

Stacked Intelligent Metasurface for End-to-End OFDM System

Yida Zhang, Qiuyan Liu, Hongtao Luo, Yuqi Xia, Qiang Wang

Abstract—Stacked intelligent metasurface (SIM) and dual-polarized SIM (DPSIM) enabled wave-domain signal processing have emerged as promising research directions for offloading baseband digital processing tasks and efficiently simplifying transceiver design. However, existing architectures are limited to employing SIM (DPSIM) for a single communication function, such as precoding or combining. To further enhance the overall performance of SIM (DPSIM)-assisted systems and achieve end-to-end (E2E) joint optimization from the transmitted bitstream to the received bitstream, we propose an SIM (DPSIM)-assisted E2E orthogonal frequency-division multiplexing (OFDM) system, where traditional communication tasks such as modulation, precoding, combining, and demodulation are performed simultaneously during electromagnetic (EM) forward propagation. Furthermore, inspired by the idea of abstracting real metasurfaces as hidden layers of a neural network, we propose the electromagnetic neural network (EMNN) to enable the control of the E2E OFDM communication system. In addition, transfer learning is introduced into the model training, and a training and deployment framework for the EMNN is designed. Simulation results demonstrate that both SIM-assisted E2E OFDM systems and DPSIM-assisted E2E OFDM systems can achieve robust bitstream transmission under complex channel conditions. Our study highlights the application potential of EMNN and SIM (DPSIM)-assisted E2E OFDM systems in the design of next-generation transceivers.

Index Terms—Stacked intelligent metasurfaces (SIM), Dual-polarized stacked intelligent metasurfaces (DPSIM), Holographic multiple-input multiple-output (HMIMO), Orthogonal frequency-division multiplexing (OFDM), electromagnetic neural network (EMNN), Reconfigurable intelligent metasurface (RIS), End-to-end (E2E).

I. INTRODUCTION

A. Research Background.

Advanced transceiver design has emerged as a critical enabler for realizing ultra-high reliability, ultra-high-speed data transmission, and ubiquitous mobile connectivity in sixth-generation (6G) wireless networks [1], [2]. Over the past decades, multiple-input multiple-output (MIMO) antenna architectures have evolved substantially to provide higher spatial

This paper was partially funded by the National Key R & D Program of China (2020YFB1806602), BUPT-China Unicom Joint Innovation Center and Fundamental Research Funds for the Central Universities (2242022k60006). (*Corresponding author: Qiang Wang, Qiuyan Liu.*)

Yida Zhang, Hongtao Luo, Yuqi Xia, Qiang Wang are with the National Engineering Research Center for Mobile Network Technologies, Beijing University of Posts and Telecommunications, Beijing 100876, China (e-mail: {zhangyida02, mashirokaze1971, xiayuqi, wangq}@bupt.edu.cn)

Qiuyan Liu is with the China United Network Communications Corporation Research Institute, Beijing 100037, China (e-mail: liuqy95@chinaunicom.cn)

Simulation code and picture:

diversity and multiplexing gains [3]. In conventional fully-digital antenna arrays, baseband digital precoding and combining are employed to mitigate inter-antenna interference (IAI), thereby ensuring the independent transmission of multiple data streams [4]. However, such an approach requires a large number of radio frequency (RF) chains, resulting in excessive power consumption and hardware cost, which becomes prohibitive for resource-constrained terminals [5]. To alleviate this issue, a well-established solution is to adopt a hybrid digital-analog architecture, wherein the number of RF chains is reduced via a phase-shifting network, thereby aiming to achieve a favorable trade-off between performance and cost [6], [7].

Recently, reconfigurable intelligent surfaces (RISs) have emerged as a pivotal technology capable of directly manipulating electromagnetic (EM) waves in the wave domain, thereby potentially offloading a substantial portion of the signal processing tasks traditionally handled by RF chains [8]. Building upon this concept, a novel antenna architecture has been proposed in which the RIS is directly integrated into the transceiver [9], [10]. By precisely controlling the EM responses of the RIS elements, the transceiver can perform joint signal processing in both the digital and wave domains. This integrated architecture, referred to as holographic multiple-input multiple-output (HMIMO) [11], effectively reduces the number of required RF chains and typically employs low-resolution digital-to-analog converters (DACs) and analog-to-digital converters (ADCs), thereby significantly lowering hardware costs and improving energy efficiency [12].

However, due to the inherent single-layer structure of RISs, their flexibility in manipulating EM waves remains limited [13]. To further enhance wave-domain signal processing capabilities, [14], [15] proposed HMIMO systems assisted by stacked intelligent metasurfaces (SIMs), which significantly outperform conventional massive MIMO systems and RIS-assisted HMIMO systems. Specifically, a SIM is a three-dimensional metasurface device that leverages cascaded multilayer metasurfaces to perform analog computing in the EM wave domain [16]. In practice, SIMs are typically integrated into the RF front end to adjust the amplitude and phase of the transmitted or received signals, thereby replacing the precoding (or combining) modules in traditional architectures and ultimately simplifying transceiver design [17]. Furthermore, [18] proposed stacking dual-polarized RISs to enable independent signal processing in two orthogonal polarization directions, thereby further improving the signal processing capability within the constraints of limited integration space.

Along with the continuous evolution of hardware architec-

tures, the end-to-end (E2E) learning-based signal processing paradigm is gradually becoming a powerful tool for enhancing the full-chain performance of next-generation wireless systems [19]. Different from the traditional modular optimization approach, E2E learning jointly models and optimizes physical layer processes such as coding, modulation, beamforming, and detection by deploying neural networks at both the transmitter and receiver, thereby approaching globally optimal performance under given hardware constraints [20]. For RIS- or SIM-assisted HMIMO transceivers, E2E learning can adaptively configure parameters in both the digital and wave domains [21], [22]. This not only helps fully exploit the potential of metasurfaces but also maintains strong robustness under hardware impairments (e.g., phase noise, limited quantization dynamic range). Therefore, it is regarded as an important direction for achieving intelligent 6G transceiver design [23], [24].

B. Related Work.

SIM and DPSIM: Since the works in [14], [15] first proposed utilizing SIMs to accomplish signal processing tasks in the wave domain, a series of follow-up studies on SIMs have emerged. The authors of [25] investigated the application of SIMs in multi-user beamforming and achieved higher spectral efficiency and lower energy consumption compared with conventional MIMO systems. The study in [26] explored the potential of SIMs in wideband HMIMO systems, enabling interference-free multicarrier transmission under frequency-selective fading. In [27], a hybrid optimization method was proposed to achieve rate maximization and minimize inter-symbol interference (ISI), demonstrating that increasing the number of metasurface layers can improve the overall system performance. The work in [28] developed a multiport network model for SIMs based on Z-parameters to address issues such as unidirectional propagation and neglected mutual coupling in conventional simplified approaches. Furthermore, SIMs have been applied to various domains, including integrated sensing and communication (ISAC) [29] and semantic communications [30]. Research on dual-polarized SIMs (DPSIMs) is also thriving. [18] first proposed a DPSIM-assisted HMIMO system and demonstrated that, under the same integration space, the DPSIM-assisted system outperforms the SIM-assisted system in terms of performance. In [31], a DPSIM transceiver architecture combined with rate-splitting (RS-DPSIM) was proposed, where common and private information are transmitted simultaneously along two polarization directions. Simulation results show that this architecture offers performance advantages in anti-interference communications.

E2E learning of OFDM system: In recent years, extensive research has focused on exploring the performance of E2E learning in various communication systems. In [32], an autoencoder (AE) was introduced into orthogonal frequency-division multiplexing (OFDM) systems for the first time, assuming that each subcarrier is processed independently, thus enabling E2E communication for OFDM. However, this approach restricts the receiver's ability to exploit the time-frequency correlation of the OFDM channel. To address this limitation, [33]

designed and optimized a neural network receiver based on fully connected layers to jointly process multiple subcarriers, significantly improving system performance. The work in [34] investigated the impact of quantization bit resolution on E2E communication in OFDM systems and verified the robustness of learning-based methods in this scenario. Moreover, [35] studied E2E learning with OFDM waveforms over frequency-and time-selective fading channels, demonstrating the performance gains achieved through joint pilot and constellation learning. In addition, novel network architectures, such as the generative adversarial network (GAN) [23] and the deep fully convolutional neural network (DeepRx) [36], have also been applied to E2E communications to further enhance system performance.

C. Motivation and Contribution.

SIM and DPSIM are capable of performing wave-domain computations at the RF frontend, thereby enabling the offloading of baseband processing tasks. Owing to their outstanding energy efficiency and negligible latency, wave-domain signal processing is envisioned as a promising research direction to simplify transceiver design [15]. However, existing studies are limited to replacing a single module in the communication link (typically precoding or combining) with SIM or DPSIM, while the baseband still undertakes other signal processing tasks such as modulation and detection. Such modular optimization makes it difficult to achieve the optimal performance of the entire system [22]. In addition, due to the architectural transformation of the transceiver, how to fairly compare the performance of SIM (DPSIM)-assisted HMIMO systems with that of conventional massive MIMO systems remains an open problem [14], [26], [18]. To address the above challenges, inspired by the RIS-assisted E2E systems [21], we propose an SIM (DPSIM)-assisted E2E OFDM system and develop a complete training and deployment framework. The main contributions of this paper are summarized as follows:

- 1) We established the mathematical models of the SIM-assisted E2E OFDM systems and the DPSIM-assisted E2E OFDM systems. Different from existing studies on SIM (DPSIM), the proposed system aims to achieve joint optimization of the communication link from the transmitted bit stream to the received bit stream. By deploying deep neural networks (DNNs) and integrating SIM (DPSIM) devices, the systems enable complete offloading of signal processing tasks from the baseband digital domains to the wave domains.
- 2) To achieve holistic optimization of the E2E system performance, we further propose the electromagnetic neural network (EMNN) by abstracting the physical metasurface as hidden layers in the network. The EMNN associates the hidden-layer parameters with the EM (DPEM) units, thereby enabling model-driven control.
- 3) To accelerate the training of EMNN, we introduce transfer learning into the model training process, where the base network is trained using statistical channel state information (CSI), and then fine-tuned based on instantaneous CSI. Finally, we propose a training and deployment framework for EMNN.

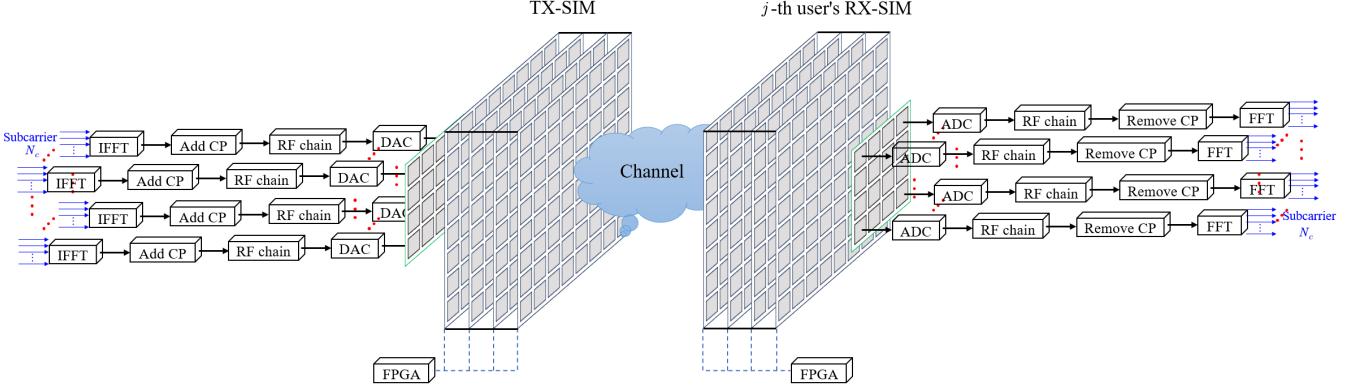


Fig. 1. Schematic diagram of the SIM-assisted E2E OFDM system.

