

Generative and Explainable AI for High-Dimensional MIMO-OFDM Channel Estimation in Time-Frequency-Space Domain

Nghia Thinh Nguyen and Tri Nhu Do

Abstract

In this paper, we address a high-dimensional (HD) geometry-based channel model (GCM) characterized in the time-frequency-space (TFS) domain for Multiple-Input Multiple-Output Orthogonal Frequency-Division Multiplexing (MIMO-OFDM) transmission systems. Specifically, HD-TFS data tensors, collected via realistic ray-tracing channel models, are sensitive to network geometric characteristic changes, i.e., varying rays and clusters in multipath, due to user mobility. Therefore, we propose an adversarial training framework where the deep learning channel generator learns an accurate variational posterior distribution (VPD) that is robust to geometry changes of the channel model. By extracting features from geometric characteristics, a novel loss function is introduced that incorporates the third moment of HD-TFS tensors. Besides, a stochastic gradient estimator based on the reparameterization trick is introduced, leveraging the distribution of channel characteristics to enhance variational posterior distribution (VPD) learning. The numerical results on realistic scenarios show that the learned VPD is not only accurate, with superior HD-TFS channel estimation in low signal-to-noise ratio (SNR) regimes, but also improves runtime efficiency through an optimized channel generator architecture design. Finally, an explanation mechanism is proposed to highlight regions of high channel gain in the HD-TFS tensor as the attention areas during model training.

Index Terms

5G NR, 3GPP, CDL, Channel estimation, MIMO, OFDM, Generative AI, GAN, VAE, Explainable AI

I. INTRODUCTION

Accurate channel estimation is important for reliable wireless communication, particularly in 5G New Radio (NR) systems operating in complex environments with *multipath propagation* and high

N. T. Nguyen and T. N. Do are with the Department of Electrical Engineering, Polytechnique Montréal, Montreal, Quebec, Canada.
Emails: {nghia-thinh.nguyen, tri-nhu.do}@polymtl.ca

This paper was presented in part at the IEEE International Conference on Communications (ICC), Montreal, Canada, May 2025 [1].

mobility. A geometry-based channel model (GCM), such as the Clustered Delay Line (CDL) model, simulates realistic wireless propagation by incorporating spatial and temporal characteristics. For example, the 3GPP (3rd Generation Partnership Project standards) CDL model (part of TR 38.901) provides predefined profiles (e.g., CDL-A for indoor, CDL-B for urban) that can be adapted for different environments by tuning cluster delays, powers, and angular distributions. These models are essential for evaluating system performance in applications like IoT networks and urban microcell scenarios.

The rapid advent of 5G and future-generation networks has been driven by the necessity to transmit high-dimensional data in increasingly complex channel environments. Many techniques [2], [3] have been developed to enhance data transmission capacity. In this context, the Orthogonal Frequency-Division Multiplexing (OFDM) technique [4] was introduced to support high-speed networks because of its ability to combat multipath fading and enable high data rates. The pilot symbols [5] facilitate channel estimation, enabling reliable communication under dynamic channel conditions. By applying the Least Squares (LS) [6] conventional channel estimation method, with interpolation methods [7] [8], the data transmission capacity is reliably maintained.

However, the complexity of modern wireless environments, including high mobility and multipath propagation, led to high computational costs and limited adaptability. Recently, many Artificial Intelligence (AI) techniques were proposed [9] to address these challenges. Researchers integrated realistic open-source simulation frameworks such as Sionna [10], [11] aiming to reduce the gap between theoretical research and practical implementation. AI-based methods, .i.e, DNN-based [12], GAN-based [13], or [14] VAE-based models, offered promising solutions to address the limitations of conventional techniques [6]. Especially, Generative AIs [1] applied the solution for domain generalization adapting to varying environmental conditions of the geometry-based channel model. However, these methods were integrated after the interpolation process, increased computational cost, and are unsuitable for real-time applications.

Geometry-based channel modeling and the estimation of such channels have been well established and are recently being applied in practical wireless measurements. In particular, [15], [16] focused on a typical number of randomly selected geometric objects. Integral geometry gives a unified approach for defining integrands over the rays, clusters, and angles, and these parameters follow several typical distributions. Compared with deterministic models [17] and statistical models [18], the GCM requires lower computations and possesses better generality, particularly in IoT applications [19]. However, these GCMs often assume a limited set of clusters and rays, neglecting the explosive growth of

parameter space when jointly modeling time, frequency, and space. This leads to prohibitive computational complexity and storage requirements as the total number of rays and clusters increases in each dimension [20]. Although deep generative models have been used to reduce pilot overhead in channel estimation [21], their integration within GCM frameworks to prune ray or cluster sets across TFS domains has not been explored. Moreover, though the Doppler shift induced by mobility has been considered [22], more complex outdoor scenarios, such as urban environments, should also be taken into account.

Our Research Problem on Generative AI for GCM Estimation: Geometry-based channel models, as standardized in 3GPP, are widely adopted in practice for wireless communication systems. These models rely on practical measurements of the physical environment, including the location, shape, and material properties of obstacles, as well as geographical features. In such models [17], [18], statistical parameters, such as the number of rays, number of clusters, delay, and angle profiles, are derived from specific measurements based on the geometry of the environment. However, these parameters become invalid when the environmental geometry changes, necessitating updated measurements to redefine the profiles, which are randomly arisen due to factors such as the movement of user terminals or the appearance of new obstacles. These dynamic changes challenge the accuracy and applicability of static channel models. Despite the progress of the Generative AIs for wireless channel modeling [23], [13], the integration of the GCM channel characteristics from the TFS domain into GAN learning, and into the VAE latent space remains underexplored.

Consider the fusing statistical properties [24]; there is potential to integrate geometry-based statistical descriptors of multipath propagation, such as the ray delays, Doppler shifts, and angular spreads, directly into the feature fusion layers of a VAE generator model. Leveraging the consistency between the learned latent distribution and the true TFS channel structure, the fusion of statistical properties enhancing latent space can improve the accuracy of channel realizations. However, the black-box nature of channel estimation of generator model causes difficulty in explaining how the model impacts or preserves geometry-based features. Reusing the more realistic channel estimation of GAN model, the VAE-GAN-based method [25], with VAE as generator, the proposal to incorporate geometry-based channel features into the attention mechanism of the GAN's critic network allows the generator model to adaptively focus on regions of the channel gain presentation that are physically meaningful.

In this paper, we propose a novel generative framework [1], which leverages the strengths of a VAE generator and GAN-based training strategy to address geometric-based channel scenarios, i.e., the CDL-B model. The proposal aims to reduce the gap in Generative AIs estimation models for

high-dimensional GCMs characterized in the TFS domain over MIMO-OFDM transmission systems. Based on the Sionna framework, we can develop and evaluate our model on realistic simulation data compliant with 3GPP 38.901 standards for 5G New Radio (NR). The key contributions of this article can be summarized as follows:

- We propose an end-to-end pipeline for OFDM transmissions over the CDL channel model using a generative channel estimation model¹. Our approach addresses highly challenging scenarios, including high user mobility, elevated noise levels, and varying geometry-based channel configurations. At the end of the pipeline, we employed a conventional evaluation metric, Bit Error Rate (BER), to assess performance in a realistic environment at Polytechnique Montréal.
- A novel framework for adversarial training is proposed with an innovative loss function that integrates the third-order moment of the channel impulse response (CIR) with reconstruction loss, thereby capturing higher-order statistics of the GCM channel characteristics. A stochastic gradient estimator based on the reparameterization trick is introduced in the generative model, leveraging the distribution of GCM features to enhance variational posterior distribution (VPD) learning. Notably, our estimation framework does not require pilot knowledges.
- An efficient VAE generative model, VAENet, is designed to optimize the computational complexity of the channel estimation process. The proposed VAENet demonstrates superior runtime efficiency compared to state-of-the-art (SOTA) conventional methods, i.e., the LS estimator combined with the Linear Minimum Mean Square Error (LMMSE) interpolator. Moreover, the proposed approach is adaptable to various HD-TFS designs in MIMO-OFDM transmissions, enhancing its flexibility for real-world deployment.
- We introduce a new explainable AI method using activation mapping for the critic network (CriticNet), termed AM4C. We observe a significant alignment between the critic's attention and the high-power gain regions in the HD-TFS data tensors, providing insights into the design of the generator's loss, pilot patterns, and other loss functions for future research.

The remainder of the paper is structured Section II outline of system model, Section III characterize mathematically the channel estimation formulation, Section IV propose our generative channel estimation framework, Section V statistic the numerical results, and conclusion in Section VI.

Notations. *Underlined notations* denote random variables, such as a random scalar $\underline{\alpha}$, a random vector \underline{x} , or a random matrix \underline{X} , while *non-underlined notations* represent realizations of these random variables, e.g., α , x , and X .

¹The source code of this paper is available at <https://github.com/TND-Lab/High-Dimensional-MIMO-OFDM-Channel-Estimation>

II. SYSTEM MODEL

A. System Components and Network Topology

In an urban microcell NLOS environment, the CDL-B NLOS configuration captures the effects of significant multipath propagation caused by reflections, diffractions, and scattering from buildings, vehicles, and other urban structures, as shown in Fig. 1. Consider an uplink system with a single user terminal (UT), where $N_{\text{ut}} = 1$, equipped with $K_{\text{ut,ant}}$ antennas, communicating with a base station (BS) with $N_{\text{bs}} = 1$, which is equipped with $K_{\text{bs,ant}}$ antennas. The spherical coordinates of the UT and BS are given by $\mathbf{u}_{\text{ut,bs}} = [d_{\text{ut,bs}}, \theta_{\text{ut,bs}}, \phi_{\text{ut,bs}}]$, where $d_{\text{ut,bs}}$ is the radial distance, $\theta_{\text{ut,bs}}$ is the zenith angle of arrival along the z -axis, and $\phi_{\text{ut,bs}}$ is the azimuth angle in the horizontal x - y plane. The UT antennas and BS antennas follow standard pattern [26, Eq. (7.1-11)], so the response vectors of each antenna can be presented as $\mathbf{a}_{\text{node}}(\theta, \phi) = [F_{\text{node},\theta} \quad F_{\text{node},\phi}]^T \odot \exp(j2\pi \frac{\mathbf{r}^T \mathbf{d}_{\text{node}}}{\lambda_0})$, where $\text{node} \in \{\text{ut, bs}\}$, $\theta \in \{\theta_{\text{ZOD}}, \theta_{\text{ZOA}}\}$ are zenith angles of departure and arrival, $\phi \in \{\phi_{\text{AOD}}, \phi_{\text{AOA}}\}$ are azimuth angles of departure and arrival, respectively. $F_{\text{node},\theta}$, $F_{\text{node},\phi}$ are the field components in the direction of the zenith angle θ , and the angle ϕ of the unit antenna, respectively. The phase shift component $\mathbf{r} \in \{\mathbf{r}_{\text{ut}}, \mathbf{r}_{\text{bs}}\}$ is the spherical unit vector of \mathbf{d}_{node} , which is the location vector of the antenna element. λ_0 is the wavelength of the carrier frequency. Both UT and BS antenna settings are equipped with a uniform planar array (UPA), i.e., a uniform rectangular panel array [10], of $K_{\text{ut,ant}}$, and $K_{\text{bs,ant}}$ cross-polarized antennas (pair antennas: red-black color) as shown in Fig 1. The UT array response vector is $\mathbf{a}_{\text{ut}}(\theta, \phi) = [\mathbf{a}_{\text{ut},1}(\theta, \phi), \dots, \mathbf{a}_{\text{ut},K_{\text{ut,ant}}}(\theta, \phi)]$ and the BS array response vector is $\mathbf{a}_{\text{bs}}(\theta, \phi) = [\mathbf{a}_{\text{bs},1}(\theta, \phi), \dots, \mathbf{a}_{\text{bs},K_{\text{bs,ant}}}(\theta, \phi)]$.

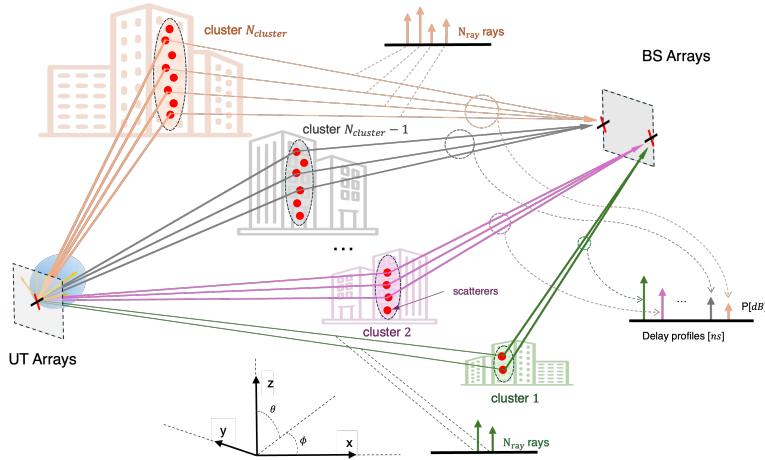


Fig. 1. Scenario of OFDM transmissions over the CDL-B channel model. The UT has $N_{\text{ut,ant}} = 2$ antennas, and the BS has $N_{\text{bs,ant}} = 4$. Multipath propagation into buildings is characterized by N_{cluster} clusters and N_{ray} rays per cluster.

B. Signal Modeling

1) *Binary Source*: The initial data transmission is the binary sequences. Let \mathbf{b}_{tx} denote the binary sequence of the spatial stream of the UT, with $\mathbf{b}_{\text{tx}} = [b_{\text{tx}}^1, b_{\text{tx}}^2, \dots, b_{\text{tx}}^{N_{\text{ut}} \times N_{\text{stream}} \times N_{\text{bit}}}]$, where $b_{\text{tx}}^i \in \{0, 1\}$ and N_{bit} is the number of bits. Each spatial stream represents an independent data transmission path. Thus, the number of streams is limited by the number of antennas at both the transmitter (UT) and the receiver (BS), i.e., $N_{\text{stream}} = \min(K_{\text{ut,ant}}, K_{\text{bs,ant}})$. In this system, the binary data transmission is given by $\mathbf{X}_{\text{bit}} \in \{0, 1\}^{N_{\text{ut}} \times N_{\text{stream}} \times N_{\text{bit}}}$, where $b_{\text{tx}}^i \in \mathbf{X}_{\text{bit}}$.

2) *Parity-Check Codes*: To enhance the reliability of data transmission over complex channels, we applied the Low-Density Parity-Check (LDPC) process that adds parity bits to the binary data transmission. The redundancy bits are used to correct errors. Define $N_{\text{ratio}} = \frac{N_{\text{bit}}}{N'_{\text{bit}}}$ with block code $(N'_{\text{bit}}, N_{\text{bit}})$, where N'_{bit} is the length of the binary data after adding redundant bits, e.g., $N_{\text{ratio}} = \frac{576}{1134} = \frac{1}{2}$ [10], and therefore $N'_{\text{bit}} = \frac{N_{\text{bit}}}{N_{\text{ratio}}}$. Consider the LDPC function $f_{\text{enc,ldpc}}(\mathbf{X}_{\text{bit}}, N_{\text{ratio}})$. The real transmit data \mathbf{X}'_{bit} can be presented as $\mathbf{X}'_{\text{bit}} = f_{\text{enc,ldpc}}(\mathbf{X}_{\text{bit}}, N_{\text{ratio}})$, where $\mathbf{X}'_{\text{bit}} \in \{0, 1\}^{N_{\text{ut}} \times N_{\text{stream}} \times N'_{\text{bit}}}$.

