A Low-Rank Approach of MIMO Optimization for Edge Smart Ports

Zechen He, Jiale Wang, Ping Feng, Jiahong Ning*, and Tingting Yang

Dalian Maritime University, Dalian, China

Email:{hzc1120221019,wangjl}@dlmu.edu.cn,747392722@163.com,jiahong.ning@mnsu.edu,yangtingting820523@163.com

***Abstract*—This study investigates the channel state information (CSI) feedback problem in large-scale multiple-input multiple-output (MIMO) systems, which is important for intelligent co-operative traffic management in smart ports. MIMO systems depend on CSI feedback for effective precoding, which enhances the system's transmission gain. While various strategies have been developed to minimize CSI feedback overhead, these have predominantly been assessed in static scenarios. Addressing the complexity and variability inherent in smart ports, this paper introduces a novel approach utilizing a Transformer-based method for CSI feedback. This method integrates the LoRA algorithm, optimizing computational efficiency during training and enabling rapid adaptation to complex, evolving environments. Experimental findings demonstrate that this approach not only requires fewer computational resources and operates more swiftly but also exhibits superior adaptability to environmental fluctuations. In addition, this approach greatly improves the robustness, stability, and resource optimization fairness of the system compared to existing CSI feedback techniques.**

***Index Terms*—CSI Feedback,Transformer,Low-Rank Adaptation**

I. Introduction

Recent advances in wireless communication have greatly influenced various industries, especially smart ports[1]. In smart ports, communication systems for cooperative transportation must be able to handle the transmission of control information and multi-channel information efficiently and reliably.Introducing large-scale MIMO systems[2] in smart ports is an unique challenge.These systems use multiple transmitting and receiving antennas to increase communication capacity and efficiency, significantly enhancing signal quality and transmission rates through parallel data stream transmission. However, the implementation of MIMO systems depends on precise CSI feedback to adjust transmission strategies, thereby ensuring efficient communication and maximizing system throughput.

With more antennas, receivers, and subcarriers, CSI feedback gets increasingly complex, especially in dynamic settings like smart ports. The heavy reliance on antennas greatly increases the need for CSI, as each antenna requires precise channel data for effective transmission strategies. Rapid changes in smart port scenarios, such as equipment movement and large-scale cargo flow, cause frequent changes in channel characteristics, requiring the CSI feedback system to adapt quickly and accurately. This increases data processing and transmission demands and requires high real-time accuracy. Additionally, these systems often manage substantial data volumes, including high-definition video and intricate control signals.This necessitates additional antennas and higher density subcarriers to support extensive data transmission, thus dramatically increasing the dimensions of CSI and adding to the complexity of data processing and transmission.In Smart Ports, many vehicles operate at various frequencies, each with different wireless signal propagation characteristics. These characteristics are influenced by factors such as distance, interference, and physical obstructions (e.g., containers and port infrastructure). When conducting remote collaborative operations of large port machinery, Channel State Information (CSI) feedback must accurately reflect the conditions of the channel, ensuring the stability of control signals even within complex electromagnetic environments. It's essential to account for these specific characteristics and environmental factors when processing CSI feedback from different frequencies, to maintain efficient and reliable wireless communication.Addressing these challenges, our research aims to reduce the overhead of CSI feedback in smart port scenarios while ensuring the timeliness and accuracy of communication, overcoming the challenges faced by existing CSI feedback technologies in complex environments while maintaining the high efficiency of MIMO systems,enhancing collaborative traffic management.

Using CSI's temporal and spatial correlations for channel compression estimation, along with Compressed Sensing (CS) technology in feedback protocol design, can effectively reduce CSI feedback overhead [3]. ErikssonT's[4] compressed sensing-based CSI feedback method improves channel information accuracy and lowers feedback load but depends greatly on prior channel structure assumptions. The CSI matrix, approximately sparse, has correlated element changes, demanding complex priors without assured recovery performance. Even advanced algorithms like BM3D-AMP[5], which apply complex priors for channel reconstruction, don't notably enhance CSI recovery quality due to its sensitivity to these priors' accuracy.

