Physics-Informed Neural Networks for Wireless Channel Estimation with Limited Pilot Signals

Seyed Alireza Javid **Nuria González-Prelcic**
University of California San Diego
La Jolla, CA
{sajavid, ngprelcic}@ucsd.edu

Abstract

Accurate wireless channel estimation is critical for next-generation wireless systems, enabling precise precoding for effective user separation, reduced interference across cells, and high resolution sensing, among other benefits. Traditional model-based channel estimation methods suffer, however, from performance degradation in complex environments with limited number of pilots, while purely data-driven approaches lack physical interpretability, require extensive data collection, and are usually site-specific. This paper presents a novel physics-informed neural network (PINN) framework that synergistically combines model-based channel estimation with a deep network to exploit prior information about environmental propagation characteristics and achieve superior performance under pilot-constrained scenarios. The proposed approach employs an enhanced U-Net architecture with transformer modules and cross-attention mechanisms to fuse initial channel estimates with received signal strength (RSS) maps to provide refined channel estimates. Comprehensive evaluation using realistic ray tracing data from urban environments demonstrates significant performance improvements, achieving over 5 dB gain in normalized mean squared error (NMSE) compared to state-of-the-art methods, with particularly strong performance in pilot-limited scenarios (achieving around -13 dB NMSE with only four pilots at signal-to-noise ratio (SNR) = 0 dB). The proposed framework maintains practical computational complexity, making it viable for massive multiple-input multiple-output (MIMO) systems in upper-mid band frequencies. Unlike black-box neural approaches, the physics-informed design provides a more interpretable channel estimation method. *Code is available at:* `https://github.com/ajavid34/PINN_channel-estimation`.

1 Introduction

Channel state information (CSI) is exploited by several functional blocks in a MIMO communication system, including precoding and combining, scheduling, link adaptation or equalization. High accuracy CSI is especially important when large arrays are exploited, many simultaneous users are served, the channel is wide-band, or the communication signal is also exploited for localization or sensing. In these scenarios, CSI inaccuracies may lead to beam misalignment, multiuser interference, poor equalization, or reduced sensing resolution.

Traditional model-based channel estimation techniques for large-array MIMO systems often rely on pilots. However, as the number of antennas increases, they require excessive training and incur high computational complexity, which leads to performance degradation under overhead constraints or in complex propagation environments [1, 2, 3, 4]. Prior work has also proposed black-box, fully data-driven approaches [5, 6, 7, 8], that often suffer from limited physical interpretability,

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Workshop: AI and ML for Next-Generation Wireless Communications and Networking (AI4NextG).

demand large-scale data collection, or tend to generalize poorly beyond the specific site where they were trained. Another approach exploits the combination of model-based and data-driven methods [9]. However, these combined approaches often utilize the black-box structure of neural networks along with simplistic channel models and unrealistic statistical data for training. In addition, the exploitation of diffusion models in the recent channel estimation literature [9, 8] will also introduce a high overhead in the backward process for generating the channels, and cannot be used in real-time communication systems.

In this work, we combine traditional wireless channel estimation methods, which are model-based, with a PINN approach. This hybrid strategy leverages environmental information to reduce overhead and enhance performance while avoiding the end-to-end black-box structure. Additionally, we evaluate our approach using ray tracing data, which closely resembles real-world data and effectively captures most characteristics of the environment [10]. Numerical results show an improvement of over 5 dB compared to the baselines, while having a practical latency for real-time applications. The contribution of this work can be summarized as follows.

- We propose a hybrid architecture that combines model-based channel estimation with deep learning by incorporating environmental propagation characteristics through RSS maps as input. This approach leverages prior physical knowledge to constrain the solution space even when pilot signals are severely limited.
- We design an enhanced U-Net architecture augmented with transformer modules and cross-attention mechanisms that effectively fuse initial channel estimates with environmental information, enabling the network to selectively incorporate physics-based environmental knowledge rather than treating channel estimation and propagation characteristics as independent modalities. To the best of our knowledge, this is the first work to integrate RSS maps into a physics-informed neural network for pilot-limited MIMO channel estimation.
- We construct several realistic ray tracing datasets spanning multiple carrier frequencies and deployment sites to evaluate the proposed approach and assess its generalization across distribution shifts. This dataset is also provided for future contributions.
- We demonstrate substantial performance improvements through comprehensive evaluation on realistic ray tracing datasets from urban environments, achieving over 5 dB gain in normalized mean squared error compared to state-of-the-art methods, with particularly strong performance in pilot-limited scenarios where our approach achieves approximately -13 dB NMSE with only four pilots at 0 dB signal-to-noise ratio.

2 Problem Statement

Physical design: Wireless signal propagation is governed by Maxwell's equations, which describe the behavior of electromagnetic waves as they interact with the environment. Under the time-harmonic assumption and appropriate simplifications, these equations can be reduced to an inhomogeneous wave equation for the electric field [11, 12]. The solution can be obtained using Green's function methods, which express the electric field as a superposition of contributions from source currents and medium inhomogeneities (see Appendix B for detailed derivations). This formulation requires knowledge of material properties such as wall thickness, relative permittivity, permeability, and electrical conductivity. After computing the electric field distribution using numerical methods, the total received power at a given location can be determined. The received power is proportional to the squared magnitude of the electric field, and accounting for all multipath components (MPC), can be expressed as [13]

$$P_R(\mathbf{r}) = \frac{\lambda^2}{8\pi\eta_0} \left| \sum_{\ell=1}^{P} E_\ell(\mathbf{r}) \right|^2, \tag{1}$$

where P denotes the number of propagation paths, λ is the wavelength, η_0 is the intrinsic impedance of free space, and E_ℓ represents the complex amplitude of the electric field associated with the ℓ-th path at the receiver location. Using $E(\mathbf{r}) = \sum_{\ell=1}^{P} E_\ell(\mathbf{r})$, this relates to the total electric field. This expression captures the coherent superposition of all path contributions, including their respective amplitudes and phases.

Communication system: We consider a communication system operating in upper-mid bands (7-24 GHz), in an urban environment. The base station (BS) equipped with a uniform rectangular array

(URA) of size $N_t = N_t^x \times N_t^y$ is at the top of a building, and the receiver vehicle operating at the same frequency is equipped with a URA consisting of $N_r = N_r^x \times N_r^y$ elements. We considered a time-domain frequency-selective channel due to multipath propagation. Hereby, the frequency selective channel consisting P paths after sampling from continues-time channel can be defined as:

$$\mathbf{H}_d = \sum_{\ell=1}^{P} \alpha_\ell f_{\mathrm{p}} \left(dT_{\mathrm{s}} - (t_\ell - t_{\mathrm{off}})\right) \mathbf{a}_{\mathrm{r}} \left(\theta_\ell^{\mathrm{az}}, \theta_\ell^{\mathrm{el}}\right) \mathbf{a}_{\mathrm{t}} \left(\phi_\ell^{\mathrm{az}}, \phi_\ell^{\mathrm{el}}\right), \tag{2}$$

where d is the channel tap index, T_{s} is the sampling interval, t_{off} is the clock offset between the transmitter and receiver, $f_{\mathrm{p}}(.)$ is the filtering function that factors in filtering effects in the system, α_ℓ and t_ℓ are the complex gain and the time-of-arrival (ToA) of the l-th path, $\mathbf{a}_{\mathrm{r}} \left(\theta_\ell^{\mathrm{az}}, \theta_\ell^{\mathrm{el}}\right)$ represents the receiver array response evaluated at the azimuth and elevation angle-of-arrival (AoA), denoted as $\theta_\ell^{\mathrm{az}}$ and $\theta_\ell^{\mathrm{el}}$ respectively, and $\mathbf{a}_{\mathrm{t}} \left(\phi_\ell^{\mathrm{az}}, \phi_\ell^{\mathrm{el}}\right)$ is the transmitter array response evaluated at the azimuth and elevation angle-of-departure (AoD), denoted as ϕ_ℓ^{az} and ϕ_ℓ^{el} respectively. The array responses can be formulated as

$$\mathbf{a}_{\mathrm{r}} \left(\theta^{\mathrm{az}}, \theta^{\mathrm{el}}\right) = \mathbf{a}\left(\theta'', \theta^{\perp}\right) = \mathbf{a}\left(\theta''\right) \otimes \mathbf{a}\left(\theta^{\perp}\right) \tag{3a}$$

$$\mathbf{a}_{\mathrm{t}} \left(\phi^{\mathrm{az}}, \phi^{\mathrm{el}}\right) = \mathbf{a}\left(\phi'', \phi^{\perp}\right) = \mathbf{a}\left(\phi''\right) \otimes \mathbf{a}\left(\phi^{\perp}\right) \tag{3b}$$

where $\theta'' = \cos\left(\theta^{\mathrm{el}}\right) \sin\left(\theta^{\mathrm{az}}\right), \theta^{\perp} = \sin\left(\theta^{\mathrm{el}}\right), \phi'' = \cos\left(\phi^{\mathrm{el}}\right) \sin\left(\phi^{\mathrm{az}}\right), \phi^{\perp} = \sin\left(\phi^{\mathrm{el}}\right)$, and $\mathbf{a}(\cdot)$ is the steering vector where $[\mathbf{a}(\vartheta)]_n = e^{-j\pi(n-1)\vartheta}$ assuming a halfwavelength element spacing.

