A Scalable Deep Learning Framework for Multi-rate CSI Feedback under Variable Antenna Ports

Yu-Chien Lin, Ta-Sung Lee, and Zhi Ding

Abstract—Channel state information (CSI) at transmitter is crucial for massive MIMO downlink systems to achieve high spectrum and energy efficiency. Existing works have provided deep learning architectures for CSI feedback and recovery at the eNB/gNB by reducing user feedback overhead and improving recovery accuracy. However, existing DL architectures tend to be inflexible and non-scalable as models are often trained according to a preset number of antennas for a given compression ratio. In this work, we develop a flexible and scalable learning framework based on a divide-and-conquer approach (DCA). This new DCA architecture can flexibly accommodate different numbers of 3GPP antenna ports and dynamic levels of feedback compression. Importantly, it also significantly reduces computational complexity and memory size by allowing UEs to feedback segmented downlink CSI. We further propose a multi-rate successive convolution encoder with fewer than 1000 parameters. Test results demonstrate superior performance, good scalability, and low complexity for both indoor and outdoor channels.

Index Terms—CSI feedback, scalability, dynamic architecture, massive MIMO, deep learning

I. INTRODUCTION

Massive Multiple-input multiple-output (MIMO) technologies play an important role in improving spectrum and energy efficiency of 5G and future generation wireless networks. The strength of massive MIMO hinges on accurate downlink channel state information (CSI) at the basestation or gNodeB (gNB). Without the benefit of uplink/downlink (UL/DL) channel reciprocity assumed in time-division duplexing (TDD) systems, frequency-division duplexing (FDD) systems typically rely on user equipment (UE) feedbacks to gNBs for DL CSI recovery. The growing large number of DL transmit antennas envisioned in millimeter wave bands or higher frequencies [1] requires a vast amount of uplink feedback information and resources such bandwidth and power. To conserve bandwidth and UE battery, efficient compression of CSI feedback is vital to broad FDD deployment of massive MIMO technologies.

From radio physics, cellular CSI exhibits limited delay spread (sparsity). Efficient UE feedback can take advantage of sparsity for CSI compression. To leverage CSI sparsity for improving feedback efficiency, a deep autoencoder framework [2] deployed encoders at UE and decoder at serving station (gNB) for CSI compression and recovery, respectively. This and other related works have demonstrated the efficacy of CSI recovery with the aids of deep learning autoencoder [3]–[6]. Physical insights with respect to slow temporal variations of propagation scenarios, similar propagation conditions of nearby UEs, and similarity of UL/DL radio paths reveal significant temporal, spatial, and spectral CSI correlations respectively. Beyond basic autoencoders, more recent works take advantages of various correlated channel information such as past CSI [3], [7], CSI of nearby UEs [8], and UL CSI [9]–[11] to aid and improve the recovery of DL CSI at base stations. Additional works considered antenna array geometry to explicitly exploit UL/DL angular reciprocity [12], [13].

To extract underlying correlated antenna information, current CSI frameworks compress and recover DL CSI over all antennas. This leads to high model complexity because of the large input size. There have attempts to directly reduce encoder’s model complexity [14]–[16] with limited success. We consider the physical insight that since geometric size of massive MIMO antenna array spans multiple wavelengths, only nearby antennas exhibit non-negligible CSI correlation. According, our low complexity framework focuses on adjacent antennas (or antenna ports) with stronger spatial correlation to compress the CSI of large array via a divide-and-conquer approach (DCA).

This works aims to systematically simplify deep learning architecture and computational complexity for DL CSI feedback while limiting the accuracy loss of CSI recovery. We develop a flexible and scalable multi-rate CSI feedback framework along with a lightweight convolutional encoder. Our contributions are summarized below.

• The DCA framework CSI is a new learning-based compression and recovery mechanism that systematically reduces the model size by exploiting both strong and weak CSI correlations among adjacent and non-adjacent antennas, respectively.
• Our framework is flexible and applicable to multi-rate CSI compression ratios, including a universal encoder for all compression ratios.
• The new framework is scalable to large antenna sizes.
• The framework incorporates a swapping encoder mechanism which generates a few extra feedback bits to further
improve the CSI recovery performance.

