On Hybrid Pilot for Channel Estimation in Massive MIMO Uplink

Jiaming Li, *Student Member IEEE*, Chau Yuen, *Senior Member IEEE*,

Dong Li, *Member IEEE*, Xianda Wu, *Student Member IEEE*,

and Han Zhang, *Member IEEE*

arXiv:1608.02112v2 [cs.IT] 22 Dec 2016

Abstract

This paper introduces a hybrid pilot-aided channel estimation technique for mitigating the effect of pilot contamination for the uplink of multi-cell multiuser massive MIMO systems. The proposed hybrid pilot is designed such that it enjoys the complementary advantages between time-multiplexed (TM) pilot and time-superimposed (TS) pilot, and thereby, allows superior solution to the conventional pilot schemes. We mathematically characterize the impact of hybrid pilot on the massive MIMO uplink by deriving a closed-form approximation for the uplink achievable rate. In large-number-of-antennas regime, we obtain the asymptotically optimal solution for hybrid pilot by jointly designing the TM pilot and the TS pilot. It is shown that either TM pilot or TS pilot has the advantages for large frame-size and limited frame-size transmission, respectively, while the hybrid pilot scheme can offer a superior performance to that employing either TM pilot or TS pilot. Numerical results demonstrate the effectiveness of the proposed design.

Index Terms

Channel estimation, hybrid pilot design, multi-cell multiuser massive MIMO system, pilot contamination, uplink achievable rate.

H. Zhang and J. M. Li are with the School of Physics & Telecommunication Engineering, South China Normal University, Guangzhou 510026, China. (e-mail: zhanghan@scnu.edu.cn).

C. Yuen is with Singapore University of Technology and Design, Singapore. (e-mail: yuenchau@sutd.edu.sg).

D. Li is with the Faculty of Information Technology, Macau University of Science and Technology, Taipa, Macau, China. (e-mail: dli@must.edu.mo).

X. D. Wu is with the Department of Electrical and Computer Engineering, University of Macau, Macau, China. (e-mail: xianda.wu@connect.umac.mo).

H. Zhang is the corresponding author.

I. INTRODUCTION

Massive multiple-input multiple-out (MIMO) technique at the base station (BS), was firstly proposed in [1], and now has attracted tremendous interest in both academia and industry. Massive MIMO (also, known as large scale MIMO) has been widely recognized as a potential candidate for the key technologies of the future wireless communication systems [2]–[4].

Compared with the conventional MIMO technique, massive MIMO with time-division duplex (TDD) exhibits several remarkable features. First, by taking advantage of channel reciprocity, additional antennas significantly increase the spectral efficiency through spatial multiplexing [5], [6]. Second, large antenna arrays enables energy efficiency in both uplink (UL) and downlink (DL) transmission through coherent combining, and hence, provide an potential for cell-size shrinking [7]. Third, when the number of BS antennas M is sufficiently large, the simplest coherent combiner and linear precoder, e.g. the matched filter (MF), turn out to be optimal [8], [9]. Although promising, the ultimate performance of TDD massive MIMO is limited by the effect of pilot contamination, an unavoidable interference caused by the reuse of pilots (or nonorthogonality of pilots) among several adjacent cells, even for the asymptotic case $M \to \infty$.

In an effort to solve the problem of pilot contamination when performing UL channel estimation, several sophisticated pilot-aided schemes have been proposed. Typically, pilots are time-multiplexed with the data during the training phase, and henceforth are referred to as time-multiplexed (TM) pilots. Relying on the coordination between neighboring cells, the second-order statistical information about the user channels of neighboring cells is involved for channel estimation [10]. Based on the singular value decomposition (SVD), blind channel estimation scheme is proposed in [11], [12], which is shown to be effective in mitigating pilot contamination. For a fixed size of training, a pilot-reuse scheme is provided in [13], aiming to maximize the UL achievable rate. In [14], a data-aided scheme is presented by employing the decision feedback information of data symbols to aid the channel estimation. In [15]–[17], the optimal designs for TM pilots by maximizing the sum spectral efficiency are proposed and discussed, and the authors in [18] propose using downlink training with pilot contamination precoding to eliminate the effect of contamination. All these studies employing TM pilots lead to a similar conclusion that the data rate will decrease with increasing pilot-size, making system throughput limited, especially for the mobility case, where the channel coherent time is limited.

As an alternative to TM pilots, time-superimposed (TS) pilots have been studied in the context of channel acquisition in massive MIMO systems [19], [20]. In comparison with TM pilots [10]–[18], TS pilots require no additional time resource reserved for pilots, and thereby, can achieve a higher spectral efficiency [21]. More recently, the analysis of TS pilots in massive MIMO systems [22] illustrates its superiority for mitigating pilot contamination. However, the mixed type of pilots suffers from co-interference from data symbols, which generally limits its performance, especially in low signal-to-noise ratio (SNR) scenarios [23], [24].

In this paper, we take a further step than the previous literatures [19]–[24], and propose a new pilot-based scheme as an alternative to the conventional pilot-aided ones for mitigating pilot contamination in massive MIMO systems. To be specific, the pilots for the UL channel estimation comprise both TM pilots and TS pilots, henceforth can be referred to as hybrid pilots. The motivation behind the proposed design is twofold.

- TM pilots with the aid of TS pilots can improve the estimation quality, while preserving transmission efficiency [19].
- TS pilots benefit from TM pilots by reducing the correlation between pilots and data [19], [22], and hence can provide substantial improvement of system performance.

Intuitively, hybrid pilot enjoys the advantages of both TM pilot and TS pilot, and thereby, is more flexible and robust to different transmission of practical relevance. The hybrid design of pilots, to the best of authors' knowledge, has not been addressed for multi-cell MIMO systems. To evaluate the proposed design, we mathematically characterize the impact of hybrid pilot on the performance of massive MIMO uplink, and demonstrate its effectiveness by deriving a closed-form approximation on signal-to-interference-plus-noise ratio (SINR) as well as cell throughput. In large-number-of-BS-antennas regime, we obtain the asymptotically optimal solutions for the hybrid design of pilots. Our result demonstrates that the time allocation between TM pilots and TS pilots, as well as power ratio between pilots and data, determine the UL spectral efficiency. Qualitative analysis and simulations show that the conventional TM pilot or TS pilot is effective for either large-frame-size or limited frame-size transmission, while the proposed hybrid pilot design can offer a superior solution than that employing conventional pilots [16], [22].

The rest of this paper is organized as follows: In Section II, we firstly describe the uplink multi-cell massive MIMO system model. In Section III, we introduce the hybrid pilots-based channel estimation technique, and then provide the analytical results for the UL achievable rate.

In Section IV, we provide an iterative data-aided solution to improve the system performance. Along with the theoretical analysis, the asymptotically optimal solutions are given in Section V, which can explain the trends observed in Section VI simulations results. Finally, Section VII summarizes the main results and insights obtained in the paper.

Notations: Boldface lower and upper case symbols represent vectors and matrices, respectively. The transpose, complex conjugate, and Hermitian transpose operations are denoted by $()^T$, $()^*$, and $()^H$, respectively. $\|\cdot\|$ denotes the Euclidian norm and $E[\cdot]$ is the statistical expectation. We use $\mathcal{CN}(a, b)$ to denote the circular symmetric complex Gaussian distribution with mean a and covariance b. $\mathcal{O}(\cdot)$ denotes the big-O notation. $\rightarrow$ denotes the convergence as $M \rightarrow \infty$.

II. SYSTEM MODEL AND PROBLEM FORMULATION

A. Multi-cell Massive MIMO Uplink

Consider a cellular network composed of L hexagonal cells, each consisting a central M-antenna BS and K ($K \leq M$) single-antenna user terminals (UTs) that share the same bandwidth. We focus on the uplink transmission without any kind of BS cooperation. The propagation channel coefficient between mth BS antenna of the jth cell and the k-th UT of the target cell, i.e., cell 1, is $h_{j,k,m} = \sqrt{\beta_{j,k}}g_{j,k,m}$, where $\{\beta_{j,k}\}$ and $\{g_{j,k,m}\}$ are large scale fading and small scale fading, respectively. Specifically, $\{\beta_{j,k}\}$ model path-loss and shadowing that change slowly and thus can be assumed to be known at receiver, while $\{g_{j,k,m}\} \sim \mathcal{CN}(0, 1)$ are identically independent distributed (i.i.d.) unknown random variables. Moreover, $\{h_{j,k,m}\}$ are assumed to be constant for the duration of T symbols in time, where T is the channel coherence time that limited by the mobility of users.

B. Effect of Pilot Contamination

Denote $\mathbf{u}_1(t) = [u_{1,1}(t), \cdots, u_{1,M}(t)]^T$ as the received signal vector at over M antennas at the BS of target cell, i.e. cell 1 at time instant t

$$\mathbf{u}_1(t) = \sum_{j=1}^{L}\sum_{k=1}^{K}\mathbf{h}_{j,k}\widetilde{x}_{j,k}(t) + \mathbf{n}_1(t) \tag{1}$$

where $\mathbf{h}_{j,k} = [h_{j,k,1}, \cdots, h_{j,k,M}]^T$, $\mathbf{n}_1(t) = [n_{1,1}(t), \cdots, n_{1,M}(t)]^T$ with $n_{1,m}(t) \sim \mathcal{CN}(0, \sigma_n^2)$ being the additive white Gaussian noise (AWGN), and $\widetilde{x}_{j,k}(t)$ denotes the transmitted signal from k-th user at j cell with unit power at time t.

In each frame of transmission, τ TM pilots are employed as the training overhead, given in the form $\mathbf{x}_{j,k} = [x_{j,k}(1), \cdots, x_{j,k}(\tau)] \in \mathbb{C}^\tau$. Then, we rewrite (1) as a matrix form $\mathbf{U}_1 = [\mathbf{u}_1(1), \cdots, \mathbf{u}_1(\tau)] \in \mathbb{C}^{M \times \tau}$, which is given by

$$\mathbf{U}_1 = \sum_{k=1}^{K} \mathbf{h}_{1,k} \mathbf{x}_{1,k} + \sum_{j \neq 1}^{L} \sum_{k=1}^{K} \mathbf{h}_{j,k} \mathbf{x}_{j,k} + \mathbf{N}_1. \tag{2}$$

The least-squares (LS) estimate on channel vector $\mathbf{h}_{1,k}$ of the k-th UT in the target cell can be obtained as [25]

$$\hat{\mathbf{h}}_{1,k} = \mathbf{h}_{1,k} + \frac{1}{\tau} \sum_{j \neq 1}^{L} \sum_{k'=1}^{K} \mathbf{h}_{j,k} \left(\mathbf{x}_{j,k} \mathbf{x}_{1,k}^H \right) + \frac{1}{\tau} \mathbf{N}_1 \mathbf{x}_{1,k}^H. \tag{3}$$

The above formula implies that the estimation on $\mathbf{h}_{1,k}$ are contaminated by the channel vectors of other cells, unless each user to be assigned a unique orthogonal pilot, i.e., $\frac{1}{\tau} \mathbf{x}_{j,k} \mathbf{x}_{1,k}^H = 0$ if $j \neq 1$. In practical TDD mode, T is limited by the mobility of users, therefore it is hard to ensure the orthogonality of pilot sequences in the multi-cell scenario as the number of overall users becomes large. Although the pilot-based schemes in [13], [26], [27] are proposed to improve the estimation quality in (3), the correlated pilot sequences in different cells, known as pilot contamination, causes capacity-limiting inter-cell interference even when $M \to \infty$.

III. HYBRID PILOT-AIDED UL CHANNEL ESTIMATION

In this section, we study a hybrid pilot-aided channel estimation scheme, where both TM pilots and TS pilots are jointly employed for channel estimation.

A. Hybrid Pilot Framework

Without loss of generality, we consider a frame-based transmission, where each frame comprises a training overhead of τ pilots and $T - \tau$ data symbols. The framework of the proposed hybrid pilots is shown in Fig. 1, where the training overhead is composed of $(1-\alpha)\tau$ TM pilots, followed by $\alpha\tau$ TS pilots. $\alpha \in [0, 1]$ denote the time fraction allocated between TM pilots and TS pilots. Note that $\alpha \to 0$ and $\alpha \to 1$ denotes that either TM pilots or TS pilots are deployed in training overhead. Therefore, the conventional methods employing only TM pilots or TS pilots is a special case of the proposed scheme.