- 4) The effectiveness of employing transfer learning to accelerate the training of EMNN is validated through simulations. In addition, a comprehensive performance evaluation of the SIM (DPSIM)-assisted E2E OFDM system is conducted, and its performance is compared with that of conventional massive MIMO systems.

D. Notations.

We adopt bold lowercase and uppercase letters to denote vectors and matrices, respectively; $(\cdot)^*$, $(\cdot)^T$, and $(\cdot)^H$ represent the conjugate, transpose, and Hermitian transpose, respectively; $|c|$, $\Re(c)$, and $\Im(c)$ refer to the magnitude, real part, and imaginary part, respectively, of a complex number c ; $\|\cdot\|_F$ is the Frobenius norm; $\|\cdot\|_1$ is the 1-norm; $\mathbb{E}(\cdot)$ stands for the expectation operation; \odot denotes the Hadamard product of matrices; $\text{diag}(\mathbf{v})$ produces a diagonal matrix with the elements of \mathbf{v} on the main diagonal; $\mathbb{C}^{x \times y}$ represents the space of $x \times y$ complex-valued matrices; $\mathbf{0}$ and $\mathbf{1}$ denote all-zero and all-one vectors, respectively, with appropriate dimensions, while $\mathbf{I}_N \in \mathbb{C}^{N \times N}$ denotes the identity matrix; the distribution of a circularly symmetric complex Gaussian (CSCG) random vector with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma} \succeq \mathbf{0}$ is denoted by $\sim \mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

E. Organization.

The remainder of this paper is organized as follows. Section II describes the system model considered in this work. Specifically, Sections II-A and II-B present the SIM-assisted E2E OFDM system and the single-polarization channel model, respectively, while Sections II-C and II-D introduce the DPSIM-assisted E2E OFDM system and the dual-polarization channel model, respectively. In Section III-A, the bit transmission problem for SIM (DPSIM)-assisted E2E OFDM systems is formulated. Sections III-B – Sections III-E detail the EMNN architecture, model training and deployment procedures, complexity analysis, and scalability analysis, respectively. The effectiveness of the proposed EMNN approach is validated via simulations in Section IV, followed by concluding remarks in Section V.

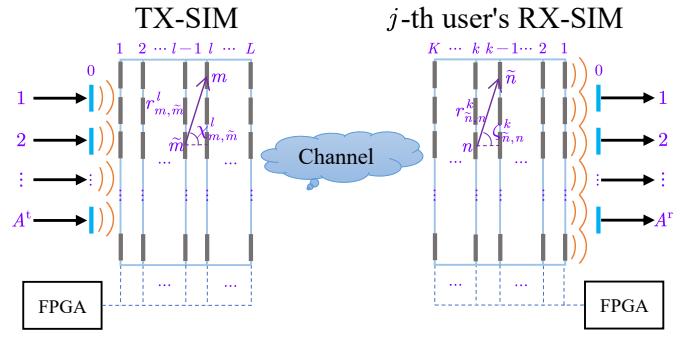


Fig. 2. TX-SIM/RX-SIM device parameter diagram.

II. SYSTEM MODEL

This section sequentially presents the SIM-assisted E2E OFDM system, the single-polarization channel model, the DPSIM-assisted E2E OFDM system, and the dual-polarization channel model.

A. SIM-assisted E2E OFDM System.

As shown in Fig. 1, we consider a J -user E2E OFDM system, where transmit SIM (TX-SIM) is integrated in the base station (BS) transmitter, and receive SIM (RX-SIM) is integrated in each user equipment (UE) receiver. To simplify the model, we assume that all UEs have the same number of receiving antennas and each integrates an identical RX-SIM device. Furthermore, all metasurface layers constituting a SIM device are identical. Let B and N^c denote the bandwidth and the number of subcarriers of the OFDM system, respectively. Moreover, as shown in Fig. 2, $A^t = A^{t^x} \times A^{t^y}$ and $A^r = A^{r^x} \times A^{r^y}$ respectively denote the number of TX antennas and the number of RX antennas per user. We further assume that TX-SIM comprises L layers of metasurfaces, with $M = M^x \times M^y$ EM units on each layer, and one RX-SIM is equipped with K layers of metasurfaces, with $N = N^x \times N^y$ EM units on each layer. For convenience of notation, we define the set of TX metasurfaces as $\mathcal{L} = \{0, 1, 2, \dots, L\}$ and the set of RX metasurfaces for the j -th user as $\mathcal{K}_j = \{0, 1, 2, \dots, K\}$, where $0 \in \mathcal{L}$ and $0 \in \mathcal{K}_j$ respectively represent the BS's TX antenna metasurface and the j -th user's RX antenna metasurface. The

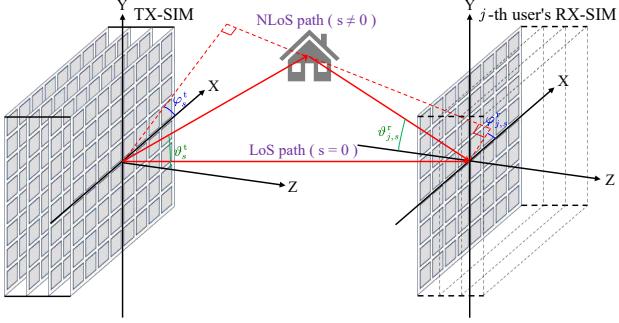


Fig. 3. Schematic diagram of single-polarization multipath channel model.

spacing between adjacent EM units in the TX-SIM and RX-SIM is denoted as d^t and d^r , respectively.

The transmission coefficient of the l -th layer for TX-SIM and the k -th layer for the j -th user's RX-SIM are defined as (1) and (2), respectively. According to the Rayleigh-Sommerfeld diffraction theory [37], the transmission coefficient from the \tilde{m} -th EM unit on the $(l-1)$ -th transmit metasurface layer to the m -th EM unit on the l -th transmit metasurface layer is expressed by (3), where $r_{m,\tilde{m}}^l$ denotes the corresponding transmission distance, $S^t = d^t \times d^t$ is the area of each EM unit in the TX-SIM, while $\chi_{m,\tilde{m}}^l$ represents the angle between the propagation direction and the normal direction of the $(l-1)$ -th transmit metasurface layer. f_i denotes the frequency of the i -th subcarrier, and c is the speed of light. Similarly, the transmission coefficient from the n -th EM unit on the k -th receive metasurface layer to the \tilde{n} -th EM unit on the $(k-1)$ -th receive metasurface layer is expressed by (4), where $r_{\tilde{n},n}^k$ denotes the corresponding transmission distance, $S^r = d^r \times d^r$ is the area of each EM unit in the RX-SIM, while $\zeta_{\tilde{n},n}^k$ represents the angle between the propagation direction and the normal direction of the k -th receive metasurface layer.

The cumulative effect generated by the layer-by-layer propagation of signals within SIM can be characterized by a chain of matrix products. For TX-SIM, the propagation coefficient matrix corresponding to the i -th subcarrier is

$$\mathbf{T}_i = \Phi^L \mathbf{V}_i^L \cdots \Phi^2 \mathbf{V}_i^2 \Phi^1 \mathbf{V}_i^1 \in \mathbb{C}^{M \times A^t}. \quad (5)$$

Similarly, for the RX-SIM of the j -th user, the propagation coefficient matrix corresponding to the i -th subcarrier is

$$\mathbf{R}_{i,j} = \mathbf{U}_i^1 \Psi_j^1 \mathbf{U}_i^2 \Psi_j^2 \cdots \mathbf{U}_i^K \Psi_j^K \in \mathbb{C}^{A^r \times N}. \quad (6)$$

In summary, the signal received by the j -th user on the i -th subcarrier is given by

$$\mathbf{y}_{i,j} = \mathbf{R}_{i,j} \mathbf{G}_{i,j} \mathbf{T}_i \mathbf{p}_i \odot \mathbf{x}_i + \mathbf{n}_{i,j}, \quad (7)$$

where $\mathbf{n}_{i,j} \in \mathbb{C}^{A^r \times 1}$ is the receiver noise vector with distribution $\mathcal{CN}(0, \sigma^2 \mathbf{I}_{A^r})$, $\mathbf{x}_i \in \mathbb{C}^{A^t \times 1}$ is the signal vector satisfying $\mathbb{E}\{\mathbf{x}_i \mathbf{x}_i^H\} = \mathbf{I}_{A^t}$, $\mathbf{p}_i \in \mathbb{C}^{A^t \times 1}$ denotes the vector of transmit powers, $\mathbf{G}_{i,j}$ is the channel coefficient matrix between the L -th layer metasurface of the TX-SIM and the K -th layer metasurface of the j -th user's RX-SIM.

Let N^{bit} denote the number of bits corresponding to an OFDM symbol transmitted by the BS, and N_j^{bit} represent the number of bits received by the j -th user within one OFDM symbol. $\mathbf{b} \in \{0, 1\}^{N^{\text{bit}} \times 1}$ and $\hat{\mathbf{b}}_j \in \{0, 1\}^{N_j^{\text{bit}} \times 1}$ denote the transmitted bit vector and the received bit vector, respectively. Then, the SIM-assisted E2E OFDM system model can be expressed as

$$\hat{\mathbf{b}}_j = \mathbf{f}_j^D \left(\begin{bmatrix} \mathbf{R}_{1,j} \mathbf{G}_{1,j} \mathbf{T}_1 \mathbf{p}_1 \odot \mathbf{f}_1^M(\mathbf{b}) + \mathbf{n}_{1,j} \\ \mathbf{R}_{2,j} \mathbf{G}_{2,j} \mathbf{T}_2 \mathbf{p}_2 \odot \mathbf{f}_2^M(\mathbf{b}) + \mathbf{n}_{2,j} \\ \vdots \\ \mathbf{R}_{N^c,j} \mathbf{G}_{N^c,j} \mathbf{T}_{N^c} \mathbf{p}_{N^c} \odot \mathbf{f}_{N^c}^M(\mathbf{b}) + \mathbf{n}_{N^c,j} \end{bmatrix} \right), \quad (8)$$

where $\mathbf{x}_i = \mathbf{f}_i^M(\mathbf{b})$, $i = 1, 2, \dots, N^c$ denotes the modulation mapping from \mathbf{b} to the signal on the i -th subcarrier. Furthermore, $\hat{\mathbf{b}}_j = \mathbf{f}_j^D \left([\mathbf{y}_{1,j}, \mathbf{y}_{2,j}, \dots, \mathbf{y}_{N^c,j}]^T \right)$, $j = 1, 2, \dots, J$ is the demodulation mapping of the received wideband signal at the j -th user to $\hat{\mathbf{b}}_j$.

