

A Data-Aided Channel Estimation Scheme for Decoupled Systems in Heterogeneous Networks

Wen Liu, *Student Member, IEEE*, Kai-Kit Wong, *Fellow, IEEE*,

Shi Jin, *Senior Member, IEEE*, and Xiaohu You, *Fellow, IEEE*

Abstract

Uplink/downlink (UL/DL) decoupling promises more flexible cell association and higher throughput in heterogeneous networks (HetNets), however, it hampers the acquisition of DL channel state information (CSI) in time-division-duplex (TDD) systems due to different base stations (BSs) connected in UL/DL. In this paper, we propose a novel data-aided (DA) channel estimation scheme to address this problem by utilizing decoded UL data to exploit CSI from received UL data signal in decoupled HetNets where a massive multiple-input multiple-output BS and dense small cell BSs are deployed. We analytically estimate BER performance of UL decoded data, which are used to derive an approximated normalized mean square error (NMSE) expression of the DA minimum mean square error (MMSE) estimator. Compared with the conventional least square (LS) and MMSE, it is shown that NMSE performances of all estimators are determined by their signal-to-noise ratio (SNR)-like terms and there is an increment consisting of UL data power, UL data length and BER values in the SNR-like term of DA method, which suggests DA method outperforms the conventional ones in any scenarios. Higher UL data power, longer UL data length and better BER performance lead to more accurate estimated channels with DA method. Numerical results verify that the analytical BER and NMSE results are close to the simulated ones and a remarkable gain in both NMSE and DL rate can be achieved by DA method in multiple scenarios with different modulations.

Index Terms

Channel estimation, data-aided, decoupled system, heterogeneous network, massive MIMO.

W. Liu, S. Jin and X. You are with the National Mobile Communications Research Laboratory, Southeast University, Nanjing 210096, China (e-mail: newen@seu.edu.cn; jinshi@seu.edu.cn; xhyu@seu.edu.cn).

K.-K. Wong is with the Department of Electronic and Electrical Engineering, University College London, London WC1E 7JE, U.K. (e-mail: kai-kit.wong@ucl.ac.uk).

I. INTRODUCTION

Heterogeneous networks (HetNets) and massive multiple-input multiple-output (MIMO) are regarded as two key technologies for 5G, promising higher coverage and spectral efficiency [1]–[4]. To meet the demands of $1000\times$ increase of throughput and $1000\times$ less energy consumption at the same time [5], [6], massive MIMO base station (BS) is recently introduced to HetNets, where small cells such as pico- or femto-cells are densely deployed. Such model is referred to as massive MIMO HetNets [7]–[10]. It is anticipated that the massive MIMO HetNet architecture will greatly improve throughput and regional coverage, and is also desirable for interference management and energy efficiency [11], [12].

Different from homogeneous networks, cell association in HetNets is a tricky business to deal with. From the perspective of user equipment (UE), the BS with maximum downlink (DL) power is probably not the one with maximum uplink (UL) power. In [7], the notion of decoupling UL and DL was first introduced and UEs can connect to the BSs with highest signal-to-noise ratio (SNR) in both directions separately. With such flexible cell association policy, throughput and coverage probability are found to be much improved especially for cell edge UEs by theoretical analysis and simulation results in [13]. Similar positive results of load balancing and energy efficiency were presented in [14]–[16]. Hence, prior literature has adopted decoupling access as a strong candidate in the next generation network, though there are still major problems to be solved before facilitating this structure, for instance, channel estimation at the DL BS, DL precoding, signal synchronization, offloading techniques and so on [17].

Channel estimation is an important issue in wireless communication and also considered as a major bottleneck for massive MIMO systems [1], [18], [19]. A standard approach in time-division-duplex (TDD) massive MIMO systems is to exploit channel reciprocity such that channels are estimated by pilot-based training in UL and then treated as the real channels for DL beamforming. Unfortunately, channel reciprocity cannot be used directly in decoupled HetNets since UEs may associate with different BSs in UL/DL. Although DL BS can estimate DL channels of decoupled UEs from the received UL training signal, but it should be noted that unlike coupled UEs, decoupled UEs are those who are neither close to its UL BS nor DL BS, and can hardly have sound channel estimation performance.

This motivates us to propose a data-aided channel estimation scheme to enable decoupling access in TDD systems, especially for massive MIMO where a large number of channel elements

have to be estimated. Using decoded data to aid channel estimation has been studied in [20]–[27] before. In [24], [25], the mutual information and a capacity lower bound for data-aided single-user MIMO systems were investigated and it was shown that data-aided methods permit the use of a very small number of pilots to achieve high spectral efficiency. An iterative joint channel estimation and data detection process was also investigated in [20], [22], [26], [27], finding that data-aided methods can effectively suppress the contamination effect in large-scale antenna systems [20], [22]. Nevertheless, prior works mainly focused on homogeneous networks and it is of significant importance to bring this idea into UL/DL decoupled HetNets for more efficient and reliable channel acquisition. To the best of authors' knowledge, it is the first time that channel estimation problem is discussed in cellular HetNets with decoupling access.

In this paper, we consider a single-cell HetNet with decoupling access where a macro base station (MBS) is in the center and small cell BSs (SBSs) are randomly but densely populated within the cell. A novel three-stage data-aided scheme is proposed to solve the DL channel acquisition problem of decoupled UEs at MBS¹. To implement the data-aided scheme, once data detection is finished at SBSs, the decoded data sequences and estimated bit error ratio (BER) values are supposed to be sent to the MBS via wired backhaul which is assumed to be error-free and latency-free. Then, the MBS utilizes known training sequences along with these decoded sequences and BER values to recover channels of decoupled UEs from received training and UL data signals. The core idea is to employ the decoded data as extended training sequences to exploit the most channel information from received UL signals although the decoded data is not orthogonal and subject to unknown errors. Different from the multi-cell model with one UE in each cell discussed in [20], [22], we consider a single-cell model² with dense small cells and multiple UEs where interference is much more severe. Furthermore, in [20], [22], Gaussian data was assumed to be received at the least-square (LS) data estimator and the data estimation error was modelled as a Gaussian variable while in our scheme, coded data is detected by a minimum mean-square-error (MMSE) data estimator and the BER is estimated by analytical derivation. However, similar conclusions are drawn that estimation performance can be greatly improved and both co-channel interference and detection error could seriously compromise performance.

More specifically, our main contribution is to develop a novel data-aided channel estimation

¹In two-tier HetNets, decoupled UEs always connect to an SBS in UL and an MBS in DL due to the user association policy.

²Here, single-cell represents single macro-cell, yet there are multiple overlapping small cells actually. This model can be generalized to multi-cell scenarios straightforwardly.

scheme for decoupled UEs in cellular decoupled HetNets by using decoded UL data with consideration of BER. We calculate the BER of BPSK-modulated UL data after the MMSE decoder with imperfect channel state information (CSI) and use this estimated BER to derive a closed-form approximated normalized mean-square-error (NMSE) expression for the proposed data-aided scheme. We further compare NMSE performance between conventional LS, MMSE and data-aided MMSE and find NMSE expressions of all estimators have the same structure and are commonly determined by an similar SNR-like term, however, compared with conventional estimators, there is an additional part consisting of UL data power, UL data length and BER values in the SNR-like term of data-aided scheme, which explicitly explains how NMSE performance benefits from the data-aided method. Average DL rate performance is also analyzed numerically with NLoS and NLoS/LoS models to verify the improvement in final DL performance and 4QAM/16QAM are utilized to suggest that the proposed scheme can be applied in multiple scenarios with different modulations.

The reminder of this paper is organized as follows. Section II introduces the system model of decoupled HetNets. A novel data-aided scheme is proposed and elaborated in Section III. The proposed scheme is compared with conventional channel estimation methods under the evaluation criteria of NMSE in Section IV. The theoretical results and some insights into this scheme are also discussed. Numerical results are presented in Section V and Section VI concludes the paper.

II. SYSTEM MODEL

We consider a single-cell scenario deployed with an MBS in the center and S SBSs along with K UEs scattering in the range of the cell randomly as shown in Fig. 1. For ease of geometric analysis, we assume the single cell as a circular area with radius of R_M . Meanwhile, the MBS equipped with M antennas and all SBSs with N antennas provide services to all single-antenna UEs in the cell coverage by fully utilizing the whole frequency band without any partitions, meaning that this model is interference limited. All communication links in the system are operating in TDD mode. Note that in this paper, we assume that there exist capacity-abundant fibers for backhaul between all SBSs and the MBS, which allows decoded data to be transmitted to the MBS without any latency nor error.

We assume that the channel coherence time is T , within which channel estimation and UL/DL transmission are processed. The channel from UEs to the MBS is denoted by $\mathbf{H} \in \mathbb{C}^{M \times K}$ while

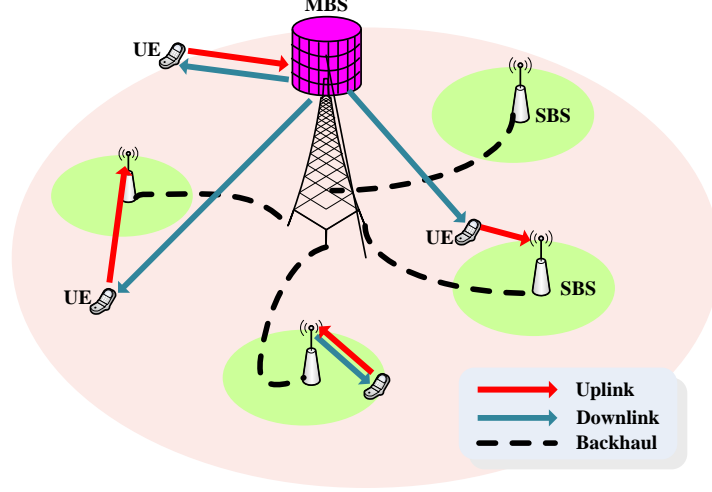


Fig. 1. Diagram for cellular HetNets with decoupled access.

the channel from UEs to the s th SBS is represented by $\mathbf{G}_s \in \mathbb{C}^{N \times K}$. Also, \mathbf{h}_k , \mathbf{g}_{sk} are the k th columns of \mathbf{H} and \mathbf{G}_s , respectively, which are of the forms:

$$\begin{aligned} \mathbf{h}_k &= \sqrt{\beta_k^M} \mathbf{h}_k^W, \\ \mathbf{g}_{sk} &= \sqrt{\beta_{sk}^S} \mathbf{g}_{sk}^W, \end{aligned} \quad (1)$$

where each element of $\mathbf{h}_k^W \in \mathbb{C}^{M \times 1}$ and $\mathbf{g}_{sk}^W \in \mathbb{C}^{N \times 1}$ follows $\mathcal{CN}(0, 1)$, β_k^M and β_{sk}^S represent the large scale fading from the k th UE to the MBS and the s th SBS, respectively, by neglecting shadowing effect and differences among antennas. The large scale fading between the k th UE to the MBS is modelled as $\beta_k^M = (d_k^M)^{-\alpha}$, where d_k^M is the distance with attenuation exponent α , and β_{sk}^S is defined similarly.

Unlike maximum average downlink receive power (MARP) [13], [16], [28] and biased cell association [14], [29] policies, due to the spectacular disparity between MBS and SBS, a new association strategy taking beamforming gain and transmit power into account is adopted in this paper. In particular, UEs perform a modified MARP policy in UL and DL respectively to achieve optimal associations in both links.

The received signal power in DL at the k th UE from a BS is given by

$$Q_{vk}^{\text{DL}} = \begin{cases} P_M \|\mathbf{h}_k\|^2, & v = 0, \\ P_S \|\mathbf{g}_{vk}\|^2, & v = 1, 2, \dots, S, \end{cases} \quad (2)$$

where P_M and P_S are the total transmit power of MBS and SBS. In addition, the subscript v represents the index of BS, 0 is for the MBS and v is for the v th SBS. Similarly, the received signal power in UL at a BS from UE k can be expressed as

$$Q_{vk}^{\text{UL}} = \begin{cases} P_D \|\mathbf{h}_k\|^2, & v = 0, \\ P_D \|\mathbf{g}_{vk}\|^2, & v = 1, 2, \dots, S, \end{cases} \quad (3)$$

where P_D stands for the transmit power of UEs.

