# Complex-Valued Differential Autoencoder for Temporal CSI Compression

**Manoj Kumar CM [1], Lakshmi N. Theagarajan[1,2]**

[1] Department of Data Science and AI, Indian Institute of Technology Madras
[2]Department of Electrical Engineering, Indian Institute of Technology Madras
Chennai, India
manoj.kumar@dsai.iitm.ac.in, lnt@ee.iitm.ac.in

## Abstract

This paper presents a complex-valued differential autoencoder architecture for efficient compression and reconstruction of channel state information (CSI) in cellular systems. This minimizes the substantial feedback overhead associated with high-dimensional CSI in massive multi-antenna multiple-input multiple-output (MIMO) and broadband orthogonal frequency division modulation (OFDM). The CSI is represented in the delay–angle domain, where it exhibits sparse and structured correlations across both spatial and temporal dimensions. Unlike conventional feedback schemes that periodically transmit full CSI instances from the user equipment (UE) to the base station (BS), the proposed model leverages temporal correlation between consecutive CSI frames to reduce feedback bandwidth. Instead of traditional models that split the real and imaginary values of the CSI into separate channels, the proposed architecture introduces a feedback driven mechanism that encodes and reconstructs only the information difference between CSI instances using complex-valued transformations. Experiments are conducted on temporally correlated delay–angle CSI sequences, generated from standard 3GPP channel models at user speeds of 40 km/h, 100 km/h, and 360 km/h to emulate varying Doppler conditions. The experimental results demonstrate that the proposed model achieves substantially higher compression efficiency and reconstruction fidelity than conventional frame-wise autoencoders.

## Introduction

The accurate and timely acquisition of channel state information (CSI) is critical for achieving high spectral efficiency and reliable communication in modern cellular systems, particularly in massive MIMO and broadband OFDM scenarios. However, the high dimensionality of CSI in these systems introduces significant feedback overhead between the user equipment (UE) and the base station (BS). To mitigate this overhead, efficient CSI compression has become essential. Transforming CSI into the delay–angle domain via a 2D discrete Fourier transform reveals that most of the channel energy is concentrated in a small subset of coefficients (Sayeed 2002). This enables compact and structured representations that are more amenable to learning based compression.



Recent deep learning-based architectures, such as CsiNet (Wen, Shih, and Jin 2018) and CRNet (Lu, Wang, and Song 2020), have shown remarkable success in exploiting spatial correlations of CSI for compression and reconstruction. CsiNet employs convolutional encoders and decoders to learn nonlinear mappings between high dimensional CSI and its compressed representation. On the other hand, CRNet extends this approach using residual blocks to enhance reconstruction accuracy. Despite their effectiveness, these architectures treat each CSI frame independently compressing the entire CSI instance at every reporting interval without leveraging the temporal correlations that naturally exist between successive frames in time varying channels. Moreover, these models typically decompose complex valued CSI into separate real and imaginary components before processing. Thus, the correlation between the real and imaginary components are not efficiently exploited in reducing redundancy.

In contrast, this paper proposes a complex-valued differential autoencoder (CVDAE) designed to exploit the temporal redundancies present in CSI sequences. In practical systems, the UE periodically transmits CSI feedback to the BS. Instead of feeding back the full CSI data at each instant, our model focuses on encoding only the difference between consecutive CSI frames, which inherently contains less information as the CSI frames are temporally correlated. The inter-frame correlation is inversely proportional to the Doppler content in the CSI.

The proposed CVDAE architecture consists of three key components:

- *Complex-valued autoencoder (CVAE)*: This learns a compact latent representation of the CSI.
- *Difference network*: a complex valued linear network that computes the residual between the current CSI matrix and the previous CSI instance.
- *Summation network*: a complex valued linear network that reconstructs the current CSI matrix by combining the decoded residual with the previously reconstructed CSI instance.

To enable feedback-driven temporal reconstruction, memory units are maintained at both the encoder and decoder. The encoder memory stores the previously observed CSI instance, while the decoder memory holds the previously reconstructed CSI. This design allows the model to learn how CSI evolves over time and to adaptively refine its representation across frames. Such memory units as opposed to GRU memory units reduce the computational complexity of the model.