B. Unpolarized Channel Model (SIM-assisted E2E OFDM System Channel Model)

As shown in Fig. 3, to precisely model the spatial and frequency-selective properties inherent in wideband propagation, the proposed SIM-assisted E2E OFDM system utilizes a geometric multi-path channel model. Assume there are S scatterers between the BS and UEs, each of which affects the signal propagation path. Let $s = 0$ denote the line-of-sight (LoS) path and $s = 1, 2, \dots, S$ denote non-LoS (NLoS) paths. We assume that the gain and delay of the i -th subcarrier from the BS to the j -th user through the s -th path are $g_{i,j,s}$ and $\tau_{i,j,s}$, respectively. Although the transmission signals between TX and RX are time-domain signals, the wideband channel model is often represented in the frequency-domain form to

$$\Phi^l = \text{diag} \left(e^{j\theta_1^l}, e^{j\theta_2^l}, \dots, e^{j\theta_M^l} \right) \in \mathbb{C}^{M \times M}, \theta_m^l \in [0, 2\pi), m \in \mathcal{M}, l \in \mathcal{L}, \quad (1)$$

$$\Psi_j^k = \text{diag} \left(e^{j\xi_1^{j,k}}, e^{j\xi_2^{j,k}}, \dots, e^{j\xi_N^{j,k}} \right) \in \mathbb{C}^{N \times N}, \xi_n^{j,k} \in [0, 2\pi), n \in \mathcal{N}, k \in \mathcal{K}, 1 \leq j \leq J. \quad (2)$$

$$[\mathbf{V}_i^l]_{m,\tilde{m}} = \frac{S^t \cos \chi_{m,\tilde{m}}^l}{r_{m,\tilde{m}}^l} \left(\frac{1}{2\pi r_{m,\tilde{m}}^l} - j \frac{f_i}{c} \right) e^{j2\pi r_{m,\tilde{m}}^l f_i / c}, m \in \mathcal{M}, \tilde{m} \in \mathcal{M}, l \in \mathcal{L}, 1 \leq i \leq N^c, \quad (3)$$

$$[\mathbf{U}_i^k]_{\tilde{n},n} = \frac{S^r \cos \zeta_{\tilde{n},n}^k}{r_{\tilde{n},n}^k} \left(\frac{1}{2\pi r_{\tilde{n},n}^k} - j \frac{f_i}{c} \right) e^{j2\pi r_{\tilde{n},n}^k f_i / c}, \tilde{n} \in \mathcal{N}, n \in \mathcal{N}, k \in \mathcal{K}, 1 \leq i \leq N^c. \quad (4)$$

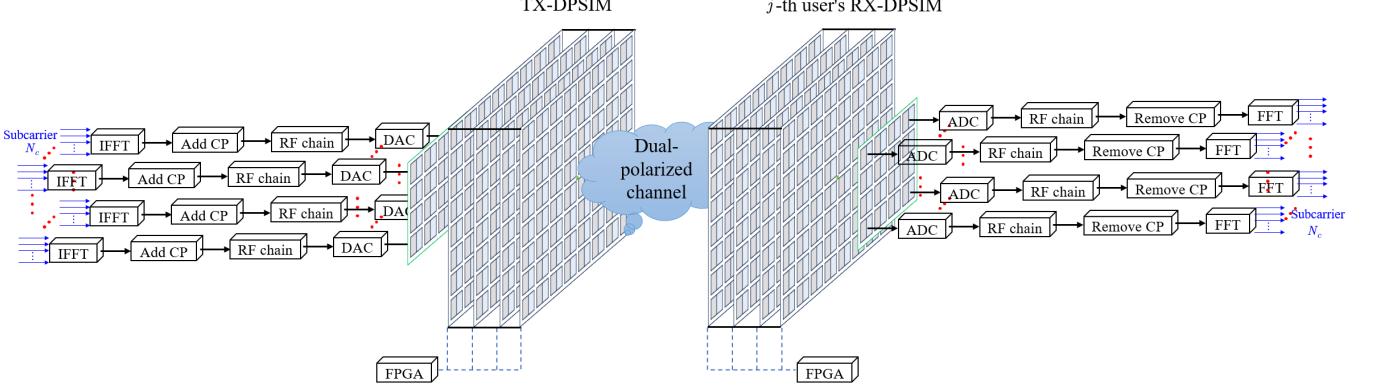


Fig. 4. Schematic diagram of the DPSIM-assisted E2E OFDM system.

simplify the IFFT-FFT progress in the beamspace [38]. For the \$i\$-th subcarrier with frequency \$f_i\$, the channel coefficient matrix \$\mathbf{G}_{i,j} \in \mathbb{C}^{N \times M}\$ is modeled as

$$\mathbf{G}_{i,j} = \sum_{s=0}^S g_{i,j,s} e^{-j2\pi f_i \tau_{i,j,s}} \boldsymbol{\alpha}_{i,j,s}^t (\boldsymbol{\alpha}_{i,s}^t)^H, \quad (9)$$

where \$\boldsymbol{\alpha}_{i,s}^t \in \mathbb{C}^{M \times 1}\$ and \$\boldsymbol{\alpha}_{i,j,s}^r \in \mathbb{C}^{N \times 1}\$ are the corresponding transmit and receive steering vectors, respectively.

Without loss of generality, we assume that the TX-SIM and RX-SIM are vertically aligned in a 3D Cartesian coordinate system. For the \$s\$-th propagation path, its elevation and azimuth angles relative to the TX-SIM are \$\vartheta_s^t \in [0, \pi/2)\$ and \$\varphi_s^t \in [0, 2\pi)\$, respectively, while those relative to the \$j\$-th user are \$\vartheta_{j,s}^r \in [0, \pi/2)\$ and \$\varphi_{j,s}^r \in [0, 2\pi)\$.

The steering vector \$\boldsymbol{\alpha}_{i,s}^t\$ and \$\boldsymbol{\alpha}_{i,j,s}^r\$ can be written as (10) and (11), respectively, where \$\boldsymbol{\alpha}_s^{t,x} (*) \in \mathbb{C}^{M^x \times 1}\$ and \$\boldsymbol{\alpha}_s^{t,y} (*) \in \mathbb{C}^{M^y \times 1}\$ are defined as follows

$$[\boldsymbol{\alpha}_s^{t,x} (*)]_{m^x} \triangleq e^{j(m^x-1)*}, m^x = 1, 2, \dots, M^x, \quad (12)$$

$$[\boldsymbol{\alpha}_s^{t,y} (*)]_{m^y} \triangleq e^{j(m^y-1)*}, m^y = 1, 2, \dots, M^y, \quad (13)$$

where \$\boldsymbol{\alpha}_s^{t,x} (*) \in \mathbb{C}^{N^x \times 1}\$ and \$\boldsymbol{\alpha}_s^{t,y} (*) \in \mathbb{C}^{N^y \times 1}\$ are defined as

$$[\boldsymbol{\alpha}_{j,s}^{t,x} (*)]_{n^x} \triangleq e^{j(n^x-1)*}, n^x = 1, 2, \dots, N^x, \quad (14)$$

$$[\boldsymbol{\alpha}_{j,s}^{t,y} (*)]_{n^y} \triangleq e^{j(n^y-1)*}, n^y = 1, 2, \dots, N^y. \quad (15)$$

C. DPSIM-assisted E2E OFDM system.

Similar to Fig. 1, Fig. 4 illustrates a DPSIM-assisted E2E OFDM system, where the TX-DPSIM and RX-DPSIM are integrated at the transmitter and receiver sides, respectively, to process the signal in the wave domain. Considering that each layer of the metasurface in DPSIM can independently control the phase for different polarization directions, the DPSIM can thus be equivalently regarded as two identical and isolated single-polarization SIMs. To facilitate understanding the model, Fig. 5 shows a schematic diagram of the HMIMO system in the two orthogonal polarization directions. We assume that \$p \in \{0, 1\}\$ denotes two orthogonal polarization directions. Consistent with the assumptions in Section II-A, each UE integrates an identical DPSIM device, and each metasurface layer constituting the DPSIM is perfectly identical. Moreover, \$A^{dp,t} = A^{dp,t^x} \times A^{dp,t^y}\$ and \$A^{dp,r} = A^{dp,r^x} \times A^{dp,r^y}\$ respectively denote the number of TX dual-polarized antennas and the number of RX dual-polarized antennas per user. We further assume that TX-DPSIM comprises \$L^{dp}\$ layers of metasurfaces, with \$M^{dp} = M^{dp,x} \times M^{dp,y}\$ DPEM units on each layer, and one RX-SIM is equipped with \$K^{dp}\$ layers of metasurfaces, with \$N^{dp} = N^{dp,x} \times N^{dp,y}\$ DPEM units on each layer. For convenience of notation, we define the set of TX metasurfaces as \$\mathcal{L}^{dp} = \{0, 1, 2, \dots, L^{dp}\}\$ and the set of RX metasurfaces for the \$j\$-th user as \$\mathcal{K}_j^{dp} = \{0, 1, 2, \dots, K^{dp}\}\$, where \$0 \in \mathcal{L}^{dp}\$ and \$0 \in \mathcal{K}_j^{dp}\$ respectively represent the BS's TX antenna metasurface and the \$j\$-th user's RX antenna metasurface. The spacing between adjacent DPEM units in

$$\boldsymbol{\alpha}_{i,s}^t = \boldsymbol{\alpha}_s^{t,x} \left(\frac{2\pi c}{f_i} d^t \sin(\vartheta_s^t) \sin(\varphi_s^t) \right) \otimes \boldsymbol{\alpha}_s^{t,y} \left(\frac{2\pi c}{f_i} d^t \cos(\vartheta_s^t) \right), \quad (10)$$

$$\boldsymbol{\alpha}_{i,j,s}^r = \boldsymbol{\alpha}_{j,s}^{t,x} \left(\frac{2\pi c}{f_i} d^r \sin(\vartheta_{j,s}^r) \sin(\varphi_{j,s}^r) \right) \otimes \boldsymbol{\alpha}_{j,s}^{t,y} \left(\frac{2\pi c}{f_i} d^r \cos(\vartheta_{j,s}^r) \right), \quad (11)$$

$$\boldsymbol{\Phi}_p^l = \text{diag} \left(e^{j\theta_{p,1}^l}, e^{j\theta_{p,2}^l}, \dots, e^{j\theta_{p,M}^l} \right) \in \mathbb{C}^{M^{dp} \times M^{dp}}, \theta_{p,m}^l \in [0, 2\pi], m \in \mathcal{M}^{dp}, l \in \mathcal{L}^{dp}, p \in \{0, 1\}, \quad (16)$$

$$\boldsymbol{\Psi}_{j,p}^k = \text{diag} \left(e^{j\xi_{p,1}^{j,k}}, e^{j\xi_{p,2}^{j,k}}, \dots, e^{j\xi_{p,N}^{j,k}} \right) \in \mathbb{C}^{N^{dp} \times N^{dp}}, \xi_{p,n}^{j,k} \in [0, 2\pi], n \in \mathcal{N}^{dp}, k \in \mathcal{K}^{dp}, p \in \{0, 1\}, 1 \leq j \leq J. \quad (17)$$

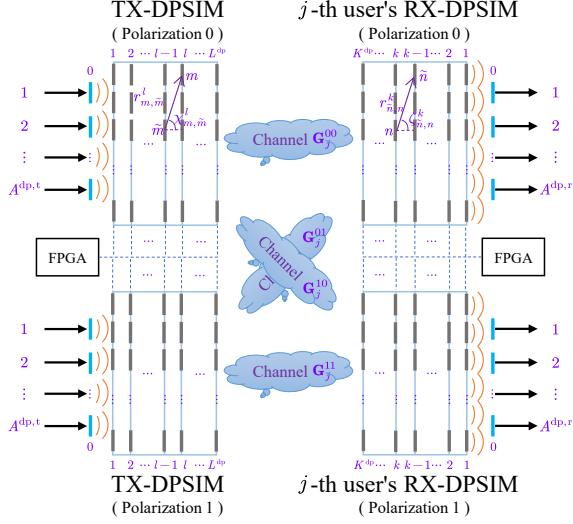


Fig. 5. Schematic diagram of the parameters of TX-DPSIM and RX-DPSIM in the two orthogonal polarization directions.

the TX-DPSIM is denoted as $d^{\text{dp},\text{t}}$, and the spacing in the RX-DPSIM is denoted as $d^{\text{dp},\text{r}}$.

The transmission coefficient of the l -th layer for TX-DPSIM and the k -th layer for the j -th user's RX-DPSIM are defined as (16) and (17), respectively. Considering that the interaction between electromagnetic waves and environmental scatterers is the main mechanism for changing their polarization state [39], we assume that the polarization state of the electromagnetic waves remains unchanged during their propagation between DPSIM layers. Therefore, the overall coefficient matrices for TX-DPSIM and the j -th user's RX-DPSIM are respectively given by (18) and (19), where \mathbf{V}_i^l , $l \in \mathcal{L}^{\text{dp}}$ and \mathbf{U}_i^k , $k \in \mathcal{K}^{\text{dp}}$ are calculated in the similar way as in Section II-A for (3) and (4).

