Optimal Pilot and Payload Power Control in Single-Cell Massive MIMO Systems

Victor Cheng, Emil Björnson and Erik G Larsson

Journal Article

N.B.: When citing this work, cite the original article.

©2016 IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE.

Victor Cheng, Emil Björnson and Erik G Larsson, Optimal Pilot and Payload Power Control in Single-Cell Massive MIMO Systems, IEEE Transactions on Signal Processing, 2017. 65(9), pp.2363-2378.
http://dx.doi.org/10.1109/TSP.2016.2641381
Postprint available at: Linköping University Electronic Press

http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-136034

LINKÖPINGS UNIVERSITET

Optimal Pilot and Payload Power Control in Single-Cell Massive MIMO Systems

Hei Victor Cheng, Emil Björnson, and Erik G. Larsson
Department of Electrical Engineering (ISY), Linköping University, Sweden
Email: {hei.cheng, emil.bjornson, erik.g.larsson}@liu.se

***Abstract*—This paper considers the jointly optimal pilot and data power allocation in single-cell uplink massive multiple-input-multiple-output (MIMO) systems. Using the spectral efficiency (SE) as performance metric and setting a total energy budget per coherence interval, the power control is formulated as optimization problems for two different objective functions: the weighted minimum SE among the users and the weighted sum SE. A closed form solution for the optimal length of the pilot sequence is derived. The optimal power control policy for the former problem is found by solving a simple equation with a single variable. Utilizing the special structure arising from imperfect channel estimation, a convex reformulation is found to solve the latter problem to global optimality in polynomial time. The gain of the optimal joint power control is theoretically justified, and is proved to be large in the low SNR regime. Simulation results also show the advantage of optimizing the power control over both pilot and data power, as compared to the cases of using full power and of only optimizing the data powers as done in previous work.**

***Index Terms*—massive MIMO, power control, power allocation, convex optimization**

I. Introduction

A. Background and Motivation

Massive MIMO communication systems have recently attracted a lot of attention [1]–[3]. The idea of massive MIMO is to use a large amount of antennas at the base station (BS) to serve multiple users in the same time and frequency resource block. The ability to increase both SE and energy efficiency makes it one of the key technologies for the 5G cellular networks. The performance analysis of massive MIMO is of vast importance and has been done in [4], [5] for single-cell systems and in [6], [7] for multi-cell systems. However the analysis has been done with the assumption of equal or arbitrary fixed power allocation among the users. Several previous papers [8]–[16] have dealt with power control and provided initial results. (For relation to our work, see below.) In order to harvest all the benefits brought by the massive antenna arrays and guarantee certain uplink system performance, power control among the users is necessary. This can be done by varying the power of different users to increase the sum SE, provide services with certain fairness, or balance between these goals.

Power control in wireless networks has been an important problem for decades, dating back to single-antenna wireless systems. Due to the interference from other users the power control is usually hard to solve optimally, in particular NP-hardness was proven in [17] for the objective of maximizing the sum performance in single-antenna wireless networks, even with single-carrier transmission. For practical use a reasonable approach is to develop suboptimal algorithms with affordable complexity while achieving an acceptable performance, as done for example in [18].

Compared to power control in single-antenna systems, power control in massive MIMO networks is a relatively new topic. Accurate channel estimates are needed at the BS for carrying out coherent linear processing, e.g. uplink detection and downlink precoding. Due to the large number of antennas in massive MIMO the instantaneous channel knowledge, which is commonly assumed to be known perfectly in the power control literature, is hard to obtain perfectly. The literature on power control for multi-user MIMO, and even jointly with optimal beamformer design, see for example [19], [20] and the references therein, did not consider the channel estimation error explicitly and the design criterion was based on SE. We want to provide power control schemes that optimize the ergodic SE based on only the large-scale fading to simplify system design, and take into account the channel estimation errors. Therefore in this work we develop a new framework for power control that matches practical systems (i.e., ergodic SE and imperfect CSI), as the methods developed in the literature cannot be applied directly for massive MIMO systems.