Experimental results show that the proposed framework effectively exploits the complex-valued structure and temporal continuity of CSI, achieving superior compression efficiency and reconstruction fidelity compared to conventional frame-wise autoencoders across varying doppler conditions.

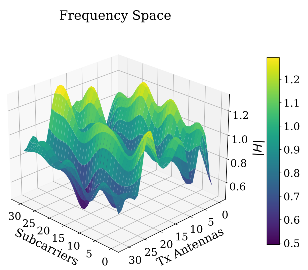


Figure 1: Sample CSI generated using NVIDIA Sionna.

## Channel Model

In this work, we consider the downlink channel of a frequency division duplex (FDD) system. The user equipment (UE) estimates the channel state (CSI) based on pilot symbols transmitted from the base station (BS), which is equipped with M antennas. The downlink employs orthogonal frequency division multiplexing (OFDM) with $N$ subcarriers. Thus, the baseband signal received at the receiver in the frequency domain is given by

$$y[f] = \mathbf{H}[f,s]^T \mathbf{x}[f,s] + z[f], \tag{1}$$

where $\mathbf{H}(f,s)$ is the wireless channel gain for the $f$th subcarrier and $s$th spatial antenna, $\mathbf{x}(f,s)$ is the transmitted signal and $z(f)$ is the additive noise.

The resulting CSI matrix $\mathbf{H}$ is essential for transmit precoding at the BS. In practical 4G and 5G systems, the UE estimates $\mathbf{H}$ using reference signals and reports it to the BS through the uplink control channel. However, due to time varying channel, the CSI remains approximately constant only over a short interval known as the *coherence time*, denoted by $T_c$. This interval is inversely related to the *Doppler spread* $f_D$, i.e., $T_c \approx \frac{1}{f_D}$. For a mobile user moving at velocity $v$, the Doppler spread is given by $f_D = \frac{v f_c}{c}$, where $f_c$ is the carrier frequency and $c$ is the speed of light. Consequently, higher mobility or carrier frequency results in faster CSI variation, requiring more frequent feedback. To reduce this overhead, the CSI is typically compressed before being fed back. Approximately, for each coherence window, a CSI matrix feedback is transmitted from the UE to the BS.

It is well established that the wireless channel exhibits a sparse structure in the *delay-angle domain* compared to the *frequency-space domain* representation (Sayeed 2002). The transformation between these two representations can be expressed as

$$G[n,m] = \frac{1}{\sqrt{MN}} \sum_{n=1}^{N} \sum_{m=1}^{M} H[n,m] e^{j2\pi\left(\frac{ml}{M} - \frac{nk}{N}\right)}, \tag{2}$$

where $H[n,m]$ and $G[n,m]$ denote the space-frequency and delay-angle CSI, respectively. An illustration of the CSI in these two domains is presented in Fig. 1. It can be seen from this illustration that the CSI is sparser in the delay-angle domain.

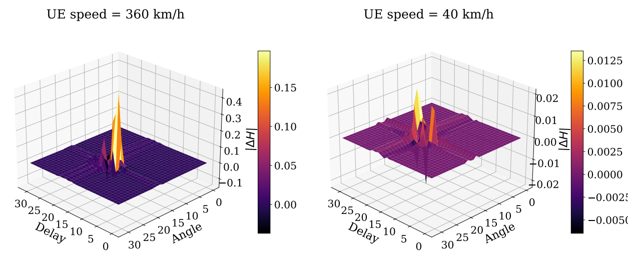


Figure 2: $|\Delta H|$ between consecutive CSI frames at different UE Speeds.

The temporal correlation between successive CSI frames at varying UE speeds is depicted in Figure 2. Here, $|\Delta H|$ denotes the magnitude of the difference between two consecutive CSI matrices over time. Figure 2 illustrates that the inter-frame correlation is inversely proportional to the Doppler content in the CSI.

