IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received February 1, 2016, accepted February 24, 2016, date of publication March 17, 2016, date of current version April 6, 2016.

Digital Object Identifier 10.1109/ACCESS.2016.2543300

Semi-Blind Channel Estimation Relying on Optimum Pilots Designed for Multi-Cell Large-Scale MIMO Systems

TIEJUN LV[1], (Senior Member, IEEE), SHAOSHI YANG[2], (Member, IEEE), AND HUI GAO[1], (Member, IEEE)

[1]School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing 100876, China
[2]School of Electronics and Computer Science, University of Southampton, Southampton SO17 1BJ, U.K.

Corresponding author: S. Yang (sy7g09@ecs.soton.ac.uk)

This work was supported by the National Natural Science Foundation of China under Grant 61271188 and Grant 61401041.

ABSTRACT Channel estimation in the presence of pilot contamination-induced inter-cell interference (ICI) is a major challenge in large-scale multiple-input multiple-output (LS-MIMO) systems. In this paper, a subspace-based semi-blind channel estimator (SBCE) relying on an optimum pilot design (in the sense of minimizing the channel estimation error covariance) is proposed for LS-MIMO systems. The proposed SBCE is capable of exploiting the asymptotic orthogonality of the channel vectors encountered in LS-MIMO systems, while taking advantage of both the optimized pilots and the data symbols. In order to ensure the best-possible performance of the proposed SBCE, we analyze the properties to be satisfied by the optimal pilots and then design these pilots relying on Zadoff–Chu sequences. As a beneficial result, the intra-cell interference is completely eliminated, and the ICI is substantially reduced. Our analytical and numerical results confirm that the performance of the proposed SBCE is superior to that of the representative state-of-the-art channel estimators in practical LS-MIMO systems, which have a finite number of base station antennas and data symbols available to be capitalized on for channel estimation.

INDEX TERMS Large-scale/massive multiple-input multiple-output (LS-MIMO/massive MIMO), semi-blind, channel estimation, pilot design, inter-cell interference (ICI).

I. INTRODUCTION

Multiple-input multiple-output (MIMO) systems offer a high flexibility for trading off between the multiplexing gain and the diversity gain [1]. Recently, multi-cell multiuser large-scale MIMO (LS-MIMO) systems have stimulated substantial research interests [2]–[14], because they hold high promise in achieving significant multiplexing gain, diversity gain and array gain by employing large antenna arrays. In LS-MIMO systems, the base station (BS) may be equipped with hundreds of antennas and simultaneously serves dozens of users in the same frequency band [2]–[8]. The channel vectors of LS-MIMO systems encountered under rich scattering conditions are approximately orthogonal, thus even low-complexity linear transmitters and receivers become near-optimal [2]–[4]. Moreover, upon assuming that the number of BS antennas tends to infinity, the effects of both the fast fading and the noise gradually become negligible [2], [3].

It is well known that in coherent MIMO systems the channel state information (CSI) is indispensable for high-reliability signal transmission and reception. Therefore, high-accuracy channel estimation [15]–[22] is vitally important for achieving their best possible performance. However, the total number of channels to be estimated in LS-MIMO systems may become excessive. Therefore, channel estimators capable of achieving high performance at the expense of low computational complexity are desirable in LS-MIMO systems. Additionally, in contrast to the situation of traditional small-scale multiuser MIMO systems, the pilot contamination effects impose a fundamental performance limitation on LS-MIMO systems [2], [3], [23]–[28], [31], [32], unless careful user scheduling [29] capable of avoiding the interference or the de-contamination precoding [30] is employed. Pilot contamination is essentially the particular inter-cell interference (ICI) encountered during the channel estimation, and it is caused by the reuse of a limited number of finite-length pilot sequences within a confined distance. Since the effect of both the small-scale fading and the noise is substantially reduced by large antenna arrays [2], [3] and the number of active users simultaneously served by each BS is increased, pilot contamination becomes the dominant performance limiting factor in LS-MIMO systems.

Recently, sophisticated methods have been proposed for addressing the channel estimation problem of LS-MIMO systems. More specifically, the authors of [33] proposed a semi-blind channel estimator (SBCE) based on the eigenvalue decomposition (EVD) method. In [33], it was assumed that the channel vectors are perfectly orthogonal, hence each channel vector can be uniquely characterized by an eigenvector having at most a multiplicative scalar ambiguity, which can be resolved with the aid of a few pilot symbols. However, in practice the channel vectors are only approximately orthogonal, hence they impose intra-cell interference and/or ICI. In [29], the authors demonstrated that when coordination between cells is possible, having statistical information about the channels is sufficient for avoiding the ICI. Although in general coordination does not necessarily require any form of information exchange amongst the cells [32], [34], the specific coordination strategy of [29] imposes an increased backhaul traffic amongst cells. In [35], an SBCE based on the singular value decomposition (SVD) was proposed, and the analysis showed that the ICI can be completely eliminated with the aid of an infinite number of BS antennas and an infinite number of data symbols, but naturally a residual ICI is encountered in realistic practical systems. Furthermore, in order to reduce the computational complexity, a compressive sensing based channel estimation approach was proposed in [36], while an L-order matrix polynomial expansion based *approximate* minimum mean-square error (MMSE) channel estimator was proposed in [37]. In [31], the transceiver's hardware impairments were incorporated into the channel model and an MMSE channel estimator was proposed. However, the purely pilot-based channel estimators of [31], [36], and [37] are even less competitive in terms of reducing the ICI than the SBCEs. Therefore, SBCEs capable of substantially mitigating the impact of ICI in practical LS-MIMO systems need further investigations.

Channel estimation designed for traditional systems has been extensively investigated. Among the existing methods, the subspace-based approach of [38] and [39] exploits the second-order statistics of the signals, and it represents one of the most popular blind channel estimation techniques owing to its attractive performance-versus-complexity tradeoff [40]. Unfortunately, as far as estimating MIMO channels is concerned, the subspace-based blind approach is subject to the matrix ambiguity problem detailed in [41]. Given that pilot contamination directly degrades the channel estimation performance, the family of SBCE methods that can achieve competitive performance by exploiting both the pilots and data symbols at a moderate extra complexity becomes promising [42]–[44]. Indeed, the classic subspace-based SBCE approach has recently been extended to traditional small-scale MIMO systems [44], [45]. However, the matrix inverse invoked in the channel estimator of [44] degrades its estimation performance in LS-MIMO systems. This is because the matrix to be inverted comprises a useful part and an interference-related part, and the matrix inverse operation may enhance the part associated with interference. Additionally, the column space of the channel matrix was estimated by solving the null space of the channel matrix, which imposes a high computational complexity [45]. Therefore, it is of high interest to develop subspace-based low-complexity SBCEs specifically for LS-MIMO systems.

On the other hand, when invoking the SBCE philosophy, it is possible for the receiver to achieve further performance improvements with the aid of carefully designed pilots [46]. It was shown in [47] that the pilot design which is optimum in the sense of minimizing the channel estimation error covariance should satisfy the Welch bound [47]. Since Zadoff-Chu sequences [48], [49] constitute the largest known Welch bound equality (WBE) signal sets [50], they have been recently recommended for pilot design in [51]–[53]. However, these pilots were designed for traditional small-scale MIMO systems and their achievable performance in LS-MIMO systems is far away from the Welch bound. The reasons for this phenomenon are twofold. Firstly, the number of pilot sequences required by traditional small-scale multiuser MIMO systems is relatively small, hence for a given sequence length, it is easy to construct a small number of orthogonal pilot sequences. By contrast, the number of orthogonal pilot sequences required by LS-MIMO systems is much larger, which is often impossible for a given sequence length, unless the sequence length is infinitely large. However, long pilot sequences may impose an excessive transmission overhead. Secondly, in contrast to the case of LS-MIMO systems, the impact of ICI as a result of pilot contamination is much lower than that of the noise in traditional small-scale multiuser MIMO systems, where the pilot contamination problem may be negligible. Therefore, traditional systems only focused on the orthogonality of the pilot sequences, and did not consider whether the pilot sequences are capable of satisfying the Welch bound or not.

Against the above background, in this paper, a low-complexity subspace-based SBCE and the specifically designed optimum pilots capable of minimizing the channel estimation error covariance are proposed for LS-MIMO systems. For the sake of reducing the computational complexity of the proposed SBCE, the eigenvectors of the sample covariance matrix of the received data symbols are relied upon, although they only approximately determine the channel matrix. As a result, there is an ambiguity between the eigenvector-based approximate and the true channel matrix. The proposed pilot symbols are then invoked for resolving the ambiguity, which completes the semi-blind estimation of the channel matrix. The main contributions of this paper are summarized as follows:

1) A subspace-based SBCE approach that exploits the asymptotic orthogonality of the channel vectors is proposed for LS-MIMO systems. Compared to that of [44], the column space of the channel matrix is directly estimated with the aid of EVD, hence the otherwise

necessary matrix inversion is avoided. Therefore, the proposed SBCE performs better than the channel estimator of [44]. Moreover, when maintaining the same channel estimation performance. the proposed SBCE imposes a lower computational complexity than that of [45]. Additionally, the number of BS antennas and data symbols is assumed to be finite in this paper, thus our analytical results are applicable to a more practical scenario in contrast to those of [35].

2) The optimal pilots are specifically designed based on Zadoff-Chu sequences for the proposed SBCE approach. The orthogonality of the circularly shifted Zadoff-Chu sequences is exploited for completely eliminating the intra-cell interference and for mitigating the ICI imposed on the channel estimate. The circularly shifted versions of the proposed pilot sequences constitute WBE sequences and they are arranged to be orthogonal within each cell. Hence, the proposed pilots result in a lower ICI than the pilots of [51]–[53]. The proposed pilots are also optimal for purely pilot-based channel estimation in LS-MIMO systems. By contrast, the design of specific pilots was not investigated in [35], thus the ICI remained relatively high.

This paper is organized as follows. In Section II, the general description of the system is presented. The derivation and the analysis of the proposed SBCE are detailed in Section III. In Section IV, the design of the optimal pilots is presented. Our simulation results and discussions are provided in Section V. Finally, our conclusions are offered in Section VI.

Notations: Lower-case (upper-case) boldface symbols denote vectors (matrices); $\mathbf{I}_K$ is the K-dimensional identity matrix, and $\mathbf{0}_{M\times K}$ represents an $(M\times K)$-element zero matrix; $(\cdot)^*$, $(\cdot)^T$, $(\cdot)^H$ and $\mathbb{E}\{\cdot\}$ denote the conjugate, the transpose, the conjugate transpose and the expectation, respectively; $[\cdot]_i$, $[\cdot]_{i,j}$, $\text{Re}(\cdot)$, $\text{tr}(\cdot)$ and $||\cdot||_F$ are the ith column, the (i,j)th entry, the real part, the trace and the Frobenius norm of a matrix, respectively; $[\mathbf{A};\mathbf{B}]$ and $[\mathbf{A},\mathbf{B}]$ represent the vertical and horizontal concatenation of the matrices $\mathbf{A}$ and $\mathbf{B}$, respectively; $\odot$ is the Hadamard product operator; $||\cdot||$ is the Euclidean norm of a vector; x mod N is the modulo-N operation for x; $\langle\cdot\rangle_n$ is the left circular shift of a vector with shift length n; $\delta(\cdot)$ is the Kronecker delta function; $\mathcal{CN}(\mu,\sigma^2)$ is a complex Gaussian random variable with mean μ and variance σ^2; and $\xrightarrow{\text{a.s.}}$ denotes the "*almost sure*" convergence.[1]

II. SYSTEM DESCRIPTION

We consider a system of L hexagonal cells, each of which comprises a BS and K user terminals (UTs), as shown in Fig. 1. Each BS is equipped with M antennas and each UT with a single antenna. Time-division duplexing (TDD) is employed in the system, thus the uplink (UL) channels are estimated using the pilots and data signals received at the BS as far as SBCE is concerned, while the downlink (DL) channels are assumed to be identical to the UL by exploiting the TDD link's reciprocity. We assume that the L cells rely on a frequency reuse factor of one. Additionally, all the UTs are assumed to be perfectly synchronized for transmitting pilots and data over frequency-flat block-fading channels in the same frequency band. The assumption of synchronized transmissions represents the worst-case scenario for channel estimation in multi-cell LS-MIMO systems [2], since it maximizes the pilot contamination. By contrast, the UL channel estimation problem would become less challenging if we assumed unsynchronized transmissions, since in this case only a fraction of the pilot-transmissions coincide and hence the impact of pilot contamination is mitigated. Nonetheless, it is worth noting that our SBCE and optimum pilots are also applicable to the less-challenging unsynchronized transmissions.[2]

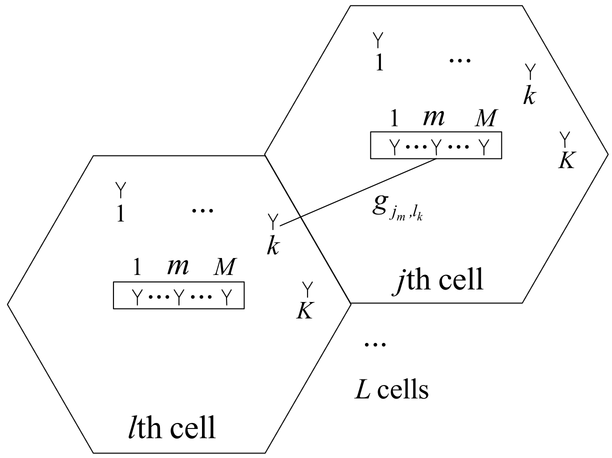

FIGURE 1. A system of L cells, each of which consists of one BS and K UTs, and adjoins the neighbouring cells. Each BS is equipped with M antennas, and each UT has one antenna. The channel from the kth UT in the lth cell to the mth antenna of the BS in the jth cell is denoted as g_{j_m,l_k}.