B. Related Work and Our Contributions

Uplink pilots are used to estimate the uplink channels. One needs to take into account both the pilot power and payload power, and hence optimal power control becomes even harder in massive MIMO compared to optimizing data power only in the single-antenna systems. Several work has tried to tackle this challenging problem. In [8] the authors optimize the data power for providing uniform service in multi-cell massive MIMO systems. In [10] and [16] the authors optimize the ratio between pilot and data power to maximize the sum SE, however each user is assumed to use the same ratio. In [11] the sum data power is minimized subject to target signal to interference plus noise ratio (SINR) constraints for multi-cell massive MIMO systems. In [9] power control is done to minimize the uplink power consumption under target SINR constraints where the authors optimize the pilot and data power iteratively to achieve local optima. In [13] data power

This work was supported by ELLIIT, the Linköping University Center for Industrial Information Technology (CENIIT), and the EU FP7 Massive MIMO for Efficient Transmission (MAMMOET) project.

control is done to maximize various objectives in multi-cell massive MIMO systems with an iterative approach which only achieve local optimal solutions. In [14] joint pilot and data power control is done to maximize the energy efficiency. The optimization is done with an approximation of the interference term which therefore does not give the optimal solution. In the conference version of this work [12] we provided a GP formulation for joint pilot and data power control in single cell massive MIMO systems with MRC. Then it was re-derived in [15] to minimize the total power consumption while meeting target uplink and downlink SINR for the users. We are not aware of any work except [12] and [15] that find the jointly global optimal pilot and payload data power for massive MIMO. The previous work either focus on power minimization with SINR constraints or only achieve local optima. The preset target SINR constraints are hard to obtain in practice and the local optima does not provide the complete information about how much can we gain by power control. In this work we address this by providing globally optimal joint power for various objectives, and the questions we want to answer are:

1) Is power control on the pilots needed for massive MIMO systems? If the answer is yes, how much can we gain from jointly optimizing the pilot power and data power, as compared to always using equal power allocation or just power control over the data power?
2) In which scenarios can we gain the most from joint optimization?
3) What intuition can be obtained from the optimal power control? This includes the pilot length, and how the pilot and payload power depend on the estimation quality and signal to noise ratio (SNR).

In this paper we provide answers to these questions in the single-cell uplink scenario with linear operation including maximum ratio combining (MRC) and zero-forcing (ZF). The single-cell scenario is considered here to gain some initial insights to the problem and the challenging extension to multi-cell is left for future work. Note that there are important scenarios when single-cell massive MIMO systems can be deployed, e.g. stadiums and rural wireless broadband access. We formulate and solve the optimization problems and compare the results with simple heuristic power control policies. Two commonly used performance objectives, namely weighted max-min SE and weighted sum SE optimization, are investigated. Our contributions are the following:

1) For the weighted max-min SE formulation, a semi-closed form solution is obtained by solving a simple equation with a single variable.
2) For the weighted sum SE formulation, which was proved to be NP-hard in general wireless networks, is transformed into a convex form in the massive MIMO setup where efficient polynomial time algorithms can be applied to find the global optimum.
3) Both theoretical and numerical results are presented to show the gains of the new framework for joint pilot and data power control.

The existing literature on power control is summarized in Table I, the ones marked with $*$ are the contributions in this work.

The rest of the paper is organized as follows. Section II presents the system model and all the necessary notations. Lower bounds on the uplink capacity are presented which are used to define the problem formulations for optimal power control. In Section III we obtain the optimal pilot length for both problem formulations. Section IV derives the solution approaches for solving the power control problems with weighted max-min SE. In Section V the weighted sum SE formulation is studied. Section VI discusses the extension of the methodology developed in this paper to correlated channel fading models. In Section VII simulation results and discussion of the results are presented. Finally in Section VIII we draw some conclusions.

II. System Model

Consider an uplink single-cell massive MIMO systems with M antennas at the BS and K single-antenna users. The K users are assigned K orthogonal pilot sequences of length τ_p for $K \leqslant \tau_p \leqslant T$, where T is the number of symbols in the coherence interval in which the channels are assumed to be constant. The channels are modeled to be independent Rayleigh fading as this matches the non-line-of-sight massive MIMO channel measurement results reported in [21]. The flat fading channel matrix between the BS and the users is denoted by $\boldsymbol{G} \in \mathbb{C}^{M\times K}$, where the kth column represents the channel response to user k and has the distribution

$$\boldsymbol{g}_k \sim CN(\mathbf{0}, \beta_k \mathbf{I}),\ k = 1, 2, \ldots, K, \tag{1}$$

which is a circularly symmetric complex Gaussian random vector. The variance $\beta_k > 0$ represents the large-scale fading including path loss and shadowing, and is normalized by the noise variance at the BS to simplify the notation. The large-scale fading coefficients are assumed to be known at the BS as they are varying slowly (in the scale of thousands of coherence intervals) and can be easily estimated. The power control proposed in this work only depends on the large-scale fading which makes it feasible to optimize the power control online.

