

# Stacked Intelligent Metasurface-Enhanced Wideband Multiuser MIMO OFDM-IM Communications

Zheao Li, *Graduate Student Member, IEEE*, Jiancheng An, *Member, IEEE*, and Chau Yuen, *Fellow, IEEE*

**Abstract**—Leveraging the multilayer realization of programmable metasurfaces, stacked intelligent metasurfaces (SIM) enable fine-grained wave-domain control. However, their wideband deployment is impeded by two structural factors: (i) a single, quasi-static SIM phase tensor must adapt to all subcarriers, and (ii) multiuser scheduling changes the subcarrier activation pattern frame by frame, requiring rapid reconfiguration. To address both challenges, we develop a SIM-enhanced wideband multiuser transceiver built on orthogonal frequency-division multiplexing with index modulation (OFDM-IM). The sparse activation of OFDM-IM confines high-fidelity equalization to the active tones, effectively widening the usable bandwidth. To make the design reliability-aware, we directly target the worst-link bit-error rate (BER) and adopt a max-min per-tone signal-to-interference-plus-noise ratio (SINR) as a principled surrogate, turning the reliability optimization tractable. For frame-rate inference and interpretability, we propose an unfolded projected-gradient-descent network (UPGD-Net) that double-unrolls across the SIM's layers and algorithmic iterations: each cell computes the analytic gradient from the cascaded precoder with a learnable per-iteration step size. Simulations on wideband multiuser downlinks show fast, monotone convergence, an evident layer-depth sweet spot, and consistent gains in worst-link BER and sum rate. By combining structural sparsity with a BER-driven, deep-unfolded optimization backbone, the proposed framework directly addresses the key wideband deficiencies of SIM.

**Index Terms**—Stacked intelligent metasurfaces (SIM), multiuser MIMO, orthogonal frequency-division multiplexing with index modulation (OFDM-IM), bit error rate (BER) minimization, deep unfolding network (DUN), wave-based beamforming,

## I. INTRODUCTION

PROGRAMMABLE metasurfaces have revolutionized the manipulations of electromagnetic (EM) fields and waves, which play a significant role in future sixth-generation (6G) wireless systems [2]–[5]. Composed of densely packed, sub-wavelength meta-atoms with tunable responses, advanced metasurfaces can be employed to support various kinds of wireless communication functions [6]–[9]. Among these, stacked intelligent metasurfaces (SIM) have emerged as a fully-analog means of executing wave-based computing directly in the EM wave domain for multiple-input multiple-output (MIMO) beamforming [10]. By cascading several metasurface layers along the propagation direction, a three-dimensional SIM can realize tunable signal processing and gain computing ability in the wave domain: each layer imposes an adjustable phase-shift profile on the impinging wavefront, and the cascade shapes the spatial transmission function [11]–[13]. In practice, SIM can be equipped on the antenna ar-

rays, functioning as an intelligent radome, offering greater flexibility in controlling EM wave propagation. This fully-analog architecture reduces the extra need for digital baseband processing, minimizes the number of radio frequency (RF) chains, and allows the use of low-resolution digital-to-analog converters (DAC) and analog-to-digital converters (ADC), resulting in lower hardware complexity and improved energy efficiency [14]. Compared with fully-digital and hybrid arrays, SIM shifts multiuser beamforming out of baseband, bringing substantial hardware/energy savings while preserving fine wavefront control. Besides, SIM addresses this limitation by providing depth with multiple sheets rather than single-layer metasurfaces, thereby providing more degrees of freedom (DoF) than traditional reconfigurable intelligent surfaces (RISs) and holographic metasurface antennas [15], [16].

Recent studies have demonstrated SIM's versatility across a range of applications. For instance, [17], [18] illustrated SIM's potential for highly controllable signal transmission and reception in holographic MIMO systems. SIM-based beamforming techniques were explored in [19]–[21] to demonstrate their superiority over conventional MIMO in energy efficiency. Multiuser communication performance enhancements by SIMs were reported in [13], [22]–[26], emphasizing their adaptability and effectiveness. SIM has also been applied to direction-of-arrival (DoA) estimation [27], satellite communication [28], communication security [29], and semantic communication [30]. In both cellular and cell-free networks, SIM has shown promise in mitigating inter-user interference (IUI) and enhancing resource allocation [31]–[33]. Moreover, acting like an artificial neural network (ANN), SIM packs multiple layers of metasurfaces for advanced computation and signal processing tasks, with the configured transmission coefficients serving as trainable network parameters. Therefore, SIM can be employed to support many beyond-communication functionalities, such as wireless sensing and localization [34]–[36], while recent efforts have investigated reinforcement learning and meta-learning frameworks to dynamically configure SIM for various scenarios [37], [38]. In summary, most previous works on SIM focus on narrowband systems with frequency-flat channels [39], [40], where analog signal processing can be easily tackled with high energy efficiency in the wave domain.

Extending SIM from narrowband prototypes to wideband orthogonal frequency-division multiplexing (OFDM) has shown clear promise. The multi- interaction of EM waves through SIM can be viewed as an analogue multi-path synthesizer, where the transmission via each meta-atom acts as a tap with adjustable channel impulse response (CIR) weight [41]. By refining the layer-wise phase shifts, the cascaded device can be configured to combat delay dispersion and enhance multi-carrier transmission. In fact, recent studies reported

This article was presented in part at the IEEE GLOBECOM 2025 [1].

Z. Li, J. An, and C. Yuen are with the School of Electrical and Electronic Engineering, Nanyang Technological University, 639798, Singapore (email: zheao001@e.ntu.edu.sg, {jiancheng.an, chau.yuen}@ntu.edu.sg). (Corresponding author: Chau Yuen.)

that fully-analog transceivers equipped with wideband SIMs improve spatial multiplexing and robustness against frequency-selective deep fades without resorting to digital precoders and combiners [41], [42]. However, two structural bottlenecks persist. First, within an OFDM symbol the SIM phase profiles are quasi-static and hence common to all subcarriers, whereas the per-tone transfer factors vary with frequency. The resulting beamformer must compromise simultaneously across all subcarriers. With moderate aperture and layer count, the performance of SIM is closely tied to the physical and EM properties of the materials used, limiting the specific operational bandwidth. Second, in multiuser MIMO OFDM with dynamic scheduling, the subcarrier allocation changes every frame. In principle, each new subcarrier allocation strategy requires recomputing the entire SIM phase-shift profile under unit-modulus constraints. These facts motivate a transmission and optimization framework that *relaxes the “equalize-all-tones” requirement imposed by cross-band coupling while enabling frame-rate reconfiguration*.

OFDM with index modulation (OFDM-IM) provides the required structural complement for wideband SIM applications from the coding and modulation side. In each OFDM-IM subblock, only a subset of subcarriers is activated, and their positions, which are determined by the incoming index bits, serve as an additional information-bearing dimension [43]–[45]. The induced spectral sparsity (i) lowers peak-to-average-power ratio (PAPR) by reducing the probability of large-symbol superposition in time [46], and (ii) requires high-quality equalization only on the active tones, thereby effectively widening the bandwidth over which a given SIM can realize fully-analog beamforming. Specifically, the reduced number of active subcarriers per user significantly alleviates the burden of jointly optimizing high-dimensional SIM phase-shift matrices across a wideband spectrum. Prior work [1] quantified this trade-off of activation pattern selections in a SIM-enhanced multiuser system, which formulated the channel fitting optimization problem with subcarrier allocation to relax the interference-mitigation constraints. In addition, RIS-aided OFDM-IM designs were proposed in [47], [48] to leverage the subcarrier allocation to transmit extra information efficiently, where phase shifts were configured by an exhaustive search-based optimization algorithm. However, no SIM-enhanced multiuser OFDM-IM transceiver benchmarked against a bit error rate (BER) metric has been established. Bridging this gap advances the SIM designs from narrowband/rate-centric formulations to a wideband, reliability-aware, fully-analog architecture.

The joint design, however, leads to a highly coupled, frame-adaptive optimization: the highly-coupled SIM phase shifts, per-tone power, and the subcarrier activation matrix determined by the current index bits must be updated every OFDM frame under unit-modulus constraints. Classical semi-definite relaxation (SDR) or interior-point methods incur high polynomial complexity and must be restarted for each new subcarrier activation strategy, violating latency targets. The black-box deep neural networks (DNNs) are fast but often ignore physical constraints, yield infeasible phases, and typically have poor interpretability [49]. A principled alternative is the *deep unfolding network* (DUN), which rewrites a finite number

of algorithmic iterations as a transparent, layer-wise neural network architecture [50], [51]. Constraints of optimization problems (e.g., unit-modulus projection) are enforced per layer, monotonic descent is preserved, and only a handful of hyper-parameters (e.g., step sizes) are learned. Because the subcarrier activation matrix enters explicitly in each unfolded stage, a single trained network naturally adapts to frame-varying activation patterns, which precisely correspond to the operating regime of OFDM-IM. Prior DUNs for downlink beamforming [52]–[54] mainly target sum rate or signal-to-noise-ratio (SNR) proxies to simplify complex tasks with lower computational complexity; explicit BER-driven SIM optimization remains fully unexplored.

To fill the above research gaps, we develop a SIM-enhanced multiuser OFDM-IM transceiver and a physics-guided deep-unfolding solver, tailored to wideband communications. A multilayer SIM performs wave-domain zero-forcing (ZF) to mitigate IUI, while OFDM-IM supplies frequency-domain sparsity that lowers PAPR and relaxes the equalization burden. By grouping nearest-neighbor error events into distance classes and introducing a single distance normalizer, we establish a compact bridge from the worst-link BER to a distance-weighted per-tone signal-to-interference-plus-noise-ratio (SINR). This converts BER minimization exactly into a max-min SINR problem that collapses to a single second-order cone programming problem, enabling joint phase-power-index design. The projected-gradient-descent (PGD) for this program is then unfolded into a double-unrolled network (iterations  $\times$  layers), preserving full interpretability. Each cell executes a closed-form PGD update with unit-modulus projection, while only the per-iteration step sizes are learned offline to avoid overfitting. And the training database can cover hundreds of channel/activation patterns.

The main contributions and innovations of this research are described as follows.

- We propose the first SIM-enhanced wideband multiuser MIMO OFDM-IM transceiver: a reconfigurable multilayer SIM realizes fully-analog ZF precoding in the wave domain, and OFDM-IM supplies frequency-domain diversity. The joint design suppresses IUI, mitigates wideband frequency-selective fading without digital precoding, and improves PAPR profile and energy efficiency.
- We derive an analytical BER-to-SINR bridge under SIM-enhanced beamforming and convert BER minimization into a distance-weighted max-min SINR problem. The problem reformulation turns a non-convex BER objective into an implementation-ready tool for joint phase, power, and index allocation, closing the gap between SINR heuristics and true error-rate optimization.
- We design a double-unrolled PGD network (UPGD-Net) aligned with SIM physics: unrolling across the SIM’s layers and iterations, using analytic worst-link SINR gradients, unit-modulus projection, and a small set of learnable step sizes per iteration. The solver preserves hard constraints and monotone descent, offers interpretability, and enables per-frame inference once trained.
- Extensive simulations on wideband multiuser downlinks demonstrate fast, monotone convergence, a clear layer-

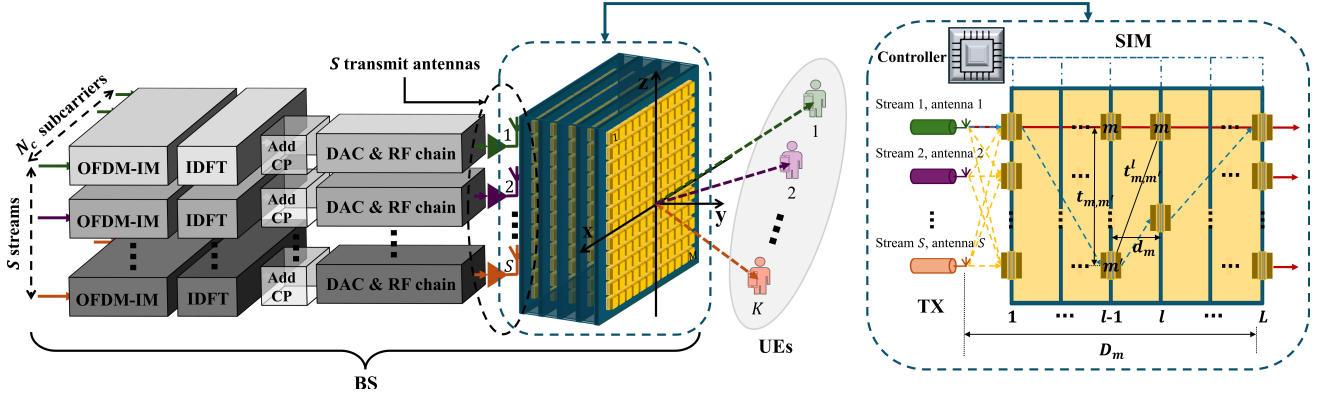


Fig. 1. Transceiver structure of the SIM-enhanced wideband multiuser MIMO OFDM-IM wideband communication system.

depth sweet spot, and consistent gains in worst-link BER and sum rate over digital ZF OFDM-IM and other baselines, which validates the practicality of the proposed model/data-hybrid framework for wideband deployments.

