IEEE INFOCOM 2024 - IEEE Conference on Computer Communications | 979-8-3503-8350-8/24/$31.00 ©2024 IEEE | DOI: 10.1109/INFOCOM52122.2024.10621411

Learning to Decompose Asymmetric Channel Kernels for Generalized Eigenwave Multiplexing

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*—Learning the principal eigenfunctions of a *kernel* is at the core of many machine-learning problems. Common methods usually deal with symmetric kernels based on Mercer's Theorem. However, in the communication systems, the channel kernel is usually asymmetric due to the inconsistencies between the uplink and the downlink propagation environment. In this paper, we propose an explainable Neural Network for extracting eigenfunctions from generic multi-dimensional asymmetric channel kernels based on a recent method called High Order Generalized Mercer's Theorem (HOGMT), by decomposing it into jointly orthogonal eigenfunctions. The proposed neural network based approach is efficient and can be easily implemented compared to the conventional SVD based solutions used for eigen decomposition. We also discuss the effect of different hyper-parameters on the training time, constraint satisfaction, and overall performance. Finally, we show that multiplexing using these eigenfunctions mitigates interference across all the available Degrees of Freedom (DoF), both mathematically as well as via neural network based system-level simulations.**

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

I. Introduction

The conventional waveform design techniques usually treat the interference separately, such as SVD-based precoding to cancel Inter-Antenna Interference (IAI) [1], OFDM modulation to avoid Inter-Symbol Interference (ISI) and OFDM detector to mitigate Inter-Carrier Interference (ICI) [2], [3]. Recently, Orthogonal Time Frequency Modulation (OTFS) [4] has been proposed to cancel joint interference in the time-frequency domain but still requires additional processing at the receiver to mitigate Inter-Doppler Interference (IDI) for rapidly time-varying channels [5], [6]. These methods explore the orthogonality in a specific domain instead of joint orthogonality across all the DoF due to the limitations of mathematical tools. Both Fourier Transfer (FT) and Symplectic Fourier Transform (SFT) can only find orthogonal bases in the time, frequency and delay-Doppler domain but cannot incorporate joint orthogonality in the space domain. Similarly, SVD-based precoding only decomposes independent channel matrices to achieve orthogonality in the space domain but fails to capture any joint interference over space-time dimensions [7].

A foundational exposition of representing multi-dimensional channels by its *asymmetric kernel* from first principles, which maps data symbols from the transmitter to the receiver, can be found in [8]–[10]. Recently, High Order Generalized Mercer's Theorem (HOGMT) [11], [12] has been proposed

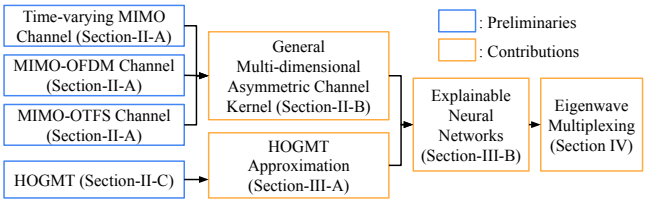


Figure 1: Paper outline: contributions and novelty

to decompose a multi-dimensional asymmetric kernel into jointly orthogonal eigenfunctions (reviewed in Section II). Specifically, it has been proven that *decomposing the channel kernel by HOGMT and multiplexing data symbols using the constituent eigenfunctions result in transmission over independent orthogonal subchannels in the eigen-domain, thereby avoiding interference from other symbols across all the DoF.*

In contrast to Fourier bases, commonly used in OFDM and OTFS, eigenfunctions are not pre-defined and are strictly unstructured [16], which makes them computationally expensive for practical purposes. In general, a multi-dimensional channel kernel is represented as a tensor instead of separate matrices to capture the joint interference across the dimensions. Decomposing this tensor by HOGMT by existing linear algebraic implementations, suffers from extremely high computational complexity [12], which limits its adoption for real-time signal processing in next Generation (xG) wireless networks.

Neural Network (NN) provides an advantage over linear techniques by inherently extracting non-linear relationships among hidden variables, while keeping the computational burden within manageable limits. This motivates us to embrace NN to implement practical HOGMT-based eigenwave multiplexing. NNs have been used to replace certain blocks in communication systems for optimized performance for a given dataset [17]–[22]. However, current NN-based eigen decomposition methods aim to learn the format and properties of the eigenvectors [23] without explaining the internal mechanisms of how the output converges to eigen components (*e.g.*, vectors, tensors or functions), which trains NNs as black-box systems. Furthermore, any random process like a multi-dimensional asymmetric channel kernel requires an infinite number of eigenfunctions for accurate characterization and reconstruction [24]. In contrast, practical implementations will only provide a finite number of eigenfunctions that approximates the kernel. Meanwhile, stringent requirements on reliability and throughput in xG wireless propagation

Authorized licensed use limited to: UNIVERSITY AT ALBANY. Downloaded on June 20,2025 at 20:59:24 UTC from IEEE Xplore. Restrictions apply.

Table I: Multi-dimensional asymmetric channel kernels in wireless communication

Channel type	Signal domain	Interference type	Input-output relation of channel kernel
MIMO [12], [13]	Space-time domain	IAI-ISI	$r(u,t) = \iint K_H(u,t;u',t')s(u',t')du'dt'$
MIMO-OFDM [7], [14]	Space-frequency domain	IAI-ICI	$r(u,f) = \iint K_B(u,f;u',f')s(u',f')du'df'$
MIMO-OTFS [14], [15]	Space-delay-Doppler domain	IAI-IDI	$r(u,\tau,\nu) = \iint K_{DD}(u,\tau,\nu;u',\tau',\nu')s(u,\tau',\nu')du'd\tau'd\nu'$
General multi-dimensional channel	All the DoF	All the DoF	$r(Z) = \int K(Z;\Gamma)s(\Gamma)\ d\Gamma$

environments [13], [25], [26] demands provable performance guarantees for transceivers over a variety of channels. Therefore, the *optimality* of kernel approximation by a finite number of eigenfunctions and the *explainability* of the NN-based eigen decomposition method are paramount design goals.

Figure 1 shows the contributions of this work and their connection to the necessary foundational concepts. In order to design a NN that is applicable to a wide variety of channels, an expression for a generic kernel is necessary. At the same time, it is imperative to prove that the output of the NN that approximates HOGMT, indeed converges to the eigenfunctions. This makes the implementation explainable and optimal. These steps ultimately allow for implementing practical multiplexing using eigenfunctions. Since the eigenfunctions represent waveforms at its DoF, it is convenient to refer to them as *eigenwaves* in the context of multiplexing. Therefore, the contributions of this paper are as follows:

- We deduce the expression for channel kernels for MIMO, MIMO-OFDM and MIMO-OTFS channels and formulate the kernel for any generic multi-dimensional channel.
- We prove that HOGMT approximation for multi-dimensional asymmetric kernel is optimal in MMSE.
- We design low-complexity, fully connected NN to decompose generic multi-dimensional asymmetric kernels into their eigenfunctions based on HOGMT.
- We propose a top-N eigenwave multiplexing method that mitigates interference across all DoF. We also give the error analysis with respect to approximate eigenfunctions and evaluate using NN based system-level simulations.

II. Asymmetric Channel Kernel

We begin by deducing the kernels for *three* known multi-dimensional channels from elementary principles of linear time-varying (LTV) channels, which is followed by a general formulation for multi-dimensional asymmetric channel kernel. These formulations are summarized in Table I. We also provide a refresher on kernel decomposition using HOGMT to contextualize the contributions of this work without ambiguity.

