity updates via principal directions, Knowl.-Based Syst. 252 (2022) 109388, https://doi.org/10.1016/j.knosys.2022.109388.

[32] A. Pietrenko-Dabrowska, S. Koziel, Low-cost design optimization of microwave passives using multi-fidelity EM simulations and selective Broyden updates, IEEE Trans. Microwave Theory Techn. 70 (11) (2022) 4765–4771, https://doi.org/10.1109/TMTT.2022.3207482.

[33] W. Na, K. Liu, W. Zhang, F. Feng, J. Zhang, H. Xie, D. Jin, Q.J. Zhang, Advanced EM optimization using adjoint-sensitivity-based multifeature surrogate for microwave filter design, IEEE Microwave Wireless Techn. Lett. 34 (1) (2024) 1–4, https://doi.org/10.1109/LMWT.2023.3329783.

[34] M. Li, Q.J. Zhang, F. Feng, J. Xue, W. Liu, J. Jin, J. Zhang, W. Zhang, Advanced mesh space mapping approach with fast coarse mesh models comprising sharpening structural processing and mesh deformation, IEEE Trans. Microwave Theory Techn. 72 (3) (2024) 1578–1590, https://doi.org/10.1109/TMTT.2023.3308106.

[35] S. Koziel, A. Pietrenko-Dabrowska, Expedited feature-based quasi-global optimization of multi-band antennas with Jacobian variability tracking, IEEE Access 8 (2020) 83907–83915, https://doi.org/10.1109/ACCESS.2020.2992134.

[36] S. Koziel, A. Pietrenko-Dabrowska, Design-oriented computationally-efficient feature-based surrogate modelling of multi-band antennas with nested kriging, AEU Int. J. Electronics Comm. 120 (2020) 153202, https://doi.org/10.1016/j.aeue.2020.153202.

[37] W. Na, T. Bai, D. Jin, H. Xie, W. Zhang, Q.J. Zhang, Advanced neural space mapping-based inverse modeling method for microwave filter design, IEEE Microwave Wireless Techn. Lett. 35 (1) (2025) 12–15, https://doi.org/10.1109/LMWT.2024.3503572.

[38] D. Spina, F. Ferranti, G. Antonini, T. Dhaene, L. Knockaert, Efficient variability analysis of electromagnetic systems via polynomial chaos and model order reduction, IEEE Trans. Comp. Packaging Manuf. Techn. 4 (6) (2014) 1038–1051, https://doi.org/10.1109/TCPMT.2014.2312455.

[39] J.E. Rayas-Sanchez, S. Koziel, J.W. Bandler, Advanced RF and microwave design optimization: a journey and a vision of future trends, IEEE J. Microwaves. 1 (1) (2021) 481–493, https://doi.org/10.1109/JMW.2020.3034263.

[40] A. Pietrenko-Dabrowska, S. Koziel, Q.J. Zhang, Cost-efficient two-level modeling of microwave passives using feature-based surrogates and domain confinement, Electronics. 12 (17) (2023) 3560, https://doi.org/10.3390/electronics12173560.

[41] A. Kumar, T. Khan, Forward and inverse modeling of multiple input multiple output microstrip antennas for satellite communication using machine learning algorithms, AEU - Int. J. Electr. Comm. 201 (2025) 155951, https://doi.org/10.1016/j.aeue.2025.155951.

[42] B. Liu, L. Xue, H. Fan, Y. Ding, M. Imran, T. Wu, An efficient and general automated power amplifier design method based on surrogate model assisted hybrid optimization technique, IEEE Trans. Microwave Theory Techn. 73 (2) (2025) 926–937, https://doi.org/10.1109/TMTT.2024.3518913.

[43] O. Oluyemi, P. Laforge, A. Bais. Space mapping technique on microwave filter circuit model using artificial neural network for parameter extraction. *IEEE Canadian Conf. Electr. Comp. Eng. (CCECE)*, Regina, SK, Canada, 2023, pp. 337-341, 2023, DOI: 10.1109/CCECE58730.2023.10288716.

