Deep Autoencoder based CSI Feedback with Feedback Errors and Feedback Delay in FDD Massive MIMO Systems

Youngrok Jang, Gyuyeol Kong, Minchae Jung, Sooyong Choi, *Member, IEEE*, and Il-Min Kim, *Senior Member, IEEE*

***Abstract*—In this paper, we study the channel state information (CSI) feedback based on the deep autoencoder (AE) considering the feedback errors and feedback delay in the frequency division duplex (FDD) massive multiple-input multiple-output (MIMO) system. We construct the deep AE by modeling the CSI feedback process, which involves feedback transmission errors and delays. The deep AE is trained by setting the delayed version of the downlink channel as the desired output. The proposed scheme reduces the impact of the feedback errors and feedback delay. Simulation results demonstrate that the proposed scheme achieves better performance than other comparable schemes.**

***Index Terms*—Autoencoder, CSI feedback, FDD massive MIMO, feedback delay, feedback errors.**

I. Introduction

Recently, the massive multiple-input multiple-output (MIMO) system has emerged as a major enabler for the next generation wireless communication systems [1]. To achieve the potential gains promised by the massive MIMO systems, the channel state information (CSI) must be known at the transmitter [1]. In practice, however, it is difficult for the transmitter to obtain the accurate CSI in the frequency division duplex (FDD) system due to a large amount of feedback overhead, which grows in proportional to the dimension of the massive MIMO channel [1]–[3].

To reduce the feedback overhead for the FDD massive MIMO system, several works have investigated the issue of reduction of CSI feedback dimensionality. In [2], the principal component analysis (PCA) based dimensionality reduction schemes were studied, which utilized only the eigenvectors of the channel correlation matrix corresponding to the dominant eigenvalues. In [3], another CSI feedback scheme was proposed based on the structure of convolutional neural network (CNN), which was more effective in reducing the dimensionality compared to other compressed sensing based algorithms.

This work was partly supported by Natural Sciences and Engineering Research Council of Canada (NSERC), Institute for Information & communications Technology Promotion (IITP) grant funded by the Korea government (MSIT) (No. 2016-0-00181, Development on the core technologies of transmission, modulation and coding with low-power and low-complexity for massive connectivity in the IoT environment), the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2017R1A2B3006025), and the International Research & Development Program of the NRF funded by the MSIT (No. 2017K1A3A1A12030076).

Y. Jang, G. Kong, M. Jung, and S. Choi (corresponding author) are with the School of Electrical and Electronic Engineering, Yonsei University, Seoul 03722, Korea (e-mail: csyong@yonsei.ac.kr).

I.-M. Kim is with the Department of Electrical and Computer Engineering, Queen's University, Kingston, ON K7L 3N6, Canada (e-mail: ilmin.kim@queensu.ca).

In the feedback channel of any practical system, however, there exist both feedback errors and feedback delay, which degrade the performance of the CSI feedback. In particular, the compressed version of the CSI feedback information is even more susceptible to such feedback impairments, compared to the uncompressed version of the CSI. To the best of our knowledge, the important problem of compressing and transmitting the CSI feedback information in the presence of both impairments of the feedback errors and feedback delay has not been studied in the literature for the massive MIMO systems. This motivated our work.

In this paper, we construct a new and effective CSI feedback scheme based on the deep AE [4] in the presence of the feedback errors and feedback delay in the FDD massive MIMO system. In our proposed scheme, the proposed AE is trained taking into account the feedback errors, which arise from the quantization of CSI and its transmission over the noisy uplink (UL) feedback channel. Furthermore, the desired output of the proposed AE is set to the delayed version of the downlink (DL) channel in order to offset the feedback delay. Once trained, the resulting CSI generated by the proposed AE is tolerant to both the feedback errors and feedback delay.