A. Channel Kernel Formulation

1) Case 1: MIMO (4-D) Channel Kernel

In LTV, the transmitted signal $s(t)$ is impacted by the underlying physics of the channel, described by path delays and Doppler shift to produce the received signal $r(t)$ [10] as,

$$r(t) = \sum_{p=1}^{P} h_p s(t-\tau_p)e^{j2\pi\nu_p t} \tag{1}$$

where h_p, τ_p and ν_p are the path attenuation factor, time delay and Doppler shift for path p, respectively. We omit the noise term for simplicity. Then (1) is expressed in terms of the overall delay τ and Doppler shift ν as

$$r(t) = \iint S_H(\nu,\tau)s(t-\tau)e^{j2\pi\nu t}\ d\tau\ d\nu \tag{2}$$

$$= \int h(t,\tau)s(t-\tau)\ d\tau \tag{3}$$

where $S_H(\nu,\tau)$ is the (Doppler-delay) spreading function, which describes the combined attenuation factor for all paths in the delay-Doppler domain. The time-varying impulse response $h(t,\tau)$ is related to $S_H(\nu,\tau)$ as,

$$h(t,\tau) = \int S_H(\nu,\tau)e^{j2\pi t\nu}\ d\nu \tag{4}$$

Extending $h(t,\tau)$ to MIMO case, $H_{u,u'}(t,\tau)$ denotes the time-varying impulse response between the u'-th transmit antenna and the u-th received antenna. Therefore, the multi-user (or multi-antenna) version of $h(t,\tau)$ [13] is a tensor,

$$\mathbf{H}(t,\tau) = \begin{bmatrix} H_{1,1}(t,\tau) & \cdots & H_{1,u'}(t,\tau) \\ \vdots & \ddots & \vdots \\ H_{u,1}(t,\tau) & \cdots & H_{u,u'}(t,\tau) \end{bmatrix} \tag{5}$$

and the received signal in (3) is extended as,

$$r_u(t) = \int \sum\nolimits_{u'} H_{u,u'}(t,\tau)s_{u'}(t-\tau)d\tau$$

$$= \int \sum\nolimits_{u'} K_{u,u'}(t,t')s_{u'}(t')dt' \tag{6}$$

where $K_{u,u'}(t,t')$=$H_{u,u'}(t,t-t')$ is the 4-D MIMO channel kernel following the definition of kernel in [8], [10]. Then, (6) can be rewritten using the kernel as[1],

$$r(u,t) = \iint K_H(u,t;u',t')s(u',t')\ du'\ dt' \tag{7}$$

In general, $\mathbf{H}(t,\tau)$ is asymmetric since the downlink and uplink channel can not be strictly the same, which makes the kernel $K_H(u,t;u',t')$ asymmetric in practice.

2) Case 2: MIMO-OFDM (4-D) Channel Kernel

Proposition 1. *The continuous form of OFDM input-output relation for LTV, in the frequency domain is given by,*

$$r(f) = \int b(f,\nu)s(f-\nu)d\nu \tag{8}$$

where $b(f,\nu)$[2] is the (frequency domain) transfer function [8]. Further, the continuous form of the input-output relation using the MIMO-OFDM kernel, $K_B(u,f;u',f')$ is given by,

$$r(u,f) = \iint K_B(u,f;u',f')s(u',f')\ du'\ df' \tag{9}$$

[1]Note that (7) represents the 2-D convolution integral over space and time DoF at the Tx and Rx, where t'=$t-\tau$ and u' are the variables at the Tx.

[2](8) is an archaic form of frequency domain input-output relation for LTV channels in [8], which defines $b(f,\nu)$ as *"Input Spectrum Output Spectrum relation"*. Since this formulation predates OFDM, we believe that it is necessary to prove (8) for MIMO-OFDM in LTV channels using contemporary formulations for OFDM in the literature based on DFT/IDFT.

Authorized licensed use limited to: UNIVERSITY AT ALBANY. Downloaded on June 20,2025 at 20:59:24 UTC from IEEE Xplore. Restrictions apply.

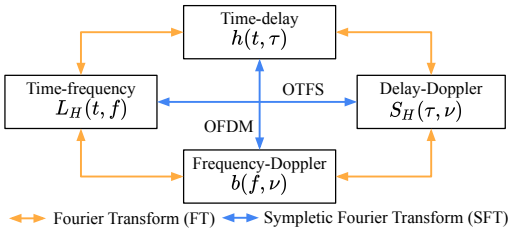


Figure 2: Representation domains of LTV channel

Proof. For notational compactness, rewrite (3) in matrix form, $\mathbf{r} = \mathbf{h}\mathbf{s}$, where $\mathbf{h}$ is the time-domain channel matrix. Then the frequency domain input-output relation is given by multiplying both sides by the M-point DFT matrix, $\mathbf{F}_M$ on both sides.

$$\underbrace{\mathbf{F}_M\mathbf{r}}_{\check{\mathbf{r}}} = \mathbf{F}_M\mathbf{h}\check{\mathbf{s}} = \underbrace{\mathbf{F}_M\mathbf{h}\mathbf{F}_M^H}_{\mathbf{b}}\underbrace{\mathbf{F}_M\mathbf{s}}_{\check{\mathbf{s}}} \implies \check{\mathbf{r}} = \mathbf{b}\check{\mathbf{s}} \tag{10}$$

Note that $\mathbf{F}_M\mathbf{h}\mathbf{F}_M^H$ is the SFT of $\mathbf{h}$ as shown in Figure 2 and in [14]. Then the continuous form of $\mathbf{b}$ is given by,

$$b(f,\nu) = \iint h(t,\tau)e^{j2\pi(t\nu - f\tau)}\, dt\, d\tau \tag{11}$$

Meanwhile, both $\check{\mathbf{r}}$ and $\check{\mathbf{s}}$ represent the signals in the frequency domain. Therefore, using (11) we can derive the OFDM input-output relation for LTV channels as,

$$r(f) = \int b(f,\nu)s(f-\nu)d\nu \tag{12}$$

Furthermore, extending $b(f,\nu)$ to the MIMO case as $B(f,\nu)$,

$$r_u(f) = \int \sum_{u'} B_{u,u'}(f,\nu)s_{u'}(f-\nu)d\nu \tag{13}$$

Similarly, the kernel input-output relation is given as

$$r(u,f) = \iint K_B(u,f;u',f')s(u',f')\, du'\, df' \tag{14}$$

This follows in the same spirit as defining (7) using (6). □

3) Case 3: MIMO-OTFS (6-D) Channel Kernel

In the literature, OTFS is often written using the Zak representation [14], [15] of (1) as,

$$\mathcal{Z}_r(\tau,\nu) = \sum\nolimits_{p=1}^{P} h_p e^{j2\pi v_p(\tau-\tau_p)}\mathcal{Z}_s(\tau-\tau_p,\nu-\nu_p) \tag{15}$$

where $\mathcal{Z}_r(\tau,\nu)$ is the Zak transform of $r(t)$ given by,

$$\mathcal{Z}_r(\tau,\nu) \triangleq \sqrt{T}\sum_k r(\tau+kT)e^{-j2\pi k\nu T} \tag{16}$$

where T is the symbol duration. Denote $r(\tau,\nu) = \mathcal{Z}_r(\tau,\nu)$ and $s(\tau,\nu) = \mathcal{Z}_s(\tau,\nu)$ and rewrite (15) as,

$$r(\tau,\nu) = \iint h_{dd}(\tau,\nu;\tau',\nu')s(\tau-\tau',\nu-\nu')\, d\tau'\, d\nu' \tag{17}$$

where $h_{dd}(\tau,\nu;\tau',\nu')d\tau'd\nu'$ describes the combined path gains for all paths in the delay and Doppler-shift range $(\tau',\tau'+d\tau')$ and $(\nu',\nu'+d\nu')$ respectively.

Following similar extensions for MIMO as in previous cases and defining the kernel, (17) is written using the MIMO-OTFS kernel in the space-delay-Doppler domain as,

$$r(u,\tau,\nu) = \iint K_{DD}(u,\tau,\nu;u',\tau',\nu')s(u',\tau',\nu')du'd\tau'd\nu' \tag{18}$$

Interpretation: The received OTFS symbol at an antenna (u) is the integral of the channel kernel, K_{DD} multiplying all the OTFS symbols over the space-delay-Doppler domain that are responsible for IAI and IDI. For a single-path, delay-Doppler channel, if the Doppler shift and delay are integer multiples of the time and frequency grids, then the kernel is 0 at other symbols, meaning there is no IDI [15]. However, this is not realistic as: 1) The perfect division of time-frequency grids is not practical; 2) For multi-path delay-Doppler channels, as the delay and Doppler shifts are different on each path, there is no common factor satisfying the integer multiple requirement. Therefore, any practical MIMO-OTFS system will have IDI and IAI and hence we include that in the kernel formulation.

B. General Multi-dimensional Asymmetric Channel Kernel

Based on the kernel formulations above, the input-output relation of *any* wireless channel (without considering noise) can be modeled as a mapping of the signal at the transmitter to the signal at the receiver by a channel kernel, K as,

$$r(\zeta_1,\ldots,\zeta_P) = \int\ldots\int K(\zeta_1,\ldots,\zeta_P;\gamma_1,\ldots,\gamma_Q)s(\gamma_1,\ldots,\gamma_Q)d\gamma_1\ldots d\gamma_Q \tag{19}$$

where $(\zeta_1,\ldots,\zeta_P)$ and $(\gamma_1,\ldots,\gamma_Q)$ are the degrees of freedom (*e.g.*, space, time, frequency, delay, Doppler, etc.) at the receiver and the transmitter, respectively. In general, for communication systems $P{=}Q$. For brevity, henceforth we denote $Z{=}(\zeta_1,\ldots,\zeta_P)$ and $\Gamma{=}(\gamma_1,\ldots,\gamma_Q)$, rewriting (19) as,

$$r(Z) = \int K(Z;\Gamma)s(\Gamma)\, d\Gamma \tag{20}$$

In general, (20) is applicable to any higher-dimensional kernel that may incorporate joint interference between any number of DoFs that are unique to a particular communication paradigm such as scattering angles, polarizations, etc. [27], [28].

