Limited and Compressed CSI Feedback System for Massive MISO Using Neural Network

Javad Ebrahimizadeh, Evgenii Vinogradov, *Senior Member, IEEE,* and Guy A. E. Vandenbosch, *Fellow, IEEE*

***Abstract*—This paper introduces a joint neural network approach to reduce the overhead in Channel State Information (CSI) feedback for massive Multiple-Input Single-Output (MISO) systems. The proposed method leverages a Feed-Forward Neural Network (FFNN) to enhance the estimation of the space–frequency channel using limited measurements. By exchanging a small subset of measured space–frequency channel elements, the joint neural network system enables accurate reconstruction of the full channel. It is shown that the joint FFNN can estimate a complex Gaussian channel with Rayleigh fading and rank $r = 1$ with an NMSE of less than 0.037 for 80% of the cases. For higher-rank channels ($r > 2$), the NMSE remains below 0.66 in 80% of instances. Additionally, for deterministic sparse channels with full rank, the joint FFNN achieves an NMSE below 0.1 in 80% of cases. The proposed FFNN method achieves a CSI compression rate of 50%.**

***Index Terms*—Channel state information, Feedforward neural network, Massive MIMO, Feedback system.**

I. Introduction

Beam management is essential in massive Multiple-Input Single-Output (MIMO) systems because these systems rely on efficient beamforming to exploit spatial diversity and maximize throughput. As the number of antennas at the Access Point (AP) increases, the beamforming capabilities improve, enabling the system to focus the transmission power toward the Use Equipment (UE) [1]. However, to properly align the beams between the AP and UE, precise Channel State Information (CSI) is required. The CSI feedback system plays a critical role in providing the necessary information about the channel characteristics, such as the angle of arrival, angle of departure, and path loss.

In Frequency Division Duplex (FDD) systems, CSI feedback is especially challenging because the uplink and downlink channels are not reciprocal, meaning that channel estimation on the uplink cannot directly be used for the downlink. This lack of reciprocity adds complexity to beam management, as accurate feedback must be sent from the UE to the AP in real-time, which can lead to overhead and increased latency. In OFDM systems, which are widely used for their robustness to multipath fading, CSI feedback is required for each subcarrier or subband. The number of measurements needed for full channel estimation can be large, further exacerbating the challenges of managing overhead and ensuring low-latency operation [2].

The acquisition and feedback of accurate CSI remains a fundamental challenge in modern MIMO systems, particularly as networks scale to massive antenna arrays and wideband operation. Traditional pilot-based channel estimation methods face significant spectral efficiency limitations due to growing overhead requirements [3]. Recent advances in deep learning have opened new possibilities for addressing these challenges through data-driven approaches. In [3], the authors demonstrated how neural networks can jointly optimize pilot design and channel estimation, employing innovative pruning techniques to reduce overhead while maintaining estimation accuracy. In [4], the authors introduced situation-aware learning for zone-specific CSI feedback, exploiting spatial channel correlations to improve feedback efficiency. The data demands of these learning-based approaches have been mitigated through digital twin frameworks [5] that generate synthetic training data from electromagnetic models. In [6], the authors proposed a radical reformulation of CSI representation through implicit neural functions, achieving unprecedented compression ratios. Together, these innovations highlight how machine learning is transforming fundamental aspects of MIMO system design, from channel acquisition to feedback compression, while addressing practical deployment challenges through techniques like digital twins and meta-learning.

In [7], the authors propsed using joint neural network as a CSI feedback system. They designed feedforward neural network to estimate the singular values of the channel with limited measurements. However, the estimation of singular values plays a role in optimizing the signal power, designing the system, and understanding the channel's rank and degrees of freedom. By analyzing these singular values, we gain insight into the system's capacity, the effectiveness of beamforming, and the Signal-to-Noise Ratio (SNR) achievable under different transmission conditions. It is still demanding to estimate the channel elements that are more general than its singular values.

In this paper, we propose a low-complexity neural network-based CSI feedback system and evaluate its performance over both statistical and deterministic channels. We demonstrate that for a full-rank complex Gaussian channel, the neural network can effectively estimate the channel. However, accurately estimating channels with a rank higher than one remains a challenge. For deterministic channels, the neural network performs well in estimating both the singular values

Corresponding author: Javad Ebrahimizadeh.

Guy A. E. Vandenbosch, Evgenii Vinogradov, and Javad Ebrahimizadeh are with the Department of Electrical Engineering, Division ESAT-WaveCoRE, Katholieke Universiteit Leuven, B-3001, Leuven, Belgium, e-mail: Javad.Ebrahimizadeh@esat.kuleuven.be

Evgenii Vinogradov is with NaNoNetworking Center in Catalunya, Universitat Politècnica de Catalunya Barcelona, Spain.

