CrossMark

Massive MIMO with multi-cell MMSE processing: exploiting all pilots for interference suppression

Xueru Li[1] , Emil Björnson[2*], Erik G. Larsson[2], Shidong Zhou[1] and Jing Wang[1]

Abstract

A new state-of-the-art multi-cell minimum mean square error (M-MMSE) scheme is proposed for massive multiple-input-multiple-output (MIMO) networks, which includes an uplink MMSE detector and a downlink MMSE precoder. Contrary to conventional single-cell schemes that suppress interference using only channel estimates for intra-cell users, our scheme shows the optimal way to suppress both intra-cell and inter-cell interference instantaneously by fully utilizing the available pilot resources. Specifically, let K and B denote the number of users per cell and the number of orthogonal pilot sequences in the network, respectively, where $\beta = B/K$ is the pilot reuse factor. Our scheme utilizes all B channel directions that can be estimated locally at each base station, to actively suppress both intra-cell and inter-cell interference. Our scheme is practical and general, since power control, imperfect channel estimation, and arbitrary pilot allocation are all accounted for. Simulations show that significant spectral efficiency (SE) gains are obtained over the conventional single-cell MMSE scheme and the multi-cell zero-forcing (ZF) scheme. Furthermore, large-scale approximations of the uplink and downlink signal-to-interference-and-noise ratios (SINRs) are derived, which are tight in the large-system limit. These approximations are easy to compute and very accurate even for small system dimensions. Using these SINR approximations, a low-complexity power control algorithm is further proposed to maximize the sum SE.

Keywords: Massive MIMO, Multi-cell MMSE, Large-scale SINR approximations, Power control

1 Introduction

Multi-user multiple-input-multiple-output (MU-MIMO) communication has drawn considerable interest in recent years. By scheduling multiple users on the same time-frequency resource, the spatial degrees of freedom offered by multiple antennas can be exploited to focus signals on intended receivers, reduce interference, and thereby increase the system throughput [1–6]. These features motivate that MU-MIMO technology is incorporated into recent and evolving wireless standards like 4G LTE-Advanced [7].

Massive MU-MIMO is an emerging 5G technology that scales up MU-MIMO by orders of magnitude [8, 9]. The idea is to employ an array comprising a hundred, or more,

antennas at the base station (BS) and serve tens of users simultaneously per cell. Compared to the contemporary cellular systems, the system throughput can be drastically increased without consuming extra bandwidth [7–9]. The uplink and downlink transmit power can also be reduced by an order of magnitude since the phase-coherent processing provides a comparable array gain [10]. In the limit of an infinite number of antennas, intra-cell interference and uncorrelated noise can be averaged out by using simple coherent precoders and detectors, and the main performance limitations are pilot contamination and the distortions from hardware impairments [8, 11].

In the uplink reception and downlink transmission, the most common processing schemes are matched filtering (MF), zero-forcing (ZF), and minimum mean square error (MMSE) processing, where the latter is referred to as single-cell MMSE (S-MMSE) in this work.[1] A key char-

*Correspondence: emil.bjornson@liu.se
[2]Department of Electrical Engineering (ISY), Linköping University, SE-58183 Linköping, Sweden
Full list of author information is available at the end of the article

Li *et al. EURASIP Journal on Wireless Communications and Networking* (2017) 2017:117

Page 2 of 15

acteristic of these schemes is that the BS only utilizes the instantaneous realizations of the channels to its own intra-cell users when creating the precoders/detectors, while users in other cells are either neglected or only considered based on their long-term statistics [12]. This is why we refer to MF, ZF, and S-MMSE as *single-cell schemes*. In the coordinated multipoint (CoMP) literature, there also exist *multi-cell schemes* that exploit the instantaneous channel realizations of the users in all cell; see [13] for an overview. However, there is no scalable solution for estimating all these channel realizations in a large system.

Massive MIMO addresses the channel estimation issue by operating in time-division duplex (TDD) mode and requiring only uplink pilots for channel estimation. Hence, the pilot overhead scales linearly with the number of users, instead of the number of BS antennas, which allows for adding additional antennas without affecting the pilot overhead [14]. The BS first listens to the uplink pilot signaling from its own cell, estimates the K intra-cell channels, and then constructs its precoders/detectors based on these channel estimates to mitigate the intra-cell interference [12, 15–17]. In principle, the BS can also estimate and utilize the channels from users in neighboring cells, but the channel estimates can be very unreliable due to pilot contamination. As shown in [17], the gains are marginal in the baseline scenario with uncorrelated Rayleigh fading channels and every pilot being reused in every cell, and a similar conclusion is drawn in [18].

In this work, we explore multi-cell scenarios where the pilot signals are not reused in every cell. Let B denote the number of orthogonal pilot sequences and K denote the number of users in each cell. The pilot reuse factor is $\beta = B/K \geq 1$, which implies that $1/\beta$ of the cells use a particular pilot sequence. In this case, a BS can estimate the channels to inter-cell users more reliably by utilizing the $B - K$ pilots that are not used in the own cell. In our previous work [19], we used these estimates to propose a multi-cell ZF detector (referred to as full-pilot ZF detector in [19]) to cancel interference from neighboring cells. Unfortunately, the gains over the single-cell schemes were marginal, partly due to the loss in array gain of B in multi-cell ZF, instead of K as with single-cell ZF. Therefore, in this work, we derive and analyze the uplink multi-cell MMSE (M-MMSE) detector and downlink M-MMSE precoder instead, under arbitrary pilot reuse and pilot allocation. This is a generalization of the M-MMSE schemes considered in [17] and [20] for the special case of $B = K$ and in [21] for the idealized case of perfect channel state information (CSI).

The main contributions of our paper are:

- A new state-of-the-art M-MMSE scheme is proposed, which includes an uplink detector and a downlink precoder. The novelty is that all B pilots are exploited at each BS to actively suppress both intra-cell and inter-cell interference. It brings significant SE gains over conventional single-cell schemes which dominate the MIMO literature. Moreover, we prove that the computational complexity of the scheme is scalable since the KL channels, in an L-cell setup, are fully represented by only B channel direction estimates. The proposed scheme is general since it accounts for imperfect channel estimation, power control, and arbitrary pilot allocation.

- Large-scale approximations of the uplink and downlink signal-to-interference-and-noise ratios (SINRs) for the proposed M-MMSE scheme are derived, which are asymptotically tight in the large-system limit. The approximations are very accurate even for small system dimensions and are easy to compute, which enables performance analysis and optimization without the need for heavy Monte-Carlo simulations.

- By utilizing the SINR approximations, a low-complexity power control algorithm for sum SE maximization is proposed. Since the SINR approximations depend only on long-term statistics, the computation complexity can be spread over time. Compared to equal power allocation, the proposed algorithm significantly improves the sum SE and provides good user fairness.

The paper is organized as follows: In Section 2, we describe the system model and the construction of the M-MMSE scheme. Large-scale approximations of the uplink and downlink SINRs are derived in Section 3. A power control algorithm is proposed in Section 4. Simulation results are provided in Section 5 before we conclude the paper in Section 6.

Notation: Boldface lower and upper case symbols represent vectors and matrices, respectively. The trace, transpose, conjugate, Hermitian transpose, and matrix inverse operators are denoted by $\text{tr}(\cdot)$, $(\cdot)^T$, $(\cdot)^*$, $(\cdot)^H$, and $(\cdot)^{-1}$, respectively. The function $\text{diag}(\cdot)$ constructs a diagonal matrix by selecting the diagonal elements of a matrix.

2 System model and transceiver design

We consider a synchronous massive MIMO cellular network with multiple cells. Each cell is assigned an index in the cell set $\mathcal{L}$, and the cardinality $|\mathcal{L}|$ is the number of cells. The BS in each cell is equipped with an array of M antennas and serves K single-antenna users within each coherence block. We assume that each time-frequency coherence block consists of T_c seconds and W_c Hz, such that T_c is smaller than the coherence time of all users and W_c is smaller than the coherence

Li *et al. EURASIP Journal on Wireless Communications and Networking* (2017) 2017:117

Page 3 of 15

bandwidth of all users. This leaves room for $S = T_c \times W_c$ transmission symbols per block, and the channels of all users remain constant within each block. Let $\mathbf{h}_{jlk}$ denote the channel response from user k in cell l to BS j within a block, and assume that it is a realization from a zero-mean circularly symmetric complex Gaussian distribution:

$$\mathbf{h}_{jlk} \sim \mathcal{CN}\left(\mathbf{0}, d_j\left(\mathbf{z}_{lk}\right)\mathbf{I}_M\right). \tag{1}$$

The vector $\mathbf{z}_{lk} \in \mathbb{R}^2$ is the geographical position of user k in cell l, and $d_j(\mathbf{z})$ is an arbitrary function that accounts for the channel attenuation (e.g., path loss and shadowing) between BS j and any user position $\mathbf{z}$. Since the user position changes relatively slowly, $d_j(\mathbf{z}_{lk})$ is assumed to be known at BS j for all l and all k.

We consider a TDD protocol in this paper, where the downlink channels are estimated by uplink pilot signaling by exploiting channel reciprocity.[2] In TDD mode, each transmission block is divided into two phases: 1) uplink channel estimation phase, where each BS acquires CSI from uplink pilot signaling which occupies B out of S symbols in each block, and 2) uplink and downlink payload data transmission phase, where each BS processes the received uplink signal and the to-be-transmitted downlink signals using the estimated CSI. Let ζ^{ul} and ζ^{dl} denote the fixed fractions allocated for uplink and downlink payload data transmission, respectively. These fractions can be selected arbitrarily under the conditions that $\zeta^{\mathrm{ul}} + \zeta^{\mathrm{dl}} = 1$ and that $\zeta^{\mathrm{ul}}(S - B)$ and $\zeta^{\mathrm{dl}}(S - B)$ are positive integers. The uplink channel estimation is first discussed to lay a foundation for the transceiver design.

2.1 Uplink channel estimation

The B pilot symbols in a coherence block are used for transmitting B-length pilot sequences. We consider a set of B orthogonal sequences with unit-modulus entries, denoted as $\mathbf{v}_1, \ldots, \mathbf{v}_B \in \mathbb{C}^B$. These sequences could, for instance, be selected as the columns of a discrete Fourier transform (DFT) matrix. By gathering the sequences in a matrix $\mathbf{V} = [\mathbf{v}_1, \ldots, \mathbf{v}_B] \in \mathbb{C}^{B \times B}$, our orthogonality and scaling assumptions lead to $\mathbf{V}^H\mathbf{V} = B\mathbf{I}_B$.

Arbitrary pilot allocation is considered in this work, with the only requirement of $B \geq K$. The parameter $\beta = B/K \geq 1$ is called the pilot reuse factor. If the pilots are allocated wisely in the network, a larger β brings a lower level of interference during the pilot transmission, also known as pilot contamination. Let $i_{lk} \in \{1, \ldots, B\}$ denote the index of the pilot sequence used by user k in cell l, which implies that the user sends the pilot sequence $\mathbf{v}_{i_{lk}}$ (i.e., the i_{lk}th column of $\mathbf{V}$).

In the uplink channel estimation phase, the collective received signal at BS j is denoted as $\mathbf{Y}_j \in \mathbb{C}^{M \times B}$. Then, $\mathbf{Y}_j$ can be expressed as

$$\mathbf{Y}_j = \sum_{l \in \mathcal{L}} \sum_{k=1}^{K} \sqrt{p_{lk}}\mathbf{h}_{jlk}\mathbf{v}_{i_{lk}}^T + \mathbf{N}_j, \tag{2}$$

where $\mathbf{h}_{jlk}$ is the channel response defined in (1), $p_{lk} \geq 0$ is the transmit power for the pilot of user k in cell l, and the additive white Gaussian noise (AWGN) term $\mathbf{N}_j \in \mathbb{C}^{M \times B}$ contains independent and identically distributed (iid) elements that are distributed as $\mathcal{CN}(0, \sigma^2)$.

Based on the received signal in (2), BS j can compute the MMSE estimate of the uplink channel $\mathbf{h}_{jlk}$ from user k in cell l as [19]

$$\hat{\mathbf{h}}_{jlk} = \sqrt{p_{lk}}d_j(\mathbf{z}_{lk})\mathbf{Y}_j\left(\mathbf{\Psi}_j^*\right)^{-1}\mathbf{v}_{i_{lk}}^*, \tag{3}$$

where $\mathbf{\Psi}_j = \sum_{\ell \in \mathcal{L}} \sum_{m=1}^{K} p_{\ell m}d_j(\mathbf{z}_{\ell m})\mathbf{v}_{i_{\ell m}}\mathbf{v}_{i_{\ell m}}^H + \sigma^2\mathbf{I}_B$. As pointed out in [19], the part $\mathbf{Y}_j\left(\mathbf{\Psi}_j^*\right)^{-1}\mathbf{v}_{i_{lk}}^*$ in (3) depends only on which pilot sequence that user k in cell l uses and is otherwise the same for all users. Consequently, users who use the same pilot sequence have parallel estimated channels at each BS, while only the amplitudes of their estimates are different (due to the factor $\sqrt{p_{lk}}d_j(\mathbf{z}_{lk})$ in (3)). To show this explicitly, define the $M \times B$ matrix

$$\hat{\mathbf{H}}_{\mathcal{V},j} = \left[\hat{\mathbf{h}}_{\mathcal{V},j1}, \ldots, \hat{\mathbf{h}}_{\mathcal{V},jB}\right] = \mathbf{Y}_j\left(\mathbf{\Psi}_j^*\right)^{-1}\left[\mathbf{v}_1^*, \ldots, \mathbf{v}_B^*\right], \tag{4}$$

which allows the channel estimate in (3) to be reformulated as

$$\hat{\mathbf{h}}_{jlk} = \sqrt{p_{lk}}d_j\left(\mathbf{z}_{lk}\right)\hat{\mathbf{H}}_{\mathcal{V},j}\mathbf{e}_{i_{lk}}, \tag{5}$$

where $\mathbf{e}_i$ denotes the ith column of the identity matrix $\mathbf{I}_B$. The property that users with the same pilot have parallel estimated channels is utilized to derive and analyze new SE expressions in the sequel.

