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.2019.2912378, IEEE Wireless Communications Letters

Deep Learning Based Channel Estimation for Massive MIMO Systems

Chang-Jae Chun, Jae-Mo Kang, *Member, IEEE*, and Il-Min Kim, *Senior Member, IEEE*

***Abstract*—In this paper, we propose a deep learning (DL) based channel estimation scheme for the massive multiple-input multiple-output (MIMO) system. Unlike existing studies, we develop the channel estimation scheme for the case that the pilot length is smaller than the number of transmit antennas. The proposed scheme takes a two-stage estimation process: a DL-based pilot-aided channel estimation and a DL-based data-aided channel estimation. In the first stage, the pilot itself and the channel estimator are jointly designed by using both a two-layer neural network (TNN) and a deep neural network (DNN). In the second stage, the accuracy of channel estimation is further enhanced by using another DNN in an iterative manner. The simulation results demonstrate that the proposed channel estimation scheme has much better performance than the conventional channel estimation scheme. We also derive an useful insight into the optimal pilot length given the number of transmit antennas.**

***Index Terms*—Channel estimation, deep learning, massive MIMO system.**

I. INTRODUCTION

Multiple-input multiple-output (MIMO) with large-scale antenna arrays, so-called the massive MIMO, is one of the most promising techniques to increase the data rate and to maintain the high communication reliability for future wireless systems [1], [2]. In the massive MIMO system, a large scale antenna array is deployed typically at the base station (BS) to provide a considerable antenna gain, and this antenna gain highly depends on the channel estimation accuracy. Thus, obtaining accurate channel estimation is very important to ensure such benefits of the massive MIMO technique.

In the literature, the issue of channel estimation has been studied for the massive MIMO systems, typically based on the linear minimum mean square error (LMMSE) method, e.g., [3]–[5]. However, the common assumption in all the literature including [3]–[5] is that the pilot length L_s is equal to or larger than the number of transmit antennas N_t, which is (very) large in the downlink of massive MIMO. Without this assumption of $L_s \geq N_t$, the channel estimation performance is substantially degraded, as demonstrated in [6] for the LMMSE channel estimator. In the massive MIMO system, however, it is hard to justify the assumption of $L_s \geq N_t$ for three main reasons. First, to ensure $L_s \geq N_t$, substantial amount of time resource should be used for pilot transmission, which results in much reduced resource for data transmission, leading to (very) low spectral efficiency. Second, the computational complexity required for channel estimation also grows as L_s increases. Last, but not least, ensuring $L_s \geq N_t$ might be even impossible at all, because L_s cannot be larger than the channel coherence interval, which is usually uncontrollable. Therefore, in the massive MIMO system, developing a channel estimation scheme for $L_s < N_t$ is one of the most important, yet very challenging, issues.

In spite of the importance of channel estimation for $L_s < N_t$ in the massive MIMO system, there was no such study in the literature. Unfortunately, developing an effective channel estimator for $L_s < N_t$ is a very difficult problem for the following reasons. In all the existing works assuming $L_s \geq N_t$, the pilot matrix $\boldsymbol{S}$ was simply designed such that its row vectors were orthogonal, which was needed to avoid the interference among the pilot sequences transmitted from different antennas. When $L_s < N_t$, however, it is not possible to design the pilot matrix $\boldsymbol{S}$ in that way.[1] This means that the channel estimation problem has to be jointly solved with the additional problem of designing the pilot $\boldsymbol{S}$, which turns out be very challenging because the optimal (or high performing) structure of $\boldsymbol{S}$ is unknown when $L_s < N_t$. Another issue is that, when $L_s < N_t$, there is no guarantee that the LMMSE is an optimal channel estimator.[2] This means that, in the design of channel estimator, one should not simply restrict the channel estimator to be linear, which implies that the problem at hand is a nonlinear (and typically nonconvex) optimization problem. Overall, individual (separate) problems of the pilot design and the channel estimation are indeed very difficult, not to mention the problem of their joint optimization, which in fact needs to be solved eventually. To the best of our knowledge, there has been no work in the literature to jointly design the pilot and the channel estimator for the range of $L_s < N_t$ in the massive MIMO. This motivated our work.