II. System Model

A. Channel Model

We consider the massive MIMO system composed of a base station (BS) with N_t antennas and a user equipment (UE) with N_r antennas. Following the spatially correlated physical channel model as in [5], the DL channel $\mathbf{H}_{d,n}$ at the n-th time slot is given by

$$\mathbf{H}_{d,n} = \sqrt{\frac{N_t N_r}{L}} \mathbf{A}_{d,r} \mathbf{D}_{d,n} \mathbf{A}_{d,t}^H, \tag{1}$$

where L is the number of propagation paths and $\mathbf{D}_{d,n} = \mathrm{diag}([\beta_{d,n}^{(1)}, ..., \beta_{d,n}^{(L)}]) \in \mathbb{C}^{L \times L}$ is the propagation path gain matrix at the n-th time slot, where $\beta_{d,n}^{(l)} \sim \mathcal{CN}(0, 1)$ is the l-th path gain. Also, $\mathbf{A}_{d,t} = \frac{1}{\sqrt{N_t}}[\mathbf{a}_{d,t}(\theta_1), ..., \mathbf{a}_{d,t}(\theta_L)] \in \mathbb{C}^{N_t \times L}$ and $\mathbf{A}_{d,r} = \frac{1}{\sqrt{N_r}}[\mathbf{a}_{d,r}(\phi_1), ..., \mathbf{a}_{d,r}(\phi_L)] \in \mathbb{C}^{N_r \times L}$ denote the array response matrices at the BS and UE, respectively. Assuming a uniform linear array (ULA), the array response vectors at the BS and UE are given by $\mathbf{a}_{d,t}(\theta_l) = [1, e^{-j2\pi \frac{d_s}{\lambda_d} \sin\theta_l}, ..., e^{-j2\pi \frac{d_s}{\lambda_d}(N_t - 1)\sin\theta_l}]^T$ and $\mathbf{a}_{d,r}(\phi_l) = [1, e^{-j2\pi \frac{d_s}{\lambda_d} \sin\phi_l}, ..., e^{-j2\pi \frac{d_s}{\lambda_d}(N_r - 1)\sin\phi_l}]^T$, respectively, where θ_l and $\phi_l \in [-\frac{\pi}{2}, \frac{\pi}{2}]$ are the angles of departure and arrival of the l-th propagation path, respectively. Also, d_s is the antenna spacing and $\lambda_d = c/f_d$ is the DL carrier wavelength given the DL carrier frequency f_d and the speed of light c. In addition, considering the Gauss-Markov model as in [6], the channel $\mathbf{H}_{d,n+n_{\text{fb}}}$ with a delay of n_{fb} time slots is determined from $\mathbf{H}_{d,n}$ as

$$\mathbf{H}_{d,n+n_{\text{fb}}} = \eta^{n_{\text{fb}}}\mathbf{H}_{d,n} + \sum_{k=1}^{n_{\text{fb}}} \eta^{n_{\text{fb}}\text{-}k}\sqrt{\frac{(1-\eta^2)N_t N_r}{L}}\mathbf{A}_{d,r}\mathbf{D}_{d,n+k}\mathbf{A}_{d,t}^H, \tag{2}$$

where $\eta = J_0(2\pi f_D T)$ is a temporal correlation coefficient, which follows the Jakes' model [6], where $J_0(\cdot)$ is the zeroth order Bessel function of the first kind, T is the channel block length, and $f_D = v f_d / c$ is the maximum Doppler frequency, where v denotes the speed of the UE. Likewise, the UL channel is also formulated from (1) and (2) with the substitution of $\mathbf{a}_{u,t}(\phi_l)$ and $\mathbf{a}_{u,r}(\theta_l)$, which are given by $[\mathbf{a}_{u,t}(\phi_l)]_k = \{[\mathbf{a}_{d,r}(\phi_l)]_k\}^{\lambda_d/\lambda_u}$ and $[\mathbf{a}_{u,r}(\theta_l)]_k = \{[\mathbf{a}_{d,t}(\theta_l)]_k\}^{\lambda_d/\lambda_u}$, where $[\mathbf{x}]_k$ denotes the k-th element of an input vector $\mathbf{x}$ and $\lambda_u = c/f_u$ denotes the UL carrier wavelength for the UL carrier frequency f_u [7].

