

# Pilot Contamination Elimination for Large-Scale Multiple-Antenna Aided OFDM Systems

Jiankang Zhang, *Member, IEEE*, Bo Zhang, *Student Member, IEEE*, Sheng Chen, *Fellow, IEEE*, Xiaomin Mu, Mohammed El-Hajjar, and Lajos Hanzo, *Fellow, IEEE*

**Abstract**—Pilot contamination constitutes a particularly significant impairment in large-scale multi-cell systems. We propose an effective pilot contamination elimination scheme for multi-cell time division duplexing based orthogonal frequency division multiplexing systems, by carefully designing a sophisticated amalgam of downlink (DL) training and ‘scheduled’ uplink (UL) training. During the DL training stage, each base station (BS) transmits the DL pilot symbols (PSs) to its mobile stations (MSs) for them to estimate their frequency-domain channel transfer functions (FDCHTFs), which are then embedded in the UL PSs by ‘predistorting’ the PSs with the estimated FDCHTFs. During the scheduled UL training, each BS’s UL receiver in turn extracts the FDCHTFs of its MSs from their received PSs by eliminating the pilot contamination imposed by the simultaneously transmitted UL PSs of all other cells. Our simulation results demonstrate that the pilot contamination is completely eliminated by the proposed scheme, even for the network consisting of a large number of unity frequency reuse cells. Most significantly, unlike many existing pilot contamination reduction schemes, our scheme does not rely on the assumption that each BS knows the second-order statistics of all the interfering UL channels.

**Index Terms**—Channel estimation, multi-cell systems, multiple antennas, pilot contamination, time division duplexing.

## I. INTRODUCTION

In order to achieve a high area-spectral efficiency, the same carrier frequencies are reused in the neighboring cells by obeying specific reuse patterns, where the base stations (BSs) employing multiple antennas are capable of simultaneously supporting a multiplicity of autonomous mobile stations (MSs) [1], [2]. However, the performance of these multi-user, multiple-antenna aided systems is critically dependent on the

accuracy of the channel state information (CSI), regardless whether the CSI is used for uplink (UL) reception or for the downlink (DL) transmit precoding (TP) or transmit beamforming (TBF) [3]. The CSI can be acquired at the BS with the aid of UL training, which may then be used for DL TBF, by exploiting the reciprocity between the UL and the DL of time-division duplexing (TDD) systems [4]. The accuracy of the channel estimate (CE) and thus the attainable system performance relies on having perfectly orthogonal pilots allocated to the different users. However, using orthogonal pilots may only be guaranteed for the users roaming within the same cell, but not for those scattered across different cells, because the convolution of the pilots with long channel impulse responses (CIRs) destroys their orthogonality. Moreover, the limited available bandwidth may not allow unique, user-specific orthogonal pilots to be employed for each user [5], especially not in multi-cell systems relying on a radical frequency reuse factor of one. Then the pilots will be gravely polluted by the adjacent cells’ users, when the BS of the serving cell carries out CE. Compared to the effect of additive white Gaussian noise (AWGN), pilot contamination constitutes a much more grave impairment that limits the system’s performance [5]–[9].

The impact of pilot contamination imposed on multi-cell, multiple-antenna systems was characterized in [5] and it was shown that the precoded DL signal of the BS in the serving cell contaminates the received signal of the users roaming in other cells [5]. A study of pilot contamination was given in [10] and a minimum mean square error (MMSE) criterion based precoding method was developed for mitigating the above-mentioned contamination. The authors of [7] analyzed the pilot contamination problems in multi-cell, multiple-input multiple-output (MIMO) systems relying on a large number of antennas at the BS, which demonstrated that pilot contamination persists in large-scale MIMO systems. In order to reduce the pilot contamination in multi-cell, multiple-antenna systems, the work [1] proposed a modified frame structure for each cell, where non-overlapping pilots were used for the different cells. The study [6] analyzed the asymptotic behavior of the signal to interference plus noise ratio (SINR) for both time aligned and time staggered pilots, which revealed that it may be possible to cancel the interference of adjacent cells, as long as the pilots do not overlap in time. However, the transmission of staggered pilots requires a central controller for managing the staggering of the pilot-intervals in all of the cells in order to safeguard their ‘orthogonality’ across different cells, which becomes a challenge for an increasing number of users and cells.

Manuscript received September 12, 2013; revised January 01, 2014; accepted February 22, 2014. Date of publication March 05, 2014; date of current version null date. This work was supported by the National Natural Science Foundation of China under Grants 61301150, 61271421, and 61172086, the RC-UK under the India-U.K. Advanced Technology Centre, the European Research Councils under its Advanced Fellow Grant, and the EU’s Concerto Project. The associate editor coordinating the review of this manuscript and approving it for publication was XXXX XXXXXXXX.

J. Zhang and X. Mu are with the School of Information Engineering, Zhengzhou University, Zhengzhou 450001, China (e-mail: jz09v@ecs.soton.ac.uk; iexmmu@zzu.edu.cn).

B. Zhang, S. Chen, M. El-Hajjar, and L. Hanzo are with Electronics and Computer Science, University of Southampton, Southampton SO17 1BJ, U.K. (e-mail: bz2g10@ecs.soton.ac.uk; sqc@ecs.soton.ac.uk; meh@ecs.soton.ac.uk; lh@ecs.soton.ac.uk).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/JSTSP.2014.2309936

Coordinated approaches were proposed for DL TBF in [11]–[13] to mitigate the inter-cell interference and the idea was extended to the coordinated CE in [2] for alleviating the effects of pilot contamination. The channel estimator adopted in the coordinated CE scheme of [2] is the linear MMSE (LMMSE) estimator [5], [10], which yields a better performance than the least squares (LS) estimator, but it requires the knowledge of the second-order statistics of all the UL channels, including those of the interfering MSs roaming in the adjacent cells. The coordinated CE scheme of [2] further exploits the known second-order statistics of all the UL channels for designing a covariance-aware pilot assignment strategy for improving the LMMSE CE. Similar to the LMMSE estimator of [5], [10], the coordinated CE scheme [2] also requires the knowledge of the second-order statistics of all the UL channels. It is worth emphasizing that each BS is estimating its in-cell channels, but not the interfering channels from the adjacent cells. Thus, it is unrealistic to assume that the BS knows the second-order statistics of all the UL channels before they were even estimated. Therefore, it is impossible to implement such a CE scheme in practical multi-cell systems.

Against the above background, we propose an effective pilot contamination elimination scheme for multi-cell TDD orthogonal frequency division multiplexing (OFDM) systems. More specifically, the proposed pilot contamination elimination scheme relies on the two processing stages of a DL training and a scheduled UL training. During the DL training stage, each BS transmits its DL pilots to the MSs supported in order for them to estimate their specific DL frequency-domain channel transfer functions (FDCHTFs). During the scheduled UL training stage, one cell at a time, the MSs of each cell predistort their UL pilot symbols (PSs) using their estimated DL-FDCHTFs and hence the uncontaminated DL-FDCHTFs are ‘encapsulated’ in their UL PSs for exploitation by the BS. Furthermore, each BS’s UL receiver in turn extracts all the DL-FDCHTFs of its MSs from the received UL signals by eliminating the pilot contamination imposed by the UL-PS transmissions of all the other cells. Unlike the schemes of [2], [5], [10], our scheme does not assume the knowledge of the channels’ second-order statistics and, therefore, it offers a practical means of alleviating the effects of pilot contamination in large-scale multi-cell MIMO systems.

Throughout our discussion,  $\mathbb{C}$  denotes the complex number field, bold fonts are used to denote matrices and vectors, where the row vector convention is adopted. Lower and upper case symbols represent the time-domain (TD) and frequency-domain (FD) signals, respectively. For any  $A \in \mathbb{C}$ , we have  $A = A_R + jA_I$ , where  $j = \sqrt{-1}$ , while  $A_R$  and  $A_I$  are the real and imaginary parts of  $A$ , respectively. The transpose and Hermitian transpose operators are denoted by  $(\cdot)^T$  and  $(\cdot)^H$ , respectively, while  $\text{tr}\{\cdot\}$  denotes the trace operation and  $\text{diag}\{\mathbf{X}\}$  denotes the diagonal matrix with the diagonal entries equal to the components of  $\mathbf{X}$ . The inverse operation is denoted by  $(\cdot)^{-1}$ , while  $\mathbb{E}\{\cdot\}$  and  $\text{var}\{\cdot\}$  stand for the expectation and variance operations, respectively. Furthermore,  $\otimes$  denotes the Kronecker tensor product and  $^*$  denotes the conjugate operation, while  $\hat{X}$  represents the estimate of  $X$ .

The rest of this paper is organized as follows. Section II describes the MIMO-aided multi-cell TDD OFDM system.

Section III is devoted to the description of the proposed pilot contamination elimination scheme, where the achievable system throughput of our scheme is also detailed. In Section IV, we present our simulation results for characterizing the proposed pilot contamination elimination scheme in various scenarios. Our conclusions are offered in Section V.

## II. MULTI-CELL TDD SYSTEM MODEL

Consider a cellular network composed of  $L$  hexagonal cells, labelled by  $l = 1, 2, \dots, L$ , where each of the  $U$  simultaneous users of each cell is equipped with a single transmit/receive antenna, while the BS of each cell employs an array of  $Q$  transmit/receive antennas. All BSs and MSs are synchronized and rely on a TDD protocol as well as unity frequency reuse (UFR). We assume that the PSs used by the single-antenna MSs in a given cell are mutually orthogonal. The employment of OFDM is justified, because it was standardized for fourth-generation wireless communications [14].

### A. Uplink Training

At the commencement of data transmission, all MSs of all cells synchronously transmit  $K$  OFDM PSs to their serving BSs. The  $k$ -th FD PS of user  $u$  in the  $l$ -th cell is given by  $\mathbf{X}_l^u[k] = [X_l^u[k, 1]X_l^u[k, 2] \cdots X_l^u[k, N]]$ , where  $N$  is the number of subcarriers and the power of each pilot  $X_l^u[k, n]$  is assumed to be unity. Let  $H_{l^*, l, q}^u[k, n]$  denote the UL-FDCHTF of the  $u$ -th user in cell  $l$  and the  $q$ -th receiver antenna of cell  $l^*$  at the  $n$ -th subcarrier of the  $k$ -th OFDM symbol. Furthermore,  $Y_{l^*, q}[k, n]$  denotes the signal received by the  $q$ -th receiver antenna element in cell  $l^*$ , which can be expressed as [1]

$$Y_{l^*, q}[k, n] = \sqrt{p_r} \sum_{u=1}^U H_{l^*, l^*, q}^u[k, n] X_{l^*}^u[k, n] + \sqrt{p_r} \sum_{l=1, l \neq l^*}^L \sum_{u=1}^U H_{l^*, l, q}^u[k, n] X_l^u[k, n] + W_{l^*, q}[k, n] \quad (1)$$

for  $1 \leq l^* \leq L$  and  $1 \leq q \leq Q$ , where  $p_r$  denotes the average power of each user and  $W_{l^*, q}[k, n]$  is the FD representation of the UL channel’s AWGN, whose power is  $\sigma_w^2$ . The set of equations constituted by (1) for  $1 \leq q \leq Q$  can be written as

$$\mathbf{Y}_{l^*}[k, n] = \underbrace{\sqrt{p_r} \mathbf{X}_{l^*}[k, n] \mathbf{H}_{l^*, l^*}[k, n]}_{\text{Desired term}} + \underbrace{\sqrt{p_r} \sum_{l=1, l \neq l^*}^L \mathbf{X}_l[k, n] \mathbf{H}_{l^*, l}[k, n]}_{\text{Inter-cell interference}} + \mathbf{W}_{l^*}[k, n], \quad (2)$$

where  $\mathbf{Y}_{l^*}[k, n] \in \mathbb{C}^{1 \times Q}$  and  $\mathbf{W}_{l^*}[k, n] \in \mathbb{C}^{1 \times Q}$  are the two row vectors hosting  $Y_{l^*, q}[k, n]$  and  $W_{l^*, q}[k, n]$  for  $1 \leq q \leq Q$ , respectively,  $\mathbf{X}_{l^*}[k, n] \in \mathbb{C}^{1 \times U}$  and  $\mathbf{X}_l[k, n] \in \mathbb{C}^{1 \times U}$  are the two row vectors hosting  $X_{l^*}^u[k, n]$  and  $X_l^u[k, n]$  for  $1 \leq u \leq U$ , respectively, while  $\mathbf{H}_{l^*, l^*}[k, n] \in \mathbb{C}^{U \times Q}$  and  $\mathbf{H}_{l^*, l}[k, n] \in \mathbb{C}^{U \times Q}$  are the two FDCHTF matrices having their  $u$ -th row and  $q$ -th column elements given by  $H_{l^*, l^*, q}^u[k, n]$  and  $H_{l^*, l, q}^u[k, n]$ , respectively.

We assume that the channel is time-invariant for the duration of  $(L + 3)K$  OFDM symbols, which allows us to drop

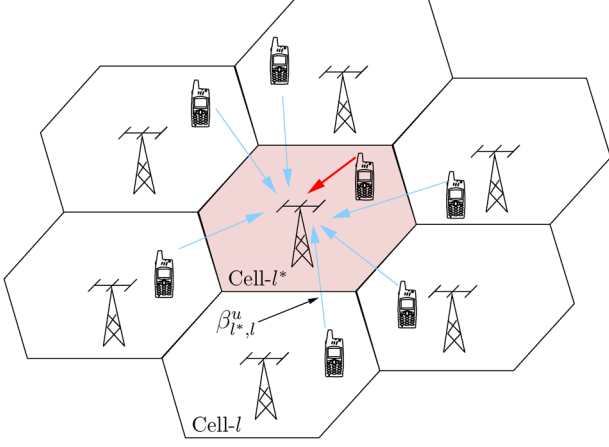


Fig. 1. Multiple-antenna aided multi-cell TDD system, where the uplink training in the cell  $l^*$  is contaminated by the MSS' transmissions from the adjacent cells which use the same pilot symbols.

the OFDM symbol index  $k$  from  $H_{l^*,l,q}^u[k,n]$ . Specifically,  $\mathbf{H}_{l^*,l,q}^u[k] = \mathbf{H}_{l^*,l,q}^u$  for  $1 \leq k \leq (L+3)K$ , where the FDCHTF row vector  $\mathbf{H}_{l^*,l,q}^u = [H_{l^*,l,q}^u[1] H_{l^*,l,q}^u[2] \cdots H_{l^*,l,q}^u[N]] \in \mathbb{C}^{1 \times N}$  represents the link between the  $u$ -th MS in the  $l$ -th cell and the  $q$ -th antenna of the target BS in the  $l^*$ -th cell over all the  $N$  OFDM subcarriers. According to [2], we set  $\mathbf{H}_{l^*,l,q}^u$  as

$$\mathbf{H}_{l^*,l,q}^u = \mathbf{h}_{l^*,l,q}^u \mathbf{F}^T, \quad (3)$$

with the TD CIR vector  $\mathbf{h}_{l^*,l,q}^u \in \mathbb{C}^{1 \times P}$  given by

$$\mathbf{h}_{l^*,l,q}^u = \left[ \alpha_{l^*,l,q,1}^u e^{-j2\pi \frac{(q-1)D}{\lambda} \cos(\theta_{l^*,l,q,1}^u)} \cdots \alpha_{l^*,l,q,P}^u e^{-j2\pi \frac{(q-1)D}{\lambda} \cos(\theta_{l^*,l,q,P}^u)} \right], \quad (4)$$

where  $D$  and  $\lambda$  are the antenna spacing at the BS and the carrier's wavelength, respectively,  $P$  is the number of resolvable paths,  $\theta_{l^*,l,q,p}^u$  is the angle of arrival (AoA) of the  $p$ -th path between the  $u$ -th MS in the  $l$ -th cell and the  $q$ -th antenna of the serving BS, while

$$\alpha_{l^*,l,q,p}^u = e^{-j\varphi_{l^*,l,q,p}^u} \sqrt{\beta_{l^*,l,q,p}^u} \quad (5)$$

is the complex-valued tap of the  $p$ -th path. Furthermore,  $\mathbf{F} \in \mathbb{C}^{N \times P}$  in (3) is the fast Fourier transform (FFT) matrix whose elements are given by  $F[n,p] = e^{-j2\pi(n-1)(p-1)/N}$  for  $1 \leq n \leq N$  and  $1 \leq p \leq P$ . The phase  $\varphi_{l^*,l,q,p}^u$  of the path is a random variable uniformly distributed in  $[0, 2\pi)$ . The path loss coefficient  $\beta_{l^*,l,q,p}^u$  accounts for the attenuation and shadow fading of the path, which change slowly as a function of distance [3], and we may assume that  $\beta_{l^*,l,q,p}^u = \beta_{l^*,l}^u$  for  $1 \leq p \leq P$  and  $1 \leq q \leq Q$ . The pilot contamination experienced during the UL training is illustrated in Fig. 1.

Since the channel is time-invariant for the duration of  $K$  symbols, we have  $\mathbf{H}_{l^*,l}[k,n] = \mathbf{H}_{l^*,l}[n] \in \mathbb{C}^{U \times Q}$  for  $1 \leq k \leq K$  in (2), where the  $u$ -th row and  $q$ -th column element of  $\mathbf{H}_{l^*,l}[n]$  is  $H_{l^*,l,q}^u[n]$ . As a benefit of the cyclic prefix, the OFDM symbols do not overlap in time and the BS processing can be carried out on a per-carrier basis. Hence, to simplify our notation, we will omit the subcarrier index  $n$  in the sequel. Assuming that a total of  $K$  consecutive OFDM symbols are dedicated to pilot subcarriers [3], the discrete-time model of the received signal associ-

ated with  $K$  OFDM symbols, namely (2) over  $1 \leq k \leq K$ , can be written as

$$\mathbf{Y}_{l^*} = \sqrt{p_r} \mathbf{X}_{l^*} \mathbf{H}_{l^*,l^*} + \sqrt{p_r} \sum_{l=1, l \neq l^*}^L \mathbf{X}_l \mathbf{H}_{l^*,l} + \mathbf{W}_{l^*}, \quad (6)$$

with  $\mathbf{Y}_{l^*} \in \mathbb{C}^{K \times Q}$ ,  $\mathbf{X}_l \in \mathbb{C}^{K \times U}$  and  $\mathbf{W}_{l^*} \in \mathbb{C}^{K \times Q}$ .

The LS CE and the LMMSE CE of the FDCHTF matrix  $\mathbf{H}_{l^*,l^*}$  are given respectively by [2], [10]

$$\hat{\mathbf{H}}_{l^*,l^*}^{\text{LS}} = \frac{1}{\sqrt{p_r}} \left( \mathbf{X}_{l^*}^H \mathbf{X}_{l^*} \right)^{-1} \mathbf{X}_{l^*}^H \mathbf{Y}_{l^*}, \quad (7)$$

$$\hat{\mathbf{H}}_{l^*,l^*}^{\text{LMMSE}} = \sqrt{p_r} \mathbf{R}_{l^*,l^*} \mathbf{X}_{l^*}^H \times \left( \sigma_w^2 \mathbf{I}_K + p_r \sum_{l=1}^L \mathbf{X}_l \mathbf{R}_{l^*,l} \mathbf{X}_l^H \right)^{-1} \mathbf{Y}_{l^*}, \quad (8)$$

where  $\mathbf{R}_{l^*,l} = \mathbb{E}\{\mathbf{H}_{l^*,l} \mathbf{H}_{l^*,l}^H\}$  is the covariance matrix of the FDCHTF matrix  $\mathbf{H}_{l^*,l}$ . The LS CE (7) is simple to implement, since it is only based on the target cell's PSs. By contrast, the LMMSE CE (8), which yields a much better estimate than the LS CE, relies on all the PSs from all the cells and, therefore, it requires the full cooperation of all the BSs to exchange their UL PSs. Thus, the LMMSE CE imposes a significantly higher complexity than the LS CE. Most critically, the LMMSE CE requires that the serving BS knows the second-order statistics  $\mathbf{R}_{l^*,l}$  of all the channels  $\mathbf{H}_{l^*,l}$  for  $1 \leq l \leq L$ .

## B. Pilot Contamination

From (7) and (8), we can see that the estimation of the desired channel  $\mathbf{H}_{l^*,l^*}$  relies on correlating the received signal with the known PSs, which suffers from significant pilot contamination due to the lack of orthogonality between the desired cell's pilots and the interfering cell's pilots [2], [6]. In particular, when the same PSs are reused in all the  $L$  cells, which is the worst case scenario [3], [5], [6], we have  $\mathbf{X}_l = \mathbf{X}$  for  $1 \leq l \leq L$ . Additionally, the PSs of the MSSs within a given cell are orthogonal<sup>1</sup> to each other, and we have  $\mathbf{X}^H \mathbf{X} = K \mathbf{I}_U$ . Now considering the LS CE (7), we arrive at [2]

$$\hat{\mathbf{H}}_{l^*,l^*}^{\text{LS}} = \mathbf{H}_{l^*,l^*} + \sum_{l=1, l \neq l^*}^L \mathbf{H}_{l^*,l} + \frac{1}{K \sqrt{p_r}} \mathbf{X}^H \mathbf{W}_{l^*}, \quad (9)$$

where the pilot contamination constituted by the second term in (9) gives rise to a significant estimation error to the desired CE, especially when the path loss coefficients between the interfering cells and the serving cell are similar to those within the serving cell [5].

For the LMMSE CE  $\hat{\mathbf{H}}_{l^*,l^*}^{\text{LMMSE}}$  of (8), again noting  $\mathbf{X}_l = \mathbf{X}$  for  $1 \leq l \leq L$  and  $\mathbf{X}^H \mathbf{X} = K \mathbf{I}_U$ , we have

$$\hat{\mathbf{H}}_{l^*,l^*}^{\text{LMMSE}} = \left( \frac{\sigma_w^2}{K p_r} \mathbf{R}_{l^*,l^*}^{-1} + \sum_{l=1}^L \mathbf{R}_{l^*,l} \mathbf{R}_{l^*,l^*}^{-1} \right)^{-1}$$

<sup>1</sup>Note that the PSs may be made orthogonal to each other on a per-carrier basis, where the  $K$ -length PSs of the users roaming in the same cell are orthogonal to each other, regardless whether they are orthogonal or nonorthogonal over the  $N$  subcarriers.

$$\times \left( \mathbf{H}_{l^*,l^*} + \sum_{l=1, l \neq l^*}^L \mathbf{H}_{l^*,l} + \frac{1}{K\sqrt{p_r}} \mathbf{X}^H \mathbf{W}_{l^*} \right). \quad (10)$$

The second term in (10) indicates that the pilot contamination imposes a significant estimation error on the desired CE. Let us investigate the asymptotic effects of the pilot contamination imposed on the LMMSE CE as the number of antennas  $Q$  tends to infinity. In order to determine the value of  $\sum_{l=1}^L \mathbf{R}_{l^*,l} \mathbf{R}_{l^*,l}^{-1}$  as  $Q \rightarrow \infty$ , we evaluate the following limit

$$\lim_{Q \rightarrow \infty} \sum_{l=1}^L \mathbf{R}_{l^*,l} \mathbf{R}_{l^*,l}^{-1} = \sum_{l=1}^L \left( \lim_{Q \rightarrow \infty} \frac{\mathbf{R}_{l^*,l}}{Q} \right) \left( \lim_{Q \rightarrow \infty} \frac{\mathbf{R}_{l^*,l^*}}{Q} \right)^{-1}, \quad (11)$$

which is valid provided that  $\lim_{Q \rightarrow \infty} (\mathbf{R}_{l^*,l^*}/Q)$  is non-singular. Based on Lemma 1 of [6] and by invoking  $\beta_{l^*,l,q,p}^u = \beta_{l^*,l}^u$  for  $1 \leq p \leq P$  and  $1 \leq q \leq Q$  again, we arrive at

$$\lim_{Q \rightarrow \infty} \frac{\mathbf{R}_{l^*,l}}{Q} = \mathbb{E} \left\{ \lim_{Q \rightarrow \infty} \frac{\mathbf{H}_{l^*,l} \mathbf{H}_{l^*,l}^H}{Q} \right\} = \text{diag} \{ [\beta_{l^*,l}^1, \beta_{l^*,l}^2, \dots, \beta_{l^*,l}^U] \}, \quad 1 \leq l \leq L. \quad (12)$$

Note furthermore that

$$\lim_{Q \rightarrow \infty} \frac{\sigma_w^2}{K p_r} \mathbf{R}_{l^*,l^*}^{-1} = \left( \lim_{Q \rightarrow \infty} \frac{\sigma_w^2}{Q K p_r} \right) \left( \lim_{Q \rightarrow \infty} \frac{\mathbf{R}_{l^*,l^*}}{Q} \right)^{-1} = 0. \quad (13)$$

Substituting (12) and (13) into (10) yields the following expression for  $\hat{\mathbf{H}}_{l^*,l^*}^{\text{LMMSE}}$  in the asymptotic case of  $Q \rightarrow \infty$

$$\hat{\mathbf{H}}_{l^*,l^*}^{\text{LMMSE}} = \text{diag} \left\{ \left[ \frac{\beta_{l^*,l^*}^1}{\sum_{l=1}^L \beta_{l^*,l}^1}, \frac{\beta_{l^*,l^*}^2}{\sum_{l=1}^L \beta_{l^*,l}^2}, \dots, \frac{\beta_{l^*,l^*}^U}{\sum_{l=1}^L \beta_{l^*,l}^U} \right] \right\} \times \left( \mathbf{H}_{l^*,l^*} + \sum_{l=1, l \neq l^*}^L \mathbf{H}_{l^*,l} + \frac{1}{K\sqrt{p_r}} \mathbf{X}^H \mathbf{W}_{l^*} \right). \quad (14)$$

Observe in (14) that the LMMSE CE strikes a tradeoff between the achievable estimation accuracy of the true FDCHTF matrix  $\mathbf{H}_{l^*,l^*}$  and the gravity of the pilot contamination imposed by the UL training of the MSs roaming in the adjacent cells. This beneficial trade-off is achieved at the expense of requiring all the path loss coefficients  $\beta_{l^*,l}^u$  for  $1 \leq u \leq U$  and  $1 \leq l \leq L$ .

### III. OUR PILOT CONTAMINATION ELIMINATION SCHEME

As shown in Section II-B, the CE relying on the UL PSs suffers from pilot contamination and the existing schemes [2], [5], [10] require the knowledge of all the channels' second-order statistics as well as the full cooperation of all the BSs to exchange their UL PSs. Additionally, having orthogonal PSs cannot be guaranteed for all of the MSs roaming in different cells for their UL training. Fortunately, the number of cells is limited compared to the number of MSs, hence it is possible to allocate orthogonal PSs to the BSs for their DL broadcasting. These BS-specific PSs can be exploited by the MSs for estimating their unique DL channels. By design, the DL training will not be contaminated by the adjacent cells, since orthogonal PSs have been allocated to the different BSs. Thus, the estimated DL channel

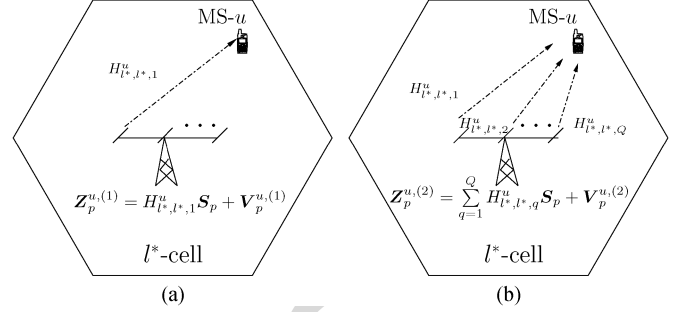


Fig. 2. Stage I: downlink training. During phase 1, the BS sends the DL pilot by its first antenna to the MSs, while during phase 2, the BS sends the pilot via its all of  $Q$  antennas to the MSs. Note that the MSs in the adjacent cells may also overhear the pilot, which is not shown in this figure for clarity. (a) Phase 1; (b) Phase 2.

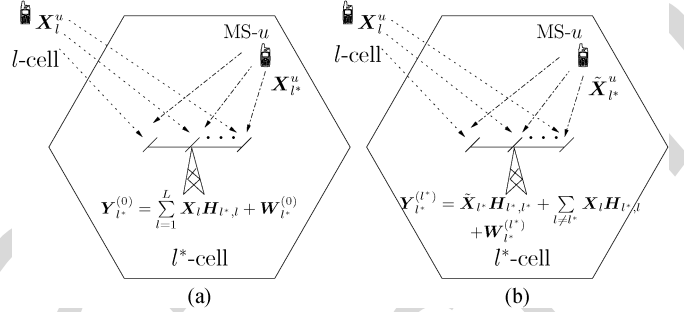


Fig. 3. Stage II: uplink training. During the initial phase, namely phase 0, all the MSs send the pre-assigned orthogonal PSs  $\mathbf{X}_l^u$  to their BSs. During the phase  $l^*$ ,  $1 \leq l^* \leq L$ , the MSs in cell  $l^*$  send their distinctive PSs  $\mathbf{X}_{l^*}^u$  to their BS, while the MSs in the other cells send the pre-assigned orthogonal PSs  $\mathbf{X}_l^u$  to their BSs. (a) Phase 0; (b) Phase  $l^*$ .

information can be encapsulated into the MSs's PSs for assisting the UL CE, as detailed below. Based on this philosophy, we propose a two-stage CE scheme consisting of the DL training and the appropriately scheduled UL training for the sake of eliminating the UL pilot contamination, which does not require any knowledge of the channels' second-order statistics.