The channel model of the system is detailed below. We denote the UL channel matrix from the UTs in the lth cell to the BS in the jth cell as $\mathbf{G}_{jl}\in\mathbb{C}^{M\times K}$, which is expressed as

$$\mathbf{G}_{jl} = \mathbf{H}_{jl}\mathbf{D}_{jl}, \tag{1}$$

where $\mathbf{H}_{jl}\in\mathbb{C}^{M\times K}$ is composed of small-scale fading coefficients, while $\mathbf{D}_{jl}\in\mathbb{R}^{K\times K}$ is a diagonal matrix that is composed of large-scale fading coefficients, which account for both pathloss and shadowing. More specifically, the channel from the kth UT in the lth cell to the mth antenna of the BS in the jth cell is defined as

$$g_{j_m,l_k} \triangleq [\mathbf{G}_{jl}]_{m,k} = h_{j_m,l_k}\beta_{j,l_k}^{1/2}, \tag{2}$$

where $h_{j_m,l_k} = [\mathbf{H}_{jl}]_{m,k}$ is the small-scale fading coefficient from the kth UT in the lth cell to the mth antenna of the BS in

[1] In probability theory, an event happens "almost surely" implies that it happens with a probability of one.

[2] In other words, for the sake of evaluating the performance of the proposed pilots and channel estimator in the worst-case scenario, a perfectly synchronized transmission mode is assumed for all the UTs in the L cells.

the jth cell. These small-scale fading coefficients are assumed to be independent and identically distributed (i.i.d.) $\mathcal{CN}(0, 1)$ variables, namely we consider Rayleigh fading. The uncorrelated channels correspond to scenarios having rich scattering and sparsely located antennas. Some contributions consider correlated channels in LS-MIMO systems [24], [29]–[32], which correspond to scenarios with low scattering or closely packed antennas. Since the antennas at the BS may be spaced sufficiently far from each other, e.g. at least a half-wavelength spacing, in this paper we only consider uncorrelated channels, and the extension of the proposed SBCE to the scenario of correlated channels may be carried out in our future work. Additionally, $\beta_{j,l_k}^{1/2} = [\mathbf{D}_{jl}]_{k,k}$ is the large-scale fading coefficient between the kth UT in the lth cell and the BS in the jth cell,[3] and it is assumed to be known *a priori* at the BS [2], [3]. This assumption is reasonable, since for a given link the large-scale fading changes very slowly and may be estimated reliably. In practical cellular systems, an effective power control mechanism may be employed for compensating the large-scale fading. However, in this paper we explicitly consider the effects of both the small-scale and large-scale fading, which represents a more generalized treatment. To elaborate a little further, the power-control based scenario may be regarded as a special case of our treatment, where the impact of the large-scale fading is eliminated at the BS and only the small-scale fading effects have to be considered. Moreover, explicitly considering the large-scale fading enables us to investigate the cell-edge performance.

Let us now describe the transmission model of the system considered. Since TDD is used and the UL channel estimation is employed at the BS, only the UL transmissions are considered. The received signal matrix of the BS in the jth cell is given by

$$\mathbf{Y}_j = \left[\mathbf{Y}_j^{\mathrm{p}}, \mathbf{Y}_j^{\mathrm{d}}\right] = \sum_{l=1}^{L} \mathbf{G}_{jl}\mathbf{S}_l^T + \mathbf{N}_j \in \mathbb{C}^{M\times N}, \tag{3}$$

where $\mathbf{Y}_j^{\mathrm{p}} \in \mathbb{C}^{M\times N_{\mathrm{p}}}$ and $\mathbf{Y}_j^{\mathrm{d}} \in \mathbb{C}^{M\times N_{\mathrm{d}}}$ are the received pilot and data matrices, respectively; N_{p} and N_{d} represent the number of pilots and data symbols, respectively, and $N = N_{\mathrm{p}} + N_{\mathrm{d}}$ is the number of symbols transmitted per UT; $\mathbf{S}_l = \sqrt{p_{\mathrm{u}}}\left[\sqrt{N_{\mathrm{p}}}\boldsymbol{\Phi}_l; \mathbf{A}_l\right] \in \mathbb{C}^{N\times K}$ is the transmitted symbol matrix of the lth cell, where p_{u} is the average transmitted power per symbol, $\sqrt{p_{\mathrm{u}}N_{\mathrm{p}}}\boldsymbol{\Phi}_l \in \mathbb{C}^{N_{\mathrm{p}}\times K}$ and $\sqrt{p_{\mathrm{u}}}\mathbf{A}_l \in \mathbb{C}^{N_{\mathrm{d}}\times K}$ are the transmitted pilot and data matrices, respectively, and the factor $\sqrt{N_{\mathrm{p}}}$ scales the average transmitted power of each pilot symbol to be the same as that of each data symbol; $\mathbf{N}_j = \left[\mathbf{N}_j^{\mathrm{p}}, \mathbf{N}_j^{\mathrm{d}}\right] \in \mathbb{C}^{M\times N}$ is the noise matrix at the receiver, where $\mathbf{N}_j^{\mathrm{p}} \in \mathbb{C}^{M\times N_{\mathrm{p}}}$ and $\mathbf{N}_j^{\mathrm{d}} \in \mathbb{C}^{M\times N_{\mathrm{d}}}$ are matrices composed of i.i.d. $\mathcal{CN}(0, 1)$ entries.

[3]Due to the assumption of co-located BS antennas, the large-scale fading coefficients of the channels between the kth UT and the antennas at BS j are the same, thus the index m is omitted.

Again, we emphasize that the goal of channel estimation is to estimate the channel matrix $\mathbf{H}_{jj}$ with the aid of the received signal matrix $\mathbf{Y}_j$, and $\mathbf{H}_{jj}$ is assumed to be constant during the transmission of the block of N symbols. For the sake of convenient exposition, we rewrite the received signal matrix $\mathbf{Y}_j$ in the jth cell as follows:

$$\mathbf{Y}_j^{\mathrm{p}} = \sqrt{p_{\mathrm{u}}N_{\mathrm{p}}}\sum_{l=1}^{L} \mathbf{G}_{jl}\boldsymbol{\Phi}_l^T + \mathbf{N}_j^{\mathrm{p}}, \tag{4}$$

$$\mathbf{Y}_j^{\mathrm{d}} = \sqrt{p_{\mathrm{u}}}\sum_{l=1}^{L} \mathbf{G}_{jl}\mathbf{A}_l^T + \mathbf{N}_j^{\mathrm{d}}. \tag{5}$$

It is assumed that the columns of the pilot matrix $\boldsymbol{\Phi}_l$ are orthogonal for the sake of completely eliminating the intra-cell interference, thus we have $\boldsymbol{\Psi}_{ll} = \mathbf{I}_K, \forall l = 1, 2, \cdots, L$, where $\boldsymbol{\Psi}_{jl} = \boldsymbol{\Phi}_l^T\boldsymbol{\Phi}_j^*$ is the correlation matrix of the pilots. Additionally, the following assumptions are made for the pilots and the data symbols. 1) The number of UTs per cell, K, should be no higher than the number of pilot symbols (also referred to as the length of pilots) N_{p}, yielding $K \le N_{\mathrm{p}}$. Thus the pilots can satisfy the assumption of $\boldsymbol{\Psi}_{ll} = \mathbf{I}_K, \forall l$. 2) The total number of active users in L adjacent cells should be higher than N_{p}, i.e. we have $KL > N_{\mathrm{p}}$, so that the pilot contamination problem may be characterized in the LS-MIMO system considered. This is because if we want to avoid ICI amongst all the KL active users of L adjacent cells, we have to maintain the orthogonality of the columns of all pilot matrices $\boldsymbol{\Phi}_l, l = 1, 2, \cdots, L$, which implies that $KL \le N_{\mathrm{p}}$ has to be satisfied. However, since the channel estimation accuracy is constrained by pilot contamination in LS-MIMO systems, $KL \le N_{\mathrm{p}}$ is not realizable. Additionally, we note that N_{p} is physically determined by the channel's coherence interval, thus N_{p} may vary as a function of the transmission environments. 3) The entries of the data matrix $\mathbf{A}_l$ are i.i.d. random variables with zero mean and unit variance, and they are uncorrelated with the noise. Under these assumptions, $p_{\mathrm{u}}\beta_{j,j_k}$ can be interpreted as the average of the signal-to-noise ratio (SNR) of each BS receiver antenna in the jth cell for the kth UT in the jth cell.

III. THE SBCE PROPOSED FOR LS-MIMO SYSTEMS

In this section, our SBCE algorithm proposed for LS-MIMO systems is presented. As mentioned in Section I, the matrix inverse involved in the channel estimator of [44] degrades the estimation performance, which we would like to circumvent in our design. As our further objective, the estimation of the column space of the channel matrix in [45] will be simplified. Finally, the channel estimator of [35] is only analyzed in the idealized scenario, where the number of BS antennas and the number of data symbols are infinite. We also want to eliminate this idealized simplifying assumption. In this section, firstly, the relevant theoretical properties of the channel matrix of LS-MIMO systems are investigated. Then, an SBCE is conceived and its properties are benchmarked against those of other SBCEs.

A. THEORETICAL PROPERTIES OF THE CHANNEL MATRIX OF LS-MIMO SYSTEMS

The SBCE takes advantage of both the pilots and the data symbols, hence it performs better than a totally blind channel estimator and may also outperform the purely pilot-based channel estimator. In particular, the subspace-based SBCE exploits the second-order statistics of the received signals and it is capable of striking an attractive performance-versus-complexity tradeoff. Below we will first calculate the covariance matrix of the received data symbols in $\mathbf{Y}_j^{\mathrm{d}}$. Then, the column space of the channel matrix $\mathbf{H}_{jj}$ may be obtained by employing the EVD of the covariance matrix. More specifically, from (5), the covariance matrix of $\left[\mathbf{Y}_j^{\mathrm{d}}\right]_n, \forall n = 1, 2, \cdots, N_{\mathrm{d}}$ can be written as

$$\begin{aligned}\mathbf{R}_{\mathbf{Y}_j^{\mathrm{d}}} &= \mathbb{E}\left\{\left[\mathbf{Y}_j^{\mathrm{d}}\right]_n \left(\left[\mathbf{Y}_j^{\mathrm{d}}\right]_n\right)^H\right\} \\ &= p_{\mathrm{u}} \sum_{l=1}^{L} \mathbf{H}_{jl}\mathbf{D}_{jl}^2\mathbf{H}_{jl}^H + \mathbf{I}_M \in \mathbb{C}^{M\times M}. \qquad (6)\end{aligned}$$

Here n is omitted in $\mathbf{R}_{\mathbf{Y}_j^{\mathrm{d}}}$ based on the assumption concerning $\mathbf{A}_l$ in Section II.

The covariance matrix of the received data symbols has to be determined for the SBCE, and it is usually approximated by the sample covariance matrix. Like in other SBCEs, the estimate of the covariance matrix $\mathbf{R}_{\mathbf{Y}_j^{\mathrm{d}}}$ in (6) can be written as

$$\hat{\mathbf{R}}_{\mathbf{Y}_j^{\mathrm{d}}} = \frac{1}{N_{\mathrm{d}}} \sum_{n=1}^{N_{\mathrm{d}}} \left[\mathbf{Y}_j^{\mathrm{d}}\right]_n \left(\left[\mathbf{Y}_j^{\mathrm{d}}\right]_n\right)^H. \qquad (7)$$

As N_{d} tends to infinity, the sample covariance matrix $\hat{\mathbf{R}}_{\mathbf{Y}_j^{\mathrm{d}}}$ in (7) converges to the true covariance matrix $\mathbf{R}_{\mathbf{Y}_j^{\mathrm{d}}}$. We will provide a range of relevant insights with the aid of asymptotic analysis. However, the main focus of this paper is on the analysis of practical systems, where both N_d and M have finite values.

Since $\hat{\mathbf{R}}_{\mathbf{Y}_j^{\mathrm{d}}}$ is a Hermitian matrix, its EVD is given by

$$\hat{\mathbf{R}}_{\mathbf{Y}_j^{\mathrm{d}}} = \mathbf{U}_j \boldsymbol{\Sigma}_j \mathbf{U}_j^H, \qquad (8)$$

where the column vectors of $\mathbf{U}_j \in \mathbb{C}^{M\times M}$ are the eigenvectors of $\hat{\mathbf{R}}_{\mathbf{Y}_j^{\mathrm{d}}}$, $\boldsymbol{\Sigma}_j \in \mathbb{R}^{M\times M}$ is a diagonal matrix of the eigenvalues of $\hat{\mathbf{R}}_{\mathbf{Y}_j^{\mathrm{d}}}$, and the first K diagonal entries of $\boldsymbol{\Sigma}_j$ from the upper left corner are the K largest singular values of $\hat{\mathbf{R}}_{\mathbf{Y}_j^{\mathrm{d}}}$. Let $\mathbf{U}_j$ be partitioned as $\mathbf{U}_j = [\mathbf{U}_j^{\mathrm{s}}, \mathbf{U}_j^{\mathrm{n}}]$, where we have $\mathbf{U}_j^{\mathrm{s}} \in \mathbb{C}^{M\times K}$ and $\mathbf{U}_j^{\mathrm{n}} \in \mathbb{C}^{M\times (M-K)}$. Then we can formulate the following proposition, which reveals the relation between $\mathbf{U}_j^{\mathrm{s}}$ and the channel matrix $\mathbf{H}_{jj}$.