978-3-903176-77-5

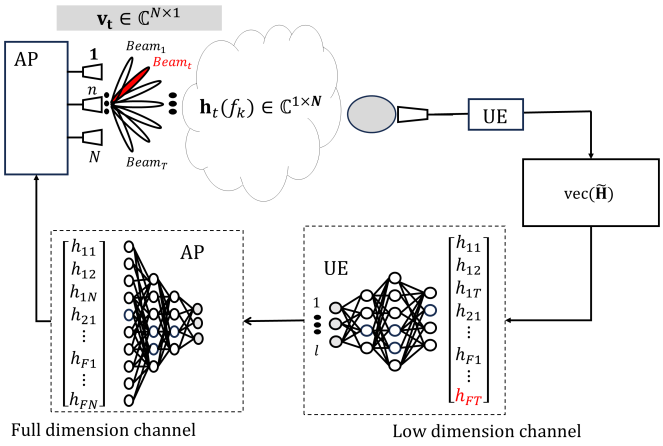


Fig. 1: The downlink beam training system model of the compressed CSI feedback system.

and the channel itself.

In this paper, a scalar x is denoted by non-bold lower-case letter and a bold lower case and upper case letter, say $\mathbf{x}$ and $\mathbf{X}$ represent a vector and a matrix, respectively. Conjugate, transpose, Hermitian transpose are denoted by $(\cdot)^*$, $(\cdot)^\top$, and $(\cdot)^\dagger$ respectively. The rest of the paper is organized as follows: Section II explains the system model, Section III and Section IV illustrate the numerical results and measurements, respectively. Finally, the paper concludes in Section V.

II. System model

In this section, we explain the signal model of beamforming in downlink wireless networks.

A. Signal model

This paper investigates downlink beam training between an AP and a UE. The AP equipped with N antennas performs beam sweeping over T time slots using a precoding vector $\mathbf{v}_t \in \mathbb{C}^{N\times 1}$ for $t \in [1,\dots,T]$. The UE is equipped with a single antenna. This setup is illustrated in Fig. 1. The channel at frequency f is represented as $\mathbf{h}(f) \in \mathbb{C}^{1\times N}$, where f belongs to the set $\{f_1,\dots,f_F\}$. The received signal in time slot t and frequency f, corresponding to the t-th beam of the AP, is given by the following expression, considering the presence of complex Gaussian noise n and a transmitted signal $s(f)$:

$$y_t(f) = \mathbf{h}(f)\mathbf{v}_t s(f) + n, \quad n \sim \mathcal{CN}(0,\sigma^2) \tag{1}$$

The total observation signal in the absence of noise is given by:

$$\mathbf{Y} = \begin{bmatrix} \mathbf{h}(f_1) \\ \vdots \\ \mathbf{h}(f_F) \end{bmatrix} \begin{bmatrix} \mathbf{v}_1, & \mathbf{v}_2, & \dots, & \mathbf{v}_T \end{bmatrix} = \tilde{\mathbf{H}}\mathbf{V} \tag{2}$$

where $\tilde{\mathbf{H}} \in \mathbb{C}^{F\times N}$ is the effective channel (consisiting of the channel and combiner) and $\mathbf{V} \in \mathbb{C}^{N\times T}$ is the precoder matrix.

III. Simulation results

This section evaluates the performance of the proposed CSI feedback system in three different channel scenarios: a complex Gaussian channel, a two-dimensional deterministic indoor L-shaped room channel, and a measured courtyard channel. The focus is on estimating the effective channel using a limited number of measurements.

Estimation of $\tilde{\mathbf{H}}$ from $\mathbf{Y}$ is generally impossible unless $\tilde{\mathbf{H}}$ has a restricted structure or $\mathbf{V}$ possesses specific properties. A neural network can successfully perform this estimation only under certain conditions. If $\tilde{\mathbf{H}}$ exhibits a structured form, such as being low-rank, sparse, or following a known distribution, the network can learn a meaningful mapping. For instance, if $\tilde{\mathbf{H}}$ is always rank-r with $r \leq m$, the neural network can learn to associate the m observable singular values of $\mathbf{Y}$ with the r singular values of $\tilde{\mathbf{H}}$. Moreover, if $\tilde{\mathbf{H}}$ follows a Gaussian ensemble, statistical properties such as the Marchenko-Pastur law can be exploited to estimate the singular value distribution more effectively.

