(Received 21 January 2025, Accepted 25 April 2025, Scheduled 6 May 2025)

Acceleration of Solving Volume Integral Equations through a Physics Driven Neural Network and Its Applications to Random Media Scattering

Jiayi Du[1], Yuanhao Cao[1], Chunzeng Luo[1], Gaoang Wang[1], and Shurun Tan[1, 2, 3, *]

[1] *Zhejiang University-University of Illinois Urbana-Champaign Institute, Zhejiang University, Haining 314400, China*
[2] *State Key Laboratory of Extreme Photonics and Instrumentation*
College of Information Science and Electronic Engineering, Zhejiang University, Hangzhou 310027, China
[3] *Department of Electrical and Computer Engineering, University of Illinois Urbana-Champaign, Urbana, IL 61801, USA*

ABSTRACT: In this paper, a novel framework is proposed which combines physical scattering models with artificial neural networks (ANNs) to solve electromagnetic scattering problems of random media through a volume integral equation formulation. The framework is applied to a snow scattering problem where snow is represented by a bicontinuous random medium. A neural network is constructed linking the random media structure to the induced dipole moments on the media. The volume integral equation (VIE) serves as a natural physical constraint on the network input-out relations and is used to guide the training of the network. A discrete dipole approximation (DDA) strategy is adopted to convert the VIE into matrix equations which also defines the loss function of the surrogate neural network. For addressing deterministic scattering problems, this represents a viable alternative to traditional iterative algorithms, providing comparable accuracy at the expense of reduced efficiency. In solving statistical scattering problems, neural networks with physics-informed loss function achieve accuracy comparable to that of data-driven models while significantly reducing the dependency on extensive precomputed training datasets. The physics-based loss function also allows the network to self-diagnose the prediction accuracy in real operations. This work demonstrates a novel strategy to effectively merge physical equations with artificial neural networks, and the idea can be inspiring to many relevant fields, especially when randomness effects are exhibited through a complicated nonlinear system.

1. INTRODUCTION

Electromagnetic modeling of wave scattering by natural random media is crucial for environmental remote sensing research [1–5]. Dry snow on land is composed of a densely packed ice particle medium against the background of air. Over the past 30 years, researchers have developed various microwave snow forward models based on different theoretical foundations [6]. Among them, numerical simulations of scattering in random media allow us to solve Maxwell's equations not limited by analytical approximations. Ding et al. (2010) [7] proposed a bicontinuous medium for simulating microwave scattering and emission in snow. This model is based on a continuous representation of interfaces between different phases or domains within the medium. Three-dimensional numerical Maxwell model (NMM3D) is used, based on a method of moments (MoM) with Discrete Dipole Approximation (DDA) to solve the electric field volume integral equation (VIE) on snow samples. This procedure is then combined with Monte Carlo simulations over random realizations of the random media in capturing the statistical properties in scattering to calculate scattering characteristics of snow in terms of its extinction coefficients, absorption coefficients, and scattering phase matrices [4].

In the field of computational electromagnetics, particularly within the scope of the MoM, discrete volume integral equations consistently result in a dense impedance matrix. Consequently, prevalent iterative algorithms, notably the Conjugate Gradient (CG) method [8] and Generalized Minimal Residual (GMRES) method [9], are conventionally employed for solving these linear matrix systems. However, a significant challenge arises due to the computational cost of matrix-vector multiplication, which escalates with the number of iterations, thereby imposing high demands on CPU and limiting the processing speed [10]. To address this issue, MoM derived matrix system is often solved using iterative techniques that incorporate acceleration methods such as through the Fast Fourier Transform (FFT) when translational symmetry is preserved in the resulting Toeplitz matrix. Nonetheless, the application of these acceleration methods may not always yield a satisfactory solution, as the impedance matrix can be poorly conditioned, leading to convergence issues in the iterative procedure. Even when convergence is achieved, the number of iterations required can be exceedingly large, rendering the overall solution inefficient. Therefore, it is crucial to explore alternative methods to replace traditional iterative techniques or strategies to reduce the number of iterations when dealing with integral equation computations. The urge is especially strong for random media scat-

* Corresponding author: Shurun Tan (srtan@intl.zju.edu.cn).

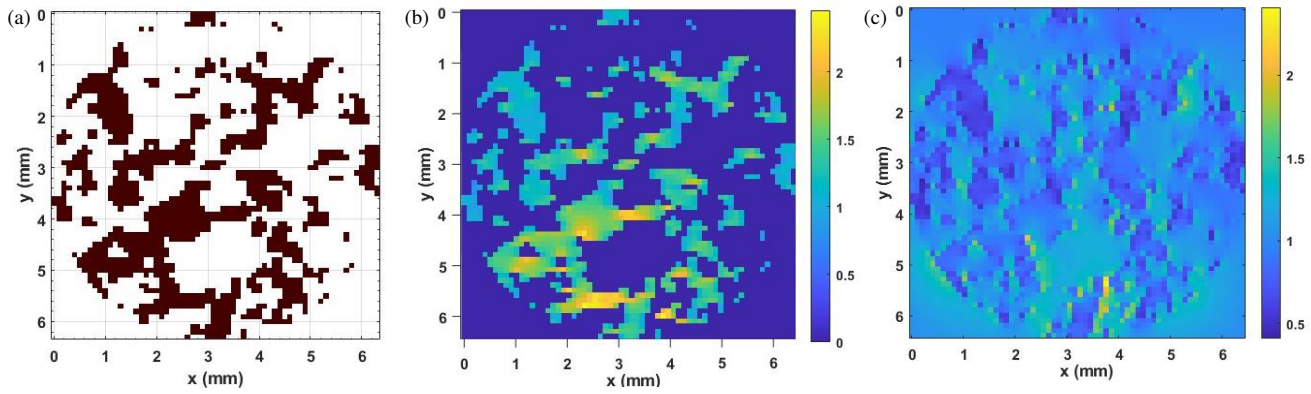

FIGURE 1. (a) Cross-section view (in xOy plane) of a bicontinuous random medium with typical parameters to represent the microstructure of dry snow. The sample is truncated into a spherical shape. (b) and (c) illustrate the induced dipole moments within the random medium sample and the corresponding near field distribution around the sample, respectively. The truncated random media sample is impinged by a plane wave propagating along $\hat{z}$ direction. Structure parameters: diameter of the truncating sphere $D = 64\,\text{mm}$, discretization resolution of the random media $d = 0.1\,\text{mm}$; bicontinuous media parameters: fractional volume $f_v = 0.3$, $b = 1.2$, $\langle\zeta\rangle = 9000\,\text{m}^{-1}$; wave frequency: 37 GHz.

tering as the time-consuming scattering calculations need to be repeated on each sample.

In electromagnetic calculation, numerical techniques, although highly accurate, often entail significant computational time. Consequently, they are ill-suited for real-time applications and optimization-based design tasks, as well as for Monte Carlo analyses that require repeated extensive full-wave simulations. Meanwhile, Deep Learning (DL) such as artificial neural network (ANN) has emerged as a promising alternative to effectively approximating the solutions to Maxwell's equations, resulting in more rapid and efficient simulations. These learning-based computational electromagnetic methods fundamentally involve the approximation of complex mappings through various machine learning techniques. These mappings are learned from a dataset during the model training process, thereby reducing the computational complexity compared to traditional computational electromagnetics methods when the trained network is applied for inference [11]. DL/ANN techniques have been successfully employed to address both forward [12] and inverse [13, 14] electromagnetic scattering problems, revolutionizing the field by providing more computationally efficient and scalable solutions.

In recent years, the adoption of Physics-Informed Neural Networks (PINNs) [15, 16] has gained notable recognition for their ability to tackle intricate computational problems intertwined with physical phenomena. This innovative framework represents a paradigm shift in addressing both forward and inverse problems associated with partial differential equations (PDEs), employing ANNs or Deep Neural Networks (DNNs) to approximate the latent solutions. Recent advancements in PINNs have demonstrated their potential in solving complex wave-based inverse problems. For instance, Jin et al. [17] proposed a Generalized Phase Retrieval Model (GPRM) that leverages DNNs combined with physical constraints to retrieve phase profiles in metasurfaces without requiring repeated retraining. Their approach highlights the advantage of embedding physical laws directly into neural network (NN) architectures for efficient and generalizable solutions. Over the past few years, PINNs have exhibited considerable promise within the realm of electromagnetics, yielding noteworthy achievements in various applications [18, 19]. While the initial intent of PINNs was to address PDEs, recently there are efforts in leveraging PINNs to solve integral equations (IEs) for electromagnetic scattering problems [20]. Empirical investigations have shown that the PINN approach can produce commendable outcomes compared to conventional methodologies like the MoM.

