




Deep Learning-Based Channel Estimation for Massive MIMO Systems With Pilot Contamination

HIROKI HIROSE ¹ (Student Member, IEEE), TOMOAKI OHTSUKI ² (Senior Member, IEEE),
AND GUAN GUI ³ (Senior Member, IEEE)

(Invited Paper)

¹Graduate School of Science, and Technology, Keio University, Kanagawa 223-8522, Japan

²Department of Information, and Computer Science, Keio University, Kanagawa 223-8522, Japan

³College of Telecommunication, and Information Engineering, Nanjing University of Posts, and Telecommunications, Nanjing 210003, China

CORRESPONDING AUTHOR: HIROKI HIROSE (e-mail: hirose hiroki@keio.jp)

ABSTRACT In a time division duplex (TDD) based massive multiple-input multiple-output (MIMO) system, a base station (BS) is required to obtain accurate estimation of channel state information (CSI) for a user terminal (UT). Because of the time-varying nature of the channel, the length of pilot signals is limited and the number of orthogonal pilot signals is finite. Hence, the same pilot signals are required to be reused in neighboring cells and thus its channel estimation performance is deteriorated by pilot contamination from the neighboring cells. The minimum mean square error (MMSE) channel estimation can be used to reduce the influence of pilot contamination. However, it needs to know the covariance matrix of channels for all the UTs, which is unknown to the BS in practice. In this paper, we propose two methods of deep learning aided channel estimation to reduce the influence of pilot contamination. One method uses a neural network consisting of fully connected layers, while the other method uses a convolutional neural network (CNN). The neural network, particularly the CNN, plays a role in extracting features of the spatial information from the contaminated signals. The former method is better in terms of the training speed, however, the latter one can estimate the channel more accurately. We evaluate the proposed methods under two scenarios, i.e., perfect timing synchronization and imperfect one. Simulation results confirm that the proposed methods are better than the LS and covariance estimation methods via normalized mean square error (NMSE). In addition, we also investigate the impact of channel aging, and show that including some expected data into training datasets can avoid the great degradation of estimation quality.

INDEX TERMS Channel estimation, deep learning, massive MIMO, pilot contamination.

I. INTRODUCTION

Massive multiple-input multiple-output (MIMO) is one of the promising technologies in the fifth generation (5G) and beyond [1]–[4]. The basic concept of the massive MIMO technique is that a base station (BS) equips with hundreds of antenna arrays to serve tens of user terminals (UTs) simultaneously. A large number of antennas at the BS enable spatial multiplexing to the UTs over the same time-frequency resource [1], which achieves high spatial efficiencies. In addition, the massive MIMO systems have a lot of benefits, such

as high data rates, high energy efficiencies, and simple linear transceiver design [5].

To fully exploit these advantages, it is necessary for the BS to have accurate channel state information (CSI). The channel estimation is an important operation since the performance of the massive MIMO systems depends on the quality of the estimation. The channel response is assumed to be constant for time and frequency domains within the coherence block, and thus the channel estimation is performed once per coherent block. When operated with time division duplex (TDD), the

characteristic of reciprocity between uplink (UL) and downlink (DL) can be exploited, so that pilot signal transmissions are conducted in the only UL. Therefore, the coherence block is divided into three categories, the UL pilot signal transmissions, the UL data signal transmissions, and the DL data signal transmissions. To increase the spectral efficiencies, the length of pilot signals should be small. However, as the number of UTs becomes larger, the length of pilot signals has to be large to keep the orthogonality to the signals of each UT. Thus, the resource for data transmission blocks is limited. To avoid this issue, the same pilot signal of its own cell is reused in neighboring cells. Therefore, the BS cannot separate the signals from the UTs using co-pilots in the other cells. Hence, the channel estimation performance degrades largely because of the inter-cell interference of the same pilot signals from the other cells, which is called pilot contamination [1]. It is hard to mitigate the effects of the pilot contamination even if the number of BS antenna arrays is infinite, unlike intra-cell interference, fast-fading, and noise at the BS [5].

A lot of approaches have been proposed to estimate channels in TDD-based massive MIMO systems under the existence of pilot contamination. In [6], a covariance-aided channel estimation is proposed, in which the minimum mean square error (MMSE) channel estimation is derived. It is shown that the pilot contamination can be removed completely when the covariance matrices satisfy a certain non-overlapping condition. With making use of the idea, a coordinated approach for assigning the pilot signals is proposed so that the covariance matrices could be transformed into the structure satisfying the required condition. However, the channel covariance matrix that expresses channel spatial correlation is needed, and the BS does not possess this information in advance [7]. The method of using extra pilots [8] is proposed to estimate the covariance matrix. It can eliminate the effects of interference to some extent. However, the available coherence block for sending data is limited because of the time varying nature of the channel. Besides, such information is also needed in the Joint Spatial Division and Multiplexing (JSDM) scheme [9], which reduces the channel estimation training overhead. The blind method proposed in [10] does not need a priori information. This scheme can distinguish the interference subspace from the desired signal subspace. It exploits the eigenvalue distribution of the sample covariance matrix calculated from the received signals, which tells the desired signals from interfering ones. Therefore, it is possible to separate blindly the desired signal subspaces and the interfering ones. To distinguish the eigenvalue groups clearly, the power margin between the signals should be large. Thus, pilot contamination can be mitigated when the power margin between the desired signals and the interfering ones is significantly large. The authors of [11] proposed a time-shifted pilot transmission protocol. When the users of a certain cell are transmitting pilot signals, the users of the neighboring cells are receiving downlink data. After every group has finished the phase of sending pilot signals, all the users send uplink data. This method can avoid pilot contamination. However,

during the phase of sending pilot signals, the users at the cell-edge receive interference from the BSs in the neighboring cells. In [12], a pilot contamination precoding method is proposed. This method makes use of multi-cell cooperation. The precoding matrix at the BS is made to minimize the sum of inter-cell and intra-cell interference. When the number of the BS antennas approaches infinity, pilot contamination can be completely removed. However, this requires a centralized processing at the BS and all BSs have to know the data of all users.

In recent years, deep learning is often applied in the fields of computer vision and natural language processing to solve complex nonlinear problems. Moreover, it is also incorporated into the area of wireless communication [13]. The nature of channel in the massive MIMO systems is complicated, and the non-linear optimization for the channel estimation requires the high computational complexity. Therefore, deep learning-based method is a candidate for addressing this problem because it has the ability to solve many non-convex and non-linear issues. In [14], the deep learning is integrated into direction-of-arrival (DoA) estimation and channel estimation in massive MIMO systems. The scheme is designed to learn the statistics of the channel model and acquire the sparsity features in angle domain. In [15], it is proposed that using neural network helps to improve the channel estimation performance. This approach is designed by the structure of the MMSE channel estimator, and the convolutional neural network (CNN) [16] is presented as a class of low-complexity channel estimator. In [17], it is shown that CNN and long short-term memory (LSTM) can be applied in the fast time-varying channel estimation. In [18], deep learning-based two-stage channel estimation including pilot-aided stage and data-aided stage is proposed. The data-aided stage has the iterative structure of signal detection and channel estimation. The estimated channel quality is improved because the length of data symbols is much longer than that of pilot symbols. However, all these methods [14], [15], [17], [18] assume a single-cell layout and ignore the existence of pilot contamination while it is quite often occurred in multi-cell layout scenarios.

To suppress the influence of pilot contamination in TDD-based massive MIMO systems, we propose two kinds of channel estimation methods based on deep learning, i.e., neural network (NN) and CNN. The NN-based estimation utilizes a neural network consisting of fully connected layers, while the CNN-based estimation utilizes a CNN. The motivation for utilizing the neural network lies that it plays a role in extracting features of the spatial information contained in the LS estimated channel. In particular, CNNs have the advantages that the spatial correlation can be exploited with the sliding convolutional filters. Moreover, it is reasonable to use neural networks that can handle a large amount of correspondences between the LS estimated channel and the desired channel. It is shown that the CNN-based estimation can achieve better estimation performance than the NN-based estimation, while the NN-based estimation takes less time to train the datasets than the CNN-based estimation.