To tackle CSI feedback challenges, this paper[6] introduces CsiNet, a Deep Neural Network (DNN) model. CsiNet uses image reconstruction techniques for CSI encoding and decoding. Unlike traditional methods, it relies on extensive data analysis instead of mathematical models for CSI compression.

While CsiNet [6] demonstrates commendable performance, its extensive parameterization results in high computational complexity. This issue becomes increasingly pronounced with the addition of more antennas, receivers, and subcarriers, leading to significant computational complexity in CSI feedback. In dealing with the rapidly changing conditions and high-density scenarios of smart ports, these challenges are further amplified. This limitation underscores the necessity for a more robust and adaptive CSI feedback method in these environments.

The Transformer model, renowned for its success in natural language processing, has recently been applied to processing CSI data [7]. In [7], the self-attention mechanism is utilized to characterize both short and long-distance dependencies between data samples. Compared to other methods, this mechanism better captures CSI features and effectively expands the feature perception area. The model's inherent ability to process sequential data makes it an ideal candidate for addressing the spatiotemporal variations in wireless channels at smart ports, although it also increases computational overhead.This may not be ideal for many resource-constrained devices on smart ports, and for those with limited computing resources, real-time communication quality may not be guaranteed.

In light of the aforementioned challenges, we have adopted an innovative CSI architecture that integrates the Low-Rank Adaptation (LoRA) algorithm with a Transformer-based framework. This combination makes our approach particularly suited for real-time applications in resource-constrained environments, significantly reducing computational overhead. Our solution not only enhances computational efficiency but also improves the model's adaptability in rapidly changing conditions, a critical attribute for smart port scenarios. Compared to approaches based on the Transformer architecture, the reduction in the number of parameters conserves computational resources, lowers the risk of overfitting, and offers better adaptability and generalization in complex electromagnetic environments such as smart ports. With this novel CSI architecture that amalgamates the Transformer and LoRA algorithms, we are effectively equipped to handle the complex communication challenges in smart ports, maintaining both the efficiency and flexibility of the system. The primary contributions of this work are summarized as follows:

•We proposed a CSI feedback scheme named "Transformer-LoRA," which effectively integrates a Transformer-based architecture with the LoRA algorithm for stable and efficient CSI feedback.The scheme is optimized for frequently changing channel characteristics in massive MIMO systems and includes both an encoder and a decoder.

•It intelligently compresses the CSI data through low-rank matrix decomposition technology, which reduces the computational resources required for processing and transmission, and at the same time improves the adaptability of the model to different channel conditions and frequencies, enabling it to quickly respond to the frequent changes in the dynamic environment of smart ports for better cooperative traffic management.

•Through comparative experiments, our model demonstrates a fast adaptation capability superior to existing techniques when tested with pre-reserved CSI data. The simulation results demonstrate the model's higher efficiency and accuracy in processing CSI data under different environmental and frequency conditions, with an average computational overhead reduction of 28.31% compared to other schemes.

II. SYSTEM MODEL

Consider a simple scenario within a Frequency Division Duplexing network utilizing large-scale single-cell MIMO, where there are $N_t(\gg 1)$ transmit antennas at the Base Station (BS) and a single receive antenna at the User Equipment (UE). Orthogonal Frequency Division Multiplexing (OFDM) technology with $\tilde{N}_c$ subcarriers is also employed. The signal received at the $n^{\text{th}}(n=1,2,\ldots,\tilde{N}_c)$ subcarrier can be represented as:

$$y_n = \tilde{h}_n^H v_n x_n + z_n \tag{1}$$

where $\tilde{h}_n \in \mathbb{C}^{N_t \times 1}$ represents the channel vector for the n^{th} subcarrier, $v_n \in \mathbb{C}^{N_t \times 1}$ is the precoding vector, $x_n \in \mathbb{C}$ denotes the transmitted data symbol in the downlink, and $z_n \in \mathbb{C}$ is the additive noise for the n^{th} subcarrier.Next, the spatial frequency domain CSI stacking matrix is defined as

$$\tilde{\mathbf{H}} = [\tilde{\mathbf{h}}_1, \ldots, \tilde{\mathbf{h}}_{\tilde{N}_c}]^H \in \mathbb{C}^{\tilde{N}_c \times N_t}. \tag{2}$$

