IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received April 9, 2019, accepted May 21, 2019, date of publication May 27, 2019, date of current version June 25, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2919133

Limited Feedforward for Channel Estimation in Massive MIMO With Cascaded Precoding

YINSHENG LIU[1,2], LI YOU[2,3], XIA CHEN[4], (Member, IEEE), AND KAI LIU[4]
[1]State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing 100044, China
[2]National Mobile Communications Research Laboratory, Southeast University, Nanjing 210096, China
[3]Purple Mountain Laboratories, Nanjing 211100, China
[4]School of Electronic and Information Engineering, Beijing Jiaotong University, Beijing 100044, China

Corresponding author: Yinsheng Liu (ys.liu@bjtu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61601018 and Grant 61801114, in part by the Open Research Fund of National Mobile Communications Research Laboratory Southeast University under Grant 2019D01, in part by the State Key Laboratory of Rail Traffic Control and Safety under Grant RCS2017ZT011, in part by the Beijing Jiaotong University, Natural Science Foundation of Jiangsu Province, under Grant BK20170688, and in part by the Fundamental Research Funds for the Central Universities.

ABSTRACT In this paper, we propose a novel channel estimation technique for frequency-division duplex (FDD)-based massive multiple-input multiple-output (MIMO) systems. Cascaded precoding has been adopted in FDD massive MIMO systems in order to reduce the dimension of physical channels so that the traditional channel estimation can be employed over a low-dimensional effective channel. However, due to the lack of *a priori* knowledge of the downlink channels, traditional channel estimation approaches can hardly achieve the minimum mean-square-error (MMSE) performance. To this end, we propose a limited feedforward strategy for downlink channel estimation based on the parametric model. In the parametric model, the channel frequency responses are represented by the path delays and the corresponding complex amplitudes. The path delays of uplink channels are first estimated and quantized at the base station, then fed forward to the user equipment (UE) through a dedicated feedforward link. In this way, the UE can obtain the *a priori* knowledge of the downlink channel under the assumption of the reciprocity between downlink and uplink path delays. Our analysis and simulation results show that the limited feedforward method can achieve near-MMSE performance.

INDEX TERMS Massive MIMO, channel estimation, parametric model, limited feedforward, cascaded precoding.

I. INTRODUCTION

Massive multiple-input multiple-output (MIMO) or large-scale MIMO has gained a lot of attention recently. As a key technology of the fifth generation mobile communication (5G), it has attracted lots of interest in both academia and industry [1]. Massive MIMO can greatly improve spectral efficiency and energy efficiency by deploying a large number of antennas at the base station (BS) [2], [3].

A critical issue in massive MIMO systems that must be addressed is the acquisition of the downlink channel state information (CSI). However, due to the dramatically increase of the number of antennas, traditional channel estimation approaches cannot be directly used in massive MIMO systems. To address this issue, cascaded precoding has been proposed in [4], which exploits the spatial correlation to reduce the dimension of the physical channels significantly, and traditional channel estimation techniques can be therefore used over a low-dimensional effective channel [5], [6]. The spatial correlation of the channels has also been used in [7], [8] to develop a closed-loop training technique to improve the performance.

Due to the lack of the *a priori* knowledge of the downlink channels, traditional channel estimation can hardly achieve the minimum mean-square-error (MMSE) performance. To achieve the MMSE bound, a novel channel estimation technique is developed based on the parametric channel model in [9]. In the parametric channel model, the channel frequency responses (CFR) are represented with the path delays and the corresponding complex-valued amplitudes, and the CFR can be regenerated by estimating the path delays and the complex amplitudes separately. To estimate

The associate editor coordinating the review of this manuscript and approving it for publication was Xueqin Jiang.

the path delays, estimation of signal parameters by rotational invariance technique (ESPRIT) [10] has been adopted in [9], which requires a long symbol sequence to obtain the frequency-domain covariance matrix (FCM). Even though improved approaches have been developed to reduce the sequence length [5], [11], the need of the relatively long symbol sequence still restricts the application of parametric model based channel estimation techniques in burst-type transmissions, such as the cellular systems.

As an extension of our previous work [12], we will exploit the large number of antennas in massive MIMO systems and the reciprocity between the downlink and the uplink path delays in this paper. Our purpose is to develop a near-MMSE strategy for downlink channel estimation in massive MIMO systems. Due to the presence of large number of antennas, the FCM, as well as the path delays, can be estimated at the BS using the channels at different antennas, and the long symbol sequence needed in traditional parametric model based approaches is therefore not required in massive MIMO systems. On the other hand, the frequency separation between the downlink and the uplink are about 5% of the center frequency [13] in FDD systems. For such small frequency separation, the downlink and the uplink channels can have many common features such as the path delays [14], [15]. In this case, the uplink path delays estimated at the BS can be directly used as the downlink ones.

In the proposed strategy, the path delays of uplink channels are first estimated at the BS using the uplink physical CFRs at different antennas, then quantized and fed forward to the UE through a dedicated feedforward link. In this way, UE can obtain the *a priori* knowledge of the downlink channel in advance because the downlink and uplink channels share identical path delays. As long as the UE has the knowledge of the path delays, the low-dimensional effective CFR can be regenerated by estimating the corresponding complex amplitudes. Theoretical analysis shows that the performance of channel estimation can be improved by increasing the accuracy of path delay estimation or using more quantization bits. Simulation results demonstrates that the proposed strategy can achieve near-MMSE performance given sufficient quantization bits and accurate path delay estimation.

It should be highlighted that the limited feedforward strategy can be also used for general downlink MIMO channel estimations. However, due to the lack of large amount of antennas, the estimation of FCM needs a long symbol sequence as in [9], which can be hardly realized in cellular systems. We therefore focus on the massive MIMO system which is more suitable for FCM estimation.

The rest of this paper is organized as follows. First, we will introduce the system model in Section II. Then, Section III focuses on the proposed channel estimation strategy. Theoretical analysis is presented in Section IV and the simulation results are given in Section V. Finally, the conclusions are drawn in Section VI.

II. SYSTEM MODEL

Consider an orthogonal frequency division multiplexing (OFDM) based massive MIMO system as in Fig. 1 (a), which consists of K subcarriers and an array of M antennas equipped at the BS. In Fig. 1 (a), $\mathbf{W} \in \mathcal{C}^{M\times D}$ indicates the inner precoder and $\mathbf{v}[k] \in \mathcal{C}^{D\times 1}$ indicates the outer precoder corresponding to the k-th subcarrier, where D is the dimension of the effective channel($D \ll M$). As in [4], the first D largest eigenvalues of the spatial covariance matrix (SCM) has captured most power of the downlink channels.

Denote $h_m[k]$ as the dowlink CFR at the k-th subcarrier corresponding to the m-th BS antenna. Due to the multipath propagation, $h_m[k]$ can be represented as

$$h_m[k] = \sum_{l=0}^{L-1} \alpha_m[l] e^{-j\frac{2\pi k}{T}\tau_l}, \tag{1}$$

where $\alpha_m[l]$ is the complex amplitude of the l-th path at the m-th antenna, τ_l is the l-th path delay, and T denotes the duration of an OFDM symbol. If assume the complex amplitudes are Gaussian distributed with zero mean and variance σ_l^2, then we have $\mathrm{E}(\alpha_m[l]) = 0$ and $\mathrm{E}(|\alpha_m[l]|^2) = \sigma_l^2$ with $\sum_{l=0}^{L-1} \sigma_l^2 = 1$. Furthermore, we assume the complex amplitudes corresponding to different paths are independently distributed in this paper.

When a huge number of antennas are deployed in a dense area, the channels at different antennas will be correlated. Accordingly, the spatial correlation function is represented as

$$r_s[m] \triangleq \mathrm{E}(h_{m+n}[k]h_n^*[k]), \tag{2}$$

or in a matrix form as

$$\mathbf{R}_s = \{r_s[m-n]\}_{m,n=0}^{M-1}. \tag{3}$$

From [4], the optimal inner precoder is represented by $\mathbf{W} = \mathbf{U}_s^*$ where $\mathbf{U}_s = (\mathbf{u}_s[0], \cdots, \mathbf{u}_s[D-1])$ consists of D eigenvectors corresponding to the largest eigenvalues of downlink spatial covariance matrix (SCM), $\mathbf{R}_s$. Due to the reciprocity, the downlink SCM is the same with the uplink SCM [16], [17]. Therefore, $\mathbf{R}_s$ is immediately obtained using the estimation of the uplink SCM, as will be introduced in Section III.