B. CSI Feedback Process

Based on (1) and (2), the CSI of the DL channel at the n-th time slot can be generated and fed back from the UE to the BS. Due to the large dimension of the DL channel, the dimensionality reduction should be performed to compress the CSI for feedback. Let $f(\cdot)$ denote an arbitrary dimensionality reduction function. Then, the reduced dimensional DL CSI at the n-th time slot is given by $\bar{\mathbf{H}}_{d,n} = f(\mathbf{H}_{d,n})$, which is composed of $N \leq N_t N_r$ complex elements. Using $\bar{\mathbf{H}}_{d,n}$, $\bar{\mathbf{x}} = \text{vec}(Q(\bar{\mathbf{H}}_{d,n})) \in \mathbb{C}^{N\times 1}$ is obtained, where $\text{vec}(\cdot)$ denotes the vectorization to stack the columns of the input matrix and $Q(\mathbf{A})$ denotes an element-wise scalar quantization in which each element is given by

$$[Q(\mathbf{A})]_{ij} = \underset{a\in\mathcal{A}}{\arg\min}\, |[\mathbf{A}]_{ij} - a|^2, \tag{3}$$

where $[\cdot]_{ij}$ denotes the element at the i-th row and the j-th column of the input matrix. Also, $\mathcal{A}$ denotes a pre-defined complex-valued constellation point set. Since only $r \leq \min(N_r, N_t)$ signals are transmitted per each time slot, at least $n_{\text{fb}} = \lceil \frac{N}{r} \rceil$ time slots are needed to feed back $\bar{\mathbf{x}}$, where $\lceil\cdot\rceil$ denotes the ceiling function. Hence, $\bar{\mathbf{x}}$ is divided into $\bar{\mathbf{x}} = [\bar{\mathbf{x}}_1^T, ..., \bar{\mathbf{x}}_{n_{\text{fb}}}^T]^T$, where $\bar{\mathbf{x}}_k \in \mathbb{C}^{r\times 1}$ denotes the transmit signal vector at the k-th time slot with $k \in \{1, ..., n_{\text{fb}}\}$. In the UL, the received signal at the k-th time slot is given by

$$\mathbf{y}_{u,k} = \mathbf{H}_{u,k}\bar{\mathbf{x}}_k + \mathbf{z}_{u,k}, \tag{4}$$

where $\mathbf{H}_{u,k} \in \mathbb{C}^{N_t \times N_r}$ denotes the UL channel at the k-th time slot and $\mathbf{z}_{u,k} \sim \mathcal{CN}(\mathbf{0}, \sigma_u^2 \mathbf{I}_{N_t}) \in \mathbb{C}^{N_t\times 1}$ is the additive white Gaussian noise (AWGN) vector at the BS. Using the minimum mean square error (MMSE) detector $\mathbf{G}_{u,k} = (\mathbf{H}_{u,k}^H\mathbf{H}_{u,k} + \sigma_u^2\mathbf{I}_{N_r})\mathbf{H}_{u,k}^H$, the detected signal vector at the k-th time slot is obtained by $\hat{\mathbf{x}}_k = Q(\mathbf{G}_{u,k}\mathbf{y}_{u,k})$. Using the n_{fb} received feedback signals, the detected version of $\bar{\mathbf{H}}_{d,n}$ is obtained by $\hat{\mathbf{H}}_{d,n} = \text{vec}^{-1}([\hat{\mathbf{x}}_1^T, ..., \hat{\mathbf{x}}_{n_{\text{fb}}}^T]^T)$, where $\text{vec}^{-1}(\cdot)$ denotes the inverse operation of $\text{vec}(\cdot)$. Let $g(\cdot)$ denote any reconstruction function. Finally, the reconstructed DL channel is obtained by $\tilde{\mathbf{H}}_{d,n} = g(\hat{\mathbf{H}}_{d,n})$.