In summary, the signal received by the j -th user on the i -th subcarrier is given by

$$\mathbf{y}_{i,j}^{\text{dp}} = \mathbf{R}_{i,j}^{\text{dp}} \mathbf{G}_{i,j}^{\text{dp}} \mathbf{T}_i^{\text{dp}} \mathbf{p}_i^{\text{dp}} \odot \mathbf{x}_i^{\text{dp}} + \mathbf{n}_{i,j}^{\text{dp}}, \quad (20)$$

where $\mathbf{n}_{i,j}^{\text{dp}} \in \mathbb{C}^{2A^{\text{dp},\text{r}} \times 1}$ is the receiver noise vector with distribution $\mathcal{CN}(0, \sigma^2 \mathbf{I}_{2A^{\text{dp},\text{r}}})$, $\mathbf{x}_i^{\text{dp}} \in \mathbb{C}^{2A^{\text{dp},\text{t}} \times 1}$ is the signal vector satisfying $\mathbb{E} \left\{ \mathbf{x}_i^{\text{dp}} (\mathbf{x}_i^{\text{dp}})^H \right\} = \mathbf{I}_{2A^{\text{dp},\text{t}}}$, $\mathbf{p}_i^{\text{dp}} \in \mathbb{C}^{2A^{\text{dp},\text{t}} \times 1}$ denotes the vector of transmit powers, $\mathbf{G}_{i,j}^{\text{dp}}$ is the channel coefficient matrix between the L^{dp} -th layer metasurface of the TX-DPSIM and the K^{dp} -th layer metasurface of the j -th user's RX-DPSIM.

Analogous to the system described in Section II-A, the model for the DPSIM-assisted E2E OFDM system is given

by

$$\hat{\mathbf{b}}_j = \mathbf{f}_j^{\text{dp},\text{D}} \left(\begin{bmatrix} \mathbf{R}_{1,j}^{\text{dp}} \mathbf{G}_{1,j}^{\text{dp}} \mathbf{T}_1^{\text{dp}} \mathbf{p}_1^{\text{dp}} \odot \mathbf{f}_1^{\text{dp,M}}(\mathbf{b}) + \mathbf{n}_{1,j}^{\text{dp}} \\ \mathbf{R}_{2,j}^{\text{dp}} \mathbf{G}_{2,j}^{\text{dp}} \mathbf{T}_2^{\text{dp}} \mathbf{p}_2^{\text{dp}} \odot \mathbf{f}_2^{\text{dp,M}}(\mathbf{b}) + \mathbf{n}_{2,j}^{\text{dp}} \\ \vdots \\ \mathbf{R}_{N^c,j}^{\text{dp}} \mathbf{G}_{N^c,j}^{\text{dp}} \mathbf{T}_{N^c}^{\text{dp}} \mathbf{p}_{N^c}^{\text{dp}} \odot \mathbf{f}_{N^c}^{\text{dp,M}}(\mathbf{b}) + \mathbf{n}_{N^c,j}^{\text{dp}} \end{bmatrix} \right), \quad (21)$$

where $\mathbf{x}_i^{\text{dp}} = \mathbf{f}_i^{\text{dp,M}}(\mathbf{b})$, $i = 1, 2, \dots, N^c$ denotes the modulation mapping from \mathbf{b} to the signal on the i -th subcarrier. Furthermore, $\hat{\mathbf{b}}_j = \mathbf{f}_j^{\text{dp,D}} \left(\left[\mathbf{y}_{1,j}^{\text{dp}}, \mathbf{y}_{2,j}^{\text{dp}}, \dots, \mathbf{y}_{N^c,j}^{\text{dp}} \right]^T \right)$, $j = 1, 2, \dots, J$ is the demodulation mapping of the received wideband signal at the j -th user to $\hat{\mathbf{b}}_j$.

D. Dual-polarization channel model (DPSIM-assisted OFDM system channel model)

Due to the presence of dual-polarization defects (polarization conversion), we use four matrices (00, 11, 10, and 01) to characterize the channel. These matrices respectively represent the channel from polarization 0 to 0, from polarization 1 to 1, from polarization 0 to 1, and from polarization 1 to 0 [40]. In addition, we introduce the cross-polarization discrimination (XPD) metric to quantify the strength of cross-polarization interference, and assume that the XPD value between the BS's TX-DPSIM and the RX-DPSIM of all users is identical. Then,

$$\text{XPD} = \frac{\mathbb{E}\{\|\mathbf{G}_{i,j}^{00}\|^2\}}{\mathbb{E}\{\|\mathbf{G}_{i,j}^{10}\|^2\}} = \frac{\mathbb{E}\{\|\mathbf{G}_{i,j}^{11}\|^2\}}{\mathbb{E}\{\|\mathbf{G}_{i,j}^{01}\|^2\}} = \frac{1-\epsilon}{\epsilon}, \quad (22)$$

where $\tilde{\mathbf{G}}_{i,j}^{qp} \in \mathbb{C}^{N^{\text{dp}} \times M^{\text{dp}}}$ denotes the channel coefficient matrix for the i -th subcarrier, from the TX-DPSIM polarization p to the j -th user's RX-DPSIM polarization q . $0 < \epsilon \leq 1$ is the proportion of radiated power converted from polarization 1 to 0 and vice versa. Therefore, the dual-polarized channel coefficient matrix $\mathbf{G}_{i,j}^{\text{dp}} \in \mathbb{C}^{N^{\text{dp}} \times M^{\text{dp}}}$ is modeled as [40], [41]

$$\mathbf{G}_{i,j}^{\text{dp}} = \begin{bmatrix} \mathbf{G}_{i,j}^{00} & \mathbf{G}_{i,j}^{01} \\ \mathbf{G}_{i,j}^{10} & \mathbf{G}_{i,j}^{11} \end{bmatrix}, \quad (23)$$

$$\mathbf{G}_{i,j}^{00} = e^{j\psi^{00}} \mathbf{G}_{i,j}, \quad (24)$$

$$\mathbf{G}_{i,j}^{01} = \frac{1}{\sqrt{\text{XPD}}} e^{j\psi^{01}} \mathbf{G}_{i,j}, \quad (25)$$

$$\mathbf{G}_{i,j}^{10} = \frac{1}{\sqrt{\text{XPD}}} e^{j\psi^{10}} \mathbf{G}_{i,j}, \quad (26)$$

$$\mathbf{G}_{i,j}^{11} = e^{j\psi^{11}} \mathbf{G}_{i,j}, \quad (27)$$

where ψ^{00} , ψ^{01} , ψ^{10} , and ψ^{11} respectively represent the phase shifts of the corresponding polarizations, $\mathbf{G}_{i,j} \in \mathbb{C}^{N^{\text{dp}} \times M^{\text{dp}}}$ can be obtained from (9) in Section II-B.

$$\mathbf{T}_i^{\text{dp}} = \begin{bmatrix} \Phi_i^L & 0 \\ 0 & \Phi_i^L \end{bmatrix} \begin{bmatrix} \mathbf{V}_i^L & 0 \\ 0 & \mathbf{V}_i^L \end{bmatrix} \dots \begin{bmatrix} \Phi_i^2 & 0 \\ 0 & \Phi_i^2 \end{bmatrix} \begin{bmatrix} \mathbf{V}_i^2 & 0 \\ 0 & \mathbf{V}_i^2 \end{bmatrix} \begin{bmatrix} \Phi_i^1 & 0 \\ 0 & \Phi_i^1 \end{bmatrix} \begin{bmatrix} \mathbf{V}_i^1 & 0 \\ 0 & \mathbf{V}_i^1 \end{bmatrix} \in \mathbb{C}^{2M^{\text{dp}} \times 2A^{\text{dp},\text{t}}}, \quad (18)$$

$$\mathbf{R}_{i,j}^{\text{dp}} = \begin{bmatrix} \mathbf{U}_i^1 & 0 \\ 0 & \mathbf{U}_i^1 \end{bmatrix} \begin{bmatrix} \Psi_{j,0}^1 & 0 \\ 0 & \Psi_{j,1}^1 \end{bmatrix} \begin{bmatrix} \mathbf{U}_i^2 & 0 \\ 0 & \mathbf{U}_i^2 \end{bmatrix} \begin{bmatrix} \Psi_{j,0}^2 & 0 \\ 0 & \Psi_{j,1}^2 \end{bmatrix} \dots \begin{bmatrix} \mathbf{U}_i^K & 0 \\ 0 & \mathbf{U}_i^K \end{bmatrix} \begin{bmatrix} \Psi_{j,0}^K & 0 \\ 0 & \Psi_{j,1}^K \end{bmatrix} \in \mathbb{C}^{2A^{\text{dp},\text{r}} \times 2N^{\text{dp}}}. \quad (19)$$

III. E2E PROBLEM-SOLVING

A. Problem posing.

We define the received bit vector for a multi-user E2E system as $\hat{\mathbf{b}} = [\hat{\mathbf{b}}_1, \hat{\mathbf{b}}_2, \dots, \hat{\mathbf{b}}_J]^T$. Our goal is to recover the transmitted vector \mathbf{b} from the received vector $\hat{\mathbf{b}}$ as accurately as possible. To this end, SIM (DPSIM) attempts to actively adjust its EM (DPEM) units on its metasurface to construct a matched channel, thereby suppressing bit error problems caused by factors such as multipath effects and frequency-selective fading in the physical channel. Specifically, the objective function is to minimize the bit error rate (BER) between \mathbf{b} and $\hat{\mathbf{b}}$, which is expressed as

$$\text{BER}(\mathbf{b}, \hat{\mathbf{b}}) = \frac{1}{N^{\text{bit}}} \sum_{b=1}^{N^{\text{bit}}} P \left\{ [\mathbf{b}]_b \neq [\hat{\mathbf{b}}]_b \right\}, \quad (28)$$

Subsequently, the optimization problem for the SIM-assisted E2E OFDM system is formulated as follows

$$\mathcal{P}1 : \min_{\Phi^l, \Psi_j^k, \mathbf{p}_i, \mathbf{x}_i} \text{BER}(\mathbf{b}, \hat{\mathbf{b}}), \quad (29a)$$

$$\text{s.t. } \|\mathbf{p}_i\|_1 = P^t, \quad (29b)$$

$$\mathbb{E}\{\mathbf{x}_i \mathbf{x}_i^H\} = \mathbf{I}_{A^t}, \quad (29c)$$

$$(1 - 15), (28). \quad (29d)$$

where P^t denotes the total transmit power of the system. Similarly, the optimization problem for the DPSIM-assisted E2E OFDM system is formulated as follows

$$\mathcal{P}2 : \min_{\Phi_p^l, \Psi_{j,p}^k, \mathbf{p}_i^{\text{dp}}, \mathbf{x}_i^{\text{dp}}} \text{BER}(\mathbf{b}, \hat{\mathbf{b}}), \quad (30a)$$

$$\text{s.t. } \|\mathbf{p}_i^{\text{dp}}\|_1 = P^t, \quad (30b)$$

$$\mathbb{E}\left\{\mathbf{x}_i^{\text{dp}} (\mathbf{x}_i^{\text{dp}})^H\right\} = \mathbf{I}_{2A^{\text{dp},t}}, \quad (30c)$$

$$(3 - 4), (9 - 28). \quad (30d)$$

Due to the strong coupling among the optimization variables and the non-convex unit-modulus constraints on the transmission coefficients of the metasurface EM (DPEM) elements, obtaining the globally optimal solutions of problems $\mathcal{P}1$ and $\mathcal{P}2$ is extremely challenging. Interestingly, we observe that the considered E2E system exhibits a structural similarity to an AE. Motivated by this observation, we propose to exploit deep learning techniques to train the system, aiming to realize flexible modulation and demodulation, while simultaneously establishing an autonomous control mechanism for the SIM (DPSIM).

B. Design of EMNN.

As illustrated in Fig. 6, the core idea of the proposed EMNN is to abstract the metasurface layer in the SIM (DPSIM) into several trainable hidden layers, while modeling the actual EM wave propagation as the forward propagation process in a deep learning network. A large number of transmit bit vectors \mathbf{b} are randomly generated, and the mini-batch stochastic gradient descent (SGD) algorithm is employed to train the EMNN.

