A Unified Frequency-Wise Electromagnetic Modeling Framework with Improved Generalizability and Scalability

Zhao Zhou, Zhaohui Wei, *Member, IEEE*, Jian Ren, *Member, IEEE*, Yingzeng Yin, *Member, IEEE*, Yue Wang, *Member, IEEE*, Tse-Tin Chan, *Member, IEEE*

***Abstract*—Electromagnetic (EM) modeling accelerates the design and optimization of EM structures by predicting their frequency-dependent features such as $|S_{11}|$. Current methodologies constrain each model to a fixed frequency range at specified points, thereby necessitating the training of separate models for different frequency conditions, which limits generalizability and scalability in real-world applications. This paper introduces a novel EM modeling framework with improved generalizability and scalability. It significantly enhances modeling accuracy for frequency conditions not previously encountered. The model can be easily scaled up by incorporating new data with variable frequency conditions, thereby further improving the generalizability. Our method incorporates a frequency-wise learning strategy that enforces a robust understanding of the frequency-dependent working mechanism of EM structures. We demonstrate the effectiveness of our approach through multiple implementations with variable frequency conditions. The comparative results validate the improved generalizability and scalability, showcasing its potential to simplify and enhance EM design processes.**

***Index Terms*—Electromagnetic modeling, generalizability, machine learning, scalability.**

I. Introduction

ELECTROMAGNETIC (EM) modeling plays a pivotal role in the design and analysis of electromagnetic structures, which are integral to a wide array of applications ranging from telecommunications to medical devices. The key challenge in this domain is to predict the frequency-dependent characteristics, such as $|S_{11}|$, according to the geometric parameters, which is essential for understanding the performance of EM structures.

Full-wave simulation is a common technique for EM modeling. However, each simulation process can require prohibitively expensive computational resources. The design and optimization of an EM structure require many modeling iterations, resulting in high computational costs. Therefore, machine learning-based approaches for EM modeling have been proposed to serve as faster alternatives to full-wave simulation for predicting EM characteristics. By training using sufficient simulation data, the models can replicate the relationship between geometric parameters of EM structures and their frequency-dependent features. Although developing these models requires significant simulation data, once trained, they enable fast and accurate modeling, significantly accelerating the long-term design and optimization processes of EM structures. Our previous works [1], [2] significantly reduce the data needed for training the surrogate models by integrating with high-quality data acquisition methods.

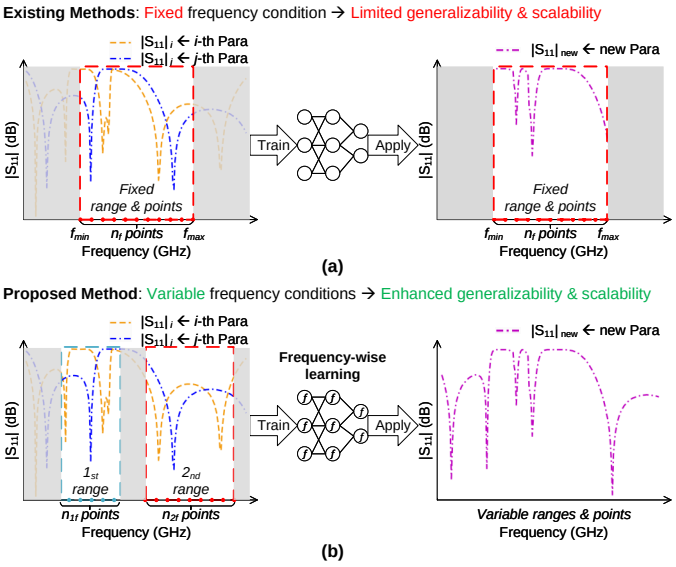


Fig. 1. Illustration of motivation. (a) Existing methods have fixed frequency ranges and points, restricting their generalizability and scalability. (b) The proposed method aims to enhance generalizability (improve the accuracy at unseen frequency outside the training frequency range) and scalability (be compatible with variable frequency ranges and points)

Multiple types of EM modeling methods have been investigated, such as Gaussian process regression [3]–[6], polynomial chaos expansion [7]–[9], kriging [10]–[13], support vector regression [14]–[16], and neural networks [17]–[35].

Gaussian process regression provides probabilistic predictions by assuming a Gaussian process with characterized mean and covariance functions. J. Jacobs used Gaussian process regression to estimate how the varying finite substrate and ground plane size affected the gain of microstrip antennas at fixed frequency points [4]. Z. Zhang *et al.* developed a two-level Gaussian process regression method for accurate surrogate modeling of antennas [5]. The first-level model imitated the relationship between the geometric parameters and EM responses over the frequency band of interest, and it was complemented by the second-level model to predict the difference between the first-level predictions and simulated

This work was supported in part by the Research Matching Grant Scheme from the Research Grants Council of Hong Kong. *(Corresponding author: Tse-Tin Chan.)*

Zhao Zhou, Yue Wang, and Tse-Tin Chan are with the Department of Mathematics and Information Technology, The Education University of Hong Kong, Hong Kong, China (e-mail: zzhou.xd@gmail.com; wyue@eduhk.hk; tsetinchan@eduhk.hk).

Zhaohui Wei, Jian Ren, and Yingzeng Yin are with the National Key Laboratory of Radar Detection and Sensing, Xidian University, Xi'an, 710071, China (e-mail: zhaohui_wei@163.com; renjianroy@gmail.com; yzyin@mail.xidian.edu.cn).

EM responses. C. Hu *et al.* proposed a nonstationary Gaussian process surrogate model to be compatible with the dynamic conditions during the optimization process [6]. Its nonstationarity was achieved by dynamically adjusting the mean and covariance functions based on the dynamic decision variables. Gaussian process regression requires careful selection of kernel functions and could be computationally expensive for large datasets.

Polynomial chaos expansion aims to formulate the uncertainty propagation features of complex systems. It represents a stochastic process as a series expansion of orthogonal polynomials, which are determined based on the probability distribution of the input random variables. J. Du *et al.* applied polynomial chaos expansion to model the relationship between random disturbances and a parsimonious representation of the far-field radiation of antennas [7]. A. Petrocchi *et al.* analyzed the residue calibration uncertainty and consequent non-linear capacitances in microwave transistor non-linear models [8]. Expanded vector spherical harmonics of the far field radiated by antennas subject to random variables were modeled through polynomial chaos expansion in [9]. Polynomial chaos expansion requires knowledge of probability distributions.

Kriging, also known as kriging interpolation, uses weighted averages of known points to predict unknown points based on spatial correlations. S. Koziel *et al.* established a co-kriging model for accurate antenna modeling [10], which was trained using sparse high-fidelity and dense low-fidelity EM simulation data. A triangulation-based constrained kriging modeling method was proposed in [11] for contemporary antenna structures. They significantly reduced the training data needed to develop the surrogate model by restricting the solving space based on a set of optimized reference designs. A similar modeling technique was introduced in [12], replacing the optimized reference designs with a small set of random observables. Integrating the performance-driven data confinement with multi-resolution simulations further enhanced the modeling performance [13]. Kriging is suitable for spatially stationary data distributions.

Support vector regression, a type of support vector machine designed for regression tasks, optimizes a function that fits the training data by minimizing deviations from the true target values within a specified margin, while maintaining model simplicity. D. Prado *et al.* applied support vector regression to model the elements of shaped-beam reflectarray antennas as a substitute for full-wave simulation [14], [15]. J. Jacobs *et al.* established a Bayesian support-vector-regression model for planar antennas. They reduced the number of required training points by exploiting coarse-discretization EM simulations. The modeling performance of support vector regression is sensitive to the choice of hyperparameters.

Inspired by the human brain, neural networks with interconnected artificial neurons organized in layers can learn complex projections and data patterns. C. Roy *et al.* employed an artificial neural network (ANN) to map the equivalent circuit model parameters to EM model geometric parameters of the target EM structure within a band of interest [24]. The target EM structure is segmented into a series of discontinuities. Each discontinuity and coupling between the discontinuities is associated with an equivalent circuit model to extract the circuit parameters over the desired band. This methodology ensures the consistency of the extracted circuit parameters over a wide frequency band. H. Kabir *et al.* presented a systematic neural network framework for inverse modeling of microwave waveguide filters [25]. The source data with non-unique multivalued solutions were separated into multiple groups with only unique solutions, which were used to train multiple inverse models. These models were integrated to improve the modeling performance by alleviating the effects of non-uniqueness. They formulated a set of neural network submodels for modeling waveguide filters in [26], decomposing the task into multiple low-dimensional problems and thus reducing the computational cost. Neural network modeling was combined with physics-informed domain confinement for small-sample antenna modeling in [27]. W. Liu *et al.* proposed model-order reduction-based neuro-impedance matrix transfer functions to enhance the modeling accuracy for two-port microwave components [28]. Similarly, a pole-residue-based transfer function was integrated with artificial neural networks to model the reflection coefficients of frequency-selective surfaces over desired frequency ranges [29].

The aforementioned modeling methods are inherently bound to fixed frequency ranges and points, which show limited generalizability and scalability, as seen in Fig. 1(a). In real-world scenarios, the existing simulation or measurement data might derive from multiple design cases at variable frequency ranges. For example, wireless communication terminal manufacturers design and update an EM structure for diverse products, such as smartphones and tablets, accumulating simulation and measurement data at variable frequency ranges. It would cause extra costs if we re-simulate or re-measure these products to include all frequency ranges. A distinct surrogate model is required for the EM structure over each frequency range with fixed points of interest. The well-trained model cannot directly predict the EM responses for unseen frequency conditions, where the frequencies fall outside the training frequency range. Although the EM similarity laws allow indirect estimations over new frequency ranges by proportioning the geometric parameters, the predictable frequency points are fixed to be proportional to the training frequency points, multiple proportioning processes are needed to make up a wide frequency range, and the modeling performance severely deteriorates. Training a multitude of models tailored to multiple frequency bands demands significant computational resources and time. To resolve this challenge, an intuitive solution is to extend the target frequency range and densify the points to cover all the desired frequency ranges and points, resulting in redundant solution space and increased complexity.

The increasing diversity of EM applications necessitates a more efficient modeling strategy that accommodates variable frequency conditions without needing multiple models. To address the challenge, this paper proposes an innovative EM modeling framework that unifies the modeling process across different frequency conditions into a single, cohesive model, as shown in Fig. 1(b). It is mainly achieved through frequency-wise learning. Unlike existing methods that attempt to imitate the relationship between geometric parameters and EM

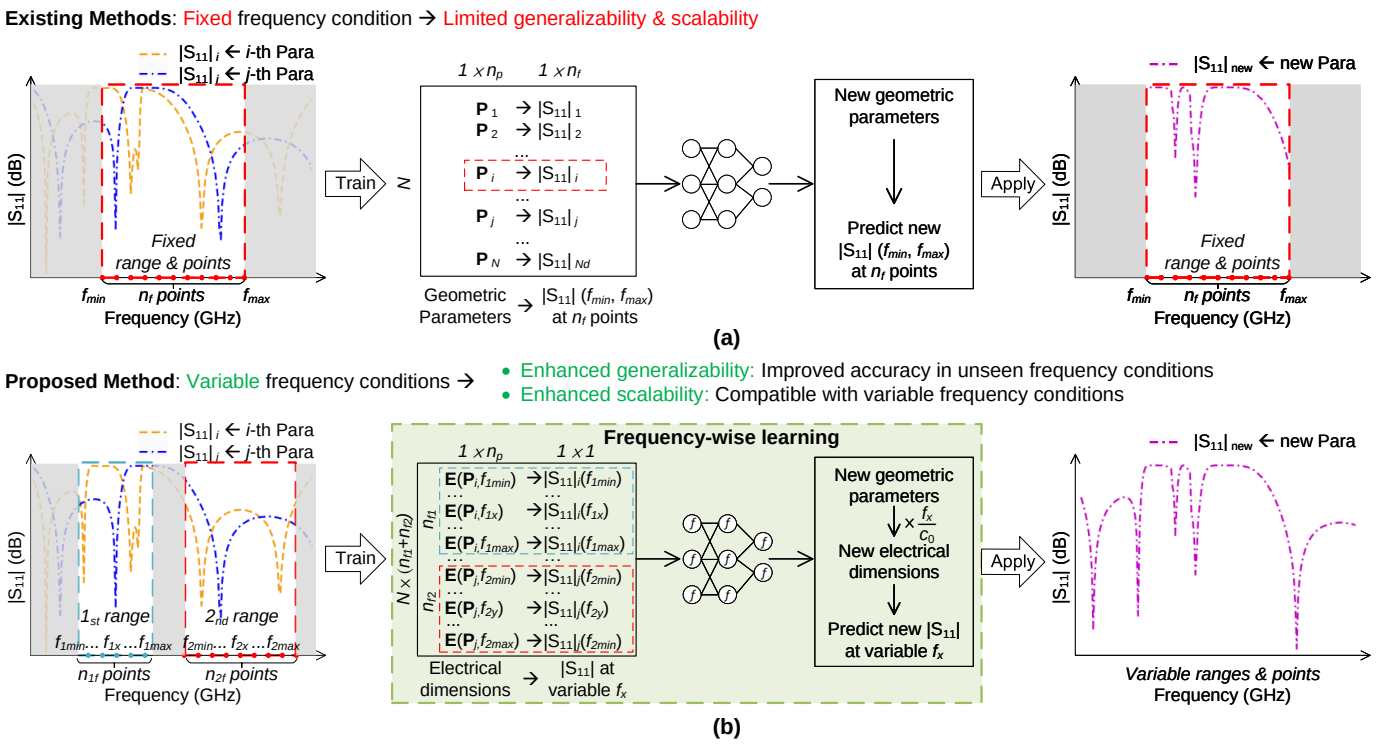


Fig. 2. Comparison of working principle between the existing and proposed methods. (a) Existing methods use fixed frequency ranges and points, limiting generalizability and scalability. (b) The proposed method integrates frequency-wise learning, leading to enhanced generalizability (improved accuracy in unseen frequency range) and scalability (compatible with variable frequency range and points).

responses over the frequency band of interest, our approach focuses on modeling EM responses at arbitrary frequency points subject to certain frequency-dependent electrical dimensions. These electrical dimensions are transformed from the geometric parameters. By leveraging frequency-wise learning, our approach dynamically adapts to varying frequency bands and points, enhancing generalizability and scalability. We conduct multiple implementations to evaluate the generalizability and scalability of the proposed method. Implementation A: Meander-Line Polarizer considers a five-dimensional modeling task with two different frequency conditions. Implementation B: Planar Metasurface Lens increases the dimensionality to ten and has three different frequency conditions. The comparative results demonstrate the improved generalizability and scalability of the proposed method. The potentials and limitations of our approach are discussed when further increasing the dimensionality and complexity in Implementation C. By addressing the generalizability and scalability issues inherent in current EM modeling practices, our work enables more efficient design and optimization of EM structures.

The main contributions of this paper are summarized as follows:

1) A novel frequency-wise modeling framework is proposed for accelerating the design and optimization of EM structures.
2) We improve the generalizability by enforcing the surrogate model to obtain a robust understanding of the EM similarity laws and non-linear proportioning characteristics of EM structures.
3) We enhance the scalability to be compatible with variable frequency conditions.
4) We conduct a comprehensive comparison of our proposed method against the existing methods through multiple implementations, which involve increased dimensionality and variable frequency conditions.

The remainder of this paper is organized as follows. Section II explains the working principle of the proposed framework. Section III evaluates the generalizability and scalability of our framework through Implementation A: Meander-Line Polarizer, which considers a five-dimensional modeling problem under two different frequency conditions. Section IV conducts further validations as the dimensionality increases to ten, and three different frequency conditions are considered, referred to as Implementation B: Planar Metasurface Lens. Section V clarifies the potentials and limitations of our method for higher-dimensional and more complex modeling applications. Section VI gives the conclusion.

II. Methodology

Modeling an EM structure aims to develop a numerical surrogate model to quickly predict its EM responses over the frequency band of interest (such as $|S_{11}|$) as its geometric parameters change. Fig. 2 illustrates and compares the working principle of the existing and proposed methods, where the modeling of $|S_{11}|$ is used as an example. Note that $|S_{11}|$ is only used as a representative of general EM responses for simplicity in Fig. 2. The theoretical analysis in Section II

is applicable to various EM responses such as EM fields, scattering parameters, and phases.