Denote $s_{j,k}(t)$, $p_{j,k}(t)$ are data and pilot symbols of the specific k-th user in cell $j, j = 1, 2, \cdots, L$ at time t, respectively, the transmitted signal within the interval of training overhead has the form
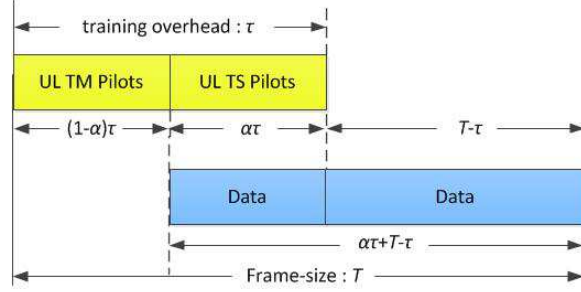
Fig. 1. UL frame structure of a hybrid pilot-aided system.

$$\mathbf{x}_{j,k} = \mathbf{s}_{j,k} + \mathbf{p}_{j,k}, \tag{4}$$

where $\mathbf{s}_{j,k}$ and $\mathbf{p}_{j,k}$ are data and hybrid pilot vectors, respectively, given by

$$\mathbf{s}_{j,k} = [0, \cdots, 0, s_{j,k}((1-\alpha)\tau + 1), \cdots, s_{j,k}(\tau)]^T, \tag{5}$$

$$\mathbf{p}_{j,k} = [\underbrace{p_{j,k}(1), \cdots, p_{j,k}((1-\alpha)\tau)}_{\text{TM Pilots}}, \underbrace{p_{j,k}((1-\alpha)\tau + 1), \cdots, p_{j,k}(\tau)}_{\text{TS Pilots}}]^T. \tag{6}$$

Similar to [23] and [24], we assume that $\{s_{j,k}(t)\}, \forall j$ contain independent identically distributed (i.i.d.) samples, and are mutually independent to $\{p_{j,k}(t)\}$. The power of data sequence and pilot sequence, respectively, are given by

$$E\{|s_{j,k}(t)|^2\} = 1 - \lambda, \tag{7}$$

$$E\{|p_{j,k}(t)|^2\} = \lambda, \tag{8}$$

where $\lambda \in (0, 1)$ is the power-allocation factor between pilots and data.

B. Hybrid Pilot-aided Channel Estimation

From (4), the received signal matrix at the BS in cell 1, denoted by, $\mathbf{Y}_1 \in \mathcal{C}^{M \times \tau}$, has the form

$$\mathbf{Y}_1 = \sum_{k=1}^{K} \mathbf{h}_{1,k}\Big(\mathbf{p}_{1,k} + \mathbf{s}_{1,k}\Big) + \sum_{j \neq 1}^{L} \sum_{k=1}^{K} \mathbf{h}_{j,k}\Big(\mathbf{p}_{j,k} + \mathbf{s}_{j,k}\Big) + \mathbf{N}_1. \tag{9}$$

Treating data as interference, the channel estimates on $\hat{\mathbf{h}}_1$ of the k-th UT in the target cell can be obtained using LS criteria

$$\hat{\mathbf{h}}_{1,k} = \arg \min_{\mathbf{h}} \left\| \mathbf{Y}_1 - \mathbf{h}_{1,k}\mathbf{p}_{1,k} \right\|_F^2. \tag{10}$$

Since the hybrid pilots comprise TS pilots superimposed onto the data symbols, the size of training overhead can be much longer than that employing only TM pilots. Therefore, provided

that $\tau \geq KL^1$, each user can be assigned a unique orthogonal pilot to avoid pilot contamination while preserving transmission efficiency. Accordingly, the estimation on $\mathbf{h}_{1,k}$ is given by

$$\hat{\mathbf{h}}_{1,k} = \mathbf{h}_{1,k} + \underbrace{\frac{1}{\|\mathbf{p}_{1,k}\|^2} \sum_{k'=1}^{K} \mathbf{h}_{1,k'} \left(\mathbf{s}_{1,k'} \mathbf{p}_{1,k}^H\right) + \frac{1}{\|\mathbf{p}_{1,k}\|^2} \sum_{j=1}^{L} \sum_{k'=1}^{K} \mathbf{h}_{j,k'} \left(\mathbf{s}_{j,k'} \mathbf{p}_{1,k}^H\right) + \frac{1}{\|\mathbf{p}_{1,k}\|^2} \mathbf{N}_1 \mathbf{p}_{1,k}^H}_{\Delta\mathbf{h}_{1,k}}. \tag{11}$$

In the above, $\Delta\mathbf{h}_{1,k}$ is the interference to channel estimation. To measure the estimation quality in (11), we derive the normalized channel mean square error (MSE) on $\hat{\mathbf{h}}_{1,k}$ as

$$\sigma_{\Delta\mathbf{h}_{1,k}}^2 = \frac{E\left[\|\Delta\mathbf{h}_{1,k}\|^2\right]}{E\left[\|\mathbf{h}_{1,k}\|^2\right]} = \frac{1}{\tau} \left(\frac{\alpha(1-\lambda)}{\lambda} \frac{\sum_{j=1}^{L} \sum_{k=1}^{K} \beta_{j,k}}{\beta_{1,k}} + \frac{\sigma_n^2}{\lambda\beta_{1,k}}\right). \tag{12}$$

As shown in (12), although orthogonal pilots have been assigned to users to eliminate pilot contamination, the performance of channel estimation suffers from data interference, and thereby, is inversely proportional to the time ratio and power of data $\alpha(1-\lambda)$. Besides, $\sigma_{\Delta\mathbf{h}_{1,k}}^2$ reduces linearly with τ and λ. This is expected, since τ independent pilots are involved for channel estimation. Later we will see that the hybrid pilot-based scheme with optimized α, λ and τ yields a substantial improvement in the average achievable rate.

C. Analysis of Achievable UL Rate

From (4), when $\alpha \neq 0$, the interval of training overhead also contains part of data symbols $s_{j,k}(t), t = (1-\alpha)\tau + 1, \cdots, \tau$. Thus, we perform data detection at k-th UT in target cell in two phases, i.e., Phase 1): data phase mixed with TS pilots (of a size $\alpha\tau$), and Phase 2): pure data phase of $T - \tau$ symbols.

Recall (1), the received signal of the above two phases at time t, denoted by $\mathbf{y}_1^I(t), t = (1-\alpha)\tau + 1, \cdots, \tau$ and $\mathbf{y}_1^{II}(t), t = \tau + 1, \cdots, T$, respectively, can be written in signal-plus-

---

[1]For high mobility case of $T \leq KL$, one can resort to channel modeling (i.e., basis expansion model) to reduce the channel unknowns, and then employ the two-step procedure to obtain channel estimates over multiple UL frames [23]. The details for the analysis of $T \leq KL$ is omitted herein due to space constraint.

interference forms,

$$\mathbf{y}_1^{\mathrm{I}}(t) = \mathbf{y}_1(t) - \hat{\mathbf{h}}_{1,k}^H p_{1,k}(t)$$

$$= \hat{\mathbf{h}}_{1,k} s_{1,k}(t) + \Delta\hat{\mathbf{h}}_{1,k} x_{1,k}(t) + \sum_{k'\neq k}^{K} \mathbf{h}_{1,k'} x_{1,k'}(t) + \sum_{j\neq 1}^{L}\sum_{k'=1}^{K} \mathbf{h}_{j,k'} x_{j,k'}(t) + \mathbf{n}_1(t) \qquad (13a)$$

$$\mathbf{y}_1^{\mathrm{II}}(t) = \hat{\mathbf{h}}_{1,k} s_{1,k}(t) + \Delta\hat{\mathbf{h}}_{1,k} s_{1,k}(t) + \sum_{k'=1}^{K} \mathbf{h}_{1,k'} s_{1,k'}(t) + \sum_{j\neq 1}^{L}\sum_{k'=1}^{K} \mathbf{h}_{j,k'} s_{j,k'}(t) + \mathbf{n}_1(t). \qquad (13b)$$

To maintain low receiver complexity, we employ a simple MF detector. The detected output are respectively given by

$$\hat{\mathbf{h}}_{1,k}^H \mathbf{y}_1^{\mathrm{I}}(t) = \underbrace{\|\hat{\mathbf{h}}_{1,k}\|^2 s_{1,k}(t)}_{S(t)} + \underbrace{\hat{\mathbf{h}}_{1,k}^H \Delta\hat{\mathbf{h}}_{1,k} x_{1,k}(t)}_{I_1(t)} + \sum_{k'\neq k}^{K} \underbrace{\hat{\mathbf{h}}_{1,k}^H \mathbf{h}_{1,k'} x_{1,k'}(t)}_{I_2(t)}$$

$$+ \sum_{j\neq 1}^{L}\sum_{k'=1}^{K} \underbrace{\hat{\mathbf{h}}_{1,k}^H \mathbf{h}_{j,k'} x_{j,k'}(t)}_{I_3(t)} + \underbrace{\hat{\mathbf{h}}_{1,k}^H \mathbf{n}_1(t)}_{N(t)}, \qquad (14a)$$

$$\hat{\mathbf{h}}_{1,k}^H \mathbf{y}_1^{\mathrm{II}}(t) = \underbrace{\|\hat{\mathbf{h}}_{1,k}\|^2 s_{1,k}(t)}_{S(t)} + \underbrace{\hat{\mathbf{h}}_1^H \Delta\hat{\mathbf{h}}_{1,k} s_{1,k}(t)}_{I_1'(t)} + \sum_{k'\neq k}^{K} \underbrace{\hat{\mathbf{h}}_{1,k}^H \mathbf{h}_{1,k'} s_{1,k'}(t)}_{I_2'(t)}$$

$$+ \sum_{j\neq 1}^{L}\sum_{k'=k}^{K} \underbrace{\hat{\mathbf{h}}_{1,k}^H \mathbf{h}_{j,k'} s_{j,k'}(t)}_{I_3'(t)} + \underbrace{\hat{\mathbf{h}}_{1,k}^H \mathbf{n}_1(t)}_{N(t)}. \qquad (14b)$$

In the given equations, the first terms on right-hand-side $S(t)$ are the desired signals, while the rest four terms are attributed to interference. In particular, since the detection is based on $\hat{\mathbf{h}}_{1,k}^H$, we treat $I_1(t)$ and $I_1'(t)$ as interference, although both terms contain part of the desired signal. Therefore, we refer to $I_1(t)$ and $I_1'(t)$ as self-interference. By similarity, we refer to $I_2(t)$, $I_2'(t)$, $I_3(t)$ and $I_3'(t)$ as cross-interference since these terms contain interference across all L cells.

According to Jensen's inequality, a lower bound on the achievable uplink rate of k-th UT can be written as

$$R_{1,k} \geq \widetilde{R}_{1,k} = \log_2\left(1 + \frac{1}{E\left[\frac{1}{\gamma}\right]}\right), \qquad (15)$$

where γ is the signal-to-interference-plus-noise ratio (SINR). From (14a) and (14b), the UL rate

for using hybrid pilot-aided channel estimation is lower-bounded

$$\widetilde{R}_{1,k}(\alpha, \tau, \lambda) = \frac{\alpha\tau}{T}\log_2\left(1 + \frac{1}{E\left[\frac{1}{\gamma^{\mathrm{I}}}\right]}\right) + \left(1 - \frac{\tau}{T}\right)\log_2\left(1 + \frac{1}{E\left[\frac{1}{\gamma^{\mathrm{II}}}\right]}\right), \tag{16}$$

where γ^{I} and γ^{II} are respectively the SINRs contained in the output of the MF detector in (14a) and (14b), which can be expressed as

$$\gamma^{\mathrm{I}} = \frac{E\left[|S(t)|^2\right]}{E\left[|I_1(t)|^2\right] + \sum_{k'\neq k}^{K} E\left[|I_2(t)|^2\right] + \sum_{j\neq1}^{L}\sum_{k'\neq k}^{K} E\left[|I_3(t)|^2\right] + E\left[|N(t)|^2\right]}, \tag{17a}$$

$$\gamma^{\mathrm{II}} = \frac{E\left[|S(t)|^2\right]}{E\left[|I_1'(t)|^2\right] + \sum_{k'\neq k}^{K} E\left[|I_2'(t)|^2\right] + \sum_{j\neq1}^{L}\sum_{k'\neq k}^{K} E\left[|I_3'(t)|^2\right] + E\left[|N(t)|^2\right]}. \tag{17b}$$

Lemma 1. *For fixed values of α and λ, when M is large, the approximate SINR in (17a) and (17b), denoted by γ_{app}^{I} and γ_{app}^{II}, respectively, are given by*

$$\gamma_{app}^{I} = \frac{(1-\lambda)\beta_{1,k}^2}{\frac{(1-\lambda)\alpha}{\lambda}\frac{1}{\tau}b_1 + \frac{1}{M}\left(b_2 + \sigma_n^2\beta_{1,k}\right)}, \tag{18a}$$

$$\gamma_{app}^{II} = \frac{(1-\lambda)\beta_{1,k}^2}{\frac{(1-\lambda)^2\alpha}{\lambda}\frac{1}{\tau}b_1 + \frac{1}{M}\left((1-\lambda)b_2 + \sigma_n^2\beta_{1,k}\right)}, \tag{18b}$$

where $b_1 = \sum_{j=1}^{L}\sum_{k=1}^{K}\beta_{j,k}^2$ *and* $b_2 = \sum_{j\neq1}^{L}\sum_{k=1}^{K}\beta_{1,k}\beta_{j,k} + \sum_{k'\neq k}^{K}\beta_{1,k}\beta_{1,k'}$.

