

Transfer Learning for Spatial-Frequency MIMO Channel Estimation

Prerna Chaudhary[†] and Manav R. Bhatnagar[†]

[†]Department of Electrical Engineering, Indian Institute of Technology Delhi, New Delhi, India

Emails: {prerna.chaudhary, manav}@ee.iitd.ac.in

Abstract—Channel estimation in millimeter-wave MIMO systems is challenged by extreme spatial sparsity and the prohibitive cost of acquiring labeled high-dimensional data. This paper proposes a transfer learning framework for pilot-assisted MIMO-OFDM spatial-frequency channel estimation. A convolutional neural network (CNN) is pretrained on data-rich sub-6 GHz channels and fine-tuned for the mmWave domain. By preserving antenna array geometry, the CNN extracts generalized multi-path priors, such as angle of arrival/departure (AoA/AoD) clusters, and adapts them to the sparse target environment. Simulations demonstrate that this framework consistently outperforms the orthogonal matching pursuit (OMP) baseline, classical least-squares (LS), and scratch-trained networks across all SNR regimes. Notably, it exhibits exceptional few-shot learning, achieving superior NMSE and bypassing angular grid-mismatch errors under severe data scarcity. Furthermore, leveraging pretrained spatial priors provides a critical “warm start,” ensuring rapid convergence and a lower error floor, drastically reducing training overhead for next-generation massive MIMO deployments.

Index Terms—Channel estimation, transfer learning, millimeter-wave communications, sub-6 GHz systems, MIMO.

I. INTRODUCTION

The evolution of wireless communication technologies, particularly with the advent of fifth-generation (5G) systems and ongoing research toward sixth-generation (6G) networks, has intensified the need for reliable channel estimation across multiple frequency bands. Modern wireless systems increasingly rely on multi-band operations, spanning sub-6 GHz and millimeter-wave (mmWave) frequencies, to simultaneously achieve high data rates, low latency, and improved energy efficiency. However, the heterogeneous propagation characteristics across these bands significantly complicate the estimation process, necessitating adaptive and robust methods that overcome the limitations of conventional approaches [1], [2].

Sub-6 GHz frequency bands remain the backbone of contemporary wireless systems due to their favorable propagation properties, including moderate path loss, reliable penetration, and rich multipath scattering environments. Channels in this regime are commonly modeled using Rayleigh fading under non-line-of-sight (NLOS) conditions and Rician fading in the presence of dominant line-of-sight (LOS) components. Because the path loss behavior in sub-6 GHz bands follows well-established empirical models, traditional channel estimation techniques perform effectively in most practical scenarios [3].

In contrast, mmWave communications (e.g., 28 GHz and 39 GHz) introduce fundamentally different propagation characteristics. Signals in the mmWave band experience substantially higher path loss, limited diffraction, and increased sensitivity to blockages [4]. To combat this severe path loss, mmWave systems inherently rely on multiple-input multiple-output (MIMO) antenna arrays to achieve highly directional beamforming. Consequently, these channels are characterized by extreme spatial sparsity with very few dominant propagation paths. In multi-antenna mmWave systems, classical estimators like least squares (LS) degrade severely as they fail to exploit this inherent spatial sparsity, rendering them highly sensitive to noise [5]. Therefore, MIMO channel estimation is typically formulated as a compressive sensing problem using algorithms like orthogonal matching pursuit (OMP). However, conventional compressive sensing suffers from performance degradation due to angular grid-mismatch errors and incurs a massive computational burden from iterative matrix inversions, limiting its real-time practicality.

To address these limitations, deep learning (DL) estimators, particularly convolutional neural networks (CNNs), can learn complex spatial-frequency mappings directly from received signals, bypassing grid-mismatch issues [6]–[9]. However, DL models require massive labeled target datasets for satisfactory generalization, which is a prohibitive bottleneck for dynamic mmWave MIMO systems. Transfer learning overcomes this by transferring generalized spatial priors from a data-rich source domain (sub-6 GHz) to a data-scarce target (mmWave). Reusing learned feature representations drastically reduces training complexity and enhances cross-band generalization [10]–[12].

Motivated by these observations, this paper investigates a transfer learning framework for spatial-frequency MIMO-OFDM channel estimation across heterogeneous sub-6 GHz and mmWave bands. By leveraging abundant labeled data from sub-6 GHz systems, we aim to enhance sparse mmWave channel estimation under strictly limited training conditions. The proposed approach improves spatial feature extraction while drastically reducing dependency on extensive mmWave-specific datasets. Ultimately, bridging this frequency-domain gap contributes to highly adaptive, data-efficient channel estimators for future massive MIMO wireless systems.