It is noteworthy that the work in [20] addresses the scattering of a fixed scatterer. There is a significant gap in handling scattering problems of scatterers with rich variability as encountered in random media scattering problems. In our proposed framework, we took a different approach to handle random media scattering. We are inspired by the fact that a random media scattering problem can be abstracted as a mapping between a random scattering structure and the resulting induced dipole distribution over the random structure or the local near fields over the random structure, c.f. Fig. 1. Therefore, we can construct a neural network linking the random media structure to the induced dipole patterns, c.f. Fig. 2. Once the dipole moments are predicted, all quantities related to scattering can be readily derived from the Maxwell's equations. Meanwhile, the VIE confining the dipole distributions can be taken as the loss function driving the training of the neural network. And the training will be performed over random realizations of the media with certain statistical characteristics. Once the network is trained, it can be applied to arbitrary configuration of the random media and predict the scattering solution efficiently in or near real time. It is a prominent feature that the VIE is only applied to test the accuracy of the network predicted solution rather than solving the dipole moments directly. Thus, the optimization of network weights replaces the traditional iterative solution of VIE. The training of the physics driven neural network also carries the potential to be more memory and computing efficient than a traditional data-driven approach where training data-pairs are to be pre-computed by solving VIEs with heavy computing burdens. Another potential advantage over traditional data-driven network approach is that the physics

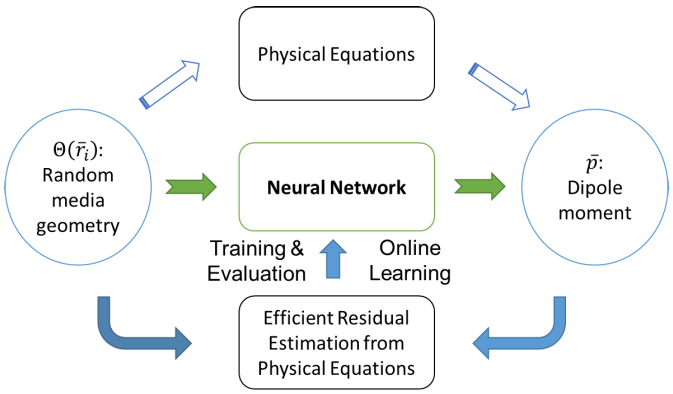

FIGURE 2. Schematic of the proposed VIE-NN framework: instead of solving the physical equations directly, the physical equations are treated as loss functions to confine the training of the neural network.

constrained loss function can be treated as indicators on the prediction accuracy of the network even in its inference phase. Further measures can be taken to refine the solution if the loss function raises an alarm, thus significantly improving the reliability of the network. This concept of prediction reliability enhancement has been recently demonstrated in [21] in a retrieval problem setting. The accuracy of the network prediction can be potentially further improved by cascading the network output to an iterative solver such as the GMRES algorithm, for which the required number of iterations to reach convergence will be substantially reduced and the overall efficiency be improved.

The structure of this paper is organized as follows. Section 2 presents the VIE-NN framework developed for solving random media scattering problems. Detailed strategies for efficiently training the network are discussed. The performance of the proposed physics-informed network is also compared with a traditional data-driven network approach to illustrate its features. In Section 3, the network performances applied to various validation sets with distinct characteristics are discussed, and the challenges associated with solving stochastic problems are addressed through comparison with the analysis of deterministic problems. Finally, conclusions are drawn in Section 4.

2. DDA-INFORMED NEURAL NETWORK FOR RANDOM MEDIA SCATTERING

2.1. The Bicontinuous Random Media Scattering Problem

In this subsection, we briefly review the bicontinuous random media scattering problem as encountered in the dense media radiative transfer (DMRT) formulation [7, 22]. The generation of the bicontinuous random media is briefly summarized in Appendix A [7]. In snow remote sensing, one can solve the DMRT equation to derive the backscatter and brightness temperature of a snowpack. There are several methods to solve the DMRT equation, such as the DISORT method [23] or iterative method [24]. The bicontinuous media — DMRT method characterizes snow scattering through its scattering phase matrix, scattering coefficient, absorption coefficient, and effective permittivity. These scattering characteristics are then fed into the radiative transfer equation and its boundary conditions to compute the microwave intensity solutions. The bicontinuous media — DDA approach characterizes the snow scattering characteristics through a Monte Carlo procedure, where multiple random media samples are generated with each confined in a spherical region as illustrated in Fig. 1(a), and then a scattering analysis is applied to each of the random sample to derive the statistical properties of scattering. In earlier development [7, 22], the electromagnetic scattering analysis is conducted through a VIE formulation with a discrete dipole approximation where a plane wave is impinged upon the random media sample to induce dipole moments on the media yielding secondary radiation known as scattering fields.

The volume integral equation on a uniform grid including N cubic elements as discretized by DDA is summarized in Eq. (1) [7].

$$\bar{q}_m = \bar{b}_m + \Theta\left(\bar{r}_m\right)\frac{k^2}{\varepsilon}\,\alpha \sum_{n=1,n\neq m}^{N} \bar{\bar{G}}\left(\bar{r}_m,\bar{r}_n\right)\cdot\bar{q}_n \qquad (1)$$

where ε is the permittivity of the background medium where the incident field propagates in; $\bar{q}_m = \frac{1}{\alpha}\bar{p}_m$ is the normalized version of the dipole moment $\bar{p}_m$ of the mth cube centered at $\bar{r}_m$; the normalization coefficient α is the unified polarizability of a discretized ice cube in the bicontinuous media; $\bar{b}_m = \Theta(\bar{r}_m)\bar{E}_{\text{inc}}(\bar{r}_m)$ is a masked version of the incident field $\bar{E}_{\text{inc}}(\bar{r}_m)$ with the masking template $\Theta(\bar{r}_m)$ being a binary indicator function that represents the bicontinuous media structure (c.f. Eq. (A3) of the Appendix A for its definition). $\bar{E}_{\text{inc}}(\bar{r}) = \hat{q}\exp(i\bar{k}_i\cdot\bar{r})$ represents a $\hat{q}$-polarized incident field propagating along $\hat{k}_i$ direction of unit amplitude where $\bar{k}_i = k\hat{k}_i$, k is the wavenumber in the background medium, and $\bar{\bar{G}}(\bar{r},\bar{r}')$ is the dyadic Green function linking a source pont $\bar{r}'$ to a field point $\bar{r}$. Note that Eq. (1) is a normalized version of Eq. (B2) in Appendix B by the ice cube polarizability α. This normalization is introduced to facilitate the neural network operation to be discussed in the following substraction. To simplify our notation in the following discussion, we further reorganize Eq. (1) into

$$\bar{\bar{A}}\bar{q} = \bar{b} \qquad (2)$$

where $\bar{\bar{A}}$ is of dimension $3N \times 3N$ whose 3×3 submatrix $\bar{\bar{A}}_{mn} = \bar{\bar{I}} - \Theta(\bar{r}_m)\frac{k^2}{\varepsilon}\alpha\bar{\bar{G}}(\bar{r}_m,\bar{r}_n)(1-\delta_{mn})$ in which $\bar{\bar{I}}$ is a 3×3 identity matrix, and δ_{mn} is the Kronnecker's Delta function that $\delta_{mn} = 1$ when $m = n$ and $\delta_{mn} = 0$ otherwise. $\bar{q} = \left[q_1^T, q_2^T, ..., q_m^T, ..., q_N^T\right]^T$ and $\bar{b} = \left[\bar{b}_1^T, \bar{b}_2^T, ..., \bar{b}_m^T, ..., \bar{b}_N^T\right]^T$ are of dimension $3N\times 1$ where the subscript T represents a transpose. The matrix $\bar{\bar{A}}$, vector $\bar{b}$, and $\bar{E}_{\text{inc}}$ are all functions of frequency f. Note in practice that the matrix $\bar{\bar{A}}$ will not be computed explicitly and stored to save memory, while the emphasis is on evaluating $\bar{\bar{A}}\bar{q}$ efficiently in the FFT based algorithms.

2.2. The DDA Informed NN Framework for Random Media Scattering

Instead of solving the DDA equation, Eq. (2), following the conventional linear algebra techniques, in this subsection we

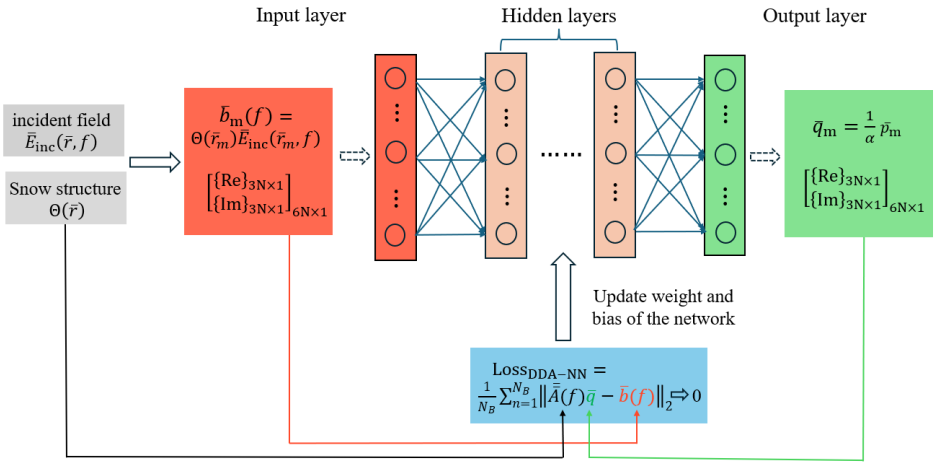

FIGURE 3. The DDA-NN framework. This framework integrates the DDA with a NN for efficient computation of electromagnetic scattering by snow structures. The input layer receives the incident field $\bar{E}_{\text{inc}}(\bar{r}, f)$ and snow structure information $\Theta(\bar{r})$, which are used to compute the vector $\bar{b}_m(f)$. The real and imaginary parts of $\bar{b}_m$ combined into a single vector of size $6N \times 1$ are input to the NN which propagates through hidden layers and produces the output vector $\bar{q}_m$. The loss function Loss_{DDA} uses the L2-norm to minimize the discrepancy between the predicted and expected outputs by updating the weights and biases of the network. The matrix $\bar{\bar{A}}$, the vector $\bar{b}$, and $\bar{E}_{\text{inc}}$ are all functions of frequency f.

present a novel model that directly approximates its solution linking the right-hand-side (RHS) $\bar{b}$ to the unknown $\bar{q}$ via an NN as illustrated in Fig. 3. Meanwhile, Eq. (3) is now utilized to build a loss function Loss_{DDA} averaged over a set of samples to train and confine the network,

$$\text{Loss}_{\text{DDA}} = \frac{1}{N_B}\sum_{n=1}^{N_B}\left\|\bar{\bar{A}}\bar{q} - \bar{b}\right\|_2 \tag{3}$$

where $\|\cdot\|_2$ represents the L2-norm of the complex error vector, and N_B represents the batch size. In computing the loss, FFT techniques can still be applied to accelerate the evaluation of the matrix vector multiplication $\bar{\bar{A}}\bar{q}$ as commonly adopted in the conventional DDA algorithm, taking advantage of the Toeplitz feature of the matrix. This DDA informed NN (DDA-NN) is a substantiation of the general idea as illustrated in Fig. 2. Note both the input to the network and the loss constraint of the network depend on the random media structure Θ. The normalization we adopted in Eq. (1) makes the input and output of the network regular on a scale comparable to 1.