Like the conventional MARP strategy, the effect of fast-fading should be averaged to let long-term parameters determine the association, which is more practical and operational in practice. Therefore, we exhibit the association results of the k th UE as an example using this modified MARP strategy:

$$\begin{aligned} \mathcal{D}_k &= \arg \max_{v=0,1,\dots,S} \left\{ \mathbb{E}_{\mathbf{h}_k^W, \mathbf{g}_{sk}^W} [Q_{vk}^{\text{DL}}] \right\} = \arg \max_{v=0,1,\dots,S} \left\{ M P_M (d_k^M)^{-\alpha}, N P_S (d_{1k}^S)^{-\alpha}, \dots, N P_S (d_{Sk}^S)^{-\alpha} \right\}, \\ \mathcal{U}_k &= \arg \max_{v=0,1,\dots,S} \left\{ \mathbb{E}_{\mathbf{h}_k^W, \mathbf{g}_{sk}^W} [Q_{vk}^{\text{UL}}] \right\} = \arg \max_{v=0,1,\dots,S} \left\{ M P_D (d_k^M)^{-\alpha}, N P_D (d_{1k}^S)^{-\alpha}, \dots, N P_D (d_{Sk}^S)^{-\alpha} \right\} \end{aligned} \quad (4)$$

where $\mathcal{D} = \{\mathcal{D}_1, \dots, \mathcal{D}_K\}$ and $\mathcal{U} = \{\mathcal{U}_1, \dots, \mathcal{U}_K\}$ are DL and UL association sets. Based on this, we can focus on the UEs with indices belonging to the set $\{k | \mathcal{D}_k \neq \mathcal{U}_k, k = 1, \dots, K\}$, who are regarded as decoupled UEs.

After performing the modified MARP strategy, UEs are able to connect to the optimal BSs in both links and ready to communicate with them. However, in decoupled HetNets, CSI acquired by UL training at the DL BS is not good enough to perform accurate DL precoding since the DL BS may not be the optimal UL BS. For the rest of this paper, we are devoted to improve the accuracy of DL channel for decoupled UEs by using decoded UL data to aid the estimation.

III. THREE-STAGE DATA-AIDED CHANNEL ESTIMATION

In this section, we elucidate our proposed data-aided channel estimation scheme for cellular HetNets with decoupled access. The whole process can be described in three stages. To illustrate the sequence of operations in each stage, the frame structure of this data-aided scheme is shown in Fig. 2. Different from conventional frame structure of a massive MIMO system, UL BSs in decoupled systems need to transmit decoded data to DL BS after completing common UL training and UL data transmission, and the DL BS is required to listen and record the UL

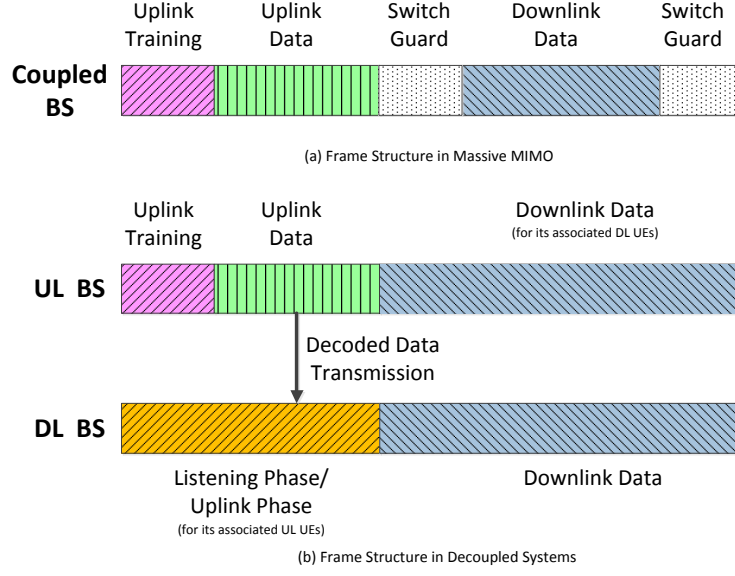


Fig. 2. The frame structure for the data-aided channel estimation scheme.

signal (including training and data phases) to perform joint channel estimation along with known training sequences and decoded data. Note that switch guard is ignored in decoupled systems for ease of understanding. The detailed implementation and signal model of the three-stage data-aided scheme are described step by step as follows.

A. Stage 1: Uplink Training

In the first stage, all UEs transmit their prescribed pilot sequences at power of P_T . The pilot sequences are orthogonal with τ_T symbols and $\mathbf{S} = [\mathbf{s}_1^T, \dots, \mathbf{s}_k^T, \dots, \mathbf{s}_K^T]^T$ denotes the pilot matrix where $\mathbf{s}_k \in \mathbb{C}^{1 \times \tau_T}$ is the pilot for the k th UE. Obviously, we have $\mathbf{S}\mathbf{S}^H = \tau_T P_T \mathbf{I}_K$ and $\tau_T \geq K$. The main task here is for all UL BSs to recover channels from the UL training signal, which is a common procedure in training based systems. To this end, we consider two conventional channel estimators, LS and MMSE.

The received signal at the v th SBS can be expressed as

$$\mathbf{Y}_v^P = \mathbf{G}_v \mathbf{S} + \mathbf{N}_v = \sum_{k=1}^K \mathbf{g}_{vk} \mathbf{s}_k + \mathbf{N}_v, \quad (5)$$

where $\mathbf{N}_v \in \mathbb{C}^{N \times \tau_T}$ denotes the additive white Gaussian noises (AWGNs) received and each element is independent identical distributed (i.i.d.), subject to $\mathcal{CN}(0, N_0)$.

Adopting the LS channel estimator, the estimated channel of the k th decoupled UE connecting to the v th SBS in UL is given as

$$\hat{\mathbf{g}}_{vk}^{\text{LS}} = \mathbf{Y}_v^{\text{P}} \mathbf{s}_k^{\text{H}} (\mathbf{s}_k \mathbf{s}_k^{\text{H}})^{-1} = \mathbf{g}_{vk} + \frac{\mathbf{N}_v \mathbf{s}_k^{\text{H}}}{\tau_{\text{T}} P_{\text{T}}}. \quad (6)$$

MMSE is another widely used channel estimator which provides better performance at the cost of higher complexity and prior statistics information in terms of correlation matrices of channels and noises. MMSE estimator is served as a benchmark and the estimated channel is

$$\hat{\mathbf{g}}_{vk}^{\text{MMSE}} = \mathbf{Y}_v^{\text{P}} \mathbf{C}_{vk}^{\text{UP,opt}} = \mathbf{G}_v \mathbf{S} \mathbf{C}_{vk}^{\text{UP,opt}} + \mathbf{N}_v \mathbf{C}_{vk}^{\text{UP,opt}}, \quad (7)$$

where $\mathbf{C}_{vk}^{\text{UP,opt}}$ is the $\tau_{\text{T}} \times 1$ linear estimation matrix for \mathbf{g}_{vk} , defined as

$$\mathbf{C}_{vk}^{\text{UP,opt}} \triangleq \arg \min_{\mathbf{C}_{vk}} \mathbb{E} [\|\mathbf{g}_{vk} - \hat{\mathbf{g}}_{vk}^{\text{MMSE}}\|^2] = \left[\sum_{i=1}^K \mathbf{s}_i^{\text{H}} \mathbf{R}_{\mathbf{g}_{vi}} \mathbf{s}_i + \mathbf{R}_{\mathbf{N}_v} \right]^{-1} \mathbf{s}_k^{\text{H}} \mathbf{R}_{\mathbf{g}_{vk}} \quad (8)$$

where $\mathbf{R}_{\mathbf{g}_{vi}} = \mathbb{E} [\mathbf{g}_{vi}^{\text{H}} \mathbf{g}_{vi}]$, $\mathbf{R}_{\mathbf{N}_v} = \mathbb{E} [\mathbf{N}_v^{\text{H}} \mathbf{N}_v]$ are the channel and noise correlation matrices.

B. Stage 2: Uplink Data Transmission

In the second stage, UEs send UL data to their associated UL BS and, the data of decoupled UEs are first decoded separately at each SBS with the estimated channels and then sent to the MBS via backhaul. However, the desired UL data signal at the SBSs is subject to high co-channel interference, and could hardly be decoded correctly using linear detectors. In this paper, discrete symbols are used to transmit data to combat severe interference. We assume that a UE transmits binary phase-shift-keying (BPSK)-coded UL data of totally τ_{D} symbols to its associated UL BS. The matrix $\mathbf{X} = [\mathbf{x}_1^{\text{T}}, \dots, \mathbf{x}_k^{\text{T}}, \dots, \mathbf{x}_K^{\text{T}}]^{\text{T}}$ is used to represent data sequences of UEs with each element randomly chosen from the set $\{P_{\text{D}}, -P_{\text{D}}\}$ and $\mathbf{x}_k \in \mathbb{C}^{1 \times \tau_{\text{D}}}$ is the UL data for the k th UE. Three linear receivers, maximal-ratio combining (MRC), zero-forcing (ZF) and MMSE are considered separately in the process of data decoding.

Firstly, the received signal at the v th SBS is expressed as

$$\mathbf{Y}_v^{\text{D}} = \mathbf{G}_v \mathbf{X} + \tilde{\mathbf{N}}_v = \sum_{k=1}^K \mathbf{g}_{vk} \mathbf{x}_k + \tilde{\mathbf{N}}_v, \quad (9)$$

where $\tilde{\mathbf{N}}_v \in \mathbb{C}^{N \times \tau_{\text{D}}}$ is AWGN with each element subject to $\mathcal{CN}(0, N_0)$ independently.

1) *ZF and MRC Detectors:* We use the estimated channel in Stage 1 to recover the UL data of the k th UE and, assume that the k th UE is associated with the v th SBS in UL. The MRC

detector only needs CSI of the k th UE, while in this scheme the SBS has the CSI knowledge of UEs served by itself to perform ZF detection. Therefore, a new channel matrix is defined as $\mathcal{G}_{vk} = [\hat{\mathbf{g}}_{vk}, \dots]$, which involves the estimated channels of UEs associated with the v th SBS, knowing that the k th UE is included at least. Thus, the combination matrix \mathbf{A} for MRC and ZF detectors are given by

$$\mathbf{A} = \begin{cases} \hat{\mathbf{g}}_{vk} (\hat{\mathbf{g}}_{vk}^H \hat{\mathbf{g}}_{vk})^{-1}, & \text{for MRC,} \\ \mathcal{G}_{vk} (\mathcal{G}_{vk}^H \mathcal{G}_{vk})^{-1}, & \text{for ZF.} \end{cases} \quad (10)$$

Then the estimated data sequence of the k th UE after applying the combination matrix becomes

$$\hat{\mathbf{x}}_k^{\text{ZF,MRC}} = \text{Decoder} \left([\mathbf{A}^H \mathbf{Y}_v^D]_1 \right), \quad (11)$$

where $[\mathbf{B}]_i$ performs the operation of taking the i th row of \mathbf{B} and $\text{Decoder}(\cdot)$ is the conventional BPSK decoder based on the maximum a posteriori probability (MAP) criterion. Note that in (10), the formation of combination matrix for ZF and MRC are similar. If an SBS only serves one UE, the matrix for ZF will degrade to that for MRC, which is the reason we take two detectors into consideration together.