Connection between physical modeling and communication channel: The electromagnetic analysis described above provides a physically grounded approach for modeling wireless propagation by explicitly computing the electric field at any location, accounting for the detailed properties of the environment. In contrast, modern communication system models represent the channel by a finite set of propagation paths, each parameterized by delay, angle, and complex gain, as encapsulated in the MIMO channel matrix $\{\mathbf{H}_d\}_{d=1}^{D}$, each corresponding to a delay tap d, with D the total number of resolvable taps. The physically computed field superposition at each receiver location thus defines the per-tap response in the wide-band MIMO channel. The received signal strength can be calculated per tap, or integrated across all taps, using the MIMO channel matrices,

$$P_R = P_T \sum_{d=1}^{D} \mathbb{E}\left[\mathrm{trace}(\|\mathbf{H}_d\|_2^2)\right], \tag{4}$$

or, equivalently, from the summed field strengths at each tap. Here P_T is the transmission power.

3 Physics-Informed Neural Network

Initial channel estimation: The first step involves evaluating a coarse channel estimation. To simplify our approach and avoid unnecessary complications, we utilize basic variations of least squares (LS) estimation. However, this PINN method can be expanded to other initial channel estimation methods. A detailed formulation of this stage can be found in Appendix C.

Physical calculation: To introduce the physically informed network, it is important to evaluate the RSS map using the numerical methods discussed earlier. A popular non-neural simulator for this computation is *Wireless Insite* [14], which can accurately describe the physical structure of the environment. Appendix E contains more information for this stage. We combine the behavior of the wireless channel with physical calculations through a physically informed neural network. We assume that the BS has a rough estimate of the user's position. This information helps locate users on the RSS map, allowing us to extract the approximate power levels in those areas for use in the network. In practical wireless deployments, RSS maps can be evaluated using digital twin technology, which creates real-time virtual representations of the electromagnetic propagation environment. Commercial digital twin platforms like VIAVI's Network Digital Twin and Ericsson's NDT solutions already provide AI-driven radio propagation modeling that can generate accurate RSS maps synchronized with physical network conditions [15], while advanced ray tracing tools such as NVIDIA's Instant Radio Maps can compute high-resolution radio maps using real-world 3D environmental data [16]. These digital twins leverage high-definition 3D maps and ray-based propagation simulation to provide real-time digital representations of electromagnetic environments.

Neural network design: We employ a physics-informed U-Net architecture with transformer enhancement as illustrated in Figure 1. The network consists of a three-layer encoder-decoder Resnet-based structure with symmetric skip connections, specifically designed for the channel estimation task with spatial dimensions (N_r, N_t). The encoder progressively downsamples the $2 \times D$-channel input through convolutional blocks, reducing the spatial resolution while increasing the feature depth. In the latent space, after flattening the extracted features, we integrate the physics-informed RSS map information through a dedicated RSS encoder that processes the cropped environmental map around the user location. Cross-modal fusion combines the flattened channel features with the RSS features via a cross-attention that learns the correlation between electromagnetic propagation characteristics and channel structure. The cross-attention mechanism integrates RSS environmental information with channel features. This process started with projection to a common hidden dimension as

$$\mathbf{X}_i = \mathbf{W}_i \mathbf{F}_i + \mathbf{b}_i \ ; \ i \in \{\text{RSS, Channel}\}, \tag{5}$$

in which $\mathbf{F}_i \in \mathbb{R}^{\text{Batch} \times D_i}$ is the extracted features from RSS and channel, $\mathbf{X}_i \in \mathbb{R}^{\text{Batch} \times D_z}$ is the projected features, $\mathbf{W}_i \in \mathbb{R}^{D_i \times D_z}$ is the projection matrices, and $\mathbf{b}_i \in \mathbb{R}^{D_z}$ is the bias vector. Here, D_z represents the dimension of the latent space, and D_i is the dimension of RSS or channel features. The cross-attention mechanism employs multi-head attention. The process is formulated as softmax $\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{D_z}}\right)\mathbf{V}$ where $\mathbf{Q} = \mathbf{X}_{\text{Channel}}$, $\mathbf{K} = \mathbf{X}_{RSS}$, and $\mathbf{V} = \mathbf{X}_{RSS}$. Setting $\mathbf{Q}$ as channel features and $\mathbf{K}$, $\mathbf{V}$ as RSS features enables the channel estimation to query what environmental information is most relevant for refinement, where the attention weights determine which RSS spatial patterns are most informative for each channel element. This cross-modal attention allows the network to selectively incorporate physics-based environmental knowledge into the channel estimation process rather than treating them as independent modalities. The fused features are then

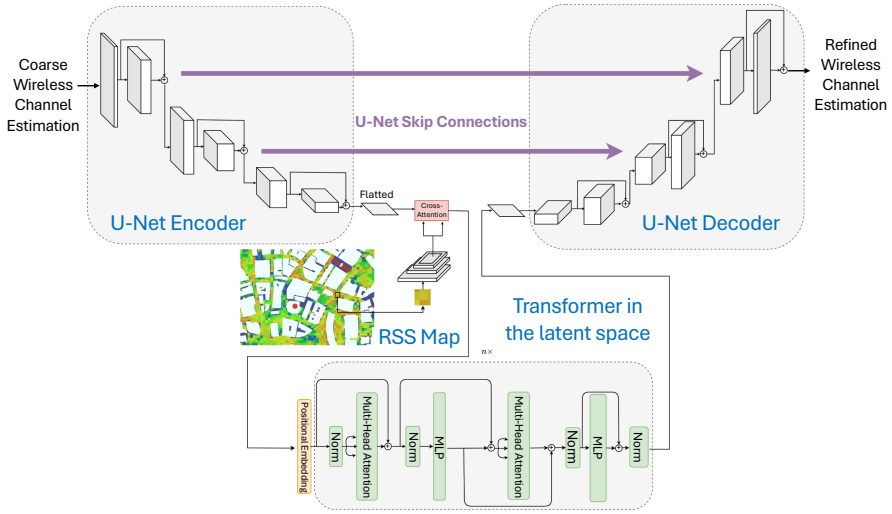


Figure 1: PINN structure for channel estimation.

processed by a transformer module consisting of several self-attention blocks, where the spatial positions are treated as a sequence. This transformer component enables the network to model long-range dependencies between different channel taps and capture the physics-informed relationships between RSS patterns and channel characteristics. The decoder mirrors the encoder structure, utilizing transposed convolutions to progressively upsample the transformed features back to the original channel dimensions. Skip connections from corresponding encoder layers preserve fine-grained details and facilitate gradient flow during training. This architecture effectively combines traditional U-Net benefits for spatial feature preservation with transformer capabilities for modeling complex channel-environment relationships in the physics-informed framework. The implementation details and exact dimensions for each part of the network are outlined in Appendix F. We also employed a physically informed loss function to combine the physical information from solving Maxwell equations using numerical methods with the estimated channel and reconstruction loss.