The User Equipment (UE) deduces $\tilde{\mathbf{H}}$ using pilot signals. The number of complex parameters is $\tilde{N}_c N_t$, proportional to the number of antennas. It is necessary to promptly feed $\tilde{\mathbf{H}}$ back to the Base Station (BS) via the feedback link to assist in generating the precoding vector. However, in large-scale MIMO systems, the substantial number of antennas leads to a high number of CSI matrix parameters. Direct feedback of the uncompressed CSI matrix would consume extensive bandwidth resources. To reduce feedback overhead, we employ the compression method proposed in [7], using a two-dimensional Discrete Fourier Transform (DFT) to sparsify $\tilde{\mathbf{H}}$ in the angle-delay domain, as shown below:

$$\mathbf{H} = \mathbf{F}_d \tilde{\mathbf{H}} \mathbf{F}_a^{\mathbf{H}} \tag{3}$$

where $\mathbf{F}_d$ and $\mathbf{F}_a$ are DFT matrices of dimensions $\tilde{N}_c \times \tilde{N}_c$ and $N_t \times N_t$, respectively. In the delay domain, due to the finite period of time delays between multiple path arrivals, the first N_c rows of $\mathbf{H}$ contain significant values. We can retain the first N_c rows of $\mathbf{H}$ to form a real-valued matrix $\mathbf{H}_a$ of dimensions $N_c \times N_t$. $\mathbf{H}_a$ is then fed into the encoder as shown in Fig.1.In this work, the encoder f_{en} transforms the channel matrix $\mathbf{H}_a$ into a codewords:

$$S = f_{en}(\mathbf{H}_a) \tag{4}$$

where $s \in \mathbb{R}^{M \times 1}$ and $M < N_c \times N_t$. The compression ratio is $M/N_c \times N_t$. Subsequently, the decoder f_{de} reconstructs the channel matrix $\mathbf{H}_b$ from the codewords S sent to the Base Station (BS) via the feedback link:

$$\mathbf{H}_b = f_{de}(S) \tag{5}$$

The general approach to CSI feedback is as follows. Initially, the channel matrix $\mathbf{H}$ is acquired on the UE side. A 2D DFT, as specified in (3), is performed to obtain the truncated matrix

$\mathbf{H}_a$. The encoder (4) is then used to generate the codeword S. Subsequently, S is sent back to the Base Station (BS) via the feedback link, where the BS employs the decoder (5) to retrieve $\mathbf{H}_b$. The final channel matrix in the spatial frequency domain is obtained by performing the inverse DFT.However, due to the dynamic and complex nature of port environments, it is challenging to find optimal encoders f_{en} and decoders f_{de} using traditional methods and some machine learning-based approaches. While there are methods involving Transformers, they also significantly increase computational overhead and burden. Below, a novel solution is proposed, combining Transformer and Lora techniques.

III. PROPOSED SCHEMES

A. Overall architecture

The general structure of CSI-LoRA is shown in Fig.1.Upon acquiring the channel matrix $\mathbf{H}$ on the UE side, a 2D DFT is first performed to obtain the truncated matrix $\mathbf{H}_a$. $\mathbf{H}_a$, a real-valued matrix of dimensions $N_c \times N_t$, undergoes convolution, batch normalization, and reshaping. It then passes through a Transformer-LoRA Layer as depicted in Fig.2, resulting in a matrix of size $S_1 \times S_2$. This matrix is then reshaped and passed through a Fully Connected Layer to produce a codeword S of length $S_1 \times S_2$.