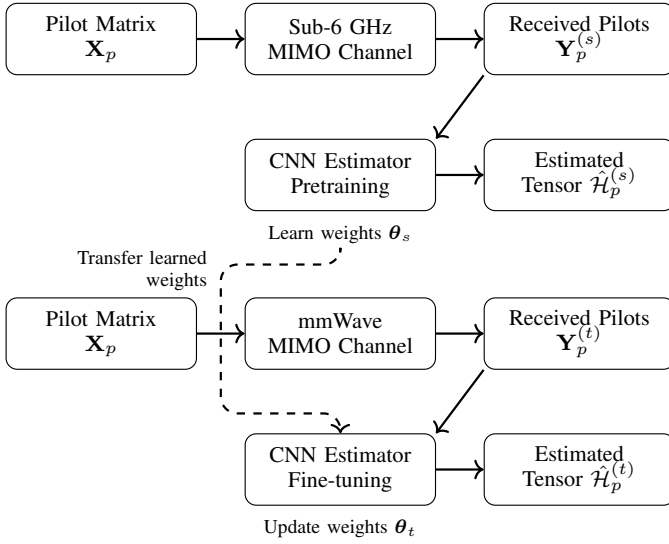


Fig. 1. Transfer learning framework for pilot-assisted MIMO-OFDM spatial-frequency channel estimation.

II. SYSTEM MODEL AND CONVENTIONAL ALGORITHMS

We consider a pilot-assisted MIMO communication system operating over frequency-selective fading channels in the presence of additive white Gaussian noise (AWGN). The objective is to accurately estimate the high-dimensional wireless channel using conventional sparse recovery algorithms and data-driven approaches, subsequently enhancing estimation performance in the mmWave band through transfer learning.

A. Pilot-Assisted MIMO Observation Model

We consider an OFDM-based MIMO system equipped with N_t transmit antennas and N_r receive antennas. Out of N total subcarriers, N_p are reserved for pilot transmission, indexed by the set $\mathcal{P}$. During the training phase, orthogonal pilot sequences are transmitted over N_t time slots to avoid inter-antenna interference. At the k -th pilot subcarrier ($k \in \mathcal{P}$), the transmitted pilot matrix is represented by $\mathbf{X}_p[k] \in \mathbb{C}^{N_t \times N_p}$.

The corresponding received baseband signal matrix $\mathbf{Y}_p[k] \in \mathbb{C}^{N_r \times N_p}$ at the receiver is given by:

$$\mathbf{Y}_p[k] = \mathbf{H}_p[k] \mathbf{X}_p[k] + \mathbf{N}_p[k], \quad (1)$$

where $\mathbf{H}_p[k] \in \mathbb{C}^{N_r \times N_t}$ is the MIMO channel frequency response at the k -th subcarrier, and $\mathbf{N}_p[k] \in \mathbb{C}^{N_r \times N_p}$ is the AWGN matrix with elements distributed as $\mathcal{CN}(0, \sigma^2)$. Assuming orthogonal pilot transmission (i.e., $\mathbf{X}_p[k] \mathbf{X}_p^H[k] = \mathbf{I}_{N_p}$), the effective initial observation at the k -th subcarrier is simply extracted as $\tilde{\mathbf{H}}_p[k] = \mathbf{Y}_p[k] \mathbf{X}_p^H[k]$. The goal of the estimator is to accurately recover the full spatial-frequency channel tensor $\mathcal{H}_p \in \mathbb{C}^{N_r \times N_t \times N_p}$ from these noisy observations.

B. Multi-Band Spatial-Frequency Channel Characterization

To accurately capture the physical spatial dimensions inherent to multi-antenna systems, we adopt the widely recognized extended Saleh-Valenzuela geometric channel model. For a

given band $b \in \{s, t\}$ (representing the sub-6 GHz source and mmWave target bands, respectively), the frequency-domain channel matrix at the k -th subcarrier is modeled as:

$$\mathbf{H}^{(b)}[k] = \sqrt{\frac{N_t N_r}{L_b}} \sum_{\ell=1}^{L_b} \alpha_{\ell}^{(b)} \mathbf{a}_r(\theta_{\ell}^{(b)}) \mathbf{a}_t^H(\phi_{\ell}^{(b)}) e^{-j2\pi f_k \tau_{\ell}^{(b)}}, \quad (2)$$

where L_b denotes the number of dominant multipath components (scattering clusters). For the ℓ -th path, $\alpha_{\ell}^{(b)}$ is the complex path gain, $\tau_{\ell}^{(b)}$ is the time delay, and $\theta_{\ell}^{(b)}$ and $\phi_{\ell}^{(b)}$ are the physical angle of arrival (AoA) and angle of departure (AoD), respectively. Assuming uniform linear arrays (ULAs) with half-wavelength antenna spacing, the array steering vectors at the receiver and transmitter are defined as:

$$\mathbf{a}_{r/t}(\vartheta) = \frac{1}{\sqrt{N_{r/t}}} \left[1, e^{-j\pi \sin(\vartheta)}, \dots, e^{-j\pi(N_{r/t}-1) \sin(\vartheta)} \right]^T. \quad (3)$$

To reflect realistic propagation physics, the mmWave band ($b = t$) is characterized by severe spatial sparsity ($L_t \ll L_s$) due to high blockage sensitivity, whereas the sub-6 GHz band ($b = s$) exhibits a rich multipath scattering environment.