Upon completion of the training, the weights of the corresponding hidden layers are extracted and mapped to the SIM (DPSIM) device, thereby enabling precise control of each EM (DPEM) unit. In summary, this approach essentially achieves an artificial intelligence (AI)-driven control mechanism. Furthermore, conventional DNNs are implemented between the transmit bit vectors and the transmitted signals, as well as between the received signals and the recovered bit vectors, to realize automatic modulation and demodulation.

Considering that signals are typically represented in the complex domain, whereas neural network computations are generally performed in the real domain, the real and imaginary parts of the transmit and receive signals for a single-polarized antenna (or one polarization of a dual-polarized antenna) are mapped to two separate weights in the hidden layer. For an EM unit (or one polarization of a DPEM unit), since its magnitude is always unity, only a bijective mapping between its phase coefficient and the corresponding hidden-layer weight is required. The EMNN can be divided into four sub-networks and one channel computation layer, which will be introduced in detail in the following subsections. The parameter configuration of the EMNN is summarized in Table I.

Sub-network 1. BS-DNN. This is a conventional DNN, which is directly deployed at the BS transmitter to realize the mapping from the transmit bit vector \mathbf{b} to the TX transmit signal $\mathbf{x} \odot \mathbf{p}$ or $\mathbf{x}^{\text{dp}} \odot \mathbf{p}^{\text{dp}}$, where

$$\mathbf{x} \odot \mathbf{p} = \left[(\mathbf{x}_1 \odot \mathbf{p}_1)^T, \dots, (\mathbf{x}_{N^c} \odot \mathbf{p}_{N^c})^T \right]^T, \quad (31)$$

$$\mathbf{x}^{\text{dp}} \odot \mathbf{p}^{\text{dp}} = \left[(\mathbf{x}_1^{\text{dp}} \odot \mathbf{p}_1^{\text{dp}})^T, \dots, (\mathbf{x}_{N^c}^{\text{dp}} \odot \mathbf{p}_{N^c}^{\text{dp}})^T \right]^T. \quad (32)$$

The DNN consists of several linear hidden layers, all employing the ReLU activation function. Finally, a power control layer, implemented as a fixed computation layer, performs proportional amplitude scaling of the signal according to the given total transmit power P^t .

Sub-network 2. TX-SIM (TX-DPSIM) NN. This is a specialized fully-connected neural network designed to establish a bijective relationship with the actual TX-SIM/TX-DPSIM hardware. The network is composed of alternating transmission layers and metasurface layers, both abstracted from the system model. The transmission layers can be directly computed from (3) and (4), and are implemented as non-trainable fixed computation layers. In contrast, the metasurface layers correspond one-to-one with the physical metasurfaces in the TX-SIM (TX-DPSIM) device and are implemented as trainable hidden layers.

Sub-network 3. Channel layer. The channel layer is implemented as a fixed computation layer, whose parameters are obtained from the physical environment through channel estimation, channel prediction, and other relevant techniques. Considering the time-varying nature of the channel, two types of channel computation layers are designed, namely, the statistical channel and the instantaneous channel. Moreover, transfer learning is employed to train the entire model, with the detailed training procedures presented in Section III-C.

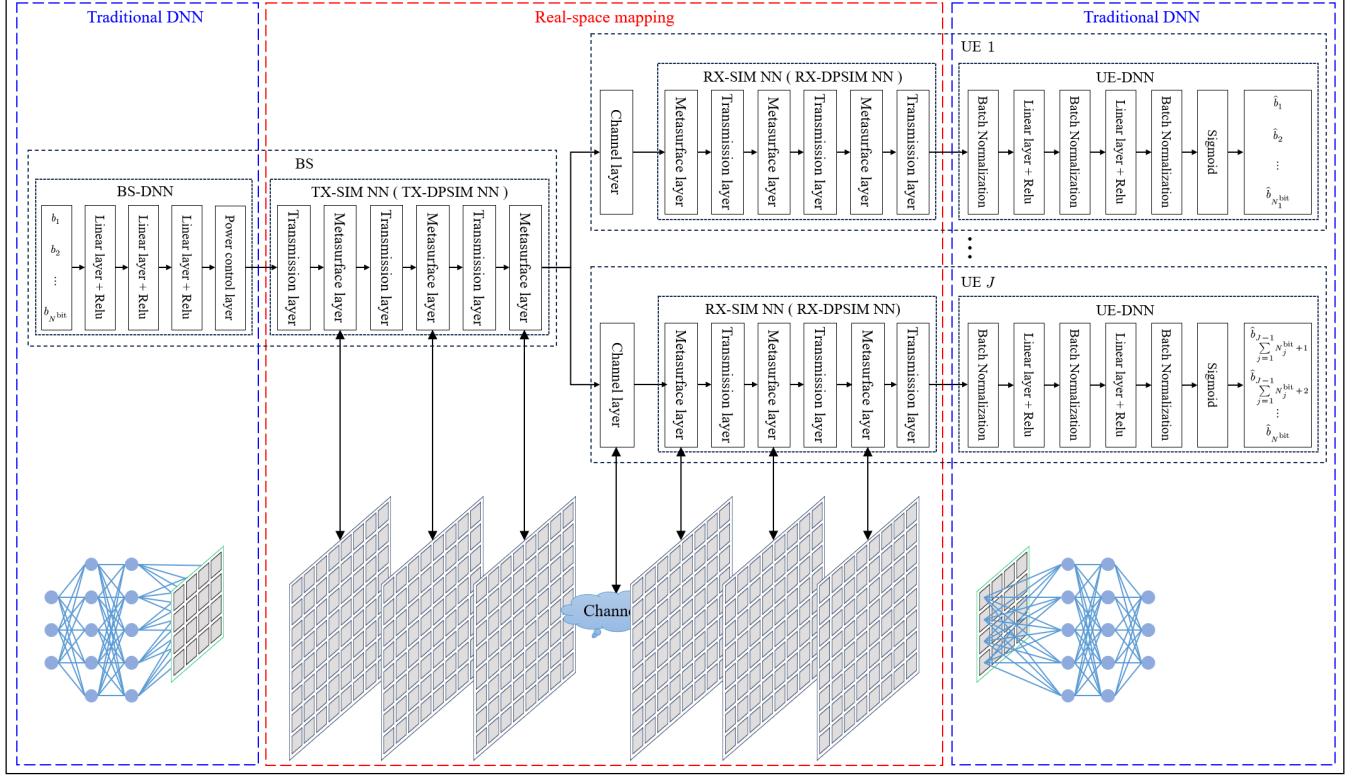


Fig. 6. Schematic diagram of the EMNN structure.

Sub-network 4. RX-SIM (RX-DPSIM) NN. Similar to the design of the TX-SIM (TX-DPSIM) NN, this is also a specialized fully-connected neural network aimed at establishing a bijective relationship with the actual RX-SIM (RX-DPSIM) hardware. The difference lies in that this sub-network is further composed of J modules, each corresponding to the RX-SIM (RX-DPSIM) device integrated at a specific UE. The design principles of the transmission layers and metasurface layers are identical to those described earlier and are therefore omitted here.

Sub-network 5. UE-DNN. Similar to the BS-DNN, this is also a conventional DNN. The difference is that this sub-network is further divided into J modules, each deployed at an individual UE to realize the mapping from the received signal \mathbf{y}_j or \mathbf{y}_j^{dp} to the detected bit vector $\hat{\mathbf{b}}_j$. The DNN alternates between batch normalization and linear layers, with all linear layers employing the ReLU activation function. Finally, a sigmoid function is applied to normalize the output to the range $[0, 1]$, followed by hard decision to recover the bit vector.

C. Model training and deployment framework.

To accurately recover the original bit vector in time-varying channels, the E2E system needs to be retrained periodically based on the CSI information. To accelerate the training

process, we adopt a transfer learning approach. Specifically, a base model is first trained using the statistical CSI, and then the model is fine-tuned in real time according to the instantaneous CSI. The loss function used during the model training is the binary cross-entropy (BCE), which can be computed as in (31), where N^{batch} denotes the training batch size. Similar to (28), BCE is employed to evaluate the consistency between the input and the output. It holds that $\mathcal{L}(\mathbf{b}, \hat{\mathbf{b}}) \geq 0$ with equality if and only if $\mathbf{b} = \hat{\mathbf{b}}$. Therefore, a smaller BCE indicates more accurate bitstream recovery and thus better system performance.

We update the weights of each subnetwork in the EMNN online via the MBSGD algorithm, with the weight update formulas for BS-DNN, TX-SIM NN, RX-SIM NN, and UE-DNN given respectively as

$$\mathbf{W}^T \leftarrow \mathbf{W}^T - \eta \nabla_{\mathbf{W}^T} \mathcal{L}(\mathbf{b}, \hat{\mathbf{b}}), \quad (32)$$

$$\boldsymbol{\Theta}^l \leftarrow \boldsymbol{\Theta}^l - \eta \nabla_{\boldsymbol{\Theta}^l} \mathcal{L}(\mathbf{b}, \hat{\mathbf{b}}), \quad 1 \leq l \leq L, \quad (33)$$

$$\boldsymbol{\Xi}^k \leftarrow \boldsymbol{\Xi}^k - \eta \nabla_{\boldsymbol{\Xi}^k} \mathcal{L}(\mathbf{b}, \hat{\mathbf{b}}), \quad 1 \leq k \leq K, \quad (34)$$

$$\mathbf{W}^R \leftarrow \mathbf{W}^R - \eta \nabla_{\mathbf{W}^R} \mathcal{L}(\mathbf{b}, \hat{\mathbf{b}}), \quad (35)$$

where η denotes the learning rate, \mathbf{W}^T and \mathbf{W}^R respectively represent the weight tensors of the BS-DNN and UE-DNN,

$$\mathcal{L}(\mathbf{b}, \hat{\mathbf{b}}) = -\frac{1}{N^{\text{batch}}} \sum_{b=1}^{N^{\text{batch}}} \sum_{b=1}^{N^{\text{bit}}} \left\{ [\mathbf{b}_b]_b \log \left([\hat{\mathbf{b}}_b]_b \right) + (1 - [\mathbf{b}_b]_b) \log \left(1 - [\hat{\mathbf{b}}_b]_b \right) \right\}. \quad (31)$$

TABLE I

OUTPUT DIMENSIONS (OD) OF EACH LAYER OF EMNN(ALL THE COMPLEX NUMBERS ARE SPLIT INTO THEIR REAL AND IMAGINARY PARTS FOR CALCULATION).