A. Problem Statement

1) Training in Fixed Frequency Conditions: The working principle of the existing methods, for example, Gaussian process regression, kriging, support vector regression, and neural networks, is shown in Fig. 2(a). Each pair of geometric parameters and EM responses over the desired frequency band is formatted as two fixed-sized vectors, $\mathbf{P}$ and $\mathbf{S}(f_{min}, f_{max}, n_f)$, which are normalized to $\overline{\mathbf{P}}$ and $\overline{\mathbf{S}}(f_{min}, f_{max}, n_f)$, respectively, between 0 and 1. Here, f_{min} and f_{max} mark the frequency range of interest, and n_f denotes the number of frequency points. Existing methods manage to estimate a surrogate model $\mathbf{F}_e(\cdot)$ that projects $\overline{\mathbf{P}}$ to $\overline{\mathbf{S}}(f_{min}, f_{max}, n_f)$,

$$\overline{\mathbf{S}_n^*}(f_{min}, f_{max}, n_f) = \mathbf{F}_e(\overline{\mathbf{P}_n}),\ n \in \{1, 2, \cdots, N\}. \quad (1)$$

$\overline{\mathbf{S}^*}(f_{min}, f_{max}, n_f)$ denotes the predicted normalized EM responses. N equals the number of training data. The surrogate model $\mathbf{F}_e(\cdot)$ is optimized by minimizing the mean squared error (MSE) between the predicted and actual normalized EM responses ($\overline{\mathbf{S}^*}(f_{min}, f_{max}, n_f)$ and $\overline{\mathbf{S}}(f_{min}, f_{max}, n_f)$, respectively) for all the N training data,

$$\begin{aligned}\mathbf{O}_e &= \min_{\mathbf{F}_e(\cdot)} \frac{1}{N}\sum_{n=1}^{N}\frac{1}{n_f}|\overline{\mathbf{S}_n^*}(f_{min}, f_{max}, n_f) \\ &\qquad -\overline{\mathbf{S}_n}(f_{min}, f_{max}, n_f)|^2 \\ &= \min_{\mathbf{F}_e(\cdot)} \frac{1}{N}\sum_{n=1}^{N}\frac{1}{n_f}|\mathbf{F}_e(\overline{\mathbf{P}_n}) - \overline{\mathbf{S}_n}(f_{min}, f_{max}, n_f)|^2.\end{aligned} \quad (2)$$

Here, $\mathbf{O}_e$ denotes the optimization metric for training the surrogate model.

2) Modeling in Fixed Frequency Conditions: During the design and optimization process, in the fixed frequency condition (from f_{min} to f_{max} with n_f points) that $\mathbf{F}_e(\cdot)$ is trained on, $\mathbf{F}_e(\cdot)$ can predict EM responses for new input geometric parameters $\mathbf{P}_{new}$, denoted as $\mathbf{S}_{new}^*(f_{min}, f_{max}, n_f)$. $\mathbf{S}_{new}^*(f_{min}, f_{max}, n_f)$ is denormalized from $\overline{\mathbf{S}_{new}^*}(f_{min}, f_{max}, n_f)$, which is obtained by

$$\overline{\mathbf{S}_{new}^*}(f_{min}, f_{max}, n_f) = \mathbf{F}_e(\overline{\mathbf{P}_{new}}). \quad (3)$$

The modeling error (L_e) is calculated by

$$\begin{aligned}L_e &= \frac{1}{n_f}|\overline{\mathbf{S}_{new}^*}(f_{min}, f_{max}, n_f) - \overline{\mathbf{S}_{new}}(f_{min}, f_{max}, n_f)|^2 \\ &= \frac{1}{n_f}\sum_{x=1}^{n_f}|\overline{\mathbf{S}_{new}^*}(f_x) - \overline{\mathbf{S}_{new}}(f_x)|^2,\end{aligned} \quad (4)$$

where $x \in \{1, 2, \cdots, n_f\}$ and $f_x \in \underbrace{\{f_{min}, \cdots, f_{max}\}}_{n_f\ points}$.

3) Modeling in Unseen Frequency Conditions: EM responses in unseen frequency conditions, $\overline{\mathbf{S}_{new}}(f_{umin}, f_{umax}, n_{uf})$, cannot be directly predicted through this surrogate model but only by combining with the EM similarity laws. The EM similarity laws are derived from the scaling properties of Maxwell's equations. There are three prerequisite conditions for the EM similarity laws: all geometric parameters are scaled by a factor k; the frequency points are scaled inversely by k; the material properties (permittivity ε, permeability μ, conductivity σ) are frequency-independent and keep unchanged. Under the prerequisite conditions, the EM fields, scattering parameters, and phases approximately remain the same at the proportionally increased frequency points if proportionally decreasing the geometric parameters,

$$\overline{\mathbf{S}}(f_{min}, f_{max}, n_f) \approx \overline{\mathbf{S}}(m \cdot f_{min}, m \cdot f_{max}, n_f) \quad (5)$$

$$\overline{\mathbf{S}}(f_{min}, f_{max}, n_f) \leftarrow \text{Simulate}(\mathbf{P}), \quad (6)$$

$$\overline{\mathbf{S}}(m \cdot f_{min}, m \cdot f_{max}, n_f) \leftarrow \text{Simulate}(\frac{1}{m} \cdot \mathbf{P}). \quad (7)$$

With assumptions of $f_{umin} = m \cdot f_{min}$, $f_{umax} = m \cdot f_{max}$, $n_{uf} = n_f$, and the n_{uf} frequency points being proportional to the n_f counterparts, $\overline{\mathbf{S}_{new}}(f_{umin}, f_{umax}, n_{uf})$ can be expressed as

$$\begin{aligned}\overline{\mathbf{S}_{new}}(f_{umin}, f_{umax}, n_{uf}) &= \overline{\mathbf{S}}(m \cdot f_{min}, m \cdot f_{max}, n_f) \\ &\approx \overline{\mathbf{S}}(f_{min}, f_{max}, n_f).\end{aligned} \quad (8)$$

Thus, $\overline{\mathbf{S}_{new}^*}(f_{umin}, f_{umax}, n_{uf})$ can be predicted by combining the surrogate model with a proportioning process,

$$\begin{aligned}\overline{\mathbf{S}_{new}^*}(f_{umin}, f_{umax}, n_{uf}) &\approx \overline{\mathbf{S}}(f_{min}, f_{max}, n_f) \\ &= \mathbf{F}_e(\overline{\mathbf{P}'_{unew}}), \quad (9) \\ \overline{\mathbf{P}'_{unew}} &= \text{normalize}(m \cdot \mathbf{P}_{unew}). \quad (10)\end{aligned}$$

Although modeling unseen frequency conditions is possible using the EM similarity laws, this approach suffers from low flexibility, high complexity, and reduced accuracy. The assumptions should hold to obtain $\overline{\mathbf{S}_{new}^*}(f_{umin}, f_{umax}, n_{uf})$ by combining the surrogate model with one proportioning process, showing low flexibility. Otherwise, multiple proportioning processes are required to make up $\overline{\mathbf{S}_{new}^*}(f_{umin}, f_{umax}, n_{uf})$, which significantly increases the computational complexity. Detailed processes of both cases will be presented in the implementations in Section III and IV. As $\overline{\mathbf{S}}(f_{min}, f_{max}, n_f)$ and $\overline{\mathbf{S}}(m \cdot f_{min}, m \cdot f_{max}, n_f)$ are not strictly equal, their complete relationship can be formulated as,

$$\overline{\mathbf{S}}(m \cdot f_{min}, m \cdot f_{max}, n_f) = \overline{\mathbf{S}}(f_{min}, f_{max}, n_f) \pm \Delta(m). \quad (11)$$

Here, $\Delta(m)$ represents the non-linear responses caused by the proportioning impedance and radiation properties. $\Delta(m)$ varies with respect to the proportioning scale m of the unseen frequency conditions. As $\Delta(m)$ increases to be noticeable ($\Delta(m) \gg \varepsilon$, where ε is a minimum threshold), the generated $\overline{\mathbf{S}_{new}^*}(f_{umin}, f_{umax}, n_{uf})$ might deviate from the actual $\overline{\mathbf{S}_{new}}(f_{umin}, f_{umax}, n_{uf})$. The modeling error ($L_{eu}$) for unseen frequency conditions is calculated by

$$\begin{aligned}L_{eu} = \frac{1}{n_{uf}}|&\overline{\mathbf{S}_{new}^*}(f_{umin}, f_{umax}, n_{uf}) \\ &-\overline{\mathbf{S}_{new}}(f_{umin}, f_{umax}, n_{uf})|^2.\end{aligned} \quad (12)$$

Replacing the right side of (12) with (9) and (11), L_{eu} is converted to

$$\begin{aligned} L_{eu} &= \frac{1}{n_f}|\overline{\mathbf{S}^*_{new}}(f_{min}, f_{max}, n_f) \\ &\quad -\overline{\mathbf{S}_{new}}(m \cdot f_{min}, m \cdot f_{max}, n_f)|^2 \\ &= \frac{1}{n_f}|\overline{\mathbf{S}^*_{new}}(f_{min}, f_{max}, n_f) \\ &\quad -[\overline{\mathbf{S}_{new}}(f_{min}, f_{max}, n_f) \pm \Delta(m)]|^2 \\ &= \frac{1}{n_f}\sum_{x=1}^{n_f}|\overline{\mathbf{S}^*_{new}}(f_x) - [\overline{\mathbf{S}_{new}}(f_x) \pm \Delta(m)]|^2, \quad (13) \end{aligned}$$

where $x \in \{1, 2, \cdots, n_f\}$ and $f_x \in \underbrace{\{f_{min}, \cdots, f_{max}\}}_{n_f \text{ points}}$. Compared with L_e in (4), L_{eu} reveals that the existing methods suffer from deteriorated modeling accuracy for the unseen frequency conditions.

4) Limited Generalizability and Scalability: Therefore, the existing methods have limited generalizability and scalability: the modeling for the unseen frequency conditions suffers from low flexibility, high complexity, and low accuracy; their working mechanism refrains them from being compatible with variable frequency conditions.

B. Unified Frequency-Wise Modeling Framework

To be compatible with variable frequency conditions in diverse real-world EM scenarios, a unified frequency-wise EM modeling framework with improved generalizability and scalability is proposed, as shown in Fig. 2(b).

1) Working Principle: The proposed unified frequency-wise EM modeling framework can be expressed as,

$$\overline{\mathbf{S}^*_n}(f_x) = \mathbf{F}_p(\mathbf{P}_n \cdot \frac{f_x}{c_0}). \quad (14)$$

Here, $n \in \{1, 2, \cdots, N\}$, $x \in \{1, 2, \cdots, n_f\}$. $f_x \in \underbrace{\{f_{min}, \cdots, f_{max}\}}_{n_f \text{ points}}$ is an arbitrary frequency within the defined frequency band. $\overline{\mathbf{S}^*}(f_x)$ denotes the predicted normalized EM response value at f_x, and $\mathbf{S}^*(f_x)$ is the denormalized value. Unlike existing methods that directly model between normalized geometric parameters ($\overline{\mathbf{P}}$) and normalized EM responses for fixed frequency range and points ($\overline{\mathbf{S}}(f_{min}, f_{max}, n_f)$), our frequency-wise modeling framework converts geometric parameters $\mathbf{P}_n$ to electrical dimensions by

$$\mathbf{E}(\mathbf{P}_n, f_x) = \mathbf{P}_n \cdot \frac{f_x}{c_0}. \quad (15)$$

Here, $\mathbf{E}(\mathbf{P}_n, f_x)$ represents the electrical dimensions of $\mathbf{P}_n$ at f_x. f_x uses the unit of Hz, $\mathbf{P}_n$ uses mm, and the speed of light in free space c_0 uses $\mathrm{mm/s}$. As $\mathbf{P}_n \leq \frac{c_0}{f_x}$ in most cases, $\mathbf{E}(\mathbf{P}_n, f_x)$ falls between 0 and 1, requiring no further normalization. $\mathbf{P}_n$ might contain both absolute and relative geometric parameters. Absolute geometric parameters directly determine the physical dimensions of EM structures, while relative geometric parameters define specific physical dimensions as a function of certain absolute geometric parameters. The definition of relative geometric parameters simplifies the adjustment of complex EM structures. The relative geometric parameters in $\mathbf{P}_n$ are converted to absolute geometric parameters before being converted to electrical dimensions. This process ensures the physical meaning of the converted electrical dimensions.

2) Training Stage: In the training process, the model is optimized by minimizing the difference between the predicted and actual normalized EM responses at each frequency f_x for all the training data. Its optimization metric $\mathbf{O}_p$ is expressed as

$$\begin{aligned} \mathbf{O}_p &= \min_{\mathbf{F}_p(\cdot)} \frac{1}{N}\sum_{n=1}^{N}\frac{1}{n_f}\sum_{x=1}^{n_f}|\overline{\mathbf{S}^*_n}(f_x) - \overline{\mathbf{S}_n}(f_x)|^2 \\ &= \min_{\mathbf{F}_p(\cdot)} \frac{1}{N}\sum_{n=1}^{N}\frac{1}{n_f}\sum_{x=1}^{n_f}|\mathbf{F}_p(\mathbf{P}_n \cdot \frac{f_x}{c_0}) - \overline{\mathbf{S}_n}(f_x)|^2. \quad (16) \end{aligned}$$

3) Modeling Stage: During the design and optimization process, the model can predict EM responses for any frequency condition $\mathbf{S}^*_{new}(f_{(u)min}, f_{(u)max}, n_{(u)f})$, which is the denormalized format of $\overline{\mathbf{S}^*_{new}}(f_{(u)min}, f_{(u)max}, n_{(u)f})$ generated by

$$\begin{aligned} \overline{\mathbf{S}^*_{new}}(f_{(u)min}, f_{(u)max}, n_{(u)f}) &= \{\overline{\mathbf{S}^*}(f_{(u)min}), \\ &\quad \cdots, \overline{\mathbf{S}^*}(f_{(u)max})\}, \quad (17) \end{aligned}$$

$$\overline{\mathbf{S}^*}(f_{new}) = \mathbf{F}_p(\mathbf{P}_{new} \cdot \frac{f_{new}}{c_0}), \quad (18)$$

where $f_{new} \in \{f_{(u)min}, \cdots, f_{(u)max}\}$.

The causality and passivity of the predicted EM responses of the proposed surrogate model are maximized through the following six aspects.

1) Data generation. The training data are sampled using Hypercube Sampling to represent the geometric solving space thoroughly and effectively. The EM responses of these data samples are generated using high-fidelity full-wave simulation to ensure data accuracy.
2) Activation function. Suitable activation functions are utilized to confine the predicted values within reasonable ranges. For example, as the EM responses are normalized between 0 and 1 in Implementations A and B, Sigmoid is utilized as the activation function for the output layer to ensure the reasonability of the predictions.
3) Model architecture optimization. The model's architecture is thoroughly optimized using Bayesian optimization to maximize its modeling accuracy, including the causality and passivity for unseen geometric parameters.
4) Model validation. A separate validation dataset is generated, which contains unseen geometric parameters. Along with the training process, the model is validated on a separate validation dataset to avoid overfitting, ensuring the modeling accuracy for unseen geometric parameters.
5) Model testing. A separate testing dataset with unseen geometric parameters is generated. The well-trained model is tested on this testing dataset to assess its accuracy for unseen geometric parameters.
6) The final EM structure design is simulated via high-fidelity full-wave simulation to validate its performance. Fabrication and measurement can be carried out for further validation if needed.

C. Generalizability

The existing methods lack generalizability because they primarily rely on the EM similarity laws to predict EM responses for unseen frequency conditions. Noticeable deviations might occur, because they ignore $\Delta(m)$ caused by the non-linear proportioning characteristics of the impedance and radiation properties. Our proposed modeling framework addresses this limitation by embedding the EM similarity laws and enforcing a robust understanding of the non-linear impedance and radiation proportioning properties, thus enhancing the model's generalizability for unseen frequency conditions.