The main contributions of this paper are summarized as follows.

- Firstly, we propose a channel estimation method using the deep learning in the presence of pilot contamination. While the conventional method [8] consumes extra time resources and its optimal parameters are unknown, the proposed methods do not require the extra signals and leverage the trained statistical information to estimate the channel end-to-end.
- Secondly, in contrast to existing researches on channel estimation using the deep learning [13]–[17], our goal is to suppress the influence of pilot contamination that occurs in practical environments. We give the explanation of how the deep learning contributes to suppress the inter-cell interference.
- Thirdly, through our computer simulation results, we show that the proposed methods improve the normalized mean square error (NMSE) of the channel compared with the conventional method [8]. Moreover, we investigate the impact of channel aging caused by the change of environments such as the movement of UTs. We consider two cases in which different statistical characteristics of datasets are used in the training and the evaluation, and the same ones are used.

The remaining of this paper is organized as follows. Section II presents the system model and describes two scenarios, i.e., timing synchronization and channel aging. Section III describes the conventional covariance estimation method for the MMSE channel estimation. Section IV proposes two methods of deep learning-based channel estimation in the presence of pilot contamination. Section V evaluates the proposed methods in terms of the channel NMSE. Section VI concludes this paper.

Notation: We use $NC(\mathbf{0}, \mathbf{R})$ as the circularly symmetric complex Gaussian distribution with zero mean and the covariance matrix \mathbf{R} . The superscripts $(\cdot)^T$, $(\cdot)^*$, and $(\cdot)^H$ denote transpose, conjugate, and the Hermitian transpose operator, respectively. The \mathbf{I}_M denotes the $M \times M$ identity matrix.

II. SYSTEM MODEL

We assume L hexagonal cells, where each cell contains one BS equipped with M antennas and K UTs with single antenna. The channel from the k -th UT of the j -th cell to the BS in the l -th cell is denoted by $\mathbf{h}_{ljk} \in \mathbb{C}^M$. Typically, uncorrelated Rayleigh fading is often used as the channel model. However, it is practically questionable that the signals are equally likely to arrive at the BS from all directions [19]. In addition, practical massive MIMO channels are spatially correlated from the measurement campaigns [20], [21]. We assume that the spatial correlated channels the correlated Rayleigh fading model as follows

$$\mathbf{h}_{ljk} \sim NC(\mathbf{0}, \mathbf{R}_{ljk}), \quad (1)$$

where \mathbf{R}_{ljk} denotes the channel covariance matrix, which describes the macroscopic propagation characteristics. When the spatial correlation is independent, it can be expressed as

$\mathbf{R}_{ljk} = \beta_{ljk} \mathbf{I}_M$, which is a diagonal matrix. β_{ljk} is the large-scale fading coefficient between the BS in the l -th cell and the k -th UT in the j -th cell. However, we consider the spatial correlated channel, whose covariance matrix is represented by non-zero off-diagonal elements and non-identical diagonal elements [22]. The covariance matrix of the channel is modeled as

$$\mathbf{R}_{ljk} = \beta_{ljk} \int_{-\pi}^{\pi} p(\theta_{ljk}) \mathbf{a}(\theta_{ljk}) \mathbf{a}(\theta_{ljk})^H d\theta_{ljk}, \quad (2)$$

where $p(\theta)$ is the power spectrum of an angle of arrival (AoA), and $\mathbf{a}(\theta)$ is the steering vector of a uniform linear array (ULA). Here, the following range of θ is excluded, $\{\theta : p(\theta) < \epsilon\}$, where ϵ is some small number. The power spectrum $p(\theta)$ is assumed to be the Laplace distribution with a standard deviation equal to 2° . This models the scattering of the received power around the center of the propagation paths. We consider two cases of perfect timing synchronization throughout the multi-cell system, and imperfect one. In addition, we deal with the case of channel aging.

A. TIMING SYNCHRONIZATION

1) CASE OF PERFECT TIMING SYNCHRONIZATION

Most multi-cell literature assumes perfect timing synchronization for simplicity. The τ -length pilot signal allocated to the k -th UT in the j -th cell is expressed as $\mathbf{s}_{jk} \in \mathbb{C}^\tau$, and $\|\mathbf{s}_{jk}\|^2 = \tau$. Ideally, the pilot signals allocated to the users within the same cell and the neighboring cells should be orthogonal to each other as

$$\mathbf{s}_{jk}^H \mathbf{s}_{j'k'} = \begin{cases} \tau, & (j = j' \text{ and } k = k') \\ 0, & (\text{otherwise}) \end{cases}. \quad (3)$$

If the pilot signals satisfy Eq. (3), the pilot contamination does not occur. That is, the BS can estimate the desired channel without the inter-cell interference. However, the number of orthogonal pilot signals is limited by the channel coherence time, and then the number of users which can be served is limited. Therefore, the same pilot signals or non-orthogonal pilot signals are used in the neighboring cells. The estimated channel becomes correlated with the channel of the users using non-orthogonal pilot signals. As a result, the accuracy of channel estimation is degraded because of the interference from the users in the neighboring cells. We consider the k -th UT in each cell using the same pilot signal \mathbf{s}_k such as

$$\mathbf{s}_k^H \mathbf{s}_{k'} = \begin{cases} \tau, & (k = k') \\ 0, & (\text{otherwise}). \end{cases} \quad (4)$$

The BS in the l -th cell receives the superposition of the pilot signals from the UTs in all the cells, and it is expressed as

$$\mathbf{Y}_l = \sum_{k=1}^K \mathbf{h}_{l1k} \mathbf{s}_k^T + \sum_{j \neq l}^L \sum_{k=1}^K \mathbf{h}_{ljk} \mathbf{s}_k^T + \mathbf{N}_l. \quad (5)$$

Here, $\mathbf{N}_l \in \mathbb{C}^{M \times \tau}$ is the additive white Gaussian noise (AWGN) with zero-mean and element-wise variance σ_n^2 . To

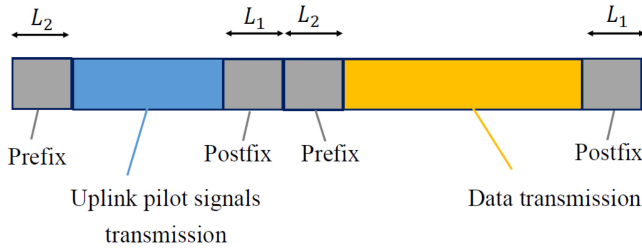


FIGURE 1. TDD scheduling protocol in the case of imperfect timing synchronization. The coherence time consists of block for uplink pilot transmission and one for uplink data transmission. A cyclic prefix of L_2 symbols and a cyclic postfix of L_1 symbols are inserted into each block.

estimate the channel, the BS in the l -th cell projects its received signal \mathbf{Y}_l on \mathbf{s}_k^* . This is called LS estimation [6], and the BS in the l -th cell can obtain

$$\hat{\mathbf{h}}_{llk}^{\text{LS}} = \mathbf{h}_{llk} + \sum_{j \neq l} \mathbf{h}_{ljk} + \frac{1}{\tau} \mathbf{N}_l \mathbf{s}_k^*. \quad (6)$$

MMSE-based channel estimator [6] is a typical method to suppress pilot contamination. The BS in the l -th cell with this method estimates the channel to the k -th user in the same cell as follows

$$\hat{\mathbf{h}}_{llk}^{\text{MMSE}} = \mathbf{R}_{llk} \mathbf{Q}_{lk}^{-1} \hat{\mathbf{h}}_{llk}^{\text{LS}}. \quad (7)$$

The channel covariance matrix is defined as $\mathbf{Q}_{lk} = \mathbb{E}[\mathbf{h}_{llk}^{\text{LS}} (\mathbf{h}_{llk}^{\text{LS}})^H]$. It is given by $\mathbf{Q}_{lk} = \sum_{j=1}^L \mathbf{R}_{ljk} + \sigma_n^2 / \tau \mathbf{I}_M$. The MMSE channel estimation exploits statistical information of the channel covariance matrix. If the BS knows the matrix \mathbf{R}_{llk} and \mathbf{Q}_{lk} in Eq. (7), the BS can calculate the estimated channel $\hat{\mathbf{h}}_{llk}$. However, in practice, the BS usually does not possess these matrices in advance.