The remainder of this paper is organized as follows: Section II details the proposed SIM-enhanced multiuser MIMO OFDM-IM system model. The BER-driven optimization problem is then formulated in Section III. Section IV introduces the proposed UPGD-Net and evaluates its complexity. The simulation results and analysis are presented in Section V. Finally, conclusions are drawn in Section VI.

**Notations:** In this paper, bold lowercase and uppercase letters are used to denote vectors and matrices, respectively;  $(\mathbf{A})^*$ ,  $(\mathbf{A})^T$ , and  $(\mathbf{A})^H$  represent the conjugate, transpose, and Hermitian transpose of matrix  $\mathbf{A}$ , respectively;  $|c|$ ,  $\Re(c)$ , and  $\Im(c)$  refer to the magnitude, real part, and imaginary part, respectively, of a complex number  $c$ ;  $\|\mathbf{A}\|_F$  denotes the Frobenius norm;  $\mathbb{E}(\mathbf{A})$  stands for the expectation operator;  $\text{diag}(\mathbf{v})$  produces a diagonal matrix with the elements of vector  $\mathbf{v}$  on the main diagonal;  $\text{vec}(\mathbf{A})$  denotes the vectorization of a matrix  $\mathbf{A}$ ;  $\mathbf{A}_{a:b,:}$  and  $\mathbf{A}_{:,c:d}$  represent the submatrices constructed by extracting rows  $a$  to  $b$  and columns  $c$  to  $d$  from matrix  $\mathbf{A}$ , respectively;  $\mathbb{C}^{x \times y}$  represents the space of  $x \times y$  complex-valued matrices;  $\partial f / \partial x$  denotes the partial derivative of a function  $f$  with respect to (w.r.t.) the variable  $x$ .

## II. THE PROPOSED SIM-ENHANCED WIDEBAND MULTIUSER MIMO OFDM-IM SYSTEM MODEL

### A. SIM-Enhanced OFDM-IM Transceiver Architecture

As illustrated in Fig. 1, we focus on the downlink of a wideband multiuser MIMO OFDM system, where the base station (BS) is equipped with a SIM front-end to assist wideband transmission from  $S$  transmit antennas to  $K$  single-antenna user equipments (UEs) within a specific frequency band. The SIM is designed by an array of  $L$  transmissive metasurfaces, each containing  $M = M_x \times M_z$  electronically reconfigurable meta-atoms, where  $M_x$  and  $M_z$  denote the number of meta-atoms along the  $x$ -axis and  $z$ -axis, respectively. A smart controller connected to the metasurfaces independently tunes the phase shift of every meta-atom, enabling the SIM to execute wave-domain computing in situ. The large aperture array of the metasurfaces provides the BS addition power margin to surpass  $S$  physical antennas at BS. Each stream is processed by one DAC-RF chain and fed to a dedicated

transmitter (TX) aperture located immediately behind the first metasurface layer. Hence only  $S$  RF chains are required, whose number is far fewer than that of a conventional fully-digital array with  $M$  active elements. Unlike conventional hybrid-based wideband OFDM architectures, which require a dedicated precoding vector for each UE, the proposed SIM-enhanced wideband OFDM transceiver architecture employs wave-based computing to enhance signal gain and achieve multiuser fully-analog spatial precoding simultaneously.

To establish multiple parallel subchannels in the physical space for interference-free transmission, the proposed SIM-enhanced fully-analog system adopts spatial division multiple access (SDMA) with  $S = K$  spatial streams for the sake of simplicity. In contrast to conventional digital SDMA, where each antenna transmits the superposition of digitally precoded symbols, resulting in a high PAPR, the proposed architecture delegates the matrix multiplication to the wave-based computing by SIM. Consequently, each data stream is radiated directly from a dedicated feed antenna, and the BS requires only  $N_{TX} = N_{RF} = K$  RF chains. The reduced RF burden, together with the inherently lower PAPR of OFDM-IM, markedly improves power-amplifier efficiency.

Let the frequency point of the  $i$ -th subcarrier be  $f_i = f_0 + (i - \frac{N_c+1}{2})\Delta f$ ,  $\Delta f = \frac{B_w}{N_c}$ ,  $i \in N_c = \{1, \dots, N_c\}$ , where  $f_0$  is the center frequency,  $B_w$  is the bandwidth, and  $N_c$  denotes the total number of OFDM subcarriers. For each OFDM symbol, the BS performs the following steps for every stream  $k$  belonging to the set of UEs  $\mathcal{K} = \{1, \dots, K\}$ :

- 1) Bit-to-symbol mapping yields a sparse OFDM-IM vector  $\mathbf{x}_k = [x_k(1), \dots, x_k(N_c)]^T \in \mathbb{C}^{N_c}$  containing  $N_c$  subcarriers in the  $k$ -th stream, following the rules in Sec. II-B.
- 2) An  $N_c$ -point inverse discrete Fourier transform (IDFT) and a cyclic prefix (CP) insertion process convert  $\mathbf{x}_k$  into the time-domain OFDM signal  $\tilde{\mathbf{x}}_k$ .
- 3) One DAC-RF chain up-converts the waveform and feeds the corresponding antenna, followed by the wave-domain beamforming implemented through the SIM structure.
- 4) At the receiver (RX), after discrete Fourier transform (DFT), each UE performs separate OFDM-IM detection.

Specifically, the spacing between the  $m$ -th and  $m'$ -th meta-atoms within the same layer of the SIM is represented as  $t_{m,m'}$ , while the interlayer spacing between the  $m$ -th meta-atom on the  $l$ -th layer and the  $m'$ -th meta-atom on the  $(l-1)$ -th layer is denoted as  $t_{m,m'}^l$ , where  $l \in \mathcal{L} = \{1, \dots, L\}$

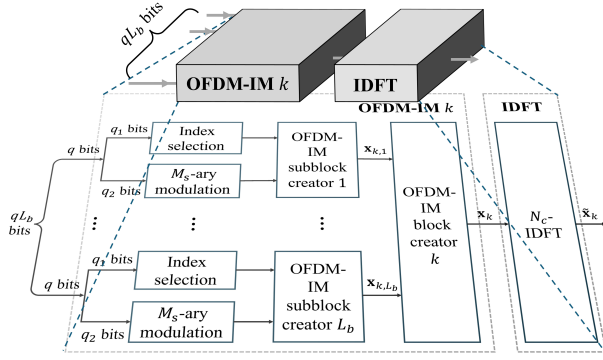


Fig. 2. OFDM index modulators at the  $k$ -th branch of the TX.

and  $m, m' \in \mathcal{M} = \{1, \dots, M\}$ . The axial interlayer distance in the SIM is  $d_m$ , with the total thickness denoted as  $D_m = d_m \times L$ . The spacing between adjacent meta-atoms within the same layer is denoted as  $r_m$ . According to the Rayleigh-Sommerfeld theory [55], the wideband transmission coefficients of the multiple layers of SIM on subcarrier  $i$  can be expressed as:

$$w_{m,m'}^l(i) = \frac{S_m d_m}{t_{m,m'}^l} \left( \frac{1}{2\pi t_{m,m'}^l} - j \frac{f_i}{c} \right) e^{j2\pi t_{m,m'}^l f_i / c}, \quad (1)$$

where  $S_m$  is the area of meta-atom and  $c$  is the speed of light.

Each meta-atom adjusts its transmission coefficient by imposing continuously adjustable phase shifts, represented by  $\phi_m^l = e^{j\theta_m^l}$ ,  $\theta_m^l \in [0, 2\pi)$  for the  $m$ -th meta-atom on the  $l$ -th layer. The phase-shift matrix of the  $l$ -th layer is denoted by  $\Phi^l = \text{diag}([\phi_1^l, \phi_2^l, \dots, \phi_M^l]^T) \in \mathbb{C}^{M \times M}$ . The cumulative effect of signal propagation through multiple layers on subcarrier  $i$  is characterized by the transmission functions:

$$\mathbf{G}(i) = \Phi^L \mathbf{W}_i^L \dots \Phi^2 \mathbf{W}_i^2 \Phi^1 \mathbf{W}_i^1 \in \mathbb{C}^{M \times K}, \quad (2)$$

where  $\mathbf{W}_i^l = [w_{m,m'}^l(i)]_{M \times M}$ ,  $l = 2, \dots, L$  denotes the wideband transmission matrix between the  $(l-1)$ -th layer and the  $l$ -th layer of SIM and  $\mathbf{W}_i^1 \in \mathbb{C}^{M \times K}$  describes the transmission from transmit antennas to the first layer of SIM.

**Remark 1.** By dynamically adjusting the phase-shift matrix of each metasurface layer, the SIM structure adjusts the EM response in the proposed system to perform two key functions:

- 1) **Enhancing beamforming gain:** The SIM, equipped at the BS, processes the emitted wavefronts, enhancing the received power at each UE via large-aperture arrays.
- 2) **Implementing multiuser precoding:** The SIM spatially separates the signals destined for different UEs, mitigating IUI and optimizing signal reception at each UE.

### B. Multiuser OFDM-IM Module Construction at the BS

Per OFDM symbol, the available  $N_c$  subcarriers are divided into  $L_b = N_c/N$  disjoint OFDM-IM subblocks. Each subblock comprises  $N$  subcarriers, among which only  $V < N$  tones are activated to carry modulated data. The remaining  $N - V$  tones remain silent, implicitly conveying index bits. As illustrated in Fig. 2, each transmit stream  $k \in \mathcal{K}$  independently maps incoming bitstreams into sparse frequency-domain subblocks via an index selector and a symbol modulator.

For a subblock  $\ell \in \mathcal{L}_b = \{1, \dots, L_b\}$  of stream  $k$ , the incoming  $q = q_1 + q_2$  bits are split as follows:

$$q_1 = \left\lceil \log_2 \binom{N}{V} \right\rceil, \quad q_2 = V \log_2 M_s, \quad (3)$$

where  $q_1$  index bits determine the positions of  $V$  tones that are active, and  $q_2$  symbol bits are mapped onto  $V M_s$ -ary modulation data symbols. These  $L_b$  subblocks in each transmit stream are aggregated into  $q L_b$  total bits, which are processed in each branch of the transmitter by the OFDM index modulators.

Instead of explicitly listing active-tone indices, all activation decisions can be encoded in a global binary activation matrix:

$$\mathbf{Z} \triangleq [\mathbf{Z}_1, \dots, \mathbf{Z}_\ell, \dots, \mathbf{Z}_{L_b}] \in \{0, 1\}^{K \times N_c}, \quad (4)$$

where  $\mathbf{Z}_\ell \in \{0, 1\}^{K \times N}$  represents the activation decisions in subblock  $\ell$ . Entry  $Z_{k,\ell,n} = 1$  indicates the stream  $k$  is active on the  $n$ -th tone of subblock  $\ell$ , and 0 otherwise. To satisfy the designed OFDM-IM structure, each row of  $\mathbf{Z}_\ell$  must satisfy:  $\sum_{n=1}^N Z_{k,\ell,n} = V$ ,  $\forall k \in \mathcal{K}$ ,  $\forall \ell \in \mathcal{L}_b$ , ensuring that each UE activates exactly  $V$  tones per subblock.

Each UE  $k$  then maps their  $V$  modulated symbols into the selected active tones. The resulting modulated subblock signal is represented compactly as

$$x_{k,\ell,n} = Z_{k,\ell,n} \cdot s_{k,\ell,n}, \quad (5)$$

where  $s_{k,\ell,n} \in \mathbb{C}$  is nonzero only when  $Z_{k,\ell,n} = 1$ .

For the  $k$ -th stream, these sparse subblock vectors  $\mathbf{x}_{k,\ell} = [x_{k,\ell,1}, \dots, x_{k,\ell,N}]^T \in \mathbb{C}^N$  are then concatenated across all  $L_b$  subblocks to produce the full frequency-domain OFDM-IM symbol:  $\mathbf{x}_k = [\mathbf{x}_{k,1}^T, \dots, \mathbf{x}_{k,L_b}^T]^T \in \mathbb{C}^{N_c}$ . Finally, an  $N_c$ -point IDFT followed by a CP insertion of  $N_{cp}$  samples transforms  $\mathbf{x}_k$  into the time-domain signal  $\tilde{\mathbf{x}}_k$ . The resulting signals are simultaneously emitted from the  $M$  SIM-equipped antennas after fully-analog precoding.

