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

Abstract

Decoupled access promises more flexible cell association and higher throughput by connecting the optimal base stations (BSs) to users in both uplink and downlink (UL/DL). However, decoupled access breaks reciprocity in time-division-duplex (TDD) operation due to associating different BSs in UL/DL, hampering the acquisition of DL channel state information (CSI) from the UL. This paper addresses this by a three-stage data-aided channel estimation method making use of the decoded UL data assembled at the DL BS combined with known training sequences to better estimate the DL channel. We consider a single-cell heterogeneous network in which a massive multiple-input multiple-output (MIMO) enabled BS (MBS) is in the center of the cell with multiple small cell BSs (SBSs) randomly deployed within the cell. We derive an approximate normalized mean square error (NMSE) expression for our proposed data-aided method using minimum mean square error (MMSE) estimator by considering the pre-estimated bit error ratio (BER) of each uplink sequence. Compared with NMSE for conventional least square (LS) and MMSE, the results demonstrate that there is an additional increment of signal-to-noise ratio (SNR)-like term in the NMSE of the data-aided MMSE estimator, introducing more degrees of freedom like power and length of UL data. In particular, higher power and longer length of UL data as well as lower UL data BER lead to more accurate channel estimates in certain ways. Simulation results verify that our analytical results of BER and NMSE are close to the simulated ones and the data-aided method offers a remarkable gain over conventional channel estimation methods in the literature.

Index Terms

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

I. Introduction

Heterogenous 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× increase of throughput and 1000× 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 decoupled access as a strong candidate in the next generation network, though there are still major problems to be solved before facilitating this structure, including 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 the UL and then treated as the real channels for the DL. Unfortunately, channel reciprocity never holds in HetNets with decoupled access since UEs may associate with different BSs in UL/DL.

This has motivated our work to propose a data-aided channel estimation scheme to enable decoupled 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 before, e.g., [20–27]. 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]. The findings were that data-aided methods can effectively suppress the contamination effect in large-scale antenna systems [20, 22]. Despite this, prior works mainly focused on

homogeneous networks. 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 a cellular HetNet with decoupled access.

In this paper, we consider a single-cell HetNet with decoupled access in which a macro base station (MBS) is in the center and small cell BSs (SBSs) are randomly but densely populated within the cell. We propose a novel three-stage data-aided scheme to solve the channel estimation problem for decoupled UEs where channel reciprocity no longer exists. The estimation process takes place in a few stages. In the first stage, all UEs send prescribed orthogonal training sequences to their serving SBSs simultaneously. Then the uplink training signal is received and proceeded to the channel estimator at each SBS while the MBS[1] needs to store the received uplink training signal for subsequent estimation. In the second stage, the SBSs receive the uplink data signals from all the UEs and try to recover the data of the UEs they serve. The uplink data signal is also recorded and stored at the MBS. At the same time, we calculate the average bit error ratio (BER) of each data stream by averaging over fast fading once the locations of a UE and its connected SBSs are determined. Then the decoded data along with the BER value are sent to the MBS via an error-free latency-free backhaul link. In the last stage, the channel estimator at the MBS estimates the decoupled UEs' channels from the combined training signals and data signals with the knowledge of training sequences and the decoded data sequences.

The core idea of this scheme is to utilize the decoded data as extended training sequences 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 model 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, pointing out that estimation performance can be greatly improved and both co-channel interference and detection error could seriously compromise performance.

One might speculate that it would be possible to directly use the training sequences received at the MBS for estimating the channel. However, it should be noted that decoupled UEs are neither in the vicinity of the MBS nor that of an SBS (because UEs near the MBS or an SBS tend to connect to that same BS in both directions), and therefore suffer the worst performance in DL/UL of the whole

[1]In the two-tier model, decoupled UEs usually connect to an SBS in the uplink and an MBS in the DL due to the user association policy.

network. Decoupled access makes it possible for those UEs to flexibly choose BSs in both directions, which promises better UL performance but does little in enhancing DL performance. As a result, there is a huge UL performance gap between DL MBS and UL SBS, and the channel estimation quality at the MBS needs to be improved. In this light, our proposed data-aided scheme is particularly attractive, capable of enhancing not only UL performance but also DL performance of decoupled UEs.

More specifically, our main contribution is a novel three-stage data-aided channel estimation scheme for decoupled UEs in cellular HetNets by using decoded uplink data with consideration of BER. We calculate the BER of modulated uplink data after the MMSE decoder with imperfect channel state information (CSI) and use this estimated BER to derived a closed-form approximated normalized mean-square-error (NMSE) expression for the proposed data-aided scheme. We further compare the NMSE performance between conventional LS, MMSE and data-aided MMSE by theoretical analysis and simulations, and find that NMSE is commonly determined by an SNR-like term. For conventional estimators, the SNR-like term is $\rho^{\text{Con}} = \frac{\tau_{\text{T}} P_{\text{T}}}{N_0}$ whereas for data-aided schemes it is expressed in the form $\rho_k^{\text{DA}} = \frac{\tau_{\text{T}} P_{\text{T}}}{N_0} + \frac{\tau_{\text{D}} P_{\text{D}} (1-2\text{BER}_{vk})^2}{\Delta S_{\mathbf{x}} + N_0}$ where UL data power, UL data length and BER are all involved in the additional second term, which is the key reason why data-aided schemes outperform the conventional ones.

The reminder of this paper is organized as follows. Section II introduces the system model of UL/DL decoupled HetNets. A novel and practical three-stage data-aided scheme for decoupled systems 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 confirm that data-aided estimation outperforms conventional ones 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 generally 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 and error.

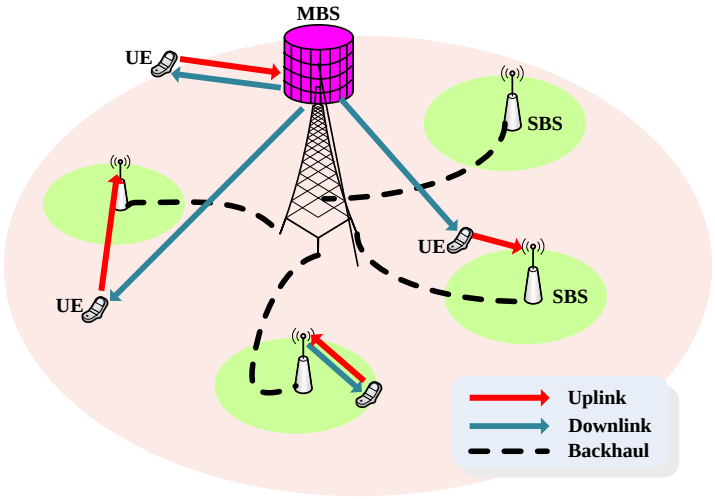

Fig. 1. System model for cellular HetNets with decoupled access.

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

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

where each element of $\mathbf{h}_k^{\mathrm{W}} \in \mathbb{C}^{M\times 1}$ and $\mathbf{g}_{sk}^{\mathrm{W}} \in \mathbb{C}^{N\times 1}$ follows $\mathcal{CN}(0,1)$, β_k^{M} and β_{sk}^{S} represent the large scale fading from the kth UE to the MBS and the sth SBS, respectively, by neglecting shadowing effect and differences among antennas. The large scale fading between the kth UE to the MBS is modelled as $\beta_k^{\mathrm{M}} = \left(d_k^{\mathrm{M}}\right)^{-\alpha}$, 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 antenna number 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 in DL at the kth UE from a BS is given by

$$Q_{vk}^{\mathrm{DL}} = \begin{cases} \frac{P_{\mathrm{M}}}{M}\|\mathbf{h}_k\|^2, & v = 0, \\ \frac{P_{\mathrm{S}}}{N}\|\mathbf{g}_{vk}\|^2, & v = 1, 2, \ldots, S, \end{cases} \tag{2}$$

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 vth SBS. Similarly, the received signal in UL at a BS from UE k can be expressed as

$$Q_{vk}^{\mathrm{UL}} = \begin{cases} P_{\mathrm{D}}\|\mathbf{h}_k\|^2, & v = 0, \\ P_{\mathrm{D}}\|\mathbf{g}_{vk}\|^2, & v = 1, 2, \ldots, S, \end{cases} \tag{3}$$

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 kth UE as an example using this modified MARP strategy:

$$\begin{aligned} \mathcal{D}_k &= \underset{v=0,1,\ldots,S}{\arg\max} \left\{ \underset{\mathbf{h}_k^{\mathrm{W}}, \mathbf{g}_{sk}^{\mathrm{W}}}{\mathbb{E}} \left[Q_{vk}^{\mathrm{UL}} \right] \right\} = \underset{v=0,1,\ldots,S}{\arg\max} \left\{ P_{\mathrm{M}} \big(d_k^{\mathrm{M}}\big)^{-\alpha}, P_{\mathrm{S}} \big(d_{1k}^{\mathrm{S}}\big)^{-\alpha}, \ldots, P_{\mathrm{S}} \big(d_{Sk}^{\mathrm{S}}\big)^{-\alpha} \right\}, \\ \mathcal{U}_k &= \underset{v=0,1,\ldots,S}{\arg\max} \left\{ \underset{\mathbf{h}_k^{\mathrm{W}}, \mathbf{g}_{sk}^{\mathrm{W}}}{\mathbb{E}} \left[Q_{vk}^{\mathrm{DL}} \right] \right\} = \underset{v=0,1,\ldots,S}{\arg\max} \left\{ M P_{\mathrm{D}} \big(d_k^{\mathrm{M}}\big)^{-\alpha}, N P_{\mathrm{D}} \big(d_{1k}^{\mathrm{S}}\big)^{-\alpha}, \ldots, N P_{\mathrm{D}} \big(d_{Sk}^{\mathrm{S}}\big)^{-\alpha} \right\}, \end{aligned} \tag{4}$$

where $\mathcal{D} = \{\mathcal{D}_1, \ldots, \mathcal{D}_K\}$ and $\mathcal{U} = \{\mathcal{U}_1, \ldots, \mathcal{U}_K\}$ are DL and UL association information sets. Based on the above results, we can focus on the UEs with indices belonging to the set $\{k | \mathcal{D}_k \neq \mathcal{U}_k, k = 1, \ldots, K\}$, who can now be 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, as mentioned previously, it is difficult for the BS with which decoupled UEs is associated in DL to acquire accurate CSI due to the absence of reciprocity. For the rest of this paper, we are devoted to addressing this issue.