Proof: See Appendix-A.   ∎

Substituting (18a) and (18b) into (16), the UL rate from UT k is given by

$$\widetilde{R}_{1,k}(\alpha, \tau, \lambda) \rightarrow \frac{\alpha\tau}{T}\log_2\left(1 + \frac{(1-\lambda)\beta_{1,k}^2}{\frac{(1-\lambda)\alpha}{\lambda}\frac{1}{\tau}b_1 + \frac{1}{M}\left(b_2 + \sigma_n^2\beta_{1,k}\right)}\right)$$

$$+ \left(1 - \frac{\tau}{T}\right)\log_2\left(1 + \frac{(1-\lambda)\beta_{1,k}^2}{\frac{(1-\lambda)^2\alpha}{\lambda}\frac{1}{\tau}b_1 + \frac{1}{M}\left((1-\lambda)b_2 + \sigma_n^2\beta_{1,k}\right)}\right). \tag{19}$$

From (19), we have the following observations:

- The UL rate from k-th UT for employing hybrid pilots is limited even when $M \rightarrow \infty$, and can be well approximated for large M as

$$\lim_{M\rightarrow\infty}\widetilde{R}_{1,k}(\alpha, \tau, \lambda) \rightarrow \frac{\alpha\tau}{T}\log_2\left(1 + \frac{\beta_{1,k}^2}{\frac{\alpha}{\lambda\tau}b_1}\right) + \left(1 - \frac{\tau}{T}\right)\log_2\left(1 + \frac{\beta_{1,k}^2}{\frac{\alpha(1-\lambda)}{\lambda\tau}b_1}\right). \tag{20}$$

The result implies that, although the hybrid pilot-based scheme cannot completely mitigate the effect of pilot contamination, it provides the potential for significant improvement in the ultimate performance in comparison with the conventional pilot-aided designs through the following three adjustment factors: (1) The time-ratio between TM pilots and TS pilots α, (2) the time allocated to hybrid pilots (training overhead) τ, and (3) the power-ratio between pilots and data λ. It will be shown in Section VI numerical results that the hybrid pilot-aided scheme with optimal α, λ and τ yields a substantial improvement in the UL achievable rate. It is also worth noting that the conventional TS pilot [19], [22] is in principle the special case of the hybrid pilot when $\tau \to T$ and $\alpha \to 1$.

D. Performance Enhancement: A Data-aided Solution

Denote $\hat{s}_{j,k}(t)$ and $\Delta s_{j,k}(t)$ as the detection and the detection error after hard-decision operation w.r.t. the data symbol $s_{j,k}(t)$ of the target user in cell $j, j = 1, 2, \cdots, L$ obtained by using (10) and MF detector (14a). Then, we have

$$\Delta s_{j,k}(t) = s_{j,k}(t) - \hat{s}_{j,k}(t), \ \ k = 1, 2, \cdots, K. \tag{21}$$

As pointed out in [14], we make the following assumptions:

1) Both $\{\hat{s}_{j,k}(t)\}$ and $\{\Delta s_{j,k}(t)\}$ are zero-mean and contain i.i.d. samples,

2) $\{\Delta s_{j,k}(t)\}$ and $\{s_{j,k}(t)\}$ are mutually independent[2].

For a given signal constellation, e.g., M-PSK ($M = 2, 4, 8, \cdots$), and consider the worst case by assuming the farthest neighbor selection when executing data decoding. Define the distance between the data signal in the target cell and its detected data as $d_k(t) = |s_{1,k}(t) - \hat{s}_{1,k}(t)|^2$. Suppose hard decision is employed, then we have

$$d_k(t) = |\Delta s_{1,k}(t)| = \begin{cases} 0, & \text{w.p. } 1 - p_{e,k} \\ 2\sqrt{1 - \lambda}, & \text{w.p. } \ \ p_{e,k} \end{cases} \tag{22}$$

leading to

$$E\Big[\Delta s_{1,k}(t_1)\Delta s_{1,k}^*(t_2)\Big] = 4(1 - \lambda)\delta(t_1 - t_2), \tag{23}$$

---

[2]We show in Appendix A-D that the correlation between self-interference $I_1(t)$ and the desired signal $S(t)$ in (15a) is inversely proportional to the training size τ. Thereby, assumption 2) is fairly accurate in scenarios with large values of τ and M.

where $p_{e,k}$ is the steady-state error probability of data decoding at the BS w.r.t. the user in the target cell.

The detected data symbols and the estimated channel of the desired user are then used in feedback to iteratively refine the estimation quality by mitigating correlation between TS pilots and data in (10). To be specific, denote $\hat{\mathbf{h}}_{1,k}^{(i)}$ as the corresponding channel estimates of the i-th iteration for using the iterative data-aided solution, we have

$$
\begin{aligned}
\hat{\mathbf{h}}_{1,k}^{(i)} &= \left(\mathbf{Y}_1 - \sum_{k'=1}^{K} \hat{\mathbf{h}}_{1,k'}^{(i-1)} \hat{\mathbf{s}}_{1,k} \right) \cdot \frac{\mathbf{p}_{1,k}^H}{\|\mathbf{p}_{1,k}\|^2} \\
&= \mathbf{h}_{1,k} + \underbrace{\frac{1}{\|\mathbf{p}_{1,k}\|^2} \left(\sum_{k'=1}^{K} \left(\mathbf{h}_{1,k'} \left(\Delta \mathbf{s}_{1,k'} \mathbf{p}_{1,k}^H \right) + \Delta \mathbf{h}_{1,k}^{(i-1)} \left(\hat{\mathbf{s}}_{1,k} \mathbf{p}_{1,k}^H \right) \right) + \sum_{j\neq 1}^{L} \sum_{k'=1}^{K} \mathbf{h}_{j,k} \left(\mathbf{s}_{j,k'} \mathbf{p}_{1,k}^H \right) + \mathbf{N}_1 \mathbf{p}_{1,k}^H \right)}_{\Delta \mathbf{h}_{1,k}^{(i)}} .
\end{aligned}
\tag{24}
$$

where $\Delta \mathbf{h}_{1,k}^{(i-1)} = \mathbf{h}_{1,k} - \hat{\mathbf{h}}_{1,k}^{(i-1)}$. Using the property $\frac{1}{\tau} E[|\Delta \mathbf{s}_{1,k} \mathbf{p}_{1,k}^H|^2] \to 4 p_{e,k} \lambda (1-\lambda)$, and assuming that $\Delta \mathbf{h}_{1,k}^{(i)} \approx \Delta \mathbf{h}_{1,k}^{(i-1)}$ when i is large, we perform data detection of the i-th iteration as

$$
\hat{s}_{1,k}^{(i)}(t) =
\begin{cases}
s_{1,k}(t) + \frac{(\hat{\mathbf{h}}_{1,k}^{(i)})^H}{\|\hat{\mathbf{h}}_{1,k}^{(i)}\|^2} \left(\sum_{k'=1}^{K} \Delta \hat{\mathbf{h}}_{1,k'} x_{1,k'}(t) + \sum_{j\neq 1}^{L} \sum_{k'=1}^{K} \mathbf{h}_{j,k'} x_{j,k'}(t) + \mathbf{n}_1(t) \right), \\
\hspace{8cm} t = (1-\alpha)\tau + 1, \cdots, \tau, \\
s_{1,k}(t) + \frac{(\hat{\mathbf{h}}_{1,k}^{(i)})^H}{\|\hat{\mathbf{h}}_{1,k}^{(i)}\|^2} \left(\sum_{k'=1}^{K} \Delta \hat{\mathbf{h}}_{1,k'} s_{1,k'}(t) + \sum_{j\neq 1}^{L} \sum_{k'=1}^{K} \mathbf{h}_{j,k'} s_{j,k'}(t) + \mathbf{n}_1(t) \right), \\
\hspace{8cm} t = \tau + 1, \cdots, T.
\end{cases}
\tag{25}
$$

Simplifying the resulting expressing similarly as in the case of initial MF detection in Section IV-B, the achievable UL rate of i-th iteration at k-th UT in the target cell is given by

$$
\begin{aligned}
\widetilde{R}_{1,k}^{(i)}(\alpha, \tau, \lambda) =& \frac{\alpha \tau}{T} \log_2 \left(1 + \frac{(1-\lambda)\beta_{1,k}^2}{\frac{(1-\lambda)\alpha}{\lambda \tau} c_1 + \frac{1}{M}\left(c_2 + \beta_{1,k}\sigma_n^2 \right)} \right) \\
&+ \left(1 - \frac{\tau}{T} \right) \log_2 \left(1 + \frac{(1-\lambda)\beta_{1,k}^2}{\frac{(1-\lambda)^2\alpha}{\lambda \tau} c_1 + \frac{1}{M}\left((1-\lambda)c_2 + \beta_{1,k}\sigma_n^2 \right)} \right).
\end{aligned}
\tag{26}
$$

where $c_1 = \sum_{k'=1}^{K} 4 p_{e,k} \cdot \beta_{1,k'}^2 + \sum_{j\neq 1}^{L} \sum_{k'=1}^{K} \beta_{j,k'}^2$ and $c_2 = \sum_{j\neq 1}^{L} \sum_{k'=1}^{K} \beta_{1,k}\beta_{j,k'} + \sum_{k'\neq k}^{K} 4 p_{e,k'}\beta_{1,k}\beta_{1,k'}$.

In summary, the iterative data-aided solution can be interpreted as follows: In each step of iteration, hybrid pilots are used to estimate the channel, by which data detection is obtained through a MF detection. Then, the detected data is employed to refine the channel estimates by

mitigating the correlation between data and pilots of the desired user within the TS pilots phase, and in turn, to improve the quality of channel and data estimation in the forthcoming step, and thereby the UL achievable rate.

IV. ASYMPTOTICALLY OPTIMAL SOLUTIONS

In this section, we optimize the variables, including the time-ratio between TM and TS pilots α, the power-ratio between pilots and data λ, and the size of training overhead τ, aiming to improve the ultimate UL rate.

A. Problem Formulation

Achieving user fairness, in order to maximize the minimum rate for all users $\widetilde{R}_{1,k}^{(i)}(\alpha, \tau, \lambda)$ in (26), we have the following problem formulation

$$(P1) : \max_{\alpha, \lambda, \tau} \quad \min_{1 \leq k \leq K} \widetilde{R}_{1,k}^{(i)}(\alpha, \lambda, \tau), \tag{27}$$

$$\text{s. t.} \quad 0 \leq \alpha \leq 1,$$

$$0 \leq \lambda \leq 1,$$

$$KL \leq \tau \leq T.$$

For fixed M, the direct optimization on (P_1) is challenging due to the nonlinear relationship between the UL rate and the variables α, λ and τ. In massive MIMO systems, these variables are coupled between the training phase and the data phase, i.e.,

- In training phase, estimation quality depends on the α, τ, as well as the power-ratio λ.
- The estimation quality affects the detection performance and the UL rate.
- Besides the estimation quality, the UL rate depends on the ratio of data phase over the frame, which depends on both α and τ.

Nevertheless, we can obtain interesting asymptotical solutions and insights in the large-M regime.

B. Asymptotical Optimization on Time-Ratio between TM and TS Pilots

Lemma 2. *The convexity or concavity of $\widetilde{R}_{1,k}$ in (P1)[3] w.r.t. α depends on the time-ratio of training overhead over the whole frame $\frac{\tau}{T}$, which includes the following three cases:*

---

[3]For simplicity, we drop the superscript of $\widetilde{R}_{1,k}^{(i)}(\alpha, \lambda, \tau)$ in (26), and denote it by $\widetilde{R}_{1,k}$ in the following of this paper.

- *Case 1: When $\frac{\tau}{T} \in (0, \frac{1}{2+2g})$, $\widetilde{R}_{1,k}$ is a convex function w.r.t. α.*
- *Case 2: When $\frac{\tau}{T} \in (\frac{1}{1+2g}, 1)$, $\widetilde{R}_{1,k}$ is a concave function w.r.t. α.*
- *Case 3: When $\frac{\tau}{T} \in [\frac{1}{2+2g}, \frac{1}{1+2g}]$, $\widetilde{R}_{1,k}$ is a convex function when $\alpha \in [0, \frac{T}{\tau} - 1 - 2g]$, and a concave function when $\alpha \in [\frac{T}{\tau} - 1 - 2g, 1]$, respectively, where $g = \frac{\lambda\tau(c_2 + \beta_{1,k}\sigma_n^2)}{M(1-\lambda)c_1}$.*

Proof: See Appendix B-A.   ■

From lemma 2, the derivative of $\widetilde{R}_{1,k}$ w.r.t. α depends on $\frac{\tau}{T}$ and g. Thus, the direct optimization on α is challenging since exhaustive search is of high complexity. To this end, we firstly evaluate the monotonicity of $\widetilde{R}_{1,k}$ w.r.t. α, and then propose an iterative bisection procedure as follows.