3) *Quadrature Amplitude Modulation*: To adapt the efficiency of the communication system, the binary data transmission has to be converted into more complex signal transmission. Consider quadrature amplitude modulation (QAM). The bit data of \mathbf{X}'_{bit} will be mapped to the complex QAM symbol $p_i = A_i^I + jA_i^Q$, where A_i^I is the in-phase amplitude and A_i^Q is the quadrature amplitude. With $M = 4$, the number of bits converted to a QAM symbol is $\log_2 M = 2$, and the value of the quadrature, based on bit information, can be presented as $00 \rightarrow \frac{1}{\sqrt{2}} + \frac{1}{\sqrt{2}}j$, $01 \rightarrow -\frac{1}{\sqrt{2}} + \frac{1}{\sqrt{2}}j$, $11 \rightarrow -\frac{1}{\sqrt{2}} - \frac{1}{\sqrt{2}}j$, $10 \rightarrow \frac{1}{\sqrt{2}} - \frac{1}{\sqrt{2}}j$. Let N_{sym} denote the number of QAM symbols, so $N_{\text{sym}} = \frac{N'_{\text{bit}}}{\log_2 M}$. Based on this definition, the transmission symbol sequences can be presented as $\mathbf{X}_{\text{qam}} \in \mathbb{C}^{N_{\text{ut}} \times N_{\text{stream}} \times N_{\text{sym}}}$. In the continuous time domain, the QAM symbol can be presented as $x_{\text{sym}}(t)$. Consider the baseband signal with a pulse-shaping filter function $f_{\text{pulse}}(t)$. We can form the symbols in baseband as $x_{\text{sym,b}}(t) = \sum_{i=0}^{N_{\text{sym}}-1} p_i f_{\text{pulse}}(t)$. Thus, the passband QAM symbols can be expressed as

$$x_{\text{sym}}(t) = \mathcal{R}\{x_{\text{sym,b}} e^{2\pi f_c t}\} = \sum_{i=0}^{N_{\text{sym}}-1} A_i^I f_{\text{sym}}(t) \cos(2\pi f_c t) + A_i^Q f_{\text{sym}}(t) \sin(2\pi f_c t). \quad (1)$$

4) *Time-Frequency Representation of OFDM Resource Grids*: In our scenario, the transmitted signal \mathbf{X}_{qam} propagates through multiple optical paths, including reflections and scatterings from buildings, before arriving at the base station. To improve the efficiency and resilience of the signal, the Orthogonal Frequency Division Multiplexing (OFDM) symbol is applied, which can be formed into an OFDM resource grid $\mathbf{X}_{\text{rg}}^{\text{ofdm}}$. The objective is to map the transmitted symbols \mathbf{X}_{qam} to the resource grid matrix $\mathbf{X}_{\text{rg}}^{\text{ofdm}}$. The resource grid matrix is a two-dimensional matrix, where N_{sub} is the number of subcarriers

representing the frequency domain, and $N_{\text{sym}}^{\text{ofdm}}$ is the number of OFDM symbols representing the time domain. The resource design follows the structure of the 5G frame, according to the 3GPP RP-38.802 and 38.804 specifications, with numerology $\mu = 0$, $N_{\text{sym}}^{\text{ofdm}} = 14$ time symbols, and the duration of each OFDM symbol satisfying $T_{\text{sym}}^{\text{ofdm}} \times N_{\text{sym}}^{\text{ofdm}} < T_{\text{slot}}$. Define the number of orthogonal sinusoid blocks as $N_{\text{sub}}^{\text{block}} = k \times N_{\text{sub}}$, where k is the number of subcarriers, and $N_{\text{sub}} = 12$. The subcarrier spacing is given by $\Delta f = 15$ kHz, N_{guards} is the number of guard subcarriers, and N_{dc} denotes the Direct Current (DC) subcarrier null. The cyclic prefix N_{cp} is a guard interval added to the beginning of each OFDM symbol in the time domain. In addition, the pilot symbols $N_{\text{pilot}} < N_{\text{sym}}^{\text{ofdm}}$, following the Kronecker pattern, are interleaved among the OFDM symbols. In our settings, the OFDM symbols are transmitted within one time slot, which means $\mathbf{X}_{\text{rg}}^{\text{ofdm}} \in \mathbb{C}^{N_{\text{ut}} \times N_{\text{stream}} \times N_{\text{sub}}^{\text{block}} \times N_{\text{sym}}^{\text{ofdm}}}$, where $N_{\text{sub}}^{\text{eff}} = N_{\text{sub}}^{\text{block}} - N_{\text{dc}} - N_{\text{guards}}$ denotes the number of effective subcarriers.

5) *Time-Domain Waveform via OFDM Modulation:* The resource grids $\mathbf{X}_{\text{rg}}^{\text{ofdm}}$ containing data are transmitted in the time domain through OFDM modulation. For a given OFDM symbol $n < N_{\text{sym}}^{\text{ofdm}}$, the number of subcarriers considered in the symbol is $N_{\text{sub}}^{\text{ofdm}} = N_{\text{sub}}^{\text{block}} + N_{\text{cp}}$. Because the transmit data has $N_{\text{sym}}^{\text{ofdm}}$ time symbols, the number of time steps in a time slot is $N_{\text{slot}}^{\text{ofdm}} = N_{\text{sub}}^{\text{ofdm}} \times N_{\text{sym}}^{\text{ofdm}}$. Applying the Inverse Fast Fourier Transform (IFFT), the discrete signal at time sample m of OFDM symbol n , $x_{\text{time}}^{\text{ofdm}}[n, m]$, in the time domain can be presented as

$$x_{\text{time}}^{\text{ofdm}}[n, m] = \sum_{k=0}^{N_{\text{sub}}^{\text{ofdm}}-1} x_{\text{rg}}^{\text{ofdm}}[n, k] e^{j2\pi km/N_{\text{sub}}^{\text{ofdm}}}, \quad (2)$$

where $n \in \{1, 2, \dots, N_{\text{sym}}^{\text{ofdm}}\}$, $m \in \{1, 2, \dots, N_{\text{sub}}^{\text{ofdm}}\}$, and $x_{\text{rg}}^{\text{ofdm}}[n, k]$ is the data symbol at carrier k of the OFDM symbol n in the resource grid matrix $\mathbf{X}_{\text{rg}}^{\text{ofdm}}$. The duration of the cyclic prefix symbol is $T_{\text{cp}}^{\text{ofdm}} = \frac{N_{\text{cp}}}{N_{\text{sub}}^{\text{block}}} \times T_{\text{sym}}^{\text{ofdm,eff}} \geq \tau_{\max}$, where $T_{\text{sym}}^{\text{ofdm,eff}} = \frac{1}{\Delta f}$ is the duration of the effective OFDM symbol, τ_{\max} is the maximum delay spread, and $T_{\text{sym}}^{\text{ofdm}} = T_{\text{cp}}^{\text{ofdm}} + T_{\text{sym}}^{\text{ofdm,eff}}$. According to (2), the OFDM transmit data in the time domain is presented as $\mathbf{X}_{\text{time}}^{\text{ofdm}} \in \mathbb{C}^{N_{\text{ut}} \times N_{\text{stream}} \times N_{\text{slot}}^{\text{ofdm}}}$.

C. Geometry-Based Cluster Delay Line Channel Modeling

We consider communication between the UT and BS using the cluster delay line (CDL) channel model, particularly in an urban microcell non-line-of-sight (NLOS) environment. The CDL-B model is selected as the primary channel model, with default settings of $N_{\text{cluster}} = 23$ delayed clusters and $N_{\text{ray}} = 20$ rays per cluster. The framework can be applied to any geometry-based channel model, and the default settings can be adjusted based on the model design requirements, as described in [26, Eq. (7.5-28), (7.5-30)]. $h_{u,s,n,m}(t)$ denotes the channel for the transmit antenna element $s \in [1, K_{\text{ut,ant}}]$,

the receive antenna element $u \in [1, K_{\text{bs,ant}}]$, for ray $m \in [1, N_{\text{ray}}]$ of cluster $n \in [1, N_{\text{cluster}}]$. Applying the network topology following Eq. 7.5-28 in the 3GPP 38.901 pattern, it can be presented as

$$h_{u,s,n,m}(t) = \sqrt{\frac{P_n}{N_{\text{ray}}}} \mathbf{a}_{\text{ut},s,n,m}(\theta_{\text{ZOD}}, \phi_{\text{AOD}}) \mathbf{a}_{\text{bs},u,n,m}(\theta_{\text{ZOA}}, \phi_{\text{AOA}}) \Phi_{n,m} e^{j2\pi \frac{\mathbf{r}_{\text{rx},n,m}^T \mathbf{v}_{\text{ut}}}{\lambda_0} t}, \quad (3)$$

where $\sqrt{\frac{P_n}{N_{\text{ray}}}}$ is the power normalization factor of each ray in the cluster, and $\exp(j2\pi \frac{\mathbf{r}_{\text{rx},n,m}^T \mathbf{v}_{\text{ut}}}{\lambda_0} t)$ is the Doppler frequency component. Let $\mathbf{v}_{\text{ut}} = v_{\text{ut}} \times [\sin(\theta_{v_{\text{ut}}}) \cos(\phi_{v_{\text{ut}}}), \sin(\theta_{v_{\text{ut}}}) \sin(\phi_{v_{\text{ut}}}), \cos(\theta_{v_{\text{ut}}})]^T$ denotes UT velocity, with speed v_{ut} at the travel azimuth angle $\phi_{v_{\text{ut}}}$ and elevation angle $\theta_{v_{\text{ut}}}$. $\mathbf{r}_{\text{rx},n,m}$ is the spherical unit vector, and the polarization coupling matrix $\Phi_{n,m}$ has values depending on angles $\phi_{n,m}$ and $\theta_{n,m}$. Because the power of each cluster P_n differs, weaker clusters have less impact on the model, resulting in lower channel power gain values as shown in Fig. 2. The impact of the proposed explanation mechanism primarily focuses on the high-power components, as illustrated in Fig. 6b.

Remark 1 (Complex-Valued Random Characteristics). *According to the default settings of the CDL-B channel model, the azimuth and elevation angles ϕ, θ are randomly selected from a set of predefined values, which depend on the number of rays N_{ray} and clusters N_{cluster} . The vector \mathbf{v}_{ut} is random because its angles $\phi_{v_{\text{ut}}}$ and $\theta_{v_{\text{ut}}}$ are random. Consequently, $h_{u,s,n,m}(t)$ is characterized as a complex-valued random variable.*

The complex-valued path coefficient $h_{u,s,n}(t)$ is the sum of the N_{ray} rays in each cluster, given by

$$h_{u,s,n}(t) = \sqrt{\frac{P_n}{N_{\text{ray}}}} \sum_{m=1}^{N_{\text{ray}}} \mathbf{a}_{\text{ut},s,n,m}(\theta_{\text{ZOD}}, \phi_{\text{AOD}}) \mathbf{a}_{\text{bs},u,n,m}(\theta_{\text{ZOA}}, \phi_{\text{AOA}}) \Phi_{n,m} e^{j2\pi \frac{\mathbf{r}_{\text{rx},n,m}^T \mathbf{v}_{\text{ut}}}{\lambda_0} t}. \quad (4)$$

From (3) and (4), the channel impulse response $h_{u,s}(t, \tau)$ of the antenna pair (u, s) is²

$$h_{u,s}(t, \tau) = \sum_{n=1}^{N_{\text{cluster}}-1} h_{u,s,n}(t) \text{sinc}(B(t - \tau_n)), \quad (5)$$

It is noteworthy that $h_{u,s,n}(t)$ in (5) corresponds to a_n in [11, Eq. (2)], but is more complex due to the inclusion of rays within clusters. Let L_{tot} denote the number of significant multipath components in a discrete-time model of the channel, i.e., the number of channel taps, where $L_{\text{tot}} = L_{\text{max}} - L_{\text{min}} + 1$, $L_{\text{max}} = B \times \tau_{\text{max}}$, L_{min} is constant, and B is the bandwidth, with $l \in \{L_{\text{min}}, \dots, L_{\text{max}}\}$. Each tap

²The relationship between the use of $\text{sinc}(t - \tau_n)$ and $\delta(t - \tau_n)$ can be elucidated as follows: The Dirac delta function $\delta(t - \tau_n)$ represents idealized multipath arrivals with discrete delays, as employed in theoretical formulations such as [26, Eq (7.5-28)], [11, Eqs. (2) and (101)]. However, this representation is physically unrealistic due to its lack of bandwidth limitation. In contrast, practical implementations, such as those in the Sionna codebase, utilize $\text{sinc}(t - \tau_n)$ to model band-limited pulse shaping, which, while conceptually equivalent, accounts for realistic bandwidth constraints.

corresponds to a delayed version of the transmitted signal arriving due to multipath propagation. The data symbols are observed up to time step $N_{\text{slot},\text{ext}}^{\text{ofdm}} = N_{\text{slot}}^{\text{ofdm}} + L_{\text{tot}} - 1$. Let the channel gain be denoted by $\mathbf{H}_{\text{gain}} \in \mathbb{C}^{N_{\text{ut}} \times K_{\text{ut,ant}} \times N_{\text{bs}} \times K_{\text{bs,ant}} \times N_{\text{cluster}} \times N_{\text{slot},\text{ext}}^{\text{ofdm}}}$, and let $h_{u,s,n}(t) \in \mathbf{H}_{\text{gain}}^{\text{time}}$. According to (5), the channel taps in the time domain can be formed as the discrete complex-baseband representation of the channel from the channel impulse response $h_{u,s}[k, l]$ at time step $k \in [1, N_{\text{slot}}^{\text{ofdm}}]$ and tap l , is

$$h_{u,s}[k, l] = \sum_{n=0}^{N_{\text{cluster}}-1} h_{u,s,n}(k) \text{sinc}(l - B \times \tau_n), \quad (6)$$

where $l \in [1, N_{\text{tot}}]$, $h_{u,s}[k, l] \in \mathbf{H}_{\text{time}}$, the channel representation in the time-space domain, with $\mathbf{H}_{\text{time}} \in \mathbb{C}^{N_{\text{ut}} \times K_{\text{ut,ant}} \times N_{\text{bs}} \times K_{\text{bs,ant}} \times N_{\text{slot},\text{ext}}^{\text{ofdm}} \times N_{\text{tot}}}$.

In OFDM channel estimation applications, the channel impulse response is estimated in the time-frequency domain. Let $h_{u,s}(t, f)$ denote the channel impulse response in the time-frequency domain for the transmit-receive antenna pair (s, u) . It can be expressed as

$$h_{u,s}(t, f) = \int_{-\infty}^{\infty} h_{u,s}(t, \tau) e^{-2\pi f \tau} d\tau = \sum_{n=1}^{N_{\text{cluster}}-1} h_{u,s,n}(t) \frac{1}{B} \text{rect}\left(\frac{f}{B}\right) e^{-2\pi f \tau_n}, \quad (7)$$

where $h_{u,s}(t, f) \in \mathbf{H}_{\text{freq}}$, and $\mathbf{H}_{\text{freq}} \in \mathbb{C}^{N_{\text{ut}} \times K_{\text{ut,ant}} \times N_{\text{bs}} \times K_{\text{bs,ant}} \times N_{\text{sym}}^{\text{ofdm}} \times N_{\text{sub}}^{\text{eff}}}$ is the channel in the TFS domain³. From (7), the channel is impacted by the geometric characteristics, i.e., ϕ , θ , N_{ray} , N_{cluster} , and \mathbf{v}_{ut} . Adapting the pattern from [26, Eqs. (7.4-1), (7.5-22), (7.5-29)], the CDL-B channel model is independent of the distance between the UT and BS, allowing the $d_{\text{ut,bs}}$ geometric parameter to be ignored. In the CDL-B model, the channel characteristics are defined by [26, Eqs. (7.5-22), (7.5-28)], corresponding to an NLOS scenario. Although [26, Eq. (7.5-29)] includes the distance between the UT and BS as defined in [26, Eq. (7.4-1)], this distance term only affects LOS components. The distance $d_{\text{ut,bs}}$ does not significantly affect the CDL-B characteristics, despite the path loss being a deterministic function of distance. In contrast, geometric characteristics have a significant impact.

D. Received Observation and Signal Modeling in Different Considered Domains

1) *Discrete Time-Domain Representation:* Based on the channel model and signal modeling, let $r_{\text{time}}[k]$ denote the received signal in the discrete time domain at the time symbol $k \in \{1, 2, \dots, N_{\text{slot},\text{ext}}^{\text{ofdm}}\}$ for the transmit antenna s and receive antenna u . It can be expressed as

$$r_{\text{time},u,s}[k] = \sum_{l=0}^{N_{\text{tot}}-1} h_{u,s}[k, l] x_{\text{time},u,s}^{\text{ofdm}}[k-l] + n_{u,s}[k], \quad (8)$$

³In the simulation, $N_{\text{ut}} = 1$, $N_{\text{bs}} = 1$ do not affect the input dimensions of our proposed estimated channel. However, they must be considered as part of the input dimensions when using the Sionna library [10], [11].

where $x_{\text{time},u,s}^{\text{ofdm}}[k-l] \in \mathbf{X}_{\text{time}}^{\text{ofdm}}$ is the transmitted symbol at time step $k-l$, and $n_{u,s}[k] \sim \mathcal{CN}(0, 1)$ is additive white Gaussian noise (AWGN) at time step k . The received matrix with multiple antennas, $\mathbf{R}_{\text{time}} \in \mathbb{C}^{N_{\text{bs}} \times K_{\text{bs,ant}} \times N_{\text{slot,ext}}^{\text{ofdm}}}$, where $r_{\text{time},u,s}[k] \in \underline{\mathbf{R}}_{\text{time}}$, at the BS, is expressed as

$$\underline{\mathbf{R}}_{\text{time}} = \underline{\mathbf{H}}_{\text{time}} \underline{\mathbf{X}}_{\text{time}}^{\text{ofdm}} + \underline{\mathbf{N}}, \quad (9)$$

where $\underline{\mathbf{N}} \in \mathbb{C}^{N_{\text{bs}} \times K_{\text{bs,ant}} \times N_{\text{slot,ext}}^{\text{ofdm}}}$ is an i.i.d. matrix of AWGN with entries $[\underline{\mathbf{N}}]_{ij} \sim \mathcal{CN}(0, 1)$.