2) IN THE CASE OF IMPERFECT TIMING SYNCHRONIZATION

Even though perfect timing synchronization is possible to achieve between a BS and its served UTs, it is difficult to achieve timing synchronization throughout the multi-cellular network [23]. The UT in the l -th cell is in perfect timing synchronization with the l -th BS. However, there is an $\epsilon_j T_s$ mismatch between the signal from the UT in the j -th cell and the one from the UT in the l -th cell. Here, T_s is a symbol interval, and $\epsilon_j \in [-0.5, 0.5)$ is a random variable and follows the uniform distribution. Note that the value ϵ_l is fixed at 0 because the UT in the l -th cell is in perfect timing synchronization.

It is assumed that pulse shaping filters are designed so that they satisfy the Nyquist criterion for avoiding the intersymbol interference (ISI). The impulse response of the filters is

$$R_p(mT_s) = \begin{cases} 0, & (m \neq 0) \\ 1, & (m = 0). \end{cases} \quad (8)$$

The scheduling protocol, when sending a pilot signal, is shown in Fig. 1. Based on the discrete-time model [23], the

BS in the l -th cell receives

$$\begin{aligned} y_l[i] &= \sum_{k=1}^K \mathbf{h}_{llk} s_k[i] \\ &+ \sum_{j \neq l} \sum_{k=1}^K \sum_{m=-L_1}^{L_2} R_p(mT_s - \epsilon_j T_s) \mathbf{h}_{ljk} s_k[i - m] + \mathbf{n}_l[i] \end{aligned} \quad (9)$$

at the time i , where $\mathbf{s}_k = [s_k[0], \dots, s_k[\tau - 1]]^T$ is used, and $\mathbf{n}_l[i]$ is the i -th column of \mathbf{N}_l . The last L_2 pilot symbols are inserted into before and after the block of UL pilot transmissions, respectively. The $R_p(mT_s - \epsilon_j T_s), \forall m \neq 0$ is no longer zero, and the term of the inter-cell interference signals are affected by the ISI due to the random ϵ_j . This is the difference between perfect timing synchronization and imperfect synchronization.

B. CHANNEL AGING

We also consider a quasi-static block fading channel model to investigate the impact of channel aging as in [24]. In general, the channel changes over time because of the movement of UTs. As a result, the channel coefficients keep constant within one symbol, but change from symbol to symbol. Therefore, the estimated channel at the BS is different from the real one when precoding the transmitted signals or combining the received signals. This is called channel aging. The time varying channel is commonly modeled by the Gauss-Markov block-fading model. Typically, the Jake's model is used, which is an autoregressive model, and assumes that the propagation path is composed of two-dimensional isotropic scatter [25]. For simplicity, we use the autoregressive model of order 1. The channel between the l -th BS and the k -th user in the j -th cell at the n -th symbol is given by

$$\mathbf{h}_{ljk}[n] = \alpha \mathbf{h}_{ljk}[n - 1] + \mathbf{e}_{ljk}[n], \quad (10)$$

where $\mathbf{e}_{ljk}[n] \sim \mathcal{CN}(\mathbf{0}, (1 - \alpha^2) \mathbf{R}_{ljk})$ is the stationary Gaussian channel error vector because of the time variation of the channel, independent of $\mathbf{h}_{ljk}[n - 1]$. The temporal correlation parameter α is given by

$$\alpha = J_0(2\pi f_D T_s), \quad (11)$$

where J_0 is the zeroth-order Bessel function of the first kind, and f_D is the maximum Doppler shift. The maximum Doppler shift f_D is given by

$$f_D = \frac{v f_c}{c}, \quad (12)$$

where v is the velocity of the UT, $c = 3.0 \times 10^8$ m/s is the speed of light, and f_c is the carrier frequency.

The relative movements between the UTs and the BS cause Doppler shift. This is a crucial cause of channel aging. The maximum Doppler shift is in proportion to the velocity of UTs according to Eq. (12). As the maximum Doppler shift increases, the temporal correlation decreases in Eq. (11). That

is, as the velocity of UTs increases, the variation of the channel coefficients becomes large. In practice, most UTs move around, and it is needed to evaluate the effects of channel aging.

III. CONVENTIONAL CHANNEL ESTIMATION METHOD

This section explains the conventional method to estimate the covariance matrices required for the MMSE channel estimation [8]. In this method, another extra pilot signal is added besides the ordinary pilot signal to make use of estimating the covariance matrix. As described in Section II, the BS in the l -th cell has to obtain the matrix \mathbf{R}_{llk} and \mathbf{Q}_{lk} for the MMSE channel estimation.

A. ESTIMATION OF \mathbf{Q}_{lk}

The covariance matrix $\mathbf{Q}_{lk} = \mathbb{E}[\hat{\mathbf{h}}_{llk}^{\text{LS}}(\hat{\mathbf{h}}_{llk}^{\text{LS}})^H]$ is estimated by some samples of the LS estimated channel for the received signals, $\hat{\mathbf{h}}_{llk}^{\text{LS}}[1], \dots, \hat{\mathbf{h}}_{llk}^{\text{LS}}[N_Q]$. The sample covariance matrix is obtained as

$$\hat{\mathbf{Q}}_{lk}^{(\text{sample})} = \frac{1}{N_Q} \sum_{n=1}^{N_Q} \hat{\mathbf{h}}_{llk}^{\text{LS}}[n] (\hat{\mathbf{h}}_{llk}^{\text{LS}}[n])^H, \quad (13)$$

where N_Q is the number of observations. However, it is more challenging to obtain the sample covariance matrix whose eigenvalues and eigenvectors are well aligned with those of the matrix \mathbf{Q}_{lk} [8]. To solve this problem, the covariance matrix is estimated as the convex combination

$$\hat{\mathbf{Q}}_{lk}(\eta) = \eta \hat{\mathbf{Q}}_{lk}^{(\text{sample})} + (1 - \eta) \hat{\mathbf{Q}}_{lk}^{(\text{diagonal})}, \quad (14)$$

where $\eta \in [0, 1]$ is a constant. Here, $\hat{\mathbf{Q}}_{lk}^{(\text{diagonal})}$ is the diagonalized sample covariance matrix $\hat{\mathbf{Q}}_{lk}^{(\text{sample})}$. When the constant η is smaller than 1, the off-diagonal elements of $\hat{\mathbf{Q}}_{lk}$ are underestimated and their values are treated as unreliable.

