

Explainable Neural Network for Joint Orthogonal Bases of Doubly Selective Channels

Zhibin Zou, Iresha Amarasekara and Aveek Dutta
 Department of Electrical and Computer Engineering
 University at Albany SUNY, Albany, NY 12222 USA
 {zzou2, iamarasekara, adutta}@albany.edu

Abstract—In this paper, we propose an explainable neural network for decomposing channel kernels into Eigenwaves and implement practical Multi-dimensional Eigenwave Multiplexing (MEM) over doubly selective channels. The quality of Eigenwave decomposition is evaluated using three key metrics: 1) eigenvalue, 2) orthogonality, and 3) duality. The eigenvalue determines the subchannel gains in the eigen domain, while orthogonality and duality impact the interference from other symbols and the distortion of the target symbol, respectively. We prove that maximizing the sum of eigenvalues is equivalent to minimizing the MSE loss function and demonstrate that the duality and orthogonality constraints not only minimize interference for multiplexing but also guide the convergence for NN. Furthermore, we show that these duality and orthogonality constraints are equivalent, allowing them to be combined for model simplification. To further enhance the adaptability of the proposed method, we introduce a second NN architecture that incorporates the Augmented Lagrangian Method (ALM). This approach eliminates the need for retraining under different MIMO scales and does not require parameter tuning. We evaluate the proposed methods under two scenarios: 1) 2-D doubly selective channels, and 2) 4-D doubly selective MIMO channels with both perfect imperfect Channel State Information (CSI) and imperfect CSI.

Keywords—Eigen-decomposition, Multi-dimensional channel, Interference cancellation, Neural Networks.

I. INTRODUCTION

Wireless channels with time-varying multipath fading cause dispersion in both time and frequency and are generally called Linear Time-Varying (LTV) channels. Specifically, multipath leads to time dispersion, resulting in frequency selectivity, while the time-varying gains of each path cause frequency dispersion, leading to time selectivity. This type of channel is referred to as a doubly selective channel [1]. Compared to Linear Time-Invariant (LTI) channels, waveform design is challenging for doubly selective channels because of joint interference across Degrees of Freedom (DoFs).

Modulation is a general approach in communication systems where data-symbols modulate a set of orthonormal basis waveforms. A classical scheme is Orthogonal Frequency-Division Multiplexing (OFDM), which divides the frequency selective frequency bands into multiple coherent subspaces to avoid Inter-Symbol Interference (ISI) caused by time dispersion. Along with a cyclic prefix, OFDM symbols are interference-free in LTI channels. However, when the channel becomes time-selective, frequency dispersion introduces Inter-Carrier Interference (ICI) [2]. Orthogonal Time Frequency Modulation (OTFS) [3], [4] has been proposed to mitigate

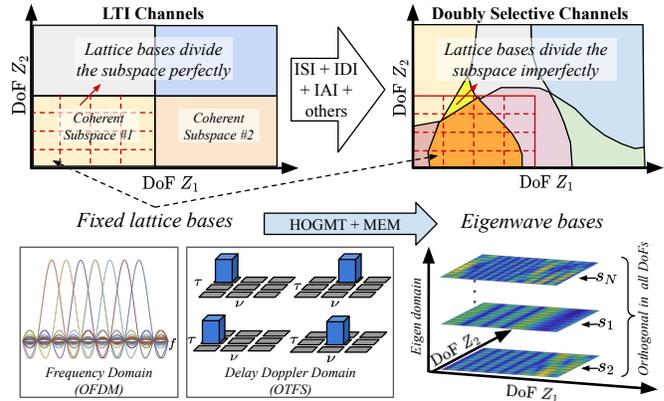


Figure 1: In LTI channels, OFDM and OTFS bases are fixed lattice bases that divide the coherent subspaces along their boundaries. However, in doubly selective channels, the subspaces are irregular and lattice bases imperfectly divide the subspaces resulting in joint interference across the DoFs. In contrast, Eigenwaves are jointly orthogonal across all DoFs, providing division-free bases that are maintain orthogonality during transmission in doubly selective channels.

joint interference in the time-frequency domain by employing orthogonal bases in the delay-Doppler domain. When the channel is rapidly time-varying, the acceleration variance causes Doppler dispersion causing the OTFS symbols to experience Inter-Doppler Interference (IDI) [5], [6].

Eigenfunctions, also referred to as Eigenwaves, are considered optimal basis functions [1] [7, Section 2.5.2.1] [8]. Since Eigenwaves are orthogonal in the eigen domain, they achieve orthogonality across all Degrees of Freedom (DoF), regardless of how the channel is represented. The channel effect on each basis function then simply reduces to a multiplicative scalar, eliminating interference between the basis functions. Figure 1 shows the differences between three types of bases. OFDM and OTFS bases are orthogonal divisions in the frequency domain and delay-Doppler domain, respectively, achieving orthogonality only within their specific domains. In contrast, *Eigenwaves are jointly orthogonal surfaces across the channel's representation domains, which are not necessarily limited to the time, frequency, and/or delay-Doppler domains.* In LTI channels, OFDM and OTFS bases can perfectly divide coherent subspaces. However, this is challenging under complex channel conditions. Longer pulses reduce ISI but are more susceptible to ICI due to time selectivity, while shorter pulses

reduce ICI but are more prone to ISI due to frequency selectivity. In OFDM and OTFS, the bases are pre-defined and thus the orthogonality is influenced by the properties of the channel. In contrast, Eigenwaves are decomposed from the *channel kernel*, a fundamental representation of the channel [9]–[11], ensuring that orthogonality is preserved during transmission. Unlike the Fourier bases of OFDM and OTFS, which are directly obtained through the Discrete Fourier Transform (DFT) and Symplectic Fourier Transform (SFT) respectively, there are no predefined Eigenwaves for doubly selective channels¹.

Mercer’s Theorem decomposes a symmetric kernels into infinite Eigenwaves. However, in wireless communication the channel kernel (2-D matrix or 4-D tensor) is often asymmetric [11]. Therefore, High Order Generalized Mercer’s Theorem (HOGMT) [12] has been proposed to generalize the decomposition for multi-dimensional asymmetric kernels (reviewed in Section III), demonstrating the existence of Eigenwaves for any wireless channels. In general, a multi-dimensional channel kernel is represented as a tensor instead of separate matrices to capture the joint interference across the dimensions. However, decomposing tensors according to HOGMT via linear algebraic methods suffer from extremely high computational overhead [12], which limits its adoption for real-time signal processing in next Generation (xG) wireless networks.

One of the advantages of Neural Network (NN) over linear methods is its ability to extract non-linear relationship among hidden variables with manageable computational complexity. This motivates us to embrace NN to implement practical HOGMT-based eigenwave multiplexing. The literature on NN-based eigen decomposition apply a black-box approach that focuses on learning the format and properties of eigenvectors [13] without explaining *how and why* its outputs converge to the eigen components (e.g., vectors, tensors, or functions). However, stringent requirements on reliability and throughput in xG wireless propagation environments [14], [15] demand performance guarantees for NN transceivers. To address this, we construct a provable trilateral relationship among multiplexing, HOGMT, and NNs. Since, multiplexing minimize interference between the DoFs while achieving diversity, eigenwaves obtained from HOGMT provide the optimal bases for doubly selective channels with eigenvalues translating to subchannel gains. By mapping these properties to the objective function of the NN, we implement HOGMT for optimal multiplexing with practical complexity. Overall, this paper answers *three open problems* discussed in the literature:

- “How to obtain eigenfunctions for the the doubly selective channels in practice?” [1] [7, Section 2.5.2.1].
- “What is the optimal training for doubly dispersive MIMO channels?” [16].
- “How can MMSE gains be leveraged through online learning of the channel statistics” [16].

Our approach provides promising solutions to these problems due to the following: 1) the *practicality* of HOGMT-

¹With cyclic prefix, OFDM bases serve as common Eigenwaves for all LTI channels, but this does not hold for doubly selective channels.

based NN, 2) the *optimality* of employing Eigenwave as bases for doubly selective MIMO channels, and 3) the *equivalence* between MMSE gains and subchannel gains in the eigen domain. To further enhance the practicality of the proposed method, we embrace ALM into our architecture which avoid the exhausting tuning for the parameters and is adaptive to varying channel types such as 2-D doubly selective channels and 4-D doubly selective MIMO channels.

The rest of the paper² is organized as follows: *Section II* reviews related work in the field, while *Section III* provides the necessary technical background on lattice bases, the general channel kernel and HOGMT. In *Section IV*, we formulate the finite Eigenwave approximation problem and show that the duality and orthogonality of Eigenwaves are essential for achieving optimal bases over doubly selective channels. In *Section V*, we show that the MSE loss function optimizes both the kernel approximation and subchannel gains, which is employed in the design of an explainable NN, HNet, for HOGMT, with outputs converging to Eigenwaves. *Section VI*, presents an adaptive architecture AHNet by incorporating ALM to adapt the NN parameters for different channel types and distributions along with complexity analysis for HNet and AHNet. AHNet is used to implement the AHNet-MEM system, where the NN outputs are subchannels or eigenwaves of the channel kernel. *Section VII* details the data generation and NN training, with channel kernel generated from QuaDriga [18] that exhibit dispersion in both time and frequency domains. The accuracy of AHNet, HNet and the performance of AHNet-MEM is evaluated in *Section VIII*, in two scenarios: 1) 2D doubly selective channels, and 2) 4D doubly selective MIMO channels under both perfect CSI and imperfect CSI. Finally, discussions about the cost, opportunities and concluding remarks are given in *Section IX* and *Section X*, respectively.

II. RELATED WORK

Lattice Bases for Waveform Design: OFDM is the most common modulation which minimizes interference in LTI channels by designing lattice bases in the frequency domain. While, OTFS is proposed to cancel interference in the time-frequency domain by designing lattice bases in the delay-Doppler domain [3]. Orthogonal short-time Fourier (STF) lattices are explored for signaling in doubly dispersive channels [1]. There are many investigations on the lattice design based on Weyl–Heisenberg and Gabor frames [19]–[22]. However, these are suboptimal alternatives for the Eigenwaves, which are

²This manuscript is a combined extension of prior publications [8], [17]. In [8], we formulated a general kernel for multi-dimensional channels and proposed an equivalent NN for implementing HOGMT, providing a performance benchmark for MEM with non-orthogonal Eigenwaves. Based on this, we introduce an adaptive NN by incorporating ALM in [17]. However, these works did not implement a practical NN-based MEM. This journal article extends the previous studies from both theoretical and system perspectives. Theoretically, we establish provable trilateral relationships among multiplexing, HOGMT, and NN, demonstrating step-by-step how the eigenvalues, duality and orthogonality of Eigenwaves relate to multiplexing performance and map to the NN’s objective function from first principles. From a systems perspective, we utilize practical 4-D channels generated by QuaDriga to assess NN convergence and evaluate MEM performance using NN outputs.

considered as the optimal waveforms, though undesirable due to the limitation of tools [1], [7]. This motivates us to design NN for decomposing eigenfunctions for practical waveforms. Modulation divides data into orthogonal subcarriers, while multiple access techniques such as Orthogonal Frequency Division Multiple Access (OFDMA) [23] and Pattern Division Multiple Access (PDMA) [24] assigns the subcarriers to different users. Unlike precoding these methods are not designed to cancel Inter-antenna-interference (IAI) in MIMO channels due to interference among subcarriers. Instead, they manage and mitigate interference in unique ways, enabling multiple users to share resources simultaneously. Similarly, MEM can be easily extended to Multi-dimensional Eigenwave Multiple Access (MEMA). Since this paper focuses on multiplexing, a detailed exploration of multiple access is beyond its scope.