2) *Time-Frequency Domain Representation via OFDM Demodulation*: From the observation data \mathbf{R}_{time} , the OFDM demodulation process converts the received data in the time domain to the frequency domain representation by the Fast Fourier Transform (FFT). Let $r_{\text{freq}}^{\text{ofdm}}[n, k]$ be the received symbol in the time-frequency (TF) domain at carrier k of time symbol n , which can be expressed as

$$r_{\text{freq}}^{\text{ofdm}}[n, k] = \sum_{m=0}^{N_{\text{sub}}^{\text{block}}-1} r_{\text{time}}[n, m] e^{-j2\pi km/N_{\text{sub}}^{\text{block}}}, \quad (10)$$

where $k \in \{1, 2, \dots, N_{\text{sub}}^{\text{block}}\}$, $n \in \{1, 2, \dots, N_{\text{sym}}^{\text{ofdm}}\}$, and $r_{\text{time}}[n, m]$ is the received symbol at time sample m of OFDM symbol n . According to (10), the received matrix in the frequency domain can be considered as $\mathbf{R}_{\text{freq}}^{\text{ofdm}} \in \mathbb{C}^{N_{\text{bs}} \times K_{\text{bs,ant}} \times N_{\text{sym}}^{\text{ofdm}} \times N_{\text{sub}}^{\text{block}}}$.

III. PROPOSED FRAMEWORK FOR STATISTICAL PROPERTY FUSION IN CHANNEL ESTIMATION

A. Channel Estimation Framework for the Considered MIMO-OFDM System

Consider the pilot information in the received signal. The Least Squares (LS) method is applied to estimate the channel. From the demodulated received signal $\mathbf{R}_{\text{freq}}^{\text{ofdm}}$, the data symbols at the pilot positions N_{pilot} are extracted to form a response pilot sequence in the frequency domain $\mathbf{R}_{\text{pilot}}^{\text{ofdm}} \in \mathbb{C}^{N_{\text{bs}} \times K_{\text{bs,ant}} \times N_{\text{pilot}} \times N_{\text{sub}}^{\text{eff}}}$. Then, the guard carriers and DC null are removed from the response matrix $\mathbf{R}_{\text{pilot}}^{\text{ofdm}}$. According to (9), the pilot observation matrix $\mathbf{R}_{\text{pilot}}^{\text{ofdm}}$ can be presented as⁴

$$\mathbf{R}_{\text{pilot}}^{\text{ofdm}} = \mathbf{H}_{\text{pilot}} \mathbf{X}_{\text{pilot}}^{\text{ofdm}} + \mathbf{N}_{\text{pilot}}, \quad (11)$$

where $\mathbf{H}_{\text{pilot}} \in \mathbb{C}^{N_{\text{ut}} \times K_{\text{ut,ant}} \times N_{\text{bs}} \times K_{\text{bs,ant}} \times N_{\text{pilot}} \times N_{\text{sub}}^{\text{eff}}}$ is the frequency channel containing the pilot information, $\mathbf{H}_{\text{pilot}} \in \mathbf{H}_{\text{freq}}$, and $\mathbf{N}_{\text{pilot}} \sim \mathcal{CN}(0, 1)$ is AWGN. Applying the LS channel model, denote $\hat{\mathbf{H}}_{\text{pilot}}$ as the estimated channel; its formulation can be expressed as

$$\hat{\mathbf{H}}_{\text{pilot}} = \mathbf{R}_{\text{pilot}}^{\text{ofdm}} \odot \frac{\mathbf{X}_{\text{pilot}}^{\text{ofdm}*}}{|\mathbf{X}_{\text{pilot}}^{\text{ofdm}}|^2}. \quad (12)$$

⁴It is noted that we do not present the precoding in the signal modeling because the element radiation pattern is set using `antenna_pattern`, which is ("omni" or "38.901") [10].

From (12), the estimated channel requires knowledge of the pilot sequences, i.e., the Kronecker pattern, to perform the estimation. Since $\hat{\mathbf{H}}_{\text{pilot}}$ only provides the channel at the pilot indices of the transmitted data, an interpolation method is applied to interpolate the channel across all subcarriers. Let $f_{\text{inter}}(\hat{\mathbf{H}}_{\text{pilot}})$ represent the interpolation method, e.g., Nearest Neighbour (NN), Linear Regression (LIN), or Linear Minimum Mean Square Error (LMMSE). Let $\hat{\mathbf{H}}_{\text{freq}} \in \mathbb{C}^{N_{\text{ut}} \times K_{\text{ut,ant}} \times N_{\text{bs}} \times K_{\text{bs,ant}} \times N_{\text{sym}}^{\text{ofdm}} \times N_{\text{sub}}^{\text{eff}}}$ denote the estimated channel after interpolation, which can be expressed as

$$\hat{\mathbf{H}}_{\text{freq}} = f_{\text{inter}}(\hat{\mathbf{H}}_{\text{pilot}}). \quad (13)$$

1) *Linear Minimum Mean Square Error Equalizer:* From (9) and (13), we can estimate the transmit symbols $\hat{\mathbf{X}}_{\text{qam}} \in \mathbb{C}^{N_{\text{ut}} \times N_{\text{stream}} \times N_{\text{sym}}}$ by applying the LMMSE equalizer method, which is presented as

$$\hat{\mathbf{X}}_{\text{qam}} = \text{diag}(\mathbf{G}\hat{\mathbf{H}}_{\text{freq}})^{-1}\mathbf{G}\mathbf{R}_{\text{freq}}^{\text{ofdm}}, \quad (14)$$

where $\mathbf{G} = \hat{\mathbf{H}}_{\text{freq}}^{\text{H}}(\hat{\mathbf{H}}_{\text{freq}}\hat{\mathbf{H}}_{\text{freq}}^{\text{H}} + \mathbf{S})^{-1}$, $\mathbf{S} = \mathbb{E}[\mathbf{N}\mathbf{N}^{\text{H}}]$, and \mathbf{N} is an AWGN matrix.

2) *Decoding Process:* For the decoding process, we consider the QAM demapper and LDPC decoder methods. According to (14), we denote $\hat{\mathbf{X}}'_{\text{bit}} \in \{0, 1\}^{N_{\text{ut}} \times N_{\text{stream}} \times N'_{\text{bit}}}$ as the estimated transmitted bit data, obtained by applying QAM demapping methods, where each symbol of $\hat{\mathbf{X}}_{\text{qam}}$ is given by $\hat{p}_i = \hat{A}_i^{\text{I}} + j\hat{A}_i^{\text{Q}}$ and can be represented as a constellation point. Since the estimated in-phase and quadrature components are affected by noise, we apply the Log-Likelihood Ratio (LLR) method to map the noisy symbols to the corresponding constellation points before transforming them into information bits, i.e., $\hat{\mathbf{X}}_{\text{qam}} \mapsto \hat{\mathbf{X}}'_{\text{bit}}$, as follows $\frac{1}{\sqrt{2}} + \frac{1}{\sqrt{2}}j \rightarrow 00, -\frac{1}{\sqrt{2}} + \frac{1}{\sqrt{2}}j \rightarrow 01, -\frac{1}{\sqrt{2}} - \frac{1}{\sqrt{2}}j \rightarrow 11, \frac{1}{\sqrt{2}} - \frac{1}{\sqrt{2}}j \rightarrow 10$. Let $f_{\text{dec,ldpc}}(\hat{\mathbf{X}}'_{\text{bit}})$ denote the LDPC decoder function, which uses redundant bits to detect and correct errors. The estimated bit data without check bits, $\hat{\mathbf{X}}_{\text{bit}} \in \{0, 1\}^{N_{\text{ut}} \times N_{\text{stream}} \times N_{\text{bit}}}$, can be expressed as

$$\hat{\mathbf{X}}_{\text{bit}} = f_{\text{dec,ldpc}}(\hat{\mathbf{X}}'_{\text{bit}}). \quad (15)$$

B. Analysis of Statistical Geometry Property in Signal Representations for Data Features

From (7), the high-dimensional Time-Frequency-Space (HD-TFS) channel representation is determined by geometric characteristics, i.e., θ , ϕ , N_{ray} , N_{cluster} , and \mathbf{v}_{ut} . In our practical simulation, the arrival and departure angles θ and ϕ are uniformly distributed within a specified boundary $[-\pi, \pi]$. These angle values are shuffled, and their number depends on the number of rays N_{ray} and clusters N_{cluster} . Therefore, the key geometric characteristics analyzed include N_{ray} , N_{cluster} , and \mathbf{v}_{ut} . As shown in Fig. 2, for each antenna, the TF channel gain is illustrated under different geometric characteristic settings. Assuming the default settings of $|\mathbf{v}_{\text{ut}}| = 0$, N_{ray} , and N_{cluster} , the channel gain decreases as

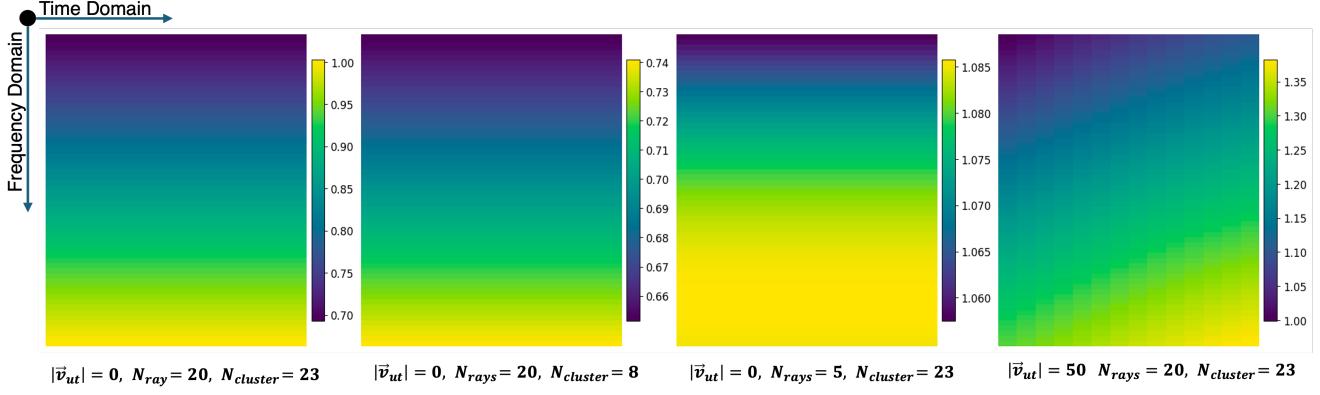


Fig. 2. Time-Frequency channel representation of an example antenna with different geometric characteristic configurations.

N_{cluster} is reduced. A smaller number of clusters leads to fewer constructive contributions, leading to a lower overall power gain compared to the default configuration. In contrast, when N_{ray} decreases and N_{cluster} remains constant, the power is distributed among fewer rays within each cluster. This increases the power gain per ray, leading to a higher overall channel gain and a brighter color representation. The most significant impact comes from the geometry of \mathbf{v}_{ut} , where the TF channel varies rapidly over time. Fig. 3 illustrates the distribution characteristics of the CDL-B channel model presented in

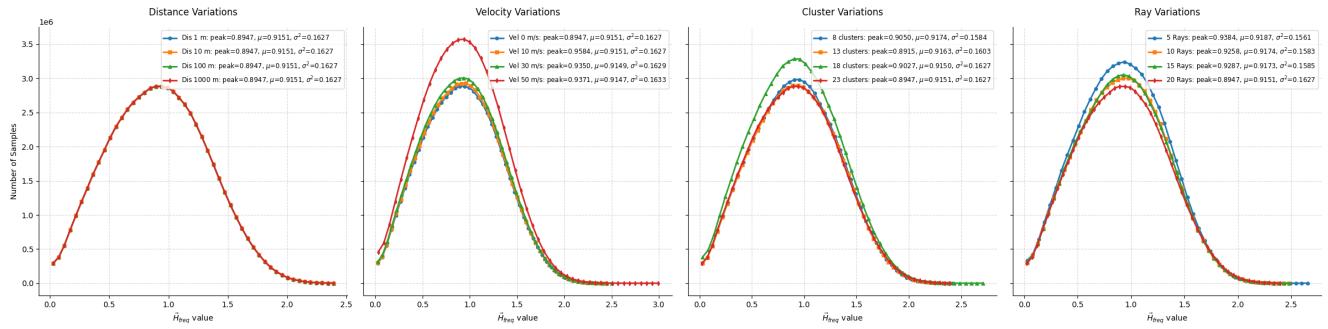


Fig. 3. The distribution of the channel model \mathbf{H}_{freq} in TFS domain over different geometric characteristic configuration settings

the HD-TFS domain under various geometric configurations. The distribution of the geometry-based channel model (GCM) remains consistent across different distances, while at higher user velocities, the distribution becomes steeper due to increased Doppler effects. Similarly, when the number of rays per cluster decreases, the power per ray increases, further sharpening the distribution. Although the power of each cluster is predefined, it can either decrease due to destructive interference or increase through constructive addition. In general, the distribution of the CDL-B channel model closely follows a normal

distribution. Given the skewness in the distribution, our proposed model leverages this distribution as an unsupervised learning solution to effectively capture the characteristics of the GCM.

C. Exploiting Probabilistic Channel Estimation Based on Geometric Characteristics

According to the channel modeling presented in the previous section, the geometric characteristics $\phi, \theta, N_{\text{ray}}, N_{\text{cluster}}, \mathbf{v}_{\text{ut}}$ are considered. From (11), let $p(\mathbf{H}_{\text{freq}}, \phi, \theta, N_{\text{ray}}, N_{\text{cluster}}, \mathbf{v}_{\text{ut}}, \mathbf{r} | \mathbf{R}_{\text{freq}}^{\text{ofdm}})$ represent the joint posterior probability of the channel \mathbf{H}_{freq} given the observation $\mathbf{R}_{\text{freq}}^{\text{ofdm}}$ and geometric characteristics, expressed as

$$p(\mathbf{H}_{\text{freq}}, \phi, \theta, N_{\text{ray}}, N_{\text{cluster}}, \mathbf{v}_{\text{ut}} | \mathbf{R}_{\text{freq}}^{\text{ofdm}}) \propto \underbrace{p(\mathbf{R}_{\text{freq}}^{\text{ofdm}} | \mathbf{H}_{\text{freq}})}_{\text{likelihood}} \underbrace{p(\mathbf{H}_{\text{freq}} | \phi, \theta, N_{\text{ray}}, N_{\text{cluster}}, \mathbf{v}_{\text{ut}})}_{\text{CDL channel model}} \\ \underbrace{p(\phi)p(\theta)p(N_{\text{ray}})p(N_{\text{cluster}})p(\mathbf{v}_{\text{ut}})}_{\text{geometry prior probabilities}}, \quad (16)$$

where $\theta \sim \mathcal{U}(-\pi, \pi)$, $\phi \sim \mathcal{U}(-\pi, \pi)$, $N_{\text{ray}} \sim \mathcal{U}(1, 10)$, $N_{\text{cluster}} \sim \mathcal{U}(1, 23)$, $\mathbf{v}_{\text{ut}} \sim \mathcal{U}(0, \mathbf{v}_{\text{max}})$, $|\mathbf{v}_{\text{max}}| = 10 \text{ m/s}$. Here, \propto denotes proportionality, indicating dependencies within the distributions. From (16), call $p(\mathbf{H}_{\text{freq}}, \gamma | \mathbf{R}_{\text{freq}}^{\text{ofdm}}) = p(\mathbf{H}_{\text{freq}}, \phi, \theta, N_{\text{ray}}, N_{\text{cluster}}, \mathbf{v}_{\text{ut}} | \mathbf{R}_{\text{freq}}^{\text{ofdm}})$ is the posterior probability given geometry γ of the channel, where $\gamma = \{\phi, \theta, N_{\text{ray}}, N_{\text{cluster}}, \mathbf{v}_{\text{ut}}\}$. The estimated channel $\hat{\mathbf{H}}_{\text{freq}} \sim p(\hat{\mathbf{H}}_{\text{freq}} | \mathbf{R}_{\text{freq}}^{\text{ofdm}})$ is effected not only by the likelihood observation $p(\mathbf{R}_{\text{freq}}^{\text{ofdm}} | \mathbf{H}_{\text{freq}})$ but also by the geometry prior probabilities.