B. ESTIMATION OF \mathbf{R}_{llk}

In addition to the ordinal pilot signal, another pilot signal is transmitted to estimate \mathbf{R}_{llk} . In this time, the l -th UT does not send pilot signals, and only the UTs in the other cells do so. The BS receives the N_R pilot signals, and computes its sample covariance matrix $\hat{\mathbf{Q}}_{lk,-k}^{(\text{sample})}$. This matrix is considered as the sum of the covariance matrix of all the UTs causing the interference of the pilot contamination. The covariance matrix $\hat{\mathbf{Q}}_{lk}^{(\text{sample})}$ is already obtained, and another one $\hat{\mathbf{R}}_{llk}^{(\text{sample})}$ is estimated as follows

$$\hat{\mathbf{R}}_{llk}^{(\text{sample})} = \hat{\mathbf{Q}}_{lk}^{(\text{sample})} - \hat{\mathbf{Q}}_{lk,-k}^{(\text{sample})}. \quad (15)$$

As in the case of the matrix $\hat{\mathbf{Q}}_{lk}$, using a constant $\mu \in [0, 1]$, the matrix $\hat{\mathbf{R}}_{llk}$ can be estimated as

$$\hat{\mathbf{R}}_{llk}(\mu) = \mu \hat{\mathbf{R}}_{llk}^{(\text{sample})} + (1 - \mu) \hat{\mathbf{R}}_{llk}^{(\text{diagonal})}. \quad (16)$$

IV. DEEP LEARNING-BASED CHANNEL ESTIMATION

This section proposes the deep learning-based channel estimation in multi-cell massive MIMO systems in which pilot

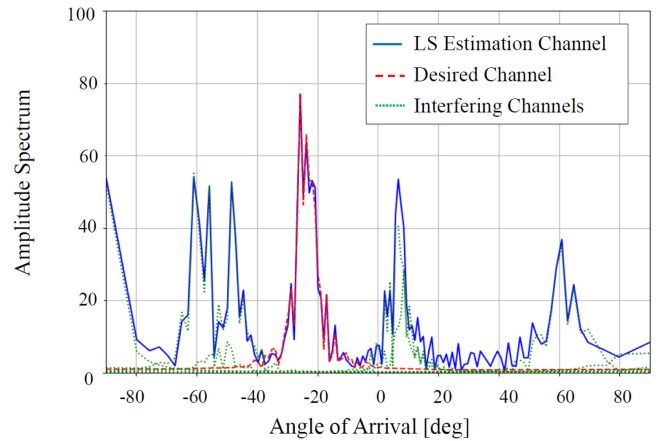


FIGURE 2. Amplitude spectrum of the angle domain for the LS estimated channel, the desired channel and the interference channels.

contamination occurs. In [15], the MMSE channel estimation is transformed into the structure of a neural network under some certain assumptions. However, their method can not be used as it is since it is designed in a single cellular network without pilot contamination. We consider the scheme of neural networks for reducing the effects of pilot contamination that can occur in a more practical environment. First it is shown how the neural network can help against inter-cell interference of pilot contamination in the multi-cell massive MIMO system. After that, we present two proposed methods: NN-based estimation and CNN-based estimation.

A. SUPPRESSION OF PILOT CONTAMINATION THROUGH LEARNING

The LS estimated channel contains the sum of the channels for all the UTs using the same pilot signal as in Eq. (6). Therefore, for accurate channel estimation, it is necessary to retrieve the desired channel from the interference-affected LS estimated channel. The channel covariance matrix \mathbf{R}_{llk} integrates the power spectrum of the AoA and the steering vector in terms of the angular domain as in Eq. (2). Therefore, by performing the discrete Fourier transform (DFT) on the channel vectors \mathbf{h}_{llk} , we can obtain the information of the AoAs and the magnitudes of signals. That is, the LS estimated channel contains such information of both the desired UT of its own cell and the interfering UTs of the other cells. Fig. 2 shows the DFT-processed data of the LS estimated channel, the desired channel, and the interference channels, assuming one desired UT and six interfering UTs. If the AoA of each signal does not completely overlap, each spectrum can be separated. To estimate the desired channel, we exploit the following information such as the statistical AoAs and the distance-dependent magnitudes of signals.

- 1) The large scale coefficient of the target cell β_{llk} is often larger than that of the other cells $\beta_{lj}, (l \neq j)$
- 2) The AoAs of the desired and interference signals are not exactly the same.

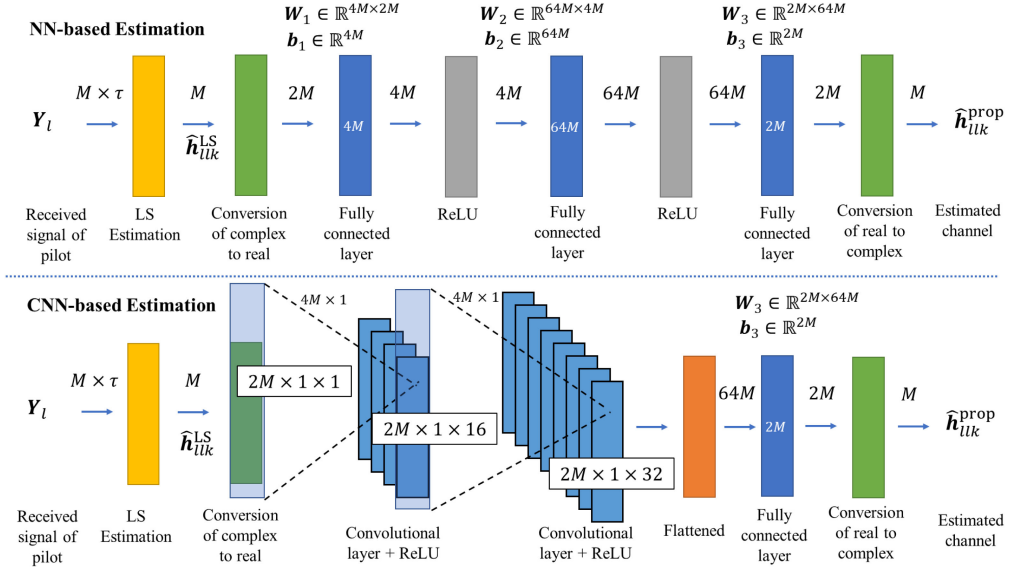


FIGURE 3. Framework for the proposed methods. The upper part and the lower part show the structure in the NN-based estimation using the fully connected layers and the CNN-based estimation using the convolutional layers, respectively.

These items are not the assumptions that must be always made since our method is not the model-based algorithm but the data-driven one. Let χ be the information related to the channel covariance matrix such as the AoA and the magnitude of signals. The MMSE estimation can be expressed as $\hat{h}_{llk}^{MMSE} = W_\chi \hat{h}_{llk}^{LS}$. However, in the practical environment, since the information χ is unknown in advance, the covariance matrix W_χ is required to be obtained from the LS estimated channel. The channel is estimated with

$$\hat{h}_{llk} = \hat{W}_\chi \hat{h}_{llk}^{LS} \quad (17)$$

$$= \mathbb{E} \left[W_\chi | \hat{h}_{llk}^{LS} \right] \hat{h}_{llk}^{LS}. \quad (18)$$

Here, \hat{W}_χ is obtained using the information contained in the LS estimated channel as follows

$$\hat{W}_\chi = \int p(\chi | \hat{h}_{llk}^{LS}) W_\chi d\chi, \quad (19)$$

where $p(\cdot)$ represents the conditional probability density function.

To realize Eq. (17), the method using neural networks is suitable since they are good at extracting features from large amounts of training data and making end-to-end inferences about unknown data. The LS estimated channel has distinguishable features, and it is necessary to assume a large amount of information χ as in Eq. (19). Therefore, we can expect that the performance of the neural network-based method approaches that of the MMSE channel estimation in which the covariance matrix information is fully known. Eq. (17) can be expressed as $\hat{h}_{llk} = f(\hat{h}_{llk}^{LS})$ because it takes the LS estimated channel as its argument. Here, f denotes a mapping from the LS estimated channel to the desired channel. Thus, the LS estimated channel can be used as the input data to the neural networks.

B. FRAMEWORK OF DEEP LEARNING-BASED CHANNEL ESTIMATION

We propose two kinds of deep learning-based channel estimation method. One is composed of fully connected layers and activation functions, the other is composed of convolutional layers and activation functions. Fig. 3 shows the framework of our proposed methods in which the neural network consists of three layers and activation functions connected with them. The upper part is a model using all three fully connected layers, and the lower part is a model of a CNN using convolutional layers. We call the method using the former model the NN-based estimation, and the latter one the CNN-based estimation.