In the following discussion, the network is chosen to be a fully connected feed-forward network with hidden layers. Since we are making use of a network structure that handles real scalar values, we have streamlined the input and output by separating the real part and imaginary part of the complex vectors, leading to the input and output layers of the network of dimension $6N \times 1$.

2.3. Network Tuning and Training

The problem of interest in this study involves the scattering behavior of random media, which is discretized into small voxels for numerical analysis. For this investigation, we consider bicontinuous media samples discretized into $5 \times 5 \times 5$ voxels, resulting in a total of 125 voxels with the structure parameter $\langle\zeta\rangle = 9000\,\text{m}^{-1}$, $b = 1.2$, and $f_v = 0.2$. The diameter of the truncating sphere $D = 2.5\,\text{mm}$, discretization resolution of the random media $d = 0.5\,\text{mm}$, and the simulation frequency is set at 37 GHz. The incident field $\bar{E}_{\text{inc}}(\bar{r})$ is chosen as a plane wave propagating along $\hat{z}$ axis. A pool of 10,000 random samples is generated for training and testing the network, with 8,000 samples used for training, 1,000 for testing, and 1,000 for validation. There is no overlapping among the three datasets.

To better understand the stochastic characteristics of the generated data, the distribution of ice-containing units across all 10,00 samples was analyzed and compared to the binomial probability mass function (PMF), as shown in Fig. 4. The binomial PMF describes the probability of having a specific number of successes (in this case, ice grains) in a fixed number of independent trials, given a fixed probability of success in each trial. In Fig. 4(a), the X-axis represents the number of ice grains per sample, while the Y-axis shows the probability of occurrence for each ice grain count. The histogram of the probability distribution for the number of ice grains across the 10,000 generated random media samples resembles the binomial PMF sharing the same peak value of occurances but with a broader distribution. This discrepancy arises because structural parameters, which are incorporated during the generation process, influence the spatial connectivity between voxels. These parameters introduce correlations between neighboring voxels, deviating from the independence assumption inherent in a binomial distribution. To further illustrate the spatial characteristics of the distribution, three representative ice grain counts (15, 25, and 35) were selected and visualized in 3D in Figs. 4(b)–(d), respectively.

2.3.1. Choosing Hyper-Parameters of the Network

The selection of model parameters is essential for optimizing network performance. Fig. 5 compares the influence on net-

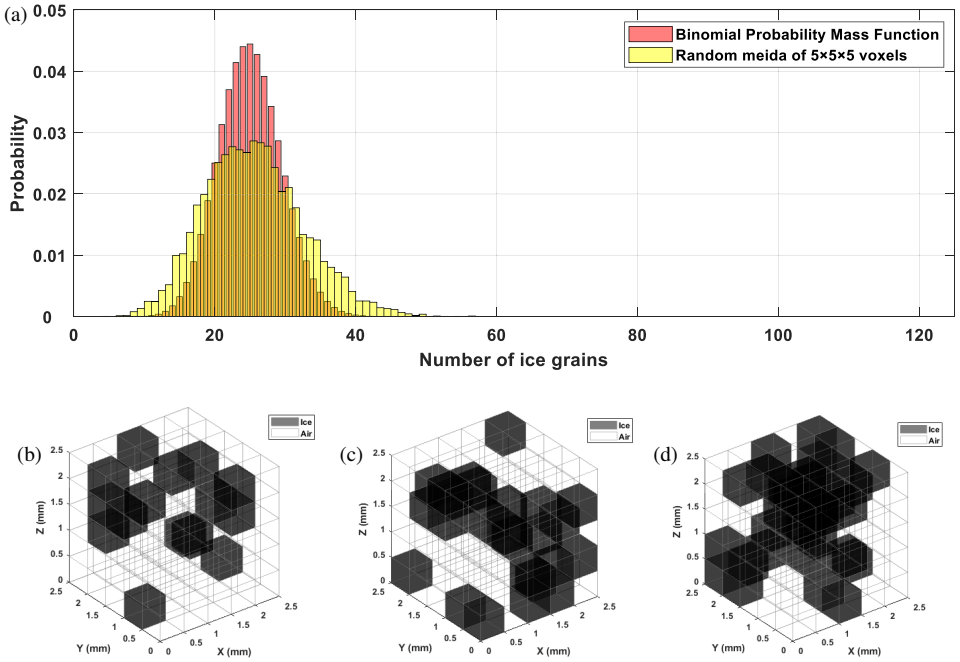

FIGURE 4. Comparison of the distribution of ice-containing units across 10,000 generated random media samples with the binomial PMF. Counting is performed on the $5 \times 5 \times 5$ voxels before truncating into a spherical shape. (a) Histogram of samples with varying numbers of ice grains. (b)–(d) 3D visualizations of random media containing 15, 25, and 35 ice grains, respectively. Black voxels represent ice, while other voxels represent air.

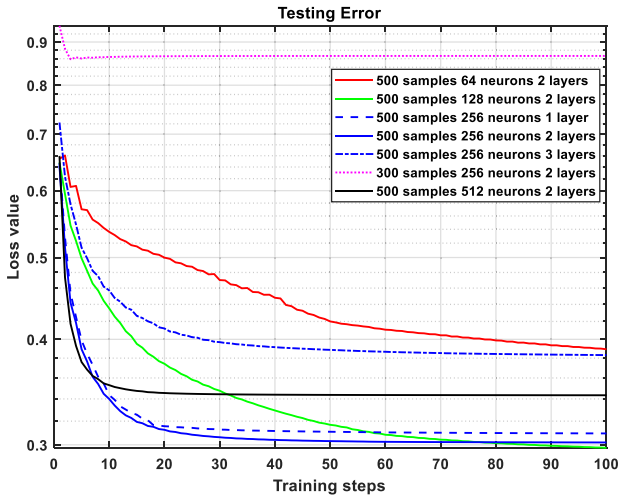

FIGURE 5. Influence of different network structures on testing results. These results are computed on $5 \times 5 \times 5$ bicontinuous media samples with $\langle \zeta \rangle = 9000\,\mathrm{m}^{-1}$, $b = 1.2$, $f_v = 0.2$. The diameter of the truncating sphere $D = 2.5\,\mathrm{mm}$, discretization resolution of the random media $d = 0.5\,\mathrm{mm}$. The simulation frequency is at 37 GHz. An incident plane wave propagating along $\hat{z}$ and polarized along $\hat{y}$ is applied. The results reported later also assume such parameters unless otherwise specified. The legend m samples n neurons and l layers represents the batch size, the number of neurons of each layer, the number of hidden layers, respectively. The y-axis represents the loss computed on the testing dataset. The x-axis denotes the training epoch. Comparison of testing losses show the superior performance when choosing a batch size 500, the number of neurons 256 on each layer and 2 hidden layers.

work convergence from several different network structures including the batch size, the number of hidden layers, and the number of neurons in each layer. Two fixed training sample sets, consisting of 500 and 300 samples each, were selected from the total training dataset of 8,000 samples. For each training step, the same 500 or 300 samples were used, with training conducted for 1,000 steps across different configurations of neurons and hidden layers. The configuration featuring 256 neurons per layer and two hidden layers exhibited the most favorable balance between convergence speed and testing loss. Although increasing the training batch size from 300 to 500 samples improves model performance on the testing dataset, it also reduces training efficiency. This trade-off highlights the need for strategies such as pre-training or adaptive training techniques to optimize both performance and efficiency across varying data distributions. In training the network, the limited-memory Broyden-Fletcher-Goldfarb-Shanno algorithm (L-BFGS) [25] is adopted to optimize the loss function.