2) *MMSE Detector*: In order to achieve better BER performance, MMSE detector is applied before decoding at each SBS. For MMSE detector with imperfect CSI, the standard way is to define MMSE channel estimation error as $\tilde{\mathbf{G}}_v = \mathbf{G}_v - \hat{\mathbf{G}}_v$, and the received signal can thus be rewritten as

$$\mathbf{Y}_v^D = \hat{\mathbf{G}}_v \mathbf{X} + \tilde{\mathbf{G}}_v \mathbf{X} + \tilde{\mathbf{N}}_v. \quad (12)$$

In what follows, we can treat the estimated channel as the real channel and the estimation error as independent noise, with the statistical property of MMSE [30]. Therefore, the MMSE combination matrix for the k th UE can be written as

$$\mathbf{C}_{vk}^{\text{UD,opt}} = \hat{\mathbf{g}}_{vk}^H \left(\hat{\mathbf{G}}_v \hat{\mathbf{G}}_v^H + \left(\sum_{k=1}^K \frac{N_0 \beta_{vk}^S}{N_0 + \beta_{vk}^S P_T \tau_T} + \frac{N_0}{P_D} \right) \mathbf{I} \right)^{-1}. \quad (13)$$

After multiplied by the combination matrix $\mathbf{C}_{vk}^{\text{UD,opt}}$, the estimated signal is sent to the BPSK decoder and the decoded data for the k th UE can be expressed as

$$\hat{\mathbf{x}}_k^{\text{MMSE}} = \text{Decoder} \left(\mathbf{C}_{vk}^{\text{UD,opt}} \mathbf{Y}_v^D \right). \quad (14)$$

Here, three detectors are considered to recover the UEs' UL data at each SBS and the decoded data is then transferred to the MBS via backhaul for the data-aided channel estimation scheme.

C. Stage 3: Data-Aided Channel Estimation

As described in the frame structure, during the last stage, the MBS needs not only to proceed with the UL channel estimation and UL data detection for its own UEs but also to listen and record both UL training and data signal. As a result, in this stage, we can utilize the recorded UL signal along with the known UL training sequences and the decoded UL data to perform channel estimation at the MBS³. With UL data signal and decoded UL data sequences, there is more information for the MBS to improve the estimated channel accuracy of decoupled UEs. First, the received signal at the MBS during the first two stages can be jointly expressed as

$$\mathbf{M}^{\text{PD}} = \mathbf{H}\mathbf{W} + \mathbf{Z} = \sum_{k=1}^K \mathbf{h}_k \mathbf{w}_k + \mathbf{Z}, \quad (15)$$

where $\mathbf{W} = [\mathbf{S}, \mathbf{X}]$ with \mathbf{w}_k being its k th row and $\mathbf{Z} \in \mathbb{C}^{M \times (\tau_{\text{T}} + \tau_{\text{D}})}$ is the AWGN noise at the MBS with each element an i.i.d. $\mathcal{CN}(0, N_0)$ random variable.

Next, the MMSE channel estimator is adopted to recover the channels of decoupled UEs, by utilizing the UL training sequences and the decoded UL data sequences. Hence, the estimated channel from the k th UE to MBS can be written as

$$\hat{\mathbf{h}}_k^{\text{MMSE}} = \mathbf{M}^{\text{PD}} \mathbf{C}_k^{\text{DA, opt}}, \quad (16)$$

where

$$\mathbf{C}_{vk}^{\text{DA, opt}} = \left[\sum_{i=1}^K \beta_k^{\text{M}} \mathbb{E} [\mathbf{w}_k^{\text{H}} \mathbf{w}_k] + N_0 \mathbb{E} [\mathbf{Z}^{\text{H}} \mathbf{Z}] \right]^{-1} \beta_k^{\text{M}} \mathbb{E} [\mathbf{w}_k^{\text{H}}], \quad (17)$$

thus, the channels from the decoupled UEs to the DL BS are obtained by this data-aided method.

Note that we only give the general expression of the combination matrix for the data-aided method in (17). In the error-free scenario, we can easily obtain the result of $\mathbb{E} [\mathbf{w}_k^{\text{H}} \mathbf{w}_k]$ in (17) from the prior information about the UL data sequences, however, the random errors in the decoded data sequences are non-negligible in a more practical scenario. Therefore, the problems of how to estimate the BER performance of decoded data and how to model the random errors in the decoded sequences will be discussed in the following section. Moreover, the specific

³When channel varies fast, it can be estimated with pilot and part of decoded UL data, herein decoded data is fully utilized.

combination matrix for the data-aided scheme and channel estimation performance of different estimators will also be analyzed.

IV. PERFORMANCE ANALYSIS

In this section, we conduct performance analysis to compare the data-aided scheme with conventional channel estimation techniques such as MMSE and LS. First of all, to evaluate the performance of different channel estimation methods, NMSE is adopted as the performance metric, which is defined as

$$\text{NMSE} = 10 \log_{10} \left(\frac{\mathbb{E} [\|\mathbf{g} - \hat{\mathbf{g}}\|^2]}{\mathbb{E} [\|\mathbf{g}\|^2]} \right) \quad (\text{in dB}), \quad (18)$$

where \mathbf{g} represents the real channel vector and $\hat{\mathbf{g}}$ is the estimated one.

A. Performance of Conventional Channel Estimation Methods

In this subsection, we use the above metric to analyze the NMSE performances of conventional channel estimation methods by only taking the UL training sequences into account. Herein, we take NMSE of the k th UE associated with the v th SBS in Stage 1 as an example. Although the NMSE expressions here are based on the channel between a UE and an SBS, they can be easily generalized to the case of a UE connected to the MBS, which will be regarded as the performance benchmarks of conventional methods.

Now, we give the performance of the LS channel estimator. From (6) and the NMSE definition in (18), the NMSE for the LS estimator of the k th UE associated with the v th SBS is

$$\mathcal{J}_{vk}^{\text{S,LS}} = 10 \log_{10} \left(\frac{N_0}{\beta_{vk}^{\text{S}} \tau_{\text{T}} P_{\text{T}}} \right). \quad (19)$$

Remark 1: From the above expression, it is observed that NMSE for the conventional LS estimator is related to N_0 , β_{vk}^{S} , τ_{T} and P_{T} . It states that once a UE is located and the BS is stationary, larger pilot power and longer pilot sequence are two means to improve channel estimation accuracy.

Before analyzing the NMSE for the MMSE estimator, we rewrite the estimated channel of the k th UE to the v th SBS as

$$\hat{\mathbf{g}}_{vk} = \frac{\mathbf{R}_{\mathbf{g}_{vk}}}{(N N_0 + \mathbf{R}_{\mathbf{g}_{vk}} P_{\text{T}} \tau_{\text{T}})} (\tau_{\text{T}} P_{\text{T}} \mathbf{g}_{vk} + \mathbf{N}_v \mathbf{s}_k^{\text{H}}), \quad (20)$$

where $\mathbf{R}_{\mathbf{g}} = \mathbb{E} [\mathbf{G}_v^{\text{H}} \mathbf{G}_v] = \text{diag} (\mathbf{R}_{\mathbf{g}_{v1}} \dots \mathbf{R}_{\mathbf{g}_{vi}} \dots \mathbf{R}_{\mathbf{g}_{vK}})$, $\mathbf{R}_{\mathbf{g}_{vi}} = \mathbb{E} [\mathbf{g}_{vi}^{\text{H}} \mathbf{g}_{vi}]$.

Observing the equation (20), \mathbf{g}_{vk} and $\mathbf{N}_v \mathbf{s}_k^H$ are independent complex Gaussian random variables, therefore, it is the property of Gaussian distribution that $\hat{\mathbf{g}}_{vk}$ is also a complex Gaussian variable following

$$\hat{\mathbf{g}}_{vk} \sim \mathcal{CN} \left(0, \frac{(\beta_{vk}^S)^2 P_T \tau_T}{N_0 + \beta_{vk}^S P_T \tau_T} \mathbf{I} \right). \quad (21)$$

Then we denote the channel estimation error of the MMSE estimator as $\tilde{\mathbf{g}}_{vk} \triangleq \mathbf{g}_{vk} - \hat{\mathbf{g}}_{vk}$, with the help of the orthogonal property of the MMSE estimator, $\tilde{\mathbf{g}}_{vk}$ is independent of $\hat{\mathbf{g}}_{vk}$. Thus, we directly obtain the channel estimation error vector, which is distributed as

$$\tilde{\mathbf{g}}_{vk} \sim \mathcal{CN} \left(0, \frac{N_0 \beta_{vk}^S}{N_0 + \beta_{vk}^S P_T \tau_T} \mathbf{I} \right). \quad (22)$$

In parallel to the LS estimator, the NMSE for the MMSE estimator of the k th UE associated with the s th SBS can be calculated by using (22)

$$\mathcal{J}_{vk}^{S, \text{MMSE}} = 10 \log_{10} \left(\frac{N_0}{N_0 + \beta_{vk}^S P_T \tau_T} \right). \quad (23)$$

Remark 2: Similar to the LS estimator, NMSE for the MMSE estimator is a function of τ_T , P_T , β_{vk}^S and N_0 . We can also increase pilot length and transmitted power to improve channel estimation performance once the locations of UE and SBS are both set. Comparing two NMSE expressions, the only difference appears to be the extra noise variance term N_0 in the numerator of NMSE for MMSE. Hence, the performance gain of the MMSE estimator over LS is obtained in the regime of low SNR and drops to 0 when SNR gets higher.

B. Performance of Data-Aided Channel Estimation Method

To evaluate our proposed data-aided method, the process of UL data transmission is analyzed and the BER is obtained since the decoded data is used to aid channel estimation and may highly affect channel estimation in stage 3. Here, we assume that the CSI acquired by the MMSE estimator is applied and the MMSE detector is performed for data decoding in Stage 2, as MMSE detector performs better in interference-limited systems. We first consider the BER performance of UL data in Stage 2 and the well-know BER expression for binary modulations is given by [31]

$$\text{BER} = \int_0^\infty f_{\text{SINR}}(x) \int_{\sqrt{x}}^\infty \frac{1}{\sqrt{2\pi}} \exp \left(-\frac{1}{2} t^2 \right) dt dx, \quad (24)$$

where $f_{\text{SINR}}(x)$ is the probability density function (p.d.f.) of SINR (signal-to-interference plus noise ratio).

As a matter of fact, the exact distribution of SINR for the MMSE detector has been derived in [32], but the distribution is too complex to obtain a closed-form expression of BER. Therefore, we resort to the well-known tractable Gamma distribution, whose p.d.f. is given as

$$f_{\text{Gamma}}(x; \alpha, \xi) = \frac{x^{\alpha-1} e^{-\frac{x}{\xi}}}{\Gamma(\alpha) \xi^\alpha} \quad (25)$$

to approximate the result.

Similar approach to obtain closed-form solutions has been applied in many previous efforts, see, e.g., [18], [31]. Following the similar procedure in [31], by determining two parameters of the approximated Gamma distribution, a closed-form expression of BER can be obtained. However, different from [18], [31], in our case, the antenna number at each SBS is assumed less than the number of UEs. In the following, two parameters of Gamma distribution are determined by moment matching.

We denote SINR_{vk} as the SINR of the k th UE served by the v th SBS with the MMSE detector. From (12)(13), SINR_{vk} can be expressed as

$$\text{SINR}_{vk} = \frac{1}{\left(\left(\mathbf{I} + \rho_v \hat{\mathbf{G}}_v^H \hat{\mathbf{G}}_v \right)^{-1} \right)_{kk}} - 1, \quad (26)$$

where $\rho_v = \left(\sum_{k=1}^K \frac{N_0 \beta_{vk}^S}{N_0 + \beta_{vk}^S P_T \tau_T} + \frac{N_0}{P_D} \right)^{-1}$ and $(\cdot)_{kk}$ represents the (k, k) th element of a matrix.

Then by applying [31, (8)], SINR_{vk} can be further expressed as

$$\text{SINR}_{vk} = \rho_v \hat{\mathbf{g}}_{vk}^H \hat{\mathbf{g}}_{vk} - \rho_v^2 \hat{\mathbf{g}}_{vk}^H \hat{\mathbf{G}}_{v(-k)} \hat{\mathbf{G}}_{v(-k)}^H \left(\mathbf{I} + \rho_v \hat{\mathbf{G}}_{v(-k)} \hat{\mathbf{G}}_{v(-k)}^H \right)^{-1} \hat{\mathbf{g}}_{vk}, \quad (27)$$

where $\hat{\mathbf{G}}_{v(-k)}$ is the matrix $\hat{\mathbf{G}}_v$ with the k th column removed and $\hat{\mathbf{g}}_{vk}$ is the k th column of $\hat{\mathbf{G}}_v$.