II. SYSTEM MODEL

We consider a single-cell MIMO FDD link in which a gNB using a \( N_a \) antenna ports communicates with single antenna UEs. Following the 3GPP specification, pilot symbols (or DMRS) for each antenna port are uniformly placed in frequency domain for downlink transmission. Assuming each subband contains \( N_f \) subcarriers with spacing of \( \Delta f \) and a downsampling rate \( \text{DR}_f \), the frequency interval between consecutive pilots is \( \text{DR}_f \cdot \Delta f \). We denote \( h_i \in \mathbb{C}^{M_f \times 1} \) as DL pilot CSI of the \( i \)-th AP where \( M_f \) is the number of pilots. By collecting DL pilot CSI of each AP, a pilot sampled DL CSI matrix \( \mathbf{H} \) is related to the full CSI matrix \( \mathbf{H} \) is given by

\[
\mathbf{H} = \mathbf{H} \mathbf{Q}_{\text{DR}_f} = \begin{bmatrix} h_1^H \\ h_2^H \\ \vdots \\ h_N_a^H \end{bmatrix} \in \mathbb{C}^{N_a \times M_f},
\]

where \( \mathbf{Q}_{\text{DR}_f} \) is a downsampling matrix with downsampling rate \( \text{DR}_f \) and superscript \((\cdot)^H\) denotes conjugate transpose.

To reduce feedback overhead, we exploit the delay sparsity of DL CSI and transform full DL CSI into delay domain by applying discrete Fourier transform (DFT) and conduct a truncation to discard those near-zero elements in the high delay region as follows:

\[
\mathbf{H} = \mathbf{H} \mathbf{F} \begin{bmatrix} 1_{N_f \times N_t} \\ \mathbf{T} \end{bmatrix},
\]

where \( \mathbf{F} \in \mathbb{C}^{M_f \times M_f} \) denotes a DFT matrix and \( M_f \times N_t \) matrix \( \mathbf{T} \) performs delay domain truncation that drop the \( M_f - N_t \) trailing columns of \( \mathbf{H} \mathbf{F} \).

Autoencoder structure has been widely adopted in several CSI feedback frameworks. An encoder UE compresses the DL CSI for uplink feedback and a decoder at gNB recovers the DL truncated time domain CSI according to the feedback from UE. Most research exploits convolutional layers to compress and recover the DL pilot CSI via

\[
\mathbf{q} = f_{\text{en}}(\mathbf{H}),
\]

\[
\hat{\mathbf{H}} = f_{\text{de}}(\mathbf{q}).
\]

III. MULTI-RATE CSI FEEDBACK FRAMEWORK WITH FLEXIBLE NUMBER OF ANTENNA PORTS

There have been notable progresses in terms of recovery performance among the recent autoencoder-based CSI feedback frameworks [8], [11], [17], [18]. Since UEs have limited resources [18], an important consideration is the reduction of complexity and required storage at UE. Unfortunately, naively following the example of autoencoders in image compression leads to the directly input of full DL CSI matrix \( \mathbf{H} \) as an image to deep learning architecture to extract pixel-wise features. Because of the inevitably large input size, fat learning models at both UE and gNB make it highly challenging to effectively reduce the model complexity and storage need.

This raises an question: is it necessary to feed full DL CSI matrix into the model for encoding the spatial features between different APs? The answer may not be yes. That is because most antenna configuration (half-wavelength antenna spacings are usually adopted) just satisfies the spatial Nyquist rate condition\(^1\) (i.e., the spatial sampling rate is at the borderline of avoiding spatial aliasing effect). In other words, there are barely margins or correlation between antennas. Thus, it is unnecessary to extract the correlation between antennas.

We can get more insights from the following preliminary results. Figs. 1 (a) and (b) show the correlation between different antennas and the statistics at different delay taps for different antennas. Obviously, there are weak correlation between antennas and the statistics at different delay taps for different antennas are similar. Thus, it is quite reasonable to use the same model to encode and decode the DL CSI for different APs.