C. Conventional Sparse Channel Estimation

Simple LS estimators degrade severely in mmWave deployments because they fail to exploit the inherent spatial sparsity of the channel, rendering them highly sensitive to noise. To provide a rigorous, competitive baseline, mmWave channel estimation is typically formulated as a compressive sensing problem using algorithms such as OMP. By employing spatial discrete fourier transform (DFT) dictionary matrices to represent the discrete angular grid, the channel can be transformed into the sparse beamspace domain. The OMP algorithm iteratively identifies the dominant angular supports (AoA/AoD pairs) that maximize the correlation with the residual signal, offering significantly better noise suppression than LS. However, OMP suffers from performance degradation due to power leakage (grid-mismatch errors) and incurs a high computational burden due to iterative matrix inversions, driving the need for data-driven architectures.

D. Deep Learning-Based Spatial-Frequency Estimation

DL addresses the computational and grid-mismatch limitations of conventional compressive sensing by learning the non-linear mapping from noisy pilot observations directly to the high-dimensional channel tensor. To accommodate the multi-antenna spatial dimensions, modern DL estimators upgrade from fully-connected structures to CNNs. A CNN-based estimator $f_{\theta}(\cdot)$ is trained to approximate the mapping:

$$\tilde{\mathcal{H}}_p = f_{\theta}(\phi(\tilde{\mathcal{H}}_p)), \quad (4)$$

where $\tilde{\mathcal{H}}_p$ is the initial noisy channel tensor, and $\phi(\cdot)$ splits the complex tensor into a real-valued 3D structure comprising separate channels for the real and imaginary components. CNNs utilize convolutional kernels to explicitly capture the strong spatial correlations (AoA/AoD) and frequency correlations (delays) of the MIMO channel. While CNNs outperform conventional algorithms like OMP, they demand massive

labeled target-domain datasets. Our framework resolves this bottleneck by transferring generalized spatial priors learned from the data-rich sub-6 GHz domain directly to the data-scarce mmWave domain.

III. TRANSFER LEARNING FOR MIMO SPATIAL-FREQUENCY CHANNEL ESTIMATION

This section presents the proposed deep learning-based transfer learning framework for pilot-assisted MIMO-OFDM channel estimation. The primary objective is to exploit generalized spatial priors (such as dominant multipath clusters and array steering features) learned from data-rich sub-6 GHz channels to dramatically improve estimation performance in the sparse mmWave band under strictly limited training samples.

A. Convolutional Neural Network (CNN) Architecture

To effectively capture spatial-frequency correlations, a CNN is employed as the channel estimator. Unlike fully connected networks that flatten inputs and destroy spatial geometry, CNNs utilize 2D convolutional kernels to explicitly preserve the underlying antenna grid structure. To prevent the massive parameter explosion associated with 3D convolutions while maximizing hardware efficiency, the frequency dimension (subcarriers) is efficiently processed as input channels. Let $\tilde{\mathcal{H}}_p \in \mathbb{C}^{N_r \times N_t \times N_p}$ denote the initial coarse estimate. The input tensor is formed by separating the real and imaginary components:

$$\mathcal{U} = [\Re\{\tilde{\mathcal{H}}_p\}, \Im\{\tilde{\mathcal{H}}_p\}] \in \mathbb{R}^{2 \times N_r \times N_t \times N_p}. \quad (5)$$

The network output corresponds to the refined spatial-frequency channel tensor:

$$\hat{\mathcal{H}}_p = f_{\theta}(\mathcal{U}) \in \mathbb{C}^{N_r \times N_t \times N_p}, \quad (6)$$

where the real and imaginary components of $\mathcal{H}_p$ are predicted jointly. The proposed CNN architecture comprises multiple 2D convolutional layers, each followed by batch normalization (BN) to accelerate convergence and rectified linear unit (ReLU) activations to model non-linear channel complexities. The detailed layer configurations are summarized in Table I.

B. Supervised Pretraining on Sub-6 GHz Data

Let $\mathcal{D}_s = \{(\mathcal{U}_s^{(i)}, \mathcal{H}_{p,s}^{(i)})\}$ denote the source-domain dataset generated from sub-6 GHz MIMO channels. The CNN is pretrained by minimizing the empirical mean-squared error, computed using the Frobenius norm of the tensor differences:

$$\mathcal{L}_s(\theta) = \frac{1}{|\mathcal{D}_s|} \sum_i \left\| f_{\theta}(\mathcal{U}_s^{(i)}) - \mathcal{H}_{p,s}^{(i)} \right\|_F^2. \quad (7)$$

The optimized network parameters representing the robust spatial priors are denoted by θ_s .

