Robust MU-MIMO Communication via Learning-Enhanced Receiver and Site-Specific Channel Modeling

Ren wei Ou, Rui si Li.

Sun Yat-sen University, New York University
ourw210@gamil.com, rl5411@nyu.edu

Abstract

As the evolution toward next-generation wireless networks gains momentum, there is a growing need for intelligent and environment-aware communication systems. This paper presents an environment-aware neural receiver for multi-user multiple-input multiple-output (MU-MIMO) uplink detection. By integrating site-specific channel modelling with a prior-enhanced deep learning architecture, our framework achieves robust signal detection under realistic propagation conditions while maintaining compatibility with existing wireless infrastructure. The proposed method incorporates linear minimum mean square error (LMMSE) derived log-likelihood ratio (LLR) priors to stabilize training convergence and employs a lightweight MobileNet-based 3D convolution (MBConv3D) network to refine channel estimates and optimize LLR outputs. Extensive simulations in urban scenarios demonstrate that our approach operates effectively in low-to-medium SNR ranges, closely approaching the performance bound of perfect channel state information (CSI). The results validate the framework's strong potential for practical 5G/6G deployments, offering a viable pathway toward intelligent receiver design for next-generation wireless systems.

Introduction

The evolution of wireless communication systems towards 5G and 6G has increasingly relied on multiple-input multiple-output (MIMO) technology to meet escalating demands for data throughput, spectral efficiency, and reliable connectivity in diverse environments (Wang et al. 2024). A cornerstone of modern MIMO-enabled networks is the multi-user MIMO (MU-MIMO) uplink, where the potential of multiple antennas at the base stations (BSs) is leveraged to serve several user equipments (UEs) simultaneously. In such configurations, BSs process received signals from multiple UEs simultaneously, which introduces significant inter-user interference (IUI), where the data streams of different UEs interfere with each other due to the simultaneous transmission, complicating signal detection and decoding at the receiver.

Conventional linear MIMO detectors, such as zero-forcing (ZF) and minimum mean square error (MMSE), are computationally simple but prone to performance degradation. Their reliance on idealized channel models makes them highly sensitive to estimation errors and real-world impairments. These limitations have motivated the exploration of data-driven alternatives. In recent years, deep learning (DL) has emerged as a powerful paradigm for wireless communication receivers, enabling end-to-end (E2E) learning that dynamically adapts to intricate signal processing demands without relying on explicit model assumptions (O'shea and Hoydis 2017). The main idea of a DL-based receiver is to employ deep neural networks (DNNs) to replace traditional functional blocks, such as channel estimation (CE), equalization, and decoding, with a single, unified data-driven model that can be trained to directly recover transmitted information bits or log-likelihood ratios (LLRs) from the raw received signal, effectively compensating for various physical impairments in a data-driven manner. Initial successes of this paradigm in single-input single-output (SISO) systems demonstrated superior bit error rate (BER) performance compared to conventional signal detectors (Honkala, Korpi, and Huttunen 2021). Extending these to MIMO, particularly MU-MIMO in orthogonal frequency-division multiplexing (OFDM) frameworks, has shown promise in approximating maximum likelihood detection with reduced complexity (Albreem et al. 2021; Goutay et al. 2021).

However, prior DL-based MU-MIMO detectors exhibit several methodological and practical limitations. From a practical deployment perspective, one limitation of most existing studies is their reliance on stochastic channel models (e.g., Rayleigh or Rician fading), which, while tractable, often fail to capture the intricate effects of real-world propagation environments, such as specific building layouts, terrain, and vegetation. Consequently, the real-world performance of these DL-based receivers in site-specific scenarios remains largely unknown. From a methodological perspective, while many DL architectures successfully unfold traditional iterative algorithms into neural layers, they inevitably inherit the underlying sensitivities to channel correlation and imperfect CSI, which often result in error propagation and instability (Liu, Thompson, and Arslan 2022; He et al. 2020). This limitation is further compounded in MU-MIMO OFDM settings, where direct end-to-end training on log-likelihood ratios may encounter convergence difficulties due to the multi-user interference and CSI estimation errors from limited pilots. More importantly, these stand-alone neural models often overlook integration with existing physical-layer hard-

ware, which limits their scalability, real-time processing capabilities, and compatibility with existing 5G infrastructure (Goutay et al. 2021).