D. Proposed Adversarial Training for MAP-Based Channel Estimation

We formulate the channel estimation problem as a Maximum A Posteriori (MAP) estimation task and propose an adversarial training framework to solve it. The target leverages the varying distribution of geometric characteristics to enhance the accuracy of GCM channel estimation in the HD-TFS domain. The framework adopts the GAN training strategy from [27], consisting of a generator G that generates synthetic channel estimates and a critic D that discriminates between the generated channel and the true channel. The generator G receives the observation $\underline{\mathbf{R}}_{\text{freq}}^{\text{ofdm}}$ after the OFDM demodulation process as input, mapping it as $G(\underline{\mathbf{R}}_{\text{freq}}^{\text{ofdm}})$ such that $G(\underline{\mathbf{R}}_{\text{freq}}^{\text{ofdm}}) : \mathbb{C}^{N_{\text{bs}} \times K_{\text{bs, ant}} \times N_{\text{sym}}^{\text{ofdm}} \times N_{\text{sub}}^{\text{block}}} \rightarrow \mathbb{C}^{N_{\text{ut}} \times K_{\text{ut, ant}} \times N_{\text{bs}} \times K_{\text{bs, ant}} \times N_{\text{sym}}^{\text{ofdm}} \times N_{\text{sub}}^{\text{eff}}}$, where $\underline{\mathbf{R}}_{\text{freq}}^{\text{ofdm}} \mapsto \hat{\mathbf{H}}_{\text{freq}}$. Thus, the instantaneous channel estimate is $\hat{\mathbf{H}}_{\text{freq}} = G(\underline{\mathbf{R}}_{\text{freq}}^{\text{ofdm}})$. Given the true random instantaneous channel $\underline{\mathbf{H}}_{\text{freq}}$, the input of the critic D can be either $\underline{\mathbf{H}}_{\text{freq}}$ or $\hat{\mathbf{H}}_{\text{freq}}$, with $D(\underline{\mathbf{H}}_{\text{freq}}^*) : \mathbb{C}^{N_{\text{ut}} \times K_{\text{ut, ant}} \times N_{\text{bs}} \times K_{\text{bs, ant}} \times N_{\text{sym}}^{\text{ofdm}} \times N_{\text{sub}}^{\text{eff}}} \rightarrow [0, 1]$, where $\underline{\mathbf{H}}_{\text{freq}}^* \mapsto \{p(\underline{\mathbf{H}}_{\text{freq}}^* | H_0), p(\underline{\mathbf{H}}_{\text{freq}}^* | H_1)\}$. Here, $\underline{\mathbf{H}}_{\text{freq}}^*$ can represent either $\underline{\mathbf{H}}_{\text{freq}}$ or $\hat{\mathbf{H}}_{\text{freq}}$. The likelihood probabilities $p(\underline{\mathbf{H}}_{\text{freq}}^* | H_0)$ and $p(\underline{\mathbf{H}}_{\text{freq}}^* | H_1)$ correspond to the hypotheses H_0 for the generated channel

and H_1 for the true channel. In the training framework formulated as a minimax game, the generative function G aims to maximize the similarity between the real and estimated characteristics of the GCM, while the critic function D attempts to minimize the distinguishability between them.

Problem Formulation for Optimizing the Generator: The shape of the observation matrix $\underline{\mathbf{R}}_{\text{freq}}^{\text{ofdm}}$ is different from $\hat{\mathbf{H}}_{\text{freq}}$, which causes difficulty in designing the adversarial model. Moreover, the $\underline{\mathbf{R}}_{\text{freq}}^{\text{ofdm}}$ contains the transmit data, which is considered as random variables, so the adversarial model has difficulty transforming the observation data to the estimated channel. According to (11), we initially transform from $\underline{\mathbf{R}}_{\text{freq}}^{\text{ofdm}}$ to $\underline{\mathbf{R}}_{\text{pilot}}^{\text{ofdm}} \in \mathbb{C}^{N_{\text{ut}} \times K_{\text{ut,ant}} \times N_{\text{bs}} \times K_{\text{bs,ant}} \times N_{\text{sym}}^{\text{ofdm}} \times N_{\text{sub}}^{\text{eff}}}$, mapping the shape of the real channel. Let $f_{\text{trans}}(\underline{\mathbf{R}}_{\text{freq}}^{\text{ofdm}})$ denote the transform function from the observation data to the transmit data containing the pilot only, which can be expressed as

$$\underline{\mathbf{R}}_{\text{pilot}}^{\text{ofdm}} = f_{\text{trans}}(\underline{\mathbf{R}}_{\text{freq}}^{\text{ofdm}}). \quad (17)$$

The input of the generative model is $\underline{\mathbf{R}}_{\text{pilot}}^{\text{ofdm}}$, and $\hat{\mathbf{H}}_{\text{freq}} = G(\underline{\mathbf{R}}_{\text{pilot}}^{\text{ofdm}})$. To enable the generator function G to converge quickly from the observation data distribution to the channel distribution, we try to minimize the Wasserstein distance between the distribution of the generated channel $p_{\hat{\mathbf{H}}_{\text{freq}}}(\hat{\mathbf{H}}_{\text{freq}}) = p(\hat{\mathbf{H}}_{\text{freq}}|\underline{\mathbf{R}}_{\text{freq}}^{\text{ofdm}})$ and that of the real channel $p_{\mathbf{H}_{\text{freq}}}(\mathbf{H}_{\text{freq}}) = p(\mathbf{H}_{\text{freq}}|\gamma)$. Let $W(p_{\hat{\mathbf{H}}_{\text{freq}}}(\hat{\mathbf{H}}_{\text{freq}}), p_{\mathbf{H}_{\text{freq}}}(\mathbf{H}_{\text{freq}}))$ denote the Wasserstein distance. If the generated channel has a distribution that largely overlaps with that of the true channel, then the mass is distributed in a similar fashion, and $W(p_{\hat{\mathbf{H}}_{\text{freq}}}(\hat{\mathbf{H}}_{\text{freq}}), p_{\mathbf{H}_{\text{freq}}}(\mathbf{H}_{\text{freq}}))$ will be small. The proposed channel estimation problem is formulated as

$$G^* = \arg \min_{(G|D)} W(p_{\hat{\mathbf{H}}_{\text{freq}}}(\hat{\mathbf{H}}_{\text{freq}}), p_{\mathbf{H}_{\text{freq}}}(\mathbf{H}_{\text{freq}})) \quad (18a)$$

$$\text{subject to } \Re\{[\mathbf{H}_{\text{freq}}]_{ij}\}, \Im\{[\mathbf{H}_{\text{freq}}]_{ij}\}, \Re\{[\hat{\mathbf{H}}_{\text{freq}}]_{ij}\},$$

$$\Im\{[\hat{\mathbf{H}}_{\text{freq}}]_{ij}\} \in [-1, 1], \forall i, \forall j, \quad (18b)$$

$$N_{\text{cp}} \geq \tau_{\max} \times \Delta f \times N_{\text{sub}}^{\text{block}}, \quad (18c)$$

$$v_{\max} \ll \frac{c\Delta f}{f_c(1 + \frac{N_{\text{cp}}}{N_{\text{sub}}^{\text{block}}})}, \quad (18d)$$

$$\Delta t_{\text{pilot}} < T_c, \quad (18e)$$

$$|p_{i,j}|^2 \leq 1, i \leq N_{\text{sym}}^{\text{ofdm}}, j \leq N_{\text{sub}}^{\text{block}}, \quad (18f)$$

where $(G|D)$ indicates that we focus on optimizing G given a specific D . Constraint (18b) shows the normalized channel in $[-1, 1]$. While (18c) is the constraint that the duration of the cyclic prefix of the data transmission $T_{\text{cp}}^{\text{ofdm}} = \frac{N_{\text{cp}}}{N_{\text{sub}}^{\text{block}}} \times \frac{1}{\Delta f} \geq \tau_{\max}$ is used to eliminate inter-symbol interference (ISI). From (4), the Doppler shift has a high impact on the GCM characteristic, so constraint (18d) indicates the

slow fading, which means the duration of symbols is much less than the coherence time, $T_{\text{sym}}^{\text{ofdm}} \ll T_c$, where $T_c = \frac{1}{f_d} = \frac{c}{f_c v_{\max}}$. Besides, the time spacing between pilot symbols, $\Delta t_{\text{pilot}} = N_{\text{pilot}} \times T_{\text{sym}}^{\text{ofdm}}$, the constraint (18e) must be less than or equal to the coherence time to allow the receiver to track time-varying channel conditions and interpolate the channel response across symbols between pilots. The pilot pattern in constraint (18f) enforces sparsity and ensures that $|p_{i,j}|^2 \leq 1$.

IV. PROPOSED AI CHANNEL ESTIMATION FRAMEWORK

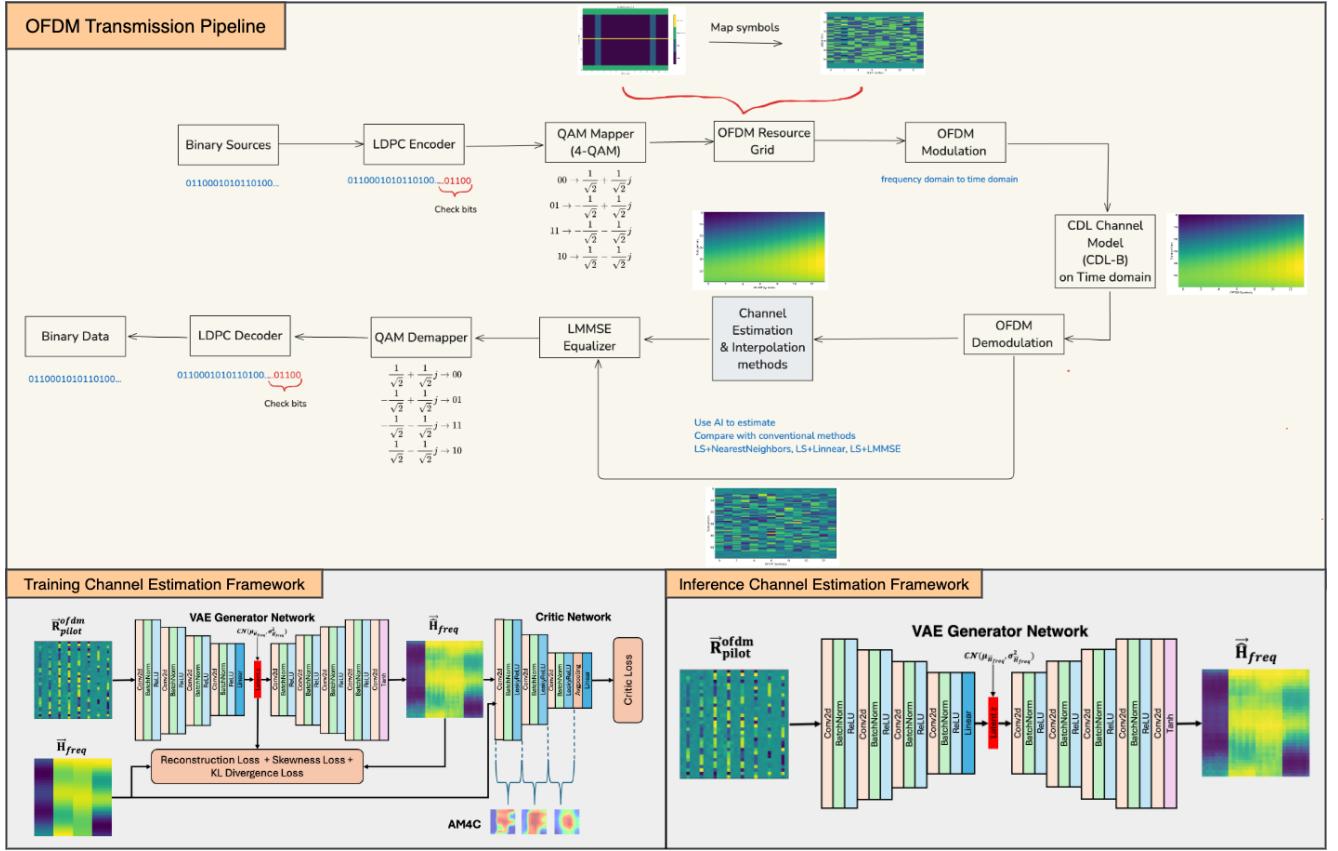


Fig. 4. Three main simulation pipeline: OFDM Transmission Pipeline over CDL-B channel model, Proposal training framework pipeline, and inference generative channel model pipeline in the real application.

The principle of our solution is that the function G can capture the correspondence of geometric features from the channel characteristics, reusing unsupervised learning to compress the TFS into latent space and shift to the TFS-GCM characteristic domain. The critic function D not only discriminates between real and generated channels but also identifies high-magnitude channel characteristics. The contribution incorporates domain knowledge (channel distribution), leverages sparsity (number of

antennas), and uses appropriate loss functions (KL divergence and skewness) tailored to the target application.

A. Mathematical Formulation of the Proposed Generator and Critic (Discriminator) Networks

1) *Proposed Generator Network for G*: Based on the work of [1], and further investigated in [28], [27], the Variational Autoencoder (VAE) provides a promising solution for the generator function G , capable of estimating the GCM channel model in the HD-TFS domain from different distributions. The mathematical function f_g of G is defined with the input being the observation data $\mathbf{R}_{\text{pilot}}^{\text{ofdm}}$, and the set of parameters Θ is updated accordingly. Drawing from the VAE architecture [29], the proposed VAE architecture, VAENet, can be divided into the encoder network (EncNet), latent space, and decoder network (DecNet). Considering the input of EncNet, the observation pilot information $\mathbf{R}_{\text{pilot}}^{\text{ofdm}}$ is designed to fit the generative model function f_g . The input of EncNet has dimensions $N_{\text{ut}} \times K_{\text{ut,ant}} \times N_{\text{bs}} \times K_{\text{bs,ant}} \times N_{\text{sym}}^{\text{ofdm}} \times N_{\text{sub}}^{\text{eff}} = 2^{K_{\text{dim}}}$, where K_{dim} is an exponent of 2. Let $f_{[\text{dim}]}(\cdot) : \mathbb{C}^{N_{\text{ut}} \times K_{\text{ut,ant}} \times N_{\text{bs}} \times K_{\text{bs,ant}} \times N_{\text{sym}}^{\text{ofdm}} \times N_{\text{sub}}^{\text{eff}}} \mapsto \mathbb{C}^{(N_{\text{ut}} \times K_{\text{ut,ant}} \times N_{\text{bs}} \times K_{\text{bs,ant}} \times N_{\text{sym}}^{\text{ofdm}}) \times N_{\text{sub}}^{\text{eff}}}$ denote a function to reduce the dimensionality of inputs, either $\mathbf{H}_{\text{freq}}^*$ or $\mathbf{R}_{\text{pilot}}^{\text{ofdm}}$. The estimated channel can be presented as $\hat{\mathbf{H}}_{\text{freq}} = f_g(f_{[\text{dim}]}(\mathbf{R}_{\text{pilot}}^{\text{ofdm}}); \Theta)$. Because $N_{\text{ut}} \times K_{\text{ut,ant}} \times N_{\text{bs}} \times K_{\text{bs,ant}} \times N_{\text{sym}}^{\text{ofdm}} \times N_{\text{sub}}^{\text{eff}} \neq 2^{K_{\text{dim}}}$, which causes difficulty in designing the VAE architecture, a bicubic interpolation function $f_{\text{resize}}(\cdot) : \mathbb{C}^{(N_{\text{ut}} \times K_{\text{ut,ant}} \times N_{\text{bs}} \times K_{\text{bs,ant}} \times N_{\text{sym}}^{\text{ofdm}}) \times N_{\text{sub}}^{\text{eff}}} \mapsto \mathbb{C}^{H_{\text{in}} \times W_{\text{in}}}$ is proposed to resize the observation before normalizing the input to $[-1, 1]$, where $H_{\text{in}}, W_{\text{in}} = 2^{K_{\text{dim}}}$. Let $f_g^{[bl_i]}$ denote the i -th *feature extraction block*, which can be represented as $f_g^{[bl_i]} = f_g^{[a_i]} \circ f_g^{[b_i]} \circ f_g^{[c_i]}$, comprising activation, batch normalization, and convolution layers. Note that for a composite function, $f \circ g = f(g(x))$. Let $\mathbf{a}_{g,\text{enc}} = f_{g,\text{enc}}(f_{[\text{dim}]}(f_{\text{resize}}(\mathbf{R}_{\text{pilot}}^{\text{ofdm}})); \Theta_{\mathbf{a}_{g,\text{enc}}}) = f_g^{[bl_1]} \circ f_g^{[bl_2]} \circ f_g^{[bl_3]} \circ f_g^{[bl_4]} \circ f_g^{[l]}$ denote the compressed vector of EncNet. The latent vector \mathbf{z} is the output of the transformation function $\mathbf{z} = f_{g,\text{trans}}(\mathbf{a}_{g,\text{enc}})$ and serves as the input to DecNet. Considering the convolution blocks design (ConvBlock), the output of DecNet is given by $\hat{\mathbf{H}}_{\text{freq}} = f_{g,\text{dec}}(\mathbf{z}; \Theta_{g,\text{dec}}) = f_g^{[bl_1]} \circ f_g^{[bl_2]} \circ f_g^{[bl_3]} \circ f_g^{[bl_4]} \circ f_g^{[\tanh]}$. The function $f_g^{[\tanh]}$ ensures that the output is normalized within $[-1, 1]$, aligning with the scaling of the true channel \mathbf{H} . The VAENet is mathematically described as

$$f_g(\mathbf{R}_{\text{pilot}}^{\text{ofdm}}; \Theta) = f_{g,\text{dec}}(f_{g,\text{trans}}(f_{g,\text{enc}}(f_{[\text{dim}]}(f_{\text{resize}}(\mathbf{R}_{\text{pilot}}^{\text{ofdm}})); \Theta_{\mathbf{a}_{g,\text{enc}}})) ; \Theta_{g,\text{dec}}). \quad (19)$$

Since the estimated channel $\hat{\mathbf{H}}_{\text{freq}} \in [-1, 1]^{H_{\text{in}} \times W_{\text{in}}}$ is not the real value of the channel, call the inverse bicubic function $f_{\text{inter}}^{\text{inv,bic}}(\cdot)$ to rescale the value of the estimated channel $\hat{\mathbf{H}}_{\text{freq}} = f_{\text{inter}}^{\text{inv,bic}}(f_g(\mathbf{R}_{\text{pilot}}^{\text{ofdm}}; \Theta))$, where $\hat{\mathbf{H}}_{\text{freq}} \in \mathbb{C}^{N_{\text{ut}} \times K_{\text{ut,ant}} \times N_{\text{bs}} \times K_{\text{bs,ant}} \times N_{\text{sym}}^{\text{ofdm}} \times N_{\text{sub}}^{\text{eff}}}$.