Proposition 1: For finite values of M and K, which satisfy $M \geq K$, $\mathbf{U}_j^{\mathrm{s}}$ can be *almost surely* expressed as

$$\mathbf{U}_j^{\mathrm{s}} = \frac{1}{\sqrt{M}} \left(\tilde{\mathbf{H}}_{jj} + \mathbf{F}_j\right) \mathbf{E}_{jj}, \qquad (9)$$

where $\mathbf{F}_j \in \mathbb{C}^{M\times K}$ is a matrix characterizing both the ICI and the intra-cell interference, while $\mathbf{E}_{jj} \in \mathbb{C}^{K\times K}$ is an unknown unitary matrix; $\tilde{\mathbf{H}}_{jj} \in \mathbb{C}^{M\times K}$ satisfies $(1/M)\tilde{\mathbf{H}}_{jj}^H\tilde{\mathbf{H}}_{jj} = \mathbf{I}_K$ (see Lemma 1 in Appendix A) and it is related to $\mathbf{H}_{jj}$ via $\mathbf{H}_{jj} = \tilde{\mathbf{H}}_{jj}\boldsymbol{\Xi}_{jj}$, where $\boldsymbol{\Xi}_{jj} \in \mathbb{C}^{K\times K}$ is formulated as

$$\boldsymbol{\Xi}_{jj} = \mathbf{I}_K + \boldsymbol{\Gamma}_{jj}, \qquad (10)$$

in which $\boldsymbol{\Gamma}_{jj} \in \mathbb{C}^{K\times K}$. Additionally, in the asymptotic scenario where $N_{\mathrm{d}} \to \infty, M \to \infty$ and K is fixed, we have $\mathbf{F}_j \xrightarrow{\text{a.s.}} \mathbf{0}_{M\times K}$.

Proof: See Appendix B. ∎

Remark 1: In our derivation, we consider the scenario where K is fixed and $M \to \infty$, which corresponds to the classic multiuser LS-MIMO scenario [54], [56], [57]. There is another scenario satisfying the multiuser LS-MIMO condition of $M \gg K$, where K scales with M at the same rate. Specifically, a constant ratio of $\frac{M}{K} = c$ is maintained, where c is typically larger than ten [24]. Asymptotic analyses for both scenarios may provide tight approximations for finite M and K [3]. It can be seen from Proposition 1 that $\mathbf{U}_j^{\mathrm{s}} = (1/\sqrt{M})\left(\mathbf{H}_{jj}\boldsymbol{\Xi}_{jj}^{-1} + \mathbf{F}_j\right)\mathbf{E}_{jj}$. We define $\tilde{\mathbf{E}}_{jj} = \boldsymbol{\Xi}_{jj}^{-1}\mathbf{E}_{jj} \in \mathbb{C}^{K\times K}$, which is termed as the ambiguity matrix. As long as $\tilde{\mathbf{E}}_{jj}$ is known, the channel matrix $\mathbf{H}_{jj}$ can be estimated from $\mathbf{U}_j^{\mathrm{s}}$.

Based on the above analysis, an SBCE is proposed in the next subsection.

B. THE PROPOSED SBCE

The pilots are invoked for generating a channel estimate, which can then be used for resolving the ambiguity matrix $\tilde{\mathbf{E}}_{jj}$ according to Proposition 1. By using (4), the received pilot symbols can be processed as follows:

$$\begin{aligned}\tilde{\mathbf{Y}}_j^{\mathrm{p}} &= \mathbf{Y}_j^{\mathrm{p}}\boldsymbol{\Phi}_j^* \\ &= \sqrt{p_{\mathrm{u}}N_{\mathrm{p}}}\left(\mathbf{G}_{jj} + \sum_{l\neq j}\mathbf{G}_{jl}\boldsymbol{\Psi}_{jl}\right) \\ &\quad + \mathbf{N}_j^{\mathrm{p}}\boldsymbol{\Phi}_j^* \in \mathbb{C}^{M\times K}. \qquad (11)\end{aligned}$$

It is assumed that the BSs are noncooperative and each BS only knows the pilots of its own cell, thus the covariance matrix $\sum_{l\neq j}\mathbf{G}_{jl}\boldsymbol{\Psi}_{jl}$ is unknown to the BS in the jth cell. Therefore, the channel matrix $\mathbf{H}_{jj}$ cannot be estimated by using the classic MMSE criterion. Instead, $\mathbf{H}_{jj}$ can be estimated by invoking $\tilde{\mathbf{Y}}_j^{\mathrm{p}}$ as

$$\begin{aligned}\hat{\mathbf{H}}_{jj}^{\mathrm{p}} &= \frac{1}{\sqrt{p_{\mathrm{u}}N_{\mathrm{p}}}}\tilde{\mathbf{Y}}_j^{\mathrm{p}}\mathbf{D}_{jj}^{-1} \qquad (12) \\ &= \mathbf{H}_{jj} + \boldsymbol{\Delta}_j \in \mathbb{C}^{M\times K}, \qquad (13)\end{aligned}$$

where

$$\boldsymbol{\Delta}_j = \left(\sum_{l\neq j}\mathbf{H}_{jl}\mathbf{D}_{jl}\boldsymbol{\Psi}_{jl} + \frac{1}{\sqrt{p_{\mathrm{u}}N_{\mathrm{p}}}}\mathbf{N}_j^{\mathrm{p}}\boldsymbol{\Phi}_j^*\right)\mathbf{D}_{jj}^{-1} \in \mathbb{C}^{M\times K} \qquad (14)$$

represents the ICI and noise contributions imposed on the received pilots.

For resolving the ambiguity matrix $\tilde{\mathbf{E}}_{jj}$ in the signal subspace $\mathbf{U}_j^{\mathrm{s}}$ and further improving the attainable estimation performance, based on the pilot-aided channel estimate $\hat{\mathbf{H}}_{jj}^{\mathrm{p}}$, we will develop a refined channel estimation approach. According to Proposition 1 and (13), we have

$$\frac{1}{\sqrt{M}}\left(\mathbf{U}_j^{\mathrm{s}}\right)^H \hat{\mathbf{H}}_{jj}^{\mathrm{p}} = \mathbf{E}_{jj}^H \boldsymbol{\Xi}_{jj} + \mathbf{E}_{jj}^H \boldsymbol{\Upsilon}_j^H = \tilde{\mathbf{E}}_{jj}^{-1} + \mathbf{E}_{jj}^H \boldsymbol{\Upsilon}_j^H, \tag{15}$$

where

$$\boldsymbol{\Upsilon}_j = \frac{1}{M}\boldsymbol{\Delta}_j^H\left(\tilde{\mathbf{H}}_{jj} + \mathbf{F}_j\right) + \frac{1}{M}\mathbf{H}_{jj}^H \mathbf{F}_j \in \mathbb{C}^{K\times K} \tag{16}$$

corresponds to the sum of the ICI, the intra-cell interference and the noise contaminating the received pilots and data. Note that $(1/\sqrt{M})\left(\mathbf{U}_j^{\mathrm{s}}\right)^H \hat{\mathbf{H}}_{jj}^{\mathrm{p}}$ in (15) contains the inverse of the ambiguity matrix $\tilde{\mathbf{E}}_{jj}$. Hence, (15) is used for resolving the ambiguity matrix in the signal subspace $\mathbf{U}_j^{\mathrm{s}}$ of the proposed SBCE. Based on Proposition 1 and (15), we have

$$\begin{aligned}\sqrt{M}\mathbf{U}_j^{\mathrm{s}}\left(\frac{1}{\sqrt{M}}\left(\mathbf{U}_j^{\mathrm{s}}\right)^H \hat{\mathbf{H}}_{jj}^{\mathrm{p}}\right) &= \mathbf{H}_{jj} + \frac{1}{M}\tilde{\mathbf{H}}_{jj}\tilde{\mathbf{H}}_{jj}^H \boldsymbol{\Delta}_j + \tilde{\mathbf{F}}_j \\ &\triangleq \hat{\mathbf{H}}_{jj}^{\mathrm{s}},\end{aligned} \tag{17}$$

where

$$\begin{aligned}\tilde{\mathbf{F}}_j = \frac{1}{M}&\left[\left(\tilde{\mathbf{H}}_{jj} + \mathbf{F}_j\right)\mathbf{F}_j^H\left(\mathbf{H}_{jj} + \boldsymbol{\Delta}_j\right) + \mathbf{F}_j\tilde{\mathbf{H}}_{jj}^H\boldsymbol{\Delta}_j\right] \\ &+ \mathbf{F}_j\boldsymbol{\Xi}_{jj} \in \mathbb{C}^{M\times K}\end{aligned} \tag{18}$$

is caused by the ICI and the intra-cell interference imposed on the received data symbols. Note that $\hat{\mathbf{H}}_{jj}^{\mathrm{s}}$ is the estimate of the channel matrix $\mathbf{H}_{jj}$ based on the proposed subspace-based SBCE approach. Upon comparing (9) to (17), it can be seen that the ambiguity matrix $\tilde{\mathbf{E}}_{jj}$ is completely resolved in the channel estimate $\hat{\mathbf{H}}_{jj}^{\mathrm{s}}$, and the residual estimation error is caused by the ICI, the intra-cell interference and the noise. It can be seen from (17) that

$$\hat{\mathbf{H}}_{jj}^{\mathrm{s}} = \mathbf{U}_j^{\mathrm{s}}\left(\mathbf{U}_j^{\mathrm{s}}\right)^H \hat{\mathbf{H}}_{jj}^{\mathrm{p}}, \tag{19}$$

which is a compact form of representing the proposed SBCE approach.

For the sake of explicit clarity, the proposed SBCE algorithm is summarized as follows.

For any estimator, the unbiased nature and the efficiency are the two salient metrics of characterizing its performance. In the next subsection, we will examine the proposed SBCE from these two aspects.

C. PERFORMANCE ANALYSIS

In this subsection, both the bias and the estimation error covariance of the proposed SBCE are analyzed. Additionally, the proposed SBCE is compared to other representative SBCEs.

Algorithm 1 Subspace-Based SBCE Approach for LS-MIMO Systems

Step 1) Compute the sample covariance matrix $\hat{\mathbf{R}}_{\mathbf{Y}_j^{\mathrm{d}}}$ *using (7).*
Step 2) Perform EVD of $\hat{\mathbf{R}}_{\mathbf{Y}_j^{\mathrm{d}}}$ *using (8).*
Step 3) Find $\mathbf{U}_j^{\mathrm{s}}$ *that corresponds to the largest* K *singular values of* $\hat{\mathbf{R}}_{\mathbf{Y}_j^{\mathrm{d}}}$*.*
Step 4) Compute the pilot-based channel estimate $\hat{\mathbf{H}}_{jj}^{\mathrm{p}}$ *using (12).*
Step 5) Obtain the subspace-based semi-blind channel estimate $\hat{\mathbf{H}}_{jj}^{\mathrm{s}}$ *using (19).*

1) THE BIAS OF THE PROPOSED SBCE

From (17), the bias matrix of $\hat{\mathbf{H}}_{jj}^{\mathrm{s}}$ is formulated as [57, p. 18]

$$\mathbb{B}\left\{\hat{\mathbf{H}}_{jj}^{\mathrm{s}}\right\} = \mathbb{E}\left\{\hat{\mathbf{H}}_{jj}^{\mathrm{s}}\right\} - \mathbf{H}_{jj} = \mathbb{E}\left\{\tilde{\mathbf{F}}_j\right\}. \tag{20}$$

It can be seen that the bias matrix is a non-zero matrix due to the ICI and the intra-cell interference imposed on the received data. According to Proposition 1, we have $\mathbf{F}_j \xrightarrow{\text{a.s.}} \mathbf{0}_{M\times K}$ as $N_{\mathrm{d}} \to \infty, M \to \infty$ and K is fixed. Hence, the proposed SBCE is asymptotically unbiased.

2) THE ESTIMATION ERROR COVARIANCE OF THE PROPOSED SBCE

The estimation error covariance is also a key performance indicator of the proposed SBCE. Let

$$\mathbf{R}_{\tilde{\mathbf{H}}_{jj}^{\mathrm{s}}} = \mathbb{E}\left\{\tilde{\mathbf{H}}_{jj}^{\mathrm{s}} \odot \left(\tilde{\mathbf{H}}_{jj}^{\mathrm{s}}\right)^*\right\} \in \mathbb{R}^{M\times K}, \tag{21}$$

where

$$\tilde{\mathbf{H}}_{jj}^{\mathrm{s}} = \hat{\mathbf{H}}_{jj}^{\mathrm{s}} - \mathbf{H}_{jj} \in \mathbb{C}^{M\times K} \tag{22}$$

is the estimation error matrix. Then $\left[\mathbf{R}_{\tilde{\mathbf{H}}_{jj}^{\mathrm{s}}}\right]_{m,n}$ is the estimation error covariance of $\left[\hat{\mathbf{H}}_{jj}^{\mathrm{s}}\right]_{m,n}$. It is shown in Appendix C that we have:

$$\begin{aligned}\left[\mathbf{R}_{\tilde{\mathbf{H}}_{jj}^{\mathrm{s}}}\right]_{m,n} &= \left[\mathbf{R}_{\tilde{\mathbf{F}}_j}\right]_{m,n} + \frac{\left\|\left[\tilde{\mathbf{H}}_{jj}^T\right]_m\right\|^2}{M\beta_{j,j_n}} \\ &\times\left(p_{\mathrm{u}}^{-1}N_{\mathrm{p}}^{-1} + \sum_{l\neq j}\sum_{k=1}^{K}\beta_{j,l_k}\left|\left[\boldsymbol{\Psi}_{jl}\right]_{k,n}\right|^2\right),\end{aligned} \tag{23}$$

where

$$\begin{aligned}\mathbf{R}_{\tilde{\mathbf{F}}_j} = \mathbb{E}&\left\{\tilde{\mathbf{F}}_j \odot \tilde{\mathbf{F}}_j^* + 2\,\mathrm{Re}\left(\left(\frac{1}{M}\tilde{\mathbf{H}}_{jj}\tilde{\mathbf{H}}_{jj}^H\boldsymbol{\Delta}_j\right)\odot\tilde{\mathbf{F}}_j^*\right)\right\} \\ &\in \mathbb{R}^{M\times K}\end{aligned} \tag{24}$$

is caused by the ICI and the intra-cell interference inflicted upon the received data symbols, $\boldsymbol{\Psi}_{jl}$ corresponds to the ICI in the pilots and the term $p_{\mathrm{u}}^{-1}N_{\mathrm{p}}^{-1}$ is caused by the noise.