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 uplink training and uplink data transmission, and the DL BS is required to listen and record the uplink 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.

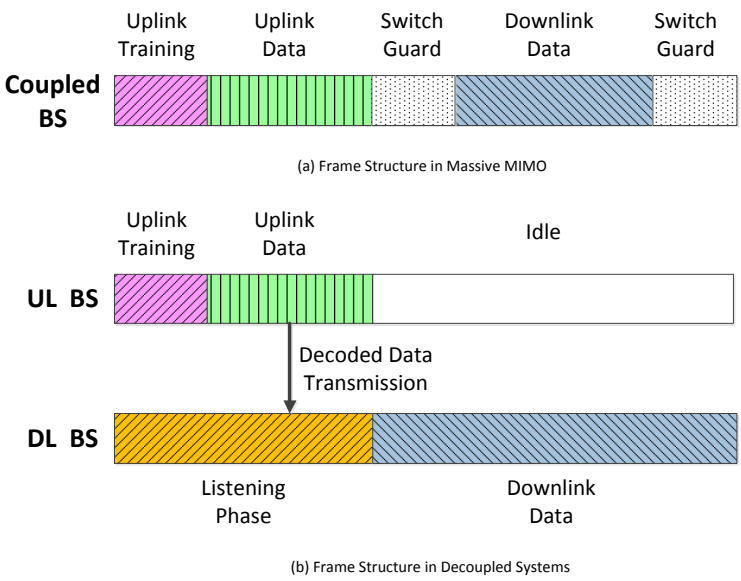

Fig. 2. The frame structure for the data-aided channel estimation scheme.

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 $\mathbf{S} = \left[\mathbf{s}_1^{\mathrm{T}}, \cdots \mathbf{s}_k^{\mathrm{T}}, \cdots, \mathbf{s}_K^{\mathrm{T}}\right]^{\mathrm{T}}$ denotes the pilot matrix where $\mathbf{s}_k \in \mathbb{C}^{1\times\tau_{\mathrm{T}}}$ is the pilot for the kth UE. Obviously, we have $\mathbf{S}\mathbf{S}^{\mathrm{H}} = \tau_{\mathrm{T}} P_{\mathrm{T}} \mathbf{I}_K$ and $\tau_{\mathrm{T}} \geq K$. The main task here is for all UL BSs to recover channels from the uplink 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 vth SBS can be expressed as

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

where $\mathbf{N}_v \in \mathbb{C}^{N\times\tau_{\mathrm{T}}}$ 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 kth decoupled UE connecting to the vth SBS in UL is given as

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

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. We

also use the MMSE channel estimator as a benchmark and the estimated channel is given by

$$\hat{\mathbf{g}}_{vk}^{\mathrm{MMSE}} = \mathbf{Y}_v \mathbf{C}_{vk}^{\mathrm{opt}} = \mathbf{G}_v \mathbf{S} \mathbf{C}_{vk}^{\mathrm{opt}} + \mathbf{N}_v \mathbf{C}_{vk}^{\mathrm{opt}}, \tag{7}$$

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

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

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

B. Stage 2: Uplink Data Transmission

In the second stage, UEs send uplink 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 uplink 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 uplink data of totally τ_{D} symbols to its associated UL BS. The matrix $\mathbf{X} = \left[\mathbf{x}_1^{\mathrm{T}}, \cdots \mathbf{x}_k^{\mathrm{T}}, \cdots, \mathbf{x}_K^{\mathrm{T}}\right]^{\mathrm{T}}$ is used to represent data sequences of UEs with each element randomly chosen from the set $\{P_{\mathrm{D}}, -P_{\mathrm{D}}\}$ and $\mathbf{x}_k \in \mathbb{C}^{1\times\tau_{\mathrm{D}}}$ is the uplink data for the kth 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 vth SBS is expressed as

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

where $\tilde{\mathbf{N}}_v \in \mathbb{C}^{N\times\tau_{\mathrm{D}}}$ 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 uplink data of the kth UE. We assume that the kth UE is associated with the vth SBS in UL. The MRC detector only needs channel information of the kth 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 $\mathcal{G}_{vk} = [\hat{\mathbf{g}}_{vk}, \cdots]$, which involves the estimated channels of UEs associated with the vth SBS, knowing that the kth UE is included at least. Thus, the combination matrix $\mathbf{A}$ for MRC and ZF detectors are given by

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

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

$$\hat{\mathbf{x}}_k^{\mathrm{ZF,MRC}} = \mathrm{Decoder}\left(\left\lceil \mathbf{A}^{\mathrm{H}}\tilde{\mathbf{Y}}_v \right\rceil_1\right), \tag{11}$$

where $\lceil \mathbf{B} \rceil_i$ performs the operation of taking the ith row of $\mathbf{B}$ and $\mathrm{Decoder}(\cdot)$ 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 $\tilde{\mathbf{G}} = \mathbf{G}_v - \hat{\mathbf{G}}_v$, and the received signal can thus be rewritten as

$$\tilde{\mathbf{Y}}_v = \hat{\mathbf{G}}_v\mathbf{X} + \tilde{\mathbf{G}}_v\mathbf{X} + \tilde{\mathbf{N}}_v. \tag{12}$$

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 kth UE can be written as

$$\mathbf{C}_{vk}^{\mathrm{opt}} = \hat{\mathbf{g}}_{vk}^{\mathrm{H}}\left(\hat{\mathbf{G}}_v\hat{\mathbf{G}}_v^{\mathrm{H}} + \left(\sum_{k=1}^{K}\frac{N_0\beta_{vk}^{\mathrm{S}}}{N_0 + \beta_{vk}^{\mathrm{S}}P_{\mathrm{T}}\tau_{\mathrm{T}}} + \frac{N_0}{P_{\mathrm{D}}}\right)\mathbf{I}\right)^{-1}. \tag{13}$$

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

$$\hat{\mathbf{x}}_k^{\mathrm{MMSE}} = \mathrm{Decoder}\left(\mathbf{C}_{vk}^{\mathrm{opt}}\tilde{\mathbf{Y}}_v\right). \tag{14}$$

Here, three detection methods are considered to detect 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 uplink training and data signal. As a result, in this stage, we can utilize the recorded uplink signal along with the known uplink training sequences and the decoded uplink data to perform channel estimation at the MBS. 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

$$\mathbf{M} = \mathbf{H}\mathbf{W} + \mathbf{Z} = \sum_{k=1}^{K} \mathbf{h}_k \mathbf{w}_k + \mathbf{Z}, \tag{15}$$

where $\mathbf{W} = [\mathbf{S}, \mathbf{X}]$ with $\mathbf{w}_k$ being its kth row and $\mathbf{Z} \in \mathbb{C}^{M \times (\tau_{\mathrm{T}} + \tau_{\mathrm{D}})}$ 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 kth UE to MBS can be expressed as

$$\hat{\mathbf{h}}_k^{\mathrm{MMSE}} = \mathbf{M}\mathbf{C}_k^{\mathrm{opt}}, \tag{16}$$

and

$$\mathbf{C}_{vk}^{\mathrm{opt}} = \left[\sum_{i=1}^{K} \beta_k^{\mathrm{M}} \mathbb{E}\left[\mathbf{w}_k^{\mathrm{H}} \mathbf{w}_k\right] + N_0 \mathbb{E}\left[\mathbf{Z}^{\mathrm{H}} \mathbf{Z}\right] \right]^{-1} \beta_k^{\mathrm{M}} \mathbb{E}\left[\mathbf{w}_k^{\mathrm{H}}\right]. \tag{17}$$

Eventually, the channels from the decoupled UEs to the DL BS are obtained by this data-aided method. In this section, we only give the expression of the combination matrix due to the complex structure of this joint estimator. Specific results of the data-aided scheme will be discussed later and the actual performance of this scheme will be analyzed and compared with the conventional estimators.