2) *Proposed Critic Network for D*: In contrast, the critic network is implemented using a Convolutional Neural Network (CNN), called CriticNet. It is denoted mathematically as f_c , has a set of parameters ω , and includes a feature extraction block defined as $f_{c,ft}^{[bl_i]} = f_c^{[a_i]} \circ f_c^{[b_i]} \circ f_c^{[c_i]}$ at the i -th block. Consider the ConvBlock of CriticNet, the extract functions $f_{c,ft}(f_{[dim]}(f_{\text{resize}}^{\text{cube}}(\mathbf{H}_{\text{freq}}^*)); \omega_{c,ft}) = f_{c,ft}^{[bl_1]} \circ f_{c,ft}^{[bl_2]} \circ f_{c,ft}^{[bl_3]}$. The function D of CriticNet can be mathematically expressed as

$$f_c(\mathbf{H}_{\text{freq}}^*; \omega) = f_c^{[l_4]}(f_c^{[\text{avg}_4]}(f_{c,ft}(f_{[dim]}(f_{\text{resize}}^{\text{cube}}(\mathbf{H}_{\text{freq}}^*)); \omega_{c,ft}); \omega_c^{[\text{avg}_4]}), \omega_c^{[l_4]}). \quad (20)$$

The output of linear function $f_c^{[l_4]}$ of CriticNet is scalar, with $f_c^{[l_4]}(X_{\underline{\mathbf{H}}_{\text{freq}}}^{[l_4]}, \omega_c^{[l_4]}) > f_c^{[l_4]}(X_{\hat{\mathbf{H}}_{\text{freq}}}^{[l_4]}, \omega_c^{[l_4]})$, where $X_{\underline{\mathbf{H}}_{\text{freq}}}^{[l_4]}$ and $X_{\hat{\mathbf{H}}_{\text{freq}}}^{[l_4]}$ are the inputs to the linear function for $D(\mathbf{H}_{\text{freq}})$ and $D(\hat{\mathbf{H}}_{\text{freq}})$, respectively.

Remark 2. *It is numerically demonstrated that the scale $[-1, 1]$ is used to limit the uncertainty of the bicubic function that effect has an impact on the final evaluation results.*

3) *Latent Distribution Learning in the VAE Network*: The novelty of our VAENet is using the true TFS-GCM characteristic to improve the latent learning; it means shifting the latent distribution close to the true channel distribution. According to the latent distribution of [1], call $q(\mathbf{z}|\mathbf{H}_{\text{freq}})$ the true posterior probability given the real channel \mathbf{H}_{freq} of DecNet. The perspective of the latent vector \mathbf{z} is to capture the GCM characteristic of the CDL-B channel at (16), $q(\mathbf{z}|\mathbf{H}_{\text{freq}}) \approx p(\gamma|\mathbf{H}_{\text{freq}})$. In the real case, EncNet cannot estimate the latent vector \mathbf{z} from \mathbf{H}_{freq} because \mathbf{H}_{freq} is not observed. From the real observation $\mathbf{R}_{\text{pilot}}^{\text{ofdm}}$, the latent vector can be estimated as $\mathbf{z} \sim p(\mathbf{z}|\mathbf{R}_{\text{pilot}}^{\text{ofdm}})$, and $p(\mathbf{z}|\mathbf{R}_{\text{pilot}}^{\text{ofdm}}) \neq q(\mathbf{z}|\mathbf{H}_{\text{freq}})$. Because $\mathbf{a}_{g,\text{enc}}$ is the compressed vector of EncNet given input $\mathbf{R}_{\text{pilot}}^{\text{ofdm}}$, with $\mathbf{z} \sim p(\mathbf{z}|\mathbf{a}_{g,\text{enc}})$. At DecNet, we want to learn a VPD robust to the geometry changes of the CDL channel model, with the estimated channel $\hat{\mathbf{H}}_{\text{freq}} \sim p(\hat{\mathbf{H}}_{\text{freq}}|\mathbf{z})$. It means the latent vector \mathbf{z} has to learn the variation of the GCM characteristic. The probabilistic problem becomes $p(\mathbf{z}|\mathbf{a}_{g,\text{enc}}) \approx p(\gamma|\mathbf{H}_{\text{freq}})$. However, the optimized latent vector in [29] practical solution tends to be closely distributed as $\mathcal{N}(0, 1)$. This causes difficulty in shifting the domain of the latent vector \mathbf{z} close to the geometry channel fusion, which has a skewed bell curve distribution as in Fig. 3. In the proposed latent space vector, assuming that the realistic channel as training data $\mathcal{D}_{\text{train}}$ is large enough to form the Gaussian distribution $\mathcal{N}(\mu_{\underline{\mathbf{H}}_{\text{freq}}}, \sigma_{\underline{\mathbf{H}}_{\text{freq}}}^2)$. Our target is estimating the latent vector $\mathbf{z} \sim p(\mathbf{z}|\mathbf{a}_{g,\text{enc}}) \approx \mathcal{N}(\mu_{\underline{\mathbf{H}}_{\text{freq}}}, \sigma_{\underline{\mathbf{H}}_{\text{freq}}}^2)$ close to the geometric characteristic distribution; it means $p(\mathbf{z}) \approx p(\gamma)$. Applying the trick parameterization $\mathbf{z} = f_{g,\text{trans}}(\mathbf{a}_{g,\text{enc}}) = \mu_{\underline{\mathbf{H}}_{\text{freq}}} + \mathbf{a}_{g,\text{enc}}\sigma_{\underline{\mathbf{H}}_{\text{freq}}}$. At the end, DecNet improves the accurate VPD of the estimated channel $\hat{\mathbf{H}}_{\text{freq}} \sim q(\hat{\mathbf{H}}_{\text{freq}}|\mathbf{z}) \approx q(\hat{\mathbf{H}}_{\text{freq}}|\gamma)$.

Remark 3. *The keys technical contributions is defining the KL divergence loss to make the latent*

distribution close to true channel distribution. Instead of make latent \mathbf{z} close to $\mathcal{N}(0, 1)$, we make $\mathbf{z} \sim \mathcal{N}(\mu_{\mathbf{H}_{\text{freq}}}, \sigma_{\mathbf{H}_{\text{freq}}}^2)$ because we assume we know $\mathcal{N}(\mu_{\mathbf{H}_{\text{freq}}}, \sigma_{\mathbf{H}_{\text{freq}}}^2)$ by sampling real channels.

B. Novel Proposed Loss Function

In the GAN training process, formulated as a minimax game, the CriticNet aims to maximize the distance between the two distributions, whereas the VAENet model seeks to minimize it.

1) *Loss of D*: The CriticNet D function seeks to maximize the Lipschitz continuity of its output by controlling the gradients of $\hat{\mathbf{H}}_{\text{freq}}$ through its parameters ω as

$$L_D = \underbrace{\mathbb{E}_{\hat{\mathbf{H}}_{\text{freq}} \sim p_{\underline{\mathbf{H}}_{\text{freq}}}(\hat{\mathbf{H}}_{\text{freq}})}[D_{\omega}(\hat{\mathbf{H}}_{\text{freq}})] - \mathbb{E}_{\mathbf{H}_{\text{freq}} \sim p_{\underline{\mathbf{H}}_{\text{freq}}}(\mathbf{H}_{\text{freq}})}[D_{\omega}(\mathbf{H}_{\text{freq}})]}_{\triangleq L_f: \text{generative loss}}}_{\triangleq L_r: \text{true loss}}. \quad (21)$$

Consider $p_{\tilde{\mathbf{H}}}(\tilde{\mathbf{H}})$ as the mixture distribution of $p_{\underline{\mathbf{H}}}(\mathbf{H})$ and $p_{\hat{\mathbf{H}}}(\hat{\mathbf{H}}_{\text{freq}})$, given by $p_{\tilde{\mathbf{H}}_{\text{freq}}}(\tilde{\mathbf{H}}_{\text{freq}}) = \epsilon p_{\underline{\mathbf{H}}_{\text{freq}}}(\mathbf{H}_{\text{freq}}) + (1 - \epsilon)p_{\hat{\mathbf{H}}_{\text{freq}}}(\hat{\mathbf{H}}_{\text{freq}})$ where $\epsilon \sim \mathcal{U}(0, 1)$. To constrain the CriticNet's update speed and preserve the 1-Lipschitz property, the gradient penalty loss can be defined as

$$L_{\text{gp}} = \lambda_{\text{gp}} \mathbb{E}_{\tilde{\mathbf{H}}_{\text{freq}} \sim p_{\underline{\mathbf{H}}_{\text{freq}}}(\tilde{\mathbf{H}}_{\text{freq}})}[(\|\nabla D_{\omega}(\tilde{\mathbf{H}}_{\text{freq}})\|_2 - 1)^2]. \quad (22)$$

From the generator loss in (24), the true loss in (21), and the gradient penalty in (22), the adopted loss L_D is given by

$$L_D = L_f - L_r + L_{\text{gp}}. \quad (23)$$

2) *Proposed Loss of G*: In contrast to D , the VAENet G function updates its parameters Θ to minimize the discrepancy between the generated channel and the real channel. Initially, the loss of G can be defined based on the negative fake loss [27] as

$$L_G = -L_f = -\mathbb{E}_{\mathbf{R}_{\text{pilot}}^{\text{ofdm}} \sim p_{\underline{\mathbf{R}}_{\text{pilot}}^{\text{ofdm}}}(\mathbf{R}_{\text{pilot}}^{\text{ofdm}})}[D(G_{\Theta}(\mathbf{R}_{\text{pilot}}^{\text{ofdm}}))]. \quad (24)$$

Taking into account the evidence lower bound of the VAENet and considering Section IV-A3, the latent distribution $p(\mathbf{z}) \approx \mathcal{N}(\mu_{\mathbf{z}, \Theta}, \sigma_{\mathbf{z}, \Theta}^2)$ and the geometry channel characteristic distribution $p(\gamma) \approx \mathcal{N}(\mu_{\mathbf{H}_{\text{freq}}}, \sigma_{\mathbf{H}_{\text{freq}}}^2)$, the KL divergence loss and reconstruction loss are defined, respectively, as

$$L_{\text{KL}} = \lambda_{\text{KL}} \sum_{i=1}^D \left(\log \frac{\sigma_{\mathbf{H}_{\text{freq}}^{[i]}}^2}{\sigma_{\mathbf{z}^{[i]}, \Theta}^2} + \frac{\sigma_{\mathbf{z}^{[i]}, \Theta}^2 + (\mu_{\mathbf{z}^{[i]}, \Theta} - \mu_{\mathbf{H}_{\text{freq}}^{[i]}})^2}{\sigma_{\mathbf{H}_{\text{freq}}^{[i]}}^2} - 1 \right), \quad (25)$$

$$L_{\text{rec}} = \mathbb{E}_{\mathbf{H}_{\text{freq}} \sim p_{\underline{\mathbf{H}}_{\text{freq}}}(\mathbf{H}_{\text{freq}}), \mathbf{R}_{\text{pilot}}^{\text{ofdm}} \sim p_{\underline{\mathbf{R}}_{\text{pilot}}^{\text{ofdm}}}(\mathbf{R})}[\|\mathbf{H}_{\text{freq}} - G_{\Theta}(\mathbf{R}_{\text{pilot}}^{\text{ofdm}})\|^2]. \quad (26)$$

According to Section III-B, the geometry channel distribution $p(\mathbf{H}_{\text{freq}} | \gamma)$ tends to a skewed bell curve shape. Let γ_r and γ_g represent the *coefficients of skewness* or the third-order moments of the true and

estimated channel distributions, respectively. The proposal's skewness loss aims to support the two distributions $p_{\underline{\mathbf{H}}_{\text{freq}}}(\mathbf{H}_{\text{freq}})$, $p_{\hat{\underline{\mathbf{H}}}_{\text{freq}}}(\hat{\mathbf{H}}_{\text{freq}})$ as close as possible. It can be expressed as

$$\gamma_r = (\mathbb{E}_{\mathbf{H}_{\text{freq}} \sim p_{\underline{\mathbf{H}}_{\text{freq}}}(\mathbf{H}_{\text{freq}})}[(\mathbf{H}_{\text{freq}} - \mu_{\mathbf{H}_{\text{freq}},r})^3]) / \sigma_{\mathbf{H}_{\text{freq}},r}^3, \quad (27)$$

$$\gamma_g = (\mathbb{E}_{\hat{\mathbf{H}}_{\text{freq}} \sim p_{\hat{\underline{\mathbf{H}}}_{\text{freq}}}(\hat{\mathbf{H}}_{\text{freq}})}[(\hat{\mathbf{H}}_{\text{freq}} - \mu_{\mathbf{H}_{\text{freq}},g})^3]) / \sigma_{\mathbf{H}_{\text{freq}},g}^3, \quad (28)$$

where $\mu_{\mathbf{H}_{\text{freq}},r}$, $\mu_{\mathbf{H}_{\text{freq}},g}$, $\sigma_{\mathbf{H}_{\text{freq}},r}$, and $\sigma_{\mathbf{H}_{\text{freq}},g}$ are the mean and the standard deviation of the true and generative distribution respectively. It is assumed that the distance of the skewness of true and generative distributions is close and continuous if the distance of two distributions is close. Thus, from (27) and (28), the third order moment-based loss can be defined as

$$L_\gamma = \lambda_\gamma |\gamma_r - \gamma_g|. \quad (29)$$

From the fake loss in (21), the reconstruction loss in (26), the KL loss in (25), and the skewness loss in (29), the proposed loss L_G is

$$L_G = -L_f + L_{\text{rec}} + L_{\text{KL}} + L_\gamma. \quad (30)$$

The proposed training channel estimation framework is outlined in the pseudocode of Algorithm 1, while the generator channel model used during the inference phase is demonstrated in Algorithm 2.