In this paper, to circumvent the challenges discussed above and to construct a very effective channel estimation mechanism (including pilot design and channel estimation) for the massive MIMO when $L_s < N_t$, we take a fundamentally different approach by proposing a deep learning (DL) based two-stage channel estimation scheme. In the first stage, the joint design of pilot and channel estimator is performed to minimize the mean square error (MSE) of channel estimation by constructing both a two-layer neural network (TNN) and a deep neural network (DNN). In the second stage, to further enhance the channel estimation performance, we adopt an iterative channel estimation technique by constructing another DNN as an additional channel estimator, in which data detection and channel estimation are iteratively performed to improve the channel estimation performance. The simulation results

This work was supported by the Natural Sciences and Engineering Research Council of Canada Discovery under Grant 2019-04727.

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

J.-M. Kang is with the School of Intelligent Mechatronics Engineering, Sejong University, Seoul 05006, South Korea (e-mail: jaemo.kang@sejong.ac.kr).

[1] If $L_s \geq N_t$ is assumed as in the literature, it is straightforward to design the pilot matrix $\boldsymbol{S}$ of size $N_t \times L_s$ such that $\boldsymbol{S}\boldsymbol{S}^H = \boldsymbol{I}_{N_t}$. However, when $L_s < N_t$, it is not possible to have $\boldsymbol{S}\boldsymbol{S}^H = \boldsymbol{I}_{N_t}$.

[2] When $L_s \geq N_t$ in the presence of additive white Gaussian noise, it is known that LMMSE is an optimal channel estimator [7].

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.2019.2912378, IEEE Wireless Communications Letters

demonstrate that the proposed channel estimation scheme considerably outperforms the state-of-the-art existing scheme.

II. Two-Stage Channel Estimation

In this section, we describe the two-stage channel estimation approach[3] that is adopted in this paper: (i) pilot-aided channel estimation and (ii) data-aided iterative channel estimation.

We consider the downlink of a massive MIMO system, in which the transmitter is equipped with N_t antennas and the receiver is equipped with N_r antennas. The channel from the transmitter to the receiver is denoted by $\boldsymbol{H} \in \mathcal{C}^{N_r \times N_t}$. In this paper, we consider the frequency division multiplexing (FDD), in which the channel is not reciprocal. We consider the block fading channel model, in which the channel is constant over a coherence block of $\bar{L}$ time slots.

A. Pilot-Aided Channel Estimation

In the first L_s time slots within a coherence fading block, pilot $\boldsymbol{S} \in \mathcal{C}^{N_t \times L_s}$ is transmitted to the receiver and the received pilot signal $\boldsymbol{Y}_s \in \mathcal{C}^{N_r \times L_s}$ is given by

$$\boldsymbol{Y}_s = \boldsymbol{H}\boldsymbol{S} + \boldsymbol{Z}_s, \tag{1}$$

where $\boldsymbol{Z}_s \in \mathcal{C}^{N_r \times L_s}$ denotes the noise matrix, whose elements are i.i.d. with $\mathcal{CN}(0, \sigma_s^2)$. Using the received pilot signal $\boldsymbol{Y}_s$ and the knowledge of the pilot $\boldsymbol{S}$, the channel $\boldsymbol{H}$ is estimated such that the MSE of channel estimation is minimized. To present the channel estimation problem in a more convenient form, we denote the matrices in vectors as $\boldsymbol{y}_s = \mathrm{vec}(\boldsymbol{Y}_s) \in \mathcal{C}^{N_r L_s \times 1}$, $\boldsymbol{h} = \mathrm{vec}(\boldsymbol{H}) \in \mathcal{C}^{N_t N_r \times 1}$, and $\boldsymbol{z}_s = \mathrm{vec}(\boldsymbol{Z}_s) \in \mathcal{C}^{N_r L_s \times 1}$. Then the received signal (1) can be rewritten as

$$\boldsymbol{y}_s = (\boldsymbol{S}^T \otimes \boldsymbol{I}_{N_r})\boldsymbol{h} + \boldsymbol{z}_s. \tag{2}$$