2.3.2. Network Training

As observed in Fig. 5, increasing the batch size yielded limited improvements in convergence accuracy. Under the selected parameters (500 batch size, 256 neurons, 2 hidden layers), loss reduction stagnated after 20 training steps and approached a minimum value of 0.3 by 40 steps. To further improve convergence, a pre-training strategy combined with random sampling was implemented: the first 40 steps used a fixed batch of 500 samples, followed by 1,960 steps with random sampling, selecting 50 random samples from 8,000 training samples per batch, resulting in a total of 2,000 training steps. The L-BFGS optimization algorithm was configured with a maximum of 50

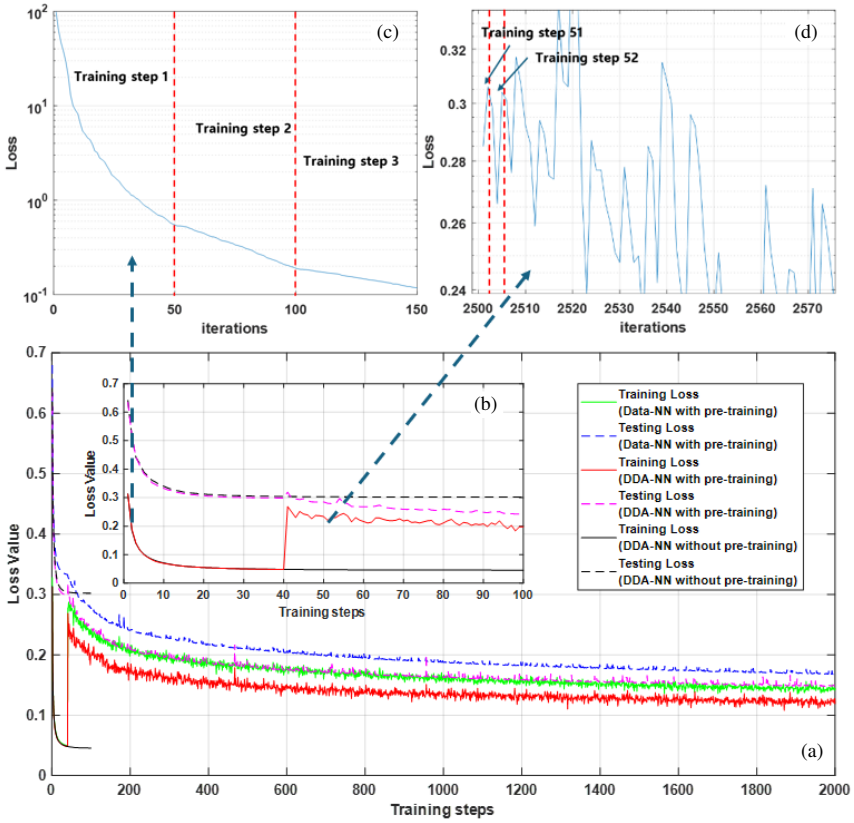

FIGURE 6. Training and testing performance with and without the pre-training strategy. The performance of DDA-NN and Data-NN is also compared. (a) Training and testing performance over 2,000 training steps for DDA-NN with and without pre-training, as well as for Data-NN with pre-training. (b) Comparison of training and testing performance during the first 100 training steps. (c) Training loss convergence behavior in the pre-training phase, where each training step includes a maximum of 50 iterations. (d) Training loss convergence behavior in the enhanced training phase, where each training step contains varying and fewer iterations indicating faster convergence.

internal iterations per training step and a optimitzation tolerance of 10^{-9}.

Figure 6 illustrates the training and testing loss curves with and without the pre-training strategy, respectively. Fig. 6(b) provides a zoomed-in view of the first 100 training steps to highlight the differences between the two training strategies of DDA-NN. During the 40 training steps, the training loss rapidly decreases to approximately 0.05, while the testing loss stabilizes around 0.3. Both training loss and testing loss stabilize and only marginally decrease in the following steps on the 500 fixed samples. When the enhanced training strategy is adopted after the pre-training phase, the introduction of new samples causes a temporary increase in training loss, but this loss decreases over time, though it may not return to the level observed during the pre-training phase. Notably, with the pre-training strategy, the testing loss for the DDA-NN model is significantly reduced to ~ 0.15 after 2,000 steps while training loss reaches ~ 0.12. This hybrid approach highlights a trade-off between training and testing losses. Although the training loss initially increases, the strategy ultimately leads to a decrease in testing loss, thereby improving the model's generalization capability and robustness. This strategy allows the model to fully digest a relatively larger consistent dataset before adapting to other patterns, enhancing the overall performance.

It is noted that the internal convergence behavior within each training step differed between the two training phases, as illustrated in Figs. 6(c)–(d). In the pre-training phase (Fig. 6(c)), each step utilized the maximum 50 iterations on the fixed 500 samples due to the abundance of data and the model's early-stage learning process. In contrast, during the enhanced training phase with random sampling (randomly 40 samples per training step) (Fig. 6(d)), steps often terminated early before the specified maximum number of iterations, with the smaller datasets satisfying the optimization tolerance more readily. Thus, incorporating both pre-training and enhanced training reduces the required sample size, accelerates convergence, and ultimately lowers the testing loss.

2.3.3. Comparison with a Data-Driven Neural Network Counterpart

To understand the features of the proposed physics informed NN, comparisons were made between the proposed network (DDA-NN) and a traditional data-driven neural network (Data-NN), as illustrated in Fig. 6. The Data-NN utilized pre-computed dipole moments derived from the iterative GMRES solver as ground truth, whose loss function is defined as:

$$\text{Loss}_{\text{Data}} = \frac{1}{N_B} \sum_{n=1}^{N_B} \|\bar{q} - \bar{q}'\|_2 \tag{4}$$

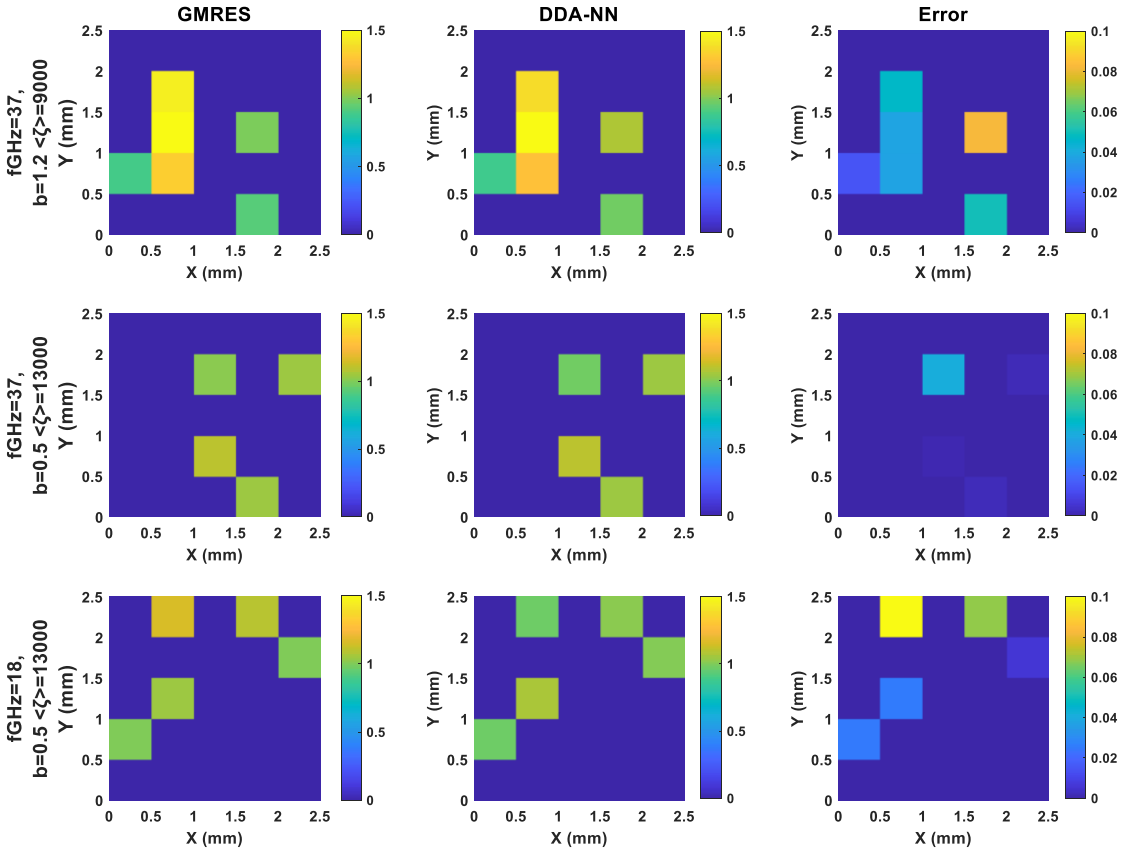

FIGURE 7. Results of $\bar{q}$ derived from DDA-NN. Each subplot represents a 5×5 data grid (the third layer of the voxel) for different scenarios. The subplot in each row corresponds to a different set of structural parameters. The subplots within each row show, from left to right, $\bar{q}'$ from GMRES (ground truth), $\bar{q}'$ estimated by DDA-NN, and the relative error between $\bar{q}$ and $\bar{q}'$, respectively.

where $\bar{q}$ is the normalized dipole moments predicted by NN, and $\bar{q}'$ is the ground truth calculated by GMRES. Both DDA-NN and Data-NN use the same network hyperparameters and the same sample space. The key difference between the two methods is that Data-NN requires precalculated dipole moment information to generate data pairs corresponding to the relevant structural information.

As shown in Fig. 6(a), the convergence behaviors of the Data-NN and DDA-NN networks are comparable, while the DDA-NN achieves a lower loss than the Data-NN. The DDA-NN and Data-NN models differ in both their loss functions and computational requirements. The loss function of DDA-NN is mathematically related to that of Data-NN, as the DDA-NN loss can be expressed as $\left\|\overline{\overline{A}}\left(\bar{q}-\bar{q}'\right)\right\|_2$ representing a transformation of the Data-NN loss. Numerical results indicate that the mapped DDA-NN loss achieves a lower value, suggesting improved performance in minimizing residuals and maintaining adherence to physical constraints. Additionally, one potential benefit of the DDA-NN lies in its adaptability across datasets with varying frequencies. Since $\overline{\overline{A}}$ is a function of both frequency f and snow structure $\Theta(\bar{r})$, the incorporation of these physical properties into the DDA-NN loss provides the model with the potential to generalize effectively to diverse application scenarios, including those with varying frequencies and snow structure parameters.