Module	Layer	OD (SIM)	OD (DPSIM)
BS DNN	Input	N^{bit}	N^{bit}
	Liner layer + Relu	$N^{\text{bit}}N^{\text{c}}$	$N^{\text{bit}}N^{\text{c}}$
	Liner layer + Relu	$N^{\text{bit}}N^{\text{c}}$	$N^{\text{bit}}N^{\text{c}}$
	Liner layer + Relu	$2A^t N^{\text{c}}$	$4A^{\text{dp},t} N^{\text{c}}$
TX-SIM (TX-DPSIM)	Power control layer	$2A^t N^{\text{c}}$	$4A^{\text{dp},t} N^{\text{c}}$
	Transmission layer 1	$2M$	$4M$
	Metasurface layer 1	$2M$	$4M$
	Transmission layer 2	$2M$	$4M$
	Metasurface layer 2	$2M$	$4M$

	Transmission layer L	$2M$	$4M$
	Metasurface layer L	$2M$	$4M$
Channle	Channle layer	$2JN$	$4JN$
RX-SIM (RX-DPSIM)	Metasurface layer K	$2JN$	$4JN$
	Transmission layer K	$2JN$	$4JN$
	Metasurface layer $K - 1$	$2JN$	$4JN$
	Transmission layer $K - 1$	$2JN$	$4JN$

	Metasurface layer 1	$2JN$	$4JN$
	Transmission layer 1	$2JA^r N^{\text{c}}$	$4JA^{\text{dp},r} N^{\text{c}}$
	Batch Normalization	$2JA^r N^{\text{c}}$	$4JA^{\text{dp},r} N^{\text{c}}$
UE DNN	Liner layer + Relu	$N^{\text{bit}}N^{\text{c}}$	$N^{\text{bit}}N^{\text{c}}$
	Batch Normalization	$N^{\text{bit}}N^{\text{c}}$	$N^{\text{bit}}N^{\text{c}}$
	Liner layer + Relu	N^{bit}	N^{bit}
	Batch Normalization	N^{bit}	N^{bit}
	Sigmoid	N^{bit}	N^{bit}
	Output	N^{bit}	N^{bit}

Θ^l and Ξ^k are the weight tensors of the corresponding hidden layers abstracted from the metasurfaces in TX-SIM and RX-SIM, respectively. Similarly, for the DPSIM-assisted system, the update formulas corresponding to (33) and (34) are given by

$$\Theta_p^l \leftarrow \Theta_p^l - \eta \nabla_{\Theta_p^l} \mathcal{L}(\mathbf{b}, \hat{\mathbf{b}}), p \in \{0, 1\}, 1 \leq l \leq L, \quad (36)$$

$$\Xi_p^k \leftarrow \Xi_p^k - \eta \nabla_{\Xi_p^k} \mathcal{L}(\mathbf{b}, \hat{\mathbf{b}}), p \in \{0, 1\}, 1 \leq k \leq K, \quad (37)$$

where Θ_p^l and Ξ_p^k are the weight tensors of the hidden layers corresponding to the different polarization directions of the metasurface in the TX-DPSIM and RX-DPSIM after abstraction, respectively.

As shown in Fig. 7, we propose a training and deployment framework for the EMNN. Importantly, the EMNN is always trained on the BS side. At regular intervals, the BS prunes and quantizes the weights of the RX-SIM (RX-DPSIM) NN and the UE-DNN, and then distributes them to the corresponding users. Finally, the BS transmitter deploys the BS-DNN in real time and dynamically controls the EM (DPEM) units according to the weights of the TX-SIM (TX-DPSIM) NN. Each user, in turn, deploys the UE-DNN based on the received weights and simultaneously adjusts the EM (DPEM) units in the RX-SIM (RX-DPSIM) dynamically. The specific training and deployment steps of the EMNN are as follows

Step 1. Pre-training of the EMNN. At the beginning of a long time period, we assume that the BS knows the statistical CSI of all service users. Therefore, before transmitting real-

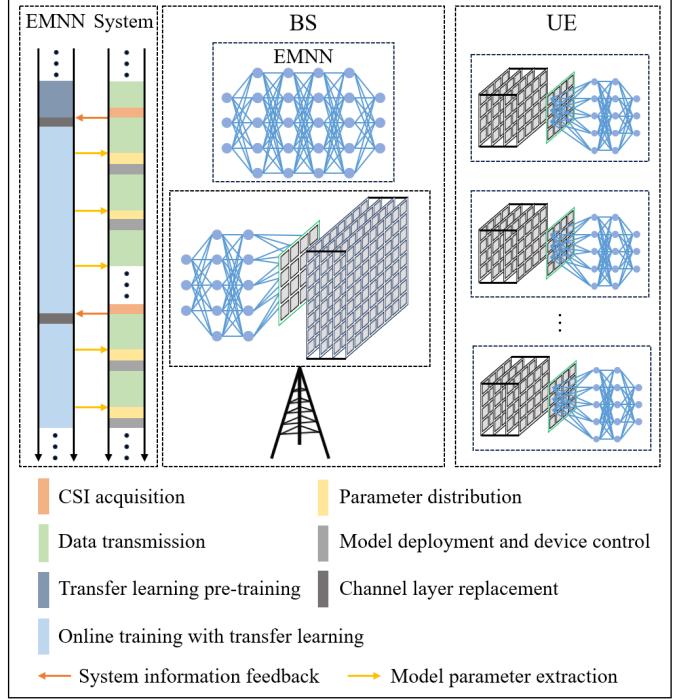


Fig. 7. Model training and deployment framework.

time data, the EMNN is pre-trained through the statistical channel layer to obtain the base model.

Step 2. Instantaneous CSI acquisition and online training of EMNN. The BS obtains the instantaneous CSI of each user through channel estimation, channel prediction, and other methods. Then, the EMNN performs real-time online training through the instantaneous channel layer.

Step 3. Parameter distribution. At regular short intervals, the BS extracts and partitions the weights of the EMNN, retaining those corresponding to the transmitter, while distributing the weights associated with the receiver to the respective users.

Step 4. Systems control. BS-DNN and UE-DNN are conventional neural networks that can be directly deployed inside the transceivers based on the distributed weight data, whereas the TX-SIM (TX-DPSIM) and RX-SIM (RX-DPSIM) devices require solving the weight data according to predefined mapping rules before enabling real-time control of the EM (DPEM) units.

D. Complexity analysis.

The overall training complexity of the proposed EMNN arises from the superposition of wave-domain signal processing in the metasurface network (SIM-NN/DPSIM-NN) and digital computation in conventional neural networks (BS-DNN/UE-DNN). For the SIM-assisted system, the complexity can be decomposed into several components. The complexity of the BS-DNN and UE-DNN depends on their network architectures. If the typical hidden layer dimension is denoted by D , the number of parameters is on the order of $\mathcal{O}(D^2 H J)$, where H denotes the number of hidden layers and J denotes the number of users. The computational complexity of forward propagation is mainly dominated by matrix multiplications and

is also on the order of $\mathcal{O}(D^2HJ)$. For the metasurface part, the number of parameters is $\mathcal{O}(ML)$ for the TX-SIM and $\mathcal{O}(NKJ)$ for all RX-SIMs at the users. The corresponding computational complexity is dominated by matrix multiplications simulating the electromagnetic transmission between analog layers. For each subcarrier, the complexity of the TX-SIM is on the order of $\mathcal{O}(M^2L)$, while that of all RX-SIMs is on the order of $\mathcal{O}(N^2KJ)$. Consequently, the total number of parameters of the SIM-assisted system is $\mathcal{O}(ML + NKJ + D^2HJ)$, and the overall computational complexity of a single forward propagation, scaling with the number of subcarriers N^c , is given by $\mathcal{O}(N^c(M^2L + N^2KJ) + D^2HJ)$.

Similarly, for the DPSIM-assisted system, its architecture can be regarded as two parallel SIMs, resulting in most of its structural complexities being approximately twice those of the SIM-assisted system. Assuming that the DNN architectures in the DPSIM-assisted system are identical to those in the SIM-assisted system, the total number of parameters for the DPSIM-assisted system is $\mathcal{O}(2(M^{dp}L^{dp} + N^{dp}K^{dp}J) + D^2HJ)$, and the overall computational complexity is approximately $\mathcal{O}(2N^c((M^{dp})^2L^{dp} + (N^{dp})^2K^{dp}J) + D^2HJ)$.

E. Model scalability.

Propagation coefficient. Due to inevitable manufacturing imperfections and modeling errors in practical hardware, the propagation coefficients between adjacent metasurfaces in Section II-A may deviate from (3) and (4). Therefore, before integrating the SIM (DPSIM) device, we can pre-calibrate the transmission coefficient matrix of the device by transmitting excitation signals and simultaneously measuring the response of the receiving panel [14].

EMNN structure. To achieve superior system performance, larger network architectures are typically required for the BS-DNN and UE-DNN as the number of UEs, subcarriers, or bits per transmission increases. The proposed EMNN architecture enables decoupling between the subnetworks, allowing the BS-DNN and UE-DNN to be directly replaced with other high-performance networks. The specific choice of such networks is beyond the scope of this work and is left for future research.

SIM (DPSIM) device. For ease of understanding, we have assumed in the preceding discussion that all UEs are equipped with identical RX-SIM (RX-DPSIM) devices, each composed of a sequence of completely identical metasurfaces. In practice, the metasurface layers in the EMNN can be abstracted from metasurfaces of different scales, and the devices integrated into different UEs do not need to be identical, as long as the correspondence between the EM (DPEM) units and the hidden-layer weights is properly established.

IV. SIMULATION RESULTS.

A. Simulation settings.

Considering that the framework proposed in Fig. 7 requires sophisticated cross-layer modeling and system-level implementation, this section focuses on verifying its core component, namely the EMNN-based transfer learning approach, in both SIM-assisted E2E OFDM systems and DPSIM-assisted

TABLE II
SIMULATION PARAMETER SETTINGS

System parameters	Value
Center frequency (f_0)	28 GHz
Center frequency wavelength (λ_0)	10.7 mm
Bandwidth (B)	100 MHz
Number of subcarriers (N^c)	32
Number of UE (J)	3
BS coordinates	(0, 0, 0) m
UE1 coordinates	(10, 0, 20) m
UE2 coordinates	(20, 0, 20) m
UE3 coordinates	(0, 0, 30) m
Number of bits for each UE in an OFDM symbol	[32, 16, 8]
Monte Carlo	100
Channel parameters	Value
Polarization conversion power ratio (ϵ)	0.2
Cross-Polarization Discrimination (XPD)	4
Number of scatterers (S)	100
Path loss reference distance (d_0)	1 m
Path loss exponent (b)	3.5
Path loss shadowing fading variance (δ)	9 dB
Receiver noise (σ^2)	-110dBm
Random generation method of scatterers	[42]
SIM parameters	Value
Number of layers of TX-SIM (L)	3
Number of layers of RX-SIM (K)	3
Number of EM units per layer of TX-SIM ($M = M^x \times M^y$)	100 = 10 × 10
Number of EM units per layer of RX-SIM ($N = N^x \times N^y$)	100 = 10 × 10
Number of TX antennas ($A^t = A^{t,x} \times A^{t,y}$)	16 = 4 × 4
Number of RX antennas per use ($A^r = A^{r,x} \times A^{r,y}$)	9 = 3 × 3
Spacing of EM units in TX-SIM (d^t)	$\lambda_0/2$
Spacing of EM units in RX-SIM (d^r)	$\lambda_0/2$
TX-SIM Layer Spacing (r^t)	$\lambda_0/2$
RX-SIM Layer Spacing (r^r)	$\lambda_0/2$
DPSIM parameters	Value
Number of layers of TX-DPSIM (L^{dp})	3
Number of layers of RX-DPSIM (K^{dp})	3
Number of DPEM units per layer of TX-DPSIM ($M^{dp} = M^{dp,x} \times M^{dp,y}$)	100 = 10 × 10
Number of DPEM units per layer of RX-DPSIM ($N^{dp} = N^{dp,x} \times N^{dp,y}$)	100 = 10 × 10
Number of TX dual-polarized antennas ($A^{dp,t} = A^{dp,t,x} \times A^{dp,t,y}$)	9 = 3 × 3
Number of RX dual-polarized antennas per use ($A^{dp,r} = A^{dp,r,x} \times A^{dp,r,y}$)	4 = 2 × 2
Spacing of DPEM units in TX-DPSIM ($d^{dp,t}$)	$\lambda_0/2$
Spacing of DPEM units in RX-DPSIM ($d^{dp,r}$)	$\lambda_0/2$
TX-DPSIM Layer Spacing ($r^{dp,t}$)	$\lambda_0/2$
RX-DPSIM Layer Spacing ($r^{dp,r}$)	$\lambda_0/2$
Training and testing parameters	Value
Loss function	BCE
Initialization	Xavier
Optimizer	AdamW
Training epoch (E)	2000
Learning rate	0.005
Batch size (N^{batch})	1000
Performance metric	BER
Test scale	100000
Learning rate decay	1 / 1.05

E2E OFDM systems. The performance of this method plays a decisive role in the overall effectiveness of the proposed framework, and thus its validation results largely reflect the feasibility of the framework. Specifically, we assume a quasi-static block fading channel. First, the EMNN is trained as a baseline model using the statistical channel. Then, several sets of instantaneous channels are randomly generated, where

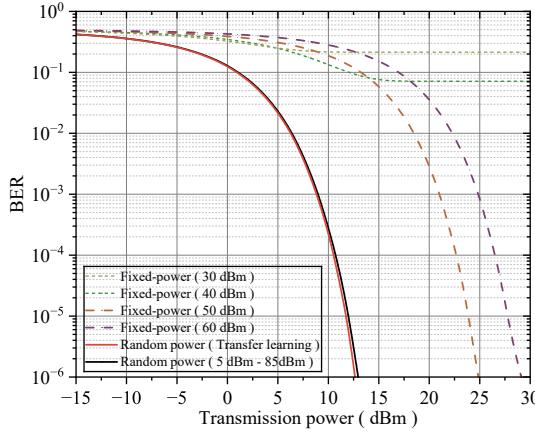


Fig. 8. Performance comparison chart of different training methods.