The channel $\boldsymbol{h}$ is estimated by a channel estimator, denoted by $\mathcal{F}(\cdot)$, which uses the received signal $\boldsymbol{y}_s$ and the knowledge of the pilot $\boldsymbol{S}$. Using $\mathcal{F}(\cdot)$, we denote the estimated channel by $\hat{\boldsymbol{h}}_s = \mathcal{F}(\boldsymbol{y}_s; \boldsymbol{S})$. Then the pilot-aided channel estimation problem, which is a joint design problem of the pilot $\boldsymbol{S}$ and the channel estimator $\mathcal{F}(\cdot)$, can be written as

$$(\text{P1}): \min_{\boldsymbol{S},\, \mathcal{F}(\cdot)} \ \mathbb{E}\left[\left\|\boldsymbol{h} - \hat{\boldsymbol{h}}_s\right\|^2\right] \ \text{s.t.}\ \mathrm{tr}\left(\boldsymbol{S}\boldsymbol{S}^H\right) \le E_s,$$

where E_s is the energy constraint for the pilot. In this paper, solving (P1) is referred to as the pilot-aided channel estimation. Note that, unlike the very common assumption in the literature, we do *not* restrict the channel estimator $\mathcal{F}(\cdot)$ to be a linear function of $\boldsymbol{y}_s$ and $\boldsymbol{S}$ because linear estimators are not necessarily optimal in our system. Now, (P1) becomes a very challenging non-linear non-convex problem.

B. Data-Aided Iterative Channel Estimation

After transmitting the pilot $\boldsymbol{S}$, data $\boldsymbol{X} \in \mathcal{C}^{N_t \times L_d}$ is transmitted in the next L_d time slots to the receiver, where $L_d = \bar{L} - L_s$. The received data signal $\boldsymbol{Y}_x \in \mathcal{C}^{N_r \times L_d}$ is given by

$$\boldsymbol{Y}_x = \boldsymbol{H}\boldsymbol{X} + \boldsymbol{Z}_x, \tag{3}$$

where $\boldsymbol{Z}_x \in \mathcal{C}^{N_r \times L_d}$ denotes the noise matrix, whose elements are i.i.d. with $\mathcal{CN}(0, \sigma_x^2)$. Using $\hat{\boldsymbol{h}}_s$ that was obtained by the pilot-aided channel estimation in the previous stage, $\boldsymbol{X}$ is first estimated (or detected) into $\hat{\boldsymbol{X}}^{(1)}$. Now, for the second stage channel estimation, we regard $\hat{\boldsymbol{X}}^{(1)}$ as a pilot and then we estimate the channel $\boldsymbol{h}$ again, which is denoted by $\hat{\boldsymbol{h}}_x^{(1)}$. This procedure is iteratively repeated to further enhance the performance of data detection and channel estimation. Specifically, at the kth iteration, $\hat{\boldsymbol{X}}^{(k)}$ denotes the estimate of $\boldsymbol{X}$, which is obtained by the LMMSE detector given by $\hat{\boldsymbol{X}}^{(k)} = \left((\hat{\boldsymbol{H}}_x^{(k-1)})^H \hat{\boldsymbol{H}}_x^{(k-1)} + \sigma_x^2 \boldsymbol{I}_{N_t}\right)^{-1} (\hat{\boldsymbol{H}}_x^{(k-1)})^H \boldsymbol{Y}_x$, where $\hat{\boldsymbol{H}}_x^{(k)} = \mathrm{vec}^{-1}(\hat{\boldsymbol{h}}_x^{(k)})$ and $\hat{\boldsymbol{h}}_x^{(0)} = \hat{\boldsymbol{h}}_s$. Also, $\hat{\boldsymbol{h}}_x^{(k)}$ denotes the estimate of $\boldsymbol{h}$ at the kth iteration, which is obtained by the channel estimator in the second stage, denoted by $\mathcal{G}^{(k)}(\cdot)$, as follows:

$$\hat{\boldsymbol{h}}_x^{(k)} = \mathcal{G}^{(k)}\left(\boldsymbol{y}_x; \hat{\boldsymbol{X}}^{(k)}\right), \ k = 1, 2, \cdots, \tag{4}$$

where $\boldsymbol{y}_x = \mathrm{vec}(\boldsymbol{Y}_x)$. At the kth iteration, the data-aided channel estimation problem can be written as

$$(\text{P2}): \min_{\mathcal{G}^{(k)}(\cdot)} \ \mathbb{E}\left[\left\|\boldsymbol{h} - \hat{\boldsymbol{h}}_x^{(k)}\right\|^2\right]. \tag{5}$$

As in the first stage, we do *not* restrict $\mathcal{G}^{(k)}(\cdot)$ to be a linear function of $\boldsymbol{y}_x$ and $\hat{\boldsymbol{X}}^{(k)}$. Thus, the problem (P2) is also a very difficult problem that is nonlinear and nonconvex.