Interpretation: At the transmitter the signal, $s(Z)$ is transmitted in Z domain (also the transmitter DoFs), which is converted to the signal in Γ domain $s(\Gamma)$ during convolution. The convolution (during propagation over LTV), projects $s(\Gamma)$ on to the channel kernel K, which transfers the signals to the Z domain (DoF at the receiver) as $r(Z)$ with the interference across Γ domain (DoF). Special cases of this kernel interpretation are seen in (7), (14) and (18).

C. Kernel Decomposition

A generalized version of Mercer's Theorem, called High Order Generalized Mercer's Theorem (HOGMT) has been recently proposed in [12]. This presents a mathematically principled approach to decompose multi-dimensional asymmetric channel kernels like in (20), into low-dimension, jointly orthogonal eigenfunctions, which is expressed as,

$$K(Z;\Gamma) = \sum_{n=1}^{\infty}\sigma_n\psi_n(Z)\phi_n(\Gamma) \tag{21}$$

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

$$\int\phi_n(\Gamma)\phi_{n'}^*(\Gamma)d\Gamma{=}\delta_{nn'} \text{ and } \int\psi_n(Z)\psi_{n'}^*(Z)dZ{=}\delta_{nn'} \tag{22}$$

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

$$\int K(Z;\Gamma)\phi_n^*(\Gamma)\, d\Gamma = \sigma_n\psi_n(Z) \tag{23}$$

This property is critical for using the eigenfunctions as

Authorized licensed use limited to: UNIVERSITY AT ALBANY. Downloaded on June 20,2025 at 20:59:24 UTC from IEEE Xplore. Restrictions apply.

independent subchannels in practice. (23) suggests that when one of the eigenfunctions is transmitted through a multi-dimensional channel kernel, it is transferred to it's dual eigenfunction, scaled only the corresponding eigenvalue (or subchannel gain). Therefore, the orthonormality and duality of the eigenfunction, unambiguously allows us to design subchannels in the eigen-domain that achieves independent flat-fading at the receiver. Note that in this context, *each subchannel is a pair of dual eigenfunctions, $\phi_n(\Gamma)$ and $\psi_n(Z)$.*

III. Learning to Decompose Channel Kernel

A. Kernel Approximation by Finite Eigenfunctions

HOGMT provides us a *mathematical tool* to decompose any kernel into an infinite number of eigenfunctions. However, in reality, we can only utilize a finite number of eigenfunctions to approximate the kernel. To achieve maximum energy efficiency, it is desirable to use the least number of eigen components to approximate (most part of) the kernel. Therefore, the general approximation problem is formulated as minimizing the number of eigenfunctions, N that limits the kernel approximation error below some threshold,

$$\arg\min_{\hat{K}} \quad N \quad \text{s.t} \quad \|K(Z;\Gamma) - \hat{K}(Z;\Gamma)\|^2 < \epsilon \tag{24}$$

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

However, in practice the number of eigenfunctions is fixed. Then (24) is equivalent to the MMSE problem for a fixed N,

$$\arg\min_{\hat{K}} \quad \mathbb{E}\big\{\|K(Z;\Gamma) - \hat{K}(Z;\Gamma)\|^2\big\} \tag{25}$$

We solve this problem by proving that the approximate kernel, $\hat{K}$ reconstructed from the eigenfunctions obtained using HOGMT is optimal in MMSE sense, which is given in Theorem 1. The choice and trade-offs regarding N are discussed in Section IV and evaluated in Section VI-B.

Theorem 1. *(Eigenfunction approximation for asymmetric kernel) If $K(Z;\Gamma)$ be an asymmetric kernel, approximating it using eigenfunctions decomposed by HOGMT solves* (25)*.*

Proof. Denote $\epsilon_N = \mathbb{E}\big\{\|K(Z;\Gamma) - \hat{K}(Z;\Gamma)\|^2\big\}$. Then the total ϵ_N across all DoF is given by,

$$\begin{aligned}
&\iint \epsilon_N \; dZ \; d\Gamma \\
&= \iint \mathbb{E}\Big\{\big[K(Z;\Gamma) - \sum_{n=1}^{N} k_n f_n(Z) g_n(\Gamma)\big]^2\Big\} \; dZ \; d\Gamma \\
&= \iint \mathbb{E}\big\{\big[K(Z;\Gamma)\big]^2\big\} \; dZ \; d\Gamma - \sum_{n=1}^{N} \mathbb{E}\big\{k_n^2\big\}
\end{aligned} \tag{26}$$

As the first term is unrelated to the choice of $\{f_n\}$ and $\{g_n\}$, the problem in (25) is equivalent to maximizing the second term in (26) with the following constraints,

$$\begin{aligned}
\arg\max_{\{f_n\},\{g_n\}} \quad & \sum_{n=1}^{N} \mathbb{E}\big\{k_n^2\big\} \\
\text{s.t.} \quad & \langle f_n, f_{n'}\rangle = \delta_{nn'}, \langle g_n, g_{n'}\rangle = \delta_{nn'}
\end{aligned} \tag{27}$$

Now, since by definition, k_n is the projection of the kernel onto the bases, we rewrite the objective function in (27) as,

$$\sum_{n=1}^{N} \mathbb{E}\Big\{\Big[\iint K(Z;\Gamma) f_n^*(Z) g_n^*(\Gamma) \; dZ \; d\Gamma\Big]^2\Big\} \tag{28}$$

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

$$\begin{aligned}
\mathcal{E} = \sum_{n=1}^{N} \mathbb{E}\Big\{\Big| \iint K(Z;\Gamma) f_n^*(Z) g_n^*(\Gamma) \; dZ \; d\Gamma \\
- \frac{1}{2}\beta_n\Big(\int f_n(Z) f_n^*(Z) \; dZ - 1\Big)\Big|\Big\}
\end{aligned} \tag{29}$$

Differentiating with respect to each f_n^* and setting the derivative to 0 yields,

$$\begin{aligned}
&\frac{\partial \mathcal{E}}{\partial f_n^*(Z)} = \\
&\mathbb{E}\Big\{\Big|\int \Big(\int K(Z;\Gamma) g_n^*(\Gamma) \; d\Gamma - \beta_n f_n(Z)\Big) \; dZ\Big|\Big\} = 0
\end{aligned} \tag{30}$$

which is satisfied when

$$\int K(Z;\Gamma) g_n^*(\Gamma) \; d\Gamma = \beta_n f_n(Z) \tag{31}$$

(31) proves that the dual eigenfunctions obtained from HOGMT with the property in (23) is indeed a solution to the problem posed in (25) [3]. □

The symmetric channel kernel is a special case of the asymmetric case. Therefore, Theorem 1 has a degenerate case for approximating symmetric kernels using the Mercer's theorem [24] as shown in Corollary 1.

Corollary 1. *(Degrade to the symmetric case) Specifically, if the kernel $K(Z;\Gamma)$ is symmetric, then decomposing by Mercer's theorem minimizes the kernel approximation error.*

Proof. For a symmetric kernel, we have the decomposition,

$$\hat{K}(Z;\Gamma) = \sum_{n=1}^{N} k_n f_n(Z) f_n(\Gamma) \tag{32}$$

Following the steps in proof of Theorem 1, we have

$$\int K(Z;\Gamma) f_n^*(\Gamma) \; d\Gamma = \beta_n f_n(Z) \tag{33}$$

which is the very definition of the eigenfunction f_n, hence suggesting (32) approximated by Mercer's theorem resulting in MMSE of the approximate kernel $\hat{K}$. □

Since, Theorem 1 generalizes Corollary 1, we always consider asymmetric kernel in our analysis, unless otherwise specified. It also shows that the approximate kernel defined by finite number of eigenfunctions decomposed by HOGMT is optimal. This satisfies the *optimality* requirement highlighted in Section I. Theorem 1 also paves the way towards a computationally tractable method to extract these eigenfunctions in practice. Unlike time-invariant channels, where complex exponentials are eigenfunctions and easily implemented by Fourier

[3] Note that introducing Lagrangian multiplier for g_n and deriving with respect to each g_n^* results in the dual form of (31) which reinforces the duality property of f_n and g_n, leading to the same proof of optimality.