A. Divide-and-Conquer (DC) Framework

Given the preliminary results of Figs. 1 (a) and (b), the inter-antenna independence and similar statistics of delay profile of different antennas herald a simpler CSI feedback framework with a divide-and-conquer manner. In this section, we propose a DC framework which divide a full DL CSI into several pieces and then compress and recover them individually.

We first define a new quantity, base number, to be the dimension of the spatial domain of the new framework input. Assuming base number is \( K \), as compared to the full DL CSI with size of \( N_a \times N_f \), we first divide the full DL CSI matrix into \( N_a/K \) secondary DL CSI matrices with the same size of \( K \times N_f \) given as follows:

\[
\mathbf{H}_i = \mathbf{P}_i \mathbf{H}, \quad \forall i = 1, 2, \ldots, N_a/K,
\]

\[
\mathbf{P}_i = \begin{bmatrix} e_{K(i-1)+1}^T \\ e_{K(i-1)+2}^T \end{bmatrix} \cdots e_{K(i-1)+\text{DR}_f}^T.
\]

\(^1\)As a rule of thumb, the channels of two antennas spaced with one wavelength are nearly independent.
Note that the input size is reduced by a factor of two. A FCDS block consists of $1 \times 7$, $1 \times 5$ and $1 \times 3$ convolutional layers with 2 channels, respectively. Note that the stride lengths are all 1 except for the last horizontal stride length in the last convolutional layer which is set as 2 to reduce the input size by a factor of two. Fig. 3 shows an example of a CSI feedback framework using $S (\equiv 4$ throughout this paper) FCDS blocks for dealing with 4 compression ratios. Specifically, the output of $i$-th block with size of $K \cdot N_f/2^i$ represents the codeword with compression ratio $= 2^i, i = 1, ..., S$. The decoder is designed individually for different compression ratios. For the $i$-th decoder, the codeword is first fed to a $K \cdot N_f$ FC layer, a $1 \times 3$ convolutional layer and activation function after reshaping for initial estimation. A RefineBlock [14] is followed for refinement. RefineBlock is a residual structure and consists of three $1 \times 3$ convolutional layers with 16, 8 and 1 channels and activation functions. Then, it is followed by a $K \cdot N_f$ FC layer for generating the real/imaginary CSI estimate. To further improve the performance, we provide another version of SCEnet, called SCEnet+. We add an additional FC layer at the end of each FCDS block which provides extra non-linearity while maintaining the same output size.

The parameters of the SCEnet are optimized according to the following criterion:

$$
\Omega_{en}, \Omega_{de} = \arg \min \sum_{i}^{D} \sum_{s=4}^{S} W_s \left\| H_i - \hat{H}_{i,s} \right\|_2^2 ,
$$

where $\Omega_{en}, \Omega_{de}$ denote the trainable parameters of encoder $f_{en}$ and decoder $f_{de}$. $D$ is the training data size. The hyper-parameters $\{W_1, W_2, W_3, W_4\}$ refers the setting $\{30/39, 6/39, 2/39, 1/39\}$ in [14].

\[ \hat{H}_{i,1}, \hat{H}_{i,2}, \hat{H}_{i,3}, \hat{H}_{i,4} = f_{de}(f_{en}(H_i)) \]

\[ \Omega_{en}, \Omega_{de} = \arg \min \sum_{i}^{D} \sum_{s=4}^{S} W_s \left\| H_i - \hat{H}_{i,s} \right\|_2^2
\]

IV. EXPERIMENTAL EVALUATIONS

A. Experiment Setup

In our numerical test, we consider both indoor and outdoor cases. Using channel model software, we position a gNB of height equal to 20 m at the center of a circular cell with a radius of 30 m for indoor and 200 m for outdoor environment. We equip the gNB with a $8 \times 4(N_H \times N_V)$ UPA for communication with single antenna UEs. UPA elements have half-wavelength uniform spacing.

