Optimal Pilot Design for Pilot Contamination Elimination/Reduction in Large-Scale Multiple-Antenna Aided OFDM Systems

Xinying Guo, Sheng Chen, *Fellow, IEEE*, Jiankang Zhang, *Member, IEEE*,
Xiaomin Mu, and Lajos Hanzo, *Fellow, IEEE*

Abstract—This paper considers the problem of pilot contamination (PC) in large-scale multi-cell multiple-input multiple-output-aided orthogonal frequency division multiplexing systems. We propose an efficient scheme relying on an optimal pilot design conceived for time-domain channel estimation, which can either completely eliminate PC or significantly reduce it, depending on the channel's coherence time. This is achieved by designing an optimal pilot set allowing us to beneficially group the users in all the cells and to assign a time-shifted pilot transmission to the different groups. Unlike the existing PC elimination schemes, which require an excessively long channel coherence time, our proposed scheme is capable of completely eliminating PC under a much shorter coherence time. Moreover, the existing PC elimination schemes can no longer be used if the channel coherent time is insufficiently large. By contrast, even for extremely short channel coherent time, our scheme can still be implemented to significantly reduce PC. This is particularly beneficial for high velocity scenarios. Our simulation results demonstrate the efficiency of the proposed scheme.

Index Terms—Multi-cell systems, large-scale multiple-input multiple-output, orthogonal frequency division multiplexing, pilot contamination, time division duplexing, time-domain channel estimation.

I. Introduction

LARGE-SCALE multiple-input multiple-output (MIMO) techniques constitute promising candidates for the fifth-generation cellular network, owing to their high spectral efficiency and energy efficiency [1]–[7]. However, the performance of large-scale MIMO systems is critically dependent on the accuracy of channel state information (CSI) estimation, regardless whether the CSI is used for uplink (UL) reception or for downlink (DL) transmission. In frequency-division duplexing (FDD) systems, the CSI is estimated by the mobile station (MS) receivers and signalled back to the base station (BS). Hence the CSI signaling overhead scales linearly with the number of DL transmit antennas deployed at the BS which renders the use of large antenna arrays impractical. By contrast, in time-division duplexing (TDD) systems, the DL CSI can be estimated at the BS with the aid of UL training by exploiting the channel's reciprocity, and the signaling overhead imposed by the acquisition of UL CSI scales linearly with the number of MSs, which is typically much lower than the number of BS antennas. However, the number of orthogonal pilot sequences available is limited since the channel coherence time is limited. The worst-case scenario is associated with the pilot reuse factor of one, when all cells use the same set of orthogonal pilot sequences at the same time. This causes severe pilot contamination (PC) [6]–[9], which results in the BS being unable to reliably differentiate the signals of different cells.

A. Motivation and Related Works

Compared to the effect of additive white Gaussian noise (AWGN), PC constitutes a much more grave impairment that limits the system's achievable performance, and hence considerable research efforts have been focused on mitigating or eliminating PC. The effects of PC in multi-cell MIMO systems deploying a large number of BS antennas was analyzed by Marzetta in [6]. Fernandes *et al.* [10] presented a method using time-staggered pilots which avoided inter-cell interference, whenever the pilots did not overlap in time. However, this approach requires a central controller to manage the staggering of pilot-aided training in all cells in order to maintain the orthogonality of the pilot transmissions of different cells, which is a challenging large-scale pilot assignment task. Moreover, such a staggered training procedure dramatically increase the training duration, which may become impractical due to the limited coherence time of the channel. The coordination between cells was exploited by Yin *et al.* in [11], where a covariance-aware pilot assignment strategy was proposed for carefully

distributing identical pilot sequences to a set of users. However, this coordinated channel estimation (CE) scheme requires the knowledge of the second-order statistics of all the UL channels. The acquisition of such a large amount of second-order statistics at the BSs is time-consuming and, moreover, sharing them requires a huge amount of back-haul transmissions. Blind CE was also proposed for mitigating the PC problem. The eigenvalue-decomposition-based approach of Ngo and Larsson [12] estimated the channel blindly based on the received data, which exploited the asymptotic orthogonality of the channel vectors in large MIMO systems. However, its performance degrades under a limited number of BS antennas, especially for a limited coherence time. Müller *et al.* [13] proposed a blind multi-user decorrelation process, which however only performed well, when the number of BS antennas was very large. Blind CE schemes are computationally costly and suffer from permutation ambiguity corresponding to reordering the estimated channel matrix columns.

Recently, a location-aware CE scheme was proposed by relying on fast Fourier transform (FFT) based post-processing following the conventional pilot aided CE [14], which is capable of distinguishing the users associated with different angles of arrival (AOAs), even if they use the same pilot. To maximally benefit from this location-aware CE algorithm, a location-aware pilot assignment was also proposed in [14] for satisfying the condition of having non-overlapping AOAs for the users of different cells adopting the same pilot. The advantage of this method is its appealing simplicity and efficiency. Moreover, this does not increase the training duration, but it requires the knowledge of the users' AOAs.

In the literature, there exist two schemes which can completely eliminate PC without the need for any knowledge of the second-order channel statistics or the users' AOAs. The strategy proposed in [15] consisted of an amalgam of DL and UL training phases, which are capable of eliminating PC at the cost of requiring a much longer training duration than the conventional simultaneous UL training. More specifically, the scheme of [15] consists of $(L+3)$ training phases for an L-cell system. Therefore, it requires that the channel's coherence interval (COHI) is no less than $(L+3)U$, where U is the length of the training sequences, which is assumed to be equal to the number of users per cell. Similarly, the scheme proposed by Vu *et al.* in [16] consisted of a conventional simultaneous UL training phase in which all the MSs simultaneously transmit UL pilots to their BSs, followed by the L consecutive pilot transmission phases in which each cell stays idle at one phase and repeatedly transmits pilot sequences in other phases. Thus, this scheme requires that the COHI is no less than $(L+1)U$, which is still very large and may not be met in practical systems. Moreover, these two scheme can no longer be used if the COHI is smaller than $(L+3)U$ and $(L+1)U$, respectively.

Indeed, PC has hampered achieving the full potential of large-scale MIMO systems. Eliminating/reducing PC while requiring a sufficiently short training duration and imposing no excessive amount of prior information is challenging and has not yet been resolved in the open literature.

B. Our Contributions

Given the above background and motivation, our goal is to develop an efficient PC elimination/reduction scheme for multi-cell TDD based orthogonal frequency division multiplexing (OFDM) systems, which only imposes a relatively short training duration and does not require any prior knowledge regarding either the MIMO channels or user information. Our contributions are summarized as follows.

1) An Efficient Signal Framework for Analyzing PC: All the existing schemes [6], [10], [11], [14]–[17] approach PC by considering the signals received by all the target BS's antennas on an individual OFDM subcarrier[1]. While considering a signal framework on an OFDM subcarrier basis is the natural choice for detection/precoding given each subcarrier's frequency-domain channel transfer function (FDCHTF) coefficient matrix, due to the orthogonality of the subcarriers, it is an inefficient one for approaching the CE or PC problem. Specifically, the pilot symbol (PS) received on each subcarrier has a duration of τ OFDM symbols, where τ is related to the number U of users per cell, by $\tau \geq U$, which allows us to have U orthogonal pilot sequences in each cell. Consequently, even a simultaneous UL training carried out for U users has the duration of U OFDM symbols, assuming that $\tau = U$.

By contrast, we approach PC by considering the signals received on all the OFDM subcarriers together on an individual BS antenna basis. At the time of writing, performing the CE based on the signals received on all the OFDM subcarriers is the standard approach of conventional OFDM systems. Instead, we propose a signal framework to perform multi-cell massive MIMO CE on an individual BS antenna basis. The substantial benefit of adopting this signal framework is that a simultaneous UL training only has the duration of a PS, i.e., that of a single OFDM symbol.

2) PC-Free Simultaneous UL Training With an Optimally Designed Pilot Set: In order to accurately estimate the CSI with the aid of limited pilot resources, we operate the CE process in the time domain (TD) by estimating the channel impulse responses (CIRs) using an optimally designed pilot set. Then the FFT is invoked for obtaining the required FDCHTFs. We prove that when sufficient subcarrier resources are available, the orthogonality of all the PSs in our optimally designed pilot set is guaranteed and, therefore, all the CIRs can be acquired free from any PC in a simultaneous UL training session that has only the duration of a single OFDM symbol. This is significant because PC-free simultaneous UL training carried out within a single OFDM symbol duration has not been reported in the open literature.

3) PC-Free/Reduction Time-Shifted UL Training With an Optimally Designed Pilot Set: While under the condition of having insufficient subcarrier resources, PC-free CE cannot be achieved with the aid of simultaneous UL training, our proposed scheme exploits what we refer to as subset orthogonality of the optimally designed pilot set by partitioning the users in all the cells and then arranging for a simple time-shifted pilot

---

[1]Some of the existing methods assume having a narrowband channel on the ground that it is equivalent to applying the OFDM technique to a dispersive channel and considering the FDCHTF matrix of each subcarrier.

transmission for the different groups. This technique is capable of either completely eliminating or significantly reducing the PC, depending on the system's COHI. Specifically, we prove that there exists an optimal number of groups, where the pilot subset assigned to the users of each group contains mutually orthogonal elements. Thus, if the COHI is no lower than the optimal number of groups, our scheme completely eliminates PC. Since this optimal number of groups is inherently smaller than the number of cells L and, therefore, it is much smaller than $(L + 1)U$, our scheme requires a much lower training duration than the scheme of [16]. Most significantly, our scheme is capable of totally eliminating PC under severe COHI conditions, which make the scheme of [16] unusable. Moreover, even when the COHI is lower than the optimal number of groups, as in high-mobility systems, our proposed scheme still remains capable of significantly reducing the PC.

It is worth emphasizing that unlike the conventional approach relying on time-staggered pilots, such as the one presented in [10], which requires huge amount of coordination amongst the cells plus costly pilot assignment, our PC-free/reduction time-shifted scheme requires no information exchange amongst the BSs and imposes no high-delay time-shifted UL training. Therefore, our scheme is capable of totally eliminating PC under severe COHI conditions, which make the scheme of [10] unusable.

It is also worth pointing out that the training duration directly impacts on the achievable system's capacity or sum-rate performance. Most of the existing works do not take into account the influence of the training duration to the effective sum-rate performance. An exception is the study [6], which explicitly evaluates the training-duration adjusted sum-rate performance. In our simulation study, we also evaluate the influence of the training duration to the system's effective sum-rate performance.

The paper is organized as follows. In Section II, we introduce the MIMO-aided multi-cell TDD OFDM system, present the optimally designed pilot set and outline the condition to be met for achieving PC-free CE using the simultaneous UL training. Section III presents our PC elimination/reduction scheme and details the conditions to be satisfied for the sake of completely eliminating PC or for significantly reducing PC, respectively, depending on the system's COHI. In Section IV, we present our simulation results for characterizing our proposed scheme as well as for comparing it with the state-of-the-art benchmarks. Our conclusions are drawn in Section V.

Throughout our discussions, $\mathbb{C}$ denotes the complex number field. For $A \in \mathbb{C}$, we have $A = A_R + \mathrm{j}A_I$, where $\mathrm{j} = \sqrt{-1}$, while A_R and A_I are the real and imaginary parts of A, respectively. Boldface upper-case symbols denote matrices, e.g., $\mathbf{X}$, while underlined boldface upper-case symbols denote column vectors, e.g., $\underline{\mathbf{X}}$. The transpose, conjugate and Hermitian transpose operators are denoted by $(\cdot)^{\mathrm{T}}$, $(\cdot)^{*}$ and $(\cdot)^{\mathrm{H}}$, respectively, while $\mathrm{diag}\{\underline{\mathbf{X}}\}$ denotes the diagonal matrix with the diagonal entries equal to the elements of $\underline{\mathbf{X}}$. $\mathbf{X}_{[n,m]}$ denotes the entity in the n-th row and m-th column of $\mathbf{X}$, and $\underline{\mathbf{X}}_{[n]}$ is the n-th element of $\underline{\mathbf{X}}$, while $\widehat{X}$ represents the estimate of X. The $(K \times K)$-element identity matrix is denoted by $\mathbf{I}_{K \times K}$, and $\mathbf{0}_{K \times K}$ denotes the $(K \times K)$-element matrix with all zero elements, while $\delta(t)$ represents the discrete Dirac delta function and $\mathbb{E}\{\cdot\}$ denotes the expectation operator. Furthermore, $\|\cdot\|_F$ denotes the Frobenius norm, while $\lfloor\cdot\rfloor$ and $\lceil\cdot\rceil$ denote the integer floor and ceiling operators, respectively.

II. MULTI-CELL TDD SYSTEM MODEL

We consider a cellular network composed of L hexagonal cells, labelled by $l = 1, 2, \cdots, L$, where each of the U MSs in each cell is equipped with a single antenna, while the BS of each cell employs an array of Q antennas. It is assumed that $Q \gg U$ and OFDM is employed, where all BSs and MSs are synchronized, relying on a TDD protocol with unity frequency reuse (UFR).