Authorized licensed use limited to: UNIVERSITY AT ALBANY. Downloaded on June 20,2025 at 20:59:24 UTC from IEEE Xplore. Restrictions apply.

methods, there are no *fixed* eigenfunctions for general LTV channels. This has been one of the impediments for practical implementations [16]. Therefore, empowered by Theorem 1, we postulate that *a neural network that minimizes the MSE of the approximate channel kernel is guaranteed to produce output that converges to finite set of eigenfunctions*. This makes the proposed NN-based implementation of HOGMT *explainable*, another desired property mentioned in Section I.

B. HOGMT−NN: Neural Network for HOGMT

As postulated above, we design a NN called HOGMT−NN whose outputs converge to finite number eigenfunctions obtained by HOGMT. This is made possible by formulation of a regularized loss function for the NN as shown in Corollary 2.

Corollary 2. *(Equivalent NN for HOGMT) HOGMT−NN minimizes the regularized MSE loss in* (34)*,*

$$\mathcal{L}=\mathcal{J}+\Omega_1+\Omega_2 \tag{34}$$

$$\textit{where,}\quad \mathcal{J}=\frac{1}{B}\sum_{b=1}^{B}\frac{\|\mathbf{K}_b-\sum_{n=1}^{N}\hat{\sigma}_{n,b}\hat{\Phi}_{n,b}\otimes\hat{\Psi}_{n,b}\|^2}{\|\mathbf{K_b}\|^2} \tag{35}$$

$$\Omega_1=\mathcal{P}_1\frac{1}{B}\sum_{b=1}^{B}\sum_{n=1}^{N}\sum_{n'\neq n}^{N}|\langle\hat{\Phi}_{n,b},\hat{\Phi}_{n',b}\rangle| \tag{36}$$

$$\Omega_2=\mathcal{P}_2\frac{1}{B}\sum_{b=1}^{B}\sum_{n=1}^{N}\sum_{n'\neq n}^{N}|\langle\hat{\Psi}_{n,b},\hat{\Psi}_{n',b}\rangle| \tag{37}$$

such that the outputs $\{\hat{\Phi}_{n,b}\}_{n=1}^{N}$ and $\{\hat{\Psi}_{n,b}\}_{n=1}^{N}$ converge to eigenfunctions of the kernel K_b, while $\hat{\lambda}_{n,b}=\hat{\sigma}_{n,b}^2$ converge to eigenvalues. $\mathcal{P}_1$ and $\mathcal{P}_2$ are the penalty coefficients and B is the batch size. $\otimes$ denotes Kronecker product.

Proof. From Theorem 1 we learned that an arbitrary set of orthogonal bases must be eigenfunctions, in order to approximate a channel kernel in MMSE sense. A careful observation reveals that the MMSE problem in (25) is equivalent to the MSE loss function, $\mathcal{J}$ and the corresponding orthogonality constraints are equivalent to the regularization factors Ω_1 and Ω_2 of HOGMT−NN. Therefore, the output of HOGMT−NN must converge to finite (N) eigenfunctions and eigenvalues of the input kernel $\mathbf{K}_b$. □

C. Complexity Analysis

Table II compares the complexity of eigen decomposition methods. SVD and DNN for SVD are designed to decompose 2-D matrices only. Given a $M_1 \times M_2$ matrix, their complexity is $\mathcal{O}(\min(M_1M_2^2, M_1^2M_2))$ and $\mathcal{O}(\max(N^2M_1M_2, N^2M_1^2, N^2M_2^2))$, respectively, where N is the number of output eigenvectors. While High-Order SVD (HOSVD) is used in HOGMT [12] and the proposed HOGMT−NN also decomposes multi-dimensional tensors. The complexity of HOSVD is approximately $\mathcal{O}(N_dL^3)$, where N_d is the order of dimensions and M is the maximum length of dimensions for the unfolding matrix [29]. Given a $M_1\times M_2\times M_3\times M_4$ tensor, N_d=4 and $M=\max(M_1M_2M_3, M_1M_2M_4, M_2M_3M_4)$. The

Table II: Comparing complexity of eigen decomposition

Eigen decomposition	Time complexity
SVD [30]	$\mathcal{O}(\min(M_1M_2^2, M_1^2M_2))$
DNN for SVD [23]	$\mathcal{O}(\max(N^2M_1M_2, N^2M_1^2, N^2M_2^2))$
HOSVD [29]	$\mathcal{O}(N_dM^3)$
HOGMT [12]	$\mathcal{O}(\min(M_1M_2(M_3M_4)^2, (M_1M_2)^2M_3M_4))$
HOGMT−NN	$\mathcal{O}(NN_LN_M)$

relationship between SVD, HOSVD and HOGMT is shown in Lemma 3 of [12], where the implementation unfolds SVD with the complexity $\mathcal{O}(\min(M_1M_2(M_3M_4)^2, (M_1M_2)^2M_3M_4))$. Since, HOGMT−NN use fully connected architecture (Section V), the time complexity is $\mathcal{O}(NN_LN_M)$, where N_L is the number of layers. N_M is the twice number of elements contained in the tensor for complex-valued NN. Given the above tensor, we have N_M=$2M_1M_2M_3M_4$. As N_M is much less than M^3, HOGMT−NN has less complexity than HOSVD. Meanwhile, for a fixed number N and N_L, we have $\mathcal{O}(NN_LN_M)$<$\mathcal{O}(\min(M_1M_2(M_3M_4)^2, (M_1M_2)^2M_3M_4))$ because the complexity of HOGMT−NN increases at a much slower rate than HOGMT by unfolding SVD with respect to the increasing size of the input tensor.

IV. Communicating using Eigenwaves

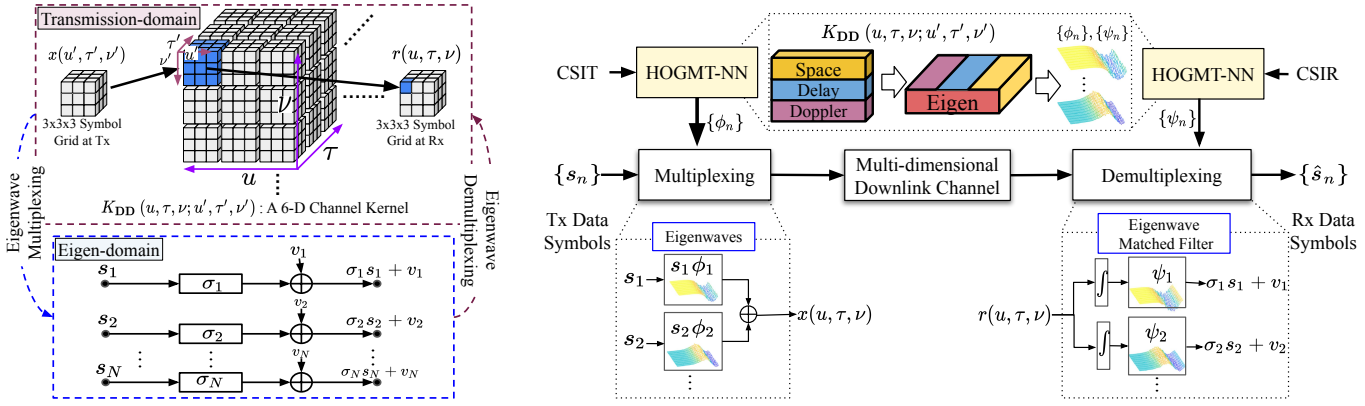
The concepts in Section II-C and the subsequent discussions provide a principled approach to combine all the DoF of the channel, embodied in its kernel, into one eigen-domain, which is simply an alternate view of the multi-dimensional channel. The eigen-domain is then divided into independent subchannels, called eigenwaves, a term introduced in Section I. Therefore, multiplexing data-symbols using eigenwaves achieve full diversity in eigen-domain, which implies that the scheme also achieves full diversity gain along each DoF.

Figure 3a shows transmitting signals over a 6-D channel kernel. The "transmission domain" may involve physical processing blocks where signals are convolved over the multi-dimensional channel. Since the DoF at the Tx and Rx are the same, the transmitted symbols, $x(u', \tau', \nu')$ (different from baseband modulated symbols) are arranged in a 3-D grid. This 3-D grid integrates over the 3-D kernel block (blue) across its DoF (u', τ', ν') during convolution and produces a (blue) cube in the received 3-D symbol grid. To obtain each cube across the DoF (u, τ, ν) at the receiver, the 3-D block kernel with DoF (u', τ', ν') should be arranged across DoF (u, τ, ν) leading to a 6-D kernel tensor.