III. DL-Based Channel Estimation for Massive MIMO

In this section, to tackle the problems (P1) and (P2), we take a fundamentally different approach by designing a new receiver structure based on DL techniques.

A. Structure of the Proposed Scheme

In this subsection, we develop a DL-based channel estimation scheme by using the neural networks. The structure of the proposed DL-based channel estimation scheme, which consists of the DL-based pilot-aided channel estimation and the DL-based data-aided iterative channel estimation, is shown in Fig. 1.

1) DL-Based Pilot-Aided Channel Estimation: In order to design the pilot $\boldsymbol{S}$, we use a TNN (as a pilot designer), which is motivated by the following observation: the signal component $(\boldsymbol{S}^T \otimes \boldsymbol{I}_{N_r})\boldsymbol{h}$ in (2) excluding the noise $\boldsymbol{z}_s$ can be exactly modeled by the output of a TNN when the input to the TNN is $\boldsymbol{h}$ and the weight matrix of the TNN is given by $\boldsymbol{S}^T \otimes \boldsymbol{I}_{N_r}$ with all zero biases and unit activation functions. Then the entire received signal $\boldsymbol{y}_s$ in (2) can be represented by the summation of the output of the TNN and the noise $\boldsymbol{z}_s$ as shown in Fig. 1. Note that the TNN will be later optimized jointly with the channel estimator $\mathcal{F}(\cdot)$ such that the MSE of channel estimation is minimized.

In order to estimate the channel $\boldsymbol{h}$ by using the received signal $\boldsymbol{y}_s$ and the pilot $\boldsymbol{S}$, we use a DNN (as a channel estimator), which is motivated by the fact that DNNs can theoretically approximate any nonlinear input-output mechanisms arbitrarily closely [9]. In our proposed scheme, a DNN is used as a *nonlinear* channel estimator $\mathcal{F}(\cdot)$ and it will be denoted by DNN-1. The weight matrix and the bias vector in the mth hidden layer (resp. the output layer) of the DNN-1 are denoted by $\boldsymbol{W}_m$ and $\boldsymbol{a}_m$ (resp. $\boldsymbol{W}_o$ and $\boldsymbol{a}_o$) for $m = 1, \cdots, M_1$, respectively. Assuming there are M_1 hidden layers, the estimated channel $\hat{\boldsymbol{h}}_s$ can be mathematically written as $\hat{\boldsymbol{h}}_s = \boldsymbol{W}_o \phi_{M_1}\left(\boldsymbol{W}_{M_1} \phi_{M_1-1}\left(\cdots \phi_1(\boldsymbol{W}_1 \boldsymbol{y}_s + \boldsymbol{a}_1) \cdots\right) + \boldsymbol{a}_{M_1}\right) + \boldsymbol{a}_o$, where $\phi_m(\cdot)$ denotes the activation functions at the nodes of the mth hidden layer. We use the rectified linear unit

[3]This two-stage approach for channel estimation has been demonstrated to be very effective in many works, e.g., [3], [8].

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.2019.2912378, IEEE Wireless Communications Letters

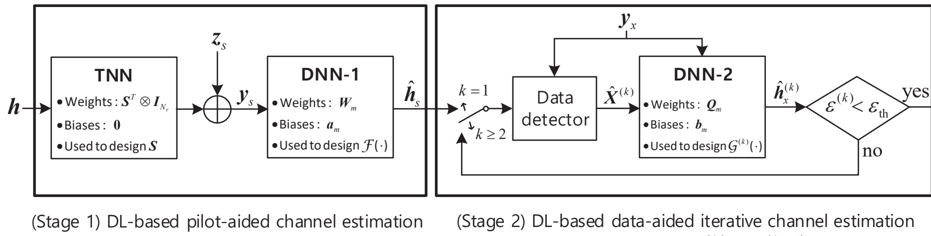


Fig. 1: The structure of the proposed scheme, where $\varepsilon^{(k)} = \|\hat{\boldsymbol{h}}_x^{(k)} - \hat{\boldsymbol{h}}_x^{(k-1)}\|^2$.