Denote $\mathbf{h}[k] = (h_0[k], \cdots, h_{M-1}[k])^{\mathrm{T}}$ as a channel vector which consists of the physical downlink CFRs from all the antennas at the k-th subcarrier. Then, for the optimal inner precoder, the low-dimensional effective CFR at the k-th subcarrier over the d-th eigenvector can be represented by

$$b_d[k] = \mathbf{u}_s^{\mathrm{H}}[d]\mathbf{h}[k]. \tag{4}$$

In this situation, the received signal at the k-th subcarrier over the effective CFR can be represented, from Fig. 1 (b), as

$$y[k] = \sum_{d=0}^{D-1} b_d[k]a_d[k] + z[k], \tag{5}$$

where $z[k]$ is the additive white Gaussian noise with zero mean and variance $\mathrm{E}(|z[k]|^2) = N_0$, and $a_d[k]$ denotes the

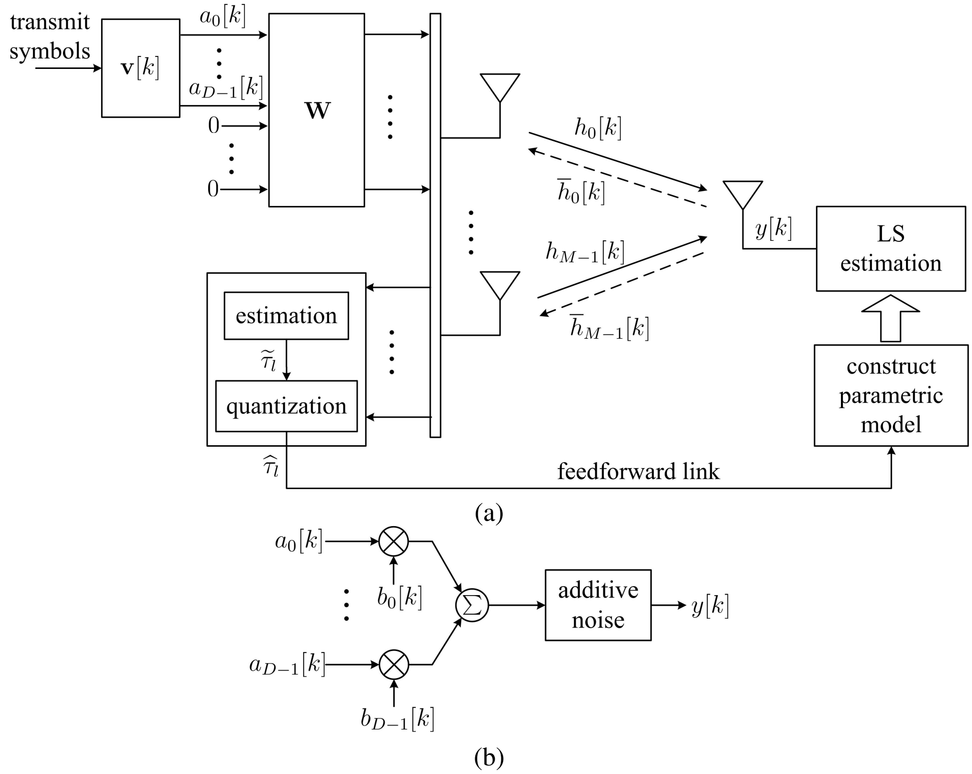

FIGURE 1. A massive MIMO system with cascaded precoding in (a) where the path delays are estimated at the BS and then fed forward to the UE through a dedicated link, and (b) an effective downlink channel model.

frequency-domain training symbol over the d-th eigenvector which is known to the UE.

Theoretically, the downlink and the uplink channels should have the same path delays since the signals will travel the same distance in both ways [18], [19]. Measurement results in [20]–[22] have shown that the power delay profiles are indeed very similar for the downlink and the uplink. It is therefore reasonable to assume identical path delays for both downlink and uplink as in [14]–[16], [23]. Experimental results in [24] have demonstrated that the assumption of identical path delays coincide with the practical measurements given small frequency separation between the downlink and the uplink.

When the downlink and the uplink channels have identical path delays, the uplink CFR at the k-th subcarrier on the m-th antenna can be expressed as

$$\overline{h}_m[k] = \sum_{l=0}^{L-1} \overline{\alpha}_m[l] e^{-j\frac{2\pi k}{T}\tau_l}, \tag{6}$$

where $\overline{\alpha}_m[l]$ is the complex amplitude of the l-th path at the m-th antenna for the uplink with zero mean and variance $\mathrm{E}(|\overline{\alpha}_m[l]|^2) = \overline{\sigma}_l^2$. Similarly, the complex amplitudes corresponding to different paths for the uplink are also assumed independent.

Since the downlink path delays in (1) and the uplink path delays in (6) are the same, the downlink path delays can be obtained at the BS by exploiting the uplink CFRs. In this way, UE can obtain *a prior* information on the downlink CFR.

III. LIMITED FEEDFORWARD CHANNEL ESTIMATION

In this section, we first investigates the parametric model based on cascaded precoding in massive MIMO systems. Then, we discuss the limited feedforward channel estimation, which includes estimation, quantization, and feedforward of the path delays, and an least-square (LS) estimation of the corresponding complex amplitudes.

A. PARAMETRIC MODEL BASED ON CASCADED PRECODING

In the parametric model, CFR is expressed via the path delays and the corresponding complex amplitude of each path [9]. Denote $\boldsymbol{\alpha}[l] = (\alpha_0[l], \cdots, \alpha_{M-1}[l])^{\mathrm{T}}$ as a complex amplitude vector which consists of the physical complex amplitudes from all the antennas at the l-th path, that is, $\alpha[l] = (\alpha_0[l], \alpha_1[l], \cdots, \alpha_{M-1}[l])^{\mathrm{T}}$. Then, from (1) and (4), the low-dimensional effective complex amplitude corresponding to the l-th path over the d-th eigenvector is obtained as

$$\beta_d[l] = \mathbf{u}_s^{\mathrm{H}}[d]\boldsymbol{\alpha}[l]. \tag{7}$$

Then, similar to (1), we can obtain the parametric channel model of the low-dimensional effective CFR as

$$b_d[k] = \sum_{l=0}^{L-1} \beta_d[l] e^{-j\frac{2\pi k}{T}\tau_l}. \tag{8}$$

From (8), the effective CFRs, $b_d[k]$'s, can be regenerated by estimating the path delays, τ_l's, and the effective complex amplitudes, $\beta_d[l]$'s, respectively.

With the parametric model in (8), the received signal in (5) can be rewritten as

$$y[k] = \sum_{d=0}^{D-1} a_d[k] \left(\sum_{l=0}^{L-1} \beta_d[l] e^{-j\frac{2\pi k}{T}\tau_l} \right) + z[k]. \quad (9)$$

Considering that there are K subcarriers in the OFDM transmission, (9) can be rewritten in a matrix form as

$$\mathbf{y} = \sum_{d=0}^{D-1} \mathbf{A}_d \mathbf{S} \boldsymbol{\beta}_d + \mathbf{z}, \quad (10)$$

where $\mathbf{y} = (y[0], \cdots, y[K-1])^{\mathrm{T}}$, $\mathbf{A}_d = \mathrm{diag}\{a_d[k]\}_{k=0}^{K-1}$, $\boldsymbol{\beta}_d = (\beta_d[0], \cdots, \beta_d[L-1])^{\mathrm{T}}$, $\mathbf{z} = (z[0], \cdots, z[K-1])^{\mathrm{T}}$, and $\mathbf{S} = [\mathbf{s}(\tau_0), \cdots, \mathbf{s}(\tau_{L-1})]$ with $\mathbf{s}(\tau_l) = [1, e^{-j\frac{2\pi}{T}\tau_l}, \cdots, e^{-j\frac{2\pi(K-1)}{T}\tau_l}]^{\mathrm{T}}$ denoting the frequency-domain steering vector. For a more tight form, (10) can be rewritten as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{z}, \quad (11)$$

where $\mathbf{X} = [\mathbf{A}_0\mathbf{S}, \cdots, \mathbf{A}_{D-1}\mathbf{S}]$ and

$$\boldsymbol{\beta} = \begin{pmatrix} \boldsymbol{\beta}_0 \\ \vdots \\ \boldsymbol{\beta}_{D-1} \end{pmatrix}. \quad (12)$$

B. PATH DELAY: ESTIMATION, QUANTIZATION, AND FEEDFORWARD

1) ESTIMATION

Since the path delays are identical for the downlink and the uplink, the estimated path delays from the uplink can be directly used as the downlink ones. The subspace-based approach, which consists of two steps, can be used for the estimation of the uplink delays at the BS [9].