#### A. The Proposed Pilot Assisted Channel Estimation

The proposed pilot contamination elimination scheme consists of a sophisticated amalgam of a DL and an UL training stage. Furthermore, the DL training contains two phases as illustrated in Fig. 2, while UL training contains  $(L+1)$  phases as illustrated in Fig. 3, where each phase occupies  $K$  OFDM symbol durations. Again, we assume that the channel is time-invariant for the duration of  $(L+3)K$  OFDM symbols, which allows the proposed scheme to acquire accurate estimates of the channel coefficients and then hence to eliminate the pilot contamination. More specifically, the operations of the DL and UL training are detailed as follows:

1) *The DL Training Stage:* It contains the two phases as illustrated in Fig. 2. We assume that there are  $L$  unique orthogonal OFDM PSs, one for each of the  $L$  cells, which is reasonable because the number of cells is far less than the total number of MSs roaming in all the cells. Specifically, the  $K$  pilots  $\mathbf{S}_{p,l} = [S_{p,l}[1] S_{p,l}[2] \dots S_{p,l}[K]]$  for cell  $l$  satisfy  $\mathbf{S}_{p,l} \mathbf{S}_{p,l}^H = K$  and  $\mathbf{S}_{p,l^*} \mathbf{S}_{p,l}^H = 0$  for  $1 \leq l, l^* \leq L$  and  $l \neq l^*$ . As seen in Fig. 2(a), during the first phase of the DL training, the  $l^*$ th BS broadcasts the  $K$  OFDM pilots  $\mathbf{S}_{p,l^*}$  using a single antenna, say the BS's

first antenna. The signal received by MS  $u$  in cell  $l^*$  can readily be expressed by<sup>2</sup>

$$\mathbf{Z}_p^{u,(1)} = \sqrt{p_f} H_{l^*,l^*,1}^u \mathbf{S}_p + \mathbf{V}_p^{u,(1)}, \quad 1 \leq u \leq U, \quad (15)$$

where  $\mathbf{Z}_p^{u,(1)} \in \mathbb{C}^{1 \times K}$ ,  $H_{l^*,l^*,1}^u$  represents the channel between the BS's 1st antenna and the MS  $u$ , while  $\mathbf{V}_p^{u,(1)} = [V_p^{u,(1)}[1] V_p^{u,(1)}[2] \dots V_p^{u,(1)}[K]]$  is the FD representation of the channel's AWGN. The unique, MS-specific DL channel information  $H_{l^*,l^*,1}^u$  is readily estimated using the LS estimator, yielding

$$\hat{H}_{l^*,l^*,1}^u = \frac{\mathbf{Z}_p^{u,(1)} \mathbf{S}_p^H}{K \sqrt{p_f}}, \quad 1 \leq u \leq U. \quad (16)$$

During the second phase of the DL training, as illustrated in Fig. 2(b), the BS activates all the DL transmit antennas to broadcast the same  $K$  OFDM pilots  $\mathbf{S}_p$  to MS  $u$ , where the received OFDM signal of MS  $u$  is given by

$$\mathbf{Z}_p^{u,(2)} = \sqrt{p_f} \sum_{q=1}^Q H_{l^*,l^*,q}^u \mathbf{S}_p + \mathbf{V}_p^{u,(2)}, \quad 1 \leq u \leq U, \quad (17)$$

where  $\mathbf{Z}_p^{u,(2)} \in \mathbb{C}^{1 \times K}$ ,  $H_{l^*,l^*,q}^u$  denotes the channel between the BS's  $q$ th antenna and the MS  $u$ , while  $\mathbf{V}_p^{u,(2)} = [V_p^{u,(2)}[1] V_p^{u,(2)}[2] \dots V_p^{u,(2)}[K]]$  is the FD representation of the channel's AWGN. Thus, MS  $u$  can estimate the sum of the channels, namely,  $H_{l^*,l^*}^{u,\text{sum}} = \sum_{q=1}^Q H_{l^*,l^*,q}^u$  for all the  $Q$  links between the serving BS's DL transmit antennas and itself, using the LS estimate of

$$\hat{H}_{l^*,l^*}^{u,\text{sum}} = \frac{\mathbf{Z}_p^{u,(2)} \mathbf{S}_p^H}{K \sqrt{p_f}}, \quad 1 \leq u \leq U. \quad (18)$$

2) *The UL Training Stage:* It consists of the  $(L+1)$  scheduled phases as depicted in Fig. 3. We assume the worst case scenario of having the same pre-assigned orthogonal PS matrix  $\mathbf{X} \in \mathbb{C}^{K \times U}$  reused in every cells, namely  $\mathbf{X}_l = \mathbf{X}$  for  $1 \leq l \leq L$  and  $\mathbf{X}^H \mathbf{X} = K \mathbf{I}_U$ , where we have  $\mathbf{X} = [(\mathbf{X}^1)^T (\mathbf{X}^2)^T \dots (\mathbf{X}^U)^T]^T$  and  $\mathbf{X}^u \in \mathbb{C}^{1 \times K}$  for  $1 \leq u \leq U$ . As illustrated in Fig. 3(a), during the initial phase 0 of the UL training, the MSs roaming in all the cells simultaneously transmit their pre-assigned orthogonal PSs  $\mathbf{X}_l^u = \mathbf{X}^u$  for  $1 \leq u \leq U$  and  $1 \leq l \leq L$  in the ULs to their BSs. The composite signal received during phase 0 at the  $l^*$ -th BS can be expressed according to (6) as

$$\mathbf{Y}_{l^*}^{(0)} = \sqrt{p_r} \mathbf{X} \mathbf{H}_{l^*,l^*} + \underbrace{\sqrt{p_r} \mathbf{X} \sum_{l=1, l \neq l^*}^L \mathbf{H}_{l^*,l}}_{\text{Pilot contamination}} + \mathbf{W}_{l^*}^{(0)}. \quad (19)$$

During the  $l^*$  phase of the UL training, where  $1 \leq l^* \leq L$ , as shown in Fig. 3(b), the MSs in cell  $l^*$  transmit their own specifically predistorted UL PSs  $\tilde{\mathbf{X}}_{l^*}^u$  given by

$$\tilde{\mathbf{X}}_{l^*}^u = \frac{\hat{H}_{l^*,l^*}^{u,\text{sum}}}{\hat{H}_{l^*,l^*,1}^u} \mathbf{X}^u, \quad 1 \leq u \leq U, \quad (20)$$

<sup>2</sup>The serving cell index  $l^*$  is dropped from the transmitted and received pilot signals for notational simplicity.

to the  $l^*$ -th BS. As  $\tilde{\mathbf{X}}_{l^*}^u$  encapsulates the DL-FDCHTF information  $\hat{H}_{l^*,l^*}^{u,\text{sum}} / \hat{H}_{l^*,l^*,1}^u$  estimated by MS  $u$  in cell  $l^*$  during the DL training, it is distinct to this MS. At the same time, the MSs roaming in all the other cells simultaneously transmit their pre-assigned orthogonal PSs  $\mathbf{X}_l^u = \mathbf{X}^u$  to their BSs, where  $1 \leq u \leq U$ ,  $1 \leq l \leq L$  and  $l \neq l^*$ . As a result, the UL signal received by the  $l^*$ -th BS's receiver is readily expressed as

$$\begin{aligned} \mathbf{Y}_{l^*}^{(l^*)} &= \sqrt{p_r} \tilde{\mathbf{X}}_{l^*} \mathbf{H}_{l^*,l^*} + \underbrace{\sqrt{p_r} \mathbf{X} \sum_{l=1, l \neq l^*}^L \mathbf{H}_{l^*,l}}_{\text{Pilot contamination}} + \mathbf{W}_{l^*}^{(l^*)}, \\ &= \sqrt{p_r} \mathbf{X} \mathbf{B}_{l^*,l^*} \mathbf{H}_{l^*,l^*} + \sqrt{p_r} \mathbf{X} \sum_{l=1, l \neq l^*}^L \mathbf{H}_{l^*,l} + \mathbf{W}_{l^*}^{(l^*)}, \end{aligned} \quad (21)$$

where  $\tilde{\mathbf{X}}_{l^*} = [(\tilde{\mathbf{X}}_{l^*}^1)^T (\tilde{\mathbf{X}}_{l^*}^2)^T \dots (\tilde{\mathbf{X}}_{l^*}^U)^T]^T = \mathbf{X} \mathbf{B}_{l^*,l^*}$  and

$$\mathbf{B}_{l^*,l^*} = \text{diag} \left\{ \left[ \frac{\hat{H}_{l^*,l^*}^{1,\text{sum}}}{\hat{H}_{l^*,l^*,1}^1} \frac{\hat{H}_{l^*,l^*}^{2,\text{sum}}}{\hat{H}_{l^*,l^*,1}^2} \dots \frac{\hat{H}_{l^*,l^*}^{U,\text{sum}}}{\hat{H}_{l^*,l^*,1}^U} \right] \right\}. \quad (22)$$

From (19) and (21), we have respectively

$$\tilde{\mathbf{Y}}_{l^*}^{(0)} = \frac{1}{K \sqrt{p_r}} \mathbf{X}^H \mathbf{Y}_{l^*}^{(0)} = \mathbf{H}_{l^*,l^*} + \sum_{l=1, l \neq l^*}^L \mathbf{H}_{l^*,l} + \tilde{\mathbf{W}}_{l^*}^{(0)}, \quad (23)$$

$$\begin{aligned} \tilde{\mathbf{Y}}_{l^*}^{(l^*)} &= \frac{1}{K \sqrt{p_r}} \mathbf{X}^H \mathbf{Y}_{l^*}^{(l^*)} \\ &= \mathbf{B}_{l^*,l^*} \mathbf{H}_{l^*,l^*} + \sum_{l=1, l \neq l^*}^L \mathbf{H}_{l^*,l} + \tilde{\mathbf{W}}_{l^*}^{(l^*)}, \end{aligned} \quad (24)$$

where  $\tilde{\mathbf{W}}_{l^*}^{(0)} = (1/K \sqrt{p_r}) \mathbf{X}^H \mathbf{W}_{l^*}^{(0)}$  and  $\tilde{\mathbf{W}}_{l^*}^{(l^*)} = (1/K \sqrt{p_r}) \mathbf{X}^H \mathbf{W}_{l^*}^{(l^*)}$ . Furthermore, the power of both  $\tilde{\mathbf{W}}_{l^*}^{(0)}$  and  $\tilde{\mathbf{W}}_{l^*}^{(l^*)}$  is  $\sigma_w^2/p_r$ . Let  $\tilde{\mathbf{Y}}_{l^*} = \tilde{\mathbf{Y}}_{l^*}^{(l^*)} - \tilde{\mathbf{Y}}_{l^*}^{(0)}$  and  $\tilde{\mathbf{W}}_{l^*} = \tilde{\mathbf{W}}_{l^*}^{(l^*)} - \tilde{\mathbf{W}}_{l^*}^{(0)}$ , where the power of  $\tilde{\mathbf{W}}_{l^*}$  is  $2\sigma_w^2/p_r$ . From (23) and (24), we readily arrive at

$$\tilde{\mathbf{Y}}_{l^*} = \mathbf{B}_{l^*,l^*} \mathbf{H}_{l^*,l^*} - \mathbf{H}_{l^*,l^*} + \tilde{\mathbf{W}}_{l^*}. \quad (25)$$

Observe in (25) that the pilot contamination is completely eliminated.

In order to extract the estimates of the MS-specific FDCHTFs  $\hat{H}_{l^*,l^*,q}^u$  for  $1 \leq u \leq U$  and  $1 \leq q \leq Q$ , we expand (25). Specifically, we do not distinguish the DL and UL FDCHTF estimates, since we assume that they are identical, and we substitute  $H_{l^*,l^*,q}^u$  in (25) by their estimates  $\hat{H}_{l^*,l^*,q}^u$  to express it in the element-based form shown in (26) at the top of the page. (See equation at bottom of next page) Observe that at the right-hand side of (26) all the entries in the  $u$ -th row have the common factor of  $(\hat{H}_{l^*,l^*}^{u,\text{sum}} / \hat{H}_{l^*,l^*,1}^u) - 1$ . Noting  $\hat{H}_{l^*,l^*}^{u,\text{sum}} = \sum_{q=1}^Q \hat{H}_{l^*,l^*,q}^u$  yields

$$\frac{\hat{H}_{l^*,l^*}^{u,\text{sum}}}{\hat{H}_{l^*,l^*,1}^u} - 1 = \frac{1}{\hat{H}_{l^*,l^*,1}^u} \sum_{q=2}^Q \hat{H}_{l^*,l^*,q}^u. \quad (27)$$

Assuming that  $(1/\hat{H}_{l^*,l^*,1}^u) \sum_{q=2}^Q \hat{H}_{l^*,l^*,q}^u \neq 0$  for  $1 \leq u \leq U$ , we have

$$\frac{\check{Y}_{l^*,1}^u}{\check{Y}_{l^*,1}^u} = \frac{\hat{H}_{l^*,l^*,q}^u}{\hat{H}_{l^*,l^*,1}^u}, \quad 2 \leq q \leq Q, \quad 1 \leq u \leq U. \quad (28)$$

With the aid of (28), (26) can be rewritten as

$$\begin{bmatrix} \check{Y}_{l^*,1}^1 & \check{Y}_{l^*,2}^1 & \cdots & \check{Y}_{l^*,Q}^1 \\ \check{Y}_{l^*,1}^2 & \check{Y}_{l^*,2}^2 & \cdots & \check{Y}_{l^*,Q}^2 \\ \vdots & \vdots & \vdots & \vdots \\ \check{Y}_{l^*,1}^U & \check{Y}_{l^*,2}^U & \cdots & \check{Y}_{l^*,Q}^U \end{bmatrix} = \begin{bmatrix} \sum_{q=2}^Q \hat{H}_{l^*,l^*,q}^1 & \frac{\hat{H}_{l^*,l^*,2}^1}{\hat{H}_{l^*,l^*,1}^1} & \cdots & \frac{\hat{H}_{l^*,l^*,Q}^1}{\hat{H}_{l^*,l^*,1}^1} \\ \sum_{q=2}^Q \hat{H}_{l^*,l^*,q}^2 & \frac{\hat{H}_{l^*,l^*,2}^2}{\hat{H}_{l^*,l^*,1}^2} & \cdots & \frac{\hat{H}_{l^*,l^*,Q}^2}{\hat{H}_{l^*,l^*,1}^2} \\ \vdots & \vdots & \vdots & \vdots \\ \sum_{q=2}^Q \hat{H}_{l^*,l^*,q}^U & \frac{\hat{H}_{l^*,l^*,2}^U}{\hat{H}_{l^*,l^*,1}^U} & \cdots & \frac{\hat{H}_{l^*,l^*,Q}^U}{\hat{H}_{l^*,l^*,1}^U} \end{bmatrix}.$$

By summing the last  $(Q-1)$  elements in each row for both the left-hand and right-hand sides of (29), we arrive at

$$\frac{\sum_{q=2}^Q \check{Y}_{l^*,q}^u}{\check{Y}_{l^*,1}^u} = \frac{\sum_{q=2}^Q \hat{H}_{l^*,l^*,q}^u}{\hat{H}_{l^*,l^*,1}^u}, \quad 1 \leq u \leq U. \quad (30)$$

Noting  $\check{Y}_{l^*,1}^u = \sum_{q=2}^Q \hat{H}_{l^*,l^*,q}^u$  for  $1 \leq u \leq U$  yields the estimates of  $\hat{H}_{l^*,l^*,1}^u$

$$\hat{H}_{l^*,l^*,1}^u = \frac{\check{Y}_{l^*,1}^u \check{Y}_{l^*,1}^u}{\sum_{q=2}^Q \check{Y}_{l^*,q}^u}, \quad 1 \leq u \leq U. \quad (31)$$

By utilizing (28) and (33), we obtain the estimates of  $H_{l^*,l^*,q}^u$

$$\hat{H}_{l^*,l^*,q}^u = \frac{\check{Y}_{l^*,1}^u \check{Y}_{l^*,q}^u}{\sum_{q=2}^Q \check{Y}_{l^*,q}^u}, \quad 1 \leq u \leq U, \quad 2 \leq q \leq Q. \quad (32)$$

*Remark 1:* Our scheme completely eliminates the pilot contamination in the estimated FDCHTFs of (31) and (32) under the worst-case UL training scenario of reusing the same orthogonal PS matrix in all the  $L$  cells. This is achieved at the cost of expanding the training session to the  $L$  appropriately scheduled periods,—one for each cell—to estimate its FDCHTF matrix. Therefore, our scheme increases the training overhead by a factor of  $L$ . However, our scheme does not require any information exchange between the BSs and, most importantly, it does not impose the unrealistic assumption of knowing the channels' second-order statistics. Consequently, unlike many existing schemes [2], [5], [10], our channel estimator has significant advantages in practical implementation.

The CE scheme proposed for the network of  $L$  cells is summarized in Algorithm 1.

---

**Algorithm 1 Pilot contamination elimination scheme.**

---

- **Stage I Simultaneous DL training:** as illustrated in Fig. 2, it consists of the two phases, a single-antenna and a multiple-antenna based channel estimation phases, during which each cell's BS simultaneously broadcasts PSs to its MSs for them to estimate their DL-FDCHTFs, as given in (16) and (18), respectively.
  - **Stage II Scheduled UL training:** as illustrated in Fig. 3, it consists of  $(L+1)$  phases.
    - (a) During the initial phase 0 of Fig. 3(a), all the MSs of all cells simultaneously transmit the pre-assigned PSs to their BSs.
    - (b) During the phase  $l^*$  of Fig. 3(b), where  $1 \leq l^* \leq L$ , the MSs of cell  $l^*$  transmit their appropriately predistorted PSs based on their estimated DL-FDCHTFs to the  $l^*$ -th BS, while the MSs of all the other cells again transmit the same pre-assigned PSs to their BSs. Given the signals received during phase 0 and phase  $l^*$ , the  $l^*$ -th BS extracts the FDCHTFs of its MSs, as seen in (31) and (32), by eliminating the pilot contamination imposed by the other cells.
- 

$$\begin{bmatrix} \check{Y}_{l^*,1}^1 & \check{Y}_{l^*,2}^1 & \cdots & \check{Y}_{l^*,Q}^1 \\ \check{Y}_{l^*,1}^2 & \check{Y}_{l^*,2}^2 & \cdots & \check{Y}_{l^*,Q}^2 \\ \vdots & \vdots & \vdots & \vdots \\ \check{Y}_{l^*,1}^U & \check{Y}_{l^*,2}^U & \cdots & \check{Y}_{l^*,Q}^U \end{bmatrix} = \begin{bmatrix} \hat{H}_{l^*,l^*,1}^1 \left( \frac{\hat{H}_{l^*,l^*,1}^{1,\text{sum}}}{\hat{H}_{l^*,l^*,1}^1} - 1 \right) & \hat{H}_{l^*,l^*,2}^1 \left( \frac{\hat{H}_{l^*,l^*,1}^{1,\text{sum}}}{\hat{H}_{l^*,l^*,1}^1} - 1 \right) & \cdots & \hat{H}_{l^*,l^*,Q}^1 \left( \frac{\hat{H}_{l^*,l^*,1}^{1,\text{sum}}}{\hat{H}_{l^*,l^*,1}^1} - 1 \right) \\ \hat{H}_{l^*,l^*,1}^2 \left( \frac{\hat{H}_{l^*,l^*,1}^{2,\text{sum}}}{\hat{H}_{l^*,l^*,1}^2} - 1 \right) & \hat{H}_{l^*,l^*,2}^2 \left( \frac{\hat{H}_{l^*,l^*,1}^{2,\text{sum}}}{\hat{H}_{l^*,l^*,1}^2} - 1 \right) & \cdots & \hat{H}_{l^*,l^*,Q}^2 \left( \frac{\hat{H}_{l^*,l^*,1}^{2,\text{sum}}}{\hat{H}_{l^*,l^*,1}^2} - 1 \right) \\ \vdots & \vdots & \vdots & \vdots \\ \hat{H}_{l^*,l^*,1}^U \left( \frac{\hat{H}_{l^*,l^*,1}^{U,\text{sum}}}{\hat{H}_{l^*,l^*,1}^U} - 1 \right) & \hat{H}_{l^*,l^*,2}^U \left( \frac{\hat{H}_{l^*,l^*,1}^{U,\text{sum}}}{\hat{H}_{l^*,l^*,1}^U} - 1 \right) & \cdots & \hat{H}_{l^*,l^*,Q}^U \left( \frac{\hat{H}_{l^*,l^*,1}^{U,\text{sum}}}{\hat{H}_{l^*,l^*,1}^U} - 1 \right) \end{bmatrix}. \quad (26)$$

### B. Mean and Variance of the Proposed Estimator

From (15) and (17), we can express  $\tilde{Y}_{l^*,q}^u$  in (25) as

$$\tilde{Y}_{l^*,q}^u = H_{l^*,l^*,q}^u \underbrace{\left( \frac{\sum_{q=1}^Q H_{l^*,l^*,q}^u + \tilde{V}_p^{u,(2)}}{H_{l^*,l^*,1}^u + \tilde{V}_p^{u,(1)}} - 1 \right)}_{\textcircled{A}} + \tilde{W}_{l^*,q}^u, \quad (33)$$

where  $\tilde{V}_p^{u,(1)} = \mathbf{V}_p^{u,(1)} \mathbf{S}_p^H / K \sqrt{p_f}$ ,  $\tilde{V}_p^{u,(2)} = \mathbf{V}_p^{u,(2)} \mathbf{S}_p^H / K \sqrt{p_f}$ , and the power of both  $\tilde{V}_p^{u,(1)}$  and  $\tilde{V}_p^{u,(2)}$  is  $\sigma_v^2 / p_f$ . The complex-valued random variables  $\tilde{V}_p^{u,(1)}$ ,  $\tilde{V}_p^{u,(2)}$  and  $\tilde{W}_{l^*,q}^u$  in (33) all obey the Gaussian distribution and they are independent of each other. As  $\tilde{Y}_{l^*,q}^u$  is a nonlinear composite of  $\tilde{V}_p^{u,(1)}$ ,  $\tilde{V}_p^{u,(2)}$  and  $\tilde{W}_{l^*,q}^u$ , the statistical analysis of the proposed estimator is challenging. We concentrate on the expectation and variance of the proposed estimator  $\hat{H}_{l^*,l^*,q}^u$ .

Invoking the expression of  $\tilde{Y}_{l^*,q}^u$  in (33), the numerator and denominator in (31) and (32) can be rewritten as follows

$$\begin{aligned} \sum_{q=2}^Q \tilde{Y}_{l^*,q}^u &= \underbrace{\left( \sum_{q=2}^Q H_{l^*,l^*,q}^u \right)}_{\textcircled{B}} + \left( \sum_{q=2}^Q \tilde{W}_{l^*,q}^u \right) \\ &= \textcircled{B} \textcircled{A} + \tilde{W}_{l^*,\text{sum}}^u, \end{aligned} \quad (34)$$

$$\tilde{Y}_{l^*,1}^u \tilde{Y}_{l^*,1}^u = (H_{l^*,l^*,1}^u)^2 \textcircled{A}^2 + 2(H_{l^*,l^*,1}^u \tilde{W}_{l^*,1}^u) \textcircled{A} + (\tilde{W}_{l^*,1}^u)^2, \quad (35)$$

$$\begin{aligned} \tilde{Y}_{l^*,1}^u \tilde{Y}_{l^*,q}^u &= (H_{l^*,l^*,1}^u H_{l^*,l^*,q}^u) \textcircled{A}^2 + (H_{l^*,l^*,q}^u \tilde{W}_{l^*,1}^u) \textcircled{A} \\ &\quad + (H_{l^*,l^*,1}^u \tilde{W}_{l^*,q}^u) \textcircled{A} + (\tilde{W}_{l^*,1}^u \tilde{W}_{l^*,q}^u), \end{aligned} \quad (36)$$

where  $\tilde{W}_{l^*,\text{sum}}^u = \sum_{q=2}^Q \tilde{W}_{l^*,q}^u$ . The expectation of  $\hat{H}_{l^*,l^*,1}^u$  is with respect to  $\tilde{W}_{l^*,1}^u$ ,  $\tilde{W}_{l^*,q}^u$ ,  $\tilde{V}_p^{u,(1)}$  and  $\tilde{V}_p^{u,(2)}$ . Therefore, we can express the expectation as follows [21, p. 117]

$$\begin{aligned} &\mathbb{E}_{\tilde{V}_p^{u,(2)}, \tilde{V}_p^{u,(1)}, \tilde{W}_{l^*,q}^u, \tilde{W}_{l^*,1}^u} \left\{ \hat{H}_{l^*,l^*,1}^u \right\} \\ &= \mathbb{E}_{\tilde{V}_p^{u,(2)}} \left\{ \mathbb{E}_{\tilde{V}_p^{u,(1)}} \left\{ \mathbb{E}_{\tilde{W}_{l^*,q}^u} \left\{ \mathbb{E}_{\tilde{W}_{l^*,1}^u} \left\{ \frac{\tilde{Y}_{l^*,1}^u \tilde{Y}_{l^*,1}^u}{\sum_{q=2}^Q \tilde{Y}_{l^*,q}^u} \right\} \right\} \right\} \right\} \\ &= \mathbb{E}_{\tilde{V}_p^{u,(2)}} \left\{ \mathbb{E}_{\tilde{V}_p^{u,(1)}} \left\{ \mathbb{E}_{\tilde{W}_{l^*,q}^u} \left\{ \frac{(H_{l^*,l^*,1}^u)^2 \textcircled{A}^2}{\textcircled{B} \textcircled{A} + \tilde{W}_{l^*,\text{sum}}^u} \right\} \right\} \right\}. \end{aligned} \quad (37)$$

Having arrived at  $(H_{l^*,l^*,1}^u)^2 \textcircled{A}^2 / \textcircled{B} \textcircled{A} + \tilde{W}_{l^*,\text{sum}}^u$  by evaluating the expectation of  $\tilde{Y}_{l^*,1}^u \tilde{Y}_{l^*,1}^u / \sum_{q=2}^Q \tilde{Y}_{l^*,q}^u$  with respect to the random variables  $\tilde{W}_{l^*,1}^u$  and to compute the expectation of  $(H_{l^*,l^*,1}^u)^2 \textcircled{A}^2 / \textcircled{B} \textcircled{A} + \tilde{W}_{l^*,\text{sum}}^u$  with respect to the random variables  $\tilde{W}_{l^*,q}^u$ , we have the following theorem.

**Theorem 1:** If a complex-valued random variable  $\gamma = a + bj = \rho e^{-j\phi}$ ,  $\rho \geq 0$ , has a zero mean, then the  $n$ th power of the function  $\gamma$ , namely  $\gamma^n$ , where  $n$  is a natural number, also has a zero mean.

*Proof:* Because  $\mathbb{E}\{\gamma\} = 0$ , we have

$$\mathbb{E}\{e^{-j\phi}\} = \int_0^{2\pi} e^{-j\phi} d\phi = 0. \quad (38)$$

Then, we can compute the mean of  $\gamma^n$  as follows

$$\begin{aligned} \mathbb{E}\{\gamma^n\} &= \mathbb{E}\{\rho^n\} \int_0^{2\pi} e^{-jn\phi} d\phi \\ &\stackrel{\varphi=n\phi}{=} \mathbb{E}\{\rho^n\} \frac{1}{n} \sum_{n'=1}^n \int_{(n'-1)2\pi}^{n'2\pi} e^{-j\varphi} d\varphi = 0. \end{aligned} \quad (39)$$

This completes the proof.  $\blacksquare$

According to the classic Taylor series theory of complex-valued random variables [22, p. 189], under the condition of  $|\tilde{W}_{l^*,\text{sum}}^u / \textcircled{B} \textcircled{A}| < 1$  with probability one,<sup>3</sup> the function of  $1 / \textcircled{B} \textcircled{A} + \tilde{W}_{l^*,\text{sum}}^u$  has the power series representation given by

$$\frac{1}{\textcircled{B} \textcircled{A} + \tilde{W}_{l^*,\text{sum}}^u} = \frac{1}{\textcircled{B} \textcircled{A}} \left( 1 + \sum_{n=1}^{\infty} (-1)^n \left( \frac{\tilde{W}_{l^*,\text{sum}}^u}{\textcircled{B} \textcircled{A}} \right)^n \right). \quad (40)$$

The series at the right-hand side of (40) is understood to converge in probability. Using Theorem 1 and (40), we obtain the expectation of  $\hat{H}_{l^*,l^*,1}^u$  as follows

$$\mathbb{E}_{\tilde{V}_p^{u,(2)}, \tilde{V}_p^{u,(1)}, \tilde{W}_{l^*,q}^u, \tilde{W}_{l^*,1}^u} \left\{ \hat{H}_{l^*,l^*,1}^u \right\} = H_{l^*,l^*,1}^u. \quad (41)$$

In the same way, the expectation of  $\hat{H}_{l^*,l^*,q}^u$ ,  $2 \leq q \leq Q$ , is given by

$$\mathbb{E}_{\tilde{V}_p^{u,(2)}, \tilde{V}_p^{u,(1)}, \tilde{W}_{l^*,q}^u, \tilde{W}_{l^*,1}^u} \left\{ \hat{H}_{l^*,l^*,q}^u \right\} = H_{l^*,l^*,q}^u. \quad (42)$$

From (41) and (42), we conclude that the proposed estimator is an unbiased one.

To provide an approximate expression for the proposed estimator's variance, we have the following approximation.

