

Comparative Analysis of CSI Feedback Transmission with Unequal Error Protection: DMD vs. TransNet

1st Fayad Haddad, 2nd Carsten Bockelmann, 3rd Armin Dekorsy

Department of Communications Engineering

University of Bremen, Germany

Email: {haddad, bockelmann, dekorsy}@ant.uni-bremen.de

Abstract—In wireless communication systems, accurate Channel State Information (CSI) is essential for base stations to perform downlink precoding. While much of the existing research primarily focus on compressing the CSI matrix, they often neglect the impact of subsequent pre-transmission processes such as quantization, channel coding, and modulation. This paper investigates two distinct approaches for CSI dimensionality reduction: TransNet, a transformer-based neural network, and Dynamic Mode Decomposition (DMD), a mathematical decomposition technique for dynamical systems. We analyze how quantization, channel coding, and modulation affect CSI feedback for both methods. Unlike TransNet, DMD can decompose the channel matrix into components (called modes) with varying significance. This decomposition allows for an effective application of Unequal Error Protection (UEP) techniques to DMD modes, which is not feasible with TransNet-based CSI. Simulation results reveal that while the compression performance of TransNet and DMD varies based on factors like target CSI size and channel estimation error, integrating UEP techniques for DMD-based CSI yields superior CSI transmission performance compared to TransNet-based CSI.

Index Terms—CSI feedback, Quantization, Polar codes, Modulation, NOMA, Dynamic mode decomposition, Transformers.

I. INTRODUCTION

Accurate downlink Channel State Information (CSI) is crucial in wireless communication systems for enabling precoding at the Base Station (BS). Typically, the channel matrix is estimated by the User Equipment (UE) and then transmitted back to the BS. Due to the large size of the channel matrix, efficient compression is necessary before transmission. To address this challenge, various compression techniques have been developed to reduce the dimensionality of CSI feedback. These techniques include Compressive Sensing (CS), matrix decomposition, and Deep Learning (DL) methods. Traditional CS methods exploit the assumed sparsity of the Channel Impulse Response (CIR) in the time domain, utilizing sparse recovery techniques such as Orthogonal Matching Pursuit (OMP) [1]. Matrix decomposition techniques, like Dynamic Mode Decomposition (DMD) [2], leverage temporal correlations in the channel, providing robust dimensionality reduction and predicting future channel states to reduce the frequency of CSI updates [3]. DL-based frameworks, such as TransNet [4], use the most recent neural network designs, namely transformers, to compress the channel matrix.

This work was funded by the German Ministry of Education and Research (BMBF) under grant 16KISK016 (Open6GHub).

Despite the extensive research on CSI compression, there is often insufficient consideration of the subsequent processes required before transmission. After compression, CSI feedback undergoes quantization, channel coding, and modulation, each of which affects CSI quality. Quantization converts continuous CSI values into discrete levels, where more quantization levels can improve accuracy but increase the data size. Channel coding adds redundancy to protect against transmission errors, though this raises overhead. Modulation maps the quantized and coded CSI data to complex symbols for transmission over the wireless channel, where higher modulation orders can improve the transmission bit rate but may also increase the bit error rate. Therefore, optimizing each process is critical to balancing the trade-off between maintaining high CSI accuracy and managing overhead and transmission bit rate.

This paper provides an extensive evaluation of CSI compression methods that accounts for the entire feedback procedure, including quantization, channel coding, modulation, and transmission. We namely focus on analyzing the transmission of CSI feedback generated by two different methods: DMD and TransNet. Both methods can compress the CSI matrix. However, the DMD output is divided into components (called modes) with varying importance, whereas the TransNet treats its output as equally important. Our comparative analysis involves utilizing Unequal Error Protection (UEP) [5] techniques in quantization, channel coding, and modulation to allocate different levels of error protection to the CSI feedback parts based on their significance. The primary objectives are:

- To assess the impact of TransNet and DMD compression on the CSI quality.
- To analyze how quantization, channel coding, and modulation affect CSI feedback.
- To investigate the benefits of using UEP with CSI compression methods.

By providing a comparative analysis of TransNet and DMD, this paper highlights which method may be more effective under various conditions, potentially improving CSI feedback accuracy and overall network performance.

Notations: Throughout this paper, we represent matrices by uppercase boldface letters, column vectors by bold lowercase letters, scalars by italic lowercase letters and numbering by italic uppercase letters. $E\{\cdot\}$ denotes the mean.