Based on the above discussion, this paper proposes a ray-tracing-assisted neural receiver framework for robust MU-MIMO uplink detection. This framework leverages deterministic ray-tracing simulations to generate site-specific channel responses, incorporating environmental effects like multipath reflections and scattering for site-specific practical deployments. Specifically, we integrated LLR priors, derived from traditional linear minimum mean square error (LMMSE) detection based on least squares (LS)-estimated CSI, into the neural receiver pipeline to stabilize training convergence and ensure seamless compatibility with current OFDM-based systems. To achieve this integration, a lightweight MobileNet-based (Howard et al. 2019) 3D convolution (MBConv3D) network was employed to refine CSI estimates using mean squared error (MSE) loss and optimize LLRs via binary cross-entropy (BCE) loss, enabling the receiver to learn environment-specific features while effectively compensating for IUI and other channel impairments to reduce BER. The main contributions of this work are summarized as follows:

- We introduce a novel DL-based receiver framework that incorporates ray-tracing for site-specific channel response with traditional priors to achieve stable, hardware-compatible operation in MU-MIMO detection;
- Extensive simulations in urban scenarios demonstrate that our proposed framework operates effectively in low-to-medium SNR ranges, closely approaches the performance bound of perfect CSI, and showcases strong potential for practical 5G/6G deployments.

Related Works

The core challenge in MU-MIMO detection lies in mitigating inter-user interference. Existing solutions can be broadly categorized into two paradigms: traditional model-based detection and data-driven deep learning approaches.

Conventional MIMO detection strategies to combat interference encompass both linear and non-linear iterative algorithms. Linear methods, such as zero-forcing (ZF), minimum mean square error (MMSE), and MMSE with successive interference cancellation (MMSE-SIC) (Liu 2009; Kong et al. 2016), have been widely adopted due to their low computational complexity and straightforward implementation. In addition, a class of non-linear iterative detectors has emerged, leveraging probabilistic inference to approximate the optimal Bayesian solution. Among them, detectors based on approximate message passing (AMP) (Jeon et al. 2015) and expectation propagation (EP) (Cespedes et al. 2014) have gained significant attention for their compelling performance-complexity trade-off. The limitation of these iterative methods, however, is that their superior performance often deteriorates in practical small-scale systems, particularly with channel correlation or imperfect CSI, where their underlying idealized assumptions often do not hold.

Driven by the limitations of model-based approaches, significant research has explored using neural networks to implement end-to-end receivers (O'shea and Hoydis 2017; He et al. 2020; Cammerer et al. 2023). MIMO detection presents a natural application for machine learning techniques, and a wide range of methods have been investigated, including fully data-driven networks, model-driven unfolded architectures, and hybrid designs. For instance, Ye and Liang (Ye and Liang 2025) introduced ChannelNet, a purely data-driven massive MIMO detector that embeds channel information into neural layers to maintain permutation equivariance across antennas. This approach enables efficient scaling to large antenna counts and high modulation orders with $\mathcal{O}(N_t N_r)$ complexity under diverse channel distributions. In another approach, deep unfolding-aided Gaussian belief propagation (BP) is employed for correlated large-scale MU-MIMO detection, where traditional BP iterations are parameterized and optimized end-to-end to accelerate convergence and mitigate the impact of spatial correlation under imperfect CSI (Shirase et al. 2020). Notably, the application of Graph Neural Networks (GNNs) to MIMO detection has become a notable trend in recent years (Cammerer et al. 2023; Chowdhury et al. 2023), primarily because they can naturally model the factor graph of a MIMO system, enabling effective message passing and inference over the underlying graphical model.

Despite the successes achieved by the above works in various scenarios, a significant gap remains in the existing body of deep learning-based receiver research: the lack of thorough investigation into their performance under site-specific channel conditions and the challenges posed by issues like pilot contamination in massive MIMO deployments. Furthermore, the complexity and sensitivity of many proposed networks to the channel characteristics assumed during training present a barrier to practical deployment. To enhance the generalization capability and practical robustness of our model, we integrate deep learning modules with physical-layer building blocks, creating a hybrid architecture that leverages the adaptability of data-driven methods.

System Model

This paper focuses on a MU-MIMO uplink communication scenario, which can be extended to the systems for both uplink and downlink transmissions. The overall system architecture is depicted in Figure 1. The transceiver design adheres to the specifications outlined by the 3rd Generation Partnership Project (3GPP) for the 5G physical uplink shared channel (PUSCH), ensuring compatibility with modern standards.

System Configuration

We consider a configuration where there are N_u users, each equipped with N_{tx} transmit antennas, communicating with a base station that has N_{rx} receive antennas configured in a uniform linear array (ULA). At the transmitter side for each user, the source information bits are first encoded using a low-density parity-check (LDPC) code to ensure reliable transmission. Then, the resulting encoded bits are sub-