The first step is to estimate the uplink FCM, which can be given as $\overline{\mathbf{R}}_f = \{\overline{r}_f[k-p]\}_{k,p=0}^{K-1}$ where $\overline{r}_f[k] \triangleq \mathrm{E}(\overline{h}_m[q+k]\overline{h}_m^*[q])$ denotes the corresponding correlation function. In massive MIMO systems, the FCM can be estimated by averaging the uplink CFRs of different antennas at the BS,

$$\widetilde{\overline{\mathbf{R}}}_f = \frac{1}{M} \sum_{m=0}^{M-1} \overline{\mathbf{h}}_m \overline{\mathbf{h}}_m^{\mathrm{H}}, \quad (13)$$

where $\overline{\mathbf{h}}_m = (\overline{h}_m[0], \cdots, \overline{h}_m[K-1])^{\mathrm{T}}$ denotes the uplink channel vector, which can be estimated via uplink channel estimation [5]. In this case, the long symbol sequence in [9], [11] is not required any more and the proposed strategy can be used in cellular systems.

For the second step, the ESPRIT algorithm can be used to obtain the estimation of the path delays as in [9]. The procedure is the same with that in [9] and thus not presented here.

The accuracy of path delay estimation can be measured by the variance, $\mathrm{E}(|\widetilde{\tau}_l - \tau_l|^2)$, where $\widetilde{\tau}_l$ denotes the corresponding estimated path delay. Although the estimation performance of ESPRIT algorithm can be improved by using more subcarriers and more antennas, it is in general difficult to obtain an analytical result [25], and we therefore use a simple notation

$$\mathrm{E}(|\widetilde{\tau}_l - \tau_l|^2) = \sigma^2, \quad (14)$$

as the performance metric of the path delay estimation, where σ^2 decreases as the numbers of subcarriers or antennas increase [25].

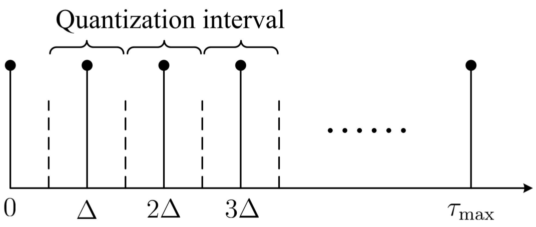

FIGURE 2. Quantization of the path delay.

2) QUANTIZATION

The estimated path delay, $\widetilde{\tau}_l$, is then quantized to, $\widehat{\tau}_l$, so that it can be fed forward to the UE. In this paper, we use a simple uniform quantization as in Fig. 2. If B bits are used for the quantization of each path delay, then the quantization interval can be given by

$$\Delta = \frac{\tau_{\max}}{2^B}, \quad (15)$$

where $\tau_{\max}$ indicates the maximum delay. In practical systems, the duration of the cyclic prefix can be used instead if the maximum delay is unknown.

For uniform quantization, the quantization error, $\widehat{\tau}_l - \widetilde{\tau}_l$, can be viewed as a uniformly distributed noise with zero mean and [26]

$$\mathrm{E}(|\widehat{\tau}_l - \widetilde{\tau}_l|^2) = \frac{\Delta^2}{12} = \frac{\tau_{\max}^2}{12 \cdot 4^B}, \quad (16)$$

which shows that the quantization performance can be improved exponentially by using more quantization bits.

3) FEEDFORWARD

After the quantization, the path delays are fed forward to the UE through a dedicated feedforward link. Similar to traditional limited feedback [6], we assume the feedforward link is error-free. On the other hand, the feedback delay in traditional limited feedback may deteriorate the system performance due to the time variation of wireless channels. In the scenario of this paper, however, the path delay depends on the surrounding scatters in typical wireless channels, which will not change for a relatively longer duration. It is therefore reasonable to assume that the path delays are constant and thus the feedback delay has no impact on the the proposed approach.

C. LS ESTIMATION OF COMPLEX AMPLITUDES

Once the UE has the knowledge of the path delays, it only needs to estimate the effective complex amplitudes. From (11), the LS estimation of the effective complex amplitudes can be given by

$$\widehat{\boldsymbol{\beta}} = (\widehat{\mathbf{X}}^{\mathrm{H}}\widehat{\mathbf{X}})^{-1}\widehat{\mathbf{X}}^{\mathrm{H}}\mathbf{y}, \tag{17}$$

where $\widehat{\mathbf{X}}$ is exactly the same with $\mathbf{X}$ except that the quantized delays, $\widehat{\tau}_l$'s, are used instead of the real ones. Note that if two or more path delays are too close to separate due to inaccurate path delay estimation or small number of quantization bits, we can view those inseparable path delays as a single one so that $\widehat{\mathbf{X}}^{\mathrm{H}}\widehat{\mathbf{X}}$ in (17) is always with full rank.

Given the quantized path delays and estimated effective complex ampltitudes, the estimated effective CFR can be regenerated, similar to (8), as

$$\widehat{b}_d[k] = \sum_{l=0}^{L-1} \widehat{\beta}_d[l]e^{-j\frac{2\pi k}{T}\widehat{\tau}_l}. \tag{18}$$

IV. PERFORMANCE ANALYSIS

In this section, we will first analyze the performance of the proposed channel estimation strategy in Section III. Then, a comparison between the proposed estimator and the MMSE estimator will be shown.

A. MSE

For cascaded precoding, the MSE can be defined over the low-dimensional effective channel, that is

$$\text{MSE} \triangleq \sum_{d=0}^{D-1}\sum_{k=0}^{K-1} \mathrm{E}(|\widehat{b}_d[k] - b_d[k]|_2^2), \tag{19}$$

where the expectation is with respect to the effective CFRs and the additive noise. If assuming the number of subcarriers is large enough, we get [27]

$$\begin{aligned}\frac{1}{K}\mathbf{s}^{\mathrm{H}}(\tau_l)\mathbf{s}(\tau_p) &= \operatorname{sinc}\left[\frac{\pi(\tau_l-\tau_p)K}{T}\right]e^{j\frac{\pi(K-1)}{T}(\tau_l-\tau_p)} \\ &\approx \begin{cases}1, & \tau_l = \tau_p \\ 0, & \tau_l \neq \tau_p\end{cases}.\end{aligned} \tag{20}$$

Using (20) and the results in Appendix A, we can obtain

$$\begin{aligned}\text{MSE} = K\sum_{l=0}^{L-1}\mathrm{Tr}\{\mathbf{U}_s^{\mathrm{H}}\mathbf{R}_{s,l}\mathbf{U}_s\}\left\{1-\operatorname{sinc}^2\left[\frac{\pi(\widehat{\tau}_l-\tau_l)K}{T}\right]\right\} \\ +N_0LD,\end{aligned} \tag{21}$$

where

$$\mathbf{R}_{s,l} = \{r_{s,l}[m-n]\}_{m,n=0}^{M-1}, \tag{22}$$

is the sub-SCM caused by the subpaths inside the l-th path with $r_{s,l}[m]$ indicating the sub-correlation-function,

$$r_{s,l}[m] \triangleq \mathrm{E}(\alpha_{n+m}[l]\alpha_n^*[l]). \tag{23}$$

From (2), it is easy to verify that $\mathbf{R}_s = \sum_{l=0}^{L-1}\mathbf{R}_{s,l}$.