II. SYSTEM AND CHANNEL MODELS

A. System Model

We consider an Orthogonal Frequency Division Multiplexing (OFDM) system with K subcarriers. Channel estimation at the UE is performed over a time segment of duration T , resulting in the estimated channel matrix $\mathbf{H} \in \mathbb{C}^{K \times T}$, with T is the number of OFDM symbols. Though we assume a single-antenna setup, this can be extended to a multi-antenna system. For such, the channel matrix is $\tilde{\mathbf{H}} \in \mathbb{C}^{K \times N_t \times N_r \times T}$, where N_t and N_r are the numbers of transmit and receive antennas, respectively. This 4D channel matrix can then be reshaped into a 2D matrix $\mathbf{H} \in \mathbb{C}^{KN_t N_r \times T}$ for processing. For simplicity, we consider $N_t = 1$ and $N_r = 1$, resulting in $\mathbf{H} \in \mathbb{C}^{K \times T}$. Moreover, \mathbf{H} is assumed to be noisy due to channel estimation errors, characterized by the Est-SNR (Signal-to-Noise Ratio). The corresponding CSI feedback is obtained by compressing \mathbf{H} , followed by quantization, coding, and modulation before transmission to the BS. The received CSI at the BS undergoes the reverse process (demodulation, decoding, dequantization and decompression), to reconstruct the channel matrix $\hat{\mathbf{H}}$. Fig. 1 illustrates the CSI feedback transmission procedure. The bubbles labeled (1), (2), and (3) divide the chain into parts that will be analyzed in the simulations.

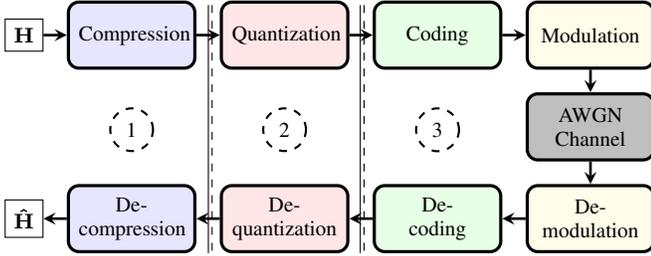


Fig. 1. System model for CSI feedback transmission.

Furthermore, due to the compression and quantization errors along with wireless channel noise, $\hat{\mathbf{H}}$ can deviate from the estimated \mathbf{H} , leading to a channel reporting error. To evaluate performance, we employ the Normalized Mean Square Error (NMSE), defined as: $\text{NMSE} = \frac{E\{\|\mathbf{H} - \hat{\mathbf{H}}\|_2^2\}}{E\{\|\mathbf{H}\|_2^2\}}$. Through this work, we use the ratio of energy per bit to noise power spectral density (Eb/N0) to represent the effects of the AWGN channel.

B. Time-Varying Channel Model

In wireless mobile networks, the movement of the UE causes in time-varying changes to the channel in the time domain [6]. One parameter used to characterize this dynamic behaviour is the coherence time τ , that represents the duration over which the channel remains temporally correlated and can be defined as:

$$\tau = \frac{c}{2vf_c}. \quad (1)$$

Where v , f_c and c denote the UE velocity, the signal carrier frequency and the speed of light, respectively. For the purposes of this paper, we assume that channel estimation occurs within this coherence time τ .

C. Channel Sparsity Model

The matrix \mathbf{H} represents the channel coefficients in frequency domain. Due to the radio propagation environment, it is accepted that the channel exhibits sparsity in the time domain, consistent with the 3GPP channel model [7]. We define:

$$\mathbf{H} = \text{DFT}(\mathbf{G}), \quad (2)$$

where $\text{DFT}(\cdot)$ represents the Discrete Fourier Transformation. Here, $\mathbf{G} \in \mathbb{C}^{K \times T}$ denotes the channel impulse response in the time domain, which is sparse along the K dimension, indicating that only a few taps are significant. According to the 3GPP model, the level of sparsity S , which represents the number of non-zero taps and their positions, remains constant during the coherence time τ . The non-zero taps of the CIR can be arranged into a matrix as follows:

$$\mathbf{S} = \mathbf{G}[\mathbf{G} \neq 0], \quad (3)$$

where $\mathbf{S} \in \mathbb{C}^{S \times T}$. To reconstruct \mathbf{H} from \mathbf{S} , we first zero-pad \mathbf{S} to form $\mathbf{G} = [\mathbf{S} \mathbf{0}]$. Then, we apply the Inverse Discrete Fourier Transformation (IDFT) to obtain: $\mathbf{H} = \text{IDFT}(\mathbf{G})$.