[44] J. Ossorio, J.C. Melgarejo, V.E. Boria, M. Guglielmi, J.W. Bandler, On the alignment of low-fidelity and high- fidelity simulation spaces for the design of microwave waveguide filters, IEEE Trans. Microwave Theory Techn. 66 (12) (2018) 5183–5196, https://doi.org/10.1109/TMTT.2018.2871022.

[45] J.C. Melgarejo, J. Ossorio, S. Cogollos, M. Guglielmi, V.E. Boria, J.W. Bandler, On space mapping techniques for microwave filter tuning, IEEE Trans. Microwave Theory Techn. 67 (12) (2019) 4860–4870, https://doi.org/10.1109/TMTT.2019.2944361.

[46] S. Koziel, S.D. Unnsteinsson, Expedited design closure of antennas by means of trust-region-based adaptive response scaling, IEEE Antennas Wireless Prop. Lett. 17 (6) (2018) 1099–1103, https://doi.org/10.1109/LAWP.2018.2834145.

[47] Y. Yu, Z. Zhang, Q.S. Cheng, B. Liu, Y. Wang, C. Guo, T.T. Ye, State-of-the-art: AI-assisted surrogate modeling and optimization for microwave filters, IEEE Trans. Microwave Theory Techn. 70 (11) (2022) 4635–4651, https://doi.org/10.1109/TMTT.2022.3208898.

[48] X. Yu, X. Hu, Z. Liu, C. Wang, W. Wang, F.M. Ghannouchi, A method to select optimal deep neural network model for power amplifiers, IEEE Microwave Wireless Comp. Lett. 31 (2) (2021) 145–148, https://doi.org/10.1109/LMWC.2020.3038821.

[49] J. Yang, F. Wang, Auto-ensemble: an adaptive learning rate scheduling based deep learning model ensembling, IEEE Access 8 (2020) 217499–217509, https://doi.org/10.1109/ACCESS.2020.3041525.

[50] J. Cai, C. Yu, L. Sun, S. Chen, J.B. King, Dynamic behavioral modeling of RF power amplifier based on time-delay support vector regression, IEEE Trans. Microwave Theory Techn. 67 (2) (2019) 533–543, https://doi.org/10.1109/TMTT.2018.2884414.

[51] Z. Zhou, Z. Wei, J. Ren, Y.X. Sun, Y. Yin, G.F. Pedersen, M. Shen, Bayesian-inspired sampling for efficient machine-learning-assisted microwave component design, IEEE Trans. Microwave Theory Techn. 72 (2) (2024) 996–1007, https://doi.org/10.1109/TMTT.2023.3298194.

[52] X. Wang, G.G. Wang, B. Song, P. Wang, Y. Wang, A novel evolutionary sampling assisted optimization method for high dimensional expensive problems, IEEE Tran. Evol. Comp. 23 (2019) 815–827, https://doi.org/10.1109/TEVC.2019.2890818.

[53] J. Ma, J. Qiao, G. Watkins, S. Dang, R. Piechocki, J. Haine, M. Beach, Revolutionizing microwave circuit design with machine learning: challenges and opportunities, IEEE Comm. Mag. 62 (8) (2024) 100–106, https://doi.org/10.1109/MCOM.001.2400209.

[54] F. Garbuglia, D. Spina, D. Deschrijver, I. Couckuyt, T. Dhaene, Bayesian optimization for microwave devices using deep GP spectral surrogate models, IEEE Trans. Microwave Theory Techn. 71 (6) (2023) 2311–2318, https://doi.org/10.1109/TMTT.2022.3228951.

[55] A. Pietrenko-Dabrowska, S. Koziel, L. Golunski, Cost-efficient globalized parameter optimization of microwave components through response-feature surrogates and nature-inspired metaheuristics, IEEE Access 12 (2024) 79051–79065, https://doi.org/10.1109/ACCESS.2024.3407978.

[56] S. Koziel, A. Pietrenko-Dabrowska, Machine-learning-based global optimization of microwave passives with variable-fidelity EM models and response features, Sci. Rep. 14 (2024) 6250, https://doi.org/10.1038/s41598-024-56823-7.

[57] M. Sedaghat, R. Trinchero, Z.H. Firouzeh, F.G. Canavero, Compressed machine learning-based inverse model for design optimization of microwave components, IEEE Trans. Microwave Theory Techn. 70 (7) (2022) 3415–3427, https://doi.org/10.1109/TMTT.2022.3166151.