Algorithm 1: Optimization of Time-Ratio between TM Pilots and TS Pilots

---

Data: The parameter from (26)

Result: Optimal Time-Ratio between TM Pilots and TS Pilots α

1: Calculate $\frac{1}{1+2g}$, $\frac{1}{2+2g}$ and $\frac{T}{\tau} - 1 - 2g$ as mentioned in Lemma 2. If $\frac{\tau}{T} < \frac{1}{2+2g}$, go to 2. Otherwise, go to 3.

2: Set $\alpha^{opt} = \arg \max\{\widetilde{R}_{1,k}(0), \widetilde{R}_{1,k}(1)\}$, since $\widetilde{R}_{1,k}$ is a convex function w.r.t. α (case 1);

3: Define the concave interval of $\widetilde{R}_{1,k}$ according to the value of $\frac{\tau}{T}$. If $\frac{\tau}{T} < \frac{1}{1+2g}$, we let $a = \frac{T}{\tau} - 1 - 2g$ and $b = 1$ (case 3), or Let $a = 0$ and $b = 1$ (case 2). Calculate $\widetilde{R}'_{1,k}(a)$ and $\widetilde{R}'_{1,k}(b)$ by (65), since $\widetilde{R}'_{1,k}$ is a monotonically decreasing function w.r.t. α. If $\widetilde{R}'_{1,k}(a) < 0$, set $\alpha^{opt} = 0$. If $\widetilde{R}'_{1,k}(b) > 0$, set $\alpha^{opt} = 1$. Otherwise, go to 4.

4: Start the iterative bisection procedure (For case 2 and 3 concave interval). Repeat the follow until $|\widetilde{R}'_{1,k}(\frac{a+b}{2})| < \epsilon$: Calculate $\widetilde{R}'_{1,k}(\frac{(a+b)}{2})$ and update the concave interval. If $\widetilde{R}'_{1,k}(\frac{a+b}{2}) > 0$, let $a = \frac{a+b}{2}$. Otherwise, let $b = \frac{a+b}{2}$. Set $\alpha^{opt} = \frac{a+b}{2}$. If $\frac{\tau}{T} < \frac{1}{1+2g}$ (case 3 concave interval), go to 5.

5: If $\widetilde{R}_{1,k}(\alpha^{opt}) < \widetilde{R}_{1,k}(0)$, set $\alpha^{opt} = 0$.

C. Asymptotical Optimization on Power-Ratio Allocated to Pilots

From (26), the asymptotically optimal power-ratio λ_{opt} is given in Lemma 3.

Lemma 3. *For fixed α and τ, the asymptotically optimal power-ratio between data and pilots is given by*

$$\lambda^{opt} = \frac{(T-\tau) + 2\alpha\tau - \sqrt{(T-\tau)^2 + 4\tau^2 f\left((T-\tau) + \alpha\tau\right)}}{2(\alpha\tau - \tau^2 f)}, \tag{28}$$

where $f = \frac{c_2 + \beta_{1,k}\sigma_n^2}{Mc_1}$.

Proof: See Appendix B-B.   ■

From (28), it can be seen that, for arbitrary α and τ, $\lambda^{opt} \to 1$ when $M \to \infty$, as pointed out in [22]. A brief explanation for this behavior is that, $M \to \infty$ results in a nearly "noise-free"

transmission background. It this case, increasing λ can *always* improve the estimation quality, and in turn, improves the UL rate without deteriorating the effective SNR.

D. Asymptotical Optimization on Time Allocated to Training Overhead

The optimization on τ^{opt} is given in Lemma 4.

Lemma 4. *The optimization on $\tau, \tau \in [KL, T]$, depends on both the power ratio of pilots λ and the time ratio between TM pilots and TS pilots α, which includes the following two cases,*

- *Case 1: For an arbitrary $\alpha \in [0, 1]$, when $\lambda \in (1 - 2^{-\frac{T}{(1+KLh)KLln2}}, 1]$, $\tau^{opt} = KL$.*
- *Case 2: For an arbitrary $\alpha \in [0, 1]$, when $\lambda \in [0, 1 - 2^{-\frac{T}{(1+KLh)KLln2}}]$, $\tau^{opt} \in (KL, T]$, where in particular $\tau^{opt} = T$ when $\alpha = 1$ and $h = \frac{\lambda}{(1-\lambda)\alpha M} \frac{c_2 + \beta_{1,k}\sigma_n^2}{c_1}$.*

Proof: See Appendix B-C.   ∎

An important consequence of $Case\ 1$ in Lemma 4 is that, in scenarios of large pilot power where λ is large enough to acquire a precise quality of channel estimation, reducing the training-size τ is beneficial to increase the efficiency of data transmission, as pointed out in [17]. In contrast, for cases of negligible pilot power λ (i.e., $Case\ 2$ of Lemma 4), where the estimation quality is not able to satisfy the detection quality, it is essential to increase τ to enhance the estimation performance, and in turn, to improve the cell rate. Particularly, we show in Appendix B-C that when $\{\lambda \in [0, 1 - 2^{-\frac{1}{(1+Th)ln2}}]\} \bigcup \{\alpha = [0, 1)\}$, $\widetilde{R}_{1,k}$ is a concave function w.r.t. τ, which implies that the global optimal τ^{opt} exists in the range $(KL, T]$. However, solving the optimization on τ is computational too expansive, because of the nonlinear relationship between α and τ. To ease the computational burden, we firstly determine the monotonicity of $\widetilde{R}_{1,k}$ w.r.t. τ, and then propose using an iterative bisection procedure to solve the optimization of τ, see Algorithm 2.

Algorithm 2: Optimization of Time Allocated to Training Overhead

---

Data: The parameter from (26)

Result: Optimal time allocated to training overhead τ

1: Calculate $\widetilde{R}'_{1,k}(KL)$ and $\widetilde{R}'_{1,k}(T)$ by (74), since $\widetilde{R}'_{1,k}$ is a monotonically decreasing function w.r.t. τ. If $\widetilde{R}'_{1,k}(KL) < 0$, set $\tau^{opt} = KL$. If $\widetilde{R}'_{1,k}(T) > 0$, set $\tau^{opt} = T$. Otherwise, go to 2.

2: Let $a = KL$ and $b = T$, and start the iterative bisection procedure. Repeat the follow until $|\widetilde{R}'_{1,k}(\frac{a+b}{2})| < \epsilon$:
Calculate $\widetilde{R}'_{1,k}(\frac{(a+b)}{2})$ and, update a and b. If $\widetilde{R}'_{1,k}(\frac{a+b}{2}) > 0$, let $a = \frac{a+b}{2}$, or let $b = \frac{a+b}{2}$. Set $\tau^{opt} = \frac{a+b}{2}$
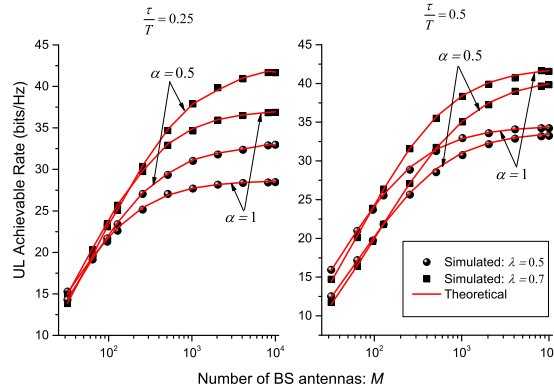

Fig. 2.   Theoretical v.s. simulated: UL rate for different variables α, λ and τ when SNR = 20dB.

V. SIMULATIONS

Consider a cellular network with $L = 7$ hexagonal cells, which consists of 1 target cell and 6 adjacent cells, with each cell $K = 10$ users. The radius of each cell (from center to vertex) is normalized, and the users are assumed to be uniformly randomly distributed. We model the pass loss of a link from k-th user in cell j to cell i as $\beta_{j,k} = \left(\frac{d_{1,k}}{d_{j,k}}\right)^{\gamma}\beta_{1,k}$, where $d_{j,k}$ denotes the distance between the user of the target cell (cell 1) and the BS of the jth cell, and γ is the pass-loss exponent. We set $\gamma = 3.8$ and $\beta_{1,k} = 1$ for all k for simplicity as in [7], and assume the channel to be quasi-static during a frame of transmission T.

A. UL Achievable Rate

Firstly, we conduct an experiment to validate the effectiveness of our theoretical analysis on the UL achievable rate described in Section III-C, where the average UL rate against different parameters (including the time-ratio α, the power-allocation factors λ and the size of training overhead τ) are plotted in Fig. 2. The solid lines are obtained by approximations derived in (19) using Monte Carlo simulations. For reference, we also simulate the approximations derived in (19). As shown in Fig. 2, the agreement between the actual values and the approximated ones demonstrates the validity of our analysis. In addition, we note that the system performance gradually saturates as M grows to infinity, i.e. $M \to 10^4$ in simulations, and the ultimate rate depends on the variables α, λ and τ, as pointed out in the analysis in Section III-C.

B. Optimal Time-Ratio between TM Pilots and TS Pilots

To explore the impact of time-ratio between TM pilots and TS pilots on the UL achievable rate, Fig. 3 plots the UL achievable rates against the time-ratio α, in order to validate the results
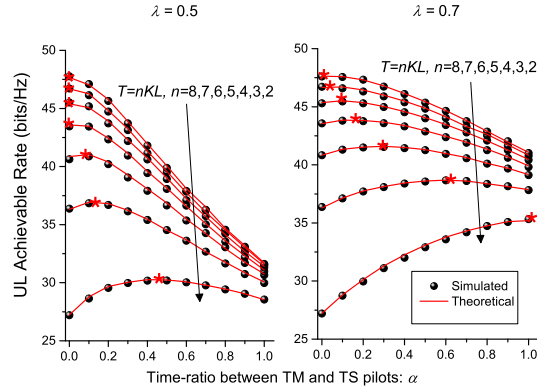
Fig. 3.   UL achievable rate for different values of α and τ for $\tau = KL$, and $\lambda = 0.5$ and 0.7, respectively.

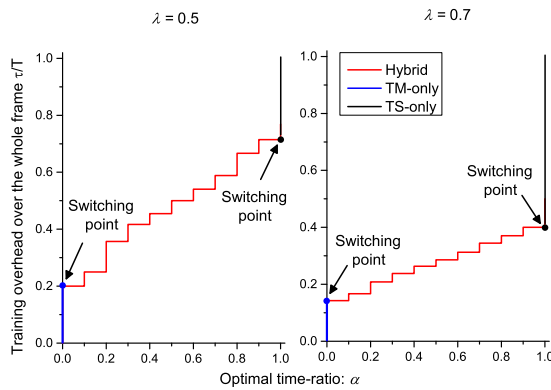

Fig. 4.   Optimal values of α for different values of $\frac{\tau}{T}$ ($M = 256$, $\lambda = 0.5, 0.7$ and SNR = 20dB).

described in Lemma 2. From Fig. 3, we can see that the optimal α depends on the time ratio $\frac{\tau}{T}$. Specifically, as the frame-size T increases (which equivalent to $\frac{\tau}{T}$ reduces for fixed τ), the optimal value of α reduces from 1 to 0. In particular, when T is small, i.e. $\frac{\tau}{T} \to 1$, $\alpha^{opt} \to 1$, which indicates that the optimal design for hybrid training overhead contains only TS pilots. This fact implies that TS pilot is more suitable for transmission of a limited frame-size T. On the contrary, when T is large, i.e., $\frac{\tau}{T} \to 0$, $\alpha^{opt} \to 0$. In this case, the optimal training overhead of the hybrid structure comprises only TM pilots, which implies that TM pilot is superior to TS pilot for large-frame based transmission. The above observations are also confirmed by the results shown in Fig. 4. Moreover, the tightness between the simulated results and the analytical ones further confirm the validity of our analysis.
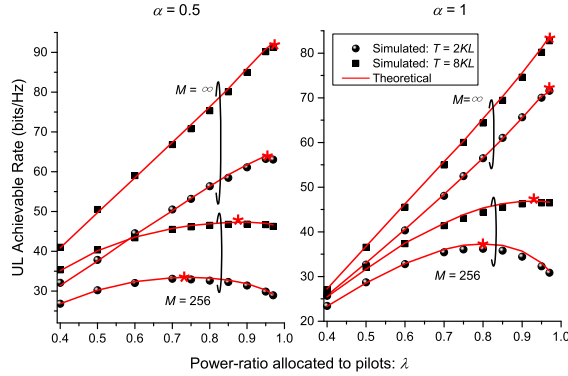
Fig. 5.   UL achievable rate versus different power-ratio between pilots and data λ when SNR = 20 dB.