Another important factor in successful estimation is the structure of $\mathbf{V}$. If $\mathbf{V}$ has a known and fixed structure, such as a partial Fourier matrix or a random projection with a well-defined distribution, the neural network can learn the compression pattern and reconstruct the channel $\tilde{\mathbf{H}}$ accordingly. Furthermore, in scenarios where only the dominant singular values of $\tilde{\mathbf{H}}$ are relevant, and their count does not exceed m, the network can be trained to recover only these significant values rather than attempting to reconstruct the entire channel.

In our study, $\mathbf{V}$ is a partial DFT matrix, and $\tilde{\mathbf{H}}$ is drawn from an i.i.d. Gaussian distribution, which allows the neural network to estimate the singular values efficiently. For the deterministic channel scenario, the estimation problem becomes more straightforward, as a deterministic relationship exists between the limited measured channel and the effective channel. This enables the neural network to reconstruct the full-rank channel with high accuracy. The architectures of the FFNNs used in each scenario are provided in Table I.

To quantify performance, we use the Normalized Mean Squared Error (NMSE), defined as :

$$\text{NMSE} \triangleq \frac{\|\mathbf{H} - \hat{\mathbf{H}}\|^2}{\|\mathbf{H}\|^2}, \tag{3}$$

where $\mathbf{H}$ and $\hat{\mathbf{H}}$ represent the exact and estimated channel matrices, respectively. The NMSE metric provides a reliable measure of how well the neural network recovers the channel information under different conditions.

A. Complex Gaussian Channel

To evaluate the performance of the proposed joint neural network for compressed CSI feedback, we consider a stochastic channel. The AP is equipped with eight discrete Fourier transform (DFT) beams, while the channel is measured using only four beams (beam indices: 2, 4, 6, and 8). Additionally, the channel is observed across eight orthogonal

TABLE I: Summary of FFNN Architectures for Different Channel Types used in simulation results

Parameter	Complex Gaussian Channel	2D Deterministic L-shape room Channel	Courtyard Channel
Hidden Layers	9 layers	8 layers	9 layers
Layer Sizes	64, 256, 128, 64, 32, 16, 68, 128, 256	64, 256, 128, 64, 32, 8, 68, 128	1024, 512, 128, 64, 128, 256, 512, 1024, 2048
Channel matrix size	8 by 8	16 by 16	32 by 32
Activation	ReLU	ReLU	ReLU
Bottleneck Layer	16 neurons	8 neurons	64 neurons
Optimizer	Adam	Adam	Adam
Loss Function	MSE	MSE	MSE
Batch Size	32	32	32
Epochs	50	50	50
Training Data	8,000 samples	1,948 samples	16,000 samples
Testing Data	2,000 samples	488 samples	4,000 samples
NMSE (80% level)	0.037 (rank 1)	0.10	0.032

frequency components. The performance of the CSI feedback system, implemented using a FFNN, is analyzed in terms of NMSE.

The proposed neural network is a fully connected deep feedforward model designed to reconstruct the channel matrix $\tilde{\mathbf{H}}$ from its compressed representation $\mathbf{Y}$. The network consists of an input layer that processes the real and imaginary components of $\mathbf{Y}$, followed by nine hidden layers with varying widths: 64, 256, 128, 64, 32, 16, 68, 128, and 256 neurons. Each hidden layer employs the ReLU activation function. The output layer reconstructs the real and imaginary components of $\tilde{\mathbf{H}}$ using a linear activation function.

The model is trained using the Adam optimizer with a Mean Squared Error (MSE) loss function, a batch size of 32, and 50 epochs. The dataset consists of 10,000 samples, with 80% allocated for training and 20% for testing. The depth and non-linearity of the network enable it to efficiently learn the inverse mapping from the compressed representation to the full channel matrix. The bottleneck layer, consisting of 16 neurons, ensures that the CSI feedback system transmits only 16 real values instead of 128, significantly reducing the overhead while utilizing limited channel measurements.

As shown in Fig. 2, the neural network converges after 50 training epochs. The Cumulative Distribution Function (CDF) of the NMSE error in channel estimation is illustrated in Fig. 3. It is observed that for a rank-1 channel, 80% of the NMSE values are below 0.037, whereas for a higher-rank channel, the NMSE increases to 0.662. To further compare the exact and estimated channel, a randomly selected example of the channel matrix elements is visualized in Fig. 4.