[58] Q. Zhou, Y. Wang, P. Jiang, X. Shao, S.-K. Choi, J. Hu, L. Cao, X. Meng, An active learning radial basis function modeling method based on self-organization maps for simulation-based design problems, Knowl.-Based Syst. 131 (2017) 10–27, https://doi.org/10.1016/j.knosys.2017.05.025.

[59] M.A. Tulum, A.S. Turk, P. Mahouti, Data driven surrogate modeling of phase array antennas using deep learning for millimetric band applications, IEEE Access 11 (2023) 114415–114423, https://doi.org/10.1109/ACCESS.2023.3324733.

[60] S. Koziel, A. Pietrenko-Dabrowska, Optimization of microwave components using machine learning and rapid sensitivity analysis, Sci. Rep. 14 (2024) 31265. DOI: s41598-024-82701-3.

[61] B.S. Chouhan, A. Nawaz, A. Das, K.M. Rohith, A. Ahmad, G. Kumar, Machine learning-driven ultra-broadband terahertz multilayer metamaterial, J. Lightwave Techn. 43 (6) (2025) 2734–2744, https://doi.org/10.1109/JLT.2024.3509492.

[62] H.G. Sarabi, G. Karimi, M. Mohammadi, D. Abbott, Ultra-wideband miniaturized Wilkinson power divider based on rectangular multi-section resonators with neural network-based enhancement, AEU – Int. J. Electr. Comm. 208 (2026) 156220, https://doi.org/10.1016/j.aeue.2026.156220.

[63] Y. He, J. Huang, W. Li, L. Zhang, S.W. Wong, Z.N. Chen, Hybrid method of artificial neural network and simulated annealing algorithm for optimizing wideband patch antennas, IEEE Trans. Antennas Propag. 72 (1) (2024) 944–949, https://doi.org/10.1109/TAP.2023.3331249.

[64] A. Touhami, S. Collardey, A. Sharaiha, A global optimization method for wideband and small supergain arrays design using Artificial Neural Network, IEEE Open J. Ant. Propag. 4 (2023) 1016–1028, https://doi.org/10.1109/OJAP.2023.3321932.

[65] Y. Liu, et al., An efficient method for antenna design based on a self-adaptive Bayesian neural network-assisted global optimization technique optimization, IEEE Trans. Antennas Propag. 70 (12) (2022) 11375–11388, https://doi.org/10.1109/TAP.2022.3211732.

[66] A. Sonker, A.K. Nayak, T. Goel, A. Patnaik, Multifunctional antenna design for wireless consumer electronic devices: a soft-computing approach, IEEE Canadian J. Electrical Computer Eng. 46 (2) (2023) 144–156, https://doi.org/10.1109/ICJECE.2023.3243994.

[67] Y. Liu, P. Chen, J. Tian, J. Xiao, S. Noghanian, Q. Ye, Hybrid ANN-GA optimization method for minimizing the coupling in MIMO antennas, AEU – Int. J. Electronics Comm 175 (2024) 155068, https://doi.org/10.1016/j.aeue.2023.155068.

[68] S.-H. Javid-Hosseini, P. Ghazanfarianpoor, V. Nayyeri, P. Colantonio, A unified neural network-based approach to nonlinear modeling and digital predistortion of RF power amplifier, IEEE Trans. Microwave Theory Techn. 72 (9) (2024) 5031–5038, https://doi.org/10.1109/TMTT.2024.3374092.

[69] J. Wang, C. Fan, Y. Liao, L. Zhou, Pattern-to-absorption prediction for multilayered metamaterial absorber based on deep learning, IEEE Microwave Wireless Techn. Lett. 34 (5) (2024) 463–466, https://doi.org/10.1109/LMWT.2024.3385982.

[70] Z.Ž. Stanković, D.I. Olćan, N.S. Dončov, B.M. Kolundžija, Consensus deep neural networks for antenna design and optimization, IEEE Trans. Antennas Propag. 70 (7) (2022) 5015–5023, https://doi.org/10.1109/TAP.2021.3138220.