**Remark 2.** The use of a unified binary activation matrix  $\mathbf{Z}$  offers a compact and flexible representation of OFDM-IM subcarrier activity across users and subblocks. This matrix-centric formulation not only simplifies the modulation and transmission process in classical OFDM-IM systems but also facilitates the joint design of analog precoding, power allocation, and user scheduling. Specifically,  $\mathbf{Z}$  serves as the foundational structure for interference-aware SINR computation, fairness-constrained scheduling, and optimization of UE-subcarrier associations in multiuser SIM-enhanced systems.

### C. Wideband Channel and Received Signal Model

We consider a wideband multipath channel model, incorporating multiple scatterers that result in frequency-selective fading effects. The tapped delay line channel accounts for  $P_k$  resolvable paths between the SIM at the BS and the  $k$ -th UE, where the complex gain and delay associated with the  $p$ -th path are denoted as  $g_p$  and  $\tau_p$ , respectively. The frequency-domain channel vector on subcarrier  $i$  is

$$\mathbf{h}_k(i) = \sum_{p=0}^{P_k} g_p(f_i) e^{-j2\pi f_i \tau_p} \boldsymbol{\alpha}_p(i)^H \in \mathbb{C}^{1 \times M}, \quad (6)$$

where  $\boldsymbol{\alpha}_p(i) \in \mathbb{C}^{M \times 1}$  is the steering vector of the uniform planar array corresponding to the elevation angle  $\vartheta_p \in [0, \pi)$  and azimuth angle  $\varphi_p \in [-\pi/2, \pi/2]$  at the SIM side. The array steering vectors  $\boldsymbol{\alpha}_p(i) = \boldsymbol{\alpha}_p^x(i) \otimes \boldsymbol{\alpha}_p^z(i)$  represent the array response of all meta-atoms on both the  $x$ -axis and  $z$ -axis in the  $L$ -th layer of SIM, which is

$$[\boldsymbol{\alpha}_p^x(i)]_{m_x} \triangleq e^{j2\pi r_m \sin(\vartheta_p) \sin(\varphi_p) (m_x - 1) f_i / c}, \quad (7)$$

$$[\boldsymbol{\alpha}_p^z(i)]_{m_z} \triangleq e^{j2\pi r_m \cos(\vartheta_p) (m_z - 1) f_i / c}, \quad (8)$$

where  $m_x = 1, 2, \dots, M_x$  and  $m_z = 1, 2, \dots, M_z$ .

Therefore, stacking the wideband channel vectors from BS to all  $K$  UEs on subcarrier  $i$  yields the global channel matrix:

$$\mathbf{H}(i) = [\mathbf{h}_1(i)^T, \dots, \mathbf{h}_K(i)^T]^T \in \mathbb{C}^{K \times M}. \quad (9)$$

In this paper, we assume perfect channel state information (CSI) at the BS, obtained via a time-division duplexing (TDD) protocol. Following the standard implementation of OFDM-IM, each subblock independently performs index and symbol mapping and is separately detected at the RX using a maximum likelihood (ML) detector. This is justified by the Hermitian Toeplitz structure of the channel correlation matrix, which ensures statistical independence of pairwise error events across subblocks, as established in [45].

To relate this global frequency index  $i$  to the OFDM-IM structure, there exists a one-to-one mapping between the global subcarrier index  $i$  and the local subblock indices  $(\ell, n)$ :

$$i = (\ell - 1) \cdot N + n, \quad \text{for } \ell \in \mathcal{L}_b, n \in \mathcal{N}. \quad (10)$$

This index mapping will be used throughout the paper to switch between global frequency index  $i$  and subblock notation  $(\ell, n)$  for clarity in system modeling. Under this setting, the received signal at the  $k$ -th UE on the  $n$ -th subcarrier of the  $\ell$ -th subblock is expressed as:

$$y_{k,\ell,n} = \sqrt{p_{k,\ell,n}} \mathbf{h}_{k,\ell,n} \mathbf{g}_{k,\ell,n} x_{k,\ell,n} + \sum_{j \neq k} \underbrace{\sqrt{p_{j,\ell,n}} \mathbf{h}_{j,\ell,n} \mathbf{g}_{j,\ell,n}}_{\text{Multiuser interference}} x_{j,\ell,n} + w_{k,\ell,n}, \quad (11)$$

where  $p_{k,\ell,n}$  is the power allocated to the  $k$ -th UE,  $\mathbf{h}_{k,\ell,n} \in \mathbb{C}^{1 \times M}$  is the channel vector from the BS to UE  $k$ , and  $\mathbf{g}_{k,\ell,n} \in \mathbb{C}^{M \times 1}$  is the  $k$ -th column of the analog precoder  $\mathbf{G}(i)$  derived from the SIM structure. The noise term  $w_{k,\ell,n} \sim \mathcal{CN}(0, \sigma_w^2)$  represents independent and identically distributed (i.i.d.) additive white Gaussian noise across all UEs, subcarriers, and OFDM-IM subblocks.

Since index bits are embedded implicitly in the nonzero positions and do not receive an extra power budget, all the RF energy is carried only by the  $V$  constellation symbols, which are determined by the activation matrix  $\mathbf{Z}$ . By treating the signals of the remaining UEs as interference according to (11), the SINR  $\gamma_{k,\ell,n}$  at the RX of the UE  $k$  on the subcarrier  $n$  of the subblock  $\ell$  is expressed as

$$\gamma_{k,\ell,n} = \frac{Z_{k,\ell,n} p_{k,\ell,n} |\mathbf{h}_{k,\ell,n} \mathbf{g}_{k,\ell,n}|^2}{\sum_{j \neq k} Z_{j,\ell,n} p_{j,\ell,n} |\mathbf{h}_{j,\ell,n} \mathbf{g}_{j,\ell,n}|^2 + \sigma_w^2}. \quad (12)$$

**Remark 3.** The design of SIM-enhanced OFDM-IM system yields a number of intertwined benefits that are difficult to realize with conventional multiuser MIMO OFDM transceivers.

- 1) **Wave-domain equalization**: Each meta-atom acts as an adjustable CIR tap; stacking  $L$  layers therefore forms a multi-tap analog equalizer. The frequency-selective fading seen by the  $N$  tones of every subblock is therefore flattened in situ, while retaining frequency diversity.
- 2) **Dual-domain interference suppression**: The SIM realizes wave-domain ZF, creating  $K$  parallel, interference-free spatial subchannels across all tones, while OFDM-IM supplies frequency-domain index diversity, together suppressing IUI effectively.

- 3) **Hardware and energy savings**: The fully-analog SIM reduces the number of RF chains that burden hybrid or digital architectures, yielding front-end power and cost savings, essential for large-aperture antenna arrays.
- 4) **Scalable optimization**: Encoding the subcarrier map in the activation matrix  $\mathbf{Z}$  confines the phase-shift optimization to a low-dimensional wave-domain optimization problem, enabling scalable design over ultra-wide bandwidths.
- 5) **Lower PAPR and PA relief**: Wave-domain ZF precoding transmits each stream via a single antenna, inherently reducing the PAPR. The  $N - V$  deliberately silent subtones per OFDM-IM subblock further reduce the probability of instantaneous peaks.

### III. OPTIMIZATION PROBLEM FORMULATION

In this section, we investigate the design of SIM phase shifts to minimize the BER performance of the proposed OFDM-IM system. Unlike conventional OFDM systems where all subcarriers are active and metasurface coefficient optimization typically targets sum rate maximization, we consider a more practical BER metric for system design. To tackle the analytical complexity of BER expressions in SIM-enhanced OFDM-IM, we next reformulate the BER optimization problem into a max-min optimization problem.

#### A. Single-Subblock BER Derivation

Let  $\mathbf{x}_{k,\ell}$  denote the transmitted symbol in the frequency domain and  $\hat{\mathbf{x}}_{k,\ell}$  denote the signal detected erroneously by the RX in valid indices and constellation symbols. For a fixed channel under the TDD protocol, the conditional pairwise error probability (CPEP) in [56] collapses to the deterministic pairwise error probability (PEP) form between two OFDM-IM blocks  $\mathbf{x}_{k,\ell}$  and  $\hat{\mathbf{x}}_{k,\ell}$  of UE  $k$  in the  $\ell$ -th subblock

$$P(\mathbf{x}_{k,\ell} \rightarrow \hat{\mathbf{x}}_{k,\ell}) = Q\left(\sqrt{\frac{1}{2} \sum_{n=1}^N \gamma_{k,\ell,n} |x_{k,\ell,n} - \hat{x}_{k,\ell,n}|^2}\right), \quad (13)$$

where  $Q$ -function is  $Q(x) = \frac{1}{\pi} \int_0^{\pi/2} \exp(-x^2/2 \sin^2 \zeta) d\zeta$ .

Let  $n_{k,\ell} = \binom{N}{V} M_s^V$  be the codebook size of the UE  $k$  to realize  $\mathbf{x}_{k,\ell}$  of subblock  $\ell$ . According to [45], the upper bound of BER can be expressed as

$$P_{b,k,\ell} = \frac{1}{q n_{k,\ell}} \sum_{\mathbf{x}_{k,\ell}} \sum_{\hat{\mathbf{x}}_{k,\ell}} e(\mathbf{x}_{k,\ell}, \hat{\mathbf{x}}_{k,\ell}) P(\mathbf{x}_{k,\ell} \rightarrow \hat{\mathbf{x}}_{k,\ell}), \quad (14)$$

where  $e(\mathbf{x}_{k,\ell}, \hat{\mathbf{x}}_{k,\ell})$  denotes the Hamming weight between the two bit labels for the corresponding pairwise error event.

Since the exact BER expression in (14) for OFDM-IM is hard to be obtained, we next derive a tight bound by keeping only the three classes with minimum distance. The others have larger Euclidean distances, and their PEP terms decay exponentially with SINR and can be neglected in the analytical upper bound. Following Table I, all distance-dominant error events fall into  $\mathcal{C} = \{\mathbf{C}_1, \mathbf{C}_2, \mathbf{C}_3\}$ , while other events typically produce a larger Euclidean distance with an extremely small probability.  $N_{k,\ell,c}$  is pair multiplicity, the number of distinct codeword pairs  $(\mathbf{x}_{k,\ell}, \hat{\mathbf{x}}_{k,\ell})$  whose Euclidean distance equals  $d_c$  and whose error pattern falls into class  $c$ . The bit error weight  $w_{k,\ell,c}$  denotes the Hamming distance between the two

TABLE I  
THREE MINIMUM-DISTANCE ERROR EVENTS IN THE PROPOSED MULTIUSER MIMO OFDM-IM SYSTEM.

| Class                                                                                                                                                                                                                                                          | Error pattern                                                                                                        | Subcarriers affected | Squared distance $d_c^2$ | Pair multiplicity $N_{k,\ell,c}$        | Bit error weight $w_{k,\ell,c}$   |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------|----------------------|--------------------------|-----------------------------------------|-----------------------------------|
| $C_1$                                                                                                                                                                                                                                                          | One active index is identified as inactive; the constellation symbol is unchanged.                                   | 1                    | $E_s$                    | $\binom{V}{1} \binom{N-V}{1}$           | $d_{\text{idx}}$                  |
| $C_2$                                                                                                                                                                                                                                                          | Constellation symbol on an active subcarrier is identified as its nearest neighbor; the index set is unchanged.      | 1                    | $d_{\min}^{(S)}$         | $\binom{V}{1} (M_s - 1)$                | $d_{\text{sym}}$                  |
| $C_3$                                                                                                                                                                                                                                                          | Within the same subcarrier: index error and the recognized symbol is the nearest neighbor of the transmitted symbol. | 1                    | $E_s + d_{\min}^{(S)}$   | $\binom{V}{1} \binom{N-V}{1} (M_s - 1)$ | $d_{\text{idx}} + d_{\text{sym}}$ |
| <b>Note:</b> $E_s = \mathbb{E} s ^2$ is the average symbol energy; $d_{\min}^{(S)}$ is the minimum constellation spacing; $d_{\text{idx}}$ (1–2 bit) and $d_{\text{sym}}$ (1 bit for Gray) are Hamming distances of an index and a symbol error, respectively. |                                                                                                                      |                      |                          |                                         |                                   |

binary labels of any (and hence all) pairs in class  $c$  and  $\beta_c = d_c^2$  is the squared Euclidean distance of that class.  $\delta_{k,\ell,n}^2$  is rewritten as  $\delta_{k,\ell,n,c}^2$ , which means the Euclidean difference on subcarrier  $n$  in class  $c$ . Hence, (14) can be rewritten as

$$P_{b,k,\ell} = \sum_{c \in \mathcal{C}} \frac{1}{q n_{k,\ell}} \sum_{(\mathbf{x}_{k,\ell}, \hat{\mathbf{x}}_{k,\ell}) \in C} e(\mathbf{x}_{k,\ell}, \hat{\mathbf{x}}_{k,\ell}) P(\mathbf{x}_{k,\ell} \rightarrow \hat{\mathbf{x}}_{k,\ell}). \quad (15)$$

Since every pair in  $C_c$  has the same squared distance  $\beta_c$ , the corresponding PEP shares one integral, denoted  $P_{\text{PEP},c}$ . Then the double sum inside (15) collapses to

$$\sum_{(\mathbf{x}_{k,\ell}, \hat{\mathbf{x}}_{k,\ell}) \in C} e(\mathbf{x}_{k,\ell}, \hat{\mathbf{x}}_{k,\ell}) P(\mathbf{x}_{k,\ell} \rightarrow \hat{\mathbf{x}}_{k,\ell}) = N_{k,\ell,c} w_{k,\ell,c} P_{\text{PEP},c}. \quad (16)$$

Using Craig's integral, (13) becomes

$$P_{\text{PEP},c} = \frac{1}{\pi} \int_0^{\pi/2} \exp\left(-\frac{\beta_c}{4 \sin^2 \zeta} \sum_{n=1}^N \gamma_{k,\ell,n} \delta_{k,\ell,n}^2\right) d\zeta, \quad (17)$$

where  $\delta_{k,\ell,n} = |x_{k,\ell,n} - \hat{x}_{k,\ell,n}|$ .