V. PROPOSED EXPLAINABLE AI FRAMEWORK AND NUMERICAL RESULTS

A. Normalized Mean Square Error (NMSE)

Consider the end-to-end pipeline, we applied two main metrics to evaluate generative model performance. According to (13), we apply Normalized Mean Square Error (NMSE) to measure the difference between the true and the generated channels in test datasets $\mathcal{D}_{\text{test}}$, which can be formulated as

$$\text{NMSE} = \frac{\sum_{i=1}^{\mathcal{D}_{\text{test}}} (\mathbf{H}_{\text{freq}}^i - \hat{\mathbf{H}}_{\text{freq}}^i)^2}{\sum_{i=1}^{\mathcal{D}} (\mathbf{H}_{\text{freq}}^i)^2}, \quad (31)$$

B. Structural Similarity (SSIM)

The SSIM metric is adapted to evaluate the structural similarity between the generated channel matrix $\hat{\mathbf{H}}_{\text{freq}}^i$ and the true channel matrix $\mathbf{H}_{\text{freq}}^i$ across the test dataset. The updated formula is:

$$\text{SSIM} = \frac{(2\mu_{\hat{\mathbf{H}}_{\text{freq}}^i} \mu_{\mathbf{H}_{\text{freq}}^i} + c_1)(2\sigma_{\hat{\mathbf{H}}_{\text{freq}}^i \mathbf{H}_{\text{freq}}^i} + c_2)}{(\mu_{\hat{\mathbf{H}}_{\text{freq}}^i}^2 + \mu_{\mathbf{H}_{\text{freq}}^i}^2 + c_1)(\sigma_{\hat{\mathbf{H}}_{\text{freq}}^i}^2 + \sigma_{\mathbf{H}_{\text{freq}}^i}^2 + c_2)}, \quad (32)$$

where $\mu_{\hat{\mathbf{H}}_{\text{freq}}^i}$ and $\mu_{\mathbf{H}_{\text{freq}}^i}$ are the mean values of the generated and true channel matrices for the i -th dataset, respectively, and $\sigma_{\hat{\mathbf{H}}_{\text{freq}}^i}$ and $\sigma_{\mathbf{H}_{\text{freq}}^i}$ are their variances. The covariance $\sigma_{\hat{\mathbf{H}}_{\text{freq}}^i \mathbf{H}_{\text{freq}}^i}$ measures the

Algorithm 1: Training Procedure of the Proposed GAN framework with VAENet model and CriticNet model

Input: Noisy received signal $\underline{\mathbf{R}}_{\text{freq}}^{\text{ofdm}} \sim p_{\underline{\mathbf{R}}_{\text{freq}}^{\text{ofdm}}}(\underline{\mathbf{R}}_{\text{freq}}^{\text{ofdm}})$. The true channel data

$$\mathbf{H}_{\text{freq}} \sim p_{\underline{\mathbf{H}}_{\text{freq}}}(\mathbf{H}_{\text{freq}})$$

Output: Trained Generator G for channel estimation $\hat{\mathbf{H}}_{\text{freq}} \sim p_{\hat{\mathbf{H}}_{\text{freq}}}(\hat{\mathbf{H}}_{\text{freq}})$

```

1 Compute  $\mu_{\underline{\mathbf{H}}_{\text{freq}}}$ ,  $\sigma_{\underline{\mathbf{H}}_{\text{freq}}}$  from training set  $\mathcal{D}_{\text{train}}$ ;
2 for  $e = 0, \dots, K_{\text{epoch}}$  do
3   Sample  $\{\mathbf{H}_{\text{freq},[i]}^{[e]}\}_{i=1}^b \sim p_{\underline{\mathbf{H}}_{\text{freq}}}(\mathbf{H}_{\text{freq}})$  a batch from true channel;
4   Sample  $\{\mathbf{R}_{\text{freq},[i]}^{\text{ofdm},[e]}\}_{i=1}^b \sim p_{\underline{\mathbf{R}}_{\text{freq}}^{\text{ofdm}}}(\mathbf{R}_{\text{freq}}^{\text{ofdm}})$  a batch from received signal;
5   for  $t = 0, \dots, n_c$  do
6      $\{\mathbf{R}_{\text{pilot},[i]}^{\text{ofdm},[e]}\}_{i=1}^b \leftarrow f_{\text{trans}}(\{\mathbf{R}_{\text{freq},[i]}^{\text{ofdm},[e]}\}_{i=1}^b);$ 
7      $\{\hat{\mathbf{H}}_{\text{freq},[i]}^{[t],[e]}\}_{i=1}^b \leftarrow f_g(\{\mathbf{R}_i^{[e]}\}_{i=1}^b; \Theta) \sim p_{\hat{\mathbf{H}}_{\text{freq}}}(\hat{\mathbf{H}}_{\text{freq}})$  predict a batch from generated
      channel;
8      $\{\tilde{\mathbf{H}}_{\text{freq},[i]}^{[t],[e]}\}_{i=1}^b \leftarrow \epsilon \{\mathbf{H}_{\text{freq},[i]}^{[e]}\}_{i=1}^b + (1 - \epsilon) \{\hat{\mathbf{H}}_{\text{freq},[i]}^{[t],[e]}\}_{i=1}^b$  a batch of mixture channel;
9     Training  $a_c^{[t],[e]} \leftarrow f_c(\{\mathbf{H}_{\text{freq},[i]}^{*,[t],[e]}\}_{i=1}^b; \omega)$ ;
10    Compute  $f_{L_D}(a_c^{[t],[e]})$ ;
11    Compute  $f_{L_{\text{gp}}}(\{\tilde{\mathbf{H}}_{\text{freq},[i]}^{[t],[e]}\}_{i=1}^b)$ ;
12     $f_{L_D}^{[t],[e]} \leftarrow f_{L_D}(a_c^{[t],[e]}) + f_{L_{\text{gp}}}(\{\tilde{\mathbf{H}}_i^{[t],[e]}\}_{i=1}^b)$ ;
13     $\omega^{[t+1],[e]} \leftarrow \text{Opt}(f_{L_D}^{[t],[e]}, \omega^{[t],[e]}, \eta, \beta_1, \beta_2)$ ;
14  end
15   $\mathbf{a}_{\text{g,enc}}^{[e]} \leftarrow f_{\text{g,enc}}(f_{[\text{dim}]}(f_{\text{cube}}(f_{\text{resize}}\{\mathbf{R}_{\text{pilot},[i]}^{\text{ofdm},[e]}\}_{i=1}^b))) \sim p(\mathbf{a}_{\text{g,enc}})$ ;
16  Reparameterize  $\mathbf{z}^{[e]} \leftarrow \mu_{\underline{\mathbf{H}}_{\text{freq}}} + \mathbf{a}_{\text{g,enc}}^{[e]} \sigma_{\underline{\mathbf{H}}_{\text{freq}}} \sim p(\mathbf{z} | \mathbf{a}_{\text{g,enc}})$ ;
17   $\{\hat{\mathbf{H}}_{\text{freq},[i]}^{[e]}\}_{i=1}^b \leftarrow f_{\text{g,dec}}(\mathbf{z}^{[e]})$ ;
18  Compute  $f_{L_f}(\{\hat{\mathbf{H}}_{\text{freq},[i]}^{[e]}\}_{i=1}^b)$ ;
19  Compute  $f_{L_{\text{rec}}}(\{\hat{\mathbf{H}}_{\text{freq},[i]}^{[e]}\}_{i=1}^b, \{\mathbf{H}_{\text{freq},[i]}^{[e]}\}_{i=1}^b)$ ;
20  Compute  $f_{L_\gamma}(\{\hat{\mathbf{H}}_{\text{freq},[i]}^{[e]}\}_{i=1}^b, \{\mathbf{H}_{\text{freq},[i]}^{[e]}\}_{i=1}^b)$ ;
21  Compute  $f_{L_{\text{KL}}}(\mathbf{z}^{[e]}, \mu_{\underline{\mathbf{H}}_{\text{freq}}}, \sigma_{\underline{\mathbf{H}}_{\text{freq}}})$ ;
22   $f_{L_G}^{[e]} \leftarrow f_{L_f}(\{\hat{\mathbf{H}}_{\text{freq},[i]}^{[e]}\}_{i=1}^b) + f_{L_{\text{rec}}}(\{\hat{\mathbf{H}}_{\text{freq},[i]}^{[e]}\}_{i=1}^b, \{\mathbf{H}_{\text{freq},[i]}^{[e]}\}_{i=1}^b) + f_{L_{\text{KL}}}(\mathbf{z}^{[e]}, \mu_{\underline{\mathbf{H}}_{\text{freq}}}, \sigma_{\underline{\mathbf{H}}_{\text{freq}}}) +$ 
     $f_{L_\gamma}(\{\hat{\mathbf{H}}_{\text{freq},[i]}^{[e]}\}_{i=1}^b, \{\mathbf{H}_{\text{freq},[i]}^{[e]}\}_{i=1}^b)$ ;
23   $\Theta^{[e+1]} \leftarrow \text{Opt}(f_{L_G}^{[e]}, \Theta^{[e]}, \eta, \beta_1, \beta_2)$ ;
24 end
25 Initialization: Initialize hyperparameters of  $G$  and  $D$ , set learning rates  $\eta$ , the batch size  $b$ ,
  the number of iterations of the CriticNet  $n_c$ , other optimizer parameters  $\beta_1$  and  $\beta_2$ 

```

Algorithm 2: Inference Procedure of the Proposed VAENet Model

Input: Noisy received signal $\mathbf{R}_{\text{freq}}^{\text{ofdm}} \sim p_{\underline{\mathbf{R}}_{\text{freq}}^{\text{ofdm}}}(\mathbf{R}_{\text{freq}}^{\text{ofdm}})$ from test set $\mathcal{D}_{\text{test}}$.

Output: Estimated channel $\hat{\mathbf{H}}_{\text{freq}} \sim p_{\hat{\mathbf{H}}_{\text{freq}}}(\hat{\mathbf{H}}_{\text{freq}})$

- 1 $\mathbf{R}_{\text{pilot}}^{\text{ofdm}} \leftarrow f_{\text{trans}}(\mathbf{R}_{\text{freq}}^{\text{ofdm}});$
 - 2 $\hat{\mathbf{H}}_{\text{freq}} = f_{\text{inter}}^{\text{inv,bic}}(f_g(\mathbf{R}_{\text{pilot}}^{\text{ofdm}}; \Theta));$
 - 3 **Initialization:** Initialize hyperparameters of G with pretrained parameters Θ
-

joint variability. Constants $c_1 = 0.01 \times \max(\hat{\mathbf{H}}_{\text{freq}}^i)$ and $c_2 = 0.03 \times \max(\mathbf{H}_{\text{freq}}^i)$ stabilize the computation. The SSIM range value at Eq. (32), [0, 1], indicates the preservation of structural information, with a higher value signifying greater similarity to the true channel matrix.

C. Bit Error Rate (BER) Analysis

In the system modeling, we utilize the Bit Error Rate (BER) method to compare the error rate between the binary transmit data \mathbf{X}_{bit} and the estimated transmit data $\hat{\mathbf{X}}_{\text{bit}}$ from (15). The BER is:

$$\text{BER} = \frac{\sum_{i=1}^{N_{\text{bit}}} \mathbb{I}(b_{\text{tx}}^i \neq \hat{b}_{\text{rx}}^i)}{N_{\text{bit}}}, \quad (33)$$

where $b_{\text{tx}}^i \in \mathbf{X}_{\text{bit}}$ and $\hat{b}_{\text{rx}}^i \in \hat{\mathbf{X}}_{\text{bit}}$ represent the bit values of the true transmit data and the estimated transmit data, respectively, and $\mathbb{I}(\cdot)$ is the indicator function, equal to 1 if $b_{\text{tx}}^i = \hat{b}_{\text{rx}}^i$ and 0 otherwise.

D. Simulation Settings and Data Generation

1) *Sionna-Based MIMO-OFDM Simulation Settings:* The HD-TFS data is generated using the Sionna library. According to Table I, the antenna settings include the number of antennas and the antenna design pattern. The CDL settings include the geometric characteristics of the CDL-B channel model and the design pattern of the transmitter and receiver configurations. The machine learning settings are used to train and evaluate our generative channel model in the system. The CDL-B channel model is generated using the CDL implementation in the Sionna library.

2) *Data Generation:* Given the OFDM transmission pipeline in Fig. 4 and the configuration settings in Table I, the CDL-B channel presented in the TFS domain is extracted after the OFDM demodulation block to serve as the input dataset for our proposed channel estimation framework. According to Algorithm 1 and Algorithm 2, the training and inference channel estimation frameworks are illustrated in Fig. 4. The training and inference data are generated separately; however, they have the same geometric characteristic distribution. Furthermore, as shown in Fig. 5, a real-world simulation is

TABLE I
SETTINGS FOR CDL SYSTEM, ANTENNA CONFIGURATION, AND MACHINE LEARNING TRAINING

CDL Settings		Antenna Settings		ML Settings	
Δf [kHz]	15	N_{sub}	12	$K_{\text{ut,ant}}$	2
$N_{\text{sub}}^{\text{block}}$	60	$N_{\text{sym}}^{\text{ofdm}}$	14	$K_{\text{bs,ant}}$	4
N_{cp}	6	N_{pilots}	[2, 11]	f_c [GHz]	2.6
N_{ray}	10	B [kHz]	900	UT Array	(1,1)
N_{ratio}	0.5	M (QAM)	4	UT Polar	Dual
$N_{\text{sub}}^{\text{eff}}$	48	N_{bit}	576	UT Type	Cross
N_{stream}	2	N_{dc}	1	BS Array	(1,2)
CDL model	B	τ_{\max} [ns]	100	BS Polar	Dual
Direction	Uplink	v_{ut} [m/s]	0	BS Type	Cross
Pilot Pattern	Kronecker	N_{cluster}	23	Patten	38.901
				λ_{γ}	1

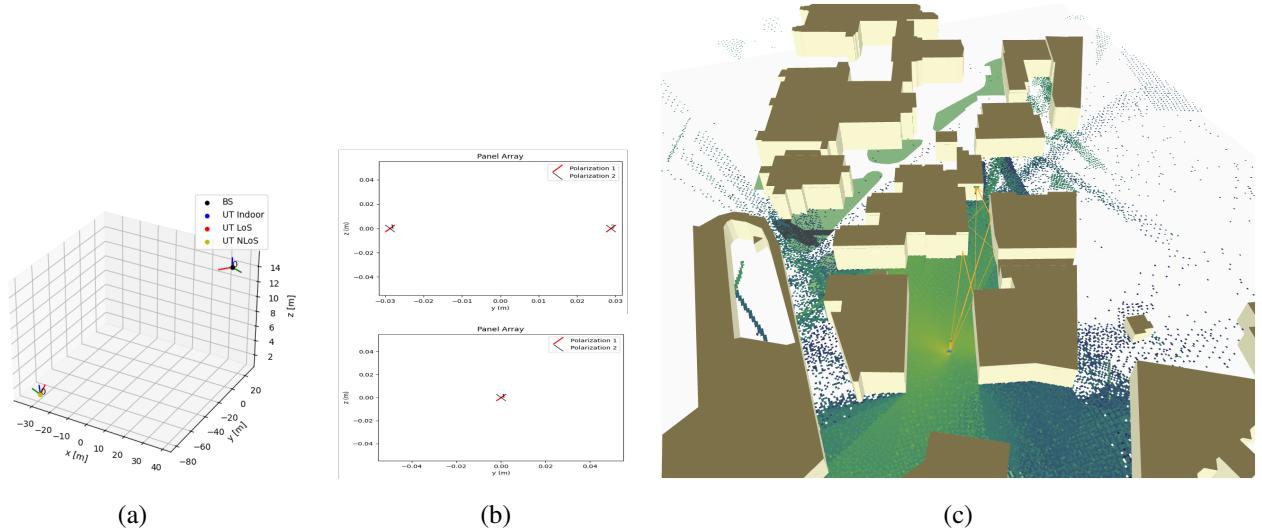


Fig. 5. (a) Network topology for generating training data (the ray setting is default by the Sionna library) (b) Antenna Pattern presentation with $K_{\text{ut,ant}} = 2, K_{\text{bs,ant}} = 4$ (c) Ray Tracings scenario at Polytechnique Montréal, in front of Lassonde Pavilion.

conducted to evaluate the system under realistic conditions. Fig. 5a shows the simulation network topology, including the locations and orientations of the BS and UT. The BS and UT antenna settings in Fig. 5b correspond to the training and simulation settings, respectively. Fig. 5 illustrates the top-view map of the actual transmitter and receiver setup at Polytechnique Montréal. The uplink directions indicate an NLOS environment due to reflections from buildings, aligning with the assumptions of the CDL-B channel model. The real channel data are not only extracted from realistic simulations but also use the same antenna configuration settings and OFDM transmission patterns as the training data.

E. Computational Complexity Analysis

The complexity of channel estimation methods can be expressed using Big- \mathcal{O} notation. From (12), the LS estimator has input $\mathbf{R}_{\text{pilot}}^{\text{ofdm}} \in \mathbb{C}^{N_{\text{ut}} \times K_{\text{ut,ant}} \times N_{\text{bs}} \times K_{\text{bs,ant}} \times N_{\text{sym}}^{\text{ofdm}} \times N_{\text{sub}}^{\text{eff}}}$; its complexity can be expressed as $\mathcal{O}(N_{\text{ut}} \times K_{\text{ut,ant}} \times N_{\text{bs}} \times K_{\text{bs,ant}} \times N_{\text{sym}}^{\text{ofdm}} \times N_{\text{sub}}^{\text{eff}})$. The NN interpolation method selects the value of the nearest known point in the grids, while the Lin interpolation estimates unknown points by the weighted average between two adjacent known points, so the complexity for each unknown point is $\mathcal{O}(1)$. Therefore, the complexity of all grids is $\mathcal{O}(N_{\text{ut}} \times K_{\text{ut,ant}} \times N_{\text{bs}} \times K_{\text{bs,ant}} \times N_{\text{sym}}^{\text{ofdm}} \times N_{\text{sub}}^{\text{eff}})$.

Considering the SOTA method, LMMSE interpolation, the complexity is analyzed in the TFS domain. From the interpolation formulation of [10], the complexity can be represented as $\mathcal{O}(N_{\text{ut}} \times K_{\text{ut,ant}} \times N_{\text{bs}} \times K_{\text{bs,ant}} \times (N_{\text{sym}}^{\text{ofdm}})^3 \times N_{\text{sub}}^{\text{eff}})$ in the time domain, $\mathcal{O}(N_{\text{ut}} \times K_{\text{ut,ant}} \times N_{\text{bs}} \times K_{\text{bs,ant}} \times N_{\text{sym}}^{\text{ofdm}} \times (N_{\text{sub}}^{\text{eff}})^3)$ in the frequency domain, and $\mathcal{O}(N_{\text{ut}} \times (K_{\text{ut,ant}})^3 \times N_{\text{bs}} \times K_{\text{bs,ant}} \times N_{\text{sym}}^{\text{ofdm}} \times N_{\text{sub}}^{\text{eff}})$ in the spatial domain.