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{NMSE}} + \zeta \, \mathcal{L}_{\text{physical}} \ , \tag{6}$$

where $\mathcal{L}_{\text{NMSE}}$ is the reconstruction loss for the channel estimation defined as

$$\mathcal{L}_{\text{NMSE}} = \mathbb{E}\left(\frac{\|\mathbf{H} - \hat{\mathbf{H}}\|_2^2}{\|\mathbf{H}\|_2^2}\right) \tag{7}$$

in which $\mathbf{H} \in \mathbb{C}^{D \times N_{\text{r}} \times N_{\text{t}}}$ and $\hat{\mathbf{H}} \in \mathbb{C}^{D \times N_{\text{r}} \times N_{\text{t}}}$ are the 3-dimensional complex tensors for accurate and estimated channels respectively. Moreover, $\mathcal{L}_{\text{physical}}$ defined as

$$\mathcal{L}_{\text{physical}} = \mathbb{E}\left(\mathbb{E}\left[\frac{\lambda^2}{8\pi\eta_0}\left|\sum_{i=1}^{P} E_i(\mathbf{r})\right|^2\right] - P_T \sum_{d=1}^{D} \mathbb{E}\left[\text{trace}(\|\mathbf{H}_d\|_2^2)\right]\right)^2, \tag{8}$$

and ζ weights the physical consistency term to prevent physics violations while maintaining reconstruction accuracy. The hyperparameter is chosen such that both loss terms contribute comparable gradient magnitudes during early training phases.

Remark. *In the pilot-limited regime, the benefit of including the physics-informed term $\mathcal{L}_{physical}$ can be understood through a simple bias–variance argument [17]. Without constraints, a data-driven estimator must search over a high-dimensional hypothesis space, leading to variance scaling as $\text{Var}[f_\theta] \propto d_{eff}/M$, where f_θ is the neural estimator with parameters θ, d_{eff} is the effective model dimension (less than the model dimension d) and M the number of pilots. By enforcing approximate physical constraints $\mathcal{P}(f_\theta(\mathbf{y})) \approx 0$, with $\mathbf{y}$ the observations (pilots and RSS), the search space is restricted to a lower-dimensional manifold, yielding*

$$\text{Var}[f_\theta] \propto d_{eff}/M, \qquad d_{eff} \ll d, \tag{9}$$

thus reducing estimation variance with only a minor bias if the physical model is imperfect. From an optimization viewpoint, the additional quadratic penalty $\lambda\mathcal{L}_{physical}$ improves the conditioning of the loss landscape. Writing a gradient step as

$$\theta_{t+1} = \theta_t - \eta\big(\nabla\mathcal{L}_{data} + \lambda\nabla\mathcal{L}_{physical} + \xi_t\big), \tag{10}$$

where ξ_t denotes stochastic noise, the Hessian gains an extra regularizing term, which lowers the effective condition number and accelerates convergence. Hence, the physics-informed loss improves both generalization and training stability.

The overall proposed method is summarized in Algorithm 1.

Algorithm 1 PINN Channel Estimation

Input: Training set $\mathcal{D} = \{(\mathbf{y}_i, \mathbf{H}_i, \text{RSS}_i, \mathbf{r}_i)\}_{i=1}^{M}$, physics weight ζ, epochs E.
Output: Trained parameters θ^*
for epoch $e = 1$ to E **do**
  Compute initial channel estimate $\tilde{\mathbf{H}} \leftarrow \text{LS}(\mathbf{y}, N_p)$
  Extract RSS features: $\mathbf{F}_{\text{RSS}} \leftarrow \text{RSSEncoder}_\theta(\text{RSS})$
  Extract channel features: $\mathbf{F}_{\text{Channel}} \leftarrow \text{Encoder}_\theta(\tilde{\mathbf{H}})$
  Fuse via cross-attention:
    $\mathbf{Q} = \mathbf{W}_{\text{Channel}}\mathbf{F}_{\text{Channel}}$, $\mathbf{K} = \mathbf{W}_{\text{RSS}}\mathbf{F}_{\text{RSS}}$, $\mathbf{V} = \mathbf{W}_{\text{RSS}}\mathbf{F}_{\text{RSS}}$
    $\mathbf{Z} \leftarrow \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{D_z}}\right)\mathbf{V}$
  Refine with transformer: $\mathbf{Z} \leftarrow \text{Transformer}_\theta(\mathbf{Z})$
  Reconstruct channel: $\hat{\mathbf{H}} \leftarrow \text{Decoder}_\theta(\mathbf{Z})$
  Compute total loss:
    $\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{NMSE}} + \zeta\mathcal{L}_{\text{physical}}$
  Minimize $\mathcal{L}_{\text{total}}$ with respect to θ
end for
Return: θ^*

4 Results

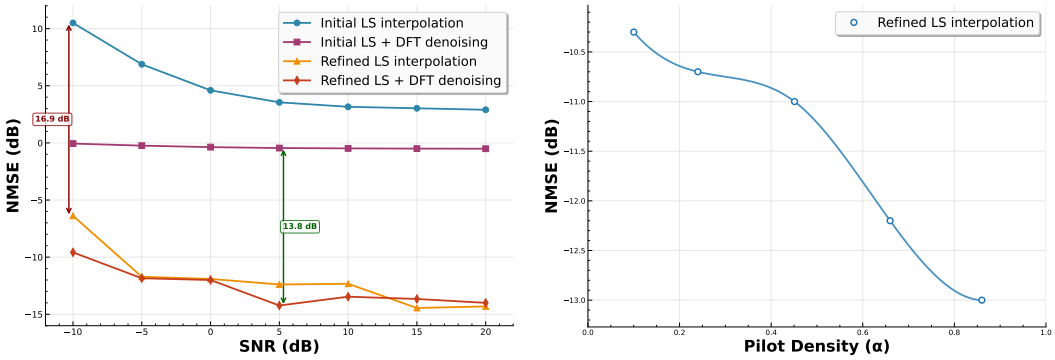
After defining our network architecture (see Appendix F), we outline in Table 1 the hyperparameter settings used during training. These choices were determined empirically to balance convergence speed and generalization.

Table 1: Training hyperparameters

Batch Size	# Epochs	Init. LR	Scheduler	Decay Step	γ (Decay)	Optimizer	Momentum	ζ
32	500	1×10^{-3}	StepLR	40	0.65	Adam	0.9	0.01

We categorize our tests into multiple main sections for presentation purposes. Additional implementation details can be found in Appendix F. In all methods, unless otherwise specified, we used $N_t = 24 \times 24$, $N_r = 2 \times 2$, and $D = 16$. We also modeled the channel at an upper-mid band frequency of $f_c = 15$ GHz, which holds significant potential for future wireless systems [18].

In the first set, we investigated the performance of less complex spatial simple LS channel estimation using interpolation and discrete Fourier transform (DFT)-based denoising. We consider different SNR values while setting the number of pilots as $N_p = 64$ ($\frac{N_p}{N_t} = 0.11$).



(a) NMSE vs. SNR.

(b) NMSE vs. pilot density.

Figure 2: NMSE performance of channel estimation using less complex initial estimation

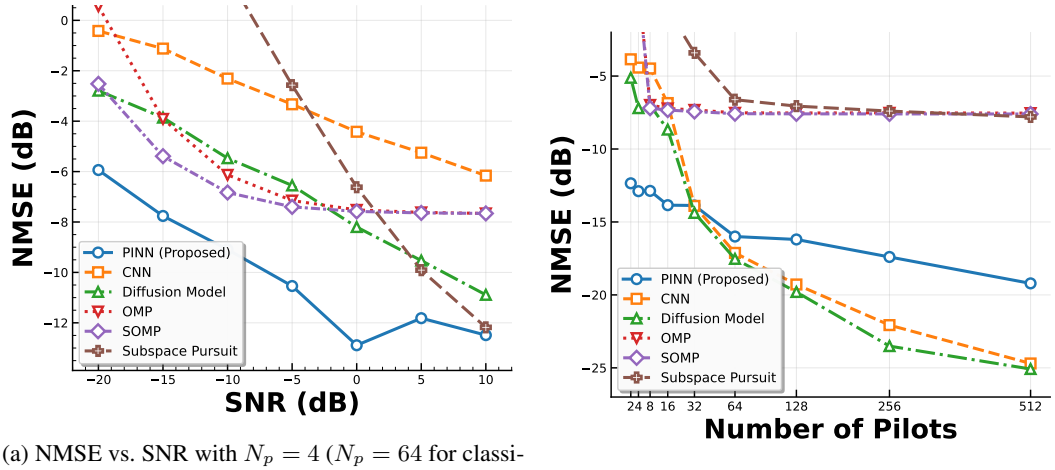
We can see in Figure 2a, applying the physics-informed neural network refinement further lowers NMSE to roughly $(-14.45, -6.38)$ dB (interpolation) and $(-14.23, -9.58)$ dB (denoising), corresponding to maximum gains of up to 16.9 dB and 13.8 dB respectively. The refined methods maintain robust performance even at $\mathrm{SNR} = -10$ dB and saturate near -14 dB NMSE at high SNRs, indicating that residual error is dominated by measurement errors that the network cannot correct. Figure 2b also indicates the performance change in different pilot densities ($\alpha = \frac{N_p}{N_t}$) at $\mathrm{SNR} = 0$ dB. We can see the robust performance of this method to pilot density. The next set of results uses orthogonal frequency-division multiplexing (OFDM) for channel estimation. We set the number of subcarriers $N = 1024$ and change the subcarrier spacing, and accordingly, the number of pilot signals. We also set the SNR as 0 dB in this simulation. Table 2 demonstrates that the PINN estimation shows a great performance in limited pilot scenarios, achieving up to NMSE = -12.89 dB with only four pilots. In the next set of experiments, we compare the introduced method with state-of-the-art

Table 2: NMSE performance of channel estimation using different numbers of pilots (SNR = 0 dB).