Both quantities  $N_{k,\ell,c}$  and  $w_{k,\ell,c}$  depend only on the codebook structure  $(N, V, M_s)$  and the specific bit mapping used for index and symbol modulation. They are independent of all optimization variables, including precoder  $\mathbf{G}$ , power allocation  $\mathbf{p}$ , and index-activation matrix  $\mathbf{Z}$ . Consequently, (15) can be rewritten as

$$P_{b,k,\ell} = \int_0^{\pi/2} \sum_{c \in \mathcal{C}} \underbrace{\frac{N_{k,\ell,c} w_{k,\ell,c}}{\pi q n_{k,\ell}}}_{W_{k,\ell,c} > 0} \exp\left(-\frac{\beta_c}{4 \sin^2 \zeta} \sum_{n=1}^N \gamma_{k,\ell,n} \delta_{k,\ell,n,c}^2\right) d\zeta, \quad (18)$$

where  $W_{k,\ell,c} = \frac{N_{k,\ell,c} w_{k,\ell,c}}{\pi q n_{k,\ell}}$  appears only as a positive scaling constant determined solely by  $(N, V, M_s)$  and the chosen Gray mapping. Since the objective (18) is a strictly decreasing function w.r.t.  $\gamma_{k,\ell,n}$ , these constants do not affect the direction of optimization.

### B. Network-Level Worst-Link BER and Distance-Weighted SINR

Denote  $P_{b,k}$  the BEP over the entire frame. Because the  $L$  subblocks are statistically independent and carry the same

number of information bits, we arrive at

$$P_{b,k} = \frac{1}{L_b} \sum_{\ell=1}^{L_b} P_{b,k,\ell} = P_{b,k,\ell}, \quad \forall \ell \in \mathcal{L}_b. \quad (19)$$

Hence, the expression per subblock (18) is already the BEP for the whole frame. All constants in Table I remain unchanged because they are defined per subblock.

We adopt the worst-link BER as the quality of service (QoS) target across the frame,

$$\begin{aligned} P_b^{\max} &= \max_k P_{b,k,\ell} \\ &= \max_k \int_0^{\pi/2} \sum_{c \in \mathcal{C}} W_{k,\ell,c} \exp\left(-\frac{\beta_c}{4 \sin^2 \zeta} \Sigma_{k,\ell,c}\right) d\zeta, \end{aligned} \quad (20)$$

where  $\Sigma_{k,\ell,c} = \sum_{n=1}^N \gamma_{k,\ell,n} \delta_{k,\ell,n,c}^2$ . Because the overall objective is a positive linear combination of monotonically decreasing exponentials, these positive constants  $W_{k,\ell,c}$  do not affect the direction of optimization in (20).

Let the error exponent be  $E_{k,\ell,c} \triangleq \exp\left(-\frac{\beta_c}{4 \sin^2 \zeta} \Sigma_{k,\ell,c}\right)$ ,  $\gamma_{\min}^{(c)} \triangleq \min_{n \in \mathcal{N}} \gamma_{k,\ell,n}$ , and  $\Upsilon_{k,\ell,c} \triangleq \sum_{n=1}^N \delta_{k,\ell,n,c}^2 > 0$ . For every UE  $k$  and class  $c$ ,  $\Sigma_{k,\ell,c} \geq \gamma_{\min}^{(c)} \Upsilon_{k,\ell,c}$ . Hence,

$$E_{k,\ell,c} \leq \exp\left(-\frac{\beta_c}{4 \sin^2 \zeta} \gamma_{\min}^{(c)} \Upsilon_{k,\ell,c}\right). \quad (21)$$

Define the largest distance  $\beta_{\max} = \max_{c \in \mathcal{C}} \beta_c$  and the normalized SINR  $\eta_c = \frac{\gamma_{\min}^{(c)}}{\beta_{\max}}$ . Multiply and divide the product  $\beta_c \gamma_{\min}^{(c)}$  by  $\beta_{\max}$ :

$$\beta_c \gamma_{\min}^{(c)} = \beta_c \frac{\beta_{\max}}{\beta_{\max}} \gamma_{\min}^{(c)} \geq \beta_{\max} \frac{\gamma_{\min}^{(c)}}{\beta_{\max}}, \quad (22)$$

Take the minimum over the four classes  $c \in \mathcal{C}$ :

$$\eta = \min_{c \in \mathcal{C}} \eta_c = \min_{c \in \mathcal{C}} \frac{\gamma_{\min}^{(c)}}{\beta_{\max}} = \frac{1}{\beta_{\max}} \min_{n \in \mathcal{N}} \gamma_{k,\ell,n}. \quad (23)$$

Every error class  $c$  contributes an error exponent  $E_{k,\ell,c}$  to the union bound. With (22) and (23), replacing  $\Sigma_{k,\ell,c}$  by its worst-link lower bound and upper-bounding the coefficient  $\beta_c \leq \beta_{\max}$  gives

$$E_{k,\ell,c} \leq \exp\left(-\frac{\beta_{\max} \eta_c}{4 \sin^2 \zeta} \Upsilon_{k,\ell,c}\right) \leq \exp\left(-\frac{\beta_{\max} \eta}{4 \sin^2 \zeta} \Upsilon_{k,\ell,c}\right). \quad (24)$$

Since  $E_{k,\ell,c}$  is decreasing monotonically in  $\eta$ , the worst-link BEP is also monotonically decreasing in  $\eta$ .

Minimizing the worst-link BEP is therefore equivalent to maximizing  $\eta$

$$\min_{\theta_m^l} P_b^{\max} \iff \max_{\theta_m^l} \eta = \frac{1}{\beta_{\max}} \min_{k,\ell,n} \gamma_{k,\ell,n}. \quad (25)$$

**Remark 4.** Without the scaling in (24), we would obtain three various class-specific squared Euclidean distances  $\eta_{C_1}, \eta_{C_2}, \eta_{C_3}$  and hence three separate inequality sets. It increases the number of SINR constraints by a factor of three while delivering the same numerical BER (the largest distance class dominates) to increase the computational complexity. The normalization factor  $\beta_{\max}$  compresses all of them into one constraint, keeps the optimization dimension minimal, and aligns the effective distance squared across classes.

**Remark 5.** All class-dependent squared Euclidean distances  $\beta_c$ , pair multiplicity  $N_{k,\ell,c}$ , and bit error weight  $w_{k,\ell,c}$  are determined solely by the modulation alphabet  $\mathbf{S}_{k,\ell}$  and the OFDM-IM activation rule  $(N, V)$ , which do not depend on channel, precoder, or power allocation. Once fixed offline, these parameters act as static scaling factors within the SINR constraints and therefore do not enlarge the online optimization dimension.

### C. Final Optimization Problem

Here we start with a BER optimization problem based on the proposed wave-based wideband multiuser OFDM-IM precoding framework, subject to (s.t.) the non-convex unit-modulus constraints and total power constraints  $P_t$ :

$$\mathcal{P}_1 : \min_{\theta_m^l} P_b^{\max} \quad (26a)$$

$$\text{s.t. } \mathbf{G}_n = \Phi^L \mathbf{W}_n^L \dots \Phi^2 \mathbf{W}_n^2 \Phi^1 \mathbf{W}_n^1, \quad (26b)$$

$$\Phi^l = \text{diag}([\phi_1^l, \phi_2^l, \dots, \phi_M^l]^T), \quad \forall l \in \mathcal{L}, \quad (26c)$$

$$|\phi_m^l| = |e^{j\theta_m^l}| = 1, \quad m \in \mathcal{M}, \quad \forall l \in \mathcal{L}, \quad (26d)$$

$$\sum_{k=1}^K \sum_{\ell=1}^{L_b} \sum_{n=1}^N p_{k,\ell,n} \leq P_t, \quad (26e)$$

Collecting the SINR constraints for all active links  $(k, \ell, n)$ , we simplify the objective function from Problem  $\mathcal{P}_1$  to Problem  $\mathcal{P}_2$ :

$$\mathcal{P}_2 : \max_{\theta_m^l} \min_{k,\ell,n} \frac{Z_{k,\ell,n} p_{k,\ell,n} |\mathbf{h}_{k,\ell,n} \mathbf{g}_{k,\ell,n}|^2}{\sum_{j \neq k} Z_{j,\ell,n} p_{j,\ell,n} |\mathbf{h}_{k,\ell,n} \mathbf{g}_{j,\ell,n}|^2 + \sigma_w^2} \quad (27a)$$

$$\text{s.t. } (26b), (26c), (26d), (26e) \quad (27b)$$

$$\sum_{n=1}^N Z_{k,\ell,n} = V, \quad \forall k \in \mathcal{K}, \quad \forall \ell \in \mathcal{L}_b, \quad (27c)$$

$$\sum_{k=1}^K Z_{k,\ell,n} \geq 1, \quad \forall \ell \in \mathcal{L}_b, \quad \forall n \in \mathcal{N}, \quad (27d)$$

$$\mathbf{Z} \in \{0, 1\}^{K \times N_c}, \quad (27e)$$

The optimization problem formulated in Section III aims to maximize the worst-link SINR across all UEs and active subcarriers by tuning the phase shifts of a fully-analog SIM-enhanced precoder. However, it is challenging to solve the max-min problem due to the nonlinear and highly coupled structure of the SINR expression w.r.t.  $\theta_m^l$ . Furthermore, in OFDM-IM, each transmitted bitstream activates a distinct subset of subcarriers according to the index bits, which corresponds to a unique subcarrier activation matrix  $\mathbf{Z}$ . In traditional solvers, every change in the subcarrier activation matrix  $\mathbf{Z}$  requires solving the full max-min SINR optimization problem in (27) from scratch. This is computationally expensive and impractical for real-time OFDM-IM transmission where  $\mathbf{Z}$  may be frequently re-optimized due to UE scheduling or adaptive resource allocation. To address this challenge, we propose a DUN that embeds the structure of the PGD method

into a trainable deep neural architecture in Section IV to obtain a SIM phase solution for problem  $\mathcal{P}_2$ .

## IV. DUN SOLUTION FOR SIM PHASE OPTIMIZATION

In this section, we propose a deep UPGD-Net to efficiently approximate the optimal phase configuration under varying subcarrier activation patterns  $\mathbf{Z}$  in our proposed SIM-enhanced multiuser OFDM-IM system. The proposed UPGD-Net acquires a data-driven mapping from the subcarrier assignment to the near-optimal SIM configuration.

### A. Motivation and Design Principle

The max-min SINR optimization problem in (27) involves a cascade of  $L \times M$  unit-modulus phase matrices and must be resolved on a subcarrier basis each time the OFDM-IM activation matrix  $\mathbf{Z}$  changes. Running conventional PGD online is therefore impractical: The iteration has to start from scratch for every new  $(\mathbf{H}, \mathbf{Z})$  pair, and its convergence speed is governed by hand-tuned step sizes.