Notice that the estimate $\hat{\mathbf{h}}_{jlk}$ is a zero-mean complex Gaussian vector and its covariance matrix is $\mathbb{E}\left\{\hat{\mathbf{h}}_{jlk}\hat{\mathbf{h}}_{jlk}^H\right\} = p_{lk}d_j(\mathbf{z}_{lk})^2\tilde{\phi}_{ji_{lk}}\mathbf{I}_M$, where

$$\tilde{\phi}_{ji_{lk}} = \frac{B}{\sum_{\ell \in \mathcal{L}}\sum_{m=1}^{K} p_{\ell m}d_j\left(\mathbf{z}_{\ell m}\right)\mathbf{v}_{i_{lk}}^H\mathbf{v}_{i_{\ell m}} + \sigma^2}. \tag{6}$$

This is straightforward to prove by utilizing the fact that[3]

$$\mathbf{v}_{i_{lk}}^H\mathbf{\Psi}_j^{-1} = \frac{\tilde{\phi}_{ji_{lk}}}{B}\mathbf{v}_{i_{lk}}^H. \tag{7}$$

According to the orthogonality principle of the MMSE estimator, the estimation error $\tilde{\mathbf{h}}_{jlk} = \mathbf{h}_{jlk} - \hat{\mathbf{h}}_{jlk}$ is independent of $\hat{\mathbf{h}}_{jlk}$ and has zero mean, and the covariance matrix

$$\mathbf{C}_{jlk} = \mathbb{E}\left\{\tilde{\mathbf{h}}_{jlk}\tilde{\mathbf{h}}_{jlk}^H\right\} = d_j\left(\mathbf{z}_{lk}\right)\left(1 - p_{lk}d_j\left(\mathbf{z}_{lk}\right)\tilde{\phi}_{ji_{lk}}\right)\mathbf{I}_M. \quad (8)$$

Finally, notice that also $\hat{\mathbf{h}}_{\mathcal{V},ji}$ is a zero-mean complex Gaussian vector and its covariance matrix is $\mathbb{E}\left\{\hat{\mathbf{h}}_{\mathcal{V},ji}\hat{\mathbf{h}}_{\mathcal{V},ji}^H\right\} = \tilde{\phi}_{ji}\mathbf{I}_M$.

2.2 Uplink M-MMSE detector

Based on the channel estimates, we will now derive the optimal M-MMSE detector in the uplink. After the uplink channel estimation, during the uplink payload data transmission phase, the received signal $\mathbf{y}_j \in \mathbb{C}^{M\times 1}$ at BS j is

$$\mathbf{y}_j = \sum_{l\in\mathcal{L}}\sum_{k=1}^{K}\sqrt{\tau_{lk}}\mathbf{h}_{jlk}x_{lk} + \mathbf{n}_j, \quad (9)$$

where τ_{lk} is the transmit power of the payload data from user k in cell l, $x_{lk}\sim\mathcal{CN}(0,1)$ is the transmitted signal from a Gaussian codebook, and $\mathbf{n}_j\sim\mathcal{CN}(\mathbf{0},\sigma^2\mathbf{I}_M)$ is AWGN. Denoting the linear detector used by BS j for an arbitrary user k in its cell as $\mathbf{g}_{jk}\in\mathbb{C}^M$, the detected signal $\hat{x}_{jk}$ is

$$\hat{x}_{jk} = \mathbf{g}_{jk}^H\mathbf{y}_j = \sqrt{\tau_{jk}}\mathbf{g}_{jk}^H\mathbf{h}_{jjk}x_{jk}$$
$$+ \mathbf{g}_{jk}^H\sum_{(l,m)\neq(j,k)}\sqrt{\tau_{lm}}\mathbf{h}_{jlm}x_{lm} + \mathbf{g}_{jk}^H\mathbf{n}_j. \quad (10)$$

By using (10), the following achievable ergodic SE can be achieved for this user [12]:

$$R_{jk}^{\mathrm{ul}} = \zeta^{\mathrm{ul}}\left(1 - \frac{B}{S}\right)\mathbb{E}_{\{\hat{\mathbf{h}}_{(j)}\}}\left\{\log_2\left(1 + \eta_{jk}^{\mathrm{ul}}\right)\right\}, \quad (11)$$

where $\mathbb{E}_{\{\hat{\mathbf{h}}_{(j)}\}}$ denotes the expectation with respect to all the channel estimates obtained at BS j, the instantaneous effective SINR η_{jk}^{ul} is

$$\eta_{jk}^{\mathrm{ul}} = \frac{\tau_{jk}\mathbf{g}_{jk}^H\hat{\mathbf{h}}_{jjk}\hat{\mathbf{h}}_{jjk}^H\mathbf{g}_{jk}}{\mathbf{g}_{jk}^H\left(\tau_{jk}\mathbf{C}_{jjk} + \displaystyle\sum_{(l,m)\neq(j,k)}\tau_{lm}\left(\hat{\mathbf{h}}_{jlm}\hat{\mathbf{h}}_{jlm}^H + \mathbf{C}_{jlm}\right) + \sigma^2\mathbf{I}_M\right)\mathbf{g}_{jk}}. \quad (12)$$

Recall that $\mathbf{C}_{jjk}$ and $\mathbf{C}_{jlk}$ are estimation error covariance matrices, defined in (8). Note that R_{jk}^{ul} is a lower bound on the uplink ergodic capacity.

The uplink SINR in (12) has the form of a generalized Rayleigh quotient. Therefore, a new M-MMSE detector can be derived to maximize this instantaneous SINR for given channel estimates:

$$\mathbf{g}_{jk}^{\mathrm{M-MMSE}} = \left(\hat{\mathbf{H}}_{\mathcal{V},j}\mathbf{\Lambda}_j\hat{\mathbf{H}}_{\mathcal{V},j}^H + \left(\sigma^2 + \varphi_j\right)\mathbf{I}_M\right)^{-1}\hat{\mathbf{h}}_{jjk}, \quad (13)$$

where $\mathbf{\Lambda}_j = \sum\limits_{l\in L}\sum\limits_{k=1}^{K}\tau_{lk}p_{lk}d_j^2(\mathbf{z}_{lk})\mathbf{e}_{i_{lk}}\mathbf{e}_{i_{lk}}^H \in \mathbb{C}^{B\times B}$ is a diagonal matrix, and its ith diagonal element λ_{ji} depends on

the large-scale fading, the pilot and payload power of the users that use the ith pilot sequence in $\mathcal{V}$. The scalar φ_j is defined as

$$\varphi_j = \sum_{l\in\mathcal{L}}\sum_{k=1}^{K}\tau_{lk}d_j(\mathbf{z}_{lk})\left(1 - p_{lk}d_j(\mathbf{z}_{lk})\tilde{\phi}_{ji_{lk}}\right), \quad (14)$$

where $\tilde{\phi}_{ji_{lk}}$ is defined in (6). As the name suggests, $\tau_{jk}\mathbf{g}_{jk}^{\mathrm{M-MMSE}}$ also minimizes the mean square error (MSE) in estimating x_{jk} [22], $\mathbb{E}\left\{\left|\hat{x}_{jk} - x_{jk}\right|^2\big|\hat{\mathbf{h}}_{(j)}\right\}$.

Remark 1 *To elaborate the advantages of our M-MMSE scheme, we compare it with the related S-MMSE detector from [12, 15, 16], which is defined as*

$$\mathbf{g}_{jk}^{\mathrm{S-MMSE}} = \left(\sum_{m=1}^{K}\tau_{jm}\hat{\mathbf{h}}_{jjm}\hat{\mathbf{h}}_{jjm}^H + \mathbf{Z}_j + \sigma^2\mathbf{I}_M\right)^{-1}\hat{\mathbf{h}}_{jjk}. \quad (15)$$

This detector only contains channel estimates from intracell users, which is why we refer to it as a single-cell scheme. The matrix $\mathbf{Z}_j\in\mathbb{C}^{M\times M}$ is zero in some prior works and otherwise equal to the covariance matrix

$$\mathbf{Z}_j = \mathbb{E}\left\{\sum_{m=1}^{K}\tau_{jm}\tilde{\mathbf{h}}_{jjm}\tilde{\mathbf{h}}_{jjm}^H + \sum_{l\neq j}\sum_{m=1}^{K}\tau_{jm}\mathbf{h}_{jlm}\mathbf{h}_{jlm}^H\right\}. \quad (16)$$

of the intra-cell estimation errors plus the inter-cell interference. When $\mathbf{Z}_j$ in (16) is used, $\tau_{jk}\mathbf{g}_{jk}^{\mathrm{S-MMSE}}$ minimizes the MSE $\mathbb{E}\{|\hat{x}_{jk} - x_{jk}|^2\big|\hat{\mathbf{h}}_{jj1},\ldots,\hat{\mathbf{h}}_{jjK}\}$ under the assumption that only estimates of the intra-cell channels are available, but we stress that this is a limiting assumption in multi-cell scenarios since also the inter-cell channels can be estimated without any additional pilot overhead. As we will show numerically in Section 5, the benefit of the M-MMSE detector over S-MMSE grows with β, since the estimates of the inter-cell channels then improve, and this allows for more efficient interference suppression.

Remark 2 *The M-MMSE detection vector in (13) involves the inversion of an M-dimensional matrix, but the dimension can be substantially reduced by some matrix*

Li *et al. EURASIP Journal on Wireless Communications and Networking* (2017) 2017:117

Page 5 of 15

algebra. By gathering the K detection vectors in cell j in matrix form, we notice that

$$\left[\mathbf{g}_{j1}^{\text{M−MMSE}}, \ldots, \mathbf{g}_{jK}^{\text{M−MMSE}}\right]$$
$$= \left(\hat{\mathbf{H}}_{\mathcal{V},j}\boldsymbol{\Lambda}_j\hat{\mathbf{H}}_{\mathcal{V},j}^H + (\sigma^2 + \varphi_j)\mathbf{I}_M\right)^{-1}$$
$$\times \hat{\mathbf{H}}_{\mathcal{V},j}\left[\sqrt{p_{j1}}d_j(\mathbf{z}_{j1})\,\mathbf{e}_{i_{j1}}, \ldots, \sqrt{p_{jK}}d_j(\mathbf{z}_{jK})\,\mathbf{e}_{i_{jK}}\right] \quad (17)$$
$$= \hat{\mathbf{H}}_{\mathcal{V},j}\left(\boldsymbol{\Lambda}_j\hat{\mathbf{H}}_{\mathcal{V},j}^H\hat{\mathbf{H}}_{\mathcal{V},j} + (\sigma^2 + \varphi_j)\mathbf{I}_B\right)^{-1}$$
$$\times \left[\sqrt{p_{j1}}d_j(\mathbf{z}_{j1})\,\mathbf{e}_{i_{j1}}, \ldots, \sqrt{p_{jK}}d_j(\mathbf{z}_{jK})\,\mathbf{e}_{i_{jK}}\right],$$

by exploiting the fact that $(\mathbf{C}_1\mathbf{C}_2 + \mathbf{I})^{-1}\mathbf{C}_1 = \mathbf{C}_1(\mathbf{C}_2\mathbf{C}_1 + \mathbf{I})^{-1}$ *for any matrices* $\mathbf{C}_1, \mathbf{C}_1$ *of compatible dimensions. Hence, only a B-dimensional matrix needs to be inverted and only once per cell and not once per user. The computation of the M-MMSE detectors in a cell requires approximately* $\frac{3}{2}B^2M$ *complex multiplications. This is greater than with the S-MMSE detector, which after similar matrix algebra requires the inversion of a K-dimensional matrix, and thus,* $\frac{3}{2}K^2M$ *complex multiplications are required.[4] In summary, the increased complexity compared to S-MMSE is about* $\frac{3}{2}\left(\beta^2 - 1\right)K^2M$ *complex multiplications. Since in massive MIMO systems* $M \gg K$ *is often assumed, the complexity increase is not a big issue when K is small or moderate, particularly since the computational efficiency of digital hardware grows rapidly and is not expected to be a bottleneck in the future. One way to reduce the complexity is to check which of the diagonal elements of* $\boldsymbol{\Lambda}_j$ *are below a certain threshold and put these values to zero, to effectively reduce the matrix to be inverted in the M-MMSE expression. This approximation can significantly reduce the complexity if only a few of the* $B - K$ *pilots that belong exclusively to other cells are used by users that cause strong interference. Note that the M-MMSE scheme can be seen as a CoMP coordinated beamforming scheme, but since there is no signaling between the BSs (BS j estimates* $\hat{\mathbf{H}}_{\mathcal{V},j}$ *from pilots), the M-MMSE scheme is fully scalable.*

2.3 Downlink M-MMSE precoder

Next, we will propose a new M-MMSE precoder for the downlink. During the downlink payload data transmission, the received signal at user k in cell j is

$$y_{jk} = \sum_{l\in\mathcal{L}}\mathbf{h}_{ljk}^H\sum_{m=1}^K\sqrt{\varrho_{lm}}\mathbf{w}_{lm}s_{lm} + n_{jk}, \quad (18)$$

where $\mathbf{w}_{lm} \in \mathbb{C}^{M\times 1}$ is the precoder used by BS l for user m in its cell, $s_{lm} \sim \mathcal{CN}(0,1)$ is the payload data symbol for user m in cell l, ϱ_{lm} is the corresponding downlink transmit power, and $n_{jk} \sim \mathcal{CN}\left(0,\sigma^2\right)$ is AWGN.

Recently, an uplink-downlink duality for massive MIMO systems was established in [19] which proves that the uplink SEs can be achieved also in the downlink if each

downlink precoder is a scaled version of the corresponding uplink detector and the downlink transmit power is selected properly. Since the M-MMSE detector proposed in the Section 2.2 is the state-of-the-art uplink scheme, we apply the same methodology for downlink precoding. The downlink M-MMSE precoder is constructed as

$$\mathbf{w}_{jk}^{\text{M−MMSE}} = \frac{\mathbf{g}_{jk}^{\text{M−MMSE}}}{\sqrt{\gamma_{jk}}}, \quad (19)$$

where $\gamma_{jk} = \mathbb{E}\left\{\left\|\mathbf{g}_{jk}^{\text{M−MMSE}}\right\|^2\right\}$ normalizes the average transmit power for the user k in cell j to $\mathbb{E}\left\{\left\|\sqrt{\varrho_{lm}}\mathbf{w}_{jk}^{\text{M−MMSE}}s_{lm}\right\|^2\right\} = \varrho_{lm}$.

