Unlocking Symbol-Level Precoding Efficiency Through Tensor Equivariant Neural Network

Jinshuo Zhang, Graduate Student Member, IEEE, Yafei Wang, Graduate Student Member, IEEE, Xinping Yi, Member, IEEE, Wenjin Wang, Member, IEEE, Shi Jin, Fellow, IEEE, Symeon Chatzinotas, Fellow, IEEE, Björn Ottersten, Fellow, IEEE

Abstract—Although symbol-level precoding (SLP) based on constructive interference (CI) exploitation offers performance gains, its high complexity remains a bottleneck. This paper addresses this challenge with an end-to-end deep learning (DL) framework with low inference complexity that leverages the structure of the optimal SLP solution in the closed-form and its inherent tensor equivariance (TE), where TE denotes that a permutation of the input induces the corresponding permutation of the output. Building upon the computationally efficient model-based formulations, as well as their known closed-form solutions, we analyze their relationship with linear precoding (LP) and investigate the corresponding optimality condition. We then construct a mapping from the problem formulation to the solution and prove its TE, based on which the designed networks reveal a specific parameter-sharing pattern that delivers low computational complexity and strong generalization. Leveraging these, we propose the backbone of the framework with an attention-based TE module, achieving linear computational complexity. Furthermore, we demonstrate that such a framework is also applicable to imperfect CSI scenarios, where we design a TE-based network to map the CSI, statistics, and symbols to auxiliary variables. Simulation results show that the proposed framework captures substantial performance gains of optimal SLP, while achieving an approximately 80-times speedup over conventional methods and maintaining strong generalization across user numbers and symbol block lengths.

Index Terms—Symbol-level precoding, deep learning, tensor equivariance, imperfect channel state information.

I. INTRODUCTION

N multi-input multi-output (MIMO) systems, precoding is a crucial technique for eliminating inter-user interference and enhancing system capacity [1], [2]. Conventional linear precoding (LP), such as zero-forcing (ZF) and minimum mean square error (MMSE) precoding [3], has low computational complexity but fails to achieve optimality. In contrast, symbollevel precoding (SLP) leverages both CSI and instantaneous transmit symbols to optimize performance on a per-symbol basis, thereby exhibiting superior performance potential while incurring higher computational complexity [4]–[25].

Jinshuo Zhang, Yafei Wang, and Wenjin Wang are with the National Mobile Communications Research Laboratory, Southeast University, Nanjing 210096, China, and also with Purple Mountain Laboratories, Nanjing 211100, China (e-mail: zhangjs@seu.edu.cn; wangyf@seu.edu.cn; wangwj@seu.edu.cn).

Xinping Yi and Shi Jin are with the National Mobile Communications Research Laboratory, Southeast University, Nanjing 210096, China (e-mail: xyi@seu.edu.cn; jinshi@seu.edu.cn).

Symeon Chatzinotas and Björn Ottersten are with the Interdisciplinary Centre for Security, Reliability and Trust (SnT), University of Luxembourg (E-mails: {symeon.chatzinotas, bjorn.ottersten}@uni.lu).

SLP achieves performance gains by effectively leveraging constructive interference (CI), which shifts the received signal constellation points away from their decision boundaries. Based on CI constraints, the symbol-level power minimization (PM) problem [9]-[12], [14], [15], [23] and signal-tointerference-plus-noise ratio (SINR) balancing problem [5], [9], [13], [15], [24] have been extensively studied. The minimization of symbol error rate (SER) was investigated in [17], [18], [25]. Notably, the authors of [26] proposed a simple and generic CI region (CIR), based on which they presented a symbol-level MMSE precoding solution. Moreover, available CSI may contain errors, making the modeling and estimation of imperfect channels an active research area [27]–[31], Consequently, the development of robust SLP techniques has garnered significant research interest in recent years [9], [19]– [22]. In particular, leveraging the a posteriori channel model [32], the work in [33], [34] investigated robust SLP designs optimizing for SINR and MMSE criteria under imperfect CSI, achieving promising performance.

Despite their superior performance, SLP algorithms suffer from inherently high computational complexity due to the dependence on the instantaneous transmit symbols to all users, which limits their practical deployment. Research on equivalent SLP formulations has progressively reduced computational complexity, with the problem evolving from a second-order cone programming (SOCP) [13] to a quadratic programming (QP) [5], [13] and finally to a non-negative least squares (NNLS) formulation [16], [26], each step yielding more efficient algorithms. However, these methods still face challenges related to the high computational cost of iterative optimization compared to LP. To address this, several lowcomplexity designs have been proposed. For the OP formulation, an unsupervised learning-based framework was proposed in [35]. Regarding the CI-constrained PM problem, the work in [36] proposed an approximation algorithm for the optimal solution of its NNLS formulation, which benefits from closedform computation and demonstrates excellent performance.

However, to the best of our knowledge, existing low-complexity schemes have not fully explored the NNLS for-mulation and its inherent properties to reduce computational complexity. Furthermore, although several low-complexity approaches have investigated approximate closed-form solutions [36], [37], they are typically restricted to specific constellations under perfect CSI. A general method applicable to multi-level QAM and PSK under both perfect and imperfect CSI remains to be explored. Meanwhile, permutation equivari-

ance/invariance (PE/PI), and their high-dimensional extension, termed tensor equivariance (TE) [38], have inspired dedicated TE neural networks (TENN). These networks have proven remarkably successful in LP design, detection, and soft demodulation [38]–[41], suggesting their potential in the design of SLP. This leads to the central research question: *How can its inherent tensor equivariance properties be exploited in designing an end-to-end learning-to-optimize solution with lower computational complexity?* This work aims to answer this question. By leveraging the NNLS formulation and its inherent TE, we develop a unified deep learning (DL)-based SLP framework supporting various modulation schemes under perfect CSI and also extending to imperfect CSI scenarios. In summary, the main contributions of this paper are as follows:

- Building upon the established NNLS formulation of typical SLP problems under SINR balancing and MMSE criteria and their known closed-form solutions, we analyze their relationship to LP and investigate the corresponding Karush-Kuhn-Tucker (KKT) conditions. Furthermore, we define a mapping from the information embedded in the KKT conditions to the optimal perturbation factors that adjust the symbols within the CIR, and then analyze the TE of this mapping, including multidimensional equivariance (MDE) and high-order equivariance (HOE).
- To leverage the properties of the formulated mapping, we propose a linear-complexity attention-based MDE (AMDE) module that achieves a strong representation power through a novel TE attention mechanism. Building on AMDE, we develop the SLPN network, which exploits the TE and NNLS formulation of the SLP problem, providing linear computational complexity, low parameter count, and inherent generalization for input sizes. The network was trained on datasets from diverse channel environments, enabling its deployment across different channel realizations. To further enhance performance, we propose a post-net refinement method which applies a lightweight computational step after SLPN to scale the obtained perturbation factors. By integrating these components, we construct an SLP framework that embodies these advantages, demonstrating applicability across diverse SLP schemes and constellations.
- For scenarios with potentially imperfect CSI, we extend the proposed SLP framework. Specifically, we design a low-complexity network RSLPN to implement the MMSE robust SLP scheme, which also leverages the inherent TE. The RSLPN estimates the auxiliary variable and perturbation factor of the closed-form solution, bypassing the required iterations while preserving the advantages afforded by TE. Simulation results show that the proposed TENN-based SLP framework retain most of the performance gains of optimal SLP, while achieving an approximately 80-fold speedup over conventional methods in typical scenarios and maintaining strong generalization across both user numbers and symbol block lengths.

The rest of the article is organized as follows: The system model and CIR are built in Section II. The low complexity SLP framework based on TENN is investigated in Section III, and Section IV subsequently extends this framework to imperfect CSI scenarios. Simulation results are provided in Section V, and Section VI concludes this article.

Notation: x, x, X, and X denote a scalar, column vector, matrix, and tensor, respectively. The operations of transpose, conjugate, conjugate transpose, and matrix inversion are denoted by $(\cdot)^T$, $(\cdot)^*$, $(\cdot)^H$, and $(\cdot)^{-1}$, respectively. $\mathbb{R}^{M\times N}$ ($\mathbb{C}^{M\times N}$) denotes the $M \times N$ dimensional real (complex) matrix space. $diag\{a\}$ denotes the diagonal matrix with a on its diagonal. $\Re(\cdot)$ and $\Im(\cdot)$ extract the real and imaginary parts of a complex quantity. $\mathcal{R}(\cdot)$ maps complex vectors/matrices to their real representations: $\mathcal{R}(\mathbf{a}) = [\Re(\mathbf{a})^T, \Im(\mathbf{a})^T]^T$ and $\mathcal{R}(\mathbf{A}) =$ $[\Re(\mathbf{A}), -\Im(\mathbf{A}); \Im(\mathbf{A}), \Re(\mathbf{A})]$. \mathbf{I}_N denotes the $N \times N$ identity matrix, and $\|\cdot\|_2$ denotes l_2 -norm. \otimes and \odot are the Kronecker and Hardmard product. $[\mathbf{X}]_{i,j}$ and $[\mathbf{x}]_i$ denotes the (i,j)-th and i-th element of X and x, respectively. $[X_1, \dots, X_K]_D$ denotes the tensor formed by stacking $\mathbf{X}_1, \dots, \mathbf{X}_K$ along the D-th dimension. $[\cdot]_{(D)}$ denotes the concatenation of tensors along the newly inserted D-th dimension. $x \succ 0$ means all the elements of x are nonnegative. $\mathcal{CN}(\mu, \sigma^2)$ denotes circularly symmetric Gaussian distribution with expectation μ and variance σ^2 . The Kronecker product of tensor and matrix is defined as $(\mathbf{X} \otimes_n \mathbf{Y})_{[m_1,\dots,m_{n-1},\dots,m_{n+2},\dots,m_N]} =$ $\mathbf{X}_{[m_1,\ldots,m_{n-1},\ldots,m_{n+2},\ldots,m_N]} \otimes \mathbf{Y}.$

II. SYSTEM MODEL AND SYMBOL-LEVEL PRECODING

A. System Model

Consider a MIMO downlink transmission system, where a base station (BS) with $N_{\rm T}$ antennas serves K single-antenna user equipments (UE). Each transmission time slot comprises two phases: (i) an uplink pilot phase for CSI acquisition, and (ii) a subsequent downlink data phase transmitting L symbols. We assume block flat fading channels, whose coefficients remain constant over a coherence interval of L symbol durations. Let $\mathbf{h}_k \in \mathbb{C}^{N_{\rm T} \times 1}$ denote the channel vector between the BS and the k-th user, which is obtained during the CSI acquisition phase. The signal received by the k-th user at the l-th downlink symbol is given by

$$y_k[l] = \mathbf{h}_k^T \mathbf{x}_c[l] + n_k[l], \ \forall k \in \mathcal{K}, \forall l \in \mathcal{L},$$
 (1)

where $\mathcal{L} \triangleq \{1, 2, ..., L\}$, $\mathcal{K} \triangleq \{1, 2, ..., K\}$, $\mathbf{x}_{c}[l] \in \mathbb{C}^{N_{\mathrm{T}} \times 1}$ denotes the precoded signal vector transmitted by the BS during the l-th symbol, and $n_{k}[l]$ denotes the noise at the k-th user, i.e., $n_{k}[l] \sim \mathcal{CN}(0, \sigma^{2})$.