[71] K. Yasmeen, K.V. Mishra, A.V. Subramanyam, S.S. Ram, Circularly polarized Fabry–Pérot cavity sensing antenna design using generative model, IEEE Sens. Lett. 7 (2) (2023) 3500304, https://doi.org/10.1109/LSENS.2023.3237366.

[72] X. Hu, S. Wen, H.K. Lam, Redundancy reduction penalty term of loss function in deep neural network, Knowl.-Based Syst. (2025) 113776, https://doi.org/10.1016/j.knosys.2025.113776.

[73] C. Yu, Q. Li, F. Feng, Q.J. Zhang, Convolutional neural network with adaptive batch-size training technique for high-dimensional inverse modeling of microwave filters, IEEE Microwave Wireless Techn. Lett. 33 (2) (2023) 122–125, https://doi.org/10.1109/LMWC.2022.3208355.

[74] A. Gupta, E.A. Karahan, C. Bhat, K. Sengupta, U.K. Khankhoje, Tandem neural network based design of multiband antennas, IEEE Trans. Antennas Propag. 71 (8) (2023) 6308–6317, https://doi.org/10.1109/TAP.2023.3276524.

[75] A. Jayasimhan, P. Pabitha, Enhancing accuracy of compressed Convolutional Neural Networks through a transfer teacher and reinforcement guided training curriculum, Knowl.-Based Syst. 306 (2024) 112719, https://doi.org/10.1016/j.knosys.2024.112719.

[76] J. Li, et al., An adaptive evolutionary neural network-based optimization design method for wideband dual-polarized antennas, IEEE Trans. Antennas Propag. 71 (10) (2023) 8165–8172, https://doi.org/10.1109/TAP.2023.3301882.

[77] G. Qi, S. Bao, Y. Wei, Y. Zhang. Accurate antenna design by deep auto-encoder surrogate model assisted particle swarm optimization. *2022 IEEE 5th Int. Conf. Electronic Information Comm. Technology (ICEICT)*, Hefei, China, 2022, pp. 875–880.

[78] J. Tan, Y. Shao, J. Zhang, J. Zhang, Efficient antenna modeling and optimization using multifidelity stacked neural network, IEEE Trans. Antennas Propag. 72 (5) (2024) 4658–4663, https://doi.org/10.1109/TAP.2024.3384758.

[79] M.R. Khan, C.L. Zekios, S. Bhardwaj, S.V. Georgakopoulos, A deep learning convolutional neural network for antenna near-field prediction and surrogate modeling, IEEE Access 12 (2024) 39737–39747, https://doi.org/10.1109/ACCESS.2024.3377219.

[80] K. Tang, C. Yu, Y. Liu, Cascaded neural network module for digital predistortion under various operating conditions, IEEE Microwave Wireless Techn. Lett. 34 (1) (2024) 96–98, https://doi.org/10.1109/LMWT.2023.3335471.

[81] Y. Su, et al., Time-domain scattering parameters-based neural network inverse model for antenna designs, IEEE Antennas Wireless Propag. Lett. 23 (7) (2024) 1976–1980, https://doi.org/10.1109/LAWP.2024.3375303.

[82] J. Jin, Q. Su, Y. Xu, Z. He, Y. Lu, Efficient radiation pattern prediction of array antennas based on complex-valued graph neural networks, IEEE Antennas Wireless Propag. Lett. 21 (12) (2022) 2467–2471, https://doi.org/10.1109/LAWP.2022.3197441.

[83] W. Liu, et al., Second-order sensitivity neural network modeling approach with applications to microwave devices, IEEE Trans. Microwave Theory Techn. 72 (7) (2024) 3980–3992, https://doi.org/10.1109/TMTT.2023.3335405.

[84] A. Faraji, S.A. Sadrossadat, J. Jin, W. Na, F. Feng, Q.-J. Zhang, Hybrid batch-normalized deep feedforward neural network incorporating polynomial regression for high-dimensional microwave modeling, IEEE Trans. Circ. Syst. i. 71 (3) (2024) 1245–1258, https://doi.org/10.1109/TCSI.2023.3340219.