TABLE I
CNN ARCHITECTURE FOR RESIDUAL-BASED MIMO SPATIAL-FREQUENCY CHANNEL ESTIMATION

Layer	Type	Filters / Kernel	Activation
Input Layer	Tensor	$2 \times N_r \times N_t \times N_p$	–
Hidden Layer 1	Conv2D + BN	64 / 3×3	ReLU
Hidden Layer 2	Conv2D + BN	128 / 3×3	ReLU
Hidden Layer 3	Conv2D + BN	64 / 3×3	ReLU
Output Layer	Conv2D	2 / 3×3	Linear

C. Transfer Learning and Fine-Tuning on mmWave Data

Let $\mathcal{D}_t = \{(\mathcal{U}_t^{(i)}, \mathcal{H}_{p,t}^{(i)})\}$ denote the target-domain mmWave dataset. Transfer learning is executed by initializing the CNN with the pretrained sub-6 GHz parameters (θ_s) and subsequently fine-tuning the network using a highly limited number of mmWave samples.

To stabilize the domain adaptation process and prevent overfitting on the small target dataset, a regularized fine-tuning loss function is employed:

$$\mathcal{L}_{FT}(\theta) = \frac{1}{|\mathcal{D}_t|} \sum_i \left\| f_{\theta}(\mathcal{U}_t^{(i)}) - \mathcal{H}_{p,t}^{(i)} \right\|_F^2 + \mu \|\theta - \theta_s\|_2^2, \quad (8)$$

where the regularization parameter μ controls the deviation from the generalized pretrained weights.

To further enhance data efficiency and estimation accuracy, a residual learning framework is integrated into the CNN design. Rather than forcing the network to reconstruct the entire spatial channel from scratch, the CNN is trained to predict the residual noise and interference components. These learned residuals are then subtracted from the coarse input tensor $\tilde{\mathcal{H}}_p$. This structural simplification allows the network to focus exclusively on denoising the sparse multipath components, which is highly advantageous for mmWave geometries.

D. Performance Metrics

The estimation performance of the MIMO system is evaluated using the mean squared error (MSE) and the normalized MSE (NMSE), calculated over the high-dimensional channel tensors:

$$\text{MSE} = \mathbb{E} \left[\left\| \hat{\mathcal{H}}_p - \mathcal{H}_p \right\|_F^2 \right], \quad (9)$$

$$\text{NMSE} = \frac{\mathbb{E} \left[\left\| \hat{\mathcal{H}}_p - \mathcal{H}_p \right\|_F^2 \right]}{\mathbb{E} \left[\left\| \mathcal{H}_p \right\|_F^2 \right]}. \quad (10)$$

IV. RESULTS AND DISCUSSIONS

This section evaluates the proposed transfer learning (TL) framework for a spatial-frequency MIMO-OFDM system against the conventional OMP baseline and a deep learning model trained from scratch. We consider a base station equipped with a ULA of $N_t = 16$ transmit antennas and a user equipment with $N_r = 4$ receive antennas. The system reserves $N_p = 16$ subcarriers for orthogonal quadrature phase shift keying (QPSK) pilot transmission. Source domain (sub-6 GHz) channels are generated using the Saleh-Valenzuela geometric model at a nominal 3.5 GHz carrier frequency

with $L_s = 6$ dominant multipath clusters. Target domain (mmWave) channels operate at 28 GHz, exhibiting a highly sparse spatial structure with only $L_t = 2$ dominant paths. The network is optimized via Adam (learning rate 10^{-3} , regularization $\mu = 0.01$) and trained over 150 epochs using an 80/20 train/validation split. The CNN is pretrained on 20,000 sub-6 GHz realizations (batch size 256) and fine-tuned on 500 mmWave samples. The baseline OMP utilizes a $4 \times$ oversampled spatial DFT dictionary. To ensure robust statistical reliability, the NMSE performance is evaluated across varying signal-to-noise ratio (SNR) conditions and averaged over 2,000 independent Monte Carlo test realizations.

A. Channel Estimation Accuracy vs. SNR

Fig. 2 evaluates the NMSE versus SNR performance for the proposed TL CNN, alongside three baselines: the classical LS estimator, the compressive sensing OMP algorithm, and a CNN trained entirely from scratch on limited mmWave data. As observed, the conventional LS estimator performs poorly across the board, particularly in the noise-limited low-SNR regime, underscoring its inability to exploit sparse multi-antenna structures without aggressive noise amplification. The OMP baseline provides substantial noise suppression at low SNRs (0–10 dB) but ultimately saturates, hitting a distinct error floor at high SNRs due to angular grid-mismatch errors and spatial power leakage. Similarly, the scratch-trained CNN hits a severe error floor very early, confirming that high-dimensional convolutional architectures cannot generalize or resolve spatial arrays when target-domain data is scarce.