1) Joint Training: Let us consider two training samples at two distinct frequency points, f_x and f_y, which originated from the same set of geometric parameters ($\mathbf{P}_n$) and the corresponding EM responses. They can be expressed as

$$\overline{\mathbf{S}_n^*}(f_x) = \mathbf{F}_p(\mathbf{P}_n \cdot \frac{f_x}{c_0}), \tag{19}$$

$$\overline{\mathbf{S}_n^*}(f_y) = \mathbf{F}_p(\mathbf{P}_n \cdot \frac{f_y}{c_0}), \tag{20}$$

where $f_x < f_y$, $f_x, f_y \in \{f_{min}, \cdots, f_{max}\}$. The proportion between these two frequency points, denoted as m, is defined by the ratio $m = \frac{f_y}{f_x}$, and m ranges within $(1, \frac{f_{max}}{f_{min}}]$. Accordingly,

$$\overline{\mathbf{S}_n^*}(f_y) = \overline{\mathbf{S}_n^*}(m \cdot f_x), \tag{21}$$

$$\begin{aligned}\mathbf{F}_p(\mathbf{P}_n \cdot \frac{f_y}{c_0}) &= \mathbf{F}_p(\mathbf{P}_n \cdot \frac{m \cdot f_x}{c_0}) \\ &= \mathbf{F}_p(m \cdot \mathbf{P}_n \cdot \frac{f_x}{c_0}).\end{aligned} \tag{22}$$

Based on (11), we have

$$\overline{\mathbf{S}_n^*}(m \cdot f_x) = \overline{\mathbf{S}_n^*}(f_x) \pm \Delta_n(m). \tag{23}$$

Replacing the right side of (21) with (23), we have

$$\overline{\mathbf{S}_n^*}(f_y) = \overline{\mathbf{S}_n^*}(f_x) \pm \Delta_n(m). \tag{24}$$

Replacing the left side component in (20) with (24) and replacing its right side component with (22), the training step in (20) is transformed to

$$\overline{\mathbf{S}_n^*}(f_x) \pm \Delta(m) = \mathbf{F}_p(m \cdot \mathbf{P}_n \cdot \frac{f_x}{c_0}). \tag{25}$$

Considering joint training of (19) and (25), the optimization metric of our approach is improved compared with $\mathbf{O}_e$ of existing methods, which is equivalent to

$$\begin{aligned}\mathbf{O}_p = \min_{\mathbf{F}_p(\cdot)} &\frac{1}{N}\sum_{n=1}^{N}\frac{1}{n_f-1}\sum_{x=1}^{n_f-1}\frac{1}{n_f-x}\sum_{y>x}^{n_f-x}\frac{1}{2} \\ &(|\overline{\mathbf{S}_n^*}(f_x) - \overline{\mathbf{S}_n}(f_x)|^2 + |\overline{\mathbf{S}_n^*}(f_y) - \overline{\mathbf{S}_n}(f_y)|^2) \\ = \min_{\mathbf{F}_p(\cdot)} &\frac{1}{N}\sum_{n=1}^{N}\frac{1}{n_f-1}\sum_{x=1}^{n_f-1}\frac{1}{n_f-x}\sum_{y>x}^{n_f-x}\frac{1}{2} \\ &(|\overline{\mathbf{S}_n^*}(f_x) - \overline{\mathbf{S}_n}(f_x)|^2 \\ &+|\overline{\mathbf{S}_n^*}(f_x) \pm \Delta(m) - \overline{\mathbf{S}_n}(f_y)|^2) \\ = \min_{\mathbf{F}_p(\cdot)} &\frac{1}{N}\sum_{n=1}^{N}\frac{1}{n_f-1}\sum_{x=1}^{n_f-1}\frac{1}{n_f-x}\sum_{y>x}^{n_f-x}\frac{1}{2} \\ &(|\mathbf{F}_p(\mathbf{P}_n \cdot \frac{f_x}{c_0}) - \overline{\mathbf{S}_n}(f_x)|^2 \\ &+|\mathbf{F}_p(m \cdot \mathbf{P}_n \cdot \frac{f_x}{c_0}) - \overline{\mathbf{S}_n}(f_y)|^2).\end{aligned} \tag{26}$$

As expressed in (26), the optimization metric of our approach integrates the propagation of $\pm\Delta(m) \leftarrow m$ ($m \in (1, \frac{f_{max}}{f_{min}}]$). The integration of $\pm\Delta(m) \leftarrow m$ ($m \in (1, \frac{f_{max}}{f_{min}}]$) enforces the model to better understand the EM similarity laws and non-linear proportioning characteristics caused by the impedance and radiation properties, hence improving the extrapolation modeling accuracy for unseen frequency ranges.

2) Improved Generalizability: In the design and optimization stage, the model can predict EM responses for variable frequency conditions (from $f_{(u)min}$ to $f_{(u)max}$ with n_{uf} points) by (17) and (18). Instead of a rough approximation through the EM similarity laws, $\overline{\mathbf{S}^*}(f_{new})$ in (18) is directly predicted by the model with a robust understanding of the EM similarity laws and non-linear proportioning characteristics. Due to its enhanced understanding of the non-linear proportioning characteristics, the modeling error $L_{p(u)}$ for variable frequency conditions is expressed as

$$\begin{aligned}L_{p(u)} = &\frac{1}{n_{(u)f}}|\overline{\mathbf{S}_{new}^*}(f_{(u)min}, f_{(u)max}, n_{(u)f}) \\ &-\overline{\mathbf{S}_{new}}(f_{(u)min}, f_{(u)max}, n_{(u)f})|^2 \\ = &\frac{1}{n_{(u)f}}\sum_{f_{new}}^{n_{(u)f}}|\overline{\mathbf{S}_{new}^*}(f_{new}) - \overline{\mathbf{S}_{new}}(f_{new})|^2.\end{aligned} \tag{27}$$

Compared with L_{eu} of existing methods in (9), the modeling error for unseen frequency conditions L_{pu} is significantly reduced, thereby greatly improving generalizability for variable frequency conditions. Note that formula (27) does not necessarily imply that the model error for unseen frequency conditions L_{pu} is reduced to a level close to L_p for the original frequency condition. In general, L_{pu} keeps greater than L_p, because m for unseen frequency conditions probably exceeds its original range $(1, \frac{f_{max}}{f_{min}}]$ that the model is trained on. Nevertheless, (27) indicates that the proposed method greatly enhances the modeling accuracy for unseen frequency conditions.

D. Scalability

Real-world applications now encompass increasingly diverse frequency conditions, necessitating enhanced scalability of the modeling framework. The existing methods require a distinct surrogate model for the specific frequency range and points of interest. Once trained on data from a certain frequency condition ranging from $f_{\alpha min}$ to $f_{\alpha max}$ with $n_{\alpha f}$ points, the surrogate model is not scalable to incorporate data from a new frequency range ranging from $f_{\beta min}$ to $f_{\beta max}$ with $n_{\beta f}$ points, when $f_{\alpha min} \neq f_{\beta min}$, $f_{\alpha max} \neq f_{\beta max}$, and $n_{\alpha f} \neq n_{\beta f}$. Some works sample multiple frequency ranges and include frequency as one of the inputs, achieving interpolation modeling along the frequency domain. However, they require complete EM responses of multiple frequency ranges for each data sample, which could be difficult to obtain in real-world scenarios. The frequency conditions are diverse in real-world scenarios. Different data samples could have different frequency ranges, and some data samples might lack EM responses at certain frequency ranges.

The unified frequency-wise modeling framework resolves this problem by enabling the integration of variable frequency conditions into a single model,

$$\begin{aligned} \overline{\mathbf{S}^*}(f_{\alpha,\cdots,\beta}) &= \mathbf{F}_p(\mathbf{P}_{\alpha,\cdots,\beta} \cdot \frac{f_{\alpha,\cdots,\beta}}{c_0}), \\ f_{\alpha x} &\in \underbrace{\{f_{\alpha min}, \cdots, f_{\alpha max}\}}_{n_{\alpha f} \text{ points}}, \\ f_{\beta x} &\in \underbrace{\{f_{\beta min}, \cdots, f_{\beta max}\}}_{n_{\beta f} \text{ points}}. \end{aligned} \tag{28}$$

Here, "$\alpha, \cdots, \beta$" represent indexes for variable frequency conditions, from which α and β index two arbitrary ones. Assuming that $f_{\alpha min} < f_{\alpha max} < f_{\beta min} < f_{\beta max}$, the sampling region of the proportioning ratio is expanded from $(1, \frac{f_{\alpha max}}{f_{\alpha min}}]$ to $(1, \frac{f_{\beta max}}{f_{\alpha min}}]$ by $\frac{f_{\beta max} - f_{\alpha max}}{f_{\alpha min}}$. With additional frequency conditions involved, the sampling region of the proportioning ratio m continuously grows. By sampling m from a progressively expanded region, the surrogate model achieves a more robust understanding of the EM similarity laws and imperfect proportioning properties, showing improved scalability.

III. Implementation A: Meander-Line Polarizer

A. Meander-Line Polarizer

Implementation A involves modeling the co-polarization transmission coefficient $|S_{21}|$ of a meander-line polarizer. Its unit cell is composed of two crossed meander microstrip lines that are etched on a dielectric substrate with a thickness of $h_0 = 0.254\,\text{mm}$ and relative permittivity of $\varepsilon_r = 2.2$, as illustrated in Fig. 3. The microstrip lines of every two adjacent unit cells along the x direction connect and form a closed loop. This polarizer converts an incident wave, linearly polarized at 45° along the z direction, into a circularly-polarized signal. The co-polarization transmission coefficient $|S_{21}|$ indicates its polarization conversion performance. $|S_{21}|$ is mainly controlled by five geometric parameters, L, W, a, d, and s. Here, a is a relative parameter that denotes the ratio of folded length along the x axis, as shown in Fig. 3. The definition of a prevents the folded length from exceeding half the width of the unit cell $0.5W$ during sampling.

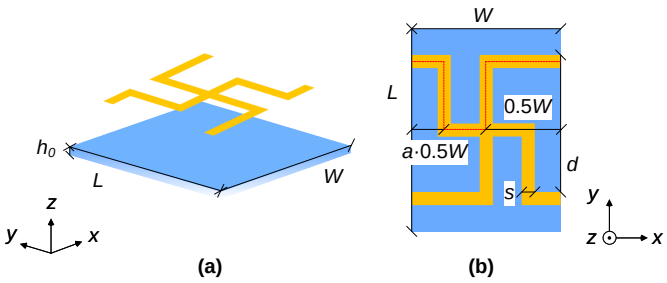


Fig. 3. Unit cell of the meander-line polarizer in Implementation A. (a) 3D view. (b) Top view.

TABLE I
Two Different Frequency Conditions and Corresponding Geometric Parameters in Implementation A: Meander-Line Polarizer

		Condition	
		$A1$	$A2$
Frequency	Range (GHz)	$[10, 25]$	$[20, 40]$
	Points	61	81
Geometric Parameters (Unit: mm; *Unit: 1)	L	$[3.6, 4.7]$	$[1.8, 2.4]$
	W	$[1.7, 2.6]$	$[0.85, 1.55]$
	$^*a$	$[0.3, 0.45]$	$[0.3, 0.45]$
	d	$[1.5, 2.4]$	$[0.7, 1.3]$
	s	$[0.2, 0.45]$	$[0.05, 0.2]$

B. Data Preparation

Two frequency conditions are defined: $A1$ from 10 GHz to 25 GHz at an interval of 0.25 GHz with 61 points; $A2$ from 20 GHz to 40 GHz at an interval of 0.25 GHz with 81 points. As shown in Table I, a distinct parameter range is assigned for the five geometric parameters in each condition based on the design experience. Under each condition, 100 combinations of geometric parameters are sampled using Latin Hypercube Sampling, and the corresponding 100 $|S_{21}|$ are generated through full-wave simulation. The respective sizes of $|S_{21}|$ in $A1$ and $A2$ are 61 and 81, respectively. In each condition, the collected 100 pairs of geometric parameters and $|S_{21}|$ are arbitrarily divided into training, validation, and testing datasets in the ratio of $2:1:2$ under a random seed: D^{A1}_{train}, D^{A1}_{val}, and D^{A1}_{test} in $A1$; D^{A2}_{train}, D^{A2}_{val}, and D^{A2}_{test} in $A2$, respectively. These datasets undergo different modifications in the existing and proposed methods. Thus, each dataset has a different size in the existing and proposed methods, but its encapsulated information remains unchanged.

C. Existing Methods

For comparison, we apply four existing methods, Gaussian process regression (A^g_1 and A^g_2), kriging (A^k_1 and A^k_2), support vector regression (A^s_1 and A^s_2), and neural networks (A^n_1 and A^n_2), for modeling the meander-line polarizer in the two conditions, $A1$ and $A2$, respectively, as listed in Table II. It is difficult for the existing methods to train an

integrated model incorporating $A1$ and $A2$ data samples, because each condition's data samples have distinct frequency ranges and the existing methods require identical frequency ranges. Each experiment is carried out in three independent runs under three random seeds. In each run, the source data are arbitrarily divided into datasets under a distinct random seed, and the experiments are conducted accordingly. During these experiments, the datasets are normalized between 0 and 1. The geometric parameters are normalized by subtracting the minimum values and dividing by the range between the maximum and minimum values. For example, L is normalized through $\frac{L-\min(L)}{\max(L)-\min(L)} = \frac{L-3.6\,\text{mm}}{4.7\,\text{mm}-3.6\,\text{mm}}$ in $A_1^{g,k,s,n}$. Here, the relative geometric parameter a and other absolute geometric parameters are normalized using this min-max scaling to fall within the 0 to 1 range. We convert the $|S_{21}|$ values from decibel (dB) format to linear format (values between 0 and 1). Under each condition, the 100 samples are separated arbitrarily to form training, validation, and testing datasets with sizes of 40, 20, and 40, respectively. Each sample has a fixed input size of 5. The output size is 61 in $A1$ and 81 in $A2$. During the training process in $A_1^{g,k,s,n}$, D_{train}^{A1} is utilized for training a surrogate model, and D_{val}^{A1} is used for validation. In $A_2^{g,k,s,n}$, D_{train}^{A2} and D_{val}^{A2} are used for training and validation, respectively. We use the $Adam$ optimizer and the mean squared error (MSE) between actual and predicted $|S_{21}|$ linear values as the loss function.

Gaussian process regression employs a combination of constant kernel and radial basis functions as its kernel function. Support vector regression uses a linear kernel. Neural networks use Sigmoid as the activation function for the output layer and ReLU for the other layers. Each neural network model's hyperparameters are optimized using Bayesian optimization, including the number of epochs, the learning rate, the batch size, the number of hidden layers, and the number of neurons in each hidden layer. The number of training epochs is determined by a quantized uniform distribution ranging from 200 to 1000, with intervals of 200. The learning rate is selected from a discrete set of values, $[0.0001,\ 0.001]$. The batch size is sampled from a quantized uniform distribution between 1 and 10, with increments of 1. The number of hidden layers is sampled from a quantized uniform distribution between 2 and 8. The number of neurons per hidden layer is chosen from the set $[16,\ 32,\ 64,\ 128,\ 256]$. The optimization process stops either after 50 consecutive iterations with no improvement or when it reaches 200 iterations. The validation loss from D_{val}^{A1} serves as the assessment metric in A_1^n, and D_{val}^{A2} in A_2^n, respectively. The optimized hyperparameters for the two neural network models are outlined in Table II. The number of gradient steps determined by the number of epochs and batch size is also shown in Table II.

Each of the eight well-trained models is tested on both D_{test}^{A1} and D_{test}^{A2} to assess its accuracy for the original and unseen frequency conditions. The test results are listed in Table II. Note that each loss value is the average result of three independent runs. To test a model's interpolation accuracy in its original frequency condition, testing on D_{test}^{A1} in $A_1^{g,k,s,n}$ or testing D_{test}^{A2} in $A_2^{g,k,s,n}$, respectively, we directly input the geometric parameters to predict the $|S_{21}|$ values. As for testing the extrapolation or extension ability in unseen frequency conditions, testing on D_{test}^{A2} in $A_1^{g,k,s,n}$ or testing on D_{test}^{A1} in $A_2^{g,k,s,n}$, respectively, the surrogate model is combined with the EM similarity laws for prediction. The EM similarity laws reveal that an EM structure with $\frac{1}{m}$ times proportioned geometric size maintains similar performance at m times higher frequency. It may require multiple proportioning processes to cover a wide target frequency range.

To describe the testing procedure for the original and unseen frequency conditions, $\mathbf{F}_{Ai}^{g,k,s,n}(\cdot)$ represents the model trained in $A_i^{g,k,s,n}$. $\mathbf{P}_j$ represents the geometric parameters in D_{test}^{Aj}, and $\overline{\mathbf{P}}_j$ denotes the normalized parameters. $\overline{\mathbf{S}_j^*}(f_{(u)min}, f_{(u)max}, n_{(u)f})$ denotes the predicted normalized $|S_{21}|$ from $f_{(u)min}$ to $f_{(u)max}$ with $n_{(u)f}$ points. $\overline{\mathbf{S}}_j(f_{(u)min}, f_{(u)max}, n_{(u)f})$ is the actual normalized $|S_{21}|$. L_{Aij} is the test loss of $\mathbf{F}_{Ai}^{g,k,s,n}(\cdot)$ on D_{test}^{Aj}, where $i, j \in \{1, 2\}$. The detailed testing procedure is introduced below.