B. Constructive Interference Region

CI refers to interference that shifts the received signal constellation point further from the decision boundaries [5], [13], [26]. For a specific transmit symbol, its CIR is defined as the area in the complex plane encompassing all signal points that undergo such CI. Fig. 1 illustrates the CIR (deep blue regions), boundary vectors, and maximum likelihood (ML) decision boundaries (dashed lines) for both PSK and QAM modulations. The typical CIR of s_k can be expressed as [26]

$$\mathcal{D}_k = \{ \tilde{s}_k | \tilde{s}_k = s_k + \delta_{\mu_k} \mu_k + \delta_{\nu_k} \nu_k, \ \delta_{\mu_k}, \delta_{\nu_k} > 0 \}, \quad (2)$$

O Constellation point Perturbed constellation point

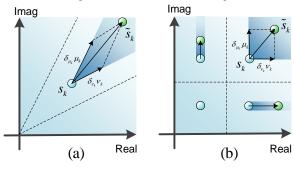


Fig. 1. CIRs and their boundary vectors of (a) PSK and (b) QAM.

where μ_k and ν_k are two complex parameters representing the boundaries of the CIR for s_k , while δ_{μ_k} and δ_{ν_k} are perturbation factors, shifting the transmit symbol within the CIR. For notational convenience, we define $\boldsymbol{\mu} = [\mu_1, \cdots, \mu_K]^T$, $\boldsymbol{\nu} = [\nu_1, \cdots, \nu_K]^T$, $\boldsymbol{\Lambda}_{\mu} = \mathrm{diag}\{\boldsymbol{\mu}\}$, $\boldsymbol{\Lambda}_{\nu} = \mathrm{diag}\{\boldsymbol{\nu}\}$, $\boldsymbol{\delta}_{\mu} = [\delta_{\mu_1}, \ldots, \delta_{\mu_K}]^T$, $\boldsymbol{\delta}_{\nu} = [\delta_{\nu_1}, \ldots, \delta_{\nu_K}]^T$, $\mathbf{s}_{\mathrm{c}} = [s_1, s_2, \ldots, s_K]^T$, and $\tilde{\mathbf{s}}_{\mathrm{c}} = [\tilde{s}_1, \tilde{s}_2, \ldots, \tilde{s}_K]^T$, then $\tilde{\mathbf{s}}_{\mathrm{c}}$ is expressed as

$$\tilde{\mathbf{s}}_{c} = \mathbf{s}_{c} + \mathbf{\Lambda}_{\mu} \boldsymbol{\delta}_{\mu} + \mathbf{\Lambda}_{\nu} \boldsymbol{\delta}_{\nu}. \tag{3}$$

Let $\mathbf{s} = \mathcal{R}(\mathbf{s}_c)$ and $\tilde{\mathbf{s}} = \mathcal{R}(\tilde{\mathbf{s}}_c)$ be the real-valued representations of \mathbf{s}_c and $\tilde{\mathbf{s}}_c$, respectively. Define $\boldsymbol{\delta} = [\delta_{\mu_1}, \dots, \delta_{\mu_K}, \delta_{\nu_1}, \dots, \delta_{\nu_K}]^T$ and

$$\mathbf{\Lambda} = \begin{bmatrix} \operatorname{diag}\{\Re(\boldsymbol{\mu})\} & \operatorname{diag}\{\Re(\boldsymbol{\nu})\} \\ \operatorname{diag}\{\Im(\boldsymbol{\mu})\} & \operatorname{diag}\{\Im(\boldsymbol{\nu})\} \end{bmatrix}, \tag{4}$$

thus the real-valued equivalent representation of $\tilde{\mathbf{s}}_c$ is given by

$$\tilde{\mathbf{s}} = \mathbf{s} + \mathbf{\Lambda} \boldsymbol{\delta}. \tag{5}$$

This representation provides an effective formulation of the CIR, establishing a foundation for analyzing SLP problems and deriving their solutions in the subsequent sections.

C. Optimization-Based SLP Design

The correct demodulation of multi-level QAM requires knowledge of the scaling factor at the receiver. Denoting the scaling factor for the l-th symbol by $\gamma[l]$, which can be sent by the BS or estimated at the receiver, the signal for demodulation at the k-th UE is expressed as [17], [42], [43]

$$\tilde{y}_k[l] = y_k[l]/\gamma[l] = \mathbf{h}_k^T \mathbf{x}_c[l]/\gamma[l] + n_k[l]/\gamma[l], \qquad (6)$$

1) CIZF: Although the CIZF precoding is essentially a CI-constrained SINR balancing (CISB) problem, its closed-form expression resembles that of ZF, so we refer to it as CIZF precoding. Specifically, it maximizes the scaling factor while satisfying the power constraint, expressed by [9], [43]

$$\max_{\mathbf{x}_{c}[l], \tilde{\mathbf{s}}_{c}[l], \gamma[l]} \gamma[l]
s.t. \mathbf{h}_{k}^{T} \mathbf{x}_{c}[l] = \gamma[l] \cdot \tilde{s}_{k}[l], \ \forall k \in \mathcal{K},
\tilde{s}_{k}[l] \in \mathcal{D}_{k}[l], \ \forall k \in \mathcal{K},
\|\mathbf{x}_{c}[l]\|_{2}^{2} \leq P_{T},$$
(7)

where $P_{\rm T}$ denotes the transmit power. This problem can be optimally addressed by first solving a convex PM problem

and then scaling the solution to meet the power constraint [16]. The PM problem is formulated as follows

$$\min_{\boldsymbol{\delta}_{\mu}[l] \succeq \mathbf{0}, \boldsymbol{\delta}_{\nu}[l] \succeq \mathbf{0}} \tilde{\mathbf{s}}_{c}^{H}[l] \boldsymbol{\Upsilon}_{\mathrm{ZF}} \tilde{\mathbf{s}}_{c}[l]
s.t. \, \tilde{\mathbf{s}}_{c}[l] = \mathbf{s}_{c}[l] + \boldsymbol{\Lambda}_{\mu}[l] \boldsymbol{\delta}_{\mu}[l] + \boldsymbol{\Lambda}_{\nu}[l] \boldsymbol{\delta}_{\nu}[l],$$
(8)

where $\Upsilon_{\mathrm{ZF}} = (\mathbf{H}\mathbf{H}^H)^{-1}$ and $\mathbf{H} = [\mathbf{h}_1, \cdots, \mathbf{h}_K]^T$. This problem is equal to the following NNLS problem

$$\min_{\boldsymbol{\delta}[l] \succ \mathbf{0}} \| \mathbf{H}_{r}^{\dagger} \boldsymbol{\Lambda}[l] \boldsymbol{\delta}[l] + \mathbf{H}_{r}^{\dagger} \mathbf{s}[l] \|_{2}^{2}, \tag{9}$$

where $\mathbf{H}_r^{\dagger} = \mathcal{R}(\mathbf{H}^{\dagger}) \in \mathbb{R}^{2N_{\mathrm{T}} \times 2K}$ and $\mathbf{H}^{\dagger} = \mathbf{H}^H(\mathbf{H}\mathbf{H}^H)^{-1}$. The definitions of $\boldsymbol{\delta}[l]$, $\boldsymbol{\Lambda}[l]$, and $\mathbf{s}[l]$ are provided in Section II-B. This NNLS problem can be solved by the active set-based algorithm [44] through an iterative procedure. After further considering the transmit power constraint, the structure of the optimal solution to problem (7) is expressed as

$$\mathbf{x}_{c}^{\star}[l] = \gamma^{\star}[l]\mathbf{H}^{\dagger}\tilde{\mathbf{s}}_{c}^{\star}[l], \ \gamma^{\star}[l] = \sqrt{P_{T}/\|\mathbf{H}^{\dagger}\tilde{\mathbf{s}}_{c}^{\star}[l]\|_{2}^{2}},$$
(10)

$$\tilde{\mathbf{s}}_{c}^{\star}[l] = \mathbf{s}_{c}[l] + \mathbf{\Lambda}_{\mu}[l]\boldsymbol{\delta}_{\mu}[l] + \mathbf{\Lambda}_{\nu}[l]\boldsymbol{\delta}_{\nu}[l]. \tag{11}$$

2) CIMMSE: The MMSE problem minimizes the mean square error (MSE) between transmit and receive symbols, and with CI constraints it is formulated as [26]

$$\min_{\mathbf{x}_{c}[l], \tilde{\mathbf{s}}_{c}[l], \gamma[l]} \mathbb{E}_{\mathbf{n}[l]} \left\{ \left\| \left(\mathbf{H} \mathbf{x}_{c}[l] + \mathbf{n}[l] \right) / \gamma[l] - \tilde{\mathbf{s}}_{c}[l] \right\|_{2}^{2} \right\} \\
\text{s.t. } \tilde{s}_{k}[l] \in \mathcal{D}_{k}[l], \ \forall k \in \mathcal{K}, \\
\|\mathbf{x}_{c}[l]\|_{2}^{2} \leq P_{T},$$
(12)

where $\mathbf{n}[l] = [n_1[l], \dots, n_K[l]]^T$. This problem can be optimally solved by applying a scaling method to the solution of its convex power-unconstrained counterpart, which is stated as follows [26]

$$\min_{\boldsymbol{\delta}_{\mu}[l] \succeq \mathbf{0}, \boldsymbol{\delta}_{\nu}[l] \succeq \mathbf{0}} \tilde{\mathbf{s}}_{c}^{H}[l] \boldsymbol{\Upsilon}_{\text{MMSE}} \tilde{\mathbf{s}}_{c}[l]
\text{s.t. } \tilde{\mathbf{s}}_{c}[l] = \mathbf{s}_{c}[l] + \boldsymbol{\Lambda}_{\mu}[l] \boldsymbol{\delta}_{\mu}[l] + \boldsymbol{\Lambda}_{\nu}[l] \boldsymbol{\delta}_{\nu}[l],$$
(13)

where $\Upsilon_{\text{MMSE}} = \left(\mathbf{H}\mathbf{H}^H + \frac{\sigma^2 K}{P_{\text{T}}}\mathbf{I}_K\right)^{-1}$. The above problem can be formulated as an NNLS problem

$$\min_{\boldsymbol{\delta}[l] \succeq \mathbf{0}} \| \mathbf{C}_{\mathbf{u}} \boldsymbol{\Lambda}[l] \boldsymbol{\delta}[l] + \mathbf{C}_{\mathbf{u}} \mathbf{s}[l] \|_{2}^{2}, \tag{14}$$

where \mathbf{C}_{u} denotes the upper triangular matrix obtained from the Cholesky decomposition of $\left(\mathbf{H}_{\mathrm{r}}\mathbf{H}_{\mathrm{r}}^{T}+\frac{\sigma^{2}K}{P_{\mathrm{r}}}\mathbf{I}_{2K}\right)^{-1}$, with $\mathbf{H}_{\mathrm{r}}=\mathcal{R}(\mathbf{H})$. This NNLS problem can be solved using a similar approach as that for problem (9). Accounting for power constraints, the solution exhibits the following structure

$$\mathbf{x}_{c}^{\star}[l] = \gamma^{\star}[l]\mathbf{H}^{H}\mathbf{\Upsilon}_{\mathrm{MMSE}}\tilde{\mathbf{s}}_{c}^{\star}[l], \tag{15}$$

$$\gamma^{\star}[l] = \sqrt{P_{\mathrm{T}}/\|\mathbf{H}^{H}\mathbf{\Upsilon}_{\mathrm{MMSE}}\tilde{\mathbf{s}}_{\mathrm{c}}^{\star}[l]\|_{2}^{2}},\tag{16}$$

$$\tilde{\mathbf{s}}_{c}^{\star}[l] = \mathbf{s}_{c}[l] + \mathbf{\Lambda}_{\mu}[l]\boldsymbol{\delta}_{\mu}[l] + \mathbf{\Lambda}_{\nu}[l]\boldsymbol{\delta}_{\nu}[l]. \tag{17}$$

As established in [26], the optimal solution to the CIMMSE problem reduces to that of the CIZF problem in the high SNR regime (i.e., as $\sigma \to 0$).

Furthermore, a block-level power reallocation scheme is introduced to maintain a uniform rescaling factor $\bar{\gamma}$ over the entire block, which substantially reduces the per-symbol

overhead [39], [43]:

$$\bar{\gamma}^{\star} = \sqrt{\frac{L}{\sum_{l=1}^{L} \frac{1}{(\gamma^{\star}[l])^{2}}}}, \quad \bar{\mathbf{x}}_{c}^{\star}[l] = \frac{\bar{\gamma}^{\star}}{\gamma^{\star}[l]} \mathbf{x}_{c}^{\star}[l], \ \forall l \in \mathcal{L}. \quad (18)$$

where $\bar{\gamma}^{\star}$ can be acquired at the UE either via the control channel or through estimation.

It is worth noting that, for the CIZF problem, the solution obtained through the above power allocation method exactly coincides with the optimal solution of the block-level CISB problem that minimizes $\bar{\gamma}$ over a symbol block of length L and subject to the total transmit energy constraint $\sum_{l=1}^{L} \|\mathbf{x}_{\mathbf{c}}[l]\|_2^2 \leq LP_{\mathrm{T}}$ [43]. Accordingly, the signal for demodulation in (6) can be reformulated as

$$\tilde{y}_k[l] = y_k[l]/\bar{\gamma} = \mathbf{h}_k^T \bar{\mathbf{x}}_c[l]/\bar{\gamma} + n_k[l]/\bar{\gamma}.$$
 (19)

D. Symbol-Level Precoding vs. Linear Precoding

1) Sources of SLP Gains: By comparing the formulations of CIZF in (10),(11), CIMMSE in (15)-(17), and LP schemes such as ZF and MMSE [3], we can observe a key difference in the general expression between SLP and LP schemes, specifically

$$LP: \mathbf{x}_{LP} = \gamma_{LP} \mathbf{P} \mathbf{s}_{c}, \tag{20}$$

SLP:
$$\mathbf{x}_{\text{SLP}} = \gamma_{\text{SLP}} \mathbf{P} (\mathbf{s}_{c} + \mathbf{\Lambda}_{\mu} \boldsymbol{\delta}_{\mu} + \mathbf{\Lambda}_{\nu} \boldsymbol{\delta}_{\nu}),$$
 (21)

where $\mathbf{P} \in \mathbb{C}^{N_{\mathrm{T}} \times K}$ is the precoding matrix, γ_{LP} and γ_{SLP} are the power normalization factors. The key distinction of SLP from LP is its intentional perturbation of the symbols to exploit the CIR, which is the primary mechanism behind its performance gains. From a multicast perspective, these gains stem from exploiting interference constructively as an additional source of useful power [11]. Consequently, the core challenge of SLP lies in determining the perturbation factors $\delta_{\mu}[l]$ and $\delta_{\nu}[l]$. With these factors obtained, the SLP problems can be solved using the closed-form solutions given by equations (10), (11), (15)-(17), and (18). Moreover, when $N_{\mathrm{T}} \gg K$, the gain introduced by SLP vanishes under most channel conditions, with $\delta_{\mu}[l]$ and $\delta_{\nu}[l]$ converging to 0 [34].