**Approximation 1:** Given that  $A = A_R + jA_I$  is a complex-valued constant and  $\zeta = \zeta_R + j\zeta_I$  is a complex-valued random variable obeying  $\mathcal{CN}(0, \sigma_\zeta^2)$  as well as assuming that

$$\Pr \left\{ \frac{A_R \zeta_R + A_I \zeta_I + |\zeta|^2}{|A|^2} < 1 \right\} = 1, \quad (43)$$

we have the following approximate formula

$$\mathbb{E} \left\{ \frac{1}{(A + \zeta)(A^\dagger + \zeta^\dagger)} \right\} \approx \frac{1}{|A|^2} + \frac{1}{|A|^4} \sigma_\zeta^2. \quad (44)$$

*Proof:* We use Taylor series theory of a real-valued random variable [23, p. 134] to express  $\mathbb{E}\{1/(A + \zeta)(A^\dagger + \zeta^\dagger)\}$  as

$$\begin{aligned} &\mathbb{E} \left\{ \frac{1}{(A + \zeta)(A^\dagger + \zeta^\dagger)} \right\} \\ &= \frac{1}{|A|^2} \mathbb{E} \left\{ \left( 1 + \sum_{n=1}^{\infty} (-1)^n \left( \frac{A_R \zeta_R + A_I \zeta_I + |\zeta|^2}{|A|^2} \right)^n \right) \right\} \\ &\approx \frac{1}{|A|^2} + \frac{1}{|A|^4} \sigma_\zeta^2, \end{aligned} \quad (45)$$

<sup>3</sup>Note that  $\Pr(|\tilde{W}_{l^*,\text{sum}}^u / \textcircled{B} \textcircled{A}| \geq 1) \approx e^{-((Q-1)\beta_{l^*,l^*} / \sigma_w^2)} \rightarrow 0$  as  $Q \rightarrow \infty$ . Therefore,  $\Pr(|\tilde{W}_{l^*,\text{sum}}^u / \textcircled{B} \textcircled{A}| < 1) = 1$  can be justified.

where the approximation arises by dropping the high-order terms  $(1/|A|^2)\mathbb{E}\{(A_R\zeta_R + A_I\zeta_I + |\zeta|^2/|A|^2)^n\}$  for  $n \geq 2$ . This completes the proof. ■

Using the approximate formula of (45) and Taylor series theory, after some further manipulations, we have the following approximate expression of the upper bound for the variance of  $\hat{H}_{l^*,1}^u$

$$\text{var}\left\{\hat{H}_{l^*,1}^u\right\} \lesssim \frac{2\left|H_{l^*,1}^u\right|^2\sigma_v^2}{p_f|\mathbb{B}|^2} + \frac{\sigma_v^2}{p_f} + \frac{(2(Q-1)\sigma_w^2)\left|H_{l^*,1}^u\right|^4}{p_r|\mathbb{B}|^4} + \frac{8\left|H_{l^*,1}^u\right|^2\sigma_w^2}{p_r|\mathbb{B}|^2}, \quad (46)$$

where  $1 \leq u \leq U$ . Similarly, for  $1 \leq u \leq U$  and  $2 \leq q \leq Q$ ,

$$\begin{aligned} \text{var}\left\{\hat{H}_{l^*,q}^u\right\} &\lesssim \frac{2\left|H_{l^*,q}^u\right|^2\sigma_v^2}{p_f|\mathbb{B}|^2} + \frac{\left|H_{l^*,q}^u\right|^2\sigma_v^2}{p_f\left|H_{l^*,1}^u\right|^2} \\ &+ \frac{2\left|H_{l^*,q}^u\right|^2\sigma_w^2}{p_r|\mathbb{B}|^2} + \frac{2\left|H_{l^*,1}^u\right|^2\sigma_w^2}{p_r|\mathbb{B}|^2} \\ &+ \frac{2\left|H_{l^*,1}^u\right|^2\left|H_{l^*,q}^u\right|^2(Q-1)\sigma_w^2}{p_r|\mathbb{B}|^4}. \end{aligned} \quad (47)$$

We point out again that  $\mathbb{B} = \sum_{q=2}^Q H_{l^*,q}^u$ .

### C. Downlink Transmission

By exploiting the reciprocity of the TDD UL and DL channels [6], [10], [15], the BS is capable of acquiring the DL CE by simply assuming that it is the same as the UL CE generated with the aid of the DL-pilots and the UL-pilots, as discussed in Section III-A and Section III-B, respectively. Then the TP matrix  $\mathbf{A}_l \in \mathbb{C}^{Q \times U}$  can be designed using the zero-forcing (ZF), the MMSE, the minimum bit error rate or the nonlinear vector precoding criteria [16]–[20]. Let the  $k$ -th information-bearing OFDM symbols transmitted to the  $U$  users of cell  $l$  be  $\mathbf{S}_l[k] = [S_l^1[k]S_l^2[k] \cdots S_l^U[k]]$ . We assume that the TP scheme ensures that  $\mathbb{E}[\mathbf{S}_l[k]] = \mathbf{0}$ ,  $\mathbb{E}[\mathbf{S}_l[k]\mathbf{S}_l^H[k]] = \mathbf{I}$ , and  $\text{tr}\{\mathbf{A}_l^H\mathbf{A}_l\} = 1$ , which implies that the average power constraint at the BS is satisfied [5], [10].

The  $l$ -th cell's BS transmits its TP signal of  $\mathbf{A}_l\mathbf{S}_l^T[k]$  in the DL to the  $U$  supported MSs. Therefore, the noisy signal  $\mathbf{Z}_{l^*}[k] \in \mathbb{C}^{1 \times U}$  received by the  $U$  MSs in cell  $l^*$  is constituted by the superimposed DL transmissions of the  $L$  cells' BSs plus the channel's AWGN, which is expressed as

$$\begin{aligned} \mathbf{Z}_{l^*}^T[k] &= \underbrace{\sqrt{p_f}\mathbf{H}_{l^*,l^*}\mathbf{A}_{l^*}\mathbf{S}_{l^*}^T[k]}_{\text{Desired term}} \\ &+ \underbrace{\sqrt{p_f}\sum_{l=1, l \neq l^*}^L \mathbf{H}_{l^*,l}\mathbf{A}_l\mathbf{S}_l^T[k]}_{\text{Inter-cell interference}} + \mathbf{V}_{l^*}^T[k], \end{aligned} \quad (48)$$

where  $p_f$  is the average transmit power of the BS and  $\mathbf{V}_{l^*} \in \mathbb{C}^{1 \times U}$  is the FD representation of the DL channel AWGN vector, whose elements have the power of  $\sigma_v^2$ . Thus the signal received by the  $u^*$ -th MS in cell  $l^*$  is given by

$$Z_{l^*}^{u^*}[k] = \sqrt{p_f}\sum_{l=1}^L\sum_{u=1}^U \mathbf{H}_{l^*,l}^{u^*}(\mathbf{A}_l^u)^T S_l^u[k] + V_{l^*}^{u^*}[k], \quad (49)$$

where  $\mathbf{H}_{l^*,l}^{u^*}$  is the  $u^*$ -th row of  $\mathbf{H}_{l^*,l}$ ,  $(\mathbf{A}_l^u)^T$  is the  $u$ -th column of  $\mathbf{A}_l$  and  $V_{l^*}^{u^*}[k]$  is the  $u^*$ -th element of  $\mathbf{V}_{l^*}[k]$ .

### D. Achievable Throughput Analysis

To further investigate the benefits of eliminating the pilot contamination imposed by the adjacent BSs on the CE, we derive the lower bound of the achievable DL throughput based on the ZF precoding matrix given by [6]

$$\mathbf{A}_l = \frac{\tilde{\mathbf{A}}_l}{\sqrt{\text{tr}\{\tilde{\mathbf{A}}_l\tilde{\mathbf{A}}_l^H\}}} \text{ with } \tilde{\mathbf{A}}_l = \hat{\mathbf{H}}_{l,l}^H(\hat{\mathbf{H}}_{l,l}\hat{\mathbf{H}}_{l,l}^H)^{-1}. \quad (50)$$

By denoting the DL data to be transmitted to the  $U$  users in cell  $l$  as  $\mathbf{S}_l = [S_l^1S_l^2 \cdots S_l^U]$ , we can rewrite the signal (49) received by the  $u^*$ -th user in the  $l^*$ -th cell as follows

$$\begin{aligned} Z_{l^*}^{u^*} &= \sqrt{p_f}\mathbb{E}\left\{\mathbf{H}_{l^*,l^*}^{u^*}(\mathbf{A}_{l^*}^{u^*})^T\right\}S_{l^*}^{u^*} \\ &+ \sqrt{p_f}\left(\mathbf{H}_{l^*,l^*}^{u^*}(\mathbf{A}_{l^*}^{u^*})^T - \mathbb{E}\left\{\mathbf{H}_{l^*,l^*}^{u^*}(\mathbf{A}_{l^*}^{u^*})^T\right\}\right)S_{l^*}^{u^*} \\ &+ \underbrace{\sqrt{p_f}\sum_{u=1, u \neq u^*}^U \mathbf{H}_{l^*,l^*}^{u^*}(\mathbf{A}_{l^*}^u)^T S_{l^*}^u}_{\text{Intra-cell interference}} \\ &+ \underbrace{\sqrt{p_f}\sum_{l=1, l \neq l^*}^L\sum_{u=1}^U \mathbf{H}_{l^*,l}^{u^*}(\mathbf{A}_l^u)^T S_l^u + \mathbf{V}_{l^*}^{u^*}}_{\text{Inter-cell interference}}. \end{aligned} \quad (51)$$

The achievable rate of the DL is the function of the signal-to-interference-plus-noise ratio (SINR) given by

$$R_{l^*}^{u^*} = \log_2\left(1 + \text{SINR}_{l^*}^{u^*}\right), \quad (52)$$

where the SINR of the  $u^*$ -th user in the  $l^*$ -th cell is given in (53) at the top of the next page. (See equation at bottom of page)

Since the objective of designing the ZF precoding matrix  $\mathbf{A}_{l^*}$  is to minimize the interference imposed by the MSs within the same serving cell, the residual intra-cell interference which is the third term in (51) is significantly lower than the inter-cell interference of the fourth term. Therefore, we may omit the term  $\sum_{u=1, u \neq u^*}^U \mathbb{E}\{|\mathbf{H}_{l^*,l^*}^{u^*}(\mathbf{A}_{l^*}^u)^T|^2\}$  imposed by the intra-cell interference in the SINR expression. Additionally, the term  $\sigma_v^2/p_f$  in the denominator of (53), which is contributed by the AWGN is also much smaller than the inter-cell interference. Thus, we



may also omit the term  $\sigma_v^2/p_f$  from (53). Then, the  $\text{SINR}_{l^*}^{u*}$  of (53) can be approximated as

$$\text{SINR}_{l^*}^{u*} \approx \frac{\left| \mathbb{E} \left\{ \mathbf{H}_{l^*,l^*}^{u*} (\mathbf{A}_{l^*}^{u*})^T \right\} \right|^2}{\text{var} \left\{ \mathbf{H}_{l^*,l^*}^{u*} (\mathbf{A}_{l^*}^{u*})^T \right\} + \sum_{l=1, l \neq l^*}^L \sum_{u=1}^U \mathbb{E} \left\{ \left| \mathbf{H}_{l^*,l}^{u*} (\mathbf{A}_l^u)^T \right|^2 \right\}}. \quad (54)$$

We will resort to using Monte Carlo simulations to evaluate  $|\mathbb{E}\{\mathbf{H}_{l^*,l^*}^{u*}(\mathbf{A}_{l^*}^{u*})^T\}|^2$ ,  $\text{var}\{\mathbf{H}_{l^*,l^*}^{u*}(\mathbf{A}_{l^*}^{u*})^T\}$  and  $\mathbb{E}\{|\mathbf{H}_{l^*,l}^{u*}(\mathbf{A}_l^u)^T|^2\}$  in the calculation of the achievable throughput (52), since there are no closed-form expressions for them.

#### IV. SIMULATION RESULTS

The default values of the various parameters for our simulated multi-cell TDD system are summarized in Table I, where  $\eta_{l^*,l,q,p}^{q_l}$  denotes the path loss coefficient for the  $p$ -th path of the link between the  $q_l$ -th antenna of the  $l$ -th BS and the  $q_{l^*}$ -th antenna of the  $l^*$ -th BS with the associated path AoA  $\theta_{l^*,l,q_{l^*},p}^{q_l}$ . Unless otherwise specified, these default parameter values were used throughout. The UFR was assumed and the same pre-assigned PSs were employed for the UL CE by all the BSs. The signal-to-noise ratio (SNR) of the system was defined as  $E_b/N_0$ , where  $E_b$  denoted the energy per bit and  $N_0$  denoted the power of the channel AWGN. All the path AoAs,  $\theta_{l^*,l,q,p}^u$  and  $\theta_{l^*,l,q_{l^*},p}^{q_l}$ , were independently and identically distributed (i.i.d.) Gaussian random variables with the mean  $\bar{\theta} = 90^\circ$  and the standard deviation  $\sigma_{\text{AoA}} = 90^\circ$ . The pilot contamination reduction schemes of the LMMSE CE using the aligned PSs and the LMMSE CE relying on the staggered PSs as well as the DL precoding relying on perfect CSI were included as the benchmarks for comparison with our proposed pilot contamination elimination scheme.

The LMMSE CE using the aligned PSs [6] refers to the estimator that relies on all the MSs of all the cells simultaneously transmitting their UL PSs, which occupies  $K$  OFDM symbol durations for its UL training. More details about the LMMSE CE using the aligned PSs can be found in reference [6]. Furthermore, observe in (14) that the LMMSE CE using aligned PSs fails to reduce the pilot contamination by employing longer OFDM PSs, although it is capable of reducing the effects of AWGN. However, the effect of AWGN on the achievable capacity can be neglected, as the number

TABLE I  
DEFAULT PARAMETERS USED IN THE SIMULATED MULTIPLE-ANTENNA AIDED MULTI-CELL TDD SYSTEM

Number of cells $L$	7
Number of MSs per-cell $U$	4
Number of antennas at each BS $Q$	50
Average transmit power at each MS $p_r$	0dB
Average transmit power at each BS $p_f$	10dB
Path loss coefficients $\beta_{l^*,l,q,p}^u = \beta_{l^*,l}, l \neq l^*$	0.3
Path loss coefficients $\beta_{l^*,l^*,q,p}^u = \beta_{l^*,l^*}$	1.0
Path loss coefficients $\eta_{l^*,l,q_{l^*},p}^{q_l} = \eta_{l^*,l}, l \neq l^*$	0.5
Mean of path AoAs $\theta$	$90^\circ$
Standard deviation of path AoAs $\sigma_{\text{AoA}}$	$90^\circ$
Antenna spacing $D$	$\frac{\lambda}{2}$
Number of paths $P$	50
Pilot length $K$	20

of antennas grows without limit [3], [6]. Hence, even if we provide  $[(L+3)L]$ -length PSs for the LMMSE CE using the aligned PSs, its performance remains almost the same as its performance associated with using  $K$ -length PSs. In our stimulated UFR system, the LMMSE CE with the aligned PSs used the same pre-assigned UL PSs in every cell, just as our scheme, and the estimate of  $\mathbf{H}_{l^*,l^*}$  provided by the  $l^*$ -th BS is given by (10), or equivalently by (14) if  $Q$  is large. The LMMSE CE relying on the aligned PSs assumes the exact knowledge of both the channel's noise variance and of all the path loss coefficients  $\beta_{l^*,l,q,p}^u$ . Naturally, this is an unrealistic assumption, since in practice it is unclear how the serving BS can acquire the second-order statistics of all the related UL channels  $\mathbf{H}_{l^*,l}$ . Nevertheless, in our simulations, the LMMSE CE using the aligned PSs was provided with all the exact path loss coefficients  $\beta_{l^*,l,q,p}^u (= \beta_{l^*,l})$ , therefore its performance represented an 'idealized' scenario, which is never encountered in reality.

For the LMMSE CE using the staggered PSs [1], [6], the MSs roaming in different cells transmit their UL PSs at non-overlapping instances. Specifically, when the MSs in cell  $l^*$  are transmitting the pre-assigned UL PSs, all the MSs of all other cells can only perform their DL data transmissions. Although the length of the UL training period remains  $K$  OFDM symbol durations for each cell, the total training period for an  $L$ -cell system spans over  $LK$  OFDM symbol durations with the adjacent cells performing their DL data transmission in a coordinated manner for LMMSE CE using the staggered PSs. This LMMSE CE using the staggered PSs, as detailed in Appendix, requires the knowledge of the FDCHTF matrices  $\mathbf{G}_{l^*,l} \in \mathbb{C}^{Q \times Q}$  between the  $l$ -th BSs for  $l \neq l^*$  and the  $l^*$ -th BS as well as the channel noise variance and all the path loss coefficients  $\beta_{l^*,l,q,p}^u$  for the serving cell. Estimating  $\mathbf{G}_{l^*,l}$  requires a huge training

$$\text{SINR}_{l^*}^{u*} = \frac{\left| \mathbb{E} \left\{ \mathbf{H}_{l^*,l^*}^{u*} (\mathbf{A}_{l^*}^{u*})^T \right\} \right|^2}{\frac{\sigma_v^2}{p_f} + \text{var} \left\{ \mathbf{H}_{l^*,l^*}^{u*} (\mathbf{A}_{l^*}^{u*})^T \right\} + \sum_{u=1, u \neq u^*}^U \mathbb{E} \left\{ \left| \mathbf{H}_{l^*,l^*}^{u*} (\mathbf{A}_{l^*}^{u*})^T \right|^2 \right\} + \sum_{l=1, l \neq l^*}^L \sum_{u=1}^U \mathbb{E} \left\{ \left| \mathbf{H}_{l^*,l}^{u*} (\mathbf{A}_l^u)^T \right|^2 \right\}}. \quad (53)$$

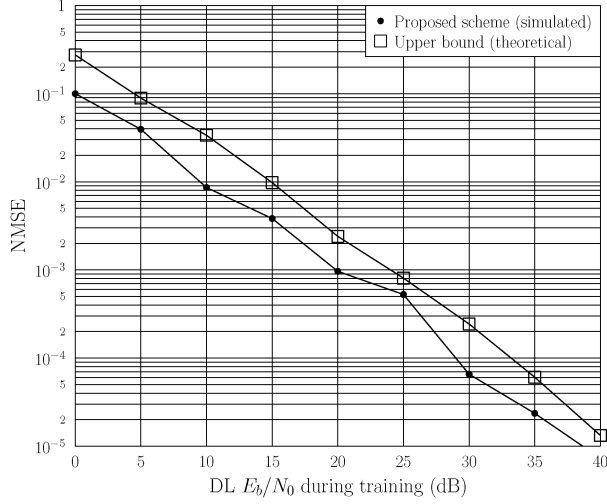


Fig. 4. Normalized mean square error performance of our proposed channel estimator as the function of the DL training SNR, given the UL training SNR of  $E_b/N_0 = 20$  dB. The simulated network is specified in Table I.

overhead, since the number of antennas  $Q$  is large, and the estimation accuracy is limited due to the inherent inter-cell interference. The path loss coefficients  $\beta_{l^*,q,p}^u$  are contained inside the FDCHTF vectors  $\mathbf{H}_{l^*,l}^u$  which have to be estimated, and it is unrealistic to assume that they are known before all the  $\mathbf{H}_{l^*,l}^u$  are estimated. In our simulations, the LMMSE CE with the staggered PSs was provided with all the true  $\mathbf{G}_{l^*,l}$  and all the true  $\beta_{l^*,l^*,q,p}^u (= \beta_{l^*,l^*}^u)$ . Therefore, its performance represented an ‘idealized’ scenario, which is difficult to approach in reality.

We firstly examined the normalized mean square error (NMSE), which is defined as

$$\text{NMSE}_{\text{simu}} = \frac{\sum_{u=1}^U \sum_{q=1}^Q |\hat{H}_{l^*,l^*,q}^u - H_{l^*,l^*,q}^u|^2}{\sum_{u=1}^U \sum_{q=1}^Q |H_{l^*,l^*,q}^u|^2}, \quad (55)$$

where  $H_{l^*,l^*,q}^u$  and  $\hat{H}_{l^*,l^*,q}^u$  are the true channel and its estimate, respectively. Because our scheme consists of the DL training and UL training stages, we began by investigating the NMSE of our estimator as a function of the DL training SNR, given the UL SNR of  $E_b/N_0 = 20$  dB. The simulation results for  $\text{NMSE}_{\text{simu}}$  obtained by averaging over 100 channel realizations with different DL training SNRs are depicted in Fig. 4, which shows standard behavior of an unbiased estimator whose estimation accuracy is determined by the signal’s SNR. To verify the accuracy of the theoretical analysis carried out in Section III-B to derive the upper bounds of our estimator variances  $\text{var}\{\hat{H}_{l^*,l^*,q}^u\}$  given in (46) and (47) for  $1 \leq u \leq U$  and  $1 \leq q \leq Q$ , we also plot the theoretical upper bound of the NMSE, which is defined as

$$\text{NMSE}_{\text{theo-ub}} = \frac{\sum_{u=1}^U \sum_{q=1}^Q \text{var}\{\hat{H}_{l^*,l^*,q}^u\}}{\sum_{u=1}^U \sum_{q=1}^Q |H_{l^*,l^*,q}^u|^2}, \quad (56)$$

in Fig. 4. Observe from Fig. 4 that  $\text{NMSE}_{\text{theo-ub}}$  is just above  $\text{NMSE}_{\text{simu}}$  and exhibits the same trend as the latter. We next

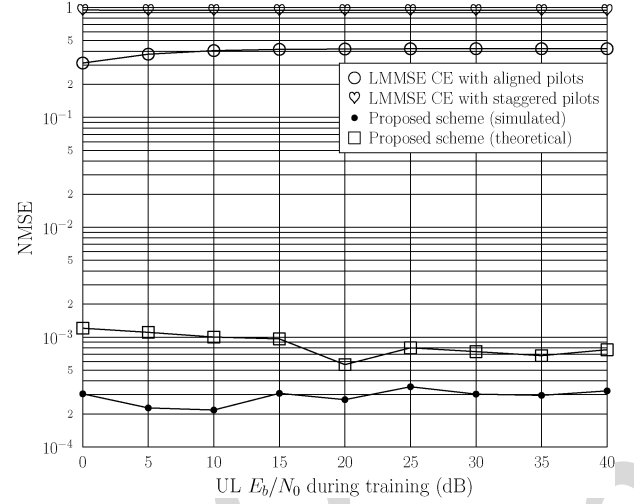


Fig. 5. Normalized mean square error performance of three channel estimators as the function of the UL training SNR. The simulated network is specified in Table I, and the DL training SNR for our proposed estimator is  $E_b/N_0 = 25$  dB.

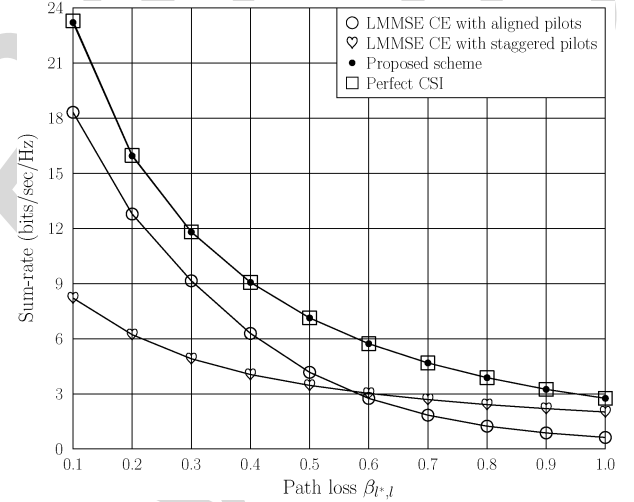


Fig. 6. Achievable sum-rate performance as the function of the path loss coefficient  $\beta_{l^*,l}$  by the three estimators, in comparison with the perfect CSI bound. The UL training SNR is  $E_b/N_0 = 20$  dB for all the three estimators and additionally the DL training SNR is  $E_b/N_0 = 25$  dB for our proposed scheme. The rest of the network parameters are specified in Table I.

studied the NMSE of our estimator as a function of the UL training SNR, given the DL SNR of  $E_b/N_0 = 25$  dB. The results obtained again by averaging over 100 channel realizations are shown in Fig. 5, where the NMSE performance of the LMMSE CE relying on aligned PSs and the LMMSE CE using staggered PSs are also given for comparison. It can be seen from Fig. 5 that the UL training SNR only has very slight influence on the achievable NMSE performance of all the three channel estimators, because the dominant factors which determine the UL estimation accuracy are the intra-cell and/or inter-cell interferences and/or the pilot contamination. Observe from Fig. 5 that our estimator significantly outperforms the other two estimators.

To investigate the achievable sum-rate performance by the three estimators under various network conditions, we set the UL training SNR to  $E_b/N_0 = 20$  dB for all the three estimators and additionally set the DL training SNR to  $E_b/N_0 = 25$  dB

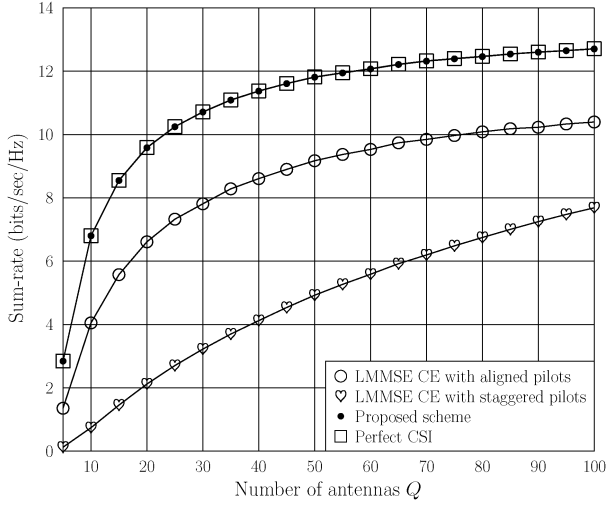


Fig. 7. Achievable sum-rate performance as the function of the number of antennas  $Q$  by the three estimators, in comparison with the perfect CSI bound. The UL training SNR is  $E_b/N_0 = 20$  dB for all the three estimators and additionally the DL training SNR is  $E_b/N_0 = 25$  dB for our proposed scheme. The rest of the network parameters are specified in Table I.

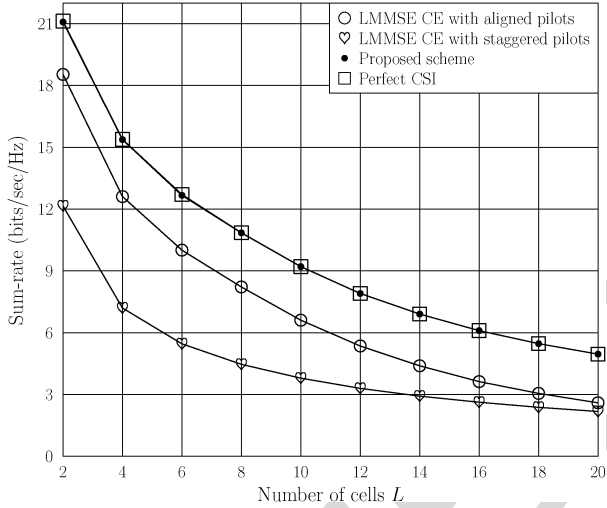


Fig. 8. Achievable sum-rate performance as the function of the number of cells  $L$  by the three estimators, in comparison with the perfect CSI bound. The UL training SNR is  $E_b/N_0 = 20$  dB for all the three estimators and additionally the DL training SNR is  $E_b/N_0 = 25$  dB for our proposed scheme. For  $L \geq 8$ , the path loss coefficients between the desired cell and its neighboring cells are set to  $\beta_{l^*,l} = 0.3$  and the path loss coefficients between the serving cell and the rest far-away cells are set to  $\beta_{l^*,l} = 0.2$ , while the rest of the network parameters are specified in Table I.

for our proposed scheme. Fig. 6 portrays the sum-rate performance versus the path loss coefficient  $\beta_{l^*,l}$  between different cells by the three estimation schemes, in comparison to the perfect CSI bound. Observe in Fig. 6 that our scheme significantly outperforms both the LMMSE CE relying on aligned PSs and the LMMSE CE relying on staggered PSs. Furthermore, the achievable sum-rate of our scheme approaches the sum-rate of the perfect CSI bound. The sum-rates achieved by the different schemes as a function of the number of antennas  $Q$  are shown in Fig. 7, where it can be seen that the sum-rate achieved by our scheme approaches the perfect CIS bound, which is substantially better than those of the LMMSE CE relying on aligned PSs and the LMMSE CE using staggered PSs. The achievable sum-rate versus the number of cells  $L$  is investigated in Fig. 8

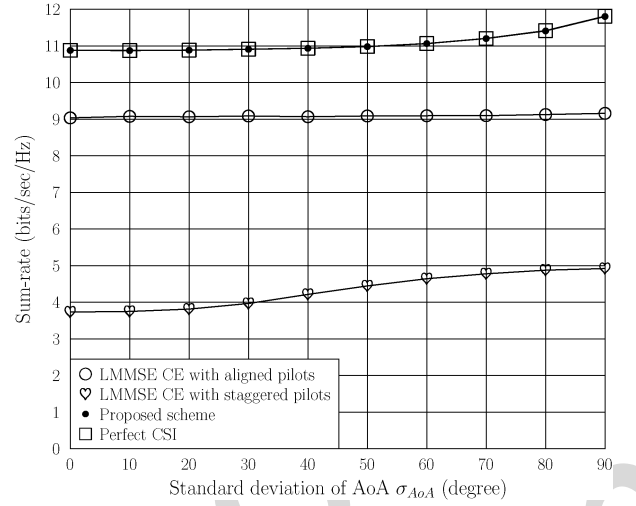


Fig. 9. Achievable sum-rate performance as the function of the path AoA's standard deviation  $\sigma_{AoA}$  by the three estimators, in comparison with the perfect CSI bound. The UL training SNR is  $E_b/N_0 = 20$  dB for all the three estimators and additionally the DL training SNR is  $E_b/N_0 = 25$  dB for our scheme. The rest of the network parameters are specified in Table I.

