Joint Pilot Design and Uplink Power Allocation in Multi-Cell Massive MIMO Systems

Trinh Van Chien, Emil Björnson and Erik G. Larsson

Tweet

LINKÖPING UNIVERSITY

Joint Pilot Design and Uplink Power Allocation in Multi-Cell Massive MIMO Systems

Trinh Van Chien *Student Member*, *IEEE*, Emil Björnson, *Senior Member*, *IEEE*, and Erik G. Larsson, *Fellow*, *IEEE*

Abstract—This paper considers pilot design to mitigate pilot contamination and provide good service for everyone in multi-cell Massive multiple input multiple output (MIMO) systems. Instead of modeling the pilot design as a combinatorial assignment problem, as in prior works, we express the pilot signals using a pilot basis and treat the associated power coefficients as continuous optimization variables. We compute a lower bound on the uplink capacity for Rayleigh fading channels with maximum ratio detection that applies with arbitrary pilot signals. We further formulate the max-min fairness problem under power budget constraints, with the pilot signals and data powers as optimization variables. Because this optimization problem is non-deterministic polynomial-time hard due to signomial constraints, we then propose an algorithm to obtain a local optimum with polynomial complexity. Our framework serves as a benchmark for pilot design in scenarios with either ideal or non-ideal hardware. Numerical results manifest that the proposed optimization algorithms are close to the optimal solution obtained by exhaustive search for different pilot assignments and the new pilot structure and optimization bring large gains over the state-of-the-art suboptimal pilot design.

Index Terms—Massive MIMO, Pilot Design, Signomial Programming, Geometric Programming, Hardware Impairments.

I. INTRODUCTION

The demands on capacity and reliability in wireless cellular networks are continuously increasing. It is known that multiple input multiple output (MIMO) techniques can improve both capacity and reliability [1]–[3], but current systems only support up to eight antennas per base station (BS). While codebook-based channel acquisition is attractive in such small-scale MIMO systems, these methods are not scalable and unable to support the fifth generation (5G) demands on spectral efficiency (SE) in non-line-of-sight conditions [4]. Massive MIMO was proposed in [5] as a possible solution and it has emerged as a key 5G technology, because it offers significant improvements in both SE and energy efficiency [4]–[8]. By equipping the BSs with hundreds of antennas, mutual interference, thermal noise, and small-scale fading can be almost eliminated by virtue of the channel hardening and favorable propagation properties [6]. The BSs only need to use linear detection schemes, such as maximum ratio (MR) or zero forcing, to achieve nearly optimal performance [9]. In addition, the SE only depends on the large-scale fading

coefficients, thus power control algorithms are easier to deploy than in small-scale MIMO systems, which are greatly affected by small-scale fading [10].

The uplink (UL) detection and downlink precoding in Massive MIMO are based on instantaneous channel state information (CSI), which the BSs obtain from UL pilot signals. Mutually orthogonal pilots are desirable, but this is impractical in multi-cell scenarios since the pilot overhead would be proportional to the total number of users in the entire system. The consequence is that the pilot signals need to be reused across cells. This leads to pilot contamination [11], [12], where users sending the same pilot degrade each others channel estimation and cause large mutual interference. Hence, the pilot design is of key importance in Massive MIMO and should be optimized to mitigate the pilot contamination effects.

The baseline scheme for mitigating pilot contamination is to introduce a pilot reuse factor f, such that each pilot is only reused in $1/f$ of the cells. This approach, which was studied in [13]–[16], can greatly reduce the pilot contamination, even if the pilots are randomly assigned within each cell. However, this gain comes at the cost of using f times more pilots than in a system reusing the pilots in every cell. For any given cell, only a few users in the neighboring cells cause most of the potential pilot contamination, thus it is most important that these potential contaminators are assigned different pilots from the users in the given cell. Algorithms for coordinated pilot assignment were proposed in [17]–[20]. A pilot reuse dictionary was defined in [17] and the corresponding pilot assignment problem was shown to be non-deterministic polynomial-time hard (NP-hard), which motivates the design of heuristic assignment mechanisms. Although [17] proposed several greedy algorithms, the optimized SE was far from that with exhaustive search over all pilot assignments. Graph theory was used for pilot assignment in [18], by exploiting variations in the large-scale fading coefficients. A method called "smart pilot assignment" was proposed in [20] to enhance the max-min fairness SE level, by optimizing a heuristic mutual interference metric. Alternatively, [19] formulated the pilot assignment problem as a potential game. The numerical results in [18]–[20] show performance that is similar to an exhaustive search, but with a substantially lower computational complexity. Moreover, the authors of [21], [22] utilized particular channel properties to reduce channel estimation errors and mitigate pilot contamination. In particular, [21] utilized the orthogonality among different channels and an assumed low-rankness of the channel covariance matrices. An adjustable phase shift pilot construction was suggested in [22] based on the relationship between channel correlations in the frequency

domain and their power angle-delay spectrum. However, all these algorithms rely on the assumption of fixed pilot and data power.

The pilot and payload data powers are usually treated as constants in the Massive MIMO literature, but it is known from [12], [23] that the performance can be much improved by using the optimal power allocation, which balances the mutual interference levels. To improve the channel estimation quality, more power might also be assigned to the pilots than to the data transmissions [24], [25]. For single-cell systems, [24] showed that a pilot-data power imbalance is especially important for cell-edge users. Moreover, the power allocation that maximizes the sum SE is much different from the one that maximizes the max-min SE. Similar behaviors for multi-cell systems were observed in [25]. The authors in [26] considered power optimization problems with pilot reuse factors. To the best of our knowledge, no prior work analyzes joint pilot design and power control in Massive MIMO systems.

In this paper, we propose a novel pilot design and optimize the UL performance in multi-cell Massive MIMO systems, using the max-min fairness utility. Our main contributions are:

- We propose a new pilot design where the pilot signals are treated as continuous variables. We demonstrate that previous pilot designs are special cases of our proposal.
- Based on the proposed pilot design, we derive closed-form expressions of the SE with Rayleigh fading channels and MR detection, for the cases of ideal hardware and with hardware impairments. These expressions explicitly demonstrate how the SE is affected by mutual interference, noise, and pilot contamination.
- We formulate the max-min fairness problem for the proposed pilot design, by treating the pilot signals, pilot powers, and data powers as optimization variables. This is an NP-hard signomial program, so we propose an algorithm that finds a local optimum in polynomial time. For comparison the optimal solution by an exhaustive search of different pilot assignments is also investigated.
- The proposed algorithms are evaluated numerically, with either ideal hardware or hardware impairments. The results show that our local solution is close to the global optimum by exhaustive search over different pilot assignments and demonstrate significant improvements over the heuristic algorithms in prior works.

A preliminary version of this work, focusing only on pilot optimization with fixed data powers, was presented in [27].

The rest of this paper is organized as follows: Section II presents our proposed pilot structure and compares it with prior works. Lower bounds on the UL ergodic SE for arbitrary pilots are derived in Section III, while Section IV formulates the max-min fairness optimization problems and provides the global and local solutions. Sections V and VI extend our research to the case of hardware impairments and correlated Rayleigh fading, respectively. Finally, Section VII gives extensive numerical results and some conclusions are provided in Section VIII.

Notations: Lower bold letters are used for vectors and upper cases are for matrices. $(\cdot)^T$ and $(\cdot)^H$ stand for regular transpose and Hermitian transpose, respectively. The superscript $*$

denotes the conjugate transpose of a complex number. $\mathbf{I}_n$ is the identity matrix of size $n \times n$. $\mathbb{C}^{m \times n}$ ($\mathbb{R}^{m \times n}$) is the space of complex (real) $m \times n$ matrices, while $\mathbb{C}^{\tau_p}$ denotes the space of τ_p-length complex vectors. $\mathbb{R}_+$ is the set of nonnegative real numbers. $\mathbb{E}\{\cdot\}$ denotes the expectation of a random variable and $\| \cdot \|$ is the Euclidean norm. Finally, $\mathcal{CN}(\cdot, \cdot)$ is the circularly symmetric complex Gaussian distribution, while $\mathcal{N}(\cdot, \cdot)$ is the normal distribution.

II. PILOT DESIGNS FOR MASSIVE MIMO SYSTEMS

We consider the UL of a multi-cell Massive MIMO system with L cells. Each cell consists of a BS equipped with M antennas that serves K single-antenna users. All tuples of cell and user indices belong to a set $\mathcal{S}$ defined as

$$\mathcal{S} = \{(i,t) : i \in \{1, \ldots, L\}, t \in \{1, \ldots, K\}\}. \tag{1}$$

The radio channels vary over time and frequency. We divide the time-frequency plane into coherence intervals, each containing τ_c samples, such that the channel between each user and each BS is static and frequency flat. In each coherence block, the pilot signaling utilizes τ_p symbols and the remaining is dedicated to data transmission. In this paper, we focus on the UL, so the fraction $(1 - \tau_p/\tau_c)$ of the coherence interval is dedicated to UL data transmission. However, it is straightforward to extend our work to the downlink by using time division duplex (TDD) and channel reciprocity. We assume $1 \leq \tau_p < \tau_c$ to keep the training process feasible and stress that the case $\tau_p < KL$ is of practical importance since it gives rise to pilot contamination and since L is large in practice.

A. Proposed Pilot Design

Let us denote the τ_p mutually orthonormal basis vectors $\{\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_{\tau_p}\}$, where $\boldsymbol{\phi}_b \in \mathbb{C}^{\tau_p}$ is a vector whose bth element has unit magnitude, and all other elements are equal to zero. The corresponding basis matrix is

$$\boldsymbol{\Phi} = [\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_{\tau_p}]. \tag{2}$$

We assume that the pilot signals of the users can span arbitrarily over the above τ_p basis vectors. We aim at designing a pilot signal collection $\{\boldsymbol{\psi}_{1,1}, \ldots, \boldsymbol{\psi}_{L,K}\}$ comprising the KL pilot signals used by all users in the network and each of them has the length of τ_p symbols. The pilot signal of user k in cell l is $\boldsymbol{\psi}_{l,k} = [\psi_{l,k}^1, \ldots, \psi_{l,k}^{\tau_p}]^T \in \mathbb{C}^{\tau_p}$ and the power that this user assigns to the bth pilot basis is denoted as $\hat{p}_{l,k}^b \geq 0$. Thus, the pilot of user k in cell l is

$$\boldsymbol{\psi}_{l,k} = \sum_{b=1}^{\tau_p} \sqrt{\hat{p}_{l,k}^b} \boldsymbol{\phi}_b, \forall l, k. \tag{3}$$

We stress that the pilot construction in (3) can be used to create any set of τ_p orthogonal pilot signals (up to a unitary transformation) and many different sets of non-orthogonal signals. [1] The total pilot power consumption utilized by user k

---

[1] The pilot signals in (3) are formed as linear combinations of basis vectors in the complex field. The new pilot design allows the use of nonorthogonal pilot signals even within a cell in order to get extra degrees of freedom to minimize the interference in the network.

in cell l is $\|\boldsymbol{\psi}_{l,k}\|^2 = \sum_{b=1}^{\tau_p} \hat{p}_{l,k}^b$ and we assume that it satisfies the power constraint

$$\frac{1}{\tau_p} \sum_{b=1}^{\tau_p} \hat{p}_{l,k}^b \leq P_{\max,l,k}, \forall l, k, \quad (4)$$

where $P_{\max,l,k}$ is the maximum pilot power for user k in cell l. The inner product of two pilot signals $\boldsymbol{\psi}_{l,k}$ and $\boldsymbol{\psi}_{i,t}$ is

$$\boldsymbol{\psi}_{l,k}^H \boldsymbol{\psi}_{i,t} = \sum_{b=1}^{\tau_p} \sqrt{\hat{p}_{l,k}^b \hat{p}_{i,t}^b}. \quad (5)$$

These pilot signals are orthogonal if the product is zero, which only happens when they allocate their powers to different subsets of basis vectors. Otherwise, they are non-orthogonal and then the two users cause pilot contamination to each other. If the square roots of the powers allocated to the K users in cell l are gathered in matrix form as

$$\boldsymbol{P}_l = \begin{bmatrix} \sqrt{\hat{p}_{l,1}^1} & \sqrt{\hat{p}_{l,2}^1} & \cdots & \sqrt{\hat{p}_{l,K}^1} \\ \sqrt{\hat{p}_{l,1}^2} & \sqrt{\hat{p}_{l,2}^2} & \cdots & \sqrt{\hat{p}_{l,K}^2} \\ \vdots & \vdots & \ddots & \vdots \\ \sqrt{\hat{p}_{l,1}^{\tau_p}} & \sqrt{\hat{p}_{l,2}^{\tau_p}} & \cdots & \sqrt{\hat{p}_{l,K}^{\tau_p}} \end{bmatrix} \in \mathbb{R}_+^{\tau_p \times K}, \quad (6)$$

then the users in cell l utilize a pilot matrix defined as

$$\boldsymbol{\Psi}_l = [\boldsymbol{\psi}_{l,1}, \ldots, \boldsymbol{\psi}_{l,K}] = \boldsymbol{\Phi} \boldsymbol{P}_l. \quad (7)$$

We now describe the difference between this new pilot structure and the prior works, for example [18], [20], [24], [25].

B. Other Pilot Designs

The works [18], [20] considered the assignment of τ_p orthogonal pilot signals under the assumption of fixed equal pilot power. Using our notation, the pilot matrix in cell l is

$$\widehat{\boldsymbol{\Psi}}_l = [\hat{\boldsymbol{\psi}}_{l,1}, \ldots, \hat{\boldsymbol{\psi}}_{l,K}] = \sqrt{\tilde{p}} \boldsymbol{\Phi} \boldsymbol{\Pi}_l, \quad (8)$$

where $0 < \tilde{p} \leq \tau_p P_{\max,l,k}$ is the equal power level of all users. $\boldsymbol{\Pi}_l \in \mathbb{R}_+^{\tau_p \times K}$ is a permutation matrix, that assigns the pilot signals to each user in cell l. The assignment is optimized in [18], [20] to minimize a heuristic mutual interference metric. Note that these works assume orthogonal pilot signals and equal power allocation, which are simplifications compared to (7). These assumptions are generally suboptimal. Apart from this, the selection of the optimal permutation matrices for cell l is a combinatorial problem, so to limit the computational complexity [18], [20] and the references therein only study the special case of $\tau_p = K$.