In the CSI feedback process, the CSI can be affected by two impairments: 1) the feedback errors due to the quantization and the transmission over the noisy UL channel, and 2) the feedback delay of length n_{fb} time slots to feed back the CSI.

III. Proposed Scheme

In this section, we present the proposed CSI feedback scheme based on the deep AE structue [4]. The purpose of the proposed scheme is to address the impacts of both the feedback errors and feedback delay as well as to effectively perform the dimensionality reduction and reconstruction.

The network architecture of the deep AE utilized for the proposed scheme is shown in Fig. 1. Note that the actual end-to-end CSI feedback process is modeled by the deep AE. Since the weights, biases, and input data are represented by the real numbers, the real-valued representation[1] is used for all of the complex variables by using the superscript "r" as shown in Fig. 1. The dimensionality reduction and reconstruction of the proposed scheme are performed by the encoder and decoder, respectively, which are represented as $\bar{\mathbf{h}}_{d,n}^r = f_{\text{enc}}(\mathbf{H}_{d,n}^r; \Psi_{\text{enc}})$ and $\tilde{\mathbf{H}}_{d,n}^r = g_{\text{dec}}(\hat{\mathbf{h}}_{d,n}^r; \Psi_{\text{dec}})$ as shown in Fig. 1, where $\mathbf{H}_{d,n}^r \in \mathbb{R}^{2N_r\times 2N_t}$ denotes the input data of the DL channel at the n-th time slot, $\hat{\mathbf{h}}_{d,n}^r \in \mathbb{R}^{2N\times 1}$ denotes the detected version of $\bar{\mathbf{h}}_{d,n}^r$, and Ψ_{enc} (resp. Ψ_{dec}) denotes the network parameter set of the encoder (resp. decoder) including the weights and biases. The input data $\mathbf{H}_{d,n}^r$ is reshaped into $\mathbf{h}_{d,n}^r = \text{vec}(\mathbf{H}_{d,n}^r) \in \mathbb{R}^{2N_t N_r\times 1}$ at the input of the encoder. Then, there are two fully connected (FC) hidden layers with an exponential linear unit (ELU) [8] activation function[2] for both the encoder and decoder, which is given by $\sigma(x) = x$ if $x \geq 0$, and $\sigma(x) = e^x - 1$ otherwise. Also, there is a FC output layer with a linear activation function. For the generating and detecting feedback signals, $Q^r(\cdot)$ performs like (3) with a predefined real-valued constellation point set. Since the generated feedback signal based on (5) is given as $\bar{\mathbf{x}}^r = Q^r(\bar{\mathbf{h}}_{d,n}^r) = [(\bar{\mathbf{x}}_1^r)^T, ..., (\bar{\mathbf{x}}_{n_{\text{fb}}}^r)^T]^T \in \mathbb{R}^{2N\times 1}$, where $\bar{\mathbf{x}}_k^r \in \mathbb{R}^{2r\times 1}$ denotes the transmit signal vector at the k-th time slot, the feedback process still requires $n_{\text{fb}} = \lceil \frac{2N}{2r} \rceil$ time slots. At the output of the decoder, the output data $\tilde{\mathbf{h}}_{d,n}^r$ is finally reshaped into $\tilde{\mathbf{H}}_{d,n}^r = \text{vec}^{-1}(\tilde{\mathbf{h}}_{d,n}^r)$.

Note that since the end-to-end CSI feedback process is modeled by the AE, which can be considered as the corruption processes as in [9], the deep AE becomes torelant to the feedback errors. Moreover, the capability of predicting the time varying channel of the proposed scheme can be explained by the principle of the denoising AE [9]. Based on (2), the

[1] $\mathbf{x} \in \mathbb{C}^{a\times 1}$ and $\mathbf{X} \in \mathbb{C}^{b\times c}$ can be represented by

$$\mathbf{x}^r = \begin{bmatrix} \Re\{\mathbf{x}\} \\ \Im\{\mathbf{x}\} \end{bmatrix} \in \mathbb{R}^{2a\times 1},\ \mathbf{X}^r = \begin{bmatrix} \Re\{\mathbf{X}\} & -\Im\{\mathbf{X}\} \\ \Im\{\mathbf{X}\} & \Re\{\mathbf{X}\} \end{bmatrix} \in \mathbb{R}^{2b\times 2c}, \tag{5}$$

where $\Re\{\cdot\}$ and $\Im\{\cdot\}$ denote the real and imaginary parts, respectively.