	NMSE (dB) vs N_p								
$\boldsymbol{N_p}$	512	256	128	64	32	16	8	4	2
Initial LS-OFDM	-14.40	-11.23	-8.25	-5.49	-3.32	-1.33	0.28	1.98	3.01
Refined LS-OFDM	-19.21	-17.40	-16.20	-16.00	-13.87	-13.85	-12.87	-12.89	-12.35
Improvement	4.81	6.17	7.95	10.51	10.55	12.52	13.15	14.87	15.36

algorithms in addition to classical orthogonal matching pursuit (OMP) methods [19, 20, 21]. These greedy compressed sensing algorithms iteratively select angular-domain dictionary atoms to exploit the sparse structure of massive MIMO channels, with SOMP handling joint multi-vector recovery and Subspace Pursuit using candidate expansion and pruning for improved performance. To have

a fair comparison, we only investigate the initial estimation for LS-OFDM. We set the number of pilots to a limited value of $N_p = 4$ (to have comparable results for the classical methods, we set that to $N_p = 64$), to emphasize the power of the introduced method across different SNRs. We implemented a convolutional neural network (CNN)-based channel estimation structure inspired by [5](using two hidden layers according to the paper), and a diffusion model-based method from [8](using the full resolution method). In this experiment, we focused on the low SNR region, which is more realistic for the next generation of wireless systems [9]. Figure 3a reveals that the introduced PINN improves the performance by a large margin (around 5 dB improvement at $\mathrm{SNR} = 0$) in the limited pilot condition. These results highlight the method's robustness when pilot signals are scarce and underscore how incorporating physical environment information significantly enhances the network's channel estimation capabilities. Figure 3b presents a comparative analysis across varying



(a) NMSE vs. SNR with $N_p = 4$ ($N_p = 64$ for classical methods).

(b) NMSE vs. number of pilot signals at $\mathrm{SNR} = 0$ dB.

Figure 3: NMSE comparison across varying SNRs and pilot counts.

pilot values at $\mathrm{SNR} = 0$. The results reveal that CNN and diffusion-based methods excel when pilot signals are abundant, whereas PINN achieves superior performance under pilot-limited conditions, often by substantial margins. This dichotomy indicates that PINN's physics-informed architecture, while potentially over-parameterized for pilot-rich scenarios, provides essential inductive bias when pilot overhead is constrained. The physical modeling component effectively compensates for sparse pilot information, validating the value of incorporating domain knowledge into the estimation framework to provide a more robust estimate with respect to pilot density. Further experimental results demonstrating the generalization performance of the proposed method across different frequency bands and environments are provided in Appendix G.

5 Conclusion

This paper introduced a PINN for accurate wireless channel estimation under pilot-constrained scenarios. By integrating initial simple least-squares estimates with RSS maps derived from Maxwell-based ray tracing, the proposed architecture leverages environmental knowledge through a U-Net backbone enhanced with transformer and cross-attention modules. Our results on realistic 15 GHz urban ray-traced data show that PINN significantly outperforms conventional and learning-based baselines, especially in low-pilot and low-SNR regimes, achieving up to 15 dB NMSE gain over initial LS estimates and over 4 dB gain compared to state-of-the art models. With low inference latency and strong interpretability, PINN offers a scalable and physically grounded solution for real-time deployment in next-generation MIMO systems.

References

[1] J. Palacios, N. González-Prelcic, and C. Rusu, "Multidimensional orthogonal matching pursuit: theory and application to high accuracy joint localization and communication at mmWave," *arXiv preprint arXiv:2208.11600*, 2022.

[2] J. Palacios and N. González-Prelcic, "Separable multidimensional orthogonal matching pursuit and its application to joint localization and communication at mmWave," in *2023 IEEE Globecom Workshops (GC Wkshps)*. IEEE, 2023, pp. 1421–1426.

[3] K. Venugopal, A. Alkhateeb, N. G. Prelcic, and R. W. Heath, "Channel estimation for hybrid architecture-based wideband millimeter wave systems," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 9, pp. 1996–2009, 2017.

[4] J. Rodríguez-Fernández, N. González-Prelcic, K. Venugopal, and R. W. Heath, "Frequency-domain compressive channel estimation for frequency-selective hybrid millimeter wave MIMO systems," *IEEE Transactions on Wireless Communications*, vol. 17, no. 5, pp. 2946–2960, 2018.

[5] M. Sattari, H. Guo, D. Gündüz, A. Panahi, and T. Svensson, "Full-Duplex Millimeter Wave MIMO Channel Estimation: A Neural Network Approach," *IEEE Transactions on Machine Learning in Communications and Networking*, 2024.

[6] I. Helmy and W. Choi, "Low-Resolution Massive MIMO Channel Estimation with LSTM Attention-Based CBDNet," *IEEE Transactions on Mobile Computing*, 2025.

[7] Q. Hu, F. Gao, H. Zhang, S. Jin, and G. Y. Li, "Deep learning for channel estimation: Interpretation, performance, and comparison," *IEEE Transactions on Wireless Communications*, vol. 20, no. 4, pp. 2398–2412, 2020.

[8] X. Zhou, L. Liang, J. Zhang, P. Jiang, Y. Li, and S. Jin, "Generative diffusion models for high dimensional channel estimation," *IEEE Transactions on Wireless Communications*, 2025.

[9] Z. Jin, L. You, D. W. K. Ng, X.-G. Xia, and X. Gao, "Near-field channel estimation for XL-MIMO: A deep generative model guided by side information," *IEEE Transactions on Cognitive Communications and Networking*, 2025.

[10] G. Skidmore, "Auto-Radar Drive Scenario Simulation: Increasing Realism with Multipath, Diffuse Scattering, and Micro-Doppler," Presentation at European Microwave Week 2021, Remcom Inc., 2021.

[11] Z. Wan, J. Zhu, and L. Dai, "Near-field channel modeling for electromagnetic information theory," *IEEE Transactions on Wireless Communications*, 2024.

[12] N. Ida, *Engineering electromagnetics*. Springer, 2015.

[13] Z. Fu, S. Mukherjee, M. T. Lanagan, P. Mitra, T. Chawla, and R. M. Narayanan, "A Fast Indoor Coverage Prediction Scheme at 60 GHz Based on Image Processing, Geometrical Optics, and Transport Theory," in *2023 IEEE International Symposium on Antennas and Propagation and USNC-URSI Radio Science Meeting (AP-S/URSI)*. Portland, OR, USA: IEEE, 2023, pp. 1231–1232.

[14] Remcom, "Wireless InSite EM Propagation Software," https://www.remcom.com/wireless-insite-em-propagation-software, 2022.

[15] Ericsson Technology Review, "Network digital twins: Outlook and opportunities," *Ericsson Technology Review*, 2023. [Online]. Available: https://www.ericsson.com/en/reports-and-papers/ericsson-technology-review/articles/network-digital-twins-outlook-and-opportunities

[16] M. Zhu, L. Cazzella, F. Linsalata, M. Magarini, M. Matteucci, and U. Spagnolini, "Toward real-time digital twins of em environments: Computational benchmark for ray launching software," *IEEE Open Journal of the Communications Society*, vol. 5, pp. 6291–6302, 2024.

[17] C. M. Bishop and N. M. Nasrabadi, *Pattern recognition and machine learning*. Springer, 2006, vol. 4, no. 4.

[18] S. Kang, M. Mezzavilla, S. Rangan, A. Madanayake, S. B. Venkatakrishnan, G. Hellbourg, M. Ghosh, H. Rahmani, and A. Dhananjay, "Cellular wireless networks in the upper mid-band," *IEEE Open Journal of the Communications Society*, 2024.