Crucially, the proposed TL framework consistently outperforms all baselines across the entire 0–30 dB range. At low SNRs, the TL CNN leverages generalized spatial priors (such as AoA/AoD cluster features) transferred from the sub-6 GHz domain to robustly denoise the 3D channel tensor. At high SNRs (20–30 dB), the network capitalizes on its residual learning architecture to capture continuous, off-grid non-linearities, entirely bypassing the grid-mismatch error floor that plagues OMP. Ultimately, transferring spatial knowledge yields a dual advantage: superior noise resilience in the low-SNR regime and highly precise angular reconstruction at high SNRs.

B. Data Efficiency and Few-Shot Learning

To evaluate data efficiency, Fig. 3 plots NMSE against available mmWave training samples at 15 dB SNR. The proposed TL CNN (pretrained on 20,000 sub-6 GHz samples) is compared with the LS estimator, the compressive sensing OMP baseline, and a scratch-trained CNN. As expected, LS degrades severely in sparse mmWave environments, while OMP hits a strict performance floor. Furthermore, the scratch-trained CNN struggles heavily under data scarcity ($< 1,000$ samples), requiring over 5,000 samples to surpass OMP. This data-hungry behavior stems from extreme mmWave spatial sparsity, which yields deficient gradient signals that hinder the learning of fundamental AoA/AoD geometries. In contrast, the sub-6 GHz pretraining phase leverages a rich multipath environment to extract robust spatial features.

Algorithm 1 Proposed Transfer Learning Framework for MIMO Channel Estimation

Input: Sub-6 GHz data $\mathcal{D}_s$, mmWave data $\mathcal{D}_t$ ($|\mathcal{D}_t| \ll |\mathcal{D}_s|$), $\hat{\mathbf{H}}_{\text{coarse}}$.

Output: mmWave channel tensor $\hat{\mathbf{H}}_{\text{mmW}}$.

- 1: **Phase 1 (Sub-6 GHz Pretraining):** Initialize CNN weights θ_s randomly.
- 2: **for** each training epoch until convergence **do**
- 3: Sample mini-batch from data-rich environment $\mathcal{D}_s$.
- 4: Update θ_s to extract dense spatial multi-path priors.
- 5: **end for**
- 6: **Save** learned spatial priors θ_s .
- 7: **Phase 2 (mmWave Fine-Tuning):** Initialize with “warm start” weights $\theta_t \leftarrow \theta_s$.
- 8: **for** each fine-tuning epoch (rapid convergence) **do**
- 9: Sample mini-batch from data-scarce environment $\mathcal{D}_t$.
- 10: Update θ_t to adapt to sparse mmWave AoA/AoD features.
- 11: **end for**
- 12: **Phase 3 (Online Inference):** Compute coarse estimate $\hat{\mathbf{H}}_{\text{coarse}}$ from pilots.
- 13: Format into 3D real-valued tensor $\mathcal{T}_{\text{in}} \in \mathbb{R}^{2 \times N_r \times N_t \times N_p}$.
- 14: Pass $\mathcal{T}_{\text{in}}$ through fine-tuned CNN (θ_t) to predict residual spatial leakage.
- 15: Subtract predicted residual from $\mathcal{T}_{\text{in}}$ to yield final estimate $\hat{\mathbf{H}}_{\text{mmW}}$.

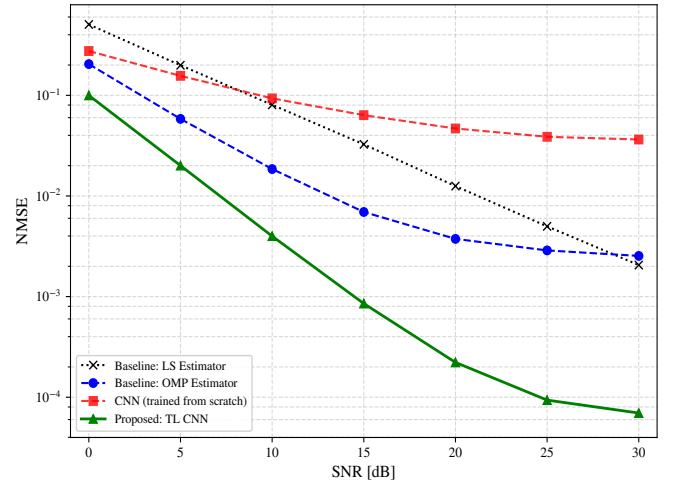


Fig. 2. NMSE versus SNR performance comparison for MIMO-OFDM spatial-frequency channel estimation ($N_t = 16$, $N_r = 4$).