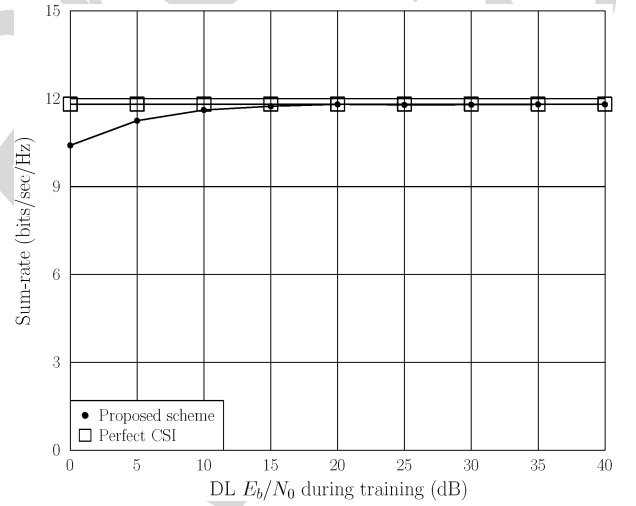


Fig. 10. Achievable sum-rate performance of our proposed channel estimator as the function of the DL training SNR, given the UL training SNR of  $E_b/N_0 = 20$  dB, in comparison with the perfect CSI bound. The simulated network is specified in Table I.

where, for  $L \geq 8$ , the path loss coefficients between the desired cell and its neighboring cells were  $\beta_{l^*,l} = 0.3$  as specified in Table I but the path loss coefficients between the serving cell and the rest far-away cells were set to  $\beta_{l^*,l} = 0.2$ . It can be seen from Fig. 8 that the proposed scheme approaches the perfect CSI bound, and it significantly outperforms both the LMMSE CE using aligned PSs and the LMMSE CE relying on staggered PSs. Fig. 9 depicts the achievable sum-rate versus the standard deviation  $\sigma_{AoA}$  of all the path AoAs. Observe from Fig. 9 that all the schemes are only slightly affected by  $\sigma_{AoA}$ . Again, our scheme approaches the perfect CSI bound and considerably outperforms the other two estimators.

In order to investigate the effect of the DL training accuracy on the achievable sum-rate of our scheme, we set the UL training SNR to  $E_b/N_0 = 20$  dB and varied the DL training SNR. The sum-rate performance attained by our scheme is shown in

Fig. 10, where it is seen that the sum-rate attained by our estimator deviates slightly from the perfect CSI bound only in the region of the DL training SNR of  $E_b/N_0 < 15$  dB.

## V. CONCLUSIONS

A novel pilot contamination elimination scheme has been proposed for multi-cell TDD and OFDM based massive MIMO systems, which relies on the two stages of the DL training and the scheduled UL training. In the DL CE stage, each BS transmits its DL PSs to its serving MSs for them to estimate the user-specific DL FDCHTFs, which are then embedded in the MSs' UL PSs to be used in the second stage of UL training. A scheduled UL training procedure allows each BS in turn to estimate its MSs' FDCHTFs accurately by eliminating the pilot contamination imposed by the surrounding cells. Unlike some existing pilot contamination reduction schemes, such as the LMMSE CE relying on aligned PSs and the LMMSE CE using staggered PSs, our scheme does not require the knowledge of the second-order statistics of the MIMO CIRs. A drawback of our scheme is that it expands the training overhead by a factor equal to the number of interfering cells. Our extensive simulation results demonstrate that the proposed estimator significantly outperforms both the LMMSE CE relying on aligned PSs and the LMMSE CE using staggered PSs. The ZF precoding aided DL data transmission based on the estimated FDCHTF matrix provided by our estimator is capable of approaching the average sum-rate bound of the perfect CSI scenario.

## APPENDIX

### THE LMMSE CE WITH THE STAGGERED PSS

The signal received at the  $l^*$ -th BS during its UL training consists of the UL pilot transmissions of the MSs in cell  $l^*$  and all the DL data transmissions from all the other BSs, which can be expressed as

$$\mathbf{Y}_{l^*} = \sqrt{p_r} \mathbf{X}_{l^*} \mathbf{H}_{l^*,l^*} + \sqrt{p_f} \sum_{l=1, l \neq l^*}^L \mathbf{D}_l^T \mathbf{G}_{l^*,l} + \mathbf{W}_{l^*}, \quad (57)$$

where  $\mathbf{D}_l = \mathbf{A}_l \bar{\mathbf{S}}_l^T \in \mathbb{C}^{Q \times K}$  is the transmitted DL data of cell  $l$  with  $\bar{\mathbf{S}}_l = [\mathbf{S}_l^T[1] \mathbf{S}_l^T[2] \cdots \mathbf{S}_l^T[K]]$  and  $\mathbf{G}_{l^*,l} \in \mathbb{C}^{Q \times Q}$  is the FDCHTF matrix connecting the BS of cell  $l$  to the BS of cell  $l^*$  on subcarrier  $n$ . The link between the  $q_l$ -th antenna of the  $l$ -th BS and the  $q_{l^*}$ -th antenna of the  $l^*$ -th BS over the  $N$  subcarriers is represented by the FDCHTF row vector  $\mathbf{G}_{l^*,l}^{q_l} = [G_{l^*,l,q_{l^*}}^{q_l}[1] G_{l^*,l,q_{l^*}}^{q_l}[2] \cdots G_{l^*,l,q_{l^*}}^{q_l}[N]]$  as

$$\mathbf{G}_{l^*,l,q_{l^*}}^{q_l} = \mathbf{g}_{l^*,l,q_{l^*}}^{q_l} \mathbf{F}^T, \quad (58)$$

with the TD CIR vector  $\mathbf{g}_{l^*,l,q_{l^*}}^{q_l} \in \mathbb{C}^{1 \times P}$  defined by

$$\mathbf{g}_{l^*,l,q_{l^*}}^{q_l} = \begin{bmatrix} \alpha_{l^*,l,q_{l^*}}^{q_l} e^{-j2\pi \frac{(q_l-1)D}{\lambda} \cos(\theta_{l^*,l,q_{l^*}}^{q_l},1)} \cdots \\ \alpha_{l^*,l,q_{l^*}}^{q_l} e^{-j2\pi \frac{(q_l-1)D}{\lambda} \cos(\theta_{l^*,l,q_{l^*}}^{q_l},P)} \end{bmatrix}, \quad (59)$$

and the complex-valued tap  $\alpha_{l^*,l,q_{l^*}}^{q_l}$  given by

$$\alpha_{l^*,l,q_{l^*}}^{q_l} = e^{-j\varphi_{l^*,l,q_{l^*}}^{q_l}} \sqrt{\eta_{l^*,l,q_{l^*}}^{q_l}}, \quad (60)$$

similar to (3) to (5). Again, we may assume that the path loss coefficients obey  $\eta_{l^*,l,q_{l^*}}^{q_l} = \eta_{l^*,l}^{q_l}$  for  $1 \leq l, q_l \leq Q$  and  $1 \leq p \leq P$ . Note that the  $q_l$ -th row and  $q_{l^*}$ -th column element of  $\mathbf{G}_{l^*,l}$  is simply  $G_{l^*,l,q_{l^*}}^{q_l}[n]$ , similar to the definition of  $\mathbf{H}_{l^*,l}$ . Recall that the  $u$ -th row of  $\mathbf{H}_{l^*,l^*}$  is denoted as  $\mathbf{H}_{l^*,l^*}^u \in \mathbb{C}^{1 \times Q}$ ,  $1 \leq u \leq U$ . Let us now define

$$\mathbf{H}_{l^*,l^*,\text{vec}}^u = \underbrace{[1 \ 1 \cdots 1]}_{K \text{ times}} \otimes (\mathbf{H}_{l^*,l^*}^u)^T \in \mathbb{C}^{1 \times KQ}, \quad (61)$$

$$\begin{aligned} \mathbf{C}_{\mathbf{H}_{l^*,l^*,\text{vec}}^u} &= \mathbb{E} \left\{ (\mathbf{H}_{l^*,l^*,\text{vec}}^u)^H \mathbf{H}_{l^*,l^*,\text{vec}}^u \right\} \\ &= \mathbb{E} \left\{ \underbrace{[1 \ 1 \cdots 1]^H}_{K \text{ times}} \underbrace{[1 \ 1 \cdots 1]}_{K \text{ times}} \right\} \\ &\quad \otimes \mathbb{E} \left\{ (\mathbf{H}_{l^*,l^*}^u)^H \mathbf{H}_{l^*,l^*}^u \right\} \\ &= \mathbf{1}_K \otimes \mathbf{C}_{\mathbf{H}_{l^*,l^*}^u} \in \mathbb{C}^{KQ \times KQ}, \end{aligned} \quad (62)$$

where  $\mathbf{1}_K$  denotes the  $(K \times K)$ -element matrix whose elements are all equal to one. Note that  $(\mathbf{X}_{l^*}^u)^T$  is the  $u$ -th column of  $\mathbf{X}_{l^*}$ ,  $1 \leq u \leq U$ , with  $\mathbf{X}_{l^*}^u \in \mathbb{C}^{1 \times K}$ , and let us denote the  $k$ -th column of  $\mathbf{D}_l$  as  $\mathbf{D}_l^T[k]$ ,  $1 \leq k \leq K$ , with  $\mathbf{D}_l[k] \in \mathbb{C}^{1 \times Q}$ . Further define

$$\Psi_{\text{diag}}^u = \text{diag} \left\{ (\mathbf{X}_{l^*}^u)^H \right\} \otimes \mathbf{I}_Q \in \mathbb{C}^{KQ \times KQ}, \quad (63)$$

$$\mathbf{D}_{l,\text{vec}} = [\mathbf{D}_l^T[1] \mathbf{D}_l^T[2] \cdots \mathbf{D}_l^T[K]] \in \mathbb{C}^{1 \times KQ}, \quad (64)$$

$$\begin{aligned} \mathbf{C}_{\mathbf{D}_{l,\text{vec}}} &= \mathbb{E} \left\{ \mathbf{D}_{l,\text{vec}}^H \mathbf{D}_{l,\text{vec}} \right\} = \mathbb{E} \left\{ \bar{\mathbf{S}}_l^H \bar{\mathbf{S}}_l \right\} \otimes \mathbb{E} \left\{ \mathbf{A}_l^H \mathbf{A}_l \right\} \\ &= \mathbf{C}_{\bar{\mathbf{S}}_l} \otimes \mathbf{C}_{\mathbf{A}_l} \in \mathbb{C}^{KQ \times KQ}, \end{aligned} \quad (65)$$

$$\mathbf{g}_{l^*,l} = \mathbf{I}_K \otimes \mathbf{G}_{l^*,l}^H \in \mathbb{C}^{KQ \times KQ}. \quad (66)$$

If we denote the  $k$ -th row of  $\mathbf{Y}_{l^*}$  by  $\mathbf{Y}_{l^*}[k] \in \mathbb{C}^{1 \times Q}$ , where  $1 \leq k \leq K$ , we can alternatively express  $\mathbf{Y}_{l^*} \in \mathbb{C}^{K \times Q}$  as

$$\mathbf{Y}_{l^*,\text{vec}} = [\mathbf{Y}_{l^*}[1] \mathbf{Y}_{l^*}[2] \cdots \mathbf{Y}_{l^*}[K]] \in \mathbb{C}^{1 \times KQ}. \quad (67)$$

Similarly, we can define  $\mathbf{W}_{l^*,\text{vec}}$  as an alternative representation of  $\mathbf{W}_{l^*}$ . Then it follows from (57) that

$$\begin{aligned} \mathbf{Y}_{l^*,\text{vec}}^H &= \sqrt{p_r} \sum_{u=1}^U \Psi_{\text{diag}}^u (\mathbf{H}_{l^*,l^*,\text{vec}}^u)^T \\ &\quad + \sqrt{p_f} \sum_{l=1, l \neq l^*}^L \mathbf{g}_{l^*,l} (\mathbf{D}_{l,\text{vec}})^T + \mathbf{W}_{l^*,\text{vec}}^H. \end{aligned} \quad (68)$$

Then the LMMSE estimator of  $\mathbf{H}_{l^*,l^*,\text{vec}}^u$  is given by [1]:

$$\left( \hat{\mathbf{H}}_{l^*,l^*,\text{vec}}^u \right)^H = \sqrt{p_r} \mathbf{C}_{\mathbf{H}_{l^*,l^*,\text{vec}}^u} (\Psi_{\text{diag}}^u)^H (\mathbf{\Upsilon}^u)^{-1} \mathbf{Y}_{l^*,\text{vec}}^H, \quad (69)$$

where  $\mathbf{\Upsilon}^u \in \mathbb{C}^{KQ \times KQ}$  is given by

$$\begin{aligned} \mathbf{\Upsilon}^u &= p_r \Psi_{\text{diag}}^u \mathbf{C}_{\mathbf{H}_{l^*,l^*,\text{vec}}^u} (\Psi_{\text{diag}}^u)^H \\ &\quad + p_f \sum_{l=1, l \neq l^*}^L \mathbf{g}_{l^*,l} \mathbf{C}_{\mathbf{D}_{l,\text{vec}}} \mathbf{g}_{l^*,l}^H + \sigma_w^2 \mathbf{I}_{KQ}. \end{aligned} \quad (70)$$

However,  $\mathbf{C}_{\mathbf{H}_{l^*,l^*,\text{vec}}^u}$ ,  $\mathbf{C}_{\bar{\mathbf{S}}_l}$  and  $\mathbf{C}_{\mathbf{A}_l}$  are unknown, and several assumptions have to be made. Specifically, it is assumed that 1)

the channel coefficients of the different channel links in the serving cell are i.i.d. with zero mean and a covariance matrix of  $\mathbf{C}_{\mathbf{H}_{l^*,l^*}^u} = \beta_{l^*,l^*} \mathbf{I}_Q$ ; 2) the DL transmitted data of the other BSs are also i.i.d. with zero mean and a covariance matrix of  $\mathbf{C}_{\mathbf{S}_l} = \mathbf{I}_K$ ; and 3) the approximation of  $\mathbb{E}\{\mathbf{A}_l^H \mathbf{A}_l\} = (1/Q) \mathbf{I}_Q$  holds for large  $Q$  [1]. Under these assumptions, we have

$$\tilde{\mathbf{C}}_{\mathbf{H}_{l^*,l^*}^u, \text{vec}} = \mathbf{1}_K \otimes (\beta_{l^*,l^*} \mathbf{I}_Q) \approx \mathbf{C}_{\mathbf{H}_{l^*,l^*}^u, \text{vec}}, \quad (71)$$

$$\tilde{\mathbf{C}}_{\mathbf{D}_l, \text{vec}} = \frac{1}{Q} \mathbf{I}_{KQ} \approx \mathbf{C}_{\mathbf{D}_l, \text{vec}}, \quad (72)$$

$$\begin{aligned} \tilde{\mathbf{Y}}^u &= p_r \Psi_{\text{diag}}^u \tilde{\mathbf{C}}_{\mathbf{H}_{l^*,l^*}^u, \text{vec}} (\Psi_{\text{diag}}^u)^H \\ &+ \frac{p_f}{Q} \sum_{l=1, l \neq l^*}^L \mathbf{g}_{l^*,l} \mathbf{g}_{l^*,l}^H + \sigma_w^2 \mathbf{I}_{KQ}. \end{aligned} \quad (73)$$

The LMMSE estimator of  $\mathbf{H}_{l^*,l^*}^u$  is then given by [1]

$$\left( \hat{\mathbf{H}}_{l^*,l^*}^u, \text{LMMSE} \right)^H = \sqrt{p_r} \tilde{\mathbf{C}}_{\mathbf{H}_{l^*,l^*}^u, \text{vec}} (\Psi_{\text{diag}}^u)^H (\tilde{\mathbf{Y}}^u)^{-1} \mathbf{Y}_{l^*, \text{vec}}^H. \quad (74)$$

The estimate  $\hat{\mathbf{H}}_{l^*,l^*}^u, \text{LMMSE}$  of  $\mathbf{H}_{l^*,l^*}^u$  is obtained by taking the first  $Q$  elements of  $\hat{\mathbf{H}}_{l^*,l^*}^u, \text{LMMSE}$ .

The LMMSE estimator (69) for MS  $u$  is contaminated by the other  $(U-1)$  MSs scattered in the same cell. To clearly see this contamination, let us consider the ‘ideal’ case of no inter-cell interference and no channel AWGN, namely, the path loss coefficients between cell  $l$  and the desired cell  $l^*$   $\eta_{l^*,l} \rightarrow 0$  and the AWGN’s power  $\sigma_w^2 \rightarrow 0$ . Invoking (58) to (60) and (66), we have  $\mathbf{g}_{l^*,l} \rightarrow \mathbf{0}_{KQ}$ , where  $\mathbf{0}_{KQ}$  represents the  $(KQ \times KQ)$ -element matrix whose elements are all equal to zero. Therefore, we can rewrite  $\mathbf{Y}^u$  of (70) as follows

$$\mathbf{Y}^u \approx p_r \Psi_{\text{diag}}^u \mathbf{C}_{\mathbf{H}_{l^*,l^*}^u, \text{vec}} (\Psi_{\text{diag}}^u)^H. \quad (75)$$

Hence, the LMMSE estimator of  $\mathbf{H}_{l^*,l^*}^u$  given in (69) can be approximated as

$$\begin{aligned} \left( \hat{\mathbf{H}}_{l^*,l^*}^u, \text{LMMSE} \right)^H &\approx \sqrt{p_r} \mathbf{C}_{\mathbf{H}_{l^*,l^*}^u, \text{vec}} (\Psi_{\text{diag}}^u)^H \\ &\times \left( p_r \Psi_{\text{diag}}^u \mathbf{C}_{\mathbf{H}_{l^*,l^*}^u, \text{vec}} (\Psi_{\text{diag}}^u)^H \right)^{-1} \mathbf{Y}_{l^*, \text{vec}}^H. \end{aligned} \quad (76)$$

Substituting  $\mathbf{Y}_{l^*, \text{vec}}^H$  of (68) and the assumption of  $\sigma_w^2 \rightarrow 0$  into (76) leads to

$$\begin{aligned} \left( \hat{\mathbf{H}}_{l^*,l^*}^u, \text{LMMSE} \right)^H &\approx (\mathbf{H}_{l^*,l^*}^u)^T \\ &+ (\Psi_{\text{diag}}^u)^{-1} \sum_{\tilde{u}=1, \tilde{u} \neq u}^U \Psi_{\text{diag}}^{\tilde{u}} (\mathbf{H}_{l^*,l^*}^{\tilde{u}})^T, \\ &= (\mathbf{H}_{l^*,l^*}^u)^T \\ &+ \sum_{\tilde{u}=1, \tilde{u} \neq u}^U \tilde{\Psi}_{\text{diag}}^{\tilde{u}} (\mathbf{H}_{l^*,l^*}^{\tilde{u}})^T, \end{aligned} \quad (77)$$

where  $\tilde{\Psi}_{\text{diag}}^{\tilde{u}} \in \mathbb{C}^{KQ \times KQ}$  is given by

$$\tilde{\Psi}_{\text{diag}}^{\tilde{u}} = \text{diag} \left\{ \left[ \begin{array}{ccc} X_{l^*}^{\tilde{u}}[1] & X_{l^*}^{\tilde{u}}[2] & \dots & X_{l^*}^{\tilde{u}}[K] \end{array} \right]^{\dagger} \right\} \otimes \mathbf{I}_Q. \quad (78)$$

Observe in (77) that the estimator  $\hat{\mathbf{H}}_{l^*,l^*}^u, \text{LMMSE}$  is actually a biased estimator. Note that the LMMSE CE using the staggered PSs of (74) requires the following information: **a)** the FDCHTF matrices  $\mathbf{G}_{l^*,l}$  between the BSs of cells  $l$  for  $l \neq l^*$  and the BS of cell  $l^*$ ; and **b)** the covariance matrices  $\mathbf{C}_{\mathbf{H}_{l^*,l^*}^u}$  of the FDCHTFs  $\mathbf{H}_{l^*,l^*}^u$ , namely, all the path loss coefficients  $\beta_{l^*,l^*}^u$  (assuming  $\beta_{l^*,l^*}^u, q, p = \beta_{l^*,l^*}^u$  for  $1 \leq p \leq P$  and  $1 \leq q \leq Q$ ) for  $1 \leq u \leq U$ .

#### ACKNOWLEDGMENT

The authors would like to express their gratitude to Haifan Yin for his help on the channel model for multi-cell systems and to Chen Dong for his helpful suggestions.

#### REFERENCES

- [1] K. Appaiah, A. Ashikhmin, and T. L. Marzetta, “Pilot contamination reduction in multi-user TDD systems,” in *Proc. IEEE Int. Conf. Commun.*, Cape Town, South Africa, May 23–27, 2010, pp. 1–5.
- [2] 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.
- [3] T. L. Marzetta, “Noncooperative cellular wireless with unlimited numbers of base station antennas,” *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.
- [4] T. L. Marzetta and B. M. Hochwald, “Fast transfer of channel state information in wireless systems,” *IEEE Trans. Signal Process.*, vol. 54, no. 4, pp. 1268–1278, Apr. 2006.
- [5] J. Jose, A. Ashikhmin, T. L. Marzetta, and S. Vishwanath, “Pilot contamination problem in multi-cell TDD systems,” in *Proc. IEEE Int. Symp. Inf. Theory*, Seoul, Korea, Jun. 28–Jul. 3 2009, pp. 2184–2188.
- [6] F. Fernandes, A. Ashikhmin, and T. L. Marzetta, “Inter-cell interference in noncooperative TDD large scale antenna systems,” *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 192–201, Feb. 2013.
- [7] H. Q. Ngo, T. L. Marzetta, and E. G. Larsson, “Analysis of the pilot contamination effect in very large multicell multiuser MIMO systems for physical channel models,” in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Prague, Czech Republic, May 22–27, 2011, pp. 3464–3467.
- [8] J. Hoydis, S. ten 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.
- [9] B. Gopalakrishnan and N. Jindal, “An analysis of pilot contamination on multi-user MIMO cellular systems with many antennas,” in *Proc. 12th IEEE Int. Workshop Signal Process. Adv. Wireless Commun.*, San Francisco, CA, USA, Jun. 26–29, 2011, pp. 381–385.
- [10] 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. 2640–2651, Aug. 2011.
- [11] T. Zhou, M. Peng, W. Wang, and H. H. Chen, “Low-complexity coordinated beamforming for downlink multicell SDMA/OFDM systems,” *IEEE Trans. Veh. Technol.*, vol. 62, no. 1, pp. 247–255, Jan. 2013.
- [12] H. Dahrouj and W. Yu, “Coordinated beamforming for the multicell multi-antenna wireless system,” *IEEE Trans. Wireless Commun.*, vol. 9, no. 5, pp. 1748–1759, May 2010.
- [13] C. B. Chae, I. S. Hwang, R. W. Heath, and V. Tarokh, “Interference aware-coordinated beamforming in a multi-cell System,” *IEEE Trans. Wireless Commun.*, vol. 11, no. 10, pp. 3692–3703, Oct. 2012.
- [14] L. Hanzo, H. Haas, S. Imre, D. O’Brien, M. Rupp, and L. Gyongyosi, “Wireless myths, realities, futures: From 3G/4G to optical and quantum wireless,” *Proc. IEEE*, vol. 100, pp. 1853–1888, May 2012, Special Centennial Issue.
- [15] L.-L. Yang, “Design of linear multiuser transmitters from linear multiuser receivers,” in *Proc. ICC ’07*, Glasgow, U.K., Jun. 24–28, 2007, pp. 5258–5263.
- [16] Y. Xiao, Y. Zhao, and M. H. Lee, “MIMO precoding of CDMA systems,” in *Proc. 8th Int. Conf. Signal Process.*, Guilin, China, Nov. 16–20, 2006, vol. 1, pp. 397–401.
- [17] B. R. Vojčić and W. M. Jang, “Transmitter precoding in synchronous multiuser communications,” *IEEE Trans. Commun.*, vol. 46, no. 10, pp. 1346–1355, Oct. 1998.

- [18] W. Yao, S. Chen, S. Tan, and L. Hanzo, "Minimum bit error rate multiuser transmission designs using particle swarm optimisation," *IEEE Trans. Wireless Commun.*, vol. 8, no. 10, pp. 5012–5017, Oct. 2009.
- [19] W. Yao, S. Chen, and L. Hanzo, "Generalised MBER-based vector precoding design for multiuser transmission," *IEEE Trans. Veh. Technol.*, vol. 60, no. 2, pp. 739–745, Feb. 2011.
- [20] W. Yao, S. Chen, and L. Hanzo, "A transceiver design based on uniform channel decomposition and MBER vector perturbation," *IEEE Trans. Veh. Technol.*, vol. 59, no. 6, pp. 3153–3159, Jul. 2010.
- [21] A. F. Karr, *Probability*. Berlin, Germany: Springer-Verlag, 1993.
- [22] J. W. Brown and R. V. Churchill, *Complex Variables and Applications*. Upper Saddle River, NJ, USA: McGraw-Hill, 2008.
- [23] R. E. Walpole, R. H. Myers, S. L. Myers, and K. Y. Keying, *Probability and Statistics for Engineers and Scientists*. Upper Saddle River, NJ, USA: Pearson, 2011.



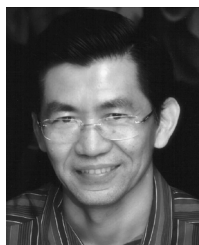
**Jiankang Zhang** (S'08–M'12) received the B.Sc. degree in Mathematics and Applied Mathematics from Beijing University of Posts and Telecommunications in 2006, and the Ph.D. degree in Communication and Information Systems from Zhengzhou University in 2012.

Since then, he has been a lecturer in School of Information Engineering, Zhengzhou University. From September 2009 to December 2011 and from January 2013 to May 2013, Dr Zhang was a visiting researcher in Electronics and Computer Science, the

University of Southampton, UK. His research interests are in the areas of wireless communications and signal processing, including channel estimation, multiuser detection, beamforming/precoding and optimization algorithms.



**Bo Zhang** received his B.S. degree in Information Engineering from National University of Defense Technology, China, in 2010. He is currently working toward the Ph.D. degree with the Communications, Signal Processing and Control Group, School of Electronics and Computer Science, University of Southampton, Southampton, UK. His research interests in wireless communications include design and analysis of cooperative communications, MIMO and network-coded systems.



**Sheng Chen** (M'1990–SM'1997–F'2008) obtained his BEng degree from the East China Petroleum Institute, Dongying, China, in January 1982, and his PhD degree from the City University, London, in September 1986, both in control engineering. In 2005, he was awarded the higher doctorate degree, Doctor of Sciences (DSc), from the University of Southampton, Southampton, UK.

From 1986 to 1999, He held research and academic appointments at the Universities of Sheffield, Edinburgh and Portsmouth, all in UK. Since 1999, he has

been with Electronics and Computer Science, the University of Southampton, UK, where he currently holds the post of Professor in Intelligent Systems and Signal Processing. Dr Chen is a Distinguished Adjunct Professor at King Abdulaziz University, Jeddah, Saudi Arabia.

Dr Chen is a Chartered Engineer (CEng) and a Fellow of IET (FIET). His recent research interests include adaptive signal processing, wireless communications, modelling and identification of nonlinear systems, neural network and machine learning, intelligent control system design, evolutionary computation

methods and optimization. He has published over 480 research papers. Dr Chen is an ISI highly cited researcher in the engineering category (March 2004).



**Xiaomin Mu** received her B.E. degree from the Beijing Institute of Technology, Beijing, China in 1982.

She is currently a full professor with the School of Information Engineering, Zhengzhou University. She has published many papers in the field of signal processing and co-authored two books. Her research interests include signal processing in communication systems, wireless communications and cognitive radio.



**Mohammed El-Hajjar** is a lecturer in the Electronics and Computer Science in the University of Southampton. He received his BEng degree in Electrical Engineering from the American University of Beirut, Lebanon in 2004. He then received an MSc in Radio Frequency Communication Systems and PhD in Wireless Communications both from the University of Southampton, UK in 2005 and 2008, respectively. Following the PhD, he joined Imagination Technologies as a design engineer, where he worked on designing and developing the

BICM peripherals in Imagination's multi-standard communications platform, which resulted in three patents. In January 2012, he joined the Electronics and Computer Science in the University of Southampton as a lecturer in the Communications, Signal Processing and Control research group. He is the recipient of several academic awards and has published a Wiley-IEEE book and in excess of 40 journal and international conference papers. His research interests are mainly in the development of intelligent communications systems for the Internet of Things including energy-efficient transceiver design, cross-layer optimization for large-scale networks, massive MIMO systems for mm-wave communications, cooperative communications and Radio over fibre systems, communications and cognitive radio.



**Lajos Hanzo** FREng, FIEEE, FIET, Fellow of EURASIP, DSc received his degree in electronics in 1976 and his doctorate in 1983. In 2009 he was awarded the honorary doctorate "Doctor Honoris Causa" by the Technical University of Budapest.

During his 37-year career in telecommunications he has held various research and academic posts in Hungary, Germany and the UK. Since 1986 he has been with Electronics and Computer Science, University of Southampton, UK, where he holds the chair in telecommunications. He has successfully

supervised 83 PhD students, co-authored 20 John Wiley/IEEE Press books on mobile radio communications totalling in excess of 10 000 pages, published 1356 research entries at IEEE Xplore, acted both as TPC and General Chair of IEEE conferences, presented keynote lectures and has been awarded a number of distinctions. Currently he is directing a 100-strong academic research team, working on a range of research projects in the field of wireless multimedia communications sponsored by industry, the Engineering and Physical Sciences Research Council (EPSRC) UK, the European Research Council and the Royal Society, UK. He is an enthusiastic supporter of industrial and academic liaison and he offers a range of industrial courses. He is also a Governor of the IEEE VTS. During 2008–2012 he was the Editor-in-Chief of the IEEE Press and a Chaired Professor also at Tsinghua University, Beijing. He was awarded the European Research Council's Senior Research Fellow Grant and the Royal Society's Wolfson Research Merit Award. For further information on research in progress and associated publications please refer to <http://www-mobile.eecs.soton.ac.uk>. Lajos has 17 000+ citations.