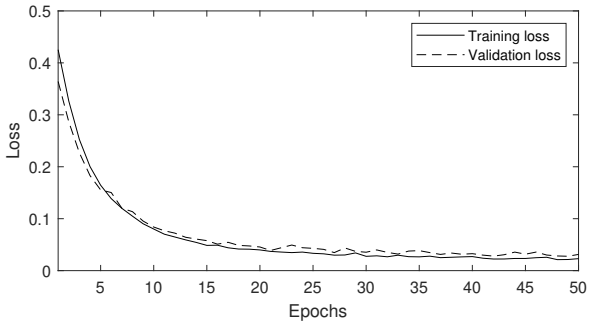


Fig. 2: The training and validation loss for FFNN.

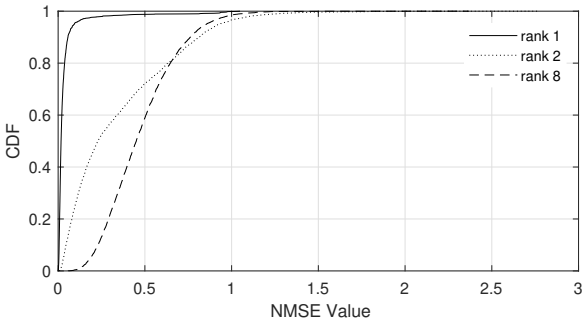


Fig. 3: NMSE error for 2000 random test data for different Gaussian cghannel matrix by considering the rank of the matrix.

B. 2D deterministic indoor channel

For the 2D indoor deterministic channel illustrated in Fig. 5, the AP is modeled as a single-antenna line source propagating within an L-shaped indoor environment. The scenario includes objects such as tables, chairs, and closets, all modeled as conductive materials. To maintain consistency with the system model in (2), the electric fields are sampled using a 16-element array, where the separation between adjacent elements is 10 cm. This sampling is performed over a frequency band of $BW \in [1, 2]$ GHz, consisting of $F = 201$ coherence subcarrier blocks.

To generate input data for training the neural network, we simulate 20 different scenarios, each corresponding to a distinct AP position (see Fig. 5). In each scenario, the AP remains fixed while the UE moves along a predefined trajectory in increments of 1 cm. This UE movement serves as a data augmentation technique. The channel is computed for each of the 20 AP positions using the commercial

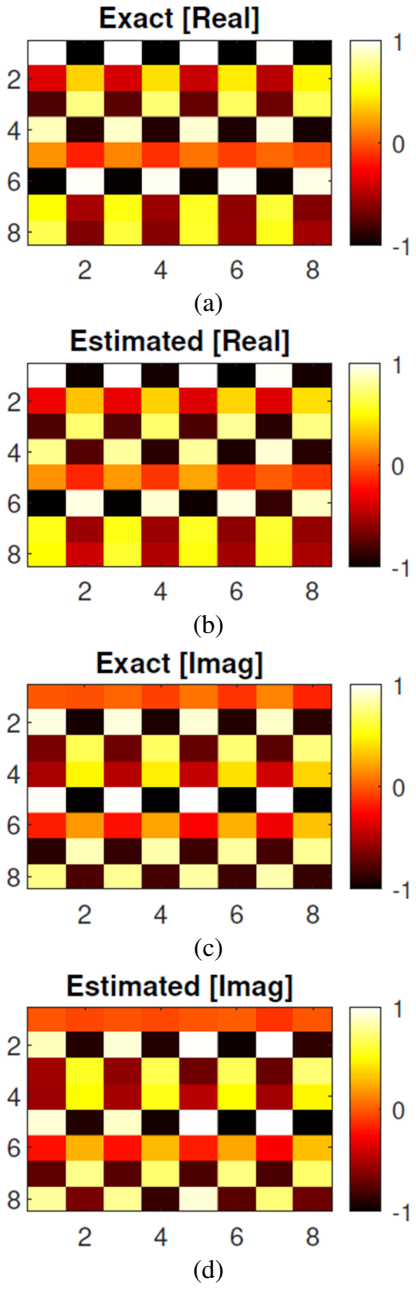


Fig. 4: An example random sample of the estimated channel after CSI feedback system for Gaussian complex channel with rank 1.

COMSOL software, leveraging the Method of Moments (MoM) technique. The simulation spans frequencies from 1 GHz to 2 GHz across 201 subcarrier frequencies [7]. In total, 13,600 channel samples are generated, obtained from 20 sampled AP positions and 680 sampled UE positions.