In (21), the overall error is composed of a quantization error and an estimation error, that is, $\widehat{\tau}_l - \tau_l = (\widehat{\tau}_l - \widetilde{\tau}_l) + (\widetilde{\tau}_l - \tau_l)$. When the number of antennas is large, the estimation error is very small [25]. Similarly, the quantization error is also very small if assuming the number of quantization bits is sufficiently large. Under those assumptions, $\widehat{\tau}_l$ and τ_l will be very close and we therefore have

$$\operatorname{sinc}\left[\frac{\pi(\widehat{\tau}_l-\tau_l)K}{T}\right] \approx 1 - \frac{\pi^2K^2}{6T^2}(\widehat{\tau}_l-\tau_l)^2. \tag{24}$$

Using (24), the MSE expression in (21) can be rewritten as

$$\begin{aligned}\text{MSE} &= \frac{\pi^2K^3}{3T^2}\sum_{l=0}^{L-1}\mathrm{Tr}\{\mathbf{U}_s^{\mathrm{H}}\mathbf{R}_{s,l}\mathbf{U}_s\}\mathrm{E}(|\widehat{\tau}_l-\tau_l|^2) + N_0\,LD \\ &= \frac{\pi^2K^3}{3T^2}\sum_{l=0}^{L-1}\mathrm{Tr}\{\mathbf{U}_s^{\mathrm{H}}\mathbf{R}_{s,l}\mathbf{U}_s\}\Big[\mathrm{E}(|\widehat{\tau}_l-\widetilde{\tau}_l|^2) \\ &\quad + \mathrm{E}(|\widetilde{\tau}_l-\tau_l|^2)\Big] + N_0LD,\end{aligned} \tag{25}$$

where we have omitted the high-order error terms since $\widehat{\tau}_l - \tau_l$ is small. The second equation in (25) follows from the fact that the estimation error and the quantization error are independent random variables. Substituting (14) and (16) into (25), we obtain

$$\text{MSE} = \frac{\pi^2K^3}{3T^2}\mathrm{Tr}\{\mathbf{U}_s^{\mathrm{H}}\mathbf{R}_s\mathbf{U}_s\}\cdot\left(\frac{\tau_{\max}^2}{12\cdot 4^B}+\sigma^2\right)+N_0LD. \tag{26}$$

From (26), we can observe that the MSE can be reduced by adopting more quantization bits or increasing the accuracy of path delay estimation. In an extreme case where the path delay estimation is ideal and the number of quantization bits is infinite, the MSE can be reduced to

$$\text{MSE} = N_0LD, \tag{27}$$

which is proportional to the number of effective complex amplitudes as well as the dimension of the effective channel.

To obtain more insights, we rewirte the term, $\mathrm{Tr}\{\mathbf{U}_s^{\mathrm{H}}\mathbf{R}_s\mathbf{U}_s\}$, in (26) as

$$\mathrm{Tr}\{\mathbf{U}_s^{\mathrm{H}}\mathbf{R}_s\mathbf{U}_s\} = L\cdot g_D\left(\frac{1}{L}\mathbf{R}_s\right), \tag{28}$$

where $g_D(\cdot)$ denotes the sum of the D largest eigenvalues of a given matrix, that is

$$g_D\left(\frac{1}{L}\mathbf{R}_s\right) = \sum_{d=0}^{D-1}\lambda_d\left(\frac{1}{L}\mathbf{R}_s\right), \tag{29}$$

where $\lambda_d(\cdot)$ denotes the d-th largest eigenvalue of a given matrix. From [28], [29], $g_D(\cdot)$ is a convex function, and

$$g_D\left(\frac{1}{L}\mathbf{R}_s\right) = g_D\left(\frac{1}{L}\sum_{l=0}^{L-1}\mathbf{R}_{s,l}\right) \leq \frac{1}{L}\sum_{l=0}^{L-1}g_D(\mathbf{R}_{s,l}), \tag{30}$$

where the equation holds when $\mathbf{R}_{s,l} = \frac{1}{L}\mathbf{R}_s$ for all l's. In this case, we have

$$\mathrm{E}\{|\beta_d[l]|^2\} = \frac{1}{L}\mathbf{u}_s^{\mathrm{H}}[d]\mathbf{R}_s\mathbf{u}_s[d], \tag{31}$$

which means that $E\{|\beta_d[l]|^2\}$'s are constant for all l's. If assuming the D largest eigenvalues can capture all the power of $\mathbf{R}_{s,l}$, then

$$g_D(\mathbf{R}_{s,l}) = \mathrm{Tr}\{\mathbf{R}_{s,l}\} = M\sigma_l^2, \tag{32}$$

and therefore (30) can be rewritten as

$$g_D\left(\frac{1}{L}\mathbf{R}_s\right) \le \frac{M}{L}\sum_{l=0}^{L-1}\sigma_l^2 = \frac{M}{L}. \tag{33}$$

As a result, by substituting (33) into (28), we can obtain

$$\mathrm{Tr}\{\mathbf{U}_s^{\mathrm{H}}\mathbf{R}_s\mathbf{U}_s\} \le M, \tag{34}$$

where the equality holds only when the effective channel has equal power among the paths. In this case, the worst channel estimation performance is obtained, which is

$$\mathrm{MSE} = \frac{\pi^2 K^3 M}{3T^2}\left(\frac{\tau_{\max}^2}{12\cdot 4^B} + \sigma^2\right) + N_0 LD. \tag{35}$$

B. COMPARISON WITH MMSE ESTIMATOR

It is necessary to make a comparison with the MMSE estimator since it can achieve the best performance. If using the MMSE estimation for the effective CFR in (5), the associated MSE can be given as [30]

$$\mathrm{MSE_{MMSE}} = \mathrm{Tr}\left\{\left(\mathbf{R}_b^{-1} + \frac{1}{N_0}\mathbf{A}^{\mathrm{H}}\mathbf{A}\right)^{-1}\right\}, \tag{36}$$

where $\mathbf{A} = (\mathbf{A}_0, \cdots, \mathbf{A}_{D-1})$ and

$$\mathbf{R}_b = \begin{pmatrix} \mathrm{E}(\mathbf{b}_0\mathbf{b}_0^{\mathrm{H}}) & \cdots & \mathrm{E}(\mathbf{b}_0\mathbf{b}_{D-1}^{\mathrm{H}}) \\ \vdots & \ddots & \vdots \\ \mathrm{E}(\mathbf{b}_{D-1}\mathbf{b}_0^{\mathrm{H}}) & \cdots & \mathrm{E}(\mathbf{b}_{D-1}\mathbf{b}_{D-1}^{\mathrm{H}}) \end{pmatrix}, \tag{37}$$

with $\mathbf{b}_d = (b_d[0], \cdots, b_d[K-1])^{\mathrm{T}}$. Similar to [31], we use $\mathrm{E}(\mathbf{A}^{\mathrm{H}}\mathbf{A}) = \mathbf{I}$ to replace $\mathbf{A}^{\mathrm{H}}\mathbf{A}$ in (36) such that the analysis can be greatly simplified. Then, from Appendix B, the MSE in (36) can be given as

$$\mathrm{MSE_{MMSE}} = N_0\sum_{i=0}^{LD-1}\frac{\lambda_b[i]}{\lambda_b[i]+N_0}, \tag{38}$$

where $\lambda_b[i]$ indicates the i-th eigenvalue of $\mathbf{R}_b$.

Since $\lambda_b[i](\lambda_b[i]+N_0)^{-1} < 1$, we can obtain that

$$\mathrm{MSE_{MMSE}} < N_0 LD, \tag{39}$$

which means the MMSE estimator is always better than the proposed strategy. However, when the signal-to-noise ratio (SNR) is large enough, $\lambda_b[i](\lambda_b[i]+N_0)^{-1} \approx 1$ and thus

$$\mathrm{MSE_{MMSE}} \approx N_0 LD. \tag{40}$$

Compared to (27), the proposed approach can achieve the performance of the MMSE estimator if accurate path delay estimation and enough quantization bits when the SNR is large enough.

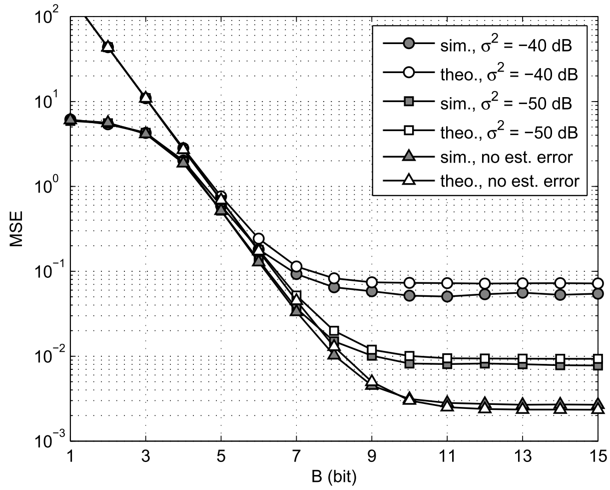

FIGURE 3. **MSE versus the quantization bit number for different variances of estimation errors with SNR = 10 dB.**

V. SIMULATION RESULTS

For the simulations, we consider an OFDM based massive MIMO system with $K = 256$ subcarriers and 15 KHz subcarrier spacing and a uniform-linear-array (ULA) with the space between neighboring antennas as half wave-length. The number of BS antennas is $M = 64$ and the dimension of the effective channel is $D = 6$ [4]. The physical channel is composed of $L = 6$ paths. An exponential power delay profile is assumed and the path delays are uniformly distributed within $[0, \tau_{\max}]$ with $\tau_{\max} = 5\mu s$. Each path has 20 unresolvable subpaths and each subpath has a random angle of departure (AoD). In practical systems, the AoDs for different paths can be distributed within a local or a rather wide range. To take various cases into account, the AoDs are assumed to be independently and uniformly distributed within a range that has a random central angle and a random angular spread uniformly distributed within $[-\pi, \pi)$ and $[0, \pi/2]$, respectively.