A. Uplink Training

All MSs of all cells commence by synchronously transmitting an OFDM PS to their serving BSs. The frequency-domain (FD) PS of user u in the l-th cell is given by $\underline{\mathbf{X}}_l^u = \left[X_l^u[1]\ X_l^u[2] \cdots\ X_l^u[N]\right]^{\mathrm{T}}$, where N is the number of subcarriers and the power of each pilot $X_l^u[n]$ is unity. Let $H_{l,l',q}^u[n]$ be the UL FDCHTF linking the u-th user in cell l to the q-th antenna of the l'-th cell's BS, at the n-th subcarrier. Furthermore, $Y_{l',q}[n]$ is the signal received by the q-th receive antenna of the l'-th BS at the n-th subcarrier, which can be expressed as

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

for $1 \leq l' \leq L$ and $1 \leq q \leq Q$, where p_r denotes the average power per subcarrier of each user and $W_{l',q}[n]$ is the FD representation of the UL channel's AWGN, denoted by $W_{l',q}[n] \sim \mathcal{CN}(0, \sigma_w^2)$ with σ_w^2 being the power of $W_{l',q}[n]$. The set of equations constituted by (1) for $1 \leq n \leq N$ can be written in the more compact form of

$$\underline{\mathbf{Y}}_{l',q} = \sqrt{p_r} \underbrace{\sum_{u'=1}^{U} \mathbf{X}_{l'}^{u'} \underline{\mathbf{H}}_{l',l',q}^{u'}}_{\text{Desired term}} + \sqrt{p_r} \underbrace{\sum_{l=1,l \neq l'}^{L} \sum_{u=1}^{U} \mathbf{X}_l^u \underline{\mathbf{H}}_{l,l',q}^u}_{\text{Inter-cell interference}}$$
$$+ \underline{\mathbf{W}}_{l',q}, \quad (2)$$

where $\mathbf{X}_l^u = \mathrm{diag}\{\underline{\mathbf{X}}_l^u\}$, while $\underline{\mathbf{Y}}_{l',q} \in \mathbb{C}^{N \times 1}$, $\underline{\mathbf{H}}_{l,l,q}^u \in \mathbb{C}^{N \times 1}$ and $\underline{\mathbf{W}}_{l',q} \in \mathbb{C}^{N \times 1}$ are the three column vectors hosting $Y_{l',q}[n]$, $H_{l,l,q}^u[n]$ and $W_{l',q}[n]$ for $1 \leq n \leq N$, respectively. It is worth emphasizing that the signal vector (2) is collected over all the N OFDM subcarriers for an individual BS antenna. This is in contrast to all the existing methods, which consider the signal vector over all the target BS's antennas for an individual subcarrier [6], [10], [11], [14]–[17]. Our approach to UL training has a significant advantage. To implement a simultaneous UL training, our approach only requires that the COHI is no less than 1, while all the existing schemes require that the COHI is no less than U, because they all expect each user to transmit a pilot sequence of length U. As mentioned

previously, our signal model is better suited for performing the CE for massive MIMOs than the one adopted in all the existing methods, since it is the natural extension of the standard CE technique of single-antenna multiuser OFDM systems.

The FDCHTF coefficient vector $\underline{\mathbf{H}}_{l,l',q}^{u}$ is the N-point FFT of the symbol-rate sampled CIR coefficient vector $\underline{\mathbf{G}}_{l,l',q}^{u} \in \mathbb{C}^{K \times 1}$ of duration K, for the channel linking the u-th user in cell l to the q-th antenna of the l'-th cell's BS, which can be expressed as

$$\underline{\mathbf{G}}_{l,l',q}^{u} = \left[G_{l,l',q}^{u}[1] \; G_{l,l',q}^{u}[2] \cdots G_{l,l',q}^{u}[K] \right]^{\mathrm{T}}, \qquad (3)$$

where K represents the maximum delay spread of the dispersive channel[2]. Moreover, $\underline{\mathbf{H}}_{l,l',q}^{u}$ and $\underline{\mathbf{G}}_{l,l',q}^{u}$ are linked by the expression

$$\underline{\mathbf{H}}_{l,l',q}^{u} = \mathbf{F}\underline{\mathbf{G}}_{l,l',q}^{u}, \qquad (4)$$

where $\mathbf{F} \in \mathbb{C}^{N \times K}$ is the FFT matrix, whose elements are given by $\mathbf{F}_{[n,k]} = \frac{1}{\sqrt{K}}e^{-j2\pi(n-1)(k-1)/N}$ for $1 \leq n \leq N$ and $1 \leq k \leq K$. The CIR (3) is valid for the generic massive MIMO system as it does not depend on the particular antenna array structure deployed. Using (4), (2) can be rewritten as

$$\underline{\mathbf{Y}}_{l',q} = \sqrt{p_r} \sum_{u'=1}^{U} \mathbf{X}_{l'}^{u'} \mathbf{F}\underline{\mathbf{G}}_{l',l',q}^{u'} + \sqrt{p_r} \sum_{l=1,l\neq l'}^{L} \sum_{u=1}^{U} \mathbf{X}_{l}^{u} \mathbf{F}\underline{\mathbf{G}}_{l,l',q}^{u} + \underline{\mathbf{W}}_{l',q}. \qquad (5)$$

In our simulation study, we will assume a uniformly spaced linear antenna array (ULA) at the BS. For the ULA, the TD CIR vector $\underline{\mathbf{G}}_{l,l',q}^{u}$ is specified by

$$G_{l,l',q}^{u}[k] = \alpha_{l,l',q,k}^{u} e^{-j2\pi \frac{(q-1)D}{\lambda} \cos\left(\theta_{l,l',q,k}^{u}\right)}, \quad 1 \leq k \leq K, \quad (6)$$

where D and λ are the antenna spacing and the carrier's wavelength, respectively. In (6), $\theta_{l,l',q,k}^{u}$ is the AOA of the k-th tap for the channel between the u-th MS in the l-th cell and the q-th antenna of the l'-th BS, while the complex-valued tap $\alpha_{l,l',q,k}^{u}$ is given by [18]

$$\alpha_{l,l',q,k}^{u} = e^{-j\varphi_{l,l',q,k}^{u}} \sqrt{\beta_{l,l',q,k}^{u}}, \qquad (7)$$

in which the phase $\varphi_{l,l',q,k}^{u}$ is a random variable uniformly distributed in $[0, \; 2\pi)$ and the path-loss coefficient $\beta_{l,l',q,k}^{u}$ accounts for the attenuation and shadow fading of the path. Since $\beta_{l,l',q,k}^{u}$ only changes slowly as a function of distance [6], we may assume that $\beta_{l,l',q,k}^{u} = \beta_{l,l'}^{u}$ for $1 \leq k \leq K$ and $1 \leq q \leq Q$.

Table I summarizes the notations used throughout our discussions.

---

[2]Let the sampling rate be $1/T_s$ and further denote the OFDM symbol duration by $T_{\mathrm{OFDM}} = (N + L_{cp})T_s$, which includes the CP of length L_{cp}. Then KT_s is the overall duration of the TD CIR. To ensure orthogonality between subcarriers, the subcarrier spacing must satisfy $\Delta f \leq 1/(NT_s)$, and to eliminate intersymbol interference, it is required that $L_{cp} \geq K - 1$. Furthermore, to maintain a high spectral efficiency we opt for, $N \gg L_{cp}$. Hence we have $N \gg K$.

TABLE I
LIST OF NOTATIONS

Notation	Definition
L	Number of cells
U	Number of MSs per-cell
Q	Number of antennas at each BS
N	Number of subcarriers
p_r	Average transmit power per subcarrier at each MS
p_f	Average transmit power per subcarrier at each BS
K	Length of symbol-rate sampled CIR
α	Complex-valued CIR tap
D	Antenna spacing
λ	Carrier's wavelength
θ	AOA
β	Path-loss coefficient
φ	Random phase of CIR tap
T_{OFDM}	OFDM symbol duration
ζ	$\lfloor \frac{N}{LU} \rfloor$
f	$\lfloor \frac{K}{\zeta} \rfloor$
n_u	$\lfloor \frac{LU}{f} \rfloor$
R	$\mathrm{Rem}\{\frac{LU}{f}\} = LU - n_u f$
B	System bandwidth
Δf	Subcarrier spacing
r	effective COHI
n_u'	$\lfloor \frac{LU}{r} \rfloor$
R'	$LU - n_u' r$

B. Time-Domain Channel Estimation

Theorem 1: Design a FD PS matrix set for all MSs in all cells according to [19]

$$\begin{aligned} \mathbf{P} &= \left\{ \mathbf{X}_{l}^{u}, 1 \leq u \leq U, 1 \leq l \leq L \right\} = \left\{ \mathbf{P}[i], 1 \leq i \leq LU \right\} \\ &= \left\{ \mathbf{X}_{1}^{1}, \mathbf{X}_{2}^{1}, \cdots, \mathbf{X}_{L}^{1}; \mathbf{X}_{1}^{2}, \mathbf{X}_{2}^{2}, \cdots, \mathbf{X}_{L}^{2}; \cdots; \right. \\ &\qquad \left. \mathbf{X}_{1}^{U}, \mathbf{X}_{2}^{U}, \cdots, \mathbf{X}_{L}^{U} \right\}, \end{aligned} \qquad (8)$$

which contains the LU diagonal PS matrices of

$$\begin{aligned} \mathbf{P}[i] &= \mathbf{P}[(u-1)L + l] = \mathbf{X}_{l}^{u}, \\ &\quad i = (u-1)L + l, \quad 1 \leq u \leq U, \; 1 \leq l \leq L. \end{aligned} \qquad (9)$$

Clearly, i is the index of the LU users. Specifically, the i-th element of the FD PS matrix set is generated from a reference $\mathbf{P}[1] = \mathbf{X}_{1}^{1}$ according to

$$\mathbf{P}[i] = \mathbf{\Phi}[i]\mathbf{P}[1], \quad 1 \leq i \leq LU, \qquad (10)$$

where the diagonal matrix

$$\mathbf{\Phi}[i] = \mathrm{diag}\left\{ e^{j2\pi \frac{(i-1)\zeta 0}{N}}, e^{j2\pi \frac{(i-1)\zeta 1}{N}}, \cdots, e^{j2\pi \frac{(i-1)\zeta(N-1)}{N}} \right\}, \\ 1 \leq i \leq LU, \qquad (11)$$

shifts the reference $\mathbf{P}[1]$ in phase with a positive integer parameter ζ.

If $\zeta = \lfloor \frac{N}{LU} \rfloor \geq K$, this FD PS matrix set offers the desired orthogonality as follows:

$$\begin{aligned} &\left(\mathbf{P}[(u_1 - 1)L + l_1]\mathbf{F} \right)^{\mathrm{H}} \left(\mathbf{P}[(u_2 - 1)L + l_2]\mathbf{F} \right) \\ &= \begin{cases} \mathbf{0}_{K \times K}, & l_1 \neq l_2 \cup u_1 \neq u_2, \\ \frac{N}{K}\mathbf{I}_{K \times K}, & l_1 = l_2 \cap u_1 = u_2, \end{cases} \end{aligned} \qquad (12)$$

where $1 \leq u_1, u_2 \leq U, 1 \leq l_1, l_2 \leq L$.

Proof: Using the PS matrix set (9) in (5), we have

$$\underline{\mathbf{Y}}_{l',q} = \sqrt{p_r} \sum_{u'=1}^{U} \mathbf{P}[(u'-1)L + l'] \mathbf{F} \underline{\mathbf{G}}_{l',l',q}^{u'}$$
$$+ \sqrt{p_r} \sum_{l=1,l\neq l'}^{L} \sum_{u=1}^{U} \mathbf{P}[(u-1)L + l] \mathbf{F} \underline{\mathbf{G}}_{l,l',q}^{u} + \underline{\mathbf{W}}_{l',q}.$$

$$(13)$$

Given $\zeta = \lfloor \frac{N}{LU} \rfloor \geq K$, we now prove (12). Let us set

$$\mathbf{T} = (\mathbf{P}[i_1]\mathbf{F})^{\mathrm{H}} (\mathbf{P}[i_2]\mathbf{F}), \quad 1 \leq i_1, i_2 \leq LU, \quad (14)$$

$$\mathbf{T}_{[k_1,k_2]} = \frac{1}{K} \sum_{n=1}^{N} \left(\mathbf{P}[i_1]_{[n,n]}\right)^* \mathbf{P}[i_2]_{[n,n]} e^{\frac{j2\pi(n-1)(k_1-k_2)}{N}},$$

$$1 \leq k_1, \quad k_2 \leq K. \quad (15)$$

Furthermore, for the integer $t = k_1 - k_2$, we have $|t| \leq K - 1$ and Eq. (15) can be rewritten as

$$\mathbf{T}_{[k_1,k_2]} = \frac{1}{K} \sum_{n=1}^{N} \left(\mathbf{P}[1]_{[n,n]}\right)^* \mathbf{P}[1]_{[n,n]} e^{\frac{-j2\pi(i_1-1)\zeta(n-1)}{N}}$$

$$\times e^{\frac{j2\pi(i_2-1)\zeta(n-1)}{N}} e^{\frac{j2\pi(n-1)t}{N}} = \frac{N}{K} \delta\big(t - (i_1 - i_2)\zeta\big). \quad (16)$$

Under the condition of $i_1 > i_2$ and $t \equiv N + t$ if $t \leq 0$, we have

$$\begin{cases} \zeta \leq (i_1 - i_2)\zeta \leq (LU - 1)\zeta, \\ 1 \leq t \leq K - 1 \text{ or } N - K + 1 \leq t \leq N. \end{cases} \quad (17)$$

To keep $\delta\big(t - (i_1 - i_2)\zeta\big) = 0$, i.e., $t \neq (i_1 - i_2)\zeta$, the inequalities in (17) should have no intersection, that is, $K - 1 < (i_1 - i_2)\zeta < N - K + 1$. In other words, we must have

$$\begin{cases} \zeta \geq K, \\ LU\zeta - \zeta \leq N - K. \end{cases} \quad (18)$$

If we restrict the range of ζ to

$$K \leq \zeta \leq \frac{N}{LU}, \quad (19)$$

which clearly meets the condition of (18), then $\mathbf{T} = \mathbf{0}_{K \times K}$ can be achieved for all $i_1 > i_2$. For $i_1 < i_2$, we can arrive at the same conclusion. Hence, $\mathbf{T} = \mathbf{0}_{K \times K}, \forall i_1 \neq i_2$, when $K \leq \zeta \leq \frac{N}{LU}$. Noting the ordered relationship of $i = (u-1)L + l$, $i_1 \neq i_2$ is equivalent to $l_1 \neq l_2 \cup u_1 \neq u_2$.