2) Gap in Computational Efficiency: The complexity of LP mainly arises from Ps multiplication, which is $\mathcal{O}(KN_{\mathrm{T}})$ and can be efficiently computed in batches. Conversely, SLP incurs a higher complexity of $\mathcal{O}(KN_{\mathrm{T}}+KN_{\mathrm{T}}N_L+N_{\mathrm{T}}N_L^3)$, where N_L denotes the number of main loop iterations of the active set-based algorithm. Furthermore, SLP requires distinct iterative computations for each symbol, which makes it challenging to directly implement efficient batch processing.

III. A LOW-COMPLEXITY SYMBOL-LEVEL PRECODING FRAMEWORK BASED ON TENSOR EQUIVARIANCE

In this section, we propose a low-complexity framework based on TENN for solving the SLP problem. We begin by analyzing the shared formulation of the CIZF and CIMMSE problems and deriving their KKT conditions. Building on this analysis, we design a mapping from the available information of the KKT conditions into perturbation factors and subsequently analyze its inherent TE. Following this, we propose a linear-complexity attention-based MDE module, which forms the foundation for the SLP framework.

A. KKT-Based SLP Mapping

As analyzed in Section II-D, the key to solve the CIZF (7) and CIMMSE (12) problems lies in obtaining $\delta_{\mu}[l]$ and $\delta_{\nu}[l]$, which can be derived by solving the problems in (8) and (13). Given the consistency in form between the two problems, we use the general notation Υ to replace $\Upsilon_{\rm ZF}$ and $\Upsilon_{\rm MMSE}$ to facilitate a unified analysis. To gain further insight into solving this problem, we derive its Lagrangian as follows

$$\mathcal{L}(\boldsymbol{\delta}_{\mu}[l], \boldsymbol{\delta}_{\nu}[l], \boldsymbol{\lambda}_{\mu}[l], \boldsymbol{\lambda}_{\nu}[l])$$

$$= \tilde{\mathbf{s}}_{c}^{H}[l] \boldsymbol{\Upsilon} \tilde{\mathbf{s}}_{c}[l] + \boldsymbol{\lambda}_{\mu}^{T}[l] \boldsymbol{\delta}_{\mu}[l] + \boldsymbol{\lambda}_{\nu}^{T}[l] \boldsymbol{\delta}_{\nu}[l],$$
(22)

where $\lambda_{\mu}[l]$ and $\lambda_{\nu}[l]$ are Lagrange multiplier vectors. Therefore, the KKT conditions can be expressed as

$$\nabla_{\boldsymbol{\delta}_{\mu}[l]} \mathcal{L}(\boldsymbol{\delta}_{\mu}[l], \boldsymbol{\delta}_{\nu}[l], \boldsymbol{\lambda}_{\mu}[l], \boldsymbol{\lambda}_{\nu}[l]) = 0, \tag{23a}$$

$$\nabla_{\boldsymbol{\delta}_{\nu}[l]} \mathcal{L}(\boldsymbol{\delta}_{\mu}[l], \boldsymbol{\delta}_{\nu}[l], \boldsymbol{\lambda}_{\mu}[l], \boldsymbol{\lambda}_{\nu}[l]) = 0, \tag{23b}$$

$$\boldsymbol{\delta}_{\mu}[l] \succeq 0, \ \boldsymbol{\lambda}_{\mu}[l] \preceq 0, \ \boldsymbol{\lambda}_{\mu}^{T}[l] \boldsymbol{\delta}_{\mu}[l] = 0,$$
 (23c)

$$\boldsymbol{\delta}_{\nu}[l] \succeq 0, \ \boldsymbol{\lambda}_{\nu}[l] \preceq 0, \ \boldsymbol{\lambda}_{\nu}^{T}[l] \boldsymbol{\delta}_{\nu}[l] = 0,$$
 (23d)

and we have

$$\nabla_{\boldsymbol{\delta}_{\mu}[l]} \mathcal{L}(\boldsymbol{\delta}_{\mu}[l], \boldsymbol{\delta}_{\nu}[l], \boldsymbol{\lambda}_{\mu}[l], \boldsymbol{\lambda}_{\nu}[l]) = \boldsymbol{\lambda}_{\mu}[l] + (24)$$

$$2\Re(\boldsymbol{\Lambda}_{\mu}^{H}[l]\boldsymbol{\Upsilon}\mathbf{s}_{c}[l] + \boldsymbol{\Lambda}_{\mu}^{H}[l]\boldsymbol{\Upsilon}\boldsymbol{\Lambda}_{\mu}[l]\boldsymbol{\delta}_{\mu}[l] + \boldsymbol{\Lambda}_{\mu}^{H}[l]\boldsymbol{\Upsilon}\boldsymbol{\Lambda}_{\nu}[l]\boldsymbol{\delta}_{\nu}[l]),$$

$$\nabla_{\boldsymbol{\delta}_{\nu}[l]} \mathcal{L}(\boldsymbol{\delta}_{\mu}[l], \boldsymbol{\delta}_{\nu}[l], \boldsymbol{\lambda}_{\mu}[l], \boldsymbol{\lambda}_{\nu}[l]) = \boldsymbol{\lambda}_{\nu}[l] + (25)$$

$$2\Re(\boldsymbol{\Lambda}_{\nu}^{H}[l]\boldsymbol{\Upsilon}\mathbf{s}_{c}[l] + \boldsymbol{\Lambda}_{\nu}^{H}[l]\boldsymbol{\Upsilon}\boldsymbol{\Lambda}_{\mu}[l]\boldsymbol{\delta}_{\mu}[l] + \boldsymbol{\Lambda}_{\nu}^{H}[l]\boldsymbol{\Upsilon}\boldsymbol{\Lambda}_{\nu}[l]\boldsymbol{\delta}_{\nu}[l]).$$

Since problems (8) and (13) can be mathematically reformulated as problems (9) and (14), respectively, the aforementioned KKT conditions also hold for problems (9) and (14), differing only in their real-valued and complex-valued representations. We then define the bias term $\mathbf{B}_{\mathrm{c}}[l]$ and the coefficient term $\mathbf{C}_{\mathrm{c}}[l]$ of the above gradient expressions, as follows:

$$\mathbf{B}_{c}[l] = \left[\mathbf{\Lambda}_{\mu}^{H}[l] \mathbf{\Upsilon} \mathbf{s}_{c}[l], \mathbf{\Lambda}_{\nu}^{H}[l] \mathbf{\Upsilon} \mathbf{s}_{c}[l] \right] \in \mathbb{C}^{K \times 2}, \tag{26}$$

$$\mathbf{C}_{c}[l] = \left[\mathbf{\Lambda}_{\mu}^{H}[l] \mathbf{\Upsilon} \mathbf{\Lambda}_{\mu}[l], \mathbf{\Lambda}_{\mu}^{H}[l] \mathbf{\Upsilon} \mathbf{\Lambda}_{\nu}[l], \right.$$

$$\left. \mathbf{\Lambda}_{\nu}^{H}[l] \mathbf{\Upsilon} \mathbf{\Lambda}_{\mu}[l], \mathbf{\Lambda}_{\nu}^{H}[l] \mathbf{\Upsilon} \mathbf{\Lambda}_{\nu}[l] \right]_{3} \in \mathbb{C}^{K \times K \times 4}, \tag{27}$$

where $\mathbf{B}_{\mathrm{c}}[l]$ and $\mathbf{C}_{\mathrm{c}}[l]$ contain the available information from the KKT conditions. We define $\mathbf{D}[l] = [\boldsymbol{\delta}_{\mu}^{\star}[l], \boldsymbol{\delta}_{\nu}^{\star}[l]] \in \mathbb{R}^{K \times 2}$, and by stacking all the relevant information of the L symbols together, we obtain

$$\mathbf{B}_{c} = \left[\mathbf{B}_{c}[1], \mathbf{B}_{c}[2], \dots, \mathbf{B}_{c}[L]\right]_{(2)} \in \mathbb{C}^{K \times L \times 2}, \tag{28}$$

$$\mathbf{C}_{c} = \left[\mathbf{C}_{c}[1], \mathbf{C}_{c}[2], \dots, \mathbf{C}_{c}[L]\right]_{(3)} \in \mathbb{C}^{K \times K \times L \times 4},$$
 (29)

$$\mathbf{D} = \left[\mathbf{D}[1], \mathbf{D}[2], \dots, \mathbf{D}[L]\right]_{(2)} \in \mathbb{R}^{K \times L \times 2}.$$
 (30)

Since (8) and (13) are convex problems, their optimal solution \mathbf{D}^{\star} can be determined by their KKT information \mathbf{B}_{c} and \mathbf{C}_{c} . We define a mapping from the available information within the KKT conditions to the optimal solution of the problem:

$$G(\mathbf{B}_c, \mathbf{C}_c) = \mathbf{D}^*. \tag{31}$$

Although the optimization problems across symbol periods are theoretically decoupled, we define a unified mapping over a block of L symbols. This design aims to enable the subsequent proposed neural network to capture the underlying statistical

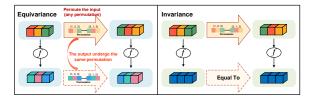


Fig. 2. Some examples of equivariance and invariance.

patterns and shared structures that persist across different symbol groups. Through offline training on channel realizations in a pre-collected dataset, the neural network is expected to learn an approximation of mapping G, enabling deployment to unseen channel realizations. Moreover, in the subsequent parts, we further exploit properties of this mapping to achieve a significant reduction in both computational complexity and parameter count, as well as strong generalization with respect to user numbers and symbol block lengths.

B. Tensor Equivariance in SLP Design

Before analyzing the properties of the mapping G, we first introduce the definition of TE. We begin with the PE in the one-dimensional (1-D) case. Consider $\mathbf{d} \in \mathbb{C}^{D\times 1}$, $f: \mathbb{C}^{D\times 1} \to \mathbb{C}^{O\times 1}$, and let \mathbb{S}_D denote the set of all permutations of D indices. For any permutation $\pi_D \in \mathbb{S}_D$, the operator $\pi_D \circ_n$ applies π_D along the n-th dimension [45]. Assuming D = O, the map f exhibits PE if [46]–[48]

$$f(\pi_D \circ_1 \mathbf{d}) = \pi_D \circ_1 f(\mathbf{d}), \ \forall \pi_D \in \mathbb{S}_D.$$
 (32)

This implies that a permutation applied to the input induces a corresponding permutation in the ordering of elements in the output. Similarly, if the mapping f exhibits PI, we have [49]

$$f(\pi_D \circ_1 \mathbf{d}) = f(\mathbf{d}), \ \forall \pi_D \in \mathbb{S}_D.$$
 (33)

This implies that permuting the indices of the input does not alter the output. Fig. 2 illustrates some examples of PE and PI. When both the input and output are tensors, and PE holds independently across multiple dimensions, the property is referred to as MDE or N-D equivariance. Conversely, if PI holds across multiple dimensions, the property is termed multidimensional invariance (MDI) or N-D invariance. When the same PE is satisfied across multiple dimensions of both the input and the output, it is defined as HOE or p-q order equivariance [50]. Furthermore, TE serves as the collective term for these higher-dimensional extensions of equivariance/invariance properties, following the detailed definitions in [38].

Proposition 1. The mapping $G(\mathbf{B}_c, \mathbf{C}_c) = \mathbf{D}^*$ (31) defined based on problems for CIZF and CIMMSE in (7) and (12) exhibits the following properties

$$G(\pi_K \circ_1 \mathbf{B}_{\mathbf{c}}, \pi_K \circ_{[1,2]} \mathbf{C}_{\mathbf{c}}) = \pi_K \circ_1 \mathbf{D}^{\star}, \ \forall \pi_K \in \mathbb{S}_K, \ (34a)$$
$$G(\pi_L \circ_2 \mathbf{B}_{\mathbf{c}}, \pi_L \circ_3 \mathbf{C}_{\mathbf{c}}) = \pi_L \circ_2 \mathbf{D}^{\star}, \ \forall \pi_L \in \mathbb{S}_L. \ (34b)$$

Proof. See Appendix A.

According to **Proposition 1**, \mathbf{B}_{c} exhibits 2-D equivariance with respect to its first and second dimensions, and \mathbf{C}_{c} exhibits

2-1-order equivariance with respect to its first and second dimensions and 1-D equivariance with respect to its third dimension.

The essence of designing neural networks for symbollevel precoding (SLP) lies in their ability to approximate the mapping G effectively. In **Proposition 1**, we have analyzed the TE of G, which suggests that networks aiming to approximate G would benefit from possessing similar TE. As established in [38], [45], [51], networks exhibiting such TE inherently incorporate distinctive parameter-sharing patterns, which lead to benefits such as reduced online computational complexity and inherent generalization for input sizes. Building on this insight, we next present the design of lightweight yet efficient TE networks with excellent generalization capabilities.

C. Attention-Based Multidimensional Equivariant Module

To achieve high performance while ensuring compliance with TE, we design an attention-based residual module, referred to as AMDE, which satisfies MDE while maintaining low online computational complexity. Since AMDE is constructed from basic TE layers [38], we begin by providing a brief introduction to them.