In each coherence interval, user k transmits its orthogonal pilot sequence with power p_p^k to enable channel estimation at the BS. We assume that minimum mean-squared error (MMSE) channel estimation is carried out at the BS to obtain the small-scale coefficients. This gives an MMSE estimate of the channel vector from user k as

$$\hat{\boldsymbol{g}}_k = \frac{\sqrt{\tau_p p_p^k}\beta_k}{1+\tau_p p_p^k \beta_k}\left(\sqrt{\tau_p p_p^k}\boldsymbol{g}_k + \boldsymbol{n}_p^k\right) \tag{2}$$

where $\boldsymbol{n}_p^k \sim CN(\mathbf{0}, \mathbf{I})$ accounts for the additive noise during the training interval. During the payload data transmission interval, the BS receive the signal

$$\boldsymbol{y} = \sum_{k=1}^{K} \boldsymbol{g}_k \sqrt{p_d^k} s_k + \boldsymbol{n} \tag{3}$$

where s_k is the zero mean and unit variance Gaussian information symbol from user k and $\boldsymbol{n} \sim CN(\mathbf{0}, \mathbf{I})$ represents the noise during the data transmission. The channel

TABLE I
EXISTING METHODS FOR POWER CONTROL PROBLEM

Problems	Massive MIMO (data)	Massive MIMO (data+pilot)	Multiuser MIMO (perfect CSI)
Max-min (MRC)	closed form [8]	*semi-closed form (Theorem 2)	convex [20]
Max-min (ZF)	closed form [8]	*semi-closed form (Theorem 2)	full power
Sum (MRC)	*virtual water-filling (Algorithm 1)	*convex (Theorem 6)	NP-hard [17]
Sum (ZF)	*virtual water-filling (Algorithm 1)	*convex (Corollary 1)	full power

estimates are used for MRC or ZF detection of the payload, which corresponds to multiplying the received signal $\boldsymbol{y}$ with $\hat{\boldsymbol{G}}^H \triangleq [\hat{\boldsymbol{g}}_1, \ldots, \hat{\boldsymbol{g}}_K]^H$ or $(\hat{\boldsymbol{G}}^H\hat{\boldsymbol{G}})^{-1}\hat{\boldsymbol{G}}^H$ to detect the symbols $s_1, \ldots, s_K$. The power control methodologies derived in this paper can be applied jointly to each subcarrier in an orthogonal frequency division multiplexing (OFDM) systems. With the channel hardening effect offered by massive MIMO, channel variations in different subcarriers can be neglected and the SE in every subcarrier will mainly depend on the large-scale fading. Therefore the whole spectrum can be allocated to every user and the same power control can be applied to all subcarriers. To make a fair comparison with the scheme with equal power allocation in which each user gives the same power to pilot and data, as done in [4] and most other previous work, we impose the following constraint on the total transmit energy over a coherence interval:

$$\tau_p p_p^k + (T - \tau_p) p_d^k \leqslant E_k, \ k = 1, \ldots, K \tag{4}$$

where E_k is the total energy budget for user k within one coherence interval. In previous work, p_p^k and p_d^k have been optimized separately or often not optimized at all in which case the massive MIMO ability to provide high SE for each user cannot be fully harvested. Therefore we consider the scenario where each user can choose freely how to allocate its energy budget on the pilots and payload. In [7], [10] p_p^k and p_d^k are set equal for every user. The work [8], [11], [13] optimized the payload power to maximize the minimum throughput, which corresponds to fixing p_p^k for every user and optimizing only over p_d^k. The work [22] adopted inverse power control for the pilot power, which corresponds to setting $p_p^k = C/\beta_k$ with a normalization constant C and the data power p_d^k are set to be equal for all users. These previous work can all be included in our framework by setting different variables to be constant. Therefore our framework of power control is the most general so far.

A. Achievable SE With Linear Detection

Since the exact ergodic capacity of the uplink multiuser channels with channel uncertainty is unknown, lower bounds on the achievable SE are often adopted as the performance metric in the massive MIMO literature. Here we present lower bounds on the capacity for arbitrary power control. The achievable SE for user k using MRC is given by the following lemma.