This difference in loss function computation also impacts training time. In Data-NN, the pre-computed $\bar{q}'$ enables rapid training, as the loss function evaluation is straightforward. Conversely, DDA-NN avoids pre-computation and dynamically calculates $\bar{q}'$ and the associated loss during each iterations. Over the entire training process, which includes about twenty thousand iterations across 8,000 total samples, individual samples are likely to be used multiple times. This leads to frequent recalculations of the loss, involving the heavy computation of the matrix vector multiplication $\overline{\overline{A}}\bar{q}'$, significantly increasing computational overhead compared to Data-NN in the training phase.

Compared to Data-NN, the key advantage of DDA-NN is its ability to self-check accuracy during predictions. By evaluating the loss function directly against physical equations, the model can dynamically validate its predictions, ensuring consistency with underlying physical principles. Additionally, DDA-NN eliminates the need for pre-computed $\bar{q}'$, reducing memory requirements and simplifying the data preparation process.

3. RESULTS AND DISCUSSION

In the previous section, the optimized DDA-NN structure and training strategy were established and compared with a Data-NN counterpart. In this section, the optimized DDA-NN is applied to the validation dataset, and the scattering characteristics

TABLE 1. RMSE comparison between $\bar{q}'$ and predicted $\bar{q}$ from two NNs for three scenarios.

Scenarios	RMSE (Data-NN)	RMSE (DDA-NN)
37 GHz, $b = 1.2$, $\langle\zeta\rangle = 9000\,\mathrm{m}^{-1}$ $D = 2.5\,\mathrm{mm}$, $d = 0.5\,\mathrm{mm}$	0.0675	0.0691
37 GHz, $b = 0.5$, $\langle\zeta\rangle = 13000\,\mathrm{m}^{-1}$ $D = 2.5\,\mathrm{mm}$, $d = 0.5\,\mathrm{mm}$	0.0662	0.0675
18 GHz, $b = 0.5$, $\langle\zeta\rangle = 13000\,\mathrm{m}^{-1}$ $D = 2.5\,\mathrm{mm}$, $d = 0.5\,\mathrm{mm}$	0.2219	0.2158

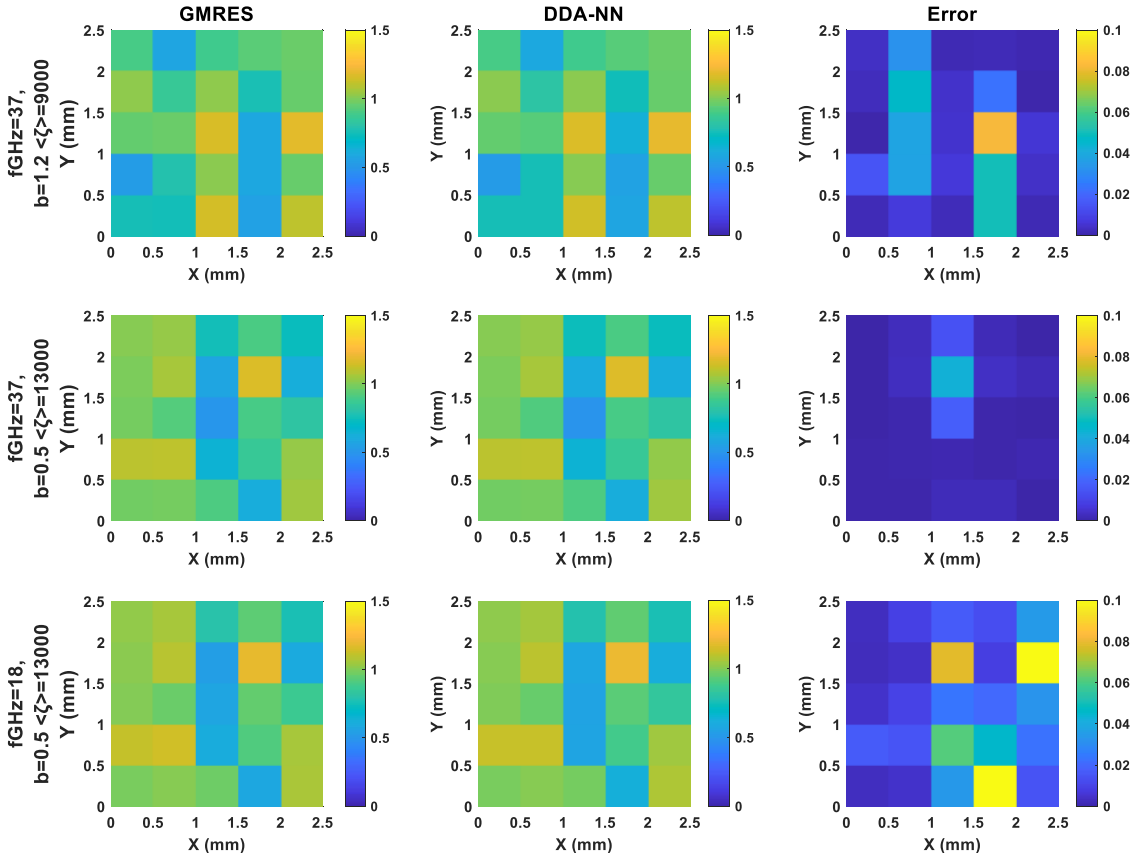

FIGURE 8. Results of near-field $\bar{E}$ derived from DDA-NN. Each subplot represents a 5×5 data grid (the third layer of the voxel) for different scenarios. The subplot in each row corresponds to a different set of structural parameters. The subplots within each row show, from left to right, $\bar{E}'$ from GMRES (ground truth), the $\bar{E}$ estimated by DDA-NN, and the relative error between $\bar{E}$ and the $\bar{E}'$, respectively.

of the bicontinuous random media are derived and illustrated. The proposed DDA-NN is applied to both stochastic and deterministic scattering problems.

3.1. Performance of NN Models for Stochastic Bicontinuous Random Media Scattering

We set three distinct scenarios to validate the performance of the NNs under different conditions. The first scenario uses a dataset which consists of 1,000 samples from the original dataset ($b = 1.2$, $\langle\zeta\rangle = 9000\,\mathrm{m}^{-1}$) as described in Section 2.3, serving as the baseline for comparison. The second scenario uses 1,000 samples generated with a modified structural parameter ($b = 0.5$, $\langle\zeta\rangle = 13000\,\mathrm{m}^{-1}$). The third scenario uses the same dataset of the second scenario but is simulated at a different frequency (18 GHz) than training the network (at 37 GHz). Results are presented across four critical aspects: the normalized dipole moment $\bar{q}$, near-field $\bar{E}$, scattering fields, and phase matrix, providing a comprehensive evaluation of the model's capability to handle diverse data inputs. The calculation of these parameters are discussed in Appendix B and C, respectively.

First, we evaluate the performance of the two NNs by predicting $\bar{q}$ for different scenarios and calculating the Root Mean Squared Error (RMSE) in $\bar{q}$ averaged over the 1,000 samples, as shown in Table 1. The results indicate that for the first two