The LS estimated channel vector $\hat{h}_{llk}^{LS} \in \mathbb{C}^M$ for the received pilot signals is given as input data of the neural network, and the desired channel vector $\hat{h}_{llk}^{prop} \in \mathbb{C}^M$ is output. The input-output correspondence is computed by the internal weight parameters of the neural network, and the underlying loss function is defined as follows

$$\mathcal{L} = \mathbb{E} \left[\|\mathbf{h}_{llk} - f(\hat{h}_{llk}^{LS})\|_2^2 \right]. \quad (20)$$

1) NN-BASED ESTIMATION

In the model of the NN-based estimation, the number of nodes is $4M$ in the first layer, $64M$ in the second layer, and $2M$ in the last layer. The rectified linear unit (ReLU), $r(x) = \max(x, 0)$, is used as the activation function. If the weight parameters at the i -th layer is denoted by W_i and b_i , the model of the NN-based estimation f_{NN} is defined as

$$f_{NN}(x) = W_3 \cdot r(W_2 \cdot r(W_1 x + b_1) + b_2) + b_3. \quad (21)$$

Here, $x \in \mathbb{R}^{2M}$ is the input for f_{NN} , and is being converted from complex into real, $x = [\text{Re}(\hat{h}_{llk}^{LS})^T, \text{Im}(\hat{h}_{llk}^{LS})^T]^T$.

Thus, the desired channel is obtained from $\hat{\mathbf{h}}_{llk}^{\text{prop}} = f(\hat{\mathbf{h}}_{llk}^{\text{LS}}) = f_{\text{NN}}(\mathbf{x})$. Here, the batch size of the input is set to one for simplicity, however in practice, the multiple batches of datasets can be handled simultaneously.

When using this estimation method, the values of the weight parameters must be determined through offline training beforehand. Given a training dataset, the weight parameters are updated in the direction that the values of the loss function Eq. (20) becomes smaller through the Adam optimizer. This training operation is performed until the values of the loss function converge. In addition, the loss function is l_2 regularized to prevent overlearning. Once the training is done, it is ready to be used online. During the pilot signal transmission of the UL, the LS estimated channel of the received signal at the BS is given as an input to the neural network. The desired channel is estimated through Eq. (21) using the trained weight parameters. The trained model contains a large number of corresponding patterns between the LS estimated channels and the desired ones. The weight parameters are determined such that Eq. (20) is small for data with the same statistical properties. Thus, even for datasets that are not identical to the training dataset, our proposed method can be used if the channel model is equal to the one in training.

In addition, it might be possible to deal with the case of hardware impairments by training with the dataset generated under such conditions. Traditional algorithms often assume a certain model, while a data-driven manner can learn how to map the LS estimated channel into more accurate channel. We would like to leave this topic as our future research.

2) CNN-BASED ESTIMATION

In the model of the CNN-based estimation, the convolutional layer is used for the first and second layers, and the fully connected layer is used for the last layer after vectorizing the feature map of matrix data. The first layer has 16 filters of $4M \times 1$, the second layer has 32 filters of $4M \times 1$, and the last layer has $2M$ nodes. The features from the input information are extracted by the convolutional layer to generate a feature map. After that, the fully connected layer converts the resulting feature map into the desired number of dimensions. As in the case of the NN-based estimation, the input \mathbf{x} is being separated into the real and imaginary parts, and thus the size of the input is $2M \times 1 \times 1$. The model in the CNN-based estimation f_{CNN} is defined as

$$f_{\text{CNN}}(\mathbf{x}) = \mathbf{W}_3 \cdot \text{vec}(r(\text{Conv}(r(\text{Conv}(\mathbf{x})))))) + \mathbf{b}_3, \quad (22)$$

where $\text{Conv}(\cdot)$ is the operation of the convolutional layer.

In general, CNNs exploit spatial local correlation with the sliding convolutional filters. As mentioned above, the LS estimated channel, which is the input data, contains information such as the AoAs and the magnitudes of signals. Therefore, the CNN plays a very powerful role in extracting these features among the input elements, and is expected to have a higher performance than the NN-based estimation. In the field of image processing, relatively small filters are used, however

TABLE 1. Simulation Parameters

Parameter	Value
Distance between BSs	2000 m
Antenna spacing	$\lambda/2$
Path loss	3.5
Cell edge SNR	0 dB
Length of pilot τ	10
Number of training data (in the proposed methods)	20000
Number of test data (in the proposed methods)	10000
Number of epochs (in the proposed methods)	20

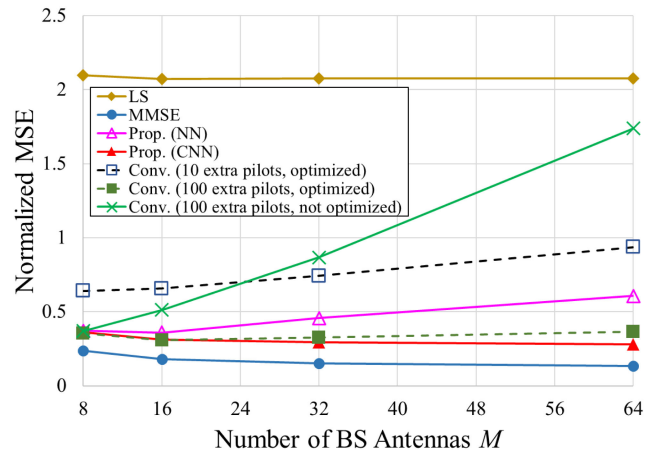


FIGURE 4. Channel NMSE versus the number of BS antennas in the case of perfect timing synchronization.

in our CNN-based estimation, the filter size is larger than the input size. This is because the input data of complex is separated into real and imaginary parts, and it is necessary to include all the elements of $2M$ in extracting the features. This point has important implications for reducing the effects of pilot contamination. Other settings and usage are the same as those of the NN-based estimation.

V. SIMULATION RESULTS

This section evaluates the performance of the proposed methods through computer simulations. The simulation parameters are listed in Table 1. We consider 7 hexagonal cells, the center of which is its own cell. The other cells are arranged around it. The UTs are located randomly, satisfying $d_{ll} \in [700, 1000]$ m, $d_{lj} \in [1000, 2000]$ m, where d_{lj} denotes the distance from the BS in the l -th cell to the UT in the j -th cell. In the simulation from Fig. 4 to Fig. 7, a single UT is distributed in each cell. Note that UTs in the same cell do not cause interference with each other since they have orthogonal pilot signals among their own cell. Thus, our results are the same even in the case that multiple UTs are distributed in each cell as long as the channel does not change so that the orthogonality among pilot signals is kept. We assume the spatial correlated channel. In this computer simulations, the channels and the corresponding signals are generated based on the system model in Section II. Each element of datasets includes the channels generated based on different covariance matrices and the corresponding received signals of pilot transmissions. That is, the noise,

channel, and covariance matrix of datasets are varying in each sample. We use the channel NMSE as the metric of a channel estimation accuracy,

$$\text{NMSE} = \frac{\mathbb{E} [\|\mathbf{h}_{lIk} - \hat{\mathbf{h}}_{lIk}\|_2^2]}{M}, \quad (23)$$

where M is the number of BS antennas. We evaluate the NMSE for the LS estimation, the MMSE estimation in which the covariance matrix is known, the method [8] in Section III, the proposed methods in Section IV.

A. CASE OF PERFECT TIMING SYNCHRONIZATION

In Fig. 4, the channel NMSE versus the number of BS antennas M is illustrated in the case of perfect timing synchronization. We compare two proposed methods (NN-based estimation and CNN-based estimation) with the LS estimation, the MMSE estimation with the known covariance matrix, and the conventional method [8]. The LS estimated channel is contaminated by the inter-cell interference due to allocating the same pilot signal as in Eq. (6), which shows the worst estimation. The MMSE estimation is conducted with the perfect covariance matrix and therefore shows the best performance as an indicator of the accuracy limit. Note here that the covariance matrix is unknown in practical environments.