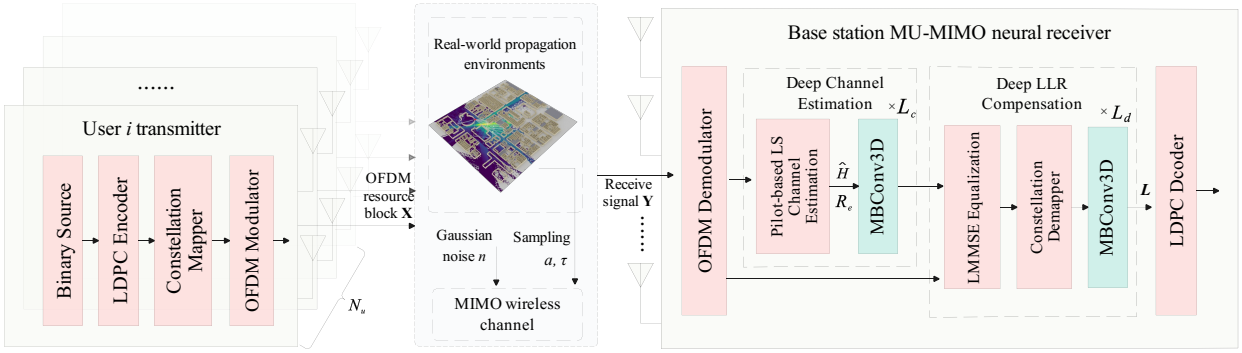


Figure 1: The architecture of the proposed MU-MIMO uplink transceiver with site-specific ray-tracing channel modeling.

sequently mapped independently to constellation points chosen from an energy-normalized constellation, which is typically M-ary quadrature amplitude modulation (M-QAM), where symbols are often Gray-coded for minimized bit error probability. For simplicity, we assume that all users employ the same modulation scheme, although in practical systems, different users may utilize varying modulation orders (e.g., 16-QAM, 64-QAM) based on their channel conditions. These data symbols, along with pilot symbols for channel estimation, are allocated to their respective resource elements (REs) within the OFDM resource block (RB). The pilots are known sequences that are predefined for both the transmitter and the receiver to facilitate accurate channel estimation.

The OFDM RB is defined with N_f subcarriers in the frequency domain and N_t symbols in the time domain. A typical value for N_t in 5G specifications is 14, corresponding to one time slot. For any index pair (t, f) within the resource block, it represents the constellation point carried on the f-th subcarrier of the t-th OFDM symbol, where t denotes the symbol index ($t = 1, 2, \ldots, N_t$) and f denotes the subcarrier index ($f = 1, 2, \ldots, N_f$). To represent the transmitted and received signals, we define the transmitted tensor $\mathbf{X} \in \mathbb{C}^{N_u \times N_{tx} \times N_t \times N_f}$, which comprises all users' data. The corresponding received signal tensor at the base station is $\mathbf{Y} \in \mathbb{C}^{N_{rx} \times N_t \times N_f}$, and the channel frequency response (CFR) is given by the five-dimensional tensor $\mathbf{H} \in \mathbb{C}^{N_u \times N_{rx} \times N_{tx} \times N_t \times N_f}$. We assume that each transmit antenna is allocated an independent RB, meaning that each user can transmit N_{tx} symbols simultaneously across N_{tx} different RBs. Let $\mathbf{x}_{t,f}^{(u)} \in \mathbb{C}^{N_{tx} \times 1}$ denote the symbol vector transmitted by user u (where $u = 1, 2, \ldots, N_u$) across its N_{tx} antennas. This vector is defined as: $\mathbf{x}_{t,f}^{(u)} = \left[x_{t,f,1}^{(u)}, \ldots, x_{t,f,n}^{(u)}, \ldots, x_{t,f,N_{tx}}^{(u)}\right]^T$, where $n = 1, 2, \ldots, N_{tx}$. Here, $x_{t,f,n}^{(u)}$ denotes the symbol transmitted by the n-th antenna of user u on the position (t, f) within its allocated RB.

The received symbol vector $\mathbf{y}_{t,f} \in \mathbb{C}^{N_{rx} \times 1}$ from all users at the base station is expressed in the frequency domain as:

$$\mathbf{y}_{t,f} = \sum_{u=1}^{N_u} \mathbf{H}_{t,f}^{(u)} \mathbf{x}_{t,f}^{(u)} + \mathbf{n}_{t,f}, \tag{1}$$

where $\mathbf{H}_{t,f}^{(u)} \in \mathbb{C}^{N_{rx} \times N_{tx}}$ is the frequency-domain channel matrix for user u at RB position (t, f), and $\mathbf{n}_{t,f} \in \mathbb{C}^{N_{rx} \times 1}$ represents additive white Gaussian noise with variance σ^2. After all symbols in the RB have been transmitted, the receiver aggregates the signals into a flattened tensor $\mathbf{Y}' \in \mathbb{C}^{N_{rx} \times (N_t N_f)}$, which can be subsequently reshaped to form $\mathbf{Y} \in \mathbb{C}^{N_{rx} \times N_t \times N_f}$. The primary objective of this paper is to recover the transmitted tensor $\mathbf{X}$ at the receiver from $\mathbf{Y}$ with the lowest possible BER, leveraging DL-based channel estimation and MIMO detection techniques.