## Deep Learning Architectures for CSI Compression

This section describes the state-of-the-art deep learning architectures (namely, CsiNet (Wen, Shih, and Jin 2018) and CRNet (Lu, Wang, and Song 2020)) and our proposed complex-valued differential autoencoder (CVDAE) for CSI compression. For comparison, we also present the complex-valued autoencoder (CVAE), which is a vanilla autoencoder architecture that does not exploit the temporal CSI correlations but learns complex-valued features. This allows us to isolate and quantify the contribution of the differential mechanism within CVDAE, providing a clearer understanding of how temporal correlations in CSI sequences can be effectively exploited.

### State-of-the-art

Convolutional neural network (CNN) based architectures have been widely adopted for CSI compression owing to their strong ability to learn spatial correlations in the channel matrix. A key work in this direction is CsiNet (Wen, Shih, and Jin 2018), which formulates CSI feedback as an image compression problem. The network employs a convolutional encoder at the UE to extract low-dimensional latent representations of the channel, and a corresponding decoder at the BS to reconstruct the CSI. CsiNet demonstrated significant performance gains over traditional compressive sensing approaches by leveraging data-driven learning of channel structures.

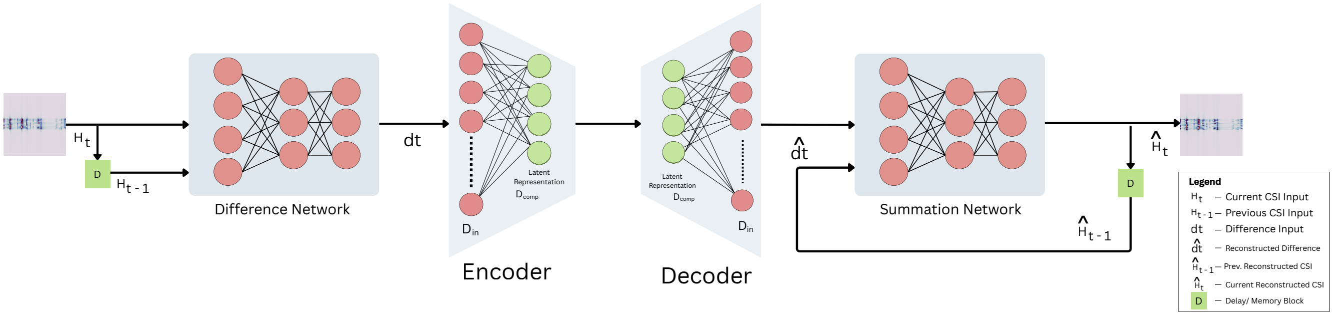


Figure 3: Complex Valued Differential Autoencoder Architecture.

Building upon CsiNet, another model named as CRNet (Lu, Wang, and Song 2020) further improved CSI reconstruction quality by introducing residual connections and a channel attention mechanism. CRNet also employs deeper convolutional layers to model non-linear channel relationships more effectively, achieving higher reconstruction accuracy at low compression ratios. These CNN-based models have established a strong baseline for CSI compression by exploiting spatial correlation and domain sparsity.

However, both CsiNet and CRNet share several limitations. First, they operate on the real-valued decomposition of the complex CSI, where the real and imaginary components are treated as independent channels. This decoupled representation discards the intrinsic amplitude–phase relationships inherent in the complex domain, potentially leading to suboptimal feature learning. Second, these models compress each CSI matrix independently, neglecting the strong temporal correlation present in sequential CSI observations when the UE periodically feeds back CSI to the BS. As a result, redundant information across consecutive matrices is repeatedly encoded, increasing feedback overhead. These limitations motivate the design of a complex-valued architecture that can jointly exploit amplitude–phase coupling and temporal redundancy for more efficient CSI compression.