C. Optimal Power-Ratio Allocated to Pilots

Next, we examine the effect of the power-ratio λ on the UL rate, in order to validate our analysis in Lemma 3. It can be seen in Fig. 5 that for finite M (e.g. $M = 64, 256$) and fixed values of α (e.g. $\alpha = 0.5$ and 1, respectively), the optimal values of λ that maximize R_1 in (27) are approximately 0.81 and 0.85 when $M = 256$ and $\alpha = 0.5$ and 1, respectively. When $M \to \infty$ (we set $M = 10^4$ in simulations), $\lambda^{opt} \to 1$. The result is expected and can be explained as follows,

1) For finite M, a larger λ leads to better estimation quality but simultaneously reduces SNR. This fact deteriorates the system performance.

2) When $M \to \infty$, the thermal noise vanishes due to the significant array gain. In such a "noise-free" scenario, increasing the λ always improves the estimation quality without reducing SNR. This leads to an increased performance directly proportional to λ.

The above results are consistent with our theoretical analysis detailed in Lemma 3. In addition, many of the simulation results generated in the course of this study (which have been removed here due to space constraints) also confirm that an excellent agreement exists between the actual values and the approximated ones.

D. Optimal Time Allocation to Training Overhead

We now move forward to investigate the performance of UL rate versus the time-ratio of training overhead $\frac{\tau}{T}$. In Fig. 6, we observe that, for fixed α, the optimal value of $\frac{\tau}{T}$ depends on the power ratio allocated to pilots λ. We also plot the optimal $\frac{\tau}{T}$ for different λ in Fig. 7. The

numerical results in both Fig. 6 and Fig. 7 agree with our analysis in Lemma 4 that $\tau \to KL$ when λ is large, whereas $\tau \to T$ for small value of λ when $\alpha = 1$. The explanation for this behavior is that, according to (12), when λ is small and not enough to acquire accurate channel estimates, it is essential to increase the training-size τ (i.e. $\frac{\tau}{T} \to 1$) to improve the estimation quality, in order to ensure the detection performance. In contrast, with the hybrid property in (7), increasing τ improves the estimation performance at the penalty of either introducing data interference or reducing the effective data rate. Thus, when λ is large enough to satisfy (78), the gain of channel estimation by increasing τ is insufficient to compensate for the corresponding loss of rate. This fact leads to $\tau^{opt} \to KL$.

E. Performance Comparison with the Conventional Pilot-based Schemes

Finally, the performance comparison between the hybrid pilot-aided scheme and the conventional methods is demonstrated in Fig. 8. In these simulations, we consider the conventional schemes (that were widely considered as benchmarks in related works) employing either TM pilots [10] or TS pilots [22] (legend by "TM pilot-aided" and "TS pilot-aided"), respectively, and perform data detection by using the MF detector in (13a). For fairness of comparison, both TM pilots and TS pilots are optimized in the sense of maximizing the UL achievable rate. To elaborate a little further, TM pilots are optimally designed and reused among cells for different coherent frame-size, while TS pilots are optimized on the aspect of power allocation between pilots and data. Clearly, the hybrid pilot-aided scheme is superior to both the TM pilot-only and TS pilot-only ones [10] and [22].

To gain an insight into the hybrid pilots, the same comparison is done in Fig. 9 for different transmission frame-size T. Clearly, the hybrid pilot-aided scheme performs superior to the conventional pilot-based schemes. It is also observed from Fig. 9 that, the TS pilot-based scheme [22] outperforms the TM pilot-only one [10] when frame-size T is limited, whereas the TM pilot-aided method performs better than that employing only TS pilots as T grows larger, i.e., when $T \gg 5KL$. This fact implies that the effectiveness for either TM pilots or TS pilots depends crucially on the frame-size T in practical scenarios. In particular, we note that the gap between the UL achievable rate of hybrid pilots scheme and TM pilots one narrows down when $T \geq 10KL$, and gradually vanishes as T further grows larger. This can be well explained by the results in Lemma 2 and Fig. 4 that when T is large, the optimal training overhead comprises
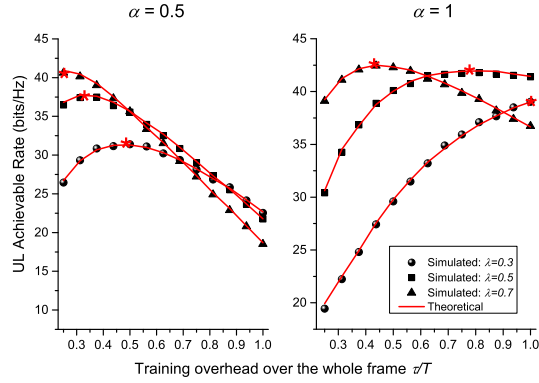
Fig. 6. UL achievable rate versus different time-ratio of training overhead $\frac{\tau}{T}$ for different values of λ when $\alpha = 0.5$ and 1, respectively ($M = 256$ and SNR = 20 dB).
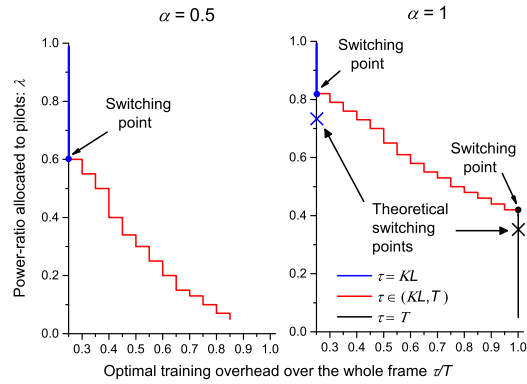

Fig. 7. Optimal time-ratio of training overhead over the whole frame $\frac{\tau}{T}$ ($M = 256$ and SNR = 20 dB).
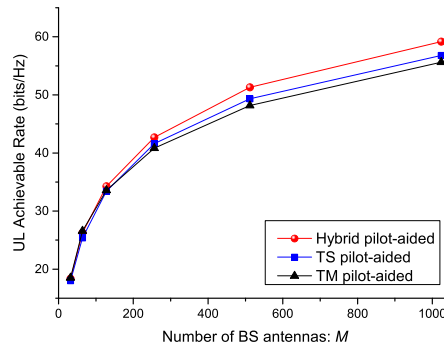

Fig. 8. Comparison of UL achievable rate for different pilot-based schemes ($M = 256$ and SNR = 20 dB).

only TM pilots, i.e., $\alpha^{opt} \to 1$. Anyway, the hybrid pilot-aided scheme performs the best among all these schemes for a wide range of T, which demonstrates the effectiveness of the proposed
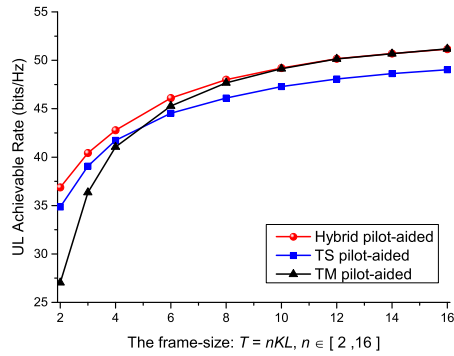
Fig. 9. UL achievable rate v.s. the transmitted frame-size T for different pilot-based schemes ($M = 256$ and SNR = 20 dB).

design.

VI. CONCLUSION

In this paper, we proposed a hybrid pilot-aided channel estimation scheme for multicell massive MIMO uplink, and analytically explored the impact of hybrid pilots on the UL achievable rate. Through an analysis of the relative importance, we show that higher UL rate can be achieved by employing both TM pilots and TS pilots, and additional performance improvements are gleaned by optimizing the time fraction between the two types of pilots, as well as the power and time ratio between pilots and data. Theoretical and numerical results demonstrate that the hybrid design enjoys mutual benefits between TM pilots and TS pilots, and thereby, offers a superior solution to the conventional pilot-based schemes in large MIMO systems.

APPENDIX A
DERIVATIONS OF UL ACHIEVABLE RATE

We discuss each term contained in (17a) and (17b), including the power of signal, noise, self-interference, and cross-interference, respectively.

A. Signal and Noise Power

From (14a) and (14b), we note that the signal and noise are identical, i.e., $\boldsymbol{S}(t) = \boldsymbol{S}'(t)$ and $\boldsymbol{N}(t) = \boldsymbol{N}'(t)$. Therefore, we only consider the power of $\boldsymbol{S}(t)$ and $\boldsymbol{N}(t)$, which can be given by

$$E\Big[|\boldsymbol{S}(t)|^2\Big] = E\Big[\|\hat{\mathbf{h}}_{1,k}\|^4\Big](1 - \lambda), \tag{29a}$$

$$E\Big[|\boldsymbol{N}(t)|^2\Big] = E\Big[\|\hat{\mathbf{h}}_{1,k}\|^2\Big]\sigma_n^2. \tag{29b}$$

We firstly evaluate the expectation $E[\|\hat{\mathbf{h}}_{1,k}\|^4]$ in the above equation. Since $\{\mathbf{p}_{1,k}\}$ and $\{\mathbf{s}_{j,k}\}, \forall j$ are fixed, $\hat{\mathbf{h}}_{1,k}$ are linear Gaussian with zero-mean and variance $\sigma_{\hat{h}_{1,k}}^2$ given by

$$\sigma_{\hat{h}_{1,k}}^2 = \beta_{1,k} + \frac{1}{(\tau\lambda)^2}\left(\sum_{j=1}^{L}\sum_{k'=1}\beta_{j,k}|\mathbf{s}_{j,k'}\mathbf{p}_{1,k}^H|^2 + \|\mathbf{p}_{1,k}\|^2\sigma_n^2\right)$$

$$= \beta_{1,k} + \frac{\sum_{j=1}^{L}\sum_{k'=1}^{K}\alpha(1-\lambda)\beta_{j,k'} + \sigma_n^2}{\tau\lambda}. \tag{30}$$

Since $\|\hat{\mathbf{h}}_{1,k}\|^2 \sim \frac{1}{2}\sigma_{\hat{h}_{1,k}}^2\chi^2(2M)$, where $\chi^2(2M)$ denotes the chi-square distribution with $2M$ degree of freedom, we can rewrite (29a) as

$$E\left[|\mathbf{S}(t)|^2\right] = \sigma_{\hat{h}_1}^4(M^2 + M)(1-\lambda). \tag{31}$$

Since $\tau > KL$, when $L \gg 1$ and $\beta_{1,k} \gg \beta_{j,k}$, $j \neq 1$, we have

$$\beta_{1,k} \gg \frac{\sum_{j=1}^{L}\sum_{k'=1}^{K}\alpha(1-\lambda)\beta_{j,k'} + \sigma_n^2}{KL\lambda}. \tag{32}$$

Taking average over the distribution $\mathbf{p}_{1,k}$ and $\mathbf{x}_{j,k}$, and omitting some intermediate derivations, it can be easily shown from (30) that

$$\sigma_{\hat{h}_{1,k}}^2 = \beta_{1,k} + \mathcal{O}\left(\frac{1}{\tau}\right), \tag{33}$$

which leads to $E[\|\hat{\mathbf{h}}_{1,k}\|^4] = M^2\beta_{1,k}^2 + \mathcal{O}\left(\frac{M^2}{\tau}\right) + \mathcal{O}\left(\frac{M^2}{\tau^2}\right)$. We obtain

$$E\left[|\mathbf{S}(t)|^2\right] = (1-\lambda)\beta_{1,k}^2 M^2 + \mathcal{O}\left(\frac{M^2}{\tau}\right) \approx (1-\lambda)\beta_{1,k}^2 M^2, \tag{34a}$$

$$E\left[|\mathbf{N}(t)|^2\right] = \sigma_n^2\beta_{1,k}M + \mathcal{O}\left(\frac{M}{\tau}\right) \approx \sigma_n^2\beta_{1,k}M. \tag{34b}$$

As shown above, the power of signal and noise are of an order $\mathcal{O}(M^2)$ and $\mathcal{O}(M)$, respectively. An explanation for this behavior is the array gain that benefits from the coherent combining in a massive MIMO system of M BS antennas.