Subsequently, the codeword S arrives at the BS side decoder via the feedback link. The decoder receives this codeword of length $S_1 \times S_2$ through a Fully Connected Layer and feeds a matrix of size $S_1 \times S_2$ into the Transformer-LoRA Layer as shown in Fig.2. After reshaping, convolution, and batch normalization, a sigmoid function is applied to constrain the output within the range $[0,1]$, thus converting the codeword S into the channel matrix $\mathbf{H}_b$ recovered by the network.

To train this feedback network, we conduct end-to-end learning on all kernels and bias values of the encoder and decoder. The network's input and output are normalized channel matrices, with elements scaled within the range $[0,1]$. The training algorithm is a type of unsupervised learning algorithm, and the parameter sets are updated via the ADAM algorithm.

Using the Transformer model in CSI feedback significantly improves communication system efficiency. The Transformer's self-attention mechanism excels in managing sequential data, especially in fast-changing wireless environments. Unlike traditional models such as RNN and LSTM, it can learn information of different levels or types in the sequence by adjusting the number of attention heads, effectively handling long-distance dependencies. This allows for a trade-off between model complexity and performance. For different tasks, various attention heads may also exhibit diverse performances. Its parallel processing feature allows for quicker and more efficient handling of large CSI data, essential in real-time systems. Additionally, its excellent generalization capability helps the model adapt to new communication scenarios, boosting overall system performance.

Although the Transformer model performs better, its greater complexity may require more computational resources. This is a critical issue for devices with limited resources, such as mobile or embedded systems, which often rely on battery power.

B. Low-Rank Adaptation

To address the limitations of the Transformer architecture, we have integrated the Low-Rank Adaptation (LoRA) algorithm. LoRA[8] allows indirect training of certain dense layers in neural networks by optimizing rank-decomposed matrices that vary during the adaptation phase, while keeping pre-trained weights frozen. This approach significantly reduces the computational effort required to update matrices during model training and enables rapid adaptation to specific domains or different environments. The core idea of LoRA is to introduce two low-rank matrices A and B in critical parts of the model, such as multi-head self-attention and feed-forward neural networks. These low-rank matrices adjust the original linear transformation matrix W, without directly modifying it. During the model's forward propagation, the original linear transformation W_x is replaced with $(W+AB)x$, where x is the input vector. In this way, LoRA enables the model to learn new representations without significantly increasing the number of model parameters. During fine-tuning, only the matrices A and B are updated, while the original W matrix remains unchanged. This approach offers multiple benefits beyond merely reducing the consumption of computational resources and ensuring model stability. By updating only a small subset of parameters, it also diminishes the risk of overfitting. Adjusting the size of low-rank matrices to control the extent of parameter updates further enhances the model's ability to resist interference.

In algorithm 1, A is initialized with random Gaussian values and B with zeros, so that $\Delta\mathbf{W} = \mathbf{BA}$ starts at zero at the beginning of training. The term $\Delta\mathbf{Wx}$ is then scaled using $\frac{\alpha}{r}$, where α is a constant in r. When optimizing with Adam, adjusting α is roughly equivalent to adjusting the learning rate.

Theoretically, LoRA can be applied to the weight matrix of any neural network, thus reducing the number of trainable parameters. Within the Transformer architecture, the self-attention module contains four weight matrices (W_q, W_k, W_v, W_o), along with two weight matrices of the MLP models. By employing various combinations of weights, superior performance can be achieved across different tasks. For tasks requiring rapid feedback, lower weights can be used to accelerate response times. Conversely, higher weight combinations can effectively reduce model error in tasks demanding high-precision feedback.