III. CSI FEEDBACK COMPRESSION

In this section, we introduce the key concepts behind two different methods for compressing CSI feedback. While both methods focus on reducing the size of the channel matrix, they employ distinct approaches and result in different forms of compressed CSI feedback.

A. DMD-based CSI Compression

Dynamic Mode Decomposition [2] is a data-driven technique used to decompose dynamical systems into spatiotemporal coherent structures that oscillate at fixed frequencies which either grow or decay at fixed rates. This method involves collecting snapshots from a dynamical system. In the context of wireless channels, the matrix \mathbf{H} comprises T channel snapshots. Specifically, $\mathbf{H} = [\mathbf{h}_1 \mathbf{h}_2 \dots \mathbf{h}_T]$, with each $\mathbf{h}_t \in \mathbb{C}^{K \times 1}$ representing the channel vector at all subcarriers over the OFDM symbol t , where $t \in [1, \dots, T]$. To use DMD, the channel vectors must be organized into two data matrices:

$$\begin{aligned} \mathbf{H}' &= [\mathbf{h}_1 \mathbf{h}_2 \dots \mathbf{h}_{T-1}] \in \mathbb{C}^{K \times T-1}, \\ \mathbf{H}'' &= [\mathbf{h}_2 \mathbf{h}_3 \dots \mathbf{h}_T] \in \mathbb{C}^{K \times T-1}. \end{aligned} \quad (4)$$

DMD defines a linear approximation to express how \mathbf{H}'' evolves from \mathbf{H}' as:

$$\mathbf{H}'' \approx \mathbf{A}\mathbf{H}', \quad (5)$$

where $\mathbf{A} \in \mathbb{C}^{K \times K}$ is an approximating linear operator, determined as: $\mathbf{A} = \mathbf{H}''\mathbf{H}'^\dagger$. This solution minimizes the Frobenius norm $\|\mathbf{H}'' - \mathbf{A}\mathbf{H}'\|_F$ functioning as a linear regression of data onto the dynamics represented by \mathbf{A} . In practice, direct analysis of the matrix \mathbf{A} may be intractable, especially when the number of subcarriers is extensive. However, the rank of \mathbf{A} is at most $T-1$, since it is constructed as a linear combination of the $T-1$ columns of \mathbf{H} . Thus, instead of solving for \mathbf{A} , DMD projects the data onto a low-rank subspace defined by at

most $T - 1$ Proper Orthogonal Decomposition (POD) modes. It then solves for a low-dimensional solution evolving on these POD modes. The DMD then uses this low-dimensional solution to find the leading M eigenvectors $\Phi \in \mathbb{C}^{K \times M}$ and eigenvalues $\Lambda \in \mathbb{C}^{M \times 1}$, which are called DMD modes and dynamics, respectively. It has been demonstrated in [2] that the snapshots are recomposed as:

$$\mathbf{h}_t \approx \Phi \Lambda^t. \quad (6)$$

Here M denotes the DMD rank truncation, indicating the number of eigendecompositions used. Eq. (6) implies that a higher M can generally improve the resolution of the recomposed \mathbf{h}_t . However, it is important to note that the eigendecompositions are sorted by significance. Thus, using few modes and dynamics may be sufficient to ensure an adequate resolution of the recomposed \mathbf{h}_t . Moreover, truncation can also contribute to noise reduction. Since DMD decomposes the channel matrix into modes that capture dominant frequencies and their growth or decay rates, and given that the channel matrix exhibits sparsity in the time domain, the resulting DMD modes will also be sparse in the time domain, reflecting these dominant frequencies. This sparse representation, denoted as $\Phi_{sp} \in \mathbb{C}^{K \times M}$, is obtained by applying an IDFT as follows: $\Phi_{sp} = \text{IDFT}(\Phi)$. The non-zero taps in Φ_{sp} are then arranged in a matrix $\Psi \in \mathbb{C}^{S \times M}$, such that: $\Psi = \Phi_{sp}[\Phi_{sp} \neq 0]$. Consequently, the CSI feedback generated by DMD comprises Ψ and Λ , resulting in a total size of $(S \times M + M)$. Thereby the CSI feedback size can be adjusted by varying the rank M .