DUN offers a principled remedy. A finite PGD trajectory can be rewritten as a feed-forward composition of deterministic operators, where the projection, the SINR gradient, and the SIM matrix product are all differentiable. The proposed UPGD-Net is trained offline by back-propagating through the  $T$  unrolled stages so that the learned step sizes minimize the statistical loss  $-\mathbb{E}[\gamma_{\min}]$  on a collection of channels and activation patterns. Once trained, UPGD-Net executes exactly  $T$  matrix-vector operations and one analytical gradient evaluation, which inherits the physical interpretability of PGD but becomes amenable to data-driven acceleration.

The proposed architecture is fundamentally different from a black-box DNN. The proposed SIM-enhanced OFDM-IM system features: (i) SIM's multilayer metasurface cascades, (ii) per-subcarrier SINR coupling, and (iii) a max-min objective that depends on the instantaneous  $\mathbf{Z}$ . Each of these traits is naturally captured by UPGD-Net:

- 1) Double unrolling (layer index  $l$  and iteration  $t$ ) matches the  $L$ -layer cascade and the  $T$  training iterations;
- 2) The worst-link SINR gradient, shared across layers, encapsulates the cross-subcarrier interference structure;
- 3) Trainable step sizes shorten the convergence path, yielding real-time fully-analog beamforming in dynamic OFDM-IM links.

In summary, UPGD-Net preserves the convergence behavior of the analytical solver, respects the hardware constraints of the multilayer SIM, and generalizes to any activation pattern without online re-optimization thanks to training on a diverse set of  $\mathbf{Z}$  matrices. It bridges the gap between model-based optimization and data-driven acceleration [50], providing an interpretable, scalable, and latency-free solution to the SIM phase problem formulated in Section III.

### B. UPGD Network Architecture

As shown in Fig. 3, the proposed UPGD-Net is organized as a two-dimensional grid: the horizontal axis unrolls the  $T$  projected-gradient iterations ( $t = 0, \dots, T-1$ ), whereas the vertical axis stacks the  $L$  physical SIM layers ( $l = 1, \dots, L$ ). A single grid cell therefore captures the micro-update for the

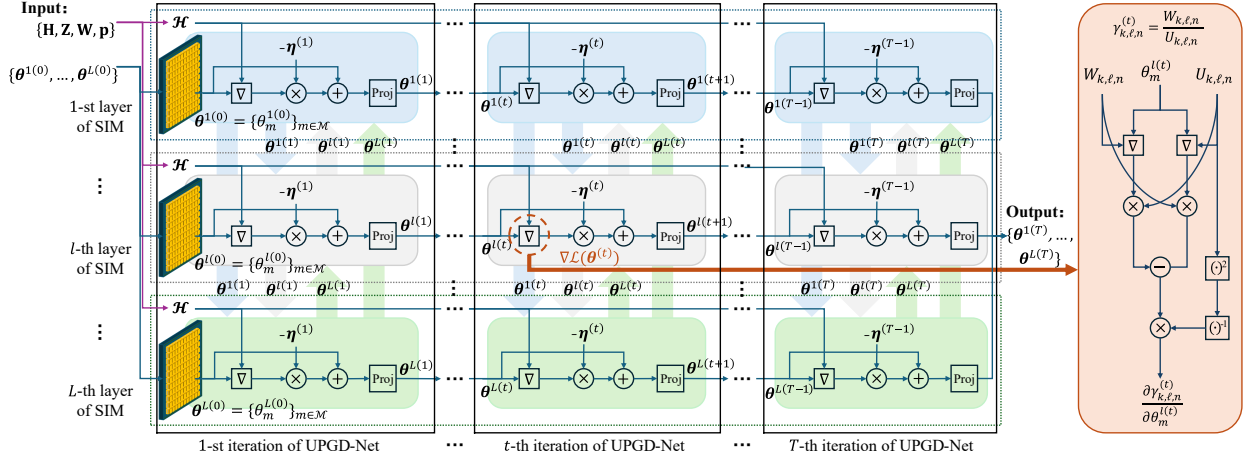


Fig. 3. UPGD-Net for jointly optimizing the phase shifts of the  $L$ -layer SIM over  $T$  projected-gradient iterations. Columns denote successive iterations  $t = 0 : T - 1$ , and rows correspond to SIM layers  $l = 1 : L$ . In each cell, the analytic gradient  $\nabla$  is scaled by the trainable step size  $-\eta^{(t)}$ , added to the current phase-shift tensors, and then projected to  $[0, 2\pi)$ . A vertical coupling bus delivers the worst-link SINR gradient, computed from the cascade precoder of all layers, to every  $\nabla$ -block within the same iteration. All modules are model-driven and fully differentiable.

layer  $l$  executed at iteration  $T$ . The forward and backward flows are described as follows.

#### 1) Forward Propagation:

The inputs of the proposed UPGD-Net are: (i) the loss function matrix  $\mathcal{H} = \{H, Z, W, p\}$  for SINR calculation, using the channel matrix  $H$ , subcarrier allocation  $Z$ , SIM transmission matrix  $W$ , and the power allocation matrix  $p$ ; (ii) the initialized phase-shift tensor  $\theta^{(0)} = \{\theta^{(0)}_m\}_{m \in \mathcal{M}}$ , where  $\theta^{(0)} = \{\theta^{(0)}_m\}_{m \in \mathcal{M}}$ . In each iteration  $t$ , the following computations are performed:

**Step 1 SIM precoder construction:** For each subcarrier  $n$  of subblock  $\ell$  and at every iteration  $t$ , the phase-shift tensor  $\{\theta^{(t)}_l\}_{l \in \mathcal{L}}$  is first cascaded to form the cascade fully-analog precoder  $G_n^{(t)} = \prod_{l=L}^1 \Phi^{(t)}_l W_n^l$ .

**Step 2 SINR computation:** Using  $G_n^{(t)}$ ,  $H$ ,  $Z$ , and  $p$ , the SINRs  $\gamma_{k,\ell,n}^{(t)}$  for different subcarriers are evaluated in parallel. The instantaneous SINR for each UE-subcarrier pair  $(k, \ell, n)$  is evaluated as (12).

**Step 3 Loss function:** Maximizing the minimum is equivalent to minimizing its negative, so the worst-link SINR is then used to compute the sub-gradient of the loss function:

$$\mathcal{L}^{(t)} = -\min_{k,\ell,n} \gamma_{k,\ell,n}^{(t)}, \quad (28)$$

where  $(k^*, \ell^*, n^*) = \arg \min_{k,\ell,n} \gamma_{k,\ell,n}^{(t)}$  denotes the worst-case UE-subcarrier pair.

**Step 4 Gradient evaluation:** The partial derivative of the SINR w.r.t. each phase shift  $\theta_m^{(t)}$  at the  $m$ -th meta-atom on the  $l$ -th layer is accorded to Lemma 1.

**Lemma 1.** Let  $W_{k,\ell,n} = Z_{k,\ell,n} p_{k,\ell,n} |s_{k,\ell,n}|^2$ ,  $U_{k,\ell,n} = \sum_{j \neq k} Z_{j,\ell,n} p_{j,\ell,n} |s_{j,\ell,n}|^2 + \sigma_w^2$  with  $s_{k,\ell,n} = \mathbf{h}_{k,\ell,n} \mathbf{g}_{k,\ell,n}$ . For the  $m$ -th meta-atom's phase of layer  $l$  at iteration  $t$ , the analytic form of the gradient is derived via chain rule:

$$\frac{\partial \gamma_{k,\ell,n}^{(t)}}{\partial \theta_m^{(t)}} = \frac{1}{U_{k,\ell,n}^2} \left( U_{k,\ell,n} \frac{\partial W_{k,\ell,n}}{\partial \theta_m^{(t)}} - W_{k,\ell,n} \frac{\partial U_{k,\ell,n}}{\partial \theta_m^{(t)}} \right). \quad (29)$$

*Proof.* Please see Appendix A.

**Step 5 PGD update and projection:** The SIM phase-shift tensors are updated using a trainable step size  $\eta^{(t)}$ , followed by projection:

$$\theta_m^{(t+1)} = \text{Proj}_{[0,2\pi)} \left[ \theta_m^{(t)} - \eta^{(t)} \cdot \frac{\partial \mathcal{L}^{(t)}}{\partial \theta_m^{(t)}} \right]. \quad (30)$$

The projection enforces the unit-modulus constraint and the updated phases enter the next iteration  $t+1$ . After  $T$  stages, the network outputs  $\theta_m^{(T)}$ , which is directly applied to the SIM controller; no run-time looping is required. After updating  $L \times M$  meta-atom phase shifts, the network outputs the optimized SIM phase configuration  $\theta^{(T)}$  for final deployment.

#### 2) Backward Propagation:

The network parameters to be trained are the per-layer step sizes  $\{\eta^{(t)}\}_{t=0}^{T-1}$ . During training, the overall loss is computed as:

$$\mathcal{L}_{\text{total}} = \mathcal{L}^{(T)} = -\min_{k,\ell,n} \gamma_{k,\ell,n}^{(T)}. \quad (31)$$

Using backpropagation, the gradients of  $\mathcal{L}_{\text{total}}$  are propagated backward through the  $T$  layers via the chain rule. Each trainable step size  $\eta^{(t)}$  is updated by:

$$\frac{\partial \mathcal{L}_{\text{total}}}{\partial \eta^{(t)}} = \sum_{m,l} \frac{\partial \mathcal{L}_{\text{total}}}{\partial \theta_m^{(t+1)}} \cdot \left( -\frac{\partial \mathcal{L}^{(t)}}{\partial \theta_m^{(t)}} \right). \quad (32)$$

The Adam optimizer is exploited to update the step sizes, while all model-driven blocks remain weight-free and fully differentiable. All updates are efficiently implemented using modern automatic differentiation frameworks, leveraging the fully differentiable structure of UPGD-Net.

**Remark 6.** Once trained, the proposed UPGD-Net acts as a generalized SINR-aware SIM phase-shift predictor: given any new bitstream with a derived  $Z$ , the network instantaneously outputs an optimized phase solution without requiring iterative solving. Our UPGD-Net embodies a hybrid design:

1) **Model-driven structure:** The architecture strictly follows the iterative PGD structure, leveraging domain

knowledge from fully-analog beamforming and wave-based SIM optimization.

- 2) **Data-driven learning:** The step sizes and update trajectories are learned from data, adapting to practical channel variations and transmission scenarios.

### C. Complexity Analysis

For each subcarrier  $n$  and iteration  $t$ , the proposed UPGD-Net executes three groups of operations: (i) *SIM-cascade construction*. Because every  $\Phi^{\ell(t)}$  is diagonal, forming the cascade  $\mathbf{G}_n^{(t)}$  amounts to  $L$  diagonal-matrix and dense-matrix products, costing  $\mathcal{O}(LM^2)$  real flops per subcarrier. (ii) *SINR calculation*. Computing  $K$  useful-signal powers and  $K(K-1)$  interference terms requires  $\mathcal{O}(K^2M)$  flops per subcarrier. Besides, calculating the partial derivative of the SINR w.r.t. each SIM phase shift  $\theta_m^\ell$  involves a SIM-Jacobian update of  $\mathcal{O}(M^2)$  and numerator/denominator derivatives of  $\mathcal{O}(K^2M)$ . (iii) *Phase update*. Element-wise scaling by  $-\eta^{(t)} \partial \gamma_{k,\ell,n}^{(t)} / \partial \theta_m^{\ell(t)}$  and  $2\pi$  projection cost  $\mathcal{O}(LM)$  flops and is negligible compared with the previous terms.

Aggregating over all  $N_c$  subcarriers and  $T$  iterations yields the forward complexity  $\mathcal{C}_{\text{fwd}} = \mathcal{O}(N_c T (LM^2 + K^2M))$ . During training, back-propagation reuses all forward intermediates and adds, per stage, only one vector inner product and a scalar update for the trainable step size  $\eta^{(t)}$ . Hence  $\mathcal{C}_{\text{bwd}} \approx \mathcal{C}_{\text{fwd}}$ , and the total complexity of UPGD-Net is approximated as  $\mathcal{O}(N_c T (LM^2 + K^2M))$ .

## V. RESULTS AND ANALYSIS

The numerical results are conducted in this section to characterize the performance of the proposed SIM-enhanced multiuser OFDM-IM communication system.