Channel Modelling

We consider time-varying channels to account for dynamic environmental effects and mobility in realistic wireless scenarios. Different from previous works that rely on statistical channel models, our approach adopts deterministic channel modelling via ray-tracing tailored to specific sites.

For each propagation path $l = 1, \ldots, L$, the complex path gain a_l and delay τ_l are calculated as:

$$a_l = \sqrt{P_t} \cdot G_t \cdot G_r \cdot \frac{\lambda}{4\pi d_l} e^{-j\frac{2\pi d_l}{\lambda}} \prod_{m=1}^{M_l} \Gamma_{l,m} \cdot \prod_{p=1}^{P_l} D_{l,p}, \tag{2}$$

where P_t denotes the transmit power; G_t and G_r represent the transmit and receive antenna gains, respectively; $\lambda = c/f_c$ is the wavelength with c being the speed of light and f_c the carrier frequency. For the l-th propagation path, d_l is the path length, $\Gamma_{l,m}$ and $D_{l,p}$ denote the reflection and diffraction coefficients, respectively, while M_l and P_l are the total numbers of reflections and diffractions along the path. It is worth mentioning that the values of these reflection and diffraction coefficients are related to the electromagnetic properties of environmental materials and the radiation patterns of the antennas. For detailed computations of these coefficients, refer to (Tse and Viswanath 2005). The path delay τ_l is determined by $\tau_l = d_l/c$. Given this, the total received signal power P_r is obtained by aggregating the power from all L paths as $P_r = P_t \sum_{l=1}^{L} |a_l|^2$. Finally, the

uplink signal-to-noise ratio (SNR) is given by: SNR $= \frac{P_r}{\sigma^2}$, where σ^2 is the noise variance.

The channel impulse response (CIR) at time t is:

$$h(\tau, t) = \sum_{l=1}^{L} a_l(t)\delta(\tau - \tau_l(t)), \tag{3}$$

where $a_l(t)$ and $\tau_l(t)$ are time-varying. Applying the Fourier transform yields the CFR:

$$H_{t,f} = \sum_{l=1}^{L} a_l(t)e^{-j2\pi f \tau_l(t)}, \tag{4}$$

which $H_{t,f}$ represents the CFR at subcarrier f and symbol t.

Channel Estimation and Equalization

LS Channel Estimation We employ the most commonly used LS method for channel estimation by leveraging predefined pilot symbols in the RB. For user u and its n-th transmit antenna, the pilot symbol at (t_p, f_p) is denoted as $x^{(u)}_{t_p,f_p,n} \in \mathbb{C}$, where t_p and f_p specifically denote the positions of pilots within the RB. We decompose the channel matrix column-wise as $\mathbf{H}^{(u)}_{t,f} = [\mathbf{h}^{(u)}_{t,f,1}, \ldots, \mathbf{h}^{(u)}_{t,f,N_{tx}}]$, where each vector $\mathbf{h}^{(u)}_{t,f,n} \in \mathbb{C}^{N_{rx}\times 1}$ represents the CFR from the n-th transmit antenna of user u to all N_{rx} receive antennas at the resource element (t, f). To avoid pilot contamination, the pilots are orthogonally allocated across all users and transmit antennas, and their power is normalized such that $|\mathbf{x}^{(u)}_{t_p,f_p,n}|^2 = 1$. The estimate for the n-th column is:

$$\hat{\mathbf{h}}^{(u)}_{t_p,f_p,n} = \mathbf{y}_{t_p,f_p} \cdot (x^{(u)}_{t_p,f_p,n})^*. \tag{5}$$

The estimation error is $\mathbf{e}^{(u)}_{t_p,f_p,n} = \hat{\mathbf{h}}^{(u)}_{t_p,f_p,n} - \mathbf{h}^{(u)}_{t_p,f_p,n} \in \mathbb{C}^{N_{rx}\times 1}$. Thus, the error covariance is:

$$\mathbf{R}_e = \mathbb{E}[\mathbf{e}\mathbf{e}^H] = \mathbb{E}[\mathbf{n}\mathbf{n}^H] = \sigma^2\mathbf{I}_{N_{rx}} \in \mathbb{C}^{N_{rx}\times N_{rx}}. \tag{6}$$

As can be observed, the orthogonal placement of pilots eliminates inter-user interference during channel estimation, resulting in an estimation error equal to the noise variance.

For non-pilot REs, we use nearest-neighbor interpolation, where the CFR at a given RE is assigned the estimate from the nearest pilot RE. Consequently, the estimation error at non-pilot positions equals that of their closest pilot RE. This process yields the full channel estimate $\hat{\mathbf{H}}^{(u)}_{t,f}$ per user, then we stack $\hat{\mathbf{H}}^{(u)}_{t,f}$ across all users to form $\hat{\mathbf{H}}$ for downstream processing.