Fig. 3 shows the MSE versus the number of quantization bits in the cases of different variances of estimation errors for path delays. The variance of estimation errors are normalized by $\tau_{\max^2}/12$. Theoretical results in (26) are also shown in Fig. 3. It can be seen that the simulated MSEs almost coincide with the theoretical ones when B is large while there exists a gap when B is small. That is because the theoretical result is derived under the assumption that the number of quantization bits is large enough. It is also observed that for $\sigma^2 = -40$ dB, the MSE cannot be improved further more when $B > 8$. The reason is that the estimation error becomes dominant in this situation. The MSE can be further reduced by adopting more quantization bits if the accuracy of the path delay estimation improves.

Fig. 4 shows the MSE versus the variances of the estimation errors for different quantization bit numbers with SNR = 10 dB. From the figure, the MSE can be hardly improved by further increasing the estimation accuracy if the number of quantization bits is small. This is because the

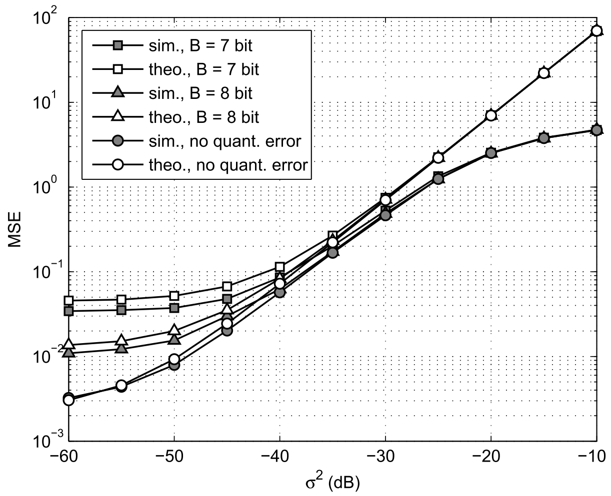

FIGURE 4. MSE versus variances of estimation errors for different quantization bit numbers with SNR = 10 dB.

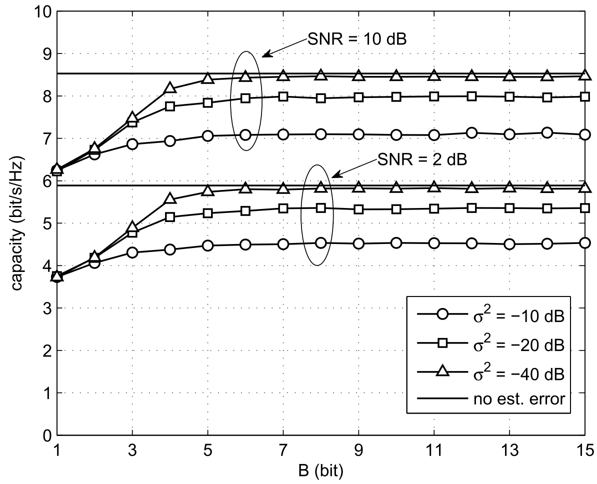

FIGURE 6. Capacity versus quantization bits with different variances of the estimation errors.

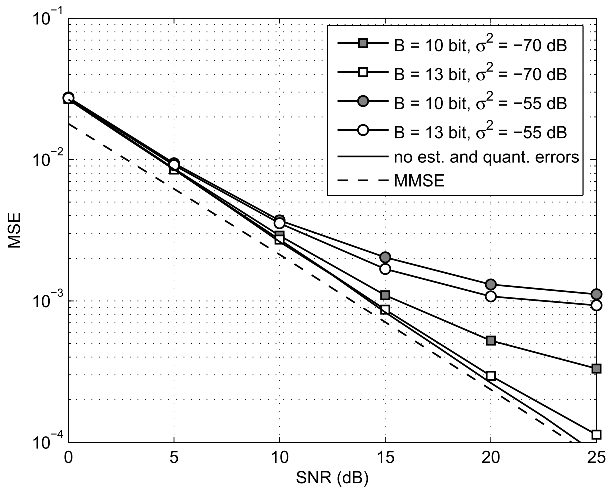

FIGURE 5. MSE versus SNR for different quantization bit numbers and variances of the estimation errors.

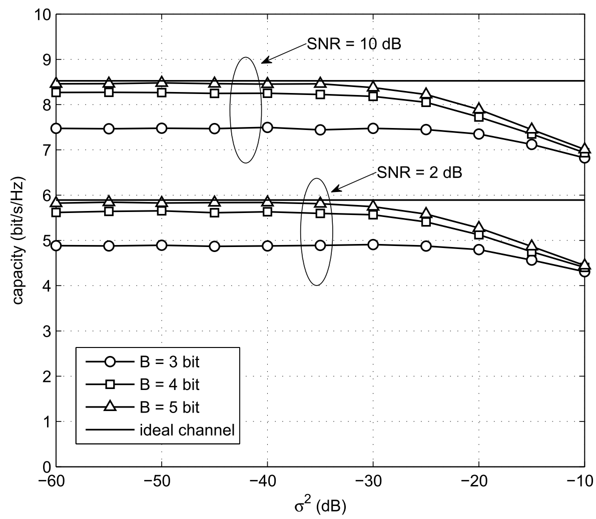

FIGURE 7. Capacity versus variances of estimation errors with different number of quantization bits.

quantization error is dominant in this situation. It can be also observed that when adopting more quantization bits, the MSE performance can be further improved as the reduction of the variance of estimation error for path delays.

Fig. 5 illustrates the MSE versus the SNR for different quantization bit numbers and variances of the estimation errors. From Fig. 5, the MSE can be improved by increasing the accuracy of path delay estimation or using more quantization bits. When the quantization bit number is large enough and the path delay estimation is accurate enough, the proposed approach can achieve near-MMSE performance in the high SNR regime. This also coincides with the previous analysis in Section IV.

Fig. 6 shows the capacity versus the number of bits for different variances of estimation errors. In this figure, the BS uses the effective CSI fed back by the UE for downlink precoding. We assume the estimated channels at the UE can be perfectly fed back to the BS such that the performance is only affected by the channel estimation error. From the figure, only 5 bits are enough to achieve the case with ideal effective CSI.

Fig. 7 shows the capacity versus variances of estimation errors with different number of quantization bits. Similar to that in Fig. 6, we still assume the estimated channels at the UE can be perfectly fed back to the BS. From the figure, the capacity can be hardly improved when $\sigma^2 < -25$ dB, which means that more accurate path delay estimation is unnecessary.

VI. CONCLUSIONS

In this paper, we have proposed a limited feedforward strategy for downlink channel estimation in FDD massive MIMO systems. Due to the reciprocity, the downlink and the uplink path delays are identical. Therefore, the path delays can be first estimated at the BS, and then quantized and fed forward to the UE through a dedicated feedforward link. In this way, the *a priori* knowledge of the downlink channel is available for UE, and the accuracy of downlink channel estimation can be therefore significantly improved. Our simulation results have shown that the proposed approach can achieve near-MMSE performance given accurate path delay estimation and sufficient quantization bits, which also coincides with our theoretical results.

APPENDIX A

The MSE for channel estimation in (19) can be rewritten in a matrix form as

$$\begin{aligned}\mathrm{MSE} &= \sum_{d=0}^{D-1} \mathrm{E}(\|\widehat{\mathbf{S}}\widehat{\boldsymbol{\beta}}_d - \mathbf{S}\boldsymbol{\beta}_d\|_2^2) \\ &= \sum_{d=0}^{D-1} \mathrm{Tr}\{\widehat{\mathbf{S}}\mathrm{E}(\widehat{\boldsymbol{\beta}}_d\widehat{\boldsymbol{\beta}}_d^{\mathrm{H}})\widehat{\mathbf{S}}^{\mathrm{H}} - \mathbf{S}\mathrm{E}(\boldsymbol{\beta}_d\widehat{\boldsymbol{\beta}}_d^{\mathrm{H}})\widehat{\mathbf{S}}^{\mathrm{H}} \\ &\quad - \widehat{\mathbf{S}}\mathrm{E}(\widehat{\boldsymbol{\beta}}_d\boldsymbol{\beta}_d^{\mathrm{H}})\mathbf{S}^{\mathrm{H}} + \mathbf{S}\mathrm{E}(\boldsymbol{\beta}_d\boldsymbol{\beta}_d^{\mathrm{H}})\mathbf{S}^{\mathrm{H}}\}. \qquad (A.1)\end{aligned}$$

Therefore, we need calculate the values of $\mathrm{E}(\boldsymbol{\beta}_d\boldsymbol{\beta}_d^{\mathrm{H}})$, $\mathrm{E}(\widehat{\boldsymbol{\beta}}_d\widehat{\boldsymbol{\beta}}_d^{\mathrm{H}})$, and $\mathrm{E}(\boldsymbol{\beta}_d\widehat{\boldsymbol{\beta}}_d^{\mathrm{H}})$, respectively.