### A. Simulation Configurations

The system operates at a center frequency  $f_0 = 28$  GHz over a bandwidth  $B_w = 60$  MHz, and the number of subcarriers is set as  $N_c = 16$ . The BS is mounted at a height of 10 m, while the  $K$  single-antenna UEs are placed 250 m away on a line-of-sight street, spaced 30 m apart. Following 3GPP TR 38.901 guidelines for mmWave outdoor micro-cells [57], we set  $P_k = 10$  dominant taps with delays. The length of CP is set to  $N_{cp} = 8$  samples (200 ns) for ensuring that  $N_{cp}$  is greater than the maximum delay  $\tau_{p-\max}$ , thus mitigating inter symbol interference. The thickness of SIM is set as  $D_m = 0.05$  m, and the spacing between the layers is set as  $d_m = D_m/L$ . The parameters of SIM are configured with  $S = K = 4$ ,  $M = 100$ ,  $L = 7$ ,  $r_m = c/(2f_0)$ , and  $S_m = c^2/(2f_0)^2$ . The transmit power  $P_t$  is set to 10 dBm and the antenna gains of the BS and UEs are set to 3 dBi and 0 dBi, respectively. The noise power spectral density is -174 dBm/Hz. The BS power allocation matrix  $\mathbf{p}$  is optimized based on the iterative water-filling algorithm [41]. For active subcarrier index selection,  $L_b = 4$ ,  $N = 4$ , and  $V = 2$  are firstly configured in Table II, where  $q_1 = 2$  bits is used to determine the indices of the two active subcarriers out of four subcarriers.

UPGD-Net implements the double-unfold grid of Fig. 3: each column corresponds to one projected-gradient  $\theta$ -update block. The step-size schedules  $\{\eta^{(t)}\}_{t=0}^{T-1}$  constitute the only trainable parameters with  $T = 30$  unfolded stages. The learning

TABLE II  
INDEX SELECTION TABLE FOR  $N = 4$ ,  $V = 2$ , AND  $q_1 = 2$ .

| Bits   | Indices $(\mathbf{I}_{s,\ell})^T$ | OFDM-IM subblocks $(\mathbf{x}_{s,\ell})^T$            |
|--------|-----------------------------------|--------------------------------------------------------|
| [0, 0] | [1, 3]                            | $[\mathbf{x}_{s,\ell,1}, 0, \mathbf{x}_{s,\ell,2}, 0]$ |
| [0, 1] | [2, 4]                            | $[0, \mathbf{x}_{s,\ell,1}, 0, \mathbf{x}_{s,\ell,2}]$ |
| [1, 0] | [1, 4]                            | $[\mathbf{x}_{s,\ell,1}, 0, 0, \mathbf{x}_{s,\ell,2}]$ |
| [1, 1] | [2, 3]                            | $[0, \mathbf{x}_{s,\ell,1}, \mathbf{x}_{s,\ell,2}, 0]$ |

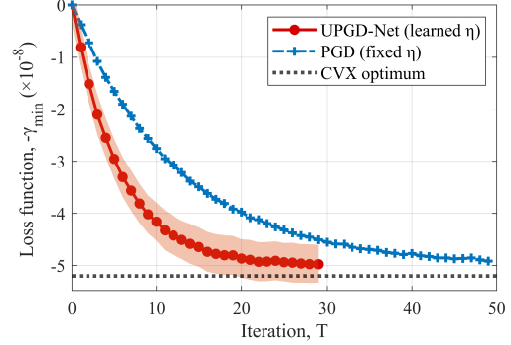


Fig. 4. Convergence performance of the worst-link loss  $\mathcal{L} = -\gamma_{\min}$  versus the iteration index  $T$ .

database comprises 5,000 pairs of independently loss function matrices  $\mathcal{H}$  for different bitstream data, while an 80%:20% split is used for training and validation. Offline optimization relies on the Adam algorithm with a learning rate of  $1 \times 10^{-3}$ , a batch size of 64, and 120 training epochs.

### B. Convergence of the Proposed UPGD-Net

UPGD-Net is benchmarked against (i) conventional PGD with a fixed step 0.15 and  $I = 50$  iterations and (ii) CVX solving the SDR reformulation with per-frame restart. All methods employ the same water-filling power allocation strategy, SIM budget, and simulation parameters.

Fig. 4 confirms the superiority of the proposed UPGD-Net design. The solid orange curve is the mean performance of the proposed UPGD-Net with learned step sizes, averaged over 1,000 independent phase-shift realizations and randomly generated OFDM-IM activation matrices  $\mathbf{Z}$ . The surrounding shaded area shows the corresponding  $\pm 1\sigma$  range, i.e., the region bounded by the 16-th and 84-th percentiles of the 1,000 trajectories, illustrating the spread induced by bitstream variability. The dashed blue curve depicts conventional PGD with an a priori tuned, fixed step, while the dotted black line marks the CVX upper bound obtained by an offline algorithm in [41]. Starting from the same initial condition of 0, UPGD-Net drives the loss below  $-4.5 \times 10^{-8}$  within 12 unfolded stages and reaches  $-5 \times 10^{-8}$  after 25 stages, less than  $0.2 \times 10^{-8}$  away from the CVX global optimum. In contrast, the fixed-step PGD requires almost 50 iterations to attain a comparable level, demonstrating a convergence speed-up of 2.5 times. The red band further highlights the generalization capability: although every realization employs a different index bitstream, with a different subcarrier activation matrix  $\mathbf{Z}$ , the learned step-size schedule keeps all trajectories tightly clustered. This behavior underpins the data-driven component of the network, while the strictly PGD-based layer structure preserves the model-driven monotonic descent, ensuring that no sample ever diverges.

Fig. 5 illustrates how the number of transmit metasurface

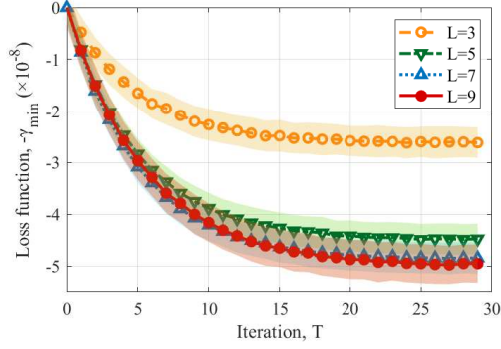


Fig. 5. Convergence performance of the worst-link loss  $\mathcal{L}$  for different numbers of SIM layers  $L$  with  $M = 100$  meta-atoms.

layers  $L$  influences the optimization carried out by the proposed UPGD-Net. With a fixed per-layer aperture with the number of meta-atoms  $M = 100$ , stacking more layers provides additional DoF, so the proposed UPGD-Net converges faster and reaches a lower loss from increasing  $L = 3$  to  $L = 7$ . When  $L$  increases further to 9, though the performance converges fast in the initial iteration, it slightly degrades in the final part. This is because neighboring layers become strongly electromagnetically coupled, so each additional layer under a fixed thickness manipulates the wavefront less effectively when exceeding a certain threshold. A larger  $L$  also enlarges the two-dimensional unfolding grid, increasing per-iteration complexity, as every projected-gradient update has to account for the phase-shift tensors of all preceding layers. Under the present hardware budget,  $L = 7$  therefore emerges as the empirical sweet spot, striking a trade-off between worst-link SINR optimization and implementation cost. Each curve represents the ensemble mean over 1,000 random OFDM-IM activation matrices, while the surrounding shaded area marks the corresponding  $\pm 1 \sigma$  interval. These results confirm that the double-unrolled schedule on training iteration and metasurface layer generalizes well across various bit-stream variability, underscoring the robustness of the proposed model-/data-hybrid design.

### C. End-to-end System Performance Evaluation

Fig. 6 depicts four average BER curves versus the total transmit power  $P_t$  on the proposed  $4 \times 4$  SIM-enhanced multiuser MIMO downlink. All schemes are normalized to the same spectral efficiency  $\eta = 2.67$  bit/s/Hz using binary phase shift keying (BPSK) modulation. The results of the proposed SIM-enhanced OFDM-IM are shown both as a Monte Carlo simulation in a red solid line with circles and as the theoretical upper bound in an orange dotted line with diamonds. As the power increases, the theoretical curve becomes very tight with the simulation curve, confirming the accuracy of the pairwise-error probability analysis in Section III. In addition, two baselines are added for reference, where OFDM-IM with digital ZF is shown in blue dash-dot triangles and classical OFDM with digital ZF is in green dashed inverted triangles. Compared with normal OFDM-IM with digital ZF, the SIM-enhanced OFDM-IM architecture enjoys a consistent 2–3 dB advantage at the target BER  $= 10^{-3}$  owing to the wave-domain gain delivered by the large-aperture SIM structure. Both

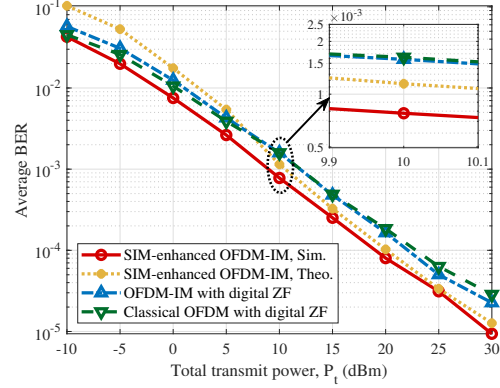


Fig. 6. Average BER versus total transmit power for the proposed SIM-enhanced OFDM-IM and three baseline schemes.

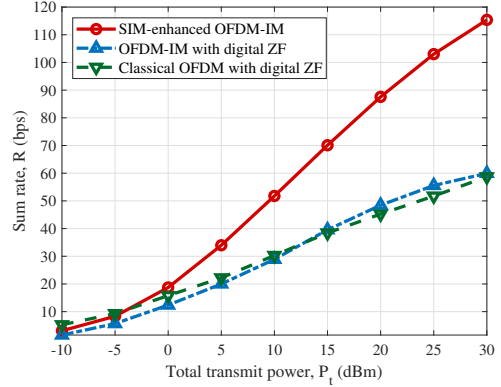


Fig. 7. Sum rate versus total transmit power for the proposed SIM-enhanced OFDM-IM and two baseline schemes.

OFDM-IM curves outperform classical OFDM at medium and high powers because the index modulation exploits the same bandwidth while concentrating energy on the  $V = 2$  active tones, thereby enlarging the minimum Euclidean distance. Overall, this figure verifies that the SIM-enhanced OFDM-IM architecture simultaneously leverages wave-domain aperture gain and index modulation diversity to achieve the lowest BER for a given spectral efficiency and processing budget.

Fig. 7 plots the energy-normalized sum rate versus the total transmit power for the proposed SIM-enhanced OFDM-IM, an OFDM-IM with digital ZF baseline, and classical OFDM with digital ZF. Across the entire transmit power sweep, the proposed SIM-enhanced OFDM-IM curve lies 40–80% above its digital ZF counterparts. The gain is attributed to the large-aperture SIM, which adds a wave-domain boost to every active subcarrier. In addition, digital ZF requires energy normalization to satisfy the total power constraint, while wave-domain precoding via SIM does not amplify coefficients and redistributes energy more effectively across space. Moreover, classical OFDM distributes the same transmit power over all sixteen subcarriers; the per-tone SNR is therefore halved w.r.t. OFDM-IM, resulting in the lowest sum-rate. In contrast, index modulation concentrates power on the  $K = 2$  active tones while injecting additional index bits through the allocation matrix  $\mathbf{Z}$ . Together with the previous BER results in Fig. 6, this joint exploitation of SIM aperture gain and index modulation diversity yields the highest sum rate for a given power

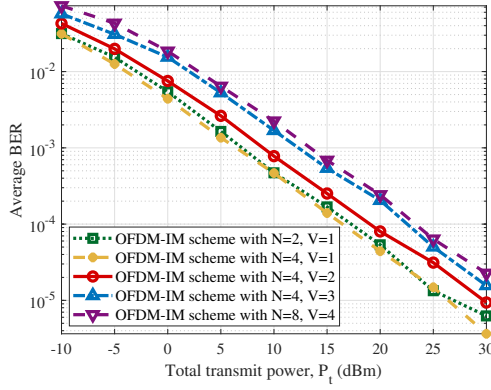


Fig. 8. Average BER versus the total transmit power  $P_t$  for five OFDM-IM configurations with different  $N$  and  $V$  values.

TABLE III  
THROUGHPUT AND ML-DETECTION COMPLEXITY PER  
OFDM-IM PATTERN.

| Pattern        | $\eta_s$ , bit/s/Hz | $L_b$ | Candidates / OFDM                |
|----------------|---------------------|-------|----------------------------------|
| $N = 2, V = 1$ | 2.67                | 8     | $8 \times (2 \times 2) = 32$     |
| $N = 4, V = 1$ | 2                   | 4     | $4 \times (4 \times 2) = 32$     |
| $N = 4, V = 2$ | 2.67                | 4     | $4 \times (6 \times 4) = 96$     |
| $N = 4, V = 3$ | 3.33                | 4     | $4 \times (4 \times 8) = 128$    |
| $N = 8, V = 4$ | 3.33                | 2     | $2 \times (70 \times 16) = 2240$ |

budget, confirming that the proposed transceiver architecture offers superior power efficiency at only a modest increase in detection complexity.