Deep Channel Estimation (DeepCE) Module Interpolation-based approximations introduce estimation errors at REs far from pilots, and LS estimation overlooks the inherent spatial correlations in MIMO systems, yielding suboptimal precision. Actually, CSI inherently contains rich spatial and temporal features that are highly exploitable for refining the coarse channel estimate. Thus, we incorporate $\hat{\mathbf{H}}$ and its associated error covariance $\mathbf{R}_e$ as priors, feeding them into a sequence of MBConv3D modules for enhanced estimation in a data-driven manner.

Specifically, we separate $\hat{\mathbf{H}} \in \mathbb{C}^{N_u\times N_{rx}\times N_{tx}\times N_t\times N_f}$ into real and imaginary components, then reshape it to $N_u \times N_t \times N_f \times (2N_{rx}N_{tx})$, and concatenate it with the covariance information. A cascade of L_c MBConv3D modules extracts latent features, followed by recombination into complex form to produce $\tilde{\mathbf{H}}$.

$$\tilde{\mathbf{H}} = \text{MBConv3D}_{L_c}\left(\left[\text{Re}(\hat{\mathbf{H}}), \text{Im}(\hat{\mathbf{H}}), \mathbf{R}_e\right]\right), \tag{7}$$

where $\text{Re}(\cdot)$ and $\text{Im}(\cdot)$ extract the real and imaginary parts, respectively. This $\tilde{\mathbf{H}}$ is subsequently utilized for signal equalization and detection.

LMMSE Channel Equalization The LMMSE equalizer is widely used in MIMO systems due to its balance between performance and complexity, effectively mitigating inter-antenna and inter-user interference while accounting for noise. Leveraging the refined channel estimate $\tilde{\mathbf{H}}$, we compute the equalization matrix to recover the transmitted symbols. Specifically, for each user u, the LMMSE equalizer is applied to $\mathbf{y}_{t,f}$ using $\tilde{\mathbf{H}}^{(u)}_{t,f}$:

$$\hat{\mathbf{x}}^{(u)}_{t,f} = \left((\tilde{\mathbf{H}}^{(u)}_{t,f})^H\tilde{\mathbf{H}}^{(u)}_{t,f} + \sigma^2\mathbf{I}_{N_{tx}}\right)^{-1}(\tilde{\mathbf{H}}^{(u)}_{t,f})^H\mathbf{y}_{t,f}. \tag{8}$$

This formulation minimizes the mean squared error between the estimated and true symbols, providing robustness against noise amplification compared to zero-forcing alternatives. Aggregating across all REs and users yields the detected tensor $\hat{\mathbf{X}} \in \mathbb{C}^{N_u\times N_{tx}\times N_t\times N_f}$, which is used for subsequent demapping and decoding.

Neural Network-based LLR Optimization

For M-QAM modulation, where each symbol carries $\log_2 M$ bits, the demapper computes the LLR for the k-th bit ($k = 1, \ldots, \log_2 M$) at resource element (t, f) as:

$$L(t, f, k) = \log\frac{P(b_k = 1 \mid \hat{x}_{t,f})}{P(b_k = 0 \mid \hat{x}_{t,f})}. \tag{9}$$

In practice, this calculation is often approximated via the max-log method for computational efficiency:

$$L(t, f, k) = \frac{1}{2\sigma^2}\left(\min_{s\in\mathcal{S}_0^k}|\hat{x}_{t,f} - s|^2 - \min_{s\in\mathcal{S}_1^k}|\hat{x}_{t,f} - s|^2\right), \tag{10}$$

where $\mathcal{S}_0^k$ and $\mathcal{S}_1^k$ are the subsets of the constellation points where the k-th bit is 0 or 1, respectively.

The initial LLRs, obtained through mathematical demapping, are fundamentally computed based on the distance between the equalized symbol and the ideal constellation points. However, this distance-based approach is inherently local and overlooks the structured nature of residual interference and noise across the spatial, temporal, and frequency dimensions. To alleviate this, our method leverages the multi-dimensional feature extraction capability of MBConv3D modules, which are particularly well-suited for capturing the joint correlations in time, frequency, and spatial domain to effectively compensate for

signal impairments.. Similarly, the initial LLR tensor $\mathbf{L} \in \mathbb{R}^{N_u \times N_{tx} \times N_t \times N_f \times \log_2 M}$ is first reshaped by merging the N_t and N_f dimensions, resulting in a 4D tensor of size $N_u \times N_{tx} \times (N_t N_f) \times \log_2 M$. This tensor is then processed through L_d cascaded MBConv3D modules, yielding refined LLRs:

$$\mathbf{L}' = \text{MBConv3D}_{L_d}(\mathbf{L}). \tag{11}$$

These refined LLRs $\mathbf{L}'$ are subsequently fed into the LDPC decoder to obtain final bit decisions and evaluate the BER performance under different SNR conditions.