Decomposing the kernel into eigenwaves and multiplexing, data-symbols, $\{s_n\}$ are transmitted over independent parallel subchannels in the "eigen-domain" as shown in the figure. Figure 3b on the other hand, presents an implementation view using HOGMT−NN for decomposing the kernel, obtained from the Channel State Information at the Transmitter (CSIT). It also shows the process of multiplexing and demultiplexing to convert the baseband data-symbols between the transmission and the eigen-domain, discussed in Sections IV-A and IV-B.

Note that, in Section III, we learned that HOGMT−NN

Authorized licensed use limited to: UNIVERSITY AT ALBANY. Downloaded on June 20,2025 at 20:59:24 UTC from IEEE Xplore. Restrictions apply.



(a) Parallel subchannels in eigen-domain

(b) Practical implementation with HOGMT−NN

Figure 3: Two views of eigenwave multiplexing: a) Transmission / Eigen-domain view of a 6-D MIMO-OTFS channel kernel with $u{=}u'{=}3, \tau{=}\tau'{=}3, \nu{=}\nu'{=}3$; b) System view for practical implementation with HOGMT−NN

approximates the kernel by a finite number of eigenwaves (N) ordered by descending eigenvalues. Therefore, transmitting signals over eigenwaves with *top-N* eigenvalues achieve the highest subchannel gains (σ_n). A large value of N will increase the complexity of HOGMT−NN and utilize lower-ranked eigenwaves (lower subchannel gain), which in turn increases the BER. On the other hand, a small N will reduce the throughput. There are multiple strategies for choosing N with respect to different priorities such as memory, complexity, BER or throughput. This trade-off is evaluated in Section VI.

A. Eigenwave Multiplexing

At the transmitter, the generic symbols $x(Z)$ is obtained by multiplexing the data symbols $\{s_n\}$ and eigenwaves $\{\phi_n^*(Z)\}$ corresponding to the Z DoF as,

$$x(Z) = \sum_{n=1}^{N} s_n \phi_n^*(Z) \tag{38}$$

The orthogonality and duality of the eigenfunctions, ensure that the data symbols remain orthogonal from each other after transmission over the multi-dimensional channel, while the corresponding eigenwaves are transferred to its dual eigenwaves by the kernel. Considering infinite eigenfunctions for the kernel (real channel) but finite number for the transmitted signal (practical case), and no noise (for brevity), the signal at the receiver with Γ DoF is given by the convolution integral (see the interpretation in Section II-B),

$$\begin{aligned} r(Z) &= \int K(Z;\Gamma) x(\Gamma)\, d\Gamma \\ &= \int \Bigg\{ \underbrace{\sum_{n=1}^{\infty} \sigma_n \psi_n(Z) \phi_n(\Gamma)}_{\text{decomposed kernel}} \underbrace{\sum_{n'=1}^{N} s_{n'} \phi_{n'}^*(\Gamma)}_{\text{transmitted symbols}} \Bigg\} d\Gamma \\ &= \int \Bigg\{ \sum_{n=1}^{N} \sigma_n s_n \psi_n(Z) \underbrace{|\phi_n(\Gamma)|^2}_{=1} \\ &\qquad + \sum_{n=1}^{\infty} \sum_{n' \neq n}^{N} \sigma_n s_{n'} \psi_n(Z) \underbrace{\phi_n(\Gamma) \phi_{n'}^*(\Gamma)}_{=0} \Bigg\} d\Gamma \\ &= \sum_{n=1}^{N} \sigma_n s_n \psi_n(Z) \end{aligned} \tag{39}$$

Therefore, (39) shows that the received signals is essentially a projection of the baseband symbols on the eigenspace spanned by the eigenfunctions $\{\phi_n\}$ and the channel transfers it to its dual eigenfunction $\{\psi_n\}$, scaled by gain of parallel subchannel $\{\sigma_n\}$. In this process the data-symbols $\{s_n\}$ are still kept independent in the eigen-domain and can be recovered by a demultiplexing process as shown in Figure 3a.

B. Eigenwave Demultiplexing

At the receiver, the received signal $r(Z)$ is projected back to the baseband signal space by multiplying with the conjugate of the dual eigenwave,

$$\begin{aligned} \hat{s}_n &= \int r(Z) \psi_n^*(Z)\, dZ = \int \sum_{n'=1}^{N} \sigma_{n'} s_{n'} \psi_{n'}(Z) \psi_n^*(Z)\, dZ \\ &= \int \sigma_n s_n |\psi_n(Z)|^2\, dZ = \sigma_n s_n \end{aligned} \tag{40}$$

The symbol $\hat{s}_n$ is an estimate of the original transmitted data symbol, scaled by the channel gain without any interference from other symbols along all the DoF. This is precisely the property of a matched filter but using eigenwaves of the CSIR decomposed by HOGMT−NN, which is highlighted in Figure 3b as "Eigenwave Matched Filter". Although, we assume perfect channel estimation and CSIT=CSIR, it is not the focus of this work. Also, the CSIT and CSIR errors are equivalent to the approximation error of $\{\hat{\phi}_n\}$ and $\{\hat{\psi}_n\}$ obtained from HOGMT−NN, which is discussed in Sections IV-C and VI.

C. Symbol Error Analysis

Ideally, all the eigenfunctions are strictly orthogonal and joint interference is fully canceled as in (40). However, in practice there exist approximation errors at the output of HOGMT−NN, resulting in a correlation (non-orthogonality) between eigenfunctions. This can also result from any error in CSIT. Consequently, using the correlated eigenfunctions for demultiplexing are not perfectly matched with the duals used

Authorized licensed use limited to: UNIVERSITY AT ALBANY. Downloaded on June 20,2025 at 20:59:24 UTC from IEEE Xplore. Restrictions apply.

for multiplexing, leading to symbol error in the transmission-domain. So, (40) can be rewritten as,

$$\hat{s}_n = \hat{\sigma}_n s_n + \sum_{n' \neq n}^{N} \hat{\sigma}_{n'} s_{n'} R_{\hat{\psi}_{n'}\hat{\psi}_n^*} \tag{41}$$

where $R_{\hat{\psi}_{n'}\hat{\psi}_n^*} = \int \hat{\psi}_{n'}(Z)\hat{\psi}_n^*(Z)dZ$ is the correlation of $\hat{\psi}_n$ and $\hat{\psi}_{n'}$, i.e., $\{\hat{\psi}_n\}$ are not orthogonal. When if $R_{\hat{\psi}_{n'}\hat{\psi}_n^*} \neq 0$ it serves as a measure of interference from other symbols. Although from (41) it may appear that the interference is only expressed by the orthogonality of $\{\hat{\psi}_n\}$, it is also related to that of $\{\hat{\phi}_n\}$. Since the approximate eigenfunctions at the output of HOGMT−NN also possess the *duality* property according to Theorem 1, it is impossible to obtain orthogonal $\{\hat{\psi}_n\}$ from non-orthogonal $\{\hat{\phi}_n\}$ and vice-versa. Therefore, like dual-orthogonality, dual non-orthogonality also impacts multiplexing and demultiplexing in a similar manner.

V. HOGMT−NN Implementation

Based on the specifications detailed in Section III-B, we implement HOGMT−NN with a clean state approach. The novelty of the underlying method does not provide any baseline NN model. So, we start with a fully-connected NN to keep computational complexity as low as possible without compromising the quality of the decomposition. It is also evident from Section III-B that the size of the input and output layers are dictated by the dimension of the channel kernel and the number of finite eigenfunctions used to perform eigenwave multiplexing. Both of these can be considered as hyperparameters and need to be decided prior to training the NN. In Section VI we explore the performance of HOGMT−NN for different values of N and kernel dimensions.