[19] W. Dai and O. Milenkovic, "Subspace pursuit for compressive sensing signal reconstruction," *IEEE Transactions on Information Theory*, vol. 55, no. 5, pp. 2230–2249, 2009.

[20] J. Lee, G.-T. Gil, and Y. H. Lee, "Channel estimation via orthogonal matching pursuit for hybrid mimo systems in millimeter wave communications," *IEEE Transactions on Communications*, vol. 64, no. 6, pp. 2370–2386, 2016.

[21] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Transactions on Information Theory*, vol. 53, no. 12, pp. 4655–4666, 2007.

[22] A. Bazzi, R. Bomfin, M. Mezzavilla, S. Rangan, T. Rappaport, and M. Chafii, "Upper mid-band spectrum for 6G: Vision, opportunity and challenges," *arXiv preprint arXiv:2502.17914*, 2025.

[23] Y. Qiu, D. Wu, Y. Zeng, Y. Tang, N. Cheng, and C. Qi, "AI-based Environment-Aware XL-MIMO Channel Estimation with Location-Specific Prior Knowledge Enabled by CKM," *arXiv preprint arXiv:2507.06066*, 2025.

[24] M. Raissi, P. Perdikaris, and G. E. Karniadakis, "Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations," *Journal of Computational Physics*, vol. 378, pp. 686–707, 2019.

[25] G. E. Karniadakis, I. G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, and L. Yang, "Physics-informed machine learning," *Nature Reviews Physics*, vol. 3, no. 6, pp. 422–440, 2021.

[26] S. Wang, H. Wang, and P. Perdikaris, "Learning the solution operator of parametric partial differential equations with physics-informed deeponets," *Science advances*, vol. 7, no. 40, p. eabi8605, 2021.

[27] X. Jin, S. Cai, H. Li, and G. E. Karniadakis, "Nsfnets (navier-stokes flow nets): Physics-informed neural networks for the incompressible navier-stokes equations," *Journal of Computational Physics*, vol. 426, p. 109951, 2021.

[28] S. Wang, S. Gao, W. Yang, Q. Zhang, T. H. Loh, Y. Yang, and F. Qin, "A physics-informed deep ray tracing network for regional channel impulse response estimation," *IEEE Transactions on Wireless Communications*, 2025.

[29] F. Jiang, T. Li, X. Lv, H. Rui, and D. Jin, "Physics-informed neural networks for path loss estimation by solving electromagnetic integral equations," *IEEE Transactions on Wireless Communications*, 2024.

[30] J. Choi, D. J. Love, and P. Bidigare, "Downlink training techniques for FDD massive MIMO systems: Open-loop and closed-loop training with memory," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 802–814, 2014.

[31] X. Rao and V. K. Lau, "Distributed compressive CSIT estimation and feedback for FDD multi-user massive MIMO systems," *IEEE Transactions on Signal Processing*, vol. 62, no. 12, pp. 3261–3271, 2014.

[32] M. Ozdemir and H. Arslan, "Channel estimation for wireless OFDM systems," *IEEE Communications Surveys and Tutorials*, vol. 9, no. 2, 2007.

[33] M. Specht, "Experimental studies on the relationship between hdop and position error in the gps system," *Metrology and Measurement Systems*, pp. 17–36, 2022.

[34] P. Medeđović, M. Veletić, and Ž. Blagojević, "Wireless insite software verification via analysis and comparison of simulation and measurement results," in *2012 Proceedings of the 35th International Convention MIPRO*. IEEE, 2012, pp. 776–781.

[35] Remcom, "Received power and delay spread comparisons in an urban environment," https://www.remcom.com/resources/examples/received-power-and-delay-spread-comparisons-in-an-urban-environment, 2025.

[36] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[37] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

A Related Works

Traditional model-based channel estimation techniques for large-array MIMO systems often rely on pilots. However, as the number of antennas increases, they require excessive training and incur high computational complexity, which leads to performance degradation under overhead constraints or in complex propagation environments [1, 2, 3, 4]. This problem is expected to intensify with the large-array configurations envisioned for FR3 systems, where wider bandwidths and higher antenna counts further amplify pilot overhead and computational demands [22]. Prior work has also proposed black-box, fully data-driven approaches [5, 6, 7, 8] that often suffer from limited physical interpretability, demand large-scale data collection, or tend to generalize poorly beyond the specific site or frequency band where they were trained.

Another line of work combines model-based priors with data-driven learning [9]. However, these combined approaches often utilize the black-box structure of neural networks along with simplistic channel models and unrealistic statistical data for training. Consequently, performance often degrades under domain (site) shift, and substantial site-specific data are required for training when the deployment location changes. In addition, the exploitation of diffusion models in the recent channel estimation literature [9, 8] will also introduce a high overhead in the backward process for generating the channels, and cannot be used in real-time communication systems. [23] also assumes narrowband single-antenna UEs and relies on location-specific denoisers and iterative MAP solvers. In contrast, we address wide-band multi-tap MIMO with physics-informed RSS maps, enabling single-shot inference and generalization without channel knowledge map (CKM) databases.

PINN have emerged as a powerful paradigm that integrates domain knowledge from physical laws into neural network architectures, enabling more interpretable and data-efficient learning [24, 25]. By incorporating governing equations, boundary conditions, and conservation laws directly into the loss function or network structure, PINNs have demonstrated remarkable success across diverse scientific computing applications, including fluid dynamics, structural mechanics, and electromagnetic wave propagation [26, 27]. Unlike purely data-driven approaches, PINNs leverage prior physical knowledge to constrain the solution space, resulting in improved generalization with limited training data and enhanced robustness to out-of-distribution scenarios. In the context of wireless communications, recent works have begun exploring physics-informed approaches for channel modeling and parameter estimation [28, 29], with some successfully incorporating detailed electromagnetic field calculations and environmental propagation characteristics through ray tracing methods and integral equation formulations. However, these approaches often focus on specific aspects of channel modeling, such as regional channel impulse response estimation or path loss prediction, while the integration of physics-informed constraints with pilot-limited channel estimation for massive MIMO systems remains relatively unexplored, particularly in scenarios requiring both high accuracy and efficiency under severe pilot constraints.

In this work, we combine traditional wireless channel estimation methods, which are model-based, with a PINN approach. This hybrid strategy leverages environmental information to reduce overhead and enhance performance while avoiding the end-to-end black-box structure. Additionally, we evaluate our approach using ray tracing data, which closely resembles real-world data and effectively captures most characteristics of the environment [10]. Numerical results show an improvement of over 5 dB compared to the baselines, while having a practical latency for real-time applications. We also demonstrated the capability of this method on different frequency bands and environments. Additionally, for mobile scenarios, we extend the decoder into a parallel multi-step temporal head that estimates L future channel snapshots from a single input, avoiding autoregressive error accumulation while preserving physical consistency with the RSS constraints.

B Electromagnetic Propagation

Maxwell's equations form the theoretical foundation for wireless signal propagation, governing the behavior of electromagnetic waves as they interact with the environment. To accurately model the

effects of environmental scatterers, one can adjust the electromagnetic properties of the medium. By applying Maxwell's equations under the time-harmonic assumption, we obtain the following form for describing electromagnetic fields [11, 12],

$$\nabla \times \mathbf{E}(\mathbf{r}) = \mathrm{j}\omega\mu(\mathbf{r})\mathbf{H}(\mathbf{r}), \tag{11a}$$

$$\nabla \times \mathbf{H}(\mathbf{r}) = -\mathrm{j}\omega\epsilon(\mathbf{r})\mathbf{E}(\mathbf{r}) + \mathbf{J}(\mathbf{r}), \tag{11b}$$

$$\nabla \cdot (\epsilon(\mathbf{r})\mathbf{E}(\mathbf{r})) = \rho(\mathbf{r}), \tag{11c}$$

$$\nabla \cdot (\mu(\mathbf{r})\mathbf{H}(\mathbf{r})) = 0, \tag{11d}$$

where $\epsilon(\mathbf{r})$, $\mu(\mathbf{r})$, and $\rho(\mathbf{r})$ denote the permittivity, permeability, and charge density at position $\mathbf{r} \in \mathbb{R}^3$, respectively. In these equations, $\mathbf{J}(\mathbf{r})$ being the current density, ω denotes the angular frequency, $\mathbf{E}(\mathbf{r})$ describes electric field, and $\mathbf{H}(\mathbf{r})$ shows the magnetic field strength. In the case of an inhomogeneous medium, these parameters are spatially varying and thus nonconstant. To facilitate analytical or numerical solutions, Maxwell's equations are often simplified. Under appropriate assumptions, the corresponding wave equation for the electric field can be expressed as,