In the scenario of estimating the channel with limited measurements for an L-shaped room using an FFNN, the architecture of the FFNN follows a structure similar to that used for a Gaussian channel. The network begins with an input layer whose size depends on the input features, followed

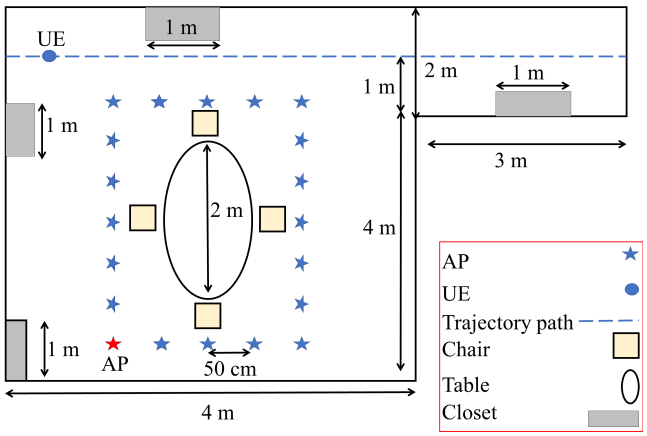


Fig. 5: Indoor propagation scenario with an L-shape room with table, chairs, and closets.

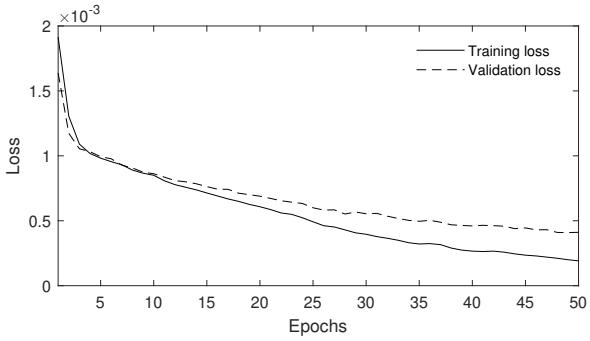


Fig. 6: The training and validation loss for FFNN for L-shape room channel.

by eight hidden layers with varying numbers of neurons: 64, 256, 128, 64, 32, 8, 68, and 128 neurons, respectively. These layers use the ReLU activation function, which introduces non-linearity to the model, allowing it to learn complex relationships between the input and the target output. The model architecture includes a bottleneck layer with eight neurons, which typically is critical for the feedback system to overcome overhead. The model is trained using the Adam optimizer with MSE loss, a batch size of 32, and 50 epochs, with 80% of the dataset used for training and 20% for testing on a dataset of 2436 samples.

As shown in Fig. 6, the NN converges after 100 epochs. The CDF of the NMSE is shown in Fig. 7. It is observed that 80% of the NMSE is less than 0.1. To compare the exact and estimated channel, an example random sample of the channel elements is shown in Fig. 8.

IV. Measurement

The scenarios include a brick facade with windows and doors, as depicted in Fig. 9, with dimensions listed in Table II. The TX antenna array is configured as a uniform rectangular planar array consisting of 64 dual-polarization patch antenna elements. The RX antenna array, on the other hand, is a circular conformal array comprising 128 dual-polarization

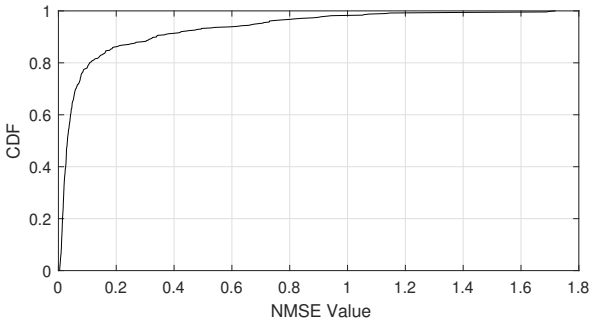


Fig. 7: NMSE error for 488 random test data for different L-shape room channel matrix.

patch antenna elements, as shown in Fig. 10. The hexagonal structure of the RX array facilitates signal reception across a full 360-degree field of view while also supporting DFT beams, enabling the formation of eight orthogonal beams [8]–[10].

The TX antenna array is positioned on the ground floor at a height of 1.7 meters and illuminates the wall with an incident angle of 50 degrees relative to the normal vector of the wall, as illustrated in Fig. 9. The RX is moved along a predefined trajectory with step increments of 10 cm. The measurements are conducted in stationary conditions, ensuring that the RX remains fixed at each position along the trajectory to mitigate Doppler effects. The trajectory spans a total distance of 10 meters, with measurements taken at 100 discrete RX positions.

To perform data augmentation, the channel between a single TX antenna and the RX array, consisting of 32 elements in a circular configuration, is utilized as the reference channel in (2). Consequently, the TX array, which comprises 64 patch antennas, and the RX array, arranged in four rows for elevation, can generate a total of 25,600 raw data samples across 100 RX positions along the trajectory path, serving as an effective data augmentation technique.