(a) L_{A11} of $\mathbf{F}_{A1}^{g,k,s,n}(\cdot)$ on D_{test}^{A1} in $A_1^{g,k,s,n}$. $\overline{\mathbf{S}_1^*}(10, 25, 61)$ is predicted by feeding $\overline{\mathbf{P}_1}$ into $\mathbf{F}_{A1}^{g,k,s,n}(\cdot)$ to calculate the MSE between $\overline{\mathbf{S}_1}(10, 25, 61)$,

$$L_{A11} = \frac{1}{61}|\overline{\mathbf{S}_1}(10, 25, 61) - \overline{\mathbf{S}_1^*}(10, 25, 61)|^2, \quad (29)$$

where

$$\overline{\mathbf{S}_1^*}(10, 25, 61) = \mathbf{F}_{A1}^{g,k,s,n}(\overline{\mathbf{P}_1}). \quad (30)$$

(b) L_{A12} of $\mathbf{F}_{A1}^{g,k,s,n}(\cdot)$ on D_{test}^{A2} in $A_1^{g,k,s,n}$. $\mathbf{P}_2$ is increased by $m = \frac{f_{min}^{A2}}{f_{min}^{A1}} = \frac{20}{10}$ times and normalized with respect to the parameter range in $A1$. The normalized $\overline{\frac{20}{10} \cdot \mathbf{P}_2}$ is input into $\mathbf{F}_{A1}^{g,k,s,n}(\cdot)$ to predict $\overline{\mathbf{S}_2^*}(10 \times \frac{20}{10} = 20, 25 \times \frac{20}{10} = 50, 61)$,

$$\begin{aligned}\overline{\mathbf{S}_2^*}(20, 50, 61) &= \overline{\mathbf{S}_2^*}(10 \times \frac{20}{10}, 25 \times \frac{20}{10}, 61), \\ &= \mathbf{F}_{A1}^{g,k,s,n}(\overline{\frac{20}{10} \cdot \mathbf{P}_2}). \end{aligned} \quad (31)$$

The last 20 $|S_{21}|$ values corresponding to frequencies above 40 GHz are cut off to fit the target frequency range in the testing condition $A2$ from 20 GHz to 40 GHz,

$$\overline{\mathbf{S}_2^*}(20, 40, 41) = \overline{\mathbf{S}_2^*}(20, 50, 61)[:, :-20]. \quad (32)$$

$\overline{\mathbf{S}_2}(20, 40, 41)$ is collected via simulation as label,

$$\overline{\mathbf{S}_2}(20, 40, 41) \leftarrow \text{Simulate}(\mathbf{P}_2). \quad (33)$$

L_{A12} equals the MSE between $\overline{\mathbf{S}_2}(20, 40, 41)$ and $\overline{\mathbf{S}_2^*}(20, 40, 41)$,

$$L_{A12} = \frac{1}{41}|\overline{\mathbf{S}_2}(20, 40, 41) - \overline{\mathbf{S}_2^*}(20, 40, 41)|^2. \quad (34)$$

(c) L_{A21} of $\mathbf{F}_{A2}^{g,k,s,n}(\cdot)$ on D_{test}^{A1} in $A_2^{g,k,s,n}$. Two proportioning processes are conducted to cover the target frequency range in the testing condition $A1$ from 10 GHz to 25 GHz. Specifically, we input $\overline{\frac{10}{20} \cdot \mathbf{P}_1}$ and $\overline{\frac{25}{40} \cdot \mathbf{P}_1}$ into

$\mathbf{F}_{A2}^{g,k,s,n}(\cdot)$ to predict $\overline{\mathbf{S}_1^*}(20\times\frac{10}{20}=10,40\times\frac{10}{20}=20,81)$ and $\overline{\mathbf{S}_1^*}(20\times\frac{25}{40}=12.5,40\times\frac{25}{40}=25,81)$, respectively,

$$\begin{aligned}\overline{\mathbf{S}_1^*}(10,20,81) &= \overline{\mathbf{S}_1^*}(20\times\frac{10}{20},40\times\frac{10}{20},81)\\ &= \mathbf{F}_{A2}^{g,k,s,n}(\overline{\frac{10}{20}\cdot\mathbf{P}_1}),\end{aligned}\tag{35}$$

$$\begin{aligned}\overline{\mathbf{S}_1^*}(12.5,25,81) &= \overline{\mathbf{S}_1^*}(20\times\frac{25}{40},40\times\frac{25}{40},81)\\ &= \mathbf{F}_{A2}^{g,k,s,n}(\overline{\frac{25}{40}\cdot\mathbf{P}_1}).\end{aligned}\tag{36}$$

The last 32 points of $\overline{\mathbf{S}_1^*}(12.5,25,81)$ from 20 GHz to 25 GHz are extracted, excluding the point at 20 GHz. The extracted values are concatenated with $\overline{\mathbf{S}_1^*}(10,20,81)$ to generate $\overline{\mathbf{S}_1^*}(10,25,113)$,

$$\overline{\mathbf{S}_1^*}(20,25,32) = \overline{\mathbf{S}_1^*}(12.5,25,81)[:,-32:],\tag{37}$$

$$\overline{\mathbf{S}_1^*}(10,25,113) = \{\overline{\mathbf{S}_1^*}(10,20,81),\overline{\mathbf{S}_1^*}(20,25,32)\}.\tag{38}$$

We collect the label $\overline{\mathbf{S}_1}(10,25,113)$ through full-wave simulation,

$$\overline{\mathbf{S}_1}(10,25,113) \leftarrow \text{Simulate}(\mathbf{P}_1).\tag{39}$$

L_{A21} equals the MSE between $\overline{\mathbf{S}_1}(10,25,113)$ and $\overline{\mathbf{S}_1^*}(10,25,113)$,

$$L_{A21} = \frac{1}{113}|\overline{\mathbf{S}_1}(10,25,113)-\overline{\mathbf{S}_1^*}(10,25,113)|^2.\tag{40}$$

(d) L_{A22} of $\mathbf{F}_{A2}^{g,k,s,n}(\cdot)$ on D_{test}^{A2} in $A_2^{g,k,s,n}$,

$$L_{A22} = \frac{1}{81}|\overline{\mathbf{S}_2}(20,40,81)-\overline{\mathbf{S}_2^*}(20,40,81)|^2,\tag{41}$$

where

$$\overline{\mathbf{S}_2^*}(20,40,81) = \mathbf{F}_{A2}^{g,k,s,n}(\overline{\mathbf{P}_2}).\tag{42}$$

Although we can generate $|S_{21}|$ within the unseen frequency ranges through the EM similarity laws, it suffers from low flexibility, high complexity, and reduced accuracy. The available frequency points are constrained to be proportional to those in the training condition. Multiple proportioning procedures are needed to make up a wide target frequency range. The surrogate model has limited performance within the unseen frequency ranges, as shown in Table II, because it lacks robustness against the non-linear characteristics of the proportioning impedance and radiation properties.

D. Proposed Method

The proposed method is validated through multiple experiments. Each experiment is repeated in three independent runs under the three random seeds, which are the same as those used in the existing methods. We reformat and convert the source datasets to be frequency-wise. The geometric parameters $\mathbf{P}_n$ are transformed into the electrical dimensions $\mathbf{E}_n(\mathbf{P}_n,f_x)$ using (15). The relative geometric parameter a in $\mathbf{P}_n$ is converted to an absolute geometric parameter by multiplying it by $0.5W$ before calculating its electrical dimension. The geometric parameters are within a wavelength of the operating frequency. Thus, the converted electrical dimensions approximately range between 0 and 1. Each $\mathbf{E}_n(\mathbf{P}_n,f_x)$ is paired with the corresponding $|S_{21}|_n$ at f_x to form a new frequency-wise sample. In A_1^p, as $|S_{21}|_n$ contains 61 frequency points in $A1$, each pair of $\mathbf{P}_n$ and $|S_{21}|_n$ is converted into 61 pairs of frequency-wise samples. The reformatted size of the generated frequency-wise training, validation, and testing dataset in $A1$ is $40\times61=2440$, $20\times61=1220$, and $40\times61=2440$, respectively. In A_2^p, each pair of $\mathbf{P}_n$ and $|S_{21}|_n$ is converted into 81 pairs of samples, forming training, validation, and testing datasets of sizes 3240, 1620, and 3240, respectively. For both A_1^p and A_2^p, each sample has a fixed input size of 5 and output size of 1.

1) Generalizability: To compare with the existing methods, the proposed method is separately conducted in each condition $A1$ or $A2$ for validating generalizability, referred to as experiments A_1^p and A_2^p, respectively, as shown in Table II. A_1^p is compared with $A_1^{g,k,s,n}$ to validate the generalizability of the proposed method, including the interpolation and extrapolation ability. Correspondingly, A_2^p is compared with $A_2^{g,k,s,n}$ for validation. In A_1^p, the model is solely trained on $A1$ using the specified training dataset D_{train}^{A1}, and it is validated on D_{val}^{A1}. In A_2^p, the model is solely trained and validated on $A2$ using D_{train}^{A2} and D_{val}^{A2}, respectively. We optimize the hyperparameters of these models using Bayesian optimization. For a fair comparison, the learning rate, the number of hidden layers, and the number of neurons in every hidden layer are sampled from the same space, and the activation function remains unchanged as those for neural networks in Section III-C. In response to the increase in dataset size by the number of frequency points 61 or 81, the sampling range of the number of epochs was modified, ranging from 500 to 2000, with intervals of 500; the optimization space for the batch size is proportionally extended, sampling from $1\times n_f$ to $10\times n_f$ with increments of n_f, where $n_f=61$ in A_1^p and $n_f=81$ in A_2^p, respectively. The validation loss on D_{val}^{A1} is taken as the optimization metric in A_1^p, and D_{val}^{A2} in A_2^p, respectively. The optimization process is designed to end after 50 iterations if no improvement occurs or at the 200th iteration. The optimized hyperparameters are given in Table II.

After training, each model is tested on both D_{test}^{A1} and D_{test}^{A2} to assess its accuracy for the original and unseen frequency conditions. For the model that trained on D_{train}^{A1} in A_1^p, testing on D_{test}^{A1} measures its interpolation accuracy for its original frequency condition $A1$, and testing on D_{test}^{A2} evaluates its extrapolation accuracy for an unseen frequency condition $A2$. The testing results are compared with existing methods in Table II.

Compared with neural networks that reach the best modeling performance, the proposed model trained in A_1^p enhances the modeling accuracy on D_{test}^{A2} for the unseen frequency range by 10.6 % ($\frac{|4.65-5.20|}{5.20}\times100\,\%=10.6\,\%$). Although its modeling error on D_{test}^{A1} for the original frequency condition increases, it remains a small value and is significantly lower than the modeling error on D_{test}^{A2}. The model trained in A_2^p improves the modeling accuracy on D_{test}^{A1} for the unseen

TABLE II
COMPARATIVE RESULTS IN IMPLEMENTATION A: MEANDER-LINE POLARIZER

Exp.	Trained on	Model	Optimized Hyperparameters						Test Loss ($\times 10^{-4}$) on	
			N_e	N_g	lr	N_b	N_h	$[N_{n1}, N_{n2}, \cdots]$	D^{A1}_{test}	D^{A2}_{test}
A^g_1	D^{A1}_{train}	**GPR**	–						0.34	5.40
A^k_1		**Kri.**	–						0.23	9.43
A^s_1		**SVR**	–						19.17	31.46
A^n_1		**NN**	1000	6000	0.0001	7	3	$[256, 256, 32]$	0.34	5.20
A^p_1		**Pro.**	1500	60000	0.001	61	6	$[256, 128, 16, 32, 32, 256]$	0.56	4.65
A^g_2	D^{A2}_{train}	**GPR**	–						175.93	0.68
A^k_2		**Kri.**	–						14.81	0.53
A^s_2		**SVR**	–						31.67	19.84
A^n_2		**NN**	800	8000	0.001	4	5	$[32, 256, 32, 16, 16]$	7.92	0.76
A^p_2		**Pro.**	1000	40000	0.001	81	2	$[64, 128]$	5.72	0.34
A^p_{12}	D^{A1}_{train} & D^{A2}_{train}	**Pro.**	1500	43500	0.001	200	2	$[256, 32]$	2.04	1.60

Note: **Exp.** refers to the experiment index; **GPR** refers to Gaussian process regression; **Kri.** refers to kriging; **SVR** refers to support vector regression; **NN** refers to neural networks; **Pro.** refers to the proposed method; N_e refers to the number of epochs; N_g refers to the number of gradient steps; lr refers to the learning rate; N_b refers to the batch size; N_h refers to the number of hidden layers; $[N_{n1}, N_{n2}, \cdots]$ refers to the number of neurons in each hidden layer.

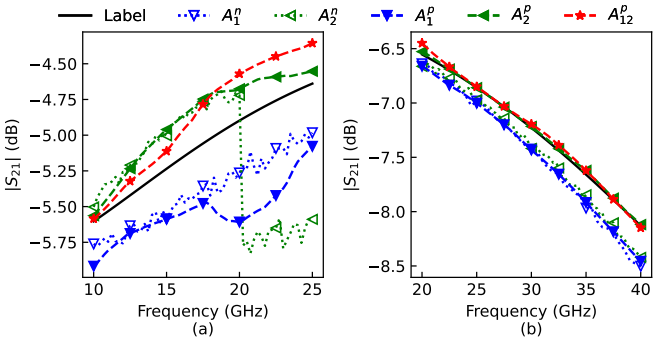


Fig. 4. Comparison of $|S_{21}|$ curves in Implementation A. (a) A test sample in D^{A1}_{test} for A^n_1, A^n_2, A^p_1, A^p_2, and A^p_{12}. (b) A test sample in D^{A2}_{test} for A^n_1, A^n_2, A^p_1, A^p_2, and A^p_{12}. ("Label" denotes the label $|S_{21}|$ generated from full-wave simulation.)

frequency condition by $27.8\,\%$ ($\frac{|5.72-7.92|}{7.92}\times 100\,\% = 27.8\,\%$), while maintaining its performance on D^{A2}_{test} in the original condition. The results in A^p_1 and A^p_2 demonstrate the enhancement of extrapolated modeling performance compared with the existing methods, indicating the improved generalizability of the proposed method.

2) Scalability: A surrogate model is jointly trained using both D^{A1}_{train} and D^{A2}_{train} to validate its scalability, referred to as an experiment A^p_{12} in Table II. The model is optimized using Bayesian optimization under the same sampling space as in Section III-D1, except that the batch size is chosen from a discrete set of values, [100, 200, 500, 1000, 1500, 2000]. Table II shows the optimized model and testing results.

The model in A^p_{12} can be considered as scaling up the model in A^p_1 by incorporating new data D^{A2}_{train} in a new condition $A2$, or scaling up A^p_2 by adding D^{A1}_{train}, respectively. It greatly improves and balances the modeling accuracy across two different frequency conditions $A1$ and $A2$. Although the modeling error in the original frequency condition slightly increases, it is still significantly lower than the maximum modeling error before scaling. To better visualize the improvement, Fig. 4 compares the $|S_{21}|$ curves generated by neural networks (A^n_1 and A^n_2) and the proposed method (A^p_1, A^p_2, and A^p_{12}). Simulation results generated from full-wave simulators are used as the ground truth, represented as black curves and denoted as "Label". Seeing A^n_1 and A^p_1 denoted as blue curves with hollow and solid down-triangle-shaped marks, respectively, the proposed method shows better extrapolation performance along out-of-range frequency $A2$ than neural networks, while the modeling accuracy in $A1$ slightly deteriorates. The proposed method reaches the best modeling performance in A^p_{12}, which is denoted as red curves with star-shaped marks, matching well with full-wave simulation results. The comparative results demonstrate the improved scalability of the proposed method to incorporate variable frequency conditions.

IV. IMPLEMENTATION B: PLANAR METASURFACE LENS MODELING

We further validate the improved generalizability and scalability of the proposed method by implementing it on a planar metasurface lens, when the solving dimensionality increases from five to ten and three different frequency conditions are considered.

A. Planar Metasurface Lens

Implementation B aims to model the reflection coefficient $|S_{11}|$ of a planar metasurface lens for millimeter-wave MIMO applications presented in [38]. Its unit cell structure is illustrated in Fig. 5. It consists of two back-to-back $h_0 = 0.254\,\mathrm{mm}$ thick $Rogers\ RT5880$ substrate layers with relative permittivity of $\varepsilon_r = 2.2$. These two substrate layers are separated by a h thick air gap layer. Two identical curved Jerusalem crosses are etched on the outer layers of the two substrate layers, which are depicted in Fig. 5. 10 geometric parameters, r_1, r_2, w, w_1, w_2, c, g_1, g, p, and h, are adjusted to

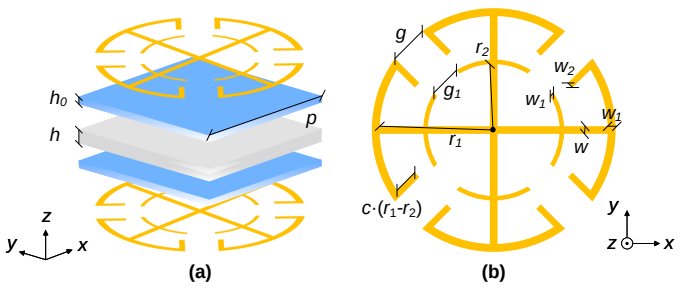


Fig. 5. Unit cell of the planar metasurface lens in Implementation B. (a) 3D view. (b) Top and bottom metal layers.

modify the $|S_{11}|$. c is a relative geometric parameter associated with the folded length of the outer arm. Including c as a design parameter prevents the outer and inner arms from overlapping during the exploration of the solution space.