The previous work [24] optimized the pilot powers to maximize functions of the SE, but the paper only considered a single cell without pilot contamination. The authors of [25] optimized the pilot powers to minimize the UL transmit power for a multi-cell system. This work assumed $\tau_p = K$ and a fixed pilot assignment. If $\tilde{p}_{l,k}$ is the pilot power of user k in cell l, the square root of the power matrix allocated to the K users in cell l is a diagonal matrix defined as

$$\widetilde{\boldsymbol{P}}_l = \text{diag}\left(\sqrt{\tilde{p}_{l,1}}, \ldots, \sqrt{\tilde{p}_{l,K}}\right), \quad (9)$$

where $\text{diag}(\mathbf{x})$ denotes the diagonal matrix with the vector $\mathbf{x}$ on the diagonal. The pilot matrix in cell l is then formulated as

$$\widetilde{\boldsymbol{\Psi}}_l = \boldsymbol{\Phi} \widetilde{\boldsymbol{P}}_l. \quad (10)$$

Similar to (4), the pilot power at user k in cell l is limited as

$$0 \leq \tilde{p}_{l,k} \leq \tau_p P_{\max,l,k}. \quad (11)$$

Since orthogonal pilots and fixed pilot assignment are assumed, this is also a special case of (7). We can combine the pilot structure in (10) and the idea of selecting a permutation matrix in (8) to jointly optimize the power allocation and pilot assignment. In particular, the pilot signals of the users in cell l are now defined as

$$\check{\boldsymbol{\Psi}}_l = [\check{\boldsymbol{\psi}}_{l,1}, \ldots, \check{\boldsymbol{\psi}}_{l,K}] = \boldsymbol{\Phi} \boldsymbol{\Pi}_l \widetilde{\boldsymbol{P}}_l. \quad (12)$$

This modified pilot design is a special case of (7) and has not been studied in prior works, but will be considered herein. In order to analyze the channel estimation, we define a pilot reuse set $\mathcal{P}_{l,k}$ including all tuples of cell and user indices that cause pilot contamination to user k in cell l:

$$\mathcal{P}_{l,k} = \{(i,t) \in \mathcal{S} : \check{\boldsymbol{\psi}}_{i,t}^H \check{\boldsymbol{\psi}}_{l,k} \neq 0\}. \quad (13)$$

We stress that designing an exhaustive search to obtain the best pilot assignment strategy is extremely computationally expensive.[2] As in prior works, we only consider the case $\tau_p = K$ when using (12) and we further assume that orthogonal pilots are used within each cell; that is, $\mathcal{P}_{l,k} \cap \mathcal{P}_{l,k'} = \emptyset$ for any user indices $k \neq k'$ in cell l. To perform an exhaustive search, we need to construct a dictionary $\mathcal{D}$, see Fig. 1, with all the possible combinations of pilot assignments in the network. Let $\chi_l^k \in \{1, \ldots, K\}$ denote the index of the pilot signal assigned to user k in cell l. It follows that $\chi_l^k \neq \chi_l^{k'}$ for $k \neq k'$ since all users within a cell use different pilots. The pilot assignment matrix $\mathbf{A} \in \{1, \ldots, K\}^{L \times K}$ containing the pilot indices of the KL users is

$$\mathbf{A} = \begin{bmatrix} \chi_1^1 & \chi_1^2 & \cdots & \chi_1^K \\ \chi_2^1 & \chi_2^2 & \cdots & \chi_2^K \\ \vdots & \vdots & \ddots & \vdots \\ \chi_L^1 & \chi_L^2 & \cdots & \chi_L^K \end{bmatrix}. \quad (14)$$

Each row of $\mathbf{A}$ contains 1 to K and there are $K!$ different combinations, each defining a permutation matrix $\boldsymbol{\Pi}_l$ for the pilot signals in (8) and (12). The dictionary $\mathcal{D} \overset{\Delta}{=} \{\mathbf{A}\}$ contains all the $(K!)^L$ pilot assignment matrices. For each $\mathbf{A} \in \mathcal{D}$, we can extract the pilot reuse sets $\mathcal{P}_{l,k}^{\mathbf{A}}, \forall l, k$ as[3]

$$\mathcal{P}_{l,k}^{\mathbf{A}} = \{(i,t) \in \mathcal{S} : \chi_l^k = \chi_i^t\}. \quad (15)$$

The dictionary $\mathcal{D}$ will be later used to obtain the pilot assignment that maximizes the SE performance.

III. Uplink Massive MIMO Transmission

This section provides ergodic SE expressions with arbitrary pilot signals, which are later used for pilot optimization.

---

[2] For the first user in the first cell ($l = 1, k = 1$), there are $(K!)^{L-1}$ possibilities of $\mathcal{P}_{1,1}$. There are then $(K-1)!^{L-1}$ possible $\mathcal{P}_{1,2}$ and so on.

[3] Each collection $\{\mathcal{P}_{l,k}^{\mathbf{A}}\}$ of pilot reuse sets is generated by $K!$ different $\mathbf{A} \in \mathcal{D}$. By eliminating the $K! - 1$ copies, the size of the dictionary $\mathcal{D}$ can be reduced to $(K!)^{L-1}$, which still grows rapidly with K and L.
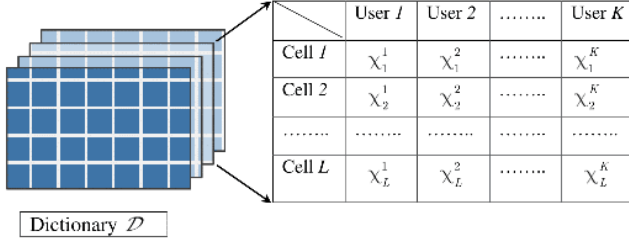
Fig. 1. The dictionary $\mathcal{D}$ contains all possible pilot assignment indices for all users in the network.

A. Channel Estimation with Arbitrary Pilots

During the UL pilot transmission, the received signal $\mathbf{Y}_l \in \mathbb{C}^{M \times \tau_P}$ at the BS of cell l is

$$\mathbf{Y}_l = \sum_{(i,t) \in \mathcal{S}} \mathbf{h}_{i,t}^l \boldsymbol{\psi}_{i,t}^H + \mathbf{N}_l, \qquad (16)$$

where $\mathbf{h}_{i,t}^l \in \mathbb{C}^M$ denotes the channel between user t in cell i and BS l. $\mathbf{N}_l \in \mathbb{C}^{M \times \tau_P}$ is the additive noise with independent elements distributed as $\mathcal{CN}(0, \sigma^2)$. Correlating $\mathbf{Y}_l$ in (16) with the pilot $\boldsymbol{\psi}_{l,k}$ of user k in cell l, we obtain

$$\mathbf{y}_{l,k} = \mathbf{Y}_l \boldsymbol{\psi}_{l,k} = \sum_{(i,t) \in \mathcal{S}} \mathbf{h}_{i,t}^l \boldsymbol{\psi}_{i,t}^H \boldsymbol{\psi}_{l,k} + \mathbf{N}_l \boldsymbol{\psi}_{l,k}. \qquad (17)$$

We consider uncorrelated Rayleigh fading since results obtained with this tractable model well matches the results obtained in non-line-of-sight measurements [28]. The channel between user t in cell i and BS l is distributed as

$$\mathbf{h}_{i,t}^l \sim \mathcal{CN}\left(\mathbf{0}, \beta_{i,t}^l \mathbf{I}_M\right), \qquad (18)$$

where the variance $\beta_{i,t}^l$ determines the large-scale fading, including geometric attenuation and shadowing. By using minimum mean squared error (MMSE) estimation, the distributions of the channel estimate and estimation error when using the pilot structure in (7) are given in Lemma 1.

Lemma 1. *If the system uses the pilot structure in* (7)*, the MMSE estimate of $\mathbf{h}_{l,k}^l$ based on $\mathbf{y}_{l,k}$ in* (17) *is computed as*

$$\hat{\mathbf{h}}_{l,k}^l = \frac{\beta_{l,k}^l \sum_{b=1}^{\tau_P} \hat{p}_{l,k}^b}{\sum_{(i,t) \in \mathcal{S}} \beta_{i,t}^l \left(\sum_{b=1}^{\tau_P} \sqrt{\hat{p}_{i,t}^b \hat{p}_{l,k}^b}\right)^2 + \sigma^2 \sum_{b=1}^{\tau_P} \hat{p}_{l,k}^b} \mathbf{y}_{l,k}. \qquad (19)$$

The channel estimate is distributed as

$$\hat{\mathbf{h}}_{l,k}^l \sim \mathcal{CN}\left(\mathbf{0}, \gamma_{l,k}^l \mathbf{I}_M\right), \qquad (20)$$

where

$$\gamma_{l,k}^l = \frac{(\beta_{l,k}^l)^2 \left(\sum_{b=1}^{\tau_P} \hat{p}_{l,k}^b\right)^2}{\sum_{(i,t) \in \mathcal{S}} \beta_{i,t}^l \left(\sum_{b=1}^{\tau_P} \sqrt{\hat{p}_{i,t}^b \hat{p}_{l,k}^b}\right)^2 + \sigma^2 \sum_{b=1}^{\tau_P} \hat{p}_{l,k}^b}. \qquad (21)$$

The estimation error $\mathbf{e}_{l,k}^l = \mathbf{h}_{l,k}^l - \hat{\mathbf{h}}_{l,k}^l$ is independent of the channel estimate and distributed as

$$\mathbf{e}_{l,k}^l \sim \mathcal{CN}\left(\mathbf{0}, \left(\beta_{l,k}^l - \gamma_{l,k}^l\right) \mathbf{I}_M\right). \qquad (22)$$

Proof. The proof follows directly from standard MMSE estimation techniques in [29]. □

Lemma 1 provides the MMSE estimator for the pilot design in (7). The pilot powers as well as inner products between pilot signals appear explicitly in the expressions. We now compute the channel estimate and estimation error of $\mathbf{h}_{l,k}^l$ when using the pilot structure in (12).

Corollary 1. *If the system uses the alternative pilot structure in* (12)*, the MMSE channel estimate in* (19) *is simplified to*

$$\hat{\mathbf{h}}_{l,k}^l = \frac{\beta_{l,k}^l}{\sum_{(i,t) \in \mathcal{P}_{l,k}^{\mathbf{A}}} \beta_{i,t}^l \tilde{p}_{i,t} + \sigma^2} \mathbf{y}_{l,k}. \qquad (23)$$

The estimate channel and estimation error are distributed as

$$\hat{\mathbf{h}}_{l,k}^l \sim \mathcal{CN}\left(\mathbf{0}, \frac{(\beta_{l,k}^l)^2 \tilde{p}_{l,k}}{\sum_{(i,t) \in \mathcal{P}_{l,k}^{\mathbf{A}}} \beta_{i,t}^l \tilde{p}_{i,t} + \sigma^2} \mathbf{I}_M\right), \qquad (24)$$

$$\mathbf{e}_{l,k}^l \sim \mathcal{CN}\left(\mathbf{0}, \beta_{l,k}^l \frac{\sum_{(i,t) \in \mathcal{P}_{l,k}^{\mathbf{A}} \backslash (l,k)} \beta_{i,t}^l \tilde{p}_{i,t} + \sigma^2}{\sum_{(i,t) \in \mathcal{P}_{l,k}^{\mathbf{A}}} \beta_{i,t}^l \tilde{p}_{i,t} + \sigma^2} \mathbf{I}_M\right). \qquad (25)$$

Proof. This follows from replacing the terms $\sum_{b=1}^{\tau_P} \hat{p}_{l,k}^b$ and $\sum_{(i,t) \in \mathcal{S}} \beta_{i,t}^l \left(\sum_{b=1}^{\tau_P} \sqrt{\hat{p}_{i,t}^b \hat{p}_{l,k}^b}\right)^2$ in Lemma 1 by $\tilde{p}_{l,k}$ and $\sum_{(i,t) \in \mathcal{P}_{l,k}^{\mathbf{A}}} \beta_{i,t}^l \tilde{p}_{i,t} \tilde{p}_{l,k}$, and then doing some algebra. □

Corollary 1 reveals that the quality of the estimated channel heavily depends on both the pilot power control and the pilot reuse set $\mathcal{P}_{l,k}^{\mathbf{A}}$. A proper selection of $\mathcal{P}_{l,k}^{\mathbf{A}}$ mitigates channel estimation errors, and will also reduce the coherent interference during data transmission. Aligned with prior works, in the special case of $\tilde{p}_{l,k} = \tilde{p}, \forall l, k$, the channel estimate and estimation error are obtained for the pilot structure in (8). We now use the distributions in Lemma 1 and Corollary 1 to derive lower bounds on the UL ergodic capacity.