In this paper, we use uplink pilots, but to limit the pilot overhead, there are no downlink pilots and we rely instead on channel hardening. Thus, the users do not know their instantaneous channel realizations. However, they can learn their average equivalent channels, $\sqrt{\varrho_{jk}}\mathbb{E}_{\{\mathbf{h}\}}\left\{\mathbf{h}_{jjk}^H\mathbf{w}_{jk}\right\}$, and the total interference variance. Then, the received signal y_{jk} in (18) can be rewritten as

$$y_{jk} = \sqrt{\varrho_{jk}}\mathbb{E}_{\{\mathbf{h}\}}\left\{\mathbf{h}_{jjk}^H\mathbf{w}_{jk}\right\}s_{jk} + \sum_{l\in\mathcal{L}}\mathbf{h}_{ljk}^H\sum_{m=1}^K\sqrt{\varrho_{lm}}\mathbf{w}_{lm}s_{lm}$$
$$- \sqrt{\varrho_{jk}}\mathbb{E}_{\{\mathbf{h}\}}\left\{\mathbf{h}_{jjk}^H\mathbf{w}_{jk}\right\}s_{jk} + n_{jk}. \quad (20)$$

Consequently, a downlink SE

$$R_{jk}^{\text{dl}} = \zeta^{\text{dl}}\left(1 - \frac{B}{S}\right)\log_2\left(1 + \eta_{jk}^{\text{dl}}\right) \quad (21)$$

can be achieved for user k in cell l [12, 19], where η_{jk}^{dl} is

$$\eta_{jk}^{\text{dl}} = \frac{\varrho_{jk}\left|\mathbb{E}_{\{\mathbf{h}\}}\left\{\mathbf{h}_{jjk}^H\mathbf{w}_{jk}\right\}\right|^2}{\sum\limits_{l\in\mathcal{L}}\sum\limits_{m=1}^K\varrho_{lm}\mathbb{E}_{\{\mathbf{h}\}}\left\{\left|\mathbf{h}_{ljk}^H\mathbf{w}_{lm}\right|^2\right\} - \varrho_{jk}\left|\mathbb{E}_{\{\mathbf{h}\}}\left\{\mathbf{h}_{jjk}^H\mathbf{w}_{jk}\right\}\right|^2 + \sigma^2}. \quad (22)$$

This downlink SINR holds for any linear precoding scheme, and we omit the superscript "M-MMSE" of $\mathbf{w}_{jk}$ for brevity. By treating $\sqrt{\varrho_{jk}}\mathbb{E}_{\{\mathbf{h}\}}\left\{\mathbf{h}_{jjk}^H\mathbf{w}_{jk}\right\}$ as the true channel, and the last three term in (20) as uncorrelated Gaussian noise, the user applies semi-coherent symbol detection and achieves the effective SINR in (22).[5] Thus, R_{jk}^{dl} is a lower bound on the downlink ergodic capacity.

By utilizing all the available estimated directions, the M-MMSE precoder can suppress intra-cell interference and also reduce the interference caused to other cells. Thus, a higher SINR is expected by our precoder than conventional single-cell precoders, at least if an appropriate power control is applied [19]. In [20], the authors also

proposed a M-MMSE precoder, but it does not account for arbitrary or optimized pilot allocation. Moreover, no closed-form performance expression is provided in [20], which makes it cumbersome to analyze the performance and optimize the power control.

Looking jointly at the uplink and downlink, the ergodic achievable SE for user k in cell j is

$$R_{jk} = \left(1 - \frac{B}{S}\right)\left(\zeta^{\mathrm{ul}}\mathbb{E}_{\{\hat{\mathbf{h}}_{(j)}\}}\left\{\log_2\left(1 + \eta_{jk}^{\mathrm{ul}}\right)\right\} + \zeta^{\mathrm{dl}}\log_2\left(1 + \eta_{jk}^{\mathrm{dl}}\right)\right). \quad (23)$$

3 Asymptotic analysis

In this section, performance analysis is conducted for the proposed M-MMSE scheme. Since the uplink SINR in (12) depends on the stochastic channel estimates in each block, the uplink SE in (11) cannot be computed in closed form. Therefore, a deterministic equivalent expression for the SINR is computed instead which is tight in the large-system limit. A large-scale approximation of the downlink SINR is also provided. The large-system limit is considered, where M and K go to infinity while keeping K/M finite and non-zero. In what follows, the notation $M \to \infty$ refers to $K, M \to \infty$ such that $K/M \to c \in (0, \infty)$. Hence, $B/M \to \beta c$. The results should be understood in the way that, for each set of system dimension parameters M, K, and B, we provide large-scale approximative expressions for the uplink SINR and downlink SINR, and the expressions are tight as M, K, and B grow large. The main feature is that they are deterministic and can be computed efficiently without the need for time-consuming Monte Carlo simulations. Almost sure convergence of a stochastic sequence is denoted by $\xrightarrow[M\to\infty]{a.s.}$, and $\xrightarrow[M\to\infty]{}$ denotes convergence of a deterministic sequence.

Before we continue with our performance analysis, a useful theorem from large random matrix theory is first recalled.

Theorem 1 ([23]) *Let* $\mathbf{T} = \mathrm{diag}\{t_1, \ldots, t_B\} \in \mathbb{R}^{B \times B}$ *be deterministic with* $t_b \geq 0$ ($b = 1, \ldots, B$) *and* $\mathbf{H} \in \mathbb{C}^{M \times B}$ *be random with independent column vectors* $\mathbf{h}_b \sim \mathcal{CN}\left(\mathbf{0}, \frac{1}{M}\mathbf{I}_b\right)$. *Assume that* $\frac{B}{M} \to \beta c \in (0, \infty)$, *then for any* $\rho > 0$,

$$\frac{1}{M}\mathrm{tr}\left(\left(\mathbf{HTH}^H + \rho\mathbf{I}_M\right)^{-1}\right) - m_o(\rho) \xrightarrow[M\to\infty]{a.s.} 0, \quad (24)$$

where $m_o(\rho)$ *is the solution of the following equation:*

$$m_o(\rho) = \left(\rho + \frac{\beta c}{B}\sum_{b=1}^{B}\frac{t_b}{1 + t_b m_o(\rho)}\right)^{-1}. \quad (25)$$

Based on Theorem 1, we obtain the following Theorem 2 which is useful in our analysis.

Theorem 2 *Under the same conditions on* $\mathbf{T}$ *and* $\mathbf{H}$ *as in Theorem* 1, *for any* $\rho > 0$,

$$\frac{1}{M}\mathrm{tr}\left(\left(\mathbf{HTH}^H + \rho\mathbf{I}_M\right)^{-1}\left(\mathbf{HTH}^H + \rho\mathbf{I}_M\right)^{-1}\right) - m_o'(\rho) \xrightarrow[M\to\infty]{a.s.} 0, \quad (26)$$

where $m_o'(\rho)$ *is defined as*

$$m_o'(\rho) = m_o^2(\rho)\left(1 - m_o^2(\rho)\frac{\beta c}{B}\sum_{b=1}^{B}\frac{t_b}{(1 + t_b m_o(\rho))^2}\right)^{-1}, \quad (27)$$

and $m_o(\rho)$ *is defined in Theorem* 1.

3.1 Large-scale approximations of the SINRs with the M-MMSE scheme

Next, we derive the deterministic equivalent $\bar{\eta}_{jk}^{\mathrm{ul}}$ of η_{jk}^{ul} with the M-MMSE detector, and the large-scale approximation $\bar{\eta}_{jk}^{\mathrm{dl}}$ of η_{jk}^{dl} with the M-MMSE precoder, such that

$$\bar{\eta}_{jk}^{\mathrm{ul}} - \eta_{jk}^{\mathrm{ul}} \xrightarrow[M\to\infty]{a.s.} 0, \quad \bar{\eta}_{jk}^{\mathrm{dl}} - \eta_{jk}^{\mathrm{dl}} \xrightarrow[M\to\infty]{} 0. \quad (28)$$

Theorem 3 *For the uplink M-MMSE detector in* (13), *we have* $\bar{\eta}_{jk}^{\mathrm{ul}} - \eta_{jk}^{\mathrm{ul}} \xrightarrow[M\to\infty]{a.s.} 0$, *where* $\bar{\eta}_{jk}^{\mathrm{ul}}$ *is given by*

$$\bar{\eta}_{jk}^{\mathrm{ul}} = \frac{\tau_{jk}p_{jk}d_j^2(\mathbf{z}_{jk})\tilde{\phi}_{ji_{jk}}\frac{\delta_j^2}{\vartheta_j}}{\tilde{\phi}_{ji_{jk}}\frac{\delta_j^2}{\vartheta_j}\sum\limits_{(l,m)\neq(j,k),i_{lm}=i_{jk}}\tau_{lm}p_{lm}d_j^2(\mathbf{z}_{lm}) + \sum\limits_{i_{lm}\neq i_{jk}}\tau_{lm}d_j(\mathbf{z}_{lm})\frac{\mu_{jlm}}{M} + \frac{\sigma^2}{M}}, \quad (29)$$

where

1. $\delta_j = m_o(\omega)$ *is given by Theorem* 1 *for* $\omega = \frac{\sigma^2 + \varphi_j}{M}$ *and* $\mathbf{T} = \mathbf{\Phi}_j\mathbf{\Lambda}_j$, *with the diagonal matrix* $\mathbf{\Phi}_j = \mathrm{diag}\left\{\tilde{\phi}_{j1}, \ldots, \tilde{\phi}_{jB}\right\}$.

2. $\vartheta_j = m_o'(\omega)$ *is given by Theorem* 2 *for* $\omega = \frac{\sigma^2 + \varphi_j}{M}$, $\mathbf{T} = \mathbf{\Phi}_j\mathbf{\Lambda}_j$ *with* $\mathbf{\Phi}_j = \mathrm{diag}\left\{\tilde{\phi}_{j1}, \ldots, \tilde{\phi}_{jB}\right\}$.

3. $\mu_{jlm} = 1 - p_{lm}d_j(\mathbf{z}_{lm})\tilde{\phi}_{ji_{lm}}\frac{\left(1 + \lambda_{ji_{lm}}\tilde{\phi}_{ji_{lm}}\delta_j\right)^2 - 1}{\left(1 + \lambda_{ji_{lm}}\tilde{\phi}_{ji_{lm}}\delta_j\right)^2}$.

Proof: See Appendix 1. □

The $\bar{\eta}_{jk}^{\mathrm{ul}}$ above not only provides a tight SINR approximation but also shows how the signal, the interference, and the noise change as M grows large. The first term of the denominator represents the interference from the pilot-sharing users, i.e., those users with $i_{lm} = i_{jk}$. This term is at the same order of magnitude as the signal power (notice the δ_{jk}^2 in both terms), since the estimated channels of these users are parallel with the target user.

Li *et al. EURASIP Journal on Wireless Communications and Networking* (2017) 2017:117

Page 7 of 15

The second term of the denominator is the interference from the non-pilot-sharing users, i.e., those users with $i_{lm} \neq i_{jk}$. Since their estimated channels are independent of the channel of the target user, their interference decreases and goes to zero as $M \to \infty$. So does the third term which represents the noise. Thus, only the signal and the interference from the pilot-sharing users remain as M grows, which is referred to as the pilot contamination effect [7–10].

Next, we provide the large-scale SINR approximation for the downlink M-MMSE precoder.

Theorem 4 *For the downlink M-MMSE precoder in (19), we have $\bar{\eta}_{jk}^{\mathrm{dl}} - \eta_{jk}^{\mathrm{dl}} \xrightarrow[M \to \infty]{} 0$, where $\bar{\eta}_{jk}^{\mathrm{dl}}$ is given by*

$$\bar{\eta}_{jk}^{\mathrm{dl}} = \frac{\varrho_{jk} p_{jk} d_j^2\left(\mathbf{z}_{jk}\right) \tilde{\phi}_{ji_{jk}} \frac{\delta_j^2}{\vartheta_j}}{p_{jk} \displaystyle\sum_{(l,m)\neq(j,k), i_{lm}=i_{jk}} \varrho_{lm} d_l^2\left(\mathbf{z}_{jk}\right) \tilde{\phi}_{li_{lm}} \frac{\delta_l^2}{\vartheta_l} + \displaystyle\sum_{i_{lm}\neq i_{jk}} \varrho_{lm} d_l\left(\mathbf{z}_{jk}\right) \frac{\mu_{ljk}}{M} + \frac{\sigma^2}{M}},$$

(30)

where δ_l, μ_{ljk}, and ϑ_l are given in Theorem 3.

Proof : See Appendix 2. □

By utilizing Theorems 3 and 4, the ergodic SEs R_{jk}^{ul} in (11) and R_{jk}^{dl} in (21), after dropping the prelog factor $\left(1 - \frac{B}{S}\right)$, converge to $\bar{R}_{jk}^{\mathrm{ul}} = \log_2\left(1 + \bar{\eta}_{jk}^{\mathrm{ul}}\right)$ and $\bar{R}_{jk}^{\mathrm{dl}} = \log_2\left(1 + \bar{\eta}_{jk}^{\mathrm{dl}}\right)$ in the large-system limit, respectively. Therefore, a large-scale approximation of the joint ergodic SE in (23) is provided by $\left(1 - \frac{B}{S}\right)\left(\zeta^{\mathrm{ul}} \bar{R}_{jk}^{\mathrm{ul}} + \zeta^{\mathrm{dl}} \bar{R}_{jk}^{\mathrm{dl}}\right)$. This approximation is easy to compute and only depends on the long-term parameters: large-scale fading, power control, and pilot allocation. As shown in Section 5, this approximation is very accurate even for small-system dimensions.