Let us now consider $i_1 = i_2$, which is equivalent to $l_1 = l_2 \cap u_1 = u_2$. If $t = k_1 - k_2 \neq 0$, then clearly $\mathbf{T}_{[k_1,k_2]} = 0$. If $(i_1 = i_2) \cap (k_1 = k_2)$, then clearly $\mathbf{T}_{[k_1,k_2]} = \frac{N}{K}$. Thus $\mathbf{T} = \frac{N}{K} \mathbf{I}_{K \times K}$ for $i_1 = i_2$, given that ζ meets the condition of (19).

It can be seen that the orthogonality of (12) actually holds for $K \leq \zeta \leq \frac{N}{LU}$. Then it also holds for $\zeta = \lfloor \frac{N}{LU} \rfloor \geq K$. ∎

1) Pilot-Contamination-Free Scenario: When we design the FD PS matrix set $\mathbf{P}$ for all the MSs in all the cells according to Theorem 1, then the identifiability of the unique TD CIR vector $\underline{\mathbf{G}}_{l',l',q}^{u'}$ is guaranteed under the condition of:

$$KLU \leq N, \quad (20)$$

which implies a PC free scenario. That is, all the LU MSs of all the L cells can simultaneously transmit their PS matrices to their serving BSs for PC-free channel estimation. The least squares (LS) CE of $\underline{\mathbf{G}}_{l',l',q}^{u'}$ is given by

$$\widehat{\underline{\mathbf{G}}}_{l',l',q}^{u'} = \frac{K}{N\sqrt{p_r}} \big(\mathbf{P}[(u'-1)L + l']\mathbf{F}\big)^{\mathrm{H}} \underline{\mathbf{Y}}_{l',q}$$
$$= \underline{\mathbf{G}}_{l',l',q}^{u'} + \frac{K}{N\sqrt{p_r}} \big(\mathbf{P}[(u'-1)L + l']\mathbf{F}\big)^{\mathrm{H}} \underline{\mathbf{W}}_{l',q}.$$

$$(21)$$

Let us now analyze the accuracy of this PC-free LS estimator in terms of its mean square error (MSE), which is defined by

$$\overline{\Xi}_0 = \mathbb{E}\left\{\frac{1}{K} \|\underline{\Xi}_0\|_F^2\right\}, \quad (22)$$

where $\underline{\Xi}_0 = \widehat{\underline{\mathbf{G}}}_{l',l',q}^{u'} - \underline{\mathbf{G}}_{l',l',q}^{u'}$ is the channel estimation error.

Theorem 2: The MSE of the PC-free LS CE (21) is given by

$$\overline{\Xi}_0 = \frac{K\sigma_w^2}{Np_r}. \quad (23)$$

Proof: See Appendix A. ∎

2) Pilot-Contamination-Existing Scenario: If the condition (20) is not met, then the orthogonality (12) does not hold, i.e., the number of the mutually orthogonal elements in $\mathbf{P}$ is less than LU, which results in PC when all MSs of all cells simultaneously transmit their PS matrices to their serving BSs. By exploiting the orthogonality of the subset PS matrices in the set (10), we propose an efficient scheme, which can either completely eliminate PC or significantly reduce it, depending on the system's COHI quantified in terms of the number of OFDM symbols. Specifically, we can always divide all the users into several groups, where the number of groups is no higher than the COHI and invoke time-shifted pilot transmission for the different groups. Specifically, the MSs of the first group simultaneously transmit their PS matrices, followed by the MSs of the second group, and so on. If the number of users in each group is no higher than the number of the mutually orthogonal elements in the PS matrix set $\mathbf{P}$, which can be guaranteed if the COHI is not too small, then PC can be completely eliminated. However, if the COHI is very small, the number of users in some of the groups will be larger than the number of the mutually orthogonal elements in $\mathbf{P}$. In such situations, although PC cannot be completely eliminated, it can be significantly mitigated.

III. PROPOSED PILOT CONTAMINATION ELIMINATION/REDUCTION SCHEME

The LU FD PS matrices $\mathbf{P}[i]$, $1 \leq i \leq LU$, of (10) are assigned to the LU users in the L cells. Based on the proof of Theorem 1, we know that when the condition (20) is not met, we cannot guarantee the orthogonality of all the FD PS matrices $\mathbf{P}[i]$, $1 \leq i \leq LU$. We first consider, how to optimally group the users for guaranteeing that the PS matrix subset associated with each user group contains mutually orthogonal elements.

TABLE II

TYPICAL SYSTEM PARAMETERS OF MULTI-CELL LARGE-SCALE MIMO TDD BASED OFDM SYSTEMS

System bandwidth B	20 MHZ
subcarrier spacing $\triangle f$	15 kHz
Number of subcarriers N	1024
Number of cells L	7
Length of CIRs K	$20 < K < 100$
Number of users per cell U	8

TABLE III

OPTIMALLY GROUPING THE USERS INTO f GROUPS GIVEN $R = 0$

$LU = n_u f + R, R = 0, i = (u-1)L + l,$ $1 \le i \le LU, 1 \le u \le U, 1 \le l \le L.$				
Group	User indexes i in each group			
1	1	$f+1$	$\cdots$	$LU - (f-1)$
2	2	$f+2$	$\cdots$	$LU - (f-2)$
$\vdots$			$\vdots$	
$f-1$	$f-1$	$2f-1$	$\cdots$	$LU - 1$
f	f	$2f$	$\cdots$	LU

A. Optimal User Grouping

Given the system parameters, namely the number of subcarriers N, length of CIRs K, number of cells L and number of MSs per cell U, we define

$$\zeta = \left\lfloor \frac{N}{LU} \right\rfloor, \quad f = \left\lceil \frac{K}{\zeta} \right\rceil, \quad n_u = \left\lfloor \frac{LU}{f} \right\rfloor,$$

$$R = \text{Rem}\left\{ \frac{LU}{f} \right\} = LU - n_u f, \qquad (24)$$

where R is the remainder of $\frac{LU}{f}$. In multi-cell MIMO based OFDM systems relying on UFR, the number of MSs served per cell U is not very high, while the length of CIRs K is large in rich scattering propagation environments. It may be readily verified that f is a small integer and n_u is larger than f, specifically,

$$2 \le f < L \text{ and } n_u > f. \qquad (25)$$

For example, given the typical system parameters of Table II, the relationships in (25) hold.

We have the following theorem specifying the optimal user grouping.

Theorem 3: Assume that the condition (20) is not met, i.e. that we have $KLU > N$. Then there exists an optimal user grouping, in which the number of groups is f or $f + 1$, and the number of users in each group is no more than $n_u + 1$. This grouping is optimal in the sense that the FD PS matrices assigned to the users of every group are mutually orthogonal, which ensures that the group-by-group time-shifted UL training is free from any PC.

Proof: Given $\zeta = \lfloor \frac{N}{LU} \rfloor < K$, we consider (16) again in order to search for optimal user grouping. Still under the condition of $i_1 > i_2$ and $t \equiv N + t$ if $t \le 0$, we have

$$\begin{cases} (i_1 - i_2)\zeta = (i_1 - i_2)\lfloor \frac{N}{LU} \rfloor, \\ 1 \le t \le K - 1 \text{ or } N - K + 1 \le t \le N. \end{cases} \qquad (26)$$

To satisfy $\delta(t - (i_1 - i_2)\zeta) = 0$, i.e., $t \neq (i_1 - i_2)\zeta$, $(i_1 - i_2)\lfloor \frac{N}{LU} \rfloor$ has to satisfy $K - 1 < (i_1 - i_2)\lfloor \frac{N}{LU} \rfloor < N - K + 1$. In other words, we must have

$$\frac{K}{\lfloor \frac{N}{LU} \rfloor} \le i_1 - i_2 \le \frac{N - K}{\lfloor \frac{N}{LU} \rfloor}, \qquad (27)$$

which holds for

$$f \le i_1 - i_2 \le \left\lfloor \frac{N - K}{\lfloor \frac{N}{LU} \rfloor} \right\rfloor. \qquad (28)$$

Hence, when a pair of user indices i_1 and i_2 satisfy (28), their assigned FD PS matrices are mutually orthogonal.

Since the minimum value of the difference between any two user indices i_1 and i_2 is f, we naturally opt for dividing the LU users into f groups. Specifically, we allocate the users whose indices are $1, f + 1, 2f + 1, \cdots$ into group 1; the users whose indices are $2, f + 2, 2f + 2, \cdots$ into group 2; and so on. Since we have $LU = n_u f + \text{Rem}\{\frac{LU}{f}\}$, depending on the value of the remainder $R = \text{Rem}\{\frac{LU}{f}\}$, the number of users in each group is either n_u or $n_u + 1$.

1) Case of $R = 0$: The user index set $I_{\bar{f}}$ for group $\bar{f}$ $(1 \le \bar{f} \le f)$ is

$$I_{\bar{f}} = \left\{ \bar{f}, \bar{f} + f, \bar{f} + 2f, \cdots, \bar{f} + (n_u - 1)f \right\}, \qquad (29)$$

and the number of users in each group is n_u. Since for any $i_1, i_2 \in I_{\bar{f}}$ and $i_1 > i_2$, $i_1 - i_2$ satisfies the condition (28), the FD PS matrices assigned to the users of each group are mutually orthogonal. This proves that the optimal user grouping is to divide the LU users into f groups, as illustrated in Table III, and the number of users in each group is n_u.

2) Case of $R \neq 0$: The remainder obeys $R \in \{1, 2, \cdots, f - 1\}$. By partitioning the users into f groups, the user index set $I_{\bar{f}}$ for group $\bar{f}$ $(1 \le \bar{f} \le f)$ is

$$I_{\bar{f}} = \begin{cases} \{\bar{f}, \bar{f} + f, \bar{f} + 2f, \cdots, \bar{f} + (n_u - 1)f\}, \\ \qquad \text{if } R - \bar{f} < 0, \\ \{\bar{f}, \bar{f} + f, \bar{f} + 2f, \cdots, \bar{f} + (n_u - 1)f, \bar{f} + n_u f\}, \\ \qquad \text{if } R - \bar{f} \ge 0. \end{cases} \qquad (30)$$

The number of users in the $\bar{f}$-th group is given by

$$\begin{cases} n_u + 0, & \text{if } R - \bar{f} < 0, \\ n_u + 1, & \text{if } R - \bar{f} \ge 0. \end{cases} \qquad (31)$$

Hence, the number of users in each group is either n_u or $n_u + 1$. This grouping is depicted in Table IV.

For $R - \bar{f} < 0$, i.e., for the groups $R + 1 \le \bar{f} \le f$, the number of users in each group is n_u and, moreover, for any $i_1, i_2 \in I_{\bar{f}}$ and $i_1 > i_2$, $i_1 - i_2$ meets the condition (28). Thus, all the n_u PS matrices in each group are mutually orthogonal.

However, for $R - \bar{f} \ge 0$, i.e., for the groups $1 \le \bar{f} \le R$, the number of users in each group is $n_u + 1$, and the number of mutually orthogonal PS matrices in group $\bar{f}$ is n_u or $n_u + 1$.