B. Power of Self-Interference

We next consider the self-interference terms in (14a) and (14b),

$$\mathbf{I}_1(t) = \hat{\mathbf{h}}_{1,k}^H\left(\mathbf{h}_{1,k} - \hat{\mathbf{h}}_{1,k}\right)x_{1,k}(t), \tag{35a}$$

$$\mathbf{I}_1'(t) = \hat{\mathbf{h}}_{1,k}^H\left(\mathbf{h}_{1,k} - \hat{\mathbf{h}}_{1,k}\right)s_{1,k}(t). \tag{35b}$$

We firstly consider $\boldsymbol{I}_1(t)$, since the extension of derivation to $\boldsymbol{I}_1'(t)$ is straightforward. For the ease of analysis, we propose decomposing $\hat{\mathbf{h}}_{1,k}$ as two independent terms by pre- and post-multiplying (11) by $\hat{\mathbf{h}}_{1,k}^H$ and $\mathbf{h}_{1,k}^H$, respectively. Then, we obtain the following new equation,

$$\mathbf{h}_{1,k} = \phi_{1,k}\hat{\mathbf{h}}_{1,k} + \mathbf{w}_{1,k}, \tag{36}$$

where

$$\phi_{1,k} = \frac{\beta_{1,k}}{\sigma_{\hat{h}_{1,k}}^2}\left(1 + \frac{1}{\tau\lambda}\mathbf{p}_{1,k}\mathbf{s}_{1,k}^H\right), \tag{37}$$

$$\mathbf{w}_{1,k} = \frac{1}{\tau\lambda}\frac{\mathbf{h}_{1,k}}{M\sigma_{\hat{h}_{1,k}}^2}\left(\sum_{k'\neq k}^{K}\left(\mathbf{p}_{1,k}\mathbf{s}_{1,k'}^H\right)\mathbf{h}_{1,k'}^H + \sum_{j\neq 1}^{L}\sum_{k'=1}^{K}\left(\mathbf{p}_{1,k}\mathbf{s}_{j,k'}^H\right)\mathbf{h}_{j,k'}^H + \mathbf{p}_{1,k}\mathbf{N}_1^H\right)\hat{\mathbf{h}}_{1,k}. \tag{38}$$

Due to the independency between $\phi_{1,k}\hat{\mathbf{h}}_{1,k}$ and $\mathbf{w}_{1,k}$, the variance of $\mathbf{w}_{1,k}$, which is denoted by $\sigma_{w_{1,k}}^2$, can be given by

$$\sigma_{w_{1,k}}^2 = \beta_{1,k}\left(1 - \frac{\beta_{1,k}\left|1 + \frac{1}{\tau\lambda}\mathbf{p}_{1,k}\mathbf{s}_{1,k}^H\right|^2}{\sigma_{\hat{h}_{1,k}}^2}\right). \tag{39}$$

Substituting (37) and (38) into (35a), we can rewrite $\boldsymbol{I}_1(t)$ as

$$\boldsymbol{I}_1(t) = (\phi_{1,k} - 1)\|\hat{\mathbf{h}}_{1,k}\|^2 x_{1,k}(t) + \hat{\mathbf{h}}_{1,k}^H\mathbf{w}_{1,k}x_{1,k}(t). \tag{40}$$

Accordingly, the average power of $\boldsymbol{I}_1(t)$ can be given by

$$E\left[|\boldsymbol{I}_1(t)|^2\right] = |\phi_{1,k} - 1|^2 E\left[\|\hat{\mathbf{h}}_{1,k}\|^4\right] + \sigma_{w_{1,k}}^2 E\left[\|\hat{\mathbf{h}}_{1,k}\|^2\right] \tag{41a}$$

$$= |\phi_{1,k} - 1|^2\sigma_{\hat{h}_{1,k}}^4(M^2 + M) + \sigma_{w_{1,k}}^2\sigma_{\hat{h}_{1,k}}^2 M. \tag{41b}$$

The equality in (41a) is due to the independency between $\hat{\mathbf{h}}_{1,k}$ and $\mathbf{w}_{1,k}$. By averaging $|\phi_{1,k}|^2\sigma_{\hat{h}_{1,k}}^4$ and $\sigma_{w_{1,k}}^2\sigma_{\hat{h}_{1,k}}^2$ in (41b) over the distribution $\{\mathbf{p}_{1,k}\}$ and $\{\mathbf{s}_{j,k}\}, \forall j$, we can rewrite (41b) as

$$
\begin{aligned}
E\left[|\boldsymbol{I}_1(t)|^2\right] &= E\left[\frac{\beta_{1,k}^2|\mathbf{p}_{1,k}\mathbf{s}_{1,k}^H|^2}{(\tau\lambda)^2} + \frac{\sum_{j=1}^{L}\sum_{k'=1}^{K}\beta_{j,k'}|\mathbf{p}_{1,k}\mathbf{s}_{j,k'}^H|^2 + \|\mathbf{p}_{1,k}\|^4\sigma_n^4}{(\tau\lambda)^4}\right](M^2 + M) \\
&\quad + E\left[\frac{\beta_{1,k}\left(\sum_{k'\neq k}^{K}|\mathbf{p}_{1,k}\mathbf{s}_{1,k'}^H|^2 + \sum_{j\neq 1}^{L}\sum_{k'=1}^{K}\beta_{j,k'}|\mathbf{p}_{1,k}\mathbf{s}_{j,k'}^H|^2 + \|\mathbf{p}_{1,k}\|^2\sigma_n^2\right)}{(\tau\lambda)^2}\right]M \\
&= \frac{(1 - \lambda)\alpha}{\lambda} \cdot \frac{M^2}{\tau} \cdot \beta_{1,k}^2 + \mathcal{O}\left(\frac{M^2}{\tau^2}\right) + \mathcal{O}\left(\frac{M}{\tau}\right) \approx \frac{(1 - \lambda)\alpha}{\lambda} \cdot \frac{M^2}{\tau} \cdot \beta_{1,k}^2. \tag{42}
\end{aligned}
$$

By similarity, according to (32), the approximate power of $\mathbf{I}'_1(t)$ for large M can be derived as

$$E\left[|\mathbf{I}'_1(t)|^2\right] = (1-\lambda)E\left[|\mathbf{I}_1(t)|^2\right] \approx \frac{\alpha(1-\lambda)^2}{\lambda} \cdot \frac{M^2}{\tau} \cdot \beta^2_{1,k}. \qquad (43)$$

The details of (43) are omitted since the derivations are similar to that of $\mathbf{I}_1$. It is worth noting that the power of self-interference and signal are both in the order of $\mathcal{O}(M^2)$. That is, if τ is fixed and $\alpha \neq 0$, self-interference imposes a limit on SINR in (17a) even when $M \to \infty$.

C. Power of Cross-Interference

In the following, we only consider the cross-interference terms $E[|I_2(t)|^2]$ and $E[|I_3(t)|^2]$, since the extension to $I'_2(t)$ and $I'_3(t)$ are straightforward. From (14a), the power of cross-interference is given by

$$E\left[|I_2(t)|^2\right] = E\left[|\hat{\mathbf{h}}^H_{1,k}\mathbf{h}_{1,k'}|^2\right], \quad k' \neq k, \qquad (44a)$$

$$E\left[|I_3(t)|^2\right] = E\left[|\hat{\mathbf{h}}^H_{1,k}\mathbf{h}_{j,k}|^2\right], \quad j \neq 1. \qquad (44b)$$

From (10), $\hat{\mathbf{h}}_{1,k}$ and $\mathbf{h}_{1,k'}$ are correlated. Again, we decompose $\mathbf{h}_{1,k'}$ as

$$\mathbf{h}_{1,k'} = \phi_{1,k'}\hat{\mathbf{h}}_{1,k} + \mathbf{w}_{1,k'}, \qquad (45)$$

where

$$\phi_{1,k'} = \frac{1}{\tau\lambda}\frac{\beta_{1,k'}}{\sigma^2_{\hat{h}_{1,k}}}\left(\mathbf{p}_{1,k}\mathbf{s}^H_{1,k'}\right), \qquad (46)$$

$$\mathbf{w}_{1,k'} = \frac{1}{\tau\lambda\sigma^2_{\hat{h}_{1,k}}}\mathbf{h}_{1,k'}\left(\sum_{k''\neq k'}^{K}\left(\mathbf{p}_{1,k}\mathbf{s}^H_{1,k''}\right)\mathbf{h}^H_{1,k''} + \sum_{j\neq 1}^{L}\sum_{k''=1}^{K}\left(\mathbf{p}_{1,k}\mathbf{s}^H_{j,k''}\right)\mathbf{h}^H_{j,k''} + \mathbf{p}_{1,k}\mathbf{N}^H_j + \mathbf{h}_{1,k}\right)\hat{\mathbf{h}}^H_{1,k}. \qquad (47)$$

Substituting (46) and (47) into (45), we obtain

$$E\left[|I_2(t)|^2\right] = E\left[\hat{\mathbf{h}}^H_{1,k}\left(\phi_{1,k'}\hat{\mathbf{h}}^H_{1,k} + \mathbf{w}_{1,k'}\right)\right]$$

$$= |\phi_{1,k'}|^2 E\left[\|\hat{\mathbf{h}}_{1,k}\|^4\right] + \sigma^2_{w_{1,k'}}E\left[\|\hat{\mathbf{h}}_{1,k}\|^2\right] \qquad (48a)$$

$$= |\phi_{1,k'}|^2\sigma^4_{\hat{h}_{1,k}}(M^2 + M) + \sigma^2_{w_{1,k'}}\sigma^2_{\hat{h}_{1,k}}M. \qquad (48b)$$

where $\sigma^2_{w_{1,k'}}$ is the variance of $\mathbf{w}_{1,k'}$ given by

$$\sigma^2_{w_{1,k'}} = \beta_{1,k'}\left(1 - \frac{\beta_{1,k'}}{(\tau\lambda)^2\sigma^2_{\hat{h}_{1,k}}}\left|\mathbf{p}_{1,k}\mathbf{s}^H_{1,k'}\right|^2\right). \qquad (49)$$

The equalities in (48a) is due to the independency between $\hat{\mathbf{h}}_{1,k}$ and $\mathbf{w}_{1,k'}$. Similar to (42), we obtain $E[|\phi_{1,k'}|^2 \sigma_{\hat{h}_{1,k}}^4]$ and $E[\sigma_{w_{1,k'}}^2 \sigma_{\hat{h}_{1,k}}^2]$ by averaging over the distribution $\{\mathbf{p}_{1,k}\}$ and $\{\mathbf{s}_{1,k}\}, \forall j$, and then, we can rewrite (48b) as

$$
\begin{aligned}
E\left[|I_2(t)|^2\right] &= E\left[|\phi_{1,k'}|^2 \sigma_{\hat{h}_{1,k}}^4\right](M^2 + M) + E\left[\sigma_{w_{1,k'}}^2 \sigma_{\hat{h}_{1,k}}^2\right]M \\
&= E\left[\frac{\beta_{1,k'}^2 |\mathbf{p}_{1,k}\mathbf{s}_{1,k'}^H|^2}{(\tau\lambda)^2}\right](M^2+M)+E\left[\beta_{1,k'}\left(\beta_{1,k}\right.\right. \\
&\quad \left.\left. + \frac{\sum_{k''\neq k'}^{K}\beta_{1,k''}|\mathbf{p}_{1,k}\mathbf{s}_{j,k''}^H|^2 + \sum_{j\neq 1}^{L}\sum_{k''=1}^{K}\beta_{j,k''}|\mathbf{p}_{1,k}\mathbf{s}_{j,k''}^H|^2+\|\mathbf{p}_{1,k}\|^2\sigma_n^2 + \beta_{1,k}}{(\tau\lambda)^2}\right)\right]M \\
&= \frac{(1-\lambda)\alpha}{\lambda} \cdot \frac{M^2}{\tau} \cdot \beta_{1,k'}^2 + M\beta_{1,k'}\beta_{1,k} + \mathcal{O}\left(\frac{M}{\tau}\right) \\
&\approx \frac{(1-\lambda)\alpha}{\lambda} \cdot \frac{M^2}{\tau} \cdot \beta_{1,k'}^2 + M\beta_{1,k'}\beta_{1,k}. \tag{50}
\end{aligned}
$$

By analogy, the power of $I_2'(t)$ can be derived as

$$
E\left[|I_2'(t)|^2\right] = (1-\lambda)E\left[|I_2(t)|^2\right] \approx \frac{(1-\lambda)^2\alpha}{\lambda} \cdot \frac{M^2}{\tau} \cdot \beta_{1,k'}^2 + (1-\lambda)M\beta_{1,k'}\beta_{1,k}. \tag{51}
$$

Applying a similar procedure, we decompose $\mathbf{h}_{j,k}$ as

$$
\mathbf{h}_{j,k} = \phi_{j,k}\hat{\mathbf{h}}_{1,k} + \mathbf{w}_{j,k}, \tag{52}
$$

where

$$
\phi_{j,k} = \frac{1}{\tau\lambda}\frac{\beta_{j,k}}{\sigma_{\hat{h}_{1,k}}^2}\left(\mathbf{p}_{1,k}\mathbf{s}_{j,k}^H\right), \tag{53}
$$

$$
\mathbf{w}_{j,k} = \frac{1}{\tau\lambda\sigma_{\hat{h}_{1,k}}^2}\mathbf{h}_{j,k}\left(\sum_{k'\neq k}^{K}\left(\mathbf{p}_{1,k}\mathbf{s}_{j,k'}^H\right)\mathbf{h}_{j,k'}^H + \sum_{i\neq j}^{L}\sum_{k'=1}^{K}\left(\mathbf{p}_{1,k}\mathbf{s}_{i,k'}^H\right)\mathbf{h}_{i,k'}^H + \mathbf{p}_{1,k}\mathbf{N}_i^H + \mathbf{h}_{1,k}\right)\hat{\mathbf{h}}_{1,k}^H. \tag{54}
$$