B. Data Preparation

We define three conditions ($B1$, $B2$, and $B3$) with different frequency ranges and points, as listed in Table III. For example, in $B1$, the planar metasurface lens operates from 15 GHz to 35 GHz, and the modeling focuses on $|S_{11}|$ at $n_f = 81$ frequency points with an interval of 0.25 GHz. Each frequency condition is associated with a distinct adjustable space for the geometric parameters. For each condition, we collect 150 pairs of geometric parameters and $|S_{11}|$ to form training, validation, and testing datasets. The 150 combinations of geometric parameters are defined using Latin Hypercube Sampling within the specific adjustable space. Note that p and h are set as constant values in each condition during sampling, but they are included as the input for modeling across different conditions. Correspondingly, 150 $|S_{11}|$ vectors of size n_f within respective frequency ranges are generated using full-wave simulation. In a ratio of $7:3:5$, we form training, validation, and testing datasets in each condition under a random seed: D^{B1}_{train}, D^{B1}_{val}, and D^{B1}_{test} for $B1$; D^{B2}_{train}, D^{B2}_{val}, and D^{B2}_{test} for $B2$; D^{B3}_{train}, D^{B3}_{val}, and D^{B3}_{test} for $B3$, respectively. The division ratio for datasets differs slightly from that of Implementation A: Meander-Line Polarizer, because the total amount of data is different. The ratio between training and validation remains similar $\frac{2}{1} \approx \frac{7}{3}$, and a relatively large amount of testing data is arranged for consistency.

C. Existing Methods

Gaussian process regression (B^g_1, B^g_2, and B^g_3), kriging (B^k_1, B^k_2, and B^k_3), support vector regression (B^s_1, B^s_2, and B^s_3), and neural networks (B^n_1, B^n_2, and B^n_3) are applied for modeling in the three frequency conditions, referred to as twelve experiments in Table IV. The existing methods require identical frequency ranges, hence refraining an integrated model from being trained with all the $B1$, $B2$, and $B3$ data samples. Each experiment is repeated in three independent runs under three random seeds, respectively, to yield an average result for consistency. We normalize the geometric parameters by subtracting the minimum values and dividing by the range between the maximum and minimum values. For example, r_1 is normalized by $\frac{r_1 - \min(r_1)}{\max(r_1) - \min(r_1)} = \frac{r_1 - 2.1\,\text{mm}}{2.7\,\text{mm} - 2.1\,\text{mm}}$ in $B^{g,k,s,n}_1$. $|S_{11}|$ values are converted from decibel (dB) format to linear format between 0 and 1. In each condition, by separating the 150 samples arbitrarily under a random seed, we form training, validation, and testing datasets of size 70, 30, and 50. In each condition $B1$, $B2$, or $B3$, each sample has an input size of 10 and an output size of 81, 101, or 121, respectively. During the training process in $B^{g,k,s,n}_1$, D^{B1}_{train} is utilized to train a surrogate model, while D^{B1}_{val} is used for validation. Accordingly, $B^{g,k,s,n}_2$ and $B^{g,k,s,n}_3$ use their corresponding datasets, respectively. The loss function is defined as the mean squared error (MSE) between actual and predicted $|S_{11}|$ linear values.

TABLE III
THREE DIFFERENT FREQUENCY CONDITIONS AND CORRESPONDING GEOMETRIC PARAMETERS IN IMPLEMENTATION B: PLANAR METASURFACE LENS

		Condition		
		$B1$	$B2$	$B3$
Frequency	Range (GHz)	$[15, 35]$	$[20, 45]$	$[30, 60]$
	Points	81	101	121
Geometric Parameters (Unit: mm; *Unit: 1)	r_1	$[2.1, 2.7]$	$[1.35, 1.75]$	$[0.9, 1.2]$
	r_2	$[1.1, 1.7]$	$[0.5, 0.9]$	$[0.45, 0.75]$
	w	$[0.1, 0.35]$	$[0.05, 0.2]$	$[0.05, 0.15]$
	w_1	$[0.1, 0.35]$	$[0.05, 0.2]$	$[0.05, 0.15]$
	w_2	$[0.1, 0.35]$	$[0.05, 0.2]$	$[0.05, 0.15]$
	$^*c$	$[0.4, 0.8]$	$[0.4, 0.8]$	$[0.4, 0.8]$
	g_1	$[0.4, 0.8]$	$[0.2, 0.6]$	$[0.2, 0.4]$
	g	$[0.6, 1.2]$	$[0.4, 0.8]$	$[0.2, 0.5]$
	p	6.2	4	2.8
	h	1.0	0.8	0.6

The kernel and activation functions for these existing methods are the same as those used in Implementation A: Meander-Line Polarizer. Each neural network's architecture is thoroughly optimized through Bayesian optimization. The key hyperparameters under exploration include the number of training epochs, the learning rate, the batch size, the number of hidden layers, and the number of neurons. Specifically, the number of training epochs is sampled from a quantized uniform distribution from 200 to 1000, with intervals of 200. The learning rate is optimized using a discrete set of values, $[0.0001,\ 0.001]$. The batch size is sampled from a quantized uniform distribution between 1 and 10, with increments of 1. The number of hidden layers is sampled from a quantized uniform distribution between 2 and 8. The number of neurons in each hidden layer is chosen from the set $[16,\ 32,\ 64,\ 128,\ 256]$. *Adam* is taken as the optimizer. The optimization is terminated after 50 iterations without improvement or upon reaching the maximum number of iterations, 300. The validation loss on the respective validation dataset serves as the assessment metric in B^n_1, B^n_2, or B^n_3, respectively. The optimized architectures of the three neural networks are listed in Table IV.

Once trained, each of the twelve models is tested on D^{B1}_{test}, D^{B2}_{test}, and D^{B3}_{test} to evaluate its accuracy and generalizability. The average test losses are listed in Table IV. For each model, testing the interpolation ability in its original frequency condition is straightforward, for example, testing the model in

$B_1^{g,k,s,n}$ on D_{test}^{B1}. We test the model's extrapolation ability under its unseen frequency conditions by combining the EM similarity laws, for example, testing the model in $B_1^{g,k,s,n}$ on D_{test}^{B2} and D_{test}^{B3}. For clarity, we define the model in $B_i^{g,k,s,n}$ as $\mathbf{F}_{Bi}^{g,k,s,n}(\cdot)$, the denormalized geometric parameters in Bj as $\mathbf{P}_j$, the normalized parameters as $\overline{\mathbf{P}}_j$, the label normalized $|S_{11}|$ from $f_{(u)min}$ to $f_{(u)max}$ with $n_{(u)f}$ points as $\overline{\mathbf{S}}_j(f_{(u)min}, f_{(u)max}, n_{(u)f})$, the predicted normalized $|S_{11}|$ as $\overline{\mathbf{S}_j^*}(f_{(u)min}, f_{(u)max}, n_{(u)f})$, and the test loss of $\mathbf{F}_{Bi}^{g,k,s,n}(\cdot)$ on D_{test}^{Bj} as L_{Bij}. Here, $i, j \in \{1, 2, 3\}$ denotes the index related to three conditions, $B1$, $B2$, and $B3$. We test each model under the following guidelines.

(a) L_{B11} of $\mathbf{F}_{B1}^{g,k,s,n}(\cdot)$ on D_{test}^{B1} in $B_1^{g,k,s,n}$. We directly input $\overline{\mathbf{P}_1}$ into $\mathbf{F}_{B1}^{g,k,s,n}(\cdot)$ to predict $\overline{\mathbf{S}_1^*}(15, 35, 81)$ and calculate the MSE between $\overline{\mathbf{S}_1}(15, 35, 81)$ and $\overline{\mathbf{S}_1^*}(15, 35, 81)$,

$$L_{B11} = \frac{1}{81}|\overline{\mathbf{S}_1}(15, 35, 81) - \overline{\mathbf{S}_1^*}(15, 35, 81)|^2, \quad (43)$$

where

$$\overline{\mathbf{S}_1^*}(15, 35, 81) = \mathbf{F}_{B1}^{g,k,s,n}(\overline{\mathbf{P}_1}). \quad (44)$$

(b) L_{B12} of $\mathbf{F}_{B1}^{g,k,s,n}(\cdot)$ on D_{test}^{B2} in $B_1^{g,k,s,n}$. We proportionally increase $\mathbf{P}_2$ by $m = \frac{f_{min}^{B2}}{f_{min}^{B1}} = \frac{20}{15}$ times, normalize it within the parameter region in $B1$, and input $\overline{\frac{20}{15} \cdot \mathbf{P}_2}$ into $\mathbf{F}_{B1}^{g,k,s,n}(\cdot)$ to predict $\overline{\mathbf{S}_2^*}(15 \times \frac{20}{15} = 20, 35 \times \frac{20}{15} \approx 46.7, 81)$,

$$\begin{aligned}\overline{\mathbf{S}_2^*}(20, 46.7, 81) &= \overline{\mathbf{S}_2^*}(15 \times \frac{20}{15}, 35 \times \frac{20}{15}, 81) \\ &= \mathbf{F}_{B1}^{g,k,s,n}(\overline{\frac{20}{15} \cdot \mathbf{P}_2}). \end{aligned} \quad (45)$$

As the target frequency range in the testing condition $B2$ is from 20 GHz to 45 GHz, the last 5 $|S_{11}|$ values out of this range are cut off,

$$\overline{\mathbf{S}_2^*}(20, 45, 76) = \overline{\mathbf{S}_2^*}(20, 46.7, 81)[:, :-5]. \quad (46)$$

Due to the misalignment of the existing $\overline{\mathbf{S}_2}(20, 45, 101)$, we simulate and collect label $\overline{\mathbf{S}_2}(20, 45, 76)$,

$$\overline{\mathbf{S}_2}(20, 45, 76) \leftarrow \text{Simulate}(\mathbf{P}_2). \quad (47)$$

L_{B12} equals the MSE between $\overline{\mathbf{S}_2}(20, 45, 76)$ and $\overline{\mathbf{S}_2^*}(20, 45, 76)$,

$$L_{B12} = \frac{1}{76}|\overline{\mathbf{S}_2}(20, 45, 76) - \overline{\mathbf{S}_2^*}(20, 45, 76)|^2. \quad (48)$$

(c) L_{B13} of $\mathbf{F}_{B1}^{g,k,s,n}(\cdot)$ on D_{test}^{B3} in $B_1^{g,k,s,n}$. $\mathbf{P}_3$ is enlarged by $m = \frac{f_{min}^{B3}}{f_{min}^{B1}} = \frac{30}{15}$ times, normalized, and input into $\mathbf{F}_{B1}^{g,k,s,n}(\cdot)$ to predict $\overline{\mathbf{S}_3^*}(15 \times \frac{30}{15} = 30, 35 \times \frac{30}{15} = 70, 81)$,

$$\begin{aligned}\overline{\mathbf{S}_3^*}(30, 70, 81) &= \overline{\mathbf{S}_3^*}(15 \times \frac{30}{15}, 35 \times \frac{30}{15}, 81) \\ &= \mathbf{F}_{B1}^{g,k,s,n}(\overline{\frac{30}{15} \cdot \mathbf{P}_3}). \end{aligned} \quad (49)$$

The last 20 $|S_{11}|$ values out over the maximum frequency of interest in the testing condition $B3$, 60 GHz, is cut off,

$$\overline{\mathbf{S}_3^*}(30, 60, 61) = \overline{\mathbf{S}_3^*}(30, 70, 81)[:, :-20]. \quad (50)$$

The label $\overline{\mathbf{S}_3}(30, 60, 61)$ is acquired through full-wave simulation,

$$\overline{\mathbf{S}_3}(30, 60, 61) \leftarrow \text{Simulate}(\mathbf{P}_3). \quad (51)$$

L_{B13} is obtained by calculating the MSE between $\overline{\mathbf{S}_3}(30, 60, 61)$ and $\overline{\mathbf{S}_3^*}(30, 60, 61)$,

$$L_{B13} = \frac{1}{61}|\overline{\mathbf{S}_3}(30, 60, 61) - \overline{\mathbf{S}_3^*}(30, 60, 61)|^2. \quad (52)$$

(d) L_{B21} of $\mathbf{F}_{B2}^{g,k,s,n}(\cdot)$ on D_{test}^{B1} in $B_2^{g,k,s,n}$. To make up the target frequency range in the testing condition $B1$ from 15 GHz to 35 GHz, we input $\overline{\frac{15}{20} \cdot \mathbf{P}_1}$ and $\overline{\frac{35}{45} \cdot \mathbf{P}_1}$ into $\mathbf{F}_{B2}^{g,k,s,n}(\cdot)$ to predict $\overline{\mathbf{S}_1^*}(20 \times \frac{15}{20} = 15, 45 \times \frac{15}{20} = 33.75, 101)$ and $\overline{\mathbf{S}_1^*}(20 \times \frac{35}{45} \approx 15.6, 45 \times \frac{35}{45} = 35, 101)$, respectively,

$$\begin{aligned}\overline{\mathbf{S}_1^*}(15, 33.75, 101) &= \overline{\mathbf{S}_1^*}(20 \times \frac{15}{20}, 45 \times \frac{15}{20}, 101) \\ &= \mathbf{F}_{B2}^{g,k,s,n}(\overline{\frac{15}{20} \cdot \mathbf{P}_1}), \end{aligned} \quad (53)$$

$$\begin{aligned}\overline{\mathbf{S}_1^*}(15.6, 35, 101) &= \overline{\mathbf{S}_1^*}(20 \times \frac{35}{45}, 45 \times \frac{35}{45}, 101) \\ &= \mathbf{F}_{B2}^{g,k,s,n}(\overline{\frac{35}{45} \cdot \mathbf{P}_1}). \end{aligned} \quad (54)$$

The last 7 points of $\overline{\mathbf{S}_1^*}(15.6, 35, 101)$ within 33.75 GHz to 35 GHz are extracted and concatenated with $\overline{\mathbf{S}_1^*}(15, 33.75, 101)$ to obtain $\overline{\mathbf{S}_1^*}(15, 35, 108)$,

$$\overline{\mathbf{S}_1^*}(33.75, 35, 7) = \overline{\mathbf{S}_1^*}(15.6, 35, 101)[:, -7:], \quad (55)$$

$$\overline{\mathbf{S}_1^*}(15, 35, 108) = \{\overline{\mathbf{S}_1^*}(15, 33.75, 101), \overline{\mathbf{S}_1^*}(33.75, 35, 7)\}. \quad (56)$$

We collect the label $\overline{\mathbf{S}_1}(15, 35, 108)$ through full-wave simulation,

$$\overline{\mathbf{S}_1}(15, 35, 108) \leftarrow \text{Simulate}(\mathbf{P}_1). \quad (57)$$

L_{B21} equals the MSE between $\overline{\mathbf{S}_1}(15, 35, 108)$ and $\overline{\mathbf{S}_1^*}(15, 35, 108)$,

$$L_{B21} = \frac{1}{108}|\overline{\mathbf{S}_1}(15, 35, 108) - \overline{\mathbf{S}_1^*}(15, 35, 108)|^2. \quad (58)$$

(e) L_{B22} of $\mathbf{F}_{B2}^{g,k,s,n}(\cdot)$ on D_{test}^{B2} in $B_2^{g,k,s,n}$,

$$L_{B22} = \frac{1}{101}|\overline{\mathbf{S}_2}(20, 45, 101) - \overline{\mathbf{S}_2^*}(20, 45, 101)|^2, \quad (59)$$

where

$$\overline{\mathbf{S}_2^*}(20, 45, 101) = \mathbf{F}_{B2}^{g,k,s,n}(\overline{\mathbf{P}_2}). \quad (60)$$

(f) L_{B23} of $\mathbf{F}_{B2}^{g,k,s,n}(\cdot)$ on D_{test}^{B3} in $B_2^{g,k,s,n}$. We input $\overline{\frac{30}{20} \cdot \mathbf{P}_3}$ into $\mathbf{F}_{B2}^{g,k,s,n}(\cdot)$ to predict $\overline{\mathbf{S}_3^*}(20 \times \frac{30}{20} = 30, 45 \times \frac{30}{20} = 67.5, 101)$,

$$\begin{aligned}\overline{\mathbf{S}_3^*}(30, 67.5, 101) &= \overline{\mathbf{S}_3^*}(20 \times \frac{30}{20}, 45 \times \frac{30}{20}, 101) \\ &= \mathbf{F}_{B2}^{g,k,s,n}(\overline{\frac{30}{20} \cdot \mathbf{P}_3}). \end{aligned} \quad (61)$$

We cut off the last 20 points of $\overline{\mathbf{S}_3^*}(30, 67.5, 101)$ over the desired maximum frequency 60 GHz in the testing condition $B3$ to form $\overline{\mathbf{S}_3^*}(30, 60, 81)$,

$$\overline{\mathbf{S}_3^*}(30, 60, 81) = \overline{\mathbf{S}_3^*}(30, 67.5, 101)[:, : -20]. \quad (62)$$

The label $\overline{\mathbf{S}_3}(30, 60, 81)$ is acquired through full-wave simulation,

$$\overline{\mathbf{S}_3}(30, 60, 81) \leftarrow \text{Simulate}(\mathbf{P}_3). \quad (63)$$