The total raw data used in this section is 20.000 and the test data size is random selection of 20% of the total data. The data are based on real and imaginary part of the channel matrix. The neural network has 11 hidden layers with varying numbers of neurons: 1024, 512, 128, 64, 128, 256, 512, 1024, 2048 neurons, respectively. It is observed that the FFNN converge after 20 Epochs as shown in Fig. 11. Also, the CDF of the NMSE error shown in Fig. 12 demonstrates that 80% of the NMSE is less than 0.032. The exact and estimated channels are shown in Fig. 13 demonstrating that the FFNN can estimate channels with high agreement.

V. Conclusion

The proposed joint neural network approach effectively reduces the overhead in CSI feedback for massive MISO systems by leveraging a feed-forward neural network for space-frequency channel estimation. The results demonstrate that the method achieves high accuracy in reconstructing complex Gaussian and deterministic sparse channels,

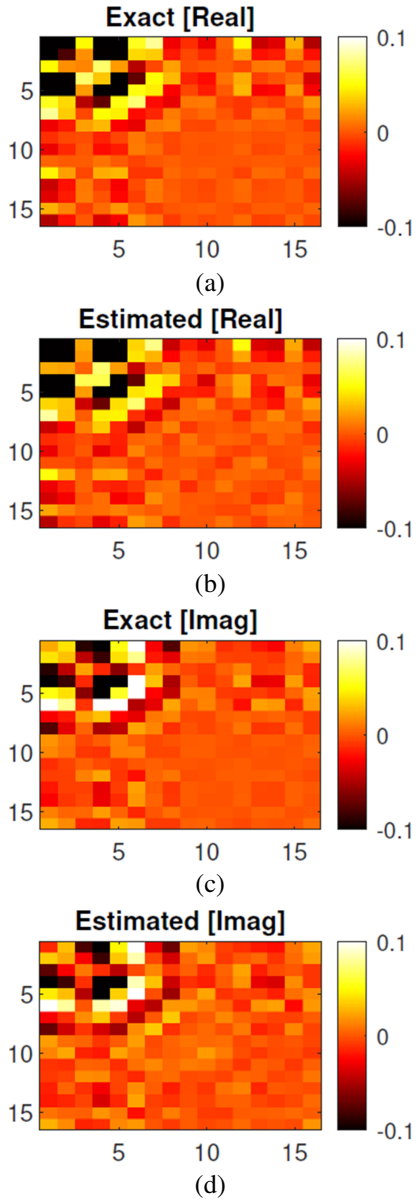
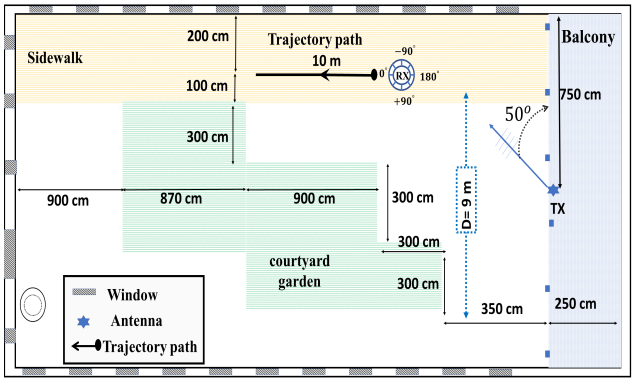


Fig. 8: An example random sample of estimated channel after CSI feedback system for L-shape room channel with energy normalized.

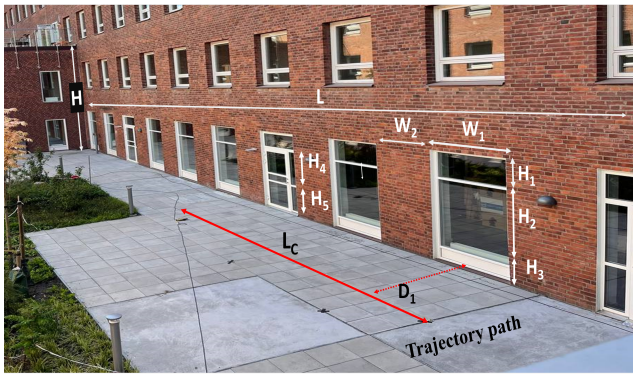
maintaining low NMSE values under various channel conditions. Furthermore, the achieved 50% CSI compression rate highlights the efficiency of the approach in reducing feedback requirements while preserving channel estimation quality. These findings, validated with measured data, suggest that the proposed method is a promising solution to improve CSI feedback in next-generation wireless communication systems.