Recall from Proposition 1 that we have $\mathbf{F}_j \xrightarrow{\text{a.s.}} \mathbf{0}_{M\times K}$ as $N_\mathrm{d} \to \infty, M \to \infty$ and K is fixed. Additionally, the term right behind the first plus sign in (23) is inversely proportional to M. Therefore, the estimation error covariance of the proposed SBCE tends *almost surely* to zero as $N_\mathrm{d} \to \infty$, $M \to \infty$ and K is fixed.

3) COMPARISON WITH OTHER REPRESENTATIVE CHANNEL ESTIMATORS

In [35], a meritorious SBCE method was proposed for LS-MIMO systems, which performs SVD of the received data matrix $\mathbf{Y}_j^\mathrm{d}$. The analysis in [35] shows that the ICI can be completely eliminated for infinite M, N_d and finite K values, which is also the case for our analysis. In contrast to [35], in this paper our goal is to analyze the estimation error for finite M and N_d values, which corresponds to practical LS-MIMO systems. Furthermore, the residual ICI will be mitigated with the aid of our optimum pilot design in the next section, which was not investigated in [35].

When the algorithm of [44] is modified for LS-MIMO systems, the semi-blind channel estimate may be formulated as:

$$\hat{\mathbf{H}}_{jj}^\mathrm{n} = M\mathbf{U}_j^\mathrm{s}\left(\hat{\mathbf{H}}_{jj}^\mathrm{p}\left(\mathbf{U}_j^\mathrm{s}\right)^H\right)^{-1}. \tag{25}$$

By substituting (9) and (13) into (25), $\hat{\mathbf{H}}_{jj}^\mathrm{n}$ is expressed as

$$\hat{\mathbf{H}}_{jj}^\mathrm{n} = \left(\mathbf{H}_{jj} + \mathbf{F}_j\boldsymbol{\Xi}_{jj}\right)\left(\boldsymbol{\Xi}_{jj} + \boldsymbol{\Gamma}_{jj}^H\boldsymbol{\Xi}_{jj} + \boldsymbol{\Upsilon}_j\boldsymbol{\Xi}_{jj}\right)^{-1}, \tag{26}$$

where $\boldsymbol{\Gamma}_{jj}$, cf. (10), corresponds to the correlation of the columns of $\mathbf{H}_{jj}$. It can be seen in (26) that the estimator modified from that of [44] becomes biased in LS-MIMO systems. Additionally, the bias is not only caused by the interference, but also by the correlation of channel vectors. By contrast, our analytical result given by (20) shows that the bias of the proposed SBCE is only caused by the interference. Therefore, we draw the conclusion that the matrix inverse operation in (25) partially loses some of the CSI inherently contained in the received signals.

The proposed SBCE has a similar computational complexity to that of the SBCEs of [35] and [44]. The computational complexity of the representative algorithms considered are compared in the asymptotic sense using the big O notation [63]. Specifically, the computational complexity of Step 1, Step 4 and Step 5 of the proposed SBCE scheme is $O\left(N_\mathrm{d}M^2\right)$, $O\left(MK^2\right)$ and $O\left(2KM^2\right)$, respectively. The complexity of the remaining steps in the proposed SBCE is $O\left(M^3\right)$. Compared to the proposed SBCE scheme, our pilot design has a negligible computational complexity, since it is an off-line design. As a result, the total computational complexity of the proposed joint channel estimation and pilot design scheme is approximately $O\left(M^3 + N_\mathrm{d}M^2 + 2KM^2 + MK^2\right)$. The computational complexity of the semi-blind approach of [44] can be calculated in a similar manner, which is $O\left(M^3 + KM^2 + 2KM\right)$. Moreover, the computational complexity of the blind channel estimation based approach of [35] is $O\left(M^3 + 2KM^2 + K^2M\right)$.

In Fig. 2, the base-10 logarithms of the computational complexities (in big O notation) versus the number of the BS antennas M for these three approaches are shown. We can see that the complexity of the proposed approach is higher than that of the benchmark approaches. Nevertheless, the proposed approach is capable of achieving better performance than the other approaches considered, as shown in Figs. (3)-(7). Therefore, the proposed approach strikes an attractive tradeoff between the achievable performance and the computational complexity.

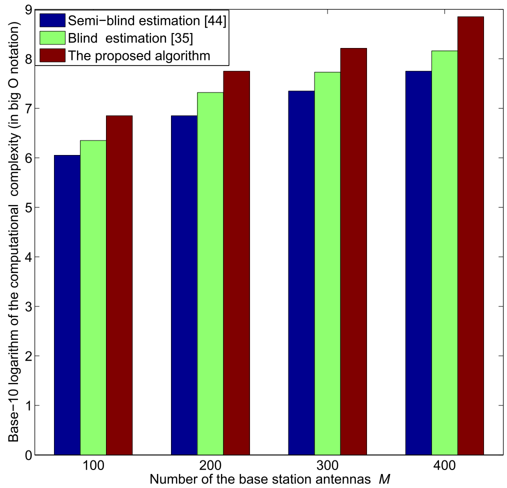

FIGURE 2. Comparison of the computational complexity of various estimation methods versus the number of the base station antennas M. Assume $N_\mathrm{d} = 300$ and $K = 11$.

Furthermore, our analysis provided in Section III.C is more general than that in [35] and the proposed SBCE outperforms the schemes advocated in [44]. Additionally, the channel estimator of [45] estimates the column space of the channel matrix with the aid of the null space of the channel matrix, which is not necessary in our system considered. The MMSE channel estimators of [31]–[36] exhibit much lower computational complexities than the proposed SBCE, but their performance is less attractive, since they are not capable of reducing the ICI.

It can be seen from (23) that $\boldsymbol{\Psi}_{jl}, l \neq j$, is the cross-correlation matrix of the pilots. Hence, using carefully designed pilots is critical for improving the performance of the proposed SBCE.

IV. PILOT DESIGN

In this section, an optimum pilot design is proposed. Although several pilot designs have been conceived for ICI cancellation [51]–[53], these pilots were not designed specifically for the proposed SBCE aided LS-MIMO systems. Hence, a pilot design criterion is proposed and Zadoff-Chu sequences based pilots are designed according to this criterion.

The properties of the pilots directly influence the estimation error covariance of our SBCE. It has been previously shown that the ICI, which corresponds to $\mathbf{R}_{\tilde{\mathbf{F}}_j}$ and $\boldsymbol{\Psi}_{jl}$ in (23), substantially degrades the performance of the proposed SBCE. Since $\tilde{\mathbf{F}}_j$, cf. (18), is caused by the ICI contaminating the received data symbols, $\mathbf{R}_{\tilde{\mathbf{F}}_j}$ in (23) is not related to the design of pilots. Then, it can be observed that the ICI caused by the pilots is related to both $\boldsymbol{\Psi}_{jl}$ and the large-scale fading coefficients $\beta_{j,l_k}, l = 1, 2, \cdots, L, k = 1, 2, \cdots, K$. Although these large-scale fading coefficients are assumed to be known *a priori* at the corresponding BS, they are typically different for different UTs and vary when the UTs move. In this paper, since we focus on optimizing the statistical performance rather than the instantaneous performance in designing the pilots, it is infeasible to exploit the instantaneous large-scale fading coefficients of all UTs. Assuming that each UT obeys the same statistical mobility model, their distances from the corresponding BSs are i.i.d random variables. Therefore, the mean of each UT's distance is the same and from a long-term perspective each UT encounters the same average large-scale fading. Additionally, since the mean of log-normal distributed shadowing is the path-loss, we may rely only on the average pathloss coefficient, which is the same for all UTs, when designing the statistically optimal pilots. Hence, the proposed pilot optimization criterion is

$$\min_{\boldsymbol{\Phi}_1, \boldsymbol{\Phi}_2, \cdots, \boldsymbol{\Phi}_L} \sum_{j=1}^{L} \sum_{l \neq j} \left|\left|\boldsymbol{\Psi}_{jl}\right|\right|_{\mathrm{F}}^2$$
$$\text{s.t. } \boldsymbol{\Psi}_{ll} = \mathbf{I}_K, \quad \forall l = 1, 2, \cdots, L, \forall N_{\mathrm{p}}, \ K \le N_{\mathrm{p}} < KL. \tag{27}$$

Note that the optimization variables in (27) are $\boldsymbol{\Phi}_l, l = 1, 2, \cdots, L$, which are related to $\boldsymbol{\Psi}_{jl}$ via $\boldsymbol{\Psi}_{jl} = \boldsymbol{\Phi}_l^T \boldsymbol{\Phi}_j^*$. Additionally, the constraint $\boldsymbol{\Psi}_{ll} = \mathbf{I}_K$ indicates that the cumulative transmit power of the N_p pilot symbols of each user is $P_{\mathrm{u}} N_{\mathrm{p}}$, cf. (4). It can be seen that the proposed pilot design criterion of (27) is derived based on the expression of our channel estimate tailored for the ICI scenario considered. We will demonstrate in Section V that this criterion is more effective than the existing pilot design criteria of [51]–[53] in the scenario considered.

Observe that the channel estimation error caused by pilot contamination is minimized when the summation in (27) is minimized, since the average pathloss coefficients are regarded to be identical. Because the autocorrelation of the pilots should satisfy $\boldsymbol{\Psi}_{ll} = \mathbf{I}_K, \forall l = 1, 2, \cdots, L$, the pilot sequences of all UTs $[\boldsymbol{\Phi}_l]_k, l = 1, 2, \cdots, L, k = 1, 2, \cdots, K$, should be unit vectors. Moreover, the assumption of $N_{\mathrm{p}} < KL$ must be satisfied, which implies that the number of entries in the vector $[\boldsymbol{\Phi}_l]_k, \forall k, l$ should be less than the total number of these vectors. Then, the summation in (27) is bounded by the Welch bound [47], i.e. we have

$$\sum_{j=1}^{L} \sum_{l \neq j} \left|\left|\boldsymbol{\Psi}_{jl}\right|\right|_{\mathrm{F}}^2 \ge \frac{(KL)^2}{N_{\mathrm{p}}} - KL. \tag{28}$$

Therefore, the objective of our pilot design is to construct sequences that satisfy both the Welch bound and the last two constraints of (27). Again, sequences that satisfy the Welch bound are known as WBE sequences. If $[\boldsymbol{\Phi}_l]_k, l = 1, 2, \cdots, L, k = 1, 2, \cdots, K$, are WBE sequences, they will satisfy

$$||\mathbf{f}||^2 = c \sum_{l=1}^{L} \left|\left|\mathbf{f}^H \boldsymbol{\Phi}_l\right|\right|^2, \quad \forall \mathbf{f} \in \mathbb{C}^{N_{\mathrm{p}}} \tag{29}$$

for some $c > 0$ [47]. Therefore, if $[\boldsymbol{\Phi}_l]_k, l = 1, 2, \cdots, L, k = 1, 2, \cdots, K$, are optimal pilot sequences, they should constitute a solution to the minimization problem of (27), which implies that they should also be WBE sequences, i.e. they should satisfy (29). It has been shown in [60] that (29) is satisfied if and only if the rows of

$$\boldsymbol{\Omega} = [\boldsymbol{\Phi}_1, \boldsymbol{\Phi}_2, \cdots, \boldsymbol{\Phi}_L] \in \mathbb{C}^{N_{\mathrm{p}} \times KL} \tag{30}$$

are orthogonal. More specifically, the pilot optimization criterion of (27) can be rewritten as

$$\min_{\boldsymbol{\Phi}_1, \boldsymbol{\Phi}_2, \cdots, \boldsymbol{\Phi}_L} \sum_{m=1}^{N_{\mathrm{p}}} \sum_{m \neq n} \left|([\boldsymbol{\Omega}^T]_m)^H [\boldsymbol{\Omega}^T]_n\right|^2,$$
$$\text{s.t. } \boldsymbol{\Psi}_{ll} = \mathbf{I}_K, \quad \forall l = 1, 2, \cdots, L, \forall N_{\mathrm{p}}, \ K \le N_{\mathrm{p}} < KL. \tag{31}$$

Among the existing sequences, the class of Zadoff-Chu sequences is the largest known set of WBE sequences [52]. Hence, Zadoff-Chu sequences are invoked for pilot design in this treatise. In traditional pilot design approaches [51]–[53], the Zadoff-Chu sequences were first designed for ensuring the orthogonality of the pilots of the UTs, and they were then refined according to other relevant criteria. For example, the pilots of [51] were first designed to be orthogonal for all of the UTs within the same cell, and then they were refined to achieve the minimum variation of the cross-correlation value $\left|\left[\boldsymbol{\Psi}_{jl}\right]_{k,n}\right|^2$ for the UTs in different cells. However, these pilots were not WBE sequences, hence the channel estimation error covariance in (23) is not minimized. By contrast, in this treatise, the Zadoff-Chu sequences are invoked for ensuring the orthogonality of the row vectors of $\boldsymbol{\Omega}$ shown in the criterion (31), and then they are refined for ensuring the orthogonality of the pilots of the UTs characterized by $\boldsymbol{\Psi}_{ll} = \mathbf{I}_K$ of (31). For the sake of ensuring the orthogonality of the rows of $\boldsymbol{\Omega}$, the nth entry of the Zadoff-Chu sequence having a length of KL may be written as [48]

$$b_n = \exp\left(i \frac{A\pi}{KL} n \left(n + (KL \bmod 2)\right)\right), \tag{32}$$

where A is an integer that is relatively prime to KL. Note that here n can be any natural number, and the vectorial form of a specific Zadoff-Chu sequence is expressed as

$$\mathbf{b} = [b_0, b_1, \cdots, b_{KL-1}]. \tag{33}$$

The autocorrelation properties of Zadoff-Chu sequences by definition ensure that the inner product of $\langle \mathbf{b} \rangle_{m'}$ and $\langle \mathbf{b} \rangle_{n'}$ is $\langle \mathbf{b} \rangle_{m'} \left(\langle \mathbf{b} \rangle_{n'}\right)^H = KL\delta(m' - n'), \ m', n' \in \{0, 1, \cdots, KL-1\}$.