L_{B23} equals the MSE between $\overline{\mathbf{S}_3}(30, 60, 81)$ and $\overline{\mathbf{S}_3^*}(30, 60, 81)$,

$$L_{B23} = \frac{1}{81}|\overline{\mathbf{S}_3}(30, 60, 81) - \overline{\mathbf{S}_3^*}(30, 60, 81)|^2. \quad (64)$$

(g) L_{B31} of $\mathbf{F}_{B3}^{g,k,s,n}(\cdot)$ on D_{test}^{B1} in $B_3^{g,k,s,n}$. To cover the frequency range of interest in the testing condition $B1$ from 15 GHz to 35 GHz, we input $\frac{15}{30} \cdot \mathbf{P}_1$ and $\frac{35}{60} \cdot \mathbf{P}_1$ into $\mathbf{F}_{B3}^{g,k,s,n}(\cdot)$ to predict $\overline{\mathbf{S}_1^*}(30 \times \frac{15}{30} = 15, 60 \times \frac{15}{30} = 30, 121)$ and $\overline{\mathbf{S}_1^*}(30 \times \frac{35}{60} = 17.5, 60 \times \frac{35}{60} = 35, 121)$, respectively,

$$\begin{aligned}\overline{\mathbf{S}_1^*}(15, 30, 121) &= \overline{\mathbf{S}_1^*}(30 \times \frac{15}{30}, 60 \times \frac{15}{30}, 121) \\ &= \mathbf{F}_{B3}^{g,k,s,n}(\overline{\frac{15}{30} \cdot \mathbf{P}_1}), \quad (65)\\ \overline{\mathbf{S}_1^*}(17.5, 35, 121) &= \overline{\mathbf{S}_1^*}(30 \times \frac{35}{60}, 60 \times \frac{35}{60}, 121) \\ &= \mathbf{F}_{B3}^{g,k,s,n}(\overline{\frac{35}{60} \cdot \mathbf{P}_1}). \quad (66)\end{aligned}$$

We concatenate $\overline{\mathbf{S}_1^*}(15, 30, 121)$ and the last 34 points of $\overline{\mathbf{S}_1^*}(17.5, 35, 121)$ within 30 GHz to 35 GHz to form $\overline{\mathbf{S}_1^*}(15, 35, 155)$,

$$\begin{aligned}\overline{\mathbf{S}_1^*}(30, 35, 34) &= \overline{\mathbf{S}_1^*}(17.5, 35, 121)[:, -34 :], \quad (67)\\ \overline{\mathbf{S}_1^*}(15, 35, 155) &= \{\overline{\mathbf{S}_1^*}(15, 30, 121), \overline{\mathbf{S}_1^*}(30, 35, 34)\}. \quad (68)\end{aligned}$$

We collect the label $\overline{\mathbf{S}_1}(15, 35, 155)$ through full-wave simulation,

$$\overline{\mathbf{S}_1}(15, 35, 155) \leftarrow \text{Simulate}(\mathbf{P}_1). \quad (69)$$

L_{B31} equals the MSE between $\overline{\mathbf{S}_1}(15, 35, 155)$ and $\overline{\mathbf{S}_1^*}(15, 35, 155)$,

$$L_{B31} = \frac{1}{155}|\overline{\mathbf{S}_1}(15, 35, 155) - \overline{\mathbf{S}_1^*}(15, 35, 155)|^2. \quad (70)$$

(h) L_{B32} of $\mathbf{F}_{B3}^{g,k,s,n}(\cdot)$ on D_{test}^{B2} in $B_3^{g,k,s,n}$. To make up the frequency range of interest in the testing condition $B2$ from 20 GHz to 45 GHz, we input $\frac{20}{30} \cdot \mathbf{P}_2$ and $\frac{45}{60} \cdot \mathbf{P}_2$ into $\mathbf{F}_{B3}^{g,k,s,n}(\cdot)$ to predict $\overline{\mathbf{S}_2^*}(30 \times \frac{20}{30} = 20, 60 \times \frac{20}{30} = 40, 121)$ and $\overline{\mathbf{S}_2^*}(30 \times \frac{45}{60} = 22.5, 60 \times \frac{45}{60} = 45, 121)$, respectively,

$$\begin{aligned}\overline{\mathbf{S}_2^*}(20, 40, 121) &= \overline{\mathbf{S}_2^*}(30 \times \frac{20}{30}, 60 \times \frac{20}{30}, 121) \\ &= \mathbf{F}_{B3}^{g,k,s,n}(\overline{\frac{20}{30} \cdot \mathbf{P}_2}), \quad (71)\\ \overline{\mathbf{S}_2^*}(22.5, 45, 121) &= \overline{\mathbf{S}_2^*}(30 \times \frac{45}{60}, 60 \times \frac{45}{60}, 121) \\ &= \mathbf{F}_{B3}^{g,k,s,n}(\overline{\frac{45}{60} \cdot \mathbf{P}_2}). \quad (72)\end{aligned}$$

We extract the last 26 points of $\overline{\mathbf{S}_2^*}(22.5, 45, 121)$ within 40 GHz to 45 GHz and integrate with $\overline{\mathbf{S}_2^*}(20, 40, 121)$ to form $\overline{\mathbf{S}_2^*}(20, 45, 147)$,

$$\begin{aligned}\overline{\mathbf{S}_2^*}(40, 45, 26) &= \overline{\mathbf{S}_2^*}(22.5, 45, 121)[:, -26 :], \quad (73)\\ \overline{\mathbf{S}_2^*}(20, 45, 147) &= \{\overline{\mathbf{S}_2^*}(20, 40, 121), \overline{\mathbf{S}_2^*}(40, 45, 26)\}. \quad (74)\end{aligned}$$

The label $\overline{\mathbf{S}_2}(20, 45, 147)$ is obtained through full-wave simulation,

$$\overline{\mathbf{S}_2}(20, 45, 147) \leftarrow \text{Simulate}(\mathbf{P}_2). \quad (75)$$

L_{B32} equals the MSE between $\overline{\mathbf{S}_2}(20, 45, 147)$ and $\overline{\mathbf{S}_2^*}(20, 45, 147)$,

$$L_{B32} = \frac{1}{147}|\overline{\mathbf{S}_2}(20, 45, 147) - \overline{\mathbf{S}_2^*}(20, 45, 147)|^2. \quad (76)$$

(i) L_{B33} of $\mathbf{F}_{B3}^{g,k,s,n}(\cdot)$ on D_{test}^{B3} in $B_3^{g,k,s,n}$,

$$L_{B33} = \frac{1}{121}|\overline{\mathbf{S}_3}(30, 60, 121) - \overline{\mathbf{S}_3^*}(30, 60, 121)|^2, \quad (77)$$

where

$$\overline{\mathbf{S}_3^*}(30, 60, 121) = \mathbf{F}_{B3}^{g,k,s,n}(\overline{\mathbf{P}_3}). \quad (78)$$

The testing results are listed in Table IV. It can be observed that only proportional frequency points can be predicted, a very complex multiple proportioning process is needed for testing when involving more variable frequency conditions, and the modeling accuracy in the unseen frequency conditions deteriorates severely.

D. Proposed Method

We carry out the proposed method in multiple experiments to validate its generalizability and scalability. The source datasets are reformatted and converted to be frequency-wise. For n-th pair of geometric parameters $\mathbf{P}_n$ and $|S_{11}|_n$ ($n \in [1,\ 150]$) in $B1$, $B2$, or $B3$, $\mathbf{P}_n$ is converted to electrical dimensions $\mathbf{E}_n(\mathbf{P}_n, f_x)$ using (15). Note that the relative geometric parameter c in $\mathbf{P}_n$ is converted to an absolute geometric parameter by $c \times (r_1 - r_2)$ before calculating its electrical dimension. Each pair of $\mathbf{E}_n(\mathbf{P}_n, f_x)$ and corresponding $|S_{11}|_n$ at f_x forms a new frequency-wise sample. The sizes of the generated frequency-wise training, validation, and testing datasets for each condition are increased by n_f: 5670, 2430, and 4050 for $B1$; 7070, 3030, and 5050 for $B2$; 8470, 3630, and 6050 for $B3$, respectively. Every sample has a fixed input size of 10 and an output size of 1.

1) Generalizability: To assess the generalizability, the proposed method is compared with the existing methods by training with data exclusively under each condition $B1$, $B2$, and $B3$, referred to as B_1^p, B_2^p, and B_3^p, respectively. Each model is optimized through Bayesian optimization. As the number of data increases proportionally with the number of frequency points n_f, the sampling ranges of the number of epochs and the batch size are modified accordingly. The number of epochs is sampled from 500 to 2000, with an

TABLE IV
COMPARATIVE RESULTS IN IMPLEMENTATION B: PLANAR METASURFACE LENS

Exp.	Trained on	Model	Optimized Hyperparameters						Test Loss ($\times 10^{-2}$) on		
			N_e	N_g	lr	N_b	N_h	$[N_{n1}, N_{n2}, \cdots]$	D_{test}^{B1}	D_{test}^{B2}	D_{test}^{B3}
B_1^g	D_{train}^{B1}	**GPR**	–						5.19	10.20	13.22
B_1^k		**Kri.**	–						6.99	8.18	8.38
B_1^s		**SVR**	–						5.37	5.71	6.49
B_1^n		**NN**	800	8000	0.001	7	4	$[256, 32, 32, 16]$	2.47	3.45	6.99
B_1^p		**Pro.**	1000	70000	0.001	81	7	$[256, 128, 256, 128, 32, 32, 128]$	1.57	3.47	6.75
B_2^g	D_{train}^{B2}	**GPR**	–						13.93	3.52	7.60
B_2^k		**Kri.**	–						10.74	3.92	8.22
B_2^s		**SVR**	–						9.64	3.69	6.84
B_2^n		**NN**	1000	70000	0.0001	1	5	$[64, 128, 256, 128, 128]$	6.59	0.80	5.58
B_2^p		**Pro.**	2000	20000	0.001	707	4	$[256, 256, 16, 32]$	5.18	0.69	2.94
B_3^g	D_{train}^{B3}	**GPR**	–						14.72	6.17	4.66
B_3^k		**Kri.**	–						10.62	7.79	6.29
B_3^s		**SVR**	–						9.53	4.71	5.73
B_3^n		**NN**	800	8000	0.001	7	6	$[256, 64, 256, 128, 128, 16]$	8.42	3.16	3.24
B_3^p		**Pro.**	1500	52500	0.001	242	7	$[32, 64, 32, 128, 128, 256, 256]$	7.19	1.91	2.02
B_{12}^p	D_{train}^{B1} & D_{train}^{B2}	**Pro.**	2000	52000	0.0001	500	2	$[256, 128]$	2.36	0.94	3.51
B_{13}^p	D_{train}^{B1} & D_{train}^{B3}	**Pro.**	2000	58000	0.001	500	3	$[64, 256, 32]$	2.15	1.15	2.39
B_{23}^p	D_{train}^{B2} & D_{train}^{B3}	**Pro.**	1500	16500	0.001	1500	3	$[256, 32, 32]$	4.79	2.17	2.92
B_{123}^p	D_{train}^{B1} & D_{train}^{B2} & D_{train}^{B3}	**Pro.**	2000	426000	0.001	100	3	$[256, 256, 128]$	2.00	0.79	1.74

Note: **Exp.** refers to the experiment index; **GPR** refers to Gaussian process regression;
Kri. refers to kriging; **SVR** refers to support vector regression;
NN refers to neural networks; **Pro.** refers to the proposed method;
N_e refers to the number of epochs; N_g refers to the number of gradient steps;
lr refers to the learning rate; N_b refers to the batch size;
N_h refers to the number of hidden layers; $[N_{n1}, N_{n2}, \cdots]$ refers to the number of neurons in each hidden layer.

interval of 500. The batch size is sampled from 1 to 10 times of n_f, with an interval of n_f. Here, n_f equals 81 for B_1^p, 101 for B_2^p, and 121 for B_3^p, respectively. The other settings remain unchanged. This modification minimizes the difference between the existing and proposed methods during the training procedures to ensure a fair comparison. By monitoring the validation loss on D_{val}^{B1}, D_{val}^{B2}, or D_{val}^{B3} for B_1^p, B_2^p, or B_3^p, respectively, the optimization ceases after 100 iterations without improvement or upon reaching 500 iterations. Table IV shows the optimized architectures. We test each model on all three test datasets, D_{test}^{B1}, D_{test}^{B2}, and D_{test}^{B3}. Let us take the model in B_1^p as an example, as it is solely trained on D_{train}^{B1}, testing on D_{test}^{B1} indicates its interpolation ability, while testing on D_{test}^{B2} and D_{test}^{B3} indicates its extrapolation ability under unseen frequency conditions. The results are listed in Table IV.

The testing results are listed in Table IV. When the model is trained on D_{train}^{B1}, it improves the accuracy on D_{test}^{B3} in its unseen frequency condition in $B3$ by $3.5\,\%$ ($\frac{|6.75-6.99|}{6.99} \times 100\,\% \approx 3.5\,\%$), while its performance for other two conditions maintains. For the model trained on D_{train}^{B2}, its performance on D_{test}^{B1} and D_{test}^{B3} in its unseen conditions are enhanced by $21.4\,\%$ ($\frac{|5.18-6.59|}{6.59} \times 100\,\% \approx 21.4\,\%$) and $47.3\,\%$ ($\frac{|2.94-5.58|}{5.58} \times 100\,\% \approx 47.3\,\%$), respectively. The model trained on D_{train}^{B3} realizes respective $14.6\,\%$ ($\frac{|7.19-8.42|}{8.42} \times 100\,\% \approx 14.6\,\%$) and $39.6\,\%$ ($\frac{|1.91-3.16|}{3.16} \times 100\,\% \approx 39.6\,\%$) enhancement of accuracy on D_{test}^{B1} and D_{test}^{B2} in its unseen frequency conditions.

The comparative results show that the proposed method significantly improves the modeling accuracy under unseen frequency conditions while maintaining its performance in the original condition, demonstrating the improved generalizability of the proposed framework.

2) Scalability: The proposed model is jointly trained using datasets from multiple frequency conditions to validate its scalability, referred to as experiments B_{12}^p, B_{13}^p, B_{23}^p, and B_{123}^p in Table IV. For example, B_{12}^p refers to training on both D_{train}^{B1} and D_{train}^{B2}. It can be considered as scaling up the model in B_1^p by incorporating new data D_{train}^{B2} in a new condition $B2$. B_{123}^p refers to training using datasets under all the three conditions, D_{train}^{B1}, D_{train}^{B2}, and D_{train}^{B3}, which can be obtained by scaling up any other models. When optimizing each model through Bayesian optimization, the batch size is chosen from a discrete set of values, $[100, 200, 500, 1000, 1500, 2000]$, and other hyperparameters are sampled from the same space in Section IV-D1. The optimized models and corresponding test results are listed in Table IV.