The proposed methods, the NN and the CNN-based estimation, achieve much better estimation performance than the LS estimation. The NMSE of the NN-based estimation is degraded as the number of BS antennas M increases, while that of the CNN-based estimation is improved. This indicates that exploiting spatial local correlation in CNNs contribute significantly to the channel estimation in pilot contamination. In the NN-based estimation, the correlation between nodes is not well captured as the number of nodes increases. However, in the CNN-based estimation, it is possible to extract the relationship between input elements using filters. On the other hand, in our computer simulations of $M = 64$, it takes 209 seconds to train the datasets in the NN-based estimation, and 498 seconds in the CNN-based estimation. Thus, the NN-based estimation has the advantages of the speed of training.

The conventional method is evaluated with the various number of pilot samples N_R . In the figure legend, ‘optimized’ means that η, μ are selected such that the least MSE is achieved, and ‘not optimized’ means that $\eta = 1, \mu = 1$. Note that the optimal η, μ are not known in advance in practical environments. The NMSE is reduced by sending the additional number of samples N_R . In the case of $\eta = 1, \mu = 1$, significant performance degradation is observed for a larger number of BS antennas.

The conventional method has to consume many coherence blocks to improve the performance accuracy, and what is worse, the optimal parameters μ, η are not known. However, the proposed methods, by capturing features such as the information of the AoAs and the magnitudes of signals, can show better estimation performance than the conventional method without sending additional pilot signals.

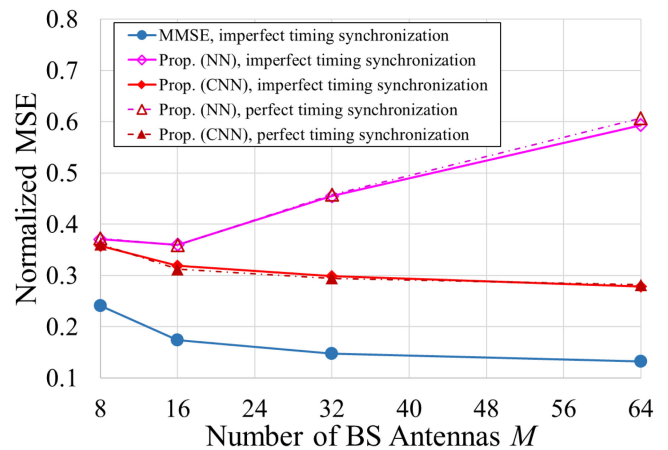


FIGURE 5. Channel NMSE versus the number of BS antennas in the case of imperfect timing synchronization.

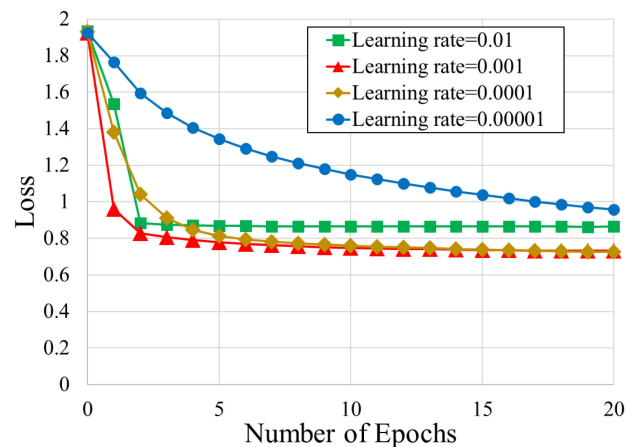


FIGURE 6. Loss value versus the number of epochs with the NN-based estimation in which the learning rate is different.

B. CASE OF IMPERFECT TIMING SYNCHRONIZATION

In Fig. 5, we show the channel NMSE versus the number of BS antennas M in the case of imperfect timing synchronization. We compare the case of imperfect timing synchronization and the perfect one with respect to the two proposed methods (NN-based estimation and CNN-based estimation). In the training, the datasets of perfect timing synchronization are used, however, the NMSE in imperfect timing synchronization is almost the same as that in perfect timing synchronization. Even if the imperfect timing synchronization occurs, our proposed methods can distinguish the desired channel from the other channels because the AoA information is kept intact.

C. HYPER PARAMETER TUNING

We show how the NMSE performance changes for the variation of the hyper parameters, i.e., the learning rates and the size of kernels. Fig. 6 shows the loss value versus the number of epochs with the NN-based estimation in which the learning rate is different. The BS antennas M is set to 64. The learning

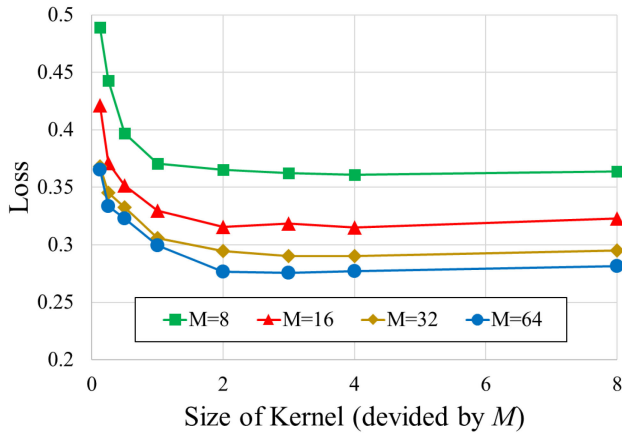


FIGURE 7. Loss value versus the size of kernel with the CNN-based estimation.

rate is the parameter to decide how much the weights of the NN is changed. In general, when the value is too large, the loss function may diverge or pass by a global minimum point. In contrast, when the value is too small, it may take a long time to converge or the loss function may stay at a local minimum point. In our result, assigning large learning rate, 0.01, converges at higher value than the others. This is because the weights of the NN fluctuate widely and do not stay a global minimum point. Fig. 6 also shows the smaller the learning rate is, the longer the time to converge takes. A similar result is obtained in the case of CNN-based estimation.

Fig. 7 shows the loss value versus the size of kernel with the CNN-based estimation. The number of BS antennas M is shown at the legend. The horizontal axis is normalized with that of BS antennas M . As the size of kernel increases, the loss value decreases drastically. However, when reaching $2M$, it reaches the bottom. This is because the input information, the LS estimated channel, is divided into real and imaginary parts. It means that handling the information of all elements simultaneously enables the CNN extracting features to improve the estimation accuracy. The slight increase in loss values between $2M$ and $8M$ is due to the fact that we simulated a fixed number of epochs, and the higher number of hyperparameters takes longer to train.

D. IMPACT OF CHANNEL AGING

We focus on the impact of channel aging in the case of perfect timing synchronization. As we mentioned in Section II-B, the channel coefficients change from symbol to symbol according to Eq. (10). The frame length is set to $N_f = 200$ symbols, and the number of UTs is $K = 2$ in this subsection. The channel varies in the part of pilot transmission and data transmission. As a result, signals from other UTs that use the same pilot signal also become interference components, even though UTs in each cell are given orthogonal pilot sequences to each other. In each frame, $\tau = 2$ symbols are used for estimating the uplink channel, and the rest of them are used for sending data. We investigate two kinds of the normalized Doppler frequency:

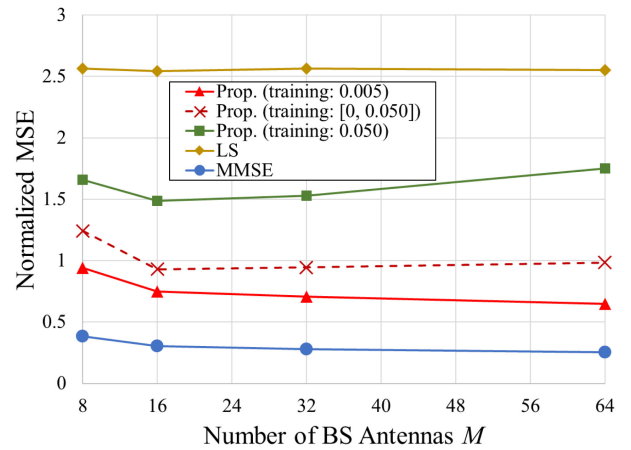


FIGURE 8. NMSE versus the number of BS antennas, which is evaluated for the channel data with the maximum Doppler shift $f_D T_s = 0.005$.