[2] The ELU is slightly different from the rectified linear unit (ReLU) in the sense that $\sigma(x)$ is non-zero for $x < 0$. It is known that the ELU is working well even without additionally using the batch normalization [8].

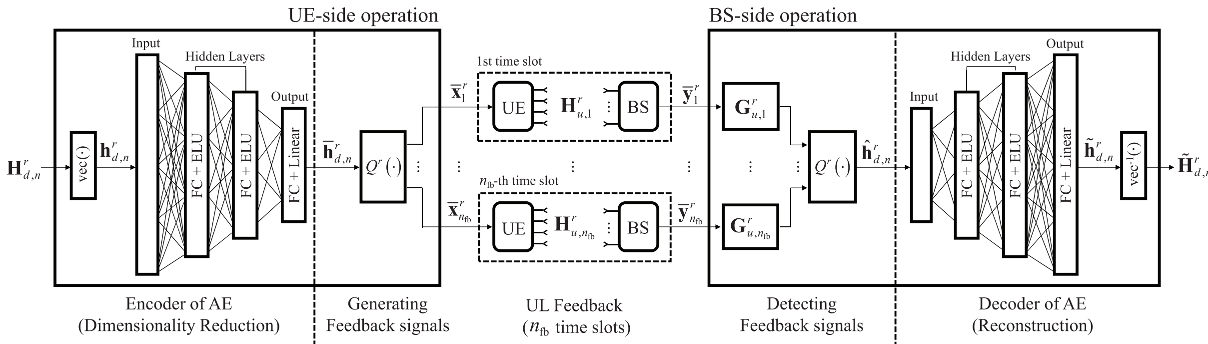


Fig. 1. Network architecture of the deep AE utilized for the proposed scheme

relationship between n_{fb} delayed channel $\mathbf{H}_{d,n+n_{\text{fb}}}$ and the current channel $\mathbf{H}_{d,n}$ can be rewritten as

$$\mathbf{H}_{d,n} = \frac{1}{\eta^{n_{\text{fb}}}} \mathbf{H}_{d,n+n_{\text{fb}}} + \mathbf{N}_{\text{del}}, \tag{6}$$

where $\mathbf{N}_{\text{del}}$[3] represents the noise due to the feedback delay. Then, $\mathbf{H}_{d,n}$ represents the noisy version of the desired input, which is used as the actual input to the proposed AE. Also, $\mathbf{H}_{d,n+n_{\text{fb}}}$ represents the desired output, which should be reconstructed at the output of the proposed AE. Hence, the loss function of the proposed AE is set to the mean square error (MSE) as follows:

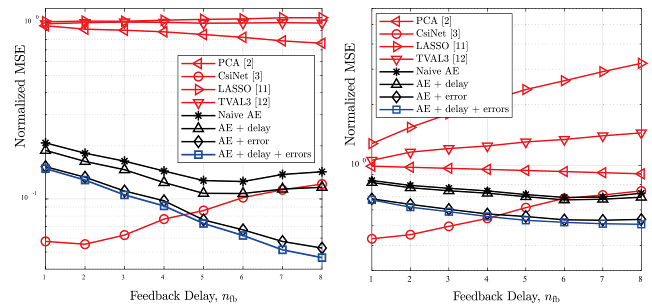
$$L\left(\Psi_{\text{enc}}, \Psi_{\text{dec}}\right) = ||\tilde{\mathbf{H}}^r_{d,n} - \mathbf{H}^r_{d,n+n_{\text{fb}}}||^2_F, \tag{7}$$

where $\|\cdot\|_F$ denotes the Frobenius norm. By minimizing (7), the weights and biases included in Ψ_{enc} and Ψ_{dec} are updated by an optimizer such as ADAM[4] [10].