$$\left(\nabla^2 + k^2(\mathbf{r})\right)\mathbf{E}(\mathbf{r}) = q(\mathbf{r}), \tag{12}$$

in which $k(\mathbf{r}) = \omega\sqrt{\epsilon(\mathbf{r})\mu(\mathbf{r})}$ denotes the position-dependent wavenumber, and $q(\mathbf{r}) = j\omega\mu(\mathbf{r})J(\mathbf{r})$ represents the source term. To solve the inhomogeneous wave equation, Green's function

$$g\left(\mathbf{r}, \mathbf{r}'\right) = \frac{e^{jk_0\|\mathbf{r}-\mathbf{r}'\|}}{4\pi\|\mathbf{r}-\mathbf{r}'\|} \tag{13}$$

can be used which is the fundamental solution to

$$\left(\nabla^2 + k_0^2\right) g\left(\mathbf{r}, \mathbf{r}'\right) = -\delta\left(\mathbf{r} - \mathbf{r}'\right), \tag{14}$$

where k_0 is the wavenumber in the background (homogeneous) medium and $\delta(\cdot)$ is the Dirac delta function. Using this Green's function, the solution for the electric field can be written as

$$E(\mathbf{r}) = -\int_{V_s} g\left(\mathbf{r}, \mathbf{r}'\right) q\left(\mathbf{r}'\right)\, \mathrm{d}\mathbf{r}' + \int_V g\left(\mathbf{r}, \mathbf{r}'\right)\left(k^2(\mathbf{r}') - k_0^2\right) E\left(\mathbf{r}'\right)\, \mathrm{d}\mathbf{r}', \tag{15}$$

where V_s is the support of the source, and V is the volume over which the medium inhomogeneity exists. To solve the wave equation in a specific environment, it is essential to have detailed knowledge of the material properties of the objects within that environment, such as walls and floors. After computing the electric field distribution using numerical methods, the total received power at a given location can be determined from the electric field. Specifically, the received power is proportional to the squared magnitude of the electric field integrated over the receiving antenna's effective area. Mathematically, the total received power at the receiver, accounting for all multipath components (MPCs), can be expressed as [13]

$$P_R(\mathbf{r}) = \frac{\lambda^2}{8\pi\eta_0}\left|\sum_{\ell=1}^{P} E_\ell(\mathbf{r})\right|^2, \tag{16}$$

where P denotes the number of propagation paths, λ is the wavelength, η_0 is the intrinsic impedance of free space, and E_i represents the complex amplitude of the electric field associated with the i-th path at the receiver location. Using $E(\mathbf{r}) = \sum_{\ell=1}^{P} E_\ell(\mathbf{r})$ this relates to the total electric field. This expression captures the coherent superposition of all path contributions, including their respective amplitudes and phases.

C Initial Wireless Channel Estimation

The frequency-selective MIMO channel is modeled as a 3-dimensional complex tensor $\mathbf{H} \in \mathbb{C}^{D\times N_\mathrm{r}\times N_\mathrm{t}}$. The received signal can be written as can be expressed as,

$$\mathbf{r}[n] = \sum_{d=0}^{D-1} \mathbf{H}_d \mathbf{s}[n-d] + \mathbf{v}[n]. \tag{17}$$

where $\mathbf{s}[n] \in \mathbb{C}^{N_\mathrm{t}}$ is the transmitted pilot vector and $\mathbf{n}_d \sim \mathcal{CN}(0, \sigma_n^2\mathbf{I})$ is the additive white Gaussian noise.

C.1 LS with Linear Interpolation

The LS with linear interpolation method performs channel estimation in three steps; LS estimation at pilot positions, magnitude and phase separation, and linear interpolation to all antenna positions.

Uniform pilot spacing is employed with adaptive step size [30],

$$\mathcal{P} = \{p_i : p_i = i \cdot \Delta_p, \quad i = 0, 1, ..., N_p - 1\}, \quad \Delta_p = \max(1, \lfloor N_\text{t}/N_p \rfloor) \tag{18}$$

At the end, linear interpolation is performed separately for magnitude and phase using unwrapped phase to avoid discontinuities. This method has a simple implementation, low complexity, and works well for slowly varying channels, while there is no explicit noise reduction.

C.2 LS with DFT Denoising

This method exploits channel sparsity in the angular domain for noise reduction. After obtaining LS estimates at pilot positions, zero-padding and DFT transformation are applied,

$$\mathbf{H}_{d,r}^{(f)} = \mathbf{F}^H \tilde{\mathbf{h}}_{d,r} \tag{19}$$

where $\mathbf{F}$ is the normalized DFT matrix and $\tilde{\mathbf{h}}_{d,r}$ is the zero-padded pilot estimate vector.

Adaptive thresholding based on noise statistics removes weak frequency components:

$$\hat{\mathbf{H}}_{d,r}^{(f)}[k] = \begin{cases} \mathbf{H}_{d,r}^{(f)}[k] & \text{if } |\mathbf{H}_{d,r}^{(f)}[k]| \geq \tau \\ 0 & \text{otherwise} \end{cases} \tag{20}$$

where the threshold is computed as,

$$\tau = 3\sqrt{\frac{\sigma_n^2 N_p}{2N_\text{t}}} \tag{21}$$

The denoised channel is reconstructed via inverse DFT,

$$\hat{\mathbf{h}}_{d,r} = \mathbf{F}\hat{\mathbf{H}}_{d,r}^{(f)} \tag{22}$$

This method effectively reduces noise by removing weak frequency components while preserving the dominant channel characteristics [31].

C.3 LS OFDM-Based Estimation

This method operates in the frequency domain using OFDM subcarrier structure. Pilot subcarrier positions are defined with uniform spacing,

$$\mathcal{S} = \{s_i : s_i = i \cdot \Delta_s, \quad i = 0, 1, ..., N_s - 1\}. \tag{23}$$

The time-domain channel is transformed to frequency domain with zero-padding as,

$$\mathbf{H}_\text{freq}[k] = \text{FFT}(\mathbf{h}_{d,r}, N_\text{subcarriers})[k] \tag{24}$$

After LS estimation at pilot subcarriers, we interpolate to data subcarriers using linear interpolation for magnitude and phase [32],

$$\hat{H}_\text{freq}[k] = \begin{cases} \hat{\mathbf{H}}_\text{pilots}[i] & k \in \mathcal{S} \\ |\text{interp}(\mathcal{S}, |\hat{\mathbf{H}}_\text{pilots}|, k)| \cdot e^{j\text{interp}(\mathcal{S}, \text{unwrap}(\angle \hat{\mathbf{H}}_\text{pilots}), k)} & k \notin \mathcal{S} \end{cases} \tag{25}$$

The time-domain channel is recovered using inverse fast Fourier transform (IFFT), retaining only the first D delay taps:

$$\hat{\mathbf{h}}_{d,r} = \text{IFFT}(\hat{\mathbf{H}}_\text{freq})[1 : D] \tag{26}$$

This method yields the best initial results compared to other methods, although its complexity is higher.

D Complexity Analysis

In this section, we analyze the computational complexity of our approach. We begin with a theoretical assessment of the initial estimation algorithms, then evaluate the complexity of the neural network.

We start with the simple **LS with linear interpolation**. This method begins with pilot extraction, which requires $\mathcal{O}(D \cdot N_{\text{r}} \cdot N_p)$ operations to select pilot positions across all delay taps and receive antennas. The subsequent phase unwrapping operation has complexity $\mathcal{O}(D \cdot N_{\text{r}} \cdot N_p)$, where phase discontinuities are resolved for each pilot measurement. Finally, linear interpolation to estimate the channel at all antenna positions requires $\mathcal{O}(D \cdot N_{\text{r}} \cdot N_{\text{t}})$ operations. The total computational complexity is therefore $\mathcal{O}(D \cdot N_{\text{r}} \cdot (N_p + N_{\text{t}}))$, making it the most computationally efficient method when $N_p \ll N_{\text{t}}$.