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

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

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 uplink training sequences into account. Herein, we take NMSE of the kth UE associated with the vth 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 consider the performance of the LS channel estimator, from (6), the numerator of NMSE for LS can be calculated as

$$\mathbb{E}\left[\left\|\mathbf{g}_{vk}-\hat{\mathbf{g}}_{vk}^{\mathrm{LS}}\right\|^2\right]=\mathbb{E}\left[\left\|\frac{1}{\tau_{\mathrm{T}}P_{\mathrm{T}}}\mathbf{N}_v\mathbf{s}_k^{\mathrm{H}}\right\|^2\right]=\frac{NN_0}{\tau_{\mathrm{T}}P_{\mathrm{T}}}, \tag{19}$$

and the denominator can be straightforwardly written as

$$\mathbb{E}\left[\left\|\mathbf{g}_{vk}\right\|^2\right]=\mathbb{E}\left[\beta_{vk}^{\mathrm{S}}\left(\mathbf{g}_{sk}^{\mathrm{W}}\right)^{\mathrm{H}}\mathbf{g}_{sk}^{\mathrm{W}}\right]=N\beta_{vk}^{\mathrm{S}}. \tag{20}$$

As a result, the NMSE for the LS estimator of the kth UE associated with the vth SBS is given by

$$\mathcal{J}_{vk}^{\mathrm{S,LS}}=10\log_{10}\left(\frac{N_0}{\beta_{vk}^{\mathrm{S}}\tau_{\mathrm{T}}P_{\mathrm{T}}}\right). \tag{21}$$

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

Before analyzing the NMSE for the MMSE estimator, we are supposed to determine a closed form of the estimated channel vector. By inserting (8) into (7), the estimated channel of the kth UE to the vth SBS can be further derived with the help of the Woodbury matrix inversion identity and given by

$$\begin{aligned}\hat{\mathbf{g}}_{vk}&=\left(\mathbf{G}_v\mathbf{S}+\mathbf{N}_v\right)\left(\mathbf{S}^{\mathrm{H}}\mathbf{R}_{\mathbf{g}}\mathbf{S}+\mathbf{R}_{\mathbf{N}_v}\right)^{-1}\mathbf{s}_k^{\mathrm{H}}\mathbf{R}_{\mathbf{g}_{vk}}\\&=\frac{\mathbf{R}_{\mathbf{g}_{vk}}}{\left(NN_0+\mathbf{R}_{\mathbf{g}_{vk}}P_{\mathrm{T}}\tau_{\mathrm{T}}\right)}\left(\tau_{\mathrm{T}}P_{\mathrm{T}}\mathbf{g}_{vk}+\mathbf{N}_v\mathbf{s}_k^{\mathrm{H}}\right),\end{aligned} \tag{22}$$

where $\mathbf{R}_{\mathbf{g}_{vi}}=\mathbb{E}\left[\mathbf{g}_{vi}^{\mathrm{H}}\mathbf{g}_{vi}\right]$, $\mathbf{R}_{\mathbf{g}}=\mathbb{E}\left[\mathbf{G}_v^{\mathrm{H}}\mathbf{G}_v\right]=\mathrm{diag}\left(\mathbf{R}_{\mathbf{g}_{v1}},\ldots,\mathbf{R}_{\mathbf{g}_{vi}},\ldots,\mathbf{R}_{\mathbf{g}_{vK}}\right)$, $\mathbf{R}_{\mathbf{N}_v}=\mathbb{E}\left[\mathbf{N}_v^{\mathrm{H}}\mathbf{N}_v\right]$.

Since the training sequences satisfy $\mathbf{S}\mathbf{S}^{\mathrm{H}}=\tau_{\mathrm{T}}P_{\mathrm{T}}\mathbf{I}_K$ and each element of $\mathbf{N}_v$ is i.i.d. $\mathcal{CN}(0,N_0)$, it can be inferred that the matrix $\mathbf{N}_v\mathbf{s}_k^{\mathrm{H}}$ is distributed as $\mathcal{CN}(0,\tau_{\mathrm{T}}P_{\mathrm{T}}N_0\mathbf{I})$. Observing the last equation of (22), $\mathbf{g}_{vk}$ and $\mathbf{N}_v\mathbf{s}_k^{\mathrm{H}}$ are independent complex Gaussian random variables, and therefore, it is the property of Gaussian distribution that $\hat{\mathbf{g}}_{vk}$ is also a complex Gaussian variable following

$$\hat{\mathbf{g}}_{vk}\sim\mathcal{CN}\left(0,\frac{\left(\beta_{vk}^{\mathrm{S}}\right)^2P_{\mathrm{T}}\tau_{\mathrm{T}}}{N_0+\beta_{vk}^{\mathrm{S}}P_{\mathrm{T}}\tau_{\mathrm{T}}}\mathbf{I}\right). \tag{23}$$

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

$$\tilde{\mathbf{g}}_{vk}\sim\mathcal{CN}\left(0,\frac{N_0\beta_{vk}^{\mathrm{S}}}{N_0+\beta_{vk}^{\mathrm{S}}P_{\mathrm{T}}\tau_{\mathrm{T}}}\mathbf{I}\right). \tag{24}$$

In parallel to the LS estimator, the NMSE for the MMSE estimator of the kth UE associated with the sth SBS can be calculated by using equations (20) and (24)

$$\mathcal{J}_{vk}^{\mathrm{S,MMSE}} = 10\log_{10}\left(\frac{\mathbb{E}\left[\left\|\mathbf{g}_{vk}-\hat{\mathbf{g}}_{vk}\right\|^2\right]}{\mathbb{E}\left[\left\|\mathbf{g}_{vk}\right\|^2\right]}\right) = 10\log_{10}\left(\frac{N_0}{N_0+\beta_{vk}^{\mathrm{S}}P_{\mathrm{T}}\tau_{\mathrm{T}}}\right). \tag{25}$$

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 the two NMSE expressions, the only difference appears to be the extra noise variance term N_0 in the numerator of NMSE for MMSE within the log operation. Hence, the performance gain of the MMSE estimator over LS is obtained in the regime of low SNR and decreases to 0 when SNR gets higher.

B. Performance of the Data-Aided Channel Estimation Method

To evaluate our proposed data-aided method, the process of uplink 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 expression below:

$$\mathrm{BER}\left(\mathrm{SINR}\right) = \int_{\sqrt{\mathrm{SINR}}}^{\infty}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}t^2\right)\mathrm{d}t. \tag{26}$$

Considering the randomness of SINR (signal-to-interference plus noise ratio), the ergodic BER for the MMSE detector can be further written as

$$\mathrm{BER} = \int_0^{\infty} f_{\mathrm{SINR}}\left(x\right)\int_{\sqrt{x}}^{\infty}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}t^2\right)\mathrm{d}t\ \mathrm{d}x, \tag{27}$$

where $f_{\mathrm{SINR}}\left(x\right)$ is the probability density function (p.d.f.) of SINR.

As a matter of fact, the exact distribution of SINR for the MMSE detector has been derived in [31], but the distribution is too complex to obtain a close-form expression of BER. Therefore, we resort to the well-known tractable Gamma distribution, whose p.d.f. is given as

$$f_{\mathrm{Gamma}}\left(x;\alpha,\xi\right) = \frac{x^{\alpha-1}e^{-\frac{x}{\xi}}}{\Gamma\left(\alpha\right)\xi^{\alpha}} \tag{28}$$

to approximate the result.

Similar approach to obtain closed-form solutions has been applied in many previous efforts, see, e.g.,

[18, 32]. Following a similar procedure in [32], by determining two parameters of the approximated Gamma distribution, a closed-form expression of BER can be obtained. However, different from [18, 32], in our case, the antenna number at each SBS is assumed to be less than the number of UEs. Thus, in the following, the two parameters of Gamma distribution are determined by moment matching.

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

$$\mathrm{SINR}_{vk} = \frac{1}{\left(\left(\mathbf{I} + \rho_v \hat{\mathbf{G}}_v^{\mathrm{H}} \hat{\mathbf{G}}_v\right)^{-1}\right)_{kk}} - 1, \tag{29}$$

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

Then by applying [32, (8)], SINR_{vk} can be further expressed as

$$\mathrm{SINR}_{vk} = \rho_v \hat{\mathbf{g}}_{vk}^{\mathrm{H}} \hat{\mathbf{g}}_{vk} - \rho_v^2 \hat{\mathbf{g}}_{vk}^{\mathrm{H}} \hat{\mathbf{G}}_{v(-k)} \hat{\mathbf{G}}_{v(-k)}^{\mathrm{H}} \left(\mathbf{I} + \rho_v \hat{\mathbf{G}}_{v(-k)} \hat{\mathbf{G}}_{v(-k)}^{\mathrm{H}}\right)^{-1} \hat{\mathbf{g}}_{vk}, \tag{30}$$