B. TransNet-based CSI Compression

TransNet [8] is an advanced neural network model originally designed to enhance the efficiency of CSI feedback compression in massive Multi-Input Multi-Output (MIMO) systems. It employs a full attention network based on the Transformer [9] architecture, known for its ability to learn long-range dependencies in data. This allows the model to focus on relevant parts of the input sequence when generating outputs, enhancing contextual understanding. As shown in Fig. 2, TransNet employs a sophisticated Encoder-Decoder architecture designed to compress and reconstruct the channel matrix. The network input consists of the non-zero taps of the channel matrix representation in time domain, denoted as \mathbf{S} . The TransNet encoder consists of two encoding layers, each incorporating multihead attention, normalization, and feed forward networks. This structure is followed by a Fully Connected (FC) layer that generates a compressed CSI feedback. The TransNet decoder mirrors the TransNet encoder's architecture to reconstruct the original \mathbf{S} matrix from the compressed data. It begins with an FC layer and proceeds with two decoder layers, each including multihead attention, normalization, and feed forward processes. The output of the TransNet decoder is the matrix $\hat{\mathbf{S}}$. This is padded with zeros to obtain $\hat{\mathbf{G}}$. Fourier transformation is then applied to obtain $\hat{\mathbf{H}}$ as can be seen in the Fig. 2. It is worth noting that the CSI feedback size can be set by tuning the compressing scale using the FC layers.

The authors in [8] consider the channel matrix $\mathbf{H}_a \in \mathbb{C}^{K \times N_t}$, which includes the subcarriers for the N_t transmitter antennas. In this case, TransNet leverages the spatial correlation between the MIMO channels. However, since we address time-varying channels, we consider the channel subcarriers over time, as discussed in Section II, exploiting the temporal correlation.

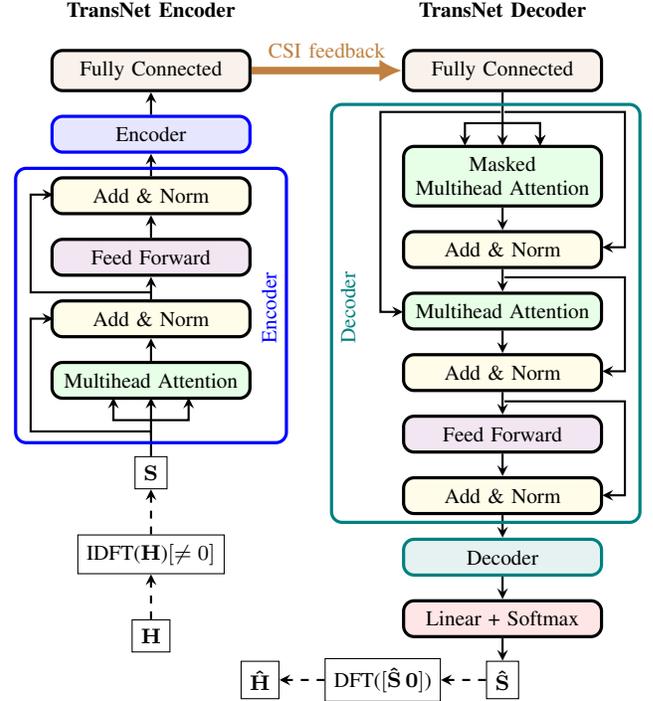


Fig. 2. TransNet network architecture for CSI compression.

IV. PRE-TRANSMISSION PROCESS

This section outlines all the essential steps required to prepare the CSI feedback for the wireless transmission.

A. Quantization

Quantization is the process of converting a continuous signal into discrete levels for a digital representation. In wireless communication, this often involves representing the signal with binary values. Considering the number of bits used to represent each level is B , then the number of the quantization levels can be determined as: $Q = 2^B$. We use uniform quantization, with the amplitude quantized in a differential manner: the maximum value is set to 1 as a reference, and the remaining values are quantized to B bits. Although quantization reduces data size, it introduces quantization noise caused by the difference between the original analog signal and the quantized digital signal. While this error can degrade signal quality, careful design typically mitigates its effect, balancing data size and quantization quality.

B. Channel Coding

Channel coding adds redundant bits to data to enhance reliability by error detection and correction. However, this introduces data overhead, which reduces throughput. Code rate indicates the efficiency of the channel coding. It is defined as

the ratio of information length in bits to the total number of transmitted bits. A lower code rate means more redundancy, which is beneficial for noisy channels but reduces bandwidth efficiency, balancing error correction and throughput.