IV. EXPERIMENTAL RESULTS

A. Data preparation

The data utilized in this study were sourced from [9] and [10]. The former encompasses 52,500 Line Of Sight (LOS) and 105,000 Non-Line Of Sight (NLOS) sample data, generated in an outdoor environment by UMa at a speed of 30 km/h. The base stations were spaced 200 meters apart, with an antenna count of $N_t = 32$and a retained sub-band number of$N_C = 13$, operating under FR1 at 2 GHz. The latter dataset originates from a Base Station (BS) equipped with a Uniform Linear Array (ULA), comprising a total of 320,000 samples. This dataset consists of data from $N_t = 32$ antennas and $N_C = 256$

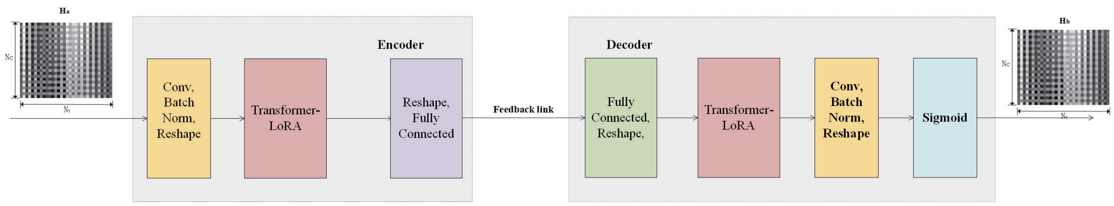


Fig. 1: Architecture of the proposed Csi_Transformer_LoRA with encoder and decoder

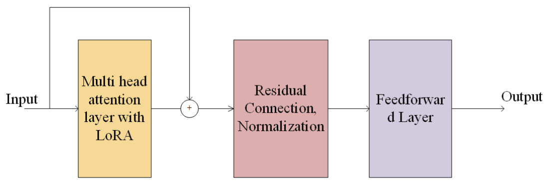


Fig. 2: Structure of the transformer layer with LoRA

Algorithm 1 Transformer Model with LoRA

Require: Pre-trained Transformer, rank r, scale α, flags for LoRA application *apply_lora*, training data, learning rate lr.
Ensure: Transformer with updated weights.

Initialize the Transformer with pre-trained weights.
for each multi-head attention layer and each head h **do**
  if *apply_lora* is true **then**
    Initialize LoRA matrices A_h and B_h.
    Set A_h and B_h as trainable parameters.
  end if
end for
Initialize the optimizer with learning rate lr.
for each training batch **do**
  Forward pass:
  for each layer and head h **do**
    if *apply_lora* **then**
      Compute $\Delta W_h = \alpha \times A_h \times B_h$.
      Adjust weights: $W_h \leftarrow W_h + \Delta W_h$.
    end if
  end for
  Compute predictions and loss.
  Backward pass for A_h and B_h gradients.
  Update A_h and B_h with the optimizer.
  Reset W_h after each batch.
end for
Evaluate and potentially revert LoRA matrices based on validation performance.
return the adapted model.

subcarriers, with only the first 16 rows retained in the angle-delay domain, operating at 3.5 GHz.

In addition to employing these datasets individually to evaluate the performance of the proposed scheme in this paper, a composite dataset, merging these three datasets, was used to simulate complex dock environments and CSI feedback scenarios involving multiple wireless devices.To reduce the risk of model overfitting and enhance model robustness and usability, a portion of the dataset with too much similarity is deleted and the dataset is disrupted after combination. All training samples exclude validation and test samples. The datasets were partitioned into training, validation, and testing sets at a ratio of 7:1:2. The epoch, learning rate, and batch size were set to 30, 0.01, and 100, respectively. The original channel data $\mathbf{H}$ and the reconstructed channel data $\hat{\mathbf{H}}$ are quantified using the NMSE as defined in 6.

$$\mathrm{NMSE} = E\left\{\frac{\|\mathbf{H}-\hat{\mathbf{H}}\|^2}{\|\mathbf{H}\|^2}\right\}. \tag{6}$$

At the same time, cosine similarity is also used to measure the quality of beamforming vectors used in different schemes, defined as follows:

$$\rho = E\left\{\frac{1}{N_c}\sum_{n=1}^{N_c}\frac{\hat{\hat{\mathbf{h}}}_n^H \hat{\mathbf{h}}_n}{\|\hat{\hat{\mathbf{h}}}_n\|_2 \|\hat{\mathbf{h}}_n\|_2}\right\}. \tag{7}$$

where $\hat{\mathbf{h}}_n$s the original channel vector of the nth subcarrier,$\hat{\hat{\mathbf{h}}}_n$is the reconstructed channel vector of the nth subcarrier.