The **LS with DFT denoising** will start with the pilot extraction with the same complexity. However, it then requires zero-padding operations with complexity $\mathcal{O}(D \cdot N_{\text{r}} \cdot N_{\text{t}})$, followed by DFT transformations requiring $\mathcal{O}(D \cdot N_{\text{r}} \cdot N_{\text{t}} \log N_{\text{t}})$ operations. The adaptive thresholding step adds $\mathcal{O}(D \cdot N_{\text{r}} \cdot N_{\text{t}})$ complexity, and the inverse DFT reconstruction requires another $\mathcal{O}(D \cdot N_{\text{r}} \cdot N_{\text{t}} \log N_{\text{t}})$ operations. The dominant term gives a total complexity of $\mathcal{O}(D \cdot N_{\text{r}} \cdot N_{\text{t}} \log N_{\text{t}})$, which is higher than linear interpolation but provides superior noise reduction capabilities.

This **frequency-domain** approach begins with transforming the time-domain channel to frequency domain using fast Fourier transform (FFT) operations, requiring $\mathcal{O}(D \cdot N_{\text{r}} \cdot N_{\text{t}} \log N_{\text{FFT}})$ complexity, where N_{FFT} is typically chosen as $N_{\text{FFT}} \geq N_{\text{t}}$ for adequate frequency resolution. Pilot subcarrier extraction across all frequency bins requires $\mathcal{O}(N_{\text{r}} \cdot N_{\text{t}} \cdot N_s)$ operations, where N_s is the number of pilot subcarriers. The interpolation in the frequency domain to estimate the channel at data subcarriers has complexity $\mathcal{O}(N_{\text{r}} \cdot N_{\text{t}} \cdot N_{\text{FFT}})$. Finally, the inverse FFT to recover the time-domain channel requires $\mathcal{O}(D \cdot N_{\text{r}} \cdot N_{\text{t}} \log N_{\text{FFT}})$ operations. The total complexity is $\mathcal{O}(D \cdot N_{\text{r}} \cdot N_{\text{t}} \log N_{\text{FFT}} + N_{\text{r}} \cdot N_{\text{t}} \cdot N_{\text{FFT}})$, which is the highest among the three methods but provides the best initial estimation performance as demonstrated in our results.

Table 3: Computational complexity of initial channel estimation methods

Method	**Complexity**
LS + linear interpolation	$\mathcal{O}\big(D\, N_r\, (N_p + N_t)\big)$
LS + DFT denoising	$\mathcal{O}\big(D\, N_r\, N_t\, \log N_t\big)$
LS-OFDM (Frequency-domain)	$\mathcal{O}\big(D\, N_r\, N_t\, \log N_{\text{FFT}} \;+\; N_r\, N_t\, N_{\text{FFT}}\big)$

Table 3 shows that the complexity hierarchy is; LS with linear interpolation $<$ LS with DFT denoising $<$ LS-OFDM, with the performance improvements justifying the increased computational cost for applications requiring high estimation accuracy.

Now, we analyze the neural network's computational requirements and compare it to a diffusion model inspired by [8], which we have adapted for our dataset and experimental setup. Table 4 highlights the key metrics for both architectures.

Table 4: Complexity, latency, and parameters comparison

	FLOPs	**Latency [ms]**	**Parameters**
(N_t, N_r, D)	$(576, 4, 16)$	$(576, 4, 16)$	$(576, 4, 16)$
PINN	70.85G	11.12	3.5×10^8
DM	130.15G	50.30	5.5×10^4

As seen in Table 4, our PINN requires significantly fewer FLOPs and achieves much lower inference latency, because of the complexity of the backward pass of the diffusion model. The diffusion architecture, however, maintains an extremely low parameter, making it more memory-efficient. In our design, most parameters reside in large linear and convolutional layers, which can benefit substantially from optimized implementations. Additionally, the use of efficient skip connections enhances our model's performance without increasing its computational cost.

E Dataset

For more realistic simulations of an urban environment in the upper-mid band, we used *Wireless Insite* [14] for generating communication channels. This ray tracing tool takes into account the geographical and morphological features of the propagation environment. It simulates the behavior of each MPCs between the transmitter and receiver by following physical principles, including free-space power loss and the interaction with various objects. This enables us to compute information for each MPCs, including complex amplitude, directions of departure and arrival, and delay.

In this setup, as we can see in Figure 4, the BS is deployed at the top of a tall building for more coverage and according to tilting. This will also help us to avoid the problems that might be caused by near-field estimation. We also used the Boston map in this software for simulations. We created static channels in the environment. The more details are mentioned in Table 5.

Table 5: Key characteristics of the Boston ray-traced channel dataset.

Parameter	**Value**
Map scale (m^2)	350×450
Cropped map scale (m^2)	10×10
Carrier frequency (GHz)	15 and 8
Bandwidth (MHz)	400 and 200
Number of samples	9877

After extracting the paths for each snapshot, the channels are created using (2). We considered a raised-cosine filter with a rolloff factor of 0.4 as the pulse shaping function. We also used *Wireless Insite* for creating the RSS map. Figure 4 shows the result of solving the Maxwell equations using the described method.

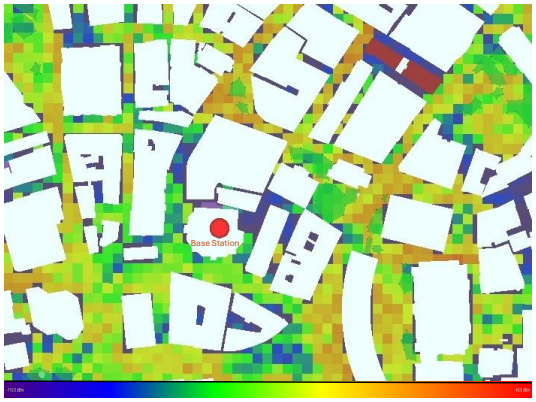


Figure 4: RSS map for the Boston environment with $P_T = 50$ dBm.

We used the location of the users at each snapshot with a horizontal Gaussian random noise of $\mathcal{N}(0,\ 9\mathbf{I})$. This assumption is valid based on the average error of the global positioning system (GPS) [33]. To provide more experimental results across different scenarios, we also used an urban canyon environment, as shown in Figure 5. The urban canyon map is usually considered a simpler map with fewer details compared to the Boston environment. In general, ray tracing tools such as *Wireless InSite* provide accurate approximations of real-world propagation, as they explicitly model electromagnetic interactions with buildings, streets, and other objects. Their ability to capture multipath, diffraction, and shadowing makes them widely accepted for validating new channel estimation methods. As shown in prior works [34, 35], the resulting synthetic channels closely mimic measured data, offering a reliable and reproducible alternative to costly field trials.

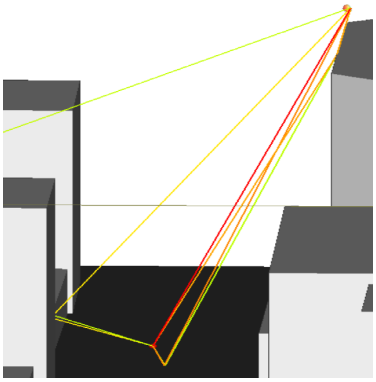


Figure 5: Urban canyon environment with $f_c = 15$ GHz. The MPCs with gains above -120 dBm are plotted.

F Implementation Details

Figure 1 highlights the three core components of our network—encoder, latent domain, and decoder—and illustrates how data flows through each stage. Although the dimensionality of these modules is flexible, Table 6 provides the exact layer sizes and overall structure we adopt in this design.

Table 6: Physics-Informed U-Net architecture and parameters. $\downarrow$ and $\uparrow$ represent the downsampling and upsampling layers, respectively

Encoder			Latent			Decoder		
#	Type	Output Size	#	Type	Output Size	#	Type	Output Size
Input	Channel	$32 \times 4 \times 576$	1	RSS Encoder	256	Output	Channel	$32 \times 4 \times 576$
1 ($\downarrow$)	ResUNetBlock	$64 \times 2 \times 288$	2	Cross-Attention	18 432	1 ($\uparrow$)	ResUNetBlock	$(128 + 128) \times 1 \times 144$
2 ($\downarrow$)	ResUNetBlock	$128 \times 1 \times 144$	3	Transformer	72×256	2 ($\uparrow$)	ResUNetBlock	$(64 + 64) \times 2 \times 288$
3 ($\downarrow$)	ResUNetBlock	$256 \times 1 \times 72$				3 ($\uparrow$)	ResUNetBlock	$32 \times 4 \times 576$

The ResUNet blocks are a pivotal component within both the encoder and decoder, merging U-Net's effective feature-preserving skip connections with ResNet's residual learning framework [36, 37]. By combining these two architectures, we achieve superior feature propagation and gradient stability—qualities that are especially advantageous for channel estimation. Table 7 lays out the precise implementation details of these blocks in both their downsampling and upsampling configurations.