Acknowledgments

The present work received funding from the European Union's Framework Programme for Research and Innovation Horizon 2020 under Grant Agreement No. 861222 (MINTS

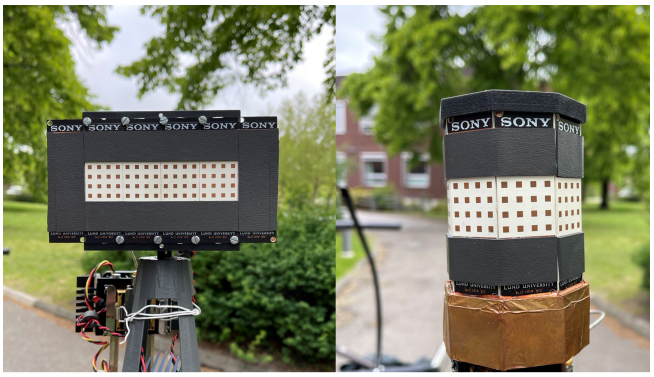


((a))



((b))

Fig. 9: The courtyard scenario includes a brick facade with windows and doors. (a) Top view (b) Perspective view [11].



(a) **(b)**

Fig. 10: The antenna array configuration. (a) TX antenna array. (b) RX antenna array.

project). We extend our sincere thanks to Prof. Fredrik Tufvesson, Hedieh Khosravi and Juan Sanchez for their invaluable assistance in the measurement of the Courtyard Scenario using LuMaMi28 Massive MIMO channel sounder at Lund University.

TABLE II

OVERVIEW OF THE COURTYARD MEASUREMENT SCENARIO.

Object	parameters	Simulation values
Frequency range	BW	27.5-29.5 GHz
Antenna	Type	Patch antenna
	TX height:	1.7 [m]
	RX height:	1.7 [m]
	Polarization	V
Windows Doors	Width : W_1	159 [cm]
	Height : H_1	60 [cm]
	Height : H_2	160 [cm]
	Height : H_3	20 [cm]
	Height : H_4	45 [cm]
	Height : H_5	80 [cm]
	Separation distance : W_2	145 [cm]
	Depth:	7 [cm]
Sidewalk	Length : L	37 [m]
	Width : D_1	3 [m]
Yard	Length : L	37 [m]
	Height : D	9 [m]
Wall	Length : L_{w}	37 [m]
	Thickness : D_{w}	40 [cm]
	Height: H	Infinite
	Relative permittivity**: $\epsilon_{\mathrm{r,w}}$	3.26
Ground	Type	Dry ground
	Relative permittivity ** : $\epsilon_{\mathrm{r,g}}$	6

**The relative permittivity is provided in [12]

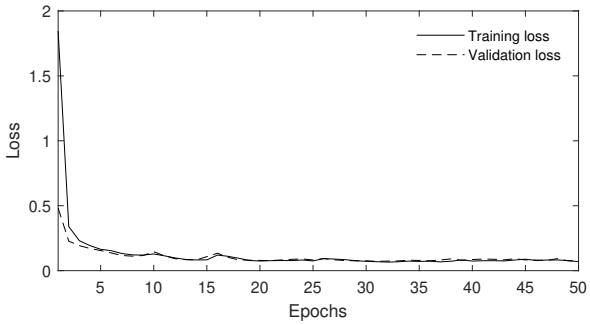


Fig. 11: The training and validation loss for FFNN for courtyard channel.

REFERENCES

[1] D. Kim, M. R. Castellanos, and R. W. Heath, "Joint band assignment and beam management using hierarchical reinforcement learning for multi-band communication," *IEEE Transactions on Vehicular Technology*, vol. 73, no. 9, pp. 13 451–13 465, Sept. 2024.

[2] M. B. Mashhadi, Q. Yang, and D. Gündüz, "Distributed deep convolutional compression for massive MIMO CSI feedback," *IEEE Transactions on Wireless Communications*, vol. 20, no. 4, pp. 2621–2633, 2020.

[3] M. B. Mashhadi and D. Gündüz, "Pruning the pilots: Deep learning-based pilot design and channel estimation for MIMO-OFDM systems," *IEEE Transactions on Wireless Communications*, vol. 20, no. 10, pp. 6315–6328, 2021.

[4] Y. Zhang and A. Alkhateeb, "Zone-specific CSI feedback for massive MIMO : A situation-aware deep learning approach," *IEEE Wireless Communications Letters*, vol. 13, no. 12, pp. 3320–3324, 2024.

[5] S. Jiang and A. Alkhateeb, "Digital twin aided massive MIMO: CSI compression and feedback," in *ICC 2024-IEEE International Conference on Communications*. IEEE, 2024, pp. 3586–3591.