B. Uplink Data Transmission

In the UL data transmission, user t in cell i transmits the signal $x_{i,t} \sim \mathcal{CN}(0, 1)$. The $M \times 1$ received signal vector at BS l is the superposition of the transmitted signals

$$\mathbf{y}_l = \sum_{(i,t) \in \mathcal{S}} \sqrt{p_{i,t}} \mathbf{h}_{i,t}^l x_{i,t} + \mathbf{n}_l, \qquad (26)$$

where $p_{i,t}$ is the transmit power corresponding to the signal $x_{i,t}$ and the additive noise is $\mathbf{n}_l \sim \mathcal{CN}(\mathbf{0}, \sigma^2 \mathbf{I}_M)$. To detect the transmitted signal, BS l selects a detection vector $\mathbf{v}_{l,k} \in \mathbb{C}^M$ and applies it to the received signal as

$$\mathbf{v}_{l,k}^H \mathbf{y}_l = \sum_{(i,t) \in \mathcal{S}} \sqrt{p_{i,t}} \mathbf{v}_{l,k}^H \mathbf{h}_{i,t}^l x_{i,t} + \mathbf{v}_{l,k}^H \mathbf{n}_l. \qquad (27)$$

A general lower bound on the UL ergodic capacity of user k in cell l is computed in [9] as

$$R_{l,k} = \left(1 - \frac{\tau_p}{\tau_c}\right) \log_2\left(1 + \text{SINR}_{l,k}\right), \qquad (28)$$

$$\text{SINR}_{l,k}^{\text{MR}} = \frac{M(\beta_{l,k}^l)^2 p_{l,k} \left(\sum_{b=1}^{\tau_p} \hat{p}_{l,k}^b \right)^2}{\left(\sum_{(i,t)\in\mathcal{S}} \beta_{i,t}^l \left(\sum_{b=1}^{\tau_p} \sqrt{\hat{p}_{i,t}^b \hat{p}_{l,k}^b} \right)^2 + \sigma^2 \sum_{b=1}^{\tau_p} \hat{p}_{l,k}^b \right) \left(\sum_{(i,t)\in\mathcal{S}} p_{i,t} \beta_{i,t}^l + \sigma^2 \right) + M \sum_{(i,t)\in\mathcal{S}\setminus(l,k)} p_{i,t}(\beta_{i,t}^l)^2 \left(\sum_{b=1}^{\tau_p} \sqrt{\hat{p}_{i,t}^b \hat{p}_{l,k}^b} \right)^2}. \quad (32)$$

$$\widetilde{\text{SINR}}_{l,k}^{\text{MR}} = \frac{M(\beta_{l,k}^l)^2 p_{l,k} \tilde{p}_{l,k}}{\left(\sum_{(i,t)\in\mathcal{P}_{l,k}^{\mathbf{A}}} \beta_{i,t}^l \tilde{p}_{i,t} + \sigma^2 \right) \left(\sum_{(i,t)\in\mathcal{S}} p_{i,t} \beta_{i,t}^l + \sigma^2 \right) + M \sum_{(i,t)\in\mathcal{P}_{l,k}^{\mathbf{A}}\setminus(l,k)} p_{i,t} \tilde{p}_{i,t}(\beta_{i,t}^l)^2}. \quad (35)$$

where the effective SINR value, $\text{SINR}_{l,k}$, is

$$\frac{p_{l,k} |\mathbb{E}\{\mathbf{v}_{l,k}^H \mathbf{h}_{l,k}^l\}|^2}{\sum_{(i,t)\in\mathcal{S}} p_{i,t} \mathbb{E}\{|\mathbf{v}_{l,k}^H \mathbf{h}_{i,t}^l|^2\} - p_{l,k} |\mathbb{E}\{\mathbf{v}_{l,k}^H \mathbf{h}_{l,k}^l\}|^2 + \sigma^2 \mathbb{E}\{\|\mathbf{v}_{l,k}\|^2\}}. \quad (29)$$

The lower bound on the UL ergodic capacity in (28) is computed by using the use-and-then-forget bounding technique [6] and its tightness compared to the other possible bounds is discussed in Appendix D in [6]. Although the channel capacity for Massive MIMO in the case of imperfect CSI is unknown, we believe that the lower bound in (28) is quite close to the actual capacity. This is because the effective noise is comprised of a sum of many uncorrelated terms, it is close to Gaussian. This agrees with the worst-case-is-Gaussian assumption made when obtaining the bound. As a contribution of this paper, we compute a closed form expression for this lower bound in the case of MR detection with

$$\mathbf{v}_{l,k} = \hat{\mathbf{h}}_{l,k}^l. \quad (30)$$

Lemma 2. *This is a highly computationally scalable detection method for Massive MIMO systems. If the system uses the pilot structure in (7) and MR detection, the SE in (28) for user k in cell l becomes*

$$R_{l,k}^{\text{MR}} = \left(1 - \frac{\tau_p}{\tau_c}\right) \log_2 \left(1 + \text{SINR}_{l,k}^{\text{MR}}\right), \quad (31)$$

where $\text{SINR}_{l,k}^{\text{MR}}$ is shown in (32).

Proof. The proof is available in Appendix A. □

From (32), we notice that it is always advantageous to add more BS antennas since the numerator grows linearly with M (and only some terms in the denominator have the same scaling). The first term in the denominator represents non-coherent interference that only depends on the number of BSs and users, while it is independent of M. The second term in the denominator represents coherent interference caused by pilot contamination and it grows linearly with M. As a consequence, as $M \to \infty$, we have

$$\text{SINR}_{l,k}^{\text{MR}} \to \frac{(\beta_{l,k}^l)^2 p_{l,k} \left(\sum_{b=1}^{\tau_p} \hat{p}_{l,k}^b \right)^2}{\sum_{(i,t)\in\mathcal{S}\setminus(l,k)} p_{i,t}(\beta_{i,t}^l)^2 \left(\sum_{b=1}^{\tau_p} \sqrt{\hat{p}_{i,t}^b \hat{p}_{l,k}^b} \right)^2}. \quad (33)$$

This limit depends only on the pilot design (i.e., inner products between pilot signals) and data power. An optimized selection of the power terms $p_{l,k}, \hat{p}_{l,k}^b, \forall l, k, b$, improves the SE by enhancing the channel estimation quality and reducing the coherent interference.

We also consider the achievable SE for the modified pilot structure in (12) as shown in Corollary 2.

Corollary 2. *If the system uses the pilot structure in (12), a lower bound on the capacity for user k in cell l with uncorrelated Rayleigh fading channels and MR detection is*

$$R_{l,k}^{\text{MR}} = \left(1 - \frac{\tau_p}{\tau_c}\right) \log_2 \left(1 + \widetilde{\text{SINR}}_{l,k}^{\text{MR}}\right), \quad (34)$$

where the SINR value, $\widetilde{\text{SINR}}_{l,k}^{\text{MR}}$, is given in (35).

Proof. This follows as a special case of Lemma 2. □

The SE in Corollary 2 depends explicitly on the choice of $\mathcal{P}_{l,k}^{\mathbf{A}}, \forall l, k$, thus the optimization of the pilot assignment is a combinatorial problem. We stress that the SINR expressions reflect the joint effects of pilot design, channel estimation quality, pilot contamination, and data power control, in contrast to the MSE that cannot distinguish between pilot contamination and noise. Hence, the SINR is a good metric to consider in the max-min fairness optimization as shown in the next section.

IV. MAX-MIN FAIRNESS OPTIMIZATION

In this section, we first utilize the SE expressions in Lemma 2 and Corollary 2 to formulate max-min fairness problems with joint pilot and data optimization. We demonstrate that these optimization problems are NP-hard and propose an algorithm to find the globally optimal solution with the pilot design in (12) by making an exhaustive search over all pilot assignments. In addition, instead of looking for the global optimum, an algorithm to obtain a locally optimal solution in polynomial time is presented when using the new pilot design in (7).

A. Problem Formulation

A key vision of Massive MIMO is to provide uniformly good quality of service for everyone in the network. We will investigate how to optimize the pilots and powers towards this goal. We consider the pilot and data powers as optimization

$$\widetilde{\mathrm{SINR}}_{l,k}^{\mathrm{MR}} = \frac{M(\beta_{l,k}^l)^2 p_{l,k} \prod_{b=1}^{\tau_p} \left(\hat{p}_{l,k}^b / \alpha_{l,k}^b\right)^{2\alpha_{l,k}^b}}{\left(\sum\limits_{(i,t)\in\mathcal{S}} \beta_{i,t}^l \left(\sum\limits_{b=1}^{\tau_p} \sqrt{\hat{p}_{i,t}^b \hat{p}_{l,k}^b}\right)^2 + \sigma^2 \sum\limits_{b=1}^{\tau_p} \hat{p}_{l,k}^b\right)\left(\sum\limits_{(i,t)\in\mathcal{S}} p_{i,t} \beta_{i,t}^l + \sigma^2\right) + M \sum\limits_{(i,t)\in\mathcal{S}\backslash(l,k)} p_{i,t}(\beta_{i,t}^l)^2 \left(\sum\limits_{b=1}^{\tau_p} \sqrt{\hat{p}_{i,t}^b \hat{p}_{l,k}^b}\right)^2}. \quad (47)$$

variables. The max-min fairness optimization problem is first formulated for the proposed pilot design in (7) as[4]

$$\begin{aligned}
\underset{\{\hat{p}_{l,k}^b, p_{l,k} \geq 0\}}{\text{maximize}} \quad & \underset{(l,k)}{\min} \log_2\left(1 + \mathrm{SINR}_{l,k}^{\mathrm{MR}}\right) \\
\text{subject to} \quad & \frac{1}{\tau_p} \sum_{b=1}^{\tau_p} \hat{p}_{l,k}^b \leq P_{\mathrm{max},l,k}, \forall l,k, \\
& p_{l,k} \leq P_{\mathrm{max},l,k}^d, \forall l,k,
\end{aligned} \quad (36)$$

where $P_{\mathrm{max},l,k}^d$ is the maximum power that users can provide for each data symbol. Note that this optimization problem jointly generates the pilot signals and performs power control on the pilot and data transmission. The epigraph-form representation of (36) is

$$\begin{aligned}
\underset{\xi,\{\hat{p}_{l,k}^b, p_{l,k} \geq 0\}}{\text{maximize}} \quad & \xi & (37a) \\
\text{subject to} \quad & \mathrm{SINR}_{l,k}^{\mathrm{MR}} \geq \xi, \forall l,k, & (37b) \\
& \frac{1}{\tau_p} \sum_{b=1}^{\tau_p} \hat{p}_{l,k}^b \leq P_{\mathrm{max},l,k}, \forall l,k, & (37c) \\
& p_{l,k} \leq P_{\mathrm{max},l,k}^d, \forall l,k. & (37d)
\end{aligned}$$

From the expression of the SINR constraints in (37b), we realize that the proposed optimization problem is a signomial program.[5] Therefore, the max-min fairness optimization problem is NP-hard in general and seeking the optimal solution has very high complexity in any non-trivial setup [31]. However, the power constraints (37c) and (37d) ensure a compact feasible domain and make the SINRs continuous functions of the optimization variables. According to Weierstrass' theorem [32], an optimal solution always exists.

For the alternative pilot design in (12), the max-min fairness optimization problem is formulated as

$$\begin{aligned}
\underset{\xi,\mathbf{A}\in\mathcal{D},\{\tilde{p}_{l,k}, p_{l,k} \geq 0\}}{\text{maximize}} \quad & \xi \\
\text{subject to} \quad & \widetilde{\mathrm{SINR}}_{l,k}^{\mathrm{MR}} \geq \xi, \forall l,k, \\
& \tilde{p}_{l,k} \leq \tau_p P_{\mathrm{max},l,k}, \forall l,k, \\
& p_{l,k} \leq P_{\mathrm{max},l,k}^d, \forall l,k.
\end{aligned} \quad (38)$$

The optimization problem (38) is non-convex since it contains a combinatorial pilot assignment selection. Fortunately the optimal solution to this problem can be obtained by looking up every instance $\mathbf{A}$ in the dictionary $\mathcal{D}$. For each $\mathbf{A}$ we attain the pilot reuse sets $\mathcal{P}_{l,k}^{\mathbf{A}}, \forall l,k$, and then convert (38) to a convex problem as shown in Corollary 3.

Corollary 3. *For a given pilot assignment matrix $\mathbf{A} \in \mathcal{D}$, (38) reduces to the geometric program*

$$\begin{aligned}
\underset{\xi,\{\tilde{p}_{l,k}, p_{l,k} \geq 0\}}{\text{maximize}} \quad & \xi \\
\text{subject to} \quad & \widetilde{\mathrm{SINR}}_{l,k}^{\mathrm{MR}} \geq \xi, \forall l,k, \\
& \tilde{p}_{l,k} \leq \tau_p P_{\mathrm{max},l,k}, \forall l,k, \\
& p_{l,k} \leq P_{\mathrm{max},l,k}^d, \forall l,k.
\end{aligned} \quad (39)$$

The optimal solution to (39) is obtained in polynomial time due to its convexity. By checking every instance $\mathbf{A}$ in the dictionary $\mathcal{D}$ and solving the corresponding problem (39), the global optimum to (38) is obtained as the highest objective value to (39).

In more detail, the globally optimal solution to (38) is obtained as shown in Algorithm 1. The ith iteration seeks the optimal solution $\xi^{(i),\mathrm{opt}}, \tilde{p}_{l,k}^{(i),\mathrm{opt}}$, and $p_{l,k}^{(i),\mathrm{opt}}, \forall l,k$ for given $\{\mathcal{P}_{l,k}^{\mathbf{A}}\}$ by considering (39) as the main cost function. The algorithm is terminated when the iteration index equals $(K!)^{L-1}$. The global optimum to the pilot and data power control together with the pilot reuse set are obtained from the maximum values of all $\{\xi^{(i),\mathrm{opt}}\}$. This is a practical issue. We are indeed able to find the solution, but it will take very long time. Algorithm 1 is computationally heavy since the number of iterations grows rapidly with K and L, but it obtains the global optimum to the max-min SE problem (38). Specifically, the main cost of each iteration in Algorithm 1 is the geometric program (40) which includes $2KL + 1$ optimization variables and $3KL$ constraints. Based on [33], in general, the computational complexity of this algorithm is of the order of

$$O\left((K!)^{L-1} \max\{(2KL+1)^3, 3KL(2KL+1)^2, F_1\}\right), \quad (41)$$

where F_1 is the cost of evaluating the first and second derivatives of the objective and constraint functions in (40). Therefore, this approach will serve as a benchmark for comparison in Section VII. For the sake of completeness, we also include another benchmark whereas the data powers are fixed at their maximum value then Algorithm 1 is solved with respect to the remaining pilot power variables, as was done in our previous work [27].

B. Local Optimality Algorithm

This subsection provides a method to obtain a local optimum to the optimization problem (37). To end this, the

---

[4]The optimization problem (36) requires coordination among the cells to be solved, but the main target in this paper is to investigate how much the max-min fairness SE can be improved in multi-cell Massive MIMO by joint pilot design and UL power control. One potential way to deal with practical limitations such as backhaul signaling, delays, and scalability is to implement the optimization problem in a distributed manner using dual/primal decomposition [30].