According to 3GPP standardization [10], polar codes [11] were adopted for channel coding in the CSI feedback transmission. Therefore, our discussion and simulations will focus specifically on polar codes. Polar codes operate by transforming a set of physical communication channels into a set of virtual channels with varying levels of reliability. This process, known as channel polarization, arranges the channels from the most reliable to the least reliable. This feature allows applying UEP technique in polar codes such that very important data can be coded with the high reliable polarized channels. Fig. 3 shows a simple example of mapping the data to the polarized channels according to their importance, with darker blue indicating higher data importance and darker green denoting more reliable channels.

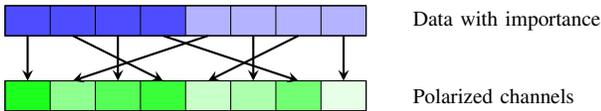


Fig. 3. An example of assigning data to polarized channels based on their importance.

C. Modulation

Modulation involves varying carrier signal characteristics, such as amplitude, frequency or phase, to represent the data bits for transmission [6]. For example, Binary Phase Shift Keying (BPSK) shifts the phase of the carrier signal between two states, offering simplicity and noise resistance but lower data rates. On the other hand, n -QAM (Quadrature Amplitude Modulation) varies both amplitude and phase, with higher n values increasing data rates by transmitting more bits per symbol but making the signal more sensitive to noise. Choosing n -QAM balances data rate and transmission reliability.

Additionally, Non-Orthogonal Multiple Access (NOMA) [12] is a technique that allows multiple users to share the same frequency band or time slot by assigning different power levels to their signals. When applied to a single user with multiple streams of varying importance, as in [13], NOMA serves as a multiplexing technique that modulates data streams to prioritize them by significance. For instance, NOMA can use QAM to manage these data streams. Such that, in a scenario with two data streams of equal length and different levels of importance, each can be modulated using 4-QAM but assigned distinct power levels. This setup allows for the simultaneous transmission of two 4-QAM symbols. From a data rate perspective, this is akin to using 16-QAM, but with the added benefit that one stream is more robust to bit errors.

V. SIMULATION SETUP AND RESULTS

In this section, we conduct numerical simulations to evaluate the effectiveness of the two CSI compression methods, DMD and TransNet, and compare their performance when preparing for transmission.

A. Simulation Setup

We employ Heterogenous Radio Mobile Simulator (HermesPy) [14] to generate the channel coefficients. The system parameters used are listed in Table I.

TABLE I
SIMULATION PARAMETERS

System Parameters	Value
Channel model	COST 259 [15]
Carrier frequency f_c	2 GHz
UE velocity v	130 Km/h
Subcarrier spacing	15 KHz
Number of subcarriers K	72
Sparsity S	8

Using (1), the coherence time is approximately $\tau \approx 2.1$ ms. With a subcarrier spacing of 15 KHz, the duration of one OFDM symbol is $66.7 \mu\text{s}$. The number of OFDM symbols T is calculated as $T = \lceil \frac{2.1 \cdot 10^{-3}}{66.7 \cdot 10^{-6}} \rceil = 32$. Thus, channel estimation is performed over 32 OFDM symbol, resulting in an estimated channel \mathbf{H} of size (72×32) complex coefficients. The matrix \mathbf{S} , representing the CIR non-zero taps, has a size of (8×32) complex coefficients, given the sparsity $S = 8$.

We trained the TransNet network using a dataset of 3000 realizations, each of the same size as \mathbf{S} . The training dataset is subjected to noise with Est-SNR values ranging between 10 and 30 dB. The training process spanned 500 epochs with a batch size of 10. All other parameters are the same as in [4].

To effectively demonstrate the simulation results, we split up the system model chain in Fig. 1 and calculate the NMSE for each process. We start by evaluating compression effect for the two mentioned methods on \mathbf{H} , as depicted in Fig. 1 part (1). We denote the reconstructed channel matrix as $\hat{\mathbf{H}}_1$, and the performance measurement as NMSE_1. Then, we incorporate the quantization process into the compression framework, as in Fig. 1 part (2). This results in $\hat{\mathbf{H}}_2$ and NMSE_2. Since the data to be quantized is complex, we quantize the real and imaginary parts separately, each with B bits. In the following step, we integrate channel coding using polar codes into the previously introduced processes, as in Fig. 1 part (3), resulting in $\hat{\mathbf{H}}_3$ and NMSE_3. To evaluate channel coding performance, introducing some bit errors to the coded data is necessary. Thus, we use BPSK modulation, which allows transmitting data through a noisy channel. Using the simple BPSK scheme enables focusing on assessing the channel coding performance. Finally, we assess the modulation process by executing the full procedure shown in Fig. 1. This results in $\hat{\mathbf{H}}$ and NMSE. For transmitting the coded CSI feedback, we use 16-QAM modulation. Additionally, we incorporate the NOMA technique to apply UEP to the DMD-based CSI, evaluating its performance by considering two data streams of different importance. Streams are formed by concatenating DMD modes as detailed in the results. NOMA is implemented with power ratios of 1 : 2, 1 : 3, and 1 : 4, with higher power allocated to more important stream. All used modulation schemes have their transmission power normalized to one.