[85] S. Koziel, A. Pietrenko-Dabrowska, L. Leifsson, Improved efficacy behavioral modeling of microwave circuits through dimensionality reduction and fast global sensitivity analysis, Sci. Rep. 14 (2024) 19465, https://doi.org/10.1038/s41598-024-70246-4.

[86] K. He, X. Zhang, S. Ren, J. Sun. Deep residual learning for image recognition. in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.* (*CVPR*), Las Vegas, NV, USA, 2016, pp. 770–778, DOI: 10.48550/arXiv.1512.03385.

[87] I. Goodfellow, Y. Bengio, A. Courville, Deep Learning, MIT Press, Cambridge, MA, USA, 2016.

[88] C. Szegedy, S. Ioffe, V. Vanhoucke, A. Alemi. Inception-v4, inception-ResNet and the impact of residual connections on learning. in *Proc. AAAI Conf. Artif. Intell.*, San Francisco, CA, USA, 2017, pp. 4278–4284, DOI: 10.48550/arXiv.1602.07261.

[89] D.P. Kingma, J. Ba. Adam: A method for stochastic optimization. in *Proc. Int. Conf. Learn. Represent. (ICLR)*, San Diego, CA, USA, 2015. [Online]. Available: https://arxiv.org/abs/1412.6980.

[90] S. Ruder. An overview of gradient descent optimization algorithms. arXiv preprint arXiv:1609.04747, 2016. Available: https://arxiv.org/abs/1609.04747.

[91] A.R. Conn, N.I.M. Gould, P.L. Toint, *Trust Region Methods*, MPS-SIAM Series on Optimization, 2000.

[92] H. Levy, F. Lessman, Finite Difference Equations, Dover Publications Inc., 1992.

[93] S. Koziel, A. Pietrenko-Dabrowska, Reliable EM-driven size reduction of antenna structures by means of adaptive penalty factors, IEEE Trans. Antenna Propag. 70 (2) (2021) 1389–1401, https://doi.org/10.1109/TAP.2021.3111285.

[94] C.G. Broyden, A class of methods for solving nonlinear simultaneous equations, Math. Comp. 19 (1965) 577–593.

[95] S. Koziel, A.T. Sigurdsson, Performance-driven modeling of compact couplers in restricted domains, Int. J. RF & Microwave CAE 28 (6) (2018), https://doi.org/10.1002/mmce.21296.

[96] C. Tseng, C. Chang, A rigorous design methodology for compact planar branch-line and rat-race couplers with asymmetrical T-structures, IEEE Trans. Microwave Theory Techn. 60 (7) (2012) 2085–2092, https://doi.org/10.1109/TMTT.2012.2195019.

[97] Z. Lin, Q.-X. Chu, „A novel approach to the design of dual-band power divider with variable power dividing ratio based on coupled-lines, Prog. Electromagn. Res. 103 (2010) 271–284, https://doi.org/10.2528/PIER10012202.
[98] B. Beachkofski, R. Grandhi. Improved distributed hypercube sampling. *American Institute of Aeronautics and Astronautics*, paper AIAA 2002-1274, 2002, DOI: 10.2514/6.2002-1274.
[99] A.I.J. Forrester, A.J. Keane, Recent advances in surrogate-based optimization, Prog. Aerospace Sci. 45 (2009) 50–79, https://doi.org/10.1016/j.paerosci.2008.11.001.
[100] C.K.I. Williams and C.E. Rasmussen. Gaussian processes for regression. in D. Touretzky, M.C. Mozer, and M. Hasselmo (Eds.) *Advances in Neural Information Proc. Syst.*, MIT Press. 8, 1995.
[101] M. Awad, R. Khanna, Support vector regression. Efficient Learning Machines, Apress, Berkeley, CA, 2015.
[102] R. Vang-Mata (Ed.), *Multilayer perceptrons*, Nova Science Pub. Inc., 2020.
[103] A.E. Hoerl, R.W. Kennard, Ridge regression: Biased estimation for nonorthogonal problems, Technometrics. 12 (1) (1970) 55–67, https://doi.org/10.1080/00401706.1970.10488634.
[104] L. Breiman, Random forests, Mach. Learn. 45 (1) (2001) 5–32, https://doi.org/10.1023/A:1010933404324.