IV. Simulation Results

In this section, we compare the proposed scheme to seven comparable state-of-the-art schemes: 1) PCA [2]; 2) CsiNet [3]; 3) LASSO [11]; 4) TVAL3 [12]; 5) the naive AE without considering the feedback errors and feedback delay; 6) the AE considering the delay only; and 7) the AE considering the error only. Recall that our proposed scheme is the AE considering both the feedback errors and feedback delay. Note that the performance comparison is conducted by using the same CSI feedback resources according to n_{fb} for all schemes.

We use the widely-adopted Gauss-Markov channel model based on (1) and (2) for all simulations except in Fig. 2(b). We set $f_d = 2.655$ GHz, $f_u = 2.535$ GHz [13], $N_t = 64$, $N_r = 4$, $d_s = \lambda_u/2$ [7], $L = 32$, $T = 0.5$ ms, $r = 4$, and $v = 10$ km/h. Also, we use 64 quadrature amplitude modulation (QAM) for $Q(\cdot)$. The training set and the test set contain 100,000 and 20,000 normalized data samples which are scaled into $[-1, 1]$, respectively. The epochs, batch size, and learning rate are set



(a) Gauss-Markov channel model

(b) COST 2100 channel model [7] for the 300 MHz outdoor scenario with 10 m/s UE speed

Fig. 2. Normalized MSE versus n_{fb} when UL SNR is set to 10 dB

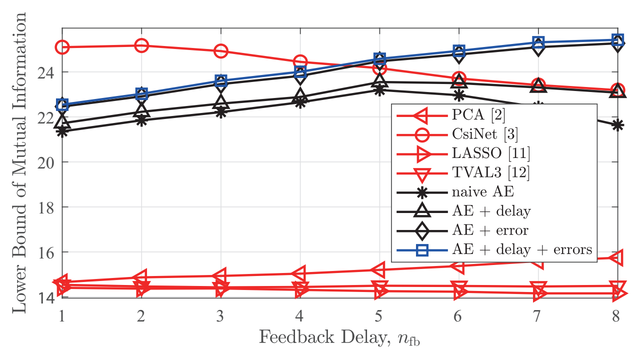


Fig. 3. Lower bound of mutual information versus n_{fb} when both of DL and UL SNR are set to 10 dB

to 5000, 2000, and 10^{-4}, respectively. In the encoder (resp. decoder), there are $2N_tN_r$, 256, 128, and $2N$ (resp. $2N$, 128, 256, and $2N_tN_r$) neurons in the input, the first hidden layer, the second hidden layer, and the output layer, respectively. The weights and biases are initialized by the Xavier initialization [14] and by the standard normal distribution, respectively.

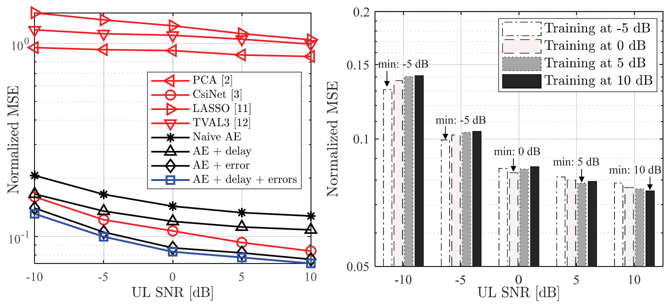
In Fig. 2, the normalized MSE values of the proposed and other comparable schemes are shown versus the feedback delay n_{fb}[5] when the UL SNR is set to 10 dB. In Fig. 2(b), we repeat the simulations using the COST 2100 channel model [15], in which the temporal correlation does not follow the Gauss-Markov model. The normalized MSE is given by $\mathbb{E}[||\tilde{\mathbf{H}} - \mathbf{H}||^2_F / \|\mathbf{H}\|^2_F]$, where $\mathbf{H}$ and $\tilde{\mathbf{H}}$ denote the desired and reconstructed channel matrices, respectively. In Fig. 3, the