# Pilot Contamination Elimination for Large-Scale Multiple-Antenna Aided OFDM Systems

Jiankang Zhang, *Member, IEEE*, Bo Zhang, *Student Member, IEEE*, Sheng Chen, *Fellow, IEEE*, Xiaomin Mu, Mohammed El-Hajjar, and Lajos Hanzo, *Fellow, IEEE*

**Abstract**—Pilot contamination constitutes a particularly significant impairment in large-scale multi-cell systems. We propose an effective pilot contamination elimination scheme for multi-cell time division duplexing based orthogonal frequency division multiplexing systems, by carefully designing a sophisticated amalgam of downlink (DL) training and ‘scheduled’ uplink (UL) training. During the DL training stage, each base station (BS) transmits the DL pilot symbols (PSs) to its mobile stations (MSs) for them to estimate their frequency-domain channel transfer functions (FDCHTFs), which are then embedded in the UL PSs by ‘predistorting’ the PSs with the estimated FDCHTFs. During the scheduled UL training, each BS’s UL receiver in turn extracts the FDCHTFs of its MSs from their received PSs by eliminating the pilot contamination imposed by the simultaneously transmitted UL PSs of all other cells. Our simulation results demonstrate that the pilot contamination is completely eliminated by the proposed scheme, even for the network consisting of a large number of unity frequency reuse cells. Most significantly, unlike many existing pilot contamination reduction schemes, our scheme does not rely on the assumption that each BS knows the second-order statistics of all the interfering UL channels.

**Index Terms**—Channel estimation, multi-cell systems, multiple antennas, pilot contamination, time division duplexing.

## I. INTRODUCTION

In order to achieve a high area-spectral efficiency, the same carrier frequencies are reused in the neighboring cells by obeying specific reuse patterns, where the base stations (BSs) employing multiple antennas are capable of simultaneously supporting a multiplicity of autonomous mobile stations (MSs) [1], [2]. However, the performance of these multi-user, multiple-antenna aided systems is critically dependent on the

accuracy of the channel state information (CSI), regardless whether the CSI is used for uplink (UL) reception or for the downlink (DL) transmit precoding (TP) or transmit beamforming (TBF) [3]. The CSI can be acquired at the BS with the aid of UL training, which may then be used for DL TBF, by exploiting the reciprocity between the UL and the DL of time-division duplexing (TDD) systems [4]. The accuracy of the channel estimate (CE) and thus the attainable system performance relies on having perfectly orthogonal pilots allocated to the different users. However, using orthogonal pilots may only be guaranteed for the users roaming within the same cell, but not for those scattered across different cells, because the convolution of the pilots with long channel impulse responses (CIRs) destroys their orthogonality. Moreover, the limited available bandwidth may not allow unique, user-specific orthogonal pilots to be employed for each user [5], especially not in multi-cell systems relying on a radical frequency reuse factor of one. Then the pilots will be gravely polluted by the adjacent cells’ users, when the BS of the serving cell carries out CE. Compared to the effect of additive white Gaussian noise (AWGN), pilot contamination constitutes a much more grave impairment that limits the system’s performance [5]–[9].

The impact of pilot contamination imposed on multi-cell, multiple-antenna systems was characterized in [5] and it was shown that the precoded DL signal of the BS in the serving cell contaminates the received signal of the users roaming in other cells [5]. A study of pilot contamination was given in [10] and a minimum mean square error (MMSE) criterion based precoding method was developed for mitigating the above-mentioned contamination. The authors of [7] analyzed the pilot contamination problems in multi-cell, multiple-input multiple-output (MIMO) systems relying on a large number of antennas at the BS, which demonstrated that pilot contamination persists in large-scale MIMO systems. In order to reduce the pilot contamination in multi-cell, multiple-antenna systems, the work [1] proposed a modified frame structure for each cell, where non-overlapping pilots were used for the different cells. The study [6] analyzed the asymptotic behavior of the signal to interference plus noise ratio (SINR) for both time aligned and time staggered pilots, which revealed that it may be possible to cancel the interference of adjacent cells, as long as the pilots do not overlap in time. However, the transmission of staggered pilots requires a central controller for managing the staggering of the pilot-intervals in all of the cells in order to safeguard their ‘orthogonality’ across different cells, which becomes a challenge for an increasing number of users and cells.

Manuscript received September 12, 2013; revised January 01, 2014; accepted February 22, 2014. Date of publication March 05, 2014; date of current version nulldate. This work was supported by the National Natural Science Foundation of China under Grants 61301150, 61271421, and 61172086, the RC-UK under the India-U.K. Advanced Technology Centre, the European Research Councils under its Advanced Fellow Grant, and the EU’s Concerto Project. The associate editor coordinating the review of this manuscript and approving it for publication was XXXX XXXXXXXX.

J. Zhang and X. Mu are with the School of Information Engineering, Zhengzhou University, Zhengzhou 450001, China (e-mail: jz09v@ecs.soton.ac.uk; iexmmu@zzu.edu.cn).

B. Zhang, S. Chen, M. El-Hajjar, and L. Hanzo are with Electronics and Computer Science, University of Southampton, Southampton SO17 1BJ, U.K. (e-mail: bz2g10@ecs.soton.ac.uk; sqc@ecs.soton.ac.uk; meh@ecs.soton.ac.uk; lh@ecs.soton.ac.uk).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/JSTSP.2014.2309936

Coordinated approaches were proposed for DL TBF in [11]–[13] to mitigate the inter-cell interference and the idea was extended to the coordinated CE in [2] for alleviating the effects of pilot contamination. The channel estimator adopted in the coordinated CE scheme of [2] is the linear MMSE (LMMSE) estimator [5], [10], which yields a better performance than the least squares (LS) estimator, but it requires the knowledge of the second-order statistics of all the UL channels, including those of the interfering MSs roaming in the adjacent cells. The coordinated CE scheme of [2] further exploits the known second-order statistics of all the UL channels for designing a covariance-aware pilot assignment strategy for improving the LMMSE CE. Similar to the LMMSE estimator of [5], [10], the coordinated CE scheme [2] also requires the knowledge of the second-order statistics of all the UL channels. It is worth emphasizing that each BS is estimating its in-cell channels, but not the interfering channels from the adjacent cells. Thus, it is unrealistic to assume that the BS knows the second-order statistics of all the UL channels before they were even estimated. Therefore, it is impossible to implement such a CE scheme in practical multi-cell systems.

Against the above background, we propose an effective pilot contamination elimination scheme for multi-cell TDD orthogonal frequency division multiplexing (OFDM) systems. More specifically, the proposed pilot contamination elimination scheme relies on the two processing stages of a DL training and a scheduled UL training. During the DL training stage, each BS transmits its DL pilots to the MSs supported in order for them to estimate their specific DL frequency-domain channel transfer functions (FDCHTFs). During the scheduled UL training stage, one cell at a time, the MSs of each cell predistort their UL pilot symbols (PSs) using their estimated DL-FDCHTFs and hence the uncontaminated DL-FDCHTFs are ‘encapsulated’ in their UL PSs for exploitation by the BS. Furthermore, each BS’s UL receiver in turn extracts all the DL-FDCHTFs of its MSs from the received UL signals by eliminating the pilot contamination imposed by the UL-PS transmissions of all the other cells. Unlike the schemes of [2], [5], [10], our scheme does not assume the knowledge of the channels’ second-order statistics and, therefore, it offers a practical means of alleviating the effects of pilot contamination in large-scale multi-cell MIMO systems.

Throughout our discussion,  $\mathbb{C}$  denotes the complex number field, bold fonts are used to denote matrices and vectors, where the row vector convention is adopted. Lower and upper case symbols represent the time-domain (TD) and frequency-domain (FD) signals, respectively. For any  $A \in \mathbb{C}$ , we have  $A = A_R + jA_I$ , where  $j = \sqrt{-1}$ , while  $A_R$  and  $A_I$  are the real and imaginary parts of  $A$ , respectively. The transpose and Hermitian transpose operators are denoted by  $(\cdot)^T$  and  $(\cdot)^H$ , respectively, while  $\text{tr}\{\cdot\}$  denotes the trace operation and  $\text{diag}\{\mathbf{X}\}$  denotes the diagonal matrix with the diagonal entries equal to the components of  $\mathbf{X}$ . The inverse operation is denoted by  $(\cdot)^{-1}$ , while  $\mathbb{E}\{\cdot\}$  and  $\text{var}\{\cdot\}$  stand for the expectation and variance operations, respectively. Furthermore,  $\otimes$  denotes the Kronecker tensor product and  $^*$  denotes the conjugate operation, while  $\hat{X}$  represents the estimate of  $X$ .

The rest of this paper is organized as follows. Section II describes the MIMO-aided multi-cell TDD OFDM system.

Section III is devoted to the description of the proposed pilot contamination elimination scheme, where the achievable system throughput of our scheme is also detailed. In Section IV, we present our simulation results for characterizing the proposed pilot contamination elimination scheme in various scenarios. Our conclusions are offered in Section V.

## II. MULTI-CELL TDD SYSTEM MODEL

Consider a cellular network composed of  $L$  hexagonal cells, labelled by  $l = 1, 2, \dots, L$ , where each of the  $U$  simultaneous users of each cell is equipped with a single transmit/receive antenna, while the BS of each cell employs an array of  $Q$  transmit/receive antennas. All BSs and MSs are synchronized and rely on a TDD protocol as well as unity frequency reuse (UFR). We assume that the PSs used by the single-antenna MSs in a given cell are mutually orthogonal. The employment of OFDM is justified, because it was standardized for fourth-generation wireless communications [14].

### A. Uplink Training

At the commencement of data transmission, all MSs of all cells synchronously transmit  $K$  OFDM PSs to their serving BSs. The  $k$ -th FD PS of user  $u$  in the  $l$ -th cell is given by  $\mathbf{X}_l^u[k] = [X_l^u[k, 1] X_l^u[k, 2] \dots X_l^u[k, N]]$ , where  $N$  is the number of subcarriers and the power of each pilot  $X_l^u[k, n]$  is assumed to be unity. Let  $H_{l^*, l, q}^u[k, n]$  denote the UL-FDCHTF of the  $u$ -th user in cell  $l$  and the  $q$ -th receiver antenna of cell  $l^*$  at the  $n$ -th subcarrier of the  $k$ -th OFDM symbol. Furthermore,  $Y_{l^*, q}[k, n]$  denotes the signal received by the  $q$ -th receiver antenna element in cell  $l^*$ , which can be expressed as [1]

$$Y_{l^*, q}[k, n] = \sqrt{p_r} \sum_{u=1}^U H_{l^*, l^*, q}^u[k, n] X_{l^*}^u[k, n] + \sqrt{p_r} \sum_{l=1, l \neq l^*}^L \sum_{u=1}^U H_{l^*, l, q}^u[k, n] X_l^u[k, n] + W_{l^*, q}[k, n] \quad (1)$$

for  $1 \leq l^* \leq L$  and  $1 \leq q \leq Q$ , where  $p_r$  denotes the average power of each user and  $W_{l^*, q}[k, n]$  is the FD representation of the UL channel’s AWGN, whose power is  $\sigma_w^2$ . The set of equations constituted by (1) for  $1 \leq q \leq Q$  can be written as

$$\mathbf{Y}_{l^*}[k, n] = \underbrace{\sqrt{p_r} \mathbf{X}_{l^*}[k, n] \mathbf{H}_{l^*, l^*}[k, n]}_{\text{Desired term}} + \underbrace{\sqrt{p_r} \sum_{l=1, l \neq l^*}^L \mathbf{X}_l[k, n] \mathbf{H}_{l^*, l}[k, n]}_{\text{Inter-cell interference}} + \mathbf{W}_{l^*}[k, n], \quad (2)$$

where  $\mathbf{Y}_{l^*}[k, n] \in \mathbb{C}^{1 \times Q}$  and  $\mathbf{W}_{l^*}[k, n] \in \mathbb{C}^{1 \times Q}$  are the two row vectors hosting  $Y_{l^*, q}[k, n]$  and  $W_{l^*, q}[k, n]$  for  $1 \leq q \leq Q$ , respectively,  $\mathbf{X}_{l^*}[k, n] \in \mathbb{C}^{1 \times U}$  and  $\mathbf{X}_l[k, n] \in \mathbb{C}^{1 \times U}$  are the two row vectors hosting  $X_{l^*}^u[k, n]$  and  $X_l^u[k, n]$  for  $1 \leq u \leq U$ , respectively, while  $\mathbf{H}_{l^*, l^*}[k, n] \in \mathbb{C}^{U \times Q}$  and  $\mathbf{H}_{l^*, l}[k, n] \in \mathbb{C}^{U \times Q}$  are the two FDCHTF matrices having their  $u$ -th row and  $q$ -th column elements given by  $H_{l^*, l^*, q}^u[k, n]$  and  $H_{l^*, l, q}^u[k, n]$ , respectively.

We assume that the channel is time-invariant for the duration of  $(L + 3)K$  OFDM symbols, which allows us to drop



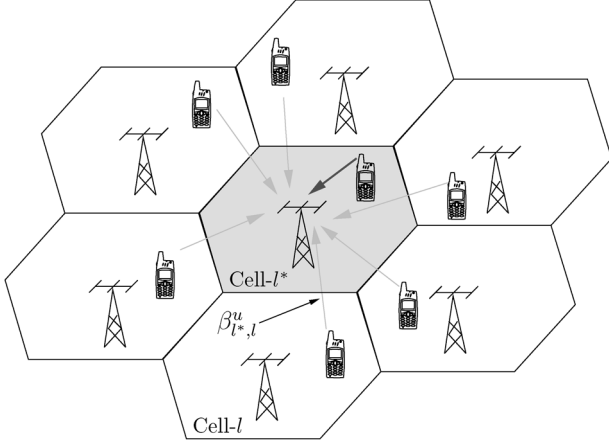


Fig. 1. Multiple-antenna aided multi-cell TDD system, where the uplink training in the cell  $l^*$  is contaminated by the MSs' transmissions from the adjacent cells which use the same pilot symbols.

the OFDM symbol index  $k$  from  $H_{l^*,l,q}^u[k,n]$ . Specifically,  $\mathbf{H}_{l^*,l,q}^u[k] = \mathbf{H}_{l^*,l,q}^u$  for  $1 \leq k \leq (L+3)K$ , where the FDCHTF row vector  $\mathbf{H}_{l^*,l,q}^u = [H_{l^*,l,q}^u[1] H_{l^*,l,q}^u[2] \cdots H_{l^*,l,q}^u[N]] \in \mathbb{C}^{1 \times N}$  represents the link between the  $u$ -th MS in the  $l$ -th cell and the  $q$ -th antenna of the target BS in the cell  $l^*$  over all the  $N$  OFDM subcarriers. According to [2], we set  $\mathbf{H}_{l^*,l,q}^u$  as

$$\mathbf{H}_{l^*,l,q}^u = \mathbf{h}_{l^*,l,q}^u \mathbf{F}^T, \quad (3)$$

with the TD CIR vector  $\mathbf{h}_{l^*,l,q}^u \in \mathbb{C}^{1 \times P}$  given by

$$\mathbf{h}_{l^*,l,q}^u = \left[ \alpha_{l^*,l,q,1}^u e^{-j2\pi \frac{(q-1)D}{\lambda} \cos(\theta_{l^*,l,q,1}^u)} \cdots \alpha_{l^*,l,q,P}^u e^{-j2\pi \frac{(q-1)D}{\lambda} \cos(\theta_{l^*,l,q,P}^u)} \right], \quad (4)$$

where  $D$  and  $\lambda$  are the antenna spacing at the BS and the carrier's wavelength, respectively,  $P$  is the number of resolvable paths,  $\theta_{l^*,l,q,p}^u$  is the angle of arrival (AoA) of the  $p$ -th path between the  $u$ -th MS in the  $l$ -th cell and the  $q$ -th antenna of the serving BS, while

$$\alpha_{l^*,l,q,p}^u = e^{-j\varphi_{l^*,l,q,p}^u} \sqrt{\beta_{l^*,l,q,p}^u} \quad (5)$$

is the complex-valued tap of the  $p$ -th path. Furthermore,  $\mathbf{F} \in \mathbb{C}^{N \times P}$  in (3) is the fast Fourier transform (FFT) matrix whose elements are given by  $F[n,p] = e^{-j2\pi(n-1)(p-1)/N}$  for  $1 \leq n \leq N$  and  $1 \leq p \leq P$ . The phase  $\varphi_{l^*,l,q,p}^u$  of the path is a random variable uniformly distributed in  $[0, 2\pi)$ . The path loss coefficient  $\beta_{l^*,l,q,p}^u$  accounts for the attenuation and shadow fading of the path, which change slowly as a function of distance [3], and we may assume that  $\beta_{l^*,l,q,p}^u = \beta_{l^*,l}^u$  for  $1 \leq p \leq P$  and  $1 \leq q \leq Q$ . The pilot contamination experienced during the UL training is illustrated in Fig. 1.

Since the channel is time-invariant for the duration of  $K$  symbols, we have  $\mathbf{H}_{l^*,l}[k,n] = \mathbf{H}_{l^*,l}[n] \in \mathbb{C}^{U \times Q}$  for  $1 \leq k \leq K$  in (2), where the  $u$ -th row and  $q$ -th column element of  $\mathbf{H}_{l^*,l}[n]$  is  $H_{l^*,l,q}^u[n]$ . As a benefit of the cyclic prefix, the OFDM symbols do not overlap in time and the BS processing can be carried out on a per-carrier basis. Hence, to simplify our notation, we will omit the subcarrier index  $n$  in the sequel. Assuming that a total of  $K$  consecutive OFDM symbols are dedicated to pilot subcarriers [3], the discrete-time model of the received signal associ-

ated with  $K$  OFDM symbols, namely (2) over  $1 \leq k \leq K$ , can be written as

$$\mathbf{Y}_{l^*} = \sqrt{p_r} \mathbf{X}_{l^*} \mathbf{H}_{l^*,l^*} + \sqrt{p_r} \sum_{l=1, l \neq l^*}^L \mathbf{X}_l \mathbf{H}_{l^*,l} + \mathbf{W}_{l^*}, \quad (6)$$

with  $\mathbf{Y}_{l^*} \in \mathbb{C}^{K \times Q}$ ,  $\mathbf{X}_l \in \mathbb{C}^{K \times U}$  and  $\mathbf{W}_{l^*} \in \mathbb{C}^{K \times Q}$ .

The LS CE and the LMMSE CE of the FDCHTF matrix  $\mathbf{H}_{l^*,l^*}$  are given respectively by [2], [10]

$$\hat{\mathbf{H}}_{l^*,l^*}^{\text{LS}} = \frac{1}{\sqrt{p_r}} \left( \mathbf{X}_{l^*}^H \mathbf{X}_{l^*} \right)^{-1} \mathbf{X}_{l^*}^H \mathbf{Y}_{l^*}, \quad (7)$$

$$\hat{\mathbf{H}}_{l^*,l^*}^{\text{LMMSE}} = \sqrt{p_r} \mathbf{R}_{l^*,l^*} \mathbf{X}_{l^*}^H \times \left( \sigma_w^2 \mathbf{I}_K + p_r \sum_{l=1}^L \mathbf{X}_l \mathbf{R}_{l^*,l} \mathbf{X}_l^H \right)^{-1} \mathbf{Y}_{l^*}, \quad (8)$$

where  $\mathbf{R}_{l^*,l} = \mathbb{E}\{\mathbf{H}_{l^*,l} \mathbf{H}_{l^*,l}^H\}$  is the covariance matrix of the FDCHTF matrix  $\mathbf{H}_{l^*,l}$ . The LS CE (7) is simple to implement, since it is only based on the target cell's PSs. By contrast, the LMMSE CE (8), which yields a much better estimate than the LS CE, relies on all the PSs from all the cells and, therefore, it requires the full cooperation of all the BSs to exchange their UL PSs. Thus, the LMMSE CE imposes a significantly higher complexity than the LS CE. Most critically, the LMMSE CE requires that the serving BS knows the second-order statistics  $\mathbf{R}_{l^*,l}$  of all the channels  $\mathbf{H}_{l^*,l}$  for  $1 \leq l \leq L$ .

### B. Pilot Contamination

From (7) and (8), we can see that the estimation of the desired channel  $\mathbf{H}_{l^*,l^*}$  relies on correlating the received signal with the known PSs, which suffers from significant pilot contamination due to the lack of orthogonality between the desired cell's pilots and the interfering cell's pilots [2], [6]. In particular, when the same PSs are reused in all the  $L$  cells, which is the worst case scenario [3], [5], [6], we have  $\mathbf{X}_l = \mathbf{X}$  for  $1 \leq l \leq L$ . Additionally, the PSs of the MSs within a given cell are orthogonal<sup>1</sup> to each other, and we have  $\mathbf{X}^H \mathbf{X} = K \mathbf{I}_U$ . Now considering the LS CE (7), we arrive at [2]

$$\hat{\mathbf{H}}_{l^*,l^*}^{\text{LS}} = \mathbf{H}_{l^*,l^*} + \sum_{l=1, l \neq l^*}^L \mathbf{H}_{l^*,l} + \frac{1}{K \sqrt{p_r}} \mathbf{X}^H \mathbf{W}_{l^*}, \quad (9)$$

where the pilot contamination constituted by the second term in (9) gives rise to a significant estimation error to the desired CE, especially when the path loss coefficients between the interfering cells and the serving cell are similar to those within the serving cell [5].

For the LMMSE CE  $\hat{\mathbf{H}}_{l^*,l^*}^{\text{LMMSE}}$  of (8), again noting  $\mathbf{X}_l = \mathbf{X}$  for  $1 \leq l \leq L$  and  $\mathbf{X}^H \mathbf{X} = K \mathbf{I}_U$ , we have

$$\hat{\mathbf{H}}_{l^*,l^*}^{\text{LMMSE}} = \left( \frac{\sigma_w^2}{K p_r} \mathbf{R}_{l^*,l^*}^{-1} + \sum_{l=1}^L \mathbf{R}_{l^*,l} \mathbf{R}_{l^*,l^*}^{-1} \right)^{-1}$$

<sup>1</sup>Note that the PSs may be made orthogonal to each other on a per-carrier basis, where the  $K$ -length PSs of the users roaming in the same cell are orthogonal to each other, regardless whether they are orthogonal or nonorthogonal over the  $N$  subcarriers.

$$\times \left( \mathbf{H}_{l^*,l^*} + \sum_{l=1, l \neq l^*}^L \mathbf{H}_{l^*,l} + \frac{1}{K\sqrt{p_r}} \mathbf{X}^H \mathbf{W}_{l^*} \right). \quad (10)$$

The second term in (10) indicates that the pilot contamination imposes a significant estimation error on the desired CE. Let us investigate the asymptotic effects of the pilot contamination imposed on the LMMSE CE as the number of antennas  $Q$  tends to infinity. In order to determine the value of  $\sum_{l=1}^L \mathbf{R}_{l^*,l} \mathbf{R}_{l^*,l}^{-1}$  as  $Q \rightarrow \infty$ , we evaluate the following limit

$$\lim_{Q \rightarrow \infty} \sum_{l=1}^L \mathbf{R}_{l^*,l} \mathbf{R}_{l^*,l}^{-1} = \sum_{l=1}^L \left( \lim_{Q \rightarrow \infty} \frac{\mathbf{R}_{l^*,l}}{Q} \right) \left( \lim_{Q \rightarrow \infty} \frac{\mathbf{R}_{l^*,l^*}}{Q} \right)^{-1}, \quad (11)$$

which is valid provided that  $\lim_{Q \rightarrow \infty} (\mathbf{R}_{l^*,l^*}/Q)$  is non-singular. Based on Lemma 1 of [6] and by invoking  $\beta_{l^*,l,q,p}^u = \beta_{l^*,l}^u$  for  $1 \leq p \leq P$  and  $1 \leq q \leq Q$  again, we arrive at

$$\lim_{Q \rightarrow \infty} \frac{\mathbf{R}_{l^*,l}}{Q} = \mathbb{E} \left\{ \lim_{Q \rightarrow \infty} \frac{\mathbf{H}_{l^*,l} \mathbf{H}_{l^*,l}^H}{Q} \right\} = \text{diag} \{ [\beta_{l^*,l}^1, \beta_{l^*,l}^2, \dots, \beta_{l^*,l}^U] \}, \quad 1 \leq l \leq L. \quad (12)$$

Note furthermore that

$$\lim_{Q \rightarrow \infty} \frac{\sigma_w^2}{K p_r} \mathbf{R}_{l^*,l^*}^{-1} = \left( \lim_{Q \rightarrow \infty} \frac{\sigma_w^2}{Q K p_r} \right) \left( \lim_{Q \rightarrow \infty} \frac{\mathbf{R}_{l^*,l^*}}{Q} \right)^{-1} = 0. \quad (13)$$

Substituting (12) and (13) into (10) yields the following expression for  $\hat{\mathbf{H}}_{l^*,l^*}^{\text{LMMSE}}$  in the asymptotic case of  $Q \rightarrow \infty$

$$\hat{\mathbf{H}}_{l^*,l^*}^{\text{LMMSE}} = \text{diag} \left\{ \left[ \frac{\beta_{l^*,l^*}^1}{\sum_{l=1}^L \beta_{l^*,l}^1} \frac{\beta_{l^*,l^*}^2}{\sum_{l=1}^L \beta_{l^*,l}^2} \dots \frac{\beta_{l^*,l^*}^U}{\sum_{l=1}^L \beta_{l^*,l}^U} \right] \right\} \times \left( \mathbf{H}_{l^*,l^*} + \sum_{l=1, l \neq l^*}^L \mathbf{H}_{l^*,l} + \frac{1}{K\sqrt{p_r}} \mathbf{X}^H \mathbf{W}_{l^*} \right). \quad (14)$$

Observe in (14) that the LMMSE CE strikes a tradeoff between the achievable estimation accuracy of the true FDCHTF matrix  $\mathbf{H}_{l^*,l^*}$  and the gravity of the pilot contamination imposed by the UL training of the MSs roaming in the adjacent cells. This beneficial trade-off is achieved at the expense of requiring all the path loss coefficients  $\beta_{l^*,l}^u$  for  $1 \leq u \leq U$  and  $1 \leq l \leq L$ .

### III. OUR PILOT CONTAMINATION ELIMINATION SCHEME

As shown in Section II-B, the CE relying on the UL PSs suffers from pilot contamination and the existing schemes [2], [5], [10] require the knowledge of all the channels' second-order statistics as well as the full cooperation of all the BSs to exchange their UL PSs. Additionally, having orthogonal PSs cannot be guaranteed for all of the MSs roaming in different cells for their UL training. Fortunately, the number of cells is limited compared to the number of MSs, hence it is possible to allocate orthogonal PSs to the BSs for their DL broadcasting. These BS-specific PSs can be exploited by the MSs for estimating their unique DL channels. By design, the DL training will not be contaminated by the adjacent cells, since orthogonal PSs have been allocated to the different BSs. Thus, the estimated DL channel

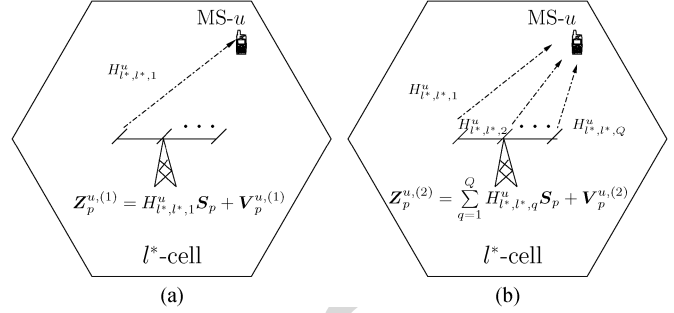


Fig. 2. Stage I: downlink training. During phase 1, the BS sends the DL pilot by its first antenna to the MSs, while during phase 2, the BS sends the pilot via its all  $Q$  antennas to the MSs. Note that the MSs in the adjacent cells may also overhear the pilot, which is not shown in this figure for clarity. (a) Phase 1; (b) Phase 2.

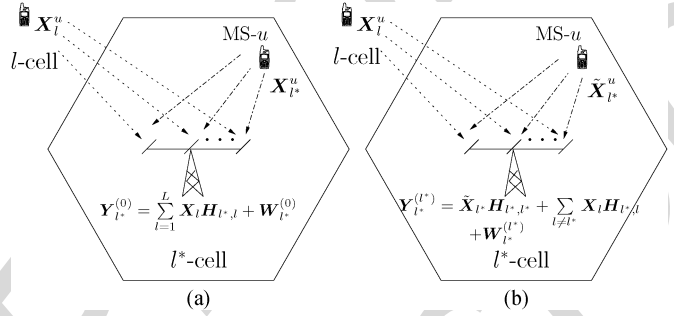


Fig. 3. Stage II: uplink training. During the initial phase, namely phase 0, all the MSs send the pre-assigned orthogonal PSs  $\mathbf{X}_l^u$  to their BSs. During the phase  $l^*$ ,  $1 \leq l^* \leq L$ , the MSs in cell  $l^*$  send their distinctive PSs  $\tilde{\mathbf{X}}_{l^*}^u$  to their BS, while the MSs in the other cells send the pre-assigned orthogonal PSs  $\mathbf{X}_l^u$  to their BSs. (a) Phase 0; (b) Phase  $l^*$ .

information can be encapsulated into the MSs's PSs for assisting the UL CE, as detailed below. Based on this philosophy, we propose a two-stage CE scheme consisting of the DL training and the appropriately scheduled UL training for the sake of eliminating the UL pilot contamination, which does not require any knowledge of the channels' second-order statistics.