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

Considering the singular value decomposition (SVD), we have $\hat{\mathbf{G}}_{v(-k)} = \mathbf{U}\mathbf{D}\mathbf{V}^{\mathrm{H}}$, $\mathbf{U} \in \mathbb{C}^{N \times N}, \mathbf{D} \in \mathbb{C}^{N \times N}, \mathbf{V}^{\mathrm{H}} \in \mathbb{C}^{N \times (K-1)}$, $\mathbf{U}\mathbf{U}^{\mathrm{H}} = \mathbf{U}^{\mathrm{H}}\mathbf{U} = \mathbf{I}_N$ and $\mathbf{V}^{\mathrm{H}}\mathbf{V} = \mathbf{I}_N$. Then applying SVD to (30), we get

$$\begin{aligned} \mathrm{SINR}_{vk} &= \rho_v \hat{\mathbf{g}}_{vk}^{\mathrm{H}} \hat{\mathbf{g}}_{vk} - \rho_v^2 \hat{\mathbf{g}}_{vk}^{\mathrm{H}} \mathbf{U}\mathbf{D}^2 \left(\mathbf{I} + \rho_v \mathbf{D}^2\right)^{-1} \mathbf{U}^{\mathrm{H}} \hat{\mathbf{g}}_{vk} \\ &= \rho_v \boldsymbol{\varphi}^{\mathrm{H}} \boldsymbol{\varphi} - \rho_v^2 \boldsymbol{\varphi}^{\mathrm{H}} \mathbf{D}^2 \left(\mathbf{I} + \rho_v \mathbf{D}^2\right)^{-1} \boldsymbol{\varphi} \\ &= \rho_v \sum_{i=1}^{N} \frac{\|\varphi_i\|^2}{1 + \rho_v d_i^2}, \end{aligned} \tag{31}$$

where $\mathbf{U}^{\mathrm{H}} \hat{\mathbf{g}}_{vk} \triangleq \boldsymbol{\varphi}$ and d_i is the ith 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 [32] 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 ESD of $\rho_v \hat{\mathbf{G}}_{v(-k)} \hat{\mathbf{G}}_{v(-k)}^{\mathrm{H}}$, denoted by $\hat{\mathbf{J}}$, converges to a measure $\mathbf{J}$, whose Stieltjes transform, denoted by $\mathcal{T}(z)$, is given as

$$\mathcal{T}(z) \triangleq \int \frac{1}{x - z} \mathbf{J}(\mathrm{d}x). \tag{32}$$

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

$$\mathcal{T}(z) \approx \left(\sum_{i \neq k}^{K} \frac{\mathbf{V}_i}{1 + \mathrm{Tr}\left(\mathbf{V}_i \mathcal{T}(z)\right)} - z\mathbf{I} \right)^{-1}, \tag{33}$$

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

$$\mathcal{T}'(z) \triangleq \int \frac{1}{(x-z)^2} \mathbf{J}(\mathrm{d}x) = \mathcal{T}^2(z) \approx \left(\sum_{i \neq k}^{K} \frac{\mathbf{V}_i}{1 + \mathrm{Tr}\left(\mathbf{V}_i \mathcal{T}(z)\right)} - z\mathbf{I} \right)^{-2}. \tag{34}$$

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

$$\frac{\mathrm{Tr}(\mathbf{\Lambda})}{N} = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{1 + \rho_v d_i^2} = \int \frac{1}{x+1} \hat{\mathbf{J}}(\mathrm{d}x) \xrightarrow{p} \int \frac{1}{x+1} \mathbf{J}(\mathrm{d}x) = \mathcal{T}(-1) \triangleq \mu, \tag{35}$$

and

$$\frac{\mathrm{Tr}\left(\mathbf{\Lambda}^2\right)}{N} = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{\left(1 + \rho_v d_i^2\right)^2} = \int \frac{1}{(x+1)^2} \hat{\mathbf{J}}(\mathrm{d}x) \xrightarrow{p} \int \frac{1}{(x+1)^2} \mathbf{J}(\mathrm{d}x) = \mathcal{T}'(-1) \triangleq \sigma^2, \tag{36}$$

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

By solving (33) and (34), 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

$$\mathbb{E}[\mathrm{SINR}_{vk}] \approx N \rho_v \hat{\beta}_{vk}^{\mathrm{S}} \mu, \tag{37}$$

$$\mathrm{Var}[\mathrm{SINR}_{vk}] \approx N \left(\rho_v \hat{\beta}_{vk}^{\mathrm{S}} \right)^2 \sigma^2. \tag{38}$$

Proof 1: 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 kth UE served by the vth SBS employing MMSE data detection with imperfect CSI (acquired by the MMSE estimator), the BER of its uplink data can be expressed as

$$\mathrm{BER}_{vk} = \frac{\Gamma\left(\alpha_{vk} + \frac{1}{2}\right)}{\Gamma(\alpha_{vk}) 2\sqrt{2\pi}} \frac{\xi_{vk}^{-\alpha_{vk}}}{\alpha_{vk} \left(\frac{1}{\xi_{vk}} + \frac{1}{2} \right)^{\alpha_{vk} + \frac{1}{2}}} {}_2F_1\left(1, \alpha_{vk} + \frac{1}{2}; \alpha_{vk} + 1; \frac{\frac{1}{\xi_{vk}}}{\frac{1}{\xi_{vk}} + \frac{1}{2}} \right), \tag{39}$$

where we have $\mathrm{SINR}_{vk} \sim \mathrm{Gamma}(\alpha_{vk}, \xi_{vk})$ with $\alpha_{vk} = N \frac{\mu^2}{\sigma^2}, \xi_{vk} = \rho_v \hat{\beta}_{vk}^{\mathrm{S}} \frac{\sigma^2}{\mu}$.

Proof 2: Based on (37) and (38), the distribution of SINR_{vk} can be determined by moment matching.

Then plugging its distribution in (27), we obtain the final result after the 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 in order to gain some insight.

Corollary 1: According to the result in (26), we find that BER is a strict concave function. Therefore, by applying Jensen's inequality, the ergodic BER can be upper bounded by

$$\mathrm{BER}_{vk}^{\mathrm{Upper}} = \mathrm{Q}\left(\sqrt{\mathbb{E}[\mathrm{SINR}_{vk}]}\right) = \mathrm{Q}\left(\sqrt{\alpha_{vk}\xi_{vk}}\right), \tag{40}$$

where $\mathrm{Q}(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}t^2\right) \mathrm{d}t$. The upper 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 the BER performance of UL data is 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

$$\mathbf{W} \triangleq \hat{\mathbf{W}} \circ \mathbf{E}, \tag{41}$$

where $\circ$ represents Hadamard product of two matrices. Note that it can further be decomposed into two parts as $\hat{\mathbf{W}} \circ \mathbf{E} = \left[\mathbf{S}, \hat{\mathbf{X}}\right] \circ [\mathbf{E}_1, \mathbf{E}_2]$. 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

$$\begin{aligned} \mathbb{E}\left[e_{ij}\right] &= 1 - 2\mathrm{BER}_{vi}, \\ \mathbb{E}\left[\left\|e_{ij}\right\|^2\right] &= 1, \end{aligned} \tag{42}$$

where $e_{ij} = [\mathbf{E}_2]_{ij}$. Intuitively, we assume that each bit in $\hat{\mathbf{X}}$ from each uplink data stream has the same probability of error, regardless of its location and what is actually sent, meaning that e_{ij} for $j = 1, \ldots, \tau_{\mathrm{D}}$ 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 kth UE at the MBS by using

data-aided channel estimation can be written as

$$\hat{\mathbf{h}}_k^{\mathrm{MMSE}} = (\mathbf{HW} + \mathbf{Z})\, \mathbf{C}_k^{\mathrm{opt}}, \tag{43}$$

$$\mathbf{C}_k^{\mathrm{opt}} = (\mathbf{P} + N_0\mathbf{I})^{-1} \left(\hat{\mathbf{w}}_k^{\mathrm{H}} \circ \mathbb{E}\left[\mathbf{e}_k^{\mathrm{H}}\right]\right) \beta_k^{\mathrm{M}}, \tag{44}$$

where $\mathbf{e}_k$ is the kth row of $\mathbf{E}$ with

$$\mathbb{E}\left[\mathbf{e}_k^{\mathrm{H}}\right] = \left(\underbrace{1 \cdots 1}_{\tau_{\mathrm{T}}} \ \underbrace{1 - 2\mathrm{BER}_{vk} \cdots 1 - 2\mathrm{BER}_{vk}}_{\tau_{\mathrm{D}}} \right)^{\mathrm{H}}, \tag{45}$$

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

$$\hat{\mathbf{P}} = \left(\left[\mathbf{S}, \hat{\mathbf{X}} \circ \mathbb{E}\left[\mathbf{E}_2\right]\right]\right)^{\mathrm{H}} \mathbf{R}_{\mathbf{H}} \left(\left[\mathbf{S}, \hat{\mathbf{X}} \circ \mathbb{E}\left[\mathbf{E}_2\right]\right]\right), \tag{46}$$

$$\Delta\mathbf{P} \triangleq \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \Delta\mathbf{P}_{\mathbf{X}} \end{pmatrix}, \tag{47}$$

where $\mathbf{R}_{\mathbf{H}} = \mathrm{diag}\left(\beta_1^{\mathrm{M}}, \ldots, \beta_k^{\mathrm{M}}, \ldots, \beta_K^{\mathrm{M}}\right), \Delta\mathbf{P}_{\mathbf{X}} = \Delta S_{\mathbf{X}} \mathbf{I}_{\tau_{\mathrm{D}}}, \Delta S_{\mathbf{X}} = P_{\mathrm{D}} \sum_{k=1}^{K} \beta_k^{\mathrm{M}} \left\{1 - (1 - 2\mathrm{BER}_{vk})^2\right\}$.

The NMSE of the channel from the kth UE to the MBS by using the data-aided channel estimation method with MMSE can be calculated asymptotically as

$$\mathcal{J}_k^{\mathrm{MMSE}} = 10 \log_{10}\left(\frac{1}{1 + \rho_k^{\mathrm{DA}} \beta_k^{\mathrm{M}}}\right), \tag{48}$$

where

$$\rho_k^{\mathrm{DA}} = \frac{\tau_{\mathrm{T}} P_{\mathrm{T}}}{N_0} + \frac{\tau_{\mathrm{D}} P_{\mathrm{D}} \left(1 - 2\mathrm{BER}_{vk}\right)^2}{\Delta S_{\mathbf{X}} + N_0}. \tag{49}$$

Proof 3: 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 uplink data transmission power, length of uplink data sequence and BER of itself and other UEs. It is anticipated that better BER, higher data transmit power and longer uplink 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 (21) and (25).