Considering the singular value decomposition (SVD), we have $\hat{\mathbf{G}}_{v(-k)} = \mathbf{U} \mathbf{D} \mathbf{V}^H$, $\mathbf{U} \in \mathbb{C}^{N \times N}$, $\mathbf{D} \in \mathbb{C}^{N \times N}$, $\mathbf{V}^H \in \mathbb{C}^{N \times (K-1)}$, $\mathbf{U} \mathbf{U}^H = \mathbf{U}^H \mathbf{U} = \mathbf{I}_N$ and $\mathbf{V}^H \mathbf{V} = \mathbf{I}_N$. Then applying SVD to (27), we get

$$\begin{aligned}
\text{SINR}_{vk} &= \rho_v \hat{\mathbf{g}}_{vk}^H \hat{\mathbf{g}}_{vk} - \rho_v^2 \hat{\mathbf{g}}_{vk}^H \mathbf{U} \mathbf{D}^2 (\mathbf{I} + \rho_v \mathbf{D}^2)^{-1} \mathbf{U}^H \hat{\mathbf{g}}_{vk} \\
&= \rho_v \boldsymbol{\varphi}^H \boldsymbol{\varphi} - \rho_v^2 \boldsymbol{\varphi}^H \mathbf{D}^2 (\mathbf{I} + \rho_v \mathbf{D}^2)^{-1} \boldsymbol{\varphi} \\
&= \rho_v \sum_{i=1}^N \frac{\|\varphi_i\|^2}{1 + \rho_v d_i^2},
\end{aligned} \tag{28}$$

where $\mathbf{U}^H \hat{\mathbf{g}}_{vk} \triangleq \boldsymbol{\varphi}$ and d_i is the i th diagonal element of \mathbf{D} . As \mathbf{U} is a unitary matrix, it can be proved that $\boldsymbol{\varphi}$ has the same distribution as $\hat{\mathbf{g}}_{vk}$.

Next, we proceed our analysis with the help of some known results of the empirical eigenvalue distribution of the product of two random matrices [33], [34]. Although these results are obtained under the limiting condition, it is shown in [31] that this approximation is, to some extent, accurate even for very small dimensions. In our case, those results are extended to the scenario where the number of antennas is less than the number of UEs, namely, $N < K$.

The empirical eigenvalue distribution (ESD) of $\rho_v \hat{\mathbf{G}}_{v(-k)} \hat{\mathbf{G}}_{v(-k)}^H$, denoted by $\hat{\mathbf{J}}$, converges to a measure \mathbf{J} , whose Stieltjes transform, denoted by $\mathcal{T}(z)$, is given as

$$\mathcal{T}(z) \triangleq \int \frac{1}{x - z} \mathbf{J}(dx). \tag{29}$$

According to the results in [35], this integral can be approximated by

$$\mathcal{T}(z) \approx \left(\sum_{i \neq k}^K \frac{\mathbf{V}_i}{1 + \text{Tr}(\mathbf{V}_i \mathcal{T}(z))} - z \mathbf{I} \right)^{-1}, \tag{30}$$

where $\mathbf{V}_i = \rho_v \hat{\beta}_{vk}^S \mathbf{I}_N$ and $\hat{\beta}_{vk}^S = \frac{(\beta_{vk}^S)^2 P_T \tau_T}{N_0 + \beta_{vk}^S P_T \tau_T}$. Similarly, the first derivative of \mathcal{T} is found as

$$\mathcal{T}'(z) \triangleq \int \frac{1}{(x - z)^2} \mathbf{J}(dx) = \mathcal{T}^2(z) \approx \left(\sum_{i \neq k}^K \frac{\mathbf{V}_i}{1 + \text{Tr}(\mathbf{V}_i \mathcal{T}(z))} - z \mathbf{I} \right)^{-2}. \tag{31}$$

Note that $\mathcal{T}(z)$ and $\mathcal{T}'(z)$ are well defined at $z = -1$ by the bounded convergence theorem [36], [37]. Then, we have

$$\frac{\text{Tr}(\boldsymbol{\Lambda})}{N} = \frac{1}{N} \sum_{i=1}^N \frac{1}{1 + \rho_v d_i^2} = \int \frac{1}{x + 1} \hat{\mathbf{J}}(dx) \xrightarrow{p} \int \frac{1}{x + 1} \mathbf{J}(dx) = \mathcal{T}(-1) \triangleq \mu, \tag{32}$$

$$\frac{\text{Tr}(\boldsymbol{\Lambda}^2)}{N} = \frac{1}{N} \sum_{i=1}^N \frac{1}{(1 + \rho_v d_i^2)^2} = \int \frac{1}{(x + 1)^2} \hat{\mathbf{J}}(dx) \xrightarrow{p} \int \frac{1}{(x + 1)^2} \mathbf{J}(dx) = \mathcal{T}'(-1) \triangleq \sigma^2, \tag{33}$$

where \xrightarrow{p} means convergence in probability and $\boldsymbol{\Lambda} \triangleq \text{Diag}(\lambda_1, \dots, \lambda_i, \dots, \lambda_N)$, $\lambda_i = \frac{1}{1 + \rho_v d_i^2}$.

By solving (30) and (31), we can obtain μ and σ^2 and have the following lemma.

Lemma 1: The first two moments of SINR_{vk} can be approximated by

$$\mathbb{E} [\text{SINR}_{vk}] \approx N \rho_v \hat{\beta}_{vk}^S \mu, \quad (34)$$

$$\text{Var} [\text{SINR}_{vk}] \approx N \left(\rho_v \hat{\beta}_{vk}^S \right)^2 \sigma^2. \quad (35)$$

Proof: See Appendix A. ■

Thus, with the first two moments of SINR_{vk} , its p.d.f. is determined and the ergodic BER expression can be presented by the proposition below.

Proposition 1: For the k th UE served by the v th SBS employing MMSE data detection with imperfect CSI (acquired by MMSE estimator), the BER of its UL data is expressed as

$$\text{BER}_{vk} = \frac{\Gamma(\alpha_{vk} + \frac{1}{2})}{\Gamma(\alpha_{vk}) 2\sqrt{2\pi}} \frac{\xi_{vk}^{-\alpha_{vk}}}{\alpha_{vk} \left(\frac{1}{\xi_{vk}} + \frac{1}{2} \right)^{\alpha_{vk} + \frac{1}{2}}} \times {}_2F_1 \left(1, \alpha_{vk} + \frac{1}{2}; \alpha_{vk} + 1; \frac{\frac{1}{\xi_{vk}}}{\frac{1}{\xi_{vk}} + \frac{1}{2}} \right), \quad (36)$$

where we have $\text{SINR}_{vk} \sim \text{Gamma}(\alpha_{vk}, \xi_{vk})$ with $\alpha_{vk} = N \frac{\mu^2}{\sigma^2}$, $\xi_{vk} = \rho_v \hat{\beta}_{vk}^S \frac{\sigma^2}{\mu}$. $\Gamma(\cdot)$ represents gamma function and ${}_2F_1(\cdot)$ is the hypergeometric function.

Proof: Based on (34) and (35), the distribution of SINR_{vk} can be determined by moment matching. Then plugging its distribution in (24), we obtain the final result after integration. ■

We note that both Gamma function and hypergeometric function are involved in the above result, which can be evaluated numerically but hardly shed any light on the BER performance. Hence, approximated expressions are derived in the following corollary.

Corollary 1: According to the result in (24), we find that BER is a strict convex function. Thus, by applying Jensen's inequality, the ergodic BER can be lower bounded by

$$\text{BER}_{vk}^{\text{Lower}} = Q \left(\sqrt{\mathbb{E}[\text{SINR}_{vk}]} \right) = Q \left(\sqrt{\alpha_{vk} \xi_{vk}} \right), \quad (37)$$

where $Q(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}t^2) dt$. The lower bound of the ergodic BER is a monotonic decreasing function of the first moment of SINR_{vk} , also of all the parameters of the first moment.

After UL data BER values are derived, we are now able to continue with the evaluation of the data-aided channel estimation, which utilizes MMSE estimation to recover the channels from the joint sequences consisting of known training sequences and decoded data with certain BER.

To do so, we model the joint sequences as follows. Recall the joint UL signal \mathbf{W} in (15), which is the combination of known training sequences \mathbf{S} and UL data sequences \mathbf{X} of all UEs. Denote the block matrix of known training sequences and decoded data as $\hat{\mathbf{W}} = [\mathbf{S}, \hat{\mathbf{X}}]$, and use

\mathbf{E} to represent the error matrix which can be defined as

$$\mathbf{W} \triangleq \hat{\mathbf{W}} \circ \mathbf{E}, \quad (38)$$

where \circ represents Hadamard product of two matrices. Note that it can further be decomposed into two parts as $\hat{\mathbf{W}} \circ \mathbf{E} = [\mathbf{S}, \hat{\mathbf{X}}] \circ [\mathbf{E}_1, \mathbf{E}_2]$. As we know the training sequences exactly, \mathbf{E}_1 is an all-one matrix, which represents an identity matrix in the Hadamard product, while \mathbf{E}_2 indicates errors in the decoded sequences with all elements from the set $\{1, -1\}$. Then we can obtain the following statistical property of \mathbf{E}_2 by utilizing the value of BER obtained previously

$$\begin{aligned} \mathbb{E}[e_{ij}] &= 1 - 2\text{BER}_{vi}, \\ \mathbb{E}[\|e_{ij}\|^2] &= 1, \end{aligned} \quad (39)$$

where $e_{ij} = [\mathbf{E}_2]_{ij}$. Intuitively, we assume that each bit in $\hat{\mathbf{X}}$ from each UL data stream has the same probability of error, regardless of its location and what is actually sent, meaning that e_{ij} for $j = 1, \dots, \tau_D$ are i.i.d. random variables, and the elements in \mathbf{E}_2 from different streams are independent. Therefore, the following proposition elucidates data-aided channel estimation with the MMSE estimator.

Proposition 2: For the MMSE estimator, the recovered channel of the k th UE at the MBS by using data-aided channel estimation can be written as

$$\hat{\mathbf{h}}_k^{\text{MMSE}} = (\mathbf{H}\mathbf{W} + \mathbf{Z}) \mathbf{C}_k^{\text{DA,opt}}, \quad (40)$$

$$\mathbf{C}_k^{\text{DA,opt}} = (\mathbf{P} + N_0 \mathbf{I})^{-1} (\hat{\mathbf{w}}_k^H \circ \mathbb{E}[\mathbf{e}_k^H]) \beta_k^M, \quad (41)$$

where \mathbf{e}_k is the k th row of \mathbf{E} with

$$\mathbb{E}[\mathbf{e}_k^H] = \left(\underbrace{1 \cdots 1}_{\tau_T} \underbrace{1 - 2\text{BER}_{vk} \cdots 1 - 2\text{BER}_{vk}}_{\tau_D} \right)^H, \quad (42)$$

and $\mathbf{P} \triangleq \hat{\mathbf{P}} + \Delta\mathbf{P}$ with

$$\hat{\mathbf{P}} = \left([\mathbf{S}, \hat{\mathbf{X}} \circ \mathbb{E}[\mathbf{E}_2]] \right)^H \mathbf{R}_H \left([\mathbf{S}, \hat{\mathbf{X}} \circ \mathbb{E}[\mathbf{E}_2]] \right), \quad (43)$$

$$\Delta\mathbf{P} \triangleq \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \Delta S_{\mathbf{X}} \mathbf{I}_{\tau_D} \end{pmatrix}, \quad (44)$$

where $\mathbf{R}_H = \text{diag}(\beta_1^M, \dots, \beta_K^M)$ and $\Delta S_{\mathbf{X}} = P_D \sum_{k=1}^K \beta_k^M \{1 - (1 - 2\text{BER}_{vk})^2\}$.