(ReLU) as the activation function at the nodes of all hidden layers, i.e., $\phi_m(i) = \max(0, i)$, which has been widely used for avoiding gradient vanishing problem and for computational efficiency [10]. Once training of the TNN (and the DNN-1) has completed, the pilot $\boldsymbol{S}$ in (P1) can be determined by simply reading the weight matrix $\left(\boldsymbol{S}^T \otimes \boldsymbol{I}_{N_r}\right)$ of the TNN.

2) DL-Based Data-Aided Iterative Channel Estimation: In order to further improve the channel estimation performance, we use another DNN, which is denoted by DNN-2, as a channel estimator in the second stage. As described in (4), the channel $\hat{\boldsymbol{h}}_x^{(k)}$ is estimated by using the received data signal $\boldsymbol{y}_x$ and the estimate of data $\hat{\boldsymbol{X}}^{(k)}$, which is estimated by using the LMMSE detector. Thus, the combined vector of $\boldsymbol{y}_x$ and $\text{vec}\big(\hat{\boldsymbol{X}}^{(k)}\big)$ is the input to the DNN-2 and the output of the DNN-2 is the estimated channel $\hat{\boldsymbol{h}}_x^{(k)}$ as presented in Fig. 1. The weight matrix and the bias vector in the mth hidden layer (resp. the output layer) of the DNN-2 are denoted by $\boldsymbol{Q}_m^{(k)}$ and $\boldsymbol{b}_m^{(k)}$ (resp. $\boldsymbol{Q}_o^{(k)}$ and $\boldsymbol{b}_o^{(k)}$) for $m = 1, \cdots, M_2$, respectively. Assuming there are M_2 hidden layers, the estimated channel $\hat{\boldsymbol{h}}_x^{(k)}$ can be mathematically written as $\hat{\boldsymbol{h}}_x^{(k)} = \boldsymbol{Q}_o^{(k)} \phi_{M_2}\big(\boldsymbol{Q}_{M_2}^{(k)} \phi_{M_2-1}\big(\cdots \phi_1\big(\boldsymbol{Q}_1^{(k)}\big[\boldsymbol{y}_x^T, \text{vec}(\hat{\boldsymbol{X}}^{(k)})^T\big]^T + \boldsymbol{b}_1^{(k)}\big) \cdots\big) + \boldsymbol{b}_{M_2}^{(k)}\big) + \boldsymbol{b}_o^{(k)}$. In the very first iteration, the input to the DNN-2 is a combined vector of $\boldsymbol{y}_x$ and $\text{vec}(\hat{\boldsymbol{X}}^{(1)})$. Recall that $\hat{\boldsymbol{X}}^{(1)}$ is the data estimate that is obtained by using the channel estimate $\hat{\boldsymbol{h}}_s$, which was obtained in the first stage at the output of the DNN-1. The iteration continues until $\varepsilon^{(k)} = ||\hat{\boldsymbol{h}}_x^{(k)} - \hat{\boldsymbol{h}}_x^{(k-1)}||^2$ is less than a threshold ε_{th}.

B. Training of the Proposed Scheme

In the first stage, we jointly train the TNN and DNN-1 for joint design of the pilot $\boldsymbol{S}$ and the channel estimator $\mathcal{F}(\cdot)$ such that the MSE of channel estimation is minimized as in the problem (P1). The loss function is set to the sample mean of squared errors as follows [11]–[13]:

$$\mathcal{J}_s = \frac{1}{|\mathcal{H}|} \sum_{\boldsymbol{h} \in \mathcal{H}} \left\| \boldsymbol{h} - \hat{\boldsymbol{h}}_s \right\|^2, \tag{6}$$

where $\mathcal{H}$ denotes the set of channel samples generated for training. All the weights and biases of the DNN-1 are updated by the stochastic gradient descent (SGD) method as $\theta_s \leftarrow \theta_s - \alpha \nabla_{\theta_s} \mathcal{J}_s$, where θ_s is the set of weights and biases of DNN-1. From the power constraint of the pilot in (P1), the weights of the TNN must satisfy $\text{tr}\big((\boldsymbol{S}^T \otimes \boldsymbol{I}_{N_r})(\boldsymbol{S}^T \otimes \boldsymbol{I}_{N_r})^H\big) \leq E_s N_r$. To this end, we use the projected gradient descent method [12] to update the weights of the TNN as follows:

$$\left(\boldsymbol{S}^T \otimes \boldsymbol{I}_{N_r}\right) \leftarrow \begin{cases} \text{vec}^{-1}(\boldsymbol{u}), & \text{if } ||\boldsymbol{u}||^2 \leq E_s N_r \\ \text{vec}^{-1}\big(\sqrt{E_s N_r}\, \boldsymbol{u} \,/\, ||\boldsymbol{u}||\big), & \text{if } ||\boldsymbol{u}||^2 > E_s N_r \end{cases}, \tag{7}$$

where $\boldsymbol{u} = \text{vec}\big(\boldsymbol{S}^T \otimes \boldsymbol{I}_{N_r}\big) - \alpha \nabla_{\boldsymbol{S}} \mathcal{J}_s$ and $\alpha > 0$ is the step size.

In the second stage, we train the DNN-2 for designing of the channel estimator $\mathcal{G}^{(k)}(\cdot)$ such that the MSE of channel estimation is minimized as in the problem (P2). That is, at the kth iteration, the loss function is set to the sample mean of squared errors as follows:

$$\mathcal{J}_x^{(k)} = \frac{1}{|\mathcal{H}|} \sum_{\boldsymbol{h} \in \mathcal{H}} \left\| \boldsymbol{h} - \hat{\boldsymbol{h}}_x^{(k)} \right\|^2. \tag{8}$$

All the weights and biases of the DNN-2 are updated by the SGD method as $\theta_x \leftarrow \theta_x - \alpha \nabla_{\theta_x} \mathcal{J}_x$, where θ_x is the set of weights and biases of DNN-2.

The iterative process at the second stage always converges because the the SGD method never increase the cost function of (8). It turns out that proposed algorithm converges within five iterations.

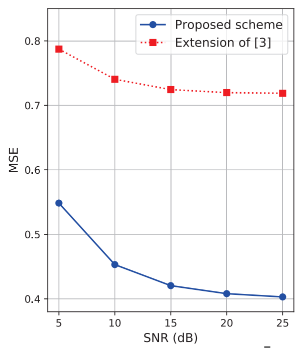
The proposed DL-based channel estimation scheme is trained offline by updating the weights and biases of all neural networks as described above. In the case of offline learning, computational complexity is a (much) less concern, because time is not strictly limited. Once the proposed scheme is trained, the pilot $\boldsymbol{S}$, DL-based pilot-aided channel estimator $\mathcal{F}(\cdot)$, and DL-based data-aided channel estimator $\mathcal{G}^{(k)}(\cdot)$ are immediately determined by the weight matrix of the TNN, the weights and biases of DNN-1, and those of DNN-2, respectively. The computational complexity of the online channel estimation of the proposed scheme can be written as $O\big((M_2{-}1)\eta\lambda_2^2\big)$, where η denotes the number of iterations and λ_i denotes the number of nodes in each hidden layers of DNN-i.

Remark 1: Recently, some DL-based channel estimation schemes have been proposed in [12] for energy harvesting systems and in [13] for multiuser systems. However, this work is different from [12] and [13] in that the issue of large transmit antenna size of the massive MIMO was not considered and the two-stage approach for channel estimation was not considered in [12] and [13].