$f_D T_s = 0.005$ and $f_D T_s = 0.050$. These values correspond to the velocity of UTs $v = 27$ km/h and 270 km/h, respectively, under the case of the carrier frequency $f_c = 2$ GHz and the symbol period $T_s = 10^{-4}$ seconds.

In this subsection, the NMSE of the channel is calculated as follows

$$\text{NMSE} = \frac{1}{M} \times \mathbb{E} \left[\frac{1}{N_f} \sum_{n=1}^{N_f} \|\mathbf{h}_{llk}[n] - \hat{\mathbf{h}}_{llk}\|_2^2 \right]. \quad (24)$$

This means that one estimated channel is compared with the all varying channels in each frame. Eq. (24) is used as the loss function. That is, the estimated channel is affected by the varying channel within a frame. The parameter of the CNN will be updated so that the MSE of channel within a frame can be small. We compare the CNN-based estimation with the LS estimation and the MMSE estimation with the perfect covariance matrix. In the evaluation of the proposed method, we use multiple training datasets with different normalized Doppler shifts. Specifically, we compare the difference among three datasets used for training, where $f_D T_s$ is 0.005, 0.050 and [0,0.050]. In the case of $f_D T_s = [0, 0.050]$, the channel between each UT and the BS is generated by using randomly selected value from 0 to 0.050.

Fig. 8 shows the NMSE versus the number of BS antennas M , which is evaluated for the dataset with $f_D T_s = 0.005$. The numbers in the legend are the normalized Doppler frequency values of the training dataset. The NMSE of the proposed method trained by the dataset with $f_D T_s = 0.050$ is degraded compared to that trained by the dataset with $f_D T_s = 0.005$. Furthermore, it can be confirmed that the estimation error increases as the number of BS antennas increases. This is due to the mismatch between the model of the datasets used in training and that used in evaluation. The channel fluctuates in the part of pilot transmission and data transmission. In the CNN-based estimation, the weight parameters are determined so that Eq. (24) becomes small during training. In other words, the channel is estimated so that the average error value of the

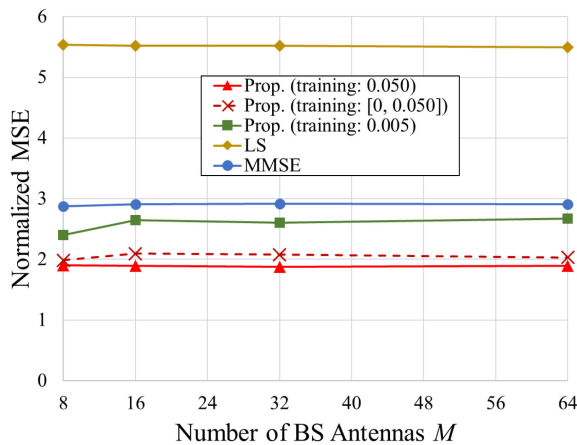


FIGURE 9. NMSE versus the number of BS antennas, which is evaluated for the channel data with the maximum Doppler shift $f_D T_s = 0.050$.

entire frame is small. If the proposed method is trained by the dataset with $f_D T_s = 0.050$, it estimates the channel matched to the variation of $f_D T_s = 0.050$, whatever the value of $f_D T_s$ in the evaluated dataset is. Therefore, if a mismatch occurs between the models of datasets in training and evaluation, the estimation accuracy deteriorates.

Fig. 9 shows the NMSE versus the number of BS antennas M , which is evaluated for the dataset with $f_D T_s = 0.050$. The channel of the dataset used in the evaluation fluctuates greatly. LS estimation and MMSE estimation try to estimate the channel by using the pilot signal. Thus, the estimated channel error increases as the channel fluctuates over time. Due to this channel aging, the NMSEs of LS and MMSE estimation evaluated by the dataset with $f_D T_s = 0.050$ are worse than those evaluated by the dataset with $f_D T_s = 0.005$. The proposed method achieves the lowest NMSE when using the dataset with $f_D T_s = 0.050$ in training. However, if trained by the dataset with $f_D T_s = 0.005$, the channel estimation accuracy deteriorates due to the mismatch between the models of datasets in training and evaluation.

In a real environment, it is difficult to perfectly match the $f_D T_s$ of the training dataset with the $f_D T_s$ of the desired channel because the speed of the UT is constantly changing. However, as Fig. 8 and Fig. 9 show, using the training dataset with $f_D T_s = [0, 0.050]$, the proposed method can bring NMSE closer to that when using the matching datasets. It is possible to maintain some estimation accuracy by training on a dataset with a mixture of various normalized Doppler shifts.

VI. CONCLUSION

In this paper, we proposed two channel estimation methods based on deep learning in TDD massive MIMO systems under the existence of pilot contamination. Simulation results confirmed that the proposed methods are better than the LS and the conventional covariance estimation in terms of the channel NMSE. Our proposed methods also have the ability

to estimate the channel in the case of imperfect timing synchronization for UTs in the neighboring cells and in the case of channel aging.