When a fully connected layer satisfies MDE or HOE, it follows a distinct parameter-sharing pattern. We use f_{MDE} and f_{HOE} to denote the functions that obey these MDE and HOE patterns, respectively. The formal proof is provided in [38], where several variants of this pattern are also introduced. In particular, the 2-D equivariant layer with input $\mathbf{X} \in \mathbb{R}^{M_1 \times M_2 \times F}$ can be expressed as [38]

$$f_{\text{MDE}}(\mathbf{X}) = \sum_{\mathcal{D} \in \mathbb{S}_{2}} \bar{\mathbf{X}}_{\mathcal{D}} \times \mathbf{W}_{\mathcal{D}} + \mathbf{1} \otimes_{2} \mathbf{b}^{T}$$

$$= \mathbf{X}_{\emptyset} \times \mathbf{W}_{\emptyset} + \bar{\mathbf{X}}_{\{1\}} \times \mathbf{W}_{\{1\}} + \bar{\mathbf{X}}_{\{2\}} \times \mathbf{W}_{\{2\}}$$

$$+ \bar{\mathbf{X}}_{\{1,2\}} \times \mathbf{W}_{\{1,2\}} + \mathbf{1} \otimes_{2} \mathbf{b}^{T},$$
(35)

where M_1 and M_2 are its equivariant dimensions, and F is the feature dimension. $\bar{\mathbf{X}}_{\mathcal{D}}$ denotes the tensor obtained by averaging \mathbf{X} along the dimensions \mathcal{D} and then broadcasting back to the original shape, and $\bar{\mathbf{X}}_{\emptyset} = \mathbf{X}$. The set $\mathbb{S}_2 = \{\emptyset, \{1\}, \{2\}, \{1, 2\}\}$ contains all possible combinations of the equivariant dimensions. $\mathbf{W}_{\mathcal{D}} \in \mathbb{R}^{F \times F_{\mathcal{O}}}$ and $\mathbf{b} \in \mathbb{R}^{F_{\mathcal{O}} \times 1}$ are learnable parameters, where $F_{\mathcal{O}}$ is the length of the output features. Furthermore, for the MDI, we adopt the module proposed in [38], which utilizes a multi-head attention mechanism to capture invariance across multiple dimensions, thereby significantly outperforming traditional methods such as mean or max pooling.

Building upon the basic TE layers, we designed the AMDE module, with its 2-D architecture illustrated in Fig. 3. The core of AMDE is a novel, lightweight decoupled attention mechanism compliant with MDE. This mechanism separately processes attention submodules along the feature and equivariant dimensions, and then serially integrates both components, following a design philosophy similar to that in [52]. The submodules are deliberately designed to ensure strict compliance with MDE, thereby enhancing representational capacity while preserving online computational efficiency. Furthermore, the residual mechanism is integrated to alleviate training

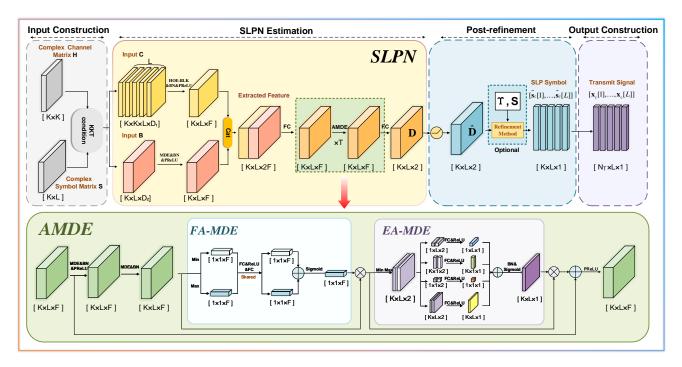


Fig. 3. The overall structure of the SLP framework.

difficulties [53]. The formulation of AMDE is given as follows

$$\mathbf{X}' = \text{PReLU}(\text{BN}(f_{\text{MDE}}(\mathbf{X}))),$$
 (36a)

$$\mathbf{X}'' = \mathrm{BN}(f_{\mathrm{MDE}}(\mathbf{X}')),\tag{36b}$$

$$\mathbf{F} = f_{\text{FA-MDE}}(\mathbf{X}'') \odot \mathbf{X}'', \tag{36c}$$

$$\mathbf{F}' = f_{\text{EA-MDE}}(\mathbf{F}) \odot \mathbf{F},\tag{36d}$$

$$\mathbf{O} = \text{PReLU}(\mathbf{F}' + \mathbf{X}), \tag{36e}$$

where o denotes element-wise multiplication during which the values are broadcast accordingly. BN refers to the batch normalization operation on the equivariant dimensions.

The $f_{\text{FA-MDE}}$ represents the feature attention MDE module, which performs pooling operations along equivariant dimensions and establishes inter-feature relationships, thereby generating a feature attention map that emphasizes the most informative components within the feature dimension of the input tensor. It can be expressed as

$$f_{\text{FA-MDE}}(\mathbf{X}) = \text{Sigmoid}\left(f(\mathbf{X}_{\mathcal{D}}^{\text{max}}) + f(\mathbf{X}_{\mathcal{D}}^{\text{mean}})\right),$$
 (37)

$$f(\mathbf{X}) = FC(ReLU(FC(\mathbf{X}))), \tag{38}$$

where $FC(\cdot)$ denotes the application of a fully connected layer along the feature dimension of the input tensor, and the set \mathcal{D} contains the dimensions for which pooling operations are performed. ${f X}^{\rm mean}_{\cal D}$ and ${f X}^{\rm max}_{\cal D}$ denote the tensors resulting from taking the mean and the maximum of **X** over the dimensions in \mathcal{D} , respectively. The two $f(\cdot)$ share the same weights. For 2-D equivariance, \mathcal{D} can be $\{1\}$, $\{2\}$, or $\{1,2\}$, and including more dimensions in \mathcal{D} reduces computational cost, whereas fewer dimensions yield higher representation power.

The $f_{\text{EA-MDE}}$ is the equivariant-dimension attention MDE module that utilizes the relationship between equivariant dimensions to generate an attention map that conforms to MDE, which can be represented as

$$f_{\text{EA-MDE}}(\mathbf{X}) = \text{Sigmoid}\left(f_{\text{RMDE}}([\mathbf{X}_{\{n_{\text{F}}\}}^{\text{max}}, \mathbf{X}_{\{n_{\text{F}}\}}^{\text{mean}}]_{n_{\text{F}}})\right), \quad (39)$$

$$f_{\text{EA-MDE}}(\mathbf{X}) = \text{Sigmoid}\left(f_{\text{RMDE}}([\mathbf{X}_{\{n_{\text{F}}\}}^{\text{max}}, \mathbf{X}_{\{n_{\text{F}}\}}^{\text{mean}}]_{n_{\text{F}}})\right), \quad (39)$$

$$f_{\text{RMDE}}(\mathbf{X}) = \sum_{\mathcal{D} \in \mathbb{S}_{D}} \text{ReLU}\left(\text{FC}(\bar{\mathbf{X}}_{\mathcal{D}})\right), \quad (40)$$

where $n_{\rm F}$ is the feature dimension. Compared to the basic MDE layer, the RMDE layer applies a ReLU activation to the dimension-averaged outputs before the summation, aiming to enhance the network's nonlinear representation power.

The proposed AMDE module is constructed from basic TE layers, which are characterized by low computational complexity [38]. Specifically, the complexity of the module is primarily dominated by its MDE layer, leading to an overall complexity of $\mathcal{O}(2^N \bar{M} F F_{\mathcal{O}})$ [38], where $\bar{M} = \prod_{n=1}^N M_n$ denotes the product of the lengths of N equivariant dimensions. The complexity thus scales linearly with \bar{M} , which confirms the computationally efficient nature of the proposed module. Furthermore, owing to its specific parameter-sharing pattern, AMDE module has a parameter count of $\mathcal{O}(2^N FF_{\mathcal{O}})$, which is independent of the lengths of the equivariant dimensions. Similarly, the basic TE layers maintain a parameter count independent of the input sizes [38]. Consequently, these modules and the networks built upon them exhibit inherent adaptability to varying input sizes [38], [46], [51].

D. TE-Based Low-Complexity SLP Framework

Building upon the basic TE layers and AMDE module, we propose a network for SLP named SLPN, which approximates the mapping $G(\mathbf{B}_c, \mathbf{C}_c) = \mathbf{D}^*$ (31). This network satisfies all the required TE while maintaining low online computational complexity, and its parameter-sharing structure offers inherent generalization for input sizes. Based on this, we further

Algorithm 1 SLP-DL Algorithm for CIZF Precoding

- 1: **Input:** $\mathbf{H}, \mathbf{S} = [\mathbf{s}_{c}[1], \dots, \mathbf{s}_{c}[L]]$
- 2: Get $\Upsilon = (\mathbf{H}\mathbf{H}^H)^{-1}$.
- 3: Get $\Lambda_{\mu}[l]$ and $\Lambda_{\nu}[l]$ according to the definition of CIR.
- 4: Determine \mathbf{B}_{c} and \mathbf{C}_{c} according to (26)-(29).
- 5: Construct **B** and **C** using (41),(42).
- 6: $\mathbf{D} = \text{SLPN}(\mathbf{B}, \mathbf{C})$.
- 7: Determine $\hat{\boldsymbol{\delta}}_{\mu}[l], \hat{\boldsymbol{\delta}}_{\nu}[l]$ using (48),(49).
- 8: Get $\rho[l]$ using (51).
- 9: Get $\tilde{\mathbf{s}}_{c}[l]$ using (52).
- 10: Get $\mathbf{x}_{c}[l]$ and $\gamma[l]$ using (10).
- 11: Get $\bar{\gamma}$, and $\bar{\mathbf{x}}_{c}[l]$ using (18).
- 12: Output: $\bar{\mathbf{x}}_{c}[l], \bar{\gamma}, \forall l = 1, \dots, L.$

develop an SLP framework, whose overall architecture is illustrated in Fig. 3. Specifically, the framework consists of the following components

1) Input Construction: Given the channel matrix $\mathbf{H} \in \mathbb{C}^{N_{\mathrm{T}} \times K}$ and the transmit symbol $\mathbf{S} \in \mathbb{C}^{K \times L}$, we first compute the matrix $\mathbf{\Upsilon}$ and apply Frobenius normalization to it. The KKT information \mathbf{B}_c and \mathbf{C}_c are then constructed according to (26)–(29). Their real and imaginary components are then separated and concatenated along the last dimension to form the input tensors \mathbf{C} and \mathbf{B} , defined as follows

$$\mathbf{C} = [\Re(\mathbf{C}_{c}), \Im(\mathbf{C}_{c})]_{4} \in \mathbb{R}^{K \times K \times L \times D_{1}}, \tag{41}$$

$$\mathbf{B} = [\Re(\mathbf{B}_{c}), \Im(\mathbf{B}_{c})]_{3} \in \mathbb{R}^{K \times L \times D_{2}}, \tag{42}$$

where $D_1 = 8$ and $D_2 = 4$.

2) SLPN Estimation: As shown in Fig. 3, the SLPN first extracts feature tensors from $\bf C$ and $\bf B$ using the 2-1 order HOE and 2-D MDE layers, respectively. Here, the HOE-BLK sequentially performs $f_{\rm HOE}$, BN, SiLU, and FC layers. These feature tensors are subsequently combined and processed by an FC layer, reducing the feature dimension from 2F to F, and it is further refined through a stack of T 2-D AMDE modules, with $\cal D$ of FA-MDE being $\{1,2\}$. Finally, a concluding FC layer is applied to reduce the feature dimension from F to 2.

$$\mathbf{C}' = \text{SiLU}(\text{BN}(f_{\text{HOE}}(\mathbf{C}))) \in \mathbb{R}^{K \times L \times F},$$
 (43)

$$\mathbf{C}'' = \text{PReLU}(\text{BN}(\text{FC}(\mathbf{C}'))) \in \mathbb{R}^{K \times L \times F},$$
 (44)

$$\mathbf{B}' = \text{PReLU}(\text{BN}(f_{\text{MDE}}(\mathbf{B}))) \in \mathbb{R}^{K \times L \times F},$$
 (45)

$$\mathbf{F} = FC([\mathbf{C}'', \mathbf{B}']_3) \in \mathbb{R}^{K \times L \times F}, \tag{46}$$

$$\mathbf{D} = FC(f_{\text{AMDF}}^{\times T}(\mathbf{F})) \in \mathbb{R}^{K \times L \times 2}, \tag{47}$$

where $f_{\text{AMDE}}^{\times T}(\cdot)$ denotes the stacking of T consecutive AMDE modules

3) Post-Net Refinement Method: To enforce non-negativity, we apply a ReLU activation to **D**, yielding

$$\hat{\mathbf{D}} = \text{ReLU}(\mathbf{D}). \tag{48}$$

Consequently, $\hat{\delta}_{\mu}[l]$ and $\hat{\delta}_{\nu}[l]$ are obtained by decomposing $\hat{\mathbf{D}}$, expressed as

$$\hat{\delta}_{\mu}[l] = \hat{\mathbf{D}}_{[:,l,1]}, \ \hat{\delta}_{\nu}[l] = \hat{\mathbf{D}}_{[:,l,2]}.$$
 (49)