Then, the transpose of the mth row of the pilot matrix $\mathbf{\Omega}$ can be expressed as

$$\left[\mathbf{\Omega}^T\right]_m = \frac{1}{\sqrt{N_\mathrm{p}}}\left(\langle\mathbf{b}\rangle_{s_m}\right)^T, \quad m = 1, 2, \cdots, N_\mathrm{p}, \tag{34}$$

where $s_m \in \{0, 1, \cdots, KL-1\}$ is the circular shift length for the mth row of $\mathbf{\Omega}$. When s_m satisfies $s_{m_1} \neq s_{m_2}, \forall m_1, m_2 \in \{1, 2, \cdots, N_\mathrm{p}\}, m_1 \neq m_2$, the rows of $\mathbf{\Omega}$ are orthogonal. It can be seen that the column vectors of $\mathbf{\Omega}$ are unit vectors, hence the resultant pilots are WBE sequences when the circular shift length s_m in (34) exhibits different values for different m. Then, the condition for the designed pilot sequences to satisfy $([\mathbf{\Omega}^T]_m)^H[\mathbf{\Omega}^T]_n = 0, \forall m \neq n$ and $\mathbf{\Psi}_{ll} = \mathbf{I}_K, \forall l = 1, 2, \cdots, L$ of (31) is given by the following proposition.

Proposition 2: When $N_\mathrm{p} = aK, a = 1, 2, \cdots, L-1, s_n = \lfloor(n-1)L/a\rfloor, n = 1, 2, \cdots, N_\mathrm{p}$, the columns of $\mathbf{\Omega}$ are WBE sequences, and $\mathbf{\Psi}_{ll} = \mathbf{I}_K, \forall l = 1, 2, \cdots, L$.

Proof: See Appendix D. ■

Remark 2: The pilots designed can only satisfy the criterion (31) when we have $N_\mathrm{p} = aK, a = 1, 2, \cdots, L-1$. More explicitly, when the number of pilot symbols N_p is an integer multiple of the number of simultaneously active UTs per cell, K, the resultant pilots satisfy the orthogonality requirement of $\mathbf{\Psi}_{ll} = \mathbf{I}_K, \forall l = 1, 2, \cdots, L$. Meanwhile, the requirement of being WBE sequences can also be satisfied, hence the sum of the cross-correlation values of the pilots, namely $\sum_{j=1}^{L}\sum_{l\neq j}\left|\left|\mathbf{\Psi}_{jl}\right|\right|_\mathrm{F}^2$ is minimized, and the estimation error covariance of $\hat{\mathbf{H}}_{jj}^\mathrm{s}$ relying on our SBCE is also minimized for a range of transmission scenarios. Hence, following the proposed design procedure is crucial for designing pilots that meet the criterion of (26). By contrast, the pilots obtained by relying on the schemes of [51]–[53] cannot satisfy this criterion. When the design criterion (31) is followed, the pilot contamination effect caused by the reuse of the pilots in different cells is also minimized. Therefore, the proposed SBCE's performance is improved.

V. SIMULATION RESULTS AND DISCUSSIONS

In this section, our simulation results are provided for validating our theoretical analysis. The system consists of $L = 7$ hexagonal cells that share the same frequency band. The central cell is surrounded by the other six cells, and the cell edges connecting the neighboring cells have the same length. The cell radius r (from center to edge) is 1000 meters, while the distance between any UT and its serving BS is no lower than $r_{\min} = 100$ meters and no higher than $r_{\max} = 800$ meters. The pathloss decay exponent is $\alpha = 3.8$. More explicitly, the large-scale fading coefficient from each UT to the BS can be expressed as $\beta_{j,l_k} = 1/(r_{j,l_k}/100)^\alpha$ when the power is measured in Watt,[4] where r_{j,l_k} is the distance between the kth UT in the lth cell to the BS in the jth cell. QPSK modulation is used for the transmission of data symbols, and the number of data symbols per transmit frame is $N_\mathrm{d} = 300$. The BS of each cell is equipped with $M = 100$ antennas, and serves $K = 11$ UTs simultaneously. In our simulations, the estimation error covariance of the proposed SBCE is defined as $(1/MKL)\sum_{j=1}^{L}\left|\left|\tilde{\mathbf{H}}_{jj}^\mathrm{s}\right|\right|_\mathrm{F}^2$, and the estimation error covariances of the benchmarking channel estimators are defined in a similar manner.

A. COMPARISON OF THE PROPOSED PILOTS TO OTHER ZADOFF-CHU SEQUENCES BASED PILOTS

The pilots proposed in [51] and [52] for the UTs of each cell are the circularly shifted versions of the sequence $\tilde{\mathbf{b}} = \left[\tilde{b}_0, \tilde{b}_1, \cdots, \tilde{b}_{N_\mathrm{p}-1}\right]$, where $\tilde{b}_n = \exp\left(i\tilde{A}\pi n\left(n + \left(N_\mathrm{p} \bmod 2\right)\right)/N_\mathrm{p}\right), n = 0, 1, \cdots, N_\mathrm{p}-1$, and the sequence parameter $\tilde{A}$ is set as follows.[5] 1) When $N_\mathrm{p} = 11, 22, 44$, we set $\tilde{A} = 1, 3, 5, 7, 9, 13, 15$. 2) For $N_\mathrm{p} = 33$, we assign $\tilde{A} = 1, 2, 4, 5, 7, 8, 10$. 3) When $N_\mathrm{p} = 55$, we have $\tilde{A} = 1, 2, 3, 4, 6, 7, 8$. 4) For $N_\mathrm{p} = 66$, $\tilde{A} = 1, 2, 5, 7, 13, 17, 19$. The pilots proposed in [53] for the UTs in one cell are the circularly shifted versions of the sequence $\bar{\mathbf{b}} = \left[\bar{b}_0, \bar{b}_1, \cdots, \bar{b}_{N_\mathrm{p}-1}\right]$, and are reused in different cells, where $\bar{b}_n = \exp(i\pi n(n + (N_\mathrm{p} \bmod 2))/N_\mathrm{p}), n = 0, 1, \cdots, N_\mathrm{p}-1$. For the proposed pilots, the sequence parameter A in (32) is set to one. It can be seen that all these pilots satisfy the requirement of $\mathbf{\Psi}_{ll} = \mathbf{I}_K, \forall l = 1, 2, \cdots, L$. Moreover, for the sake of avoiding any performance-dependence on the network topology, the proposed pilots are allocated randomly to each cell in each trial. The simulations are conducted in a noiseless environment, thus the channel estimation error covariances explicitly reflect the impact of the interference.

In Fig. 3, the sum of correlations $\sum_{j=1}^{L}\sum_{l=1}^{L}\left|\left|\mathbf{\Psi}_{jl}\right|\right|_\mathrm{F}^2$ of different pilots was evaluated for $N_\mathrm{p} = 11, 22, 33, 44, 55, 66$. We can see that the proposed pilots satisfy the Welch bound and their sum of correlations is significantly lower than that of the benchmarkers. In Fig. 4, the estimation error covariances of the proposed SBCE invoking different pilots are shown for $N_\mathrm{p} = 11, 22, 33, 44, 55, 66$. It is observed that for the proposed SBCE the estimation error covariance corresponding to the proposed pilots is significantly lower than that of the benchmarking pilots of [51]–[53]. In Fig. 4, for the sake of clarity, the curves are obtained without considering the noise, which explicitly reflects the impact of pilot designs on the ICI. Jointly observing Fig. 3 and Fig. 4, we can conclude that if the proposed pilots enjoy the minimum sum of correlations, i.e., satisfying the Welch bound, the ICI can also be minimized in the multi-cell scenario. Therefore, the influence of pilot contamination can be more effectively mitigated as compared to the schemes in [51]–[53].

[4] This means that the large-scale fading has boils down to the average pathloss from the long-term perspective, since the shadowing effect has been averaged out statistically.

[5] In each setup considered, $\tilde{A}$ has seven values, each of which is for one cell.

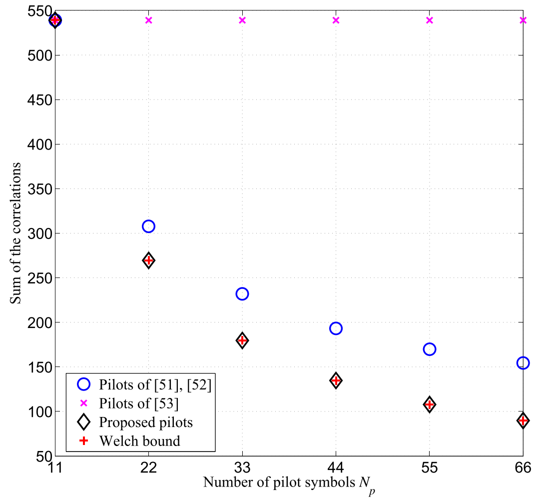

FIGURE 3. Sum of the correlations versus the number of pilot symbols for different pilots, and they are benchmarked against the Welch bound. The number of UTs per cell is $K = 11$, the number of BS antennas is $M = 100$, and the number of cells is $L = 7$. Noiseless flat Rayleigh distributed block-fading channels are considered.

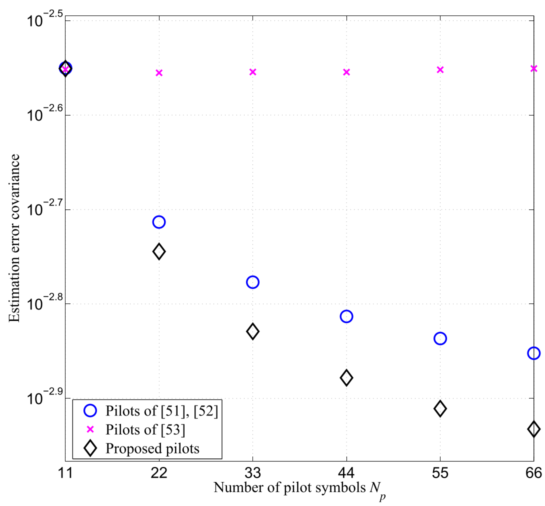

FIGURE 4. The estimation error covariances of the proposed SBCE versus the number of pilot symbols for different pilots. Noiseless flat Rayleigh distributed block-fading channels are considered. The number of UTs per cell is $K = 11$, the number of BS antennas is $M = 100$, and the number of cells is $L = 7$.

B. PERFORMANCE COMPARISON OF THE PROPOSED SBCE AND OTHER CHANNEL ESTIMATORS WHEN USING THE SAME AND SHORT PILOTS

In these illustrations, the number of pilot symbols is set to $N_\mathrm{p} = K$. Note that for the sake of fair comparison, the same pilots obtained from our design are employed for all the channel estimators considered. In this case, the channel estimators of [35] and [45] and the proposed SBCE exhibit the same performance. Hence, the performance of the channel estimators of [35] and [45] is not shown in the simulations.