[5]A function $f(x_1, \ldots, x_{N_1}) = \sum_{n=1}^{N_2} c_n \prod_{m=1}^{N_1} x_m^{a_{n,m}}$ defined in $\mathbb{R}_+^{N_1}$ is signomial with N_2 terms ($N_2 \geq 2$) if the exponents $a_{n,m}$ are real numbers and the coefficients c_n are also real but at least one must be negative. In case all $c_n, \forall n$, are positive, $f(x_1, \ldots, x_{N_1})$ is a posynomial function.

Algorithm 1 Global solution to (38) by exhaustive search

Input: Set $i = 1$; Select the initial values of $P_{\max,l,k}$ and $P^d_{\max,l,k}$ for $\forall k, l$; Set up the dictionary $\mathcal{D}$.

1. *Iteration i:*

    1.1. Assign the reuse pilot set index $\mathcal{P}^{\mathbf{A}}_{l,k}, \forall l, k$, by an instance $\mathbf{A} \in \mathcal{D}$.

    1.2. Solve the following geometric program to obtain $\xi^{(i),\text{opt}}, p^{(i),\text{opt}}_{l,k}$, and $\tilde{p}^{(i),\text{opt}}_{l,k}, \forall l, k$:

$$\underset{\xi^{(i)}, \{\tilde{p}^{(i)}_{l,k}, p^{(i)}_{l,k} \geq 0\}}{\text{maximize}} \quad \xi^{(i)}$$

$$\text{subject to} \quad \widetilde{\text{SINR}}^{(i),\text{MR}}_{l,k} \geq \xi^{(i)}, \forall l, k, \quad (40)$$

$$\tilde{p}^{(i)}_{l,k} \leq \tau_p P_{\max,l,k}, \forall l, k,$$

$$p^{(i)}_{l,k} \leq P^d_{\max,l,k}, \forall l, k.$$

2. If $i = (K!)^{L-1} \to$ Stop. Otherwise, go to Step 3.
3. Restore $\xi^{(i),\text{opt}}, \tilde{p}^{(i),\text{opt}}_{l,k}$, and $p^{(i),\text{opt}}_{l,k}$. Set $i = i + 1$, then go to Step 1.

Output: Set $i^{\text{opt}} = \underset{i}{\arg\max}\{\xi^{(i),\text{opt}}\}$, then the optimal solutions:

$\xi^{\text{opt}} = \xi^{(i^{\text{opt}}),\text{opt}}, \tilde{p}^{\text{opt}}_{l,k} = \tilde{p}^{(i^{\text{opt}}),\text{opt}}_{l,k}$, and $p^{\text{opt}}_{l,k} = p^{(i^{\text{opt}}),\text{opt}}_{l,k}, \forall l, k$.

---

signomial SINR constraints are converted to monomial ones by using the weighted arithmetic mean-geometric mean inequality [34] stated in Lemma 3.[6]

Lemma 3. *[34, Lemma 1] Assume that a posynomial function $g(x)$ is defined from the set of τ_p monomials $\{u_1(x), \ldots, u_{\tau_p}(x)\}$ as*

$$g(x) = \sum_{b=1}^{\tau_p} u_b(x), \quad (42)$$

then it is lower bounded by a monomial function $\tilde{g}(x)$ as

$$g(x) \geq \tilde{g}(x) = \prod_{b=1}^{\tau_p} (u_b(x)/\alpha_b)^{\alpha_b}, \quad (43)$$

where α_b is a non-negative weight corresponding to $u_b(x)$. We say that $\tilde{g}(x_0)$ is the best approximation to $g(x_0)$ near the point x_0 in the sense of the first order Taylor expansion, if the weight α_b is selected as

$$\alpha_b = \frac{u_b(x_0)}{\sum_{b=1}^{\tau_p} u_b(x_0)}. \quad (44)$$

By using this lemma, the max-min fairness optimization problem (37) is converted to a geometric program by bounding the term $\sum_{b=1}^{\tau_p} \hat{p}^b_{l,k}$ in the numerators of the SINR constraints:

$$\sum_{b=1}^{\tau_p} \hat{p}^b_{l,k} \geq \prod_{b=1}^{\tau_p} \left(\hat{p}^b_{l,k}/\alpha^b_{l,k}\right)^{\alpha^b_{l,k}}, \quad (45)$$

where $\alpha^b_{l,k}$ is the weight value corresponding to $\hat{p}^b_{l,k}$. This leads to a lower bound on the SINR value for user k in cell l obtained as

$$\text{SINR}^{\text{MR}}_{l,k} \geq \widetilde{\text{SINR}}^{\text{MR}}_{l,k}, \quad (46)$$

---

[6] A function $f(x_1, \ldots, x_{N_1}) = c \prod_{m=1}^{N_1} x_m^{a_m}$ defined in $\mathbb{R}^{N_1}_+$ is monomial if the coefficient $c > 0$ and the exponents $a_m, \forall m$, are real numbers.

---

where the $\widetilde{\text{SINR}}^{\text{MR}}_{l,k}$ value is presented in (47).

The optimal solution ξ to the max-min SE optimization problem (37) is lower bounded by solving the geometric program

$$\underset{\xi, \{\hat{p}^b_{l,k}, p_{l,k} \geq 0\}}{\text{maximize}} \quad \xi$$

$$\text{subject to} \quad \widetilde{\text{SINR}}^{\text{MR}}_{l,k} \geq \xi, \forall l, k,$$

$$\frac{1}{\tau_p} \sum_{b=1}^{\tau_p} \hat{p}^b_{l,k} \leq P_{\max,l,k}, \forall l, k, \quad (48)$$

$$p_{l,k} \leq P^d_{\max,l,k}, \forall l, k.$$

By virtue of the successive approximation technique [35], a locally optimal Karush-Kuhn-Tucker (KKT) point to the max-min fairness optimization problem (37) can be obtained if we solve (48) iteratively as shown in Theorem 1.

Theorem 1. *Selecting an initial point $\hat{p}^{b,(0)}_{l,k}, \forall l, k, b$, in the feasible domain and solving (48) in an iterative manner by consecutively updating the weight values $\alpha^b_{l,k}$ from the optimal powers of the previous iteration, the solution will converge to a KKT local point to (37).*

Proof. The proof is adapted from the general framework in [35] and is sketched in Appendix B. $\square$

In particular, we first select the initial powers $\hat{p}^{b,(0)}_{l,k}, \forall l, k, b$ that satisfy $\hat{p}^{b,(0)}_{l,k} \geq 0, \sum_{b=1}^{\tau_p} \hat{p}^{b,(0)}_{l,k} \leq \tau_p P_{\max,l,k}$. Then the corresponding weight values are computed as in (44). Furthermore, in each iteration, the SINR constraints are converted to the corresponding monomials by bounding the pilot power of user k in cell l as in (47), by using the weight values computed from the optimal pilot powers in the previous iteration. The pilot and data allocation solution is obtained by solving the geometric program (48) before the weight values are updated again at the end of each iteration. We repeat the procedure until this algorithm has converged to a KKT point. The convergence can be declared, for example, when the variation between two consecutive iterations is sufficient small. The proposed algorithm for obtaining a locally optimal solution is summarized in Algorithm 2. Note that one can also fix the data powers and only optimize the pilot signals in Algorithm 2, as was done in our previous work [27]. Algorithm 2 involves optimization with $KL(\tau_p+1)+1$ variables and $3KL$ constraints, and it has a computational complexity of the order of [33] [7]

$$O\left(N \max\{(KL(\tau_p + 1) + 1)^3, 3KL(KL(\tau_p + 1) + 1)^2, F_2\}\right), \quad (49)$$

where F_2 is the cost of evaluating the first and second derivatives of the objective and constraint functions in (48). N is the number of iterations needed for this algorithm to converge to the KKT point. Even though each iteration in Algorithm 2 is more costly than in Algorithm 1 since we carefully design powers for all pilot signals, the successive approximation approach converges after only a few iterations.

---

[7] The exact complexity or the runtime of the proposed algorithms are not suitable metrics since they depend significantly on the computer configuration and how much time is spent to optimize the implementations. However (41) and (49) give basic insights into the general computational complexity scaling.

Algorithm 2 Successive approximation algorithm for (37)

Input: Set $i = 1$; Select the maximum powers $P_{\max,l,k}$ and $P^d_{\max,l,k}$ for $\forall l, k$; Select the initial values of powers $\hat{p}^{b,(0)}_{l,k}$ for $\forall l, k, b$; Compute the weight values $\alpha^{b,(1)}_{l,k} = \hat{p}^{b,(0)}_{l,k} / \sum_{b=1}^{\tau_p} \hat{p}^{b,(0)}_{l,k}, \forall l, k, b$.

1. *Iteration i*:
   1.1. Solve the geometric program (48) with $\alpha^b_{l,k} = \alpha^{b,(i)}_{l,k}$ to get the optimal values $\xi^{(i),\mathrm{opt}}, \hat{p}^{b,(i),\mathrm{opt}}_{l,k}, \forall l, k, b$, and $p^{(i),\mathrm{opt}}_{l,k}, \forall l, k$.
   1.2. Update the weight values: $\alpha^{b,(i+1)}_{l,k} = \hat{p}^{b,(i),\mathrm{opt}}_{l,k} / \sum_{b=1}^{\tau_p} \hat{p}^{b,(i),\mathrm{opt}}_{l,k}, \forall l, k, b$.
2. If Stopping criterion satisfied $\rightarrow$ Stop. Otherwise, go to Step 3.
3. Set $\xi^{\mathrm{opt}} = \xi^{(i),\mathrm{opt}}, \hat{p}^{b,\mathrm{opt}}_{l,k} = \hat{p}^{b,(i),\mathrm{opt}}_{l,k}, \forall l, k, b$, and $p^{\mathrm{opt}}_{l,k} = p^{(i),\mathrm{opt}}_{l,k}, \forall l, k$; Set $i = i + 1$, go to Step 1.

Output: The solutions $\xi^{\mathrm{opt}}, \hat{p}^{b,\mathrm{opt}}_{l,k}, \forall l, k, b$, and $p^{\mathrm{opt}}_{l,k}, \forall l, k$.

V. PILOT OPTIMIZATION FOR CELLULAR MASSIVE MIMO SYSTEMS WITH HARDWARE IMPAIRMENTS

The previous sections considered Massive MIMO with ideal transceiver hardware. However practical transceivers are non-ideal in the sense of creating distortions that can have a substantial impact on the SE. Thanks to the non-coherent combining of the independent distortion caused at each of the BS antennas, the hardware impairments at the BS can be neglected in Massive MIMO [36]. However, the distortion caused by the single-antenna users leads to coherent self-interference, which can be viewed as pilot contamination that the user causes to itself. The pilot optimization problem is fundamentally different when accounting for self-interference, because pilot contamination from distant users can now be neglected when it is substantially weaker than the self-interference. In this section, we investigate how the distortions from hardware impairments at the users affect the proposed pilot design and the optimization problems.

A. Channel Estimation under Hardware Impairments

We model the distortion caused by the hardware impairments at a user as a reduction of the signal amplitude by $\sqrt{1 - \epsilon^2}$ and the addition of Gaussian distortion with a power that equals the reduction in signal power [36]. We refer to ϵ as the impairment level. The received pilot signal at BS l is

$$\mathbf{Y}_l = \sum_{(i,t)\in\mathcal{S}} \mathbf{h}^l_{i,t} \left(\sqrt{1 - \epsilon^2}\boldsymbol{\psi}^H_{i,t} + \boldsymbol{\varepsilon}^H_{p,i,t} \right) + \mathbf{N}_l. \qquad (50)$$

Similar to [37], we assume that the UL distortion term of user t in cell i is distributed as

$$\boldsymbol{\varepsilon}_{p,i,t} \sim \mathcal{CN}\left(\mathbf{0}, \boldsymbol{\Lambda}_{i,t}\right), \qquad (51)$$

where $\boldsymbol{\Lambda}_{i,t} = \epsilon^2\mathrm{diag}(\hat{p}^1_{i,t}, \ldots, \hat{p}^{\tau_p}_{i,t})$. Since the distortion term is multiplied with the unknown channel, just as the pilot signals,

the channel estimation is more complicated in this case. An estimate $\hat{\mathbf{h}}_{l,k}$ of $\mathbf{h}_{l,k}$ channel can be obtained from

$$\mathbf{y}_{l,k} = \mathbf{Y}_l\boldsymbol{\psi}_{l,k} = $$
$$\sum_{(i,t)\in\mathcal{S}} \mathbf{h}^l_{i,t} \left(\sqrt{1 - \epsilon^2}\boldsymbol{\psi}^H_{i,t}\boldsymbol{\psi}_{l,k} + \boldsymbol{\varepsilon}^H_{p,i,t}\boldsymbol{\psi}_{l,k} \right) + \mathbf{N}_l\boldsymbol{\psi}_{l,k}, \qquad (52)$$

where the received signal in (50) is correlated with the pilot sequence used by the user of interest. We note that the MMSE estimator is intractable, but we can derive the linear minimum mean square error (LMMSE) estimator as shown in Lemma 4.