Eigen Approximation Problem: There are some theoretical works analyzing eigen approximation problems. Karhunen–Loève Theorem (KLT) approximation is proven to be optimal for the random process approximation by finite eigenfunctions in [25]. Eckart–Young–Mirsky Theorem shows that SVD is optimal for low-rank matrix approximation [26]. Nyström approximation [27] shows rank-k approximation using SVD is optimal for Symmetric Positive Semi-Definite (SPSD) matrix. However, there is no optimality analysis for eigenfunction approximation for multi-dimensional asymmetric kernels in the literature. For the implementation of eigen approximation, [13] present black-box NNs for SVD decomposition, which is applicable for matrices. [28] proposed a NN-based method for extracting eigenfunctions based on Mercer’s Theorem. However, it fails to show the optimality of eigen approximation and is only applicable for symmetric kernels. [12] generalize Mercer’s Theorem to multi-dimensional asymmetric kernels. However, its implementation using linear Algebra suffers from high computational complexity.

Explainable Neural Networks in Communications: The explainability of NN-based transceivers is very rare in the literature, attributed primarily to their black-box implementation. [29] combines an autoencoder with a classical OFDM system to inherit its advantages, including robustness to synchronization errors and multipath equalization. The model-driven OFDM receiver in [30] combines DL with expert knowledge and implements two subnets for channel estimation and symbol detection modular structure similar to classical systems. Model-driven NN shows a competitive performance with data-driven approaches but with a lower number of trainable variables than the black box architecture. However, these approaches are limited by the flexibility and accuracy of the underlying model and the trained weights, which limits their applicability to specific channel conditions. In contrast, we ensure performance guarantees for the NN by explaining its internal mechanisms that achieves equivalence to HOGMT. Additionally, the designed NN does not directly produce the transmitted signal; instead, it extracts Eigenwaves from the kernel, following NN’s feature extraction principle.

III. PRELIMINARIES AND BACKGROUND

A. Lattice Bases

In general, the channel effect is caused by both multipath delay τ and Doppler shifts ν . The received signal $r(t)$ can be given by a transmitted signal $x(t)$ undergoing a channel operator \mathcal{H} [11] as

$$r(t) = \mathcal{H}(x(t)) = \iint S_H(\tau, \nu) x(t-\tau) e^{j2\pi\nu t} d\tau d\nu \quad (1)$$

where $S_H(\tau, \nu)$ is the (*delay-Doppler*) *spreading function*, which describes the combined attenuation factor for all paths in the delay-Doppler domain.

Let $\{g_{l,m}(t)\}$ be orthonormal time-frequency lattice bases with time slot index l and frequency slot index m (for OTFS, the lattice is in the delay-Doppler domain), orthogonal multiplexing methods modulate data symbols $\{s_{l,m}\}$ onto these bases by

$$x(t) = \sum_{l,m} s_{l,m} g_{l,m}(t) \quad (2)$$

Given the signaling duration time T , time separation T_0 , bandwidth B and frequency separation B_0 , the dimension of the signal subspace is approximately $N = \lceil \frac{TF}{T_0 B_0} \rceil$, where $\lceil x \rceil$ is the least not less than x . The received data symbol at (l,m) by match filter is given [1] as

$$r_{l,m} = \sum_{l',m'} c_{lm,l'm'} s_{l',m'} + v_{lm} \quad (3)$$

where $c_{lm,l'm}$ is a coefficient that measures the degree of matching between the transmit basis $g_{l',m'}(t)$ and the receive basis $g_{l,m}(t)$, given by

$$c_{lm,l'm} = \langle \mathcal{H}(g_{l',m'}(t)), g_{l,m}(t) \rangle \quad (4)$$

where $\langle a(t), b(t) \rangle = \int a(t) b^*(t) dt$ is the inner product.

Coherence time T_{coh} and coherence bandwidth B_{coh} are measurements for the variation of the channel in time and frequency domain, respectively. When $T_{coh} B_{coh} \geq 1$, the channel called *underspread*, otherwise called *overspread*. Ideally, the bases should be subdivided within the coherence subspace as shown in, so that the bases would not undergo interference from each other, i.e., $c_{lm,l'm} = 0$ for $l \neq l'$ and $m \neq m'$.

B. General Channel Kernel

The doubly selective channel can be represented in different domains. The time domain continuous input-output relations [11] is given by,

$$r(t) = \int h(t;\tau) x(t-\tau) d\tau = \int k(t;t') x(t') dt' \quad (5)$$

where $h(t;\tau)$ is the time-varying impulse response. $k(t;t') = h(t;t-\tau)$ is the channel kernel defined in [9], [11]. In OFDM, OTFS, and MIMO systems, the channel is represented in the frequency, delay-Doppler, and spatial domains, respectively. Since Eigenwaves are bases in a unified eigen domain, we formulate the general input-output relation for doubly selective channels as,

$$r(Z) = \int H(Z;\Gamma) x(Z-\Gamma) d\Gamma = \int K(Z;Z') x(Z') d\Gamma \quad (6)$$

where $H(Z;\Gamma)$ is the multi-dimensional transfer function. $K(Z;Z') = H(Z;Z-\Gamma)$ is the general channel kernel. Z and

Table I: Eigen decomposition methods

Method	Formulation	Applications
SVD	$H=U\Sigma V^*$	Precoding
KLT	$e(t)=\sum_{n=1}^{\infty}\sigma_n\phi_n(t)$	Signal detection
Mercer's Theorem	$e(t,t')=\sum_{n=1}^{\infty}\lambda_n\phi_n(t)\phi_n(t')$	Feature extraction
HOGMT	$K(Z;Z')=\sum_{n=1}^{\infty}\sigma_n\psi_n(Z)\phi_n(Z')$	General to above

Z' can be seen as the signal domains at the receiver and transmitter, respectively. In the time-frequency domain, $(Z;Z')=(t,f;t',f')$, while in the space-delay-Doppler domain, $(Z;Z')=(u,\tau,\nu;u',\tau',\nu')$, where u and u' denote the antenna index at the receiver and transmitter, respectively. Note that $K(Z;Z')$ is considered asymmetric for the general channel, i.e., $K(X;Y)\neq K(Y;X)$.

C. High Order Generalized Mercer's Theorem

A symmetric kernel $e(t,t')$ can be decomposed into one Eigenwave set in two domains by Mercer's theorem as,

$$e(t;t')=\sum_{n=1}^{\infty}\lambda_n\phi_n(t)\phi_n(t') \quad (7)$$

where, ϕ_n denotes the Eigenwave and λ_n is the corresponding eigenvalue. HOGMT generalizes Mercer's Theorem for multi-dimensional asymmetric kernels in (6), into two different, low-dimension, jointly orthogonal Eigenwave sets as³,

$$K(Z;Z')=\sum_{n=1}^{\infty}\sigma_n\psi_n(Z)\phi_n(Z') \quad (8)$$

where $\mathbb{E}\{\sigma_n\sigma_{n'}\}=\lambda_n\delta_{nn'}$. λ_n is the n -th eigenvalue and $\psi_n(Z)$ and $\phi_n(Z')$ are orthonormal Eigenwaves, i.e.,

$$\int\phi_n(Z')\phi_{n'}^*(Z')dZ'=\delta_{nn'}; \int\psi_n(Z)\psi_{n'}^*(Z)dZ=\delta_{nn'} \quad (9)$$

These Eigenwaves are referred as *dual Eigenwaves* that exhibit the important *duality* property,

$$\int K(Z;Z')\phi_n^*(Z')dZ'=\sigma_n\psi_n(Z) \quad (10)$$

Table I shows the existing eigen-decomposition methods. SVD is limited to 2-D matrices and is commonly applied in SVD-based precoding for MIMO channels. KLT, while useful for signal detection, cannot decompose high-dimensional kernels into lower-dimensional Eigenwaves. Mercer's theorem, applicable only to symmetric kernels, is widely used for feature extraction in machine learning. As a generalized version of Mercer's Theorem, HOGMT can decompose any multi-dimensional kernel into low-dimensional Eigenwaves, providing theoretical evidence that eigenwaves exist for any wireless channel. The signal space at the transmitter, with DoFs Z' is spanned by Eigenwaves $\{\phi_n\}$ while the signal space at the receiver, with DoFs Z is spanned by Eigenwaves $\{\psi_n\}$. According to the duality of Eigenwaves, the two eigen spaces can be mapped onto each other, enabling the transmission of symbols in the two eigen spaces by multiplexing and demultiplexing. It is important to note that the orthogonality of Eigenwaves is not affected by the general kernel, making them suitable as basis functions for doubly selective channels.

³In 1907, Schmidt defined eigenfunctions for both symmetric and asymmetric (then termed "unsymmetric") kernels. However, his work only demonstrated the existence of eigenfunctions, not a eigen decomposition [31].

IV. EIGENWAVES AS OPTIMAL BASES

A. Finite Eigenwave Approximation Problem

HOGMT is a *mathematical principle* to decompose any kernel into infinite number of Eigenwaves. However, in reality, we can only utilize a finite number of those to approximate the kernel. To achieve maximum energy efficiency, it is desirable to use minimum number of eigenwaves to approximate (most part of) the kernel. Therefore, the general approximation problem is formulated as minimizing the number of Eigenwaves N for a fixed kernel approximation error ϵ as in (11),

$$\operatorname{argmin}_N N \text{ s.t. } \|K(Z;Z')-\hat{K}(Z;Z')\|^2<\epsilon \quad (11)$$

where, $\hat{K}(Z;Z')=\sum_{n=1}^N k_n f_n(Z)g_n(Z')$ is the approximate kernel, $\{f_n\}$ and $\{g_n\}$ are two sets of arbitrary orthonormal bases and k_n is the projection of the kernel onto the bases.

However, for fixed N (11) is equivalent to the MMSE problem in (12),

$$\operatorname{argmin}_{\hat{K}} \mathbb{E}\{\|K(Z;Z')-\hat{K}(Z;Z')\|^2\} \quad (12)$$

This problem is solved by proving that the approximate kernel, \hat{K} , reconstructed from the Eigenwaves obtained using HOGMT is optimal in MMSE sense. The choice and trade-offs regarding N are evaluated in Section VI-C and Section VIII.

B. Duality and Joint Orthogonality of Eigenwaves

The duality in (10) indicates that the Eigenwaves are transferred to their dual Eigenwaves scaled only by the channel gains when transmitting over the channel and hence immune to any joint interference across all DoFs.

Proposition 1. *The average distortion for given bases at the transmitter and the receiver $g_{tx}(Z')$ and $g_{rx}(Z)$ is*

$$\epsilon=\mathbb{E}\{\|\mathcal{H}(g_{tx}(Z'))-a g_{rx}(Z)\|^2\} \quad (13)$$

where $\|\cdot\|$ denote Frobenius norm. When the transmit and the receive bases are a pair of dual Eigenwaves, i.e., $a=\sigma_n$, $g_{tx}(Z')=\phi_n(Z')$ and $g_{rx}(Z)=\psi_n(Z)$, we have $\epsilon=0$.