A. CALCULATION OF $\mathrm{E}(\boldsymbol{\beta}_d\boldsymbol{\beta}_d^{\mathrm{H}})$

From (7), the correlation function of $\beta_d[l]$ can be given by $\mathrm{E}(\beta_d[l]\beta_d^*[l_1]) = \mathbf{u}_s^{\mathrm{H}}[d]\mathbf{R}_{s,l}\mathbf{u}_s[d]\delta[l-l_1]$. Therefore, $\mathbf{R}_{\beta,d} \triangleq \mathrm{E}(\boldsymbol{\beta}_d\boldsymbol{\beta}_d^{\mathrm{H}})$ is a diagonal matrix with the (l, l)-th entry given by

$$[\mathbf{R}_{\beta,d}]_{(l,l)} = \mathbf{u}_s^{\mathrm{H}}[d]\mathbf{R}_{s,l}\mathbf{u}_s[d]. \qquad (A.2)$$

B. CALCULATION OF $\mathrm{E}(\widehat{\boldsymbol{\beta}}_d\widehat{\boldsymbol{\beta}}_d^{\mathrm{H}})$ and $\mathrm{E}(\boldsymbol{\beta}_d\widehat{\boldsymbol{\beta}}_d^{\mathrm{H}})$

For the calculation of $\mathrm{E}(\widehat{\boldsymbol{\beta}}_d\widehat{\boldsymbol{\beta}}_d^{\mathrm{H}})$ and $\mathrm{E}(\boldsymbol{\beta}_d\widehat{\boldsymbol{\beta}}_d^{\mathrm{H}})$, from (17), we have

$$\widehat{\boldsymbol{\beta}} = \left(\frac{1}{K}\widehat{\mathbf{X}}^{\mathrm{H}}\widehat{\mathbf{X}}\right)^{-1}\frac{1}{K}\widehat{\mathbf{X}}^{\mathrm{H}}\mathbf{X}\boldsymbol{\beta} + \left(\frac{1}{K}\widehat{\mathbf{X}}^{\mathrm{H}}\widehat{\mathbf{X}}\right)^{-1}\frac{1}{K}\widehat{\mathbf{X}}^{\mathrm{H}}\mathbf{z}. \qquad (A.3)$$

1) CALCULATION OF $\frac{1}{K}\widehat{\mathbf{X}}^{\mathrm{H}}\widehat{\mathbf{X}}$

Note that

$$\frac{1}{K}\widehat{\mathbf{X}}^{\mathrm{H}}\widehat{\mathbf{X}} = \begin{pmatrix} \frac{1}{K}\widehat{\mathbf{S}}^{\mathrm{H}}\widehat{\mathbf{S}} & \cdots & \frac{1}{K}\widehat{\mathbf{S}}^{\mathrm{H}}\mathbf{A}_0^{\mathrm{H}}\mathbf{A}_{D-1}\widehat{\mathbf{S}} \\ \vdots & \ddots & \vdots \\ \frac{1}{K}\widehat{\mathbf{S}}^{\mathrm{H}}\mathbf{A}_{D-1}^{\mathrm{H}}\mathbf{A}_0\widehat{\mathbf{S}} & \cdots & \frac{1}{K}\widehat{\mathbf{S}}^{\mathrm{H}}\widehat{\mathbf{S}} \end{pmatrix} \qquad (A.4)$$

For the diagonal submatrices in (A.4), we have

$$\frac{1}{K}\widehat{\mathbf{S}}^{\mathrm{H}}\widehat{\mathbf{S}} = \mathbf{I}, \qquad (A.5)$$

where we have used the identity in (20).

For the (d_1, d_2)-th off-diagonal submatrix $(d_1 \neq d_2)$ in (A.4),

$$\begin{aligned}&\frac{1}{K}\widehat{\mathbf{S}}^{\mathrm{H}}\mathbf{A}_{d_1}^{\mathrm{H}}\mathbf{A}_{d_2}\widehat{\mathbf{S}} \\ &= \frac{1}{K}\begin{bmatrix} \mathbf{s}^{\mathrm{H}}(\widehat{\tau}_0)\mathbf{A}_{d_1}^{\mathrm{H}} \\ \vdots \\ \mathbf{s}^{\mathrm{H}}(\widehat{\tau}_{L-1})\mathbf{A}_{d_1}^{\mathrm{H}} \end{bmatrix}[\mathbf{A}_{d_2}\mathbf{s}(\widehat{\tau}_0), \cdots, \mathbf{A}_{d_2}\mathbf{s}(\widehat{\tau}_{L-1})], \qquad (A.6)\end{aligned}$$

the (l_1, l_2)-th entry is given by

$$\frac{1}{K}\mathbf{s}^{\mathrm{H}}(\widehat{\tau}_{l_1})\mathbf{A}_{d_1}^{\mathrm{H}}\mathbf{A}_{d_2}\mathbf{s}(\widehat{\tau}_{l_2}) = \frac{1}{K}\sum_{k=0}^{K-1} e^{j\varphi[k]}, \qquad (A.7)$$

where $\varphi[k] = \phi[k] + \frac{2\pi k}{T}(\widehat{\tau}_{l_1} - \widehat{\tau}_{l_2})$ with $\phi[k] = \phi_{d_2}[k] - \phi_{d_1}[k]$. Since $\phi_d[k]$'s are uniformly distributed within $[-\pi, \pi)$ and mutually independent for different subcarriers and different eigen-beams, $\phi[k]$'s with $k = 0, 1, \cdots, K-1$ can be viewed as independently identically distributed random variables with zero means and a common probability density function (pdf)

$$p_X(x) = \begin{cases} \frac{1}{2\pi} - \frac{|x|}{4\pi^2} & |x| < 2\pi \\ 0 & \text{otherwise} \end{cases}. \qquad (A.8)$$

Accordingly, $\varphi[k]$'s with $k = 0, 1, \cdots, K-1$ are also independently distributed random variables but with mean $\mathrm{E}(\varphi[k]) = \frac{2\pi k}{T}(\widehat{\tau}_{l_1} - \widehat{\tau}_{l_2})$. Since the functions of independent random variables are still independent, $e^{j\varphi[k]}$'s are therefore independent for different subcarriers. Then, following the law of large numbers, we have

$$\begin{aligned}\frac{1}{K}\sum_{k=0}^{K-1} e^{j\varphi[k]} &= \frac{1}{K}\sum_{k=0}^{K-1}\mathrm{E}(e^{j\varphi[k]}) \\ &= \frac{1}{K}\sum_{k=0}^{K-1} e^{j\frac{2\pi k}{T}(\widehat{\tau}_{l_1}-\widehat{\tau}_{l_2})}\mathrm{E}(e^{j\phi[k]}). \qquad (A.9)\end{aligned}$$

Using the pdf in (A.8), we have

$$\mathrm{E}(e^{j\phi[k]}) = \int_{-2\pi}^{2\pi}\left(\frac{1}{2\pi} - \frac{|\phi[k]|}{4\pi^2}\right)e^{j\phi[k]}d\phi[k] = 0. \qquad (A.10)$$

Substituting (A.10) into (A.9), we obtain

$$\frac{1}{K}\mathbf{s}^{\mathrm{H}}(\widehat{\tau}_{l_1})\mathbf{A}_{d_1}^{\mathrm{H}}\mathbf{A}_{d_2}\mathbf{s}(\widehat{\tau}_{l_2}) = \frac{1}{K}\sum_{k=0}^{K-1} e^{j\varphi[k]} = 0, \qquad (A.11)$$

and therefore,

$$\frac{1}{K}\widehat{\mathbf{S}}^{\mathrm{H}}\mathbf{A}_{d_1}^{\mathrm{H}}\mathbf{A}_{d_2}\widehat{\mathbf{S}} = \mathbf{0}, \qquad (A.12)$$

for $d_1 \neq d_2$. In other words, the off-diagonal submatrices in (A.4) are all zeros when the number of subcarriers is large enough.