Lemma 4. *Under hardware impairments, the LMMSE channel estimate of $\mathbf{h}^l_{l,k}$ at BS l is*

$$\hat{\mathbf{h}}^l_{l,k} = \frac{\sqrt{1 - \epsilon^2}\beta^l_{l,k}\sum_{b=1}^{\tau_p}\hat{p}^b_{l,k}}{\sum_{(i,t)\in\mathcal{S}}\beta^l_{i,t}\kappa_{i,t} + \sigma^2\sum_{b=1}^{\tau_p}\hat{p}^b_{l,k}}\mathbf{y}_{l,k}, \qquad (53)$$

where

$$\kappa_{i,t} = (1 - \epsilon^2)\left(\sum_{b=1}^{\tau_p}\sqrt{\hat{p}^b_{i,t}\hat{p}^b_{l,k}}\right)^2 + \epsilon^2\sum_{b=1}^{\tau_p}\hat{p}^b_{i,t}\hat{p}^b_{l,k}. \qquad (54)$$

The channel estimate $\hat{\mathbf{h}}^l_{l,k}$ and estimation error $\mathbf{e}^l_{l,k}$ are uncorrelated, but not independent, have zero mean, and the covariance matrices as

$$\mathrm{Cov}\{\hat{\mathbf{h}}^l_{l,k}, \hat{\mathbf{h}}^l_{l,k}\} = \frac{(1 - \epsilon^2)(\beta^l_{l,k})^2\left(\sum_{b=1}^{\tau_p}\hat{p}^b_{l,k}\right)^2}{\sum_{(i,t)\in\mathcal{S}}\beta^l_{i,t}\kappa_{i,t} + \sigma^2\sum_{b=1}^{\tau_p}\hat{p}^b_{l,k}}\mathbf{I}_M, \qquad (55)$$

$$\mathrm{Cov}\{\mathbf{e}^l_{l,k}, \mathbf{e}^l_{l,k}\} = \frac{\sum_{(i,t)\in\mathcal{S}\setminus(l,k)}\beta^l_{i,t}\kappa_{i,t} + \sigma^2\sum_{b=1}^{\tau_p}\hat{p}^b_{l,k} + \epsilon^2\sum_{b=1}^{\tau_p}(\hat{p}^b_{l,k})^2}{\sum_{(i,t)\in\mathcal{S}}\beta^l_{i,t}\kappa_{i,t} + \sigma^2\sum_{b=1}^{\tau_p}\hat{p}^b_{l,k}}\mathbf{I}_M. $$
$$(56)$$

Proof. The proof follows the standard LMMSE estimation technique as described in [29]. $\qquad\square$

This lemma shows that the variance of the estimation error grows with the impairment level. In the special case of $\epsilon = 0$, we obtain ideal hardware and the estimate is the same as in Lemma 1. In contrast, the estimate is zero (equal to the mean value) if $\epsilon = 1$. We now use the statistics in Lemma 4 to derive a lower bound on the UL achievable SE and formulate the corresponding optimization problems.

B. UL Data Transmission and Max-min Fairness Optimization under Hardware Impairments

Similar to [36], the received signal at BS l during data transmission for the case of hardware impairments at the users is modeled as

$$\mathbf{y}_l = \sum_{(i,t)\in\mathcal{S}} \mathbf{h}^l_{i,t} \left(\sqrt{(1 - \epsilon^2)p_{i,t}}x_{i,t} + \varepsilon_{i,t} \right) + \mathbf{n}_l. \qquad (57)$$

We assume that the distortion caused by user t in cell i is worst-case Gaussian distributed with $\varepsilon_{i,t} \sim \mathcal{CN}(0, \epsilon^2 p_{i,t})$. A

lower bound on the UL ergodic capacity of user k in cell l is obtained in Theorem 2.

Theorem 2. *Under hardware impairments, if the system uses the pilot structure in* (7) *and MR detection, the SE in* (28) *for user k in cell l becomes*

$$R_{l,k} = \left(1 - \frac{\tau_p}{\tau_c}\right) \log_2 \left(1 + \overline{\text{SINR}}_{l,k}^{\text{MR}}\right), \tag{58}$$

where the effective SINR value, $\overline{\text{SINR}}_{l,k}^{\text{MR}}$ *, is*

$$\frac{M(1-\epsilon^2)^2 p_{l,k}(\beta_{l,k}^l)^2 \left(\sum\limits_{b=1}^{\tau_p} \hat{p}_{l,k}^b\right)^2}{\left(\sum\limits_{(i,t)\in\mathcal{S}} \beta_{i,t}^l \kappa_{i,t} + \sigma^2 \sum\limits_{b=1}^{\tau_p} \hat{p}_{l,k}^b\right) \left(\sum\limits_{(i,t)\in\mathcal{S}} p_{i,t}\beta_{i,t}^l + \sigma^2\right) + \eta_{l,k}}, \tag{59}$$

where

$$\eta_{l,k} = M \sum_{(i,t)\in\mathcal{S}\backslash(l,k)} \kappa_{i,t} p_{i,t}(\beta_{i,t}^l)^2 + M\epsilon^2 p_{l,k}(\beta_{l,k}^l)^2 \sum_{b=1}^{\tau_p}(\hat{p}_{l,k}^b)^2$$
$$+ M\epsilon^2(1-\epsilon^2) \left(\sum_{b=1}^{\tau_p} \hat{p}_{l,k}^b\right)^2 p_{l,k}(\beta_{l,k}^l)^2. \tag{60}$$

Proof. The proof is given in Appendix C. □

In comparison to having ideal hardware, the hardware impairments reduce the coherent gain in the numerator of the SINR by a factor $(1-\epsilon^2)^2$. There is now coherent self-interference from user k, which behave similarly to pilot contamination. This contamination can be relatively large since the user generally has a stronger channel than the pilot-contaminating interferers. In the asymptotic limit when $M \to \infty$, only the signal term in the numerator and the term $\eta_{l,k}$ with self-interference and coherent interference from pilot contamination remain.

Based on the SE expression with hardware impairments, we consider the max-min fairness optimization problem

$$\begin{aligned}
\underset{\xi, \{p_{l,k}, \hat{p}_{l,k}^b \geq 0\}}{\text{maximize}} \quad & \xi \\
\text{subject to} \quad & \overline{\text{SINR}}_{l,k}^{\text{MR}} \geq \xi, \forall l, k, \\
& \frac{1}{\tau_p} \sum_{b=1}^{\tau_p} \hat{p}_{l,k}^b \leq P_{\text{max},l,k}, \forall l, k, \\
& p_{l,k} \leq P_{\text{max},l,k}^d, \forall l, k.
\end{aligned} \tag{61}$$

We stress that (61) is a generalization of (37), but all algorithms we proposed for the case of ideal hardware can be readily extended. Similar to (37), by utilizing the successive approximation method similar to Algorithm 2, a local optimum to (61) is obtained in polynomial time. The performance with heuristic pilot designs and power allocation is obtained directly from Theorem 2 by using the corresponding values on $\hat{p}_{l,k}^b$ and for example, in the case of using the combinatorial pilot structure in (12), similar to the procedures in Algorithm 1, the global max-min SE solution under hardware impairments is obtained by jointly optimizing pilot and data powers with exhaustive search over all $\mathbf{A} \in \mathcal{D}$.

VI. GENERALIZATION TO CORRELATED RAYLEIGH FADING

Since the propagation channels may be spatially correlated in practice, we now consider a correlation model where the channel between user t in cell i and BS l is modeled as

$$\mathbf{h}_{i,t}^l \sim \mathcal{CN}\left(\mathbf{0}, \mathbf{R}_{i,t}^l\right), \tag{62}$$

where $\mathbf{R}_{i,t}^l \in \mathbb{C}^{M \times M}$ is the covariance matrix with equal diagonal elements denoted by $\beta_{i,t}^l$ since the BS antennas are co-located. This assumption also leads to convex optimization problems. Meanwhile the non-zero off-diagonal elements represent the spatial correlation. By using the pilot transmission model in (16) and element-wise MMSE estimation [24], [38], the channel estimate of $\mathbf{h}_{l,k}^l$ is

$$\hat{\mathbf{h}}_{l,k}^l = \varrho_{l,k}^l \mathbf{y}_{l,k}, \tag{63}$$

where

$$\varrho_{l,k}^l = \frac{\beta_{l,k}^l \sum\limits_{b=1}^{\tau_p} \hat{p}_{l,k}^b}{\sum\limits_{(i,t)\in\mathcal{S}} \beta_{i,t}^l \left(\sum\limits_{b=1}^{\tau_p} \sqrt{\hat{p}_{i,t}^b \hat{p}_{l,k}^b}\right)^2 + \sigma^2 \sum\limits_{b=1}^{\tau_p} \hat{p}_{l,k}^b}. \tag{64}$$

After that, a closed-form expression of the achievable SE is obtained by computing the moments of Gaussian distributions in (29):

$$R_{l,k}^{\text{MR}} = \left(1 - \frac{\tau_p}{\tau_c}\right) \log_2 \left(1 + \widehat{\text{SINR}}_{l,k}^{\text{MR}}\right), \tag{65}$$

where the effective SINR, denoted by $\widehat{\text{SINR}}_{l,k}^{\text{MR}}$, is

$$\widehat{\text{SINR}}_{l,k}^{\text{MR}} = \frac{M(\beta_{l,k}^l)^2 p_{l,k} \left(\sum_{b=1}^{\tau_p} \hat{p}_{l,k}^b\right)^2}{\mathcal{I}_{\text{non}} + \mathcal{I}_{\text{coh}}}. \tag{66}$$

By denoting the trace of a matrix as $\text{tr}(\cdot)$, the non-coherent interference and thermal noise $\mathcal{I}_{\text{non}}$, and the coherent interference $\mathcal{I}_{\text{coh}}$ in (66) are respectively given as

$$\mathcal{I}_{\text{non}} = \sum_{(i,t)\in\mathcal{S}} \frac{p_{i,t}}{M} \sum_{(i',t')\in\mathcal{S}} \text{tr}\left(\mathbf{R}_{i,t}^l \mathbf{R}_{i',t'}^l\right) \left(\sum_{b=1}^{\tau_p} \sqrt{\hat{p}_{i',t'}^b \hat{p}_{l,k}^b}\right)^2$$
$$+ \sigma^2 \sum_{b=1}^{\tau_p} \hat{p}_{l,k}^b \sum_{(i,t)\in\mathcal{S}} p_{i,t}\beta_{i,t}^l + \frac{\beta_{l,k}^l \sigma^2 \sum_{b=1}^{\tau_p} \hat{p}_{l,k}^b}{\varrho_{l,k}^l}, \tag{67}$$

$$\mathcal{I}_{\text{coh}} = M \sum_{(i,t)\in\mathcal{S}\backslash(l,k)} p_{i,t}(\beta_{i,t}^l)^2 \left(\sum_{b=1}^{\tau_p} \sqrt{\hat{p}_{i,t}^b \hat{p}_{l,k}^b}\right)^2. \tag{68}$$

From the SINR expression in (66), the correlation between channels only effects the non-coherent interference while the coherent interference remains the same as in the case of uncorrelated Rayleigh fading. We now can formulate the max-min fairness optimization problem as

$$\begin{aligned}
\underset{\xi, \{p_{l,k}, \hat{p}_{l,k}^b \geq 0\}}{\text{maximize}} \quad & \xi \\
\text{subject to} \quad & \widehat{\text{SINR}}_{l,k}^{\text{MR}} \geq \xi, \forall l, k, \\
& \frac{1}{\tau_p} \sum_{b=1}^{\tau_p} \hat{p}_{l,k}^b \leq P_{\text{max},l,k}, \forall l, k, \\
& p_{l,k} \leq P_{\text{max},l,k}^d, \forall l, k.
\end{aligned} \tag{69}$$

The optimization problem (69) has the same general structure as the problems considered in Section IV. For example, with the proposed pilot design, we can directly use Algorithm 2 to obtain a local optimum, while the global solution of the max-min fairness optimization can obtained with the pilot design in (12) by making simple modifications to Algorithm 1.

VII. NUMERICAL RESULTS

In this section, we use numerical simulations to quantify and discuss the effectiveness of the proposed pilot designs, using the exact closed-form expressions of the SE in (31), (34), (58), and (65). A Massive MIMO system with a coverage area 1 km^2 comprising of four square cells is considered. In each cell, a BS is located at the center, while the K users are uniformly distributed at distance not closer to the BS than 35 m. To even out interference, the coverage area is wrapped around, and therefore one BS has eight neighbors. The coherence interval contains 200 symbols and the system bandwidth is 20 MHz. The noise variance is -96 dBm. We deploy the 3GPP LTE model from [39] where the large-scale fading coefficient $\beta_{i,t}^l$ [dB] is

$$\beta_{i,t}^l = -148.1 - 37.6 \log_{10} d_{i,t}^l + z_{i,t}^l, \tag{70}$$

where, $d_{i,t}^l$ denotes the distance in km between user t in cell i and BS l. The shadow fading $z_{i,t}^l$ has a Gaussian distribution with zero mean and the standard derivation 7 dB.[8] The maximum pilot and data power constraints are $P_{\max,l,k} = P_{\max,l,k}^d = 200$ mW, $\forall l, k$.

For Algorithm 2 and its modification with fixed data power (which was considered in [27]), we observe better performance with a hierarchical initialization of $p_{l,k}^{b,(0)}$ than with an all-equal initialization. Consequently, we initialize $p_{l,k}^{b,(0)}$ as uniformly distributed over the range $[0; P_{\max,l,k}]$.[9] Algorithm 2 converges quite fast, so the stopping criteria was specified in number of iterations (e.g., 15 iterations). The proposed algorithms are compared with related works and exhaustive search:

(i) *Random pilot assignment*, as considered in [11], [16]. The same pilots are reused in every cell and assigned randomly to the users within the cell. Equal pilot and data powers of 200 mW are used by all users. This method is denoted as *Random P. assignment* in the figures.

(ii) *Smart pilot assignment*, as proposed in [20]. Orthogonal pilots are assumed within a cell and reused in every cell. They are assigned to the users based on the mutual interference information, determined by the large-scale fading coefficients. Equal pilot and data powers 200 mW are used by all users. This method is denoted as *Smart P. assignment* in the figures.

(iii) *Pilot power optimization with exhaustive search*, as proposed in [27], utilizes the pilot structure in (12) and the optimal solution to the max-min SE is obtained by a

[8]Shadow fading realizations were sometimes regenerated to ensure that the home BS has the largest large-scale fading to its users (i.e., $\beta_{l,k}^l$ is the maximum over all $\beta_{i,k}^l$, $i = 1, \ldots, L$). This makes sure that the coverage area of each BS is a square also with shadow fading, while retaining the macro-diversity towards shadow fading that exists in practice.