An optimization factor $\rho[l]$ is introduced to scale $\hat{\delta}_{\mu}[l]$ and $\hat{\delta}_{\nu}[l]$, with the goal of further enhancing performance. The corresponding optimization problem is formulated as follows

$$\min_{\rho[l] \ge 0} \sum_{l=1}^{L} (\mathbf{s}_{c}[l] + \rho[l]\mathbf{p}[l])^{H} \Upsilon(\mathbf{s}_{c}[l] + \rho[l]\mathbf{p}[l])$$
s.t.
$$\mathbf{p}[l] = \mathbf{\Lambda}_{\mu}[l]\hat{\boldsymbol{\delta}}_{\mu}[l] + \mathbf{\Lambda}_{\nu}[l]\hat{\boldsymbol{\delta}}_{\nu}[l].$$
(50)

The optimal solution is given by

$$\rho[l] = \begin{cases} \max\left\{0, -\frac{\mathbf{s}_{c}^{H}[l]\Upsilon\mathbf{p}[l] + \mathbf{p}^{H}[l]\Upsilon\mathbf{s}_{c}[l]}{2\mathbf{p}^{H}[l]\Upsilon\mathbf{p}[l]}\right\} & \text{if } \mathbf{p}[l] \neq \mathbf{0} \\ 0 & \text{if } \mathbf{p}[l] = \mathbf{0}. \end{cases}$$
(51)

Finally, the transmit symbols are given by

$$\tilde{\mathbf{s}}_{c}[l] = \mathbf{s}_{c}[l] + \rho[l](\mathbf{\Lambda}_{\mu}[l]\hat{\boldsymbol{\delta}}_{\mu}[l] + \mathbf{\Lambda}_{\nu}[l]\hat{\boldsymbol{\delta}}_{\nu}[l]). \tag{52}$$

The proposed post-net refinement method is a heuristic approach that can be selectively integrated to enhance performance, depending on the specific optimization problem and training methodology.

4) Output Construction: For CIZF, the solutions $\bar{\mathbf{x}}_{c}[l]$ and $\bar{\gamma}$ are obtained from (10) and (18), whereas for CIMMSE they are determined by (15), (16), and (18).

By integrating the above framework, we propose the SLP-DL algorithm, with its CIZF version detailed in Algorithm 1. For CIMMSE, the algorithm requires two modifications: (i) In Step 2, Υ is modified to $\left(\mathbf{H}\mathbf{H}^H + \frac{\sigma^2 K}{P_{\mathrm{T}}}\mathbf{I}_K\right)^{-1}$; (ii) In Step 10, $\mathbf{x}_{\mathrm{c}}[l]$ and $\gamma[l]$ are obtained using equations (15) and (16).

We employ supervised learning, utilizing the following mean squared error loss for training

$$Loss = \frac{1}{N_{\rm sp}} \sum_{s=1}^{N_{\rm sp}} MSE(\mathbf{D}[s], \mathbf{D}^{\star}[s]), \qquad (53)$$

where $\mathbf{D}[s]$ and $\mathbf{D}^{\star}[s]$ denote the predicted and target tensors for the s-th sample in the dataset, respectively, and $N_{\rm sp}$ is the total number of samples. The MSE is computed as the average of squared differences over all tensor elements.

IV. LOW-COMPLEXITY ROBUST SYMBOL-LEVEL PRECODING FRAMEWORK AGAINST CHANNEL AGING

The CSI may contain errors due to channel estimation errors [54] and channel aging [27], necessitating the development of robust precoding strategies to mitigate the associated performance degradation. Conventional robust SLP approaches, which rely on optimization-based methods, often suffer from excessive computational complexity. To overcome this challenge, we first derive a closed-form robust SLP solution under the MMSE criterion, then analyze the inherent TE of the problem, and subsequently propose a tensor-equivariant and low-complexity DL framework for the efficient design of robust SLP.

A. Robust SLP Transmission

We utilize the *a posteriori* channel model to characterize the propagation channels [32]. Let $\mathbf{h}_k[l] \in \mathbb{C}^{N_{\mathrm{T}} \times 1}$ denote the channel vector between the BS and the k-th user during the l-th downlink symbol interval, and let $\mathbf{h}_k[0]$ denote the

channel estimated during the uplink pilot transmission phase, then $\mathbf{h}_k[l]$ can be represented as follows [33]

$$\mathbf{h}_{k}[l] = \alpha_{k}[l]\mathbf{h}_{k}[0] + \sqrt{1 - \alpha_{k}^{2}[l]}\mathbf{V}_{T}^{*}(\mathbf{m}_{k} \odot \mathbf{w}_{k}[l]).$$
 (54)

In this formulation, the temporal evolution of the channel is modeled as a first-order Markov process. The term $\alpha_k[l]$ represents the time correlation coefficient of the l-th symbol interval for user k, which are characterized using Jakes' autocorrelation model; The matrix $\mathbf{V}_{\mathrm{T}} \in \mathbb{C}^{N_{\mathrm{T}} \times N_F N_{\mathrm{T}}}$ comprises a partial discrete Fourier transform (DFT) matrix, with $N_F \in \mathbb{N}^+$ being the fine factor that enhances model accuracy; The $\mathbf{m}_k \in \mathbb{R}^{N_F N_{\mathrm{T}} \times 1}$ is a sparse and nonnegative vector, whose elements remain constant for a long time; The elements of $\mathbf{w}_k[l] \in \mathbb{C}^{N_F N_{\mathrm{T}} \times 1}$ are independent and identically distributed (i.i.d.), following $\mathcal{CN}(0,1)$.

By defining $\bar{\mathbf{h}}_k[l] = \alpha_k[l]\mathbf{h}_k[0]$, substituting (54) into (1) and subsequently simplifying the resulting expression, we obtain [34]

$$y_k[l] = \bar{\mathbf{h}}_k^T[l]\mathbf{x}_c[l] + \bar{n}_k[l], \tag{55}$$

where $\bar{n}_k[l] = (\beta_k[l]\mathbf{V}_k\mathbf{x}_c[l])^T\mathbf{w}_k[l] + n_k[l]$, with $\beta_k[l] = \sqrt{1 - \alpha_k^2[l]}$; $\mathbf{V}_k = [\mathbf{m}_k \odot \mathbf{v}_1 \cdots \mathbf{m}_k \odot \mathbf{v}_{N_T}]$, with \mathbf{v}_i being the i-th row of \mathbf{V}_T . It can be verified that $\bar{n}_k[l] \sim \mathcal{CN}(0, \beta_k^2[l] \|\mathbf{V}_k\mathbf{x}_c[l]\|_2^2 + \sigma^2)$ [34]. Thus, the signal for demodulation can be expressed as

$$\tilde{y}_k[l] = y_k[l]/\gamma_k[l] = \bar{\mathbf{h}}_k^T[l]\mathbf{x}_c[l]/\gamma_k[l] + \bar{n}_k[l]/\gamma_k[l], \quad (56)$$

where $\gamma_k[l]$ is a user-specific rescaling factor of the SLP, which enhances robustness against imperfect CSI but is only applicable to PSK modulation.

The robust SLP MMSE problem is formulated as

$$\min_{\substack{\mathbf{x}_{\mathrm{c}}[l],\,\tilde{\mathbf{s}}_{\mathrm{c}}[l],\\\mathbf{\Gamma}[l],\forall l\in\mathcal{L}}} \sum_{l=1}^{L} \mathbb{E}_{\mathbf{n}} \left\{ \left\| \mathbf{\Gamma}^{-1}[l] \left(\bar{\mathbf{H}}[l]\mathbf{x}_{\mathrm{c}}[l] + \bar{\mathbf{n}}[l] \right) - \tilde{\mathbf{s}}_{\mathrm{c}}[l] \right\|_{2}^{2} \right\}$$

s.t.
$$\|\mathbf{x}_{c}[l]\|_{2}^{2} \leq P_{T}, \ \forall l \in \mathcal{L},$$

 $\tilde{s}_{k}[l] \in \mathcal{D}_{k}[l], \ \forall k \in \mathcal{K}, \forall l \in \mathcal{L},$
 $\gamma_{k}[l] > 0, \ \forall k \in \mathcal{K}, \forall l \in \mathcal{L},$

$$(57)$$

where $\Gamma[l] = \operatorname{diag}\{\gamma_1[l], ..., \gamma_K[l]\}, \bar{\mathbf{H}}[l] = [\bar{\mathbf{h}}_1[l], ..., \bar{\mathbf{h}}_K[l]]^T$, and $\bar{\mathbf{n}}[l] = [\bar{n}_1[l], ..., \bar{n}_K[l]]^T$. This problem can be decoupled into per-symbol optimization problems, which are then solved via an alternating optimization approach [34], yielding a closed-form solution

$$\mathbf{x}_{c}^{\star}[l] = \eta^{\star}[l]\mathbf{P}[l]\tilde{\mathbf{s}}_{c}[l], \quad \eta^{\star}[l] = \sqrt{P_{T}/\|\mathbf{P}[l]\tilde{\mathbf{s}}_{c}[l]\|_{2}^{2}}, \quad (58)$$

with

$$\mathbf{P}[l] = \left(\bar{\mathbf{H}}^{H}[l]\mathbf{\Psi}^{2}[l]\bar{\mathbf{H}}[l] + \mathbf{\Phi}[l] + \frac{\sigma^{2}\sum_{k=1}^{K}\psi_{k}^{2}[l]}{P_{T}}\mathbf{I}_{N_{T}}\right)^{-1}$$

$$\bar{\mathbf{H}}^{H}[l]\mathbf{\Psi}[l], \tag{59}$$

$$\mathbf{\Phi}[l] = \sum_{k=1}^{K} \psi_k^2[l] \beta_k^2[l] \mathbf{E}_k, \tag{60}$$

$$\tilde{\mathbf{s}}_{c}[l] = \mathbf{s}_{c}[l] + \mathbf{\Lambda}_{\mu}[l]\boldsymbol{\delta}_{\mu}[l] + \mathbf{\Lambda}_{\nu}[l]\boldsymbol{\delta}_{\nu}[l], \tag{61}$$

where $\Psi[l] = \text{diag}\{\psi_1[l], \dots, \psi_K[l]\}, \mathbf{E}_k = \mathbf{V}_k^T \mathbf{V}_k$, and

 $\frac{\psi_k[l]}{\eta[l]}=\frac{1}{\gamma_k[l]}.$ When $\Psi[l]$ is obtained, problem (57) can be reformulated as [34]

$$\min_{\boldsymbol{\delta}_{\mu}[l] \succeq \mathbf{0}, \boldsymbol{\delta}_{\nu}[l] \succeq \mathbf{0}} \sum_{l=1}^{L} \tilde{\mathbf{s}}_{c}^{H}[l] \boldsymbol{\Upsilon}[l] \tilde{\mathbf{s}}_{c}[l]
\text{s.t. } \tilde{\mathbf{s}}_{c}[l] = \mathbf{s}_{c}[l] + \boldsymbol{\Lambda}_{\mu}[l] \boldsymbol{\delta}_{\mu}[l] + \boldsymbol{\Lambda}_{\nu}[l] \boldsymbol{\delta}_{\nu}[l], \forall l \in \mathcal{L},
\boldsymbol{\Upsilon}[l] = (\mathbf{I}_{K} - \boldsymbol{\Psi}[l] \bar{\mathbf{H}}[l] \mathbf{P}[l]), \forall l \in \mathcal{L}.$$
(62)

Given its strong resemblance to problems (8) and (13), this problem can also be solved using the aforementioned SLPN.

When the key variables $\psi_k[l]$, $\delta_{\nu}[l]$, and $\delta_{\mu}[l]$ are determined, the optimal transmit vector $\mathbf{x}_{c}^{\star}[l]$ can be obtained in closed form via (58)-(61). Consequently, the subsequent challenge is to design a low-complexity neural network for the estimation of these variables.