For our proposed model and other competing models, we set the number of epochs to 1,000. We use batch size of 200. For our model, we start with learning rate of 0.001 before switching to $5 \times 10^{-4}$ after the 300-th epoch. Using the channel simulator, We generate several indoor and outdoor datasets, each containing 100,000 random channels. 57,143 and 28,571 random channels are for training and validation. The remaining 14,286 channels are test data for performance evaluation. For both indoor and outdoor, we use the QuaDRIGa simulator [19] using the scenario features given in 3GPP.
Fig. 3. DC Framework and SCE Network Architecture. The input is first divided into real and imaginary parts and separated into secondary CSI matrices. These matrices are parallelly fed into the SCE network and recovered after combining the estimates of secondary CSI matrices. Note that, at encoder, after each FCDS block, the total size of input is reduced by a factor of two and those fully convoluted FCDS blocks share parameters. We also provide another version of encoder, named as SCE-Dense encoder, where a FC layer is attached at the end of each FCDS block for performance improvement.

Fig. 4. Pilot placement illustration (Note that the red lines indicate the time-frequency resources to be placed pilot symbols. AF and AD stand for antenna-frequency and antenna-delay domains, respectively).

TR 38.901 Indoor and 3GPP TR 38.901 UMa at 5.1-GHz and 5.3-GHz, and 300 and 330 MHz of UL and DL with LOS paths, respectively. Here, we assume UEs are capable of perfect channel estimation. For each data channel, we consider $N_f = 1024$ subcarriers with $15K$-Hz spacing and place $M_f = 86$ pilots with downsampling ratio $DR_f = 12$ as illustrated in the Fig. 4. We set antenna type to *omni*. We use normalized MSE as the performance metric

$$\frac{1}{D} \sum_{d=1}^{D} \left\| \mathbf{H}_{\hat{d}}^{f} - \mathbf{H}_{d}^{f} \right\|_{F}^{2} / \left\| \mathbf{H}_{d}^{f} \right\|_{F}^{2},$$

where the number $D$ and subscript $d$ denote the total number and index of channel realizations, respectively. $\mathbf{H}_{\hat{d}}^{f} \in \mathbb{C}^{N_a \times N_f}$ denotes the estimated DL CSI after padding zeros back to its full size. $\mathbf{H}_{d}^{f} \in \mathbb{C}^{N_a \times N_f}$ denotes the true DL CSI. For comparison, other than the proposed models, SCEnet and SCEnet-Dense, we also include two multi-rate CSI feedback alternatives as follows:

- **CsiNet-SM**: Fig. 2 (a) shows the general architecture. Note that the same decoders as the proposed models for different compression ratios are adopted but the convolutional filter size is a two-dimensional version (i.e., $(3,3)$, $(5,5)$ and $(7,7)$).

- **CsiNet-PM**: Fig. 2 (b) shows the general architecture. Note that CsiNet-PM is a more compact model than CsiNet-SM but suffers slightly performance degradation in general.

Note that all the above multi-rate CSI frameworks focus on the compression ratios $= 2, 4, 8$ and $16$.

B. SCEnet vs. SCEnet-Dense

Figs. 5 (a) and (b) demonstrate the NMSE performance of the two proposed models at different compression ratios for indoor and outdoor scenarios, respectively. We first discover that a decent performance improvement by adding extra FC layers if the number of RefineBlocks is small. However, we can also find that the additional FC layers do not contribute obvious performance improvement if the number of RefineBlock is 5. Previous works usually adopt FC layer at encoder to provide enough non-linearity. Yet, this result implies that it is not necessary to adopt the FC at encoder if a powerful decoder is utilized. As compared to UEs, gNB usually does not suffer from computational and storage limitations. Therefore, it is more reasonable to use a simple encoder with the help of a complicated decoder.