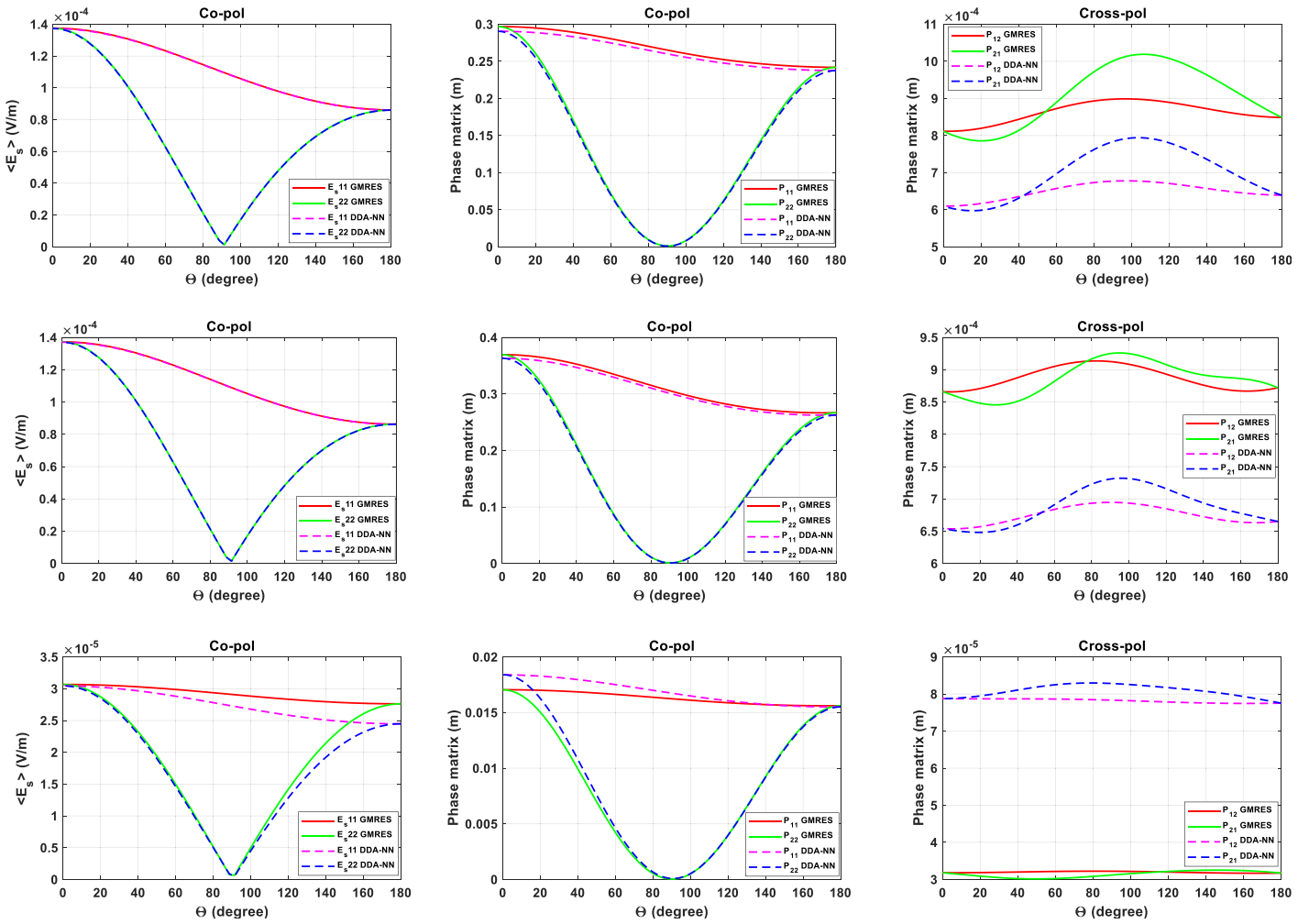

FIGURE 9. Results of the coherent scattering amplitude and phase matrix derived from DDA-NN. Each row represents a different set of structural parameters. The first row shows the validation set with $f = 37\,\text{GHz}$, $b = 1.2$, $\langle\zeta\rangle = 9000\,\text{m}^{-1}$; The middle row shows the validation set with $f = 37\,\text{GHz}$, $b = 0.5$, $\langle\zeta\rangle = 13000\,\text{m}^{-1}$; The third row shows the validation set with $f = 18\,\text{GHz}$, $b = 0.5$, $\langle\zeta\rangle = 13000\,\text{m}^{-1}$. The first column shows the co-polarization coherent scattering field, the second column shows the co-polarization phase matrix, and the third column shows the cross-polarization phase matrix.

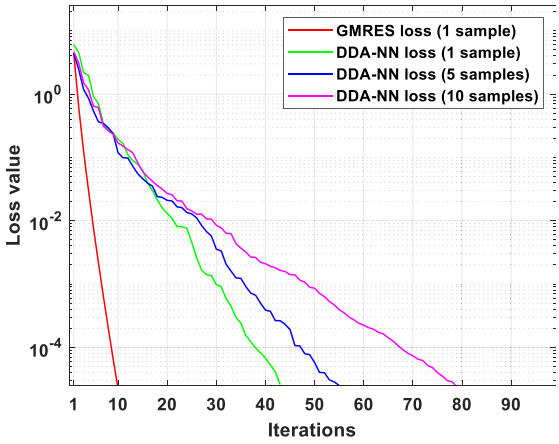

FIGURE 10. Convergence curves comparing the GMRES method applied to a single random media sample with the DDA-NN applied to both single and multiple random media samples.

scenarios (at 37 GHz), both networks demonstrate strong generalization ability across varying snow structure parameters. The RMSE of Data-NN is close to that of DDA-NN. However, when being applied to samples with a different frequency (18 GHz), the RMSE of both models increases with DDA-NN outperforming the Data-NN, suggesting an improved capability of DDA-NN to adapt to variation in frequencies.

We then took the samples from scenario 1 as an example to illustrate the $\bar{q}$ spatial distribution predicted by DDA-NN in a 2D plane. Specifically, we focus on a sample containing 15 ice crystals (counted after spherical truncation on the $5 \times 5 \times 5$ voxels), which corresponds to the most diverse random media configuration. The 2D plane corresponds to the middle plane of the random media voxels along the X and Y axes. Fig. 7 presents the visualization of $\bar{q}$ and compares it to the ground truth $\bar{q}'$. The first column shows the ground truth $\bar{q}'$ for different datasets, while the middle column displays the results estimated by DDA-NN. The last column highlights the relative error between $\bar{q}$ and $\bar{q}'$, i.e., $\frac{|\bar{q}-\bar{q}'|}{|\bar{q}|} * 100\%$. The relative errors are shown to be consistently below 0.1, demonstrating the DDA-NN's relative good performance. Fig. 8 shows the derived near-field $\bar{E}$ (following Eqs. (B5), (B6), (B7) in Ap-

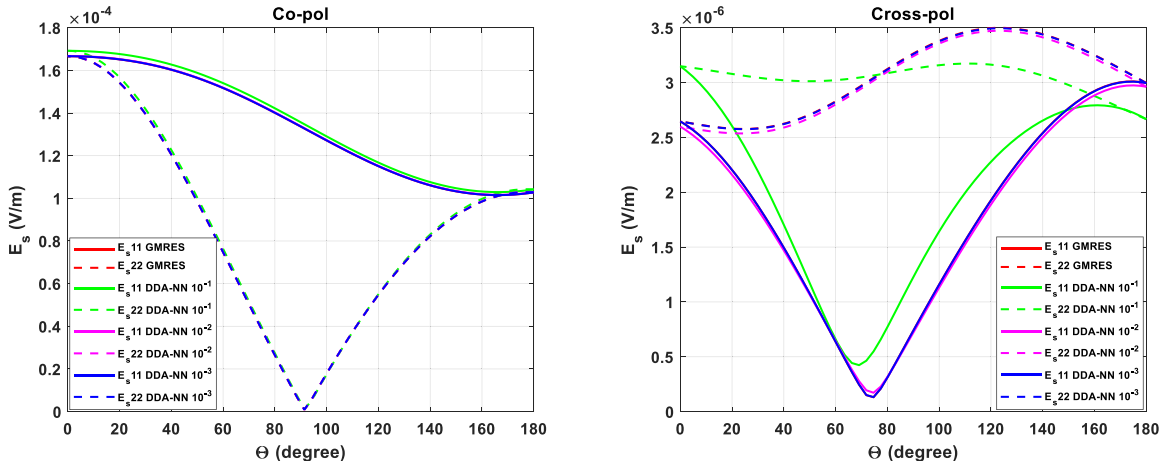

FIGURE 11. The far-field scattering amplitude for a single sample, comparing the results of GMRES and DDA-NN under different convergence precisions. GMRES results are computed with a convergence precision of 10^{-3}, while DDA-NN is evaluated at convergence precisions of 10^{-1}, 10^{-2}, and 10^{-3}, respectively.

pendix B) on the same plane, with relative errors also consistently below 0.1.

Next, we calculate the coherent scattering field for co-polarization (following Eq. (B8) in Appendix B) and the phase matrix for both co- and cross-polarizations (following Eq. (C4) in Appendix C) using the DDA-NN model across the three scenarios, as shown in Fig. 9. The coherent scattering field and phase matrix are linked to the first and 2nd order moments of the random media scattering characteristics. These statistics are computed over the 1000 samples for each scenario. The results show that as the structural parameters $\langle\zeta\rangle$ and b change, the co-polarization results align well with the ground truth for both the scattering field and phase matrix. In contrast, the performance for cross-polarization is less accurate when predicting scattering characteristics at a different frequency, and the consistency in co-polarization results for both the scattering field and phase matrix decreases, though the model still captures the overall data pattern.

The results indicate that the DDA-NN struggles to accurately predict cross-polarization results. This limitation stems from the non-satisfactory network testing accuracy with a loss value of approximately 0.14 (as shown in Fig. 6). A detailed discussion on the impact of the DDA-NN convergence behavior on its scattering prediction accuracy is provided in Section 3.2.

3.2. DDA-NN to Solve Deterministic Problems

The proposed DDA-NN can be tuned to solve both deterministic and stochastic scattering problems. In this subsection, we examine its features in solving deterministic problems given a specific random media structure sample Θ and an incident field $\bar{E}_{\text{inc}}(\bar{r})$, and compare the accuracy between DDA-NN and GMRES on scattering field calculation. In this situation, the network training procedure becomes a process of seeking the solution to the DDA equation. The hyper parameters of DDA-NN used here (256 neurons and 2 hidden layers) are consistent with those specified in Section 2.3. The deterministic problem is evaluated by comparing the performance of DDA-NN with the traditional GMRES iterative algorithm, using a $5 \times 5 \times 5$ voxel sample with parameters $b = 1.2$, $\langle\zeta\rangle = 9000\,\text{m}^{-1}$, at 37 GHz.

Figure 10 compares the loss values against iterations from the GMRES method when it is applied to a single random media sample with those from the DDA-NN when it is applied to both single and multiple random media samples. When solving a single sample, the DDA-NN requires significantly more iterations (43 iterations) than GMRES (10 iterations) to achieve the same convergence. This difference primarily arises from the different optimization targets in each method. GMRES optimizes the solution vector $\bar{q}'$ to minimize the residual $\overline{\overline{A}}\bar{q} - \bar{b}$, whereas in DDA-NN the L-BFGS algorithm optimizes the network weights and biases to update the prediction in $\bar{q}'$. When being extended to 5 and 10 random samples, the iterations required by DDA-NN further increase to 55 and 79, respectively. In Fig. 10, the loss functions adopted for GMRES and the DDA-NN are the same.