3.2 Uplink-downlink duality

It is pointed out in [19] that when the precoder is a scaled version of the detector, the same per-user SEs as in the uplink can be achieved in the downlink by properly selecting the downlink payload power. We establish this uplink-downlink duality also for our M-MMSE scheme, using the large-scale SINR approximations given by Theorem 3 and Theorem 4.

Theorem 5 *For the proposed M-MMSE scheme, if $\bar{\eta}_{jk}^{\mathrm{ul}}$ in (29) is achievable in the uplink for user k in cell j, then a downlink power control policy $\{\varrho_{jk}\}$ can be obtained by transforming the corresponding uplink power $\{\tau_{jk}\}$, such that the total transmitted power remains the same, i.e.,*

$\sum_{j \in \mathcal{L}} \sum_{k=1}^{K} \tau_{jk} = \sum_{j \in \mathcal{L}} \sum_{k=1}^{K} \varrho_{jk}$, *and that the same SE is achieved in the downlink, i.e., $\bar{\eta}_{jk}^{\mathrm{dl}} = \bar{\eta}_{jk}^{\mathrm{ul}}$. The transformation is*

$$\boldsymbol{\varrho} = \left(\mathbf{D} - \boldsymbol{\Psi}\mathbf{F}^T\right)^{-1} (\mathbf{D} - \mathbf{E}\mathbf{F}) \boldsymbol{\tau},$$

(31)

where $\boldsymbol{\tau} = [\tau_{11}, \ldots, \tau_{LK}]^T \in \mathbb{R}^{LK \times 1}$, $\boldsymbol{\varrho} = [\varrho_{11}, \ldots, \varrho_{LK}]^T \in \mathbb{R}^{LK \times 1}$, and $\mathbf{E} = \mathrm{diag}\left\{\bar{\eta}_{11}^{\mathrm{ul}}, \ldots, \bar{\eta}_{LK}^{\mathrm{ul}}\right\} \in \mathbb{R}^{LK \times LK}$. The matrix $\mathbf{F} \in \mathbb{R}^{LK \times LK}$ and the diagonal matrix $\mathbf{D} \in \mathbb{R}^{LK \times LK}$ are defined as

$$\mathbf{F}_{u,v} = \begin{cases} \frac{\delta_j^2 p_{lm} d_j^2(\mathbf{z}_{lm})}{\vartheta_j}, & \text{if } i_{lm} = i_{jk}, \\ \frac{d_j(\mathbf{z}_{lm})\mu_{jlm}}{M}, & \text{if } i_{lm} \neq i_{jk}, \end{cases} \quad \mathbf{D} = \mathrm{diag}\,(\mathbf{F}),$$

(32)

where $u = k + (j-1)K$, $v = m + (l-1)K$. The symbol $[\cdot]_{i,j}$ represents the element in the ith row and the jth column of the corresponding matrix.

Proof : The proof follows the same lines as the duality proof in [19] and is thus omitted. □

Remark 3 *By utilizing the large-scale SINR approximations, Theorem 5 provides a powerful tool to obtain a judicious downlink power allocation whenever the same SEs are desired in both uplink and downlink. However, a certain level of BS coordination is required for this downlink power control policy. Specifically, LK elements in $\mathbf{E}$, LK elements in $\boldsymbol{\tau}$, and $2KL^2$ elements in $\mathbf{F}$ need to be exchanged ($\mathbf{F}$ can be represented by $2KL^2$ elements from its definition). Therefore, the exchange overhead is $2KL(L+1)$ elements. Fortunately, this overhead is acceptable since the exchanged elements are long-term statistical parameters.*

4 Iterative power control

Power control for sum SE maximization has been widely studied in cellular networks [13, 24–30]. However, the power control with the M-MMSE scheme is complicated since the detector/precoder depend on the power control parameters and since the SINRs can not be computed in closed form. In this section, we provide a key application of the results from Theorem 3: joint uplink payload power control for sum SE maximization in multi-cell network. Since the downlink payload power can be obtained according to Theorems 4 and 5, the optimized uplink SEs can also be achieved in the downlink.

Define $\boldsymbol{r} = [r_1, \ldots, r_{LK}] = \left[\bar{\eta}_{11}^{\mathrm{ul}}, \ldots, \bar{\eta}_{LK}^{\mathrm{ul}}\right]^T \in \mathbb{R}^{LK \times 1}$ and suppose the uplink pilot powers are given. We want to find the uplink payload powers $\{\tau_{jk}\}$ that maximize the weighted uplink SE. The problem is called $\mathcal{P}$ in the following, where P_{max} is the maximum radiated

Li *et al. EURASIP Journal on Wireless Communications and Networking* (2017) 2017:117

Page 8 of 15

transmit power of each user and $\xi_l > 0$ is the weight for the corresponding user.

$$\mathcal{P}: \quad \underset{\boldsymbol{\tau}}{\text{maximize}} \quad \sum_{l=1}^{LK} \xi_l \log_2 (1 + r_l)$$
$$s.t. \quad 0 \le \tau_l \le P_{max}, \ \forall l.$$

Power control problems for sum SE maximization are strongly NP-hard [31]. Thus, lower bounding of $\log_2(1 + r_l)$ by $\log_2(r_l)$ is often used to construct an approximative problem [32, 33]. This approximative problem can be further turned into a geometric programming (GP) problem for fixed **F** and **D**, by introducing the auxiliary vector **q** with its lth element $q_l \le r_l^{\xi_l}$. The corresponding GP problem is shown as $\mathcal{P}_1$.

$$\mathcal{P}_1: \quad \underset{\boldsymbol{\tau}, \mathbf{q}}{\text{maximize}} \quad \prod_{l=1}^{LK} q_l$$
$$s.t. \quad q_l^{\frac{1}{\xi_l}} \left(\sum_{j=1}^{K} F_{lj}\tau_j + \frac{\sigma^2}{M} \right) \tau_l^{-1} D_{l,l}^{-1} \le 1, \ \forall l,$$
$$0 \le \tau_l \le P_{max}, \ \forall l.$$

It can be solved numerically with the convex optimization toolbox in MATLAB, and a low-complexity fixed point iteration method is also proposed in [33] to solve the problem of the same type. With our notation, the power coefficient τ_l is updated as

$$\tau_l (t + 1) = \min \left\{ \xi_l \Big/ \left(\sum_{j=1}^{LK} \frac{\xi_j F_{j,l} r_j (t)}{D_j \tau_j (t)} \right), P_{max} \right\}, \quad (33)$$

where t is the iteration index in the fixed point algorithm, for $t = 0, 1, \dots$. It is proved in [33] that starting from the initial point $\tau_l(0) = P_{max}$ for all l, the above algorithm converges at a geometric rate to the optimal solution of $\mathcal{P}_1$ (for fixed **F** and **D**).

In our case, however, **F** and **D** are not fixed since δ_j and ϑ_j will change as τ_l changes. Hence, $\mathcal{P}_1$ in our work is not a pure GP. Therefore, Algorithm 1 is proposed to iterate between solving $\mathcal{P}_1$ for fixed **F** and **D** and updating **F** and **D** using the current $\boldsymbol{\tau}$.

The rigorous proof of convergence of $R(t)$ is intractable, since **D** and **F** depend in a very complicated way on the powers τ_{lm} of all users, and we update **D** and **F** after each

---

Algorithm 1 : Approximated Sum SE Maximization Power Control Algorithm

1: Initialize $\boldsymbol{\tau}(0) = P_{max}$ for $t = 0$ and select $\epsilon > 0$. Calculate $\mathbf{F}(0)$, $\mathbf{D}(0)$ and $R(0) = \sum_{l=1}^{LK} \xi_l \log_2(r_l)$ using $\boldsymbol{\tau}(0)$.

2: Let $i = 0$, and $R^{(i)}(t + 1) = R(t)$. Do:
   update $\boldsymbol{\tau}^{(i+1)}(t + 1)$ using $\mathbf{F}(t)$ and $\mathbf{D}(t)$ by (33);
   calculate $R^{(i+1)}(t + 1)$ based on $\boldsymbol{\tau}^{(i+1)}(t + 1)$, $\mathbf{F}(t)$ and $\mathbf{D}(t)$;
   update i with $i + 1$;
   until $\left| R^{(i+1)}(t + 1) - R^{(i)}(t + 1) \right| \le \epsilon$.
   Let $\boldsymbol{\tau}(t+1) = \boldsymbol{\tau}^{(i+1)}(t+1)$ and $R(t+1) = R^{(i+1)}(t+1)$

3: Update $\mathbf{F}(t+1)$ and $\mathbf{D}(t+1)$ with $\boldsymbol{\tau}(t+1)$, and update the time slot index t with $t + 1$.

4: Repeat step $2 - 4$ until $R(t)$ converges.

---

iteration. However, numerical results testify the fast convergence: about five iterations are enough. Therefore, our algorithm can converges to some local optimal solution of $\mathcal{P}_1$, and the involved information exchange overhead is acceptable. Moreover, since only long-term parameters need to be exchanged, the exchange overhead can be spread over time.

5 Simulation results

In this section, we illustrate the analytical contributions by simulation results for a symmetric hexagonal network topology. We apply the classic 19-cell-wrap-around structure to avoid edge effects and guarantee consistent simulated performance for all cells; see Fig. 1. Each hexagonal cell has a radius of $r = 500$ m and is surrounded by six interfering cells in the first tier and 12 in the second tier. To achieve a symmetric pilot allocation in this network, the pilot reuse factor can be $\beta \in \{1, 3, 4, 7\}$. For each pilot reuse policy, the same subset of pilots are allocated to the cells with the same color, and pilots in each cell are allocated randomly to the users.

The user locations are generated independently and uniformly at random in the cells, but the distance between each user and its serving BS is at least $0.14r$. For each
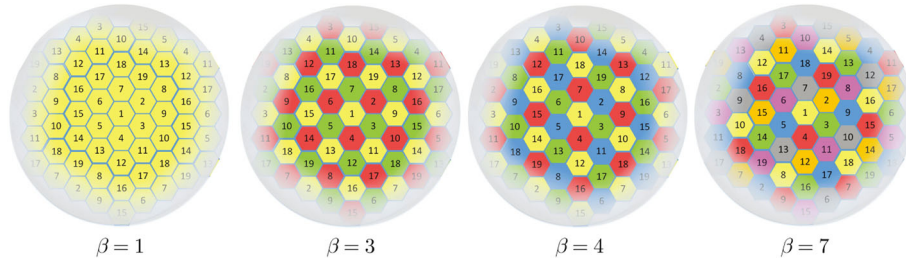

Fig. 1 The 19-cell-wrap-around hexagonal network topology for $\beta = 1$, $\beta = 3$, $\beta = 4$, and $\beta = 7$

user location $\mathbf{z} \in \mathbb{R}^2$, a classic pathloss model is considered, where the variance of the channel attenuation is $d_j(\mathbf{z}) = \frac{C_{(\mathbf{z})}}{\|\mathbf{z}-\mathbf{b}_j\|^\kappa}$. The vector $\mathbf{b}_j \in \mathbb{R}^2$ is the location of the BS in cell j, κ is the pathloss exponent, and $\|\cdot\|$ denotes the Euclidean norm. $C_{(\mathbf{z})} > 0$ is independent shadow fading for some user location $\mathbf{z}$ with $10 \log_{10}\left(C_{(\mathbf{z})}\right) \sim \mathcal{N}\left(0, \sigma_{sf}^2\right)$. In the simulation, we assume $\kappa = 3.7$, $\sigma_{sf}^2 = 5$ and the coherence block length $S = 1000$.[6]
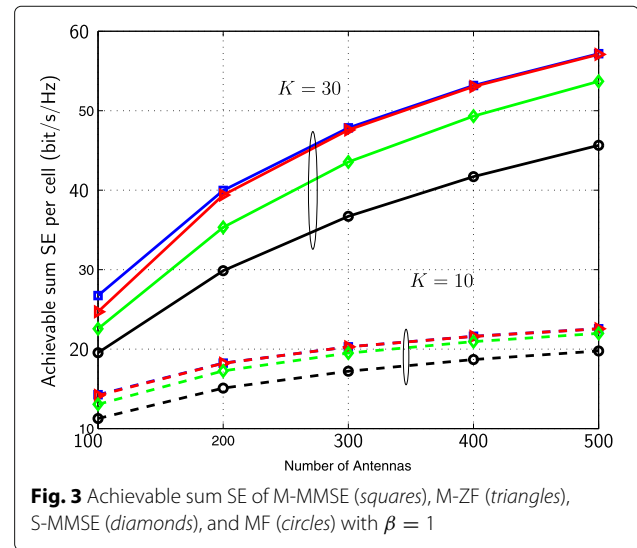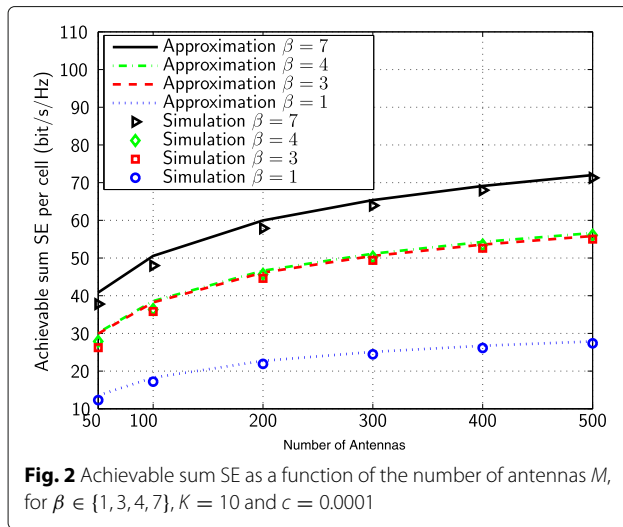
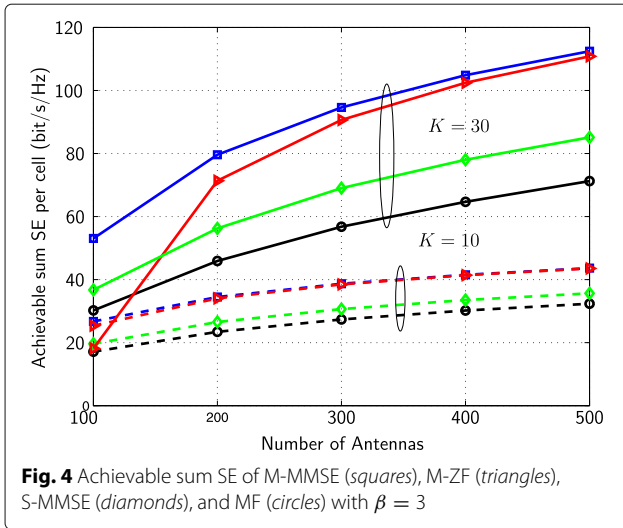### 5.1 Benefits of the proposed M-MMSE scheme

In this subsection, we show the benefits of our M-MMSE scheme over the conventional alternatives. Statistical channel inversion power control is applied to both pilot and uplink payload data, i.e., $p_{lk} = \tau_{lk} = \frac{\rho}{d_l(\mathbf{z}_{lk})}$ [19]. Thus, during the uplink phase, the average effective channel gain between users and their serving BSs is constant: $\mathbb{E}\left\{p_{lk}\|\mathbf{h}_{llk}\|^2\right\} = \mathbb{E}\left\{\tau_{lk}\|\mathbf{h}_{llk}\|^2\right\} = M\rho$. Then, the average uplink SNR per antenna and user at its serving BS is ρ/σ^2. This is a simple but effective policy to avoid near-far blockage and, to some extent, guarantee a uniform user performance in the uplink. For downlink payload data transmission, the transmit power ϱ_{lk} is selected according to Theorem 5 to achieve the same downlink SE at each user as in the uplink. In our simulation, ρ/σ^2 is set to 0 dB to allow for decent channel estimation accuracy, and the time proportions for the uplink and downlink are set to $\zeta^{\mathrm{ul}} = \zeta^{\mathrm{dl}} = \frac{1}{2}$.