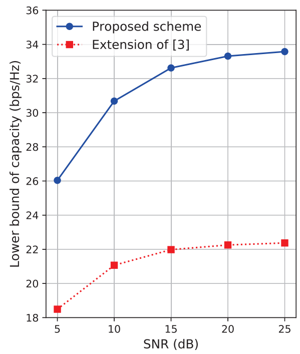
IV. Numerical Results

In the simulation, we consider the massive MIMO system with $N_t = 64$, $N_r = 16$, and $\bar{L} = 64$. Also, the pilot energy constraint is assumed to be given by $E_s = L_s T P_s$, where T denotes the length of a time slot and P_s denotes the transmit power. In the simulation, we set $T = 1$ and $P_s = 1$. In the proposed scheme, we set $M_1 = M_2 = 5$, $\alpha = 0.001$, $\lambda_1 = 2NrL_s$, and $\lambda_2 = 2(N_t + N_r)L_d$. We consider the spatially correlated channel, which is modeled by using the Kronecker channel model as $\boldsymbol{H} = \boldsymbol{\Sigma}_r^{1/2} \boldsymbol{H}_w (\boldsymbol{\Sigma}_t^{1/2})^H$. The (i, j)-th element of $\boldsymbol{\Sigma}_r$ and $\boldsymbol{\Sigma}_t$ are given by $\rho_r^{|i-j|}$ and $\rho_t^{|i-j|}$, respectively, where the correlation coefficients ρ_r and ρ_t are set to be 0.5. We generate each element of $\boldsymbol{H}_w$ independently according to $\mathcal{CN}(0, 1)$. The signal-to-noise ratio (SNR) is $\gamma = \frac{1}{\sigma^2}$, where $\sigma^2 = \sigma_s^2 = \sigma_x^2$. We train the proposed scheme using 10^5 channel samples

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.2019.2912378, IEEE Wireless Communications Letters

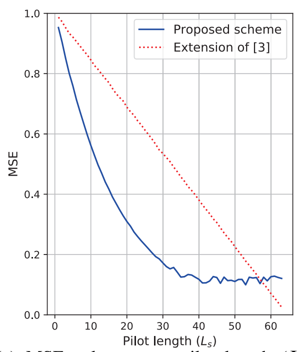


(a) MSE value versus SNR when $\bar{L} = 64$ and $L_s = 18$.

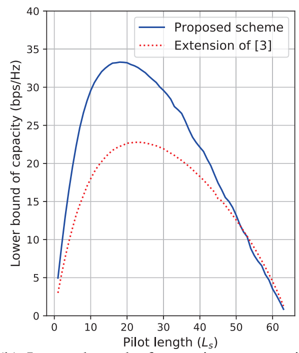


(b) Lower bound of capacity versus SNR when $\bar{L} = 64$ and $L_s = 18$.

Fig. 2: MSE and lower bound of capacity comparison between the proposed scheme and the LMMSE based iterative channel estimation scheme, i.e., extension of [3].



(a) MSE value versus pilot length (L_s) when SNR = 20 dB.



(b) Lower bound of capacity versus pilot length (L_s) when SNR = 20 dB.

Fig. 3: Lower bound of capacity versus pilot length L_s for the proposed scheme and the LMMSE based iterative channel estimation scheme, i.e., extension of [3].

and 10^5 noise samples over 20 epochs. Also, we use 10^4 test samples apart from the training samples for performance evaluation.

In order to compare the proposed scheme with the state-of-the-art existing scheme, we extend the scheme of [3] to the MIMO channel.[4] Specifically, the estimated channel based on the LMMSE channel estimators for the pilot-aided channel estimation and for the data-aided channel estimation are respectively given by [7]

$$\hat{\boldsymbol{h}}_s = \boldsymbol{C}_h(\boldsymbol{S}^*\otimes\boldsymbol{I}_{N_r})\left[(\boldsymbol{S}^T\otimes\boldsymbol{I}_{N_r})\boldsymbol{C}_h(\boldsymbol{S}^*\otimes\boldsymbol{I}_{N_r})+\boldsymbol{C}_{z,s}\right]^{-1}\boldsymbol{y}_s \quad (9)$$

$$\hat{\boldsymbol{h}}_x^{(k)} = \boldsymbol{C}_h\bar{\boldsymbol{X}}^{(k)}\left[(\bar{\boldsymbol{X}}^{(k)})^H\boldsymbol{C}_h\bar{\boldsymbol{X}}^{(k)}+\boldsymbol{C}_{z,x}\right]^{-1}\boldsymbol{y}_x, \quad (10)$$

where $\boldsymbol{C}_h = \mathbb{E}(\boldsymbol{h}\boldsymbol{h}^H)$, $\boldsymbol{C}_{z,s} = \mathbb{E}(z_s z_s^H)$, $\boldsymbol{C}_{z,x} = \mathbb{E}(z_x z_x^H)$, and $\bar{\boldsymbol{X}}^{(k)} = (\hat{\boldsymbol{X}}^{(k)})^*\otimes\boldsymbol{I}_{N_r}$. Also, the pilot $\boldsymbol{S}$ is set to be an $N_t \times L_s$ semi-unitary matrix, i.e., the column vectors of $\boldsymbol{S}$ are orthogonal with one another.