TABLE II
INFERENCE TIME OF CHANNEL ESTIMATION MODELS WITH DIFFERENT ANTENNA CONFIGURATIONS (s)

Antenna Configuration	VAE Net	LS+NN	LS+LIN	LS+LMMSE
$K_{\text{ut,ant}} = 1, K_{\text{bs,ant}} = 1$	0.0496	0.0566	0.0594	0.3960
$K_{\text{ut,ant}} = 1, K_{\text{bs,ant}} = 4$	0.0518	0.0588	0.0616	0.3971
$K_{\text{ut,ant}} = 2, K_{\text{bs,ant}} = 4$	0.0545	0.0655	0.0641	0.6600

Consider the proposed channel estimation model, VAE Net . The complexity of each ConvBlock $f_g^{[\text{bl}_i]}$ is given by $\mathcal{O}(C_{\text{in}}^{[\text{bl}_i]} \times C_{\text{out}}^{[\text{bl}_i]} \times (K^{[\text{bl}_i]})^2 \times H_{\text{out}}^{[\text{bl}_i]} \times W_{\text{out}}^{[\text{bl}_i]})$, where $i \in \{1, 2, 3, 4\}$, $C_{\text{in}}, C_{\text{out}}$ are the input and output channels, and $H_{\text{out}}, W_{\text{out}}$ are the feature map dimensions after each ConvBlock. The total complexity of the EncNet is $\mathcal{O}\left(\sum_{i=1}^4 C_{\text{in}}^{[\text{bl}_i]} \times C_{\text{out}}^{[\text{bl}_i]} \times (K^{[\text{bl}_i]})^2 \times H_{\text{out}}^{[\text{bl}_i]} \times W_{\text{out}}^{[\text{bl}_i]}\right)$. In the network architecture, the feature map size after each ConvBlock is given by $H_{\text{out}}^{[\text{bl}_i]} = \frac{H_{\text{in}}}{2^i}, W_{\text{out}}^{[\text{bl}_i]} = \frac{W_{\text{in}}}{2^i}$, where $H_{\text{in}} \times W_{\text{in}} = (N_{\text{ut}} \times K_{\text{ut,ant}} \times N_{\text{bs}} \times K_{\text{bs,ant}} \times N_{\text{sym}}^{\text{ofdm}}) \times N_{\text{sub}}^{\text{eff}}$ represents the reduced feature map of $\mathbf{R}_{\text{pilot}}^{\text{ofdm}}$. Furthermore, the input and output channels follow $C_{\text{in}}^{[\text{bl}_i]} = C_{\text{in}} \times 2^{i-1}, C_{\text{out}}^{[\text{bl}_i]} = C_{\text{in}} \times 2^i$, where C_{in} is the initial number of channels of the EncNet. Substituting these into the expression, the complexity of the EncNet becomes $\mathcal{O}\left(\sum_{i=1}^4 C_{\text{in}} \times 2^{i-1} \times C_{\text{in}} \times 2^i \times (K^{[\text{bl}_i]})^2 \times \frac{H_{\text{in}}}{2^i} \times \frac{W_{\text{in}}}{2^i}\right) = \mathcal{O}\left(\sum_{i=1}^4 \frac{1}{2} \times C_{\text{in}}^2 \times (K^{[\text{bl}_i]})^2 \times H_{\text{in}} \times W_{\text{in}}\right)$. Because C_{in} and $K^{[\text{bl}_i]}$ are fixed, the final complexity of the EncNet simplifies to $\mathcal{O}(H_{\text{in}} \times W_{\text{in}}) = \mathcal{O}(N_{\text{ut}} \times K_{\text{ut,ant}} \times N_{\text{bs}} \times K_{\text{bs,ant}} \times N_{\text{sym}}^{\text{ofdm}} \times N_{\text{sub}}^{\text{eff}})$. The complexity of the DecNet is the same as that of the EncNet, so the overall complexity of VAE Net is $\mathcal{O}(N_{\text{ut}} \times K_{\text{ut,ant}} \times N_{\text{bs}} \times K_{\text{bs,ant}} \times N_{\text{sym}}^{\text{ofdm}} \times N_{\text{sub}}^{\text{eff}})$. As shown in Table II, the runtime of the end-to-

end pipeline for each approach is measured on our device (MacBook M1), with VAENet consistently achieving the fastest inference across all configurations. The Big- \mathcal{O} complexity analysis of the channel estimation methods numerically demonstrates that the proposed VAENet is computationally more efficient than SOTA interpolation methods, including the LMMSE, as well as the NN and LIN methods.

F. Proposed Explainable AI Framework: Activation Mapping for the CriticNet (AM4C)

To demonstrate how the proposed channel model can effectively estimate the characteristics of the CDL-B channel, we introduce an explainable AI method, AM4C [1], designed to analyze the interpretability of CriticNet during the training process, as presented in Fig. 6a. The method uses a propagation mechanism to generate attention heatmaps, highlighting the most salient (high-value) regions of the channel representation across the TFS domain. Denote the activation map of the k -th feature map at the spatial location (i, j) from the convolution function $f_c^{[c_n]}$, where $n \in \{1, 2, 3\}$, given by $f_c^{[c_n]}(\mathbf{X}_c^{[c_n]}; \boldsymbol{\omega}_c^{[c_n]}) = \{\mathbf{A}_{k,i,j}^{c_n} \mid k \in \{1, \dots, K_{\text{spatial}}\}\}$, where K_{spatial} is the number of spatial locations and $\mathbf{X}_c^{[c_n]}$ is the input of the convolution layer. Next, we compute the gradient of the CriticNet output $f_c(\mathbf{H}_{\text{freq}}^*; \boldsymbol{\omega})$ with respect to each feature map's activation $\frac{\partial f_c(\mathbf{H}_{\text{freq}}^*; \boldsymbol{\omega})}{\partial f_c^{[c_n]}(\mathbf{X}_c^{[c_n]}; \boldsymbol{\omega}_c^{[c_n]})}$. Let a_c^k be the weight representing the overall influence of the k -th feature map on the CriticNet decision value, given by $a_c^k = \frac{1}{K_{\text{spatial}}} \sum_{i,j} \frac{\partial f_c(\mathbf{H}_{\text{freq}}^*; \boldsymbol{\omega})}{\partial f_c^{[c_n]}(\mathbf{X}_c^{[c_n]}; \boldsymbol{\omega}_c^{[c_n]})}$. Consider ReLU an activation function to remove negative values. The activation map $M_D^{c_n}(i, j)$ is

$$M_D^{c_n}(i, j) = \text{ReLU}\left((1/K_{\text{spatial}}) \sum_k \sum_{i,j} \frac{\partial f_c(\mathbf{H}_{\text{freq}}^*; \boldsymbol{\omega})}{\partial f_c^{[c_n]}(\mathbf{X}_c^{[c_n]}; \boldsymbol{\omega}_c^{[c_n]})} f_c^{[c_n]}(\mathbf{X}_c^{[c_n]}; \boldsymbol{\omega}_c^{[c_n]})\right). \quad (34)$$

The resulting activation map $M_D^{c_n}(i, j)$ provides a visual and quantitative representation of the spatial regions in the input channel data that are most influential to the CriticNet D . High values in the activation map indicate areas with substantial impact, thereby serving as the *attention regions* that the model focuses on during channel estimation.

Remark 4. After training, the attention areas identified by the CriticNet D consistently align with regions of high magnitude in the channel's TRF, as illustrated in Fig. 6b.

G. Performance Evaluation and Distribution Learning Illustration

In this section, the performance of the VAENet model is compared with conventional methods, specifically the LS channel estimator combined with interpolation techniques such as Nearest Neighbor (NN), Linear (LIN), and LMMSE. First, in Fig. 7, the channel is presented in the TFS domain, with the spatial domain defined by the number of antennas $K_{\text{ut,ant}} \times K_{\text{bs,ant}} = 8$. From the observation,

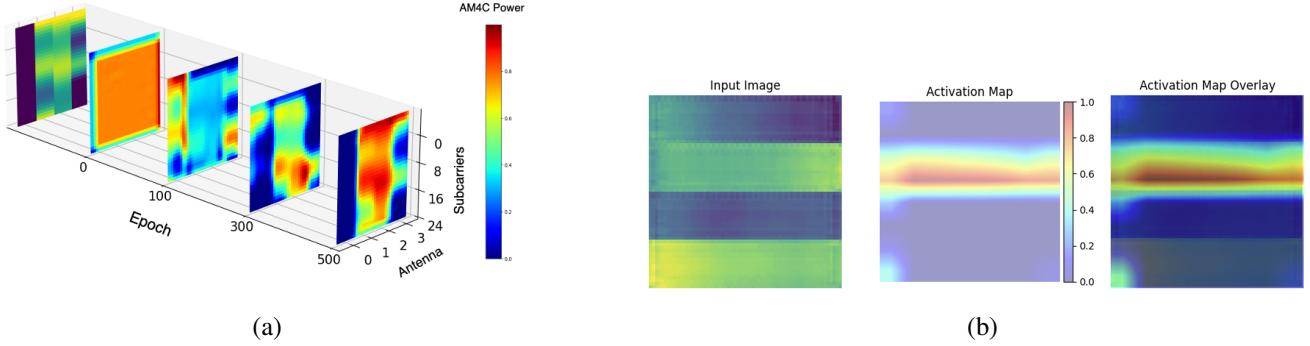


Fig. 6. (a) Evolution of the attention area, i.e., the learning focus, on the critic's data features over training epochs. (b) Final learned attention area overlaid on the flattened channel representation in the TRF domain ($K_{\text{ut,ant}} = 1, K_{\text{bs,ant}} = 4$).

the pilot matrix $\mathbf{R}_{\text{pilot}}^{\text{ofdm}}$ is extracted and given as input to the generative model, and the estimated channel $\hat{\mathbf{H}}_{\text{freq}}$ exhibits a normalized signal strength color similar to that of the real channel \mathbf{H}_{freq} . In addition, Fig. 8 numerically demonstrates the distribution of pixel values in the proposed model.

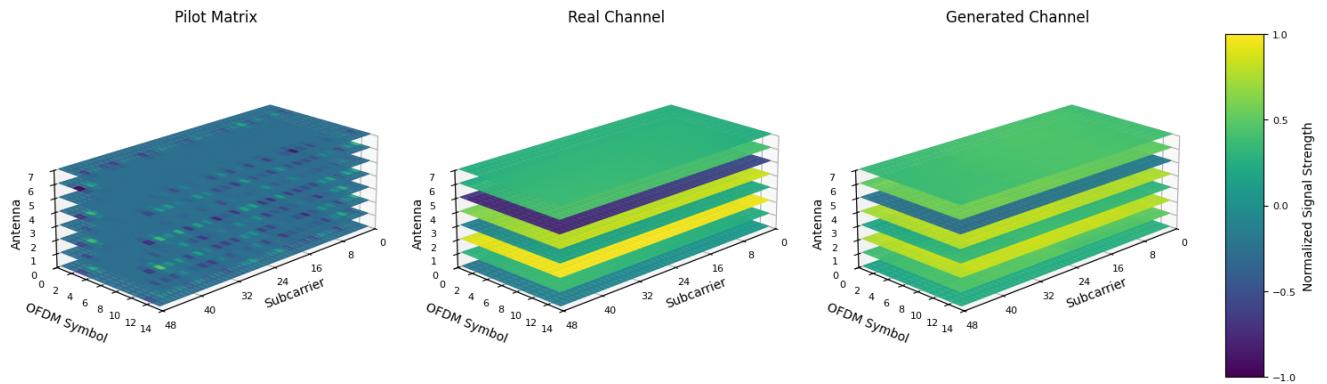


Fig. 7. HD-TFS representation of the pilot matrix $\mathbf{R}_{\text{pilot}}^{\text{ofdm}}$, the true GCM channel matrix \mathbf{H}_{freq} , and the estimated channel matrix $\hat{\mathbf{H}}_{\text{freq}}$.

The observation data $\mathbf{R}_{\text{pilot}}^{\text{ofdm}} \sim p_{\mathbf{R}_{\text{pilot}}^{\text{ofdm}}}(\mathbf{R}_{\text{pilot}}^{\text{ofdm}})$ are given as input to VAENet. The figure empirically illustrates that the distribution of the latent vector is transformed to closely resemble the CDL-B channel distribution shown in Fig. 3, with $p(\mathbf{z}|\mathbf{a}_{\text{g,enc}}) \approx \mathcal{N}(\mu_{\mathbf{H}_{\text{freq}}}, \sigma_{\mathbf{H}_{\text{freq}}}^2)$ after the means are shifted to zero by the scaling function. Based on the latent vector \mathbf{z} , the decoder estimates the channel $\hat{\mathbf{H}}_{\text{freq}} \sim q(\hat{\mathbf{H}}_{\text{freq}}|\mathbf{z})$, which exhibits a pixel distribution similar to that of the real channel $\mathbf{H}_{\text{freq}} \sim p(\mathbf{H}_{\text{freq}}|\mathbf{z})$. According to Eq. (31), the NMSE scores of the channel estimation models are shown in Fig. 9a. In the experiments, the antenna configuration is fixed with $K_{\text{ut,ant}} = 2, K_{\text{bs,ant}} = 4$, while the geometric channel characteristics are varied. In general, the LS+LMMSE model with t-f ordering achieves the best performance across all scenarios. At low SNRs, from -15dB to 0dB , the proposed VAE model

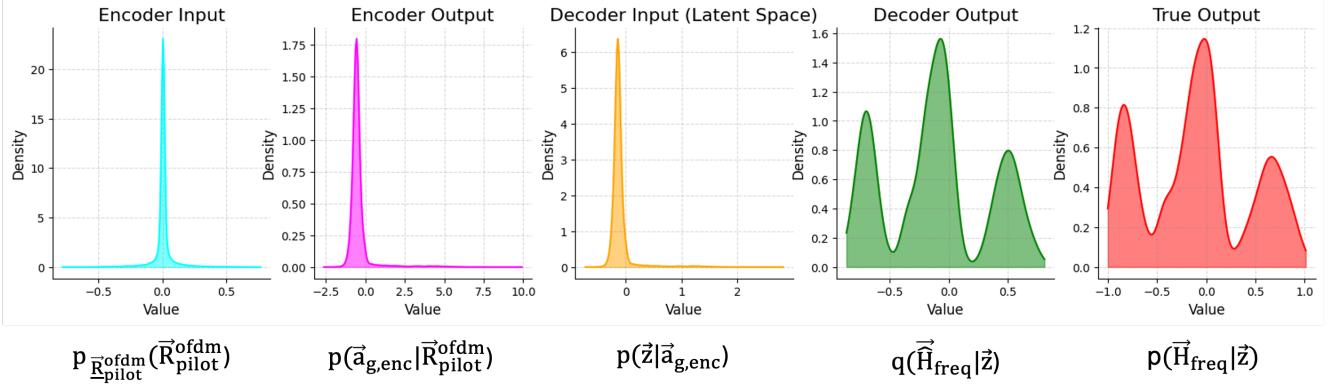
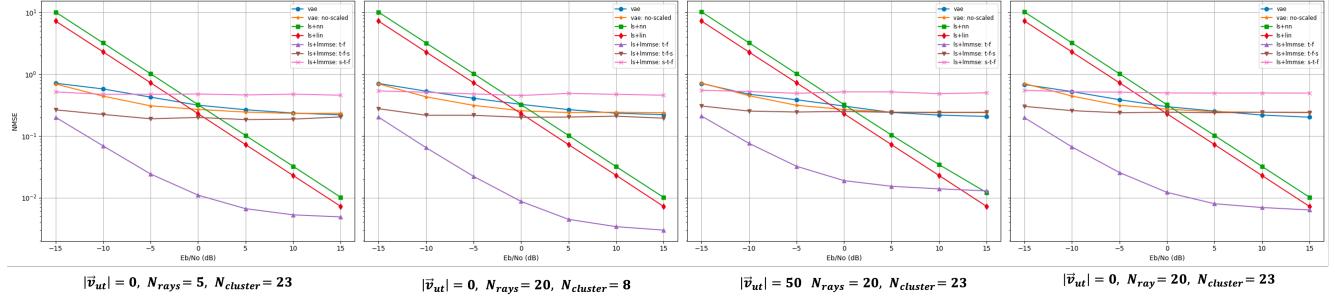


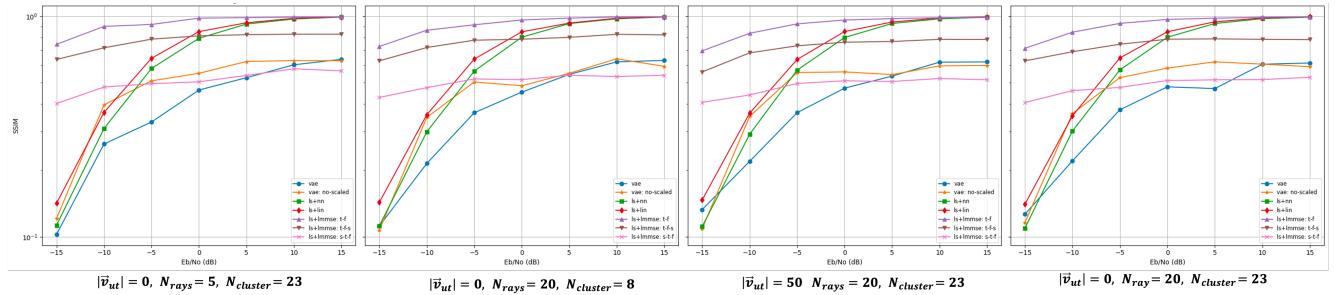
Fig. 8. Distribution of input and output of the encoder model, $p_{\underline{\mathbf{R}}_{\text{pilot}}^{\text{ofdm}}}(\vec{\mathbf{R}}_{\text{pilot}}^{\text{ofdm}})$, $p(\mathbf{a}_{g,\text{enc}}|\vec{\mathbf{R}}_{\text{pilot}}^{\text{ofdm}})$, the latent space $p(\mathbf{z}) \sim p(\mathbf{z}|\mathbf{a}_{g,\text{enc}})$, and decode output $q(\hat{\mathbf{H}}_{\text{freq}}|\mathbf{z}) \approx p(\mathbf{H}_{\text{freq}}|\mathbf{z})$

achieves lower NMSE values than LS+LIN and LS+NN. It is observed that the VAE without scaling to $[-1, 1]$ fluctuates more due to boundary effects during channel estimation, but still provides better NMSE at low SNR. Furthermore, the Doppler shift effect, \mathbf{v}_{ut} , has no impact on the conventional estimation methods. To clarify, Fig. 9b presents SSIM scores to visualize the structural similarity between the generated and real channels. Among the methods, the LS+LMMSE approach achieves the highest performance, while the proposed model yields the lowest similarity scores. However, a more comprehensive evaluation using the end-to-end BER metric reveals that the proposed channel estimation model emphasizes the most impactful characteristics, rather than strictly reproducing the full channel representation.