## Complex Valued Feedforward Neural Network for CSI Compression

Existing architectures such as CsiNet (Wen, Shih, and Jin 2018) and CRNet (Lu, Wang, and Song 2020) treat CSI as real valued data and compress each frame independently. The proposed model directly operates on complex values and exploits temporal correlations between consecutive CSI instances. In practical FDD systems, the UE periodically feeds back CSI to the BS, and consecutive CSI samples exhibit strong temporal similarity within the channel coherence time. Motivated by this, we encode only the incremental difference between successive CSI frames instead of transmitting the full CSI at every time instant, thereby reducing feedback redundancy. The proposed complex-valued differential autoencoder (CVDAE) comprises of three modules: a complex-valued autoencoder (CVAE), a difference network, and a summation network. Each module is implemented using fully connected layers (Goodfellow, Bengio, and Courville 2016) with complex weights and activations to preserve both amplitude and phase dependencies in the CSI matrices.

**Complex-Valued Autoencoder (CVAE)** : The CVAE forms the central compression module of the framework. It consists of a complex-valued encoder and decoder, each implemented as the following transformation:

$$\mathbf{z}_t = f(\mathbf{W}_e \mathbf{x}_t + \mathbf{b}_e), \quad \hat{\mathbf{x}}_t = \mathbf{W}_d \mathbf{z}_t + \mathbf{b}_d, \tag{3}$$

where $\mathbf{x}_t \in \mathbb{C}^{1024}$ denotes the input CSI vector at time $t$ obtained by flattening the CSI matrix $\mathbf{H}_t$, $\mathbf{z}_t \in \mathbb{C}^h$ is the encoded latent vector of dimension $h$, and $f(\cdot)$ is a non-linear activation function. In this work, we adopt the *complex-valued hyperbolic tangent* (tanh) as the activation, which has been shown to stabilize magnitude growth and maintain phase continuity in complex-valued transformations.

$$\tanh(a + ib) = \frac{\sinh(2a) + i\sin(2b)}{\cosh(2a) + \cos(2b)}. \tag{4}$$

The encoder reduces the CSI dimensionality, achieving compression, while the decoder reconstructs the full-scale CSI from the latent embedding. All operations are performed using `complex64` precision using pytorch (Paszke et al. 2019), ensuring consistent complex arithmetic throughout the network.

**Difference Network** : The difference network computes the complex residual between the current CSI instance $\mathbf{H}_t$ and the previous value $\mathbf{H}_{t-1}$ in the memory. Thus, the difference network computes only those components that are not present in the previous CSI matrix.

$$\mathbf{d}_t = \text{DiffNet}(\mathbf{H}_t, \mathbf{H}_{t-1}). \tag{5}$$

The difference network comprises of two fully connected layers with complex-valued weights. The input is the concatenation of $\mathbf{H}_t$ and $\mathbf{H}_{t-1}$, forming a $2d$-dimensional vector. The difference network projects it to $d$ dimensions through a learnable transformation. Unlike conventional subtraction, this learned differencing allows the network to model complex-valued temporal relationships adaptively, capturing both amplitude and phase shifts that arise from Doppler induced variations.

**Summation Network** : The summation network performs the complementary operation at the decoder side, reconstructing the full CSI $\hat{\mathbf{H}}_t$ from the decoded residual $\hat{\mathbf{d}}_t$ and the previous reconstructed CSI $\hat{\mathbf{H}}_{t-1}$:

$$\hat{\mathbf{H}}_t = \text{SumNet}(\hat{\mathbf{d}}_t, \hat{\mathbf{H}}_{t-1}). \tag{6}$$

It has an identical structure to the difference network, employing two complex-valued linear layers. By learning to merge the reconstructed residual with the decoder memory, the summation network ensures smooth temporal evolution and continuity across CSI frames. This adaptive cumulation mechanism generalizes beyond element wise summation, allowing the network to learn optimal phase alignment and amplitude scaling between consecutive reconstructions.

**Memory Mechanism** : Two memory buffers are maintained across time steps. One at the encoder to store the previous CSI input $\mathbf{H}_{t-1}$, and another at the decoder to store the previously reconstructed CSI $\hat{\mathbf{H}}_{t-1}$. These memories are dynamically updated at each iteration, creating a feedback loop. During inference, the first CSI frame is encoded and reconstructed directly, while subsequent frames are processed through the differential pathway.