To verify the accuracy of the large-scale approximations from Section 3, 10,000 independent Monte-Carlo channel realizations are generated to numerically calculate the joint achievable SE in (23). The numerical results and their approximations from Theorems 3 and 4 are shown in Fig. 2. As shown in the figure, the achievable sum SE increases with β for the considered range of values.

This is because a larger β results in a lower level of pilot contamination, which contributes to a higher channel estimation accuracy, and thereby increases the achievable SE. Moreover, a larger β provides more estimated channel directions in the construction of the M-MMSE scheme; thus, a higher inter-cell interference suppression can be achieved.[7] Figure 2 shows that the numerical results and the large-scale approximations match very well, even for small M and small K.

To show explicitly the advantages of our M-MMSE scheme, simulation results for the MF scheme from [8], the multi-cell ZF (M-ZF) scheme from [19], and the S-MMSE scheme from [12, 15, 16] (and given in (15)) are provided for comparison. The same downlink power allocation from Theorem 5 and normalization based on (19) are applied for all precoders. Notice that $M - \beta K > 0$ is needed for the M-ZF scheme; thus, the minimum value of M for M-ZF is $\beta K + 1$. Simulation results are shown in Figs. 3–4 for $\beta = 1$ and $\beta = 3$, respectively. The MF scheme always achieves the lowest performance since it does not suppress any interference. Compared to S-MMSE, our proposed M-MMSE always achieves a higher sum SE, and the advantage becomes more significant as β and/or K increases. For $\beta = 3$ and $M = 200$, the SEs of M-MMSE are 30% and 42% higher than those of S-MMSE for $K = 10$ and $K = 30$, respectively. For $\beta = 7$, the gains increase to 42% and 82% for $K = 10$ and $K = 30$, respectively (the related figure is omitted for brevity). The higher performance gain at larger K or β comes from the fact that more channel directions can be learned and utilized for interference suppression by M-MMSE, while S-MMSE always uses K directions regardless of β. The advantage of M-MMSE over M-ZF is minor for small β and small K, but the gain becomes notable as β and K grow. Since the complexity of our M-MMSE scheme is the same as for M-ZF,

Fig. 2 Achievable sum SE as a function of the number of antennas M, for $\beta \in \{1, 3, 4, 7\}$, $K = 10$ and $c = 0.0001$

Fig. 3 Achievable sum SE of M-MMSE (*squares*), M-ZF (*triangles*), S-MMSE (*diamonds*), and MF (*circles*) with $\beta = 1$

Fig. 4 Achievable sum SE of M-MMSE (*squares*), M-ZF (*triangles*), S-MMSE (*diamonds*), and MF (*circles*) with $\beta = 3$

and M-ZF can sometimes achieve very low SE for small M, in general, our scheme is the a better choice if high system SE is desirable.

Since the optimal pilot reuse factor may be different for different schemes, we further compare the performance when each scheme uses its own separately optimized $\beta_o \in \{1, 3, 4, 7\}$. The results are shown in Fig. 5. We notice that our M-MMSE scheme prefers a higher pilot reuse policy $\beta_o = 7$ while S-MMSE prefers $\beta_o = 3$. Moreover, our M-MMSE achieves a significantly higher performance than S-MMSE, also when considering separately optimized pilot reuse factors.

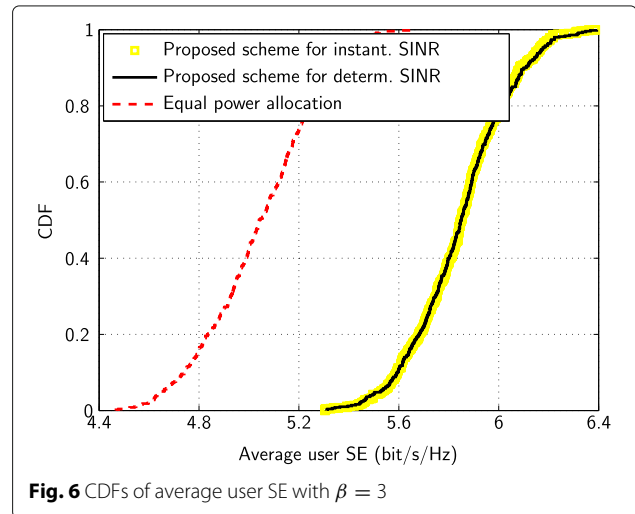### 5.2 Effectiveness of joint power control
In this subsection, the effectiveness of the proposed power control scheme is testified. Statistical power control $p_{lk} =$
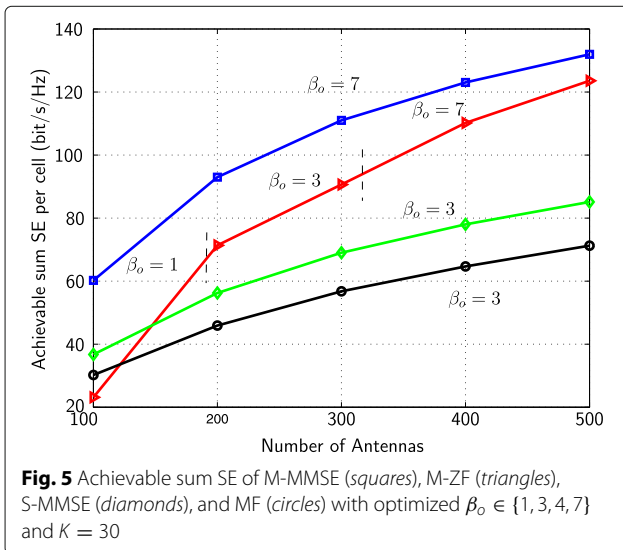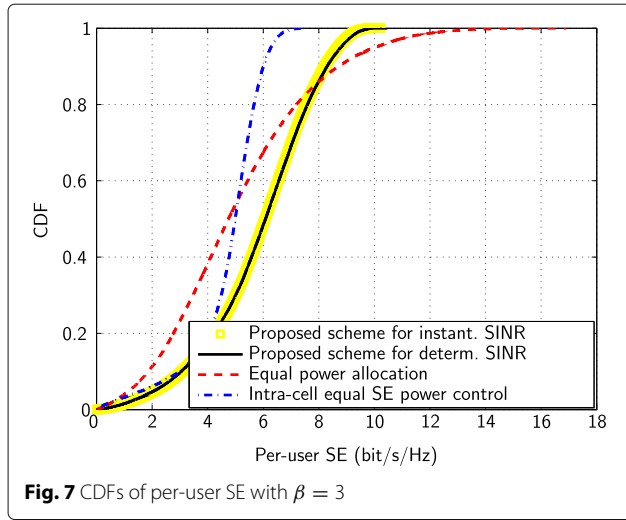
$\frac{\rho}{d_l(\mathbf{z}_{lk})}$ is still applied for pilots, while the uplink payload power τ_{jk} is optimized. ρ/σ^2 is still set to 0 dB, and the maximal transmit power P_{max} in $\mathcal{P}$ is selected as in Section 5.1. Results for equal maximum power allocation (i.e., $\tau_{lk} = P_{max}$) are provided as a baseline. We also apply Algorithm 1 to the instantaneous SINR in (12) for comparison. The following results are obtained for $M = 300$ and $K = 10$. After generating user locations and shadow fading realizations, the 9 users with the worst channel conditions in the whole network are dropped to provide 95% coverage.

We first consider the average per-user SE which is calculated as the network sum SE divided by the number of served users. The cumulative distribution functions (CDFs) over user locations are shown in Fig. 6 for $\beta = 3$. As seen from the figure, the CDF curve with long-term power control based on Algorithm 1 coincides with those with short-term power control optimized for the instantaneous SINR at every coherence block, which validates that there is negligible loss associated with our power control based on the large-scale SINR approximation. Furthermore, compared with the equal power allocation policy, the average user SEs can be significantly improved by our power control scheme. At the 50th percentile, 16% increase can be achieved by our scheme.

We also analyze how the per-user SE at different parts of the cells is affected by our power control. Results are provided for the power control proposed in [35], which tries to provide equal SE for users in the same cell so that, to some extent, intra-cell user fairness is guaranteed. The CDF of the per-user SE is shown in Fig. 7. With our algorithm, in contrast to equal power allocation, the majority of the users can enjoy higher SEs at the cost of a small degradation for the users with the strongest channels. This is because our algorithm assigns

Fig. 5 Achievable sum SE of M-MMSE (*squares*), M-ZF (*triangles*), S-MMSE (*diamonds*), and MF (*circles*) with optimized $\beta_o \in \{1, 3, 4, 7\}$ and $K = 30$

Fig. 6 CDFs of average user SE with $\beta = 3$

Li *et al. EURASIP Journal on Wireless Communications and Networking* (2017) 2017:117

Page 11 of 15

Fig. 7 CDFs of per-user SE with $\beta = 3$

lower transmit powers to the few users close to the center of the cell and higher powers to the many users further away, so that the interference caused by the former to the latter is reduced. In this way, our algorithm suppresses interference. Compared with the power control from [35], our algorithm provides essentially the same SE for the weakest users, while pushing the SE of the majority of the users to higher values. Despite the larger SE variations, the proposed power control brings a better type of user fairness than the scheme from [35] since the strong users get higher SEs without degrading for the weakest ones.

6 Conclusions

In this paper, a new state-of-the-art M-MMSE scheme is proposed, which includes an uplink M-MMSE detector and a downlink M-MMSE precoder. It brings very promising sum SE gains over S-MMSE and other single-cell schemes by actively suppressing both intra-cell and inter-cell interference. Since imperfect CSI is accounted for in our scheme, the gains obtained by our scheme are likely to be achievable in practical systems. Furthermore, large-scale approximations of the uplink and downlink SINRs are derived for the proposed M-MMSE scheme. The approximations are very accurate even for small system dimensions and are easy to compute since they only depend on long-term statistics. Hence, the expressions can be utilized for efficient performance analysis, without the need for Monte-Carlo simulations. The SINR approximations can further be used for power control design, and a low-complexity power control algorithm for sum SE maximization is proposed. The proposed algorithm brings a notable sum SE gain and also provides good user fairness compared to the equal power allocation policy. Since the SINR approximations depend only

on long-term statistics, the complexity of the algorithm can be spread over a long time period.

Endnotes

[1] These schemes have several names in the literature: MF is also known as maximum ratio combining/transmission; ZF is also known as channel-inversion; and regularized ZF (RZF) is a simple variation on S-MMSE.

[2] In practice, only the propagation channels are reciprocal, while the hardware used for uplink and downlink communication is not. This requires reciprocity-calibration of the hardware, but there are many algorithms for this, and the variations are slow so the calibration overhead is negligible [34].

[3] Notice that $\frac{1}{\sqrt{B}}\mathbf{V} = \frac{1}{\sqrt{B}}[\mathbf{v}_1, \ldots, \mathbf{v}_B] \in \mathbb{C}^{B \times B}$ is an orthogonal basis for a B dimensional space. Therefore, a singular value decomposition of $\mathbf{\Psi}_j$ is $\mathbf{\Psi}_j = \frac{1}{B}\mathbf{V}\mathbf{A}_j\mathbf{V}^H$, where $\mathbf{A}_j$ is a diagonal matrix with its bth element as $a_{jb} = B/\tilde{\phi}_{jb}$. Then, (7) is obtained.

[4] Only multiplications are counted in the complexity comparison, since additions and subtractions have a negligible complexity in comparison.

[5] This method works well in massive MIMO systems due to channel hardening—the effective channel is relatively close to its mean, while the performance loss would be large in a small-scale MIMO system.

[6] This coherence block can, for example, have the dimensions of $T_c = 10$ ms and $W_c = 100$ kHz.

[7] One should notice that K and β cannot be increased indefinitely due to the prelog loss in the achievable SE.

