On the Impact of Training Mismatch in Deep-Learning-Based Channel Denoising

Sungyoung Ha and Yo-Seb Jeon
Department of Electrical Engineering, POSTECH, Pohang, Gyeongbuk 37673, South Korea
Email: {sungyoungha, yoseb.jeon}@postech.ac.kr

***Abstract*—To further improve the accuracy of channel estimation in multiple-input multiple-output orthogonal frequency-division multiplexing (MIMO-OFDM) systems, we study deep-learning (DL)-based channel denoising. We train a denoising neural network offline on simulated realizations from a single baseline channel scenario and evaluate the pre-trained model—without adaptation—across multiple online deployment environments. When the deployment environment aligns with the training scenario, the denoiser provides the strongest gains; however, under distribution shift (e.g., differing delay/Doppler characteristics or signal-to-noise ratio regimes), estimation accuracy degrades, revealing limited cross-scenario generalization. These findings highlight the sensitivity of DL-based denoising to channel mismatch and motivate further investigation into online robustness to sustain performance in dynamic wireless settings.**

I. Introduction

Accurate channel estimation plays a crucial role in multiple-input multiple-output orthogonal frequency-division multiplexing (MIMO-OFDM) systems, as reliable channel state information at the receiver directly impacts both reliability and spectral [1]. In practical systems, pilot-aided estimation is performed using demodulation reference signals (DM-RSs) sparsely embedded in the time–frequency resource grid [2]. Linear estimators such as least squares (LS) is applied at DM-RS positions, followed by linear interpolation to reconstruct the full channel response. However, due to the limited pilot density and measurement noise, these estimates can be noticeably inaccurate.

To mitigate this issue, channel denoising is employed as a post-processing step that refines pilot and interpolation-based estimates by suppressing noise. Recently, deep learning (DL)-based denoising approaches have shown strong potential. Neural networks can learn mappings from noisy to clean channel estimates using large simulated datasets without explicit statistical modeling. Motivated by the two-dimensional time-frequency structure of the channel, many studies treat the channel estimate as an image and apply convolutional architectures. Representative designs include complex-valued residual denoisers and two-stage models that combine super-resolution with denoising [3], [4].

This work was supported in part by the Institute of Information & communications Technology Planning & Evaluation (IITP) Grant Funded by MSIT (Intelligent 6G Wireless Access System) under Grant RS-2021-II210467 and in part by a grant of Korean ARPA-H Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare (No. RS-2025-25454431).

Most existing DL-based denoisers are trained offline on simulated or pre-collected data and then deployed without adaptation. We therefore examine how well pre-trained models retain performance when deployment conditions diverge from those seen in training. In this work, we train a denoising neural network offline on a specific baseline channel scenario and evaluate the fixed model without online adaptation across multiple deployment environments. The denoiser performs best when deployment matches training. Under distribution shift, such as changes in delay and Doppler statistics or other channel characteristics, accuracy degrades, indicating limited cross-scenario generalization.

Our experiments provide an empirical characterization of channel-mismatch sensitivity and clarify the limitations of offline-only training for DL-based channel denoising. The results point to a need for robustness mechanisms during deployment to maintain performance under nonstationary channels.

II. System Model

We consider a MIMO-OFDM system with N_t transmit antennas, N_r receive antennas, and K subcarriers. Let (n,k) index the OFDM symbol and subcarrier, which together define a resource element (RE). The RE grid is partitioned into a pilot set $\mathcal{I}_p$ that carries DM-RS and a data set $\mathcal{I}_d$ that carries user symbols. The two sets need not be disjoint. At REs in $\mathcal{I}_p \setminus \mathcal{I}_d$ a known pilot vector $\mathbf{x}_p[n,k] \in \mathbb{C}^{N_t}$ is transmitted. At REs in $\mathcal{I}_d \setminus \mathcal{I}_p$ a data vector $\mathbf{x}_d[n,k] \in \mathcal{X}^{N_t}$ is transmitted, where $\mathcal{X}$ denotes the modulation constellation. Symbols are normalized per transmit antenna so that $\mathbb{E}[|x_{p,i}[n,k]|^2] = \mathbb{E}[|x_{d,i}[n,k]|^2] = 1$ for $i = 1, \ldots, N_t$. Then, the received signal is given by

$$\mathbf{y}[n,k] = \mathbf{H}[n,k]\mathbf{x}[n,k] + \mathbf{v}[n,k], \tag{1}$$

where $\mathbf{H}[n,k] \in \mathbb{C}^{N_r \times N_t}$ denotes the channel frequency response (CFR) and $\mathbf{v}[n,k] \sim \mathcal{CN}(\mathbf{0}, \sigma^2 \mathbf{I}_{N_r})$ denotes additive noise. Measurements on pilot REs are used to estimate $\mathbf{H}[n,k]$ and data REs are demodulated using the channel estimate $\hat{\mathbf{H}}[n,k]$.

III. DL-Based Channel Denoising

We adopt a DL-based channel denoising approach to mitigate errors in pilot-aided channel estimation. The key idea is to view the CFR over the time-frequency plane as a two-dimensional image and to apply a denoising neural network to recover a cleaner CFR. To control computational cost, the denoiser operates only on a subsampled CFR grid defined by the DM-RS positions, forming an $M_t \times M_f$ sub-grid of OFDM symbols and subcarriers.