[3]When there exists channel estimation error (i.e., $\tilde{\mathbf{H}}_{d,n} = \mathbf{H}_{d,n} + \mathbf{N}_{\text{est}}$, where $\mathbf{N}_{\text{est}}$ represents the channel estimation error), eq. (6) can be modified as $\tilde{\mathbf{H}}_{d,n} = \frac{1}{\eta^{n_{\text{fb}}}}\mathbf{H}_{d,n+n_{\text{fb}}} + \mathbf{N}_{\text{del}} + \mathbf{N}_{\text{est}}$. Then, the erroneously estimated current channel $\tilde{\mathbf{H}}_{d,n}$ represents the noisy version of the desired input, which is used as the actual input to the proposed AE.

[4]Note that in the real scenario, the proposed AE should be trained separately for various practical and common scenarios, e.g., different UE speeds, typical indoor/outdoor propagation environments, different number of antennas at the UE, different operating SNR values, etc.

[5]The maximum value of n_{fb} is set to 8 and T=0.5 ms. Then, the maximum feedback delay is equal to 4 ms , which is the same as the feedback delay of the periodic CSI reporting in the Long Term Evolution (LTE) standard [13].

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/LWC.2019.2895039, IEEE Wireless Communications Letters



(a) For each operating UL SNR, the deep AE of the proposed scheme is trained at the corresponding (i.e., the same) UL SNR value.

(b) The proposed scheme is robust to the change of the operating UL SNR.

Fig. 4. Normalized MSE versus UL SNR when $n_{\text{fb}} = 5$

lower bounds of mutual information of the proposed and other comparable schemes are shown versus the feedback delay n_{fb} when both of DL and UL SNRs are set to 10 dB. The lower bounds of mutual information are formulated as in [16, eq. (2)] with a given estimated channel knowledge. From Figs. 2 and 3, when the delay is small, the CsiNet [3] is better than the proposed scheme, because the CsiNet can extract the spatial features well based on the CNN. However, since the CsiNet does not have any capability of mitigating the effect of the feedback delay[6], the proposed scheme performs the best when the feedback delay is not small.

In Figs. 2 and 3, it is interesting and important to see that n_{fb} has both the positive and negative impacts on the (MSE and capacity) performance. First, increasing n_{fb} has a positive impact, since longer n_{fb} means that more resources are used for the CSI feedback (i.e., the dimension of the CSI feedback is larger). On the other hand, increasing n_{fb} has also a negative impact, since longer n_{fb} means that the feedback delay is longer and more feedback signals are exposed to the feedback errors. For these reasons, when n_{fb} increases up to a certain value, the performances of both naive AE and AE+delay improve. However, when n_{fb} further increases beyond that value, the performance starts to deteriorate, which means that the negative impact starts to outweigh the positive impact. In sharp contrast to this result, the performance of the proposed scheme improves continuously and monotonically. This implies that both issues of the feedback errors and feedback delay are well addressed by the proposed scheme, and the positive impact always outweighs the negative one in the proposed scheme.

In Fig. 4, the normalized MSE is shown versus the UL SNR when $n_{\text{fb}} = 5$. From Fig. 4(a), it can be seen that the proposed scheme considerably outperforms other comparable schemes. In Fig. 4(a), to obtain the MSE value of the proposed scheme for each operating UL SNR, the deep AE is trained at the corresponding (i.e., the same) UL SNR value. In practical systems, however, it might be difficult to *re-train* the deep AE whenever the operating UL SNR changes. Fortunately, it turns out that the proposed scheme is robust to the change of the operating UL SNR. Fig. 4(b) shows the normalized MSE value for each operating UL SNR when the deep AE is trained at four different UL SNRs. One can see that the proposed scheme performs almost the same even when the deep AE is trained at an UL SNR that is different from the actual operating UL SNR. This makes the proposed scheme more suitable for practical applications.