B. Simulation Results

Fig. 4 compares the performance of DMD and the TransNet with varying CSI sizes, where each method is denoted in the legend along with its corresponding output size, measured in complex coefficients. In this analysis, a size of 9 coefficients corresponds to a DMD rank of $M = 1$, while a size of 27 coefficients corresponds to $M = 3$. The TransNet-based CSI size can be adjusted to be equal to the DMD output size by tuning the FC layer parameters. The x-axis represent the channel estimation error as an Est-SNR value in dB.

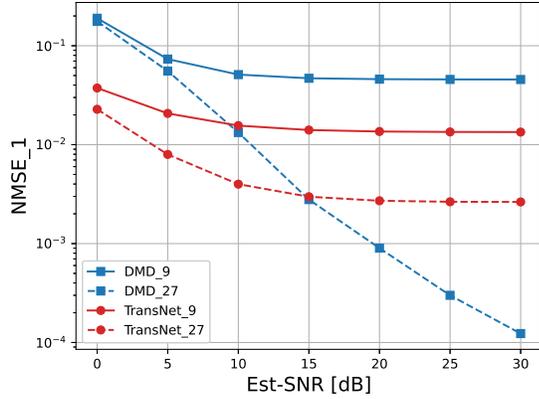


Fig. 4. Comparison of CSI compression performance for TransNet and DMD, in terms of channel estimation Est-SNR and for two output sizes.

For a CSI size of 9 coefficients, TransNet consistently outperforms DMD by approximately 5 dB, regardless of the Est-SNR value. However, when the CSI feedback size is increased to 27 coefficients, DMD initially performs about 9 dB worse than TransNet but improves rapidly with increasing Est-SNR, nearly matching TransNet’s NMSE₁ at an Est-SNR of 15 dB. At a high Est-SNR of 30 dB, DMD surpasses TransNet by approximately 13 dB. These results suggest that under specific conditions, such as high channel estimation performance and ample system resources for overhead, DMD can achieve highly accurate channel compression, while TransNet shows better performance in other scenarios. While DMD reduces noise through rank truncation, see III-A, TransNet performs better in high-noise scenarios due to its self-attention mechanism, which focuses on relevant data and ignores noise.

Fig. 5, illustrates the quantization performance NMSE₂ for various number of bits B . We consider the CSI feedback size at 27 coefficients, where both DMD and TransNet perform reasonably well, see Fig. 4. Simulation results in subfigure (a), show that increasing B from 4 to 6 significantly improves the NMSE₂ for both DMD and TransNet. However, further increasing B from 6 to 8 bits yields no substantial enhancement. Furthermore, we leverage the unique characteristic of DMD outputs, which consist of three distinct modes of different importance. Different quantization levels are applied to these modes to optimize performance. Subfigure (b) provides insights, where the legend indicates the number of bits allocated to each mode in order of importance. It is evident that the

configuration $B = (8, 6, 4)$ achieves superior performance by allowing the first mode to be constructed with high accuracy. Conversely, the $B = (2, 8, 8)$ configuration results in poor NMSE₂ due to inadequate reconstruction of the dominant mode. These findings underscore the potential of using UEP techniques to safeguard the critical data within DMD outputs, thereby enhancing reconstruction accuracy. It is important to note that despite varying quantization levels, the overall output size remains constant at $(27 \times 2 \times 6 = 324)$ bits in all scenarios. For the next simulations we take always $B = (8, 6, 4)$ for DMD, which results in 144 bits for the first mode, 108 bits for the second mode and 72 bits for the third mode.