Lemma 1. *The capacity of user k with MRC detection is lower bounded by the achievable ergodic SE*

$$R_k = \left(1 - \frac{\tau_p}{T}\right) \log_2(1 + \mathrm{SINR}_k) \tag{5}$$

where pilot and payload powers are arbitrary,

$$\mathrm{SINR}_k = \frac{M p_d^k \gamma_k}{1 + \sum_{j=1}^K \beta_j p_d^j} \tag{6}$$

and $\gamma_k = \frac{\tau_p p_p^k \beta_k^2}{1 + \tau_p p_p^k \beta_k}$.

For ZF, an achievable ergodic SE of user k is given by the following lemma.

Lemma 2. *The capacity of user k with ZF detection is lower bounded by the achievable ergodic SE*

$$R_k = \left(1 - \frac{\tau_p}{T}\right) \log_2(1 + \mathrm{SINR}_k) \tag{7}$$

where pilot and payload powers are arbitrary,

$$\mathrm{SINR}_k = \frac{(M - K) p_d^k \gamma_k}{1 + \sum_{j=1}^K p_d^j (\beta_j - \gamma_j)} \tag{8}$$

and $\gamma_j = \frac{\tau_p p_p^j \beta_j^2}{1 + \tau_p p_p^j \beta_j}$. *$M > K$ needs to be satisfied for ZF detector to work.*

The proofs of Lemmas 1 and 2 can be obtained by adding corresponding indices for different users' pilot power in the proofs of [6]. Note that these achievable rates are valid for any number of antennas at the BS. However, they are only close to the capacity when there is substantial channel hardening, which is the case when M is large, i.e. in the massive MIMO regime.

These achievable SEs are the performance metric commonly used in the massive MIMO literature. Therefore it is used throughout the paper, where τ_p, p_p^k and p_d^k are the variables to be optimized (for $k = 1, \ldots, K$). The optimization can be done at the BS, which can then inform the users about the pilot length, the amount of power to be spent on pilots, and the amount of power to be spent on payload data. The aim is to maximize a given utility function $U(R_1, \ldots, R_K)$ where $U(\cdot)$ can be any function that is monotonically increasing in every argument. The utility function characterizes the performance and fairness that we provide to the users. Examples of commonly used utility functions are the max-min fairness, sum performance, and proportional fairness [20]. The general problem we address for both MRC and ZF is:

$$\begin{aligned} \underset{\tau_p, \{p_p^k\}, \{p_d^k\}}{\textbf{maximize}} \quad & U\ (R_1, \ldots, R_K) \\ \textbf{subject to} \quad & \tau_p p_p^k + (T - \tau_p) p_d^k \leqslant E_k, \ \forall k, \\ & p_p^k \geqslant 0, p_d^k \geqslant 0, \ \forall k, \\ & K \leqslant \tau_p \leqslant T. \end{aligned} \tag{9}$$

III. Optimal Pilot Length

In this section we derive the optimal length of the pilot sequences in (9) in closed form. First we provide the following lemma:

Lemma 3. *For any monotonically increasing utility function with MRC or ZF detection, the energy constraint* (4) *is satisfied with equality for every user at the optimal solution, i.e.,*

$$\tau_p p_p^k + (T - \tau_p) p_d^k = E_k, \ k = 1, \ldots, K \tag{10}$$

at the optimal point of (9).

Proof. We prove this by contradiction. The SINRs in (6) and (8) for MRC and ZF are monotonically increasing in p_p^k for every user k, and independent of the other users' pilot powers. Suppose some users do not use the full energy budget in the optimal power allocation, they can each increase their pilot power to improve their own SINR without lowering any other user's SINR. Therefore we create a solution which is better than or equal to the optimal one, which is a contradiction to our assumption. Therefore the energy constraint is satisfied with equality. □

Then we state the following theorem which gives the optimal length of training interval in closed form.

Theorem 1. *For any monotonically increasing utility function $U(R_1, \ldots, R_k)$, the problem* (9) *has $\tau_p = K$ at the optimal solution.*

Proof. The proof can be found in Appendix A. □

Using Theorem 1, we can reduce the number of variables involved in (9) and this enables us to find the optimal solutions for certain utility functions in the following sections. Also from Theorem 1 we know that the optimal training period τ_p is equal to the number of users being served, and is the same for every user. Therefore there is no need for assigning pilot sequences of different lengths to different users.