The NMSE of the channel from the k th UE to the MBS by using the data-aided channel

TABLE I
NMSE FOR DIFFERENT CHANNEL ESTIMATION METHODS

Conventional		Data-Aided
LS	MMSE	MMSE
$10 \log_{10} \left(\frac{1}{\rho^{\text{Con}} \beta_k^{\text{M}}} \right)$	$10 \log_{10} \left(\frac{1}{1 + \rho^{\text{Con}} \beta_k^{\text{M}}} \right)$	$10 \log_{10} \left(\frac{1}{1 + \rho_k^{\text{DA}} \beta_k^{\text{M}}} \right)$
$\rho^{\text{Con}} = \frac{\tau_{\text{T}} P_{\text{T}}}{N_0}$		$\rho_k^{\text{DA}} = \frac{\tau_{\text{T}} P_{\text{T}}}{N_0} + \frac{\tau_{\text{D}} P_{\text{D}} (1 - 2\text{BER}_{vk})^2}{\Delta S_{\text{X}} + N_0}$ $\Delta S_{\text{X}} = P_{\text{D}} \sum_k^K \beta_k^{\text{M}} \left\{ 1 - (1 - 2\text{BER}_{vk})^2 \right\}$

estimation method with MMSE can be calculated asymptotically as

$$\mathcal{J}_k^{\text{MMSE}} = 10 \log_{10} \left(\frac{1}{1 + \rho_k^{\text{DA}} \beta_k^{\text{M}}} \right), \quad (45)$$

where

$$\rho_k^{\text{DA}} = \frac{\tau_{\text{T}} P_{\text{T}}}{N_0} + \frac{\tau_{\text{D}} P_{\text{D}} (1 - 2\text{BER}_{vk})^2}{\Delta S_{\text{X}} + N_0}. \quad (46)$$

Proof: See Appendix B. ■

Remark 3: From *Proposition 2*, it is observed that the proposed data-aided channel estimation method introduces more parameters into NMSE, such as UL data transmission power, length of UL data sequence and BER of itself and other UEs. It is anticipated that better BER, higher data transmit power and longer UL data length will have positive effects on NMSE of data-aided estimation.

C. Comparison between Conventional and Data-Aided Channel Estimation

Here, we first summarize the NMSE expressions derived previously in TABLE I. Note that results in the table describe NMSE performances of the estimated channel from decoupled UEs to the associated MBS with different channel estimators. The NMSE results in the first two columns are directly transformed from that of UEs associated with the SBSs in (19) and (23).

In general, the table demonstrates that all NMSE results are parameterized by large-scale fading and a SNR-like term. Furthermore, after we combine these two parameters, it can be viewed as an effective signal power divided by noise power at the receiver, which corresponds to the meaning of an effective SNR. Thus, NMSE can be interpreted as the function of reverse effective SNR as a whole.

Next, we will examine the differences among these estimators and try to reveal some insights. To begin with, compared with the LS estimator, we can clearly observe that MMSE outperforms LS at low SNR, but performs nearly equally as LS when ρ_{Con} is high. Secondly, different from conventional methods, the data-aided channel estimation method makes full use of UL data signal to explore channel information it carries. In addition, in our two-tier network model, decoded UL data streams at the SBSs will be assembled at the central MBS via capacity-rich backhaul links. As such, we can utilize the collected UL data streams as a kind of asymptotically orthogonal pilot sequences, although these sequences have a certain probability of error. As a matter of fact, the data-aided method offers more information of the desired channels and introduces more degrees of freedom in the process of estimation, such as UL data power, UL data length and BER performance. Furthermore, the advantage of our proposed scheme can be clearly shown by an increment in the SNR-like term ρ_{DA} in TABLE I, which implies that the data-aided method does help with elevating channel estimation performance in all cases.

Some intuitions can be gained on how to obtain better channel for downlink transmission in decoupled systems using the data-aided channel estimation method. According to the NMSE expression for the data-aided method, we can gain several insights into practical design.

- Generally, when we increase training power, training length, UL data length and UL power, NMSE will monotonically decrease and better channel estimation can be achieved.
- When we reduce our generalized model to the scenario without UL data BER, it basically corresponds to the case where all UL data are decoded correctly and $\rho_{\text{DA}} = \left(\frac{\tau_{\text{T}} P_{\text{T}} + \tau_{\text{D}} P_{\text{D}}}{N_0} \right)$. As a result, NMSE is determined by the total energy within training slot and UL data transmission slot, which is different from some previous literatures where power and length distribution within two slots can be optimized by assuming a fixed total energy [25], [38], [39]. However, in our model, NMSE is a constant regardless how to allocate power and length between the two slots. Also, the performance of this zero-BER case can be regarded as the performance upper bound.
- If the BER performance is extremely poor in certain circumstances, say close to 0.5, the benefit from data-aided estimation in the SNR-like term will disappear and $\rho_k^{\text{DA}} = \frac{\tau_{\text{T}} P_{\text{T}}}{N_0}$. The data-aided method will be degraded to the conventional method. Hence, BER is a significant factor to exploit the most of channel information from data sequences. Further, NMSE of a typical UE is not only affected by its own UL data BER, but also by BER of co-channel UEs due to non-orthogonality property of coded data sequences although both sequences

are orthogonal asymptotically.

- As observed, we can improve the NMSE for all the estimators listed in TABLE I by increasing training power. However, note that the SNR-like term for data-aided estimation would eventually saturate to

$$\frac{\tau_D (1 - 2\text{BER}_{v1})^2}{\sum_k^K \beta_k^M \{1 - (1 - 2\text{BER}_{vk})^2\}} \quad (47)$$

when P_D approaches infinity. Therefore, there is a performance floor when increasing data power, although in practice the maximum transmit power is fixed.

V. NUMERICAL RESULTS

In this section, we perform simulations to validate the analytical results and demonstrate the potential of data-aided channel estimation for decoupled UEs. As described in the system model, we consider a single cell consisting of an MBS in the center and S SBSs uniformly placed in the cell. Also, K UEs are uniformly distributed in random within the cell and connect to the BSs according to the modified MARP with decoupling access. As a result, there are a number of decoupled UEs who connect to different BSs in UL/DL. Unless otherwise specified, numerical results provided focus on and are averaged over all decoupled UEs, and the parameters in this section are listed in TABLE II. All points in simulations were obtained via 100 association patterns and 1000 independent channel realizations. In this section, PO corresponds to “pilot only” and DA represents the “data-aided” method.

Two main performance metrics are used to evaluate the proposed DA method. One is the NMSE defined in Section IV and the second performance metric is the average DL rate by assuming standard zero-forcing beamforming (ZFBF) based on the estimated channels at all BSs. The average DL rate is defined as

$$\mathcal{R} = \mathbb{E} [\log_2 (1 + \text{SINR}_{\text{UE}}^{\text{DL}})] , \quad (48)$$

where $\text{SINR}_{\text{UE}}^{\text{DL}}$ is the received DL SINR of UEs. Note that the maximum UL power of UEs in some figures is extended to 43 dBm to exhibit the whole picture of impacts on metrics, but in practice UL power would not be raised to that high .

In order to verify data-aided scheme can work in other scenarios, NLoS/LoS path loss model is also adopted in simulations. Due to page restrictions, only rate performance is shown and

TABLE II
SIMULATION PARAMETERS

Parameters	Values
Radius of macro cell R_M	1000 m
Number of SBSs S	30
Number of UEs K	30
Noise power spectral density N_0	-174 dBm/Hz
Attenuation exponent α	4
Antenna number of MBS M	256
Antenna number of SBS N	8
Transmit power of MBS P_M	46 dBm
Transmit power of SBS P_S	24 dBm
Training power of UEs P_T	-7~23 dBm
Data power of UEs P_D	-7~23 dBm
Training length τ_T	30
Data length τ_D	128

NMSE performance of this model is similar to all-NLoS model. According to 3GPP [40] and previous work [41], NLoS/LoS path loss model is defined as

$$\begin{aligned}\beta_k^{M,NL} &= A_M^{NL} (d_k^M)^{-\alpha_M^{NL}}, \\ \beta_{sk}^{S,L} &= A_S^L (d_{sk}^S)^{-\alpha_S^L}\end{aligned}\tag{49}$$

where A_M^{NL} and A_S^L are the path loss of NLoS path between a MBS and a UE and the path loss of LoS path between a SBS and a UE at a reference distance of 1. α_M^{NL} and α_S^L are path loss exponents for their corresponding paths. According to the recommendations of 3GPP, we adopt the parameters as $\alpha_M^{NL} = 3.75$, $\alpha_S^L = 2.09$, $A_M^{NL} = 10^{-14.54}$, $A_S^L = 10^{-10.38}$.

In Fig. 3 and Fig. 4, the average UL data BER performance of decoupled UEs with different detectors versus training power and data power are demonstrated. Numerical results reveal that ZF enjoys better performance due to interference cancelation with CSI at the SBSs, but both detectors are not able to obtain as low BER as MMSE when training or data power grows. All the MMSE estimators with estimated CSI can achieve BER performance better than 10^{-3} under severe co-channel interference and the curves for those estimators saturate to the one for the MMSE detector with perfect CSI. The estimated BER with imperfect CSI in (36) is plotted with plus signs and is pretty close to the simulated one, while a gap between BER simulations and its lower bound (in circles) is observed in both figures. Hence, the estimated BER is accurate enough to be utilized to represent the real BER performance of the decoded data in the later DA

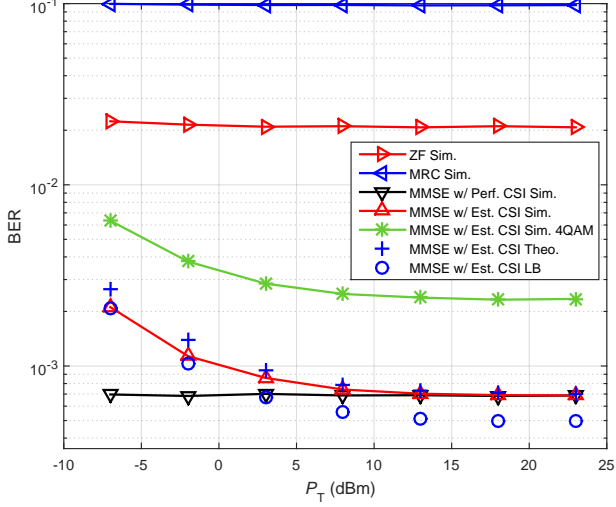


Fig. 3. Average uplink data BER of decoupled UEs versus P_T with $P_D = 23$ dBm for different data estimators.

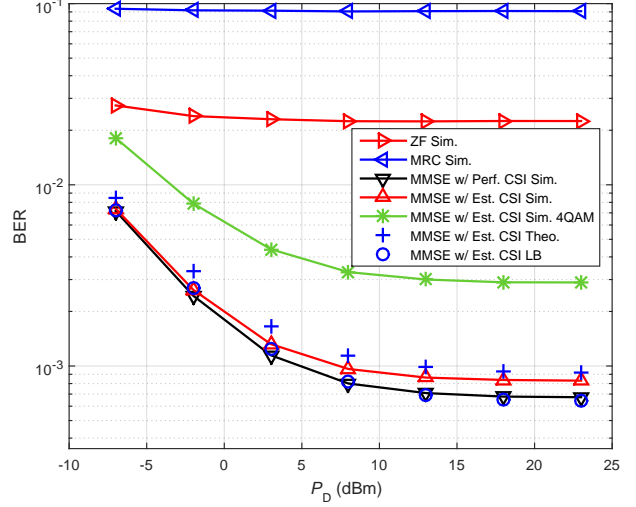


Fig. 4. Average uplink data BER of decoupled UEs versus P_D with $P_T = 3$ dBm for different data estimators.

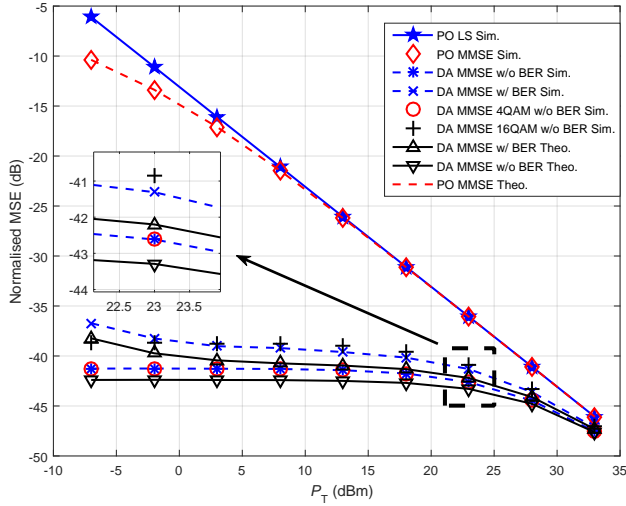


Fig. 5. Comparison of average NMSE of decoupled UEs for P_T with $P_D = 23$ dBm.