Proof. See Appendix A □

Proposition 1 shows that for each pair of dual Eigenwaves the channel exhibits flat-fading. Meanwhile, when multiplexing with Eigenwaves, the interference from other bases can be avoided by joint orthogonality according to (14)

Proposition 2. *The received symbol is given as*

$$r_n=\sigma_n s_n + \sum_{n'\neq n}^N s_{n'} \langle \mathcal{H}(\phi_n(Z')), \psi_{n'}(Z) \rangle \quad (14)$$

The joint orthogonality of Eigenwaves across multi-dimension domains Z and Z' makes the second term zero, e.g., avoid interference from other bases.

Proof. See Appendix B. □

Note that ISI, ICI, IDI and IAI represents interference from basis functions in different domains. Since Eigenwaves are orthogonal in the eigen domain, they inherently avoid all joint interference across its DoFs.

V. EXPLAINABLE NN FOR DUAL EIGENWAVES

A. Loss Function Design: Maximizing the sum of Eigenvalues

In eigen-decomposition the eigenvalue is a measure for *quality* of the decomposition. For a fixed number of Eigenwaves, larger sum of eigenvalues means that more information is extracted from the kernel. Therefore, maximizing this sum must be part of the objective function for the NN.

Theorem 1. *If $K(Z;Z')$ is approximated by a finite number of Eigenwaves as $\widehat{K}(Z;Z')=\sum_{n=1}^N\sigma_n\psi_n(Z)\phi_n(Z')$, then maximizing the sum of eigenvalues,*

$$\operatorname{argmax}_{\{\phi_n\},\{\psi_n\}} \sum_{n=1}^N \mathbb{E}\{\sigma_n^2\} \quad (15)$$

is equivalent to MMSE,

$$\operatorname{argmin}_{\{\phi_n\},\{\psi_n\}} \mathbb{E}\{\|K(Z;Z')-\widehat{K}(Z;Z')\|^2\} \quad (16)$$

Proof. Maximizing $\sum_{n=1}^N \mathbb{E}\{\sigma_n^2\}$ is equivalent to minimizing its negative. Furthermore, $\iint \mathbb{E}\{\|K(Z;Z')\|^2\} dZ dZ'$ is a constant that only depends on the total transmission gains of the channel, which is independent of the choice of $\{\psi_n\}$ and $\{\phi_n\}$. Therefore, (15) is rewritten as

$$\operatorname{argmin}_{\{\phi_n\},\{\psi_n\}} \iint \mathbb{E}\{\|K(Z;Z')\|^2\} dZ dZ' - \sum_{n=1}^N \mathbb{E}\{\sigma_n^2\} \quad (17)$$

Now, notice that,

$$\begin{aligned} & \iint \mathbb{E}\{K(Z;Z')\widehat{K}(Z;Z')^*\} dZ dZ' \\ &= \iint \mathbb{E}\left\{\sum_{n=1}^{\infty} \sigma_n \psi_n(Z) \phi_n(Z') \sum_{n'=1}^N \sigma_{n'} \psi_{n'}^*(Z) \phi_{n'}^*(Z')\right\} dZ dZ' \\ &= \iint \mathbb{E}\left\{\sum_{n=1}^N \underbrace{\sigma_n^2}_{=1} \underbrace{\|\psi_n(Z)\|^2}_{=1} \underbrace{\|\phi_n(Z')\|^2}_{=1}\right. \\ & \left. + \sum_{n \neq n'} \sigma_n \sigma_{n'} \underbrace{\psi_n(Z) \psi_{n'}^*(Z)}_{=0, \text{ by (9)}} \underbrace{\phi_n(Z') \phi_{n'}^*(Z')}_{=0, \text{ by (9)}}\right\} dZ dZ' = \sum_{n=1}^N \mathbb{E}\{\sigma_n^2\} \end{aligned} \quad (18)$$

Furthermore, note that

$$\begin{aligned} & \iint \mathbb{E}\{\|\widehat{K}(Z;Z')\|^2\} dZ dZ' \\ &= \iint \mathbb{E}\left\{\sum_{n=1}^N \underbrace{\sigma_n^2}_{=1} \underbrace{\|\psi_n(Z)\|^2}_{=1} \underbrace{\|\phi_n(Z')\|^2}_{=1}\right\} dZ dZ' = \sum_{n=1}^N \mathbb{E}\{\sigma_n^2\} \end{aligned} \quad (19)$$

Then (17) is rewritten as,

$$\begin{aligned} & \iint \mathbb{E}\{\|K(Z;Z')\|^2\} dZ dZ' - 2 \sum_{n=1}^N \mathbb{E}\{\sigma_n^2\} + \sum_{n=1}^N \mathbb{E}\{\sigma_n^2\} \\ &= \iint \mathbb{E}\{\|K(Z;Z')\|^2\} - 2 \underbrace{\mathbb{E}\{K(Z;Z')\widehat{K}(Z;Z')^*\}}_{\text{by (18)}} \\ & \quad + \underbrace{\mathbb{E}\{\|\widehat{K}(Z;Z')\|^2\}}_{\text{by (19)}} dZ dZ' \\ &= \iint \mathbb{E}\{\|K(Z;Z')-\widehat{K}(Z;Z')\|^2\} dZ dZ' \end{aligned} \quad (20)$$

Substituting (20) in (17), we have (15), which implies that minimizing the integral of MSE is equivalent to MMSE in (16). From an inverse deduction, we can also obtain (16) implies (15), so that (15) and (16) are equivalent. \square

Theorem 1 shows that the loss function in (16) maximizes the sum of eigenvalues as well as address the finite Eigenwaves approximation problem discussed in Section IV-A.

B. NN Constraints: Duality and Orthogonality

The MMSE loss function is equivalent to maximize the sum of eigenvalues, while by the nature of NN, constraints are responsible for maintaining orthogonality and duality, which are critical properties of Eigenwaves for multiplexing, as discussed in Proposition 1 and Proposition 2. However, both orthogonality and duality are desired for effective multiplexing rather than being specifically designed to improve training. Before employing them as constraints, two potential challenges must be addressed: (a) *the constraints should guide the outputs to converge with the loss function*, and (b) *the constraints need to be simplified to reduce the model complexity*.

Lemma 1. *(Necessity of duality) Duality constraint is a necessity for achieving the maximum sum of eigenvalues.*

Proof. Since by definition, $\lambda_n = \mathbb{E}\{\sigma_n^2\}$, where σ_n is the projection of the kernel onto the Eigenwaves, we have

$$\sum_{n=1}^N \lambda_n = \sum_{n=1}^N \mathbb{E}\left\{\left[\iint K(Z;Z') \psi_n^*(Z) \phi_n^*(Z') dZ dZ'\right]^2\right\} \quad (21)$$

Since maximizing a squared term is equivalent to maximizing its absolute value, we introduce a Lagrange multiplier $\frac{1}{2}\beta_n$ associated with the constraint for ψ_n and maximize \mathcal{E} as,

$$\begin{aligned} \mathcal{E} = & \sum_{n=1}^N \mathbb{E}\left\{\left|\iint K(Z;Z') \psi_n^*(Z) \phi_n^*(Z') dZ dZ'\right.\right. \\ & \left.\left. - \frac{1}{2}\beta_n \left(\int \psi_n(Z) \psi_n^*(Z) dZ - 1\right)\right|\right\} \end{aligned} \quad (22)$$

Differentiating (22) w.r.t each ψ_n^* and setting it to 0 yields,

$$\begin{aligned} & \frac{\partial \mathcal{E}}{\partial \psi_n^*(Z)} = \\ & \mathbb{E}\left\{\left|\int \left(\int K(Z;Z') \phi_n^*(Z') dZ' - \beta_n \psi_n(Z)\right) dZ\right|\right\} = 0 \end{aligned} \quad (23)$$

which is satisfied when,

$$\int K(Z;Z') \phi_n^*(Z') dZ' = \beta_n \psi_n(Z) \quad (24)$$

(24) is indeed the duality constraint. (21) implies (24) indicating duality constraint is necessary for maximizing the sum of eigenvalues. \square

Lemma 1 shows that duality not only reduce the channel distortion for multiplexing but also ensure the convergence for NN, which partially addresses the first challenge stated above.

Lemma 2. *(Equivalence between duality and orthogonality) Given $K(Z;Z')=\sum_{n=1}^{\infty}\sigma_n\psi_n(Z)\phi_n(Z')$, where $\{\psi_n\}$ and $\{\phi_n\}$ are arbitrary normalized bases. If they are dual, then they are orthogonal, and vice versa.*

Proof. The proof is provided in Appendix C. \square

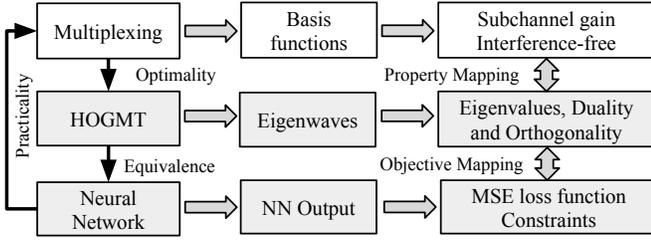


Figure 2: Trilateral relationship of Multiplexing, HOGMT and NN: Multiplexing is the primary goal for communication; HOGMT is mathematical foundation for an optimal solution; NN is a low complexity implementation of HOGMT, which in turn achieves optimal multiplexing in practice.

Lemma 1 and Lemma 2 indicate that both duality and orthogonality constraints guide the convergence of NN, and they can be combined into a single constraint to reduce the complexity of the model. Therefore, two challenges mentioned above are addressed. Figure 2 shows the trilateral relationship among multiplexing, HOGMT and NN. For multiplexing, basis functions are employed to achieve subchannel gains and cancel interference. Eigenwaves, decomposed by HOGMT, are shown to be the optimal bases, where the eigenvalues, duality and orthogonality correspond to subchannels and interference cancellation, respectively. Theorem 1, Lemma 1, and Lemma 2 demonstrate that these properties can be realized by an equivalent neural network with the MMSE loss function and orthogonality constraints, thus providing a practical implementation of multiplexing for doubly selective channels.

C. HNet: NN Implementation of HOGMT

To obtain the unstructured Eigenwaves for doubly selective channels and implement optimal multiplexing in practice, we design an equivalent NN for HOGMT, HNet.