To assess the quality of the proposed channel model, realistic evaluation data are used based on the ray-tracing simulation settings shown in Fig. 5. Considering $|\mathbf{v}_{\text{ut}}| = 0 \text{ m/s}$, the scenario distribution depicted in Fig. 10a closely resembles the simulated CDL-B channel characteristics. This leads to the evaluation results presented in Fig. 10b, where the proposed model achieves lower NMSE compared to the LS+NN and LS+LIN methods. Despite obtaining the lowest SSIM scores, the proposed VAENet model still delivers competitive performance relative to the LS+LMMSE approach. This outcome confirms that the AI model focuses on the most impactful channel characteristics, as highlighted by the AM4C explainability method. However, when considering the most impactful factor the Doppler shift with $|\mathbf{v}_{\text{ut}}| = 50 \text{ m/s}$, the distribution of the real channel scenario slightly differs from that of the simulated CDL-B channel model shown in Fig. 3. This discrepancy may be attributed to the influence of geometric characteristics, specifically the locations of the UT and BS in the real-world scenario, which are not captured by the CDL-B simulation model.



(a) NMSE↓ Scores of MIMO system with different geometric settings



(b) SSIM↑ Scores of MIMO system with different geometric settings at

Fig. 9. NMSE and SSIM evaluation scores of channel estimation models in MIMO-OFDM systems with $K_{\text{ut,ant}} = 2$ and $K_{\text{bs,ant}} = 4$ across different geometric channel configurations in HD-TFS.

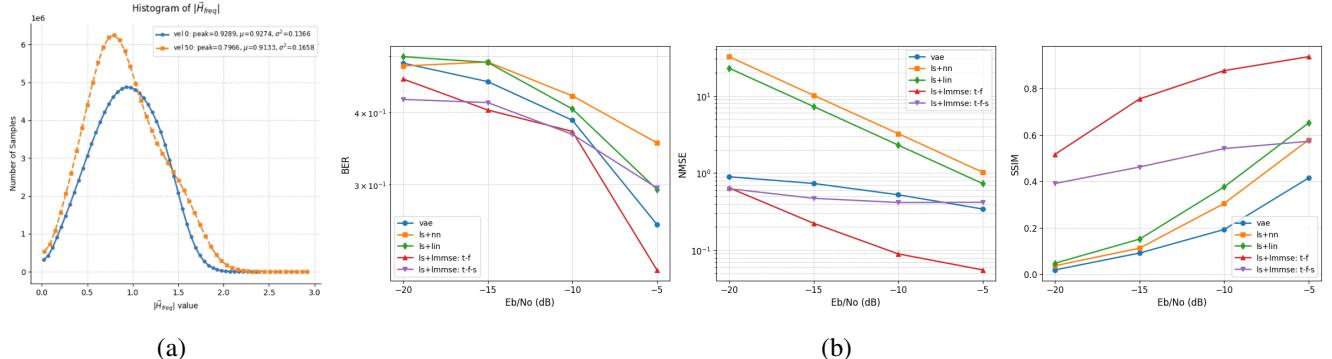


Fig. 10. (a) Ray tracing channel distribution at $|v_{\text{ut}}| = 0 \text{ m/s}$ and $|v_{\text{ut}}| = 50 \text{ m/s}$. (b) BER, NMSE, and SSIM results for the MIMO-OFDM simulation scenario with $K_{\text{ut,ant}} = 2$ and $K_{\text{bs,ant}} = 4$.

The similarity between the real and estimated channels is insufficient to fully assess the effectiveness of a channel estimation method, as the decoder can often recover the imperfections in the estimated channel. Therefore, an end-to-end pipeline metric, the BER score is presented in Table III. The evaluation focuses on the most significant geometric factor affecting performance: the Doppler shift at low SNR. The results under the CDL-B simulation scenario demonstrate that the LS+LMMSE method consistently outperforms other approaches in terms of BER. However, the generative model, VAE-Net,

surpasses the performance of the LS+NN and LS+LIN methods. Additionally, the VAE model achieves BER scores comparable to those of LS+LMMSE at a low SNR of -20 dB. In contrast, the unscaled VAE model performs poorly in terms of BER, confirming that the $[-1, 1]$ scaling is essential for effective generative channel estimation. When evaluated in a real world scenario at Polytechnique Montréal, all estimation methods demonstrate strong BER performance. This highlighted strengths of the proposed methods by maintaining consistent performance in real world settings when evaluated on a pretrained CDL-B simulated channel model.

TABLE III

BIT ERROR RATES (BER \downarrow) FOR ANTENNA CONFIGURATION $N_{\text{ut,ant}} = 2$, $N_{\text{bs,ant}} = 4$, AT A SPEED OF 50 m/s.

Scenario	SNR (dB)	VAE	VAE (no-scaled)	LS+NN	LS+LIN	LS+LMMSE (T-F)	LS+LMMSE (T-F-S)	LS+LMMSE (S-T-F)
CDL-B	-20	0.4702	0.4797	0.5122	0.5174	0.4826	0.4714	0.4592
CDL-B	-15	0.4291	0.4443	0.4714	0.4887	0.4158	0.4045	0.4019
CDL-B	-10	0.3561	0.3776	0.3976	0.4010	0.3307	0.3290	0.2786
CDL-B	-5	0.1867	0.2619	0.3090	0.2717	0.1502	0.2101	0.3177
Polymtl	-20	0.4618	0.4827	0.4965	0.4861	0.4453	0.4635	0.5243
Polymtl	-15	0.4269	0.4637	0.5017	0.4756	0.3784	0.4357	0.3984
Polymtl	-10	0.3723	0.4144	0.4227	0.3689	0.3107	0.3532	0.3211
Polymtl	-5	0.2533	0.3344	0.3064	0.2578	0.1067	0.1996	0.0529

VI. CONCLUSIONS

In this paper, we proposed an end-to-end pipeline for OFDM transmissions over a CDL channel model. Specifically, we designed a novel GAN-based generative framework that improves the reliability of geometry-based channel estimation in complex environments. Our results demonstrated that the proposed VAENet model, with the fusion of statistical geometric properties into the latent space, achieved performance comparable to conventional methods across three key evaluation metrics. Notably, the framework was evaluated on a real-world simulation scenario at Polytechnique Montréal and achieved outstanding performance. The implementation pipeline provided valuable insights into the inference runtime efficiency of VAENet compared to the advanced LS+LMMSE channel estimation model. Importantly, the explainability framework AM4C highlighted the attention areas of the CriticNet in the high-power gain regions of the HD-TFS data. With their flexibility and robustness, the proposed methods present promising opportunities for the continued development of wireless communication in increasingly complex environments.

REFERENCES

- [1] N. T. Nguyen and T. N. Do, "Generative and Explainable AI for high-dimensional channel estimation," 2025. [Online]. Available: <https://arxiv.org/abs/2504.10775>
- [2] J. Wu and P. Fan, "A survey on high mobility wireless communications: Challenges, opportunities and solutions," *IEEE Access*, vol. 4, p. 450–476, 2016. [Online]. Available: <http://dx.doi.org/10.1109/ACCESS.2016.2518085>
- [3] P. Angueira, I. Val, J. Montalban, O. Seijo, E. Iradier, P. S. Fontaneda, L. Fanari, and A. Arriola, "A survey of physical layer techniques for secure wireless communications in industry," *IEEE Communications Surveys; Tutorials*, vol. 24, no. 2, p. 810–838, 2022. [Online]. Available: <http://dx.doi.org/10.1109/COMST.2022.3148857>
- [4] P. Moose, "A technique for orthogonal frequency division multiplexing frequency offset correction," *IEEE Transactions on Communications*, vol. 42, no. 10, p. 2908–2914, 1994. [Online]. Available: <http://dx.doi.org/10.1109/26.328961>
- [5] S. Lee, C. Lee, and D. Sim, "Performance analysis for optimal pilot spacing in MIMO-OFDM systems," *IEEE Transactions on Vehicular Technology*, p. 1–14, 2024. [Online]. Available: <http://dx.doi.org/10.1109/TVT.2024.3520949>
- [6] X. Lu, Y. Lu, J. Xu, and G. Lin, "Least Square channel estimation for MIMO-OFDM system," in *2009 5th International Conference on Wireless Communications, Networking and Mobile Computing*. IEEE, Sep. 2009, p. 1–4. [Online]. Available: <http://dx.doi.org/10.1109/WICOM.2009.5302824>
- [7] X. Dong, W.-s. Lu, and A. Soong, "Linear interpolation in pilot symbol assisted channel estimation for OFDM," *IEEE Transactions on Wireless Communications*, vol. 6, no. 5, p. 1910–1920, May 2007. [Online]. Available: <http://dx.doi.org/10.1109/TWC.2007.360392>
- [8] S. Coleri, M. Ergen, A. Puri, and A. Bahai, "Channel estimation techniques based on pilot arrangement in OFDM systems," *IEEE Transactions on Broadcasting*, vol. 48, no. 3, p. 223–229, Sep. 2002. [Online]. Available: <http://dx.doi.org/10.1109/TBC.2002.804034>
- [9] A. Alhammadi, I. Shayea, A. A. El-Saleh, M. H. Azmi, Z. H. Ismail, L. Kouhalvandi, and S. A. Saad, "Artificial intelligence in 6G wireless networks: Opportunities, applications, and challenges," *International Journal of Intelligent Systems*, vol. 2024, p. 1–27, Mar. 2024. [Online]. Available: <http://dx.doi.org/10.1155/2024/8845070>
- [10] J. Hoydis, S. Cammerer, F. Ait Aoudia, M. Nimier-David, L. Maggi, G. Marcus, A. Vem, and A. Keller, "Sionna," 2022, <https://nvlabs.github.io/sionna/>.
- [11] F. Aït Aoudia, J. Hoydis, M. Nimier-David, S. Cammerer, and A. Keller, "Sionna RT: Technical report," *arXiv:2504.21719*, Apr 2025. [Online]. Available: <https://arxiv.org/abs/2504.21719>
- [12] A. Melgar, A. de la Fuente, L. Carro-Calvo, O. Barquero-Perez, and E. Morgado, "Deep Neural Network: An alternative to traditional channel estimators in massive MIMO systems," *IEEE Transactions on Cognitive Communications and Networking*, vol. 8, no. 2, p. 657–671, Jun. 2022. [Online]. Available: <http://dx.doi.org/10.1109/TCCN.2022.3164888>
- [13] Haider et al., "GAN-based channel estimation for IRS-aided communication systems," *IEEE Transactions on Vehicular Technology*, vol. 73, no. 4, p. 6012–6017, Apr. 2024.
- [14] M. Baur, B. Fesl, and W. Utschick, "Leveraging Variational Autoencoders for parameterized MMSE Estimation," *IEEE Transactions on Signal Processing*, vol. 72, p. 3731–3744, 2024. [Online]. Available: <http://dx.doi.org/10.1109/TSP.2024.3439097>
- [15] L. A. Santalo and M. Kac, "Editor's statement," in *Integral Geometry and Geometric Probability*. Cambridge: Cambridge University Press, Oct. 2004, pp. xi–xii.
- [16] F. Matheron, *Random sets and integral geometry*, ser. Probability and Mathematical Statistics S. Nashville, TN: John Wiley and Sons, Feb. 1975.
- [17] D. He, B. Ai, K. Guan, L. Wang, Z. Zhong, and T. Kürner, "The design and applications of high-performance ray-tracing simulation platform for 5G and beyond wireless communications: A tutorial," *IEEE Communications Surveys andamp; Tutorials*, vol. 21, no. 1, p. 10–27, 2019. [Online]. Available: <http://dx.doi.org/10.1109/COMST.2018.2865724>

- [18] A. Saleh and R. Valenzuela, "A statistical model for indoor multipath propagation," *IEEE Journal on Selected Areas in Communications*, vol. 5, no. 2, p. 128–137, Feb. 1987. [Online]. Available: <http://dx.doi.org/10.1109/JSAC.1987.1146527>
- [19] W. Gu, Y. Liu, C.-X. Wang, W. Xu, Y. Yu, W.-J. Lu, and H.-B. Zhu, "A general 3-D geometry-based stochastic channel model for B5G mmWave IIoT," *IEEE Internet of Things Journal*, vol. 11, no. 2, p. 3362–3376, Jan. 2024. [Online]. Available: <http://dx.doi.org/10.1109/JIOT.2023.3297621>
- [20] M. Wang, X. Chen, S. Wang, and W. Shi, "Three-dimensional geometry-based channel modeling and simulations for reconfigurable intelligent surface-assisted uav-to-ground mimo communications," *IET Communications*, vol. 19, no. 1, Jan. 2024. [Online]. Available: <http://dx.doi.org/10.1049/cmu2.12724>
- [21] E. Balevi, A. Doshi, A. Jalal, A. Dimakis, and J. G. Andrews, "High dimensional channel estimation using deep generative networks," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 1, p. 18–30, Jan. 2021. [Online]. Available: <http://dx.doi.org/10.1109/JSAC.2020.3036947>
- [22] L. Zhang, C.-X. Wang, Z. Zhou, L. Xin, J. Huang, S. Yang, and E.-H. M. Aggoune, "A novel geometry-based stochastic model for indoor scenarios incorporating dense multipath components towards standardization," *IEEE Transactions on Vehicular Technology*, p. 1–16, 2025. [Online]. Available: <http://dx.doi.org/10.1109/TVT.2025.3548077>
- [23] H. Ye, L. Liang, G. Y. Li, and B.-H. Juang, "Deep learning-based end-to-end wireless communication systems with conditional gans as unknown channels," *IEEE Transactions on Wireless Communications*, vol. 19, no. 5, p. 3133–3143, May 2020. [Online]. Available: <http://dx.doi.org/10.1109/TWC.2020.2970707>
- [24] B. J. B. Deutschmann, E. Leitinger, and K. Witrisal, "Geometry-based channel estimation, prediction, and fusion," 2025. [Online]. Available: <https://arxiv.org/abs/2503.17868>
- [25] M. Razghandi, H. Zhou, M. Erol-Kantarci, and D. Turgut, "Smart home energy management: VAE-GAN synthetic dataset generator and Q-Learning," *IEEE Transactions on Smart Grid*, vol. 15, no. 2, p. 1562–1573, Mar. 2024. [Online]. Available: <http://dx.doi.org/10.1109/TSG.2023.3288824>
- [26] ETSI Technical Report, "Study on channel model for frequencies from 0.5 to 100 GHz (3GPP TR 38.901 version 16.1.0 Release 16)," https://www.etsi.org/deliver/etsi_tr/138900_138999/138901/16.01.00_60/tr_138901v160100p.pdf, [Accessed 31-07-2025].
- [27] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. Courville, "Improved training of Wasserstein GANs," in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, 2017, p. 5769–5779.
- [28] M. Arjovsky and L. Bottou, "Towards principled methods for training Generative Adversarial Networks," in *Proceedings of the International Conference on Learning Representations*, Apr. 2017.
- [29] D. P. Kingma and M. Welling, "Auto-Encoding Variational Bayes," *CoRR*, vol. abs/1312.6114, 2013. [Online]. Available: <https://api.semanticscholar.org/CorpusID:216078090>