The entire network is trained to minimize the mean squared error (MSE) loss between the reconstructed CSI $\hat{\mathbf{H}}_t$ and the ground truth CSI $\mathbf{H}_t$:

$$\mathcal{L}_{\text{MSE}} = \frac{1}{L}\sum_{t=1}^{L}\left\|\mathbf{H}_t - \hat{\mathbf{H}}_t\right\|_2^2. \tag{7}$$

where $L$ denotes the total number of training samples. This objective encourages the network to learn compact yet faithful representations of the CSI data. Overall, this complex-valued formulation preserves the physical characteristics of the CSI while learning efficient representations of its temporal evolution. This leads to improved compression efficiency and reconstruction accuracy across different Doppler conditions.

## Experimental Results

To evaluate the proposed complex-valued differential autoencoder, we conduct experiments on temporally correlated CSI sequences generated using NVIDIA Sionna (Hoydis et al. 2022) in compliance with 3GPP channel models. The proposed model is benchmarked against the complex valued autoencoder, CRNet and CsiNet baselines across compression rates ranging from 50% (low compression) to 99% (high compression). The neural networks were trained using a batch size of 64 and for a maximum of 20 epochs.

### CSI Dataset

For training, CSI data is generated using the clustered delay line (CDL) channel models implemented in NVIDIA Sionna (Hoydis et al. 2022). The dataset adheres to 3GPP specifications, encompassing diverse CDL profiles, delay spreads, angular spreads, and path losses, simulated for a 32-antenna BS. To emulate urban macrocell environments and assess the model's robustness under mobility, the dataset is generated at varying Doppler frequencies corresponding to UE speeds of 40 km/h, 100 km/h, and 360 km/h. Each CSI instance is sampled at an interval of 34 $\mu$s, and 64 such consecutive instances are captured to form a complete CSI sample. An example of the generated CSI instance is shown in Fig. 1.

### Performance Analysis

To quantitatively evaluate the reconstruction performance, we consider the following two metrics: (i) *Average Cosine Similarity* and (ii) *Normalized Mean Squared Error (NMSE)*.

**Average Cosine Similarity** : To assess the structural alignment between the reconstructed and ground-truth CSI, we compute the cosine similarity along the columns of each CSI matrix and report their average value. Formally, it is defined as

$$\rho = \mathbf{E}\left\{\frac{1}{L}\sum_{n=1}^{L}\frac{|\hat{\mathbf{h}}_n^H\mathbf{h}_n|}{\|\hat{\mathbf{h}}_n\|_2\,\|\mathbf{h}_n\|_2}\right\}, \tag{8}$$

where $\mathbf{h}_n$ and $\hat{\mathbf{h}}_n$ denote the $n$-th column of the ground-truth and reconstructed CSI matrices, respectively. Higher values of cosine similarity denote better performance.

**Normalized Mean Squared Error** : The NMSE measures the reconstruction distortion normalized by the signal power of the ground-truth CSI, and is defined as

$$\text{NMSE} = \mathbf{E}\left\{\frac{\|\mathbf{H}-\hat{\mathbf{H}}\|_2^2}{\|\mathbf{H}\|_2^2}\right\}. \tag{9}$$

For evaluation, NMSE values are expressed in decibels (dB).

$$\text{NMSE}_{\text{dB}} = 10\log_{10}(\text{NMSE}). \tag{10}$$

Lower NMSE values (in dB) indicate better reconstruction fidelity.

| Rate (%) | CVAE | CRNet | CsiNet | CVDAE |
|---|---|---|---|---|
| 50 | 4.196M | 7.571M | 6.005M | 29.360M |
| 75 | 2.097M | 5.474M | 3.908M | 27.264M |
| 80 | 1.671M | 5.052M | 3.490M | 26.836M |
| 90 | 0.835M | 4.213M | 2.646M | 26.000M |
| 95 | 0.409M | 3.795M | 2.228M | 25.576M |
| 99 | 0.081M | 3.459M | 1.892M | 25.248M |