[6] H. Wu, M. Zhang, Y. Shao, K. Mikolajczyk, and D. Gündüz, "MIMO channel as a neural function: Implicit neural representations for extreme

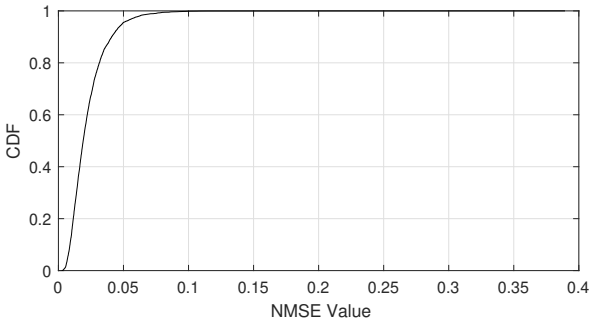


Fig. 12: NMSE error for 4000 random test data for different courtyard channel matrix.

CSI compression," in *ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2025, pp. 1–5.

[7] J. Ebrahimizadeh, F. Rusek, E. Bengtsson, J. Flordelis, E. Vinogradov, and G. A. E. Vandenbosch, "Limited CSI feedback system for massive MISO: A neural network-aided SVD approach," *IEEE Transactions on Antennas and Propagation*, vol. 73, no. 7, pp. 4891–4902, 2025.

[8] X. Cai, M. Zhu, A. Fedorov, and F. Tufvesson, "Enhanced effective aperture distribution function for characterizing large-scale antenna arrays," *IEEE Transactions on Antennas and Propagation*, vol. 71, no. 8, pp. 6869–6877, 2023.

[9] A. Al-Ameri, J. Park, J. Sanchez, X. Cai, and F. Tufvesson, "A hybrid antenna switching scheme for dynamic channel sounding," in *2023 IEEE 97th Vehicular Technology Conference (VTC2023-Spring)*. IEEE, 2023, pp. 1–6.

[10] X. Cai, E. L. Bengtsson, O. Edfors, and F. Tufvesson, "A switched array sounder for dynamic millimeter-wave channel characterization: Design, implementation and measurements," *IEEE Transactions on Antennas and Propagation*, vol. 72, no. 7, pp. 5985–5999, 2024.

[11] J. Ebrahimizadeh, V. Khorashadi-zadeh, X. Cai, F. Tufvesson, and G. A. Vandenbosch, "Millimeter-wave scattering from building facade: A simulation and verification study," in *2024 18th European Conference on Antennas and Propagation (EuCAP)*. IEEE, 2024, pp. 1–5.

[12] Y. Pinhasi, A. Yahalom, and S. Petnev, "Propagation of ultra wide-band signals in lossy dispersive media," in *2008 IEEE International Conference on Microwaves, Communications, Antennas and Electronic Systems*. IEEE, 2008, pp. 1–10.

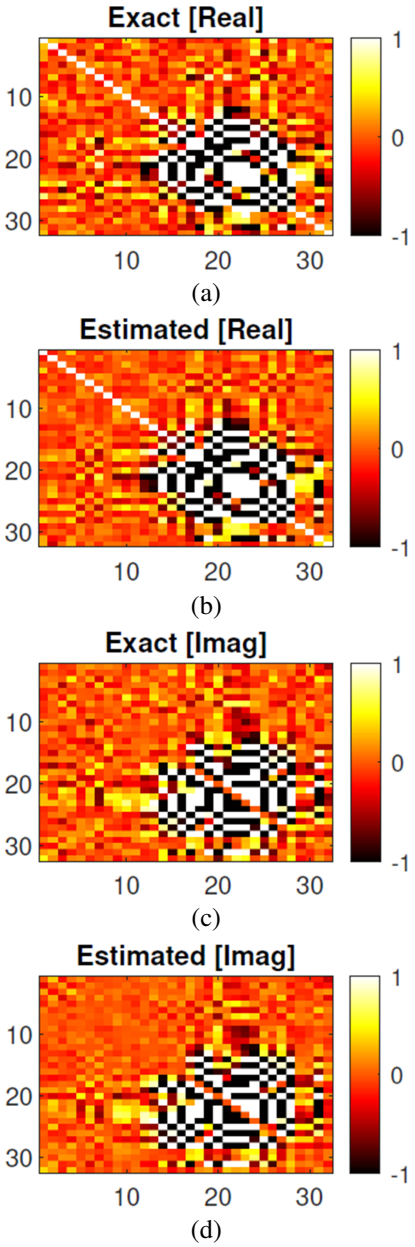


Fig. 13: An example random sample of estimated channel after CSI feedback system for measured courtyard channel with energy normalized.