TABLE I
NMSE FOR DIFFERENT CHANNEL ESTIMATION METHODS

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

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 uplink data signal to explore channel information it carries. In addition, in our two-tier network model, decoded uplink data streams at the SBSs will be assembled at the central MBS via capacity-rich backhaul links. As such, we can utilize the collected uplink 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 uplink data power, uplink 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 intuition 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, uplink data length and uplink power, NMSE will monotonically decrease and better channel estimation can be achieved.
- When we reduce our generalized model to the scenario without uplink data BER, it basically corresponds to the case where all uplink 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 uplink 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 $\rho_k^{\mathrm{DA}} = \frac{\tau_{\mathrm{T}} P_{\mathrm{T}}}{N_0}$. 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 uplink 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 can be written as

$$\frac{\tau_{\mathrm{D}} \left(1 - 2\mathrm{BER}_{v1}\right)^2}{\sum\limits_{k}^{K} \beta_k^{\mathrm{M}} \left\{1 - \left(1 - 2\mathrm{BER}_{vk}\right)^2\right\} + \frac{N_0}{P_{\mathrm{D}}}} \tag{50}$$

and this term would eventually saturate to

$$\frac{\tau_{\mathrm{D}} \left(1 - 2\mathrm{BER}_{v1}\right)^2}{\sum\limits_{k}^{K} \beta_k^{\mathrm{M}} \left\{1 - \left(1 - 2\mathrm{BER}_{vk}\right)^2\right\}} \tag{51}$$

when P_{D} approaches infinity. Therefore, there is a performance upper bound 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 association policy mentioned previously. As a result, there are a number of decoupled UEs who connect to different BSs in UL/DL. Note that the numerical results provided focus on and are averaged over all decoupled UEs, and all the points the simulations were obtained via 100 association patterns and 1000 independent channel realizations. Unless otherwise specified, the parameters in this section are those listed in TABLE

II. In the results, PO corresponds to "pilot only" and DA represents the "data-aided" method.

TABLE II
SIMULATION PARAMETERS

Parameters	Values
Radius of macro cell R_{M}	1000 m
Number of SBSs S	30
Number of UEs K	30
Bandwidth W	20 MHz
Noise power spectral density N_0	-174 dBm/Hz
Antenna number of MBS M	128
Antenna number of SUEs N	4/8
Transmit power of MBS P_{M}	46 dBm
Transmit power of SBS P_{S}	24 dBm
Training power of UEs P_{T}	-7$\sim$23 dBm
Data power of UEs P_{D}	-7$\sim$23 dBm
Training length τ_{T}	30
Data length τ_{D}	128

We define the worst SNR, which represents the horizontal axis, given by

$$\mathrm{SNR}_x = \frac{P_x \beta^y_{\mathrm{worst}}}{N_0}, \tag{52}$$

where $x \in \{\mathrm{T}, \mathrm{D}\}$ represents types of power, $y \in \{\mathrm{S}, \mathrm{M}\}$ indicates types of BS, P_x is the UE transmit power of x-type, and β^y_{worst} is large scale fading between a y-type BS and the UE at the cell edge.

Fig. 3 represents the NMSE performance of conventional methods and as we can see, the simulations match well with our analysis of LS and MMSE estimators for decoupled UEs. Also, MMSE outperforms LS when $\mathrm{SNR}_{\mathrm{T}}$ is small and becomes almost identical as $\mathrm{SNR}_{\mathrm{T}}$ grows. Results in this figure further verify that the uplink performance for decoupled UEs associated with the SBSs (blue line) is much better (roughly 10dB) than those connecting to the MBS (red line). This explains not only why those decoupled UEs connect to the SBSs in uplink according to the decoupled association policy but also the major reason that channel information quality at the MBS should be improved for better DL performance.

Fig. 4 shows the average uplink data BER performance of decoupled UEs with growing training power (or uplink data SNR) by utilizing different data detectors. Two lines on the top of this figure indicate the BER performances of MRC and ZF. Results reveal that ZF enjoys better performance due to superior interference cancelation with CSI at the SBSs, but both detectors are not able to obtain as low BER as MMSE when training power grows. As can be seen, the BER performance of MMSE is better than 10^{-2} under severe co-channel interference and can improve with growing training power but tends to saturate

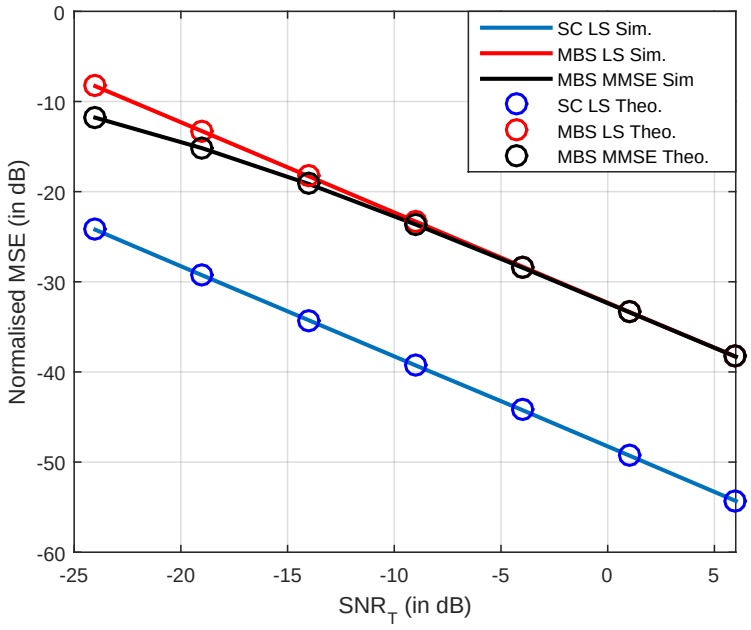

Fig. 3. NMSE versus $\mathrm{SNR_T}$ with $P_{\mathrm{D}} = 23$ dBm for LS and MMSE conventional channel estimation methods.

to the black line, which represents the BER for the MMSE detector under perfect CSI. In this figure, our theoretical analysis of BER is plotted in squares and is shown to be pretty close to the simulated ones, while there is a gap between BER simulations and its lower bound.

Fig. 5 demonstrates the relationship between BER and $\mathrm{SNR_D}$ with different data detectors. Results illustrate that all detectors have better performances as $\mathrm{SNR_D}$ (or data power) increases, but their performances tend to saturate when $\mathrm{SNR_D}$ keeps increasing. In addition, the MMSE detector performs best and MRC is the worst among the three estimators. The gap between perfect CSI and the estimated one is small because the training power is switched to the maximum. The theoretical BER performance with imperfect CSI matches with the simulated one and the lower bound is tight as well.

We next verify the NMSE performance of data-aided estimation and compare it with conventional channel estimation methods. Fig. 6 shows how $\mathrm{SNR_T}$ affects the NMSE of decoupled UEs. Note that the maximum training power in this figure is extended to 43 dBm in order to see the complete trend of NMSE. From the simulations, it is observed that data-aided channel estimation is always better than LS and MMSE estimators. In particular, when $\mathrm{SNR_T}$ is low, the data-aided method outperforms the conventional ones (over 30dB in this scenario) and the results for the data-aided method tend to approach to that for MMSE and LS estimators as $\mathrm{SNR_T}$ increases but will saturate eventually. This is because

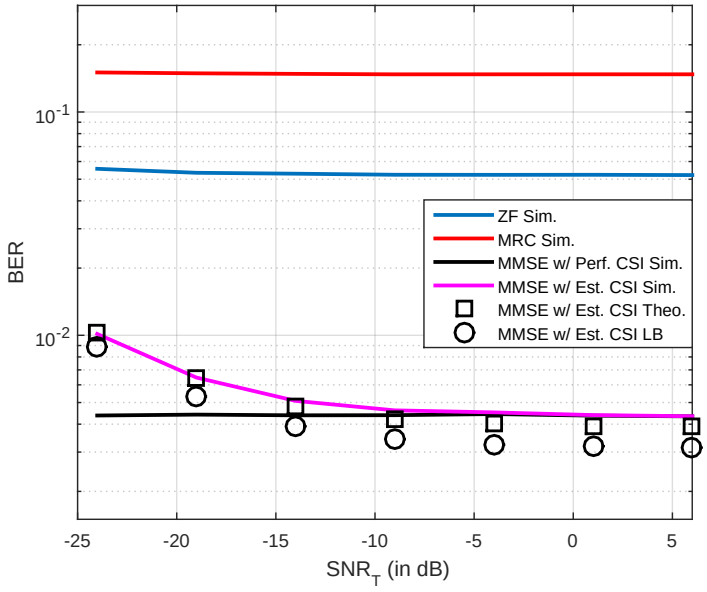

Fig. 4. Average uplink data BER of decoupled UEs versus $\mathrm{SNR_T}$ with $P_\mathrm{D} = 23$ dBm for different data estimators.