B. *Comparative analysis*

The experiment compares the performance of three schemes: Csi_net[6], Transformer_Net[7], and the proposed Transformer_LoRA_Net, under both LOS and NLOS conditions. The optimal results are presented in Tables I , II and III.

TABLE I: PERFORMANCE COMPARISON IN LOS

Compression Rate	Scheme	NMSE	ρ
1/8	Transformer_Net	**0.0332**	**0.991**
	Transformer_LoRA_Net	0.0676	0.960
	Csi_Net	0.0591	0.966
1/16	Transformer_Net	**0.0932**	**0.947**
	Transformer_LoRA_Net	0.1276	0.923
	Csi_Net	0.1083	0.941
1/32	Transformer_Net	**0.2043**	**0.875**
	Transformer_LoRA_Net	0.2296	0.843
	Csi_Net	0.2944	0.806

TABLE II: PERFORMANCE COMPARISON IN NLOS

Compression Rate	Scheme	NMSE	ρ
1/8	Transformer_Net	**0.0472**	**0.975**
	Transformer_LoRA_Net	0.0672	0.959
	Csi_Net	0.0541	0.963
1/16	Transformer_Net	**0.0972**	**0.951**
	Transformer_LoRA_Net	0.1172	0.938
	Csi_Net	0.1041	0.942
1/32	Transformer_Net	**0.2166**	**0.869**
	Transformer_LoRA_Net	0.2459	0.862
	Csi_Net	0.2813	0.820

From the above table, it can be found that at low compression rates, the performance scenarios of the three schemes are excellent, but at high compression rates, the performance of Csi_net[6] is somewhat deficient compared to the other two

TABLE III: PERFORMANCE COMPARISON IN 3.5GHz

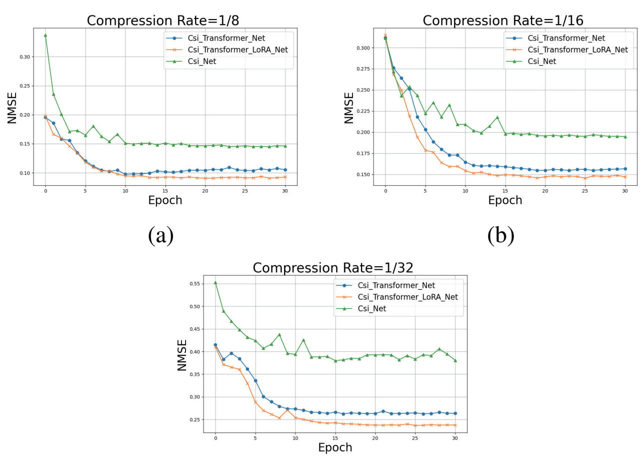
Compression Rate	Scheme	NMSE	ρ
1/8	Transformer_Net	**0.0371**	**0.981**
	Transformer_LoRA_Net	0.0382	0.980
	Csi_Net	0.0618	0.964
1/16	Transformer_Net	**0.1162**	**0.932**
	Transformer_LoRA_Net	0.1123	0.929
	Csi_Net	0.1652	0.903
1/32	Transformer_Net	**0.2340**	**0.867**
	Transformer_LoRA_Net	0.2448	0.861
	Csi_Net	0.3422	0.803

schemes.CsiBut in complex environments, see Table IV, the performance scenarios of the three systems have a certain degree of degradation, and NMSE is elevated for all of them. Due to the different data formats of the two datasets.