Choice of the kernel for evaluation: We know that the transmission-domain for both MIMO-OTFS and MIMO-OFDM is ultimately in the space-time domain:

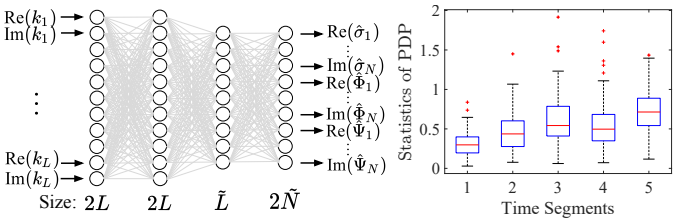
$$\text{MIMO-OTFS: } s(u,\tau,\nu) \xrightarrow{\text{SFT}} s(u,t,f) \xrightarrow{\text{IDFT}} s(u,t)$$

$$\text{MIMO-OFDM: } s(u,f) \xrightarrow{\text{IDFT}} s(u,t)$$

Therefore, for generality, we choose to evaluate HOGMT−NN using the *"rapidly time-varying MIMO channels"*, with the kernel, $K_H(u,t;u',t')$ given in the first row of Table I. Since, HOGMT−NN can handle tensors, HOGMT−NN is able to learn the variation in the time domain. Also, to understand the decomposition with changing statistics [23] the kernel is also non-stationary. This is further explained in the dataset generation section below. Note that the choice of $K_H(u,t;u',t')$ is primarily for practicality. In general, HOGMT−NN does not constrain the dimensions of the input kernel as per Theorem 1.

A. Neural Network Architecture

Figure 4a shows the architecture of HOGMT−NN for decomposing the channel kernel $K_H(u,t;u',t') \in \mathbb{C}^{N_u \times N_t \times N_{u'} \times N_{t'}}$. It consists of 4 fully connected layers with number of parameters, $L \triangleq N_u \times N_t \times N_{u'} \times N_{t'}$. Prior to input we split the real and imaginary parts and arrange as vectors in $\mathbb{R}^{2L}$. For N eigenwaves, the total number of parameters for the output components $\{\hat{\sigma}_n, \hat{\Phi}_n, \hat{\Psi}_n\}_{n=1}^N$ is $\tilde{N} \triangleq N(1+N_uN_t+N_{u'}N_{t'})$. The sizes of the input, first hidden, second hidden, and the output layers are $2L$, $2L$, $\tilde{L} \triangleq L+\tilde{N}$, and $2\tilde{N}$, respectively. To capture negative values of the real and imaginary parts of the kernel, LeakyRelu activation function is used. HOGMT−NN is trained using the Adam optimizer with a learning rate of 1×10^{-5} and mini-batch size is $B=16$ for all tests. For each data sample, the real and imaginary parts are combined and reshaped to 2-D matrix to conform to the dimension of the transmission-domain. Then, the loss function $\mathcal{J}$ and constraints Ω_1 and Ω_2 are calculated according to (35), (36) and (37), respectively.



(a) HOGMT−NN Architecture (b) PDP distribution

Figure 4: NN architecture and statistics of channel between the first antenna pair (u=u'=1) in the dataset

B. Data Generation

To train the complex-valued channel kernel $K_H(u,t;u',t')$, we generated 5120 time-varying MIMO channel kernels. Each kernel is with $N_u \times N_{u'}$ antennas for N_t=$N_{t'}$ moments (symbols), where N_u=$N_{u'}$=4 and N_t=$N_{t'}$=5 in our work. The delay taps are randomly generated in the interval [3 5]. The distribution of Power Delay Profile (PDP) at each moment of the generated kernel changes over time as shown in Figure 4b for one antenna pair. A training-test split of 80%-20% is used to evaluate the model. The code is available at *https://anonymous.4open.science/r/NNs-based-EM-D6CD*.

VI. Results

A. HOGMT−NN Accuracy

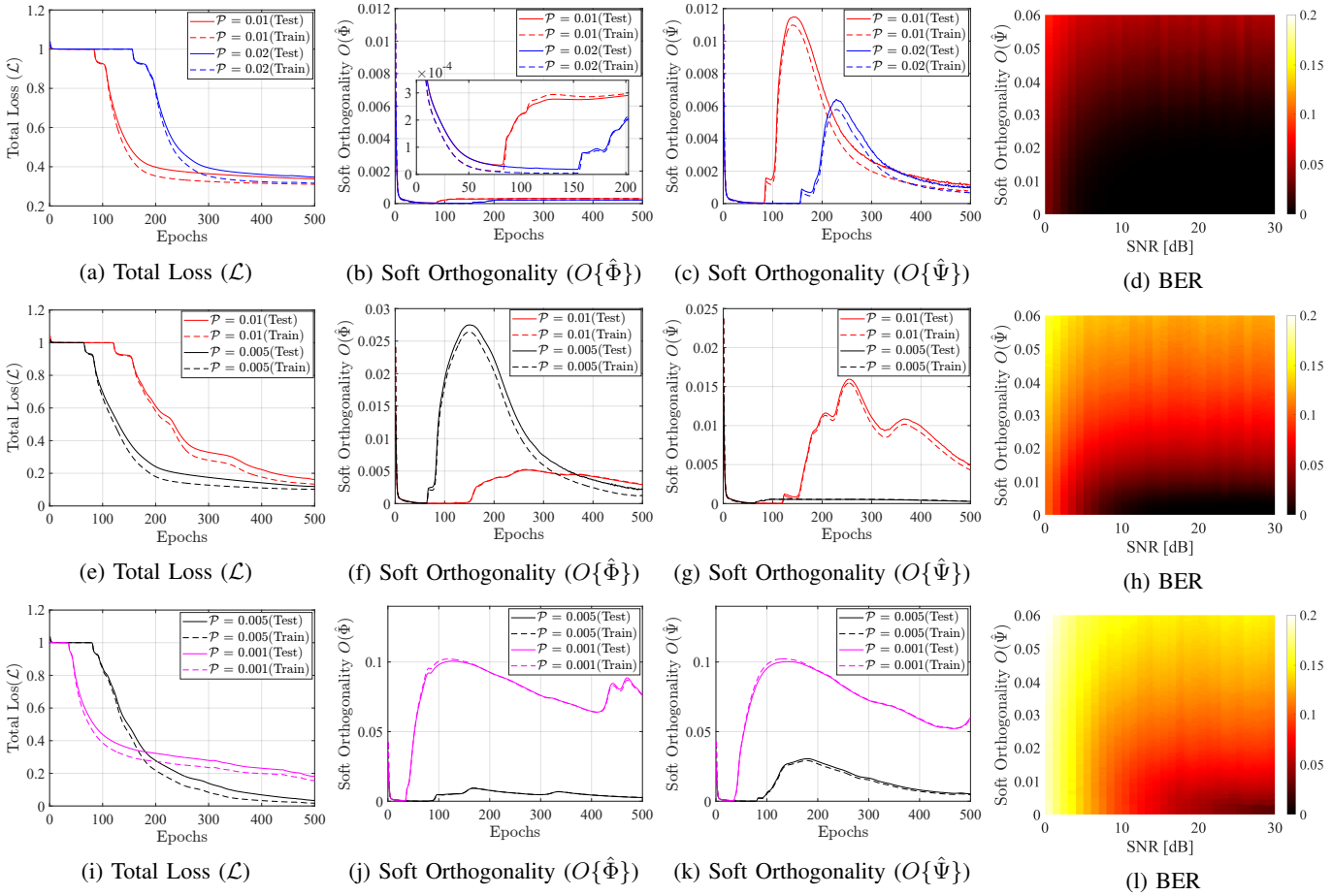
For implementation, we set the same penalty for the regularization terms, i.e., $\mathcal{P}_1$=$\mathcal{P}_2$=$\mathcal{P}$. Unlike algebraic eigen-decomposition methods that result in strict orthogonal eigenvectors, eigenwaves obtained by HOGMT−NN only have soft orthogonality, based on the efficacy of the NN. So, we define a measure for soft orthogonality,

Definition 1. *Soft orthogonality of the normalized basis function* $\{\mathbf{e}_n\}$ *is defined as*

$$O(\mathbf{e}) = \frac{1}{N(N-1)} \sum_{n=1}^{N} \sum_{n' \neq n}^{N} |\langle \mathbf{e}_n, \mathbf{e}_{n'} \rangle| \tag{42}$$

where the best case is $O(\mathbf{e})$=0*, meaning the bases is strictly orthogonal and the worst case is* $O(\mathbf{e})$=1*, indicating all the bases are the same. Specifically, we denote* $O(\hat{\Phi})$ *and* $O(\hat{\Psi})$ *as the soft orthogonality of* $\{\hat{\Phi}_n\}$ *and* $\{\hat{\Psi}_n\}$*, respectively.*

Authorized licensed use limited to: UNIVERSITY AT ALBANY. Downloaded on June 20,2025 at 20:59:24 UTC from IEEE Xplore. Restrictions apply.