[6]The time varying version of [3] was proposed in [17] based on the recurrent convolutional neural network (RCNN), which appears to outperform our proposed scheme. However, the computational complexity and the required memory are much higher due to adoption of the recurrent neural network (RNN), e.g., 200 times of computation and 180 times of required memory.

V. Conclusion

We studied the deep AE based CSI feedback scheme considering the feedback errors and feedback delay in the FDD massive MIMO system. By modeling the CSI feedback process by the deep AE and setting the desired output of the deep AE to the delayed version of the DL channel, the proposed scheme became torelant to the impact of both the feedback errors and feedback delay. We demonstrated that the proposed scheme outperformed other comparable schemes in terms of the normalized MSE and the mutual information.

References

[1] F. Rusek *et al.*, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, Jan. 2013.

[2] Z. Jiang, A. F. Molisch, G. Caire, and Z. Niu, "Achievable rates of FDD massive MIMO systems with spatial channel correlation," *IEEE Trans. Wireless Commun.*, vol. 14, no. 5, pp. 2868–2882, May 2015.

[3] C. K. Wen, W. T. Shih, and S. Jin, "Deep learning for massive MIMO CSI feedback," *IEEE Wireless Commun. Lett.*, vol. 7, no. 5, pp. 748–751, Mar. 2018.

[4] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.

[5] V. Raghavan and A. M. Sayeed, "Sublinear capacity scaling laws for sparse MIMO channels," *IEEE Trans. Info. Theory*, vol. 57, no. 1, pp. 345–364, Jan. 2011.

[6] M. Vu and A. Paulraj, "On the capacity of MIMO wireless channels with dynamic CSIT," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 7, pp. 1269–1283, Sept. 2007.

[7] M. B. Khalilsarai, S. Haghighatshoar, X. Yi, and G. Caire, "FDD massive MIMO via UL/DL channel covariance extrapolation and active channel sparsification," *arXiv preprint arXiv: 1803.05754*, 2018.

[8] D.-A. Clevert, T. Unterthiner, and S. Hochreiter, "Fast and accurate deep network learning by exponential linear units (ELUs)," *arXiv preprint arXiv: 1511.07289*, 2015.

[9] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P. Manzagol, "Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion," *J. Mach. Learn. Res.*, vol. 11, pp. 3371–3408, Dec. 2010.

[10] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv: 1412.6980*, 2017.

[11] I. Daubechies, M. Defrise, and C. D. Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Commun. Pure Appl. Math.*, vol. 57, no. 11, pp. 1413–1457, 2004.

[12] C. Li, W. Yin, and Y. Zhang. (2009). *TVAL3: TV Minimization by Augmented Lagrangian and Alternating Direction Algorithms.* [Online]. Available: http://www.caam.rice.edu/∼optimization/L1/TVAL3/

[13] *ETSI, LTE; Evolved Universal Terrestrial Radio Access (E-UTRA); User Equipment (UE) Radio Transmission and Reception*, 3GPP TS 36.101, ver. 14.5.0, Release 14, 2017.

[14] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proc. Conf. Artificial Intelligence and Statistics*, 2010, pp. 249–256.

[15] L. Liu, C. Oestges, J. Poutanen, K. Haneda, P. Vainikainen, F. Quitin, F. Tufvesson, and P. Doncker, "The COST 2100 MIMO channel model," *IEEE Wireless Commun.*, vol. 19, no. 6, pp. 92–99, Dec. 2012.

[16] T. Yoo and A. J. Goldsmith, "Capacity and power allocation for fading MIMO channels with channel estimation error," *IEEE Trans. Inf. Theory*, vol. 52, no. 5, pp. 2203–2214, May 2006.

[17] T. Wang, C. K. Wen, S. Jin, and G. Y. Li, "Deep learning-based CSI feedback approach for time-varying massive MIMO channels," *IEEE Wireless Commun. Lett., early access*, 2018.