Model Optimization

The DeepCE module optimizes via MSE loss:

$$\mathcal{L}_{ce} = \frac{1}{N_u N_{rx} N_{tx} N_t N_f} \sum |\tilde{\mathbf{H}} - \mathbf{H}|^2, \tag{12}$$

minimizing deviation from ground-truth channels element-wise. The DeepLLR refinement module employs BCE loss:

$$\mathcal{L}_{llr} = -\frac{1}{N_{bits}} \sum_{i=1}^{N_{bits}} \left[b_i \log(\sigma(L_i')) + (1 - b_i) \log(1 - \sigma(L_i')) \right], \tag{13}$$

where $b_i \in \{0, 1\}$ are true bits, $\sigma(\cdot)$ the sigmoid function, and N_{bits} the total bit count.

Experimental Results and Analysis

Simulation Setting

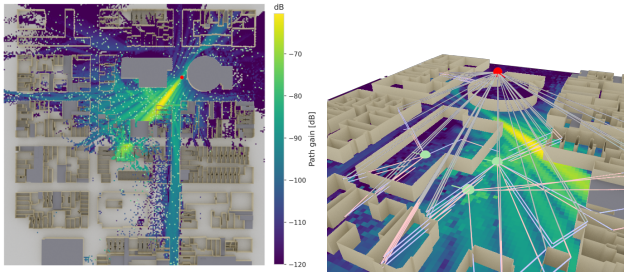
In our study, we consider site-specific MU-MIMO simulations. For realistic training and evaluation, a 500 × 500 m urban area in New York City was selected from OpenStreetMap[1]. The buildings within this region were imported into Blender, where material properties were configured, as illustrated in Figure 2. Then ray-tracing simulations were performed in this region using Sionna RT (Hoydis et al. 2023), an open-source, differentiable ray tracer designed for radio propagation modelling. A BS was deployed at a suitable location with a height of 30 m and a transmit power of $P_t = 40$ dBm, equipped with a 32-element ($N_{rx} = 32$) ULA. Based on this setup, we used the ray tracer to generate a path gain radio map across the simulated area, while multiple user terminals were randomly placed within the BS coverage to sample channel realizations during communications, which refer to path gains and path delays, as illustrated in Figure 3(a). These realizations are highly environment-dependent, capturing realistic propagation effects such as reflection and shadowing. Figure 3(b) shows the radio wave propagation trace for an example user. The number of active users was set to $N_u = 8$, each equipped with $N_{tx} = 2$ antennas and positioned at a height of 1.5 m. Each user moved randomly at speeds between 2 m/s and 20 m/s. The OFDM RB configuration followed the 5G NR standard. Each transmitted OFDM RB comprised $N_f = 72$ subcarriers in the frequency domain and $N_t = 14$ OFDM symbols in the time domain, with a center frequency of 3.5 GHz and a subcarrier spacing of 15 kHz.

[1]https://www.openstreetmap.org

During training, each iteration began with all users generating independent random bit streams. These bits were first mapped to complex-valued constellation symbols and subsequently modulated into OFDM waveforms before being transmitted through their respective antennas. The transmitted signals were then convolved with the sampled channel responses to generate the received signals. After OFDM demodulation at the receiver, the resulting signals were fed into our DL-based signal detection pipeline. The parameters of both the DeepCE and DeepLLR modules were iteratively updated based on their outputs, allowing the model to adapt to site-specific channel conditions. During evaluation, LDPC encoding with a code rate of 0.5 was introduced to enhance noise robustness, and the BER was used as the primary performance metric. All models were trained with a batch size of 16. The MBConv3D modules utilized depthwise convolution kernels with sizes of 3×3 and 5×5. Model training was conducted over 50,000 iterations within our constructed urban scenario. After training, we evaluated it over 100 iterations across different SNR levels and modulation schemes, recording the corresponding mean BER for each configuration.

(a) (b)

Figure 2: Simulation scenario setup: (a) a $500m \times 500m$ urban area selected from OpenStreetMap in New York City, and (b) the corresponding 3D scene with material properties configured in Blender.



(a) (b)

Figure 3: Simulated wireless environment: (a) Path gain radio map with BS (red) and UE (green) locations; (b) Example ray propagation trace showing transmitted (red) and reflected (blue) paths.