Fig. 5 shows the estimation error covariances of different channel estimators versus the received SNR $\gamma = p_\mathrm{u}\beta M$ of the UT farthest away from its BS, where $\beta = 1/(r_{\max}/100)^\nu$

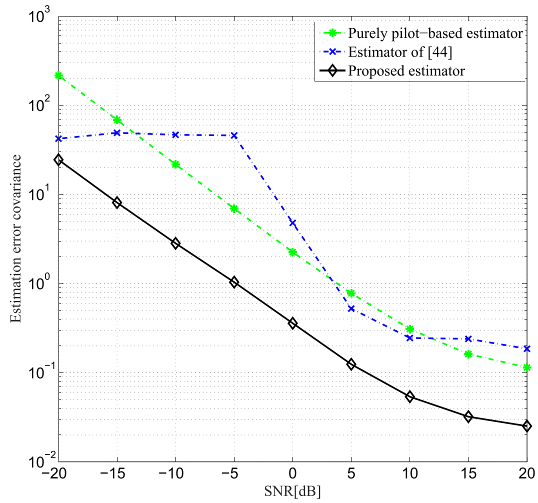

FIGURE 5. The estimation error covariances versus the received SNR γ at the BS for various channel estimators using the same and short pilots obtained from our design. The estimation error covariance is measured for the UT farthest away from its BS. Flat Rayleigh distributed block-fading channels are considered. The number of pilot symbols is $N_\mathrm{p} = K = 11$, where K is the number of UTs per cell. The number of BS antennas is $M = 100$, and the number of cells is $L = 7$.

is the minimal possible pathloss coefficient between the UTs and their serving BSs. In Fig. 5, it is observed that when the received SNR is high, the estimation error covariance of the proposed SBCE is significantly lower than that of the other channel estimators considered. The interference gradually starts to dominate the estimation error covariances of these channel estimators as the SNR increases and the noise becomes negligible. The benchmark SBCE developed from that of [44] becomes biased due to both the interference and the correlation of the channel vectors, which implies that it partially loses the CSI inherently encapsulated in the received signals. Hence the estimation error covariance of the benchmark SBCE inspired by [44] is significantly higher than that of the proposed SBCE. These results confirm that the ICI and intra-cell interference contaminating the estimated signal subspace makes all the benchmark estimators biased, and the ICI cannot be completely eliminated when the estimator of [35] is employed. Moreover, it can be observed that the estimation error covariance of the benchmark SBCE developed from [44] first slightly increases and then slightly decreases when the SNR increases. This phenomenon can be explained by analyzing $\hat{\mathbf{H}}_{jj}^{\mathrm{n}}$ in (26), which corresponds to the benchmark SBCE adapted from [44]. When the SNR is low, the matrix inverse in (26) is dominated by the term $(1/\sqrt{p_\mathrm{u}N_\mathrm{p}})\mathbf{N}_j^{\mathrm{p}}\mathbf{\Phi}_j^{*}\mathbf{D}_{jj}^{-1}$ in (14). Hence, we have $\hat{\mathbf{H}}_{jj}^{\mathrm{n}} \approx M\sqrt{p_\mathrm{u}N_\mathrm{p}}(\mathbf{\Phi}_j^{T}(\mathbf{N}_j^{\mathrm{p}})^{H})^{-1}\mathbf{D}_{jj}$, and the estimation error covariance of the benchmark SBCE adapted from [44] increases when the SNR increases. When the SNR is high and the number of BS antennas M is large, the matrix inverse in (26) is dominated by $\mathbf{\Xi}_{jj}$. Consequently, the matrix inverse can be approximated by its Neumann series [61, p. 55]. Hence, we have $\hat{\mathbf{H}}_{jj}^{\mathrm{n}} \approx (\mathbf{H}_{jj} + \mathbf{F}_j\mathbf{\Xi}_{jj})(\mathbf{I}_K - \mathbf{\Gamma}_{jj} - \mathbf{\Gamma}_{jj}^{H}\mathbf{\Xi}_{jj} - \mathbf{\Upsilon}_j\mathbf{\Xi}_{jj})$, and the estimation error covariance of the benchmark SBCE adapted from [44] decreases when the SNR increases.

Fig. 6 shows the BERs of the different channel estimators versus the received SNR γ, when considering the UT that is farthest away from its BS. The MMSE detector is employed for detecting data symbols, along with the respective channel estimators. It is observed that the BER of the proposed SBCE is significantly lower than that of the benchmark estimator of [44] in the medium- and high-SNR region. These results also validate that the proposed SBCE mitigates the ICI more effectively than the estimator of [44], thus it performs better.

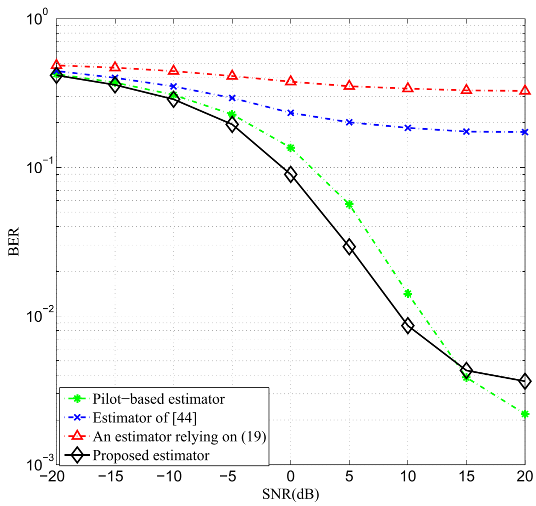

FIGURE 6. **BERs versus the received SNR γ at the BS for various channel estimators. The BER is measured for the UT farthest away from its BS. Flat Rayleigh distributed block-fading channels are considered. The number of pilot symbols is $N_p = K = 11$, where K is the number of UTs per cell. The number of BS antennas is $M = 100$, and the number of cells is $L = 7$. The pilots employed are the same for all the channel estimators considered.**

It can be seen from Fig. 5 and Fig. 6 that the estimation error of the proposed SBCE is lower than that of the purely pilot-based channel estimator, albeit the BER of the former is higher than that of the latter in the high-SNR region. The reason for this phenomenon is that the ICI imposed on the purely pilot-based channel estimator becomes less significant at high SNRs than at low SNRs. More specifically, according to our definition of the received SNR in Sec. V-B, high SNRs imply that the UT considered is near the BS. As a result, the advantage of the proposed SBCE over the purely pilot-based channel estimator quantified in terms of its ICI mitigation becomes less critical.

On the other hand, it can be seen from Fig. 5 and Fig. 6 that the estimation error covariances and BERs of all these channel estimators saturate to some degree in the high-SNR region. The results imply that although the proposed SBCE is capable of mitigating the ICI more effectively than the existing channel estimators, it still fails to completely eliminate the ICI. It is observed from Fig. 6 that the BER of the proposed SBCE based scheme is significantly lower than that of the benchmark estimator of [44] and the estimator of (19) (represented by the triangle-marked red line), which uses the columns of $\mathbf{H}_{jj}$ instead of the subspace $\mathbf{U}_j^s$ in the medium and high SNR regime. The columns of the true signal subspace $\mathbf{U}_j^s$ are highly correlated with the columns of $\mathbf{H}_{jj}$ (the red line), as demonstrated in [33] and [35], but they are not identical. To elaborate a little further, as shown in [33], each column of the channel matrix $\mathbf{H}_{jj}$ can be estimated from the corresponding eigenvector of the covariance matrix $\mathbf{R}_{Y_j^d}$ as M becomes large, i.e., $\mathbf{U}_j = \mathbf{H}_{jj}\tilde{\mathbf{E}}_{jj}$, where $\tilde{\mathbf{E}}_{jj}$ is the ambiguity matrix. As long as $\tilde{\mathbf{E}}_{jj}$ is known, the channel matrix $\mathbf{H}_{jj}$ can be estimated from $\mathbf{U}_j$, where $\mathbf{U}_j = \left[\mathbf{U}_j^s, \mathbf{U}_j^n\right]$.

C. PERFORMANCE COMPARISON OF THE PROPOSED SBCE AND OTHER CHANNEL ESTIMATORS WHEN USING DIFFERENT AND LONG PILOTS

In these investigations, the number of pilot symbols is set to $N_{\mathrm{p}} = 5K$. In this case, the channel estimator of [35] and [45] uses the known pilots of [51]–[53] and the proposed SBCE employs our newly designed pilots. The first benchmark scheme is constituted by the channel estimator of [35] and [45] using the pilots of [51] and [52], while the second one is the channel estimator of [35] and [45] using the pilots of [53].

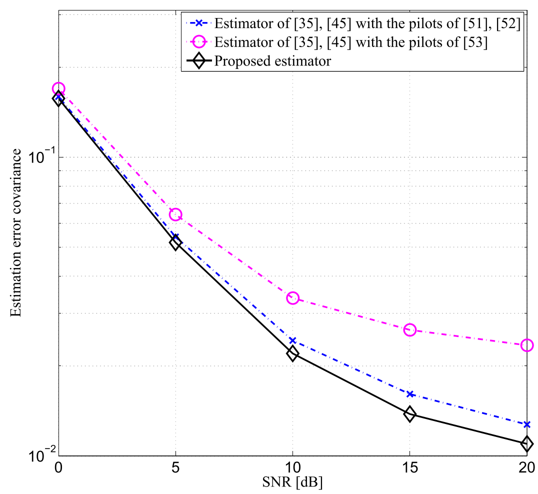

FIGURE 7. **The estimation error covariances versus the received SNR γ at the BS for various channel estimators using different and long pilots. The estimation error covariance is measured for the UT farthest away from its BS. Flat Rayleigh distributed block-fading channels are considered. The number of pilot symbols is $N_{\mathrm{p}} = 5K = 55$, where $K = 11$ is the number of UTs per cell. The number of BS antennas is $M = 100$, and the number of cells is $L = 7$. The proposed SBCE uses our newly designed pilots.**

More specifically, Fig. 7 shows the estimation error covariances of different channel estimators versus the received SNR γ of the UT that is farthest away from its BS. We can see from Fig. 7 that the proposed SBCE outperforms the other two approaches. The simulations demonstrate that the estimation performance of the proposed SBCE is better than that of [35] and [45], where the pilots were not specifically designed for reducing the ICI of LS-MIMO systems. In this regard, Fig. 7 also confirms that the proposed SBCE is more generically applicable than that of [35] and [45], although neither the channel estimator of [35] nor the proposed SBCE can completely eliminate ICI, when the number of BS antennas and the number of data symbols are finite.

VI. CONCLUSIONS AND FUTURE WORK

A semi-blind channel estimator (SBCE) was proposed for improving the performance of LS-MIMO systems. The proposed SBCE first estimates the column space of the channel matrix, and then estimates the channel matrix with the aid of the optimum pilots in the sense of minimizing the channel estimation error covariance. It was shown that the proposed SBCE approach exhibits a significantly better performance than other state-of-the-art SBCEs in LS-MIMO systems. Our theoretical analysis revealed that the optimal pilots should satisfy the Welch bound and be orthogonal to each other within each cell. Inspired by this insight, Zadoff-Chu sequences with the best possible circular shifts were designed, which were shown to be optimal for both the proposed SBCE and the purely pilot-based channel estimator.

Since in this paper we only considered uncorrelated channels, the extension of the proposed SBCE to the scenario of correlated channels may be carried out in our future work. Additionally, since the length of the proposed pilot sequences is constrained to be an integer multiple of the number of simultaneously active UTs per cell, pilot sequences with arbitrary lengths will also be designed in our future work. We can further relax some of the constraints (including the number of the cells, the number of the BS antennas and the number of users) imposed on the optimal pilot sequences by exploiting a simple user scheduling strategy depending on the angle-of-arrival distribution of the selected users. Meanwhile, we can also mitigate or eliminate pilot contamination with other approaches, such as pilot contamination precoding, pilot reuse and directional antennas.

APPENDIX A
PROOF OF LEMMA 1

Lemma 1: Consider a matrix $\mathbf{X} \in \mathbb{C}^{M\times K}$ composed of i.i.d. $\mathcal{CN}(0,1)$ entries. For $M \geq K$, $\mathbf{X}$ can *almost surely* be written as

$$\mathbf{X} = \tilde{\mathbf{X}}\boldsymbol{\Xi}, \tag{35}$$

where $\tilde{\mathbf{X}} \in \mathbb{C}^{M\times K}$ satisfies $(1/M)\tilde{\mathbf{X}}^H\tilde{\mathbf{X}} = \mathbf{I}_K$, and $\boldsymbol{\Xi} \in \mathbb{C}^{K\times K}$ can be expressed as

$$\boldsymbol{\Xi} = \mathbf{I}_K + \boldsymbol{\Gamma}, \tag{36}$$

where $\boldsymbol{\Gamma} \in \mathbb{C}^{K\times K}$. Furthermore, we have $\mathbf{X} \xrightarrow{\text{a.s.}} \tilde{\mathbf{X}}$ when $M \to \infty$.

Proof: For $M \geq K$, $\mathbf{X}$ defined in Lemma 1 *almost surely* has full rank [62, p. 22]. The QR decomposition of $\mathbf{X}$ can be expressed as $\mathbf{X} = (\tilde{\mathbf{X}}/\sqrt{M})(\sqrt{M}\boldsymbol{\Xi}^{-1})$, where $\tilde{\mathbf{X}}$ satisfies $(1/M)\tilde{\mathbf{X}}^H\tilde{\mathbf{X}} = \mathbf{I}_K$, and $\boldsymbol{\Xi}$ is an upper triangular matrix given by $\boldsymbol{\Xi} = \mathbf{I}_K + \boldsymbol{\Gamma}$.

According to the law of large numbers, for a massive MU-MIMO system the following relation holds true [5]:

$$\frac{1}{M}\mathbf{X}^H\mathbf{X} \xrightarrow{\text{a.s.}} \mathbf{I}_K, \quad \text{as } M \to \infty. \tag{37}$$

Hence, we have $\boldsymbol{\Xi} \xrightarrow{\text{a.s.}} \mathbf{I}_K$ as $M \to \infty$, which means $\mathbf{X} \xrightarrow{\text{a.s.}} \tilde{\mathbf{X}}$ as $M \to \infty$. ∎

APPENDIX B
PROOF OF PROPOSITION 1

The proof of Proposition 1 relies on the Lemma 1 given above, while the proof of this lemma is given in Appendix A.

According to Lemma 1, for $M \geq K$, the channel matrix $\mathbf{H}_{jl}$ can *almost surely* be written as

$$\mathbf{H}_{jl} = \tilde{\mathbf{H}}_{jl}\boldsymbol{\Xi}_{jl}, \tag{38}$$

where $\tilde{\mathbf{H}}_{jl} \in \mathbb{C}^{M\times K}$ satisfies $(1/M)\tilde{\mathbf{H}}_{jl_1}^H\tilde{\mathbf{H}}_{jl_2} = \delta(l_1 - l_2)\mathbf{I}_K$, and $\boldsymbol{\Xi}_{jl} \in \mathbb{C}^{K\times K}$ can be expressed as

$$\boldsymbol{\Xi}_{jl} = \mathbf{I}_K + \boldsymbol{\Gamma}_{jl}, \tag{39}$$

in which $\boldsymbol{\Gamma}_{jl} \in \mathbb{C}^{K\times K}$.