#### A. The Proposed Pilot Assisted Channel Estimation

The proposed pilot contamination elimination scheme consists of a sophisticated amalgam of a DL and an UL training stage. Furthermore, the DL training contains two phases as illustrated in Fig. 2, while UL training contains  $(L + 1)$  phases as illustrated in Fig. 3, where each phase occupies  $K$  OFDM symbol durations. Again, we assume that the channel is time-invariant for the duration of  $(L + 3)K$  OFDM symbols, which allows the proposed scheme to acquire accurate estimates of the channel coefficients and then hence to eliminate the pilot contamination. More specifically, the operations of the DL and UL training are detailed as follows:

1) *The DL Training Stage:* It contains the two phases as illustrated in Fig. 2. We assume that there are  $L$  unique orthogonal OFDM PSs, one for each of the  $L$  cells, which is reasonable because the number of cells is far less than the total number of MSs roaming in all the cells. Specifically, the  $K$  pilots  $\mathbf{S}_{p,l} = [S_{p,l}[1] S_{p,l}[2] \dots S_{p,l}[K]]$  for cell  $l$  satisfy  $\mathbf{S}_{p,l}^H \mathbf{S}_{p,l} = K$  and  $\mathbf{S}_{p,l^*}^H \mathbf{S}_{p,l} = 0$  for  $1 \leq l, l^* \leq L$  and  $l \neq l^*$ . As seen in Fig. 2(a), during the first phase of the DL training, the  $l^*$ th BS broadcasts the  $K$  OFDM pilots  $\mathbf{S}_{p,l^*}$  using a single antenna, say the BS's

first antenna. The signal received by MS  $u$  in cell  $l^*$  can readily be expressed by<sup>2</sup>

$$\mathbf{Z}_p^{u,(1)} = \sqrt{p_f} H_{l^*,1}^u \mathbf{S}_p + \mathbf{V}_p^{u,(1)}, \quad 1 \leq u \leq U, \quad (15)$$

where  $\mathbf{Z}_p^{u,(1)} \in \mathbb{C}^{1 \times K}$ ,  $H_{l^*,1}^u$  represents the channel between the BS's 1st antenna and the MS  $u$ , while  $\mathbf{V}_p^{u,(1)} = [V_p^{u,(1)}[1] V_p^{u,(1)}[2] \dots V_p^{u,(1)}[K]]$  is the FD representation of the channel's AWGN. The unique, MS-specific DL channel information  $H_{l^*,1}^u$  is readily estimated using the LS estimator, yielding

$$\hat{H}_{l^*,1}^u = \frac{\mathbf{Z}_p^{u,(1)} \mathbf{S}_p^H}{K \sqrt{p_f}}, \quad 1 \leq u \leq U. \quad (16)$$

During the second phase of the DL training, as illustrated in Fig. 2(b), the BS activates all the DL transmit antennas to broadcast the same  $K$  OFDM pilots  $\mathbf{S}_p$  to MS  $u$ , where the received OFDM signal of MS  $u$  is given by

$$\mathbf{Z}_p^{u,(2)} = \sqrt{p_f} \sum_{q=1}^Q H_{l^*,q}^u \mathbf{S}_p + \mathbf{V}_p^{u,(2)}, \quad 1 \leq u \leq U, \quad (17)$$

where  $\mathbf{Z}_p^{u,(2)} \in \mathbb{C}^{1 \times K}$ ,  $H_{l^*,q}^u$  denotes the channel between the BS's  $q$ th antenna and the MS  $u$ , while  $\mathbf{V}_p^{u,(2)} = [V_p^{u,(2)}[1] V_p^{u,(2)}[2] \dots V_p^{u,(2)}[K]]$  is the FD representation of the channel's AWGN. Thus, MS  $u$  can estimate the sum of the channels, namely,  $H_{l^*,l^*}^{u,\text{sum}} = \sum_{q=1}^Q H_{l^*,q}^u$  for all the  $Q$  links between the serving BS's DL transmit antennas and itself, using the LS estimate of

$$\hat{H}_{l^*,l^*}^{u,\text{sum}} = \frac{\mathbf{Z}_p^{u,(2)} \mathbf{S}_p^H}{K \sqrt{p_f}}, \quad 1 \leq u \leq U. \quad (18)$$

2) *The UL Training Stage:* It consists of the  $(L + 1)$  scheduled phases as depicted in Fig. 3. We assume the worst case scenario of having the same pre-assigned orthogonal PS matrix  $\mathbf{X} \in \mathbb{C}^{K \times U}$  reused in every cells, namely  $\mathbf{X}_l = \mathbf{X}$  for  $1 \leq l \leq L$  and  $\mathbf{X}^H \mathbf{X} = K \mathbf{I}_U$ , where we have  $\mathbf{X} = [(\mathbf{X}^1)^T (\mathbf{X}^2)^T \dots (\mathbf{X}^U)^T]^T$  and  $\mathbf{X}^u \in \mathbb{C}^{1 \times K}$  for  $1 \leq u \leq U$ . As illustrated in Fig. 3(a), during the initial phase 0 of the UL training, the MSs roaming in all the cells simultaneously transmit their pre-assigned orthogonal PSs  $\mathbf{X}_l^u = \mathbf{X}^u$  for  $1 \leq u \leq U$  and  $1 \leq l \leq L$  in the ULs to their BSs. The composite signal received during phase 0 at the  $l^*$ -th BS can be expressed according to (6) as

$$\mathbf{Y}_{l^*}^{(0)} = \sqrt{p_r} \mathbf{X} \mathbf{H}_{l^*,l^*} + \underbrace{\sqrt{p_r} \mathbf{X} \sum_{l=1, l \neq l^*}^L \mathbf{H}_{l^*,l}}_{\text{Pilot contamination}} + \mathbf{W}_{l^*}^{(0)}. \quad (19)$$

During the  $l^*$  phase of the UL training, where  $1 \leq l^* \leq L$ , as shown in Fig. 3(b), the MSs in cell  $l^*$  transmit their own specifically pre-distorted UL PSs  $\tilde{\mathbf{X}}_{l^*}^u$  given by

$$\tilde{\mathbf{X}}_{l^*}^u = \frac{\hat{H}_{l^*,l^*}^{u,\text{sum}}}{\hat{H}_{l^*,l^*,1}^u} \mathbf{X}^u, \quad 1 \leq u \leq U, \quad (20)$$

<sup>2</sup>The serving cell index  $l^*$  is dropped from the transmitted and received pilot signals for notational simplicity.

to the  $l^*$ -th BS. As  $\tilde{\mathbf{X}}_{l^*}^u$  encapsulates the DL-FDCHTF information  $\hat{H}_{l^*,l^*}^{u,\text{sum}} / \hat{H}_{l^*,l^*,1}^u$  estimated by MS  $u$  in cell  $l^*$  during the DL training, it is distinct to this MS. At the same time, the MSs roaming in all the other cells simultaneously transmit their pre-assigned orthogonal PSs  $\mathbf{X}_l^u = \mathbf{X}^u$  to their BSs, where  $1 \leq u \leq U$ ,  $1 \leq l \leq L$  and  $l \neq l^*$ . As a result, the UL signal received by the  $l^*$ -th BS's receiver is readily expressed as

$$\begin{aligned} \mathbf{Y}_{l^*}^{(l^*)} &= \sqrt{p_r} \tilde{\mathbf{X}}_{l^*} \mathbf{H}_{l^*,l^*} + \underbrace{\sqrt{p_r} \mathbf{X} \sum_{l=1, l \neq l^*}^L \mathbf{H}_{l^*,l}}_{\text{Pilot contamination}} + \mathbf{W}_{l^*}^{(l^*)}, \\ &= \sqrt{p_r} \mathbf{X} \mathbf{B}_{l^*,l^*} \mathbf{H}_{l^*,l^*} + \sqrt{p_r} \mathbf{X} \sum_{l=1, l \neq l^*}^L \mathbf{H}_{l^*,l} + \mathbf{W}_{l^*}^{(l^*)}, \end{aligned} \quad (21)$$

where  $\tilde{\mathbf{X}}_{l^*} = [(\tilde{\mathbf{X}}_{l^*}^1)^T (\tilde{\mathbf{X}}_{l^*}^2)^T \dots (\tilde{\mathbf{X}}_{l^*}^U)^T]^T = \mathbf{X} \mathbf{B}_{l^*,l^*}$  and

$$\mathbf{B}_{l^*,l^*} = \text{diag} \left\{ \left[ \frac{\hat{H}_{l^*,l^*}^{1,\text{sum}}}{\hat{H}_{l^*,l^*,1}^1} \frac{\hat{H}_{l^*,l^*}^{2,\text{sum}}}{\hat{H}_{l^*,l^*,1}^2} \dots \frac{\hat{H}_{l^*,l^*}^{U,\text{sum}}}{\hat{H}_{l^*,l^*,1}^U} \right] \right\}. \quad (22)$$

From (19) and (21), we have respectively

$$\tilde{\mathbf{Y}}_{l^*}^{(0)} = \frac{1}{K \sqrt{p_r}} \mathbf{X}^H \mathbf{Y}_{l^*}^{(0)} = \mathbf{H}_{l^*,l^*} + \sum_{l=1, l \neq l^*}^L \mathbf{H}_{l^*,l} + \tilde{\mathbf{W}}_{l^*}^{(0)}, \quad (23)$$

$$\begin{aligned} \tilde{\mathbf{Y}}_{l^*}^{(l^*)} &= \frac{1}{K \sqrt{p_r}} \mathbf{X}^H \mathbf{Y}_{l^*}^{(l^*)} \\ &= \mathbf{B}_{l^*,l^*} \mathbf{H}_{l^*,l^*} + \sum_{l=1, l \neq l^*}^L \mathbf{H}_{l^*,l} + \tilde{\mathbf{W}}_{l^*}^{(l^*)}, \end{aligned} \quad (24)$$

where  $\tilde{\mathbf{W}}_{l^*}^{(0)} = (1/K \sqrt{p_r}) \mathbf{X}^H \mathbf{W}_{l^*}^{(0)}$  and  $\tilde{\mathbf{W}}_{l^*}^{(l^*)} = (1/K \sqrt{p_r}) \mathbf{X}^H \mathbf{W}_{l^*}^{(l^*)}$ . Furthermore, the power of both  $\tilde{\mathbf{W}}_{l^*}^{(0)}$  and  $\tilde{\mathbf{W}}_{l^*}^{(l^*)}$  is  $\sigma_w^2/p_r$ . Let  $\tilde{\mathbf{Y}}_{l^*} = \tilde{\mathbf{Y}}_{l^*}^{(l^*)} - \tilde{\mathbf{Y}}_{l^*}^{(0)}$  and  $\tilde{\mathbf{W}}_{l^*} = \tilde{\mathbf{W}}_{l^*}^{(l^*)} - \tilde{\mathbf{W}}_{l^*}^{(0)}$ , where the power of  $\tilde{\mathbf{W}}_{l^*}$  is  $2\sigma_w^2/p_r$ . From (23) and (24), we readily arrive at

$$\tilde{\mathbf{Y}}_{l^*} = \mathbf{B}_{l^*,l^*} \mathbf{H}_{l^*,l^*} - \mathbf{H}_{l^*,l^*} + \tilde{\mathbf{W}}_{l^*}. \quad (25)$$

Observe in (25) that the pilot contamination is completely eliminated.

In order to extract the estimates of the MS-specific FDCHTFs  $\hat{H}_{l^*,l^*,q}^u$  for  $1 \leq u \leq U$  and  $1 \leq q \leq Q$ , we expand (25). Specifically, we do not distinguish the DL and UL FDCHTF estimates, since we assume that they are identical, and we substitute  $H_{l^*,l^*,q}^u$  in (25) by their estimates  $\hat{H}_{l^*,l^*,q}^u$  to express it in the element-based form shown in (26) at the top of the page. (See equation at bottom of next page) Observe that at the right-hand side of (26) all the entries in the  $u$ -th row have the common factor of  $(\hat{H}_{l^*,l^*}^{u,\text{sum}} / \hat{H}_{l^*,l^*,1}^u) - 1$ . Noting  $\hat{H}_{l^*,l^*}^{u,\text{sum}} = \sum_{q=1}^Q \hat{H}_{l^*,l^*,q}^u$  yields

$$\frac{\hat{H}_{l^*,l^*}^{u,\text{sum}}}{\hat{H}_{l^*,l^*,1}^u} - 1 = \frac{1}{\hat{H}_{l^*,l^*,1}^u} \sum_{q=2}^Q \hat{H}_{l^*,l^*,q}^u. \quad (27)$$

Assuming that  $(1/\hat{H}_{l^*,1}^u) \sum_{q=2}^Q \hat{H}_{l^*,q}^u \neq 0$  for  $1 \leq u \leq U$ , we have

$$\frac{\check{Y}_{l^*,1}^u}{\check{Y}_{l^*,1}^u} = \frac{\hat{H}_{l^*,q}^u}{\hat{H}_{l^*,1}^u}, \quad 2 \leq q \leq Q, \quad 1 \leq u \leq U. \quad (28)$$

With the aid of (28), (26) can be rewritten as

$$\begin{bmatrix} \check{Y}_{l^*,1}^1 & \check{Y}_{l^*,2}^1 & \cdots & \check{Y}_{l^*,Q}^1 \\ \check{Y}_{l^*,1}^2 & \check{Y}_{l^*,2}^2 & \cdots & \check{Y}_{l^*,Q}^2 \\ \vdots & \vdots & \ddots & \vdots \\ \check{Y}_{l^*,1}^U & \check{Y}_{l^*,2}^U & \cdots & \check{Y}_{l^*,Q}^U \end{bmatrix} = \begin{bmatrix} \sum_{q=2}^Q \hat{H}_{l^*,1}^1 & \frac{\hat{H}_{l^*,2}^1}{\hat{H}_{l^*,1}^1} & \cdots & \frac{\hat{H}_{l^*,Q}^1}{\hat{H}_{l^*,1}^1} \\ \sum_{q=2}^Q \hat{H}_{l^*,1}^2 & \frac{\hat{H}_{l^*,2}^2}{\hat{H}_{l^*,1}^2} & \cdots & \frac{\hat{H}_{l^*,Q}^2}{\hat{H}_{l^*,1}^2} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{q=2}^Q \hat{H}_{l^*,1}^U & \frac{\hat{H}_{l^*,2}^U}{\hat{H}_{l^*,1}^U} & \cdots & \frac{\hat{H}_{l^*,Q}^U}{\hat{H}_{l^*,1}^U} \end{bmatrix}.$$

By summing the last  $(Q-1)$  elements in each row for both the left-hand and right-hand sides of (29), we arrive at

$$\frac{\sum_{q=2}^Q \check{Y}_{l^*,q}^u}{\check{Y}_{l^*,1}^u} = \frac{\sum_{q=2}^Q \hat{H}_{l^*,q}^u}{\hat{H}_{l^*,1}^u}, \quad 1 \leq u \leq U. \quad (30)$$

Noting  $\check{Y}_{l^*,1}^u = \sum_{q=2}^Q \hat{H}_{l^*,q}^u$  for  $1 \leq u \leq U$  yields the estimates of  $\hat{H}_{l^*,1}^u$

$$\hat{H}_{l^*,1}^u = \frac{\check{Y}_{l^*,1}^u \check{Y}_{l^*,1}^u}{\sum_{q=2}^Q \check{Y}_{l^*,q}^u}, \quad 1 \leq u \leq U. \quad (31)$$

By utilizing (28) and (33), we obtain the estimates of  $\hat{H}_{l^*,q}^u$

$$\hat{H}_{l^*,q}^u = \frac{\check{Y}_{l^*,1}^u \check{Y}_{l^*,q}^u}{\sum_{q=2}^Q \check{Y}_{l^*,q}^u}, \quad 1 \leq u \leq U, \quad 2 \leq q \leq Q. \quad (32)$$

*Remark 1:* Our scheme completely eliminates the pilot contamination in the estimated FDCHTFs of (31) and (32) under the worst-case UL training scenario of reusing the same orthogonal PS matrix in all the  $L$  cells. This is achieved at the cost of expanding the training session to the  $L$  appropriately scheduled periods,—one for each cell—to estimate its FDCHTF matrix. Therefore, our scheme increases the training overhead by a factor of  $L$ . However, our scheme does not require any information exchange between the BSs and, most importantly, it does not impose the unrealistic assumption of knowing the channels' second-order statistics. Consequently, unlike many existing schemes [2], [5], [10], our channel estimator has significant advantages in practical implementation.

The CE scheme proposed for the network of  $L$  cells is summarized in Algorithm 1.

---

**Algorithm 1 Pilot contamination elimination scheme.**

---

- **Stage I Simultaneous DL training:** as illustrated in Fig. 2, it consists of the two phases, a single-antenna and a multiple-antenna based channel estimation phases, during which each cell's BS simultaneously broadcasts PSs to its MSs for them to estimate their DL-FDCHTFs, as given in (16) and (18), respectively.
  - **Stage II Scheduled UL training:** as illustrated in Fig. 3, it consists of  $(L+1)$  phases.
    - (a) During the initial phase 0 of Fig. 3(a), all the MSs of all cells simultaneously transmit the pre-assigned PSs to their BSs.
    - (b) During the phase  $l^*$  of Fig. 3(b), where  $1 \leq l^* \leq L$ , the MSs of cell  $l^*$  transmit their appropriately predistorted PSs based on their estimated DL-FDCHTFs to the  $l^*$ -th BS, while the MSs of all the other cells again transmit the same pre-assigned PSs to their BSs. Given the signals received during phase 0 and phase  $l^*$ , the  $l^*$ -th BS extracts the FDCHTFs of its MSs, as seen in (31) and (32), by eliminating the pilot contamination imposed by the other cells.
- 

$$\begin{bmatrix} \check{Y}_{l^*,1}^1 & \check{Y}_{l^*,2}^1 & \cdots & \check{Y}_{l^*,Q}^1 \\ \check{Y}_{l^*,1}^2 & \check{Y}_{l^*,2}^2 & \cdots & \check{Y}_{l^*,Q}^2 \\ \vdots & \vdots & \ddots & \vdots \\ \check{Y}_{l^*,1}^U & \check{Y}_{l^*,2}^U & \cdots & \check{Y}_{l^*,Q}^U \end{bmatrix} = \begin{bmatrix} \hat{H}_{l^*,1}^1 \left( \frac{\hat{H}_{l^*,1}^{1,\text{sum}}}{\hat{H}_{l^*,1}^1} - 1 \right) & \hat{H}_{l^*,2}^1 \left( \frac{\hat{H}_{l^*,1}^{1,\text{sum}}}{\hat{H}_{l^*,1}^1} - 1 \right) & \cdots & \hat{H}_{l^*,Q}^1 \left( \frac{\hat{H}_{l^*,1}^{1,\text{sum}}}{\hat{H}_{l^*,1}^1} - 1 \right) \\ \hat{H}_{l^*,1}^2 \left( \frac{\hat{H}_{l^*,1}^{2,\text{sum}}}{\hat{H}_{l^*,1}^2} - 1 \right) & \hat{H}_{l^*,2}^2 \left( \frac{\hat{H}_{l^*,1}^{2,\text{sum}}}{\hat{H}_{l^*,1}^2} - 1 \right) & \cdots & \hat{H}_{l^*,Q}^2 \left( \frac{\hat{H}_{l^*,1}^{2,\text{sum}}}{\hat{H}_{l^*,1}^2} - 1 \right) \\ \vdots & \vdots & \ddots & \vdots \\ \hat{H}_{l^*,1}^U \left( \frac{\hat{H}_{l^*,1}^{U,\text{sum}}}{\hat{H}_{l^*,1}^U} - 1 \right) & \hat{H}_{l^*,2}^U \left( \frac{\hat{H}_{l^*,1}^{U,\text{sum}}}{\hat{H}_{l^*,1}^U} - 1 \right) & \cdots & \hat{H}_{l^*,Q}^U \left( \frac{\hat{H}_{l^*,1}^{U,\text{sum}}}{\hat{H}_{l^*,1}^U} - 1 \right) \end{bmatrix}. \quad (26)$$

### B. Mean and Variance of the Proposed Estimator

From (15) and (17), we can express  $\tilde{Y}_{l^*,q}^u$  in (25) as

$$\tilde{Y}_{l^*,q}^u = H_{l^*,l^*,q}^u \underbrace{\left( \frac{\sum_{q=1}^Q H_{l^*,l^*,q}^u + \tilde{V}_p^{u,(2)}}{H_{l^*,l^*,1}^u + \tilde{V}_p^{u,(1)}} - 1 \right)}_{\textcircled{A}} + \tilde{W}_{l^*,q}^u, \quad (33)$$

where  $\tilde{V}_p^{u,(1)} = \mathbf{V}_p^{u,(1)} \mathbf{S}_p^H / K \sqrt{p_f}$ ,  $\tilde{V}_p^{u,(2)} = \mathbf{V}_p^{u,(2)} \mathbf{S}_p^H / K \sqrt{p_f}$ , and the power of both  $\tilde{V}_p^{u,(1)}$  and  $\tilde{V}_p^{u,(2)}$  is  $\sigma_v^2 / p_f$ . The complex-valued random variables  $\tilde{V}_p^{u,(1)}$ ,  $\tilde{V}_p^{u,(2)}$  and  $\tilde{W}_{l^*,q}^u$  in (33) all obey the Gaussian distribution and they are independent of each other. As  $\tilde{Y}_{l^*,q}^u$  is a nonlinear composite of  $\tilde{V}_p^{u,(1)}$ ,  $\tilde{V}_p^{u,(2)}$  and  $\tilde{W}_{l^*,q}^u$ , the statistical analysis of the proposed estimator is challenging. We concentrate on the expectation and variance of the proposed estimator  $\hat{H}_{l^*,l^*,q}^u$ .

Invoking the expression of  $\tilde{Y}_{l^*,q}^u$  in (33), the numerator and denominator in (31) and (32) can be rewritten as follows

$$\sum_{q=2}^Q \tilde{Y}_{l^*,q}^u = \underbrace{\left( \sum_{q=2}^Q H_{l^*,l^*,q}^u \right)}_{\textcircled{B}} + \left( \sum_{q=2}^Q \tilde{W}_{l^*,q}^u \right) = \textcircled{B} + \tilde{W}_{l^*,\text{sum}}^u, \quad (34)$$

$$\tilde{Y}_{l^*,1}^u \tilde{Y}_{l^*,1}^u = (H_{l^*,l^*,1}^u)^2 \textcircled{A}^2 + 2(H_{l^*,l^*,1}^u \tilde{W}_{l^*,1}^u) \textcircled{A} + (\tilde{W}_{l^*,1}^u)^2, \quad (35)$$

$$\tilde{Y}_{l^*,1}^u \tilde{Y}_{l^*,q}^u = (H_{l^*,l^*,1}^u H_{l^*,l^*,q}^u) \textcircled{A}^2 + (H_{l^*,l^*,q}^u \tilde{W}_{l^*,1}^u) \textcircled{A} + (H_{l^*,l^*,1}^u \tilde{W}_{l^*,q}^u) \textcircled{A} + (\tilde{W}_{l^*,1}^u \tilde{W}_{l^*,q}^u), \quad (36)$$

where  $\tilde{W}_{l^*,\text{sum}}^u = \sum_{q=2}^Q \tilde{W}_{l^*,q}^u$ . The expectation of  $\hat{H}_{l^*,l^*,1}^u$  is with respect to  $\tilde{W}_{l^*,1}^u$ ,  $\tilde{W}_{l^*,q}^u$ ,  $\tilde{V}_p^{u,(1)}$  and  $\tilde{V}_p^{u,(2)}$ . Therefore, we can express the expectation as follows [21, p. 117]

$$\begin{aligned} & \mathbb{E}_{\tilde{V}_p^{u,(2)}, \tilde{V}_p^{u,(1)}, \tilde{W}_{l^*,q}^u, \tilde{W}_{l^*,1}^u} \left\{ \hat{H}_{l^*,l^*,1}^u \right\} \\ &= \mathbb{E}_{\tilde{V}_p^{u,(2)}} \left\{ \mathbb{E}_{\tilde{V}_p^{u,(1)}} \left\{ \mathbb{E}_{\tilde{W}_{l^*,q}^u} \left\{ \mathbb{E}_{\tilde{W}_{l^*,1}^u} \left\{ \frac{\tilde{Y}_{l^*,1}^u \tilde{Y}_{l^*,1}^u}{\sum_{q=2}^Q \tilde{Y}_{l^*,q}^u} \right\} \right\} \right\} \right\} \\ &= \mathbb{E}_{\tilde{V}_p^{u,(2)}} \left\{ \mathbb{E}_{\tilde{V}_p^{u,(1)}} \left\{ \mathbb{E}_{\tilde{W}_{l^*,q}^u} \left\{ \frac{(H_{l^*,l^*,1}^u)^2 \textcircled{A}^2}{\textcircled{B} + \tilde{W}_{l^*,\text{sum}}^u} \right\} \right\} \right\}. \end{aligned} \quad (37)$$

Having arrived at  $(H_{l^*,l^*,1}^u)^2 \textcircled{A}^2 / (\textcircled{B} + \tilde{W}_{l^*,\text{sum}}^u)$  by evaluating the expectation of  $\tilde{Y}_{l^*,1}^u \tilde{Y}_{l^*,1}^u / \sum_{q=2}^Q \tilde{Y}_{l^*,q}^u$  with respect to the random variables  $\tilde{W}_{l^*,1}^u$  and to compute the expectation of  $(H_{l^*,l^*,1}^u)^2 \textcircled{A}^2 / (\textcircled{B} + \tilde{W}_{l^*,\text{sum}}^u)$  with respect to the random variables  $\tilde{W}_{l^*,q}^u$ , we have the following theorem.

**Theorem 1:** If a complex-valued random variable  $\gamma = a + bj = \rho e^{-j\phi}$ ,  $\rho \geq 0$ , has a zero mean, then the  $n$ th power of the function  $\gamma$ , namely  $\gamma^n$ , where  $n$  is a natural number, also has a zero mean.

*Proof:* Because  $\mathbb{E}\{\gamma\} = 0$ , we have

$$\mathbb{E}\{e^{-j\phi}\} = \int_0^{2\pi} e^{-j\phi} d\phi = 0. \quad (38)$$

Then, we can compute the mean of  $\gamma^n$  as follows

$$\begin{aligned} \mathbb{E}\{\gamma^n\} &= \mathbb{E}\{\rho^n\} \int_0^{2\pi} e^{-jn\phi} d\phi \\ &\stackrel{\varphi \triangleq n\phi}{=} \mathbb{E}\{\rho^n\} \frac{1}{n} \sum_{n'=1}^n \int_0^{n'2\pi} e^{-j\varphi} d\varphi = 0. \end{aligned} \quad (39)$$

This completes the proof.  $\blacksquare$

According to the classic Taylor series theory of complex-valued random variables [22, p. 189], under the condition of  $|\tilde{W}_{l^*,\text{sum}}^u / \textcircled{B}| < 1$  with probability one,<sup>3</sup> the function of  $1 / (\textcircled{B} + \tilde{W}_{l^*,\text{sum}}^u)$  has the power series representation given by

$$\frac{1}{\textcircled{B} + \tilde{W}_{l^*,\text{sum}}^u} = \frac{1}{\textcircled{B}} \left( 1 + \sum_{n=1}^{\infty} (-1)^n \left( \frac{\tilde{W}_{l^*,\text{sum}}^u}{\textcircled{B}} \right)^n \right). \quad (40)$$

The series at the right-hand side of (40) is understood to converge in probability. Using Theorem 1 and (40), we obtain the expectation of  $\hat{H}_{l^*,l^*,1}^u$  as follows

$$\mathbb{E}_{\tilde{V}_p^{u,(2)}, \tilde{V}_p^{u,(1)}, \tilde{W}_{l^*,q}^u, \tilde{W}_{l^*,1}^u} \left\{ \hat{H}_{l^*,l^*,1}^u \right\} = H_{l^*,l^*,1}^u. \quad (41)$$

In the same way, the expectation of  $\hat{H}_{l^*,l^*,q}^u$ ,  $2 \leq q \leq Q$ , is given by

$$\mathbb{E}_{\tilde{V}_p^{u,(2)}, \tilde{V}_p^{u,(1)}, \tilde{W}_{l^*,q}^u, \tilde{W}_{l^*,1}^u} \left\{ \hat{H}_{l^*,l^*,q}^u \right\} = H_{l^*,l^*,q}^u. \quad (42)$$

From (41) and (42), we conclude that the proposed estimator is an unbiased one.

To provide an approximate expression for the proposed estimator's variance, we have the following approximation.