Theorem 2. (HNet) *The neural network with the equality-constrained objective function (25) makes the outputs converge to the optimal bases for doubly selective channels.*

$$\mathcal{L} = \mathcal{J} + \sum_{i=1}^2 \alpha_i \Omega_i \quad (25)$$

where \mathcal{J} is the MSE optimization in (26)

$$\mathcal{J} \triangleq \frac{1}{B} \sum_{b=1}^B \frac{\|\mathbf{K}_b - \sum_{n=1}^N \sigma_{n,b} \Psi_{n,b} \otimes \Phi_{n,b}\|^2}{\|\mathbf{K}_b\|^2} \quad (26)$$

where \mathbf{K}_b is the input kernel. B is the batch size. Ω_1 and Ω_2 are the regularizations for the orthogonality constraints in (9),

$$\Omega_1 \triangleq \frac{1}{B} \sum_{b=1}^B \sum_{n=1}^N \sum_{n'=1, n' \neq n}^N \|\langle \Psi_{n,b}, \Psi_{n',b} \rangle\| \quad (27)$$

$$\Omega_2 \triangleq \frac{1}{B} \sum_{b=1}^B \sum_{n=1}^N \sum_{n'=1, n' \neq n}^N \|\langle \Phi_{n,b}, \Phi_{n',b} \rangle\| \quad (28)$$

where, α_i are penalties for Ω_i . For HNet, $\sigma_1 = \sigma_2 = \sigma$.

Proof. The optimal bases for doubly selective channels must maximize subchannel gains while minimizing channel distortion and interference. According to (10), the subchannel gains are equivalent to eigenvalues. Proposition 1 and Proposition 2

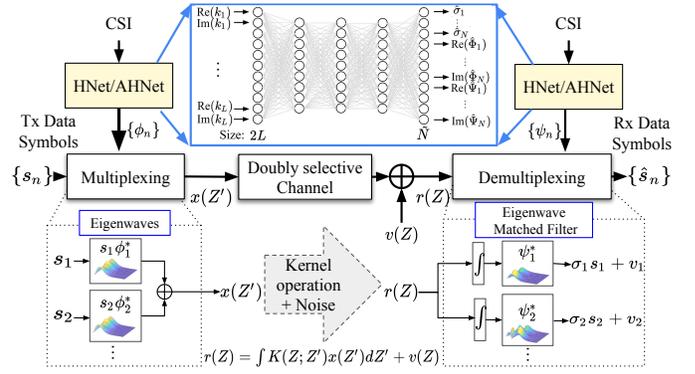


Figure 3: Practical system view of HNet/AHNet-MEM: HNet/AHNet decomposes CSI into dual Eigenwaves $\{\phi_n\}$ and $\{\psi_n\}$, which are utilized for multiplexing at the transmitter and demultiplexing at the receiver, respectively.

show that, with duality and orthogonality as constraints, channel distortion and interference are eliminated. By Lemma 1, these constraints also guide the convergence. Therefore, the optimization of the bases for the doubly selective channel with kernel $K(Z; Z')$ is formulated as,

$$\operatorname{argmax}_{\{\phi_n\}, \{\psi_n\}} \sum_{n=1}^N \mathbb{E}\{\sigma_n^2\} \quad \text{s.t. (9), (10)} \quad (29)$$

By Theorem 1, (15) is equivalent to (26). By Lemma 2, (9) and (10) are simplified to (9). Therefore, (29) is rewritten as,

$$\operatorname{argmax}_{\{\phi_n\}, \{\psi_n\}} \sum_{n=1}^N \mathbb{E}\{\|K(Z; Z') - \hat{K}(Z; Z')\|^2\} \quad \text{s.t. (9)} \quad (30)$$

Introducing Lagrange multipliers α_i as penalties and rewriting (30) in tensor form, the corresponding Lagrangian function is identical to the objective function (25). \square

Theorem 2 proves that HNet answers the three open questions posed in the Section I as follows,

- HNet provides a practical method for obtaining non-predefined and unstructured Eigenwaves for doubly selective channels, with convergence proven by Theorem 2, and low complexity discussed in the Section VI-B.
- Since Eigenwaves decomposed by HOGMT are optimal bases for doubly selective MIMO channels, HNet serves as an optimal training method for doubly selective MIMO channels with respect to multiplexing.
- The MMSE gains are equivalent to subchannel gains, as established by Theorem 1. By using the outputs of HNet as bases, these subchannel gains can be achieved.

Figure 3 presents a practical system view of HNet/AHNet-MEM, offering a high-level understanding of how neural networks (NN) are integrated into our system. Detailed explanations of AHNet, the corresponding architecture and AHNet-MEM can be found in Section VI, Section VI-A and Section VI-C, respectively. The trained HNet/AHNet model is deployed at both the transmitter and receiver. With CSI as input, the NN outputs dual Eigenwaves, which serve as subcarriers for multiplexing at the transmitter and demultiplexing at the receiver.

VI. ADAPTIVE NEURAL NETWORK

In practice, both the required number of Eigenwaves and the channel environment (such as MIMO scales) may vary, leading to different kernel sizes and N , which require changes in the NN architecture and tuning of the Lagrange multiplier (penalty), α . The fundamental reason is that any α , is no longer optimal for the Lagrangian function when its optimal solution changes for a different input. Therefore, training HNet with a fixed penalty for different kernel types does not guarantee the optimal performance, as it requires dynamic adaptation of the Lagrange multipliers during the training. To address this, we propose an adaptive neural network architecture, AHNet, by incorporating the ALM [32] into HNet as in Corollary 1.

Corollary 1. (AHNet) *The adaptive neural network for implementing HOGMT is achieved by modifying the objective function (25) to (31), which does not lose the convergence.*

$$\mathcal{L}' = \mathcal{J} + \sum_{i=1}^2 A_i^\top \Omega'_i + \frac{\mu}{2} \sum_{i=1}^2 \|\Omega'_i\|^2 \quad (31)$$

where, A_i is the vector containing the Lagrange multipliers.

$$A_i \triangleq [\alpha_{i,1}, \alpha_{i,2}, \dots, \alpha_{i,K}]^\top \in \mathbb{R}^{2K} \quad (32)$$

Ω'_1 and Ω'_2 are orthogonality constraints as

$$\Omega'_1 \triangleq \frac{1}{B} \sum_{b=1}^B [\tilde{\Phi}_{1,b}, \tilde{\Phi}_{2,b}, \dots, \tilde{\Phi}_{K,b}]^\top \quad (33)$$

$$\Omega'_2 \triangleq \frac{1}{B} \sum_{b=1}^B [\tilde{\Psi}_{1,b}, \tilde{\Psi}_{2,b}, \dots, \tilde{\Psi}_{K,b}]^\top$$

where, $\tilde{\Phi}_{k,b}$ and $\tilde{\Psi}_{k,b}$ are the inner products of one pair of $(\Phi_{n,b}, \Phi_{n',b})$ and $(\Psi_{n,b}, \Psi_{n',b})$ for $n \neq n'$ respectively. There are $K = N(N-1)$ pairs for such Eigenwave set. During the training, A_i is updated as follows,

$$A_i^{k+1} = A_i^k + \mu^k \Omega'_i \quad (34)$$

where, μ is used to control the learning rate of the model.

Proof. ALM [32] is an adaptive technique to solve an equality-constrained optimization problem. Consider an optimization problem under M constraints

$$\underset{x}{\operatorname{argmin}} F(x) \quad \text{s.t. } c_i(x) = 0, \forall i = 1, 2, \dots, M \quad (35)$$

The corresponding Lagrangian function is,

$$L(x, \alpha) \triangleq F(x) + \sum_{i=1}^M \alpha_i c_i(x) \quad (36)$$

where $\alpha \triangleq [\alpha_1, \alpha_2, \dots, \alpha_M]^\top$ are the Lagrange multipliers. ALM modifies (35) in to the following problem,

$$\underset{x}{\operatorname{argmin}} F(x) + \frac{\mu}{2} \|c(x)\|^2 \quad (37)$$

$$\text{s.t. } c_i(x) = 0, \forall i = 1, 2, \dots, M$$

where, $c(x) \triangleq [c_1(x), c_2(x), \dots, c_M(x)]^\top$. μ is the penalty parameter. Therefore, the Lagrangian function of (37) is,

$$L_\mu(x, \alpha) \triangleq F(x) + \frac{\mu}{2} \|c(x)\|^2 + \sum_{i=1}^M \alpha_i c_i(x) \quad (38)$$

(38) is called the augmented Lagrangian function. It is shown that, both problems (35) and (37) share the same optimal solution x^* and the optimal Lagrange multipliers α^* [32].

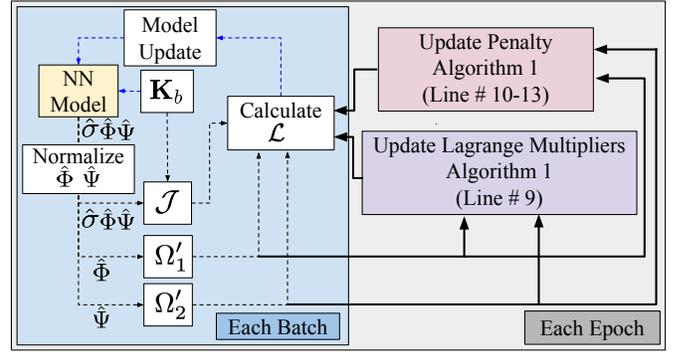


Figure 4: Adaptive Training of AHNet: The outer loop (“per epoch”) updates the Lagrange multipliers A_i and the penalty μ . The inner loop (“per batch”) updates the NN parameters.

Therefore, ALM transforms the constrained optimization in (35) into an unconstrained problem in (38). By following the same deduction steps in the proof of Theorem 2 and substituting (30) in (35), (38) is converted into (31). \square

Remark 1. In HNet, the Lagrange multipliers α_i have to be tuned manually. To reduce the model’s complexity, we set only two multipliers, each corresponding to the orthogonality constraint for one Eigenwave set. In contrast, in AHNet, the Lagrange multipliers are automatically updated according to (34), providing the flexibility to set penalties for each inner product of the Eigenwaves.

The ALM theory requires dynamic update of μ in order to adapt to the current constraint. Otherwise, it leads to either too-small or too-large update rate for A_i . [32] provides a dynamic update criteria for μ . This update step depends on the current conditions of the constraint. Predefined parameters $\beta > 1$ and $\gamma < 1$ ensure that μ increases when the constraints Ω'_1 and Ω'_2 does not decrease over the iterations. Therefore, by coupling the NN training with the update of the Lagrange multipliers in (34) and the parameter μ , AHNet can ensure that the Lagrange multipliers are always optimized towards the optimal NN model, which solves the problem of fixed penalty in HNet mentioned above.

A. AHNet Architecture

To decompose a spatio-temporal channel kernel, $K(u, t; u', t') \in \mathbb{C}^{L_u \times L_t \times L_{u'} \times L_{t'}}$ with $L \triangleq L_u L_t L_{u'} L_{t'}$ elements, the HNet and AHNet is designed with 5 fully connected, feed-forward layers as shown in Figure 3. However, this model requires real inputs and outputs only. As a result, the L elements of the kernel are split into real and imaginary parts at the input. The dimension of each layer is as follows: input and output layers have $2L$ and $\tilde{N} \triangleq N + 2N(L_u L_t + L_{u'} L_{t'})$ nodes, respectively, while the hidden layers have 1024 nodes each. The *LeakyRelu* activation function with a negative slope of 0.01 is used for the hidden linear layers. Output nodes corresponding to σ parameters have soft plus activation, and the remaining output nodes have linear activation.