In general, the results in the experiments B_{12}^p, B_{13}^p, and B_{23}^p indicate that the proposed model is scalable to incorporate

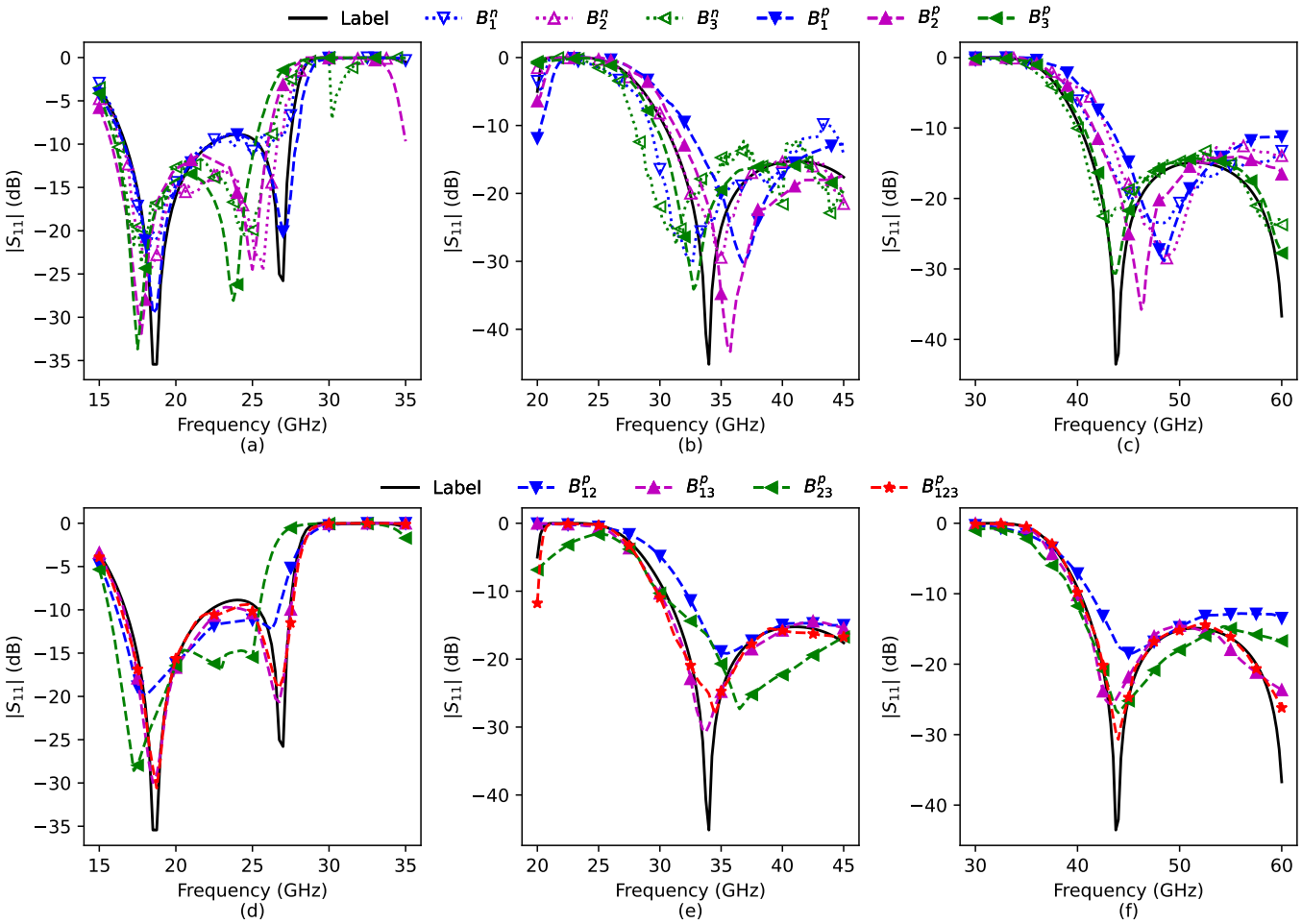


Fig. 6. Comparison of $|S_{11}|$ curves in Implementation B. (a) A test sample in D_{test}^{B1} for B_1^n, B_2^n, B_3^n, B_1^p, B_2^p, and B_3^p. (b) A test sample in D_{test}^{B2} for B_1^n, B_2^n, B_3^n, B_1^p, B_2^p, and B_3^p. (c) A test sample in D_{test}^{B3} for B_1^n, B_2^n, B_3^n, B_1^p, B_2^p, and B_3^p. (d) A test sample in D_{test}^{B1} for B_{12}^p, B_{13}^p, B_{23}^p, and B_{123}^p. (e) A test sample in D_{test}^{B2} for B_{12}^p, B_{13}^p, B_{23}^p, and B_{123}^p. (f) A test sample in D_{test}^{B3} for B_{12}^p, B_{13}^p, B_{23}^p, and B_{123}^p. ("Label" denotes the label $|S_{11}|$ generated from full-wave simulation.)

variable frequency conditions, leading to further improved generalizability. For example, as the model scaling up from B_1^p to B_{12}^p, it further enhances the accuracy on D_{test}^{B3} in its unseen condition $B3$ by $48.0\,\%$ ($\frac{|3.51-6.75|}{6.75} \times 100\,\% = 48.0\,\%$). Scaling up from B_1^p to B_{13}^p by incorporating D_{train}^{B3} enhances the accuracy on D_{test}^{B2} under its unseen condition by $66.9\,\%$ ($\frac{|1.15-3.47|}{3.47} \times 100\,\% = 66.9\,\%$). Based on B_2^p, scaling up to B_{23}^p reduces the modeling error on D_{test}^{B1} in its unseen condition by $7.5\,\%$ ($\frac{|4.79-5.18|}{5.18} \times 100\,\% = 7.5\,\%$). $39.8\,\%$ ($\frac{|1.15-1.91|}{1.91} \times 100\,\% = 39.8\,\%$) and $33.4\,\%$ ($\frac{|4.79-7.19|}{7.19} \times 100\,\% = 33.4\,\%$) improvements are achieved by scaling up from B_3^p to B_{13}^p and B_{23}^p, respectively.

The largest-scaled model B_{123}^p incorporates all the three frequency conditions, $B1$, $B2$, and $B3$, which can be obtained by scaling up any other models, realizing superior modeling performance across variable frequency conditions. $|S_{11}|$ curves generated by neural networks and the proposed method are compared in Fig. 6. Here, black curves marked as "Label" are simulation results from the full-wave simulator CST, which indicate the ground truth. The proposed method achieves higher extrapolation accuracy along frequency than neural networks. As the model is scaled up from single to triple conditions, the modeling accuracy of $|S_{11}|$ curves enhances and the maximum modeling error across variable conditions reduces gradually, matching well the full-wave simulation results, which demonstrates further enhanced generalizability.

Overall, the comparative results in Table IV validate the improved generalizability and scalability of the proposed modeling framework as the solving dimensionality increases from five to ten under three different frequency conditions. Besides, the modeling process is more flexible and straightforward than the existing methods. This improvement is attributed to the frequency-wise learning strategy that enforces the model's robust understanding of the EM similarity laws and non-linear proportioning characteristics.

V. Discussion

A. Potentials and Limitations

The enhanced generalizability and scalability of our proposed approach are attributed to its embedding of the EM

similarity laws and robust understanding of non-linear proportioning characteristics. Therefore, our approach primarily works in the applications with frequency-independent material properties (permittivity ε, permeability μ, conductivity σ), where the EM similarity laws hold true. Besides, its application region is also limited by the dimensionality and data availability. Theoretically, the proposed method can handle more complex and higher-dimensional modeling because its working principle is independent of dimensionality, as long as sufficient training data samples are provided. As the dimensionality grows, the amount of training data samples required increases exponentially. Therefore, the highest dimensionality is mainly limited by the available computational resources for data generation. The model's capability of handling high-dimensional modeling can be further improved by utilizing advanced sampling strategies, data augmentation techniques, and small-sample machine learning approaches. In future work, we will investigate small-sample frequency-wise learning techniques to reduce the computational costs and further enhance the benefits of the proposed method. Tuning the frequency sampling density is another meaningful direction to explore. A basic strategy is to increase the frequency sampling points where the EM responses vary heavily. An optimized dynamic sampling strategy can be developed to balance the computational costs and modeling performance along the frequency domain.

Compared with the conventional full-wave-simulator-based modeling methods, the proposed method significantly accelerates each modeling process and reduces the computational costs. Although initial data collection and model training require time and computational costs, the proposed method shows superior long-term efficiency as the number of design tasks increases, outperforming the existing methods. The traditional full-wave-simulator-based modeling approaches repetitively perform meshing and solve Maxwell's equations to model an EM structure for each setting of geometric parameters, incurring redundant time and computational costs. In contrast, the proposed method effectively resolves the modeling of the EM structure by developing a surrogate model trained on representative simulation data, thereby enhancing the long-term efficiency by eliminating the unnecessary repetitive time and computation costs.

B. Increasing Complexity and Dimensionality

The modeling of a multi-resonant bandpass filter with increasing complexity and dimensionality is investigated to validate the proposed method.

A multi-resonant bandpass filter presented by Yang *et al.* in [39] is modeled in Implementation C to validate the flexibility and robustness of the proposed method. As shown in Fig. 7, it consists of three metal layers, which are etched on two substrate layers of $Rogers\ RO4003C$ with relative permittivity of $\varepsilon_r = 3.38$ and loss tangent of $tan\delta = 0.0027$. The top and bottom metal layers are two centrosymmetric feeding structures. Each feeding structure is an open-ended two-order microstrip line. The middle layer is composed of a cross-shaped slot etched on a rectangular metal plane. A wide

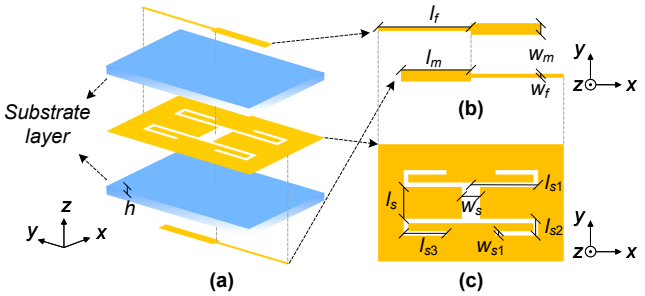


Fig. 7. Multi-resonant bandpass filter in Implementation C. (a) 3D view. (b) Top and bottom metal layers. (c) Middle metal layer.

TABLE V
TWO DIFFERENT FREQUENCY CONDITIONS AND CORRESPONDING GEOMETRIC PARAMETERS IN IMPLEMENTATION C: MULTI-RESONANT BANDPASS FILTER

		Condition	
		$C1$	$C2$
Frequency	Range (GHz)	$[0.5, 3.5]$	$[4, 28]$
	Points	61	241
Geometric Parameters (Unit: mm)	l_s	$[9.1, 10.7]$	$[1.3, 2.3]$
	l_m	$[19.8, 21.2]$	$[1.9, 3.0]$
	w_m	$[3.3, 4.5]$	$[0.7, 1.3]$
	l_{s1}	$[11.5, 12.5]$	$[1.3, 2.1]$
	l_{s2}	$[2, 3]$	$[1.1, 1.7]$
	l_{s3}	$[9.5, 10.5]$	$[1.2, 1.8]$
	w_s	$[1.2, 1.8]$	$[0.2, 0.5]$
	w_{s1}	$[0.3, 0.7]$	$[0.2, 0.4]$
	h	1.5	0.5
	l_x	45	7
	l_y	35	6
	w_f	1.5	0.5

passband filtering performance with multiple resonances is generated. The filter's reflection coefficient $|S_{11}|$ is optimized by tuning 12 geometric variables (L_s, L_m, W_m, L_{s1}, L_{s2}, L_{s3}, W_s, W_{s1}, h, l_x, l_y, and w_f), with other geometric parameters fixed at constant values.

Table V exhibits two different frequency conditions, $C1$ from 0.5 GHz to 3.5 GHz with 61 points and $C2$ from 4 GHz to 28 GHz. Correspondingly, distinct tuning ranges are assigned for the geometric parameters in $C1$ and $C2$. 180 pairs of geometric parameters and $|S_{11}|$ are collected through Latin Hypercube Sampling and full-wave simulation. Each $|S_{11}|$ sample is of size 61 in $C1$ and 241 in $C2$, respectively. The 180 data samples under each condition are arbitrarily divided into training, validation, and testing datasets in the ratio of $10:3:5$ under a random seed: D^{C1}_{train}, D^{C1}_{val}, and D^{C1}_{test} in $C1$; D^{C2}_{train}, D^{C2}_{val}, and D^{C2}_{test} in $C2$, respectively.

Four existing methods, Gaussian process regression, kriging, support vector regression, and neural networks, are utilized to model the filter for comparison. The existing methods have difficulty in establishing an integrated model that incorporates samples in both $C1$ and $C2$. Separate experiments and models are assigned: C^g_1, C^k_1, C^s_1, and C^n_1 in $C1$; C^g_2, C^k_2, C^s_2, and C^n_2 in $C2$, as listed in Table VI. During these experiments, the geometric parameters and $|S_{11}|$ are normalized. Each data sample has the output size of 61 in $C1$ and 241 in $C2$, while the input size is fixed at 12. Similarly, each neural network

TABLE VI
COMPARATIVE RESULTS IN IMPLEMENTATION C: MULTI-RESONANT BANDPASS FILTER

Exp.	Trained on	Model	Optimized Hyperparameters						Test Loss ($\times 10^{-2}$) on	
			N_e	N_g	lr	N_b	N_h	$[N_{n1}, N_{n2}, \cdots]$	D_{test}^{C1}	D_{test}^{C2}
C_1^g	D_{train}^{C1}	**GPR**	–						0.55	23.87
C_1^k		**Kri.**	–						0.62	10.62
C_1^s		**SVR**	–						0.24	14.24
C_1^n		**NN**	1000	13000	0.0001	8	4	$[32, 32, 256, 16]$	0.51	17.56
C_1^p		**Pro.**	600	7200	0.001	549	4	$[16, 256, 32, 256]$	0.65	9.78
C_2^g	D_{train}^{C2}	**GPR**	–						19.46	0.96
C_2^k		**Kri.**	–						17.63	1.01
C_2^s		**SVR**	–						15.83	1.34
C_2^n		**NN**	600	7200	0.001	9	3	$[256, 64, 256]$	6.21	2.00
C_2^p		**Pro.**	1000	13000	0.001	1928	4	$[256, 128, 256, 128]$	8.31	2.36
C_{12}^p	D_{train}^{C1} & D_{train}^{C2}	**Pro.**	1500	43500	0.001	200	2	$[256, 32]$	1.75	2.20

Note: **Exp.** refers to the experiment index; **GPR** refers to Gaussian process regression; **Kri.** refers to kriging; **SVR** refers to support vector regression; **NN** refers to neural networks; **Pro.** refers to the proposed method; N_e refers to the number of epochs; N_g refers to the number of gradient steps; lr refers to the learning rate; N_b refers to the batch size; N_h refers to the number of hidden layers; $[N_{n1}, N_{n2}, \cdots]$ refers to the number of neurons in each hidden layer.

model's hyperparameters are optimized using Bayesian optimization. Table VI records the test losses of the eight models on D_{test}^{C1} and D_{test}^{C2}. Testing $C_1^{g,k,s,n}$ on D_{test}^{C1} or testing $C_2^{g,k,s,n}$ on D_{test}^{C2} assesses the interpolation accuracy; testing $C_1^{g,k,s,n}$ on D_{test}^{C2} or testing $C_2^{g,k,s,n}$ on D_{test}^{C1} measures the extrapolation accuracy. The testing procedure is similar to that of Implementation A and is therefore omitted for simplicity. As shown in Table VI, the kriging model (C_1^k) under the $C1$ condition realizes the highest extrapolation accuracy on D_{test}^{C2} under the $C2$ condition. However, the kriging model's performance (C_2^k) under the $C2$ condition severely deteriorates when extrapolating on D_{test}^{C1} under the $C1$ condition. Conversely, the neural network model (C_2^n) under the $C2$ condition obtains the highest extrapolation accuracy on D_{test}^{C1} under the $C1$ condition, but C_1^n under the $C1$ condition exhibits significant degradation when extrapolating on D_{test}^{C2} under the $C2$ condition.

In experiments C_1^p and C_2^p, two proposed models are separately trained under $C1$ and $C2$, respectively. Their hyperparameters are optimized within similar tuning ranges using Bayesian optimization. The proposed method's generalizability is validated by comparing C_1^p with $C_1^{g,k,s,n}$ and comparing C_2^p with $C_2^{g,k,s,n}$, respectively. As shown in Table VI, the existing methods sometimes perform well under one condition, but their extrapolation ability might severely deteriorate when switching to another condition. Compared with them, the proposed method achieves balanced extrapolation accuracy under both conditions $C1$ and $C2$, with comparative performance with kriging under $C1$ and with neural networks under $C2$.

Superior to the existing approaches, the proposed method enables an integrated model that leverages training samples under both conditions, D_{train}^{C1} and D_{train}^{C2}, referred to as C_{12}^p in Table VI. As it scales up from C_1^p or C_2^p, C_{12}^p significantly reduces the maximum modeling error, which demonstrates enhanced scalability for variable frequency conditions.

VI. CONCLUSION

This paper introduces a unified frequency-wise electromagnetic (EM) modeling framework to enhance generalizability and scalability for variable frequency conditions. The proposed method addresses the limitations inherent in existing modeling techniques, which suffer from deteriorated accuracy under unseen frequency conditions and require multiple separate models for different frequency conditions. Integrating a novel frequency-wise learning strategy, our approach enforces a robust understanding of the EM similarity and non-linear proportioning characteristics, hence improving generalizability and scalability for variable frequency conditions. The effectiveness of the proposed framework is demonstrated through multiple implementations involving increased solving dimensionality and variable frequency conditions. Compared with the existing methods, including Gaussian process regression, kriging, support vector regression, and neural networks, the proposed framework outperforms with respect to generalizability and scalability. It has the potential to develop a powerful large-scale EM model by incorporating significant frequency conditions, thereby greatly accelerating the design and optimization processes in diverse EM applications. Future work may leverage its capabilities to encompass other EM problems and extend generalizability and scalability for diverse geometric topologies.

REFERENCES

[1] Z. Zhou, Z. Wei, A. Tahir, J. Ren, Y. Yin, G. F. Pedersen and M. Shen, "A high-quality data acquisition method for machine learning-based design and analysis of electromagnetic structures," *IEEE Trans. Microw. Theory Techn.*, vol. 71, no. 10, pp. 4295-4306, Mar. 2023, doi: 10.1109/TMTT.2023.3235066.

[2] Z. Zhou, Z. Wei, J. Ren, Y. Yin, G. F. Pedersen and M. Shen, "Bayesian-inspired sampling for efficient machine-learning-assisted microwave component design," *IEEE Trans. Microw. Theory Techn.*, vol. 72, no. 2, pp. 996-1007, Aug. 2023, doi: 10.1109/TMTT.2023.3298194.