**Approximation 1:** Given that  $A = A_R + jA_I$  is a complex-valued constant and  $\zeta = \zeta_R + j\zeta_I$  is a complex-valued random variable obeying  $\mathcal{CN}(0, \sigma_\zeta^2)$  as well as assuming that

$$\Pr \left\{ \frac{A_R \zeta_R + A_I \zeta_I + |\zeta|^2}{|A|^2} < 1 \right\} = 1, \quad (43)$$

we have the following approximate formula

$$\mathbb{E} \left\{ \frac{1}{(A + \zeta)(A^\dagger + \zeta^\dagger)} \right\} \approx \frac{1}{|A|^2} + \frac{1}{|A|^4} \sigma_\zeta^2. \quad (44)$$

*Proof:* We use Taylor series theory of a real-valued random variable [23, p. 134] to express  $\mathbb{E}\{1/(A + \zeta)(A^\dagger + \zeta^\dagger)\}$  as

$$\begin{aligned} & \mathbb{E} \left\{ \frac{1}{(A + \zeta)(A^\dagger + \zeta^\dagger)} \right\} \\ &= \frac{1}{|A|^2} \mathbb{E} \left\{ \left( 1 + \sum_{n=1}^{\infty} (-1)^n \left( \frac{A_R \zeta_R + A_I \zeta_I + |\zeta|^2}{|A|^2} \right)^n \right) \right\} \\ &\approx \frac{1}{|A|^2} + \frac{1}{|A|^4} \sigma_\zeta^2, \end{aligned} \quad (45)$$

<sup>3</sup>Note that  $\Pr(|\tilde{W}_{l^*,\text{sum}}^u / \textcircled{B}| \geq 1) \approx e^{-((Q-1)\beta_{l^*,l^*} / \sigma_w^2)} \rightarrow 0$  as  $Q \rightarrow \infty$ . Therefore,  $\Pr(|\tilde{W}_{l^*,\text{sum}}^u / \textcircled{B}| < 1) = 1$  can be justified.

where the approximation arises by dropping the high-order terms  $(1/|A|^2)\mathbb{E}\{(A_R\zeta_R + A_I\zeta_I + |\zeta|^2/|A|^2)^n\}$  for  $n \geq 2$ . This completes the proof. ■

Using the approximate formula of (45) and Taylor series theory, after some further manipulations, we have the following approximate expression of the upper bound for the variance of  $\hat{H}_{l^*,l^*,1}^u$

$$\text{var} \left\{ \hat{H}_{l^*,l^*,1}^u \right\} \lesssim \frac{2 \left| H_{l^*,l^*,1}^u \right|^2 \sigma_v^2}{p_f |\mathbb{B}|^2} + \frac{\sigma_v^2}{p_f} + \frac{(2(Q-1)\sigma_w^2) \left| H_{l^*,l^*,1}^u \right|^4}{p_r |\mathbb{B}|^4} + \frac{8 \left| H_{l^*,l^*,1}^u \right|^2 \sigma_w^2}{p_r |\mathbb{B}|^2}, \quad (46)$$

where  $1 \leq u \leq U$ . Similarly, for  $1 \leq u \leq U$  and  $2 \leq q \leq Q$ ,

$$\begin{aligned} \text{var} \left\{ \hat{H}_{l^*,l^*,q}^u \right\} \lesssim & \frac{2 \left| H_{l^*,l^*,q}^u \right|^2 \sigma_v^2}{p_f |\mathbb{B}|^2} + \frac{\left| H_{l^*,l^*,q}^u \right|^2 \sigma_v^2}{p_f \left| H_{l^*,l^*,1}^u \right|^2} \\ & + \frac{2 \left| H_{l^*,l^*,q}^u \right|^2 \sigma_w^2}{p_r |\mathbb{B}|^2} + \frac{2 \left| H_{l^*,l^*,1}^u \right|^2 \sigma_w^2}{p_r |\mathbb{B}|^2} \\ & + \frac{2 \left| H_{l^*,l^*,1}^u \right|^2 \left| H_{l^*,l^*,q}^u \right|^2 (Q-1) \sigma_w^2}{p_r |\mathbb{B}|^4}. \end{aligned} \quad (47)$$

We point out again that  $\mathbb{B} = \sum_{q=2}^Q H_{l^*,l^*,q}^u$ .

### C. Downlink Transmission

By exploiting the reciprocity of the TDD UL and DL channels [6], [10], [15], the BS is capable of acquiring the DL CE by simply assuming that it is the same as the UL CE generated with the aid of the DL-pilots and the UL-pilots, as discussed in Section III-A and Section III-B, respectively. Then the TP matrix  $\mathbf{A}_l \in \mathbb{C}^{Q \times U}$  can be designed using the zero-forcing (ZF), the MMSE, the minimum bit error rate or the nonlinear vector precoding criteria [16]–[20]. Let the  $k$ -th information-bearing OFDM symbols transmitted to the  $U$  users of cell  $l$  be  $\mathbf{S}_l[k] = [S_l^1[k] S_l^2[k] \cdots S_l^U[k]]$ . We assume that the TP scheme ensures that  $\mathbb{E}[\mathbf{S}_l[k]] = \mathbf{0}$ ,  $\mathbb{E}[\mathbf{S}_l[k] \mathbf{S}_l^H[k]] = \mathbf{I}$ , and  $\text{tr}\{\mathbf{A}_l^H \mathbf{A}_l\} = 1$ , which implies that the average power constraint at the BS is satisfied [5], [10].

The  $l$ -th cell's BS transmits its TP signal of  $\mathbf{A}_l \mathbf{S}_l^T[k]$  in the DL to the  $U$  supported MSs. Therefore, the noisy signal  $\mathbf{Z}_{l^*}[k] \in \mathbb{C}^{1 \times U}$  received by the  $U$  MSs in cell  $l^*$  is constituted by the superimposed DL transmissions of the  $L$  cells' BSs plus the channel's AWGN, which is expressed as

$$\begin{aligned} \mathbf{Z}_{l^*}^T[k] = & \underbrace{\sqrt{p_f} \mathbf{H}_{l^*,l^*} \mathbf{A}_{l^*} \mathbf{S}_{l^*}^T[k]}_{\text{Desired term}} \\ & + \underbrace{\sqrt{p_f} \sum_{l=1, l \neq l^*}^L \mathbf{H}_{l^*,l} \mathbf{A}_l \mathbf{S}_l^T[k]}_{\text{Inter-cell interference}} + \mathbf{V}_{l^*}^T[k], \end{aligned} \quad (48)$$

where  $p_f$  is the average transmit power of the BS and  $\mathbf{V}_{l^*} \in \mathbb{C}^{1 \times U}$  is the FD representation of the DL channel AWGN vector, whose elements have the power of  $\sigma_v^2$ . Thus the signal received by the  $u^*$ -th MS in cell  $l^*$  is given by

$$Z_{l^*}^{u^*}[k] = \sqrt{p_f} \sum_{l=1}^L \sum_{u=1}^U \mathbf{H}_{l^*,l}^{u^*} (\mathbf{A}_l^u)^T S_l^u[k] + V_{l^*}^{u^*}[k], \quad (49)$$

where  $\mathbf{H}_{l^*,l}^{u^*}$  is the  $u^*$ -th row of  $\mathbf{H}_{l^*,l}$ ,  $(\mathbf{A}_l^u)^T$  is the  $u$ -th column of  $\mathbf{A}_l$  and  $V_{l^*}^{u^*}[k]$  is the  $u^*$ -th element of  $\mathbf{V}_{l^*}[k]$ .

### D. Achievable Throughput Analysis

To further investigate the benefits of eliminating the pilot contamination imposed by the adjacent BSs on the CE, we derive the lower bound of the achievable DL throughput based on the ZF precoding matrix given by [6]

$$\mathbf{A}_l = \frac{\tilde{\mathbf{A}}_l}{\sqrt{\text{tr}\{\tilde{\mathbf{A}}_l \tilde{\mathbf{A}}_l^H\}}} \text{ with } \tilde{\mathbf{A}}_l = \hat{\mathbf{H}}_{l,l}^H \left( \hat{\mathbf{H}}_{l,l} \hat{\mathbf{H}}_{l,l}^H \right)^{-1}. \quad (50)$$

By denoting the DL data to be transmitted to the  $U$  users in cell  $l$  as  $\mathbf{S}_l = [S_l^1 S_l^2 \cdots S_l^U]$ , we can rewrite the signal (49) received by the  $u^*$ -th user in the  $l^*$ -th cell as follows

$$\begin{aligned} Z_{l^*}^{u^*} = & \sqrt{p_f} \mathbb{E} \left\{ \mathbf{H}_{l^*,l^*}^{u^*} (\mathbf{A}_{l^*}^{u^*})^T \right\} S_{l^*}^{u^*} \\ & + \sqrt{p_f} \left( \mathbf{H}_{l^*,l^*}^{u^*} (\mathbf{A}_{l^*}^{u^*})^T - \mathbb{E} \left\{ \mathbf{H}_{l^*,l^*}^{u^*} (\mathbf{A}_{l^*}^{u^*})^T \right\} \right) S_{l^*}^{u^*} \\ & + \underbrace{\sqrt{p_f} \sum_{u=1, u \neq u^*}^U \mathbf{H}_{l^*,l^*}^{u^*} (\mathbf{A}_{l^*}^u)^T S_{l^*}^u}_{\text{Intra-cell interference}} \\ & + \underbrace{\sqrt{p_f} \sum_{l=1, l \neq l^*}^L \sum_{u=1}^U \mathbf{H}_{l^*,l}^{u^*} (\mathbf{A}_l^u)^T S_l^u + V_{l^*}^{u^*}}_{\text{Inter-cell interference}}. \end{aligned} \quad (51)$$

The achievable rate of the DL is the function of the signal-to-interference-plus-noise ratio (SINR) given by

$$R_{l^*}^{u^*} = \log_2 \left( 1 + \text{SINR}_{l^*}^{u^*} \right), \quad (52)$$

where the SINR of the  $u^*$ -th user in the  $l^*$ -th cell is given in (53) at the top of the next page. (See equation at bottom of page)

Since the objective of designing the ZF precoding matrix  $\mathbf{A}_{l^*}$  is to minimize the interference imposed by the MSs within the same serving cell, the residual intra-cell interference which is the third term in (51) is significantly lower than the inter-cell interference of the fourth term. Therefore, we may omit the term  $\sum_{u=1, u \neq u^*}^U \mathbb{E}\{|\mathbf{H}_{l^*,l^*}^{u^*} (\mathbf{A}_{l^*}^u)^T|^2\}$  imposed by the intra-cell interference in the SINR expression. Additionally, the term  $\sigma_v^2/p_f$  in the denominator of (53), which is contributed by the AWGN is also much smaller than the inter-cell interference. Thus, we

may also omit the term  $\sigma_v^2/p_f$  from (53). Then, the  $\text{SINR}_{l^*}^{u^*}$  of (53) can be approximated as

$$\text{SINR}_{l^*}^{u^*} \approx \frac{\left| \mathbb{E} \left\{ \mathbf{H}_{l^*,l^*}^{u^*} (\mathbf{A}_{l^*}^{u^*})^T \right\} \right|^2}{\text{var} \left\{ \mathbf{H}_{l^*,l^*}^{u^*} (\mathbf{A}_{l^*}^{u^*})^T \right\} + \sum_{l=1, l \neq l^*}^L \sum_{u=1}^U \mathbb{E} \left\{ \left| \mathbf{H}_{l^*,l}^{u^*} (\mathbf{A}_l^u)^T \right|^2 \right\}}. \quad (54)$$

We will resort to using Monte Carlo simulations to evaluate  $\left| \mathbb{E} \left\{ \mathbf{H}_{l^*,l^*}^{u^*} (\mathbf{A}_{l^*}^{u^*})^T \right\} \right|^2$ ,  $\text{var} \left\{ \mathbf{H}_{l^*,l^*}^{u^*} (\mathbf{A}_{l^*}^{u^*})^T \right\}$  and  $\mathbb{E} \left\{ \left| \mathbf{H}_{l^*,l}^{u^*} (\mathbf{A}_l^u)^T \right|^2 \right\}$  in the calculation of the achievable throughput (52), since there are no closed-form expressions for them.

#### IV. SIMULATION RESULTS

The default values of the various parameters for our simulated multi-cell TDD system are summarized in Table I, where  $\eta_{l^*,l,q,p}^{q_l}$  denotes the path loss coefficient for the  $p$ -th path of the link between the  $q_l$ -th antenna of the  $l$ -th BS and the  $q_{l^*}$ -th antenna of the  $l^*$ -th BS with the associated path AoA  $\theta_{l^*,l,q,p}^{q_l}$ . Unless otherwise specified, these default parameter values were used throughout. The UFR was assumed and the same pre-assigned PSs were employed for the UL CE by all the BSs. The signal-to-noise ratio (SNR) of the system was defined as  $E_b/N_0$ , where  $E_b$  denoted the energy per bit and  $N_0$  denoted the power of the channel AWGN. All the path AoAs,  $\theta_{l^*,l,q,p}^u$  and  $\theta_{l^*,l,q,p}^{q_l}$ , were independently and identically distributed (i.i.d.) Gaussian random variables with the mean  $\bar{\theta} = 90^\circ$  and the standard deviation  $\sigma_{\text{AoA}} = 90^\circ$ . The pilot contamination reduction schemes of the LMMSE CE using the aligned PSs and the LMMSE CE relying on the staggered PSs as well as the DL precoding relying on perfect CSI were included as the benchmarks for comparison with our proposed pilot contamination elimination scheme.

The LMMSE CE using the aligned PSs [6] refers to the estimator that relies on all the MSs of all the cells simultaneously transmitting their UL PSs, which occupies  $K$  OFDM symbol durations for its UL training. More details about the LMMSE CE using the aligned PSs can be found in reference [6]. Furthermore, observe in (14) that the LMMSE CE using aligned PSs fails to reduce the pilot contamination by employing longer OFDM PSs, although it is capable of reducing the effects of AWGN. However, the effect of AWGN on the achievable capacity can be neglected, as the number

TABLE I  
DEFAULT PARAMETERS USED IN THE SIMULATED MULTIPLE-ANTENNA AIDED MULTI-CELL TDD SYSTEM

Number of cells $L$	7
Number of MSs per-cell $U$	4
Number of antennas at each BS $Q$	50
Average transmit power at each MS $p_r$	0dB
Average transmit power at each BS $p_f$	10dB
Path loss coefficients $\beta_{l^*,l,q,p}^u = \beta_{l^*,l}, l \neq l^*$	0.3
Path loss coefficients $\beta_{l^*,l,q,p}^{q_l} = \beta_{l^*,l^*}$	1.0
Path loss coefficients $\eta_{l^*,l,q_{l^*},p}^{q_l} = \eta_{l^*,l}, l \neq l^*$	0.5
Mean of path AoAs $\theta$	$90^\circ$
Standard deviation of path AoAs $\sigma_{\text{AoA}}$	$90^\circ$
Antenna spacing $D$	$\frac{\lambda}{2}$
Number of paths $P$	50
Pilot length $K$	20

of antennas grows without limit [3], [6]. Hence, even if we provide  $[(L+3)L]$ -length PSs for the LMMSE CE using the aligned PSs, its performance remains almost the same as its performance associated with using  $K$ -length PSs. In our stimulated UFR system, the LMMSE CE with the aligned PSs used the same pre-assigned UL PSs in every cell, just as our scheme, and the estimate of  $\mathbf{H}_{l^*,l^*}$  provided by the  $l^*$ -th BS is given by (10), or equivalently by (14) if  $Q$  is large. The LMMSE CE relying on the aligned PSs assumes the exact knowledge of both the channel's noise variance and of all the path loss coefficients  $\beta_{l^*,l,q,p}^u$ . Naturally, this is an unrealistic assumption, since in practice it is unclear how the serving BS can acquire the second-order statistics of all the related UL channels  $\mathbf{H}_{l^*,l}$ . Nevertheless, in our simulations, the LMMSE CE using the aligned PSs was provided with all the exact path loss coefficients  $\beta_{l^*,l,q,p}^u (= \beta_{l^*,l})$ , therefore its performance represented an 'idealized' scenario, which is never encountered in reality.

For the LMMSE CE using the staggered PSs [1], [6], the MSs roaming in different cells transmit their UL PSs at non-overlapping instances. Specifically, when the MSs in cell  $l^*$  are transmitting the pre-assigned UL PSs, all the MSs of all other cells can only perform their DL data transmissions. Although the length of the UL training period remains  $K$  OFDM symbol durations for each cell, the total training period for an  $L$ -cell system spans over  $LK$  OFDM symbol durations with the adjacent cells performing their DL data transmission in a coordinated manner for LMMSE CE using the staggered PSs. This LMMSE CE using the staggered PSs, as detailed in Appendix, requires the knowledge of the FDCHTF matrices  $\mathbf{G}_{l^*,l} \in \mathbb{C}^{Q \times Q}$  between the  $l$ -th BSs for  $l \neq l^*$  and the  $l^*$ -th BS as well as the channel noise variance and all the path loss coefficients  $\beta_{l^*,l,q,p}^u$  for the serving cell. Estimating  $\mathbf{G}_{l^*,l}$  requires a huge training

$$\text{SINR}_{l^*}^{u^*} = \frac{\left| \mathbb{E} \left\{ \mathbf{H}_{l^*,l^*}^{u^*} (\mathbf{A}_{l^*}^{u^*})^T \right\} \right|^2}{\frac{\sigma_v^2}{p_f} + \text{var} \left\{ \mathbf{H}_{l^*,l^*}^{u^*} (\mathbf{A}_{l^*}^{u^*})^T \right\} + \sum_{u=1, u \neq u^*}^U \mathbb{E} \left\{ \left| \mathbf{H}_{l^*,l^*}^{u^*} (\mathbf{A}_{l^*}^u)^T \right|^2 \right\} + \sum_{l=1, l \neq l^*}^L \sum_{u=1}^U \mathbb{E} \left\{ \left| \mathbf{H}_{l^*,l}^{u^*} (\mathbf{A}_l^u)^T \right|^2 \right\}}. \quad (53)$$

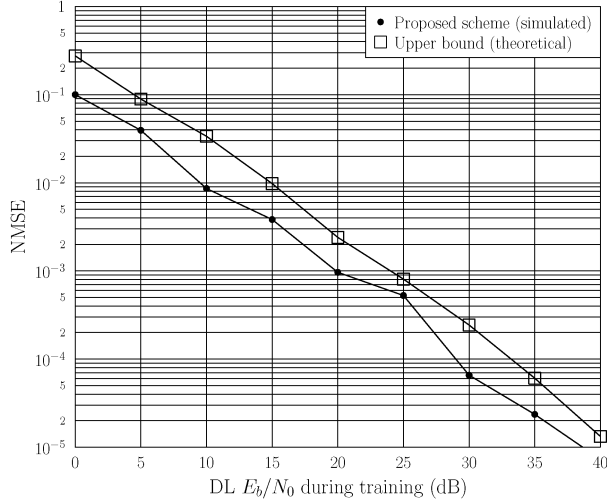


Fig. 4. Normalized mean square error performance of our proposed channel estimator as the function of the DL training SNR, given the UL training SNR of  $E_b/N_0 = 20$  dB. The simulated network is specified in Table I.

overhead, since the number of antennas  $Q$  is large, and the estimation accuracy is limited due to the inherent inter-cell interference. The path loss coefficients  $\beta_{l^*,l^*,q,p}^u$  are contained inside the FDCHTF vectors  $\mathbf{H}_{l^*,l^*}^u$  which have to be estimated, and it is unrealistic to assume that they are known before all the  $\mathbf{H}_{l^*,l^*}^u$  are estimated. In our simulations, the LMMSE CE with the staggered PSs was provided with all the true  $\mathbf{G}_{l^*,l^*}$  and all the true  $\beta_{l^*,l^*,q,p}^u (= \beta_{l^*,l^*}^u)$ . Therefore, its performance represented an ‘idealized’ scenario, which is difficult to approach in reality.

We firstly examined the normalized mean square error (NMSE), which is defined as

$$\text{NMSE}_{\text{simu}} = \frac{\sum_{u=1}^U \sum_{q=1}^Q \left| \hat{H}_{l^*,l^*,q}^u - H_{l^*,l^*,q}^u \right|^2}{\sum_{u=1}^U \sum_{q=1}^Q \left| H_{l^*,l^*,q}^u \right|^2}, \quad (55)$$

where  $H_{l^*,l^*,q}^u$  and  $\hat{H}_{l^*,l^*,q}^u$  are the true channel and its estimate, respectively. Because our scheme consists of the DL training and UL training stages, we began by investigating the NMSE of our estimator as a function of the DL training SNR, given the UL SNR of  $E_b/N_0 = 20$  dB. The simulation results for  $\text{NMSE}_{\text{simu}}$  obtained by averaging over 100 channel realizations with different DL training SNRs are depicted in Fig. 4, which shows standard behavior of an unbiased estimator whose estimation accuracy is determined by the signal’s SNR. To verify the accuracy of the theoretical analysis carried out in Section III-B to derive the upper bounds of our estimator variances  $\text{var}\{\hat{H}_{l^*,l^*,q}^u\}$  given in (46) and (47) for  $1 \leq u \leq U$  and  $1 \leq q \leq Q$ , we also plot the theoretical upper bound of the NMSE, which is defined as

$$\text{NMSE}_{\text{theo-ub}} = \frac{\sum_{u=1}^U \sum_{q=1}^Q \text{var}\{\hat{H}_{l^*,l^*,q}^u\}}{\sum_{u=1}^U \sum_{q=1}^Q \left| H_{l^*,l^*,q}^u \right|^2}, \quad (56)$$

in Fig. 4. Observe from Fig. 4 that  $\text{NMSE}_{\text{theo-ub}}$  is just above  $\text{NMSE}_{\text{simu}}$  and exhibits the same trend as the latter. We next

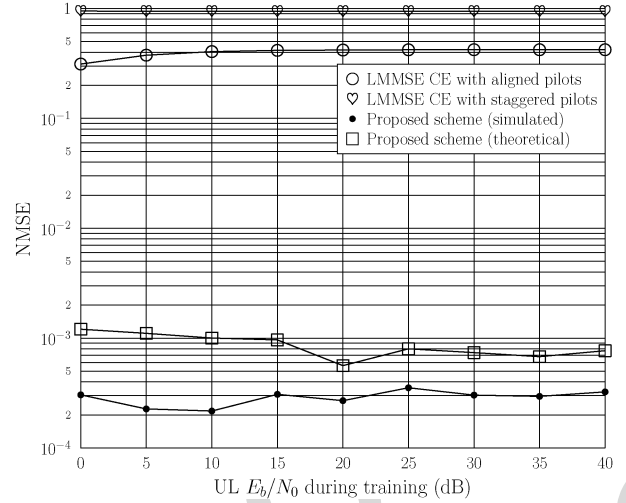


Fig. 5. Normalized mean square error performance of three channel estimators as the function of the UL training SNR. The simulated network is specified in Table I, and the DL training SNR for our proposed estimator is  $E_b/N_0 = 25$  dB.

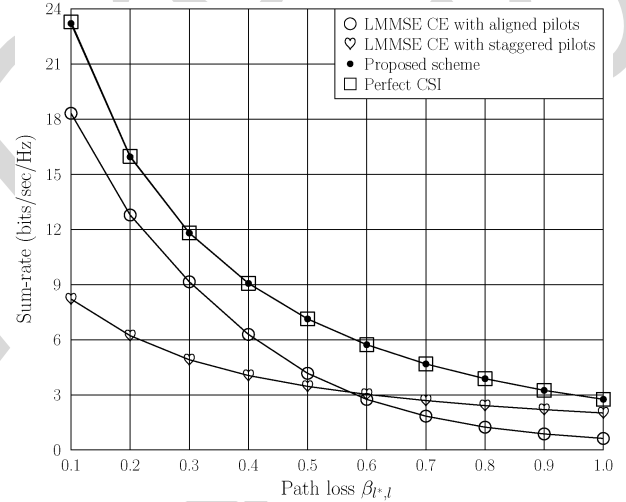


Fig. 6. Achievable sum-rate performance as the function of the path loss coefficient  $\beta_{l^*,l^*}$  by the three estimators, in comparison with the perfect CSI bound. The UL training SNR is  $E_b/N_0 = 20$  dB for all the three estimators and additionally the DL training SNR is  $E_b/N_0 = 25$  dB for our proposed scheme. The rest of the network parameters are specified in Table I.

studied the NMSE of our estimator as a function of the UL training SNR, given the DL SNR of  $E_b/N_0 = 25$  dB. The results obtained again by averaging over 100 channel realizations are shown in Fig. 5, where the NMSE performance of the LMMSE CE relying on aligned PSs and the LMMSE CE using staggered PSs are also given for comparison. It can be seen from Fig. 5 that the UL training SNR only has very slight influence on the achievable NMSE performance of all the three channel estimators, because the dominant factors which determine the UL estimation accuracy are the intra-cell and/or inter-cell interferences and/or the pilot contamination. Observe from Fig. 5 that our estimator significantly outperforms the other two estimators.

To investigate the achievable sum-rate performance by the three estimators under various network conditions, we set the UL training SNR to  $E_b/N_0 = 20$  dB for all the three estimators and additionally set the DL training SNR to  $E_b/N_0 = 25$  dB



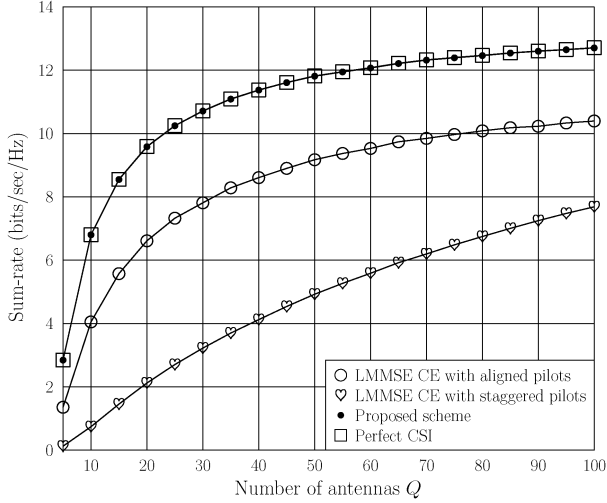


Fig. 7. Achievable sum-rate performance as the function of the number of antennas  $Q$  by the three estimators, in comparison with the perfect CSI bound. The UL training SNR is  $E_b/N_0 = 20$  dB for all the three estimators and additionally the DL training SNR is  $E_b/N_0 = 25$  dB for our proposed scheme. The rest of the network parameters are specified in Table I.

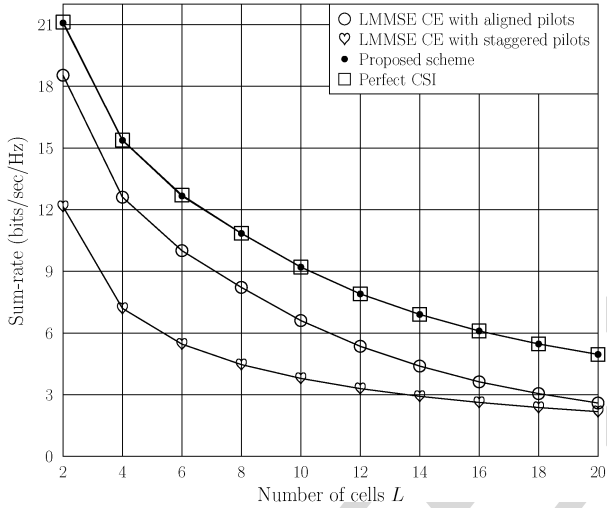


Fig. 8. Achievable sum-rate performance as the function of the number of cells  $L$  by the three estimators, in comparison with the perfect CSI bound. The UL training SNR is  $E_b/N_0 = 20$  dB for all the three estimators and additionally the DL training SNR is  $E_b/N_0 = 25$  dB for our proposed scheme. For  $L \geq 8$ , the path loss coefficients between the desired cell and its neighboring cells are set to  $\beta_{l^*,l} = 0.3$  and the path loss coefficients between the serving cell and the rest far-away cells are set to  $\beta_{l^*,l} = 0.2$ , while the rest of the network parameters are specified in Table I.

for our proposed scheme. Fig. 6 portrays the sum-rate performance versus the path loss coefficient  $\beta_{l^*,l}$  between different cells by the three estimation schemes, in comparison to the perfect CSI bound. Observe in Fig. 6 that our scheme significantly outperforms both the LMMSE CE relying on aligned PSs and the LMMSE CE relying on staggered PSs. Furthermore, the achievable sum-rate of our scheme approaches the sum-rate of the perfect CSI bound. The sum-rates achieved by the different schemes as a function of the number of antennas  $Q$  are shown in Fig. 7, where it can be seen that the sum-rate achieved by our scheme approaches the perfect CIS bound, which is substantially better than those of the LMMSE CE relying on aligned PSs and the LMMSE CE using staggered PSs. The achievable sum-rate versus the number of cells  $L$  is investigated in Fig. 8

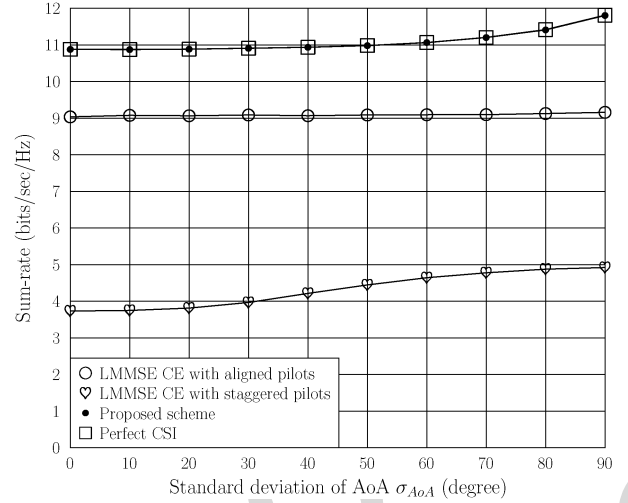


Fig. 9. Achievable sum-rate performance as the function of the path AoA's standard deviation  $\sigma_{AoA}$  by the three estimators, in comparison with the perfect CSI bound. The UL training SNR is  $E_b/N_0 = 20$  dB for all the three estimators and additionally the DL training SNR is  $E_b/N_0 = 25$  dB for our scheme. The rest of the network parameters are specified in Table I.