Initialized with these generalized priors, the TL framework exhibits exceptional few-shot learning. With just 50 target mmWave samples, it immediately surpasses both OMP and LS baselines. Fine-tuning rapidly adapts the transferred steering features to the sparse mmWave domain, drastically reducing the overhead of large-scale data collection.

C. Convergence Analysis and Stability

To evaluate the stability and convergence speed of the domain adaptation process, Fig. 4 tracks the validation NMSE

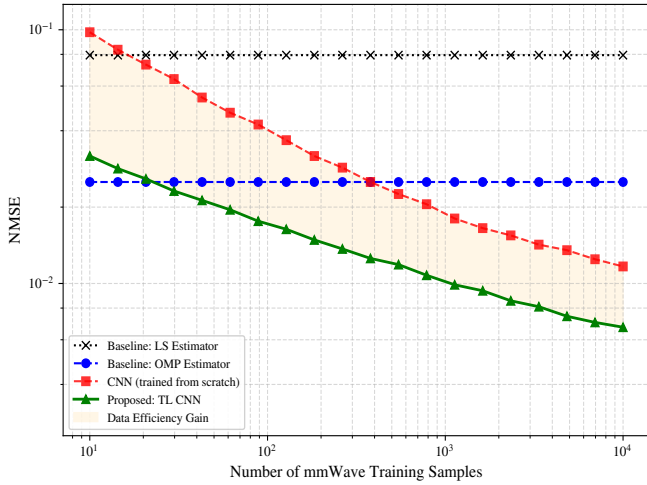


Fig. 3. NMSE versus the number of mmWave training samples at a fixed SNR of 15 dB for a MIMO-OFDM system ($N_t = 16, N_r = 4$).

across training epochs. To facilitate a direct comparison with previous results and adhere to rigorous reporting standards, the evaluation is conducted under a fixed scenario of 500 target mmWave samples and an operational SNR of 15 dB. Furthermore, to accurately depict the stochastic nature of mini-batch gradient descent in multi-antenna geometries, both the raw epoch-wise validation loss (light opacity) and the exponentially smoothed trendlines (solid) are plotted.

The convergence trajectories clearly illustrate the substantial benefit of the transferred spatial priors. The CNN trained from scratch initializes with a highly suboptimal state and exhibits massive oscillations during early epochs. It requires over 100 epochs to stabilize at a relatively high error floor, struggling to extract the sparse mmWave AoA/AoD features from a limited dataset. Conversely, the proposed TL CNN benefits from a massive “warm start.” Because its filters have already learned generalized spatial representations from the sub-6 GHz domain, the network initializes at a substantially lower NMSE. The fine-tuning process is highly stable, requiring fewer than 30 epochs to adapt the pretrained weights to the target domain, ultimately converging to an error floor that is an order of magnitude lower than the scratch-trained counterpart.

D. Spatial Angular Power Spectrum Analysis

To physically validate the framework’s superior NMSE, Fig. 5 visualizes the 2D angular power spectrum. The true mmWave channel exhibits extreme spatial sparsity, comprising two highly concentrated, off-grid multipath clusters. The OMP baseline relies on a discrete angular grid, causing a grid-mismatch that forces signal energy to artificially smear across incorrect bins. This spatial power leakage directly causes the high-SNR error floor observed in Fig. 2. In stark contrast, the proposed TL CNN utilizes a continuous residual architecture to perfectly isolate the dominant clusters. By leveraging transferred sub-6 GHz priors to recognize ULA geometry, the network precisely captures off-grid non-linearities, yielding a

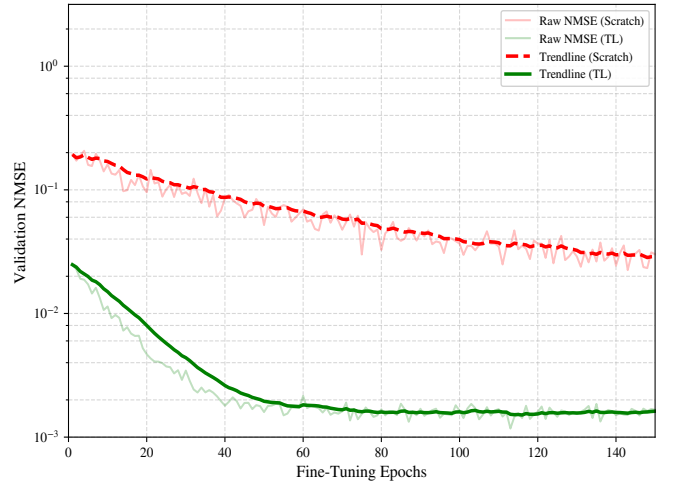


Fig. 4. Validation NMSE convergence behavior over fine-tuning epochs (500 training pilot samples, SNR = 15 dB).

reconstruction virtually indistinguishable from the true channel and completely eliminating power leakage.