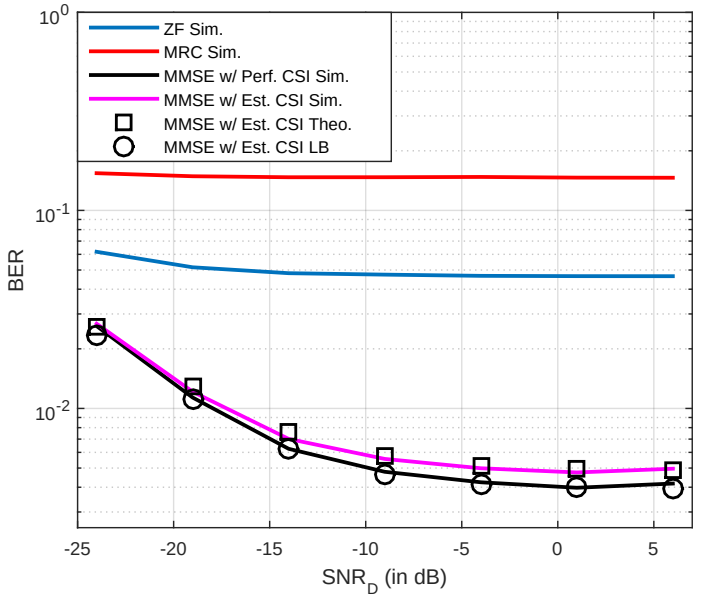

Fig. 5. Average uplink data BER of decoupled UEs versus $\mathrm{SNR_D}$ with $P_\mathrm{T} = 3$ dBm for different data estimators.

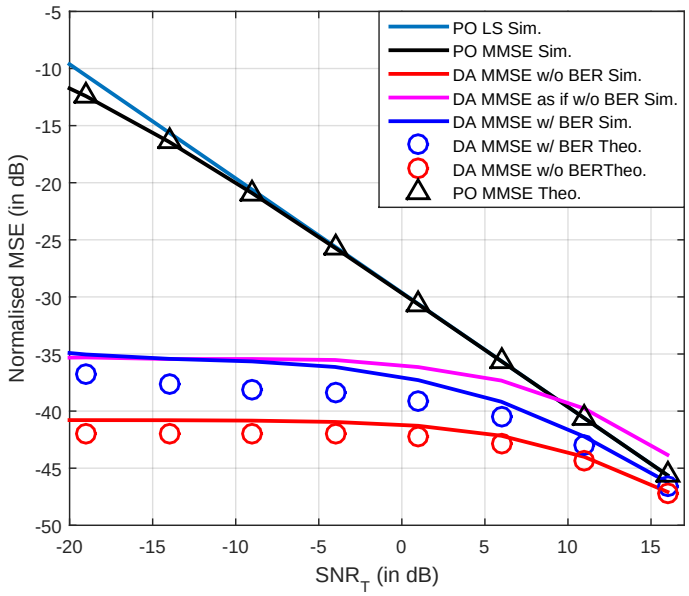

Fig. 6. Comparison of average NMSE of decoupled UEs for $\mathrm{SNR_T}$ with $P_\mathrm{D} = 23$ dBm.

the performance of MMSE or LS is improved so quickly when the training power goes high, benefitting from the aided data (not as remarkable for low training power scenarios). Nevertheless, in practice, the training power would not be raised that high, so the data-aided methods would be an effective means to ensure the quality of the estimated channel and may also reduce the training power to save energy and mitigate pilot contamination across cells. Lines in magenta represent the NMSE by treating the sequences without BER. As expected, the performance of this method is almost the same as that considering the BER effect when $\mathrm{SNR_T}$ is low and becomes worse, and will not saturate to the conventional MMSE.

The relationship between NMSE and $\mathrm{SNR_D}$ (also data power) is presented in Fig. 7. First, the gap between conventional methods and the data-aided estimator is getting larger by increasing data power for the reason that data power is not involved in the conventional methods. Hence, the results remain same when the power varies while for the data-aided method, higher data power would not only improve BER but also directly increase the SNR-like term, leading to better NMSE performance. Secondly, the NMSE of the data-aided method without BER can be improved nearly log-linearly as $\mathrm{SNR_D}$ grows while that with BER tends to decrease slower and reaches a performance limit as predicted previously.

Finally, we investigate the NMSE performance of our proposed method with different data length, $\tau_\mathrm{D} = 64, 128, 256, 512$. We observe that when the UL data length gets longer, the NMSE for conventional MMSE remains constant but the NMSE for the data-aided method enjoys nearly log-linear elevation.

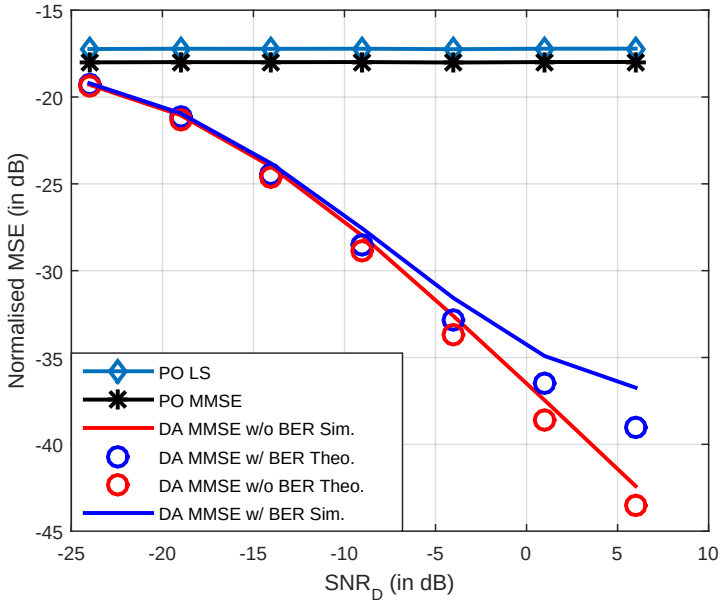

Fig. 7. Comparison of average NMSE of decoupled UEs for $\mathrm{SNR_D}$ with $P_{\mathrm{T}} = 3$ dBm.

Recall that we make the approximation that $\hat{\mathbf{X}}\hat{\mathbf{X}}^{\mathrm{H}} \to \tau_{\mathrm{D}} P_{\mathrm{D}} \mathbf{I}_k$ when τ_{D} is large. Fig. 8 indicates that for the data-aided method without BER, the gap between simulations and the approximated result is getting smaller when τ_{D} becomes larger, while for the case with BER, the approximation is getting worse. This is because there may be more erroneous bits in a long sequence. As a consequence, the elements of $\hat{\mathbf{X}}\mathbf{\Lambda}_{\mathbf{X}}\hat{\mathbf{X}}^{\mathrm{H}}$ except the diagonal ones, less likely tend to zeros.

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 from UEs to the MBS. The ergodic BER of UL data has been analyzed to model the decoded data and the approximated NMSE for the data-aided method with MMSE estimation 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 channel quality 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.

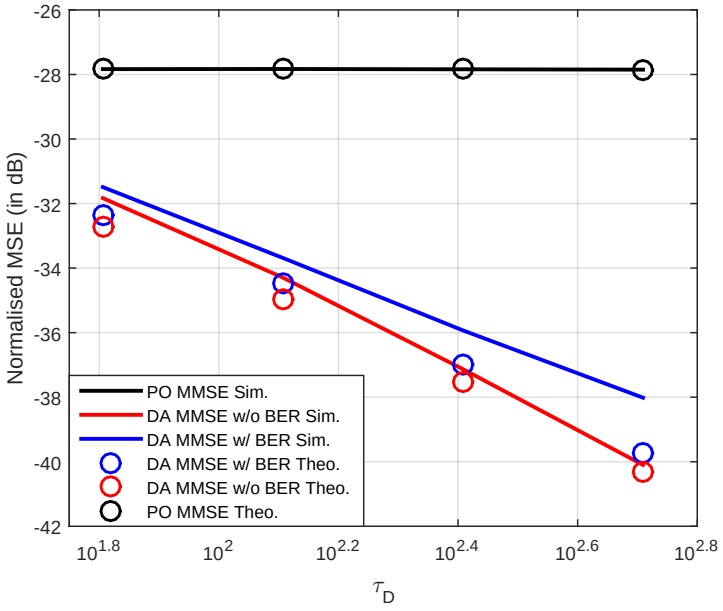

Fig. 8. Comparison of average NMSE of decoupled UEs for $\tau_{\mathrm{D}} = 64, 128, 256, 512$ with $P_{\mathrm{T}} = 13$ dBm and $P_{\mathrm{D}} = 13$ dBm.

APPENDIX A
PROOF OF LEMMA 1

Proof 4: From the expression of SINR_{vk} in (31), we have

$$\begin{aligned}
\mathbb{E}\left[\mathrm{SINR}_{vk}\right] &= \mathbb{E}\left[\rho_v \sum_{i=1}^{N} \frac{\|\varphi_i\|^2}{1+\rho_v d_i^2}\right] \\
&= \mathbb{E}\left[\rho_v \sum_{i=1}^{N} \lambda_i \mathbb{E}\left[\|\varphi_i\|^2\right]\right] \\
&= N\rho_v \hat{\beta}_{vk}^{\mathrm{S}} \mathbb{E}\left[\left(\frac{1}{N}\sum_{i=1}^{N}\lambda_i\right)\right] \\
&= N\rho_v \hat{\beta}_{vk}^{\mathrm{S}} \mathbb{E}\left[\frac{\mathrm{Tr}\left(\mathbf{\Lambda}\right)}{N}\right] \\
&\approx N\rho_v \hat{\beta}_{vk}^{\mathrm{S}} \mu,
\end{aligned} \tag{53}$$

where the above approximation uses the definition in (35).