Appendix 1
Proof of Theorem 3
Define $\mathbf{\Sigma}_j = \left(\hat{\mathbf{H}}_{\mathcal{V},j}\mathbf{\Lambda}_j\hat{\mathbf{H}}_{\mathcal{V},j}^H + \left(\sigma^2 + \varphi_j\right)\mathbf{I}_M\right)^{-1}$, then the M-MMSE detector in (13) is $\mathbf{g}_{jk} = \mathbf{\Sigma}_j\hat{\mathbf{h}}_{jjk}$. We omit the superscript "M-MMSE" in the proof for brevity. In the following proof, we use $\asymp$ to denote the almost sure convergence such that $a \asymp b$ represents $a - b \xrightarrow[M \to \infty]{a.s.} 0$. Define

1. $\hat{\mathbf{H}}_{\mathcal{V},jlk} = \left[\hat{\mathbf{h}}_{\mathcal{V},j1}, \ldots, \hat{\mathbf{h}}_{\mathcal{V},j(i_{lk}-1)}, \hat{\mathbf{h}}_{\mathcal{V},j(i_{lk}+1)}, \ldots, \hat{\mathbf{h}}_{\mathcal{V},jB}\right]$,
2. $\mathbf{\Lambda}_{jlk} = \text{diag}\left\{\lambda_{j1}, \ldots \lambda_{j(i_{lk}-1)}, \lambda_{j(i_{lk}+1)}, \ldots, \lambda_{jB}\right\}$,
3. $\mathbf{\Phi}_j = \text{diag}\{\tilde{\phi}_{j1}, \ldots, \tilde{\phi}_{jB}\}$,
4. $\mathbf{\Sigma}_{jjk} = \left(\hat{\mathbf{H}}_{\mathcal{V},jjk}\mathbf{\Lambda}_{jjk}\hat{\mathbf{H}}_{\mathcal{V},jjk}^H + \left(\sigma^2 + \varphi_j\right)\mathbf{I}_M\right)^{-1}$,
5. $\mathbf{\Sigma}_j' = M\mathbf{\Sigma}_j$ and $\mathbf{\Sigma}_{jjk}' = M\mathbf{\Sigma}_{jjk}$,

then we have the following lemma.

Li *et al. EURASIP Journal on Wireless Communications and Networking* (2017) 2017:117

Page 12 of 15

Lemma 1 *Let $\hat{\mathbf{h}}_{jlk}$ and $\tilde{\mathbf{h}}_{jlk}$ denote the MMSE estimate of $\mathbf{h}_{jlk}$ as in* (5) *and its estimation error, respectively, then*

$$\hat{\mathbf{h}}_{jjk}^H \Sigma_j \hat{\mathbf{h}}_{jjk} - \frac{p_{jk} d_j^2 (\mathbf{z}_{jk}) \tilde{\phi}_{ji_{jk}} \delta_j}{1 + \lambda_{ji_{jk}} \tilde{\phi}_{ji_{jk}} \delta_j} \xrightarrow[M\to\infty]{a.s.} 0, \tag{34}$$

$$\hat{\mathbf{h}}_{jjk}^H \Sigma_j \tilde{\mathbf{h}}_{jlm} \xrightarrow[M\to\infty]{a.s.} 0. \tag{35}$$

Proof Let $x = \hat{\mathbf{h}}_{jjk}^H \Sigma_j \hat{\mathbf{h}}_{jjk}$, then

$$x = \hat{\mathbf{h}}_{jjk}^H \left(\Sigma_{jjk}^{-1} + \lambda_{ji_{jk}} \hat{\mathbf{h}}_{\mathcal{V},ji_{jk}} \hat{\mathbf{h}}_{\mathcal{V},ji_{jk}}^H \right)^{-1} \hat{\mathbf{h}}_{jjk} \stackrel{(a)}{=} \frac{p_{jk} d_j^2 (\mathbf{z}_{jk}) \hat{\mathbf{h}}_{\mathcal{V},ji_{jk}}^H \Sigma_{jjk} \hat{\mathbf{h}}_{\mathcal{V},ji_{jk}}}{1 + \lambda_{ji_{jk}} \hat{\mathbf{h}}_{\mathcal{V},ji_{jk}}^H \Sigma_{jjk} \hat{\mathbf{h}}_{\mathcal{V},ji_{jk}}}$$

$$= \frac{\frac{1}{M} p_{jk} d_j^2 (\mathbf{z}_{jk}) \hat{\mathbf{h}}_{\mathcal{V},ji_{jk}}^H \Sigma'_{jjk} \hat{\mathbf{h}}_{\mathcal{V},ji_{jk}}}{1 + \lambda_{ji_{jk}} \frac{1}{M} \hat{\mathbf{h}}_{\mathcal{V},ji_{jk}}^H \Sigma'_{jjk} \hat{\mathbf{h}}_{\mathcal{V},ji_{jk}}} \stackrel{(b)}{\asymp} \frac{\frac{1}{M} p_{jk} d_j^2 (\mathbf{z}_{jk}) \tilde{\phi}_{ji_{jk}} \operatorname{tr} \left(\Sigma'_{jjk} \right)}{1 + \frac{1}{M} \lambda_{ji_{jk}} \tilde{\phi}_{ji_{jk}} \operatorname{tr} \left(\Sigma'_{jjk} \right)}$$

$$\stackrel{(c)}{\asymp} \frac{\frac{1}{M} p_{jk} d_j^2 (\mathbf{z}_{jk}) \tilde{\phi}_{ji_{jk}} \operatorname{tr} \left(\Sigma'_j \right)}{1 + \frac{1}{M} \lambda_{ji_{jk}} \tilde{\phi}_{ji_{jk}} \operatorname{tr} \left(\Sigma'_j \right)} \stackrel{(d)}{=} \frac{p_{jk} d_j^2 (\mathbf{z}_{jk}) \tilde{\phi}_{ji_{jk}} \delta_j}{1 + \lambda_{ji_{jk}} \tilde{\phi}_{ji_{jk}} \delta_j},$$

where (*a*) follows from Lemma 1 in [12] and $\hat{\mathbf{h}}_{jjk} = \sqrt{p_{jk}} d_j(\mathbf{z}_{jk}) \hat{\mathbf{h}}_{\mathcal{V},ji_{jk}}$ and (*b*) follows from Lemma 12 in [36], which can be applied since Σ'_{jjk} has uniformly bounded spectral norm with respect to M, because φ_j scales as K and $\frac{K}{M} > 0$ by assumption; thus, $\frac{\varphi_j}{M} > 0$ for all M. (*c*) follows from Lemma 14.3 in [37]. In step (*d*), we define $\delta_j = m_o \left(\frac{\sigma^2 + \varphi_j}{M} \right)$, which is obtained by Theorem 1 for $\mathbf{T} = \Phi_j \Lambda_j$ and $\rho = \frac{\sigma^2 + \varphi_j}{M}$.

Let $y = \hat{\mathbf{h}}_{jjk}^H \Sigma_j \tilde{\mathbf{h}}_{jlm}$, then

$$y = \hat{\mathbf{h}}_{jjk}^H \left(\Sigma_{jjk}^{-1} + \lambda_{ji_{jk}} \hat{\mathbf{h}}_{\mathcal{V},ji_{jk}} \hat{\mathbf{h}}_{\mathcal{V},ji_{jk}}^H \right)^{-1} \tilde{\mathbf{h}}_{jlm}$$

$$\stackrel{(a)}{=} \frac{\sqrt{p_{jk}} d_j(\mathbf{z}_{jk}) \hat{\mathbf{h}}_{\mathcal{V},ji_{jk}}^H \Sigma_{jjk} \tilde{\mathbf{h}}_{jlm}}{1 + \lambda_{ji_{jk}} \hat{\mathbf{h}}_{\mathcal{V},ji_{jk}}^H \Sigma_{jjk} \hat{\mathbf{h}}_{\mathcal{V},ji_{jk}}}$$

$$\stackrel{(b)}{=} \frac{\sqrt{p_{jk}} d_j(\mathbf{z}_{jk}) \frac{1}{M} \hat{\mathbf{h}}_{\mathcal{V},ji_{jk}}^H \Sigma'_{jjk} \tilde{\mathbf{h}}_{jlm}}{1 + \lambda_{ji_{jk}} \frac{1}{M} \hat{\mathbf{h}}_{\mathcal{V},ji_{jk}}^H \Sigma'_{jjk} \hat{\mathbf{h}}_{\mathcal{V},ji_{jk}}} \stackrel{(c)}{\asymp} 0, \tag{36}$$

where (*a*) and (*b*) follow from Lemma 1 in [12] and Lemma 12 in [36], respectively. □

We use Lemma 1 in the following to determine the asymptotic behavior of each term in (12).

Signal power
Since $\mathbf{g}_{jk}^H \hat{\mathbf{h}}_{jjk} = \hat{\mathbf{h}}_{jjk}^H \Sigma_j \hat{\mathbf{h}}_{jjk}$, then according to Lemma 1, it is obvious that

$$\mathbf{g}_{jk}^H \hat{\mathbf{h}}_{jjk} - \frac{p_{jk} d_j^2 (\mathbf{z}_{jk}) \tilde{\phi}_{ji_{jk}} \delta_j}{1 + \lambda_{ji_{jk}} \tilde{\phi}_{ji_{jk}} \delta_j} \xrightarrow[M\to\infty]{a.s.} 0. \tag{37}$$

By the continuous mapping theorem [38], we further obtain

$$\left| \mathbf{g}_{jk}^H \hat{\mathbf{h}}_{jjk} \right|^2 - \left(\frac{p_{jk} d_j^2 (\mathbf{z}_{jk}) \tilde{\phi}_{ji_{jk}} \delta_j}{1 + \lambda_{ji_{jk}} \tilde{\phi}_{ji_{jk}} \delta_j} \right)^2 \xrightarrow[M\to\infty]{a.s.} 0. \tag{38}$$

Channel uncertainty
According to Lemma 1, $\mathbf{g}_{jk}^H \tilde{\mathbf{h}}_{jjk} = \hat{\mathbf{h}}_{jjk}^H \Sigma_j \tilde{\mathbf{h}}_{jjk} \xrightarrow[M\to\infty]{a.s.} 0$. Thus, by the dominated convergence theorem [39] and the continuous mapping theorem, we have

$$\mathbb{E} \left\{ \tau_{jk} \left| \mathbf{g}_{jk}^H \tilde{\mathbf{h}}_{jjk} \right|^2 \middle| \hat{\mathbf{h}}_{(j)} \right\} \xrightarrow[M\to\infty]{a.s.} 0. \tag{39}$$

Interference power
Since $\mathbf{g}_{jk}^H = \Sigma_j \hat{\mathbf{h}}_{jjk}$, the interference power from user m in cell l is

$$\mathbb{E}_{\{\mathbf{h}\}} \left\{ \left| \mathbf{g}_{jk}^H \mathbf{h}_{jlm} \right|^2 \middle| \hat{\mathbf{h}}_{(j)} \right\} = \mathbb{E} \left\{ \left| \hat{\mathbf{h}}_{jjk}^H \Sigma_j \mathbf{h}_{jlm} \right|^2 \middle| \hat{\mathbf{h}}_{(j)} \right\} \tag{40}$$

The computation depends on which pilots that are used.

$i_{lm} = i_{jk} = i_0$
In this case, user k in cell j use the same pilot sequence as user m in cell j, and there will be coherence pilot contaminated interference. Since

$$\hat{\mathbf{h}}_{jlm} = \sqrt{p_{lm}} d_j(\mathbf{z}_{lm}) \hat{\mathbf{h}}_{\mathcal{V},ji_0} = \sqrt{\frac{p_{lm}}{p_{jk}}} \frac{d_j(\mathbf{z}_{lm})}{d_j(\mathbf{z}_{jk})} \hat{\mathbf{h}}_{jjk}, \tag{41}$$

we have

$$\hat{\mathbf{h}}_{jjk}^H \Sigma_j \mathbf{h}_{jlm} = \sqrt{\frac{p_{lm}}{p_{jk}}} \frac{d_j(\mathbf{z}_{lm})}{d_j(\mathbf{z}_{jk})} \hat{\mathbf{h}}_{jjk}^H \Sigma_j \hat{\mathbf{h}}_{jjk} + \hat{\mathbf{h}}_{jjk}^H \Sigma_j \tilde{\mathbf{h}}_{jlm}$$

$$\stackrel{(a)}{\asymp} d_j(\mathbf{z}_{jk}) d_j(\mathbf{z}_{lm}) \frac{\sqrt{p_{jk} p_{lm}} \tilde{\phi}_{ji_{jk}} \delta_j}{1 + \lambda_{ji_{jk}} \tilde{\phi}_{ji_{jk}} \delta_j}, \tag{42}$$

where in step (*a*) the first term remains and the second term vanishes according to Lemma 1. Indicated by the dominated convergence theorem and the continuous mapping theorem, we have

$$\mathbb{E} \left\{ \left| \hat{\mathbf{h}}_{jjk}^H \Sigma_j \mathbf{h}_{jlm} \right|^2 \middle| \hat{\mathbf{h}}_{(j)} \right\} - d_j^2 (\mathbf{z}_{jk}) d_j^2 (\mathbf{z}_{lm})$$

$$\times \frac{p_{jk} p_{lm} \tilde{\phi}_{ji_{jk}}^2 \delta_j^2}{\left(1 + \lambda_{ji_{jk}} \tilde{\phi}_{ji_{jk}} \delta_j \right)^2} \xrightarrow[M\to\infty]{a.s.} 0. \tag{43}$$

Li *et al. EURASIP Journal on Wireless Communications and Networking* (2017) 2017:117