Fig. 5. NMSE performance at different compression ratios for SCEnet and SCEnet-Dense in indoor and outdoor scenarios.
| CR | Scen. | SANE | CsiNet-SM | CsiNet-PM |
|----|-------|------|-----------|-----------|
|    |       |      | K=1       | K=2       | K=8       | K=32      |
| 2  | Indoor| -33.5| -34.8     | -35.4     | -33.0     | -31.6     | -31.7     |
|    | Outdoor| -17.4| -18.2     | -21.2     | -21.2     | -20.0     | -19.8     |
| 4  | Indoor| -28.6| -31.0     | -30.0     | -27.2     | -26.3     | -26.4     |
|    | Outdoor| -13.8| -14.0     | -16.6     | -18.1     | -18.2     | -15.2     |
| 8  | Indoor| -21.2| -21.3     | -21.2     | -21.6     | -21.2     | -20.2     |
|    | Outdoor| -10.7| -10.9     | -12.6     | -14.0     | -12.8     | -11.7     |
| 16 | Indoor| -12.2| -12.8     | -12.4     | -13.5     | -13.1     | -12.6     |
|    | Outdoor| -8.9 | -8.9      | -9.7      | -11.0     | -10.6     | -9.6      |

Fig. 6. Comparison of (a) numbers of FLOPs and (b) model parameters at various base numbers, K. Note that K = 32 for CsiNet-SM and CsiNet-PM. (K = 1, 2, 4, 8, 16, 32)

C. Comparison with SOTAs and scalability of SCEnet

Table I shows the NMSE performance at different compression ratios and base number (K) for SCEnet, CsiNet-SM and CsiNet-PM in both indoor and scenarios. Fig. 6 shows the FLOP number and model size of encoder at base number (K) for SCEnet, SCEnet-Dense, CsiNet-SM and CsiNet-PM. We can first find that the SCEnet when K = 32 generally outperforms CsiNet-SM and CsiNet-PM while requiring far less FLOP number and storage at UE side. On the other hand, by introducing the DC framework, when we utilize a base number K which can divide Na, we can enjoy lower complexity and less required storage at UEs while suffering slight performance degradation. Moreover, by adopting K = 2, SCEnet can be a universal CSI feedback framework which can be applied to different number of antenna ports (according to the 3GPP specification, 2, 4, 8, 16, 32 are possible antenna port number).

Fig. 7 shows the NMSE performance at different compression ratios, propagation channels, and number of antennas being used for SCEnet. We can find that there are no obvious performance difference when considering arrays with different numbers of antennas. This result demonstrates the scalability of the proposed DC framework.

V. CONCLUSIONS

This work proposes a divide-and-conquer (DC) CSI feedback framework which is universal for different number of antenna ports defined in 3GPP specification, but also lowers the computational complexity and model storage requirement at UE side by allowing UE parallel feedbacks segmented DL CSI with respect to the low correlation between antennas (or antenna ports). The framework consists of a multi-rate successive convolutional encoder (SCE) without any FC layer which usually is the major reason causing a fat model. In numerical results, the proposed framework generally outperform than the SOTAs, CsiNet-SM and CsiNet-PM, while requiring much lower computation and storage requirements for UEs, which usually have tight resource budgets.

VI. ACKNOWLEDGEMENT

The authors would like to acknowledge Mason del Rosario for his useful discussions which helped the authors better understand of pilot placement and channel truncation.