#### D. Trade-off of the Index Selection in OFDM-IM

Fig. 8 compares the average BER of five activation patterns to analyze the trade-off of index selection. All curves retain the BPSK  $1/P_t$  slope expected under SIM-enhanced wave-domain ZF, sweeping the total transmit power from  $-10$  to  $30$  dBm. Table III summarizes the spectral efficiency of MIMO OFDM-IM  $\eta_s = (K q L_b)/(N_c + N_{cp})$  and the per-OFDM symbol detector size  $L_b \binom{N}{V} 2^V$ , where the ML detector enumerates  $\binom{N}{V} 2^V$  candidates per block for BPSK.

When considering equal spectral efficiency, i.e.,  $\eta_s = 2.67$  bit/s/Hz, the performance of eight 2-subcarrier subblocks with  $V = 1$  outperforms  $N = 4, V = 2$  by  $1.2$  dB at  $\text{BER} = 10^{-3}$ , because every active subcarrier enjoys the full symbol power  $P_t/V$ . However, it also introduces twice as many subblocks, increasing the index-mapping and ML search overhead linearly with  $L_b$ . Raising the throughput to  $\eta_s = 3.33$  bit/s/Hz using scheme  $N = 4, V = 3$  or scheme  $N = 8, V = 4$  incurs a  $\sim 3$  dB penalty. This is because increasing the active subcarrier from 2 to 3 will further increase the complexity of SIM phase-shift optimization with high-dimensional activation matrix  $\mathbf{Z}$ . Moreover, the latter scheme further explodes the ML search space, demonstrating that merely enlarging the subblock while increasing  $V$  offers no spectral-efficiency gain over the original scheme but markedly higher complexity. Finally, the low-rate scheme  $N = 4, V = 1$  with  $\eta_s = 2$  bit/s/Hz confirms that BER performance is monotonic in  $V$  for fixed symbol power, while the number of subblocks  $L_b$  affects only implementation overhead. These results highlight a three-way trade-off: reliability favors small- $V$ /small- $N$  patterns ( $N = 2, V = 1$ )

but more OFDM-IM modules; throughput favors moderate- $V$  choices ( $N = 4, V = 2$ ) with considerate reliability; and computational practicality disfavors large-subblock, large- $V$  configurations ( $N = 8, V = 4$ ). The subsequent optimization therefore focuses on the moderate  $N = 4, V = 2$  operating points, which balance BER, spectral efficiency, and detector complexity under the perfect CSI assumption guaranteed by TDD reciprocity.

#### E. Comparison with Existing Transmission Technologies

Finally, Fig. 9 illustrates the PAPR of the time-domain waveforms at four TX antennas with RF chains for three transmission schemes: (i) SIM-enhanced OFDM-IM with wave-domain ZF, (ii) OFDM-IM with conventional digital ZF, and (iii) classical full-subcarrier BPSK-OFDM with digital ZF. For clarity, the maximum and average amplitudes are indicated by dashed lines in each subfigure, such that the vertical gap directly visualizes the PAPR. It is evident that the proposed SIM-enhanced OFDM-IM consistently achieves the lowest PAPR across all TX antennas, with values around  $3.5$ – $3.9$  dB. In contrast, conventional OFDM-IM yields moderately reduced PAPR levels of  $5$ – $6$  dB, while traditional OFDM exhibits the largest fluctuations with PAPR close to  $8.3$  dB.

These results confirm that by jointly exploiting index modulation and wave-domain processing enabled by SIM, the proposed transceiver significantly suppresses envelope peaks, thereby mitigating the intrinsic high-PAPR drawback of OFDM systems. Specifically, the digital baseband directly feeds raw signals into the RF chains, which are subsequently processed by the SIM. This fully-analog wave-domain ZF precoding avoids symbol-level superposition across data streams, while the superposition caused by digital ZF precoding across four data streams of OFDM-IM and OFDM systems further increases constructive peak values. Additionally, index modulation activates only a subset of subcarriers within each subblock, which reduces the instantaneous signal power variance in the time domain, which mitigates large signal spikes and thus suppresses the PAPR. By contrast, the conventional full-tone OFDM with digital ZF generates the most fluctuating waveforms, since every subcarrier is active and carries independent BPSK symbols. Moreover, the superposition caused by digital ZF precoding across four data streams of OFDM-IM and OFDM systems further increases peak values. Overall, the reduction of  $2$ – $5$  dB by the proposed SIM-enhanced architecture over the baselines translates into lower back-off requirements for the power amplifier, improved energy efficiency, and enhanced feasibility of practical deployment.

#### VI. CONCLUSIONS

This paper proposed a SIM-enhanced multiuser OFDM-IM transceiver together with a physics-guided DUN tailored for wideband communications. On the system side, OFDM-IM is integrated as a structural complement to SIM: activating only a subset of tones widens the effective equalization bandwidth attainable by a practical  $L$ -layer SIM while simultaneously lowering the RF-output PAPR. On the objective side, we established a rigorous bridge from the worst-link BER to a distance-weighted per-tone SINR, thereby converting reliability optimization exactly into a max-min SINR problem

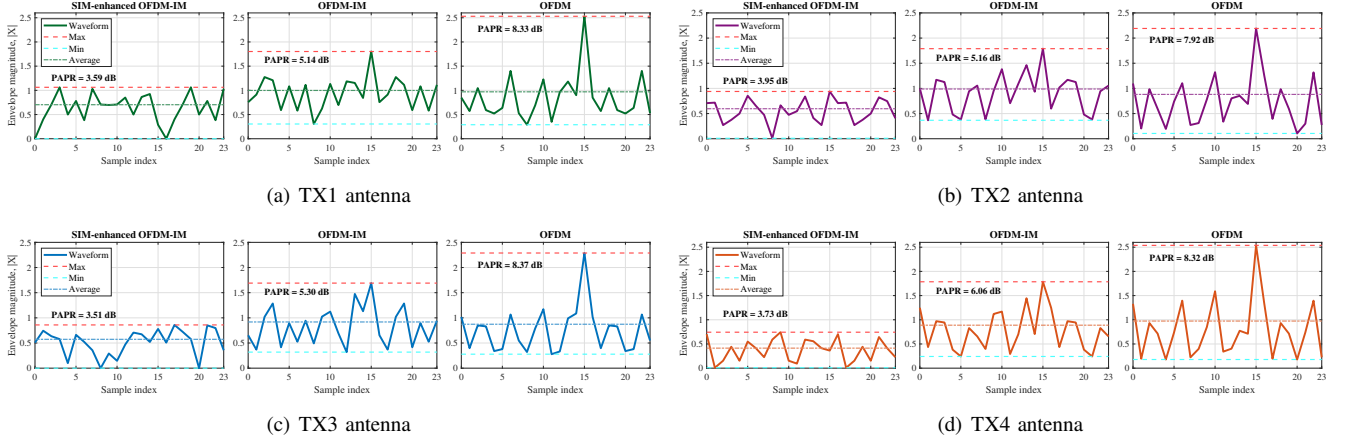


Fig. 9. Instantaneous envelope magnitudes of time-domain waveforms at four TX antennas with RF chains for three schemes, where all cases employ BPSK modulation and the same average transmit power.

with a single second-order cone programming backbone that enables joint phase-power-index design. On the solver side, we unfolded PGD into the proposed UPGD-Net, which enforces unit-modulus feasibility and monotone descent layer-wise and learns only a small set of per-iteration step sizes to support frame-rate reconfiguration under varying activation matrices.

Comprehensive simulations have corroborated these design choices. The proposed framework converges rapidly and stably, typically requiring  $2-3\times$  fewer iterations than fixed-step PGD for comparable accuracy, and exhibits a clear layer-depth sweet spot. Compared with conventional OFDM-IM with digital ZF, the SIM-enhanced OFDM-IM achieves a consistent  $2-3$  dB advantage at the target  $\text{BER}=10^{-3}$ , while its sum rate curve lies  $80\%$  above the digital ZF counterpart across the tested SNR range. An activation-layout sweep further shows that the operating point  $(N, V) = (4, 2)$  strikes a favorable balance among BER, spectral efficiency, and detector complexity under perfect CSI. Finally, time-domain waveform analysis over the baselines confirms a reduced  $2-5$  dB PAPR, underscoring the RF efficiency of the proposed architecture.

Beyond validating feasibility, this paper yields two design insights. First, structure design can broaden the bandwidth of analog processors: frequency-domain sparsity from OFDM-IM expands the usable equalization bandwidth of a finite-depth SIM. Second, the model-based network gains enhanced understanding capability of data: embedding exact PGD dynamics in a learnable scaffold (UPGD-Net) delivers near-optimal performance at frame-level latency while preserving interpretability and constraint satisfaction. These results highlight a practical path to reliability-aware, hardware-efficient wideband multiuser MIMO OFDM systems based on programmable metasurfaces.

#### APPENDIX A PROOF OF LEMMA 1

For the  $m$ -th meta-atom in the  $l$ -th layer, introduce the selection matrix  $\mathbf{E}_m = \text{diag}(\mathbf{e}_m)$ , where  $\mathbf{e}_m \in \mathbb{R}^M$  is the  $m$ -th standard basis vector. Hence  $\frac{\partial \Phi^l}{\partial \theta_m^l} = j \Phi^l \mathbf{E}_m$ . Then, split the chain around layer  $l$ :  $\mathbf{G}_n = \mathbf{G}_n^{(>l)} \Phi^l \mathbf{W}_n^l \mathbf{G}_n^{(<l)}$ , where  $\mathbf{G}_n^{(>l)} = \prod_{\ell=L}^{l+1} \Phi^\ell \mathbf{W}_n^\ell$  and  $\mathbf{G}_n^{(<l)} = \mathbf{W}_n^{l-1} \prod_{\ell=l-2}^1 \Phi^\ell \mathbf{W}_n^\ell$ .

Taking the derivative, we arrive at

$$\frac{\partial \mathbf{G}_n}{\partial \theta_m^l} = j \mathbf{G}_n^{(>l)} \mathbf{E}_m \mathbf{W}_n^l \mathbf{G}_n^{(<l)}. \quad (33)$$

Recall that the row channel times SIM-enhanced precoder is  $s_{k,\ell,n} = \mathbf{h}_{k,\ell,n} \mathbf{g}_{k,\ell,n} \in \mathbb{C}$ . The useful-signal numerator of SINR is

$$\frac{\partial W_{k,\ell,n}}{\partial \theta_m^l} = Z_{k,\ell,n} p_{k,\ell,n} \frac{\partial |s_{k,\ell,n}|^2}{\partial \theta_m^l}, \quad (34)$$

where  $\frac{\partial |s_{k,\ell,n}|^2}{\partial \theta_m^l} = 2 \Re \left\{ s_{k,\ell,n}^* \mathbf{h}_{k,\ell,n} \frac{\partial \mathbf{g}_{k,\ell,n}}{\partial \theta_m^l} \right\}$ .

The interference-plus-noise denominator of SINR is

$$\frac{\partial U_{k,\ell,n}}{\partial \theta_m^l} = \sum_{j \neq k} Z_{j,\ell,n} p_{j,\ell,n} 2 \Re \left\{ s_{j,\ell,n}^* \mathbf{h}_{k,\ell,n} \frac{\partial \mathbf{g}_{j,\ell,n}}{\partial \theta_m^l} \right\}. \quad (35)$$

Using the quotient rule,

$$\frac{\partial \gamma_{k,\ell,n}}{\partial \theta_m^l} = \frac{1}{U_{k,\ell,n}^2} \left( U_{k,\ell,n} \frac{\partial W_{k,\ell,n}}{\partial \theta_m^l} - W_{k,\ell,n} \frac{\partial U_{k,\ell,n}}{\partial \theta_m^l} \right). \quad (36)$$

Upon plugging (34) and (35), the expression for (36) yields the closed-form gradient used in the PGD update.