Substituting (53) and (54) into (52), we obtain

$$
\begin{aligned}
E\left[|I_3(t)|^2\right] &= E\left[\hat{\mathbf{h}}_{1,k}^H\left(\phi_{j,k}\hat{\mathbf{h}}_{1,k}^H + \mathbf{w}_{j,k}\right)\right] \\
&= |\phi_{j,k}|^2 E\left[\|\hat{\mathbf{h}}_{1,k}\|^4\right] + \sigma_{w_{j,k}}^2 E\left[\|\hat{\mathbf{h}}_{1,k}\|^2\right] \tag{55a} \\
&= |\phi_{j,k}|^2 \sigma_{\hat{h}_{1,k}}^4(M^2 + M) + \sigma_{w_j}^2 \sigma_{\hat{h}_{1,k}}^2 M. \tag{55b}
\end{aligned}
$$

where $\sigma^2_{w_{j,k}}$ is the variance of $\mathbf{w}_{j,k}$ given by

$$\sigma^2_{w_{j,k}} = \beta_{j,k}\left(1 - \frac{\beta_{j,k}}{(\tau\lambda)^2\sigma^2_{\hat{h}_{1,k}}}\left|\mathbf{p}_{1,k}\mathbf{s}^H_{j,k}\right|^2\right). \tag{56}$$

The equalities in (55a) and (56) are due to the independency between $\hat{\mathbf{h}}_{1,k}$ and $\mathbf{w}_{j,k}$. Similar to (42), we obtain $E[|\phi_{j,k}|^2\sigma^4_{\hat{h}_{1,k}}]$ and $E[\sigma^2_{w_{j,k}}\sigma^2_{\hat{h}_{1,k}}]$ by averaging over the distribution $\{\mathbf{p}_{1,k}\}$ and $\{\mathbf{s}_{j,k}\}, \forall j$, and then, we can rewrite (55b) as

$$E\left[|I_3(t)|^2\right] = E\left[\frac{\beta^2_{j,k}|\mathbf{p}_{1,k}\mathbf{s}^H_{j,k}|^2}{(\tau\lambda)^2}\right](M^2+M)+E\left[\beta_{j,k}\left(\beta_{1,k}\right.\right.$$

$$\left.\left.+\frac{\sum^K_{k'\neq k}\beta_{j,k'}|\mathbf{p}_{1,k}\mathbf{s}^H_{j,k'}|^2 + \sum^L_{i\neq j}\sum^K_{k'=1}\beta_{j,k'}|\mathbf{p}_{1,k}\mathbf{s}^H_{j,k'}|^2+\|\mathbf{p}_{1,k}\|^2\sigma^2_n + \beta_{1,k}}{(\tau\lambda)^2}\right)\right]M$$

$$= \frac{(1-\lambda)\alpha}{\lambda}\cdot\frac{M^2}{\tau}\cdot\beta^2_{j,k} + M\beta_{j,k}\beta_{1,k} + \mathcal{O}\left(\frac{M}{\tau}\right)$$

$$\approx \frac{(1-\lambda)\alpha}{\lambda}\cdot\frac{M^2}{\tau}\cdot\beta^2_{j,k} + M\beta_{j,k'}\beta_{1,k}. \tag{57}$$

By analogy, the power of $I'_3(t)$ can be derived as

$$E\left[|I'_3(t)|^2\right] = (1-\lambda)E\left[|I_3(t)|^2\right] \approx \frac{(1-\lambda)^2\alpha}{\lambda}\cdot\frac{M^2}{\tau}\cdot\beta^2_{j,k} + (1-\lambda)M\beta_{j,k}\beta_{1,k}. \tag{58}$$

D. Correlation of Signal and Self-Interference

From (17a) and (17b), we observe that the signal $S(t)$ and self-interference $I_1(t)$ (I'_1 in (14b)) are correlated. This correlation may complicate the performance analysis of the data detection. To evaluate the effect of correlation, we quantify the correlation using following criteria [28]

$$\zeta = \frac{\left|E[S^*(t)I_1(t)]\right|^2}{E\left[|S(t)|^2\right]E\left[|I_1(t)|^2\right]}. \tag{59}$$

Since $E[|S(t)|^2]$ and $E[|I_1(t)|^2]$ in the denominator of (59) have been derived in (34a) and (42), respectively, we consider $|E[S^*(t)I_1(t)]|^2$ in the numerator of (59). Following the same procedure as in deriving self-interference, we have

$$E[S^*(t)I_1(t)] = E\left[\|\hat{\mathbf{h}}_{1,k}\|\;\hat{\mathbf{h}}^H_{1,k}\left(\mathbf{h}_{1,k} - \hat{\mathbf{h}}_{1,k}\right)\right](1-\lambda)$$

$$= E\left[(\phi_{1,k} - 1)\|\hat{\mathbf{h}}_{1,k}\|^4\right](1-\lambda) + E\left[\|\hat{\mathbf{h}}_{1,k}\|\left(\hat{\mathbf{h}}^H_{1,k}\mathbf{w}_{1,k}\right)\right](1-\lambda)$$

$$= (\phi_{1,k} - 1)\sigma^4_{\hat{h}_{1,k}}(M^2 + M)(1-\lambda). \tag{60}$$

Again, we derive $E[(\phi_{1,k} - 1)\sigma^4_{\hat{h}_{1,k}}]$ by averaging $(\phi_{1,k} - 1)\sigma^4_{\hat{h}_{1,k}}$ over the distribution $\{\mathbf{p}_{1,k}\}$ and $\{\mathbf{s}_{j,k}\}, \forall j$

$$
\begin{aligned}
E\left[(\phi_{1,k} - 1)\sigma^4_{\hat{h}_{1,k}}\right] &= E\left[\left(\frac{\mathbf{p}_{1,k}\mathbf{s}^H_{1,k}}{\tau\lambda}\beta_{1,k} - \sum_{j=1}^{L}\sum_{k'=1}^{K}\frac{|\mathbf{p}_{1,k}\mathbf{s}^H_{j,k'}|^2}{(\tau\lambda)^2}\beta_{j,k'} - \frac{\|\mathbf{p}_{1,k}\|^2\sigma^2_n}{(\tau\lambda)^2}\right)\right. \\
&\quad\times \left.\left(\beta_{1,k} + \sum_{j=1}^{L}\sum_{k'=1}^{K}\frac{|\mathbf{p}_{1,k}\mathbf{s}^H_{1,k'}|^2}{(\tau\lambda)^2}\beta_{j,k'} + \frac{\|\mathbf{p}_{1,k}\|^2\sigma^2_n}{(\tau\lambda)^2}\right)\right] \\
&= \frac{E[\mathbf{p}_{1,k}\mathbf{s}^H_{1,k}]\beta^2_{1,k}}{\lambda}\cdot\frac{1}{\tau} + \mathcal{O}\left(\frac{1}{\tau^2}\right).
\end{aligned}
\tag{61}
$$

From (32), and (60)-(61), when M is large, we obtain

$$
\begin{aligned}
\left|E[S^*(t)I_1(t)]\right|^2 &= \left|E\left[(\phi_{1,k} - 1)\sigma^4_{\hat{h}_{1,k}}\right](M^2 + M)(1 - \lambda)\right|^2 \\
&= \frac{(1 - \lambda)^2\beta^4_{1,k}\left|E[\mathbf{p}_{1,k}\mathbf{s}^H_{1,k}]\right|^2}{\lambda^2}\cdot\frac{M^4}{\tau^2} + \mathcal{O}\left(\frac{M^4}{\tau^4}\right) + \mathcal{O}\left(\frac{M^2}{\tau^2}\right).
\end{aligned}
\tag{62}
$$

From (34a) and (42), we have

$$
E\left[|S(t)|^2\right]E\left[|I_1(t)|^2\right] = \frac{\alpha(1 - \lambda)^2\beta^4_{1,k}}{\lambda}\cdot\frac{M^4}{\tau} + \mathcal{O}\left(\frac{M^4}{\tau^3}\right) + \mathcal{O}\left(\frac{M^3}{\tau^3}\right).
\tag{63}
$$

Substituting (62) and (63) into the (59), we arrive at the following result

$$
\zeta = \frac{\frac{(1-\lambda)^2\beta^4_{1,k}|E[\mathbf{p}_{1,k}\mathbf{s}^H_{1,k}]|^2}{\lambda^2}\cdot\frac{M^4}{\tau^2} + \mathcal{O}\left(\frac{M^4}{\tau^4}\right) + \mathcal{O}\left(\frac{M^2}{\tau^2}\right)}{\frac{\alpha(1-\lambda)^2\beta^4_{1,k}}{\lambda}\cdot\frac{M^4}{\tau} + \mathcal{O}\left(\frac{M^4}{\tau^3}\right) + \mathcal{O}\left(\frac{M^3}{\tau^3}\right)} \approx \frac{|E[\mathbf{p}_{1,k}\mathbf{s}^H_{1,k}]|^2}{\alpha\lambda\cdot\tau} \ll 1.
\tag{64}
$$

Since $\{s_{1,k}(t)\}$ and $\{p_{1,k}(t)\}$ are mutually independent, i.e. $\frac{1}{\tau}E[\mathbf{p}_{1,k}\mathbf{s}^H_{1,k}] \to 0$, it is reasonable to ignore the correlation between the signal and self-interference when both M and τ are large.

APPENDIX B

ASYMPTOTICALLY OPTIMAL SOLUTIONS

A. Proof of Lemma 2

Proof: The derivative of $\widetilde{R}_{1,k}$ in (26) w.r.t. α is given by

$$
\frac{\partial\widetilde{R}_{1,k}}{\partial\alpha} = -\left(1 - \frac{\tau}{T} + \frac{\alpha\tau}{T}\right)\cdot\frac{1}{ln2}\cdot\frac{1}{\alpha + g} + \frac{\tau}{T}\log_2\left(1 + \frac{(1 - \lambda)\beta^2_{1,k}}{\frac{(1-\lambda)\alpha}{\lambda\tau}c_1 + \frac{1}{M}\left(c_2 + \beta_{1,k}\sigma^2_n\right)}\right),
\tag{65}
$$

where $g = \frac{\lambda\tau(c_2+\beta_{1,k}\sigma_n^2)}{M(1-\lambda)c_1}$. From (65), we note that the linearity of $\widetilde{R}_{1,k}$ w.r.t. α depends on $\frac{\tau}{T}$. By deriving the second order derivative of $\widetilde{R}_{1,k}$ w.r.t. α, we obtain

$$
\begin{aligned}
\frac{\partial^2 \widetilde{R}_{1,k}}{\partial \alpha^2} &= -\frac{2\tau}{T} \cdot \frac{1}{ln2} \cdot \frac{1}{\alpha+g} + \left(1 - \frac{\tau}{T} + \frac{\alpha\tau}{T}\right) \cdot \frac{1}{ln2} \cdot \frac{1}{(\alpha+g)^2} \\
&= \left(\frac{1}{ln2} \cdot \frac{1}{\alpha+g}\right)\left(-\frac{2\tau}{T} + \left(1 - \frac{\tau}{T} + \frac{\alpha\tau}{T}\right) \cdot \frac{1}{(\alpha+g)}\right).
\end{aligned}
\tag{66}
$$

Firstly, let $\frac{\partial^2 \widetilde{R}_{1,k}}{\partial \alpha^2} > 0$, we easily obtain

$$
\frac{\tau}{T} < \frac{1}{1+\alpha+2g} \le \frac{1}{2+2g}.
\tag{67}
$$

This implies that when $\frac{\tau}{T} < \frac{1}{1+\alpha+2g} \le \frac{1}{2+2g}$, $\widetilde{R}_{1,k}$ is a convex function w.r.t. α. By similarity, let $\frac{\partial^2 \widetilde{R}_{1,k}}{\partial \alpha^2} < 0$, we obtain

$$
\frac{\tau}{T} > \frac{1}{1+\alpha+2g} \ge \frac{1}{1+2g},
\tag{68}
$$

which means that $\widetilde{R}_{1,k}$ is a concave function w.r.t. α when $\frac{\tau}{T} > \frac{1}{1+\alpha+2g} \ge \frac{1}{1+2g}$. Finally, we consider the case $\frac{\tau}{T} \in [\frac{1}{2+2g}, \frac{1}{1+2g}]$. Let $\frac{\partial^2 \widetilde{R}_{1,k}}{\partial \alpha^2} > 0$, we obtain

$$
\alpha > \frac{T}{\tau} - 1 - 2g.
\tag{69}
$$

This implies that $\widetilde{R}_{1,k}$ is a convex function for $\alpha \in [0, \frac{T}{\tau} - 1 - 2g)$ when $\frac{\tau}{T} \in [\frac{1}{2+2g}, \frac{1}{1+2g}]$. By analogy, let $\frac{\partial^2 \widetilde{R}_{1,k}}{\partial \alpha^2} \le 0$ yields

$$
\alpha \le \frac{T}{\tau} - 1 - 2g.
\tag{70}
$$

This fact leads to $\widetilde{R}_{1,k}$ a concave function for $\alpha \in [\frac{T}{\tau} - 1 - 2g, 1]$ when $\frac{\tau}{T} \in [\frac{1}{2+2g}, \frac{1}{1+2g}]$.