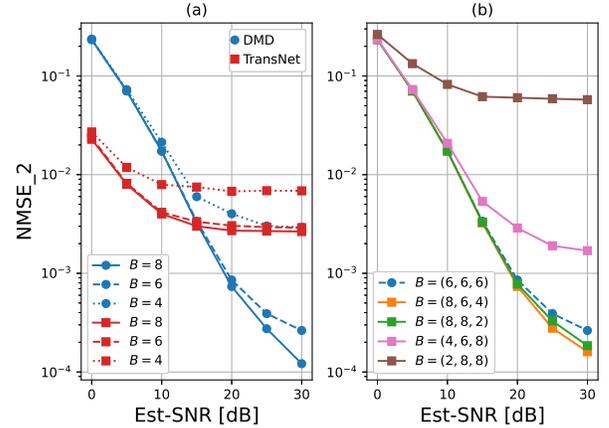


Fig. 5. CSI quantization performance with different values B in terms of Est-SNR. (a) compares TransNet and DMD. (b) compares applying different values of B on the DMD modes.

Fig. 6 illustrates the performance of the polar coding applied to TransNet and DMD utilizing the UEP technique. To ensure a fair comparison, we set the channel estimation Est-SNR to be 15 dB, as both methods performs comparably under this condition, see Fig.4. Further, we use the quantization configuration $B = (8, 6, 4)$ for DMD and $B = 6$ for TransNet, resulting in the same size of CSI. Simulation results show comparable performance for DMD and TransNet, and both methods converge to the optimal (error-free) decoding when $E_b/N_0 = 4$ dB. However, by utilizing UEP techniques with polar codes, DMD data can be optimized by assigning bits from dominant modes to the most reliable polar code channels. This configuration is labeled as “DMD_optimized” in the legend of the figure, and demonstrates improved performance within the E_b/N_0 range of 0 to 4 dB compared to merely transmitting DMD data through the polar code channels without considering data significance and channel reliability. Additionally, we apply the polar codes for each mode independently, employing different code rates based on the importance of each DMD mode: $(\frac{144}{250})$ for the first mode, $(\frac{108}{150})$ for the second mode, and $(\frac{72}{100})$ for the third mode. The codeword lengths are also selected to accommodate the next step, involving NOMA, as discussed later. As depicted in Fig. 6, applying polar codes to the DMD modes separately, denoted as “DMD_separated,” enhances NMSE₃ and achieves convergence to optimal decoding at $E_b/N_0 = 2$ dB.

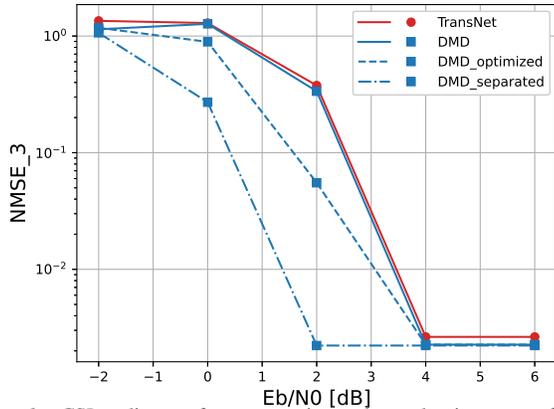


Fig. 6. CSI coding performance using polar codes in terms of Est-SNR, with a focus on examining the application of UEP techniques in both DMD_optimized and DMD_separated.

Finally, we evaluate the performance of modulation with and without UEP on the DMD modes, as depicted in Fig. 7. For NOMA, the first stream consists of the first mode with highest protection priority (250 bits). The second stream is created by concatenating the second and third modes, also totaling 250 bits. This concatenation provides both second and third modes an equal protection, even though their importance differs, as the lengths of the two NOMA input streams must match. Notably, a power ratio of 1 : 2 in NOMA shows better performance, achieving convergence to optimal transmission at $E_b/N_0 = 11$ dB, compared to 13 dB for QAM. This improvement is attributed to the enhanced protection of the first vector, which contains the most critical mode.

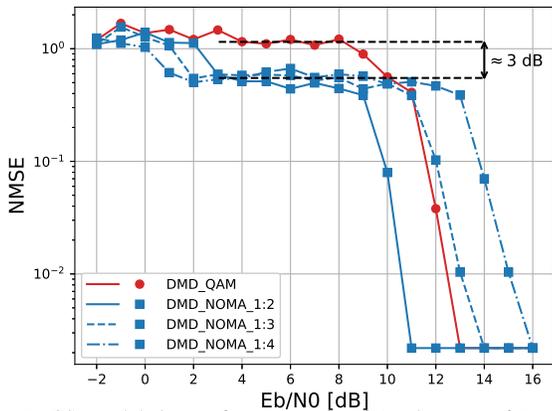


Fig. 7. CSI modulation performance using QAM in terms of Est-SNR, with a focus on examining the application of NOMA UEP technique.