Substituting (5) and (38) into (7), we have

$$\begin{aligned}\hat{\mathbf{R}}_{\mathbf{Y}_j^{\mathrm{d}}} &= \frac{1}{N_{\mathrm{d}}}\left(\sqrt{p_{\mathrm{u}}}\sum_{l=1}^{L}\mathbf{G}_{jl}\mathbf{A}_l^T + \mathbf{N}_j^{\mathrm{d}}\right)\\ &\quad\times\left(\sqrt{p_{\mathrm{u}}}\sum_{l=1}^{L}\mathbf{G}_{jl}\mathbf{A}_l^T + \mathbf{N}_j^{\mathrm{d}}\right)^H\\ &= \frac{p_{\mathrm{u}}}{N_{\mathrm{d}}}\tilde{\mathbf{H}}_{jj}\mathbf{P}_{jj}\tilde{\mathbf{H}}_{jj}^H + \mathbf{Q}_j\end{aligned} \tag{40}$$

where $\mathbf{P}_{jj} = \boldsymbol{\Xi}_{jj}\mathbf{D}_{jj}\mathbf{A}_j^T\mathbf{A}_j^*\mathbf{D}_{jj}\boldsymbol{\Xi}_{jj}^H \in \mathbb{C}^{K\times K}$ and $\mathbf{Q}_j \in \mathbb{C}^{M\times M}$ represents the remaining part of the summation in (40). Let us formulate the EVD of $\mathbf{P}_{jj}$ as $\mathbf{P}_{jj} = \mathbf{E}_{jj}\boldsymbol{\Lambda}_{jj}\mathbf{E}_{jj}^H$, where $\mathbf{E}_{jj} \in \mathbb{C}^{K\times K}$ is the matrix of the eigenvectors of $\mathbf{P}_{jj}$, and $\boldsymbol{\Lambda}_{jj} \in \mathbb{R}^{K\times K}$ is a diagonal matrix of the eigenvalues of $\mathbf{P}_{jj}$. Since $\mathbf{P}_{jj}$ is Hermitian, $\mathbf{E}_{jj}$ is a unitary matrix. Then, the K eigenvectors of $\hat{\mathbf{R}}_{\mathbf{Y}_j^{\mathrm{d}}}$ corresponding to the largest K eigenvalues of $\hat{\mathbf{R}}_{\mathbf{Y}_j^{\mathrm{d}}}$ form the matrix $\mathbf{U}_j^{\mathrm{s}} = \left((1/\sqrt{M})\tilde{\mathbf{H}}_{jj}\mathbf{E}_{jj} + \mathbf{B}_j\right) = (1/\sqrt{M})\left(\tilde{\mathbf{H}}_{jj} + \sqrt{M}\mathbf{B}_j\mathbf{E}_{jj}^H\right)\mathbf{E}_{jj}$, where $\mathbf{B}_j \in \mathbb{C}^{M\times K}$ corresponds to both the ICI and the intra-cell interference. Hence, we have

$$\mathbf{U}_j^{\mathrm{s}} = \frac{1}{\sqrt{M}}\left(\tilde{\mathbf{H}}_{jj} + \mathbf{F}_j\right)\mathbf{E}_{jj}, \tag{41}$$

where $\mathbf{F}_j = \sqrt{M}\mathbf{B}_j\mathbf{E}_{jj}^H \in \mathbb{C}^{M\times K}$ also corresponds to the ICI and the intra-cell interference.

Additionally, when $N_{\mathrm{d}} \to \infty$, $M \to \infty$ and K is fixed, $\hat{\mathbf{R}}_{\mathbf{Y}_j^{\mathrm{d}}}$ in (40) tends *almost surely* to $\mathbf{R}_{\mathbf{Y}_j^{\mathrm{d}}}$ in (6) and $\mathbf{H}_{jj} \xrightarrow{\text{a.s.}} \tilde{\mathbf{H}}_{jj}$, and we have

$$\hat{\mathbf{R}}_{\mathbf{Y}_j^{\mathrm{d}}} \xrightarrow{\text{a.s.}} p_{\mathrm{u}}\sum_{l=1}^{L}\tilde{\mathbf{H}}_{jl}\mathbf{D}_{jl}^2\tilde{\mathbf{H}}_{jl}^H + \mathbf{I}_M. \tag{42}$$

$$\begin{aligned}
&\mathbb{E}\left\{[\boldsymbol{\Delta}_j]_n\left([\boldsymbol{\Delta}_j]_n\right)^H\right\} \\
&= \beta_{j,j_n}^{-1}\mathbb{E}\left\{\operatorname{tr}\left(\sum_{l\neq j}\mathbf{D}_{jl}\left[\boldsymbol{\Psi}_{jl}\right]_n\left([\boldsymbol{\Psi}_{jl}]_n\right)^H\mathbf{D}_{jl}+\frac{1}{p_{\mathrm{u}}N_{\mathrm{p}}}\left[\boldsymbol{\Phi}_j^*\right]_n\left(\left[\boldsymbol{\Phi}_j^*\right]_n\right)^H\right)\right\}\mathbf{I}_M \\
&= \sigma_1\mathbf{I}_M
\end{aligned} \tag{44}$$

According to the statement above (3), we know that $\left[\mathbf{D}_{jj}\right]_{k_1,k_1} > \left[\mathbf{D}_{jl}\right]_{k_2,k_2}, \forall l \neq j,\ \forall k_1, k_2$. Therefore, we have

$$\begin{aligned}
\hat{\mathbf{R}}_{\mathbf{Y}_j^{\mathrm{d}}} &\xrightarrow{a.s} p_{\mathrm{u}}\sum_{l=1}^{L}\tilde{\mathbf{H}}_{jl}\mathbf{D}_{jl}^2\tilde{\mathbf{H}}_{jl}^H+\mathbf{I}_M \\
&\approx p_{\mathrm{u}}\tilde{\mathbf{H}}_{jj}\mathbf{D}_{jj}^2\tilde{\mathbf{H}}_{jj}^H+\mathbf{I}_M.
\end{aligned}$$

In other words, the effects of the ICI and the intra-cell interference on $\hat{\mathbf{R}}_{\mathbf{Y}_j^{\mathrm{d}}}$ approximately vanish. Since we have $\frac{1}{M}\tilde{\mathbf{H}}_{jl_1}^H\tilde{\mathbf{H}}_{jl_2} = \delta(l_1 - l_2)\,\mathbf{I}_K$ as $M \to \infty$ (see Lemma 1 in Appendix A), we can conclude that the largest K eigenvalues of $\hat{\mathbf{R}}_{\mathbf{Y}_j^{\mathrm{d}}}$ tend *almost surely* to be the diagonal elements of $\mathbf{D}_{jj}^2$, and the corresponding eigenvectors tend *almost surely* to be in the column space of $\tilde{\mathbf{H}}_{jj}$. Hence, $\mathbf{F}_j$ in (41) tends *almost surely* to a zero matrix.

APPENDIX C
THE ESTIMATION ERROR COVARIANCE OF THE PROPOSED SBCE

Here the estimation error covariance of the proposed SBCE is derived. By substituting (17) into (22), the channel estimation error matrix can be written as

$$\tilde{\mathbf{H}}_{jj}^{\mathrm{s}} = \frac{1}{M}\tilde{\mathbf{H}}_{jj}\tilde{\mathbf{H}}_{jj}^H\boldsymbol{\Delta}_j+\tilde{\mathbf{F}}_j. \tag{43}$$

Using (14), we obtain (44), as shown at the top of this page, where $\sigma_1 = p_{\mathrm{u}}^{-1}N_{\mathrm{p}}^{-1}\beta_{j,j_n}^{-1}+\beta_{j,j_n}^{-1}\sum_{l\neq j}\sum_{k=1}^{K}\beta_{j,l_k}\left|[\boldsymbol{\Psi}_{jl}]_{k,n}\right|^2$. Then, the expectation is given by

$$\begin{aligned}
&\mathbb{E}\left\{\left|\left[\frac{1}{M}\tilde{\mathbf{H}}_{jj}\tilde{\mathbf{H}}_{jj}^H\boldsymbol{\Delta}_j\right]_{m,n}\right|^2\right\} \\
&= \frac{\sigma_1}{M^2}\left(\left[\tilde{\mathbf{H}}_{jj}^T\right]_m\right)^T\tilde{\mathbf{H}}_{jj}^H\tilde{\mathbf{H}}_{jj}\left(\left[\tilde{\mathbf{H}}_{jj}^T\right]_m\right)^* \\
&= \frac{\sigma_1\sigma_2}{M},
\end{aligned} \tag{45}$$

where $\sigma_2 = \left\|\left[\tilde{\mathbf{H}}_{jj}^T\right]_m\right\|^2$. By substituting (43) and (45) into (21), $\left[\mathbf{R}_{\tilde{\mathbf{H}}_{jj}^{\mathrm{s}}}\right]_{m,n}$ can be expressed as (23).

APPENDIX D
PROOF OF PROPOSITION 2

Since $s_n = \lfloor (n-1)L/a \rfloor, n = 1, 2, \cdots, N_{\mathrm{p}}, a = 1, 2, \cdots, L-1$, we obtain $s_{n+a} = s_n + L,\ n = 1, 2, \cdots, N_{\mathrm{p}} - a$. When $N_{\mathrm{p}} = aK,\ 0 \leq s_n < KL$ holds. Because $a < L$, we have $s_{m_1} \neq s_{m_2}$ for $m_1 \neq m_2$, which means that the rows of $\boldsymbol{\Omega}$ are orthogonal. Therefore, $N_{\mathrm{p}} = aK, a = 1, 2, \cdots, L-1$ implies that the designed pilots are WBE sequences.

Then, the relationship of $\boldsymbol{\Psi}_{ll} = \mathbf{I}_K, \forall l$ is proven as follows. Using (32), we arrive at $b_n = b_{n+KL}$. By substituting (33) into (34), the column vector $[\boldsymbol{\Omega}]_m$ can be written as

$$([\boldsymbol{\Omega}]_m)^T = \frac{1}{\sqrt{N_{\mathrm{p}}}}\left[b_{s_1+m-1}, b_{s_2+m-1}, \cdots, b_{s_{N_{\mathrm{p}}}+m-1}\right], \tag{46}$$

where $m = 1, 2, \cdots, KL$. Then, the inner product of $[\boldsymbol{\Omega}]_m$ and $[\boldsymbol{\Omega}]_{m+j}$ is given by

$$\begin{aligned}
([\boldsymbol{\Omega}]_m)^T([\boldsymbol{\Omega}]_{m+j})^* &= \frac{1}{N_{\mathrm{p}}}\sum_{n=1}^{N_{\mathrm{p}}}b_{s_n+m-1}b_{s_n+m+j-1}^* \\
&= \eta_m\sum_{n=1}^{N_{\mathrm{p}}}\exp\left(i\frac{-2\pi A}{KL}js_n\right),
\end{aligned} \tag{47}$$

where we have

$$\eta_m = \frac{1}{N_{\mathrm{p}}}\exp\left(i\frac{\pi A}{KL}\left(-j^2-2j(m-1)-j\,(KL \bmod 2)\right)\right).$$

Using $N_{\mathrm{p}} = aK, a = 1, 2, \cdots, L-1$, (47) can be reformulated as

$$\begin{aligned}
([\boldsymbol{\Omega}]_m)^T([\boldsymbol{\Omega}]_{m+j})^* &= \eta_m\sum_{k=0}^{K-1}\sum_{n=ka+1}^{(k+1)a}\exp\left(i\frac{-2\pi A}{KL}js_n\right) \\
&= \zeta_m K\delta(j \bmod K),
\end{aligned} \tag{48}$$

where

$$\zeta_m = \eta_m\sum_{n=1}^{a}\exp\left(i\frac{-2\pi A}{KL}js_n\right).$$

Based on the definition of $\boldsymbol{\Omega}$ in (30), it can be seen that the inner product of any two different columns of $\boldsymbol{\Phi}_l$ is zero, which implies that $\boldsymbol{\Psi}_{ll} = \mathbf{I}_K, \forall l$.

REFERENCES

[1] D. Gesbert, M. Kountouris, R. W. Heath, Jr., C.-B. Chae, and T. Sälzer, "Shifting the MIMO paradigm," *IEEE Signal Process. Mag.*, vol. 24, no. 5, pp. 36–46, Sep. 2007.

[2] 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.

[3] 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.

[4] S. Yang and L. Hanzo, "Fifty years of MIMO detection: The road to large-scale MIMOs," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 4, pp. 1941–1988, Fourth Quarter 2015.

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

[6] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Uplink power efficiency of multiuser MIMO with very large antenna arrays," in *Proc. 49th Annu. Allerton Conf. Commun., Control, Comput. (Allerton)*, Monticello, IL, USA, Sep. 2011, pp. 1272–1279.

[7] A. Hu, T. Lv, H. Gao, Z. Zhang, and S. Yang, "An ESPRIT-based approach for 2-D localization of incoherently distributed sources in massive MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 996–1011, Oct. 2014.

[8] M. Cheng, S. Yang, and X. Fang, "Adaptive antenna-activation based beamforming for large-scale MIMO communication systems of high speed railway," *China Commun.*, 2016, accepted to publish. [Online]. Available: http://eprints.soton.ac.uk/365444/

[9] J. Zhang, C. Yuen, C.-K. Wen, S. Jin, and X. Q. Gao, "Ergodic secrecy sum-rate for multiuser downlink transmission via regularized channel inversion: Large system analysis," *IEEE Commun. Lett.*, vol. 18, no. 9, pp. 1627–1630, Sep. 2014.

[10] X. Gao, L. Dai, C. Yuen, and Z. Wang, "Turbo-like beamforming based on tabu search algorithm for millimeter-wave massive MIMO systems," *IEEE Trans. Veh. Technol.*, 2016, accepted to publish. [Online]. Available: http://dx.doi.org/10.1109/TVT.2015.2461440

[11] L. Liu, C. Yuen, Y. L. Guan, Y. Li, and Y. Su, "A low-complexity Gaussian message passing iterative detector for massive MU-MIMO systems," in *Proc. IEEE Int. Conf. Inf., Commun. Signal Process. (ICICS)*, Singapore, Dec. 2015, pp. 1–5.