Figure 11 evaluated the DDA-NN model's accuracy under different convergence conditions by comparing its scattering filed results at convergence precisions of 10^{-1}, 10^{-2} and 10^{-3}, respectively, with the sufficiently converged GMRES benchmark (10^{-3}). The results indicate that co-polarization is more robust than cross-polarization to convergence errors, as it achieves reasonable accuracy even at a convergence level 10^{-1}, whereas cross-polarization requires a more rigid convergence level of 10^{-3} to align closely with the GMRES benchmark. This also explains why the trained model is more efficient in calculating co-polarization than cross-polarization when addressing stochastic scattering problems.

3.3. Computing Efficiency

Table 2 summarizes the CPU time required for solving both deterministic and stochastic problems using the proposed methods against the traditional GMRES methods as a comparison baseline. The CPU time is measured on a workstation with Intel Core i7-13700KF supporting 24 threads multi-threading. It is

TABLE 2. CPU time record for solving deterministic and stochastic problems: GMRES vs. DDA-NN and Data-NN.

Problem	Method	1 sample	100 samples	1000 samples	Training times
Deterministic	GMRES	0.036 s	0.51 s	4.95 s	
	DDA-NN	9.87 s	178.53 s	not measured	
Stochastic	DDA-NN	0.00154 s	0.00205 s	0.01278 s	101091 s
	Data-NN	0.00148 s	0.00223 s	0.01319 s	9895 s

noted that for solving a deterministic problem, the GMRES significantly outperforms the DDA-NN approach. This confirms the supremacy in convergence speed of the GMRES method over the network training approach. Meanwhile, it is also noted that the measured CPU time does not strictly linearly increase with the sample numbers, potentially due to the embedded multi-threading and computing optimization processes, such as shared computing overhead over multiple samples. After training, both DDA-NN and Data-NN exhibited around two orders of magnitudes efficiency improvement compared to GMRES in the inference stage (for stochastic problems) over 1,000 random media samples. Meanwhile, the training of the networks takes tremendous CPU resources as a one-time overhead. The DDA-NN approach takes about 10 more times in training the network compared to the Data-NN approach, consistent with the repeated matrix-vector multiplication evaluation procedures in the DDA-NN network training, sacrificing CPU time for flexibility. The training time was recorded for 2,000 training steps with the loss convergence curves being illustrated in Fig. 6.

4. CONCLUSIONS

This paper presents a novel framework that seamlessly integrates a physical scattering model with ANNs to solve electromagnetic scattering problems in random media through a VIE formulation. The proposed DDA-NN framework establishes a direct mapping between the random media structure and its dipole moment distribution. The loss function is derived from the DDA equation, ensuring that the network outputs adhere to Maxwell's equations.

A key advantage of the DDA-NN over a purely data-driven network (Data-NN) is that it eliminates the need for extensive pre-computed datasets. Unlike Data-NN, which requires prepared data pairs (linking the random media structure and the induced dipole distribution) as model input and output, DDA-NN only takes the random media structure as inputs and makes use of physical relations to tune the network for predicting the dipole moments. This not only reduces memory usage but also enables DDA-NN to self-check its prediction accuracy through its loss function, allowing the physical constraints to be checked even during inference. Additionally, the matrix $\overline{\overline{A}}$ involved in the loss function is a function of both frequency f and random media structure $\Theta(\bar{r})$, giving DDA-NN the ability to generalize across a wide range of scattering scenarios. Furthermore, the pre-training strategy adopted for DDA-NN demonstrates a trade-off between training and testing losses. This hybrid approach improves the model's generalization and robustness by allowing it to stabilize first on a representative and consistent dataset before adapting to more complex patterns.

When being applied to solve deterministic scattering problems, the proposed DDA-NN demonstrates lower efficiency than the traditional GMRES method. Yet it demonstrates a general framework to make nonlinear mapping incorporating physical constraints. When being applied to stochastic random media scattering, the trained DDA-NN model predicts $\bar{q}$ much faster than the GMRES method in its inference phase. The DDA-NN demonstrates better performance in predicting co-polarization scattering characteristics, including the coherent scattering amplitude and incoherent scattering phase matrix, compared to cross-polarization counterparts, linked to the limited convergence level when training the network.

The DDA-NN framework provides a promising paradigm for solving random media scattering problems combining physical constraints and artificial intelligence. The network predictions could be naturally fed into a traditional iterative solver as a well-positioned initial guess to further enhance the prediction accuracy with reduced efforts, a handy solution to improve the cross-polarization scattering prediction precision. In addition, the NN adapted in this paper is trained using real-valued data instead of complex-valued data, potentially limiting its ability to capture the intricate phase relationships inherent in the scattering properties. Future improvement of the model could incorporate complex-valued NNs. Furthermore, advanced NN architectures, such as convolutional neural networks (CNNs), could be investigated. With their ability to process large scale 3D data, CNNs better capture spatial correlations, potentially improving the model's capacity to handle scattering of electrical large and complicated random structures.

APPENDIX A. GENERATION OF THE BICONTINUOUS MEDIA

The numerical procedures for generating the bicontinuous random media are summarized in the Appendix following [7]. We define a zero mean random field S of position r by superimposing a large number of stochastic standing waves,

$$S(\bar{r}) = \frac{1}{\sqrt{N}} \sum_{n=1}^{N} \cos\left(\bar{\zeta}_n \cdot \bar{r} + \phi_n\right) \tag{A1}$$

where N is a sufficiently large number ($N = 10^4$ is used in this paper); ϕ_n is a random phase distributed uniformly between 0 and 2π; $\bar{\zeta}_n$ is a random wave vector $\bar{\zeta}_n = \zeta_n \hat{\zeta}_n$. The unit vector $\hat{\zeta}_n$ is uniformly distributed in all directions. The wavenumber

ζ_n follows a gamma distribution $p(\xi)$ as specified in Eq. (A2), whose mean value $\langle\zeta\rangle$ is inversely related to the snow grain size, while b is associated with the uniformity of the generated random structure.

$$p(\zeta)=\frac{b+1}{\Gamma(b+1)\langle\zeta\rangle}\left[(b+1)\frac{\zeta}{\langle\zeta\rangle}\right]^b\exp\left(-(b+1)\frac{\zeta}{\langle\zeta\rangle}\right) \quad \text{(A2)}$$

The bicontinuous random media structure $\Theta(\bar{r})$ is obtained by level cutting the random field $S(\bar{r})$,

$$\Theta_\alpha(S(\bar{r}))=\begin{cases}1 & S(\bar{r})>\alpha\\ 0 & \text{otherwise}\end{cases} \quad \text{(A3)}$$

where α is the cutting level related to the volume fraction f_v of the random media, c.f. Eq. (A4), where erfc$(\cdot)$ is the complementary error function.

$$f_v=\langle\Theta_\alpha(S(\bar{r}))\rangle=\frac{\text{erfc}(\alpha)}{2} \quad \text{(A4)}$$

APPENDIX B. THE VOLUME INTEGRAL EQUATION AND THE DISCRETE DIPOLE APPROXIMATION

The DDA equation, as derived from the volume integral equation,

$$\bar{E}(\bar{r})=\bar{E}_{\text{inc}}(\bar{r})+\frac{k^2}{\varepsilon}\int_V d\bar{r}'\overline{\overline{G}}(\bar{r},\bar{r}')\cdot[\varepsilon_p(\bar{r})-\varepsilon]\bar{E}(\bar{r}') \quad \text{(B1)}$$

is summarized in Eq. (B2) [7],

$$\bar{p}_m=\alpha_m\bar{E}_{\text{inc}}+\alpha_m\sum\nolimits_{n=1,n\neq m}^{N}\frac{k^2}{\varepsilon}\overline{\overline{G}}(\bar{r}_m,\bar{r}_n)\cdot\bar{p}_n \quad \text{(B2)}$$

where

$$\alpha_m=\frac{\alpha_m^c}{1-\frac{\alpha_m^c}{4\pi\varepsilon\Delta V}\left[\left(\frac{4\pi}{3}\right)^{\frac{1}{3}}(kd)^2+i\frac{2}{3}(kd)^3\right]} \quad \text{(B3)}$$

$$\alpha_m^c=3\varepsilon\Delta V\frac{\varepsilon_p(\bar{r}_m)-\varepsilon}{\varepsilon_p(\bar{r}_m)+2\varepsilon} \quad \text{(B4)}$$

in which $\Delta V=\frac{V}{N}=d^3$ is the equal volume of each cube, where V is the total region volume; N is the number of voxels; d is the voxel edge length. $\varepsilon_p(\bar{r}_m)$ represents the permittivity distribution.

In the air, the near-field $\bar{E}$ amounts to the sum of the incident field $\bar{E}_{\text{inc}}$ and the scattering field $\bar{E}_s$,

$$\bar{E}=\bar{E}_{\text{inc}}+\bar{E}_s \quad \text{(B5)}$$

where $\bar{E}_s$ at $\bar{r}_m$ in the air is linked to the dipole moments $\bar{p}$

$$\bar{E}_s(\bar{r}_m)=\sum\nolimits_{n=1,n\neq m}^{N}\frac{k^2}{\varepsilon}\overline{\overline{G}}(\bar{r}_m,\bar{r}_n)\cdot\bar{p}_n \quad \text{(B6)}$$

In ice particles, the near-field $\bar{E}$ is linked to the dipole moment $\bar{p}$ directly through

$$\bar{p}=\Delta V(\varepsilon_p-\varepsilon)\bar{E} \quad \text{(B7)}$$

The dipole moments $\bar{p}$ and near-fields $\bar{E}$ corresponding to a particular bicontinuous media sample as impinged by a plane incidence wave are illustrated in Figs. 1(b) and (c), respectively.