B. Tensor Equivariance in Robust SLP Design

As formulated in (57), the $\psi_k[l]$ is dependent on the inputs $\bar{\mathbf{H}}[l]$, $\mathbf{s}_{\mathrm{c}}[l]$, $\alpha_k[l]$, σ^2 , and \mathbf{m}_k . Therefore, we can define a mapping to represent this relationship. For notational compactness, we first define

$$\mathbf{A} \in \mathbb{R}^{K \times L}, \text{ where } [\mathbf{A}]_{k,l} = \alpha_k[l],$$

$$\mathbf{\Psi} = [\boldsymbol{\psi}[1], \dots, \boldsymbol{\psi}[L]] \in \mathbb{R}^{K \times L},$$

$$\boldsymbol{\psi}[l] = [\psi_1[l], \dots, \psi_K[l]]^T \in \mathbb{R}^{K \times 1},$$

$$\mathbf{M} = [\mathbf{m}_1, \dots, \mathbf{m}_K]^T \in \mathbb{R}^{K \times N_T},$$

$$\mathbf{U} = \mathbf{M}\mathbf{V}_T^H \in \mathbb{C}^{K \times N_T},$$

$$\mathbf{S} = [\mathbf{s}_c[1], \dots, \mathbf{s}_c[L]] \in \mathbb{C}^{K \times L}.$$

$$(63)$$

Then, we stack these components to obtain

$$\mathbf{Q} = [\mathbf{H}, \mathbf{U}]_3 \in \mathbb{C}^{K \times N_{\mathrm{T}} \times 2},$$

$$\mathbf{G} = [\Re{\{\mathbf{S}\}}, \Im{\{\mathbf{S}\}}, \mathbf{A}]_3 \in \mathbb{R}^{K \times L \times 3}.$$
 (64)

Finally, we define the mapping as

$$G_{\rm P1}(\mathbf{Q}, \mathbf{G}, \sigma^2) = \mathbf{\Psi}^*. \tag{65}$$

This mapping satisfies the following property

Proposition 2. When (57) has a unique optimal solution for Ψ , i.e., $G_{\text{P1}}(\mathbf{Q}, \mathbf{G}, \sigma^2) = \Psi^*$, for any $\pi_K \in \mathbb{S}_K$, $\pi_{N_{\text{T}}} \in \mathbb{S}_{N_{\text{T}}}$, and $\pi_L \in \mathbb{S}_L$, the following equations hold

$$G_{\text{P1}}(\pi_K \circ_1 \mathbf{Q}, \pi_K \circ_1 \mathbf{G}, \sigma^2) = \pi_K \circ_1 \mathbf{\Psi}^*,$$
 (66a)

$$G_{\rm P1}(\pi_{N_{\rm TD}} \circ_2 \mathbf{Q}, \mathbf{G}, \sigma^2) = \mathbf{\Psi}^{\star},$$
 (66b)

$$G_{\text{P1}}(\mathbf{Q}, \pi_L \circ_2 \mathbf{G}, \sigma^2) = \pi_L \circ_2 \mathbf{\Psi}^*.$$
 (66c)

Proof. The proof follows a similar procedure to that in **Proposition 1**.

As for the mapping to $\delta_{\nu}[l]$ and $\delta_{\mu}[l]$, since (62) and (8) share nearly identical forms, we can first employ the same method as in III-A to derive its KKT conditions, and we define

$$\mathbf{B}_{c}[l] = \left[\mathbf{\Lambda}_{\mu}^{H}[l] \mathbf{\Upsilon}[l] \mathbf{s}_{c}[l], \mathbf{\Lambda}_{\nu}^{H}[l] \mathbf{\Upsilon}[l] \mathbf{s}_{c}[l] \right] \in \mathbb{C}^{K \times 2},$$

$$\mathbf{C}_{c}[l] = \left[\mathbf{\Lambda}_{\mu}^{H}[l] \mathbf{\Upsilon}[l] \mathbf{\Lambda}_{\mu}[l], \mathbf{\Lambda}_{\mu}^{H}[l] \mathbf{\Upsilon}[l] \mathbf{\Lambda}_{\nu}[l],$$

$$\mathbf{\Lambda}_{\nu}^{H}[l] \mathbf{\Upsilon}[l] \mathbf{\Lambda}_{\mu}[l], \mathbf{\Lambda}_{\nu}^{H}[l] \mathbf{\Upsilon}[l] \mathbf{\Lambda}_{\nu}[l] \right]_{2} \in \mathbb{C}^{K \times K \times 4}.$$
(68)

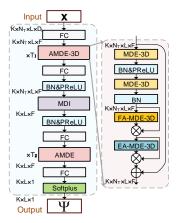


Fig. 4. The architecture of RSLPN-A.

Then \mathbf{B}_{c} and \mathbf{C}_{c} are definde according to (28) and (29), and \mathbf{D} is defined as (30). Finally, we define the mapping as follows

$$G_{\rm P2}(\mathbf{B}_{\rm c}, \mathbf{C}_{\rm c}) = \mathbf{D}^{\star},\tag{69}$$

which is exactly the same as (31) and also satisfies the TE described in (34a) and (34b).

C. TE-Based Low-Complexity Robust SLP Framework

Leveraging the TE analyzed in the previous subsection, we then propose a DL-based low-complexity framework for robust SLP, with the following two networks to determine the auxiliary variables Ψ and the perturbation factors \mathbf{D} :

1) Network for Auxiliary Variables: As shown in Fig. 4, we design the network RSLPN-A to approximate the mapping $G_{\rm P1}$, ensuring that the network is structured to fully exploit the TE of $G_{\rm P1}$. The input is constructed as follows

$$\mathbf{X} = [\Re(\tilde{\mathbf{Q}}), \Im(\tilde{\mathbf{Q}}), \tilde{\mathbf{G}}, \sigma^2 \mathbf{1}]_4 \in \mathbb{R}^{K \times N_{\mathrm{T}} \times L \times D_I},$$

$$\tilde{\mathbf{G}} = {\mathbf{G}}_{2,N_{\mathrm{T}}} \in \mathbb{R}^{K \times N_{\mathrm{T}} \times L \times 3},$$

$$\tilde{\mathbf{Q}} = {\mathbf{Q}}_{3,L} \in \mathbb{C}^{K \times N_{\mathrm{T}} \times L \times 2},$$
(70)

where $D_I = 8$, and $\{\cdot\}_{p,q}$ denotes the operation that replicates the input tensor q times along a newly inserted p-th dimension. To process the 4-D input tensor with TE across its first three dimensions, the network employs a specialized 3-D AMDE module, denoted as AMDE-3D. Within this module, the MDE and RMDE layers are designed to enforce equivariance along the first three dimensions, while the FA-MDE component is configured with $\mathcal{D} = \{1,2\}$. The transmit antenna dimension is processed by the MDI layer in [38] to capture its invariance. In addition, A softplus layer is used as the output layer to ensure that Ψ satisfies the positivity constraint.

- 2) Network for Perturbation Factors: Since the mapping $G_{\rm P2}$ has the same form as the mapping G in (31) and satisfies the same TE, it can be approximated using the SLPN proposed in Section III-D. For distinction, we refer to this network as RSLPN-B in this section.
- 3) Overall Structure of the RSLP Framework: The proposed framework operates in two sequential stages: first, RSLPN-A provides an estimate of the matrix Ψ ; second, this estimate is used to construct the input for RSLPN-B, which

subsequently yields the final output **D**. In detail, **Q** and **G** are computed from (63)-(64) and employed to derive **X** using (70). This **X** serves as the input to the RSLPN-A network, which generates the estimate Ψ . Following this, the matrix $\mathbf{P}[l]$ is obtained using (59)-(60). Then, $\Upsilon[l]$ is computed via (62), which is then normalized by its Frobenius norm. The terms \mathbf{B}_c and \mathbf{C}_c are then obtained from (67), (68), (28), (29), and **B** and **C** are defined in (41)-(42). Then **D** is determined by RSLPN-B, and the variables $\hat{\delta}_{\mu}[l]$ and $\hat{\delta}_{\nu}[l]$ are determined from (48)-(49), while $\tilde{\mathbf{s}}_c[l]$ is calculated via (61). Finally, the transmit vector $\mathbf{x}_c[l]$ is given by (58). The MSE loss function, as defined in (53), is utilized for training both subnetworks.

V. NUMERICAL RESULTS

In this section, we evaluate the performance of the proposed methods via Monte Carlo simulations. To generate channels compliant with 3GPP standards, we utilize the widely adopted QuaDRiGa channel simulator [55]. We consider a MIMO system where the BS is equipped with an $N_{\mathrm{T}} \times 1$ uniform linear array (ULA) of type '3gpp-3d', serving K singleantenna users, each equipped with an 'omni' antenna. Users are randomly located within a 500-meter radius in a 120degree sector facing the BS. The carrier frequency is 3.5 GHz, and the propagation environment corresponds to the '3GPP_38.901_UMa_NLOS' scenario [56]. Shadow fading and path loss are not considered. The channel is normalized such that $\mathbb{E}\{\operatorname{tr}(\mathbf{H}\mathbf{H}^H)\}=KN_{\mathrm{T}}$. The SNR is defined as $SNR = P_T/\sigma^2$. The imperfect channels were generated according to the widely adopted a posteriori model in (54), where $\{\mathbf{m}_k\}_{k=1}^K$ was estimated from the channel realizations generated by the QuaDRiGa channel simulator [32]. For simplicity, we assume $\alpha_{k,l} = \alpha, \forall k \in \mathcal{K}, \forall l \in \mathcal{L}$, i.e., the same α is shared across all users and symbols within a transmission block. Under the predefined channel scenario configuration, each channel realization in the dataset is independently generated by QuaDRiGa, with random user distributions, scatterer distributions, and channel statistics to ensure diversity across channel realizations.

This section compares the following schemes:

- 'ZF' and 'MMSE' [5], [57]: The ZF and MMSE precoding schemes with symbol-level power constraints.
- 'CIZF' and 'CIMMSE' [16], [26]: The conventional SLP solutions to problems (7) and (12), respectively.
- 'CIZF-CF' [36]: The approximate closed-form solution for CIZF problem.
- 'CIZF-DL' and 'CIMMSE-DL': The TENN-based SLP framework proposed in Section III. The network is configured with $T=4,\,F=4$.
- 'RCIMMSE': The solutions to the robust MMSE SLP problems (57) proposed in [34].
- 'RCIMMSE-DL': The TENN-based SLP neural networks proposed in Section IV-C. The RSLPN-A is configured with $T_1=2, T_2=2, F=16$, and the RSLPN-B is configured with T=2, F=16.

A. Training Details

For the SLPN in Section III, the dataset consisted of 100,000 independent channel realizations, with 90,000 used for training

TABLE I
PER SYMBOL COMPUTATIONAL COMPLEXITY

Methods	Complexity Order		
ZF	$\mathcal{O}(KN_{\mathrm{T}})$		
MMSE	$\mathcal{O}(KN_{\mathrm{T}})$		
CIZF	$\mathcal{O}(KN_{\mathrm{T}} + KN_{\mathrm{T}}N_L + N_{\mathrm{T}}N_L^3)$		
CIZF-CF	$\mathcal{O}ig(KN_{\mathrm{T}}+M_{\mathrm{CF}}^{3}ig)$		
CIMMSE	$\mathcal{O}(KN_{\mathrm{T}} + K^2N_L + KN_L^3)$		
CIZF-DL	$\mathcal{O}(KN_{\mathrm{T}}+KF^2)$		
CIMMSE-DL	$\mathcal{O}ig(KN_{\mathrm{T}}+KF^2ig)$		
RCIMMSE	$\mathcal{O}ig(N_{\mathrm{T}}^3 + K N_L^3) \hat{I}ig)$		
RCIMMSE-DL	$\mathcal{O}(N_{\mathrm{T}}^3 + K N_{\mathrm{T}} F^2)$		

and 10,000 for testing, each associated with a transmit symbol sequence of length L=100. The corresponding perturbation factors under the CIZF and CIMMSE criteria were then computed as labels. To improve robustness across SNRs, the CIMMSE labels were generated over a set of SNR values $\{0, 5, 10, 15, 20, 25, 30\}$ dB. Separate networks were trained for CIZF and CIMMSE for 800 and 400 epochs, respectively. For the RSLPN in Section IV-C, a dataset of 60,000 distinct channel realizations (55,000 for training and 5,000 for testing) was generated, each associated with the corresponding M and 50 different transmit symbols. The corresponding auxiliary variables and perturbation factors were used as labels, generated over SNR values {10, 15, 20, 25, 30, 35, 40} dB. The training was performed sequentially: RSLPN-A was first trained to estimate Ψ , which is then used to train RSLPN-B. The two subnets of RSLPN were each trained for 300 epochs.

All networks were trained using the Adam optimizer with a batch size of 400. A two-stage learning rate schedule was adopted: 5×10^{-3} for the first half of the training epochs and 5×10^{-4} for the second half. Although the networks were trained on datasets with a fixed number of users K and symbol block length L, their design inherently provides generalization capability to other configurations. Moreover, the test set consists of channel realizations that were not encountered during training, characterized by different channel environments, user locations, and distinct transmit symbols.

B. Complexity Comparison

Table I presents the dominant order complexity for various precoding algorithms. Let K, $N_{\rm T}$, and N_L denote the number of UEs, BS antennas, and main loop iterations for the active set algorithm, respectively. $M_{\rm CF}$ is the estimated size of the inactive constraint set for CIZF-CF, F represents the number of hidden layer neurons, and I is the number of RCIMMSE iterations (which stops when the MSE difference between two consecutive iterations falls below 10^{-4} or after 400 iterations). We assume $L \gg N_{\rm T} \geq K$, and $2K > N_L, M_{\rm CF}$. In the non-robust scenario, K > F, while in the robust scenario, $N_{\rm T} \approx F$. The proposed CIZF-DL and CIMMSE-DL frameworks employ the low-complexity SLPN, thereby avoiding the iterations required for solving NNLS problems. As a result, they overcome the high-order $N_{\rm T}N_L^3$ complexity and achieve online computational efficiency that scales linearly with both K and $N_{\rm T}$, consistent with LP. The complexity of

TABLE II
AVERAGE EXECUTION TIME PER SYMBOL (S)

Methods	CPU	GPU
RCIMMSE	1.7×10^{-1}	3.5×10^{-1}
RCIMMSE-DL	3.1×10^{-4}	2.3×10^{-4}

RCIMMSE is primarily determined by the number of iterations I, whereas the proposed RCIMMSE-DL eliminates this and significantly reduces computational load. Furthermore, the proposed method generalizes effectively across both K and L, allowing a single trained network to adapt to different K and L configurations, thereby avoiding the overhead of training and storing multiple networks.