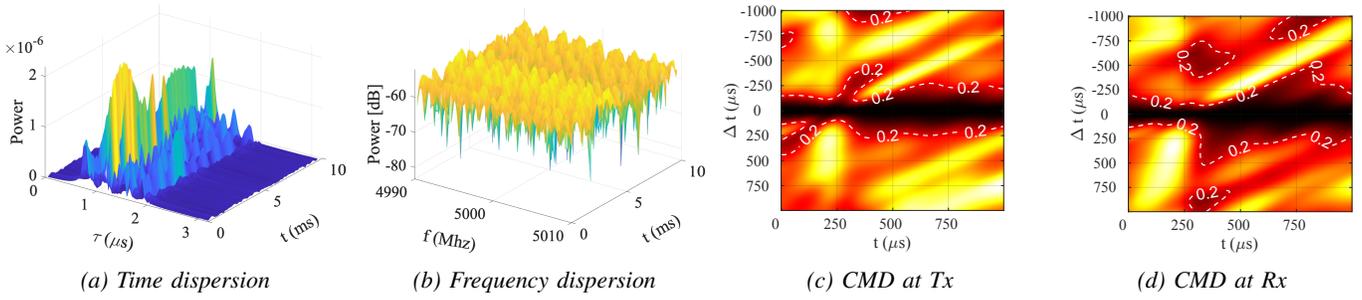


Figure 5: Doubly selective MIMO channel profile in the time, frequency and space domains

Table II: Complexity of eigen-decomposition methods: AHNet and HNet both have lowest complexity. N : Number of eigen-components. N_d : channel dimension, N_L : Number of layers.

Channel Type	Method	Complexity
Spatial (H)	SVD	$\mathcal{O}(\min(N_t N_r^2, N_t^2 N_r))$
	SVD-DNN [13]	$\mathcal{O}(\max(2N^2 N_t N_r, 2N^2 N_t^2, 2N^2 N_r^2))$
Higher Dim. Kernel (K)	HOSVD [33]	$\mathcal{O}(N_d \max(L_u L_t L_{u'}, L_u L_t L_{t'}, L_t L_{u'} L_{t'})^3)$
	HOGMT [12]	$\mathcal{O}(\min(L_u L_t (L_{u'} L_{t'})^2, (L_u L_t)^2 L_{u'} L_{t'}))$
	HNet	$\mathcal{O}(2N N_L L_u L_t L_{u'} L_{t'})$
	AHNet	$\mathcal{O}(2N N_L L_u L_t L_{u'} L_{t'})$

B. Complexity Analysis

The computational complexity for eigen-decomposition is the crucial challenging for practical MEM. Table II provides the complexity comparison for HNet and AHNet with SoTA. Both SVD and DNN-SVD are designed to decompose 2-D matrices only. While High-Order SVD (HOSVD), HOGMT and proposed methods decompose multi-dimensional tensors. The relationship between SVD, HOSVD and HOGMT can be found in Lemma 3 of [12]. Since both HNet and AHNet use fully connected architectures (Section VII), their time complexities are the same, which depends on the number of layers (N_L), the size of the input ($2L_u L_t L_{u'} L_{t'}$ for a complex-valued NN) and the size of the output (N). It is clear that proposed NNs have less complexity than HOSVD. Meanwhile, for a fixed N and N_L , the complexity of both HNet and AHNet increase at a much slower rate than HOGMT with increasing size of the input tensor.

C. AHNet-MEM: NN-based MEM

Theorem 2 and Corollary 1 show that with CSI as inputs, the outputs of HNet and AHNet converge to the optimal bases, Eigenwaves. This provides the practical implementation of MEM for doubly selective channels. Figure 3 shows the practical system of AHNet-MEM. Decomposing the CSI into Eigenwaves by HNet/AHNet, the transmitted signal $x(Z')$ is obtained by multiplexing data symbols $\{s_n\}$ onto $\{\phi_n^*\}$.

$$x(Z') = \sum_n^N s_n \phi_n^*(Z') \quad (39)$$

Transmitting it over the channel, the received signal is

$$r(Z) = \int K(Z; Z') x(Z') dZ' + v(Z)$$

$$\begin{aligned} &= \int K(Z; Z') \sum_n^N s_n \phi_n^*(Z') dZ' + v(Z) \\ &= \sum_{n=1}^N \underbrace{\sigma_n s_n \psi_n(Z)}_{\text{by (10)}} + v(Z) \end{aligned} \quad (40)$$

Estimate symbols $\{\hat{s}_n\}$ are obtained by demultiplexing the received signal $r(Z)$ using "Matched Filter" with dual Eigenwaves $\{\psi_n^*\}$ at the receiver,

$$\begin{aligned} \hat{s}_n &= \int r(Z) \psi_n^*(Z) dZ = \int \sum_{n'=1}^N \sigma_{n'} s_{n'} \psi_{n'}(Z) \psi_n^*(Z) dZ \\ &+ \int v(Z) \psi_n^*(Z) dZ = \sum_{n'=1}^N \sigma_{n'} s_{n'} \underbrace{\delta_{n'n}}_{\text{by (9)}} + v_n = \sigma_n s_n + v_n \end{aligned} \quad (41)$$

(41) shows that there is no interference across demodulated symbols. During the transmission, $\{s_n\}$ can be seen as transmitted through independent subchannels in the eigen domain with subchannel gains $\{\sigma_n\}$. The error probability for M-QAM modulated symbols is given as [34],

$$Pr \approx 4Q \left(\sqrt{3\sigma_n^2 / \zeta_n (M-1)} \right) \quad (42)$$

where ζ_n is power of v_n . (42) indicates that, with a fixed ζ_n , low σ_n results in high BER. Since eigenvalues are descending ordered, selecting an appropriate number, N , of Eigenwaves as bases is crucial to minimize the BER.

VII. AHNET IMPLEMENTATION AND TRAINING

The proposed method is validated in two scenarios: Case-1: 4-D doubly selective MIMO channels. Since SoTA are proposed for 2-D channels, we compare them in the degeneration case: Case-2: 2-D doubly selective channels. To avoid redundancy, we focus on the implementation details for the Case-2, as Case-1 is a special cases of Case-2.

Data Generation: The doubly selective channel is generated using 3GPP 38.901 UMa NLOS scenario built on QuaDriga in Matlab. The channel parameters and the layout of the base station (BS) and the user equipment (UE) are shown in Table III. Figure 5a and 5b show the channel profiles in time-delay domain and time-frequency domain for the first user, respectively. The time-varying dispersion in both delay and frequency domain indicates the doubly selective channels. Correlation Matrix Distance (CMD) is a measure of the time-varying correlation in the space domain [35]. Figure 5c and 5d show the CMD at the transmitter and the receiver, respectively,

Table III: Parameters of the doubly selective MIMO channel

Parameter	Value
Channel model	3GPP 38.901 UMa NLOS [36]
Array type	BS: 3GPP 3-D [37]; UE: Vehicular [38]
BS antenna	Height $h_b=10$ m; Number $N_u=10$
UE antenna	Height $h_u=1.5$ m; Number $N_u=4$
UE speed	$v \in [100, 150]$ km/h
Bandwidth	Bw = 20 MHz; Center frequency: $f_c=5$ GHz
Channel size	Each segments: $\mathbf{H}(t, \tau) \in \mathbb{C}^{4 \times 64 \times 4 \times 64}$

where they are presented over time instead of distance because the varying mobility profiles of the transceiver lead to different distances over time. Noticing that normalizing a channel does not change its structure, distribution and characteristics, we normalize the above channel and generate 10000 samples by $\hat{\mathbf{H}}(t, \tau) = \mathbf{H}(t, \tau) + \Delta \mathbf{H}(t, \tau)$ with $\Delta \mathbf{H}(t, \tau) \sim \mathcal{N}(0, \Delta \sigma^2)$, where $\Delta \sigma^2$ represents the variation of samples. Since our methods are validated across multiple scenarios with varying channel sizes and distributions, a unified metric, the expected Error Vector Magnitude (EVM), η , is used to quantify the relative difference between samples and the centered channel,

$$\eta = \mathbb{E} \left\{ \frac{\|\hat{\mathbf{H}}(t, \tau) - \mathbf{H}(t, \tau)\|}{\|\mathbf{H}(t, \tau)\|} \right\} \quad (43)$$

For MEM, we evaluate the performance under both perfect and imperfect CSI scenarios. In the perfect CSI case, the transmission channel matches the input channel sample exactly, where η indicates the divergence of samples. In the imperfect CSI scenario, a new sample is generated as the transmission channel, and thus η also indicates the level of CSI error.

Training: For the generated samples, 80% are used for training and 20% are used for validation. The mini-batch training is adopted to train both HNet and AHNet. The training data set is split into 32 mini-batches and trained using Adam optimizer with a learning rate of 1×10^{-5} . Figure 4 shows how various training-related parameters are updated in every *batch* and *epoch* while training AHNet. Figure 4 shows that there are some adaptive parameters involved in the training, such as A_1 , A_2 and μ , that depend on Ω'_1 and Ω'_2 . These parameters are updated during each epoch, relative to the Ω'_1 and Ω'_2 of the last mini-batch of the previous epoch using (34) for A_1 and A_2 and Algorithm 1 lines 10-13 for μ with $\gamma = 0.25$ and $\beta = 1.01$ and initial value for μ as 1. The initial values of A_i are set to 0 as it provides flexibility in choosing the rest of the parameters. Algorithm 1 is used for training AHNet with an initialization of the inputs, A_1 , A_2 , μ and γ . Lines 3-8 are steps for kernel processing and parameter calculation in each batch according to (31)-(33). Line 9 updates the Lagrange multipliers according to the constraints and μ , where μ is further updated in lines 10-13.

VIII. EVALUATION AND RESULTS

The total loss is a general evaluation metric for NNs, which depends on the sum of all eigenvalues for eigen-decomposition [28]. While in the communication systems, duality and orthogonality are the most critical metrics for interference cancellation. However, any NN-based method

Algorithm 1 AHNet Training

```

1: Inputs  $A_1^{[0]}, A_2^{[0]}, \mu^{[0]}, \gamma$ ;
2: for  $i \leftarrow 0$  to training epochs do
3:   for  $mini\ batches \leftarrow 0$  to data size/batch size do
4:      $\sigma_{n,b}, \Phi_{n,b}, \Psi_{n,b} = \text{AHNet\_Model}(\mathbf{K}_b)$ ;
5:     Normalize  $\Phi_{n,b}$  and  $\Psi_{n,b}$ ;
6:     Calculate  $\mathcal{J}$  according to (26);
7:     Calculate  $\Omega'_1$  and  $\Omega'_2$  according to (33);
8:   end for
9:    $A_1^{[i+1]} \leftarrow A_1^{[i]} + \mu \Omega'_1$ ;  $A_2^{[i+1]} \leftarrow A_2^{[i]} + \mu \Omega'_2$ ;
10:  if  $\|\Omega'_1\|^{[i]} + \|\Omega'_2\|^{[i]} > \gamma (\|\Omega'_1\|^{[i-1]} + \|\Omega'_2\|^{[i-1]})$  then
11:     $\mu^{[i+1]} \leftarrow \beta \mu^{[i]}$ ;
12:  else  $\mu^{[i+1]} \leftarrow \mu^{[i]}$ ;
13:  end if
14: end for

```

cannot achieve strict duality and orthogonality as the hard decision will limit the freedom of the output space and lead to over-fitting. Therefore, we define metrics, *soft duality* and *soft orthogonality* as follows,

Definition 1. According to (10), *soft duality* of two Eigenwave sets $\{\Psi\}$ and $\{\Phi\}$ with respect to kernel \mathbf{K}_b is defined as

$$d \triangleq \frac{1}{N} \sum_{n=1}^N \|\langle \mathbf{K}_b, \Phi_n \rangle - \sigma_n \Psi_n\|^2 \quad (44)$$

where $d=0$ indicates that $\{\Psi\}$ and $\{\Phi\}$ are strict dual.