As a result,

$$\frac{1}{K}\widehat{\mathbf{X}}^{\mathrm{H}}\widehat{\mathbf{X}} = \mathbf{I}. \qquad (A.13)$$

2) CALCULATION OF $\frac{1}{K}\widehat{\mathbf{X}}^{\mathrm{H}}\mathbf{X}$

Similar to the derivation above, we can obtain that

$$\frac{1}{K}\widehat{\mathbf{X}}^{\mathrm{H}}\mathbf{X} = \begin{pmatrix} \frac{1}{K}\widehat{\mathbf{S}}^{\mathrm{H}}\mathbf{S} & & \\ & \ddots & \\ & & \frac{1}{K}\widehat{\mathbf{S}}^{\mathrm{H}}\mathbf{S} \end{pmatrix}, \quad \text{(A.14)}$$

since the off-diagonal submatrices are zeros. If we assume the quantized delays, $\widehat{\tau}_l$'s, are close to the real delays, τ_l's, then $\frac{1}{K}\widehat{\mathbf{S}}^{\mathrm{H}}\mathbf{S}$ can be approximated by

$$\frac{1}{K}\widehat{\mathbf{S}}^{\mathrm{H}}\mathbf{S} = \Lambda, \quad \text{(A.15)}$$

where Λ is an $L \times L$ diagonal matrix with the (l, l)-th entry

$$[\Lambda]_{(l,l)} = \mathrm{sinc}\left[\frac{\pi(\tau_l - \widehat{\tau}_l)K}{T}\right] e^{j\frac{\pi(K-1)(\tau_l - \widehat{\tau}_l)}{T}}, \quad \text{(A.16)}$$

and therefore

$$\frac{1}{K}\widehat{\mathbf{X}}^{\mathrm{H}}\mathbf{X} = \begin{pmatrix} \Lambda & & \\ & \ddots & \\ & & \Lambda \end{pmatrix}. \quad \text{(A.17)}$$

Using (A.13) and (A.17), (A.3) can be simplified as

$$\widehat{\boldsymbol{\beta}}_d = \Lambda\boldsymbol{\beta}_d + \frac{1}{K}\mathbf{S}^{\mathrm{H}}\mathbf{A}_d\mathbf{z}. \quad \text{(A.18)}$$

It is therefore easy to obtain that

$$\mathrm{E}(\widehat{\boldsymbol{\beta}}_d\widehat{\boldsymbol{\beta}}_d^{\mathrm{H}}) = \Lambda\mathbf{R}_{\beta,d}\Lambda^{\mathrm{H}} + \frac{N_0}{K}\mathbf{I}, \quad \text{(A.19)}$$

$$\mathrm{E}(\boldsymbol{\beta}_d\widehat{\boldsymbol{\beta}}_d^{\mathrm{H}}) = \mathbf{R}_{\beta,d}\Lambda^{\mathrm{H}}. \quad \text{(A.20)}$$

Using (A.2), (A.19) and (A.20), (A.1) can be rewritten as

$$\begin{aligned}\mathrm{MSE} &= \sum_{d=0}^{D-1}\mathrm{Tr}\left\{\widehat{\mathbf{S}}\left(\Lambda\mathbf{R}_{\beta,d}\Lambda^{\mathrm{H}} + \frac{N_0}{K}\mathbf{I}\right)\widehat{\mathbf{S}}^{\mathrm{H}} - \mathbf{S}\mathbf{R}_{\beta,d}\Lambda^{\mathrm{H}}\widehat{\mathbf{S}}^{\mathrm{H}} \right. \\ &\quad \left. -\widehat{\mathbf{S}}\Lambda\mathbf{R}_{\beta,d}\mathbf{S}^{\mathrm{H}} + \mathbf{S}\mathbf{R}_{\beta,d}\mathbf{S}^{\mathrm{H}}\right\} \\ &= \sum_{d=0}^{D-1}\left\{K\sum_{l=0}^{L-1}[\mathbf{R}_{\beta,d}]_{(l,l)}(1 - |[\Lambda]_{(l,l)}|^2) + N_0 L\right\} \\ &= K\sum_{l=0}^{L-1}\sum_{d=0}^{D-1}[\mathbf{R}_{\beta,d}]_{(l,l)}\left\{1 - \mathrm{sinc}^2\left[\frac{\pi(\widehat{\tau}_l - \tau_l)K}{T}\right]\right\} \\ &\quad + N_0 LD. \end{aligned} \quad \text{(A.21)}$$

From (A.2), we have

$$\sum_{d=0}^{D-1}[\mathbf{R}_{\beta,d}]_{(l,l)} = \mathrm{Tr}\{\mathbf{U}_s^{\mathrm{H}}\mathbf{R}_{s,l}\mathbf{U}_s\}. \quad \text{(A.22)}$$

As a result, the MSE of channel estimation can be finally obtained as

$$\mathrm{MSE} = K\sum_{l=0}^{L-1}\mathrm{Tr}\{\mathbf{U}_s^{\mathrm{H}}\mathbf{R}_{s,l}\mathbf{U}_s\}\left\{1 - \mathrm{sinc}^2\left[\frac{\pi(\widehat{\tau}_l - \tau_l)K}{T}\right]\right\} + N_0 LD. \quad \text{(A.23)}$$

APPENDIX B

Replace $\mathbf{A}^{\mathrm{H}}\mathbf{A}$ with $\mathrm{E}(\mathbf{A}^{\mathrm{H}}\mathbf{A}) = \mathbf{I}$, then, similar to the derivation in [31], the MSE in (36) can be expressed by

$$\mathrm{MSE}_{\mathrm{MMSE}} = N_0\sum_{i=0}^{DK-1}\frac{\lambda_b[i]}{\lambda_b[i] + N_0}. \quad \text{(B.1)}$$

From (8), we have $\mathbf{b}_d = \mathbf{S}\boldsymbol{\beta}_d$ and thus $\mathbf{R}_b$ can be rewritten as

$$\mathbf{R}_b = (\mathbf{I}\otimes\mathbf{S})\mathbf{R}_\beta(\mathbf{I}\otimes\mathbf{S}^{\mathrm{H}}), \quad \text{(B.2)}$$

where $\otimes$ denotes the Kronecker product and

$$\mathbf{R}_\beta = \begin{pmatrix} \mathrm{E}(\boldsymbol{\beta}_0\boldsymbol{\beta}_0^{\mathrm{H}}) & \cdots & \mathrm{E}(\boldsymbol{\beta}_0\boldsymbol{\beta}_{D-1}^{\mathrm{H}}) \\ \vdots & \ddots & \vdots \\ \mathrm{E}(\boldsymbol{\beta}_{D-1}\boldsymbol{\beta}_0^{\mathrm{H}}) & \cdots & \mathrm{E}(\boldsymbol{\beta}_{D-1}\boldsymbol{\beta}_{D-1}^{\mathrm{H}}) \end{pmatrix}. \quad \text{(B.3)}$$

From Appendix A, $\mathrm{E}(\boldsymbol{\beta}_{d_1}\boldsymbol{\beta}_{d_2}^{\mathrm{H}})$ is a diagonal matrix and thus $\mathrm{rank}\{\mathrm{E}(\boldsymbol{\beta}_{d_1}\boldsymbol{\beta}_{d_2}^{\mathrm{H}})\} = L$. Accordingly,

$$\mathrm{rank}\{\mathbf{R}_\beta\} = DL \quad \text{(B.4)}$$

Actually, rank deficiency of $\mathbf{R}_\beta$ is a very strong condition that can be hardly achieved in practical engineering problems, and therefore $\mathbf{R}_\beta$ is always with full rank.

When the number of subcarriers is very large, the relation in (20) means that the columns of $\mathbf{S}$ are mutually orthogonal and therefore $\mathbf{S}$ is full column rank matrix. As a result, we can obtain from (B.2) that

$$\mathrm{rank}\{\mathbf{R}_b\} = \mathrm{rank}\{\mathbf{R}_\beta\} = DL. \quad \text{(B.5)}$$

In other words, there are only DL significant eigenvalues for $\mathbf{R}_b$ while the others are very small and thus can be omitted. As a result, the MSE in (B.1) is reduced to

$$\mathrm{MSE}_{\mathrm{MMSE}} = N_0\sum_{i=0}^{DL-1}\frac{\lambda_b[i]}{\lambda_b[i] + N_0}, \quad \text{(B.6)}$$

which is exactly (38).