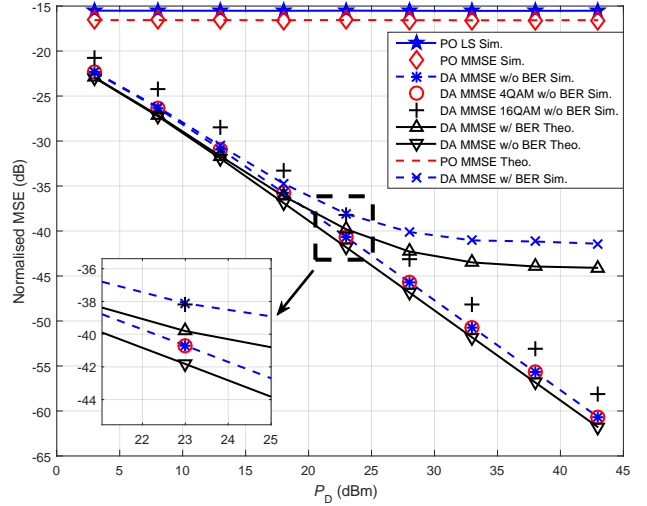


Fig. 6. Comparison of average NMSE of decoupled UEs for P_D with $P_T = 3$ dBm.

channel estimation. Note that all the curves are for BPSK modulation except the one in asterisk is for the BER performance of 4QAM modulation, which has higher BER than BPSK since the power of UL symbols remains the same in different modulations.

Then, we verify the NMSE performance of the DA method and compare it with the PO LS and MMSE. Fig. 5 depicts the relation between training power and NMSE of decoupled UEs. It is observed that the DA channel estimation is always better than LS and MMSE estimators, in

particular, when training power is low, the DA method outperforms the conventional ones over 30dB. The curves for the DA method tend to approach to and eventually saturate to the one for the PO MMSE and LS estimators as training power increases, which can be explain with the NMSE expression for DA in TABLE I that when training power grows higher, the first term with training power dominates the SNR-like term which is highly related to the NMSE performance and the benefit from DA method is not as remarkable as in the low training power scenarios. Moreover, the squares represent the NMSE by treating a decoded sequence with errors as if there were no error in it. Its performance is almost the same as the one considering the BER effect at low training power regime and become worse as training power grows. Interestingly, in the error-free scenario, we find the NMSE performance of 4QAM is exactly the same as that of BPSK and both two modulations performs a little better than 16QAM, which provides a solid proof for the insight concluded in Section IV-C that NMSE of the DA method is determined by the total energy of training and data slots since all symbols of 4QAM and BPSK have the same power P_T while symbols of 16QAM have several power levels with the maximum power of P_T .

The NMSE versus data power is shown in Fig. 6. The gap between PO and the DA methods gets larger by increasing data power for the reason that the NMSE performances of PO methods has no relation with data power while for the DA method, higher data power will not only improve BER performance but also directly increase the SNR-like term. Furthermore, the NMSE of the DA method without BER can be improved almost log-linearly as data power grows while the case with BER tends to decrease slower and reaches a performance limit as analyzed in Section IV-C. Similar results of 4QAM and 16QAM can be observed in this figure.

The relation between NMSE and data lengths is also investigated. In Fig. 7, it is observed that when the UL data length gets longer, NMSE for the PO MMSE remains constant but the NMSE for the DA method enjoys nearly log-linear improvement. Recall that we make the approximation that $\hat{\mathbf{X}}\hat{\mathbf{X}}^H \rightarrow \tau_D P_D \mathbf{I}_k$ when τ_D is large. Fig. 7 indicates that for the DA method without BER, the gap between simulations and the approximated result is getting smaller quickly when τ_D becomes larger, while for the case with BER, the approximation gap reduces slowly. In general, as the result of BER effect, the DA methods without BER performs better than the method with BER and the performance gap between two methods gets larger as the length of data grows.

Next, we examine the impacts of parameters on the final rate metric and compare the average DL rate performance between three types of UEs, which are decoupled UEs, SUEs (served by SBSs in UL/DL) and MUEs (served by the MBS in UL/DL), also between two different path loss

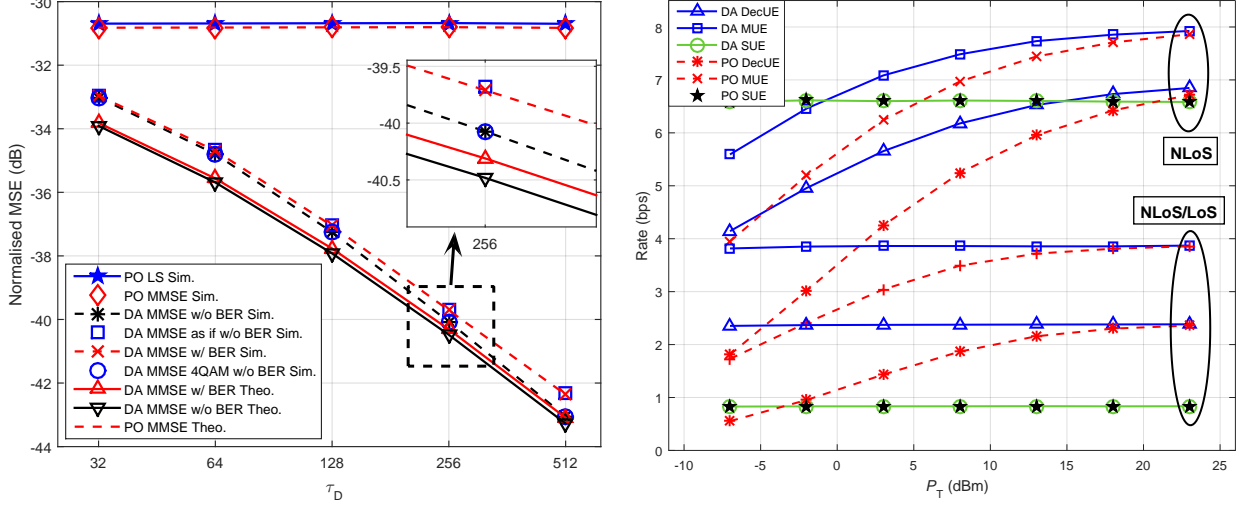


Fig. 7. Comparison of average NMSE of decoupled UEs for Fig. 8. Average DL rate versus P_T with $P_D = 23$ dBm for the $\tau_D = 64, 128, 256, 512$ with $P_T = 13$ dBm and $P_D = 13$ dBm. comparison between DA and PO.

models. From Fig. 8 and Fig. 9, we observe that DA method can improve DL rate of both MUEs and decoupled UEs by using two different path loss models within the whole range of training and data power, especially in high data power and low training power regimes. By utilizing DA method, the MBS has more accurate channels of decoupled UEs and the accuracy of DL beamforming is also improved, thus, the inter-UE interference at both MUEs and decoupled UEs can be suppressed effectively and higher rate can be obtained. The average rate for SUEs remains almost the same because the UL channel estimation at SBSs is so good that the improvement in NMSE by increasing training power would have little benefit to the DL rate, while it is easy to figure out that changing UL data power has no effect on DL rate of SUEs. Therefore, the DA methods can not only improve the average DL rate of the whole cell but also reduce rate gaps between different types of UEs, which promises better fairness in the cell. Moreover, although DL rates for three types of UEs have dropped by using NLoS/LoS model, the data-aided scheme could still provide remarkable promotion in rate performances of MUEs and decoupled UEs.

Fig. 10 shows that the DL rate gain achieved by the DA method grows with the increase of UL data length. Similarly, the DL rates of both MUEs and decoupled UEs are improved by DA method and it is observed that decoupled UEs enjoys larger rate growth than MUEs.

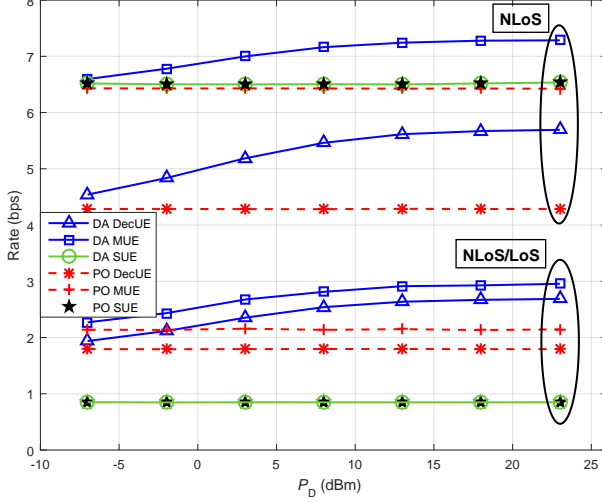


Fig. 9. Average DL rate versus P_D with $P_T = 3$ dBm for the comparison between DA and PO.

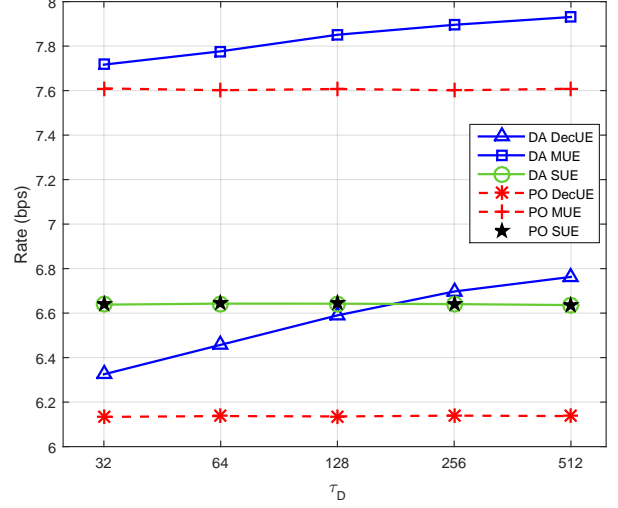


Fig. 10. Average DL rate versus τ_T with $P_T = 3$ dBm and $P_D = 23$ dBm for the comparison between DA and PO.

VI. CONCLUSION

In this paper, we proposed a data-aided channel estimation method for decoupled UEs in HetNets. In this method, the decoded UL data and known training sequence are jointly utilized to better estimate the DL channels at the MBS. The ergodic BER of UL data has been analyzed to model the decoded data with errors and the approximated NMSE for the data-aided MMSE estimator was derived and compared with the conventional estimators. It has been proved by both theoretic results and simulations that the data-aided method can greatly improve the accuracy of estimated channels of decoupled UEs by introducing more degrees of freedom. Although the BER may affect the effectiveness of this method, the BER could be controlled at a very low level by applying more advanced data detection techniques in the future. DL rate performance is also investigated numerically to show the true benefit to this final metric and the results of 4QAM/16QAM and NLoS/NLoS verify the proposed method can be applied in multiple scenarios with different modulations.

APPENDIX A
PROOF OF LEMMA 1

Proof: From the expression of SINR_{vk} in (28), we have

$$\begin{aligned}\mathbb{E} [\text{SINR}_{vk}] &= \mathbb{E} \left[\rho_v \sum_{i=1}^N \frac{\|\varphi_i\|^2}{1 + \rho_v d_i^2} \right] \\ &= N \rho_v \hat{\beta}_{vk}^S \mathbb{E} \left[\left(\frac{1}{N} \sum_{i=1}^N \lambda_i \right) \right] \\ &= N \rho_v \hat{\beta}_{vk}^S \mathbb{E} \left[\frac{\text{Tr}(\Lambda)}{N} \right] \approx N \rho_v \hat{\beta}_{vk}^S \mu,\end{aligned}\tag{50}$$

where the above approximation uses the definition in (32).