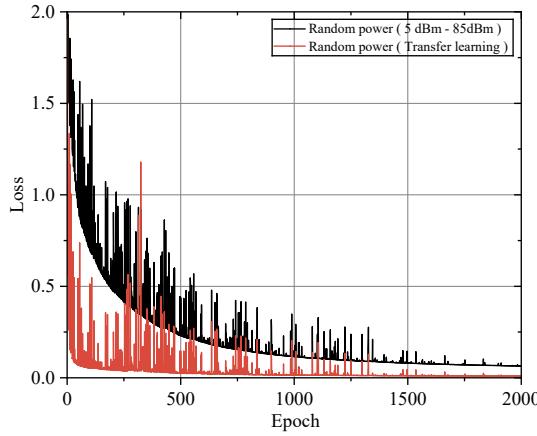


Fig. 9. Iteration graph of EMNN training loss.

the channel layer in the EMNN is accordingly replaced, and the model is fine-tuned via transfer learning. Finally, the average testing performance under different channel conditions is adopted as the system performance metric.

The path loss between the transmitter and the receiver is modeled by [43]

$$\text{PL}(d) = \text{PL}(d_0) + 10b \log_{10} \left(\frac{d}{d_0} \right) + X_\delta, \quad d \geq d_0, \quad (38)$$

where $\text{PL}(d_0) = 20 \log_{10} \left(\frac{4\pi d_0}{\lambda} \right)$ dB is the free space path loss at the reference distance d_0 , b represents the path loss exponent, X_δ is a zero mean Gaussian random variable with a standard deviation δ , characterizing the large-scale signal fluctuations of shadow fading. Furthermore, Table II sequentially summarizes the system parameter settings, channel parameter settings, SIM (DPSIM) parameter settings, and model training/testing parameter settings used for simulation. Unless otherwise specified, the subsequent simulation results are obtained based on the above settings.

B. Model Training Comparative Analysis.

As shown in Fig. 8, we compare the performance of the EMNN trained with different strategies. It can be observed that when trained with a larger fixed transmit power, the EMNN tends to focus on performance at high transmit powers

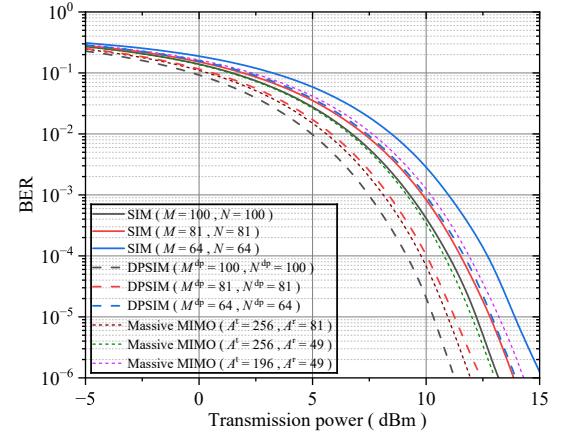


Fig. 10. Performance comparison chart of different EM (DPEM) unit quantities.

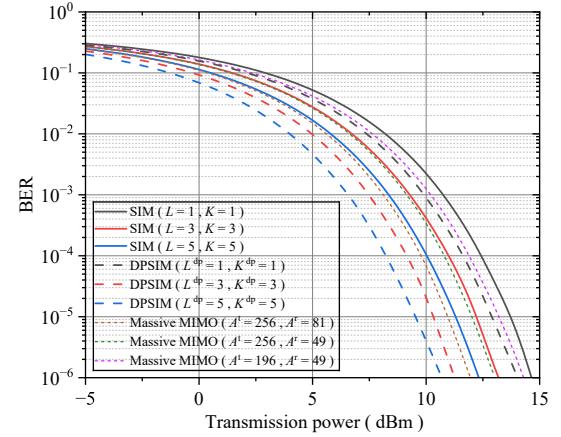


Fig. 11. Performance comparison chart of different metasurface layer unit quantities.

(e.g., 15 dBm–30 dBm), but performs poorly under low transmit power conditions. Conversely, when trained with a smaller fixed transmit power, the EMNN emphasizes low-power performance (e.g., 0 dBm–10 dBm), but degrades at higher transmit powers. In contrast, when the EMNN is trained with transmit powers randomly generated according to a Beta distribution, it can simultaneously maintain robust performance under both high and low transmit powers, thereby achieving an overall lower BER. Furthermore, as shown in Fig. 9, we observe that the transfer learning-based training method can achieve comparable performance to direct training, while significantly reducing the training time.

C. Performance comparison analysis of different specifications of SIM (DPSIM).

Fig. 10 – Fig. 12 depict the BER curves with varying numbers of EM (DPEM) units, metasurface layers, and transceiver antennas, respectively. It can be observed that enlarging the EM (DPEM) array size, increasing the number of metasurface layers, and employing more transmit/receive antennas all contribute to effectively reducing the system BER. Moreover, under all configurations, the DPSIM-assisted E2E OFDM system consistently achieves significantly better BER performance than its SIM-assisted counterpart. Particularly, we

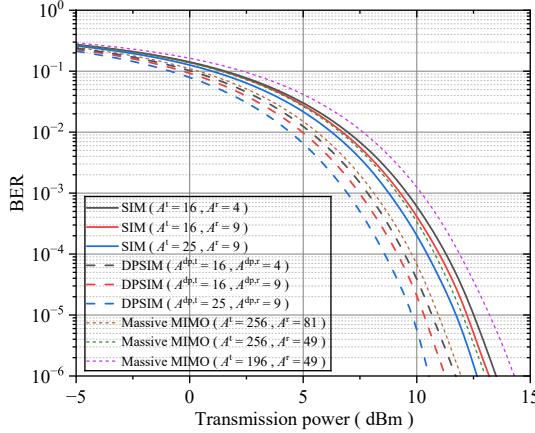


Fig. 12. Performance comparison chart of different antenna quantities.

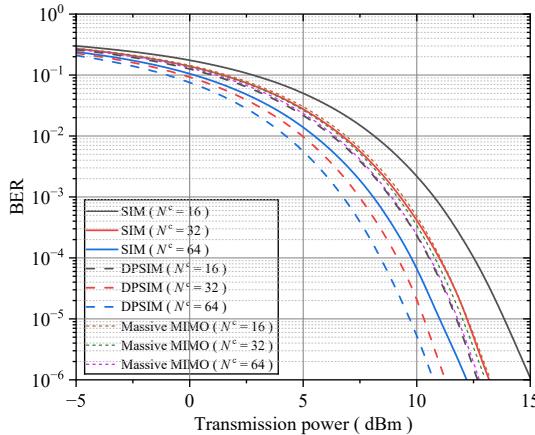


Fig. 13. Performance comparison chart of different subcarrier quantities. Massive MIMO (256T49R).

observe that by integrating the SIM (DPSIM) device, an E2E OFDM system with 16T9R can achieve comparable performance to a massive MIMO OFDM system with 256T49R. This validates the potential of incorporating SIM (DPSIM) devices for wave-domain signal processing to offload baseband processing tasks, thereby simplifying the transceiver design and reducing hardware cost.

D. Comparative analysis of different numbers of subcarriers, transmission bit lengths and polarization defect intensities.

Fig. 13 presents the BER performance with different numbers of subcarriers. The results indicate that increasing the number of subcarriers can effectively reduce the system BER, and that the DPSIM-assisted E2E OFDM system achieves a significant performance improvement compared to the SIM-assisted counterpart. Fig. 14 shows the BER curves for three users within one OFDM symbol period when each user transmits a different number of bits. It can be observed that as the total number of bits per OFDM symbol decreases, the system BER decreases accordingly. In all the considered cases, the DPSIM-assisted E2E OFDM system consistently outperforms the SIM-assisted system. Fig. 15 illustrates the BER performance under different radiated power conversed factors. The results demonstrate that a smaller ϵ (i.e., less energy coupling between orthogonal polarization directions)

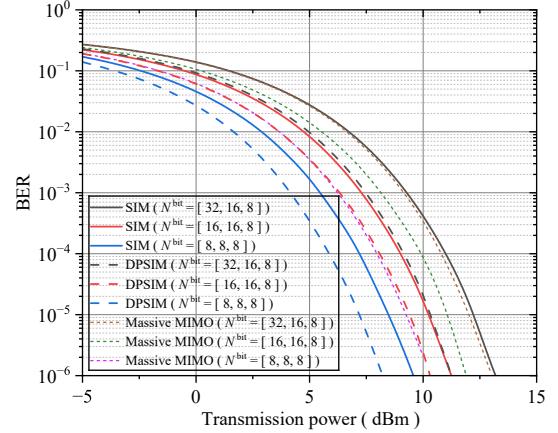


Fig. 14. Performance comparison chart of different transmission bit numbers. Massive MIMO (256T49R).

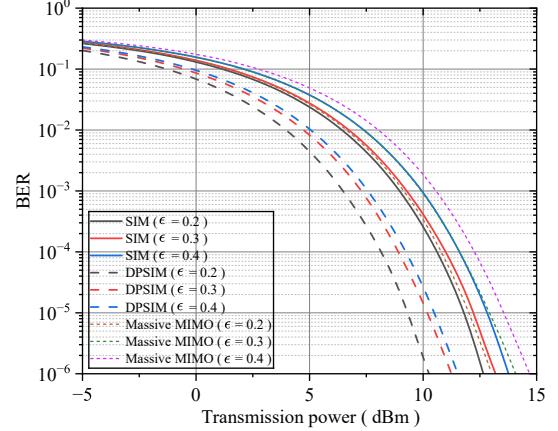


Fig. 15. Performance comparison chart of different polarization radiation factors. Massive MIMO (256T49R).

leads to a lower system BER. Owing to its capability of simultaneously processing signals from two polarization directions, the DPSIM-assisted E2E OFDM system exhibits more pronounced performance gains over the SIM-assisted counterpart in such scenarios.

V. CONCLUSION

To further enhance the overall performance of SIM (DPSIM)-assisted communication systems, we propose an SIM (DPSIM)-assisted E2E OFDM system architecture and design a transfer-learning-based EMNN training and deployment framework for E2E joint optimization from transmitted bit streams to received bit streams. Simulation results demonstrate that, under complex channel conditions, both the SIM-assisted and DPSIM-assisted E2E OFDM systems can achieve robust bit-stream transmission. It is noteworthy that the DPSIM-assisted system exhibits superior BER performance compared with its SIM-assisted counterpart. By integrating wave-domain signal processing with deep-learning-based control, the proposed system significantly reduces the reliance on digital baseband processing capability, thereby simplifying transceiver design and improving energy efficiency. This study highlights the great potential of EMNN combined with SIM (DPSIM)-assisted E2E OFDM systems for next-generation intelligent transceiver design.