Table 7: ResUNetBlock architecture details

Downsampling Block			Upsampling Block		
#	Layer	Parameters	#	Layer	Parameters
1	Conv2d	3×3, stride=1, pad=1	1	ConvTranspose2d	3×3, stride=2, pad=1
2	GroupNorm	8 groups	2	GroupNorm	8 groups
3	LeakyReLU	$\alpha = 0.2$	3	ReLU	-
4	Conv2d	3×3, stride=2, pad=1	4	Conv2d	3×3, stride=1, pad=1
5	GroupNorm	8 groups	5	GroupNorm	8 groups
6	LeakyReLU	$\alpha = 0.2$	6	ReLU	-
Residual Connection			**Residual Connection**		
1	Conv2d	1×1, stride=2, pad=0	1	ConvTranspose2d	1×1, stride=2, pad=0
2	GroupNorm	8 groups	2	GroupNorm	8 groups

To process RSS inputs, we employ a compact CNN-based encoder that efficiently distills spatial RSS patterns into a fixed-length feature vector for downstream channel estimation. As detailed in Table 8, the encoder begins with a series of 3×3 convolutions and ReLU activations interleaved with 2×2 max-pooling to gradually increase channel depth while reducing spatial dimensions. A final adaptive

average pooling and flatten operation produces a 256-dimensional embedding that captures the most salient RSS features.

Table 8: RSS encoder architecture details

Stage	Operations	Output Size
Input	RSS Map input	$1 \times 30 \times 30$
Block 1	Conv(3×3, 32) $\rightarrow$ ReLU $\rightarrow$ MaxPool(2×2)	$32 \times 15 \times 15$
Block 2	Conv(3×3, 64) $\rightarrow$ ReLU $\rightarrow$ MaxPool(2×2)	$64 \times 7 \times 7$
Block 3	Conv(3×3, 128) $\rightarrow$ ReLU $\rightarrow$ MaxPool(2×2)	$128 \times 3 \times 3$
Block 4	Conv(3×3, 256) $\rightarrow$ ReLU	$256 \times 3 \times 3$
Pooling	AdaptiveAvgPool(1×1) $\rightarrow$ Flatten	256

After defining hyperparameters from Table 1, we proceed with training. The dataset is partitioned into 80% for training, 10% for validation, and 10% for testing. Prior to training, we normalize every sample by a fixed constant to ensure consistency when minimizing the NMSE reconstruction loss. An identical normalization—using the same constant—is applied to the RSS-derived power values so that all inputs and targets share the same scale. Some of the training loss curves are shown in Figure 6. All experiments were run on an NVIDIA GeForce RTX 4090 GPU.

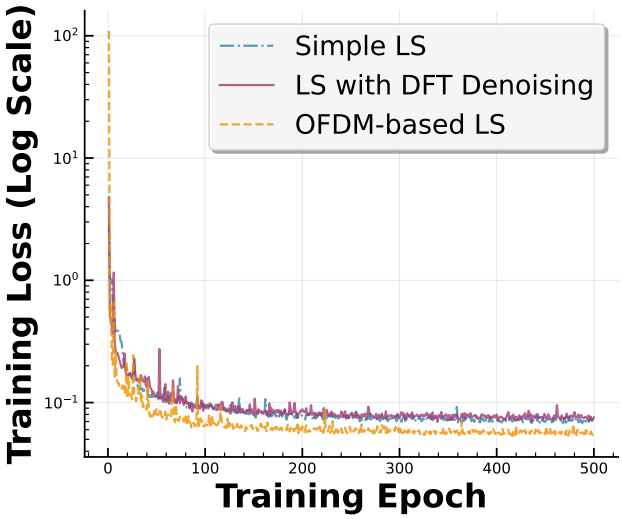


Figure 6: Log–scale comparison of training loss for the selected models.

Figure 6 clearly demonstrates the convergence of this method for different initial estimations.

G Additional Experiments

We begin by conducting an ablation study to determine the optimal weighting parameter ζ in (6). Figure 7 presents the NMSE performance as a function of ζ for the Boston dataset using LS-OFDM initial estimates with $N_p = 4$ pilots at $SNR = 0$ dB. The results indicate that $\zeta = 0.01$ yields optimal performance, balancing the reconstruction accuracy (NMSE term) with the physics-based power constraint. This value is adopted for all subsequent experiments.

It is important to note that both the initial channel estimates and RSS maps undergo global normalization prior to network input and loss computation by a fixed number (the global maximum value across all the channels). This normalization ensures numerical stability during training and meaningful gradient propagation across different physical quantities.

To demonstrate the generalization capability of our physics-informed approach, we conducted transfer learning experiments from 15 GHz to 8 GHz, using initial LS-OFDM estimation and four pilot signals at $\mathrm{SNR} = 0$ dB. We also changed the bandwidth for the new frequency band as 200 MHz. The model was initially trained on the 15 GHz dataset and then fine-tuned on 8 GHz data with varying amounts

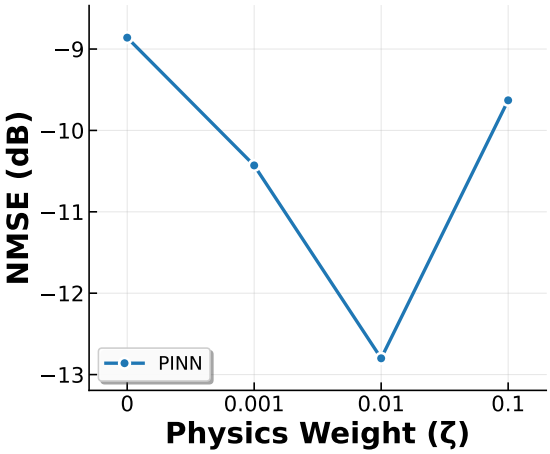
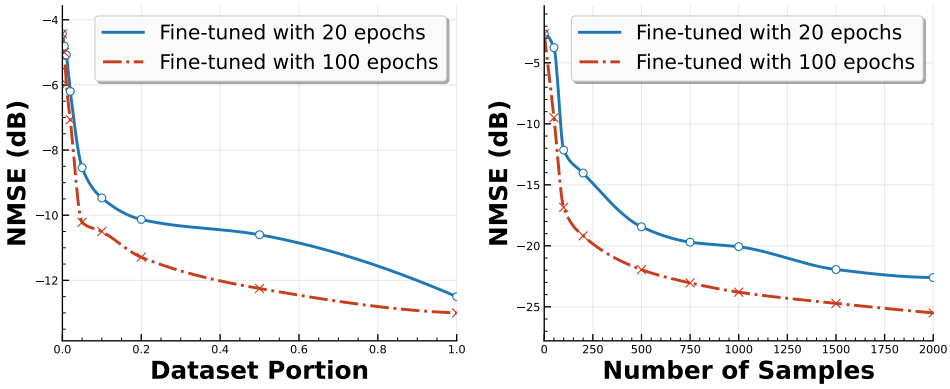


Figure 7: Ablation study of weight for physical component in loss function.

of training samples and different fine-tuning durations. Figure 8a shows the performance across different dataset portions. The results show that even with only 10% of the 8 GHz training data, the fine-tuned model achieves NMSE values of approximately -4.5 dB to -5.0 dB. Performance improves consistently as more data is used, reaching around -13 dB with the full dataset which aligns with the experiments for 15 GHz. To further validate the adaptability of our physics-informed approach, we conducted transfer learning experiments from the Boston environment to an urban canyon scenario. Figure 8b shows the NMSE performance as a function of the number of training samples for both 20 and 100 epoch fine-tuning.



(a) From 15 GHz to 8 GHz in Boston environment (b) From Boston to urban canyon environment

Figure 8: Transfer learning performance: NMSE vs. dataset for 20 and 100 epoch fine-tuning using $N_p = 4$ and $\text{SNR} = 0$.

The results demonstrate rapid adaptation with limited data: using only 100 samples, the model achieves NMSE values of approximately -4 dB (20 epochs) and -9 dB (100 epochs). Performance continues to improve substantially with more samples, reaching -22 dB and -25 dB respectively with 2000 samples. This confirms that our PINN framework effectively captures environment-agnostic electromagnetic principles, enabling efficient adaptation to different urban propagation scenarios with minimal training. We can also conclude that when the network is trained on a more complex dataset, like the Boston map, it learns generalizable electromagnetic propagation principles that enable effective adaptation to different propagation scenarios, such as urban canyons, with minimal additional training data, achieving a better performance than the original dataset.