Table 1: Computation FLOPs at Different Compression Rates

### Observation

The results of the experiments are listed in Table 2, and Figures 4, 5, and 6. It can be observed that the proposed CVDAE consistently outperforms all other state-of-the-art models across varying UE speeds and compression ratios. As the UE speed decreases from 360 km/h to 40 km/h, the cosine similarity consistently improves across all compression rates, indicating enhanced reconstruction stability under slower mobility conditions. CVDAE exhibits a noticeable improvement in NMSE approximately 1 dB at 90%

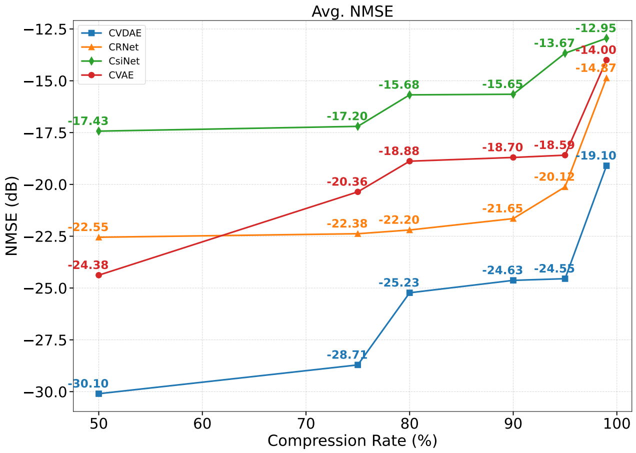


Figure 4: Average NMSE (dB) at a UE speed of 360 km/h.

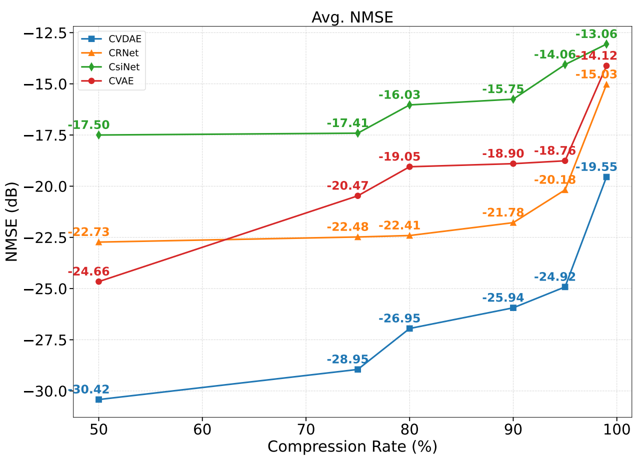


Figure 5: Average NMSE (dB) at a UE speed of 100 km/h.

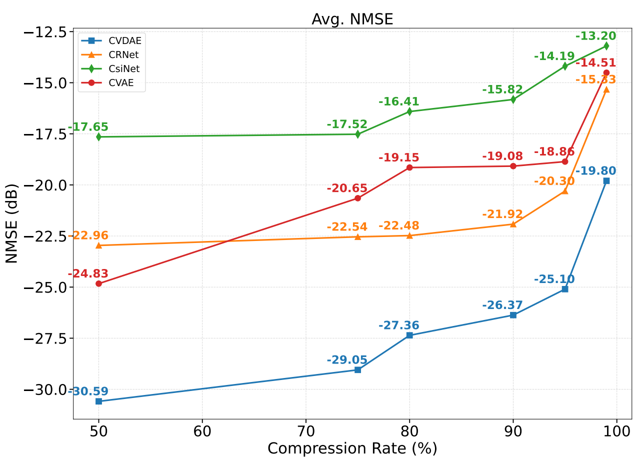


Figure 6: Average NMSE (dB) at a UE speed of 40 km/h.

compression demonstrating its robustness under lower mobility due to increased temporal correlation in the channel.