TABLE IV
GROUPING THE USERS INTO f GROUPS GIVEN $R \neq 0$

$LU = n_u f + R, R \in \{1,2,\cdots,f-1\}, i = (u-1)L+l, 1 \le i \le LU, 1 \le u \le U, 1 \le l \le L.$						
Group	User indexes i in each group					
1	1	$f+1$	$\cdots$	$(n_u-(R-1))f+1$	$\cdots$	$(n_u-1)f+1$	$n_u f+1$
2	2	$f+2$	$\cdots$	$(n_u-(R-1))f+2$	$\cdots$	$(n_u-1)f+2$	$n_u f+2$
$\vdots$				$\vdots$			
$R-1$	$R-1$	$f+R-1$	$\cdots$	$(n_u-(R-1))f+R-1$	$\cdots$	$(n_u-1)f+R-1$	$n_u f+R-1$
R	R	$f+R$	$\cdots$	$\mathbf{(n_u-(R-1))f+R}$	$\cdots$	$\mathbf{(n_u-1)f+R}$	$\mathbf{n_u f+R}$
$R+1$	$R+1$	$f+R+1$	$\cdots$	$(n_u-(R-1))f+R+1$	$\cdots$	$(n_u-1)f+R+1$	
$\vdots$				$\vdots$			
f	f	$2f$	$\cdots$	$(n_u-(R-2))f$	$\cdots$	$n_u f$	

TABLE V
OPTIMALLY GROUPING THE USERS INTO $f+1$ GROUPS GIVEN $R \neq 0$ AND $n_u f = LU - R > \left\lfloor \frac{N-K}{\lfloor \frac{N}{LU} \rfloor} \right\rfloor$

$LU = n_u f + R, R \in \{1,2,\cdots,f-1\}, i = (u-1)L+l, 1 \le i \le LU, 1 \le u \le U, 1 \le l \le L.$					
Group	User indexes i in each group				
1	1	$\cdots$	$(n_u-(R-1))f+1$	$(n_u-(R-2))f+2$	$\cdots$	$(n_u-1)f+R-1$
2	2	$\cdots$	$(n_u-(R-1))f+2$	$(n_u-(R-2))f+3$	$\cdots$	$(n_u-1)f+R+1$
$\vdots$			$\vdots$			
$R-1$	$R-1$	$\cdots$	$(n_u-(R-1))f+R-1$	$(n_u-(R-2))f+R+1$	$\cdots$	$(n_u-1)f+2R-2$
R	R	$\cdots$	$(n_u-(R-1))f+R+1$	$(n_u-(R-2))f+R+2$	$\cdots$	$(n_u-1)f+2R-1$
$R+1$	$R+1$	$\cdots$	$(n_u-(R-1))f+R+2$	$(n_u-(R-2))f+R+3$	$\cdots$	$(n_u-1)f+2R$
$\vdots$			$\vdots$			
f	f	$\cdots$	$(n_u-(R-2))f+1$	$(n_u-(R-3))f+2$	$\cdots$	$n_u f+R-1$
$f+1$	$(n_u-(R-1))f+R$	$\cdots$	$(n_u-1)f+R$	$n_u f+R$		

2.1) $R \neq 0$ and $n_u f = LU - R \le \left\lfloor \frac{N-K}{\lfloor \frac{N}{LU} \rfloor} \right\rfloor$: For a pair of user indices $i_1, i_2 \in I_{\bar{f}}$ and $i_1 > i_2$, where $1 \le \bar{f} \le R$, the condition (28) holds, and thus the number of mutually orthogonal PS matrices in group $\bar{f}$ is $n_u + 1$. Therefore, the grouping scheme of Table IV is optimal, since the PS matrices assigned to the users of every group are mutually orthogonal.

2.2) $R \neq 0$ and $n_u f = LU - R > \left\lfloor \frac{N-K}{\lfloor \frac{N}{LU} \rfloor} \right\rfloor$: For a pair of user indices $i_1, i_2 \in I_{\bar{f}}$ and $i_1 > i_2$, the condition (28) holds, except for the pair of the first and last users, whose indices $\bar{f}$ and $n_u f + \bar{f}$ satisfy $(n_u f + \bar{f}) - \bar{f} > \left\lfloor \frac{N-K}{\lfloor \frac{N}{LU} \rfloor} \right\rfloor$, i.e., do not meet the condition (28). In this case, we may partition the LU users into $f + 1$ groups, so that the number of users in each group is no more than n_u. Noting that $R \le f - 1$ and $f < n_u$, this is definitely achievable. Moreover, we must guarantee that any pair of user indices in each group meets the condition (28) to ensure that the PS matrices assigned to each group are mutually orthogonal. We now show how to construct such an optimal grouping of $(f + 1)$ user groups based on the f user groups of Table IV.

Step 0. We extract the R users $\mathbf{(n_u - (R - 1))f + R}, \cdots, (n_u - 1)f + R, n_u f + R$ from group R in Table IV, which are marked as boldface, to construct group $f + 1$. Clearly, this $(f+1)$-st group has R users, which is less than n_u, and the R users' FD PS matrices are mutually orthogonal.

The remaining $n_u f$ users in the f groups of Table IV are rearranged as follows. *Step 1*. At the column of $\mathbf{(n_u - (R - 1))f + R}$, move the elements of rows $R + 1$ to f one row up. This releases the position at the last f-th row of this column. *Step 2*. At the next column, move the elements of rows 1 to $R-1$ one row up, and the shifted-out 1st-row element fills the empty position of the previous column at the f-th row, while the elements of rows $R+1$ to f are moved two rows up, releasing the two positions at rows $f-1$ and f. The procedure continues until it reaches *Step R*, which moves the elements of rows 1 to $R - 1$ at the column of $\mathbf{(n_u f + R)}$ $(R - 1)$ rows up, i.e., shifts them out to fill the positions of the preceding column in rows $f-(R-2)$ to f. This procedure rearranges the f groups of Table IV into the $(f + 1)$ groups of Table V. It is clear that group $\bar{f}$ of Table V has n_u users, where $1 \le \bar{f} \le f$. We have to prove that the difference $i_1 - i_2$ between any two user indices i_1 and i_2 in each group meets the condition (28), assuming $i_1 > i_2$.

Note that in the original f groups of Table IV, the user indices in each group are in ascending order with the increment f, while at each column, the user indices are in ascending order with an increment of 1. Furthermore, we always have $i_1-i_2 \ge f$ for any two user indices $i_1 > i_2$ in any group.

From the construction procedure of *Step 1* to *Step R*, clearly, any two user indices $i_1 > i_2$ in group $\bar{f}$ of Table V satisfy $i_1 - i_2 \ge f$, where $1 \le \bar{f} \le f$. Moreover, the user indices in any group of Table V remain in ascending order and, therefore,
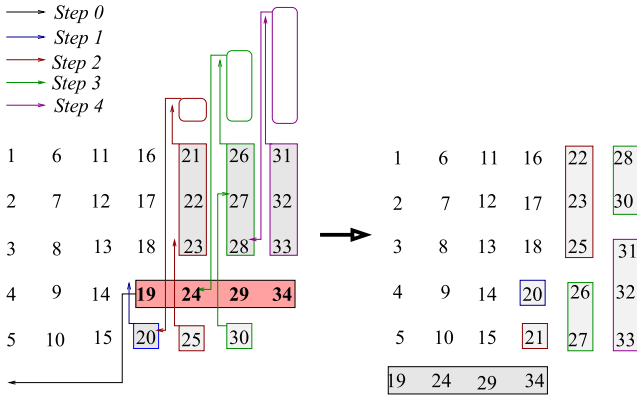
Fig. 1. Illustrative example of optimally grouping the users into $f+1$ groups given $R \neq 0$ and $n_u f = LU - R > \left\lfloor \frac{N-K}{\left\lfloor \frac{N}{LU} \right\rfloor} \right\rfloor$.

we only have to check that the difference of the last user index and the first user index is no higher than $\left\lfloor \frac{N-K}{\zeta} \right\rfloor$. For $\bar{f} = 1$, the maximum difference is $(n_u - 1)f + R - 2$, and for $2 \leq \bar{f} \leq f$, the maximum difference is $(n_u - 1)f + R - 1$. Obviously, $(n_u - 1)f + R - 1 \leq \left\lfloor \frac{N-K}{\zeta} \right\rfloor$. Thus, the condition (28) is met and the grouping scheme of Table V is optimal. ∎

We now provide an simple illustrative example for *Case 2.2)*. Consider the system associated with $N = 206$, $LU = 34$ and $K = 29$. Clearly, this is a system satisfying $N < KLU$ with

$$\zeta = \left\lfloor \frac{N}{LU} \right\rfloor = 6, \quad f = \left\lceil \frac{K}{\zeta} \right\rceil = 5,$$

$$n_u = \left\lfloor \frac{LU}{f} \right\rfloor = 6, \quad R = \mathrm{Rem}\left\{ \frac{LU}{f} \right\} = 4.$$

Furthermore,

$$n_u f = 30 > \left\lfloor \frac{N-K}{\zeta} \right\rfloor = 29,$$

and this is the case of $R \neq 0$ and $n_u f > \left\lfloor \frac{N-K}{\zeta} \right\rfloor$. Grouping the users into the f groups is depicted in the left side of Fig. 1, where it can be seen that for the group $\bar{f}$, where $1 \leq \bar{f} \leq R = 4$, the difference between the last user index and the first user index is $30 > \left\lfloor \frac{N-K}{\zeta} \right\rfloor = 29$. Hence, the PS matrices assigned to these two users are not orthogonal. The steps 0 to $R = 4$ of the rearranging procedure is also illustrated in Fig. 1, leading to the optimal grouping of $f + 1 = 6$ groups, as shown at the right side of Fig. 1. Clearly, the condition (28) is met for each of these $f + 1$ groups, and thus the PS matrices assigned to each group are mutually orthogonal.

Fig. 2 illustrates the TDD protocol's frame structure and its relationship to the channel's coherence time. Let us denote the channel's coherence time as t_{coherent}. We can define the overall COHI as the ratio of t_{coherent} over T_{OFDM}

$$r' = \left\lfloor \frac{t_{\mathrm{coherent}}}{T_{\mathrm{OFDM}}} \right\rfloor, \tag{32}$$
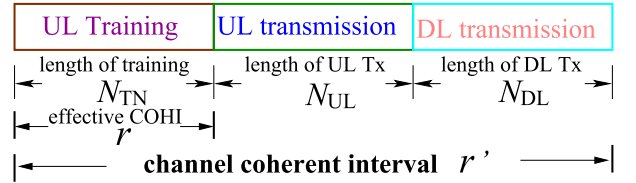

Fig. 2. TDD protocol frame structure and its relationship with the channel coherence time.

which specifies the maximum number of OFDM symbols during the duration of which the CIRs remain near constant. Since the system has to carry out the training and the uplink and/or downlink data transmission within r', we define the effective COHI for performing CE as

$$r = r' - (N_{\mathrm{UL}} + N_{\mathrm{DL}}), \tag{33}$$

where N_{UL} and N_{DL} are the numbers of OFDM symbols transmitted during the uplink and downlink data transmissions, respectively. Clearly, the training duration N_{TN} must satisfy $N_{\mathrm{TN}} \leq r$.

By partitioning the users into a number of groups that are no more than r, we can implement the CE based on a time-shifted pilot transmission procedure for the different groups. Explicitly, the MSs of the first group simultaneously transmit their PS matrices, followed by the MSs of the second group, and so on. According to Theorem 3, depending on the system's parameters, we can optimally partition the users into f or $f+1$ groups, as indicated in Tables III to V. Thus, we consider the cases of $r \geq f + 1$, $r = f$ and $r < f$.

B. Pilot Contamination Elimination

1) $r \geq f + 1$: Clearly, the time-shifted pilot transmission procedure for f or $f + 1$ groups can be realized. For $R = 0$, we can implement the optimal grouping scheme of Table III with f groups; for $R \neq 0$ and $LU - R \leq \left\lfloor \frac{N-K}{\zeta} \right\rfloor$ we can implement the optimal grouping scheme of Table IV with f groups; while for $R \neq 0$ and $LU - R > \left\lfloor \frac{N-K}{\zeta} \right\rfloor$, we can implement the optimal grouping scheme of Table V with $f+1$ groups. The CE based on the time-shifted pilot transmission is completely free from PC, and the MSE of the CE is as given in Theorem 2.

2) $r = f$: For $R = 0$ as well as $R \neq 0$ and $LU - R \leq \left\lfloor \frac{N-K}{\zeta} \right\rfloor$, we can implement the corresponding optimal grouping schemes in conjunction with f groups to achieve PC-free CE.