REFERENCES

- [1] X. You, C. X. Wang, J. Huang *et al.*, “Towards 6g wireless communication networks: vision, enabling technologies, and new paradigm shifts,” *Sci. China Inf. Sci.*, vol. 64, p. 110301, Nov. 2021.
- [2] Q. Cui, X. You, N. Wei *et al.*, “Overview of ai and communication for 6g network: fundamentals, challenges, and future research opportunities,” *Sci. China Inf. Sci.*, vol. 68, p. 171301, Apr. 2025.
- [3] X. Zhai, Y. Cai, Q. Shi, M. Zhao, G. Y. Li, and B. Champagne, “Joint transceiver design with antenna selection for large-scale mu-mimo mmwave systems,” *IEEE J. Sel. Areas Commun.*, vol. 35, no. 9, pp. 2085–2096, Apr. 2017.
- [4] W. Liu, L. L. Yang, and L. Hanzo, “SVD-assisted multiuser transmitter and multiuser detector design for MIMO systems,” *IEEE Trans. Veh. Technol.*, vol. 58, no. 2, pp. 1016–1021, June 2009.
- [5] Z. Pi, “An introduction to millimeter-wave mobile broadband systems,” *IEEE Commun. Mag.*, vol. 49, no. 6, pp. 101–107, June 2011.
- [6] A. F. Molisch, V. V. Ratnam, S. Han, Z. Li, S. L. H. Nguyen, L. Li, and K. Haneda, “Hybrid beamforming for massive mimo: A survey,” *IEEE Commun. Mag.*, vol. 55, no. 9, pp. 134–141, Sept. 2017.
- [7] S. Park, A. Alkhateeb, and R. W. Heath, “Dynamic subarrays for hybrid precoding in wideband mmwave mimo systems,” *IEEE Trans. Wireless Commun.*, vol. 16, no. 5, pp. 2907–2920, May 2017.
- [8] Y. Liu, X. Liu, X. Mu, T. Hou, J. Xu, M. Di Renzo, and N. Al-Dhahir, “Reconfigurable intelligent surfaces: Principles and opportunities,” *IEEE Commun. Surv. Tutor.*, vol. 23, no. 3, pp. 1546–1577, May 2021.
- [9] W. Tang, J. Y. Dai, M. Z. Chen, K.-K. Wong, X. Li, X. Zhao, S. Jin, Q. Cheng, and T. J. Cui, “Mimo transmission through reconfigurable intelligent surface: System design, analysis, and implementation,” *IEEE J. Sel. Areas Commun.*, vol. 38, no. 11, pp. 2683–2699, Jul. 2020.
- [10] Y. Sun, K. An, Y. Zhu, G. Zheng, K.-K. Wong, S. Chatzinotas, H. Yin, and P. Liu, “Ris-assisted robust hybrid beamforming against simultaneous jamming and eavesdropping attacks,” *IEEE Trans. Wireless Commun.*, vol. 21, no. 11, pp. 9212–9231, May 2022.
- [11] Y. Chen, X. Guo, G. Zhou, S. Jin, D. W. K. Ng, and Z. Wang, “Unified far-field and near-field in holographic mimo: A wavenumber-domain perspective,” *IEEE Commun. Mag.*, vol. 63, no. 1, pp. 30–36, Jan. 2025.
- [12] T. Gong, P. Gavrilidis, R. Ji, C. Huang, G. C. Alexandropoulos, L. Wei, Z. Zhang, M. Debbah, H. V. Poor, and C. Yuen, “Holographic MIMO communications: Theoretical foundations, enabling technologies, and future directions,” *IEEE Commun. Surv. Tuts.*, vol. 26, no. 1, pp. 196–257, Aug. 2024.
- [13] A. Papazafeiropoulos, J. An, P. Kourtessis, T. Ratnarajah, and S. Chatzinotas, “Achievable rate optimization for stacked intelligent metasurface-assisted holographic mimo communications,” *IEEE Trans. Wireless Commun.*, vol. 23, no. 10, pp. 13 173–13 186, May 2024.
- [14] J. An, C. Xu, D. W. K. Ng, G. C. Alexandropoulos, C. Huang, C. Yuen, and L. Hanzo, “Stacked intelligent metasurfaces for efficient holographic MIMO communications in 6G,” *IEEE J. Sel. Areas Commun.*, vol. 41, no. 8, pp. 2380–2396, June 2023.
- [15] J. An, M. Debbah, T. J. Cui, Z. N. Chen, and C. Yuen, “Emerging technologies in intelligent metasurfaces: Shaping the future of wireless communications,” *IEEE Trans. Antennas Propag.*, pp. 1–1, May 2025.
- [16] C. Liu, Q. Ma, Z. J. Luo *et al.*, “A programmable diffractive deep neural network based on a digital-coding metasurface array,” *Nat. Electron.*, vol. 5, pp. 113–122, Feb. 2022.
- [17] J. An, C. Yuen, C. Xu, H. Li, D. W. K. Ng, M. Di Renzo, M. Debbah, and L. Hanzo, “Stacked intelligent metasurface-aided mimo transceiver design,” *IEEE Wirel. Commun.*, vol. 31, no. 4, pp. 123–131, Jul. 2024.
- [18] Y. Zhang, Q. Liu, H. Luo, Y. Xia, and Q. Wang, “Dual-polarization stacked intelligent metasurfaces for holographic mimo,” *IEEE Trans. Veh. Technol.*, pp. 1–6, Jul. 2025.
- [19] S. Dörner, S. Cammerer, J. Hoydis, and S. t. Brink, “Deep learning based communication over the air,” *IEEE J. Sel. Top. Signal Process.*, vol. 12, no. 1, pp. 132–143, Dec. 2018.
- [20] T. O’Shea and J. Hoydis, “An introduction to deep learning for the physical layer,” *IEEE Trans. Cogn. Commun. Netw.*, vol. 3, no. 4, pp. 563–575, Oct. 2017.
- [21] H. Jiang, L. Dai, M. Hao, and R. MacKenzie, “End-to-end learning for ris-aided communication systems,” *IEEE Trans. Veh. Technol.*, vol. 71, no. 6, pp. 6778–6783, Mar. 2022.
- [22] X. Teng, W. Tang, X. Li, and S. Jin, “Baseband-free end-to-end communication system based on diffractive deep neural network,” Jun. 2025. [Online]. Available: <https://arxiv.org/abs/2506.02411>
- [23] H. Ye, L. Liang, G. Y. Li, and B.-H. Juang, “Deep learning-based end-to-end wireless communication systems with conditional gans as unknown channels,” *IEEE Trans. Wireless Commun.*, vol. 19, no. 5, pp. 3133–3143, Feb. 2020.
- [24] H. Ye, G. Y. Li, and B.-H. Juang, “Deep learning based end-to-end wireless communication systems without pilots,” *IEEE Trans. Cogn. Commun. Netw.*, vol. 7, no. 3, pp. 702–714, Feb. 2021.
- [25] J. An, M. D. Renzo, M. Debbah, H. Vincent Poor, and C. Yuen, “Stacked intelligent metasurfaces for multiuser downlink beamforming in the wave domain,” *IEEE Trans. Wireless Commun.*, pp. 1–1, Mar. 2025.
- [26] Z. Li, J. An, and C. Yuen, “Stacked intelligent metasurfaces-enhanced MIMO OFDM wideband communication systems,” Mar. 2025. [Online]. Available: <https://arxiv.org/abs/2503.00368>
- [27] E. E. Bahingayi, N. S. Perović, and L.-N. Tran, “Scaling achievable rates in SIM-aided MIMO systems with metasurface layers: A hybrid optimization framework,” *IEEE Wirel. Commun. Lett.*, pp. 1–1, June 2025.
- [28] A. Abrardo, G. Bartoli, and A. Toccafondi, “A novel comprehensive multiport network model for stacked intelligent metasurfaces (SIM) characterization and optimization,” *IEEE Trans. Commun.*, pp. 1–1, June 2025.
- [29] S. Li, F. Zhang, T. Mao, R. Na, and Z. Wang, “Transmit beamforming design for ISAC with stacked intelligent metasurfaces,” *IEEE Trans. Veh. Technol.*, vol. 74, no. 4, pp. 6767–6772, July 2025.
- [30] G. Huang, J. An, Z. Yang, L. Gan, M. Bennis, and M. Debbah, “Stacked intelligent metasurfaces for task-oriented semantic communications,” *IEEE Wirel. Commun. Lett.*, vol. 14, no. 2, pp. 310–314, Feb. 2025.
- [31] Y. Sun, K. An, M. Yu, Y. Hu, Y. Zhu, Z. Lin, M. Xiao, N. Al-Dhahir, D. Niyato, and J. Wang, “Dual-polarized stacked metasurface transceiver design with rate splitting for next-generation wireless networks,” *IEEE J. Sel. Areas Commun.*, vol. 43, no. 3, pp. 811–833, Jan. 2025.
- [32] A. Felix, S. Cammerer, S. Dörner, J. Hoydis, and S. Ten Brink, “Ofdm-autencoder for end-to-end learning of communications systems,” in *Proc. IEEE 19th Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Aug. 2018, pp. 1–5.
- [33] H. Ye, G. Y. Li, and B.-H. Juang, “Power of deep learning for channel estimation and signal detection in ofdm systems,” *IEEE Wireless Commun. Lett.*, vol. 7, no. 1, pp. 114–117, Sept. 2018.
- [34] E. Balevi and J. G. Andrews, “One-bit ofdm receivers via deep learning,” *IEEE Trans. Commun.*, vol. 67, no. 6, pp. 4326–4336, Mar. 2019.
- [35] F. Ait Aoudia and J. Hoydis, “End-to-end learning for ofdm: From neural receivers to pilotless communication,” *IEEE Trans. Wireless Commun.*, vol. 21, no. 2, pp. 1049–1063, Aug. 2022.
- [36] M. Honkala, D. Korpi, and J. M. J. Huttunen, “Deeprx: Fully convolutional deep learning receiver,” *IEEE Trans. Wireless Commun.*, vol. 20, no. 6, pp. 3925–3940, Feb. 2021.
- [37] X. Lin, Y. Rivenson, N. T. Yardimci, M. Veli, Y. Luo, M. Jarrahi, and A. Ozcan, “All-optical machine learning using diffractive deep neural networks,” *Science*, vol. 361, no. 6406, pp. 1004–1008, Sept. 2018.
- [38] R. W. Heath, N. González-Prelcic, S. Rangan, W. Roh, and A. M. Sayeed, “An overview of signal processing techniques for millimeter wave mimo systems,” *IEEE J. Sel. Top. Signal Process.*, vol. 10, no. 3, pp. 436–453, Feb. 2016.
- [39] V. Degli-Esposti, V.-M. Kolmonen, E. M. Vitucci, and P. Vainikainen, “Analysis and modeling on co- and cross-polarized urban radio propagation for dual-polarized mimo wireless systems,” *IEEE Trans. Antennas Propag.*, vol. 59, no. 11, pp. 4247–4256, Aug. 2011.
- [40] Y. Han, X. Li, W. Tang, S. Jin, Q. Cheng, and T. J. Cui, “Dual-polarized RIS-assisted mobile communications,” *IEEE Trans. Wireless Commun.*, vol. 21, no. 1, pp. 591–606, July 2022.
- [41] S. Zeng, H. Zhang, B. Di, Z. Han, and H. Vincent Poor, “Dual-polarized reconfigurable intelligent surface-based antenna for holographic mimo communications,” *IEEE Trans. Wireless Commun.*, vol. 23, no. 11, pp. 17 339–17 353, Sept. 2024.
- [42] E. Basar, I. Yildirim, and F. Kilinc, “Indoor and outdoor physical channel modeling and efficient positioning for reconfigurable intelligent surfaces in mmwave bands,” *IEEE Trans. Commun.*, vol. 69, no. 12, pp. 8600–8611, Sept. 2021.
- [43] T. S. Rappaport, G. R. MacCartney, M. K. Samimi, and S. Sun, “Wideband millimeter-wave propagation measurements and channel models for future wireless communication system design,” *IEEE Trans. Commun.*, vol. 63, no. 9, pp. 3029–3056, Sept. 2015.