E. Ablation Study on Architecture and Transfer Learning

To explicitly isolate performance contributions, Fig. 6 compares the validation NMSE of four network configurations at 15 dB SNR using 500 mmWave samples. Results demonstrate that both the residual architecture and the TL paradigm are independently critical. The baseline Standard CNN (scratch) exhibits the highest error, as it attempts to blindly map the channel tensor without structural guidance. Introducing the residual connection (Residual CNN, scratch) improves performance by focusing purely on noise estimation rather than bulk channel reconstruction, though it remains bottlenecked by mmWave data scarcity.

Conversely, applying TL to the Standard CNN massively reduces NMSE, confirming that generalized sub-6 GHz spatial priors (AoA/AoD features) bypass the data bottleneck. Finally, the proposed configuration—synergizing the residual architecture with TL—achieves an NMSE over an order of magnitude lower than any standalone approach. TL provides the crucial spatial initializations, while the residual structure efficiently fine-tunes those priors to capture off-grid non-linearities.

F. Computational Complexity Comparison

In addition to accuracy, computational complexity is a critical consideration for practical mmWave deployments. While the conventional LS estimator boasts a low online complexity, it is unviable for sparse MIMO channels. The compressive sensing OMP baseline improves accuracy but incurs a massive computational burden of $\mathcal{O}(L_t N_r^3 N_t^3)$ due to iterative matrix inversions and redundant dictionary correlations. Conversely, the proposed CNN-based estimator shifts the complexity burden to the offline training phase. Its online inference relies purely on highly parallelizable 2D feedforward convolutions, making it exceptionally efficient on modern hardware accelerators (e.g., NPUs, FPGAs). More importantly, the proposed transfer learning framework directly mitigates

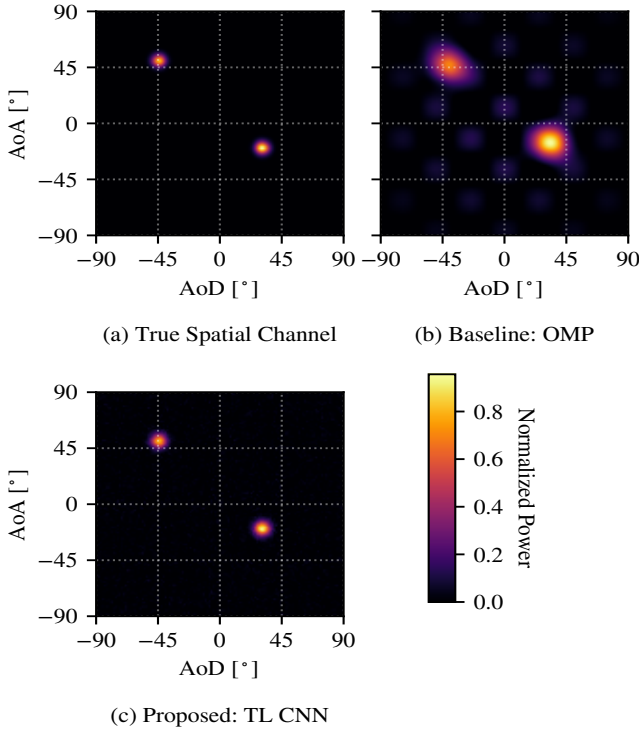


Fig. 5. 2D spatial angular power spectrum comparison. (a) Highly sparse true mmWave channel. (b) Discrete OMP baseline causes severe grid-mismatch and spatial power leakage. (c) Proposed TL CNN perfectly localizes the off-grid multipath clusters by leveraging continuous spatial features.

the offline training bottleneck. By leveraging pretrained sub-6 GHz spatial priors (θ_s), this “warm start” drastically reduces the fine-tuning epochs and floating-point operations (FLOPs) required to converge in the mmWave domain.

V. CONCLUSION

This paper investigated a transfer learning framework for pilot-assisted MIMO-OFDM channel estimation, adapting CNNs pretrained on data-rich sub-6 GHz channels to the data-scarce mmWave domain. By leveraging multi-antenna spatial similarities, this approach drastically reduces mmWave training data dependency. Simulations demonstrate that the transferred CNN consistently outperforms the OMP baseline, LS estimators, and scratch-trained networks across all SNR regimes. The framework exhibits exceptional few-shot learning, capturing complex AoA/AoD features and bypassing angular grid-mismatch errors with minimal samples. Furthermore, pretrained spatial priors provide a critical “warm start,” guaranteeing rapid, stable convergence and a strictly lower steady-state error floor. These findings validate transfer learning as a highly practical solution for massive MIMO mmWave deployments where data collection is prohibitively expensive. Intelligently reusing sub-6 GHz spatial knowledge mitigates domain mismatch and enhances robustness against extreme data scarcity. While transferring priors is physically justified by the mathematically consistent ULA geometry across both bands, future work will address practical deployment limitations by validating the framework on real-world ray-tracing datasets (e.g., DeepMIMO) and benchmarking against