Definition 2. According to (9), *soft orthogonality* of the Eigenwave set $\{\Psi\}$ is defined as

$$O(\Psi) \triangleq \frac{1}{N(N-1)} \sum_{n=1}^N \sum_{n'=1, n' \neq n}^N \|\langle \Psi_n, \Psi_{n'} \rangle\| \quad (45)$$

where $O(\Psi)=0$ indicates that the Eigenwaves obtained from the NN, $\{\Psi\}$ are strictly orthogonal. Similarly, $O(\Phi)$ denotes the soft orthogonality for Eigenwaves $\{\Phi\}$.

A. Case-1: 4-D Doubly Selective MIMO Channels

1) Performance of Neural Networks: AHNet and HNet

Figure 6 compares the performance of AHNet and HNet for 4D channels. Notably, AHNet does not require parameter tuning, while HNet does. With different penalties, $\alpha=0.1$ and $\alpha=0.01$, HNet shows varying convergence patterns, and both settings converge more slowly than AHNet, as shown in Figures 6a-6d. HNet achieves a small gain in $O\{\hat{\Phi}\}$ and $O\{\hat{\Psi}\}$, while AHNet show a better performance in both \mathcal{J} and duality. Given that AHNet is more adaptive and stable—avoiding parameter tuning and converging faster than HNet—we consider AHNet a more practical and effective architecture than HNet.

We further evaluate the adaptivity of AHNet in Figures 7a-7d, where the length of the time dimension is fixed at $L_t=L'_t=64$ and the length of the space dimension is set to $L_u=L'_u=2, 4$ and 6, corresponding to $N=128, 256$ and 384, respectively. For different N , MSE converges to the similar value $\mathcal{J} \approx 4 \times 10^{-2}$, indicating the loss remains consistent across varying N . With larger N , AHNet achieves lower d , $O\{\hat{\Phi}\}$ and $O\{\hat{\Psi}\}$. This occurs because duality and soft orthogonality

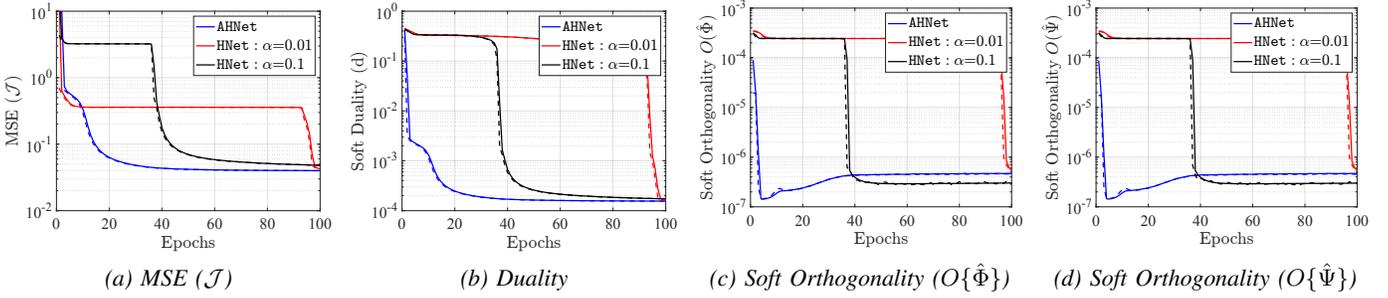


Figure 6: Performance comparison between AHNet and HNet for a 4-D channel kernel: (a)-(d) show the MSE \mathcal{J} , duality, soft orthogonality $O(\hat{\Phi})$ and $O(\hat{\Psi})$ for $N=256$ ($L_u=L'_u=4$ and $L_t=L'_t=64$) at $\eta=0.2$, respectively.

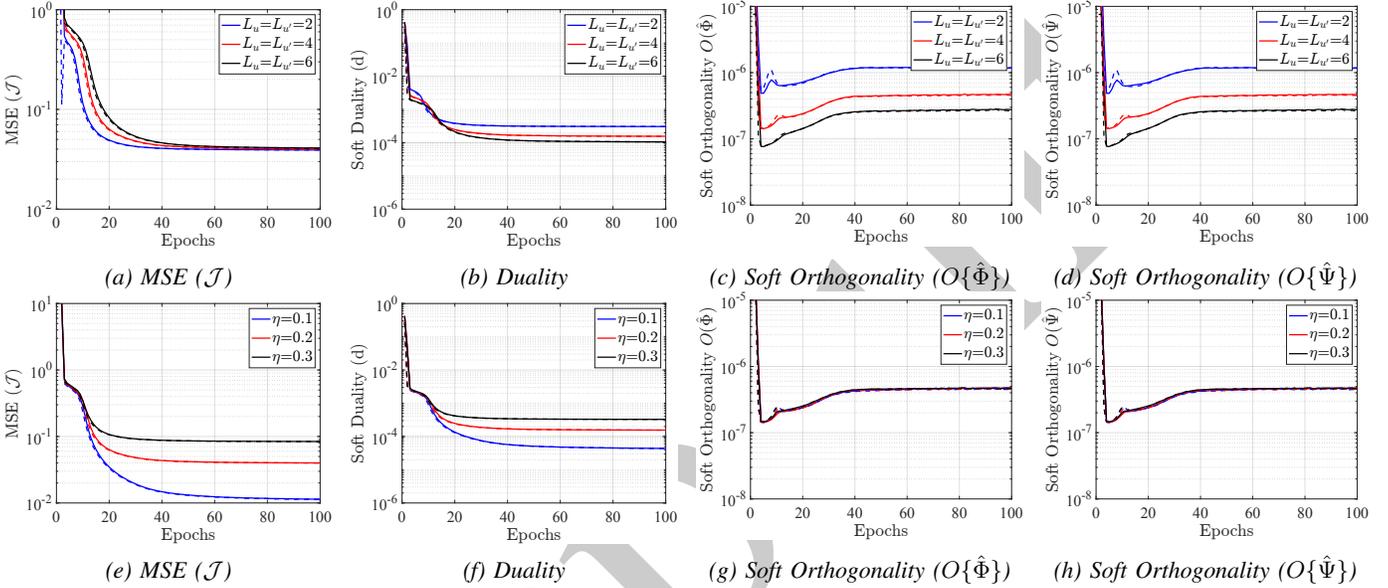


Figure 7: AHNet for 4-D channel kernel: (a)-(d) MSE \mathcal{J} , duality, soft orthogonality $O(\hat{\Phi})$ and $O(\hat{\Psi})$ for $N=128, 256$ and 384 ($L_u=L'_u=2, 4$ and 6 with $L_t=L'_t=64$) at $\eta=0.2$. (e)-(h) Performance for $\eta=0.1, 0.2$ and 0.3 with $N=256$, respectively.

are calculated by averaging the non-dual and non-orthogonal errors, respectively, which are inversely proportional to N . Nonetheless, even with the smallest $N=128$, AHNet can still achieve duality below 4×10^{-4} and the soft orthogonality approximating at only 2×10^{-6} .

Figures 7e-7h evaluate the robustness of AHNet for $N=256$ at $\eta=0.1, 0.2$ and 0.3 , respectively. With increasing η , both MSE and the duality rise. Specifically, in the worst case $\eta=0.3$, we have $\mathcal{J} < 10^{-1}$ and $d < 4 \times 10^{-4}$. In contrast, the soft orthogonality is consistent across different η , converging to approximately 5×10^{-7} . Such low soft orthogonality will significantly reduce the interference for multiplexing. The performance of decomposed Eigenwaves will be further evaluated by multiplexing in Section VIII-A2.

2) Performance of Multiplexing: AHNet-MEM

MEM with perfect CSI is considered as the ideal modulation with respect to interference cancellation, since data symbols are transmitting through orthogonal subchannels in the eigen domain, where eigenvalue affects the error rate as shown in (42). In this subsection, N denotes the number of top Eigenwaves are employed for multiplexing. Figure 8a com-

pare the BER of AHNet-MEM and MEM with $N=160, 192, 224$ and 256 for the perfect CSI, respectively, where 4-D channel is $\mathbf{H}(t, \tau) \in \mathbb{C}^{4 \times 64 \times 4 \times 64}$ as given in Table III. For $N=192, 224$ and 256 , the BER of both AHNet-MEM and MEM is higher than 10^{-2} , as eigenvalues of last few eigenwaves are vary small for this complex channel. When $N=160$, AHNet-MEM achieves BER at around 10^{-3} and the gap between it and the ideal MEM is less than 1 order of magnitude. Figure 8b shows the throughput of AHNet-MEM and MEM for different N , where throughput is calculated as the total transmitted bits minus the error bits. With larger N , the throughput is higher. This is because more data symbols are multiplexed for each transmission. Therefore, there is a clear trade-off between BER and the throughput with respect to N . Figure 8c and 8d compares their performance for imperfect CSI with a fixed $N=192$ and $\eta=0.1, 0.2$ and 0.3 , respectively. In this case, the performance of MEM degrades over η as the Eigenwaves decomposed from the imperfect CSI are not strict orthogonal over the transmission channel. However, AHNet-MEM shows strong robustness to the CSI error and achieves better performance than MEM across all the η . This is because, by using

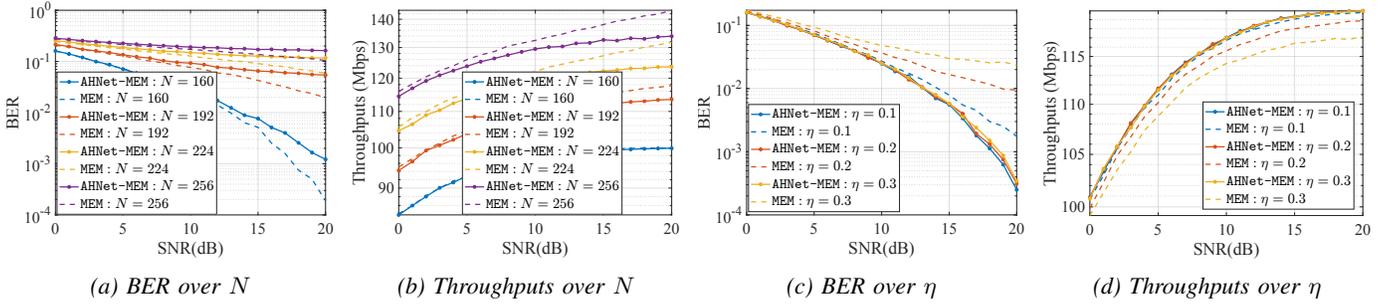


Figure 8: AHNet-MEM and MEM for 4-D doubly selective MIMO channels: (a)-(b) BER and throughput for the perfect CSI with $N=160, 192, 224$ and 256 ; (c)-(d) BER and throughput for imperfect CSI with fixed $N=192$ at $\eta=0.1, 0.2$ and 0.3 .

imperfect CSI as inputs, AHNet make the outputs converge to optimal Eigenwaves across the CSI errors in MMSE sense during the training. There is a trade-off between optimality and robustness for MEM and AHNet-MEM. Considering perfect CSI is unattainable in practice and that MEM has significantly higher complexity, AHNet-MEM presents a more practical approach.