[3] J. P. Jacobs and S. Koziel, "Two-stage framework for efficient Gaussian process modeling of antenna input characteristics," *IEEE Trans. Antennas Propag.*, vol. 62, no. 2, pp. 706-713, Feb. 2014, doi: 10.1109/TAP.2013.2290121.

[4] J. P. Jacobs, "Characterisation by Gaussian processes of finite substrate size effects on gain patterns of microstrip antennas," *IET Microw., Antennas Propag.*, vol. 10, no. 11, pp. 1189–1195, Aug. 2016, doi: 10.1049/iet-map.2015.0621.

[5] Z. Zhang, F. Jiang, Y. Jiao, and Q. S. Cheng, "Low-cost surrogate modeling of antennas using two-level Gaussian process regression method," *Int. J. Numer. Model., Electron. Netw., Devices Fields*, vol. 34, no. 5, pp. 1-16, Aug. 2021, doi: 10.1002/jnm.2886.

[6] C. Hu, S. Zeng, C. Li, and F. Zhao, "On nonstationary Gaussian process model for solving data-driven optimization problems," *IEEE Trans. Cybern.*, vol. 53, no. 4, pp. 2440-2453, April 2023, doi: 10.1109/TCYB.2021.3120188.

[7] J. Du and C. Roblin, "Statistical modeling of disturbed antennas based on the polynomial chaos expansion," *IEEE Antennas Wireless Propag. Lett.*, vol. 16, pp. 1843–1846, 2017, doi: 10.1109/LAWP.2016.2609739.

[8] A. Petrocchi, A. Kaintura, G. Avolio, D. Spina, T. Dhaene, and A. Raffo, "Measurement uncertainty propagation in transistor model parameters via polynomial chaos expansion," *IEEE Microw. Wireless Compon. Lett.*, vol. 27, no. 6, pp. 572–574, Jun. 2017, doi: 10.1109/LMWC.2017.2701334.

[9] J. Du and C. Roblin, "Stochastic surrogate models of deformable antennas based on vector spherical harmonics and polynomial chaos expansions: Application to textile antennas," *IEEE Trans. Antennas Propag.*, vol. 66, no. 7, pp. 3610–3622, Jul. 2018, doi: 10.1109/TAP.2018.2829820.

[10] S. Koziel, S. Ogurtsov, I. Couckuyt, and T. Dhaene, "Variable-fidelity electromagnetic simulations and co-kriging for accurate modeling of antennas," *IEEE Trans. Antennas Propag.*, vol. 61, no. 3, pp. 1301-1308, March 2013, doi: 10.1109/TAP.2012.2231924.

[11] S. Koziel and A. T. Sigurðsson, "Triangulation-based constrained surrogate modeling of antennas," *IEEE Trans. Antennas Propag.*, vol. 66, no. 8, pp. 4170-4179, Aug. 2018, doi: 10.1109/TAP.2018.2839759.

[12] S. Koziel, A. Pietrenko-Dabrowska, and U. Ullah, "Low-cost modeling of microwave components by means of two-stage inverse/forward surrogates and domain confinement," *IEEE Trans. Microw. Theory Techn.*, vol. 69, no. 12, pp. 5189-5202, Dec. 2021, doi: 10.1109/TMTT.2021.3112156.

[13] S. Koziel and A. Pietrenko-Dabrowska, "Expedited variable-resolution surrogate modeling of miniaturized microwave passives in confined domains," *IEEE Trans. Microw. Theory Techn.*, vol. 70, no. 11, pp. 4740-4750, Nov. 2022, doi: 10.1109/TMTT.2022.3191327.

[14] D. R. Prado, J. A. López-Fernández, M. Arrebola, and G. Goussetis, "Support vector regression to accelerate design and crosspolar optimization of shaped-beam reflectarray antennas for space applications," *IEEE Trans. Antennas Propag.*, vol. 67, no. 3, pp. 1659-1668, Mar. 2019, doi: 10.1109/TAP.2018.2889029.

[15] D. R. Prado, J. A. López-Fernández, M. Arrebola, M. R. Pino, and G. Goussetis, "Wideband shaped-beam reflectarray design using support vector regression analysis," *IEEE Antennas Wireless Propag. Lett.*, vol. 18, no. 11, pp. 2287-2291, Nov. 2019, doi: 10.1109/LAWP.2019.2932902.

[16] J. P. Jacobs, S. Koziel, and S. Ogurtsov, "Computationally efficient multi-fidelity Bayesian support vector regression modeling of planar antenna input characteristics," *IEEE Trans. Antennas Propag.*, vol. 61, no. 2, pp. 980-984, Feb. 2013, doi: 10.1109/TAP.2012.2220513.

[17] B. Liu, H. Aliakbarian, Z. Ma, G. A. E. Vandenbosch, G. Gielen, and P. Excell, "An efficient method for antenna design optimization based on evolutionary computation and machine learning techniques," *IEEE Trans. Antennas Propag.*, vol. 62, no. 1, pp. 7-18, Jan. 2014, doi: 10.1109/TAP.2013.2283605.

[18] L.-Y. Xiao, W. Shao, X. Ding, and B.-Z. Wang, "Dynamic adjustment kernel extreme learning machine for microwave component design," *IEEE Trans. Microw. Theory Techn.*, vol. 66, no. 10, pp. 4452-4461, Oct. 2018, doi: 10.1109/TMTT.2018.2858787.

[19] Z. Zhou, Z. Wei, J. Ren, Y. Yin, G. F. Pedersen, and M. Shen, "Two-order deep learning for generalized synthesis of radiation patterns for antenna arrays," *IEEE Trans. Artif. Intell.*, vol. 4, no. 5, pp. 1359-1368, Oct. 2023, doi: 10.1109/TAI.2022.3192505.

[20] Z. Wei, Z. Zhou, P. Wang, J. Ren, Y. Yin, G. F. Pedersen, and M. Shen, "Equivalent circuit theory-assisted deep learning for accelerated generative design of metasurfaces," *IEEE Trans. Antennas Propag.*, vol. 70, no. 7, pp. 5120-5129, Feb. 2022, doi: 10.1109/TAP.2022.3152592.

[21] Q. Wu, H. Wang, and W. Hong, "Multistage collaborative machine learning and its application to antenna modeling and optimization," *IEEE Trans. Antennas Propag.*, vol. 68, no. 5, pp. 3397-3409, May 2020, doi: 10.1109/TAP.2019.2963570.

[22] Z. Zhou, Z. Wei, J. Ren, Y. Yin, G. F. Pedersen, and M. Shen, "Transfer learning assisted multi-element calibration for active phased antenna arrays," *IEEE Trans. Antennas Propag.*, vol. 71, no. 2, pp. 1982-1987, Oct. 2022, doi: 10.1109/TAP.2022.3216548.

[23] Z. Wei, Z. Zhou, P. Wang, J. Ren, Y. Yin, G. F. Pedersen, and M. Shen, "Fully automated design method based on reinforcement learning and surrogate modeling for antenna array decoupling," *IEEE Trans. Antennas Propag.*, vol. 71, no. 1, pp. 660-671, Jan. 2023, doi: 10.1109/TAP.2022.3221613.

[24] C. Roy and K. Wu, "Surrogate model-based filter optimization by a field-circuit model mapping," *IEEE Trans. Microw. Theory Techn.*, vol. 72, no. 5, pp. 3144-3157, May 2024, doi: 10.1109/TMTT.2023.3318692.

[25] H. Kabir, Y. Wang, M. Yu, and Q. -J. Zhang, "Neural network inverse modeling and applications to microwave filter design," *IEEE Trans. Microw. Theory Techn.*, vol. 56, no. 4, pp. 867-879, April 2008, doi: 10.1109/TMTT.2008.919078.

[26] H. Kabir, Y. Wang, M. Yu, and Q. -J. Zhang, "High-dimensional neural-network technique and applications to microwave filter modeling," *IEEE Trans. Microw. Theory Techn.*, vol. 58, no. 1, pp. 145-156, Jan. 2010, doi: 10.1109/TMTT.2009.2036412.

[27] S. Koziel, N. Çalık, P. Mahouti, and M. A. Belen, "Low-cost and highly accurate behavioral modeling of antenna structures by means of knowledge-based domain-constrained deep learning surrogates," *IEEE Trans. Antennas Propag.*, vol. 71, no. 1, pp. 105-118, Jan. 2023, doi: 10.1109/TAP.2022.3216064.

[28] W. Liu, Q. J. Zhang, F. Feng, X. Wang, J. Xue, J. Zhang, K. Ma, "Electromagnetic Parametric Modeling Using MOR-Based Neuro-Impedance Matrix Transfer Functions for Two-Port Microwave Components," *IEEE Trans. Microw. Theory Techn.*, vol. 72, no. 5, pp. 2973-2989, May 2024, doi: 10.1109/TMTT.2023.3330766.

[29] Y. Ma, W. Wu, W. Yuan, J. Huang, X. Zhang, Z. Huang, and N. Yuan, "Multiparameter modeling technique based on TF-ANN for fss design," *IEEE Trans. Antennas Propag.*, vol. 71, no. 10, pp. 8384-8389, Oct. 2023, doi: 10.1109/TAP.2023.3291491.

[30] S. Koziel and M. Abdullah, "Machine-learning-powered EM-based framework for efficient and reliable design of low scattering metasurfaces," *IEEE Trans. Microw. Theory Techn.*, vol. 69, no. 4, pp. 2028-2041, Mar. 2021, doi: 10.1109/TMTT.2021.3061128.

[31] Y. He, J. Huang, W. Li, L. Zhang, S. Wong, and Z. Chen "Hybrid method of artificial neural network and simulated annealing algorithm for optimizing wideband patch antennas," *IEEE Trans. Antennas Propag.*, vol. 72, no. 1, pp. 944-949, Jan. 2024, doi: 10.1109/TAP.2023.3331249.

[32] Z. Zhou, Z. Wei, J. Ren, Y. Yin, G. F. Pedersen, and M. Shen, "Representation learning-driven fully automated framework for the inverse design of frequency-selective surfaces," *IEEE Trans. Microw. Theory Techn.*, vol. 71, no. 6, pp. 2409-2421, Jun. 2023, doi: 10.1109/TMTT.2023.3235066.

[33] J. Li, W. Qiu, P. Xiao, Z. Liu, G. Li, W. T. Joines, and S. Liao, "An adaptive evolutionary neural network-based optimization design method for wideband dual-polarized antennas," *IEEE Trans. Antennas Propag.*, vol. 71, no. 10, pp. 8165-8172, Oct. 2023, doi: 10.1109/TAP.2023.3301882.

[34] Z. Wei, Z. Zhou, P. Wang, J. Ren, Y. Yin, G. F. Pedersen, and M. Shen, "Automated antenna design via domain knowledge-informed reinforcement learning and imitation learning," *IEEE Trans. Antennas Propag.*, vol. 71, no. 7, pp. 5549-5557, Jul. 2023, doi: 10.1109/TAP.2023.3266051.

[35] Q. Wu, W. Chen, C. Yu, H. Wang, and W. Hong, "Machine learning-assisted optimization for antenna geometry design," *IEEE Trans. Antennas Propag.*, vol. 72, no. 3, pp. 2083-2095, Mar. 2024, doi: 10.1109/TAP.2023.3346493.

[36] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Int. Conf. Learn. Repr.*, Y. Bengio and Y. LeCun, Eds., 2015, pp. 1-15.

[37] P. Fei, Z. Shen, X. Wen, and F. Nian, "A Single-Layer Circular Polarizer Based on Hybrid Meander Line and Loop Configuration," *IEEE Trans. Antennas Propag.*, vol. 63, no. 10, pp. 4609-4614, Oct. 2015, doi: 10.1109/TAP.2015.2462128.

[38] Z. Wani, M. P. Abegaonkar, and S. K. Koul, "Thin planar metasurface lens for millimeter-wave MIMO applications," *IEEE Trans. Antennas Propag.*, vol. 70, no. 1, pp. 692-696, Jan. 2022, doi: 10.1109/TAP.2021.3098571.

[39] L. Yang, L. Zhu, R. Zhang, J. Wang, W. Choi, K. Tam, and R. Gómez-García, "Novel multilayered ultra-broadband bandpass filters on high-impedance slotline resonators," *IEEE Trans. Microw. Theory Techn.*, vol. 67, no. 1, pp. 129–139, Jan. 2019, doi: 10.1109/TMTT.2018.2873330.

Zhao Zhou was born in Shaanxi, China. He received the B.Sc. in electronic information engineering and M.Eng. degrees in electromagnetic field and microwave technology from Xidian University, Xi'an, China, in 2017 and 2020, respectively, and the Ph.D. degree in wireless communications from Aalborg University, Aalborg, Denmark, in 2024.

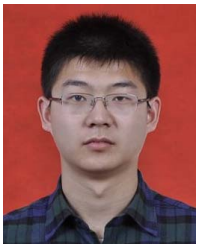
He conducted this research when he was a Postdoctoral Fellow at The Education University of Hong Kong from 2024 to 2025. He is currently a Postdoctoral Fellow at The Hong Kong Polytechnic University. His research interests include machine learning-driven electromagnetic structure design, optimization, and other applications.

Dr. Zhou has served as a reviewer for IEEE Transactions on Antennas and Propogation, IEEE Transactions on Circuits and Systems I: Regular Papers, IEEE Transactions on Artificial Intelligence, IEEE Antennas and Wireless Propogation Letters, and IEEE Microwave and Wireless Technology Letters.

Zhaohui Wei (Member, IEEE) was born in Shandong, China. He received the B.Sc. and M.Eng. degrees in electronic engineering from Xidian University, Xi'an, China, in 2017 and 2020, respectively, and the Ph.D. degree in wireless communications from Aalborg University, Aalborg, Denmark, in 2024.

He joined the School of Electronic Engineering, Xidian University, in 2024, where he is currently an Associate Professor. His research interests include filtering antennas, frequency selective surfaces, and deep learning-based methods for design and analysis of antenna systems.

Dr. Wei has served as a reviewer for IEEE Transactions on Antennas and Propagation, IEEE Transactions on Artificial Intelligence, and IEEE Antennas and Wireless Propagation Letters.



Jian Ren (Member, IEEE) was born in Shandong, China. He received the B.Sc. and M.Eng. degrees in electronic engineering from Xidian University, Xi'an, China, in 2012 and 2015, respectively, and the Ph.D. degree in electronic engineering from the City University of Hong Kong, Hong Kong, in 2018.

In 2019, he joined the National Key Laboratory of Antennas and Microwave Technology, Xidian University, where he is currently an Associate Professor. Since 2015, he has been a Research Assistant with the City University of Hong Kong. He has authored or co-authored over 40 refereed journal articles. His current research interests include dielectric resonator antenna, millimeter-wave antennas, and metamaterials.

Dr. Ren received the Honorable Mention at the Student Best Paper Competition at the 2018 IEEE 7th Asia–Pacific Conference on Antennas and Propagation (APCAP). He has served as a reviewer for different peer-reviewed journals, including the IEEE Transactions on Antennas and Propagation, the IEEE Antennas and Wireless Propagation Letters, the Journal of Physics D: Applied Physics, and the IET Microwaves, Antennas & Propagation.

Yingzeng Yin (Member, IEEE) received the B.S., M.S., and Ph.D. degrees in electromagnetic wave and microwave technology from Xidian University, Xi'an, China, in 1987, 1990, and 2002, respectively.

From 1990 to 1992, he was a Research Assistant and an Instructor with the Institute of Antennas and Electromagnetic Scattering, Xidian University, where he was an Associate Professor with the Department of Electromagnetic Engineering, from 1992 to 1996, and has been a Professor, since 2004. His current research interests include the design of microstrip antennas, feeds for parabolic reflectors, artificial magnetic conductors, phased array antennas, base-station antennas, and computer-aided design for antennas.

Yue Wang (Member, IEEE) received the B.S. degree in Electronics and M.E. in Information Science from Peking University, Beijing China, and the Ph.D. degree in Industrial Engineering from The Hong Kong University of Science and Technology. He is an Associate Professor in the Department of Mathematics and Information Technology, The Education University of Hong Kong. His research interests include machine learning, natural language processing, design and manufacturing informatics, healthcare informatics, and supply chain management.

Tse-Tin Chan (Member, IEEE) received the B.Eng. and Ph.D. degrees in Information Engineering from The Chinese University of Hong Kong (CUHK), Hong Kong, China, in 2014 and 2020, respectively.

From 2020 to 2022, he was an Assistant Professor with the Department of Computer Science, The Hang Seng University of Hong Kong (HSUHK), Hong Kong. He is currently an Assistant Professor with the Department of Mathematics and Information Technology, The Education University of Hong Kong (EdUHK), Hong Kong. His research interests include wireless communications and networking, the Internet of Things (IoT), age of information (AoI), and artificial intelligence (AI)-native wireless communications.