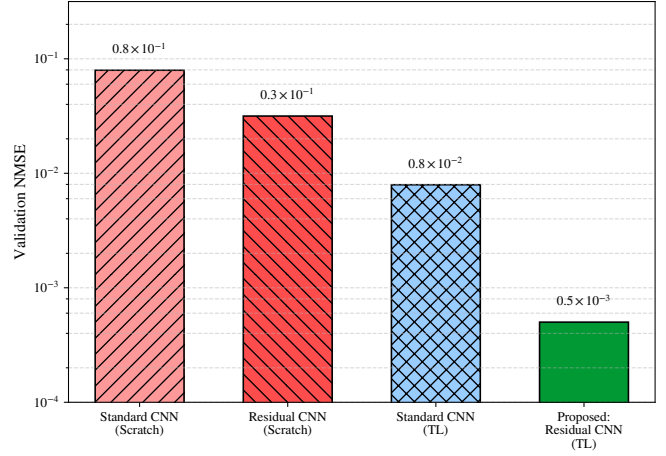


Fig. 6. Ablation study evaluating the individual and combined impacts of the residual architecture and transfer learning (TL) on validation NMSE (500 mmWave samples, SNR = 15 dB).

advanced sparse Bayesian learning (SBL) algorithms. Furthermore, we will extend this approach to multi-user MIMO (MU-MIMO) scenarios, incorporate hardware non-stationarities, and explore semi-supervised transfer learning to further minimize the total data acquisition costs across both bands.

REFERENCES

- [1] M. Alrabeiah and A. Alkhateeb, “Deep learning for mmWave beam and blockage prediction using sub-6 GHz channels,” *IEEE Trans. Commun.*, vol. 68, no. 9, pp. 5504–5518, Sep. 2020.
- [2] W. Deng, M. Li, Y. Liu, M.-M. Zhao, and M. Lei, “Enhancing mmWave beam prediction through deep learning with sub-6 GHz channel estimate,” in *Proc. IEEE Wireless Communications and Networking Conference (WCNC)*, Apr., 2024, pp. 1–6.
- [3] C.-X. Wang, J. Huang, H. Wang, X. Gao, X. You, and Y. Hao, “6G wireless channel measurements and models: Trends and challenges,” *IEEE Vehicular Technology Magazine*, vol. 15, no. 4, pp. 22–32, Dec. 2020.
- [4] W. Hong, Z. H. Jiang, C. Yu, D. Hou, H. Wang, C. Guo *et al.*, “The role of Millimeter-Wave technologies in 5G/6G wireless communications,” *IEEE Journal of Microwaves*, vol. 1, no. 1, pp. 101–122, Jan. 2021.
- [5] A. S. Ahmed, M. M. Hamdi, M. S. Abood, A. M. Khaleel, M. Fathy, and S. H. Khaleefah, “Channel estimation using LS and MMSE channel estimation techniques for MIMO-OFDM systems,” in *2022 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA)*, June, 2022, pp. 1–6.
- [6] M. Belgiovine, K. Sankhe, C. Bocanegra, D. Roy, and K. R. Chowdhury, “Deep learning at the edge for channel estimation in beyond-5G massive MIMO,” *IEEE Wireless Commun.*, vol. 28, no. 2, pp. 19–25, Apr. 2021.
- [7] Z. Liu, L. Zhang, and Z. Ding, “Overcoming the channel estimation barrier in massive mimo communication via deep learning,” *IEEE Wireless Commun.*, vol. 27, no. 5, pp. 104–111, Oct. 2020.
- [8] B. Lim, W. J. Yun, J. Kim, and Y.-C. Ko, “Joint pilot design and channel estimation using deep residual learning for multi-cell massive MIMO under hardware impairments,” *IEEE Trans. on Veh. Technol.*, vol. 71, no. 7, pp. 7599–7612, July 2022.
- [9] W. Alves, I. Correa, N. González-Prelcic, and A. Klautau, “Deep transfer learning for site-specific channel estimation in low-resolution mmWave MIMO,” *IEEE Wireless Commun. Lett.*, vol. 10, no. 7, pp. 1424–1428, July 2021.
- [10] S. J. Pan and Q. Yang, “A survey on transfer learning,” *IEEE Trans. on Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010.
- [11] K. Weiss, T. M. Khoshgoftaar, and D. Wang, “A survey of transfer learning,” *Springer Link*, vol. 3, no. 9, May 2016.
- [12] A. Hosna, E. Merry, J. Gyalmo, Z. Alom, Z. Aung, and M. A. Azim, “Transfer learning: a friendly introduction,” *Springer Link*, vol. 9, no. 102, Oct. 2022.