[9]The maximum power settings indicate relatively low median SNRs at the cell-edge users of the network.
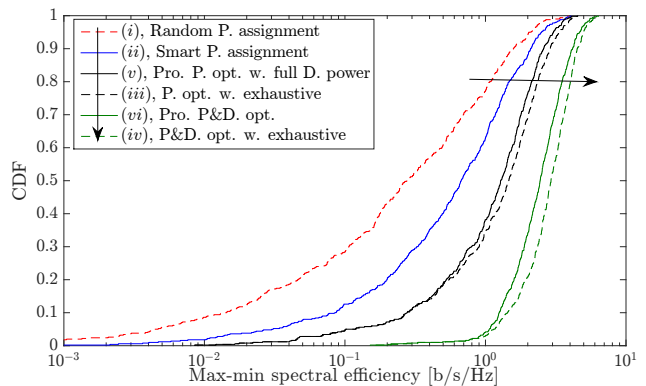

Fig. 2. Cumulative distribution function (CDF) versus the max-min SE [b/s/Hz] with $M = 300$, $K = \tau_p = 2$.
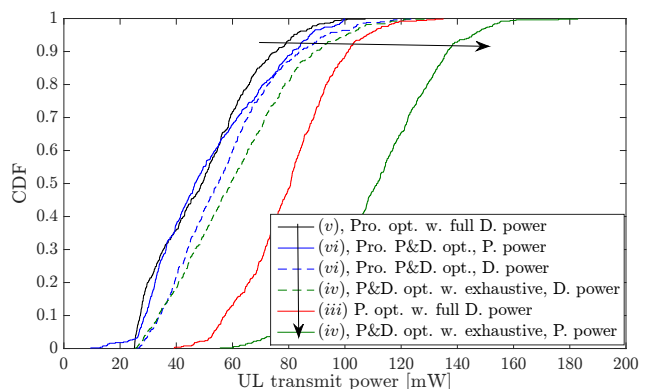

Fig. 3. Cumulative distribution function (CDF) versus the UL transmit power [mW] with $M = 300$, $K = \tau_p = 2$.

modification of Algorithm 1 with fixed data powers of $p_{l,k} = 200$ mW at all users. This method is denoted as *P. opt. w. exhaustive* in the figures.

(iv) *Joint pilot and data power optimization with exhaustive search* utilizes the pilot structure in (12) and the optimal solution to the max-min SE is obtained as shown in Algorithm 1. This method is denoted as *P&D opt. w. exhaustive* in the figures.

(v) *Proposed pilot optimization with full data power*, as presented in [27], utilizes the pilot structure in (7) and the local optimum to the max-min SE is obtained by a modification of Algorithm 2 with fixed data powers of $p_{l,k} = 200$ mW at all users. This method is denoted as *Pro. P. opt. w. full D. power* in the figures.

(vi) *Proposed joint pilot and data power optimization* utilizes the pilot structure in (7) and the local optimum to the max-min SE is obtained as shown in Algorithm 2. This method is denoted as *Pro. P&D opt.* in the figures.

The SE is measured over different random user locations and shadow fading realizations. The SE achieved by (i)–(iv) are also averaged over different pilot reuse locations, while for (v) and (vi) the SE is also averaged over different initializations of $\hat{p}_{l,k}^{b,(0)}, \forall l, k, b$. The solutions to the optimization problems are obtained by utilizing the MOSEK solver [40] in CVX [41]. Fig. 2 shows the cumulative distribution
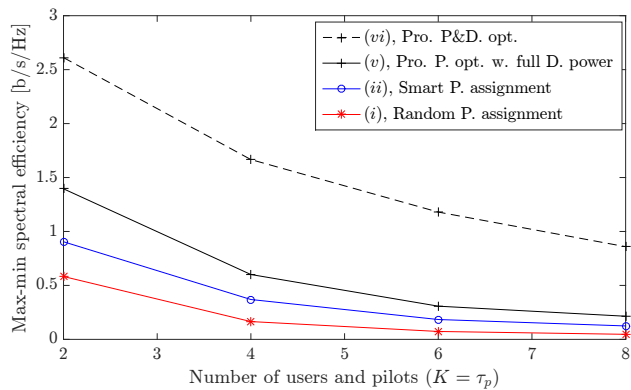
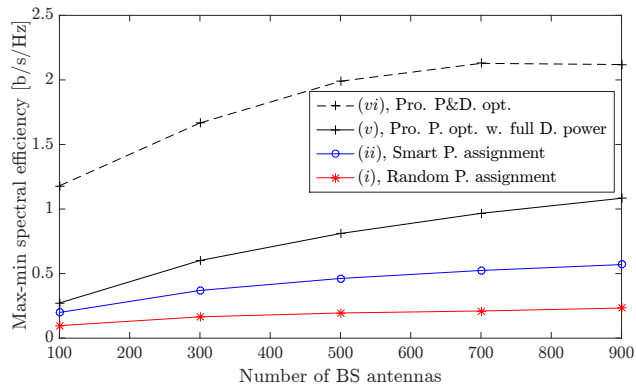Fig. 4. Max-min SE [b/s/Hz] versus the number of users and pilots, $M = 300$.

Fig. 6. Max-min SE [b/s/Hz] versus the length of pilot signals, $K = 4$.

Fig. 5. Max-min SE [b/s/Hz] versus the number of BS antennas, $K = \tau_p = 4$.

Fig. 7. Max-min SE [b/s/Hz] versus the impairment level, $M = 300, K = \tau_p = 2$.

function (CDF) of the max-min SE level [b/s/Hz] for the case $K = \tau_p = 2$ and $M = 300$. Random pilot assignment yields the worst performance due to the pilot contamination and mutual interference. At the 95%-likely point, smart pilot assignment brings significant improvements: it is about 6.86× better since it attempts to reduce the mutual interference between the users [20]. Although the performance of smart pilot assignment is very close to optimal pilot assignment with exhaustive search for a predetermined pilot power level [20], by jointly optimizing power and allocation of the pilot signals, the proposed method outperforms smart pilot assignment by providing an additional 1.53× gain in average max-min SE. Fig. 2 also demonstrates the superiority of jointly optimizing both data and pilot powers. It yields 1.86× higher average SE than optimizing only the pilots. Furthermore, the similar performance between the proposed pilot design and exhaustive search confirms the effectiveness of the proposed algorithms for finding locally optimal solutions. The proposed approach for joint optimization pilot and data power on average achieves 84% of the SE with the optimal solution, with a substantially lower computational complexity.

Fig. 3 plots the CDF of power allocated to each pilot/data symbol with $K = \tau_p = 2$ and $M = 300$. We do not include random pilot assignment and smart pilot assignment because they consume full power 200 mW. We observe a vast energy saving for the remaining methods. For the pilot and data power control with exhaustive search, the power
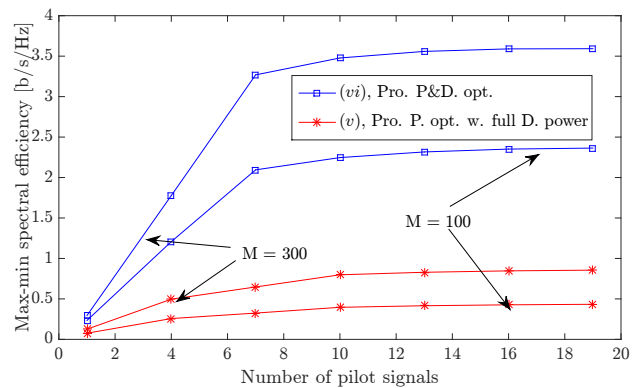
consumption on pilot signaling is on average only 111 mW. The proposed design without data power control, each pilot symbol on average only spends 50 mW. In comparison to the previous pilot design (e.g., the max-min SE in Fig. 2 and the corresponding power consumption in Fig. 3), we conclude that the proposed pilot structure can reduce the transmit power more significantly than the prior works while producing better max-min SE. This is because the prior works mitigate pilot contamination by assigning the orthogonal pilot signals to the users; that is, each user assigns power to only one of the basis vectors. In contrast, our design allows for non-orthogonal pilot signals with increasing feasible domain to minimize the total coherent interference in the network. By counting 95% of the pilot energy, the average number of non-zero pilot symbols in each pilot signal is on average about 1.02 and 1.32 for $\tau_p = K = 2$ and $\tau_p = K = 4$, respectively. Hence, some users have non-orthogonal pilots. The pilot power optimization with exhaustive search has a computational complexity that grows exponentially in K. When increasing the number of users, we henceforth only compare the three pilot designs: proposed pilot design, smart pilot assignment, and random pilot assignment. Fig. 4 displays the max-min SE as a function of the number of users and pilots with $M = 300, K = \tau_p$. Our proposed design yields the highest performance among the related algorithms. Specifically, in comparison to random pilot assignment, the improvement varies from 2.39× with $K = \tau_p = 2$ to 4.58×
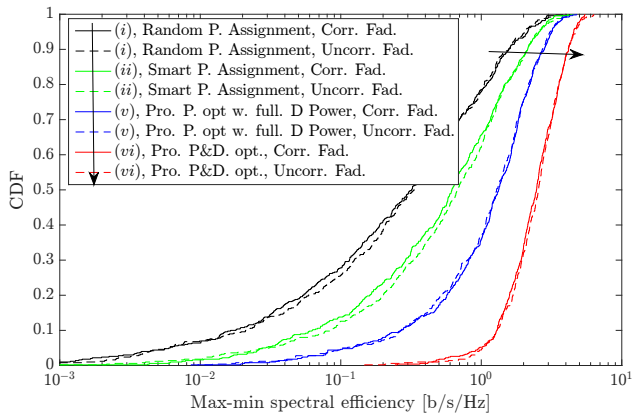
Fig. 8. Max-min SE [b/s/Hz] for the cases of uncorrelated Rayleigh fading and correlated Rayleigh fading, $M = 300$, $K = \tau_p = 2$.

with $K = \tau_p = 8$. Even though smart pilot assignment also performs much better than random pilot assignment, our proposed scheme provides substantial improvements over the state-of-the-art (e.g., we gain $1.72\times$ with $K = \tau_p = 8$ for fixed data power). The figure also demonstrates a significant improvement with jointly optimizing pilot and data power for all the tested cases of the user and pilot number. In comparison to the case of only optimizing the pilots, the performance gain is from $1.87\times$ to $4.01\times$ as the number of users and pilots increases from 2 to 8. In addition, we observe a dramatic reduction of the max-min SE of all the pilot designs when the number of users increases due to stronger mutual interference. According to Fig. 4, the performances of the different pilot designs and power control algorithms are about the same for large numbers of users.

Fig. 5 shows the max-min SE in the network as a function of the number of BS antennas. Among the three pilot structures, we still observe the worst performance with random pilot assignment, for which the max-min SE only increases from 0.10 [b/s/Hz] to 0.23 [b/s/Hz], when the number of antennas increases from 100 to 900 antennas. Our proposed pilot design obtains better performance than smart pilot assignment and the gap grows as more antennas BSs are added. For example, with 100 BS antennas, the smart pilot assignments yields a max-min SE 0.20 [b/s/Hz], while optimizing pilot signaling for fixed data powers gives 0.27 [b/s/Hz] and the joint optimization of pilot and data powers yields 1.18 [b/s/Hz]. The max-min SE is respectively 0.60, 1.08, and 2.12 [b/s/Hz] for the three methods when the BSs are equipped with 900 antennas. Hence the proposed pilot structure is especially suitable for Massive MIMO systems with the large ratios M/K. Moreover, due to the pilot contamination, a saturation of the SE at about 2.10 [b/s/Hz] is observed with the proposed joint optimization of pilot and data powers.

Only the special case $K = \tau_p$ has been investigated so far due to the high computational complexity of the pilot assignments from [17], [20]. In contrast, our algorithms can be applied for arbitrary lengths of the pilots. This is demonstrated in Fig. 6, which displays the max-min SE as a function of the length of the pilot signals. We observe noticeable gains when

the pilot length increases for both (v) and (vi). For instance, when the pilot length per coherence interval increases from 1 to 19, we observe a gain up to $12.17\times$ with the proposed joint pilot and data power optimization. Moreover, the pilot signals with lengths $\tau_p > KL$ still bring an increasing SE since it produces better channel estimation quality. However the gain is not significant, and therefore $\tau_p = KL$ is a good choice to eliminate pilot contamination and maximize the minimum SE if we have enough time-frequency resources.

The impact of hardware impairments is shown in Fig. 7. Increasing the impairment level leads to a decrease in max-min SE, but the reduction is not dramatic since we are operating at low SE where the performance is already interference limited, so the additional self-interference has no dominant impact. When the impairment level reaches 0.2, there is only a reduction of about $15.04\%, 16.51\%, 18.88\%$, and 28.87% for random pilot assignment, smart pilot assignment, proposed pilot design with full data power, and proposed joint pilot and data power optimization respectively. This impairment level is hardly found in practice. For a more realistic impairment level of $\epsilon = 0.1$, the reduction only varies from 3.61% to 7.27%. Furthermore that hardware impairments reduce more the performance of the algorithms giving higher SEs. Finally, we stress that our proposed algorithms provide substantial performance gains also under hardware impairments.

Motivated by the fact that practical channels may be spatially correlated, Fig. 8 shows the max-min SE of the pilot design (please refer to Section VI for the analysis) where the channel between user i in cell t and BS l is modeled by using the exponential correlation model [42]. In particular, $\mathbf{h}_{i,t}^l \sim \mathcal{CN}(\mathbf{0}, \mathbf{R}_{i,t}^l)$ and $\mathbf{R}_{i,t}^l$ is defined as

$$\mathbf{R}_{i,t}^l = \beta_{i,t}^l \begin{bmatrix} 1 & r_{i,t}^{l,*} & \cdots & (r_{i,t}^{l,*})^{M-1} \\ r_{i,t}^l & 1 & \cdots & (r_{i,t}^{l,*})^{M-2} \\ \vdots & \vdots & \ddots & \vdots \\ (r_{i,t}^l)^{M-1} & (r_{i,t}^l)^{M-2} & \cdots & 1 \end{bmatrix}, \quad (71)$$

where the correlation coefficient $r_{i,t}^l = \rho_{i,t}^l e^{\theta_{i,t}^l}$ and $\theta_{i,t}^l$ is the angle between the BS-user vector and the horizontal line. The coefficient magnitude $\rho_{i,t}^l$ is in the range of $[0, 1]$ and it equals 0.5 in our simulation. The figure demonstrates that our pilot design provides much higher max-min SE than the other pilot designs. Although the spatial correlation affects the user channels, the differences in SE are not significant compared to uncorrelated Rayleigh fading.