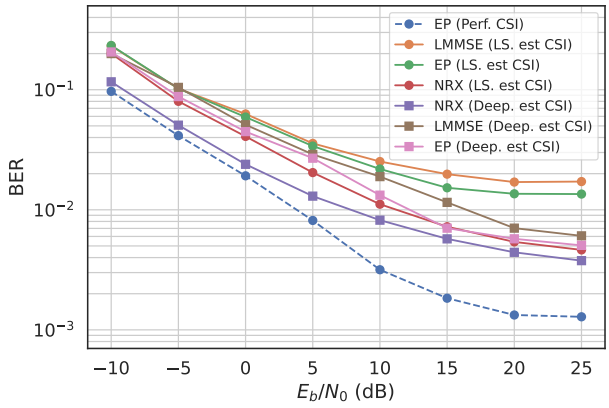
To validate the effectiveness of our proposed model, we compared it with two representative MIMO detection methods: LMMSE (linear method) and EP (non-linear method). These baselines are widely adopted in conventional MIMO receivers and provide a solid benchmark for evaluating detection performance under the same simulated environment. In addition, ablation studies were conducted to evaluate the individual contributions of the DeepCE and DeepLLR modules to the overall system performance.

Simulation Analysis

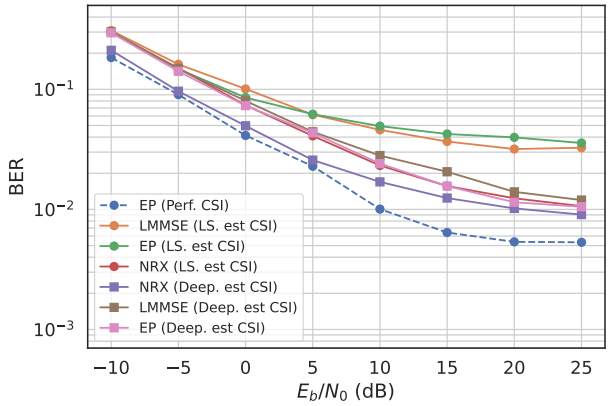
We conducted experiments under the most commonly used two modulation schemes: 16-QAM and 64-QAM, to evaluate the performance of our proposed neural receiver (NRX) pipeline in realistic site-specific environments. Under these configurations, a comprehensive evaluation was conducted, and the BER simulation results of all baselines are illustrated in Figure 4. To establish a benchmark for the best achievable performance under these conditions, we also simulated the EP detector with perfect CSI. It should be noted, however, that perfect CSI is unavailable in practice due to estimation inaccuracies and channel noise; thus, this configuration represents only a theoretical upper bound and is not feasible in actual deployments. To validate the effectiveness of individual components in our pipeline, we performed ablation studies by comparing variants with LS estimated CSI against those incorporating our DeepCE module. As can be seen, comparing the BER curves for LMMSE (LS est CSI) and LMMSE (Deep est CSI) at both 16QAM and 64QAM modulation, a noticeable reduction in BER is observed when employing DeepCE-refined CSI, with improvements becoming more pronounced at higher SNR where estimation accuracy critically impacts detection. Similar trends hold for the EP detector, where EP (Deep est CSI) consistently outperforms EP (LS est CSI) across the tested SNR range, demonstrating approximately an order of magnitude lower BER at SNR = 25 dB in both modulation schemes. These results demonstrate the effectiveness of our DeepCE module in leveraging spatial correlations and prior error covariances to yield more accurate CSI, thereby enhancing overall receiver robustness in site-specific channels.

Regarding our proposed NRX, the results also show its superior performance over traditional baselines. When using LS-estimated CSI, NRX significantly outperforms both LMMSE and EP detectors. For example, at SNR = 10 dB under 16-QAM, NRX achieves a BER of approximately 10^{-2}, representing a reduction of approximately 50% compared to the EP detector. When integrating DeepCE-estimated CSI, NRX further excels, particularly in low-to-medium SNR regimes. Although at high SNR conditions (greater than 20 dB), NRX converges closely to the EP method with negligible differences, NRX demonstrates substantial gains at lower SNRs. For instance, under 16-QAM at SNR = 0 dB, NRX (Deep est CSI) yields a BER near 2×10^{-2}, approaching the perfect CSI benchmark and outperforming EP (Deep est CSI) by 50%. In 64-QAM at SNR = 5 dB, similar trends show NRX closing the gap to perfect CSI more effectively. Notably, although the EP method also gains performance boosts from DeepCE, our DL-based LLR refinement provides an advantage in noisy conditions by learning site-specific priors.

Overall, these findings illustrate the advantages of our DeepLLR optimization in capturing non-linear interference patterns and environment-specific features, which embody the power of deep learning in end-to-end refinement of LLRs, enabling better mitigation of inter-user interference in MU-MIMO systems.



(a) 16QAM



(b) 64QAM

Figure 4: The comparison of BER performance under 16-QAM and 64-QAM modulation in our selected realistic urban scenario.

CONCLUSION

In this paper, we propose a novel ray-tracing-assisted neural network framework for MU-MIMO uplink systems. By leveraging deterministic site-specific channel modelling through ray-tracing, our method achieves environment-aware training that bridges the gap between simulated statistical channels and real-world deployments. Furthermore, integrating LLR priors from traditional LMMSE detection ensures compatibility with existing wireless infrastructure, while the MBConv3D-based modules refine channel estimates and LLRs to minimize BER. Experimental results under 16-QAM and 64-QAM modulations also demonstrate

the superiority of the proposed method over baselines like LMMSE and EP detectors. Future work will explore extending the proposed framework to various downstream tasks, including semantic communications, integrated sensing and communication, and physical-layer security, to further enhance its applicability in next-generation wireless systems.