#### REFERENCES

- [1] Z. Li, J. An, and C. Yuen, "Fundamental trade-off in wideband stacked intelligent metasurface assisted OFDMA systems," Accepted to *IEEE GLOBECOM*'25, Taipei, Taiwan, Dec. 2025, pp. 1-6.
- [2] E. Björnson *et al.*, "Towards 6G MIMO: Massive spatial multiplexing, dense arrays, and interplay between electromagnetics and processing," *arXiv:2401.02844*, Jan. 2024.
- [3] M. Di Renzo *et al.*, "Smart radio environments empowered by reconfigurable AI meta-surfaces: an idea whose time has come," *EURASIP J. Wireless Commun. Net.*, vol. 2019, no. 129, pp. 1-20, May 2019.
- [4] M. Di Renzo, *et al.*, "Smart radio environments empowered by reconfigurable intelligent surfaces: How it works, state of research, and the road ahead," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 11, pp. 2450-2525, Nov. 2020.
- [5] Q. Wu and R. Zhang, "Towards smart and reconfigurable environment: Intelligent reflecting surface aided wireless network," *IEEE Commun. Mag.*, vol. 58, no. 1, pp. 106-112, Jan. 2020.
- [6] C. Huang, A. Zappone, G. C. Alexandropoulos, M. Debbah, and C. Yuen, "Reconfigurable intelligent surfaces for energy efficiency in wireless communication," *IEEE Trans. Wireless Commun.*, vol. 18, no. 8, pp. 4157-4170, Aug. 2019.
- [7] E. Basar, M. Di Renzo, J. De Rosny, M. Debbah, M. -S. Alouini, and R. Zhang, "Wireless communications through reconfigurable intelligent surfaces," *IEEE Access*, vol. 7, pp. 116753-116773, 2019.
- [8] M. Di Renzo, F. H. Danufane, and S. Tretyakov, "Communication models for reconfigurable intelligent surfaces: From surface electromagnetics to wireless networks optimization," *Proc. IEEE*, vol. 110, no. 9, pp. 1164-1209, Sept. 2022.

- [9] Y. Liu *et al.*, “Reconfigurable intelligent surfaces: Principles and opportunities,” *IEEE Commun. Surveys Tuts.*, vol. 23, no. 3, pp. 1546–1577, thirdquarter 2021.
- [10] J. An *et al.*, “Stacked intelligent metasurface-aided MIMO transceiver design,” *IEEE Wireless Commun.*, vol. 31, no. 4, pp. 123–131, Aug. 2024.
- [11] C. Liu *et al.*, “A programmable diffractive deep neural network based on a digital-coding metasurface array,” *Nature Electron.*, vol. 5, no. 2, pp. 113–122, Feb. 2022.
- [12] N. Chamanara, Y. Vahabzadeh, and C. Caloz, “Stacked metasurface slab,” in *Proc. 12th Int. Congr. Artif. Mater. Novel Wave Phenomena (Metamaterials)*, Aug. 2018, pp. 70–72.
- [13] J. An, M. Di Renzo, M. Debbah, and C. Yuen, “Stacked intelligent metasurfaces for multiuser beamforming in the wave domain,” in *Proc. IEEE ICC*, Rome, Italy, May 2023, pp. 1–6.
- [14] N. S. Perović, E. E. Bahingayi, and L. Tran, “Energy-efficient designs for SIM-based broadcast MIMO systems,” *arXiv:2409.00628*, Sept. 2024.
- [15] J. An, C. Yuen, L. Dai, M. Di Renzo, M. Debbah, and L. Hanzo, “Near-field communications: Research advances, potential, and challenges,” *IEEE Wireless Commun.*, vol. 31, no. 3, pp. 100–107, June 2024.
- [16] 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. 13173–13186, Oct. 2024.
- [17] J. An *et al.*, “Stacked intelligent metasurfaces for efficient holographic MIMO communications in 6G,” *IEEE J. Sel. Areas Commun.*, vol. 41, no. 8, pp. 2380–2396, Aug. 2023.
- [18] N. S. Perović and L. -N. Tran, “Mutual information optimization for SIM-based holographic MIMO systems,” *IEEE Commun. Letters*, vol. 28, no. 11, pp. 2583–2587, Nov. 2024.
- [19] M. Nerini and B. Clerckx, “Physically consistent modeling of stacked intelligent metasurfaces implemented with beyond diagonal RIS,” *IEEE Commun. Letters*, vol. 28, no. 7, pp. 1693–1697, July 2024.
- [20] N. U. Hassan, J. An, M. Di Renzo, M. Debbah, and C. Yuen, “Efficient beamforming and radiation pattern control using stacked intelligent metasurfaces,” *IEEE Open J. Commun. Society*, vol. 5, pp. 599–611, Jan. 2024.
- [21] A. Papazafeiropoulos, P. Kourtessis, S. Chatzinotas, D. I. Kaklamani, and I. S. Venieris, “Near-field beamforming for stacked intelligent metasurfaces-assisted MIMO networks,” *IEEE Wireless Commun. Letters*, vol. 13, no. 11, pp. 3035–3039, Nov. 2024.
- [22] Q. -U. -A. Nadeem, J. An, and A. Chaaban, “Hybrid digital-wave domain channel estimator for stacked intelligent metasurface enabled multi-user MISO systems,” in *Proc. IEEE WCNC*, Dubai, United Arab Emirates, Apr. 2024, pp. 1–6.
- [23] D. Darsena, F. Verde, I. Iudice, and V. Galdi, “Design of stacked intelligent metasurfaces with reconfigurable amplitude and phase for multiuser downlink beamforming,” *arXiv:2408.16606*, Aug. 2024.
- [24] A. Papazafeiropoulos, P. Kourtessis, and S. Chatzinotas, “Performance of double-stacked intelligent metasurface-assisted multiuser massive MIMO communications in the wave domain,” *arXiv:2402.16405*, Feb. 2024.
- [25] J. An *et al.*, “Stacked intelligent metasurfaces for multiuser downlink beamforming in the wave domain,” *IEEE Trans. Wireless Commun.*, vol. 24, no. 7, pp. 5525–5538, July 2025.
- [26] X. Yao, J. An, L. Gan, M. Di Renzo, and C. Yuen, “Channel estimation for stacked intelligent metasurface-assisted wireless networks,” *IEEE Wireless Commun. Letters*, vol. 13, no. 5, pp. 1349–1353, May 2024.
- [27] J. An *et al.*, “Two-dimensional direction-of-arrival estimation using stacked intelligent metasurfaces,” *IEEE J. Sel. Areas Commun.*, vol. 42, no. 10, pp. 2786–2802, Oct. 2024.
- [28] S. Lin *et al.*, “Stacked intelligent metasurfaces enabled LEO satellite communications relying on statistical CSI,” *IEEE Wireless Commun. Letters*, vol. 13, no. 5, pp. 1295–1299, May 2024.
- [29] H. Niu, X. Lei, J. An, L. Zhang, and C. Yuen, “On the efficient design of stacked intelligent metasurfaces for secure SISO transmission,” *IEEE Trans. Inf. Forensics Secur.*, vol. 20, pp. 60–70, Nov. 2024.
- [30] G. Huang, J. An, Z. Yang, L. Gan, M. Bennis, and M. Debbah, “Stacked intelligent metasurfaces for task-oriented semantic communications,” *IEEE Wireless Commun. Letters*, vol. 14, no. 2, pp. 310–314, Feb. 2025.
- [31] Q. Li *et al.*, “Stacked intelligent metasurfaces for holographic MIMO aided cell-free networks,” *IEEE Trans. Commun.*, vol. 72, no. 11, pp. 7139–7151, Nov. 2024.
- [32] E. Shi, J. Zhang, J. An, G. Zhang, Z. Liu, C. Yuen, and B. Ai, “Joint AP-UE association and precoding for SIM-aided cell-free massive MIMO systems,” *IEEE Trans. Wireless Commun.*, vol. 24, no. 6, pp. 5352–5367, June 2025.
- [33] Y. Sun *et al.*, “Dual-polarized stacked metasurface transceiver design with rate splitting for next generation wireless networks,” *IEEE J. Sel. Areas Commun.*, Jan. 2025, doi: 10.1109/JSAC.2025.3531526.
- [34] H. Liu *et al.*, “Stacked intelligent metasurfaces for wireless sensing and communication: Applications and challenges,” *IEEE Wireless Commun.*, vol. 32, no. 4, pp. 46–53, Aug. 2025.
- [35] Z. Wang *et al.*, “Multi-user ISAC through stacked intelligent metasurfaces: New algorithms and experiments,” in *Proc. GLOBECOM’24*, Cape Town, South Africa, Dec. 2024, pp. 4442–4447.
- [36] H. Niu, J. An, A. Papazafeiropoulos, L. Gan, S. Chatzinotas, and M. Debbah, “Stacked intelligent metasurfaces for integrated sensing and communications,” *IEEE Wireless Commun. Letters*, vol. 13, no. 10, pp. 2807–2811, Oct. 2024.
- [37] H. Liu, J. An, G. C. Alexandropoulos, D. W. K. Ng, C. Yuen, and L. Gan, “Multi-user MISO with stacked intelligent metasurfaces: A DRL-based sum-rate optimization approach,” *IEEE Trans. Cognitive Commun. Netw.*, doi: 10.1109/TCCN.2025.3558008.
- [38] A. Mohammadzadeh *et al.*, “Meta reinforcement learning empowered orchestration of SIM and RIS for downlink multiuser communications,” *arXiv:Preprint*, Sept. 2024.
- [39] 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.*, doi: 10.1109/TAP.2025.3571069.
- [40] M. D. Renzo, “State of the art on stacked intelligent metasurfaces: Communication, sensing and computing in the wave domain,” in *Proc. EuCAP*, Stockholm, Sweden, Apr. 2025, pp. 1–3.
- [41] Z. Li, J. An, and C. Yuen, “Stacked intelligent metasurfaces-enhanced MIMO OFDM wideband communication systems,” *arXiv:2503.00368*, Mar. 2025.
- [42] Z. Li, J. An, and C. Yuen, “Stacked intelligent metasurfaces for fully-analog wideband beamforming design,” in *Proc. IEEE VTS APWCS*, Singapore, Aug. 2024, pp. 1–5.
- [43] E. Basar, “Index modulation techniques for 5G wireless networks,” *IEEE Commun. Mag.*, vol. 54, no. 7, pp. 168–175, July 2016.
- [44] E. Basar, M. Wen, R. Mesleh, M. Di Renzo, Y. Xiao, and H. Haas, “Index modulation techniques for next-generation wireless networks,” *IEEE Access*, vol. 5, pp. 16693–16746, Aug. 2017.
- [45] E. Basar, U. Aygolu, E. Panayirci, and H. V. Poor, “Orthogonal frequency division multiplexing with index modulation,” *IEEE Trans. Sig. Process.*, vol. 61, no. 22, pp. 5536–5549, Nov. 2013.
- [46] T. Mao, Q. Wang, Z. Wang, and S. Chen, “Novel index modulation techniques: A Survey,” *IEEE Commun. Surveys Tuts.*, vol. 21, no. 1, pp. 315–348, Firstquarter 2019.
- [47] E. K. Hidir, E. Basar, and H. A. Cirpan, “On Practical RIS-Aided OFDM With Index Modulation,” *IEEE Access*, vol. 11, pp. 13113–13120, Feb. 2023.
- [48] E. K. Hidir, E. Basar, and H. A. Cirpan, “Over-the-Air OFDM-IM Through Frequency Mixing and Modulating Reconfigurable Intelligent Surfaces,” *IEEE Open J. Commun. Society*, vol. 5, pp. 4531–4540, July 2024.
- [49] K. Gregor and Y. LeCun, “Learning fast approximations of sparse coding,” in *Proc. ICML*, Madison, WI, USA: Omnipress, 2010, p. 399–406.
- [50] V. Monga, Y. Li, and Y. C. Eldar, “Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing,” *IEEE Signal Process. Mag.*, vol. 38, no. 2, pp. 18–44, Mar. 2021.
- [51] S. Deka, K. Deka, N. T. Nguyen, S. Sharma, V. Bhatia, and N. Rajatheva, “Comprehensive review of deep unfolding techniques for next-generation wireless communication systems,” *arXiv:2502.05952*, July 2025.
- [52] L. Pellaco, M. Bengtsson, and J. Jaldén, “Matrix-inverse-free deep unfolding of the weighted MMSE beamforming algorithm,” *IEEE Open J. Commun. Soc.*, vol. 3, pp. 65–81, Dec. 2021.
- [53] L. Schynol and M. Pesavento, “Coordinated sum-rate maximization in multicell MU-MIMO with deep unrolling,” *IEEE J. Sel. Areas Commun.*, vol. 41, no. 4, pp. 1120–1134, Feb. 2023.
- [54] C. Yi *et al.*, “Distributed multiuser precoding via unfolding WMMSE for multicell MIMO downlinks,” *IEEE Trans. Cognitive Commun. Netw.*, vol. 11, no. 4, pp. 2185–2200, Aug. 2025.
- [55] X. Lin, Y. Rivenon, N. T. Yardimci, M. Veli, Y. Luo, M. Jarrahi, and A. Ozcan, “All-optical machine learning using diffractive deep neural networks,” *Sci.*, vol. 361, no. 6406, pp. 1004–1008, July 2018.
- [56] H. Jafarkhani, *Space-Time Coding*. Cambridge, UK: Cambridge Univ. Press, 2005.
- [57] 3GPP, *Study on channel model for frequencies from 0.5 To 100 GHz (Release 17)*, document TR 38.901, V17.0.0, Mar. 2022.