VIII. CONCLUSION

This paper proposed a novel pilot design for cellular Massive MIMO systems and combines the pilot assignment and uplink power allocation into a unified optimization framework. A key difference from prior work is that we treat the pilot signals as continuous optimization variables, instead of predefined vectors that should be assigned combinatorially. We used the proposed pilot structure to compute a new SE on the UL ergodic SE for Rayleigh fading channels with MR detection, both with ideal hardware and hardware impairments. A general max-min SE optimization problem was formulated. Although

finding the globally optimal solution is NP-hard, we developed an algorithm that finds a local optimum that outperforms the previous state-of-the-art algorithms for pilot assignment and is close to the optimum obtained by the exhaustive search. The numerical results demonstrate the importance of jointly optimizing the pilots and transmit powers in order to improve the max-min SE.

APPENDIX

A. Proof of Lemma 2

The SINR value in the theorem is obtained by computing the expectations in (29). Due to the independence of the channel estimate and estimation error, the numerator of (29) is computed as

$$
\begin{aligned}
p_{l,k} |\mathbb{E}\{\mathbf{v}_{l,k}^H \mathbf{h}_{l,k}^l\}|^2 &= p_{l,k} |\mathbb{E}\{||\hat{\mathbf{h}}_{l,k}^l||^2\}|^2 \\
&= p_{l,k} \left(M \beta_{l,k}^l \varrho_{l,k}^l \sum_{b=1}^{\tau_p} \hat{p}_{l,k}^b \right)^2.
\end{aligned}
\tag{72}
$$

Note that this term also appears in the denominator of (29). The first expectation in the denominator is reformulated as

$$
\begin{aligned}
\mathbb{E}\{|\mathbf{v}_{l,k}^H \mathbf{h}_{i,t}^l|^2\} / (\varrho_{l,k}^l)^2 &= \mathbb{E}\left\{ |\mathbf{y}_{l,k}^H \mathbf{h}_{i,t}^l|^2 \right\} \\
&= \mathbb{E}\left\{ \left| \sum_{(i',t') \in \mathcal{S}} \sum_{b=1}^{\tau_p} \sqrt{\hat{p}_{l,k}^b \hat{p}_{i',t'}^b} (\mathbf{h}_{i',t'}^l)^H \mathbf{h}_{i,t}^l \right|^2 \right\} \\
&\quad + \mathbb{E}\left\{ \left| (\mathbf{N}_l \boldsymbol{\psi}_{l,k})^H \mathbf{h}_{i,t}^l \right|^2 \right\}
\end{aligned}
\tag{73}
$$

by utilizing the independence of the noise and the channels. Decomposing the first expectation into two parts, one related to $\mathbf{h}_{i,t}^l$ and the other comprises of the remaining channels, we obtain

$$
\begin{aligned}
&\mathbb{E}\left\{ \left| \sum_{(i',t') \in \mathcal{S}} \sum_{b=1}^{\tau_p} \sqrt{\hat{p}_{l,k}^b \hat{p}_{i',t'}^b} (\mathbf{h}_{i',t'}^l)^H \mathbf{h}_{i,t}^l \right|^2 \right\} \\
&= \mathbb{E}\left\{ \left| \sum_{b=1}^{\tau_p} \sqrt{\hat{p}_{l,k}^b \hat{p}_{i,t}^b} (\mathbf{h}_{i,t}^l)^H \mathbf{h}_{i,t}^l \right|^2 \right\} \\
&\quad + \mathbb{E}\left\{ \left| \sum_{(i',t') \in \mathcal{S}_{i,t} \setminus (i,t)} \sum_{b=1}^{\tau_p} \sqrt{\hat{p}_{l,k}^b \hat{p}_{i',t'}^b} (\mathbf{h}_{i',t'}^l)^H \mathbf{h}_{i,t}^l \right|^2 \right\} \\
&\overset{(a)}{=} M(M+1) \left(\sum_{b=1}^{\tau_p} \sqrt{\hat{p}_{l,k}^b \hat{p}_{i,t}^b} \right)^2 (\beta_{i,t}^l)^2 \\
&\quad + M \sum_{(i',t') \in \mathcal{S}_{i,t} \setminus (i,t)} \left(\sum_{b=1}^{\tau_p} \sqrt{\hat{p}_{l,k}^b \hat{p}_{i',t'}^b} \right)^2 \beta_{i',t'}^l \beta_{i,t}^l \\
&= M^2 \left(\sum_{b=1}^{\tau_p} \sqrt{\hat{p}_{l,k}^b \hat{p}_{i,t}^b} \right)^2 (\beta_{i,t}^l)^2 \\
&\quad + M \sum_{(i',t') \in \mathcal{S}} \left(\sum_{b=1}^{\tau_p} \sqrt{\hat{p}_{l,k}^b \hat{p}_{i',t'}^b} \right)^2 \beta_{i',t'}^l \beta_{i,t}^l,
\end{aligned}
\tag{74}
$$

where (a) is obtained by using Lemma 2.9 in [43] to compute the fourth-order moment $\mathbb{E}\{||\mathbf{h}_{i,t}^l||^4\}$. The second expectation

in (73) is computed from the independence property between additive noise and the original channel $\mathbf{h}_{i,t}^l$ as

$$
\mathbb{E}\{| (\mathbf{N}_l \boldsymbol{\psi}_{l,k})^H \mathbf{h}_{i,t}^l|^2\} = M \beta_{i,t}^l \sigma^2 \sum_{b=1}^{\tau_p} \hat{p}_{l,k}^b.
\tag{75}
$$

Plugging (74) and (75) into (73), the first expectation in the denominator of (29) is

$$
\begin{aligned}
\mathbb{E}\{|\mathbf{v}_{l,k}^H \mathbf{h}_{i,t}^l|^2\} = M (\varrho_{l,k}^l)^2 \sum_{(i,t) \in \mathcal{S}} p_{i,t} \beta_{i,t}^l \times \\
\left(\frac{\beta_{l,k}^l \sum_{b=1}^{\tau_p} \hat{p}_{l,k}^b}{\varrho_{l,k}^l} + M \left(\sum_{b=1}^{\tau_p} \sqrt{\hat{p}_{l,k}^b \hat{p}_{i,t}^b} \right)^2 \beta_{i,t}^l \right).
\end{aligned}
\tag{76}
$$

The last expectation in the denominator of (29) is

$$
\mathbb{E}\{||\mathbf{v}_{l,k}||^2\} = M \beta_{l,k}^l \varrho_{l,k}^l \sum_{b=1}^{\tau_p} \hat{p}_{l,k}^b,
\tag{77}
$$

by using the definition of MR detection in (30) and the estimated channel distribution in Lemma 1. Plugging (72), (76), and (77) into (29), we obtain the SINR value as shown in the theorem after some simple algebra.

B. Proof of Theorem 1

Let us denote the feasible set of the optimization problem (37) as

$$
\begin{aligned}
\mathcal{F} = \Big\{ p_{l,k}, \hat{p}_{l,k}^b, \forall l, k, b : \; & p_{l,k}, \hat{p}_{l,k}^b \in \mathbb{R}_+, p_{l,k} \le P_{\max,l,k}^d, \\
& \frac{1}{\tau_p} \sum_{b=1}^{\tau_p} \hat{p}_{l,k}^b \le P_{\max,l,k} \Big\}.
\end{aligned}
\tag{78}
$$

The optimal solution set to the optimization problem (48) in the ith iteration is denoted as

$$
\mathcal{I}^{(i)} = \left\{ p_{l,k}^{(i),\text{opt}}, \hat{p}_{l,k}^{b,(i),\text{opt}}, \forall l, k, b \right\}.
\tag{79}
$$

Let $\text{SINR}_{l,k}^{\text{MR}}(\tilde{f})$ denote the SINR value computed from (32) for any feasible point $\tilde{f} \in \mathcal{F}$, while $\widetilde{\text{SINR}}_{l,k}^{(i),\text{MR}}(\mathcal{I}^{(i)})$ denotes the approximated SINR value at the ith iteration computed from (47) using the solution $\mathcal{I}^{(i)}$ obtained in the ith iteration. By using the approximation in (46), our family of SINR functions satisfies the following properties [34], $\forall l, k, b$:

$$
\text{SINR}_{l,k}^{\text{MR}}\left(\tilde{f} \right) \ge \widetilde{\text{SINR}}_{l,k}^{(i),\text{MR}}\left(\tilde{f} \right), \forall \tilde{f} \in \mathcal{F}, \tag{80}
$$

$$
\text{SINR}_{l,k}^{\text{MR}}\left(\mathcal{I}^{(i)} \right) = \widetilde{\text{SINR}}_{l,k}^{(i+1),\text{MR}}\left(\mathcal{I}^{(i)} \right), \tag{81}
$$

$$
\frac{\partial \text{SINR}_{l,k}^{\text{MR}}\left(\mathcal{I}^{(i)} \right)}{\partial \hat{p}_{l,k}^b} = \frac{\partial \widetilde{\text{SINR}}_{l,k}^{(i+1),\text{MR}}\left(\mathcal{I}^{(i)} \right)}{\partial \hat{p}_{l,k}^b}, \tag{82}
$$

$$
\frac{\partial \text{SINR}_{l,k}^{\text{MR}}\left(\mathcal{I}^{(i)} \right)}{\partial p_{l,k}} = \frac{\partial \widetilde{\text{SINR}}_{l,k}^{(i+1),\text{MR}}\left(\mathcal{I}^{(i)} \right)}{\partial p_{l,k}}. \tag{83}
$$

The property (80) implies that the globally objective value to the geometric program (48) is also feasible for the signomial program (37) and we can construct the following chain of inequalities:

$$
\begin{aligned}
\dots = \widetilde{\text{SINR}}_{l,k}^{(i),\text{MR}}\left(\mathcal{I}^{(i-1)} \right) &\overset{(a)}{\le} \widetilde{\text{SINR}}_{l,k}^{(i),\text{MR}}\left(\mathcal{I}^{(i)} \right) \\
&\overset{(b)}{\le} \text{SINR}_{l,k}^{\text{MR}}\left(\mathcal{I}^{(i)} \right) \overset{(c)}{=} \widetilde{\text{SINR}}_{l,k}^{(i+1),\text{MR}}\left(\mathcal{I}^{(i)} \right) \le \dots,
\end{aligned}
\tag{84}
$$

$$\overline{\text{SINR}}_{l,k}^{\text{MR}} = \frac{(1-\epsilon^2)p_{l,k}|\mathbb{E}\{\mathbf{y}_{l,k}^H\mathbf{h}_{l,k}^l\}|^2}{\sum\limits_{(i,t)\in\mathcal{S}} p_{i,t}\mathbb{E}\{|\mathbf{y}_{l,k}^H\mathbf{h}_{i,t}^l|^2\} - (1-\epsilon^2)p_{l,k}|\mathbb{E}\{\mathbf{y}_{l,k}^H\mathbf{h}_{l,k}^l\}|^2 + \sigma^2\mathbb{E}\{\|\mathbf{y}_{l,k}\|^2\}}. \tag{85}$$

where (a) is obtained by solving the geometric optimization problem (48). (b) and (c) follow by (80) and (81), respectively. Thus, if $\xi^{(i),\text{opt}}$ is the optimal objective value of (48), then we obtain $\xi^{(i+1),\text{opt}} \geq \xi^{(i),\text{opt}}$. The objective function is non-decreasing with the iteration index i, while the pilot and data power ranges make the SINR expressions continuous functions and they are bounded from above (i.e., $\widetilde{\widetilde{\text{SINR}}}_{l,k}^{\text{MR}} < \infty, \forall l, k$) ensuring that (48) converges. If the convergence holds at the ith iteration (i.e., $\xi^{(i+1),\text{opt}} = \xi^{(i),\text{opt}}$), then the optimal solution set $\mathcal{I}^{(i)}$ must also be a solution in the $(i+1)$th iteration (otherwise, it leads to $\xi^{(i+1),\text{opt}} > \xi^{(i),\text{opt}}$). Hence (48) converges to a limit point. Furthermore, our constraint functions in (48) satisfy Slater's condition [33] and ensure that the KKT conditions of (37) and (48) coincide, if we use (82) and (83) to do a matching procedure as in the proof of Theorem 1 in [35]. Consequently, the limit point obtained when solving (48) in an iterative manner is a KKT local point to (37).

C. Proof of Theorem 2

We use Lemma 4 [16] to obtain the effective SINR of user k in cell l with MR detection as in (85). It remains to compute the expectations, which is similar to proof of Lemma 2, but more complicated since the channel estimates and estimation errors are neither Gaussian nor independent. Therefore, we summarize the main steps to compute these expectations.