C. Pilot Contamination Reduction

In the presence of PC, the LS CE of $\underline{\mathbf{G}}_{l',l',q}^{u'}$ for the desired user $i_1 = (u' - 1)L + l'$ is given by

$$\widehat{\underline{\mathbf{G}}}_{l',l',q}^{u'} = \underline{\mathbf{G}}_{l',l',q}^{u'} + \sum_{i_2 : \mathbf{T} \neq \mathbf{0}_{K \times K}} \frac{K}{N} \left(\mathbf{P}[i_1]\mathbf{F} \right)^{\mathrm{H}} \mathbf{P}[i_2] \mathbf{F} \underline{\mathbf{G}}_{l,l',q}^{u}$$

$$+ \frac{K}{N\sqrt{p_r}} \left(\mathbf{P}[i_1]\mathbf{F} \right)^{\mathrm{H}} \underline{\mathbf{W}}_{l',q}, \tag{34}$$

where $i_2 = (u-1)L + l$, and $\mathbf{T} = (\mathbf{P}[i_1]\mathbf{F})^{\mathrm{H}}\mathbf{P}[i_2]\mathbf{F}$ as defined in (14). Hence the channel estimation error can be expressed as $\underline{\boldsymbol{\Xi}}_0 = \widehat{\underline{\mathbf{G}}}_{l',l',q}^{u'} - \underline{\mathbf{G}}_{l',l',q}^{u'} = \underline{\boldsymbol{\Xi}}_0^{\mathbf{G}} + \underline{\boldsymbol{\Xi}}_0^{\mathbf{W}}$, where

$$\underline{\boldsymbol{\Xi}}_0^{\mathbf{G}} = \sum_{i_2 : \mathbf{T} \neq \mathbf{0}_{K \times K}} \frac{K}{N} (\mathbf{P}[i_1]\mathbf{F})^{\mathrm{H}} \mathbf{P}[i_2]\mathbf{F}\underline{\mathbf{G}}_{l,l',q}^u, \quad (35)$$

$$\underline{\boldsymbol{\Xi}}_0^{\mathbf{W}} = \frac{K}{N\sqrt{p_r}} (\mathbf{P}[i_1]\mathbf{F})^{\mathrm{H}} \underline{\mathbf{W}}_{l',q}. \quad (36)$$

Since $\underline{\boldsymbol{\Xi}}_0^{\mathbf{G}}$ and $\underline{\boldsymbol{\Xi}}_0^{\mathbf{W}}$ are independent of each other, the MSE of the CE $\widehat{\underline{\mathbf{G}}}_{l',l',q}^{u'}$ can be expressed as

$$\overline{\Xi}_0 = \mathbb{E}\left\{ \frac{1}{K} \|\underline{\boldsymbol{\Xi}}_0\|_F^2 \right\} = \overline{\Xi}_0^{\mathbf{G}} + \overline{\Xi}_0^{\mathbf{W}}, \quad (37)$$

where $\overline{\Xi}_0^{\mathbf{G}} = \mathbb{E}\left\{ \frac{1}{K}\|\underline{\boldsymbol{\Xi}}_0^{\mathbf{G}}\|^2 \right\}$ and $\overline{\Xi}_0^{\mathbf{W}} = \mathbb{E}\left\{ \frac{1}{K}\|\underline{\boldsymbol{\Xi}}_0^{\mathbf{W}}\|_F^2 \right\}$. Obviously, $\overline{\Xi}_0^{\mathbf{W}} = \frac{K\sigma_w^2}{Np_r}$ as given in (23).

1) $r = f$ and $R \neq 0$, $LU - R > \left\lfloor \frac{N-K}{\zeta} \right\rfloor$: Clearly, the time-shifted pilot transmission procedure conceived for f groups can be realized. Hence, we can implement the grouping scheme of Table IV with f groups. This is not an optimal grouping, and the PC cannot be completely eliminated. However, only the PS matrices assigned to the last and first users in group $\bar{f}$ are not orthogonal for $1 \leq \bar{f} \leq R$. Therefore, the time-shifted pilot transmission procedure is capable of significantly reducing the PC, compared to the scenario where all the users simultaneously transmit their PS matrices, at the cost of increasing the training time by a small factor of f.

Theorem 4: Given $r = f$ as well as $R \neq 0$ and $LU - R > \left\lfloor \frac{N-K}{\zeta} \right\rfloor$, by implementing the grouping scheme of Table IV with f groups, the MSE of the CE based on the time-shifted pilot transmission procedure is as follows.

i) If the user $i_1 = (u'-1)L + l'$ is the last or the first user of group $\bar{f}$, where $1 \leq \bar{f} \leq R$, the MSE of the CE for this user is given by

$$\overline{\Xi}_0 = \frac{(K + (|(i_1 - i_2)\zeta| - N))\beta_{l,l'}^u}{K} + \frac{K\sigma_w^2}{Np_r}, \quad (38)$$

where $i_2 = (u-1)L + l$ denotes the index of the first or the last user of group $\bar{f}$.

ii) For any user other than the $2R$ users specified in i), the MSE of its CE is given by

$$\overline{\Xi}_0 = \frac{K\sigma_w^2}{Np_r}. \quad (39)$$

Proof: See Appendix B. ∎

2) $2 \leq r < f$: We can only realize the time-shifted pilot transmission procedure consisting of r groups. Hence we partition the LU users into the r groups of Table VI, where

$$n_u' = \left\lfloor \frac{LU}{r} \right\rfloor \text{ and } R' = \mathrm{Rem}\left\{ \frac{LU}{r} \right\}. \quad (40)$$

For any pair of user indices $i_1 = (u_1 - 1)L + l_1$ and $i_2 = (u_2 - 1)L + l_2$ of group $\bar{r}$, where $1 \leq \bar{r} \leq r$, we consider two situations:

TABLE VI
GROUPING THE USERS INTO r GROUPS GIVEN $r < f$

$LU = n_u'r + R'$, $i = (u-1)L + l$, $1 \leq i \leq LU$, $1 \leq u \leq U$, $1 \leq l \leq L$.					
Group	User indexes i in each group				
1	1	$r+1$	$\cdots$	$LU - (2r-2)$	$LU - (R'-1)$
2	2	$r+2$	$\cdots$	$LU - (2r-3)$	$LU - (R'-2)$
$\vdots$			$\vdots$		
R'	R'	$R'+r$	$\cdots$	$R' + r(n_u'-1)$	LU
$\vdots$			$\vdots$		
$r-1$	$r-1$	$2r-1$	$\cdots$	$LU - (R'+1)$	
r	r	$2r$	$\cdots$	$LU - R'$	

i) $i_1 > i_2$: If the condition (28) is met, their PS matrices are mutually orthogonal. Therefore, the user pairs associated with non-orthogonal PS matrices are specified by

$$1 \leq i_1 - i_2 < f \text{ or } \left\lfloor \frac{N-K}{\zeta} \right\rfloor < i_1 - i_2 \leq LU - 1. \quad (41)$$

ii) $i_1 < i_2$: If

$$-\left\lfloor \frac{N-K}{\zeta} \right\rfloor \leq i_1 - i_2 \leq -f \quad (42)$$

is met, their PS matrices are mutually orthogonal. Therefore, the user pairs having non-orthogonal PS matrices are specified by

$$-f < i_1 - i_2 \leq -1 \text{ or } -LU+1 \leq i_1 - i_2 < -\left\lfloor \frac{N-K}{\zeta} \right\rfloor. \quad (43)$$

Regarding the MSE of the CE obtained by this time-shifted pilot transmission procedure of r groups, we have the following theorem.

Theorem 5: Given $r < f$, by implementing the grouping scheme of Table VI with r groups, the MSE of the CE based on the time-shifted pilot transmission procedure for user $i_1 = (u_1 - 1)L + l_1$ of group $\bar{r}$ is

$$\begin{aligned}
\overline{\Xi}_0 = & \sum_{(i_1 - i_2)\zeta \in \{1, \cdots, K-1\}} \frac{(K - (i_1 - i_2)\zeta)\beta_{l_2, l_1}^{u_2}}{K} \\
& + \sum_{(i_1 - i_2)\zeta \in \{N-K+1, \cdots, N\}} \frac{(K + ((i_1 - i_2)\zeta - N))\beta_{l_2, l_1}^{u_2}}{K} \\
& + \sum_{(i_1 - i_2)\zeta \in \{-K+1, \cdots, -1\}} \frac{(K + (i_1 - i_2)\zeta)\beta_{l_2, l_1}^{u_2}}{K} \\
& + \sum_{(i_1 - i_2)\zeta \in \{-N, \cdots, -N+K-1\}} \frac{(K - ((i_1 - i_2)\zeta + N))\beta_{l_2, l_1}^{u_2}}{K} \\
& + \frac{K\sigma_w^2}{Np_r},
\end{aligned} \quad (44)$$

where user $i_2 = (u_2 - 1)L + l_2$ is in group $\bar{r}$ and $1 \leq \bar{r} \leq r$.
Proof: See Appendix C. ∎

D. Comparison With Existing Schemes

As mentioned previously, all the existing schemes [6], [10], [11], [14]–[17] consider the received signal vector over all the target BS's antennas for an individual subcarrier, and each user transmits an UL pilot sequence of length U for UL training. Therefore, even to implement a conventional simultaneous UL training, which inevitably suffers from serious PC with the exception of the location-aware CE scheme [14], all these schemes require the effective COHI to be no shorter than U. By contrast, our approach considers the received signal vector over all the N OFDM subcarriers for an individual BS antenna, and it only requires the minimum effective COHI of 1 to implement a simultaneous UL training. Moreover, if we have sufficient subcarrier resources, namely, the condition (20) is met, our design guarantees a PC-free CE even with a simultaneous UL training. Only when the condition (20) is not met would our design suffer from PC in a simultaneous UL training regime. In the case of having insufficient subcarrier resources, i.e., for $N < KLU$, our approach can invoke the time-shifted pilot transmission procedure of f or $f + 1$ groups for completely eliminating PC, which requires that the effective COHI is no shorter than $f + 1$.

Many existing schemes rely on user related features, such as user locations or users' channel covariance matrices, to combat PC. Therefore, they must acquire these statistics, which is generally costly and requires considerable information exchange between cells. After the acquisition of this user-related knowledge, typically complex centralized pilot assignment is needed to match the user-related information. Moreover, these user-related parameters are time-varying. When they change, the whole process has to be repeated again. By contrast, our proposed scheme does not require any of these. Explicitly, no user related features or statistics are needed, no information exchange is needed between cells, no pilot assignment is required, and above all, our optimal design does not change at all. More specifically, our design is produced at the network planning stage, as it only relies on the network parameters, namely, the number of cells L, the number of antennas at each BS Q, the maximum number of MSs supported per cell U, the number of subcarriers N, the maximum delay spread or the maximum length of the CIRs K, and the effective COHI r. Since our design does not depend on any user related features or statistics that are changing with time, it remains unchanged during the entire network operational life time.

The scheme of [16] also does not rely on user related statistics but it requires that the effective COHI is no shorter than $(L + 1)U$. Since typically $f < L$, we have $f + 1 \ll (L + 1)U$, and the significant advantage of our proposed scheme over the existing scheme of [16] becomes apparent. Additionally, given $r \geq (L + 1)U$, although the scheme of [16] can also completely eliminate PC, the associated signal cancellation operations of this scheme amplify the noise. Therefore, we expect that the achievable MSE of our scheme will be better than that of [16]. Most importantly, if the effective COHI is shorter than $(L + 1)U$, the scheme of [16] cannot be used at all. By contrast, for any $2 \leq r \leq f$, our approach can implement the time-shifted pilot transmission

procedure consisting of r groups for significantly reducing the PC. Thus the efficiency of our design in combating PC naturally depends on the COHI encountered. In the worst-case of having an effective COHI equal to 1 and insufficient subcarrier resources, our design 'degrades' gracefully to a simultaneous UL training. It is worth emphasizing that this simultaneous UL training can be implemented under the minimum effective COHI of $r = 1$. Furthermore, as a benefit of an optimal pilot design, our simultaneous UL training remains more immune to PC than the conventional simultaneous UL training scheme, which incidentally can only be realized under the condition of $r \geq U > 1$.

It is worth pointing our that training reduces the effective bandwidth efficiency and, therefore, the training duration has a significant impact on the effective sum rate achievable. Let C_{UL} and C_{DL} denote the ideal UL sum-rate and ideal DL sum-rate for the UL and DL transmissions, respectively, without taking into account the training overhead. Referring to Fig. 2, since training reduces the system's effective throughput, the effective UL sum-rate C_{UL}^{ef} and effective DL sum-rate C_{DL}^{ef} are obtained respectively as

$$C_{\mathrm{UL}}^{ef} = \frac{N_{\mathrm{UL}}}{\frac{1}{2}N_{\mathrm{TN}} + N_{\mathrm{UL}}} C_{\mathrm{UL}}, \qquad (45)$$

$$C_{\mathrm{DL}}^{ef} = \frac{N_{\mathrm{DL}}}{\frac{1}{2}N_{\mathrm{TN}} + N_{\mathrm{DL}}} C_{\mathrm{DL}}, \qquad (46)$$

by a training overhead adjustment $\frac{1}{2}N_{\mathrm{TN}}$, where the factor $\frac{1}{2}$ is owing to the fact that the channel estimate obtained by UL training is used in both UL and DL transmissions. Since both our scheme and the scheme of [16] can eliminate PC, assuming a sufficiently long r, both will achieve similar C_{UL} and C_{DL}, which are significantly higher than those attained by the scheme of conventional simultaneous UL training. However, the actual effective sum-rates, C_{UL}^{ef} and C_{DL}^{ef}, achieved by our scheme are much higher than those achieved by the scheme of [16], since the former requires a much shorter training duration. In fact, even the conventional simultaneous UL training may attain higher C_{UL}^{ef} and C_{DL}^{ef} than the scheme of [16], as will be confirmed by the simulation results.