For each transmit–receive antenna pair (r, t), let $\mathbf{M}^{(r,t)} \in \mathbb{C}^{M_t \times M_f}$ denote the true sub-CFR on this grid and let $\hat{\mathbf{M}}^{(r,t)} \in \mathbb{C}^{M_t \times M_f}$ be its noisy estimate obtained from pilot-aided channel estimation. A denoising network $f_{\text{Dn}}(\cdot; \theta)$ maps the noisy input to the denoised output

$$\tilde{\mathbf{M}}^{(r,t)} = f_{\text{Dn}}(\hat{\mathbf{M}}^{(r,t)}; \theta). \tag{2}$$

Complex-valued inputs may be processed by stacking real and imaginary parts as separate channels, the formulation does not depend on any specific network architecture. The denoised sub-grid is then expanded to the full time-frequency grid using standard two-dimensional interpolation.

Most existing DL-based denoisers are trained offline on datasets intended to represent deployment conditions and are subsequently used for inference without adaptation. Because wireless channels are inherently dynamic, test conditions are difficult to anticipate. Consequently, pre-trained models often exhibit performance degradation under distribution mismatch or out-of-distribution scenarios between training and deployment. Therefore, adaptation during deployment is necessary to maintain denoising performance under distribution shift.

IV. Simulation Results

We evaluate a MIMO-OFDM system at a carrier frequency of 3.5 GHz with a subcarrier spacing of 15 kHz and $K = 512$ subcarriers serving four users. Each slot comprises 14 OFDM symbols, and each resource block spans 12 adjacent subcarriers. Modulation is 4-quadrature amplitude modulation. The antenna configuration is $N_t = 2$ transmit and $N_r = 16$ receive, and low-density parity check code with rate $1/2$ is applied. LS channel estimation and linear two-dimensional interpolation are applied.

For channel denoising, we use a denoising convolutional neural network (DnCNN) proposed in [5] model trained offline and kept fixed at test time (no online adaptation). This architecture is trained to learn a residual mapping $\mathcal{R}(\hat{\mathbf{M}}^{(r,t)}; \theta)$ that estimates the noise component in the input. To this end, we define the loss function following the formulation

$$\ell(\hat{\mathbf{M}}, \mathbf{M}; \theta) = \left\| \mathcal{R}(\hat{\mathbf{M}}; \theta) - (\hat{\mathbf{M}} - \mathbf{M}) \right\|_{\text{F}}^2, \tag{3}$$

where $\| \cdot \|_{\text{F}}$ denotes Frobenious norm. A total of $N_{\text{total}} = 4$ slots are processed, all for inference. The network input window is $(M_t, M_f) = (8, 58)$, corresponding to the numbers of OFDM symbols and subcarriers in each sub-CFR grid.

Fig. 1 plots frame error rate (FER) versus signal-to-noise ratio (SNR) for a denoiser trained offline on clustered delay line (CDL)-B with a 300 ns delay spread and 30 km/h mobility, then evaluated without adaptation across multiple test environments. The matched condition (CDL-B, 300 ns, 30 km/h) attains the lowest FER and the steepest decay with SNR. In contrast, mismatched channel models (CDL-A/C/D/E) yield noticeably

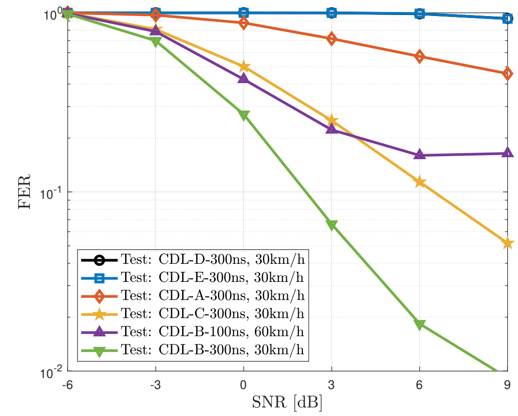

Fig. 1. FER comparison under various channel environments.

higher FER across SNR, with some curves exhibiting clear error floors indicative of severe distribution shift. Even within the same model family, changes in delay spread or user speed reduce performance. For example, CDL-B with 100 ns delay spread and 60 km/h performs worse than the matched baseline. Overall, the results show sensitivity to channel and mobility mismatch and limited cross-scenario generalization of the offline-trained denoiser.

V. Conclusion

This paper examined DL-based channel denoising for MIMO-OFDM when a denoiser trained offline on a specific baseline scenario is deployed without adaptation in diverse environments. We applied DnCNN on DM-RS subgrids and evaluated performance across multiple channel models, delay spreads, and user speeds. The results showed a consistent pattern. Performance was strongest when deployment conditions matched training. Under distribution shift, including changes within the same model family, FER increased and error floors appeared. These outcomes indicate limited cross-scenario generalization and high sensitivity to channel and mobility mismatch. Although the absolute figures depend on the chosen system configuration, the trend remained stable across the tested settings.

Future work will examine approaches that improve robustness at deployment, including broader training distributions and online adaptation designed to sustain denoising performance under nonstationary channels.

References

[1] H. Bolcskei, D. Gesbert, and A. J. Paulraj, "On the capacity of OFDM-based spatial multiplexing systems," *IEEE Trans. Commun.*, vol. 50, no. 2, pp. 225–234, Feb. 2002.

[2] "Physical channels and modulation," 3GPP, TS 38.211, V16.2.0, 2020.

[3] Q. Zhao, X. Zeng, Z. Fan, Q. Zhang, and W. Li, "Channel estimation for FDD massive MIMO with complex residual denoising network," *IEEE Wireless Commun. Letters*, vol. 13, no. 8, pp. 2070–2074, Aug. 2024.

[4] M. Soltani, V. Pourahmadi, A. Mirzaei, and H. Sheikhzadeh, "Deep learning-based channel estimation," *IEEE Commun. Lett.*, vol. 23, no. 4, pp. 652–655, Apr. 2019.

[5] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising," *IEEE Trans. Image Process.*, vol. 26, no. 7, pp. 3142–3155, July 2017.