Conversely, adjusting the power ratio to 1 : 3 or 1 : 4 results in worse performance than QAM, due to the severe reduced protection afforded to the second vector, which includes the second DMD mode. This shows that excessive error protection for the first mode, at the expense of the second and third modes, can degrade overall system performance. Therefore, it is crucial to balance data protection across modes. It is worth noting that the transmission power for all discussed cases is normalized. Additionally, the figure illustrates that within the E_b/N_0 range of 3 to 9 dB, all NOMA-based transmissions, irrespective of power ratio, outperform QAM by approximately

3 dB. This can be attributed to the substantial protection given to the first mode, even under poor channel conditions.

VI. CONCLUSION

This study evaluates two CSI compression methods, TransNet and DMD, and how pre-transmission processes affect CSI quality. Compressed CSI must undergo three main processes before wireless transmission: quantization, channel coding, and modulation. These processes can impact CSI quality and introduce unwanted overhead. While TransNet compresses CSI by treating all data uniformly, DMD decomposes the channel matrix into modes with varying importance, allowing for the application of UEP techniques to provide greater protection to the most critical modes of the CSI feedback. Our results indicate that, under conditions where DMD and TransNet compression perform similarly, DMD-based CSI transmission with UEP techniques outperforms TransNet-based CSI transmission. The ability to prioritize modes in DMD enables more effective use of UEP techniques, particularly when paired with polar codes' robust error correction and NOMA's stream-prioritized enhancements. In conclusion, although both methods compress CSI effectively, DMD-based CSI supports UEP integration and delivers superior performance, boosting the reliability of CSI feedback.

REFERENCES

- [1] R. Ahmed, E. Visotsky, and T. Wild, "Explicit csi feedback design for 5g new radio phase ii," in *WSA 2018; 22nd International ITG Workshop on Smart Antennas*, pp. 1–5, VDE, 2018.
- [2] J. N. Kutz, S. L. Brunton, B. W. Brunton, and J. L. Proctor, *Dynamic mode decomposition: data-driven modeling of complex systems*. SIAM, 2016.
- [3] F. Haddad, C. Bockelmann, and A. Dekorsy, "A dynamical model for csi feedback in mobile mimo systems using dynamic mode decomposition," in *ICC 2023-IEEE International Conference on Communications*, pp. 5265–5271, IEEE, 2023.
- [4] Y. Cui, A. Guo, and C. Song, "Transnet: Full attention network for csi feedback in fdd massive mimo system," *IEEE Wireless Communications Letters*, vol. 11, no. 5, pp. 903–907, 2022.
- [5] S. Borade, B. Nakiboğlu, and L. Zheng, "Unequal error protection: An information-theoretic perspective," *IEEE Transactions on Information Theory*, vol. 55, no. 12, pp. 5511–5539, 2009.
- [6] T. S. Rappaport, *Wireless communications: principles and practice*. Cambridge University Press, 2024.
- [7] "ETSI TR 138 901.," https://www.etsi.org/deliver/etsi_tr/138900_138999/138901/16.01.00_60/tr_138901v160100p.pdf.
- [8] W. Han, X. Ma, D. Tang, and N. Zhao, "Study of ser and ber in noma systems," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 4, pp. 3325–3340, 2021.
- [9] T. Lin, Y. Wang, X. Liu, and X. Qiu, "A survey of transformers," *AI open*, vol. 3, pp. 111–132, 2022.
- [10] "ETSI TS 138 212.," https://www.etsi.org/deliver/etsi_ts/138200_138299/138212/18.02.00_60/ts_138212v180200p.pdf.
- [11] E. Arikan, "Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels," *IEEE Transactions on information Theory*, vol. 55, pp. 3051–3073, 2009.
- [12] H. Yahya, A. Ahmed, E. Alsusa, A. Al-Dweik, and Z. Ding, "Error rate analysis of noma: Principles, survey and future directions," *IEEE Open Journal of the Communications Society*, 2023.
- [13] N. Bulk, C. Bockelmann, and A. Dekorsy, "Enhancing multi-service transmission efficiency in high-density scenarios with service-based noma and unequal error protection," 28. *VDE-ITG-Fachtagung Mobilkommunikation (MKT'24)*, May 2024.
- [14] "HermesPy," <https://www.barkhauseninstitut.org/en/results/hermespy>.
- [15] "ETSI TR 125 943.," <https://www.etsi.org/deliver/etsitr/125900125999/125943/06.00.0060/tr125943v060000p.pdf>.