Similarly, we obtain the second moment by

$$\begin{aligned}\text{Var}\left[\text{SINR}_{vk}\right] &= N^2\rho_v^2\text{Var}\left[\frac{1}{N}\sum_{i=1}^{N}\lambda_i\left\|\phi_i\right\|^2\right]\\ &= N^2\rho_v^2\text{Var}\left[\frac{1}{N}\sum_{i=1}^{N}\lambda_i\mathbb{E}\left[\left\|\phi_i\right\|^2|\hat{\mathbf{G}}_{v(-k)}\right]\right]+N^2\rho_v^2\mathbb{E}\left[\frac{1}{N}\sum_{i=1}^{N}\lambda_i^2\text{Var}\left[\left\|\phi_i\right\|^2|\hat{\mathbf{G}}_{v(-k)}\right]\right]\\ &\overset{(a)}{=} N^2\rho_v^2\left(\hat{\beta}_{vk}^{\text{S}}\right)^2\text{Var}\left[\frac{\text{Tr}\left(\mathbf{\Lambda}\right)}{N}\right]+N^2\rho_v^2\left(\hat{\beta}_{vk}^{\text{S}}\right)^2\mathbb{E}\left[\frac{\text{Tr}\left(\mathbf{\Lambda}^2\right)}{N}\right]\\ &\approx N^2\rho_v^2\left(\hat{\beta}_{vk}^{\text{S}}\right)^2\sigma^2,\end{aligned} \tag{54}$$

where the first term of (a) can be proved to converge to 0 by the results from [32].

APPENDIX B
PROOF OF PROPOSITION 2

Proof 5: First, denote the mean-square-error (MSE) as $\mathcal{J}\left(\mathbf{C}_k\right)=\mathbb{E}\left[\left\|\hat{\mathbf{h}}_k^{\text{MMSE}}-\mathbf{h}_k\right\|^2\right]$. Then by applying the MMSE rule directly, the problem is formulated as

$$\min_{\mathbf{C}_k}\mathcal{J}\left(\mathbf{C}_k\right)=\mathbb{E}\left[\left(\left(\mathbf{H}\left(\hat{\mathbf{W}}\circ\mathbf{E}\right)+\mathbf{Z}\right)\mathbf{C}_k^{\text{opt}}-\mathbf{h}_k\right)^{\text{H}}\left(\left(\mathbf{H}\left(\hat{\mathbf{W}}\circ\mathbf{E}\right)+\mathbf{Z}\right)\mathbf{C}_k^{\text{opt}}-\mathbf{h}_k\right)\right], \tag{55}$$

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

$$\left(\mathbb{E}\left[\left(\hat{\mathbf{W}}\circ\mathbf{E}\right)^{\text{H}}\mathbf{H}^{\text{H}}\mathbf{H}\left(\hat{\mathbf{W}}\circ\mathbf{E}\right)+\mathbf{Z}^{\text{H}}\mathbf{Z}\right]\right)\mathbf{C}_k=\mathbb{E}\left[\left(\hat{\mathbf{W}}\circ\mathbf{E}\right)^{\text{H}}\mathbf{H}^{\text{H}}\mathbf{h}_k\right]. \tag{56}$$

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

$$\mathbf{R}_{\mathbf{H}}\triangleq\frac{1}{M}\mathbb{E}\left[\mathbf{H}^{\text{H}}\mathbf{H}\right]=\text{diag}\left(\beta_1^{\text{M}},\ldots,\beta_k^{\text{M}},\ldots,\beta_K^{\text{M}}\right), \tag{57}$$

and

$$\left(\mathbb{E}\left[\left(\hat{\mathbf{W}}\circ\mathbf{E}\right)^{\text{H}}\mathbf{R}_{\mathbf{H}}\left(\hat{\mathbf{W}}\circ\mathbf{E}\right)\right]+N_0\mathbf{I}\right)\mathbf{C}_k=\left(\hat{\mathbf{w}}_k^{\text{H}}\circ\mathbb{E}\left[\mathbf{e}_k^{\text{H}}\right]\right)\beta_k^{\text{M}}. \tag{58}$$

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

$$\mathbf{P}\triangleq\mathbb{E}\left[\left(\left[\mathbf{S}\circ\mathbf{E}_1,\hat{\mathbf{X}}\circ\mathbf{E}_2\right]\right)^{\text{H}}\mathbf{R}_{\mathbf{H}}\left(\left[\mathbf{S}\circ\mathbf{E}_1,\hat{\mathbf{X}}\circ\mathbf{E}_2\right]\right)\right]=\begin{pmatrix}\mathbf{P}_{11}&\mathbf{P}_{12}\\ \mathbf{P}_{21}&\mathbf{P}_{22}\end{pmatrix}, \tag{59}$$

with

$$\mathbf{P}_{11} = \mathbf{S}^{\mathrm{H}}\mathbf{R}_{\mathbf{H}}\mathbf{S}, \tag{60}$$

$$\mathbf{P}_{12} = \mathbf{S}^{\mathrm{H}}\mathbf{R}_{\mathbf{H}}\left(\hat{\mathbf{X}} \circ \mathbb{E}\left[\mathbf{E}_2\right]\right), \tag{61}$$

$$\mathbf{P}_{21} = \left(\hat{\mathbf{X}} \circ \mathbb{E}\left[\mathbf{E}_2\right]\right)^{\mathrm{H}}\mathbf{R}_{\mathbf{H}}\mathbf{S}, \tag{62}$$

$$\mathbf{P}_{22} = \mathbb{E}\left[\left(\hat{\mathbf{X}} \circ \mathbf{E}_2\right)^{\mathrm{H}}\mathbf{R}_{\mathbf{H}}\left(\hat{\mathbf{X}} \circ \mathbf{E}_2\right)\right]. \tag{63}$$

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

$$\begin{aligned}
\left[\mathbf{P}_{22}\right]_{ij,i\neq j} &= \sum_{k}^{K}\sum_{l}^{K}\left[\mathbf{R}_{\mathbf{H}}\right]_{kl}\left(\hat{x}_{ki}\right)'\hat{x}_{lj}\mathbb{E}\left[\left(\left[\mathbf{E}_2\right]_{ki}\right)'\left[\mathbf{E}_2\right]_{lj}\right] \\
&= \sum_{k}^{K}\beta_k^{\mathrm{M}}\left(\hat{x}_{ki}\right)'\hat{x}_{kj}\left(1-2\mathrm{BER}_{vk}\right)^2 \qquad (64) \\
\left[\mathbf{P}_{22}\right]_{ij,i=j} &= \sum_{k}^{K}\beta_k^{\mathrm{M}}\left(\hat{x}_{ki}\right)'\hat{x}_{ki}\mathbb{E}\left[\left\|\left[\mathbf{E}_2\right]_{ki}\right\|^2\right] \\
&= P_{\mathrm{D}}\sum_{k}^{K}\beta_k^{\mathrm{M}}. \qquad (65)
\end{aligned}$$

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

$$\left(\hat{\mathbf{X}} \circ \mathbb{E}\left[\mathbf{E}_2\right]\right)^{\mathrm{H}}\mathbf{R}_{\mathbf{H}}\left(\hat{\mathbf{X}} \circ \mathbb{E}\left[\mathbf{E}_2\right]\right) + \Delta\mathbf{P}_{\mathbf{X}}, \tag{66}$$

where $\Delta\mathbf{P}_{\mathbf{X}} = \Delta S_{\mathbf{X}}\mathbf{I}_{\tau_{\mathrm{D}}}$ and $\Delta S_{\mathbf{X}} = P_{\mathrm{D}}\sum\limits_{k=1}^{K}\beta_k^{\mathrm{M}} - P_{\mathrm{D}}\sum\limits_{k=1}^{K}\beta_k^{\mathrm{M}}\left(1-2\mathrm{BER}_{vk}\right)^2$. 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

$$\hat{\mathbf{P}} = \left(\left[\mathbf{S}, \hat{\mathbf{X}} \circ \mathbb{E}\left[\mathbf{E}_2\right]\right]\right)^{\mathrm{H}}\mathbf{R}_{\mathbf{H}}\left(\left[\mathbf{S}, \hat{\mathbf{X}} \circ \mathbb{E}\left[\mathbf{E}_2\right]\right]\right) \tag{67}$$

and

$$\Delta\mathbf{P} = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \Delta\mathbf{P}_{\mathbf{X}} \end{pmatrix}. \tag{68}$$