In Fig. 2(a), the MSE of channel estimation is shown versus SNR for $L_s = 18$. The MSE of channel estimation of the proposed scheme is obtained by applying the test channel samples to the proposed channel estimation process. It can be seen that the proposed scheme considerably outperforms the conventional scheme for the entire SNR range. This means that the pilot and channel estimator are well designed by our proposed scheme, and the channel estimation performance is substantially enhanced by our proposed DL-based channel estimation scheme. In Fig. 2(b), the lower bound of capacity is shown versus the SNR for $L_s = 18$. The lower bound of capacity is given by [6] $C_{\text{low}} = \frac{\bar{L}-L_s}{\bar{L}}\mathbb{E}\{\log(\boldsymbol{I}_{N_t} + \frac{1}{\sigma^2+\varepsilon}\hat{\boldsymbol{H}}^H\hat{\boldsymbol{H}})\}$, where $\hat{\boldsymbol{H}}$ is the estimated channel and ε is the MSE of channel estimation. This lower bound is very close to the exact capacity [6] (i.e., very accurate approximation of capacity). It can be seen that, in terms of C_{low}, the proposed scheme also outperforms the conventional scheme for the entire SNR range.

In Fig. 3(a), MSE is shown versus L_s. In the range of very large L_s values (e.g., $L_s > 50$), the MSE value of the conventional scheme is lower than the proposed scheme. However, when L_s is very large, the data transmission time L_d available for data transmission becomes too small, which leads to a smaller capacity, as shown in Fig. 3(b). In Fig. 3(b), C_{low} is shown versus L_s. It can be seen that the lower bound of capacity is maximized when $L_s = 18$. For both schemes, it can be seen that the data rate performance deteriorates either if L_s is too short (due to too much channel estimation errors) or if L_s is too long (due to too little resource left available for data transmission).

The computational complexity of [3] is $O(\eta N_r^3 N_t^2 L_d)$, which turns out be slightly higher than the complexity of the proposed scheme when the simulation parameters are substituted.

V. Conclusion

We proposed a DL-based effective channel estimation scheme for the massive MIMO system with better performance than the conventional channel estimation scheme.

References

[1] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.

[2] H. Xie, B. Wang, F. Gao, and S. Jin, "A full-space spectrum-sharing strategy for massive MIMO cognitive radio systems," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 10, pp. 2537–2549, Oct. 2016.

[3] J. Ma and L. Ping, "Data-aided channel estimation in large antenna systems," *IEEE Trans. Signal Process.*, vol. 62, no. 12, pp. 3111–3124, Jun. 2014.

[4] Y. Li, C. Tao, G. Seco-Granados, A. Mezghani, A. L. Swindlehurst, and L. Liu, "Channel estimation and performance analysis of one-bit massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 65, no. 15, pp. 4075–4089, Aug. 2017.

[5] C. Huang, G. C. Alexandropoulos, A. Zappone, C. Yuen, and M. Dbbah, "Deep learning for UL/DL channel calibration in generic massive MIMO systems," *arXiv:1903.02875*, 2019.

[6] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 951–963, Apr. 2003.

[7] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory.* Englewood Cliffs, NJ, USA: Prentice-Hall, 1993.

[8] S. Buzzi, M. Lops, and S. Sardellitti, "Performance of iterative data detection and channel estimation for single-antenna and multiple-antennas wireless communications," *IEEE Trans. Veh. Technol.*, vol. 53, no. 4, pp. 1085–1104, Jul. 2004.

[9] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," *Neural Networks*, vol. 2, no. 5, pp. 359–366, 1989.

[10] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015.

[11] 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, Oct. 2018.

[12] J.-M. Kang, C.-J. Chun, and I.-M. Kim, "Deep learning based channel estimation for wireless energy transfer," *IEEE Commun. Lett.*, vol. 22, no. 11, pp. 2310–2313, Nov. 2018.

[13] C.-J. Chun, J.-M. Kang, and I.-M. Kim, "Deep learning based joint pilot design and channel estimation for multiuser MIMO channels," *arXiv:1812.04120*, 2018.

[4] In [3], the authors proposed the LMMSE based the iterative channel estimation scheme for the multiuser single-input multiple-output system.