The scattering far-field $\bar{E}_s(\bar{r})$ is computed from Eq. (B8) through a far-field approximation of the dyadic Green's function [7],

$$\bar{E}_s(\bar{r})=\frac{\exp(ikr)}{4\pi r}\frac{k^2}{\varepsilon}\left(\hat{1}_s\hat{1}_s+\hat{2}_s\hat{2}_s\right)\cdot\sum\nolimits_n\exp\left(-i\bar{k}_s\cdot\bar{r}_n\right)\bar{p}_n \quad \text{(B8)}$$

where $\hat{1}_i=\hat{1}_s=\frac{\hat{k}_s\times\hat{k}_i}{|\hat{k}_s\times\hat{k}_i|}$, $\hat{2}_i=\hat{k}_i\times\hat{1}_i$, $\hat{2}_s=\hat{k}_s\times\hat{1}_s$ defines the 1–2 polarization system; $\hat{k}_i$ and $\hat{k}_s$ represent the propagation directions of the incident field and scattered field, respectively. The scattering field $\bar{E}_s(\bar{r})$ is linked to the scattering amplitude $f(\hat{k}_s,\hat{k}_i)$ through

$$\bar{E}_s(\bar{r})=\hat{e}_s f\left(\hat{k}_s,\hat{k}_i\right)E_0\frac{\exp(ikr)}{r} \quad \text{(B9)}$$

where E_0 is the complex amplitude of the incident field $\bar{E}_i(\bar{r})$

$$\bar{E}_i(\bar{r})=\hat{e}_i E_0\exp\left(ik\hat{k}_i\cdot\bar{r}\right) \quad \text{(B10)}$$

APPENDIX C. CALCULATION OF THE COHERENT SCATTERING FIELD AND THE PHASE MATRIX

The scattered field is averaged over random media realizations to obtain the coherent scattered wave [7],

$$\langle\bar{E}_s\rangle=\frac{1}{N_r}\sum\nolimits_{\tau=1}^{N_r}\bar{E}_s^\tau \quad \text{(C1)}$$

where τ is the index of realization with $\tau=1,2,...,N_r$, and $\bar{E}_s^\tau$ is the scattered field of the τth sample from the ensemble. The incoherent scattered field is taken as the fluctuation over the mean field,

$$\bar{\mathcal{E}}_s^\tau=\bar{E}_s^\tau-\langle\bar{E}_s\rangle \quad \text{(C2)}$$

The bistatic scattering cross section is linked to the averaged incoherent scattered wave intensity normalized by the incident wave intensity [7].

$$\sigma_{bi}\left(\hat{k}_s,\hat{k}_i\right)=\lim_{r\to\infty}\frac{r^2}{\left|\bar{E}^{\text{inc}}\right|^2}\left(\frac{1}{N_r}\sum\nolimits_{\tau=1}^{N_r}\left|\bar{\mathcal{E}}_s^\tau\right|^2\right) \quad \text{(C3)}$$

The phase matrix $P(\hat{k}_s,\hat{k}_i)$ is defined to be the averaged bistatic scattering cross section per unit volume [7].

$$P\left(\hat{k}_s,\hat{k}_i\right)=\frac{1}{V}\sigma_{bi}\left(\hat{k}_s,\hat{k}_i\right) \quad \text{(C4)}$$

Note that in random media scattering, the above stochastic scattering characteristics is physically meaningful when the random media sample is large enough to fully capture the multiple scattering effects within the sample. This in general requires the bicontinuous media to be on the order of a wavelength for snow scattering. This condition has not been fully met in the coherent scattering amplitude and incoherent phase matrix results presented in the paper due to the limited scale of the implemented network.

ACKNOWLEDGEMENT

This work was supported by the National Key Research and Development Program of China under Grant 2021YFB3900104. It was also supported by the startup funds from Zhejiang University. This work was led by the corresponding author and Principal Supervisor Dr. Shurun Tan.

REFERENCES

[1] Ishimaru, A., *Wave Propagation and Scattering in Random Media*, Vol. 1, Academic Press, New York, 1978.

[2] Tsang, L., J. A. Kong, and R. T. Shin, *Theory of Microwave Remote Sensing*, Wiley Interscience, New York, 1985.

[3] Tsang, L., J. A. Kong, and K.-H. Ding, *Scattering of Electromagnetic Waves*, Vol. 1, Wiley Interscience, New York, 2000.

[4] Tsang, L., J. A. Kong, K.-H. Ding, and C. O. Ao, *Scattering of Electromagnetic Waves*, Vol. 2, Wiley-Interscience, New York, 2001.

[5] Tang, L. and J. A. Kong, *Scattering of Electromagnetic Waves*, Vol. 3, Wiley-Interscience, New York, 2001.

[6] Shi, J., C. Xiong, and L. Jiang, "Review of snow water equivalent microwave remote sensing," *Science China Earth Sciences*, Vol. 59, 731–745, 2016.

[7] Ding, K.-H., X. Xu, and L. Tsang, "Electromagnetic scattering by bicontinuous random microstructures with discrete permittivities," *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 48, No. 8, 3139–3151, Aug. 2010.

[8] Saad, Y., *Iterative Methods for Sparse Linear Systems*, PWS Publishing Company, Boston, 1996.

[9] Saad, Y. and M. H. Schultz, "GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems," *SIAM Journal on Scientific and Statistical Computing*, Vol. 7, No. 3, 856–869, 1986.

[10] Xue, B.-W., R. Guo, M.-K. Li, S. Sun, and X.-M. Pan, "Deep-learning-equipped iterative solution of electromagnetic scattering from dielectric objects," *IEEE Transactions on Antennas and Propagation*, Vol. 71, No. 7, 5954–5966, 2023.

[11] Ma, Z., K. Xu, R. Song, C.-F. Wang, and X. Chen, "Learning-based fast electromagnetic scattering solver through generative adversarial network," *IEEE Transactions on Antennas and Propagation*, Vol. 69, No. 4, 2194–2208, 2021.

[12] Abdelrahman, M. A., A. Gupta, and W. A. Deabes, "A feature-based solution to forward problem in electrical capacitance tomography of conductive materials," *IEEE Transactions on Instrumentation and Measurement*, Vol. 60, No. 2, 430–441, Feb. 2011.

[13] Caorsi, S. and P. Gamba, "Electromagnetic detection of dielectric cylinders by a neural network approach," *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 37, No. 2, 820–827, Mar. 1999.

[14] Bermani, E., S. Caorsi, A. Massa, and M. Raffetto, "On the training patterns of a neural network for target localization in the spatial domain," *Microwave and Optical Technology Letters*, Vol. 28, No. 3, 207–209, 2001.

[15] Raissi, M. and G. E. Karniadakis, "Hidden physics models: Machine learning of nonlinear partial differential equations," *Journal of Computational Physics*, Vol. 357, 125–141, 2018.

[16] Raissi, M., P. Perdikaris, and G. E. Karniadakis, "Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations," *Journal of Computational Physics*, Vol. 378, 686–707, 2019.

[17] Jin, L., J. Xie, B. Pan, and G. Luo, "Generalized phase retrieval model based on physics-inspired network for holographic metasurface (Invited Paper)," *Progress In Electromagnetics Research*, Vol. 178, 103–110, 2023.

[18] Zhang, P., Y. Hu, Y. Jin, S. Deng, X. Wu, and J. Chen, "A Maxwell's equations based deep learning method for time domain electromagnetic simulations," *IEEE Journal on Multiscale and Multiphysics Computational Techniques*, Vol. 6, 35–40, 2021.

[19] Son, S., H. Lee, D. Jeong, K.-Y. Oh, and K. H. Sun, "A novel physics-informed neural network for modeling electromagnetism of a permanent magnet synchronous motor," *Advanced Engineering Informatics*, Vol. 57, 102035, 2023.

[20] Zhang, J.-B., D.-M. Yu, and X.-M. Pan, "Physics-informed neural networks for the solution of electromagnetic scattering by integral equations," in *2022 International Applied Computational Electromagnetics Society Symposium (ACES-China)*, 1–2, Xuzhou, China, 2022.

[21] Bai, X. and S. Tan, "Layered soil remote sensing with multichannel passive microwave observations using a physics-embedded artificial intelligence framework: A theoretical study," *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 61, 1–12, 2023.

[22] Xu, X., L. Tsang, and S. Yueh, "Electromagnetic models of co/cross polarization of bicontinuous/DMRT in radar remote sensing of terrestrial snow at X- and Ku-band for CoReH2O and SCLP applications," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, Vol. 5, No. 3, 1024–1032, Jun. 2012.

[23] Tsang, L., J. Pan, D. Liang, Z. Li, D. W. Cline, and Y. Tan, "Modeling active microwave remote sensing of snow using dense media radiative transfer (DMRT) theory with multiple-scattering effects," *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 45, No. 4, 990–1004, Apr. 2007.

[24] Tan, S., W. Chang, L. Tsang, J. Lemmetyinen, and M. Proksch, "Modeling both active and passive microwave remote sensing of snow using dense media radiative transfer (DMRT) theory with multiple scattering and backscattering enhancement," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, Vol. 8, No. 9, 4418–4430, Sep. 2015.

[25] Nocedal, J. and S. J. Wright, *Numerical Optimization*, Springer, 1999.