REFERENCES

[1] F. Rusek, D. Perrsson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, Jan. 2013.

[2] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Energy and spectral efficiency of very large multiuser MIMO systems," *IEEE Trans. Commun.*, vol. 61, no. 4, pp. 1436–1449, Apr. 2013.

[3] 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.

[4] A. Adhikary, J. Nam, J.-Y. 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.

[5] Y. Liu, Z. Tan, H. Hu, L. J. Cimini, and G. Y. Li, "Channel estimation for OFDM," *IEEE Commun. Survey Tuts.*, vol. 16, no. 4, pp. 1891–1980, 4th Quart., 2014.

[6] D. J. Love, R. W. Heath, Jr., V. K. Lau, D. Gesbert, B. D. Rao, and M. Andrews, "An overview of limited feedback in wireless communication systems," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 8, pp. 1341–1365, Oct. 2008.

[7] J. Choi, D. J. Love, and P. Bidigare, "Downlink training techniques for FDD massive MIMO systems: Open-loop and closed-loop training with memory," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 802–814, Oct. 2014.

[8] A. J. Duly, T. Kim, D. J. Love, and J. V. Krogmeier, "Closed-loop beam alignment for massive MIMO channel estimation," *IEEE Commun. Lett.*, vol. 18, no. 8, pp. 1439–1442, Aug. 2015.

[9] B. Yang, K. B. Letaief, R. S. Cheng, and Z. Cao, "Channel estimation for OFDM transmission in multipath fading channels based on parametric channel modeling," *IEEE Trans. Commun.*, vol. 49, no. 3, pp. 467–479, Mar. 2001.

[10] R. Roy and T. Kailath, "Esprit-estimation of signal parameters via rotational invariance techniques," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 7, pp. 984–995, Jul. 1989.

[11] M. R. Raghavendra and K. Giridhar, "Improving channel estimation in OFDM systems for sparse multipath channels," *IEEE Signal Process. Lett.*, vol. 12, no. 1, pp. 52–55, Jan. 2005.

[12] K. Liu, C. Tao, L. Liu, Y. Lu, T. Zhou, and J. Qiu, "Analysis of downlink channel estimation based on parametric model in massive MIMO systems," in *Proc. 12th Int. Symp. Antennas, Propag. EM Theory (ISAPE)*, pp. 1–4, Dec. 2018.

[13] *Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA); Physical Channels Modulation*, document TS 36.211, 3GPP, Sep. 2008.

[14] A. J. Paulraj and B. C. Ng, "Space-time modems for wireless personal communications," *IEEE Pers. Commun.*, vol. 1, no. 5, pp. 36–48, Feb. 1998.

[15] A. J. Paulraj and C. B. Papadias, "Space-time processing for wireless communications," *IEEE Signal Process. Mag.*, vol. 14, no. 6, pp. 49–83, Nov. 1997.

[16] C. Sun, X. Q. Gao, S. Jin, M. Matthaiou, Z. Ding, and C. Xiao, "Beam division multiple access transmission for massive MIMO communications," *IEEE Trans. Commun.*, vol. 63, no. 6, pp. 2170–2184, Jun. 2015.

[17] G. Barriac and U. Madhow, "Space-time communication for OFDM with implicit channel feedback," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3111–3129, Dec. 2004.

[18] Y. I. Wu and K. T. Wong, "A geometrical model for the toa distribution of uplink/downlink multipaths, assuming scatterers with a conical spatial density," *IEEE Antennas Propag. Mag.*, vol. 50, no. 6, pp. 196–205, Dec. 2008.

[19] A. Y. Olenko, K. T. Wong, S. A. Qasmi, and J. Ahmadi-Shokouh, "Analytically derived uplink/downlink TOA and 2-D-DOA distributions with scatterers in a 3-D hemispheroid surrounding the mobile," *IEEE Trans. Antennas Propag.*, vol. 54, no. 9, pp. 2446–2454, Sep. 2006.

[20] S. Salous and H. Gokalp, "Medium- and large-scale characterization of UMTS-allocated frequency division duplex channels," *IEEE Trans. Veh. Technol.*, vol. 56, no. 5, pp. 2831–2843, Sep. 2007.

[21] S. Salous, "Urban and suburban frequency division duplex measurements for UMTS," in *Proc. 13th IEEE Int. Symp. Pers., Indoor Mobile Radio Commun.*, vol. 2, Sep. 2002, pp. 624–628.

[22] S. Salous and H. Gokalp, "Dual-frequency sounder for UMTS frequency-division duplex channels," *IEE Proc.-Commun.*, vol. 149, no. 2, pp. 117–122, Apr. 2002.

[23] G. G. Raleigh and V. K. Jones, "Adaptive antenna transmission for frequency duplex digital wireless communication," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Montreal, QC, USA, Jun. 1997, pp. 641–646.

[24] N. Promsuvana and P. Uthansakul, "Feasibility of adaptive 4×4 MIMO system using channel reciprocity in FDD mode," in *Proc. 14th Asia–Pacific Conf. Commun.*, Oct. 2008, pp. 1–5.

[25] F. Li, H. Liu, and R. J. Vaccaro, "Performance analysis for DOA estimation algorithms: Unification, simplification, and observations," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 29, no. 4, pp. 1170–1184, Oct. 1993.

[26] B. Widrow and I. Kollar, and M.-C. Liu, "Statistical theory of quantization," *IEEE Trans. Instrum. Meas.*, vol. 45, no. 2, pp. 353–361, Apr. 1996.

[27] M. Viberg, B. Ottersten, and A. Nehorai, "Performance analysis of direction finding with large arrays and finite data," *IEEE Trans. Signal Process.*, vol. 43, no. 2, pp. 469–477, Feb. 1995.

[28] A. Ben-Tal and A. Nemirovski. *Convex Optimization in Engineering: Modeling, Analysis, Algorithms*. [Online]. Available: http://www.st.ewi.tudelft.nl/roos/courses/WI4218/tud00r.pdf

[29] W. Watkins, "Convex matrix functions," *Proc. Amer. Math. Soc.*, vol. 44, no. 1, pp. 31–34, May 1974.

[30] S. Haykin, *Adaptive Filter Theory*, 4th ed. Beijing, China: Electronics Industry, 2010.

[31] O. Edfors, M. Sandell, J. J. van de Beek, S. K. Wilson, and P. O. Börjesson, "OFDM channel estimation by singular value decomposition," *IEEE Trans. Commun.*, vol. 46, no. 7, pp. 931–939, Jul. 1998.

YINSHENG LIU received the B.E. and Ph.D. degrees in communication and information systems from Beijing Jiaotong University, Beijing, China, in 2009 and 2015, respectively. From 2012 to 2013, he was a Visiting Scholar with the Georgia Institute of Technology, Atlanta, GA, USA. He is currently an Associate Professor with the State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University. He has authored over ten journal papers published in international journals as the first author. His research interests include channel estimation and equalization, new waveform design in next generation cellular systems, and generally signal processing with the emphasis on wireless communications. He was a recipient of the IBM Excellent Student Award and the Excellent Graduate Award of Beijing Jiaotong University, and the BJTU graduate scholarship during his master and Ph.D. programs.

LI YOU received the B.E. and M.E. degrees in electrical engineering from the Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2009 and 2012, respectively, and the Ph.D. degree in electrical engineering from Southeast University, Nanjing, in 2016. From 2014 to 2015, he conducted visiting research at the Center for Pervasive Communications and Computing, University of California at Irvine, Irvine, CA, USA. Since 2016, he has been with the Faculty of the National Mobile Communications Research Laboratory, Southeast University. His research interests include the general areas of communications, signal processing, and information theory, with the current emphasis on massive MIMO communications.

XIA CHEN (M'14) received the B.Eng. and Ph.D. degrees from Beijing Jiaotong University, in 1997 and 2003, respectively. She has been with Beijing Jiaotong University, since 2003, where she is currently a Lecturer with the School of Electronic and Information Engineering. Her research interests include mobile channel modeling, multicarrier transmission, and radio resource management of wireless networks. She is a member of the Chinese Institute of Electronics.

KAI LIU received the B.E. degree from Dalian Nationalities University, Dalian, China, in 2009 and the M.S. degree from the China University of Geosciences, Beijing, China, in 2013. He is currently pursuing the Ph.D. degree with the School of Electronic and Information Engineering, Beijing Jiaotong University. His research interests include signal processing and channel modeling in massive MIMO systems.