Page 13 of 15

$i_{lm} \neq i_{jk}$

In this case, two users have different pilots, such that

$$\left| \hat{\mathbf{h}}_{jjk}^H \boldsymbol{\Sigma}_j \mathbf{h}_{jlm} \right|^2$$

$$\overset{(a)}{=} \frac{p_{jk} d_j^2 \left(\mathbf{z}_{jk} \right) \frac{1}{M^2} \hat{\mathbf{h}}_{\mathcal{V}, ji_{jk}}^H \boldsymbol{\Sigma}_{jjk}' \mathbf{h}_{jlm} \mathbf{h}_{jlm}^H \boldsymbol{\Sigma}_{jjk}' \hat{\mathbf{h}}_{\mathcal{V}, ji_{jk}}}{\left(1 + \lambda_{ji_{jk}} \hat{\mathbf{h}}_{\mathcal{V}, ji_{jk}}^H \boldsymbol{\Sigma}_{jjk} \hat{\mathbf{h}}_{\mathcal{V}, ji_{jk}} \right)^2}$$

$$\overset{(b)}{\asymp} \frac{p_{jk} d_j^2 \left(\mathbf{z}_{jk} \right) \tilde{\phi}_{ji_{jk}} \frac{1}{M^2} \mathrm{tr} \left(\boldsymbol{\Sigma}_{jjk}' \mathbf{h}_{jlm} \mathbf{h}_{jlm}^H \boldsymbol{\Sigma}_{jjk}' \right)}{\left(1 + \lambda_{ji_{jk}} \tilde{\phi}_{ji_{jk}} \delta_j \right)^2}$$

$$= \frac{p_{jk} d_j^2 \left(\mathbf{z}_{jk} \right) \tilde{\phi}_{ji_{jk}} \mathrm{tr} \left(\boldsymbol{\Sigma}_{jjk} \mathbf{h}_{jlm} \mathbf{h}_{jlm}^H \boldsymbol{\Sigma}_{jjk} \right)}{\left(1 + \lambda_{ji_{jk}} \tilde{\phi}_{ji_{jk}} \delta_j \right)^2}$$

$$= \frac{p_{jk} d_j^2 \left(\mathbf{z}_{jk} \right) \tilde{\phi}_{ji_{jk}} \mathbf{h}_{jlm}^H \boldsymbol{\Sigma}_{jjk} \boldsymbol{\Sigma}_{jjk} \mathbf{h}_{jlm}}{\left(1 + \lambda_{ji_{jk}} \tilde{\phi}_{ji_{jk}} \delta_j \right)^2}, \qquad (44)$$

where step (*a*) follows from Lemma 1 in [12] and the definition of $\boldsymbol{\Sigma}_{jjk}'$. Step (*b*) follows from Lemma 12 in [36], Lemma 14.3 in [37], and Theorem 1. It remains to obtain a deterministic equivalent of the numerator in (44). Define $\boldsymbol{\Sigma}_{j,jk,lm} = \left(\boldsymbol{\Sigma}_{jjk}^{-1} - \lambda_{ji_{lm}} \hat{\mathbf{h}}_{\mathcal{V}, ji_{lm}} \hat{\mathbf{h}}_{\mathcal{V}, ji_{lm}}^H \right)^{-1}$, then according to Lemma 2 in [12] we have

$$\boldsymbol{\Sigma}_{jjk} = \boldsymbol{\Sigma}_{j,jk,lm} - \frac{\boldsymbol{\Sigma}_{j,jk,lm} \lambda_{ji_{lm}} \hat{\mathbf{h}}_{\mathcal{V}, ji_{lm}} \hat{\mathbf{h}}_{\mathcal{V}, ji_{lm}}^H \boldsymbol{\Sigma}_{j,jk,lm}}{1 + \lambda_{ji_{lm}} \hat{\mathbf{h}}_{\mathcal{V}, ji_{lm}}^H \boldsymbol{\Sigma}_{j,jk,lm} \hat{\mathbf{h}}_{\mathcal{V}, ji_{lm}}}. \quad (45)$$

Plugging (45) into the numerator of (44), we obtain

$$\mathbf{h}_{jlm}^H \boldsymbol{\Sigma}_{jjk} \boldsymbol{\Sigma}_{jjk} \mathbf{h}_{jlm} = \mathbf{h}_{jlm}^H \boldsymbol{\Sigma}_{j,jk,lm} \boldsymbol{\Sigma}_{j,jk,lm} \mathbf{h}_{jlm} \qquad \rightarrow \text{(intf. 1)}$$

$$- 2 \,\mathrm{Re} \left\{ \frac{\lambda_{ji_{lm}} \mathbf{h}_{jlm}^H \boldsymbol{\Sigma}_{j,jk,lm} \boldsymbol{\Sigma}_{j,jk,lm} \hat{\mathbf{h}}_{\mathcal{V}, ji_{lm}} \hat{\mathbf{h}}_{\mathcal{V}, ji_{lm}}^H \boldsymbol{\Sigma}_{j,jk,lm} \mathbf{h}_{jlm}}{1 + \lambda_{ji_{lm}} \hat{\mathbf{h}}_{\mathcal{V}, ji_{lm}}^H \boldsymbol{\Sigma}_{j,jk,lm} \hat{\mathbf{h}}_{\mathcal{V}, ji_{lm}}} \right\} \rightarrow \text{(intf. 2)}$$

$$+ \left| \lambda_{ji_{lm}} \right|^2 \frac{\left| \mathbf{h}_{jlm}^H \boldsymbol{\Sigma}_{j,jk,lm} \hat{\mathbf{h}}_{\mathcal{V}, ji_{lm}} \right|^2 \hat{\mathbf{h}}_{\mathcal{V}, ji_{lm}}^H \boldsymbol{\Sigma}_{j,jk,lm} \boldsymbol{\Sigma}_{j,jk,lm} \hat{\mathbf{h}}_{\mathcal{V}, ji_{lm}}}{\left| 1 + \lambda_{ji_{lm}} \hat{\mathbf{h}}_{\mathcal{V}, ji_{lm}}^H \boldsymbol{\Sigma}_{j,jk,lm} \hat{\mathbf{h}}_{\mathcal{V}, ji_{lm}} \right|^2} \rightarrow \text{(intf. 3)}.$$

Deterministic equivalent of (intf. 1)*:* Define $\boldsymbol{\Sigma}_{j,jk,lm}' = M \boldsymbol{\Sigma}_{j,jk,lm}$, then following similar procedures as before, it is straightforward to show that

$$\mathbf{h}_{jlm}^H \boldsymbol{\Sigma}_{j,jk,lm} \boldsymbol{\Sigma}_{j,jk,lm} \mathbf{h}_{jlm} \asymp \frac{d_j(\mathbf{z}_{lm})}{M^2} \mathrm{tr} \left(\boldsymbol{\Sigma}_{j,jk,lm}' \boldsymbol{\Sigma}_{j,jk,lm}' \right)$$

$$\asymp \frac{d_j(\mathbf{z}_{lm})}{M^2} \mathrm{tr} \left(\boldsymbol{\Sigma}_j' \boldsymbol{\Sigma}_j' \right) \asymp \frac{d_j(\mathbf{z}_{lm})}{M} \vartheta_j, \qquad (46)$$

where $\vartheta_j = m_o' \left(\frac{\sigma^2 + \varphi_j}{M} \right)$ is given by Theorem 2 for $\rho = \frac{\sigma^2 + \varphi_j}{M}$ and $\mathbf{T} = \boldsymbol{\Phi}_j \boldsymbol{\Lambda}_j$.

Deterministic equivalent of (intf. 2)*:* Instead of tackling the expression in (intf. 2) directly, we derive the deterministic equivalents of its numerator and denominator,

respectively. Plugging $\mathbf{h}_{jlm} = \hat{\mathbf{h}}_{jlm} + \tilde{\mathbf{h}}_{jlm}$ and $\hat{\mathbf{h}}_{jlm} = \sqrt{p_{lm}} d_j(\mathbf{z}_{lm}) \hat{\mathbf{h}}_{\mathcal{V}, ji_{lm}}$ into the numerator, we have that

$$\mathbf{h}_{jlm}^H \boldsymbol{\Sigma}_{j,jk,lm} \boldsymbol{\Sigma}_{j,jk,lm} \hat{\mathbf{h}}_{\mathcal{V}, ji_{lm}} \overset{(a)}{\asymp} \frac{\sqrt{p_{lm}} d_j(\mathbf{z}_{lm})}{M^2} \tilde{\phi}_{ji_{lm}} \mathrm{tr} \left(\boldsymbol{\Sigma}_j' \boldsymbol{\Sigma}_j' \right)$$

$$\asymp \frac{\sqrt{p_{lm}} d_j(\mathbf{z}_{lm})}{M} \tilde{\phi}_{ji_{lm}} \vartheta_j. \qquad (47)$$

Step (*a*) follows from Lemma 12 in [36] and Lemma 14.3 in [37]. Similarly, we have

$$\hat{\mathbf{h}}_{\mathcal{V}, ji_{lm}}^H \boldsymbol{\Sigma}_{j,jk,lm} \mathbf{h}_{jlm} \asymp \frac{\sqrt{p_{lm}} d_j(\mathbf{z}_{lm}) \tilde{\phi}_{ji_{lm}}}{M} \mathrm{tr} \left(\boldsymbol{\Sigma}_j \right)$$

$$= \sqrt{p_{lm}} d_j(\mathbf{z}_{lm}) \tilde{\phi}_{ji_{lm}} \delta_j, \qquad (48)$$

$$\hat{\mathbf{h}}_{\mathcal{V}, ji_{lm}}^H \boldsymbol{\Sigma}_{j,jk,lm} \hat{\mathbf{h}}_{\mathcal{V}, ji_{lm}}^H \asymp \tilde{\phi}_{ji_{lm}} \delta_j, \qquad (49)$$

where δ_j is given in Lemma 1. Based on (47) − (49), the equivalents of the denominator and numerator are given as $1 + \lambda_{ji_{lm}} \tilde{\phi}_{ji_{lm}} \delta_j$ and $\frac{1}{M} \lambda_{ji_{lm}} p_{lm} d_j^2(\mathbf{z}_{lm}) \tilde{\phi}_{ji_{lm}}^2 \vartheta_j \delta_j$, respectively. According to the continuous mapping theorem,

$$\text{(intf. 2)} - \frac{-2 \tilde{\phi}_{ji_{lm}}^2 \vartheta_j \delta_j}{1 + \lambda_{ji_{lm}} \tilde{\phi}_{ji_{lm}} \delta_j} \frac{p_{lm} d_j^2(\mathbf{z}_{lm}) \lambda_{ji_{lm}}}{M} \xrightarrow[M \to \infty]{a.s.} 0. \qquad (50)$$

Deterministic equivalent of (intf. 3)*:* Based on the techniques used to characterize (intf. 1) and (intf. 2), it is straightforward to show that

$$\text{(intf. 3)} - \frac{\left| \lambda_{ji_{lm}} \right|^2 \tilde{\phi}_{ji_{lm}}^3 \delta_j^2 \vartheta_j}{\left(1 + \lambda_{ji_{lm}} \tilde{\phi}_{ji_{lm}} \delta_j \right)^2} \frac{p_{lm} d_j^2(\mathbf{z}_{lm})}{M} \xrightarrow[M \to \infty]{a.s.} 0. \qquad (51)$$

Plugging (46), (50), and (51) into (44), we have that

$$\left| \hat{\mathbf{h}}_{jjk}^H \boldsymbol{\Sigma}_j \mathbf{h}_{jlm} \right|^2 = \frac{p_{jk} \tilde{\phi}_{ji_{jk}} d_j^2(\mathbf{z}_{jk}) d_j(\mathbf{z}_{lm}) \mu_{lm}}{\left(1 + \lambda_{ji_{jk}} \tilde{\phi}_{ji_{jk}} \delta_j \right)^2 M}, \qquad (52)$$

where $\mu_{jlm} = \vartheta_j - p_{lm} d_j(\mathbf{z}_{lm}) \lambda_{ji_{lm}} \vartheta_j \delta_j \tilde{\phi}_{ji_{lm}}^2 \frac{2 + \tilde{\phi}_{ji_{lm}} \lambda_{ji_{lm}} \delta_j}{\left(1 + \lambda_{ji_{lm}} \tilde{\phi}_{ji_{lm}} \delta_j \right)^2}$ is defined. Consequently, we have by the dominated convergence theorem that

$$\mathbb{E} \left\{ \left| \hat{\mathbf{h}}_{jjk}^H \boldsymbol{\Sigma}_j \mathbf{h}_{jlm} \right|^2 \middle| \hat{\mathbf{h}}_{(j)} \right\} - \frac{p_{jk} \tilde{\phi}_{ji_{jk}} d_j^2(\mathbf{z}_{jk}) d_j(\mathbf{z}_{lm}) \mu_{lm}}{\left(1 + \lambda_{ji_{jk}} \tilde{\phi}_{ji_{jk}} \delta_j \right)^2 M} \xrightarrow[M \to \infty]{a.s.} 0. \qquad (53)$$

Noise power

The noise term in (12) is scaled by $\left\|\mathbf{g}_{jk}\right\|^2$ for which we have that

$$
\begin{aligned}
\left\|\mathbf{g}_{jk}\right\|^2 &= \hat{\mathbf{h}}_{jjk}^H \mathbf{\Sigma}_j \mathbf{\Sigma}_j \hat{\mathbf{h}}_{jjk} \stackrel{(a)}{=} p_{jk} d_j^2\left(\mathbf{z}_{jk}\right) \\
&\frac{\frac{1}{M^2} \hat{\mathbf{h}}_{\mathcal{V},ji_{jk}}^H \mathbf{\Sigma}_{jjk}' \mathbf{\Sigma}_{jjk}' \hat{\mathbf{h}}_{\mathcal{V},ji_{jk}}}{\left(1+\lambda_{ji_{jk}} \hat{\mathbf{h}}_{\mathcal{V},ji_{jk}}^H \mathbf{\Sigma}_{jjk} \hat{\mathbf{h}}_{\mathcal{V},ji_{jk}}\right)^2} \\
&\stackrel{(b)}{\asymp} \frac{p_{jk} d_j^2\left(\mathbf{z}_{jk}\right) \tilde{\phi}_{ji_{jk}} \vartheta_j}{\left(1+\lambda_{ji_{jk}} \tilde{\phi}_{ji_{jk}} \delta_j\right)^2 M},
\end{aligned} \tag{54}
$$

where step (a) follows from Lemma 1 in [12] and step (b) follows from Lemma 12 in [36], Lemma 14.3 in [37], and Theorem 2. Then, by the dominated convergence theorem, we have

$$
\mathbb{E}\left\{\left\|\mathbf{g}_{jk}^H\right\|^2 \left|\hat{\mathbf{h}}_{(j)}\right.\right\} - \frac{p_{jk} d_j^2\left(\mathbf{z}_{jk}\right) \tilde{\phi}_{ji_{jk}} \vartheta_j}{\left(1+\lambda_{ji_{jk}} \tilde{\phi}_{ji_{jk}} \delta_j\right)^2 M} \xrightarrow[M\to\infty]{a.s.} 0. \tag{55}
$$

Finally, by the continuous mapping theorem, we arrive at the expression in (29).