(a) Total Loss ($\mathcal{L}$) (b) Soft Orthogonality ($O\{\hat{\Phi}\}$) (c) Soft Orthogonality ($O\{\hat{\Psi}\}$) (d) BER

(e) Total Loss ($\mathcal{L}$) (f) Soft Orthogonality ($O\{\hat{\Phi}\}$) (g) Soft Orthogonality ($O\{\hat{\Psi}\}$) (h) BER

(i) Total Loss ($\mathcal{L}$) (j) Soft Orthogonality ($O\{\hat{\Phi}\}$) (k) Soft Orthogonality ($O\{\hat{\Psi}\}$) (l) BER

Figure 5: Performance of HOGMT−NN and *top-N* eigenwave multiplexing: (a)-(c) show the total loss $\mathcal{L}$, soft orthogonality $O(\hat{\Phi})$ and $O(\hat{\Psi})$ defined by (42) for N=10 with penalties $\mathcal{P}$=0.01 and 0.02 over 500 epochs. (d) BER of *top-N* eigenwave multiplexing for N=10 over SNR and $O(\hat{\Psi})$; (e)-(g) $\mathcal{L}$, $O(\hat{\Phi})$ and $O(\hat{\Psi})$ for N=15 with $\mathcal{P}$=0.005 and 0.01; (h) BER for N=15; (i)-(l) $\mathcal{L}$, $O(\hat{\Phi})$ and $O(\hat{\Psi})$ for N=20 with $\mathcal{P}$=0.0001 and 0.005; (h) BER for $N = 20$.

Figures 5a-5c, Figures 5e-5g and Figures 5i-5k compare the performance of HOGMT−NN. For all the different values of N and $\mathcal{P}$, HOGMT−NN first learns to satisfy the orthogonality constraints of eigenwaves, which quickly fall down to near 0 over the first few epochs as shown in Figures 5b, 5c, 5f, 5g, 5j and 5k. Since, $\mathcal{P}$ is small, this sudden and large drop leads to a small decrease in the total loss in Figures 5a, 5e and 5i. Following this, the total loss and soft orthogonality measures for all the cases remain constant for a while until $O(\hat{\Phi})$ and $O(\hat{\Psi})$ increase while $\mathcal{L}$ drops at the same time as shown, which indicates that the orthogonality constraints are being relaxed to lower $\mathcal{L}$. Subsequently, both constraints follow similar behavior; either both are orthogonal or both non-orthogonal, which validates the discussion in Section IV-C.

When $\mathcal{P}$ is small in (34), $\mathcal{J}\approx\mathcal{L}$, since the total loss is dominated by $\mathcal{J}$. Also, from (27) we know that minimizing $\mathcal{J}$ is equivalent to maximizing the sum of eigenvalues. This negative proportionality between $\mathcal{J}$ and the total sum of the eigenvalues, leads to an interesting observation. In Figure 5a, $\mathcal{J}\approx\mathcal{L}$ converges to $\approx$0.3. This indicates that the sum of the eigenvalues, also a measure of the total transmit power, of the eigenwaves produced by HOGMT−NN in these two cases is approximately 70% of the maximum possible value. Therefore, from an eigenwave multiplexing stand-point, this means that by transmitting signals with less power we incur a SNR penalty of $\approx 10\log_{10}(0.7) = -1.5$ dB at the receiver to achieve the same BER, compared to the ideal case of transmitting infinite eigenwaves. We believe that this small penalty is negligible compared to the gains obtained from higher throughput and model/receiver complexity (reduced processing, memory, and power), which would not have been possible by any other contemporary methods in a non-stationary LTV channel. We also find that for N=15 and 20, $\mathcal{L}$ is extremely small but it also increases the soft orthogonality measure, subsequently increasing the residual interference as shown in (41). This indicates that the choice of N provides a clear trade-off between penalties of SNR and residual interference.

B. Top-N Eigenwaves Multiplexing

To evaluate the accuracy of eigenwave multiplexing using the kernel in Section V-B. The throughput is calculated for B_w=20 Mhz and data-symbols are QPSK modulated. Fig-

Authorized licensed use limited to: UNIVERSITY AT ALBANY. Downloaded on June 20,2025 at 20:59:24 UTC from IEEE Xplore. Restrictions apply.

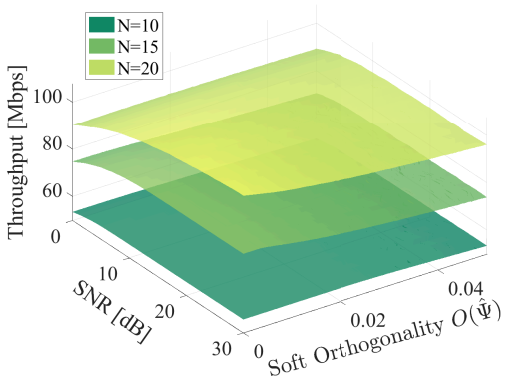


Figure 6: Throughput for N=10, 15, 20

ures 5d, 5h and 5l compare the BER for N=10, 15 and 20 with varying SNR and $O(\hat{\Psi})$. Note that though $\{\hat{\Phi}_n\}$ affect $\{\hat{\Psi}_n\}$ and have a same behavior, $\{\hat{\Psi}_n\}$ directly contributes to estimate symbols $\{\hat{s}_n\}$ as in (41). Therefore we use $O(\hat{\Psi})$ instead of $O(\hat{\Phi})$ as the reference. For both cases, the BER decreases with SNR but increases with $O(\hat{\Psi})$ because higher values of $O(\hat{\Psi})$ indicate more residual interference from other symbols. The ideal case for eigenwave multiplexing is the horizontal line at $O(\hat{\Psi})$=0, meaning there is no interference and the estimated symbol is the symbol scaled by the channel gain with AWGN as shown in (40). Meanwhile, with larger N, the BER performance is worse. For the same channel kernel, as the eigenwaves are descending ordered, the top-ranked eigenwaves have the higher eigenvalues. Large values of N will force the multiplexing to use lower-ranked eigenwaves with lower eigenvalues as well, which enhances the noise at the receiver and consequently increase the BER.

Figure 6 compares the throughput for N=10, 15 and 20. For each N, the throughput has the same trend as the BER, with respect to SNR and $O(\hat{\Psi})$. However, the throughput increases with a N, which has the opposite trend from BER. The reason being that by using more eigenwaves, more symbols are multiplexed and transmitted during the same time interval. Therefore, there is clear trade-off between the BER and the throughput with respect to N.

VII. Discussions

HOGMT−NN has to retrain for different N. Theoretically, N is adaptive for different scenarios. For example, the error rate for M-QAM modulated symbol transmitted over the eigen subchannel with channel gain σ_N is given as

$$Pr(M, \gamma_s, \sigma_N) \approx 4Q\left(\sigma_N \sqrt{\frac{3\gamma_s}{M-1}}\right) \tag{43}$$

where γ_s is the SNR. Then N is chosen as the largest integer satisfy the constraint $Pr(M, \gamma_s, \sigma_N) < \beta$, i.e.,

$$\sigma_N > \frac{Q^{-1}(\frac{\beta}{4})}{\sqrt{\frac{3\gamma_s}{M-1}}} \tag{44}$$

which means the error probability of the symbol that is transmitted over the N-th eigenwave should be less than the desired quality β. This is akin to adaptive modulation and coding in modern systems. This requires an *"adaptive NNs"* that will output eigenwaves until the eigenvalues are less than a threshold in (44) that is set based on γ_s, M and β.

VIII. Conclusion

In this paper, we formulate a generic multi-dimensional asymmetric channel kernel for MIMO, MIMO-OFDM and MIMO-OTFS channels and prove the optimality of HOGMT approximation for this generic kernel. Based on this, we design low-complexity, fully connected and explainable HOGMT−NN for decomposing the channel kernel into eigenwaves jointly orthogonal cross all the DoF. Then we proposed an eigenwvae multiplexing method, where symbols multiplexed by decomposed eigenwaves can independently transmit over the subchannels in the eigen-domain, thereby mitigating the interference across all the DoF. The performance with respect to the approximation error is analyzed and evaluated. The result shows that eigenwaves decomposed by HOGMT−NN can achieve around 0.001 soft orthogonality for N=10. Multiplexing using those achieves near-ideal BER and Throughput.

IX. Related Work

Communications in multi-dimensional channels: The input-output relations and analysis of MIMO-OTFS systems can be found in [31], [32], which can be transferred to the kernel expression as deduced in Section II-A3. The MIMO-OTFS channel estimation is investigated in [33]–[36]. These methods provide the CSI as the input to HOGMT−NN. The equalization techniques for MIMO-OTFS are discussed in [37]–[39]. However, those techniques treat the interference in the space domain and delay-Doppler domain separately, failing to achieve joint orthogonality. A jointly spatio-temporal precoding for multi-dimensional channels is presented in [12]. However, this technique utilizes eigenfunctions to construct the whole signal instead of transmitting symbols separately, which makes it highly energy-consuming and sensitive to CSI errors.