By utilizing the uncorrelation property among the noise, the distortion and the propagation channels, the numerator of (85) is computed as

$$(1-\epsilon^2)p_{l,k}|\mathbb{E}\{\mathbf{y}_{l,k}^H\mathbf{h}_{l,k}^l\}|^2 = (1-\epsilon^2)^2 p_{l,k} M^2 (\beta_{l,k}^l)^2 \|\boldsymbol{\psi}_{l,k}\|^4$$
$$= (1-\epsilon^2)^2 p_{l,k} M^2 (\beta_{l,k}^l)^2 \left(\sum_{b=1}^{\tau_P} \hat{p}_{l,k}^b\right)^2. \tag{86}$$

Similarly, thanks to the uncorrelation of the noise, the propagation channels, and the distortion term, the first expectation in the denominator of (85) is computed as

$$\mathbb{E}\{|\mathbf{y}_{l,k}^H\mathbf{h}_{i,t}^l|^2\} =$$
$$M\beta_{i,t}^l \left(\sum_{(i',t')\in\mathcal{S}} \beta_{i',t'}^l \kappa_{i',t'} + \|\boldsymbol{\psi}_{l,k}\|^2\sigma^2\right) + M^2 \kappa_{i,t}(\beta_{i,t}^l)^2, \tag{87}$$

where $\kappa_{i,t} = (1-\epsilon^2)|\boldsymbol{\psi}_{i,t}^H\boldsymbol{\psi}_{l,k}|^2 + \boldsymbol{\psi}_{l,k}^H\boldsymbol{\Lambda}_{i,t}\boldsymbol{\psi}_{l,k}$, and then by utilizing (5) we obtain the alternative expression for $\kappa_{i,t}$ that is shown in (54). In the same manner, the last expectation in the denominator of (85) is

$$\mathbb{E}\{\|\mathbf{y}_{l,k}\|^2\} = M\left(\sum_{(i,t)\in\mathcal{S}} \beta_{i,t}^l \kappa_{i,t} + \sigma^2\|\boldsymbol{\psi}_{l,k}\|^2\right). \tag{88}$$

Plugging (86)–(88) into (85), we obtain the SINR given in the theorem by using the following identities:

$$\|\boldsymbol{\psi}_{l,k}\|^4 = \left(\sum_{b=1}^{\tau_P} \hat{p}_{l,k}^b\right)^2 \tag{89}$$

$$|\boldsymbol{\psi}_{i,t}^H\boldsymbol{\psi}_{l,k}|^2 = \left(\sum_{b=1}^{\tau_P} \sqrt{\hat{p}_{i,t}^b\hat{p}_{l,k}^b}\right)^2 \tag{90}$$

$$\boldsymbol{\psi}_{l,k}^H\boldsymbol{\Lambda}_{i,t}\boldsymbol{\psi}_{l,k} = \epsilon^2\sum_{b=1}^{\tau_P} \hat{p}_{i,t}^b\hat{p}_{l,k}^b. \tag{91}$$

REFERENCES

[1] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge University Press, 2005.

[2] M. Jiang and L. Hanzo, "Multiuser MIMO-OFDM for next-generation wireless systems," *Proc. IEEE*, vol. 95, no. 7, pp. 1430–1469, 2007.

[3] Q. Li, G. Li, W. Lee, M.-I. Lee, D. Mazzarese, B. Clerckx, and Z. Li, "MIMO techniques in WiMAX and LTE: A feature overview," *IEEE Commun. Mag.*, vol. 48, pp. 86–92, 2010.

[4] E. Björnson, E. G. Larsson, and T. L. Marzetta, "Massive MIMO: 10 myths and one critical question," *IEEE Commun. Mag.*, vol. 54, no. 2, pp. 114 – 123, 2016.

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

[6] T. L. Marzetta, E. G. Larsson, H. Yang, and H. Q. Ngo, *Fundamentals of Massive MIMO*. Cambridge University Press, 2016.

[7] L. Wang, H. Q. Ngo, M. Elkashlan, T. Q. Duong, Y. Yang, and K.-K. Wong, "Massive MIMO in spectrum sharing networks: Achievable rate and power efficiency," *IEEE Systems Journal*, vol. 11, no. 1, pp. 20–31, 2017.

[8] S. Mohammed, "Impact of transceiver power consumption on the energy efficiency of zero-forcing detector in Massive MIMO systems," *IEEE Trans. Commun.*, vol. 62, no. 11, pp. 3874–3890, 2014.

[9] T. V. Chien and E. Björnson, *Massive MIMO Communications*. Springer International Publishing, 2017, pp. 77–116.

[10] T. V. Chien, E. Björnson, and E. G. Larsson, "Joint power allocation and user association optimization for Massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 15, no. 9, pp. 6384 – 6399, 2016.

[11] J. Jose, A. Ashikhmin, T. L. Marzetta, and S. Vishwanath, "Pilot contamination and precoding in multi-cell TDD systems," *IEEE Trans. Commun.*, vol. 10, no. 8, pp. 2640–2651, 2011.

[12] T. E. Bogale and L. Le, "Pilot optimization and channel estimation for multiuser Massive MIMO systems," in *Proc. IEEE CISS*, 2014.

[13] H. Yang and T. L. Marzetta, "Performance of pilot reuse in multi-cell Massive MIMO," in *Proc. IEEE BlackSeaCom*, 2015.

[14] X. Li, E. Björnson, and E. G. Larsson, "Massive MIMO with multi-cell MMSE processing: Exploiting all pilots for interference suppression," *EURASIP J. Wirel. Commun. Netw.*, vol. 1, no. 1, pp. 1–15, 2017.

[15] L. You, X. Gao, X. G. Xia, N. Ma, and Y. Peng, "Pilot reuse for Massive MIMO transmission over spatially correlated rayleigh fading channels," *IEEE Trans. Wireless Commun.*, vol. 14, no. 6, pp. 3352–3366, 2015.

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

[17] S. Jin, M. Li, Y. Huang, Y. Du, and X. Gao, "Pilot scheduling schemes for multi-cell Massive multiple-input-multiple-output transmission," *IET Communications*, vol. 9, no. 5, pp. 689–700, 2015.

[18] X. Zhu, L. Dai, and Z. Wang, "Graph coloring based pilot allocation to mitigate pilot contamination for multi-cell Massive MIMO systems," *IEEE Commun. Letters*, vol. 9, no. 10, pp. 1842 – 1845, October 2015.

[19] H. Ahmadi, A. Farhang, N. Marchetti, and A. MacKenzie, "A game theoretic approach for pilot contamination avoidance in Massive MIMO," *IEEE Wireless Commun. Lett.*, vol. 5, no. 1, pp. 12–15, 2016.

[20] X. Zhu, Z. Wang, L. Dai, and C. Qian, "Smart pilot assignment for Massive MIMO," *IEEE Commun. Letters*, vol. 19, no. 9, pp. 1644 – 1647, 2015.

[21] H. Yin, L. Cottatellucci, D. Gesbert, R. R. Müller, and G. He, "Robust pilot decontamination based on joint angle and power domain discrimination," *IEEE Trans. Signal Process.*, vol. 64, no. 11, pp. 2990–3003, 2016.

[22] L. You, X. Gao, A. L. Swindlehurst, and W. Zhong, "Channel acquisition for Massive MIMO-OFDM with adjustable phase shift pilots," *IEEE Trans. Signal Process.*, vol. 64, no. 6, pp. 1461–1476, 2016.

[23] L. Zhao, H. Zhao, F. Hu, K. Zheng, and Z. Zhang, "Energy efficient power allocation algorithm for downlink Massive MIMO with MRT precoding," in *Proc. IEEE VTC-Fall*, 2013.

[24] H. V. Cheng, E. Björnson, and E. G. Larsson, "Optimal pilot and payload power control in single-cell Massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 65, no. 9, pp. 2363–2378, 2017.

[25] K. Guo, Y. Guo, G. Fodor, and G. Ascheid, "Uplink power control with MMSE receiver in multi-cell MU-Massive-MIMO systems," in *Proc. IEEE ICC*, 2014, pp. 5184–5190.

[26] P. Liu, T. Jiang, Q. Zhang, and M. Matthaiou, "Pilot power allocation through user grouping in multi-cell Massive MIMO systems," *IEEE Trans. Commun.*, vol. 65, no. 4, pp. 1561 – 1574, 2017.

[27] T. V. Chien, E. Björnson, and E. G. Larsson, "Joint pilot sequence design and power control for max-min fairness in uplink Massive MIMO," in *Proc. IEEE ICC*, 2017.

[28] X. Gao, O. Edfors, F. Rusek, and F. Tufvesson, "Massive MIMO performance evaluation based on measured propagation data," *IEEE Trans. Wireless Commun.*, vol. 14, no. 7, pp. 3899–3911, 2015.

[29] S. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Prentice Hall, 1993.

[30] E. Björnson and E. Jorswieck, "Optimal resource allocation in coordinated multi-cell systems," *Foundations and Trends in Communications and Information Theory*, vol. 9, no. 2-3, pp. 113–381, 2013.

[31] G. Y. Li, Z. Xu, C. Xiong, C. Yang, S. Zhang, Y. Chen, and S. Xu, "MM algorithms for geometric and signomial programming," *Mathematical Programming*, vol. 143, no. 1, pp. 339 – 356, 2014.

[32] R. Horn and C. Johnson, *Matrix Analysis*. Cambridge University Press, 2003.

[33] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[34] M. Chiang, C. W. Tan, D. P. Palomar, D. O. Neill, and D. Julian, "Power control by geometric programming," *IEEE Trans. Wireless Commun.*, vol. 6, no. 7, pp. 2640–2651, 2007.

[35] B. R. Marks and G. P. Wright, "A general inner approximation algorithm for nonconvex mathematical programs," *Operations Research*, vol. 26, no. 4, pp. 681–683, 1978.

[36] 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, 2014.

[37] W. Zhang, "A general framework for transmission with transceiver distortion and some applications," *IEEE Trans. Commun.*, vol. 60, no. 2, pp. 384–399, 2012.

[38] M. Biguesh and A. Gershman, "Training-based MIMO channel estimation: a study of estimator tradeoffs and optimal training signals," *IEEE Trans. Signal Process.*, vol. 54, no. 3, pp. 884–893, 2006.

[39] *Further advancements for E-UTRA physical layer aspects (Release 9)*. 3GPP TS 36.814, Mar. 2010.

[40] E. D. Andersen, B. Jensen, J. Jensen, R. Sandvik, and U. Worsøe, "MOSEK version 6," 2009, Technical Report TRC2009C3, MOSEK.

[41] CVX Research Inc., "CVX: Matlab software for disciplined convex programming, academic users," http://cvxr.com/cvx, 2015.

[42] S. Loyka, "Channel capacity of MIMO architecture using the exponential correlation matrix," *IEEE Commun. Lett.*, vol. 5, no. 9, pp. 369–371, 2001.

[43] A. Tulino and S. Verdú, "Random Matrix Theory and Wireless Communications," *Foundations and Trends in Communications and Information Theory*, vol. 1, no. 1, pp. 1–182, 2004.

PLACE PHOTO HERE

Trinh Van Chien (S'16) received the B.S. degree in Electronics and Telecommunications from Hanoi University of Science and Technology (HUST), Vietnam, in 2012. He then received the M.S. degree in Electrical and Computer Enginneering from Sungkyunkwan University (SKKU), Korea, in 2014. He is currently working towards the Ph.D. degree in Communication Systems from Linköping University (LiU), Sweden. His interest lies in convex optimization problems for wireless communications and image & video processing. He was an IEEE wireless communications letters exemplary reviewer for 2016. He also received the award of scientific excellence in the first year of the 5Gwireless project funded by European Union Horizon's 2020.

PLACE PHOTO HERE

Emil Björnson (S'07, M'12, SM'17) received the M.S. degree in Engineering Mathematics from Lund University, Sweden, in 2007. He received the Ph.D. degree in Telecommunications from KTH Royal Institute of Technology, Sweden, in 2011. From 2012 to mid 2014, he was a joint postdoc at the Alcatel-Lucent Chair on Flexible Radio, SUPELEC, France, and at KTH. He joined Linkping University, Sweden, in 2014 and is currently Associate Professor and Docent at the Division of Communication Systems.

He performs research on multi-antenna communications, Massive MIMO, radio resource allocation, energy-efficient communications, and network design. He is on the editorial board of the IEEE TRANSACTIONS ON COMMUNICATIONS and the IEEE TRANSACTIONS ON GREEN COMMUNICATIONS AND NETWORKING. He is the first author of the textbooks *Massive MIMO Networks: Spectral, Energy, and Hardware Efficiency* (2017) and *Optimal Resource Allocation in Coordinated Multi-Cell Systems* (2013). He is dedicated to reproducible research and has made a large amount of simulation code publicly available.

Dr. Björnson has performed MIMO research for more than ten years and has filed more than ten related patent applications. He received the 2016 Best PhD Award from EURASIP, the 2015 Ingvar Carlsson Award, and the 2014 Outstanding Young Researcher Award from IEEE ComSoc EMEA. He has co-authored papers that received best paper awards at WCSP 2017, IEEE ICC 2015, IEEE WCNC 2014, IEEE SAM 2014, IEEE CAMSAP 2011, and WCSP 2009.

PLACE PHOTO HERE

Erik G. Larsson (S'99, M'03, SM'10, F'16) received the Ph.D. degree from Uppsala University, Uppsala, Sweden, in 2002.

He is currently Professor of Communication Systems at Linkping University (LiU) in Linkping, Sweden. He was with the Royal Institute of Technology (KTH) in Stockholm, Sweden, the University of Florida, USA, the George Washington University, USA, and Ericsson Research, Sweden. In 2015 he was a Visiting Fellow at Princeton University, USA, for four months. His main professional interests are within the areas of wireless communications and signal processing. He has co-authored some 130 journal papers on these topics, he is co-author of the two Cambridge University Press textbooks Space-Time Block Coding for Wireless Communications (2003) and Fundamentals of Massive MIMO (2016). He is co-inventor on 16 issued and many pending patents on wireless technology.

He is a member of the IEEE Signal Processing Society Awards Board during 20172019. From 2015 to 2016 he served as chair of the IEEE Signal Processing Society SPCOM technical committee. From 2014 to 2015 he 1246 was chair of the steering committee for the IEEE Wireless Communications Letters. He was the General Chair of the Asilomar Conference on Signals, Systems and Computers in 2015, and its Technical Chair in 2012. He was Associate Editor for, among others,the IEEE Transactions on Communications (2010-2014) and the IEEE Transactions on Signal Processing (2006-2010). He received the IEEE Signal Processing Magazine Best Column Award twice, in 2012 and 2014, the IEEE ComSoc Stephen O. Rice Prize in Communications Theory in 2015, and the IEEE ComSoc Leonard G. Abraham Prize in 2017.