B. Case-2: 2-D Doubly Selective Channels

1) Performance of Neural Networks: AHNet and HNet

Figure 9 compares the training performance among AHNet, HNet and SVD-DNN for 2-D doubly selective channels with a rank of $N=64$ and a relative CSI error of $\eta=0.2$. The penalties for orthogonality constraints defined by (25) are both set to $\alpha=0.1$. Figure 9a shows the training and testing losses, where both AHNet and HNet reach $\mathcal{J} < 5 \times 10^{-2}$. The SVD-DNN achieves the least training loss with $\mathcal{J} \approx 4 \times 10^{-3}$ but the test loss diverges after 20 epochs, stabilizing at a similar order of magnitude as AHNet and HNet. Figure 9b compares the duality, with both AHNet and HNet achieving $d \approx 6 \times 10^{-4}$, while SVD-DNN shows divergence in duality during training, approaching $d \approx 5 \times 10^{-2}$. Consequently, both AHNet and HNet achieve 2 orders of magnitude improvement over SVD-DNN. Figure 9c and 9d show the soft orthogonality of the decomposed Eigenwaves, where $O\{\hat{\Phi}\}$ and $O\{\hat{\Psi}\}$ show a similar trends as their orthogonality influences each other due to the duality. Both HNet and AHNet achieve approximate 2 orders of magnitude gains than SVD-DNN. While HNet performs slightly better than AHNet, it converges about 20 epochs more slowly. Overall, SVD-DNN achieves a slight improvement in testing loss at the expense of: 1) both duality and orthogonality, which are key properties for multiplexing as discussed in Section IV-B; 2) higher complexity than both HNet and AHNet as shown in Table II, and 3) adaptivity to 4-D tensors. Further, since SVD-DNN is not explainable, its performance is not guaranteed.

2) Performance of Multiplexing: AHNet-MEM

In this case, we compare the performance of AHNet-MEM with MEM and OTFS. For a fair comparison, OTFS is equipped with the Time-Frequency Single Tap (TFST) [39] detector and Zero-Padded Maximal Ratio Combining (ZP-MRC) [40] detector, respectively, which also leverage the perfect CSI at the receiver. The size of delay-Doppler grid for

OTFS is 4×16 and the rank of the channel is 64. Figure 10a and 10b show their performance under perfect CSI. Even with $N=64$, AHNet-MEM and MEM achieve lower BER than OTFS. Because OTFS subcarriers can not maintain orthogonal in doubly selective channels. AHNet-MEM and MEM with $N=64$ achieve the highest throughput. With $N=60$, their throughput is similar as OTFS with TFST, as 4 Eigenwaves are not utilized. OTFS with ZP-MRC has the lowest throughput due to the requirement of zero padding, set to $1/8$ length of symbols. Figure 10c and 10d show their performance under imperfect CSI with a fixed $N=60$ and $\eta=0.1, 0.2$ and 0.3 . Due to space constraints, we only show the performance of OTFS at $\eta=0.1$. Even with $\eta=0.3$, AHNet-MEM achieves the lower BER and higher throughput than both MEM and OTFS. Overall, AHNet-MEM shows a near-ideal performance under perfect CSI and a robust performance under imperfect CSI.

IX. DISCUSSION

Conventional modulation methods require CSI only at the receiver side for detection. In contrast, AHNet-MEM requires CSI at both the transmitter and receiver, which is a significant cost. However, since precoding techniques also require CSI at the transmitter for spatial interference cancellation, we believe this cost does not impose an additional burden from a system-level perspective. Meanwhile, from a results perspective, regardless of channel conditions or the number of DoFs involved, AHNet-MEM converts all channel interference into errors arising from imperfect CSI and neural network inaccuracies, both of which are thoroughly evaluated in this paper. As an NN-based approach, performance is inevitably influenced by the distribution of the training and validation data. This implies that for non-stationary channels, the performance cannot be guaranteed. In the future, we plan to further explore equalization and multiple access techniques based on AHNet-MEM in the eigen domain. Further, based on HOGMT, AHNet can decompose a general multi-dimensional asymmetrical kernel into Eigenwaves, where the kernel is not limited to be the channel. This flexibility and generality enables AHNet to support a broader range of applications in other domains that require eigen decomposition.

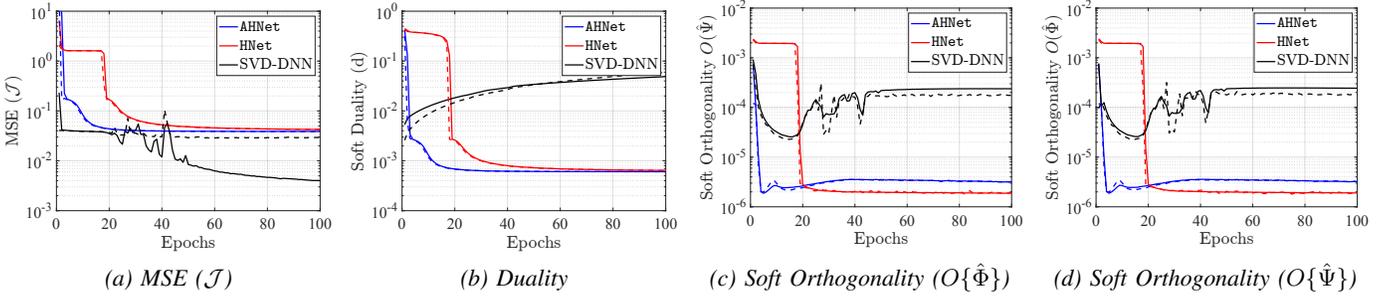


Figure 9: Performance comparison among AHNet, HNet and SVD-DNN for a 2-D channel kernel: (a)-(d) show the MSE \mathcal{J} , duality, soft orthogonality $O\{\hat{\Phi}\}$ and $O\{\hat{\Psi}\}$ for $N=64$ at $\eta=0.2$, respectively.

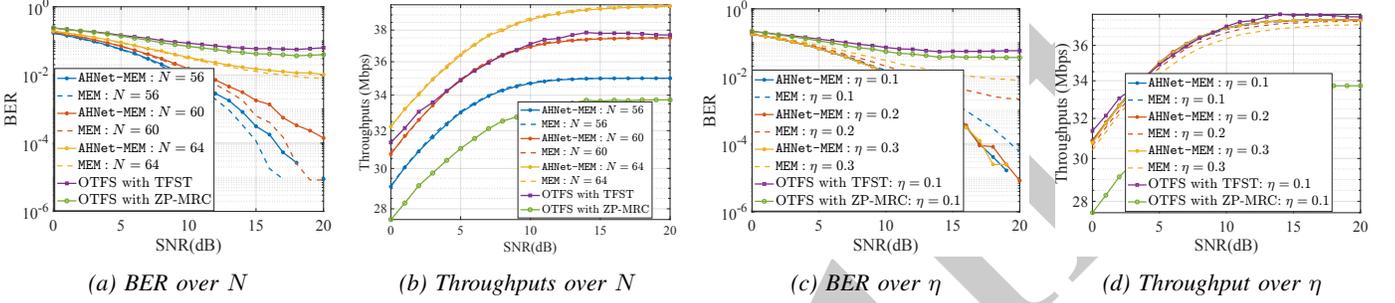


Figure 10: Comparing AHNet-MEM, MEM and OTFS for 2-D doubly selective channels: (a)-(b) BER and throughput for perfect CSI with $N=56, 60$ and 64 ; (c)-(d) BER and throughput for imperfect CSI with fixed $N=60$ at $\eta=0.1, 0.2$ and 0.3 , respectively.

X. CONCLUSION

In this paper, we show that Eigenwaves are the optimal bases for doubly selective channels due to their inherent properties of duality and orthogonality, with eigenvalues translate to subchannel gains. To address the open problem of obtaining Eigenwaves with practical complexity, we establish a trilateral relationship among multiplexing, HOGMT, and NN, which forms the basis of an explainable NN, HNet. The loss function and regularization terms are derived from first principles to ensure convergence, generality, and low computational complexity. The HNet is enhanced to AHNet, with the ALM framework to avoid parameter tuning, with both demonstrating lower complexity than conventional eigen decomposition methods. This adaptive NN implements AHNet-MEM system, which minimizes inference for doubly selective channels. We validate our theory by examining NN convergence and BER in both 2-D doubly selective channels and 4-D doubly selective MIMO channels. Additionally, we test the robustness of AHNet-MEM with imperfect CSI, achieving nearly 2 orders of magnitude improvement in BER than OTFS at 20 dB SNR.

APPENDIX A

PROOF OF PROPOSITION 1

Proof. Let $a=\sigma_n$, $g_{tx}(Z')=\phi_n^*(t,f)$ and $g_{rx}(Z)=\psi_n(Z)$. Therefore, we have,

$$\begin{aligned} \epsilon &= \mathbb{E} \left\{ \left\| \mathcal{H}(\phi_n(Z')) - \sigma_n \psi_n(Z) \right\|^2 \right\} \\ &= \mathbb{E} \left\{ \left\| \underbrace{\iint K(Z;Z') \phi_n^*(Z') dZ'}_{\text{Duality}} - \sigma_n \psi_n(Z) \right\|^2 \right\} \\ &= \mathbb{E} \left\{ \left\| \sigma_n \psi_n(Z) - \sigma_n \psi_n(Z) \right\|^2 \right\} = 0 \quad \square \end{aligned}$$

APPENDIX B

PROOF OF PROPOSITION 2

Proof. The second term in (14) is

$$\begin{aligned} & \sum_{n' \neq n}^N s_{n'} \underbrace{\langle \mathcal{H}(\phi_n(Z')), \psi_{n'}(Z) \rangle}_{\text{by (10)}} = \sum_{n' \neq n}^N s_{n'} \langle \sigma_n \psi_n(Z), \psi_{n'}(Z) \rangle \\ &= \sum_{n' \neq n}^N s_{n'} \sigma_n \underbrace{\langle \psi_n(Z), \psi_{n'}(Z) \rangle}_{\text{by (9)}} = 0 \quad \square \end{aligned}$$

APPENDIX C

PROOF OF LEMMA 2

Proof. By duality, $\langle \psi_n, \psi_{n'} \rangle$ can be rewritten as in (46), where the residual terms must be 0. Careful observation reveals that $\langle \phi_n, \phi_{n'} \rangle = 0$ for any $n \neq n'$, confirming that ϕ_n are orthogonal. Interchanging ϕ_n and ψ_n yields the same conclusion. Through inverse deduction, it follows that orthogonality implies duality. Thus, duality and orthogonality are equivalent constraints. \square