Similarly, we obtain the second moment by

$$\begin{aligned}\text{Var} [\text{SINR}_{vk}] &= N^2 \rho_v^2 \text{Var} \left[\frac{1}{N} \sum_{i=1}^N \lambda_i \|\phi_i\|^2 \right] \\ &= N^2 \rho_v^2 \text{Var} \left[\frac{1}{N} \sum_{i=1}^N \lambda_i \mathbb{E} [\|\phi_i\|^2 | \hat{\mathbf{G}}_{v(-k)}] \right] + N^2 \rho_v^2 \mathbb{E} \left[\frac{1}{N} \sum_{i=1}^N \lambda_i^2 \text{Var} [\|\phi_i\|^2 | \hat{\mathbf{G}}_{v(-k)}] \right] \\ &\stackrel{(a)}{=} N^2 \rho_v^2 \left(\hat{\beta}_{vk}^S \right)^2 \text{Var} \left[\frac{\text{Tr}(\Lambda)}{N} \right] + N^2 \rho_v^2 \left(\hat{\beta}_{vk}^S \right)^2 \mathbb{E} \left[\frac{\text{Tr}(\Lambda^2)}{N} \right] \\ &\approx N^2 \rho_v^2 \left(\hat{\beta}_{vk}^S \right)^2 \sigma^2,\end{aligned}\tag{51}$$

where the first term of (a) can be proved to converge to 0 by the results from [31]. ■

APPENDIX B
PROOF OF PROPOSITION 2

Proof: First, denote the mean-square-error (MSE) as $\mathcal{J}(\mathbf{C}_k) = \mathbb{E} \left[\left\| \hat{\mathbf{h}}_k^{\text{MMSE}} - \mathbf{h}_k \right\|^2 \right]$. Then by applying the MMSE rule directly, the problem is formulated as

$$\min_{\mathbf{C}_k} \mathcal{J}(\mathbf{C}_k) = \mathbb{E} \left[\left(\left(\mathbf{H} (\hat{\mathbf{W}} \circ \mathbf{E}) + \mathbf{Z} \right) \mathbf{C}_k - \mathbf{h}_k \right)^H \left(\left(\mathbf{H} (\hat{\mathbf{W}} \circ \mathbf{E}) + \mathbf{Z} \right) \mathbf{C}_k - \mathbf{h}_k \right) \right], \tag{52}$$

and we take the derivative of \mathcal{J} in terms of \mathbf{C}_k and let the derivative be 0 to find the optimal solution

$$\left(\mathbb{E} \left[\left(\hat{\mathbf{W}} \circ \mathbf{E} \right)^H \mathbf{H}^H \mathbf{H} (\hat{\mathbf{W}} \circ \mathbf{E}) + \mathbf{Z}^H \mathbf{Z} \right] \right) \mathbf{C}_k = \mathbb{E} \left[\left(\hat{\mathbf{W}} \circ \mathbf{E} \right)^H \mathbf{H}^H \mathbf{h}_k \right]. \tag{53}$$

Then according to the statistical properties of \mathbf{H} , \mathbf{E} and the independency between two random matrices, we can obtain

$$\mathbf{R}_\mathbf{H} \triangleq \frac{1}{M} \mathbb{E} [\mathbf{H}^H \mathbf{H}] = \text{diag} (\beta_1^M, \dots, \beta_k^M, \dots, \beta_K^M), \quad (54)$$

and

$$\left(\mathbb{E} \left[\left(\hat{\mathbf{W}} \circ \mathbf{E} \right)^H \mathbf{R}_\mathbf{H} \left(\hat{\mathbf{W}} \circ \mathbf{E} \right) \right] + N_0 \mathbf{I} \right) \mathbf{C}_k = \left(\hat{\mathbf{w}}_k^H \circ \mathbb{E} [\mathbf{e}_k^H] \right) \beta_k^M. \quad (55)$$

The above expectation can be defined by \mathbf{P} and decomposed into four block matrices

$$\mathbf{P} \triangleq \mathbb{E} \left[\left(\left[\mathbf{S} \circ \mathbf{E}_1, \hat{\mathbf{X}} \circ \mathbf{E}_2 \right] \right)^H \mathbf{R}_\mathbf{H} \left(\left[\mathbf{S} \circ \mathbf{E}_1, \hat{\mathbf{X}} \circ \mathbf{E}_2 \right] \right) \right] = \begin{pmatrix} \mathbf{P}_{11} & \mathbf{P}_{12} \\ \mathbf{P}_{21} & \mathbf{P}_{22} \end{pmatrix}, \quad (56)$$

with $\mathbf{P}_{11} = \mathbf{S}^H \mathbf{R}_\mathbf{H} \mathbf{S}$, $\mathbf{P}_{12} = \mathbf{S}^H \mathbf{R}_\mathbf{H} \left(\hat{\mathbf{X}} \circ \mathbb{E} [\mathbf{E}_2] \right)$, $\mathbf{P}_{21} = \left(\hat{\mathbf{X}} \circ \mathbb{E} [\mathbf{E}_2] \right)^H \mathbf{R}_\mathbf{H} \mathbf{S}$, $\mathbf{P}_{22} = \mathbb{E} \left[\left(\hat{\mathbf{X}} \circ \mathbf{E}_2 \right)^H \mathbf{R}_\mathbf{H} \left(\hat{\mathbf{X}} \circ \mathbf{E}_2 \right) \right]$.

The results of \mathbf{P}_{11} , \mathbf{P}_{12} and \mathbf{P}_{21} are straightforward. Hence, we focus on the expectation of \mathbf{P}_{22} . Two cases are discussed separately by using the definition of matrix multiplication and independency, as

$$\begin{aligned} [\mathbf{P}_{22}]_{ij, i \neq j} &= \sum_k^K \sum_l^K [\mathbf{R}_\mathbf{H}]_{kl} (\hat{x}_{ki})' \hat{x}_{lj} \mathbb{E} \left[([\mathbf{E}_2]_{ki})' [\mathbf{E}_2]_{lj} \right] \\ &= \sum_k^K \beta_k^M (\hat{x}_{ki})' \hat{x}_{kj} (1 - 2\text{BER}_{vk})^2 \\ [\mathbf{P}_{22}]_{ij, i=j} &= \sum_k^K \beta_k^M (\hat{x}_{ki})' \hat{x}_{ki} \mathbb{E} [\|[\mathbf{E}_2]_{ki}\|^2] = P_D \sum_k^K \beta_k^M. \end{aligned}$$

Therefore, $[\mathbf{P}_{22}]_{i \neq j}$ can be generally written as $\left[\left(\left[\mathbf{S}, \hat{\mathbf{X}} \circ \mathbb{E} [\mathbf{E}_2] \right] \right)^H \mathbf{R}_\mathbf{H} \left(\left[\mathbf{S}, \hat{\mathbf{X}} \circ \mathbb{E} [\mathbf{E}_2] \right] \right) \right]_{i \neq j}$, but diagonal elements of \mathbf{P}_{22} are $P_D \sum_k^K \beta_k^M$. Hence, for brevity and ease of calculation, we can rewrite \mathbf{P}_{22} as

$$\left(\hat{\mathbf{X}} \circ \mathbb{E} [\mathbf{E}_2] \right)^H \mathbf{R}_\mathbf{H} \left(\hat{\mathbf{X}} \circ \mathbb{E} [\mathbf{E}_2] \right) + \Delta \mathbf{P}_\mathbf{X}, \quad (57)$$

where $\Delta \mathbf{P}_\mathbf{X} = \Delta S_\mathbf{X} \mathbf{I}_{\tau_D}$, $\Delta S_\mathbf{X} = P_D \sum_k^K \beta_k^M \left\{ 1 - (1 - 2\text{BER}_{vk})^2 \right\}$. As a result, the block matrices can be reunited as the sum of two parts denoted as $\mathbf{P} \triangleq \hat{\mathbf{P}} + \Delta \mathbf{P}$, with

$$\hat{\mathbf{P}} = \left(\left[\mathbf{S}, \hat{\mathbf{X}} \circ \mathbb{E} [\mathbf{E}_2] \right] \right)^H \mathbf{R}_\mathbf{H} \left(\left[\mathbf{S}, \hat{\mathbf{X}} \circ \mathbb{E} [\mathbf{E}_2] \right] \right) \quad (58)$$

and

$$\Delta \mathbf{P} = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \Delta \mathbf{P}_{\mathbf{x}} \end{pmatrix}. \quad (59)$$

By plugging these results into (55), the optimal combination matrix for the data-aided channel estimation method with the MMSE estimator is obtained. After the combination matrix is obtained, we can continue to evaluate NMSE of the data-aided method. From the definition in (18), we can first calculate numeric term inside logarithm as

$$\begin{aligned} & \mathbb{E} \left[\left\| \hat{\mathbf{h}}_k^{\text{MMSE}} - \mathbf{h}_k \right\|^2 \right] \\ &= \mathbb{E} \left[\left(\left(\mathbf{H} \left(\hat{\mathbf{W}} \circ \mathbf{E} \right) + \mathbf{Z} \right) \mathbf{C}_k - \mathbf{h}_k \right)^H \left(\left(\mathbf{H} \left(\hat{\mathbf{W}} \circ \mathbf{E} \right) + \mathbf{Z} \right) \mathbf{C}_k - \mathbf{h}_k \right) \right] \\ &= M \beta_k^{\text{M}} - M \left(\beta_k^{\text{M}} \right)^2 \mathbb{E} \left[\left(\hat{\mathbf{w}}_k \circ \mathbb{E} [\mathbf{e}_k] \right) \left(\mathbf{P} + N_0 \mathbf{I} \right)^{-1} \left(\hat{\mathbf{w}}_k^H \circ \mathbb{E} [\mathbf{e}_k^H] \right) \right] \end{aligned} \quad (60)$$

Now, we denote

$$\mathbf{P}_{\text{Aux}} \triangleq \Delta \mathbf{P} + N_0 \mathbf{I} = \begin{pmatrix} N_0 \mathbf{I}_{\tau_{\text{T}}} & \mathbf{0} \\ \mathbf{0} & (\Delta S_{\mathbf{x}} + N_0) \mathbf{I}_{\tau_{\text{D}}} \end{pmatrix} \quad (61)$$

and

$$\bar{\mathbf{W}} \triangleq \hat{\mathbf{W}} \circ \mathbb{E} [\mathbf{E}]. \quad (62)$$

Thus, we can derive the expectation part in the last equation of (60) as

$$\begin{aligned} & \mathbb{E} \left[\left(\hat{\mathbf{w}}_k \circ \mathbb{E} [\mathbf{e}_k] \right) \left(\mathbf{P} + N_0 \mathbf{I} \right)^{-1} \left(\hat{\mathbf{w}}_k^H \circ \mathbb{E} [\mathbf{e}_k^H] \right) \right] \\ &= \mathbb{E} \left[\left[\bar{\mathbf{W}} \left(\bar{\mathbf{W}}^H \mathbf{R}_{\mathbf{H}} \bar{\mathbf{W}} + \mathbf{P}_{\text{Aux}} \right)^{-1} \bar{\mathbf{W}}^H \right]_{kk} \right] \\ &\stackrel{(a)}{=} \mathbb{E} \left[\left[\bar{\mathbf{W}} \mathbf{P}_{\text{Aux}}^{-1} \bar{\mathbf{W}}^H - \bar{\mathbf{W}} \mathbf{P}_{\text{Aux}}^{-1} \bar{\mathbf{W}}^H \left(\mathbf{R}_{\mathbf{H}}^{-1} + \bar{\mathbf{W}} \mathbf{P}_{\text{Aux}}^{-1} \bar{\mathbf{W}}^H \right)^{-1} \bar{\mathbf{W}} \mathbf{P}_{\text{Aux}}^{-1} \bar{\mathbf{W}}^H \right]_{kk} \right] \\ &\stackrel{(b)}{=} \left[\mathbf{R}_{\mathbf{H}}^{-1} - \mathbf{R}_{\mathbf{H}}^{-1} \mathbb{E} \left[\left(\bar{\mathbf{W}} \mathbf{P}_{\text{Aux}}^{-1} \bar{\mathbf{W}}^H + \mathbf{R}_{\mathbf{H}}^{-1} \right)^{-1} \right] \mathbf{R}_{\mathbf{H}}^{-1} \right]_{kk}, \end{aligned} \quad (63)$$

where (a) follows the Woodbury matrix inversion identity while (b) uses the matrix identity

$$\mathbf{A} - \mathbf{A} (\mathbf{A} + \mathbf{B})^{-1} \mathbf{A} = \mathbf{B} - \mathbf{B} (\mathbf{A} + \mathbf{B})^{-1} \mathbf{B}. \quad (64)$$