| Rate (%) | CVDAE | CRNet | CsiNet | CVAE |
|---|---|---|---|---|
| **UE Speed: 360 km/h** | | | | |
| 50 | 0.9997 | 0.9972 | 0.9918 | 0.9987 |
| 75 | 0.9995 | 0.9968 | 0.9893 | 0.9970 |
| 80 | 0.9986 | 0.9966 | 0.9881 | 0.9950 |
| 90 | 0.9985 | 0.9964 | 0.9852 | 0.9941 |
| 95 | 0.9985 | 0.9928 | 0.9762 | 0.9938 |
| 99 | 0.9825 | 0.9738 | 0.9681 | 0.9826 |
| **UE Speed: 100 km/h** | | | | |
| 50 | 0.9998 | 0.9973 | 0.9920 | 0.9988 |
| 75 | 0.9996 | 0.9969 | 0.9898 | 0.9971 |
| 80 | 0.9991 | 0.9967 | 0.9884 | 0.9953 |
| 90 | 0.9990 | 0.9966 | 0.9852 | 0.9945 |
| 95 | 0.9989 | 0.9929 | 0.9784 | 0.9940 |
| 99 | 0.9830 | 0.9745 | 0.9694 | 0.9831 |
| **UE Speed: 40 km/h** | | | | |
| 50 | 0.9998 | 0.9978 | 0.9930 | 0.9989 |
| 75 | 0.9997 | 0.9972 | 0.9905 | 0.9972 |
| 80 | 0.9994 | 0.9970 | 0.9884 | 0.9954 |
| 90 | 0.9992 | 0.9969 | 0.9853 | 0.9946 |
| 95 | 0.9990 | 0.9945 | 0.9794 | 0.9949 |
| 99 | 0.9850 | 0.9768 | 0.9700 | 0.9834 |

Table 2: Average Cosine Similarity ($\rho$) at Different Compression Rates and UE Speeds

It is observed that at higher compression rates ($\geq 95\%$), the complex-valued autoencoder (CVAE) outperforms CRNet in terms of cosine similarity and achieves comparable NMSE, while requiring nearly $9\times$ fewer FLOPs. It highlights the efficiency of complex valued representations, which preserve the inherent amplitude–phase relationships of the CSI without resorting to separate real and imaginary processing.

The proposed CVDAE further improves reconstruction quality by approximately 6 dB across all compression rates compared to CVAE. The modest increase in computational cost (table 1) arises from the difference and summation sub networks each comprising two fully connected complex valued linear layers that process concatenated CSI vectors to capture temporal dependencies. At 95% compression rate, CVDAE provides the same or better NMSE and cosine similarity performance compared to that of other models at 50% compression rate.

## Conclusion

This work introduced a novel *complex-valued differential autoencoder* (CVDAE) for compressing temporally correlated CSI in FDD systems. By learning differential representations in the complex domain, CVDAE effectively preserves amplitude–phase relationships and temporal consistency. It was showed that the proposed CVDAE model achieves superior reconstruction accuracy, particularly at high compression rates. Although the complex-valued linear layers incur a moderately higher computational footprint, this is well compensated by drastic performance gains both in terms of cosine similarity and NMSE. The experiments demonstrated the efficacy of proposed CVDAE model performance in wireless CSI compression at different Doppler frequencies. As future work, we plan to extend this framework to complex-valued convolutional and graph neural network architectures, which can exploit the spatial and structural correlations in CSI more efficiently.

## References


Goodfellow, I.; Bengio, Y.; and Courville, A. 2016. *Deep Learning*. MIT Press. http://www.deeplearningbook.org.

Hoydis, J.; Cammerer, S.; Ait Aoudia, F.; Nimier-David, M.; Maggi, L.; Marcus, G.; Vem, A.; and Keller, A. 2022. Sionna. Https://nvlabs.github.io/sionna/.

Lu, Z.; Wang, J.; and Song, J. 2020. Multi-resolution CSI feedback with deep learning in massive MIMO system. In *ICC 2020-2020 IEEE international conference on communications (ICC)*, 1–6. IEEE.

Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; et al. 2019. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32.

Sayeed, A. M. 2002. Deconstructing multiantenna fading channels. *IEEE Transactions on Signal processing*, 50(10): 2563–2579.

Wen, C.-K.; Shih, W.-T.; and Jin, S. 2018. Deep Learning for Massive MIMO CSI Feedback. *IEEE Wireless Communications Letters*, 7(5): 748–751.