IV. SIMULATION RESULTS

The default values of the various parameters used for our simulated multi-cell TDD system are listed in Table VII. Unless otherwise specified, these default parameter values were used throughout. The UFR regime was assumed and the pre-assigned PSs in (10) were employed for the UL CE by all the BSs. The signal-to-noise ratio (SNR) of the system was defined as E_s/N_0, where E_s denoted the energy per symbol and $N_0 = \sigma_w^2$ denoted the power of the AWGN. The ULA was deployed at each BS. All the AOAs $\theta_{l',l,q,k}^u$ were i.i.d. Gaussian random variables with a mean of $\bar{\theta} = 90°$ and the standard deviation of $\sigma_{\mathrm{AOA}} = 90°$.

Clearly, this is a system with $N < KLU$, and we have

$$\zeta = \left\lfloor \frac{N}{LU} \right\rfloor = 18, \quad f = \left\lceil \frac{K}{\zeta} \right\rceil = 3,$$

$$n_u = \left\lfloor \frac{LU}{f} \right\rfloor = 18, \quad R = \mathrm{Rem}\left\{ \frac{LU}{f} \right\} = 2.$$

TABLE VII

DEFAULT PARAMETERS USED IN THE SIMULATED MULTIPLE-ANTENNA AIDED AND TDD BASED OFDM SYSTEM

Number of cells L	7
Radius of each cell	1000 m
Number of MSs per-cell U	8
Number of antennas at each BS Q	100
Average transmit power per subcarrier at each MS p_r	0 dB
Average transmit power per subcarrier at each BS p_f	10 dB
Path loss exponent	3
Mean of path AOAs θ	90°
Standard deviation of path AOAs σ_{AOA}	90°
Antenna spacing D	$\frac{\lambda}{2}$
Length of CIRs K	54
Number of subcarriers N	1024

Furthermore, since we have

$$n_u f = 54 > \left\lfloor \frac{N-K}{\zeta} \right\rfloor = 53,$$

this is the case of $R \neq 0$ and $n_u f > \left\lfloor \frac{N-K}{\zeta} \right\rfloor$. The corresponding optimal grouping is defined in Table V with $f + 1 = 4$ groups. If the effective COHI obeys $r \geq f + 1 = 4$, our 4-group based optimal UL training scheme can be invoked for completely eliminating the PC. If $r = 3$, our 3-group based UL training scheme can be employed for significantly mitigating the PC effect. Even under the worst-case condition of $r = 1$, we can still realize our 1-group based simultaneous UL training procedure.

A. Estimation Performance

We first verify the effectiveness of our proposed design by examining the normalized MSE (NMSE) of the CE, which is defined as

$$\text{NMSE} = \frac{\sum_{l=1}^{L}\sum_{u=1}^{U}\sum_{q=1}^{Q}\sum_{n=1}^{N}\left|\widehat{H}_{l,l,q}^{u}[n] - H_{l,l,q}^{u}[n]\right|^2}{\sum_{l=1}^{L}\sum_{u=1}^{U}\sum_{q=1}^{Q}\sum_{n=1}^{N}\left|H_{l,l,q}^{u}[n]\right|^2}. \quad (47)$$

Fig. 3 investigates the NMSE of our estimator as a function of the UL training SNR under different effective COHI conditions, where the simulated $\text{NMSE}_{\text{simu}}$ results are obtained by averaging over 100 channel realizations and they agree well with the theoretical $\text{NMSE}_{\text{theo}}$ results as given in Theorems 2 to 5, respectively. Specifically, if the effective COHI is no smaller than 4, our 4-group based optimal scheme is capable of completely eliminating PC. Under the condition of $r = 3$, our 3-group based scheme can be employed for significantly reducing PC resulting in an NMSE floor of 4×10^{-5}, as seen from Fig. 3. If the effective COHI is reduced to $r = 2$, our 2-group based scheme has an NMSE floor of 0.01. Under the worst operating condition of $r = 1$, our 1-group based simultaneous training scheme has an NMSE floor of 0.05. Observe that the conventional simultaneous training scheme can only be implemented if the effective COHI is no shorter than 8 OFDM symbols. Furthermore, it has a higher NMSE floor of 0.1. Vu *et al.* [16] proposed two schemes both of
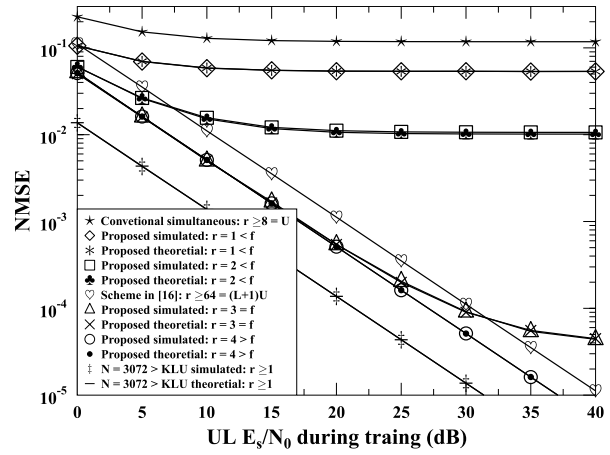

Fig. 3. NMSE performance of different channel estimators as the functions of the UL training SNR.

which are capable of eliminating PC and both require that the effective COHI is no shorter than 64. This should be compared to our scheme, which only requires that $r \geq 4$. We use scheme 2 of [16] in our comparison, which has a better performance than the other one. It can be seen from Fig. 3 that scheme 2 of [16] eliminates PC with no NMSE floor, but its performance is 2 dB worse than our 4-group based optimal scheme, because the signal cancellation operations required in this scheme amplify the noise.

Additionally, our optimal pilot design is always capable of attaining a PC-free estimate in the simultaneous training, if $N \geq KLU$. Thus, we also consider the case of $N = 3072 > KLU$, while all the other system parameters remain the same, as in Table VII. With a sufficient subcarrier resource of $N = 3072 > KLU$, the FD PS matrices designed for all the users in all the cells retain the desired orthogonality, and the corresponding NMSE performance obtained by our simultaneous training procedure is also depicted in Fig. 3, which is about 4.5 dB lower than the PC-free case of $N = 1024$, since the MSE is inversely proportional to N.

B. Ideal Sum-Rate Performance

We next investigate the achievable ideal UL sum-rate without considering the impact of training duration based on various channel estimators. Specifically, the channel estimate is first obtained during the UL training. Then during the UL transmission, the system's SNR is maintained to the same UL training SNR value, and each BS performs the maximum-ratio combining (MRC) of its received signal by multiplying it with the conjugate-transpose of the channel estimate. The ideal per-cell UL sum-rate is defined by

$$C_{\text{UL}} = \frac{1}{L}\sum_{l=1}^{L}\sum_{u=1}^{U}\log_2\left(1 + \text{SINR}_{l,u}\right), \quad (48)$$

where $\text{SINR}_{l,u}$ is the desired signal-to-interference-plus-noise ratio of user u in cell l, obtained by the MRC based on the estimated channels. Fig. 4 portrays the ideal per-cell UL sum-rate performance. The results of Fig. 4 are consistent with the results of Fig. 3, i.e., a better NMSE CE performance generally leads to a higher ideal UL sum-rate. Although the
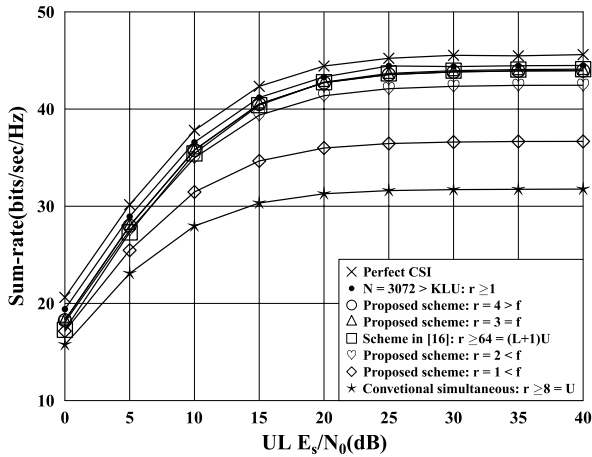
Fig. 4.   Achievable ideal per-cell UL sum-rate performance as the functions of the UL system's SNR by different estimators with the UL training SNR equal to the UL system's SNR.
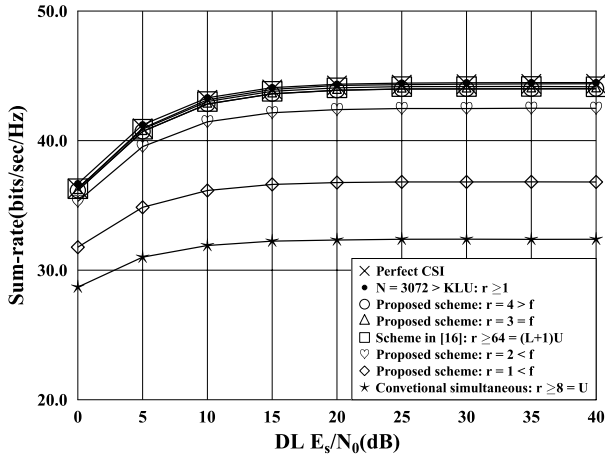

Fig. 5.   Achievable ideal per-cell DL sum-rate performance as the functions of the DL system's SNR by different estimators where the UL training SNR is fixed to $E_s/N_0 = 20$ dB.

achievable ideal UL sum-rate of the scheme proposed in [16] is very close to that of our optimal-grouping scheme, it is worth emphasizing again that our optimal scheme can operate under the network condition of an effective COHI $r \geq 4$, while the scheme of [16] can only operate for an effective COHI of $r \geq 64$. Also from Fig. 4 it can be seen that our proposed simultaneous training scheme, which only requires the minimum effective COHI of $r = 1$, achieves a 5 bits/s/Hz higher sum-rate than the conventional simultaneous training scheme, which requires an effective COHI of $r = 8$.

We then study the achievable ideal DL sum-rate performance. Specifically, given the channel estimate obtained under the UL training SNR of $E_s/N_0 = 20$ dB, each BS carries out DL transmission by invoking zero-forcing (ZF) transmit precoding based on the UL CE. The per-cell DL sum-rate, C_{DL}, is defined similarly to C_{UL} of (48). Fig. 5 portrays the achievable ideal per-cell DL sum-rate versus the DL system's SNR. It can be seen from Fig. 5 that the ideal per-cell DL sum-rate achieved by our simultaneous training scheme with sufficient subcarrier resource of $N = 3072$ approaches the perfect-CSI bound, while our 4-group based scheme with $N = 1024$ and the scheme of [16] perform
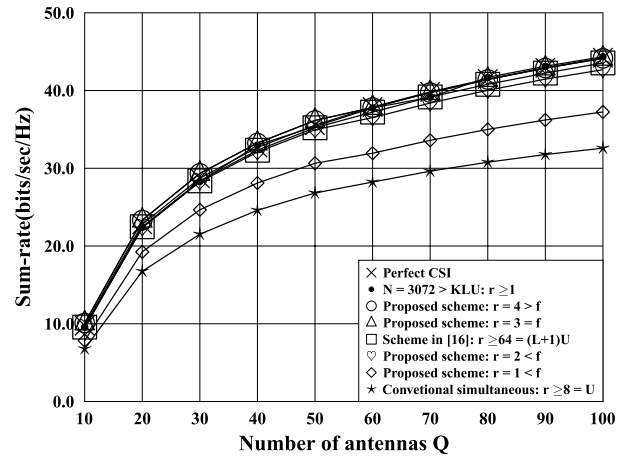

Fig. 6.   Achievable ideal per-cell DL sum-rate performance as the functions of the number of antennas Q by different estimators where both the UL training SNR and the DL system's SNR are fixed to 20 dB.

very close to the perfect-CSI bound. Observe furthermore that our simultaneous training scheme associated with $N = 1024$ achieves a 5 bits/s/Hz higher sum-rate than the conventional simultaneous training scheme.

We also investigate the effect of the number of antennas Q. In particular, we fix the UL training SNR to 20 dB, vary the number of antennas Q, and estimate the corresponding channel matrix. Then each BS carries out ZF transmit precoding based DL transmission using the obtained CE under the DL system's SNR of $E_s/N_0 = 20$ dB. Fig. 6 portrays the ideal per-cell DL sum-rate versus the number of antennas Q. As expected, increasing the number of antennas enhances the achievable sum-rate. It can be seen again from Fig. 6 that the ideal per-cell DL sum-rates achieved by the three PC-free schemes, namely, our scheme using $N = 3072 > KLU$, our optimal grouping scheme with $N = 1024 < KLU$ and the scheme of [16], perform very close to the perfect-CSI bound. Our simultaneous UL training scheme relying on $N = 1024 < KLU$ is substantially better than the conventional simultaneous UL training scheme.