Appendix 2
Proof of Theorem 4

Except for the channel variance $\operatorname{var}\left\{\mathbf{h}_{jjk}^H \mathbf{w}_{jk}\right\} = \mathbb{E}\left\{\left|\mathbf{h}_{jjk}^H \mathbf{w}_{jk} - \mathbb{E}\left\{\mathbf{h}_{jjk}^H \mathbf{w}_{jk}\right\}\right|^2\right\}$, large-scale approximations of the signal power and the interference in (22) can be calculated by following similar procedures as in Appendix 1. Thus, only the channel variance is considered here.

Define $c = \hat{\mathbf{h}}_{jjk}^H \mathbf{\Sigma}_j \hat{\mathbf{h}}_{jjk}$, $\bar{c} = \mathbb{E}\left\{\hat{\mathbf{h}}_{jjk}^H \mathbf{\Sigma}_j \hat{\mathbf{h}}_{jjk}\right\}$, and $b = \tilde{\mathbf{h}}_{jjk}^H \mathbf{\Sigma}_j \hat{\mathbf{h}}_{jjk}$, then

$$
\begin{aligned}
\operatorname{var}\left\{\mathbf{h}_{jjk}^H \mathbf{w}_{jk}\right\} &= \frac{1}{\gamma_{jk}} \mathbb{E}\left\{|c - \bar{c} + b|^2\right\} \\
&= \frac{1}{\gamma_{jk}} \mathbb{E}\left\{(c - \bar{c})(c + \bar{c})\right\} + \frac{1}{\gamma_{jk}} \mathbb{E}\left\{|b|^2\right\},
\end{aligned} \tag{56}
$$

where the last step is due to the fact that $\hat{\mathbf{h}}_{jjk}$ is independent of $\tilde{\mathbf{h}}_{jjk}$ and that $\mathbb{E}\{b\} = 0$.

From step (a) of Eq. (36), we have

$$
\begin{aligned}
c &= \frac{p_{jk} d_j^2\left(\mathbf{z}_{jk}\right) \hat{\mathbf{h}}_{\mathcal{V},ji_{jk}}^H \mathbf{\Sigma}_{jjk} \hat{\mathbf{h}}_{\mathcal{V},ji_{jk}}}{1+\lambda_{ji_{jk}} \hat{\mathbf{h}}_{\mathcal{V},ji_{jk}}^H \mathbf{\Sigma}_{jjk} \hat{\mathbf{h}}_{\mathcal{V},ji_{jk}}} \\
&\leq \frac{p_{jk} d_j^2\left(\mathbf{z}_{jk}\right) \hat{\mathbf{h}}_{\mathcal{V},ji_{jk}}^H \mathbf{\Sigma}_{jjk} \hat{\mathbf{h}}_{\mathcal{V},ji_{jk}}}{\lambda_{ji_{jk}} \hat{\mathbf{h}}_{\mathcal{V},ji_{jk}}^H \mathbf{\Sigma}_{jjk} \hat{\mathbf{h}}_{\mathcal{V},ji_{jk}}} \\
&\leq \frac{p_{jk} d_j^2\left(\mathbf{z}_{jk}\right)}{\lambda_{ji_{jk}}} \triangleq \theta.
\end{aligned} \tag{57}
$$

Therefore, $c \leq \theta$ and same bound also holds for $\bar{c}$. Thus, we have

$$
\operatorname{var}\left\{\mathbf{h}_{jjk}^H \mathbf{w}_{jk}\right\} \leq \frac{2\theta}{\gamma_{jk}} \mathbb{E}\{|c - \bar{c}|\} + \frac{1}{\gamma_{jk}} \mathbb{E}\left\{|b|^2\right\}. \tag{58}
$$

It is shown by Lemma 1 that $c - \frac{d_j^2(\mathbf{z}_{jk})\delta_{jk}}{1+\lambda_{ji_{jk}}\delta_{jk}} \xrightarrow[M\to\infty]{a.s.} 0$. Since c and $\bar{c}$ are bounded, this implies by the dominated convergence theorem that $\mathbb{E}\{|c - \bar{c}|\} \to 0$ as $M \to \infty$. Furthermore,

$$
\begin{aligned}
\mathbb{E}\left\{|b|^2\right\} &= \mathbb{E}\left\{\hat{\mathbf{h}}_{jjk}^H \mathbf{\Sigma}_j \tilde{\mathbf{h}}_{jjk} \tilde{\mathbf{h}}_{jjk}^H \mathbf{\Sigma}_j \hat{\mathbf{h}}_{jjk}\right\} \\
&= \mathbb{E}\left\{\hat{\mathbf{h}}_{jjk}^H \mathbf{\Sigma}_j \mathbf{C}_{jjk} \mathbf{\Sigma}_j \hat{\mathbf{h}}_{jjk}\right\} \\
&\stackrel{(a)}{\leq} \frac{1}{\varphi_j^2} \mathbb{E}\left\{\hat{\mathbf{h}}_{jjk}^H \mathbf{C}_{jjk} \hat{\mathbf{h}}_{jjk}\right\} \\
&= \frac{1}{\varphi_j^2} \operatorname{tr}\left(\hat{\mathbf{\Phi}}_{jjk} \mathbf{C}_{jjk}\right),
\end{aligned} \tag{59}
$$

where step (a) holds because $\mathbf{\Sigma}_j \preceq \frac{1}{\varphi_j} \mathbf{I}_M$ (where $\mathbf{A} \preceq \mathbf{B}$ means that $\mathbf{B} - \mathbf{A}$ is positive semi-definite). Since φ_j^2 scales as K^2 or equivalently as M^2, and $\operatorname{tr}(\hat{\mathbf{\Phi}}_{jjk}\mathbf{C}_{jjk})$ scales as M, we have that $\mathbb{E}\{|b|^2\} \to 0$ as $M \to \infty$. Consequently,

$$
\operatorname{var}\left\{\mathbf{h}_{jjk}^H \mathbf{w}_{jk}\right\} \xrightarrow[M\to\infty]{} 0. \tag{60}
$$

Authors' contributions

All authors contributed to this work, and the authors are listed in the order of their contribution. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

[1] Department of Electronic Engineering, Research Institute of Information Technology, Tsinghua University, Beijing 100084, China. [2] Department of Electrical Engineering (ISY), Linköping University, SE-58183 Linköping, Sweden.

References

1. G Caire, N Jindal, M Kobayashi, N Ravindran, Multiuser MIMO achievable rates with downlink training and channel state feedback. IEEE Trans. Inf. Theory. **56**(6), 2845–2866 (2010)
2. G Caire, S Shamai, On the achievable throughput of a multiantenna Gaussian broadcast channel. IEEE Trans. Inf. Theory. **49**(7), 1691–1706 (2003)
3. Y Wei, JM Cioffi, Sum capacity of Gaussian vector broadcast channels. IEEE Trans. Inf. Theory. **50**(9), 1875–1892 (2004)

Li *et al. EURASIP Journal on Wireless Communications and Networking*   (2017) 2017:117

Page 15 of 15

4. D Gesbert, M Kountouris, RW Heath, CB Chae, T Salzer, From single user to multiuser communications: shifting the MIMO paradigm. IEEE Signal Process. Mag. **24**(5), 36–46 (2007)

5. V Stankovic, M Haardt, Generalized design of multiuser MIMO precoding matrices. IEEE Trans. Wireless Commun. **7**(3), 953–961 (2008)

6. T Yoo, A Goldsmith, On the optimality of multiantenna broadcast scheduling using zero-forcing beamforming. IEEE J. Sel. Areas Commun. **24**(3), 1912–1921 (2006)

7. EG Larsson, O Edfors, F Tufvesson, TL Marzetta, Massive MIMO for next generation wireless systems. IEEE Commun. Mag. **52**(2), 186–195 (2014)

8. TL Marzetta, Noncooperative cellular wireless with unlimited numbers of base station antennas. IEEE Trans. Wireless Commun. **9**(1), 3590–3600 (2010)

9. F Rusek, D Persson, KL Buon, EG Larsson, TL Marzetta, O Edfors, F Tufvesson, Scaling up MIMO: opportunities and challenges with very large arrays. IEEE Trans. Signal Process. **30**(1), 40–60 (2013)

10. HQ Ngo, EG Larsson, TL Marzetta, Energy and spectral efficiency of very large multiuser MIMO systems. IEEE Trans. Commun. **61**(4), 1436–1449 (2013)

11. E Björnson, J Hoydis, M Kountouris, M Debbah, Massive MIMO systems with non-ideal hardware: energy efficiency, estimation, and capacity limits. IEEE Trans. Inf. Theory. **60**(11), 7112–7139 (2014)

12. J Hoydis, S ten Brink, M Debbah, Massive MIMO in the UL/DL of cellular networks: how many antennas do we need? IEEE J. Sel. Areas Commun. **31**(2), 160–171 (2013)

13. E Björnson, E Jorswieck, Optimal resource allocation in coordinated multi-cell systems. Found. Trends Commun. Inf. Theory. **9**(2–3), 113–381 (2013)

14. TL Marzetta, in *Proc. IEEE Asilomar Conference on Signals, Systems and Computers*. How much training is required for multiuser MIMO? (IEEE, Pacific Grove, 2006), pp. 359–363. 29 Oct.-1 Nov. 2006

15. KF Guo, Y Guo, G Fodor, G Ascheid, in *Proc. IEEE International Conference on Communications (ICC)*. Uplink power control with MMSE receiver in multi-cell MU-massive-MIMO systems (IEEE, Sydney, 2014), pp. 5184–5190

16. N Krishnan, RD Yates, NB Mandayam, Uplink linear receivers for multi-cell multiuser MIMO with pilot contamination: large system analysis. IEEE Trans. Wireless Commun. **13**(8), 4360–4373 (2014)

17. HQ Ngo, M Matthaiou, EG Larsson, in *2012 Swedish Communication Technologies Workshop (Swe-CTW)*. Performance analysis of large scale MU-MIMO with optimal linear receivers (IEEE, Lund, 2012), pp. 59–64

18. J Hoydis, S ten Brink, M Debbah, in *Proc. of 49th Allerton*. Massive MIMO: How many antennas do we need? (IEEE, Monticello, 2011), pp. 545–550

19. E Björnson, EG Larsson, M Debbah, Massive MIMO for maximal spectral efficiency: How many users and pilots should be allocated? IEEE Trans. Wireless Commun. **15**(2), 1293–1308 (2016)

20. J Jose, A Ashikhmin, TL Marzetta, S Vishwanath, Pilot contamination and precoding in multi-cell TDD systems. IEEE Trans. Wireless Commun. **10**(8), 2640–2651 (2011)

21. KF Guo, G Ascheid, in *Proc. IEEE Wireless Communications and Networking Conference (WCNC)*. Performance analysis of multi-cell MMSE based receivers in MU-MIMO systems with very large antenna arrays (IEEE, Shanghai, 2013), pp. 3175–3179

22. D Tse, P Viswanath, *Fundamentals of Wireless Communication*. (Cambridge University Press, New York, 2005)

23. JW Silverstein, ZD Bai, On the empirical distribution of eigenvalues of a class of large dimensional random matrices. J. Multivariate Anal. **54**(2), 175–192 (1995)

24. M Chiang, P Hande, T Lan, CW Tan, Power control in wireless cellular networks. Foundations Trends® Netw. **2**(4), 381–533 (2008)

25. D Gesbert, SG Kiani, A Gjendemsjo, GE Oien, Adaptation, coordination, and distributed rresource allocation in interference-limited wireless networks. Proc. IEEE. **95**(12), 2393–2409 (2007)

26. ZQ Luo, W Yu, An introduction to convex optimization for communications and signal processing. IEEE J. Sel. Areas Commun. **24**(8), 1426–1438 (2006)

27. M Chiang, Balancing transport and physical layers in wireless multihop networks: jointly optimal congestion control and power control. IEEE J. Sel. Areas Commun. **23**(1), 104–116 (2005)

28. IC Paschalidis, W Lai, D Starobinski, Asymptotically optimal transmission policies for large-scale low-power wireless sensor networks. IEEE/ACM Trans. Netw. **15**(1), 105–118 (2007)

29. K Kumaran, L Qian, Uplink scheduling in CDMA packet-data systems. Wireless Netw. **12**(1), 33–43 (2006)

30. M Charafeddine, A Sezgin, A Paulraj, in *Proc. of 45th Allerton*. Rate region frontiers for *n*-user interference channel with interference as noise (Allerton House, UIUC, Illinois, 2007). September 26–28, 2007

31. ZQ Luo, SZ Zhang, Dynamic spectrum management: complexity and duality. IEEE J. Sel. Topics Signal Process. **2**(1), 57–73 (2008)

32. M Chiang, CW Tan, DP Palomar, D O'Neill, D Julian, Power control by geometric programming. IEEE Trans. Wireless Commun. **6**(7), 2640–2651 (2007)

33. CW Tan, M Chiang, R Srikant, in *Proc. IEEE INFOCOM*. Fast algorithms and performance bounds for sum rate maximization in wireless networks (IEEE, Rio de Janeiro, 2009), pp. 1350–1358

34. J Vieira, S Malkowsky, K Nieman, Z Miers, N Kundargi, L Liu, IC Wong, V Öwall, O Edfors, F Tufvesson, in *Proc. IEEE GLOBECOM Workshop*. A flexible 100-antenna testbed for massive MIMO, (2014), pp. 287–293

35. H Yang, TL Marzetta, in *Proc. IEEE Vehicular Technology Conference (VTC Fall)*. A macro cellular wireless network with uniformly high user throughputs (IEEE, Vancouver, 2014), pp. 1–5

36. J Hoydis, Random matrix methods for advanced communication system. Ph.D dissertation, Supélec, Gif-Sur-Yvette, France (2012)

37. R Couillet, M Debbah, *Random Matrix Methods for Wireless Communications*. (Cambridge University Press, New York, 2011)

38. AW van der Vaart, *Asymptotic Statistics (Cambridge Series in Statistical and Probabilistic Mathematics)*. (Cambridge University Press, New York, 2000)

39. P Billingsley, *Probability and Measure*, 3rd ed. edn. (John Wiley & Sons, Inc., New York, 1995)