C. Performance of SLP with Perfect CSI

Fig. 5 compares the SER performance of various schemes. As can be observed, both CIZF-DL and CIMMSE-DL consistently outperform conventional ZF and MMSE, while also achieving gains over CIZF-CF. Specifically, CIMMSE-DL achieves a lower SER than CIZF in the low-SNR regime, while CIZF-DL surpasses MMSE at high SNRs. For instance, with $N_{\rm T}=K=12$ and 4-QAM, CIMMSE-DL requires approximately 6 dB less SNR than MMSE to achieve an SER of 10^{-2} . Under the same conditions, CIZF-DL delivers about 3 dB of SNR gain over MMSE and more than 6 dB over CIZF-CF. These results affirm that the proposed SLP framework effectively retains the performance benefits of SLP.

Figs. 6 and 7 depict the total transmit power as a function of the SINR threshold, defined as γ^2/σ^2 . Under a given SINR threshold, a reduction in transmit power corresponds to improved energy efficiency. As evidenced, the proposed CIZF-DL scheme achieves substantial power gains of approximately 11.3 dB over ZF precoding and 5.3 dB over CIZF-CF in the $N_{\rm T}=12$, K=12 scenario. Furthermore, it delivers transmit power comparable to the optimal CIZF solution in the $N_{\rm T}=14$, K=12 scenario. These results clearly demonstrate the superior power efficiency of our CIZF-DL approach, which is a critical advantage for practical implementations.

The SER of the proposed network generally decreases as the size of the hidden layer increases, converging towards the optimal achievable performance. Fig. 8 depicts the SER of CIMMSE-DL as a function of its hidden layer size F. With increasing F, the SER steadily decreases, approaching the performance of CIMMSE and eventually reaching a closely comparable level. This trend demonstrates the capability of our proposed architecture to effectively approximate sophisticated SLP solutions. Fig. 9 and Fig. 10 demonstrate the generalization capability of the proposed approach. We train our network in scenario $N_{\rm T}=14,\,K=12,\,L=100,$ and directly apply it to various scenarios. As can be observed, the proposed method exhibits consistently outstanding performance, highlighting its robust practical utility.

Fig. 11 shows the average execution times implemented in PyTorch on Intel(R) Xeon(R) Platinum 8336C CPU and NVIDIA GeForce RTX 4090 GPU. The proposed CIZF-DL

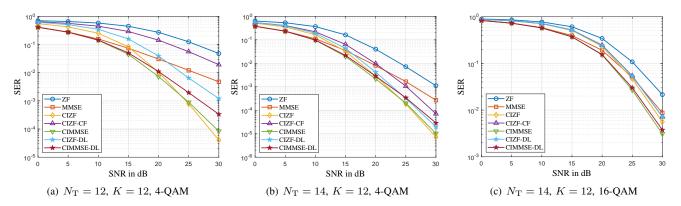


Fig. 5. The SER versus SNR under different antenna configurations and modulation schemes, L=100.

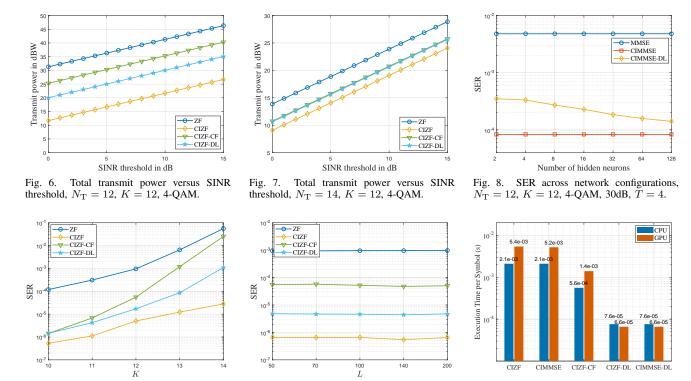


Fig. 9. Generalization performance with respect to K for the network trained under scenario $N_{\rm T}{=}14,~K{=}12,~{\rm and}~L{=}100.$

Fig. 10. Generalization performance with respect to L for the network trained under scenario $N_{\rm T}{=}14,~K{=}12,$ and $L{=}100.$

Fig. 11. Average execution time per symbol, $N_{\rm T}=14,~K=12,~L=100,~{\rm SNR}=30{\rm dB},$ 4-QAM.

and CIMMSE-DL methods exhibit significantly lower execution times on both platforms compared to conventional SLP precodings. Specifically, CIZF-DL consumes merely 3.62% of the CPU runtime and 1.22% of the GPU runtime required by CIZF. Similarly, the CIMMSE-DL demands only approximately 3.62% of the CPU and 1.27% of the GPU execution time relative to CIMMSE. This substantial speedup is attributed to the ability of the proposed methods to process transmit symbols in blocks, which greatly enhances operational efficiency.

D. Performance of Robust SLP

Fig. 12 and 13 compare the MSE and SER performance, respectively, for QPSK downlink transmission with $\alpha=0.995$, $N_{\rm T}=14$, and K=12. Fig. 12 shows that the RCIMMSE-

DL achieves near-optimal MSE performance, consistently outperforming the other compared methods across the evaluated SNR range. Fig. 13 indicates that the SER performance of RCIMMSE-DL nearly matches that of RCIMMSE. Specifically, at an SNR of 40 dB, the SER for the CIMMSE precoding is approximately 4.7×10^{-3} , while RCIMMSE-DL achieves a significantly lower SER of 3.2×10^{-4} , very close to the RCIMMSE result of 2.4×10^{-4} . This substantial improvement becomes increasingly pronounced at higher SNRs.

Table II compares the average execution times of the RCIMMSE and RCIMMSE-DL methods. The simulations were configured with $N_{\rm T}=14,\,K=12,\,L=50,\,\alpha=0.995,\,{\rm SNR}=30{\rm dB},$ and QPSK modulation. A significant advantage in execution time for the proposed RCIMMSE-DL is clearly observable. Specifically, compared to the RCIMMSE method,

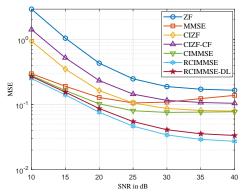


Fig. 12. MSE vs SNR, $N_{\rm T}=14,\,K=12,\,\alpha=0.995,\,{\rm QPSK}.$

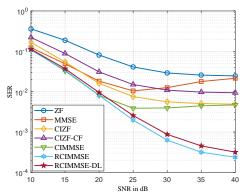


Fig. 13. SER vs SNR, $N_{\rm T} = 14$, K = 12, $\alpha = 0.995$, QPSK.

the execution times of RCIMMSE-DL on the CPU and GPU are merely 0.18% and 0.07%, respectively. This demonstrates that our proposed method drastically reduces online computation time while maintaining excellent performance.

VI. CONCLUSION

This paper proposed a framework based on TENN to address the high computational complexity of SLP problems. Leveraging the NNLS forms of the SLP problems and their closed-form solutions, we analyzed the optimality condition to establish a mapping from the available information to the perturbation factors, and then revealed the inherent TE within this mapping. Utilizing this TE, we proposed an attentionbased tensor-equivariant neural network, based on which a low-complexity SLP framework is developed. Subsequently, considering imperfect CSI scenarios, we extended the proposed framework to the robust MMSE problem. Simulation results demonstrate the effectiveness of the proposed frameworks in terms of transmit power, SER, execution time, and online computational complexity. Future work includes the joint design of SLP and receiver demodulation, aiming for further enhancements in system performance.

APPENDIX A PROOF OF PROPOSITION 1

Define $M_l(\mathbf{B}_c, \mathbf{C}_c, \mathbf{D}) = \nabla_{\boldsymbol{\delta}_{\mu}[l]} \mathcal{L}(\boldsymbol{\delta}_{\mu}[l], \boldsymbol{\delta}_{\nu}[l], \boldsymbol{\lambda}_{\mu}[l], \boldsymbol{\lambda}_{\nu}[l])$, we first prove

$$M_l(\pi_K \circ_1 \mathbf{B}_c, \pi_K \circ_{[1,2]} \mathbf{C}_c, \pi_K \circ_1 \mathbf{D})$$

$$= \pi_K \circ_1 M_l(\mathbf{B}_c, \mathbf{C}_c, \mathbf{D}). \tag{71}$$

We define the permutation matrix Π to represent the permutation π in the first dimension of a tensor. Each row and column of the matrix Π contains a single 1, with all other entries being 0, and $\Pi\Pi^T = \mathbf{I}$. Then, for $\pi_K \circ_1 \mathbf{D}$, $\pi_K \circ_{[1,2]} \mathbf{C}_c$, $\pi_K \circ_1 \mathbf{B}_c$, the variables during the l-th symbols can be expressed as $\mathbf{s}_c'[l] = \mathbf{\Pi}\mathbf{s}_c[l]$, $\boldsymbol{\delta}_\mu'[l] = \mathbf{\Pi}\boldsymbol{\delta}_\mu[l]$, $\boldsymbol{\delta}_\nu'[l] = \mathbf{\Pi}\boldsymbol{\delta}_\nu[l]$, $\boldsymbol{\Lambda}_\mu'[l] = \mathbf{\Pi}\boldsymbol{\Lambda}_\mu[l]\boldsymbol{\Pi}^T$, $\boldsymbol{\Lambda}_\nu'[l] = \mathbf{\Pi}\boldsymbol{\Lambda}_\nu[l]\boldsymbol{\Pi}^T$, $\boldsymbol{H}' = \mathbf{\Pi}\mathbf{H}$. On this basis, we have $\boldsymbol{\Upsilon}' = \mathbf{\Pi}\boldsymbol{\Upsilon}\boldsymbol{\Pi}^T$, $\tilde{\mathbf{s}}_c'[l] = \mathbf{\Pi}\tilde{\mathbf{s}}_c[l]$. For simplicity, we temporarily drop l, and thus we have

$$\begin{split} &M_{l}(\pi_{K} \circ_{1} \mathbf{B}_{c}, \pi_{K} \circ_{[1,2]} \mathbf{C}_{c}, \pi_{K} \circ_{1} \mathbf{D}) \\ &= 2 \mathbf{\Pi} \boldsymbol{\Lambda}_{\mu}^{H} \mathbf{\Pi}^{T} \mathbf{\Pi} \boldsymbol{\Upsilon} \mathbf{\Pi}^{T} \mathbf{\Pi} \mathbf{s}_{c} + 2 \mathbf{\Pi} \boldsymbol{\Lambda}_{\mu}^{H} \mathbf{\Pi}^{T} \mathbf{\Pi} \boldsymbol{\Upsilon} \mathbf{\Pi}^{T} \mathbf{\Pi} \boldsymbol{\Lambda}_{\mu} \mathbf{\Pi}^{T} \mathbf{\Pi} \boldsymbol{\delta}_{\mu} \\ &+ 2 \mathbf{\Pi} \boldsymbol{\Lambda}_{\mu}^{H} \mathbf{\Pi}^{T} \mathbf{\Pi} \boldsymbol{\Upsilon} \mathbf{\Pi}^{T} \mathbf{\Pi} \boldsymbol{\Lambda}_{\nu} \mathbf{\Pi}^{T} \mathbf{\Pi} \boldsymbol{\delta}_{\nu} \\ &= 2 \mathbf{\Pi} \boldsymbol{\Lambda}_{\mu}^{H} \boldsymbol{\Upsilon} \mathbf{s}_{c} + 2 \mathbf{\Pi} \boldsymbol{\Lambda}_{\mu}^{H} \boldsymbol{\Upsilon} \boldsymbol{\Lambda}_{\mu} \boldsymbol{\delta}_{\mu} + 2 \mathbf{\Pi} \boldsymbol{\Lambda}_{\mu}^{H} \boldsymbol{\Upsilon} \boldsymbol{\Lambda}_{\nu} \boldsymbol{\delta}_{\nu} \\ &= \mathbf{\Pi} M_{l}(\mathbf{B}_{c}, \mathbf{C}_{c}, \mathbf{D}). \end{split}$$
(72)

If $M_l(\mathbf{B}_c, \mathbf{C}_c, \mathbf{D}) = \mathbf{0}$, then $M_l(\pi_K \circ_1 \mathbf{B}_c, \pi_K \circ_{[1,2]} \mathbf{C}_c, \pi_K \circ_1 \mathbf{D}) = \mathbf{0}$ also holds. Therefore, Equation (23a) remains valid under arbitrary permutation along the K dimension. Following the same methodology, we can prove that Equation (23b) holds. Consequently, permutation operations along the K dimension satisfy the KKT conditions and likewise yield optimal solutions. This completes the proof of Equation (34a).

We further consider permutations along the L dimension. Since the optimization problem is solved independently for each $l \in \mathcal{L}$, permuting along the L dimension does not alter the solution at any individual l, but merely applies the same permutation to the order of the solutions, which clearly preserves TE. This completes the proof of Equation (34b).