REFERENCES

- [1] K. Liu, T. Kadous, and A. Sayeed, "Orthogonal time-frequency signaling over doubly dispersive channels," *IEEE Transactions on Information Theory*, vol. 50, no. 11, pp. 2583–2603, 2004.
- [2] A. Stamoulis, S. N. Diggavi, and N. Al-Dahir, "Inter-carrier interference in MIMO OFDM," *IEEE Transactions on signal processing*, vol. 50, no. 10, pp. 2451–2464, 2002.
- [3] R. Hadani, S. Rakib, M. Tsatsanis, A. Monk, A. J. Goldsmith, A. F. Molisch, and R. Calderbank, "Orthogonal time frequency space modulation," in *2017 IEEE Wireless Communications and Networking Conference (WCNC)*. IEEE, 2017, pp. 1–6.
- [4] R. Hadani, S. Rakib, S. Kons, M. Tsatsanis, A. Monk, C. Ibars, J. Delfeld, Y. Hebron, A. J. Goldsmith, A. F. Molisch, and A. R. Calderbank, "Orthogonal time frequency space modulation," *CoRR*, vol. abs/1808.00519, 2018. [Online]. Available: <http://arxiv.org/abs/1808.00519>

$$\begin{aligned}
\langle \psi_n, \psi_{n'} \rangle &= \int \psi_n(Z) \psi_{n'}^*(Z) dZ = \int \left(\frac{1}{\sigma_n} \int K(Z; Z') \phi_n^*(Z') dZ' \right) \left(\frac{1}{\sigma_{n'}} \int K^*(Z; Z') \phi_{n'}(Z') dZ' \right) dZ \\
&= \int \left(\frac{1}{\sigma_n} \int \sum_{i=1}^{\infty} \sigma_i \psi_i(Z) \phi_i(Z') \phi_n^*(Z') dZ' \right) \left(\frac{1}{\sigma_{n'}} \int \sum_{j=1}^{\infty} \sigma_j \psi_j^*(Z) \phi_j^*(Z') \phi_{n'}(Z') dZ' \right) dZ \\
&= \int \left(\frac{1}{\sigma_n} \int \sigma_n \psi_n(Z) \phi_n(Z') \phi_n^*(Z') + \sum_{i \neq n} \sigma_i \psi_i(Z) \phi_i(Z') \phi_n^*(Z') dZ' \right) \left(\frac{1}{\sigma_{n'}} \int \sigma_{n'} \psi_{n'}^*(Z) \phi_{n'}^*(Z') + \sum_{j \neq n'} \sigma_j \psi_j^*(Z) \phi_j^*(Z') \phi_{n'}(Z') dZ' \right) dZ \\
&= \int \left(\psi_n(Z) + \sum_{i \neq n} \frac{\sigma_i}{\sigma_n} \psi_i(Z) \int \phi_i(Z') \phi_n^*(Z') dZ' \right) \left(\psi_{n'}(Z) + \sum_{j \neq n'} \frac{\sigma_j}{\sigma_{n'}} \psi_j^*(Z) \int \phi_j^*(Z') \phi_{n'}(Z') dZ' \right) dZ \\
&= \underbrace{\langle \psi_n, \psi_{n'} \rangle + \sum_{i \neq n} \frac{\sigma_i}{\sigma_n} \langle \psi_i, \psi_{n'} \rangle \langle \phi_i, \phi_n \rangle + \sum_{j \neq n'} \frac{\sigma_j}{\sigma_{n'}} \langle \psi_j, \psi_n \rangle \langle \phi_j, \phi_{n'} \rangle + \sum_{i \neq n} \sum_{j \neq n'} \frac{\sigma_i \sigma_j}{\sigma_n \sigma_{n'}} \langle \psi_i, \psi_j \rangle \langle \phi_i, \phi_n \rangle \langle \phi_j, \phi_{n'} \rangle}_{=0} \quad (46)
\end{aligned}$$

- [5] P. Raviteja, K. T. Phan, Y. Hong, and E. Viterbo, "Interference Cancellation and Iterative Detection for Orthogonal Time Frequency Space Modulation," *IEEE Transactions on Wireless Communications*, vol. 17, no. 10, pp. 6501–6515, 2018.
- [6] A. Tusha and H. Arslan, "Low Complex Inter-Doppler Interference Mitigation for OTFS Systems via Global Receiver Windowing," *IEEE Transactions on Vehicular Technology*, 2023.
- [7] S. S. Das and R. Prasad, *Orthogonal Time Frequency Space Modulation: OTFS a Waveform for 6G*. CRC Press, 2022.
- [8] Z. Zou, I. Amarasekara, and A. Dutta, "Learning to Decompose Asymmetric Channel Kernels for Generalized Eigenwave Multiplexing," in *IEEE INFOCOM 2024 - IEEE Conference on Computer Communications*, 2024, pp. 1341–1350.
- [9] P. Bello, "Characterization of Randomly Time-Variant Linear Channels," *IEEE Transactions on Communications Systems*, vol. 11, no. 4, pp. 360–393, 1963.
- [10] L. A. Zadeh, "Frequency analysis of variable networks," *Proceedings of the IRE*, vol. 38, no. 3, pp. 291–299, 1950.
- [11] F. Hlawatsch and G. Matz, *Wireless Communications Over Rapidly Time-Varying Channels*, 1st ed. USA: Academic Press, Inc., 2011.
- [12] Z. Zou, M. Careem, A. Dutta, and N. Thawdar, "Joint Spatio-Temporal Precoding for Practical Non-Stationary Wireless Channels," *IEEE Transactions on Communications*, vol. 71, no. 4, pp. 2396–2409, 2023.
- [13] T. Peken, S. Adiga, R. Tandon, and T. Bose, "Deep learning for SVD and hybrid beamforming," *IEEE Transactions on Wireless Communications*, vol. 19, no. 10, pp. 6621–6642, 2020.
- [14] J. J. Jaime-Rodríguez, C. A. Gómez-Vega, C. A. Gutiérrez, J. M. Luna-Rivera, D. U. Campos-Delgado, and R. Velázquez, "A non-wssus channel simulator for v2x communication systems," *Electronics*, vol. 9, no. 8, 2020. [Online]. Available: <https://www.mdpi.com/2079-9292/9/8/1190>
- [15] A. A. Khuwaja, Y. Chen, N. Zhao, M.-S. Alouini, and P. Dobbins, "A survey of channel modeling for uav communications," 2018.
- [16] G. Matz and F. Hlawatsch, "Time-varying communication channels: Fundamentals, recent developments, and open problems," in *2006 14th European Signal Processing Conference*, 2006, pp. 1–5.
- [17] I. Amarasekara, Z. Zou, and A. Dutta, "Adaptive Neural Network for Eigen-Decomposition of Multi-dimensional Channel Kernels," in *IEEE Vehicular Technology Conference*. IEEE, 2024.
- [18] 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 Transactions on Antennas and Propagation*, vol. 62, no. 6, pp. 3242–3256, 2014.
- [19] W. Kozek and A. Molisch, "Nonorthogonal pulseshapes for multicarrier communications in doubly dispersive channels," *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 8, pp. 1579–1589, 1998.
- [20] T. Strohmer, "Approximation of dual gabor frames, window decay, and wireless communications," *Applied and Computational Harmonic Analysis*, vol. 11, no. 2, pp. 243–262, 2001.
- [21] W. Kozek, "Adaptation of weyl-heisenberg frames to underspread environments," in *Gabor Analysis and Algorithms: Theory and Applications*. Springer, 1998, pp. 323–352.
- [22] R. Haas and J.-C. Belfiore, "A time-frequency well-localized pulse for multiple carrier transmission," *Wireless personal communications*, vol. 5, pp. 1–18, 1997.
- [23] H. Yin and S. Alamouti, "Ofdma: A broadband wireless access technology," in *2006 IEEE sarnoff symposium*. IEEE, 2006, pp. 1–4.
- [24] S. Chen, B. Ren, Q. Gao, S. Kang, S. Sun, and K. Niu, "Pattern division multiple access—a novel nonorthogonal multiple access for fifth-generation radio networks," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 4, pp. 3185–3196, 2016.
- [25] J. Brown, Jr, "Mean square truncation error in series expansions of random functions," *Journal of the Society for Industrial and Applied Mathematics*, vol. 8, no. 1, pp. 28–32, 1960.
- [26] G. W. Stewart, *Matrix Algorithms: Volume II: Eigensystems*. SIAM, 2001.
- [27] C. Williams and M. Seeger, "Using the Nyström Method to Speed Up Kernel Machines," in *Advances in Neural Information Processing Systems*, T. Leen, T. Dietterich, and V. Tresp, Eds., vol. 13. MIT Press, 2000. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2000/file/19de10adbaa1b2ee13f77f679fa1483a-Paper.pdf
- [28] Z. Deng, J. Shi, and J. Zhu, "Neuralef: Deconstructing kernels by deep neural networks," in *International Conference on Machine Learning*. PMLR, 2022, pp. 4976–4992.
- [29] A. Felix, S. Cammerer, S. Dörner, J. Hoydis, and S. Ten Brink, "OFDM-autoencoder for end-to-end learning of communications systems," in *2018 IEEE 19th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*. IEEE, 2018, pp. 1–5.
- [30] X. Gao, S. Jin, C.-K. Wen, and G. Y. Li, "ComNet: Combination of deep learning and expert knowledge in OFDM receivers," *IEEE Communications Letters*, vol. 22, no. 12, pp. 2627–2630, 2018.
- [31] G. W. Stewart, "On the early history of the singular value decomposition," *SIAM review*, vol. 35, no. 4, pp. 551–566, 1993.
- [32] D. P. Bertsekas, *Nonlinear Programming*. Athena Scientific, 2016.
- [33] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM review*, vol. 51, no. 3, pp. 455–500, 2009.
- [34] M. K. Simon and M.-S. Alouini, *Digital communication over fading channels*. New York: Wiley, 2001.
- [35] O. Renaudin, V.-M. Kolmonen, P. Vainikainen, and C. Oestges, "Non-stationary narrowband mimo inter-vehicle channel characterization in the 5-ghz band," *IEEE Transactions on Vehicular Technology*, vol. 59, no. 4, pp. 2007–2015, 2010.
- [36] 3GPP TR 38.901 v16.1.0, "5G; Study on channel model for frequencies from 0.5 to 100 GHz," 3rd Generation Partnership Project (3GPP), Tech. Rep., 11 2020.
- [37] 3GPP TR 37.885 v15.1.0, "Technical specification group radio access network; study on evaluation methodology of new vehicle-to-everything v2x use cases for lte and nr," 3rd Generation Partnership Project (3GPP), Tech. Rep., 09.
- [38] 3GPP TR 36.873 v12.5.0, "Technical specification group radio access network; study on 3d channel model for lte," 3rd Generation Partnership Project (3GPP), Tech. Rep., 06.
- [39] Y. Hong, T. Thaj, and E. Viterbo, *Delay-Doppler Communications: Principles and Applications*, 1st ed. USA: Elsevier Science., 2022.
- [40] T. Thaj and E. Viterbo, "Low complexity iterative rake decision feedback equalizer for zero-padded OTFS systems," *IEEE transactions on vehicular technology*, vol. 69, no. 12, pp. 15 606–15 622, 2020.