TABLE IV: PERFORMANCE COMPARISON IN FLEXIBLE

Compression Rate	Scheme	NMSE	ρ
1/8	Transformer_Net	0.0748	0.959
	Transformer_LoRA_Net	**0.0619**	**0.960**
	Csi_Net	0.1157	0.932
1/16	Transformer_Net	0.1566	0.907
	Transformer_LoRA_Net	**0.1470**	**0.915**
	Csi_Net	0.1837	0.895
1/32	Transformer_Net	0.2941	0.0.810
	Transformer_LoRA_Net	**0.2686**	**0.847**
	Csi_Net	0.4093	0.784

From 3, it can be seen that under the three compression rates, Transformer_LoRA_Net is the fastest convergence among the three schemes and the performance of the system is optimal.



(a) (b)

(c)

Fig. 3: NMSE performance of Csi_Net, Transformer_Net and Transformer_LoRA_Net during training as a function of elapsed time for different compression rates

C. System Performance

Finally, a comparison of Transformer_Net[7] and Transformer_LoRA_Net was made to obtain the maximum amount of video memory utilized during their operation using the test dataset at different compression rates, and it is shown in Table IV that the Transformer_LoRA_Net scheme can drastically reduce the amount of video memory required.

TABLE V: MAX DISPLAY MEMORY USED

Compression Rate	Scheme	MAX DISPLAY MEMORY(MB)
1/16	Transformer_Net	189.8
	Transformer_LoRA_Net	136.4
1/32	Transformer_Net	185.3
	Transformer_LoRA_Net	133.2
1/64	Transformer_Net	186.9
	Transformer_LoRA_Net	131.5

V. CONCLUSION

In summary, this study introduces an innovative CSI feedback method in large-scale MIMO systems. The design and implementation of a CSI feedback system that utilizes both Transformer and LoRA algorithms is a technical challenge, necessitating a deep understanding of the working principles and interactions of these two technologies. Furthermore, maintaining the system's robustness and high performance while optimizing computational efficiency and adaptability.Our approach opens up new perspectives and possibilities for the application of CSI feedback techniques in dynamic and complex environments. It focuses on maintaining communication efficiency and stability in smart ports' dynamic environments while conserving computational resources. However, specific scenarios on smart ports may require further customization of feedback frameworks to address their unique challenges effectively. This research is not only important for cooperative traffic management in the field of smart ports, but also provides valuable insights for other applications of large-scale multiple-input multiple-output systems.

REFERENCES

[1] Kok-Lim Alvin Yau, Shuhong Peng, Junaid Qadir, Yeh-Ching Low, and Mee Hong Ling. Towards smart port infrastructures: Enhancing port activities using information and communications technology. *IEEE Access*, 8:83387–83404, 2020.

[2] Erik G Larsson, Ove Edfors, Fredrik Tufvesson, and Thomas L Marzetta. Massive mimo for next generation wireless systems. *IEEE communications magazine*, 52(2):186–195, 2014.

[3] Xiongbin Rao and Vincent K. N. Lau. Distributed compressive CSIT estimation and feedback for FDD multi-user massive MIMO systems. *CoRR*, abs/1405.2786, 2014.

[4] Thomas Eriksson and Tony Ottosson. Compression of feedback for adaptive transmission and scheduling. *Proceedings of the IEEE*, 95(12):2314–2321, 2007.

[5] Christopher A Metzler, Arian Maleki, and Richard G Baraniuk. From denoising to compressed sensing. *IEEE Transactions on Information Theory*, 62(9):5117–5144, 2016.

[6] Chao-Kai Wen, Wan-Ting Shih, and Shi Jin. Deep learning for massive mimo csi feedback. *IEEE Wireless Communications Letters*, 7(5):748–751, 2018.

[7] Xiaojun Bi, Shuo Li, Changdong Yu, and Yu Zhang. A novel approach using convolutional transformer for massive mimo csi feedback. *IEEE Wireless Communications Letters*, 11(5):1017–1021, 2022.

[8] Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. Lora: Low-rank adaptation of large language models, 2021.

[9] Open Data Platform for mobile communication. Csi feedback datasets. https://www.mobileai-dataset.cn/, 2022-12-01.

[10] National Mobile Communications Research Laboratory. Southeast university. http://www.china-ai.ac.cn/, 2020.