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 [40]. Eckart–Young–Mirsky Theorem shows that SVD is optimal for low-rank matrix approximation [41]. Nyström approximation [42] 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, [23] present black-box NNs for SVD decomposition, which is applicable for matrices. [43] 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.

Authorized licensed use limited to: UNIVERSITY AT ALBANY. Downloaded on June 20,2025 at 20:59:24 UTC from IEEE Xplore. Restrictions apply.

REFERENCES

[1] M. Vu and A. Paulraj, "MIMO wireless linear precoding," *IEEE Signal Processing Magazine*, vol. 24, no. 5, pp. 86–105, 2007.

[2] J. Armstrong, "Analysis of new and existing methods of reducing intercarrier interference due to carrier frequency offset in OFDM," *IEEE transactions on communications*, vol. 47, no. 3, pp. 365–369, 1999.

[3] A. Stamoulis, S. N. Diggavi, and N. Al-Dhahir, "Intercarrier interference in MIMO OFDM," *IEEE Transactions on signal processing*, vol. 50, no. 10, pp. 2451–2464, 2002.

[4] 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.

[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] Y. S. Cho, J. Kim, W. Y. Yang, and C. G. Kang, *MIMO-OFDM wireless communications with MATLAB*. John Wiley & Sons, 2010.

[8] P. Bello, "Characterization of Randomly Time-Variant Linear Channels," *IEEE Transactions on Communications Systems*, vol. 11, no. 4, pp. 360–393, 1963.

[9] L. A. Zadeh, "Frequency analysis of variable networks," *Proceedings of the IRE*, vol. 38, no. 3, pp. 291–299, 1950.

[10] F. Hlawatsch and G. Matz, *Wireless Communications Over Rapidly Time-Varying Channels*, 1st ed. USA: Academic Press, Inc., 2011.

[11] Z. Zou, M. Careem, A. Dutta, and N. Thawdar, "Unified Characterization and Precoding for Non-Stationary Channels," in *ICC 2022 - IEEE International Conference on Communications*, 2022, pp. 5140–5146.

[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] P. Almers, E. Bonek, A. Burr, N. Czink, M. Debbah, V. Degli-esposti, H. Hofstetter, P. Kyosti, D. Laurenson, G. Matz, A. F. Molisch, C. Oestges, and H. Ozcelik, *Survey of channel and radio propagation models for wireless MIMO systems*. EURASIP Journal on Wireless Communications and Net-working, 2007.

[14] Y. Hong, T. Thaj, and E. Viterbo, *Delay-Doppler Communications: Principles and Applications*. Academic Press, 2022.

[15] S. K. Mohammed, "Derivation of OTFS Modulation From First Principles," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 8, pp. 7619–7636, 2021.

[16] 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.

[17] Q. Hu, Y. Cai, Q. Shi, K. Xu, G. Yu, and Z. Ding, "Iterative algorithm induced deep-unfolding neural networks: Precoding design for multiuser MIMO systems," *IEEE Transactions on Wireless Communications*, vol. 20, no. 2, pp. 1394–1410, 2020.

[18] W. Ma, C. Qi, Z. Zhang, and J. Cheng, "Sparse channel estimation and hybrid precoding using deep learning for millimeter wave massive MIMO," *IEEE Transactions on Communications*, vol. 68, no. 5, pp. 2838–2849, 2020.

[19] T. Wang, C. Wen, H. Wang, F. Gao, T. Jiang, and S. Jin, "Deep learning for wireless physical layer: opportunities and challenges," *China Communications,vol.*, vol. 14, no. 11, pp. 92–111, 2017.

[20] H. He, S. Jin, C. Wen, F. Gao, G. Y. Li, and Z. Xu, "Model Driven Deep Learning for Physical Layer Communications," *IEEE Wireless Communications*, vol. 26, no. 5, pp. 77–83, 2019.

[21] H. Kim, Y. Jiang, R. Rana, S. Kannan, S. Oh, and P. Viswanath, "Communication algorithms via deep learning," *arXiv preprint arXiv:1805.09317*, 2018.

[22] X. Zhang, M. Vaezi, and T. J. O'Shea, "SVD-embedded deep autoencoder for MIMO communications," in *ICC 2022-IEEE International Conference on Communications*. IEEE, 2022, pp. 5190–5195.

[23] 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.

[24] J. Mercer, "Functions of Positive and Negative Type, and their Connection with the Theory of Integral Equations," *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, vol. 209, pp. 415–446, 1909. [Online]. Available: http://www.jstor.org/stable/91043

[25] 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

[26] A. A. Khuwaja, Y. Chen, N. Zhao, M.-S. Alouini, and P. Dobbins, "A survey of channel modeling for uav communications," 2018.

[27] C. Guo, F. Liu, S. Chen, C. Feng, and Z. Zeng, "Advances on Exploiting Polarization in Wireless Communications: Channels, Technologies, and Applications," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 1, pp. 125–166, 2017.

[28] A. Ali, E. De Carvalho, and R. W. Heath, "Linear receivers in non-stationary massive MIMO channels with visibility regions," *IEEE Wireless Communications Letters*, vol. 8, no. 3, pp. 885–888, 2019.

[29] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM review*, vol. 51, no. 3, pp. 455–500, 2009.

[30] G. H. Golub and C. F. van Loan, *Matrix Computations*, 4th ed. JHU Press, 2013. [Online]. Available: http://www.cs.cornell.edu/cv/GVL4/golubandvanloan.htm

[31] M. Kollengode Ramachandran and A. Chockalingam, "MIMO-OTFS in High-Doppler Fading Channels: Signal Detection and Channel Estimation," in *2018 IEEE Global Communications Conference (GLOBECOM)*, 2018, pp. 206–212.

[32] A. RezazadehReyhani, A. Farhang, M. Ji, R. R. Chen, and B. Farhang-Boroujeny, "Analysis of Discrete-Time MIMO OFDM-Based Orthogonal Time Frequency Space Modulation," in *2018 IEEE International Conference on Communications (ICC)*, 2018, pp. 1–6.

[33] S. Srivastava, R. K. Singh, A. K. Jagannatham, and L. Hanzo, "Delay-Doppler and angular domain 4D-sparse CSI estimation in OTFS aided MIMO systems," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 12, pp. 13 447–13 452, 2022.

[34] R. Bomfin, M. Chafii, A. Nimr, and G. Fettweis, "Channel estimation for MIMO space time coded OTFS under doubly selective channels," in *2021 IEEE International Conference on Communications Workshops (ICC Workshops)*. IEEE, 2021, pp. 1–6.

[35] S. Srivastava, R. K. Singh, A. K. Jagannatham, A. Chockalingam, and L. Hanzo, "OTFS transceiver design and sparse doubly-selective CSI estimation in analog and hybrid beamforming aided mmWave MIMO systems," *IEEE Transactions on Wireless Communications*, vol. 21, no. 12, pp. 10 902–10 917, 2022.

[36] D. Shi, W. Wang, L. You, X. Song, Y. Hong, X. Gao, and G. Fettweis, "Deterministic pilot design and channel estimation for downlink massive MIMO-OTFS systems in presence of the fractional Doppler," *IEEE Transactions on Wireless Communications*, vol. 20, no. 11, pp. 7151–7165, 2021.

[37] H. Qu, G. Liu, M. A. Imran, S. Wen, and L. Zhang, "Efficient channel equalization and symbol detection for MIMO OTFS systems," *IEEE Transactions on Wireless Communications*, vol. 21, no. 8, pp. 6672–6686, 2022.

[38] B. C. Pandey, S. K. Mohammed, P. Raviteja, Y. Hong, and E. Viterbo, "Low complexity precoding and detection in multi-user massive MIMO OTFS downlink," *IEEE transactions on vehicular technology*, vol. 70, no. 5, pp. 4389–4405, 2021.

[39] B. Cao, Z. Xiang, and P. Ren, "Low complexity transmitter precoding for MU MIMO-OTFS," *Digital Signal Processing*, vol. 115, p. 103083, 2021.

[40] 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.

[41] G. W. Stewart, *Matrix Algorithms: Volume II: Eigensystems*. SIAM, 2001.

[42] 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

[43] 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.

Authorized licensed use limited to: UNIVERSITY AT ALBANY. Downloaded on June 20,2025 at 20:59:24 UTC from IEEE Xplore. Restrictions apply.