C. Effective Sum-Rate Performance

The ideal UL sum-rate C_{UL} and ideal DL sum-rate C_{DL} shown in Section IV-B do not take into account the training overhead imposed. We now investigate the effective UL sum-rate C_{UL}^{ef} and effective DL sum-rate C_{DL}^{ef} achievable by various schemes. We consider a very slow fading system associated with the COHI $r' = 84$ so that the scheme of [16] can be implemented in conjunction with $N_{\text{TN}} = 64$ and $N_{\text{UL}} = N_{\text{DL}} = 10$ OFDM symbols. Since our proposed 4-group, 3-group, 2-group and 1-group based schemes require $N_{\text{TN}} = 4, 3, 2$ and 1, respectively, they can implement the UL and DL transmissions with the aid of $N_{\text{UL}} = N_{\text{DL}} = 40, 40.5, 41, 41.5$, respectively. On the other hand, the conventional simultaneous UL training scheme requires $N_{\text{TN}} = 8$, and it can support the UL and DL transmissions with $N_{\text{UL}} = N_{\text{DL}} = 38$.

Under the identical experimental conditions of Figs. 4 to 6, the effective sum-rate performance which properly consider the impact of training overhead are depicted in Figs. 7 to 9, respectively. Observe from Figs. 7 to 9 that our proposed
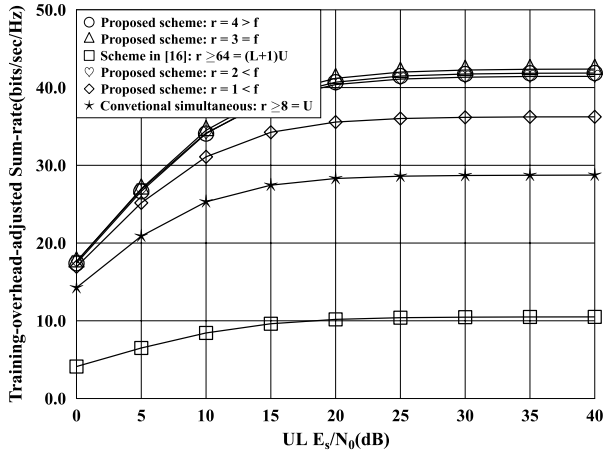
Fig. 7. Achievable effective per-cell UL sum-rate performance as the functions of the UL system's SNR by different estimators with the UL training SNR equal to the UL system's SNR. The COHI is $r' = 84$.
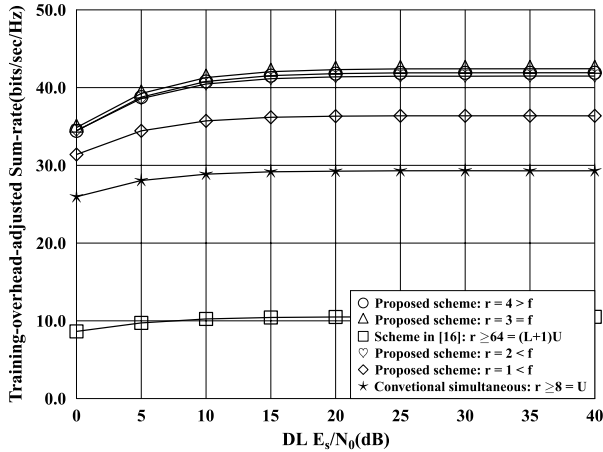

Fig. 8. Achievable effective per-cell DL sum-rate performance as the functions of the DL system's SNR by different estimators where the UL training SNR is fixed to $E_s/N_0 = 20$ dB. The COHI is $r' = 84$.
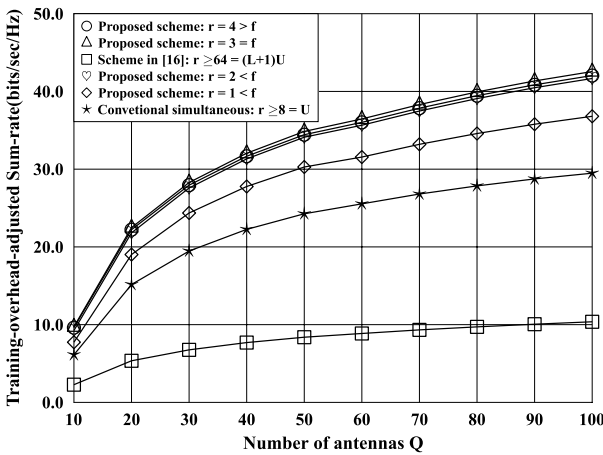

Fig. 9. Achievable effective per-cell DL sum-rate performance as the functions of the number of antennas Q by different estimators where both the UL training SNR and the DL system's SNR are fixed to 20 dB. The COHI is $r' = 84$.

4-group, 3-group and 2-group based schemes outperform the scheme of [16] by more than 30 bits/sec/Hz, while our proposed simultaneous UL training scheme outperforms the

the scheme of [16] by more than 25 bits/sec/Hz, in terms of effective sum-rate performance. In fact, even the conventional simultaneous UL training scheme achieves higher effective sum-rate performance than the scheme of [16] by almost 20 bits/sec/Hz, as the benefit of requiring a much shorter training duration, even though it suffers from serious PC while the latter is PC-free.

V. CONCLUSIONS

A novel PC elimination scheme has been proposed for multi-cell TDD and OFDM based massive MIMO systems, which relies on an optimal pilot design for TD CE. It has been shown that when sufficient subcarrier resources are available in conjunction with our optimal pilot design, all users in all cells can simultaneously perform UL training completely free from PC, even under the worst-case effective COHI of 1. When the subcarrier resources are insufficient, users in all the cells can be grouped into a very small optimal number of groups, and the pilots associated with each group are guaranteed to be orthogonal. If the system's effective COHI is no smaller than the optimal number of groups, the pilot transmissions destined for different groups can be implemented in a time-shifted manner, which completely eliminates pilot contamination. Moreover, even when the effective COHI is smaller than the optimal number of groups, our time-shifted scheme can still be implemented according to the actual system's effective COHI for significantly reducing the PC. Our proposed scheme requires no user related features or statistics, no information exchange between cells and no pilot assignment. After the design at the network planning stage, it remains unchanged during the entire network operational life time.

Pilot contamination has been a stumbling block in the way of realizing massive MIMO systems. The existing PC elimination schemes either require user related features that are changing with time or need excessively long effective COHIs, and therefore are impractical. The significance of our design is that it solves this challenging PC problem, since it provides an effective and practical technique of eliminating PC in multi-cell TDD and OFDM based massive MIMO systems.

APPENDIX

A. Proof of Theorem 2

From (21) we have

$$\underline{\Xi}_0 = \frac{K}{N\sqrt{p_r}} \big(\mathbf{P}[(u'-1)L+l']\mathbf{F}\big)^{\mathrm{H}} \underline{\mathbf{W}}_{l',q}. \tag{49}$$

The k-th element of $\underline{\Xi}_0$ is

$$\underline{\Xi}_{0_{[k]}} = \frac{\sqrt{K}}{N\sqrt{p_r}} \sum_{n=1}^{N} e^{\frac{\mathrm{j}2\pi(n-1)(k-1)}{N}} \big(\mathbf{P}[1]_{[n,n]}\big)^*$$
$$\times e^{\frac{-\mathrm{j}2\pi\big((u'-1)L+l'-1\big)\zeta(n-1)}{N}} W_{l',q}[n]. \tag{50}$$

Since the noise obeys $W_{l',q}[n] \sim \mathcal{CN}(0, \sigma_w^2)$, $\underline{\Xi}_{0_{[k]}}$ is also Gaussian distributed with zero mean and the power given by

$$\mathbb{E}\big\{\big|\underline{\Xi}_{0_{[k]}}\big|^2\big\} = \frac{K}{N^2 p_r} N\sigma_w^2 = \frac{K\sigma_w^2}{Np_r}. \tag{51}$$

Clearly, the MSE of the LS CE (21) satisfies the Cramer-Rao lower bound (CRLB) [20].

B. Proof of Theorem 4

i) Only the PS matrices assigned to the first and last users of group $\bar{f}$, where $1 \leq \bar{f} \leq R$, are non-orthogonal. If the user $i_1 = (u' - 1)L + l'$ is the last user and $i_2 = (u - 1)L + l$ is the first user of group $\bar{f}$, we have $i_1 - i_2 = n_u f > 0$, and $(i_1 - i_2)\zeta = n_u f \zeta = \lfloor \frac{LU}{f} \rfloor f \lfloor \frac{N}{LU} \rfloor \leq N$. Because $LU - R = n_u f > \lfloor \frac{N-K}{\zeta} \rfloor$, we also have $(i_1 - i_2)\zeta = n_u f \zeta > \lfloor \frac{N-K}{\zeta} \rfloor \zeta$ and $(i_1 - i_2)\zeta \geq N - K + 1$, for $(i_1 - i_2)\zeta \in \{N-K+1, \cdots, N\}$.

Consider $\mathbf{T} = (\mathbf{P}[i_1]\mathbf{F})^H \mathbf{P}[i_2]\mathbf{F}$ of (14) and its element $\mathbf{T}_{[k_1,k_2]}$ of (16). For $(i_1 - i_2)\zeta = t + N = k_1 - k_2 + N \in \{N-K+1, \cdots, N\}$, $\mathbf{T}_{[k_1,k_2]} = \frac{N}{K}$. Then there are $K + \big((i_1 - i_2)\zeta - N\big)$ pairs of (k_1, k_2), specifically, $\big(1, 1 - ((i_1 - i_2)\zeta - N)\big), \big(2, 2 - ((i_1 - i_2)\zeta - N)\big), \cdots, \big(K + ((i_1 - i_2)\zeta - N), K\big)$, for which $\mathbf{T}_{[k_1,k_2]} = \frac{N}{K}$, while the rest of the elements in $\mathbf{T}$ are all equal to 0. Hence, for $1 \leq k \leq K + \big((i_1 - i_2)\zeta - N\big)$, the k-th element of $\underline{\boldsymbol{\Xi}}_0^{\mathbf{G}}$ is

$$\underline{\boldsymbol{\Xi}}_{0_{[k]}}^{\mathbf{G}} = G_{l,l',q}^u[k]. \tag{52}$$

Because $G_{l,l',q}^u[k]$ is an independently identically distributed (i.i.d.) complex-valued Gaussian random variable with zero mean and power $\beta_{l,l'}^u$, i.e., $G_{l,l',q}^u[k] \sim \mathcal{CN}(0, \beta_{l,l'}^u)$, $\underline{\boldsymbol{\Xi}}_{0_{[k]}}^{\mathbf{G}}$ is also a complex-valued Gaussian variable with zero mean and the power given by

$$\mathbb{E}\big\{|\underline{\boldsymbol{\Xi}}_{0_{[k]}}^{\mathbf{G}}|^2\big\} = \beta_{l,l'}^u, \quad 1 \leq k \leq K + \big((i_1 - i_2)\zeta - N\big). \tag{53}$$

On the other hand, for $K + \big((i_1 - i_2)\zeta - N\big) + 1 \leq k \leq K$, the k-th element of $\underline{\boldsymbol{\Xi}}_0^{\mathbf{G}}$ is 0. Thus we arrive at

$$\overline{\boldsymbol{\Xi}}_0^{\mathbf{G}} = \mathbb{E}\Big\{\frac{1}{K}\|\underline{\boldsymbol{\Xi}}_{0_{[k]}}^{\mathbf{G}}\|_F^2\Big\} = \frac{\big(K + ((i_1 - i_2)\zeta - N)\big)\beta_{l,l'}^u}{K}, \tag{54}$$

and hence the MSE of (38).

If the user $i_1 = (u' - 1)L + l'$ is the first user and $i_2 = (u - 1)L + l$ is the last user of group $\bar{f}$, we can also arrive at (38) in a similar way.

ii) Proof of (39) is straightforward.

C. Proof of Theorem 5

i.a) For two users $i_1 = (u_1 - 1)L + l_1$ and $i_2 = (u_2 - 1)L + l_2$ of group $\bar{r}$ with $i_1 - i_2 > 0$, where $1 \leq \bar{r} \leq r$, if $(i_1 - i_2)\zeta \in \{1, \cdots, K - 1\}$, the PS matrices assigned to these two users are non-orthogonal.