References

Albreem, M. A.; Alhabbash, A. H.; Shahabuddin, S.; and Juntti, M. 2021. Deep learning for massive MIMO uplink detectors. *IEEE Communications Surveys & Tutorials*, 24(1): 741–766.

Cammerer, S.; Aït Aoudia, F.; Hoydis, J.; Oeldemann, A.; Roessler, A.; Mayer, T.; and Keller, A. 2023. A neural receiver for 5G NR multi-user MIMO. In *2023 IEEE Globecom Workshops (GC Wkshps)*, 329–334. IEEE.

Cespedes, J.; Olmos, P. M.; Sánchez-Fernández, M.; and Perez-Cruz, F. 2014. Expectation propagation detection for high-order high-dimensional MIMO systems. *IEEE Transactions on Communications*, 62(8): 2840–2849.

Chowdhury, A.; Verma, G.; Swami, A.; and Segarra, S. 2023. Deep graph unfolding for beamforming in MU-MIMO interference networks. *IEEE Transactions on Wireless Communications*, 23(5): 4889–4903.

Goutay, M.; Ait Aoudia, F.; Hoydis, J.; and Gorce, J.-M. 2021. Machine learning for MU-MIMO receive processing in OFDM systems. *IEEE Journal on Selected Areas in Communications*, 39(8): 2318–2332.

He, H.; Wen, C.-K.; Jin, S.; and Li, G. Y. 2020. Model-driven deep learning for MIMO detection. *IEEE Transactions on Signal Processing*, 68: 1702–1715.

Honkala, M.; Korpi, D.; and Huttunen, J. M. 2021. DeepRx: Fully convolutional deep learning receiver. *IEEE Transactions on Wireless Communications*, 20(6): 3925–3940.

Howard, A.; Sandler, M.; Chu, G.; Chen, L.-C.; Chen, B.; Tan, M.; Wang, W.; Zhu, Y.; Pang, R.; Vasudevan, V.; et al. 2019. Searching for mobilenetv3. In *Proceedings of the IEEE/CVF international conference on computer vision*, 1314–1324.

Hoydis, J.; Aït Aoudia, F.; Cammerer, S.; Nimier-David, M.; Binder, N.; Marcus, G.; and Keller, A. 2023. Sionna RT: Differentiable ray tracing for radio propagation modeling. In *2023 IEEE Globecom Workshops (GC Wkshps)*, 317–321. IEEE.

Jeon, C.; Ghods, R.; Maleki, A.; and Studer, C. 2015. Optimality of large MIMO detection via approximate message passing. In *2015 IEEE International Symposium on Information Theory (ISIT)*, 1227–1231. IEEE.

Kong, D.; Zeng, J.; Su, X.; Rong, L.; and Xu, X. 2016. Multiuser detection algorithm for PDMA uplink system based on SIC and MMSE. In *2016 IEEE/CIC International Conference on Communications in China (ICCC)*, 1–5. IEEE.

Liu, C.; Thompson, J.; and Arslan, T. 2022. A deep unfolding network for massive multi-user MIMO-OFDM detection. In *2022 IEEE Wireless Communications and Networking Conference (WCNC)*, 2405–2410. IEEE.

Liu, T.-H. 2009. Some results for the fast MMSE-SIC detection in spatially multiplexed MIMO systems. *IEEE Transactions on Wireless Communications*, 8(11): 5443–5448.

O'shea, T.; and Hoydis, J. 2017. An introduction to deep learning for the physical layer. *IEEE Transactions on Cognitive Communications and Networking*, 3(4): 563–575.

Shirase, D.; Takahashi, T.; Ibi, S.; Muraoka, K.; Ishii, N.; and Sampei, S. 2020. Deep unfolding-aided Gaussian belief propagation for correlated large MIMO detection. In *GLOBECOM 2020-2020 IEEE Global Communications Conference*, 1–6. IEEE.

Tse, D.; and Viswanath, P. 2005. *Fundamentals of wireless communication*. Cambridge university press.

Wang, Z.; Zhang, J.; Du, H.; Niyato, D.; Cui, S.; Ai, B.; Debbah, M.; Letaief, K. B.; and Poor, H. V. 2024. A tutorial on extremely large-scale MIMO for 6G: Fundamentals, signal processing, and applications. *IEEE Communications Surveys & Tutorials*, 26(3): 1560–1605.

Ye, H.; and Liang, L. 2025. On purely data-driven massive MIMO detectors. *IEEE Transactions on Signal Processing*.