REFERENCES

- [1] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin et al., "An overview of massive MIMO: Benefits and challenges," *IEEE J. Sel. Top. Sign. Proces.*, vol. 8, no. 5, pp. 742–758, Oct. 2014.
- [2] W. Wang, Y. Zhu, Y. Wang, R. Ding et al., "Toward mobile satellite internet: The fundamental limitation of wireless transmission and enabling technologies," Engineering, Early Access, 2025.
- [3] M. Joham, W. Utschick, and J. Nossek, "Linear transmit processing in MIMO communications systems," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 2700–2712, Aug. 2005.
- [4] M. Alodeh, D. Spano, A. Kalantari, C. G. Tsinos et al., "Symbol-level and multicast precoding for multiuser multiantenna downlink: A stateof-the-art, classification, and challenges," *IEEE Commun. Surv. Tutor.*, vol. 20, no. 3, pp. 1733–1757, May 2018.
- [5] A. Li, C. Masouros, B. Vucetic, Y. Li et al., "Interference exploitation precoding for multi-level modulations: Closed-form solutions," *IEEE Trans. Commun.*, vol. 69, no. 1, pp. 291–308, Jan. 2021.
- [6] C. Masouros and E. Alsusa, "Dynamic linear precoding for the exploitation of known interference in MIMO broadcast systems," *IEEE Trans. Wireless Commun.*, vol. 8, no. 3, pp. 1396–1404, Mar. 2009.
- [7] A. Li, D. Spano, J. Krivochiza, S. Domouchtsidis et al., "A tutorial on interference exploitation via symbol-level precoding: Overview, stateof-the-art and future directions," vol. 22, no. 2, pp. 796–839, 2nd Quart 2020.
- [8] C. Masouros, "Correlation rotation linear precoding for MIMO broadcast communications," *IEEE Trans. Signal Process.*, vol. 59, no. 1, pp. 252– 262, Jan. 2011.
- [9] C. Masouros and G. Zheng, "Exploiting known interference as green signal power for downlink beamforming optimization," *IEEE Trans.* Signal Process., vol. 63, no. 14, pp. 3628–3640, Jul. 2015.
- [10] M. Alodeh, S. Chatzinotas, and B. Ottersten, "Constructive interference through symbol level precoding for multi-level modulation," in *Proc. IEEE Glob. Commun. Conf. (GLOBECOM)*, Dec. 2015, pp. 1–6.
- [11] —, "Symbol-level multiuser MISO precoding for multi-level adaptive modulation," *IEEE Trans. Wireless Commun.*, vol. 16, no. 8, pp. 5511– 5524, Aug. 2017.

- [12] S. Domouchtsidis, C. G. Tsinos, S. Chatzinotas, and B. Ottersten, "Symbol-level precoding for low complexity transmitter architectures in large-scale antenna array systems," *IEEE Trans. Wireless Commun.*, vol. 18, no. 2, pp. 852–863, Feb. 2019.
- [13] A. Li and C. Masouros, "Interference exploitation precoding made practical: Optimal closed-form solutions for PSK modulations," *IEEE Trans. Wireless Commun.*, vol. 17, no. 11, pp. 7661–7676, Nov. 2018.
- [14] A. Haqiqatnejad, F. Kayhan, and B. Ottersten, "Constructive interference for generic constellations," *IEEE Signal Process. Lett.*, vol. 25, no. 4, pp. 586–590, Apr. 2018.
- [15] ——, "Symbol-level precoding design based on distance preserving constructive interference regions," *IEEE Trans. Signal Process.*, vol. 66, no. 22, pp. 5817–5832, Nov. 2018.
- [16] —, "An approximate solution for symbol-level multiuser precoding using support recovery," in *Proc. IEEE Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Jul. 2019, pp. 1–5.
- [17] Y. Wang, H. Hou, W. Wang, and X. Yi, "Symbol-level precoding for average SER minimization in multiuser MISO systems," *IEEE Wireless Commun. Lett.*, vol. 13, no. 4, pp. 1103–1107, Apr. 2024.
- [18] M. Shao, Q. Li, and W.-K. Ma, "Minimum symbol-error probability symbol-level precoding with intelligent reflecting surface," *IEEE Wireless Commun. Lett.*, vol. 9, no. 10, pp. 1601–1605, Oct. 2020.
- [19] A. Mohammad, C. Masouros, and Y. Andreopoulos, "An unsupervised deep unfolding framework for robust symbol-level precoding," *IEEE Open J. Commun. Soc.*, vol. 4, pp. 1075–1090, Apr. 2023.
- [20] A. Haqiqatnejad, F. Kayhan, and B. Ottersten, "Robust design of power minimizing symbol-level precoder under channel uncertainty," in *Proc. IEEE Glob. Commun. Conf. (GLOBECOM)*, Dec. 2018, pp. 1–6.
- [21] G. Hegde, C. Masouros, and M. Pesavento, "Interference exploitation-based hybrid precoding with robustness against phase errors," *IEEE Trans. Wireless Commun.*, vol. 18, no. 7, pp. 3683–3696, Jul. 2019.
- [22] A. Haqiqatnejad, F. Kayhan, and B. Ottersten, "Robust SINR-constrained symbol-level multiuser precoding with imperfect channel knowledge," *IEEE Trans. Signal Process.*, vol. 68, pp. 1837–1852, Mar. 2020.
- [23] M. Alodeh, S. Chatzinotas, and B. Ottersten, "Constructive multiuser interference in symbol level precoding for the MISO downlink channel," *IEEE Trans. Signal Process.*, vol. 63, no. 9, pp. 2239–2252, May 2015.
- [24] K. L. Law and C. Masouros, "Constructive interference exploitation for downlink beamforming based on noise robustness and outage probability," in *IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2016, pp. 3291–3295.
- [25] —, "Symbol error rate minimization precoding for interference exploitation," *IEEE Trans. Commun.*, vol. 66, no. 11, pp. 5718–5731, Nov. 2018
- [26] Y. Wang, W. Wang, L. You, C. G. Tsinos et al., "Weighted MMSE precoding for constructive interference region," *IEEE Wireless Commun. Lett.*, vol. 11, no. 12, pp. 2605–2609, Dec. 2022.
- [27] J. Zheng, J. Zhang, E. Björnson, and B. Ai, "Impact of channel aging on cell-free massive MIMO over spatially correlated channels," *IEEE Trans. Wireless Commun.*, vol. 20, no. 10, pp. 6451–6466, Oct. 2021.
- [28] H. Hou, Y. Wang, Y. Zhu, X. Yi et al., "A tensor-structured approach to dynamic channel prediction for massive MIMO systems with temporal non-stationarity," arXiv preprint arXiv:2412.06713, 2024.
- [29] J. Zhuang, X. He, Y. Wang, J. Liu et al., "Covnet: Covariance information-assisted CSI feedback for FDD massive MIMO systems," *IEEE Wireless Commun. Lett.*, vol. 14, no. 3, pp. 641–645, Mar. 2025.
- [30] H. Hou, Y. Wang, X. Yi, W. Wang et al., "Joint beam alignment and doppler estimation for fast time-varying wideband mmwave channels," *IEEE Trans. Wireless Commun.*, vol. 23, no. 9, pp. 10895–10910, Sept. 2024
- [31] J. Zhuang, Y. Wang, H. Hou, Y. Han et al., "Extract the best, discard the rest: CSI feedback with offline large AI models," arXiv preprint arXiv:2505.08566, 2025.
- [32] A.-A. Lu, X. Gao, W. Zhong, C. Xiao et al., "Robust transmission for massive MIMO downlink with imperfect CSI," *IEEE Trans. Commun.*, vol. 67, no. 8, pp. 5362–5376, Aug. 2019.
- [33] Y. Wang, X. Yi, H. Hou, and W. Wang, "Robust symbol-level precoding for MIMO downlink transmission with channel aging," in *IEEE Glob. Commun. Conf. (GLOBECOM)*, Dec. 2023, pp. 5763–5768.
- [34] Y. Wang, X. Yi, H. Hou, W. Wang *et al.*, "Robust symbol-level precoding for massive MIMO communication under channel aging," *IEEE Trans. Wireless Commun.*, vol. 23, no. 9, pp. 10864–10878, Sept. 2024.
- [35] A. Mohammad, C. Masouros, and Y. Andreopoulos, "An unsupervised learning-based approach for symbol-level-precoding," in *Proc. IEEE Glob. Commun. Conf. (GLOBECOM)*, Dec. 2021, pp. 1–6.

- [36] A. Haqiqatnejad, F. Kayhan, and B. Ottersten, "Power minimizer symbol-level precoding: A closed-form suboptimal solution," *IEEE Signal Process. Lett.*, vol. 25, no. 11, pp. 1730–1734, Nov. 2018.
- [37] J. Krivochiza, J. Merlano-Duncan, S. Andrenacci, S. Chatzinotas et al., "Closed-form solution for computationally efficient symbol-level precoding," in *Proc. IEEE Glob. Commun. Conf. (GLOBECOM)*, Dec. 2018, pp. 1–6.
- [38] Y. Wang, H. Hou, X. Yi, W. Wang et al., "Towards unified AI models for MU-MIMO communications: A tensor equivariance framework," *IEEE Trans. Wireless Commun.*, Early Access, 2025.
- [39] Y. Wang, H. Hou, W. Wang, X. Yi et al., "Soft demodulator for symbol-level precoding in coded multiuser MISO systems," *IEEE Trans. Wireless Commun.*, vol. 23, no. 10, pp. 14819–14835, Oct. 2024.
- [40] K. Pratik, B. D. Rao, and M. Welling, "RE-MIMO: Recurrent and permutation equivariant neural MIMO detection," *IEEE Trans. Signal Process.*, vol. 69, pp. 459–473, Dec. 2020.
- [41] Y. Wang, V. N. Ha, K. Ntontin, H. Yan et al., "Statistical CSI-based distributed precoding design for OFDM-cooperative multi-satellite systems," arXiv preprint arXiv:2505.08038, 2025.
- [42] Y. Choi, J. Lee, M. Rim, and C. G. Kang, "Constructive interference optimization for data-aided precoding in multi-user MISO systems," *IEEE Trans. Wireless Commun.*, vol. 18, no. 2, pp. 1128–1141, Jan. 2019.
- [43] A. Li, F. Liu, X. Liao, Y. Shen et al., "Symbol-level precoding made practical for multi-level modulations via block-level rescaling," in *IEEE Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Sept. 2021, pp. 71–75.
- [44] C. L. Lawson and R. J. Hanson, Solving least squares problems. Philadelphia, PA, USA: SIAM, 1995.
- [45] M. Zaheer, S. Kottur, S. Ravanbakhsh, B. Poczos et al., "Deep sets," Neural Inf. Proces. Syst. (NeurIPS), vol. 30, Dec. 2017.
- [46] J. Hartford, D. Graham, K. Leyton-Brown, and S. Ravanbakhsh, "Deep models of interactions across sets," in *Int. Conf. Mach. Learn. (ICML)*, Jul. 2018, pp. 1909–1918.
- [47] C. Yun, S. Bhojanapalli, A. S. Rawat, S. J. Reddi et al., "Are transformers universal approximators of sequence-to-sequence functions?" in Int. Conf. Learn. Represent. (ICLR), Apr. 2019.
- [48] J. Kim, D. Nguyen, S. Min, S. Cho et al., "Pure transformers are powerful graph learners," Neural Inf. Proces. Syst. (NeurIPS), vol. 35, pp. 14582–14595, Nov. 2022.
- [49] J. Lee, Y. Lee, J. Kim, A. Kosiorek et al., "Set transformer: A framework for attention-based permutation-invariant neural networks," in *Int. Conf. Mach. Learn. (ICML)*, Jun. 2019, pp. 3744–3753.
- [50] N. Keriven and G. Peyré, "Universal invariant and equivariant graph neural networks," *Neural Inf. Proces. Syst. (NeurIPS)*, vol. 32, Dec. 2019
- [51] H. Pan and R. Kondor, "Permutation equivariant layers for higher order interactions," in *Proc. Int. Conf. Artif. Intell. Statist. (AISTATS)*, Mar. 2022, pp. 5987–6001.
- [52] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, September 2018.
- [53] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog. (CVPR)*, June 2016.
- [54] H. Hou, Y. Wang, X. Yi, W. Wang et al., "Tensor-structured bayesian channel prediction for upper mid-band XL-MIMO systems," arXiv preprint arXiv:2508.08491, 2025.
- [55] S. Jaeckel, L. Raschkowski, K. Börner, and L. Thiele, "Quadriga: A 3-D multi-cell channel model with time evolution for enabling virtual field trials," *IEEE Trans. Antennas Propag.*, vol. 62, no. 6, pp. 3242–3256, Mar. 2014.
- [56] 3GPP, "Tr 38.901 v19.0.0: Study on channel model for frequencies from 0.5 to 100 GHz," 3GPP, Tech. Rep. TR 38.901 V19.0.0, Jun 2025.
- [57] C. Peel, B. Hochwald, and A. Swindlehurst, "A vector-perturbation technique for near-capacity multiantenna multiuser communication-part I: Channel inversion and regularization," *IEEE Trans. Commun.*, vol. 53, no. 1, pp. 195–202, Jan. 2005.