[12] T. Lv, F. Tan, H. Gao, and S. Yang, "A beamspace approach for 2-D localization of incoherently distributed sources in massive MIMO systems," *Signal Process.*, vol. 121, pp. 30–45, Apr. 2016.

[13] S. Yang, C. Zhou, T. Lv, and L. Hanzo, "Large-scale MIMOs are capable of eliminating power-thirsty channel coding for wireless transmission of HEVC/H.265 video," *IEEE Wireless Commun.*, 2016, accepted to publish. [Online]. Available: http://arxiv.org/pdf/1601.06684v1.pdf

[14] H. Liu, H. Gao, S. Yang, and T. Lv, "Low-complexity downlink user selection for massive MIMO systems," *IEEE Syst. J.*, 2016, accepted to publish. [Online]. Available: http://dx.doi.org/10.1109/JSYST.2015.2422475

[15] Z. Gao, L. Dai, Z. Lu, C. Yuen, and Z. Wang, "Super-resolution sparse MIMO-OFDM channel estimation based on spatial and temporal correlations," *IEEE Commun. Lett.*, vol. 18, no. 7, pp. 1266–1269, Jul. 2014.

[16] Z. Gao, L. Dai, C. Yuen, and Z. Wang, "Asymptotic orthogonality analysis of time-domain sparse massive MIMO channels," *IEEE Commun. Lett.*, vol. 19, no. 10, pp. 1826–1829, Oct. 2015.

[17] O. Rabaste and T. Chonavel, "Estimation of multipath channels with long impulse response at low SNR via an MCMC method," *IEEE Trans. Signal Process.*, vol. 55, no. 4, pp. 1312–1325, Apr. 2007.

[18] N. L. Pedersen, C. N. Manchón, and B. H. Fleury, "A fast iterative Bayesian inference algorithm for sparse channel estimation," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Budapest, Hungary, Jun. 2013, pp. 4591–4596.

[19] F. Poree, O. Rosec, T. Chonavel, and J. M. Boucher, "Two Bayesian methods for multipath propagation parameters estimation," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Istanbul, Turkey, Jun. 2000, pp. 69–72.

[20] T. Datta, N. A. Kumar, A. Chockalingam, and B. S. Rajan, "A novel Monte-Carlo-sampling-based receiver for large-scale uplink multiuser MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 62, no. 7, pp. 3019–3038, Sep. 2013.

[21] L. Wan, G. Han, J. J. P. C. Rodrigues, W. Si, and N. Feng, "An energy efficient DOA estimation algorithm for uncorrelated and coherent signals in virtual MIMO systems," *Telecommun. Syst.*, vol. 59, no. 1, pp. 93–110, May 2015.

[22] L. Wan, G. Han, J. Jiang, J. J. P. C. Rodrigues, N. Feng, and T. Zhu, "DOA estimation for coherently distributed sources considering circular and noncircular signals in massive MIMO systems," *IEEE Syst. J.*, 2016, accepted to publish. [Online]. Available: http://dx.doi.org/10.1109/JSYST.2015.2445052

[23] 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.

[24] 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.

[25] 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 (ISIT)*, Seoul, South Korea, Jun./Jul. 2009, pp. 2184–2188.

[26] 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.

[27] K. Appaiah, A. Ashikhmin, and T. L. Marzetta, "Pilot contamination reduction in multi-user TDD systems," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Cape Town, South Africa, May 2010, pp. 1–5.

[28] 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. (ICASSP)*, Prague, Czech Republic, May 2011, pp. 3464–3467.

[29] 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.

[30] J. Zuo, J. Zhang, C. Yuen, W. Jiang, and W. Luo, "Multi-cell multi-user massive MIMO transmission with downlink training and pilot contamination precoding," *IEEE Trans. Veh. Technol.*, 2016, accepted to publish. [Online]. Available: http://dx.doi.org/10.1109/TVT.2015.2475284

[31] E. Björnson, J. Hoydis, M. Kountouris, and M. Debbah, "Massive MIMO systems with non-ideal hardware: Energy efficiency, estimation, and capacity limits," *IEEE Trans. Inf. Theory*, vol. 60, no. 11, pp. 7112–7139, Nov. 2014.

[32] H. Huh, G. Caire, H. C. Papadopoulos, and S. A. Ramprashad, "Achieving 'massive MIMO' spectral efficiency with a not-so-large number of antennas," *IEEE Trans. Wireless Commun.*, vol. 11, no. 9, pp. 3226–3239, Sep. 2012.

[33] H. Q. Ngo and E. G. Larsson, "EVD-based channel estimation in multicell multiuser MIMO systems with very large antenna arrays," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Kyoto, Japan, Mar. 2012, pp. 3249–3252.

[34] J. Hoydis, K. Hosseini, S. ten Brink, and M. Debbah, "Making smart use of excess antennas: Massive MIMO, small cells, and TDD," *Bell Labs Tech. J.*, vol. 18, no. 2, pp. 5–21, Sep. 2013.

[35] 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.

[36] S. L. H. Nguyen and A. Ghrayeb, "Compressive sensing-based channel estimation for massive multiuser MIMO systems," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Shanghai, China, Apr. 2013, pp. 2890–2895.

[37] N. Shariati, E. Björnson, M. Bengtsson, and M. Debbah, "Low-complexity polynomial channel estimation in large-scale MIMO with arbitrary statistics," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 815–830, Oct. 2014.

[38] L. Tong, G. Xu, and T. Kailath, "Blind identification and equalization based on second-order statistics: A time domain approach," *IEEE Trans. Inf. Theory*, vol. 40, no. 2, pp. 340–349, Mar. 1994.

[39] L. Tong and S. Perreau, "Multichannel blind identification: From subspace to maximum likelihood methods," *Proc. IEEE*, vol. 86, no. 10, pp. 1951–1968, Oct. 1998.

[40] S. Abdallah and I. N. Psaromiligkos, "Widely linear versus conventional subspace-based estimation of SIMO flat-fading channels: Mean squared error analysis," *IEEE Trans. Signal Process.*, vol. 60, no. 3, pp. 1307–1318, Mar. 2012.

[41] K. Abed-Meraim, P. Loubaton, and E. Moulines, "A subspace algorithm for certain blind identification problems," *IEEE Trans. Inf. Theory*, vol. 43, no. 2, pp. 499–511, Mar. 1997.

[42] V. Buchoux, O. Cappé, É. Moulines, and A. Gorokhov, "On the performance of semi-blind subspace-based channel estimation," *IEEE Trans. Signal Process.*, vol. 48, no. 6, pp. 1750–1759, Jun. 2000.

[43] E. de Carvalho and D. T. M. Slock, "Blind and semi-blind FIR multichannel estimation: (Global) identifiability conditions," *IEEE Trans. Signal Process.*, vol. 52, no. 4, pp. 1053–1064, Apr. 2004.

[44] B. Muquet, M. de Courville, and P. Duhamel, "Subspace-based blind and semi-blind channel estimation for OFDM systems," *IEEE Trans. Signal Process.*, vol. 50, no. 7, pp. 1699–1712, Jul. 2002.

[45] Y. Zeng and T.-S. Ng, "A semi-blind channel estimation method for multiuser multiantenna OFDM systems," *IEEE Trans. Signal Process.*, vol. 52, no. 5, pp. 1419–1429, May 2004.

[46] L. Hanzo, M. Münster, B. J. Choi, and T. Keller, *OFDM and MC-CDMA for Broadband Multi-User Communications, WLANs and Broadcasting*. Chichester, U.K.: Wiley, 2003.
[47] S. Waldron, "Generalized Welch bound equality sequences are tight frames," *IEEE Trans. Inf. Theory*, vol. 49, no. 9, pp. 2307–2309, Sep. 2003.
[48] D. C. Chu, "Polyphase codes with good periodic correlation properties," *IEEE Trans. Inf. Theory*, vol. 18, no. 4, pp. 531–532, Jul. 1972.
[49] R. L. Frank, "Comments on 'polyphase codes with good correlation properties,'" *IEEE Trans. Inf. Theory*, vol. 19, no. 2, p. 244, Mar. 1973.
[50] B. M. Popovic, "Generalized chirp-like polyphase sequences with optimum correlation properties," *IEEE Trans. Inf. Theory*, vol. 38, no. 4, pp. 1406–1409, Jul. 1992.
[51] G. Kang, P. Hasselbach, Y. Yang, P. Zhang, and A. Klein, "Pilot design for inter-cell interference mitigation in MIMO OFDM systems," *IEEE Commun. Lett.*, vol. 11, no. 3, pp. 237–239, Mar. 2007.
[52] J. W. Kang, Y. Whang, H. Y. Lee, and K. S. Kim, "Optimal pilot sequence design for multi-cell MIMO-OFDM systems," *IEEE Trans. Wireless Commun.*, vol. 10, no. 10, pp. 3354–3367, Oct. 2011.
[53] W. Wang, X. Wu, and G. Kang, "Pilot sequence design scheme for inter-cell interference mitigation in OFDM systems under time-varying channels," in *Proc. IEEE Veh. Technol. Conf. (VTC Spring)*, Singapore, May 2008, pp. 314–318.
[54] 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.
[55] J. Zhang, X. Yuan, and L. Ping, "Hermitian precoding for distributed MIMO systems with individual channel state information," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 241–250, Feb. 2013.
[56] H. Yang and T. L. Marzetta, "Performance of conjugate and zero-forcing beamforming in large-scale antenna systems," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 172–179, Feb. 2013.
[57] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1993.
[58] L. Tong, B. M. Sadler, and M. Dong, "Pilot-assisted wireless transmissions: General model, design criteria, and signal processing," *IEEE Signal Process. Mag.*, vol. 21, no. 6, pp. 12–25, Nov. 2004.
[59] H. L. V. Trees, "Classical detection and estimation theory," in *Detection Estimation and Modulation Theory, Part I: Detection, Estimation, and Filtering Theory*. New York, NY, USA: Wiley, 2001.
[60] R. Reams and S. Waldron, "Isometric tight frames," *Electron. J. Linear Algebra*, vol. 9, pp. 122–128, Jul. 2002.
[61] G. W. Stewart, "Matrices, algebra, and analysis," in *Matrix Algorithms*, vol. 1. Philadelphia, PA, USA: SIAM, 1998.
[62] A. M. Tulino and S. Verdú, *Random Matrix Theory and Wireless Communications*. Hanover, MA, USA: Now Publishers, 2004.
[63] G. H. Golub and C. F. Van Loan, *Matrix Computations*. Baltimore, MD, USA: The Johns Hopkins Univ. Press, 2012.

TIEJUN LV (M'08–SM'12) received the M.S. and Ph.D. degrees in electronic engineering from the University of Electronic Science and Technology of China, Chengdu, China, in 1997 and 2000, respectively.

He was a Post-Doctoral Fellow with Tsinghua University, Beijing, China, from 2001 to 2003. In 2005, he became a Full Professor with the School of Information and Communication Engineering, Beijing University of Posts and Telecommunications. From 2008 to 2009, he was a Visiting Professor with the Department of Electrical Engineering, Stanford University, Stanford, CA, USA. He has authored over 200 published technical papers on the physical layer of wireless mobile communications. His current research interests include signal processing, communications theory, and networking.

Dr. Lv is also a Senior Member of the Chinese Electronics Association. He was a recipient of the Program for New Century Excellent Talents in University Award from the Ministry of Education, China, in 2006.

SHAOSHI YANG (S'09–M'13) received the B.Eng. degree in information engineering from the Beijing University of Posts and Telecommunications (BUPT), China, in 2006, the Ph.D. degree in electronics and electrical engineering from the University of Southampton, U.K., in 2013, and the Ph.D. degree in signal and information processing from BUPT, in 2014. Since 2013, he has been a Post-Doctoral Research Fellow with the University of Southampton, U.K, and from 2008 to 2009, he was an Intern Research Fellow with Intel Labs China, Beijing, where he focused on channel quality indicator channel design for mobile WiMAX (802.16 m). He has authored in excess of 30 research papers in the IEEE journals. His research interests include MIMO signal processing, green radio, heterogeneous networks, cross-layer interference management, and convex optimization and its applications.

He has received a number of academic and research awards, including the PMC-Sierra Telecommunications Technology Scholarship at BUPT, the Electronics and Computer Science Scholarship of the University of Southampton, and the Best Ph.D. Thesis Award of BUPT. He serves as a TPC Member of a number of IEEE conferences and journals, including the IEEE ICC, PIMRC, ICCVE, HPCC, and the IEEE Journal on Selected Areas in Communications. He is also a Junior Member of the Isaac Newton Institute for Mathematical Sciences, Cambridge University, U.K. For his recent research progress, please refer to https://sites.google.com/site/shaoshiyang.

HUI GAO (S'10–M'13) received the B.Eng. degree in information engineering and the Ph.D. degree in signal and information processing from the Beijing University of Posts and Telecommunications (BUPT), Beijing, China, in 2007 and 2012, respectively. From 2009 to 2012, he served as a Research Assistant with the Wireless and Mobile Communications Technology R&D Center, Tsinghua University, Beijing. In 2012, he visited the Singapore University of Technology and Design (SUTD), Singapore, as a Research Assistant. From 2012 to 2014, he was a Post-Doctoral Researcher with SUTD. He is with the School of Information and Communication Engineering, BUPT, as an Assistant Professor. His research interests include massive MIMO systems, cooperative communications, and ultrawideband wireless communications.