Consider $\mathbf{T} = (\mathbf{P}[i_1]\mathbf{F})^H \mathbf{P}[i_2]\mathbf{F}$ and its element $\mathbf{T}_{[k_1,k_2]}$. Clearly, for $(i_1 - i_2)\zeta = t = k_1 - k_2 \in \{1, 2, \cdots, K - 1\}$, $\mathbf{T}_{[k_1,k_2]} = \frac{N}{K}$. That is, there are $K - (i_1 - i_2)\zeta$ pairs of (k_1, k_2), specifically, $\big((i_1 - i_2)\zeta + 1, 1\big), \big((i_1 - i_2)\zeta + 2, 2\big), \cdots, \big(K, K - (i_1 - i_2)\zeta\big)$, for which $\mathbf{T}_{[k_1,k_2]} = \frac{N}{K}$, while the rest of the elements in $\mathbf{T}$ are all equal to 0. Hence, for $1 \leq k \leq (i_1 - i_2)\zeta$, $\underline{\boldsymbol{\Xi}}_{0_{[k]}}^{\mathbf{G}} = 0$, while

$$\underline{\boldsymbol{\Xi}}_{0_{[k]}}^{\mathbf{G}} = G_{l_2,l_1,q}^{u_2}[k], \quad (i_1 - i_2)\zeta + 1 \leq k \leq K, \tag{55}$$

which leads to

$$\mathbb{E}\big\{|\underline{\boldsymbol{\Xi}}_{0_{[k]}}^{\mathbf{G}}|^2\big\} = \beta_{l_2,l_1}^{u_2}, \quad (i_1 - i_2)\zeta + 1 \leq k \leq K. \tag{56}$$

Thus we have

$$\overline{\boldsymbol{\Xi}}_0^{\mathbf{G}} = \mathbb{E}\Big\{\frac{1}{K}\|\underline{\boldsymbol{\Xi}}_0^{\mathbf{G}}\|_F^2\Big\} = \frac{\big(K - (i_1 - i_2)\zeta\big)\beta_{l_2,l_1}^{u_2}}{K},$$
$$(i_1 - i_2)\zeta \in \{1, \cdots, K - 1\}. \tag{57}$$

i.b) For the pair of users $i_1 = (u_1 - 1)L + l_1$ and $i_2 = (u_2 - 1)L + l_2$ of group $\bar{r}$ and $i_1 - i_2 > 0$, where $1 \leq \bar{r} \leq r$, if $(i_1 - i_2)\zeta \in \{N - K + 1, \cdots, N\}$, the PS matrices assigned to these two users are non-orthogonal. Hence, for $(i_1 - i_2)\zeta = t + N = k_1 - k_2 + N \in \{N - K + 1, \cdots, N\}$, $\mathbf{T}_{[k_1,k_2]} = \frac{N}{K}$. That is, there are $K + \big((i_1 - i_2)\zeta - N\big)$ pairs of (k_1, k_2), specifically, $\big(1, 1 - ((i_1 - i_2)\zeta - N)\big), \big(2, 2 - ((i_1 - i_2)\zeta - N)\big), \cdots, \big(K + ((i_1 - i_2)\zeta - N), K\big)$, for which $\mathbf{T}_{[k_1,k_2]} = \frac{N}{K}$, while the rest of the elements in $\mathbf{T}$ are all equal to 0. Hence

$$\underline{\boldsymbol{\Xi}}_{0_{[k]}}^{\mathbf{G}} = G_{l_2,l_1,q}^{u_2}[k], \quad 1 \leq k \leq K + \big((i_1 - i_2)\zeta - N\big), \tag{58}$$

or

$$\mathbb{E}\big\{|\underline{\boldsymbol{\Xi}}_{0_{[k]}}^{\mathbf{G}}|^2\big\} = \beta_{l_2,l_1}^{u_2}, \quad 1 \leq k \leq K + \big((i_1 - i_2)\zeta - N\big), \tag{59}$$

while for $K + \big((i_1 - i_2)\zeta - N\big) + 1 \leq k \leq K$, $\underline{\boldsymbol{\Xi}}_{0_{[k]}}^{\mathbf{G}} = 0$. Therefore, we arrive at

$$\overline{\boldsymbol{\Xi}}_0^{\mathbf{G}} = \frac{\big(K + ((i_1 - i_2)\zeta - N)\big)\beta_{l_2,l_1}^{u_2}}{K},$$
$$(i_1 - i_2)\zeta \in \{N - K + 1, \cdots, N\}. \tag{60}$$

ii.a) For the pair of users $i_1 = (u_1 - 1)L + l_1$ and $i_2 = (u_2 - 1)L + l_2$ of group $\bar{r}$ and $i_1 - i_2 < 0$, where $1 \leq \bar{r} \leq r$, if $(i_1 - i_2)\zeta \in \{-K + 1, \cdots, -1\}$, the PS matrices assigned to these two users are non-orthogonal. Clearly, for $(i_1 - i_2)\zeta = t = k_1 - k_2 \in \{-K + 1, \cdots, -1\}$, $\mathbf{T}_{[k_1,k_2]} = \frac{N}{K}$. That is, there are $K + (i_1 - i_2)\zeta$ pairs of (k_1, k_2), specifically, $\big(1, 1 - (i_1 - i_2)\zeta\big), \big(2, 2 - (i_1 - i_2)\zeta\big), \cdots, \big(K + (i_1 - i_2)\zeta, K\big)$, for which $\mathbf{T}_{[k_1,k_2]} = \frac{N}{K}$, while the rest of the elements in $\mathbf{T}$ are all equal to 0. Hence we have

$$\overline{\boldsymbol{\Xi}}_0^{\mathbf{G}} = \frac{\big(K + (i_1 - i_2)\zeta\big)\beta_{l_2,l_1}^{u_2}}{K},$$
$$(i_1 - i_2)\zeta \in \{-K + 1, \cdots, -1\}. \tag{61}$$

ii.b) For the pair of users $i_1 = (u_2 - 1)L + l_2$ and $i_2 = (u_1 - 1)L + l_1$ of group $\bar{r}$ and $i_1 - i_2 < 0$, where $1 \leq \bar{r} \leq r$, if $(i_1 - i_2)\zeta \in \{-N, \cdots, -N + K - 1\}$, the PS matrices assigned to these two users are non-orthogonal. Thus, for $(i_1 - i_2)\zeta = t - N = k_1 - k_2 - N \in \{-N, \cdots, -N + K - 1\}$, $\mathbf{T}_{[k_1,k_2]} = \frac{N}{K}$. That is, there are $K - \big((i_1 - i_2)\zeta + N\big)$ pairs of (k_1, k_2), specifically, $\big(1 + ((i_1 - i_2)\zeta + N), 1\big), \big(2 + ((i_1 - i_2)\zeta + N), 2\big), \cdots, \big(K, K - ((i_1 - i_2)\zeta + N)\big)$, for which $\mathbf{T}_{[k_1,k_2]} = \frac{N}{K}$, while the rest of elements in $\mathbf{T}$ are all equal to 0. Hence we have

$$\overline{\boldsymbol{\Xi}}_0^{\mathbf{G}} = \frac{\big(K - ((i_1 - i_2)\zeta + N)\big)\beta_{l_2,l_1}^{u_2}}{K},$$
$$(i_1 - i_2)\zeta \in \{-N, \cdots, -N + K - 1\}. \tag{62}$$

Combining (57), (60), (61) and (62) as well as $\overline{\boldsymbol{\Xi}}_0^{\mathbf{W}} = \frac{K\sigma_w^2}{Np_r}$ leads to the MSE of (44).

References

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

[2] D. Wang, Y. Zhang, H. Wei, X. You, X. Gao, and J. Wang, "An overview of transmission theory and techniques of large-scale antenna systems for 5G wireless communications," *Sci. China Inf. Sci.*, vol. 59, no. 8, pp. 1–18, Aug. 2016.

[3] L. Zhao, K. Li, K. Zheng, and M. O. Ahmad, "An analysis of the tradeoff between the energy and spectrum efficiencies in an uplink massive MIMO-OFDM system," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 62, no. 3, pp. 291–295, Mar. 2015.

[4] E. Björnson, L. Sanguinetti, J. Hoydis, and M. Debbah, "Optimal design of energy-efficient multi-user MIMO systems: Is massive MIMO the answer?" *IEEE Trans. Wireless Commun.*, vol. 14, no. 6, pp. 3059–3075, Jun. 2015.

[5] T. L. Marzetta, "Massive MIMO: An introduction," *Bell Labs Tech. J.*, vol. 20, pp. 11–22, Mar. 2015.

[6] 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.

[7] F. Rusek *et al.*, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, Jan. 2013.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] H. Q. Ngo and E. G. Larsson, "EVD-based channel estimation in multicell multiuser MIMO systems with very large antenna arrays," in *Proc. ICASSP*, Kyoto, Japan, Mar. 2012, pp. 3249–3252.

[13] R. R. Müller, L. Cottatellucci, and M. Vehkaperä, "Blind pilot decontamination," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 773–786, Oct. 2014.

[14] Z. Wang, C. Qian, L. Dai, J. Chen, C. Sun, and S. Chen, "Location-based channel estimation and pilot assignment for massive MIMO systems," in *Proc. ICC Workshop*, London, U.K., Jun. 2015, pp. 1264–1268.

[15] J. Zhang, B. Zhang, S. Chen, X. Mu, M. El-Hajjar, and L. Hanzo, "Pilot contamination elimination for large-scale multiple-antenna aided OFDM systems," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 759–772, Oct. 2014.

[16] T. X. Vu, T. A. Vu, and T. Q. S. Quek, "Successive pilot contamination elimination in multiantenna multicell networks," *IEEE Wireless Commun. Lett.*, vol. 3, no. 6, pp. 617–620, Dec. 2014.

[17] A. Ashikhmin and T. Marzetta, "Pilot contamination precoding in multi-cell large scale antenna systems," in *Proc. ISIT*, Cambridge, MA, USA, Jul. 2012, pp. 1137–1141.

[18] T. S. Rappaport, *Wireless Communications: Principles and Practice*. Upper Saddle River, NJ, USA: Prentice-Hall, 1996.

[19] Y. Li, "Simplified channel estimation for OFDM systems with multiple transmit antennas," *IEEE Trans. Wireless Commun.*, vol. 1, no. 1, pp. 67–75, Jan. 2002.

[20] S. M. Kay, *Fundamentals of Statistical Signal Processing: Detection Theory*, vol. 2. Upper Saddle River, NJ, USA: Prentice-Hall, 1998.

Xinying Guo received the B.Sc. degree in electronic and information engineering from the North University of China, Taiyuan, China, in 2011. She is currently pursuing the Ph.D. degree with the School of Information Engineering, Zhengzhou University. Since 2014, she has been a Visiting Ph.D. Student with Prof. L. Hanzo and Prof. S. Chen with the School of Electronics and Computer Science, University of Southampton, U.K. Her research interests are in the areas of wireless communications and signal processing, including channel estimation, multiuser detection, beamforming/precoding, and pilot contamination elimination.

Sheng Chen (M'90–SM'97–F'08) received the B.Eng. degree in control engineering from the East China Petroleum Institute, Dongying, China, in 1982, the Ph.D. degree in control engineering from City University, London, U.K., in 1986, and the D.Sc. degree from the University of Southampton, Southampton, U.K., in 2005. He held research and academic appointments with the University of Sheffield, the University of Edinburgh, and the University of Portsmouth, U.K., from 1986 to 1999. Since 1999, he has been with the Electronics and Computer Science Department, University of Southampton, where he is currently a Professor of intelligent systems and signal processing. His research interests include adaptive signal processing, wireless communications, modeling and identification of nonlinear systems, neural network and machine learning, intelligent control system design, evolutionary computation methods, and optimization. He has authored over 550 research papers. He is a fellow of the United Kingdom Royal Academy of Engineering, a fellow of IET, and a Distinguished Adjunct Professor with King Abdulaziz University, Jeddah, Saudi Arabia. He was an ISI Highly Cited Researcher in engineering in 2004.

Jiankang Zhang (S'08–M'12) received the B.Sc. degree in mathematics and applied mathematics from the Beijing University of Posts and Telecommunications in 2006, and the Ph.D. degree in communication and information systems from Zhengzhou University in 2012. He has been a Lecturer with the School of Information Engineering, Zhengzhou University. From 2009 to 2011 and from 2013 to 2013, he was a Visiting Researcher in electronics and computer science with the University of Southampton, U.K. His research interests are in the areas of wireless communications and signal processing, including channel estimation, multiuser detection, beamforming/precoding, and optimization algorithms.

Xiaomin Mu received the 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 authored 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.

Lajos Hanzo (F'04) received the D.Sc. degree in electronics, the Ph.D. degree and the Doctor Honoris Causa degree from the Technical University of Budapest, in 1976, 1983, and 2009, respectively. During his 38-year career in telecommunications, he has held various research and academic positions in Hungary, Germany, and the U.K. Since 1986, he has been with the School of Electronics and Computer Science, University of Southampton, U.K., as the Chair in telecommunications. He has successfully supervised 100 Ph.D. students, co-authored 20 John Wiley/IEEE Press books in mobile radio communications totaling in excess of 10 000 pages, authored over 1500 research entries on IEEE Xplore, acted as the TPC Chair and the General Chair of IEEE conferences, presented keynote lectures, and received a number of distinctions. He is directing a 100 strong academic research team, where he is involved in a range of research projects in the field of wireless multimedia communications sponsored by the industry, the Engineering and Physical Sciences Research Council, U.K., the European Research Councils Advanced Fellow Grant, and the Royal Society's, Wolfson Research Merit Award. He is an enthusiastic supporter of industrial and academic liaison and offers a range of industrial courses. He is also a fellow of the Royal Academy of Engineering, the Institution of Engineering and Technology, and the European Association for Signal Processing. He is also a Governor of the IEEE VTS. From 2008 to 2012, he was the Editor-in-Chief of the IEEE PRESS and a Chaired Professor with Tsinghua University, Beijing. He has over 24 000 citations.