With the definitions of

$$\tilde{\mathbf{E}}_2 \triangleq \text{diag}((1 - 2\text{BER}_{v1}), \dots, (1 - 2\text{BER}_{vK})) \quad (65)$$

and

$$\mathbf{\Lambda}_{\mathbf{X}} \triangleq \frac{1}{\Delta S_{\mathbf{X}} + N_0} \hat{\mathbf{X}} \hat{\mathbf{X}}^H, \quad (66)$$

the expectation part in (63) can be simplified as

$$\begin{aligned} & \mathbb{E} \left[\left(\bar{\mathbf{W}} \mathbf{P}_{\text{Aux}}^{-1} \bar{\mathbf{W}}^H + \mathbf{R}_{\mathbf{H}}^{-1} \right)^{-1} \right] \\ &= \mathbb{E} \left[\left(\left[\mathbf{S}, \hat{\mathbf{X}} \circ \mathbb{E} [\mathbf{E}_2] \right] \Delta \mathbf{P}_{\text{Aux}}^{-1} \begin{bmatrix} \mathbf{S}^H \\ \left(\hat{\mathbf{X}} \circ \mathbb{E} [\mathbf{E}_2] \right)^H \end{bmatrix} + \mathbf{R}_{\mathbf{H}}^{-1} \right)^{-1} \right] \\ &= \mathbb{E} \left[\left(\frac{\tau_{\text{T}} P_{\text{T}}}{N_0} \mathbf{I}_k + \tilde{\mathbf{E}}_2 \mathbf{\Lambda}_{\mathbf{X}} \tilde{\mathbf{E}}_2^H + \mathbf{R}_{\mathbf{H}}^{-1} \right)^{-1} \right] \\ &\approx \left(\frac{\tau_{\text{T}} P_{\text{T}}}{N_0} \mathbf{I}_k + \frac{\tau_{\text{D}} P_{\text{D}}}{\Delta S_{\mathbf{X}} + N_0} \tilde{\mathbf{E}}_2^2 + \mathbf{R}_{\mathbf{H}}^{-1} \right)^{-1} \\ &= \text{diag} \left(\frac{\beta_1^{\text{M}}}{1 + \rho_1^{\text{DA}} \beta_1^{\text{M}}}, \dots, \frac{\beta_K^{\text{M}}}{1 + \rho_K^{\text{DA}} \beta_K^{\text{M}}} \right), \end{aligned} \quad (67)$$

where

$$\rho_k^{\text{DA}} = \frac{\tau_{\text{T}} P_{\text{T}}}{N_0} + \frac{\tau_{\text{D}} P_{\text{D}} (1 - 2\text{BER}_{vk})^2}{\Delta S_{\mathbf{X}} + N_0}, \quad (68)$$

and the approximation above is due to the fact that the UL data length is usually long enough to hold that the two different data stream are uncorrelated. This means that $\hat{\mathbf{X}} \hat{\mathbf{X}}^H \rightarrow \tau_{\text{D}} P_{\text{D}} \mathbf{I}_k$ when τ_{D} is large. Hence, utilizing the above results, we have

$$\mathbb{E} \left[\left\| \hat{\mathbf{h}}_k^{\text{MMSE}} - \mathbf{h}_k \right\|^2 \right] = \frac{M \beta_k^{\text{M}}}{1 + \rho_k^{\text{DA}} \beta_k^{\text{M}}}. \quad (69)$$

Finally, the NMSE for the data-aided method with the MMSE estimator can be obtained based on the definition of NMSE

$$\mathcal{J}_k^{\text{MMSE}} = 10 \log_{10} \left(\frac{1}{1 + \rho_k^{\text{DA}} \beta_k^{\text{M}}} \right), \quad (70)$$

which proves *Proposition 2*. ■

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, 2010.
- [2] F. Boccardi, R. W. Heath, A. Lozano, T. L. Marzetta, and P. Popovski, "Five disruptive technology directions for 5G," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 74–80, 2014.

- [3] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. Soong, and J. C. Zhang, "What will 5G be?" *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, 2014.
- [4] C.-X. Wang, F. Haider, X. Gao, X.-H. You, Y. Yang, D. Yuan, H. M. Aggoune, H. Haas, S. Fletcher, and E. Hepsaydir, "Cellular architecture and key technologies for 5G wireless communication networks," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 122–130, 2014.
- [5] Qualcomm, "The 1000x mobile data challenge," [Online]. Available: <https://www.qualcomm.com/invention/1000x>, 2012.
- [6] Huawei, "5G white paper," [Online]. Available: <http://www.huawei.com/5gwhitepaper/>, 2013.
- [7] J. G. Andrews, "Seven ways that HetNets are a cellular paradigm shift," *IEEE Commun. Mag.*, vol. 51, no. 3, pp. 136–144, 2013.
- [8] A. Adhikary, H. S. Dhillon, and G. Caire, "Massive-MIMO meets HetNet: Interference coordination through spatial blanking," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 6, pp. 1171–1186, 2015.
- [9] J. Hoydis, K. Hosseini, S. t. Brink, and M. Debbah, "Making smart use of excess antennas: Massive MIMO, small cells, and TDD," *Bell Labs Technical Journal*, vol. 18, no. 2, pp. 5–21, 2013.
- [10] E. Björnson, M. Kountouris, and M. Debbah, "Massive MIMO and small cells: Improving energy efficiency by optimal soft-cell coordination," in *Telecommunications (ICT), 2013 20th International Conference on*. IEEE, 2013, pp. 1–5.
- [11] W. Liu, S. Han, C. Yang, and C. Sun, "Massive MIMO or small cell network: Who is more energy efficient?" in *Wireless Communications and Networking Conference Workshops (WCNCW), 2013 IEEE*. IEEE, 2013, pp. 24–29.
- [12] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang, "An overview of massive MIMO: Benefits and challenges," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 742–758, 2014.
- [13] K. Smiljkovic, P. Popovski, and L. Gavrilovska, "Analysis of the decoupled access for downlink and uplink in wireless heterogeneous networks," *IEEE Commun. Lett.*, vol. 4, no. 2, pp. 173–176, 2015.
- [14] S. Singh, X. Zhang, and J. G. Andrews, "Joint rate and SINR coverage analysis for decoupled uplink-downlink biased cell associations in hetnets," *IEEE Trans. Wireless Commun.*, vol. 14, no. 10, pp. 5360–5373, 2015.
- [15] H. Elshaer, F. Boccardi, M. Dohler, and R. Irmer, "Load & backhaul aware decoupled downlink/uplink access in 5G systems," in *2015 IEEE International Conference on Communications (ICC)*. IEEE, 2015, pp. 5380–5385.
- [16] K. Smiljkovic, L. Gavrilovska, and P. Popovski, "Efficiency analysis of downlink and uplink decoupling in heterogeneous networks," in *2015 IEEE International Conference on Communication Workshop (ICCW)*. IEEE, 2015, pp. 125–130.
- [17] F. Boccardi, J. Andrews, H. Elshaer, M. Dohler, S. Parkvall, P. Popovski, and S. Singh, "Why to decouple the uplink and downlink in cellular networks and how to do it," *IEEE Commun. Mag.*, vol. 54, no. 3, pp. 110–117, 2016.
- [18] 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, 2013.
- [19] 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, 2013.
- [20] J. Ma and P. Li, "Data-aided channel estimation in large antenna systems," in *2014 IEEE International Conference on Communications (ICC)*. IEEE, 2014, pp. 4626–4631.
- [21] J. Baltersee, G. Fock, and H. Meyr, "Achievable rate of MIMO channels with data-aided channel estimation and perfect interleaving," *IEEE J. Sel. Areas Commun.*, vol. 19, no. 12, pp. 2358–2368, 2001.
- [22] J. Ma and P. Li, "Data-aided channel estimation in large antenna systems," *IEEE Trans. Signal Process.*, vol. 62, no. 12, pp. 3111–3124, 2014.
- [23] M.-A. Khalighi, J. J. Boutros, and J.-F. Héland, "Data-aided channel estimation for turbo-PIC MIMO detectors," *IEEE Commun. Lett.*, vol. 10, no. 5, pp. 350–352, 2006.

- [24] M. Coldrey and P. Bohlin, "Training-based MIMO systems: Part II- improvements using detected symbol information," *IEEE Trans. Signal Process.*, vol. 56, no. 1, pp. 296–303, 2008.
- [25] N. Jindal, A. Lozano, and T. L. Marzetta, "What is the value of joint processing of pilots and data in block-fading channels?" in *2009 IEEE International Symposium on Information Theory (ISIT)*, 2009, pp. 2189–2193.
- [26] H. Schoeneich and P. A. Hoeher, "Iterative pilot-layer aided channel estimation with emphasis on interleave-division multiple access systems," *EURASIP Journal on Applied Signal Processing*, vol. 2006, pp. 250–250, 2006.
- [27] M. Zhao, Z. Shi, and M. C. Reed, "Iterative turbo channel estimation for OFDM system over rapid dispersive fading channel," *IEEE Trans. Wireless Commun.*, vol. 7, no. 8, pp. 3174–3184, 2008.
- [28] H. Elshaer, F. Boccardi, M. Dohler, and R. Irmer, "Downlink and uplink decoupling: A disruptive architectural design for 5G networks," in *Global Communications Conference (GLOBECOM), 2014 IEEE*, pp. 1798–1803.
- [29] H.-S. Jo, Y. J. Sang, P. Xia, and J. G. Andrews, "Heterogeneous cellular networks with flexible cell association: A comprehensive downlink sinr analysis," *IEEE Trans. Wireless Commun.*, vol. 11, no. 10, pp. 3484–3495, 2012.
- [30] M. K. Steven, "Fundamentals of statistical signal processing, volume I: estimation theory," *PTR Prentice-Hall, Englewood Cliffs, NJ*, 1993.
- [31] P. Li, D. Paul, R. Narasimhan, and J. Cioffi, "On the distribution of SINR for the MMSE MIMO receiver and performance analysis," *IEEE Trans. Inf. Theory*, vol. 52, no. 1, pp. 271–286, 2006.
- [32] H. Gao, P. J. Smith, and M. V. Clark, "Theoretical reliability of MMSE linear diversity combining in Rayleigh-fading additive interference channels," *IEEE Trans. Commun.*, vol. 46, no. 5, pp. 666–672, 1998.
- [33] J. W. Silverstein, "Strong convergence of the empirical distribution of eigenvalues of large dimensional random matrices," *J. Multivariate Anal.*, vol. 55, no. 2, pp. 331–339, 1995.
- [34] Z. D. Bai, J. W. Silverstein *et al.*, "CLT for linear spectral statistics of large-dimensional sample covariance matrices," *Ann. Probab.*, vol. 32, no. 1A, pp. 553–605, 2004.
- [35] J. Hoydis, M. Kobayashi, and M. Debbah, "Asymptotic performance of linear receivers in network MIMO," in *Signals, Systems and Computers (ASILOMAR), 2010 Conference Record of the Forty Fourth Asilomar Conference on.* IEEE, 2010, pp. 942–948.
- [36] Z. Bai and J. W. Silverstein, *Spectral analysis of large dimensional random matrices.* Springer, 2010, vol. 20.
- [37] R. Durrett, *Probability: theory and examples.* Cambridge university press, 2010.
- [38] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 951–963, 2003.
- [39] N. Jindal and A. Lozano, "A unified treatment of optimum pilot overhead in multipath fading channels," *IEEE Commun. Lett.*, vol. 58, no. 10, pp. 2939–2948, 2010.
- [40] "Further enhancements to LTE time division duplex (TDD) for downlink-uplink (DL-UL) interference management and traffic adaptation," 3GPP, TR 36.828, Jun 2012, version 11.0.0.
- [41] M. Ding, P. Wang, D. López-Pérez, G. Mao, and Z. Lin, "Performance impact of LoS and NLoS transmissions in dense cellular networks," *IEEE Trans. Wireless Commun.*, vol. 15, no. 3, pp. 2365–2380, 2016.