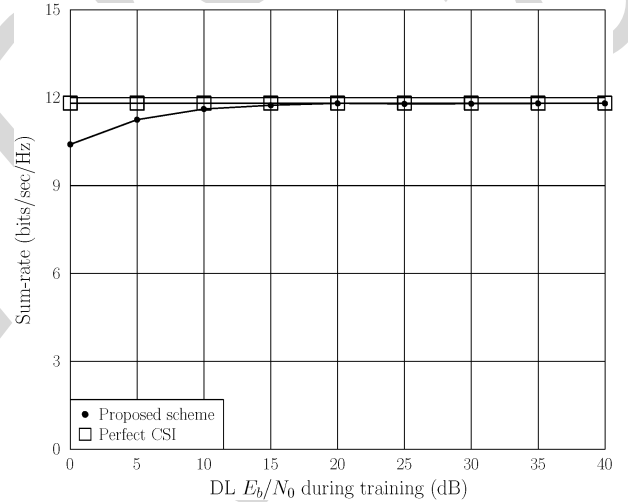


Fig. 10. Achievable sum-rate performance of our proposed channel estimator as the function of the DL training SNR, given the UL training SNR of  $E_b/N_0 = 20$  dB, in comparison with the perfect CSI bound. The simulated network is specified in Table I.

where, for  $L \geq 8$ , the path loss coefficients between the desired cell and its neighboring cells were  $\beta_{l^*,l} = 0.3$  as specified in Table I but the path loss coefficients between the serving cell and the rest far-away cells were set to  $\beta_{l^*,l} = 0.2$ . It can be seen from Fig. 8 that the proposed scheme approaches the perfect CSI bound, and it significantly outperforms both the LMMSE CE using aligned PSs and the LMMSE CE relying on staggered PSs. Fig. 9 depicts the achievable sum-rate versus the standard deviation  $\sigma_{AoA}$  of all the path AoAs. Observe from Fig. 9 that all the schemes are only slightly affected by  $\sigma_{AoA}$ . Again, our scheme approaches the perfect CSI bound and considerably outperforms the other two estimators.

In order to investigate the effect of the DL training accuracy on the achievable sum-rate of our scheme, we set the UL training SNR to  $E_b/N_0 = 20$  dB and varied the DL training SNR. The sum-rate performance attained by our scheme is shown in

Fig. 10, where it is seen that the sum-rate attained by our estimator deviates slightly from the perfect CSI bound only in the region of the DL training SNR of  $E_b/N_0 < 15$  dB.

## V. CONCLUSIONS

A novel pilot contamination elimination scheme has been proposed for multi-cell TDD and OFDM based massive MIMO systems, which relies on the two stages of the DL training and the scheduled UL training. In the DL CE stage, each BS transmits its DL PSs to its serving MSs for them to estimate the user-specific DL FDCHTFs, which are then embedded in the MSs' UL PSs to be used in the second stage of UL training. A scheduled UL training procedure allows each BS in turn to estimate its MSs' FDCHTFs accurately by eliminating the pilot contamination imposed by the surrounding cells. Unlike some existing pilot contamination reduction schemes, such as the LMMSE CE relying on aligned PSs and the LMMSE CE using staggered PSs, our scheme does not require the knowledge of the second-order statistics of the MIMO CIRs. A drawback of our scheme is that it expands the training overhead by a factor equal to the number of interfering cells. Our extensive simulation results demonstrate that the proposed estimator significantly outperforms both the LMMSE CE relying on aligned PSs and the LMMSE CE using staggered PSs. The ZF precoding aided DL data transmission based on the estimated FDCHTF matrix provided by our estimator is capable of approaching the average sum-rate bound of the perfect CSI scenario.

## APPENDIX

### THE LMMSE CE WITH THE STAGGERED PSS

The signal received at the  $l^*$ -th BS during its UL training consists of the UL pilot transmissions of the MSs in cell  $l^*$  and all the DL data transmissions from all the other BSs, which can be expressed as

$$\mathbf{Y}_{l^*} = \sqrt{p_r} \mathbf{X}_{l^*} \mathbf{H}_{l^*,l^*} + \sqrt{p_f} \sum_{l=1, l \neq l^*}^L \mathbf{D}_l^T \mathbf{G}_{l^*,l} + \mathbf{W}_{l^*}, \quad (57)$$

where  $\mathbf{D}_l = \mathbf{A}_l \bar{\mathbf{S}}_l^T \in \mathbb{C}^{Q \times K}$  is the transmitted DL data of cell  $l$  with  $\bar{\mathbf{S}}_l = [\mathbf{S}_l^T[1] \mathbf{S}_l^T[2] \cdots \mathbf{S}_l^T[K]]$  and  $\mathbf{G}_{l^*,l} \in \mathbb{C}^{Q \times Q}$  is the FDCHTF matrix connecting the BS of cell  $l$  to the BS of cell  $l^*$  on subcarrier  $n$ . The link between the  $q_l$ -th antenna of the  $l$ -th BS and the  $q_{l^*}$ -th antenna of the  $l^*$ -th BS over the  $N$  subcarriers is represented by the FDCHTF row vector  $\mathbf{G}_{l^*,l}^{q_l} = [G_{l^*,l,q_{l^*}}^{q_l}[1] G_{l^*,l,q_{l^*}}^{q_l}[2] \cdots G_{l^*,l,q_{l^*}}^{q_l}[N]]$  as

$$\mathbf{G}_{l^*,l,q_{l^*}}^{q_l} = \mathbf{g}_{l^*,l,q_{l^*}}^{q_l} \mathbf{F}^T, \quad (58)$$

with the TD CIR vector  $\mathbf{g}_{l^*,l,q_{l^*}}^{q_l} \in \mathbb{C}^{1 \times P}$  defined by

$$\mathbf{g}_{l^*,l,q_{l^*}}^{q_l} = \begin{bmatrix} \alpha_{l^*,l,q_{l^*}}^{q_l} e^{-j2\pi \frac{(q_{l^*}-1)D}{\lambda} \cos(\theta_{l^*,l,q_{l^*}}^{q_l})} \cdots \\ \alpha_{l^*,l,q_{l^*},P}^{q_l} e^{-j2\pi \frac{(q_{l^*}-1)D}{\lambda} \cos(\theta_{l^*,l,q_{l^*},P}^{q_l})} \end{bmatrix}, \quad (59)$$

and the complex-valued tap  $\alpha_{l^*,l,q_{l^*}}^{q_l}$  given by

$$\alpha_{l^*,l,q_{l^*}}^{q_l} = e^{-j\varphi_{l^*,l,q_{l^*}}^{q_l}} \sqrt{\eta_{l^*,l,q_{l^*}}^{q_l}}, \quad (60)$$

similar to (3) to (5). Again, we may assume that the path loss coefficients obey  $\eta_{l^*,l,q_{l^*}}^{q_l} = \eta_{l^*,l}$  for  $1 \leq l, q_{l^*} \leq Q$  and  $1 \leq p \leq P$ . Note that the  $q_l$ -th row and  $q_{l^*}$ -th column element of  $\mathbf{G}_{l^*,l}$  is simply  $G_{l^*,l,q_{l^*}}^{q_l}[n]$ , similar to the definition of  $\mathbf{H}_{l^*,l}$ . Recall that the  $u$ -th row of  $\mathbf{H}_{l^*,l^*}$  is denoted as  $\mathbf{H}_{l^*,l^*}^u \in \mathbb{C}^{1 \times Q}$ ,  $1 \leq u \leq U$ . Let us now define

$$\mathbf{H}_{l^*,l^*,\text{vec}}^u = \underbrace{[1 \ 1 \cdots 1]}_{K \text{ times}} \otimes (\mathbf{H}_{l^*,l^*}^u)^{\dagger} \in \mathbb{C}^{1 \times KQ}, \quad (61)$$

$$\begin{aligned} \mathbf{C}_{\mathbf{H}_{l^*,l^*,\text{vec}}^u} &= \mathbb{E} \left\{ (\mathbf{H}_{l^*,l^*,\text{vec}}^u)^H \mathbf{H}_{l^*,l^*,\text{vec}}^u \right\} \\ &= \mathbb{E} \left\{ \underbrace{[1 \ 1 \cdots 1]^H}_{K \text{ times}} \underbrace{[1 \ 1 \cdots 1]}_{K \text{ times}} \right\} \\ &\quad \otimes \mathbb{E} \left\{ (\mathbf{H}_{l^*,l^*}^u)^H \mathbf{H}_{l^*,l^*}^u \right\} \\ &= \mathbf{1}_K \otimes \mathbf{C}_{\mathbf{H}_{l^*,l^*}^u} \in \mathbb{C}^{KQ \times KQ}, \end{aligned} \quad (62)$$

where  $\mathbf{1}_K$  denotes the  $(K \times K)$ -element matrix whose elements are all equal to one. Note that  $(\mathbf{X}_{l^*}^u)^T$  is the  $u$ -th column of  $\mathbf{X}_{l^*}$ ,  $1 \leq u \leq U$ , with  $\mathbf{X}_{l^*}^u \in \mathbb{C}^{1 \times K}$ , and let us denote the  $k$ -th column of  $\mathbf{D}_l$  as  $\mathbf{D}_l^T[k]$ ,  $1 \leq k \leq K$ , with  $\mathbf{D}_l[k] \in \mathbb{C}^{1 \times Q}$ . Further define

$$\Psi_{\text{diag}}^u = \text{diag} \left\{ (\mathbf{X}_{l^*}^u)^H \right\} \otimes \mathbf{I}_Q \in \mathbb{C}^{KQ \times KQ}, \quad (63)$$

$$\mathbf{D}_{l,\text{vec}} = [\mathbf{D}_l^T[1] \ \mathbf{D}_l^T[2] \cdots \mathbf{D}_l^T[K]] \in \mathbb{C}^{1 \times KQ}, \quad (64)$$

$$\begin{aligned} \mathbf{C}_{\mathbf{D}_{l,\text{vec}}} &= \mathbb{E} \left\{ \mathbf{D}_{l,\text{vec}}^H \mathbf{D}_{l,\text{vec}} \right\} = \mathbb{E} \left\{ \bar{\mathbf{S}}_l^H \bar{\mathbf{S}}_l \right\} \otimes \mathbb{E} \left\{ \mathbf{A}_l^H \mathbf{A}_l \right\} \\ &= \mathbf{C}_{\bar{\mathbf{S}}_l} \otimes \mathbf{C}_{\mathbf{A}_l} \in \mathbb{C}^{KQ \times KQ}, \end{aligned} \quad (65)$$

$$\mathbf{g}_{l^*,l} = \mathbf{I}_K \otimes \mathbf{G}_{l^*,l}^H \in \mathbb{C}^{KQ \times KQ}. \quad (66)$$

If we denote the  $k$ -th row of  $\mathbf{Y}_{l^*}$  by  $\mathbf{Y}_{l^*}[k] \in \mathbb{C}^{1 \times Q}$ , where  $1 \leq k \leq K$ , we can alternatively express  $\mathbf{Y}_{l^*} \in \mathbb{C}^{K \times Q}$  as

$$\mathbf{Y}_{l^*,\text{vec}} = [\mathbf{Y}_{l^*}[1] \ \mathbf{Y}_{l^*}[2] \cdots \mathbf{Y}_{l^*}[K]] \in \mathbb{C}^{1 \times KQ}. \quad (67)$$

Similarly, we can define  $\mathbf{W}_{l^*,\text{vec}}$  as an alternative representation of  $\mathbf{W}_{l^*}$ . Then it follows from (57) that

$$\begin{aligned} \mathbf{Y}_{l^*,\text{vec}}^H &= \sqrt{p_r} \sum_{u=1}^U \Psi_{\text{diag}}^u (\mathbf{H}_{l^*,l^*,\text{vec}}^u)^T \\ &\quad + \sqrt{p_f} \sum_{l=1, l \neq l^*}^L \mathbf{g}_{l^*,l} (\mathbf{D}_{l,\text{vec}})^T + \mathbf{W}_{l^*,\text{vec}}^H. \end{aligned} \quad (68)$$

Then the LMMSE estimator of  $\mathbf{H}_{l^*,l^*,\text{vec}}^u$  is given by [1]:

$$\left( \hat{\mathbf{H}}_{l^*,l^*,\text{vec}}^u \right)^H = \sqrt{p_r} \mathbf{C}_{\mathbf{H}_{l^*,l^*,\text{vec}}^u} (\Psi_{\text{diag}}^u)^H (\mathbf{\Upsilon}^u)^{-1} \mathbf{Y}_{l^*,\text{vec}}^H, \quad (69)$$

where  $\mathbf{\Upsilon}^u \in \mathbb{C}^{KQ \times KQ}$  is given by

$$\begin{aligned} \mathbf{\Upsilon}^u &= p_r \Psi_{\text{diag}}^u \mathbf{C}_{\mathbf{H}_{l^*,l^*,\text{vec}}^u} (\Psi_{\text{diag}}^u)^H \\ &\quad + p_f \sum_{l=1, l \neq l^*}^L \mathbf{g}_{l^*,l} \mathbf{C}_{\mathbf{D}_{l,\text{vec}}} \mathbf{g}_{l^*,l}^H + \sigma_w^2 \mathbf{I}_{KQ}. \end{aligned} \quad (70)$$

However,  $\mathbf{C}_{\mathbf{H}_{l^*,l^*,\text{vec}}^u}$ ,  $\mathbf{C}_{\bar{\mathbf{S}}_l}$  and  $\mathbf{C}_{\mathbf{A}_l}$  are unknown, and several assumptions have to be made. Specifically, it is assumed that  $\mathbf{1}$

the channel coefficients of the different channel links in the serving cell are i.i.d. with zero mean and a covariance matrix of  $\mathbf{C}_{\mathbf{H}_{l^*,l^*}}^u = \beta_{l^*,l^*} \mathbf{I}_Q$ ; **2**) the DL transmitted data of the other BSs are also i.i.d. with zero mean and a covariance matrix of  $\mathbf{C}_{\mathbf{S}_l} = \mathbf{I}_K$ ; and **3**) the approximation of  $\mathbb{E}\{\mathbf{A}_l^H \mathbf{A}_l\} = (1/Q) \mathbf{I}_Q$  holds for large  $Q$  [1]. Under these assumptions, we have

$$\tilde{\mathbf{C}}_{\mathbf{H}_{l^*,l^*},\text{vec}}^u = \mathbf{1}_K \otimes (\beta_{l^*,l^*} \mathbf{I}_Q) \approx \mathbf{C}_{\mathbf{H}_{l^*,l^*},\text{vec}}^u, \quad (71)$$

$$\tilde{\mathbf{C}}_{\mathbf{D}_{l,\text{vec}}} = \frac{1}{Q} \mathbf{I}_{KQ} \approx \mathbf{C}_{\mathbf{D}_{l,\text{vec}}}, \quad (72)$$

$$\begin{aligned} \tilde{\mathbf{Y}}^u &= p_r \mathbf{\Psi}_{\text{diag}}^u \tilde{\mathbf{C}}_{\mathbf{H}_{l^*,l^*},\text{vec}}^u (\mathbf{\Psi}_{\text{diag}}^u)^H \\ &+ \frac{p_f}{Q} \sum_{l=1, l \neq l^*}^L \mathbf{g}_{l^*,l} \mathbf{g}_{l^*,l}^H + \sigma_w^2 \mathbf{I}_{KQ}. \end{aligned} \quad (73)$$

The LMMSE estimator of  $\mathbf{H}_{l^*,l^*}^u$  is then given by [1]

$$\left( \hat{\mathbf{H}}_{l^*,l^*}^{u,\text{LMMSE}} \right)^H = \sqrt{p_r} \tilde{\mathbf{C}}_{\mathbf{H}_{l^*,l^*},\text{vec}}^u (\mathbf{\Psi}_{\text{diag}}^u)^H \left( \tilde{\mathbf{Y}}^u \right)^{-1} \mathbf{Y}_{l^*,\text{vec}}^H. \quad (74)$$

The estimate  $\hat{\mathbf{H}}_{l^*,l^*}^{u,\text{LMMSE}}$  of  $\mathbf{H}_{l^*,l^*}^u$  is obtained by taking the first  $Q$  elements of  $\hat{\mathbf{H}}_{l^*,l^*}^{u,\text{LMMSE}}$ .

The LMMSE estimator (69) for MS  $u$  is contaminated by the other  $(U-1)$  MSs scattered in the same cell. To clearly see this contamination, let us consider the ‘ideal’ case of no inter-cell interference and no channel AWGN, namely, the path loss coefficients between cell  $l$  and the desired cell  $l^*$   $\eta_{l^*,l} \rightarrow 0$  and the AWGN’s power  $\sigma_w^2 \rightarrow 0$ . Invoking (58) to (60) and (66), we have  $\mathbf{g}_{l^*,l} \rightarrow \mathbf{0}_{KQ}$ , where  $\mathbf{0}_{KQ}$  represents the  $(KQ \times KQ)$ -element matrix whose elements are all equal to zero. Therefore, we can rewrite  $\mathbf{Y}^u$  of (70) as follows

$$\mathbf{Y}^u \approx p_r \mathbf{\Psi}_{\text{diag}}^u \mathbf{C}_{\mathbf{H}_{l^*,l^*},\text{vec}}^u (\mathbf{\Psi}_{\text{diag}}^u)^H. \quad (75)$$

Hence, the LMMSE estimator of  $\mathbf{H}_{l^*,l^*}^u$  given in (69) can be approximated as

$$\begin{aligned} \left( \hat{\mathbf{H}}_{l^*,l^*}^{u,\text{LMMSE}} \right)^H &\approx \sqrt{p_r} \mathbf{C}_{\mathbf{H}_{l^*,l^*},\text{vec}}^u (\mathbf{\Psi}_{\text{diag}}^u)^H \\ &\times \left( p_r \mathbf{\Psi}_{\text{diag}}^u \mathbf{C}_{\mathbf{H}_{l^*,l^*},\text{vec}}^u (\mathbf{\Psi}_{\text{diag}}^u)^H \right)^{-1} \mathbf{Y}_{l^*,\text{vec}}^H. \end{aligned} \quad (76)$$

Substituting  $\mathbf{Y}_{l^*,\text{vec}}^H$  of (68) and the assumption of  $\sigma_w^2 \rightarrow 0$  into (76) leads to

$$\begin{aligned} \left( \hat{\mathbf{H}}_{l^*,l^*}^{u,\text{LMMSE}} \right)^H &\approx (\mathbf{H}_{l^*,l^*}^u)^T \\ &+ (\mathbf{\Psi}_{\text{diag}}^u)^{-1} \sum_{\tilde{u}=1, \tilde{u} \neq u}^U \mathbf{\Psi}_{\text{diag}}^{\tilde{u}} (\mathbf{H}_{l^*,l^*}^{\tilde{u}})^T, \\ &= (\mathbf{H}_{l^*,l^*}^u)^T \\ &+ \sum_{\tilde{u}=1, \tilde{u} \neq u}^U \tilde{\mathbf{\Psi}}_{\text{diag}}^{\tilde{u}} (\mathbf{H}_{l^*,l^*}^{\tilde{u}})^T, \end{aligned} \quad (77)$$

where  $\tilde{\mathbf{\Psi}}_{\text{diag}}^{\tilde{u}} \in \mathbb{C}^{KQ \times KQ}$  is given by

$$\tilde{\mathbf{\Psi}}_{\text{diag}}^{\tilde{u}} = \text{diag} \left\{ \left[ \begin{array}{ccc} X_{l^*}^{\tilde{u}}[1] & X_{l^*}^{\tilde{u}}[2] & \dots & X_{l^*}^{\tilde{u}}[K] \\ X_{l^*}^{\tilde{u}}[1] & X_{l^*}^{\tilde{u}}[2] & \dots & X_{l^*}^{\tilde{u}}[K] \end{array} \right]^{\dagger} \right\} \otimes \mathbf{I}_Q. \quad (78)$$

Observe in (77) that the estimator  $\hat{\mathbf{H}}_{l^*,l^*}^{u,\text{LMMSE}}$  is actually a biased estimator. Note that the LMMSE CE using the staggered PSs of (74) requires the following information: **a**) the FDCHTF matrices  $\mathbf{G}_{l^*,l}$  between the BSs of cells  $l$  for  $l \neq l^*$  and the BS of cell  $l^*$ ; and **b**) the covariance matrices  $\mathbf{C}_{\mathbf{H}_{l^*,l^*}}^u$  of the FDCHTFs  $\mathbf{H}_{l^*,l^*}^u$ , namely, all the path loss coefficients  $\beta_{l^*,l^*}^u$  (assuming  $\beta_{l^*,l^*,p}^u = \beta_{l^*,l^*}^u$  for  $1 \leq p \leq P$  and  $1 \leq q \leq Q$ ) for  $1 \leq u \leq U$ .

#### ACKNOWLEDGMENT

The authors would like to express their gratitude to Haifan Yin for his help on the channel model for multi-cell systems and to Chen Dong for his helpful suggestions.

#### REFERENCES

- [1] K. Appaiah, A. Ashikhmin, and T. L. Marzetta, “Pilot contamination reduction in multi-user TDD systems,” in *Proc. IEEE Int. Conf. Commun.*, Cape Town, South Africa, May 23–27, 2010, pp. 1–5.
- [2] 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.
- [3] T. L. Marzetta, “Noncooperative cellular wireless with unlimited numbers of base station antennas,” *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.
- [4] T. L. Marzetta and B. M. Hochwald, “Fast transfer of channel state information in wireless systems,” *IEEE Trans. Signal Process.*, vol. 54, no. 4, pp. 1268–1278, Apr. 2006.
- [5] J. Jose, A. Ashikhmin, T. L. Marzetta, and S. Vishwanath, “Pilot contamination problem in multi-cell TDD systems,” in *Proc. IEEE Int. Symp. Inf. Theory*, Seoul, Korea, Jun. 28–Jul. 3 2009, pp. 2184–2188.
- [6] F. Fernandes, A. Ashikhmin, and T. L. Marzetta, “Inter-cell interference in noncooperative TDD large scale antenna systems,” *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 192–201, Feb. 2013.
- [7] H. Q. Ngo, T. L. Marzetta, and E. G. Larsson, “Analysis of the pilot contamination effect in very large multicell multiuser MIMO systems for physical channel models,” in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Prague, Czech Republic, May 22–27, 2011, pp. 3464–3467.
- [8] J. Hoydis, S. ten 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.
- [9] B. Gopalakrishnan and N. Jindal, “An analysis of pilot contamination on multi-user MIMO cellular systems with many antennas,” in *Proc. 12th IEEE Int. Workshop Signal Process. Adv. Wireless Commun.*, San Francisco, CA, USA, Jun. 26–29, 2011, pp. 381–385.
- [10] 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. 2640–2651, Aug. 2011.
- [11] T. Zhou, M. Peng, W. Wang, and H. H. Chen, “Low-complexity coordinated beamforming for downlink multicell SDMA/OFDM systems,” *IEEE Trans. Veh. Technol.*, vol. 62, no. 1, pp. 247–255, Jan. 2013.
- [12] H. Dahrouj and W. Yu, “Coordinated beamforming for the multicell multi-antenna wireless system,” *IEEE Trans. Wireless Commun.*, vol. 9, no. 5, pp. 1748–1759, May 2010.
- [13] C. B. Chae, I. S. Hwang, R. W. Heath, and V. Tarokh, “Interference aware-coordinated beamforming in a multi-cell system,” *IEEE Trans. Wireless Commun.*, vol. 11, no. 10, pp. 3692–3703, Oct. 2012.
- [14] L. Hanzo, H. Haas, S. Imre, D. O’Brien, M. Rupp, and L. Gyongyosi, “Wireless myths, realities, futures: From 3G/4G to optical and quantum wireless,” *Proc. IEEE*, vol. 100, pp. 1853–1888, May 2012, Special Centennial Issue.
- [15] L.-L. Yang, “Design of linear multiuser transmitters from linear multiuser receivers,” in *Proc. ICC ’07*, Glasgow, U.K., Jun. 24–28, 2007, pp. 5258–5263.
- [16] Y. Xiao, Y. Zhao, and M. H. Lee, “MIMO precoding of CDMA systems,” in *Proc. 8th Int. Conf. Signal Process.*, Guilin, China, Nov. 16–20, 2006, vol. 1, pp. 397–401.
- [17] B. R. Vojčić and W. M. Jang, “Transmitter precoding in synchronous multiuser communications,” *IEEE Trans. Commun.*, vol. 46, no. 10, pp. 1346–1355, Oct. 1998.

- [18] W. Yao, S. Chen, S. Tan, and L. Hanzo, "Minimum bit error rate multiuser transmission designs using particle swarm optimisation," *IEEE Trans. Wireless Commun.*, vol. 8, no. 10, pp. 5012–5017, Oct. 2009.
- [19] W. Yao, S. Chen, and L. Hanzo, "Generalised MBER-based vector precoding design for multiuser transmission," *IEEE Trans. Veh. Technol.*, vol. 60, no. 2, pp. 739–745, Feb. 2011.
- [20] W. Yao, S. Chen, and L. Hanzo, "A transceiver design based on uniform channel decomposition and MBER vector perturbation," *IEEE Trans. Veh. Technol.*, vol. 59, no. 6, pp. 3153–3159, Jul. 2010.
- [21] A. F. Karr, *Probability*. Berlin, Germany: Springer-Verlag, 1993.
- [22] J. W. Brown and R. V. Churchill, *Complex Variables and Applications*. Upper Saddle River, NJ, USA: McGraw-Hill, 2008.
- [23] R. E. Walpole, R. H. Myers, S. L. Myers, and K. Y. Keying, *Probability and Statistics for Engineers and Scientists*. Upper Saddle River, NJ, USA: Pearson, 2011.



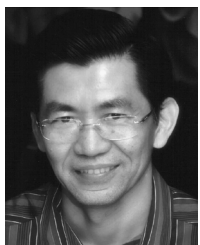
**Jiankang Zhang** (S'08–M'12) received the B.Sc. degree in Mathematics and Applied Mathematics from Beijing University of Posts and Telecommunications in 2006, and the Ph.D. degree in Communication and Information Systems from Zhengzhou University in 2012.

Since then, he has been a lecturer in School of Information Engineering, Zhengzhou University. From September 2009 to December 2011 and from January 2013 to May 2013, Dr Zhang was a visiting researcher in Electronics and Computer Science, the

University of Southampton, UK. His research interests are in the areas of wireless communications and signal processing, including channel estimation, multiuser detection, beamforming/precoding and optimization algorithms.



**Bo Zhang** received his B.S. degree in Information Engineering from National University of Defense Technology, China, in 2010. He is currently working toward the Ph.D. degree with the Communications, Signal Processing and Control Group, School of Electronics and Computer Science, University of Southampton, Southampton, UK. His research interests in wireless communications include design and analysis of cooperative communications, MIMO and network-coded systems.



**Sheng Chen** (M'1990–SM'1997–F'2008) obtained his BEng degree from the East China Petroleum Institute, Dongying, China, in January 1982, and his PhD degree from the City University, London, in September 1986, both in control engineering. In 2005, he was awarded the higher doctorate degree, Doctor of Sciences (DSc), from the University of Southampton, Southampton, UK.

From 1986 to 1999, He held research and academic appointments at the Universities of Sheffield, Edinburgh and Portsmouth, all in UK. Since 1999, he has

been with Electronics and Computer Science, the University of Southampton, UK, where he currently holds the post of Professor in Intelligent Systems and Signal Processing. Dr Chen is a Distinguished Adjunct Professor at King Abdulaziz University, Jeddah, Saudi Arabia.

Dr Chen is a Chartered Engineer (CEng) and a Fellow of IET (FIET). His recent research interests include adaptive signal processing, wireless communications, modelling and identification of nonlinear systems, neural network and machine learning, intelligent control system design, evolutionary computation

methods and optimization. He has published over 480 research papers. Dr Chen is an ISI highly cited researcher in the engineering category (March 2004).



**Xiaomin Mu** received her B.E. degree from the Beijing Institute of Technology, Beijing, China in 1982.

She is currently a full professor with the School of Information Engineering, Zhengzhou University. She has published many papers in the field of signal processing and co-authored two books. Her research interests include signal processing in communication systems, wireless communications and cognitive radio.



**Mohammed El-Hajjar** is a lecturer in the Electronics and Computer Science in the University of Southampton. He received his BEng degree in Electrical Engineering from the American University of Beirut, Lebanon in 2004. He then received an MSc in Radio Frequency Communication Systems and PhD in Wireless Communications both from the University of Southampton, UK in 2005 and 2008, respectively. Following the PhD, he joined Imagination Technologies as a design engineer, where he worked on designing and developing the

BICM peripherals in Imagination's multi-standard communications platform, which resulted in three patents. In January 2012, he joined the Electronics and Computer Science in the University of Southampton as a lecturer in the Communications, Signal Processing and Control research group. He is the recipient of several academic awards and has published a Wiley-IEEE book and in excess of 40 journal and international conference papers. His research interests are mainly in the development of intelligent communications systems for the Internet of Things including energy-efficient transceiver design, cross-layer optimization for large-scale networks, massive MIMO systems for mm-wave communications, cooperative communications and Radio over fibre systems, communications and cognitive radio.



**Lajos Hanzo** FREng, FIEEE, FIET, Fellow of EURASIP, DSc received his degree in electronics in 1976 and his doctorate in 1983. In 2009 he was awarded the honorary doctorate "Doctor Honoris Causa" by the Technical University of Budapest.

During his 37-year career in telecommunications he has held various research and academic posts in Hungary, Germany and the UK. Since 1986 he has been with Electronics and Computer Science, University of Southampton, UK, where he holds the chair in telecommunications. He has successfully

supervised 83 PhD students, co-authored 20 John Wiley/IEEE Press books on mobile radio communications totalling in excess of 10 000 pages, published 1356 research entries at IEEE Xplore, acted both as TPC and General Chair of IEEE conferences, presented keynote lectures and has been awarded a number of distinctions. Currently he is directing a 100-strong academic research team, working on a range of research projects in the field of wireless multimedia communications sponsored by industry, the Engineering and Physical Sciences Research Council (EPSRC) UK, the European Research Council and the Royal Society, UK. He is an enthusiastic supporter of industrial and academic liaison and he offers a range of industrial courses. He is also a Governor of the IEEE VTS. During 2008–2012 he was the Editor-in-Chief of the IEEE Press and a Chaired Professor also at Tsinghua University, Beijing. He was awarded the European Research Council's Senior Research Fellow Grant and the Royal Society's Wolfson Research Merit Award. For further information on research in progress and associated publications please refer to <http://www-mobile.ecs.soton.ac.uk>. Lajos has 17 000+ citations.