By plugging these results into (58), 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

$$\begin{aligned}
&\mathbb{E}\left[\left\|\hat{\mathbf{h}}_k^{\text{MMSE}}-\mathbf{h}_k\right\|^2\right]\\
=&\mathbb{E}\left[\mathbf{C}_k^{\text{H}}\left(\mathbf{H}\left(\hat{\mathbf{W}}\circ\mathbf{E}\right)+\mathbf{Z}\right)^{\text{H}}\left(\mathbf{H}\left(\hat{\mathbf{W}}\circ\mathbf{E}\right)+\mathbf{Z}\right)\mathbf{C}_k+\mathbf{h}_k^{\text{H}}\mathbf{h}_k-\mathbf{C}_k^{\text{H}}\left(\hat{\mathbf{W}}\circ\mathbf{E}\right)^{\text{H}}\mathbf{H}^{\text{H}}\mathbf{h}_k-\mathbf{h}_k^{\text{H}}\mathbf{H}\left(\hat{\mathbf{W}}\circ\mathbf{E}\right)\mathbf{C}_k\right]\\
=&\mathbb{E}\left[M\left(\beta_k^{\text{M}}\right)^2\left(\hat{\mathbf{w}}_k\circ\mathbb{E}\left[\mathbf{e}_k\right]\right)\left(\mathbf{P}+N_0\mathbf{I}\right)^{-1}\left(\hat{\mathbf{w}}_k^{\text{H}}\circ\mathbb{E}\left[\mathbf{e}_k^{\text{H}}\right]\right)+M\beta_k^{\text{M}}-2M\left(\beta_k^{\text{M}}\right)^2\left(\hat{\mathbf{w}}_k\circ\mathbb{E}\left[\mathbf{e}_k\right]\right)\left(\mathbf{P}+N_0\mathbf{I}\right)^{-1}\left(\hat{\mathbf{w}}_k^{\text{H}}\circ\mathbb{E}\left[\mathbf{e}_k^{\text{H}}\right]\right.\right.\\
=&M\beta_k^{\text{M}}-M\left(\beta_k^{\text{M}}\right)^2\mathbb{E}\left[\left(\hat{\mathbf{w}}_k\circ\mathbb{E}\left[\mathbf{e}_k\right]\right)\left(\mathbf{P}+N_0\mathbf{I}\right)^{-1}\left(\hat{\mathbf{w}}_k^{\text{H}}\circ\mathbb{E}\left[\mathbf{e}_k^{\text{H}}\right]\right)\right]
\end{aligned}\tag{69}$$

Now, we denote

$$\mathbf{P}_{\text{Aux}}\triangleq\Delta\mathbf{P}+N_0\mathbf{I}=\begin{pmatrix}N_0\mathbf{I}_{\tau_{\text{T}}} & \mathbf{0}\\ \mathbf{0} & \left(\Delta S_{\mathbf{X}}+N_0\right)\mathbf{I}_{\tau_{\text{D}}}\end{pmatrix}\tag{70}$$

and

$$\bar{\mathbf{W}}=\hat{\mathbf{W}}\circ\mathbb{E}\left[\mathbf{E}\right].\tag{71}$$

Thus, we can derive the expectation part in the last equation of (69) as

$$\begin{aligned}
&\mathbb{E}\left[\left(\hat{\mathbf{w}}_k\circ\mathbb{E}\left[\mathbf{e}_k\right]\right)\left(\mathbf{P}+N_0\mathbf{I}\right)^{-1}\left(\hat{\mathbf{w}}_k^{\text{H}}\circ\mathbb{E}\left[\mathbf{e}_k^{\text{H}}\right]\right)\right]\\
=&\mathbb{E}\left[\left[\bar{\mathbf{W}}\left(\bar{\mathbf{W}}^{\text{H}}\mathbf{R}_{\mathbf{H}}\bar{\mathbf{W}}+\mathbf{P}_{\text{Aux}}\right)^{-1}\bar{\mathbf{W}}^{\text{H}}\right]_{kk}\right]\\
\overset{(a)}{=}&\mathbb{E}\left[\left[\bar{\mathbf{W}}\mathbf{P}_{\text{Aux}}^{-1}\bar{\mathbf{W}}^{\text{H}}-\bar{\mathbf{W}}\mathbf{P}_{\text{Aux}}^{-1}\bar{\mathbf{W}}^{\text{H}}\left(\mathbf{R}_{\mathbf{H}}^{-1}+\bar{\mathbf{W}}\mathbf{P}_{\text{Aux}}^{-1}\bar{\mathbf{W}}^{\text{H}}\right)^{-1}\bar{\mathbf{W}}\mathbf{P}_{\text{Aux}}^{-1}\bar{\mathbf{W}}^{\text{H}}\right]_{kk}\right]\\
\overset{(b)}{=}&\left[\mathbf{R}_{\mathbf{H}}^{-1}-\mathbf{R}_{\mathbf{H}}^{-1}\mathbb{E}\left[\left(\bar{\mathbf{W}}\mathbf{P}_{\text{Aux}}^{-1}\bar{\mathbf{W}}^{\text{H}}+\mathbf{R}_{\mathbf{H}}^{-1}\right)^{-1}\right]\mathbf{R}_{\mathbf{H}}^{-1}\right]_{kk},
\end{aligned}\tag{72}$$

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

$$\mathbf{A}-\mathbf{A}\left(\mathbf{A}+\mathbf{B}\right)^{-1}\mathbf{A}=\mathbf{B}-\mathbf{B}\left(\mathbf{A}+\mathbf{B}\right)^{-1}\mathbf{B}.\tag{73}$$

With the definitions of

$$\tilde{\mathbf{E}}_2=\text{diag}\left(\left(1-2\text{BER}_{v1}\right),\ldots,\left(1-2\text{BER}_{vK}\right)\right)\tag{74}$$

and

$$\mathbf{\Lambda}_{\mathbf{X}}=\tilde{\mathbf{E}}_2\Delta\mathbf{P}_{\mathbf{X}}^{-1}\tilde{\mathbf{E}}_2^{\text{H}}=\frac{1}{\Delta S_{\mathbf{X}}+N_0}\text{diag}\left(\left(1-2\text{BER}_{v1}\right)^2,\ldots,\left(1-2\text{BER}_{vK}\right)^2\right),\tag{75}$$

the expectation part in (72) can be simplified as

$$\begin{aligned}
&\mathbb{E}\left[\left(\bar{\mathbf{W}}\mathbf{P}_{\text{Aux}}^{-1}\bar{\mathbf{W}}^{\text{H}}+\mathbf{R}_{\bar{\mathbf{H}}}^{-1}\right)^{-1}\right]\\
=&\mathbb{E}\left[\left(\left[\mathbf{S},\hat{\mathbf{X}}\circ\mathbb{E}\left[\mathbf{E}_2\right]\right]\Delta\mathbf{P}_{\mathbf{X}}^{-1}\begin{bmatrix}\mathbf{S}^{\text{H}}\\ \left(\hat{\mathbf{X}}\circ\mathbb{E}\left[\mathbf{E}_2\right]\right)^{\text{H}}\end{bmatrix}+\mathbf{R}_{\mathbf{H}}^{-1}\right)^{-1}\right]\\
=&\mathbb{E}\left[\left(\frac{\tau_{\text{T}}P_{\text{T}}}{N_0}\mathbf{I}_k+\left(\hat{\mathbf{X}}\tilde{\mathbf{E}}_2\right)\begin{pmatrix}\frac{1}{N_0}\mathbf{I}_{\tau_{\text{T}}} & \mathbf{0}\\ \mathbf{0} & \Delta\mathbf{P}_{\mathbf{X}}^{-1}\end{pmatrix}\left(\hat{\mathbf{X}}\tilde{\mathbf{E}}_2\right)^{\text{H}}+\mathbf{R}_{\mathbf{H}}^{-1}\right)^{-1}\right]\\
=&\mathbb{E}\left[\left(\frac{\tau_{\text{T}}P_{\text{T}}}{N_0}\mathbf{I}_k+\hat{\mathbf{X}}\mathbf{\Lambda}_{\mathbf{X}}\hat{\mathbf{X}}^{\text{H}}+\mathbf{R}_{\mathbf{H}}^{-1}\right)^{-1}\right]\\
\approx&\left(\frac{\tau_{\text{T}}P_{\text{T}}}{N_0}\mathbf{I}_k+\frac{\tau_{\text{D}}P_{\text{D}}}{\Delta S_{\mathbf{X}}+N_0}\text{diag}\left(\left(1-2\text{BER}_{v1}\right)^2,\ldots,\left(1-2\text{BER}_{vK}\right)^2\right)+\mathbf{R}_{\mathbf{H}}^{-1}\right)^{-1}\\
=&\text{diag}\left(\frac{\beta_1^{\text{M}}}{1+\rho_1^{\text{DA}}\beta_1^{\text{M}}},\ldots,\frac{\beta_K^{\text{M}}}{1+\rho_K^{\text{DA}}\beta_K^{\text{M}}}\right),
\end{aligned}\tag{76}$$

where

$$\rho_k^{\text{DA}}=\frac{\tau_{\text{T}}P_{\text{T}}}{N_0}+\frac{\tau_{\text{D}}P_{\text{D}}\left(1-2\text{BER}_{vk}\right)^2}{\Delta S_{\mathbf{X}}+N_0},\tag{77}$$

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

$$\mathbb{E}\left[\left\|\hat{\mathbf{h}}_k^{\text{MMSE}}-\mathbf{h}_k\right\|^2\right]=\frac{M\beta_k^{\text{M}}}{1+\rho_k^{\text{DA}}\beta_k^{\text{M}}}.\tag{78}$$

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

$$\mathcal{J}_k^{\text{MMSE}}=10\log_{10}\left(\frac{\mathbb{E}\left[\left\|\hat{\mathbf{h}}_k^{\text{MMSE}}-\mathbf{h}_k\right\|^2\right]}{\mathbb{E}\left[\left\|\mathbf{h}_k\right\|^2\right]}\right)=10\log_{10}\left(\frac{1}{1+\rho_k^{\text{DA}}\beta_k^{\text{M}}}\right),\tag{79}$$

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. Smiljkovikj, 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. Smiljkovikj, 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 L. Ping, "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 L. Ping, "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. Hélard, "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] 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.

[32] 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.

[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.