This concludes the proof. ∎

B. Proof of Lemma 3

Proof: Taking the first derivative of $\widetilde{R}_{1,k}$ w.r.t. λ, and omitting some intermediate derivations, we have

$$
\frac{\partial \widetilde{R}_{1,k}}{\partial \lambda} = \frac{\alpha\tau}{T} \cdot \frac{1}{ln2} \cdot \frac{-\left(\frac{\alpha}{\lambda\tau} + \frac{f}{(1-\lambda)}\right) + \frac{\alpha}{\lambda^2\tau}}{\frac{(1-\lambda)\alpha}{\lambda\tau} + f} + \left(1 - \frac{\tau}{T}\right) \cdot \frac{1}{ln2} \cdot \frac{\frac{\alpha}{\lambda^2\tau}}{\frac{(1-\lambda)\alpha}{\lambda\tau} + f},
\tag{71}
$$

where $f = \frac{c_2+\beta_{1,k}\sigma_n^2}{Mc_1}$. Simplifying some intermediate derivations and setting (71) to zero, we arrive at the following second-order polynomial of λ,

$$
\left(\alpha\tau - \tau^2 f\right)\lambda^2 - \left(T - \tau + 2\alpha\tau\right)\lambda + \left(T - \tau + \alpha\tau\right) = 0.
\tag{72}
$$

It is easily to show that $\widetilde{R}_{1,k}(\lambda)$ is a concave function. Hence, the global maximum can be obtained by solving the above formula. The optimal λ that maximizes $\widetilde{R}_{1,k}$ can be given by

$$\lambda^{opt} = \frac{(T - \tau) + 2\alpha\tau - \sqrt{(T - \tau)^2 + 4\tau^2 f\left((T - \tau) + \alpha\tau\right)}}{2(\alpha\tau - \tau^2 f)}. \tag{73}$$

This concludes the proof. ∎

C. Proof of Lemma 4

Proof: The first derivative of $\widetilde{R}_{1,k}$ w.r.t. τ is given by

$$\frac{\partial \widetilde{R}_{1,k}}{\partial \tau} = \frac{\alpha}{T}\log_2\left(1 + \frac{(1 - \lambda)\beta_{1,k}^2}{\frac{(1-\lambda)\alpha}{\lambda\tau}c_1 + \frac{c_2+\beta_{1,k}\sigma_n^2}{M}}\right) - \frac{1}{T}\log_2\left(1 + \frac{\beta_{1,k}^2}{\frac{(1-\lambda)\alpha}{\lambda\tau}c_1 + \frac{c_2+\beta_{1,k}\sigma_n^2}{M}}\right)$$

$$+ \left(\frac{1}{\tau} - \frac{1-\alpha}{T}\right) \cdot \frac{1}{ln2} \cdot \frac{\frac{(1-\lambda)\alpha}{\lambda\tau}c_1}{\frac{(1-\lambda)\alpha}{\lambda\tau}c_1 + \frac{c_2+\beta_{1,k}\sigma_n^2}{M}}. \tag{74}$$

Let $h = \frac{\lambda}{(1-\lambda)\alpha M}\frac{c_2+\beta_{1,k}\sigma_n^2}{c_1}$, the second order derivative of $\widetilde{R}_{1,k}$ w.r.t. τ can be given by

$$\frac{\partial^2 \widetilde{R}_{1,k}}{\partial \tau^2} = -\frac{1}{ln2}\left(\underbrace{\frac{1-\alpha}{T} \cdot \frac{1}{\tau(1+h\tau)}}_{\geq 0} + \underbrace{\frac{T - (1-\alpha)\tau}{T\tau} \cdot \frac{h}{(1+h\tau)^2}}_{\geq 0} + \underbrace{\frac{1}{(1+h\tau)\tau^2}}_{>0}\right). \tag{75}$$

From the above, the second order derivative on τ is strictly negative, implying that $\widetilde{R}_{1,k}$ is a concave function respect to τ. From (74), we have the following observations:

- $C1$: When $\tau = KL$, $\frac{\partial \widetilde{R}_{1,k}}{\partial \tau} < 0$, which implies that $\widetilde{R}_{1,k}$ is a monotonically decreasing function with $\tau^{opt} = KL$.
- $C2$: When $\frac{\partial \widetilde{R}_{1,k}}{\partial \tau}|_{\tau=T} > 0$, $\widetilde{R}_{1,k}$ is a monotonically increasing function with $\tau^{opt} = T$.
- $C3$: When $\frac{\partial \widetilde{R}_{1,k}}{\partial \tau}|_{\tau<T} \geq 0$, $\widetilde{R}_{1,k}$ is a concave function with $\tau^{opt} \in (KL, T)$.

Next, we focus on the necessary and sufficient conditions of the above cases. From (74), we note that $\frac{\partial \widetilde{R}_{1,k}}{\partial \tau}$ depends on α, and thus, take the first order derivative of (74) w.r.t. α as

$$\frac{\partial^2 \widetilde{R}_{1,k}}{\partial \tau \partial \alpha} = \frac{1}{T}\log_2\left(1 + \frac{(1 - \lambda)\beta_{1,k}^2}{\frac{(1-\lambda)\alpha}{\lambda\tau}c_1 + \frac{1}{M}\left(c_2 + \beta_{1,k}\sigma_n^2\right)}\right)$$

$$+ \frac{1}{ln2} \cdot \frac{1}{(\alpha+h)T}\left((1-\alpha) + \frac{\alpha^2 + 2\alpha h}{\alpha + h} + \frac{(T-\tau)h}{(\alpha+h)\tau}\right) > 0. \tag{76}$$

As a consequence, $\frac{\partial \widetilde{R}_{1,k}}{\partial \tau}$ is an increasing function respect to α. Thus, $\alpha = 1$ maximizes $\frac{\partial \widetilde{R}_{1,k}}{\partial \tau}$, which can be denoted by $\max\{\frac{\partial \widetilde{R}_{1,k}}{\partial \tau}\}|_{\alpha=1}$. Provided that $\max\{\frac{\partial \widetilde{R}_{1,k}}{\partial \tau}\}|_{\alpha=1} < 0$, $\widetilde{R}_{1,k}$ is a monotonically decreasing function w.r.t. $\tau, \tau \in [KL, T]$, which implies that $\tau^{opt} = KL$ is the optimal time-allocation to the training overhead that maximizes the UL rate $\widetilde{R}_{1,k}$. This result is identical to that of $(C1)$ with the corresponding necessary and sufficient condition as

$$\max\{\widetilde{R}_{1,k}\}\Big|_{\substack{\alpha=1 \\ \tau=KL}} = \frac{1}{T}\log_2(1 - \lambda) + \frac{1}{ln2} \cdot \frac{1}{(1 + KLh)KL} < 0, \tag{77}$$

which is equivalent to

$$1 > \lambda > 1 - \frac{1}{2^{\frac{T}{(1+KLh) \cdot KLln2}}}. \tag{78}$$

Using the same argument, the necessary and sufficient condition of $(C2)$ is given by

$$\max\{\widetilde{R}_{1,k}\}\Big|_{\substack{\alpha=1 \\ \tau=KL}} = \frac{1}{T}\log_2(1 - \lambda) + \frac{1}{ln2} \cdot \frac{1}{(1 + Th)KL} > 0, \tag{79}$$

which is equivalent to

$$0 < \lambda < 1 - \frac{1}{2^{\frac{T}{(1+KLh) \cdot KLln2}}}. \tag{80}$$

∎

This concludes the proof.

REFERENCES

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

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

[3] E. Larsson, O. Edfors, F. Tufvesson, and T. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186−195, Feb. 2014.

[4] J. Andrews, et.al., "What will 5G be?" *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065−1082, Jun. 2014.

[5] J. Hoydis, S. Brink, and M. Debbah, "Massive MIMO in the UL/DL of cellular networks: how many antennas do we need?" *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 160−171, Feb. 2013.

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

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

[8] H. Wang, Y. Huang, S. Jin, F. Yu, and L. Yang, "Performance analysis on precoding and pilot scheduling in very large MIMO multi-cell systems," *in Proc. IEEE WCNC*, Apr. 2013, pp. 1−5.

[9] J. Jose, A. Ashikhmin, T. L. Marzetta, and S. Vishwanath, "Pilot contamination and precoding in multi-cell TDD systems," *IEEE Trans. Wireless Commun.*, vol. 10, no. 8, pp. 2640−2611, Aug. 2011.

[10] H. Yin, D. Gesbert, M. Filippou, and Y. Liu, "A coordinated approach to channel estimation in large-scale multiple-antenna systems," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 264−273, Feb. 2013.

[11] H. Q. Ngo and E. Larsson, "Evd-based channel estimation in multicell multiuserMIMO systems with very large antenna arrays," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 2012, pp. 3249−3252.

[12] R. Muller, L. Cottatellucci, and M. Vehkapera, "Blind pilot decontamination," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 773−786, Oct. 2014.

[13] Y. Li, Y. H. Nam, B. L. Ng, and J. Z. Zhang, "A non-asymptotic throughput for massive MIMO cellular uplink with pilot reuse," in *Proc. IEEE Globalcom*, Anaheim, CA, USA, Dec. 2012, pp. 4500−4504.

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

[15] T. E. Bogale and L. B. Le, "Pilot optimization and channel estimation for multiuser massive MIMO systems," in *Proc. CISS*, Montreal, QC, Canada, 2014, pp. 1−6.

[16] S. Noh, M. D. Zoltowski, Y. Sung, and D. J. Love, "Pilot beam pattern design for channel estimation in massive MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 787−801, Oct. 2014.

[17] H. Q. Ngo, M. Matthaiou, and E. G. Larsson, "Massive MIMO With Optimal Power and Training Duration Allocation" *IEEE Wireless Commun. Lett.*, vol. 3, no. 6, pp. 605−608, Dec. 2014.

[18] J. Zuo, J. Zhang, C. Yuen, W. Jiang, W. Luo, "Multi-Cell multi-user massive MIMO transmission with downlink training and pilot contamination precoding," *IEEE Trans. Vehi. Tech.*, Aug. 2015.

[19] K. Upadhya, S. A. Vorobyov, M. Vehkaperä, "Superimposed pilots are superior for mitigating pilot contamination in massive MIMO Part I: theory and channel estimation," arXiv:1603.00648v1.

[20] J. M. Li, H. Zhang, D. Li and H. B. Chen, "On the performance of wireless-energy-transfer-enabled massive MIMO systems with superimposed pilot-aided channel estimation," *IEEE Access*, vol. 3, pp. 2014−2027, 2015.

[21] H. Zhang, D. Pan, H. Cui and F. Gao, "Superimposed training for channel estimation of OFDM modulated amplify-and-forward relay networks," *Sci. China Info. Science*, vol. 56, no. 10, pp. 1−12, July 2013.

[22] H. Zhang, S. Gao, D. Li, H. B. Chen and L. Yang, "On superimposed pilot for channel estimation in multi-cell multiuser MIMO uplink: large system analysis," *IEEE Trans. Vehi. Tech.*, vol. 65, no. 3, pp. 1492−1505, March 2016.

[23] X. Dai, H. Zhang, and D. Li, "Linearly time-varying channel estimation for MIMO/OFDM systems using superimposed training," *IEEE Trans. Commun.*, vol. 58, no. 2, pp. 681−693, Feb. 2010.

[24] W. Huang, C. Li, and H. Li, "On the power allocation and system capacity of OFDM systems using superimposed training schemes," *IEEE Trans. Veh. Technol.*, vol. 58, no. 4, pp. 1731−1380, May 2009.

[25] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Upper Saddle River, NJ, USA: Prentice Hall PTR, 1993.

[26] E. Björnson, E. G. Larsson, and M. Debbah, "Massive MIMO for maximal spectral efficiency: how many users and pilots should be allocated?" *IEEE Trans. Wireless Communi.*, vol. 15, no. 2, pp. 1293−1308, Feb. 2016.

[27] S. Jin, X. Wang, Z. Li, K. K. Wong, Y. M. Huang, and X. Y. Tang, "On massive MIMO zero-forcing transceiver using time-shifted pilots" *IEEE Trans. Vehi. Tech.*, vol. 65, no. 1, pp. 59−74, Jan. 2016.

[28] S. Ross, A First Course in Probability. Englewood Cliffs, NJ, USA: Prentice-Hall, 2005.