REFERENCES

[1] C.-H. Lin, S.-C. Lin, and E. Blasch, “TULVCAN: Terahertz Ultra-broadband Learning Vehicular Channel-aware Networking,” in IEEE INFOCOM workshop, May 2021, pp. 1–6.
[2] C. Wen, W. Shih, and S. Jin, “Deep Learning for Massive MIMO CSI Feedback,” IEEE Wirel. Commun. Lett., vol. 7, no. 5, pp. 748–751, 2018.
[3] J. Guo et al., “Convolutional Neural Network-Based Multiple-Rate Compressive Sensing for Massive MIMO CSI Feedback: Design, Simulation, and Analysis,” IEEE Trans. Wirel. Commun., vol. 19, no. 4, pp. 2827–2840, 2020.
[4] Z. Lu, J. Wang, and J. Song, “Multi-resolution CSI Feedback with Deep Learning in Massive MIMO System,” in IEEE Intern. Conf. Communications (ICC), 2020, pp. 1–6.
[5] Q. Yang, M. B. Mashjadi, and D. Gündüz, “Deep Convolutional Compression For Massive MIMO CSI Feedback,” in IEEE Intern. Workshop Mach. Learning for Signal Process. (MLSP), 2019, pp. 1–6.
[6] S. Ji and M. Li, “CLNet: Complex Input Lightweight Neural Network Designed for Massive MIMO CSI Feedback,” IEEE Wirel. Commun. Lett., vol. 10, no. 10, pp. 2318–2322, 2021.
[7] Z. Liu, M. Rosario, and Z. Ding, “A Markovian Model-Driven Deep Learning Framework for Massive MIMO CSI Feedback,” IEEE Trans. Wirel. Commun., 2021, early access.
[8] J. Guo et al., “DL-based CSI Feedback and Cooperative Recovery in Massive MIMO,” arXiv preprint arXiv:2003.03303, 2020.
[9] Z. Liu, L. Zhang, and Z. Ding, “An Efficient Deep Learning Framework for Low Rate Massive MIMO CSI Reporting,” IEEE Trans. Wirel. Commun., vol. 68, no. 8, pp. 4761–4772, 2020.
[10] ——, “Exploiting Bi-Directional Channel Reciprocity in Deep Learning for Low Rate Massive MIMO CSI Feedback,” IEEE Wirel. Commun. Lett., vol. 8, no. 3, pp. 889–892, 2019.

[11] Y.-C. Lin, Z. Liu, T.-S. Lee, and Z. Ding, “Deep Learning Phase Compression for MIMO CSI Feedback by Exploiting FDD Channel Reciprocity,” IEEE Wireless Commun. Lett., vol. 10, no. 10, pp. 2200–2204, 2021.

[12] Y. Ding and B. D. Rao, “Dictionary Learning-based Sparse Channel Representation and Estimation for FDD Massive MIMO Systems,” IEEE Trans. Wirel. Commun., vol. 17, no. 8, pp. 5437–5451, 2018.

[13] X. Zhang, L. Zhong, and A. Sabharwal, “Directional Training for FDD Massive MIMO,” IEEE Trans. Wirel. Commun., vol. 17, no. 8, pp. 5183–5197, 2018.

[14] J. Guo, C.-K. Wen, S. Jin, and G. Y. Li, “Convolutional Neural Network-Based Multiple-Rate Compressive Sensing for Massive MIMO CSI Feedback: Design, Simulation, and Analysis,” IEEE Trans. Wirel. Commun., vol. 19, no. 4, pp. 2827–2840, 2020.

[15] H. Tang, J. Guo, M. Matthaiou, C.-K. Wen, and S. Jin, “Knowledge-distillation-aided Lightweight Neural Network for Massive MIMO CSI Feedback,” in IEEE Veh. Technol. Conf. (VTC2021-Fall), 2021, pp. 1–5.

[16] J. Guo, J. Wang, C.-K. Wen, S. Jin, and G. Y. Li, “Compression and Acceleration of Neural Networks for Communications,” IEEE Trans. Wirel. Commun., vol. 27, no. 4, pp. 110–117, 2020.

[17] J. Guo, C.-K. Wen, and S. Jin, “CAnet: Uplink-aided Downlink Channel Acquisition in FDD Massive MIMO using Deep Learning,” IEEE Trans. Commun., 2021, early access.

[18] Y. Sun, W. Xu, L. Liang, N. Wang, G. Y. Li, and X. You, “A Lightweight Deep Network for Efficient CSI Feedback in Massive MIMO Systems,” IEEE Wirel. Commun. Lett., vol. 10, no. 8, pp. 1840–1844, 2021.

[19] S. Jaeckel et al., “QuaDRiGa: A 3-D Multi-Cell Channel Model with Time Evolution for Enabling Virtual Field Trials,” IEEE Trans. Antennas and Propag., vol. 62, no. 6, pp. 3242–3256, 2014.