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] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, Feb. 2014.
- [3] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang, "An overview of massive MIMO: Benefits and challenges," *IEEE J. Sel. Top. Signal Process.*, vol. 8, no. 5, pp. 742–758, Oct. 2014.
- [4] E. Björnson, E. G. Larsson, and M. Debbah, "Massive MIMO for maximal spectral efficiency: How many users and pilots should be allocated?" *IEEE Trans. Wireless Commun.*, vol. 15, no. 2, pp. 1293–1308, Feb. 2016.
- [5] O. Elijah, C. Y. Leow, T. A. Rahman, S. Nunoo, and S. Z. Iliya, "A comprehensive survey of pilot contamination in massive MIMO-5G system," *IEEE Commun. Surv. Tuts.*, vol. 18, no. 2, pp. 905–923, Apr.-Jun. 2016.
- [6] H. Yin, D. Gesbert, M. Filippou, and Y. Liu, "A coordinated approach to channel estimation in large-scale multiple-antenna systems," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 264–273, Feb. 2013.
- [7] S. Haghghatshoar and G. Caire, "Massive MIMO pilot decontamination and channel interpolation via wideband sparse channel estimation," *IEEE Trans. Wireless Commun.*, vol. 16, no. 12, pp. 8316–8332, Dec. 2017.
- [8] E. Björnson, L. Sanguinetti, and M. Debbah, "Massive MIMO with imperfect channel covariance information," in *Proc. 50th Asilomar Conf. Signals, Syst., Comput.*, Nov. 2016, pp. 974–978.
- [9] A. Adhikary, J. Nam, J. Ahn, and G. Caire, "Joint spatial division and multiplexing—the large-scale array regime," *IEEE Trans. Inf. Theory*, vol. 59, no. 10, pp. 6441–6463, Oct. 2013.
- [10] R. R. Müller, L. Cottatellucci, and M. Vehkaperä, "Blind pilot decontamination," *IEEE J. Sel. Top. Signal Process.*, vol. 8, no. 5, pp. 773–786, Oct. 2014.
- [11] F. Fernandes, A. Ashikhmin, and T. L. Marzetta, "Inter-cell interference in noncooperative TDD large scale antenna systems," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 192–201, Feb. 2013.
- [12] J. Jose, A. Ashikhmin, T. L. Marzetta, and S. Vishwanath, "Pilot contamination and precoding in multi-cell TDD systems," *IEEE Trans. Wirel. Commun.*, vol. 10, no. 8, pp. 2640–2651, Aug. 2011.
- [13] H. Huang *et al.*, "Deep learning for physical-layer 5G wireless techniques: Opportunities, challenges and solutions," *IEEE Wireless Commun.*, vol. 27, no. 1, pp. 214–222, Feb. 2020.
- [14] H. Huang, J. Yang, H. Huang, Y. Song, and G. Gui, "Deep learning for super-resolution channel estimation and DOA estimation based massive MIMO system," *IEEE Trans. Veh. Technol.*, vol. 67, no. 9, pp. 8549–8560, Sep. 2018.
- [15] D. Neumann, T. Wiese, and W. Utschick, "Learning the MMSE channel estimator," *IEEE Trans. Signal Process.*, vol. 66, no. 11, pp. 2905–2917, Jun. 2018.
- [16] Y. LeCun, P. Haffner, L. Bottou, and Y. Bengio, *Object Recognition With Gradient-Based Learning*. Berlin, Germany: Springer-Verlag, 1999.
- [17] Y. Liao, Y. Hua, X. Dai, H. Yao, and X. Yang, "ChanEstNet: A deep learning based channel estimation for high-speed scenarios," in *Proc. IEEE Int. Conf. Commun.*, May 2019, pp. 1–6.
- [18] C. Chun, J. Kang, and I. Kim, "Deep learning-based channel estimation for massive MIMO systems," *IEEE Wireless Commun. Lett.*, vol. 8, no. 4, pp. 1228–1231, Aug. 2019.
- [19] L. Sanguinetti, E. Björnson, and J. Hoydis, "Toward massive MIMO 2.0: Understanding spatial correlation, interference suppression, and pilot contamination," *IEEE Trans. Commun.*, vol. 68, no. 1, pp. 232–257, Jan. 2020.
- [20] X. Gao, O. Edfors, F. Tufvesson, and E. G. Larsson, "Massive MIMO in real propagation environments: Do all antennas contribute equally?" *IEEE Trans. Commun.*, vol. 63, no. 11, pp. 3917–3928, Nov. 2015.
- [21] X. Gao, O. Edfors, F. Rusek, and F. Tufvesson, "Massive MIMO performance evaluation based on measured propagation data," *IEEE Trans. Wireless Commun.*, vol. 14, no. 7, pp. 3899–3911, Jul. 2015.

- [22] E. Björnson, J. Hoydis, and L. Sanguinetti, "Massive MIMO has unlimited capacity," *IEEE Trans. Wirel. Commun.*, vol. 17, no. 1, pp. 574–590, Jan. 2018.
- [23] A. Pitarokoilis, E. Björnson, and E. G. Larsson, "On the effect of imperfect timing synchronization on pilot contamination," in *Proc. IEEE Int. Conf. Commun.*, May 2017, pp. 1–6.
- [24] K. T. Truong and R. W. Heath, "Effects of channel aging in massive MIMO systems," *J. Commun. Netw.*, vol. 15, no. 4, pp. 338–351, Aug. 2013.
- [25] W. C. Jakes, *Microwave Mobile Communications*. Hoboken, NJ, USA: Wiley, 1974.



HIROKI HIROSE (Student Member) was born in Gifu, Japan, in 1995. He received the B.E. degree with Faculty of Science and Technology from Keio University in 2019. He is a master's student with Graduate School, Keio University. His research interest focuses on deep learning based physical-layer wireless communications. He is a member of IEICE.



TOMOAKI OHTSUKI (Senior Member, IEEE) received the B.E., M.E., and Ph. D. degrees in electrical engineering from Keio University, Yokohama, Japan in 1990, 1992, and 1994, respectively. From 1994 to 1995, he was a Postdoctoral Fellow and a Visiting Researcher with Electrical Engineering, Keio University. From 1993 to 1995, he was a Special Researcher of Fellowships of the Japan Society for the Promotion of Science for Japanese Junior Scientists. From 1995 to 2005, he was with the Science University of Tokyo. In 2005, he joined

Keio University. He is currently a Professor with Keio University. From 1998 to 1999, he was with the Department of Electrical Engineering and Computer Sciences, University of California, Berkeley. He is engaged in research on wireless communications, optical communications, signal processing, and information theory. Dr. Ohtsuki is the recipient of the 1997 Inoue Research Award for Young Scientist, the 1997 Hiroshi Ando Memorial Young Engineering Award, Ericsson Young Scientist Award 2000, 2002 Funai Information and Science Award for Young Scientist, IEEE the 1st Asia-Pacific Young Researcher Award 2001, the 5th International Communication Foundation (ICF) Research Award, 2011 IEEE SPCE Outstanding Service Award, the 27th TELECOM System Technology Award, ETRI Journal's 2012 Best Reviewer Award, and 9th International Conference on Communications and Networking in China 2014 (CHINACOM '14) Best Paper Award. He has authored or coauthored more than 190 journal papers and 400 international conference papers.

He was the Chair of IEEE Communications Society, Signal Processing for Communications and Electronics Technical Committee. He was a Technical Editor of the *IEEE Wireless Communications Magazine* and an Editor of *Elsevier Physical Communications*. He is currently an Area Editor of the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY and an Editor of the IEEE COMMUNICATIONS SURVEYS AND TUTORIALS. He was General Co-Chair, Symposium Co-Chair, and TPC Co-Chair of many conferences, including IEEE GLOBECOM 2008, SPC, IEEE ICC 2011, CTS, IEEE GLOBECOM 2012, SPC, IEEE ICC 2020, SPC, IEEE APWCS, IEEE SPAWC, and IEEE VTC. He gave tutorials and keynote speech at many international conferences including IEEE VTC, IEEE PIMRC, and so on. He was the Vice President and the President of Communications Society of the IEICE. He is a fellow of the IEICE.



GUAN GUI (Senior Member, IEEE) received the Ph.D. degree from the University of Electronic Science and Technology of China, Chengdu, China, in 2012. From 2009 to 2014, he joined the Tohoku University as a Research Assistant as well as a Postdoctoral Research Fellow, respectively. From 2014 to 2015, he was an Assistant Professor with the Akita Prefectural University. Since 2015, he has been a Professor with the Nanjing University of Posts and Telecommunications, Nanjing, China. He has authored or coauthored more than 200 IEEE

Journal/Conference papers. His recent research interests include artificial intelligence, deep learning, non-orthogonal multiple access, wireless power transfer, and physical layer security. He won several Best Paper Awards, e.g., ICC 2017, ICC 2014 and VTC 2014-Spring. He was the recipient of the Member and Global Activities Contributions Award in 2018, the Top Editor Award of IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY in 2019, the Exemplary Reviewer Award of IEEE COMMUNICATIONS LETTERS in 2017. He was also selected as for the Jiangsu Specially-Appointed Professor in 2016, the Jiangsu High-level Innovation and Entrepreneurial Talent in 2016, the Jiangsu Six Top Talent in 2018, the Nanjing Youth Award in 2018. He is serving or served on the editorial boards of several journals, including IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, *IEICE Transactions on Communications*, *Physical Communication*, *Wireless Networks*, *IEEE ACCESS*, *Security and Communication Networks*, *IEICE Communications Express*, and *KSII Transactions on Internet and Information Systems*, *Journal on Communications*. In addition, he served as the TPC Chair of WiMob 2020, Track Chairs of VTC 2020 spring, ISNCC 2020 and ICC 2020, Award Chair of PIMRC 2019, and TPC member of many IEEE international conferences, including GLOBECOM, ICC, WCNC, PIMRC, VTC, and SPAWC.