IV. Joint Power Control of Pilots and Payload to Maximize Weighted Minimum SE

In this section we solve the power control problem (9) for the class of max-min fairness problem. The max-min fairness problem is selected to provide the same quality-of-service to all users in the cell. The two cases with MRC and ZF will be discussed separately since the SINR expressions are different. With max-min fairness we aim at serving every user with equal weighted SE according to their priorities and make this value as large as possible. We choose $U(\tilde{R}_1, \ldots, \tilde{R}_K) = \min_k \tilde{R}_k$ with $\tilde{R}_k = (1 - \frac{\tau_p}{T}) \log_2(1 + w_k \text{SINR}_k)$ where $w_k > 0$ are weighting factors to prioritize different users and enable us to achieve any point on the Pareto boundary of the achievable rate region $(R_1, \ldots, R_K)$ by varying the weights [23]. It is trivial to extend Theorem 1 to this case and prove that the optimal length of training equal to K. Since $(1 - \frac{\tau_p}{T}) \log_2(1 + w_k \text{SINR}_k)$ is monotonically increasing in $w_k \text{SINR}_k$, it is equivalent to choose objective as $\min_k w_k \text{SINR}_k$.

A. Max-Min for MRC

With MRC, the power control problem becomes

$$\begin{aligned} \underset{\{p_p^k\},\ \{p_d^k\}}{\text{maximize}} \quad & \min_k \frac{w_k M p_d^k \gamma_k}{1 + \sum_{j=1}^K \beta_j p_d^j} \\ \text{subject to} \quad & \tau_p p_p^k + (T - \tau_p) p_d^k \leqslant E_k, \forall k \\ & p_p^k \geqslant 0, p_d^k \geqslant 0, \forall k. \end{aligned} \tag{11}$$

1) Geometric Program Formulation: Using the epigraph form of (11) we have the following equivalent problem formulation:

$$\begin{aligned} \underset{\{p_p^k\},\{p_d^k\},\ \lambda}{\text{maximize}} \quad & \lambda \\ \text{subject to} \quad & w_k M p_d^k \tau_p p_p^k \beta_k^2 \geqslant \\ & \lambda(1 + \sum_{j=1}^K \beta_j p_d^j + \tau_p p_p^k \beta_k \\ & + \tau_p p_p^k \beta_k \sum_{j=1}^K \beta_j p_d^j), \forall k \\ & \tau_p p_p^k + (T - \tau_p) p_d^k \leqslant E_k, \forall k \\ & p_p^k \geqslant 0, p_d^k \geqslant 0, \forall k. \end{aligned} \tag{12}$$

This problem is non-convex as it is formulated here, however we recognize it as a geometric program (GP). The GP formulation has been considered in the conference version of this paper [12]. Since we next present a new semi-closed form solution with much lower complexity, the GP details are omitted here and we refer the interested readers to [12].

2) Explicit Solution: Next we develop a semi-closed form solution to the max-min fairness problem. Before we present the solution, we need the following lemma:

Lemma 4. *At the optimal point, all $w_k p_d^k \gamma_k$ are equal, i.e.,*

$$w_k p_d^k \gamma_k = w_j p_d^j \gamma_j, \ \forall \ j, k = 1, \ldots, K. \tag{13}$$

Proof. First we need the key observation that at the optimal solution, all weighted SINR_k are equal. We prove this by contradiction. Assume that at the optimal solution, there is at least one user k that has a higher weighted SINR than the others. Denote the minimum weighted SINR at the optimal solution as SINR^*. We can then construct a new solution by decreasing p_d^k by $\delta > 1$ while maintaining that $w_k \text{SINR}_k > \text{SINR}^*$. Since $w_k \text{SINR}_k$ is a continuous increasing function in p_d^k, we can always find such $\delta > 1$. Keeping the other users' powers fixed, we have increased all other users' weighted SINRs. Then we have $w_j \text{SINR}_j > \text{SINR}^*, \ \forall j$, hence we constructed a solution that is better than the optimal solution, which is a contradiction to the initial assumption. Therefore at the optimal solution all weighted SINR_k are equal, and we have

$$w_k \text{SINR}_k = \frac{w_k M p_d^k \gamma_k}{1 + \sum_{j=1}^K \beta_j p_d^j} = \overline{\text{SINR}}, \forall k = 1, \ldots, K, \tag{14}$$

where $\overline{\text{SINR}}$ is the common weighted SINR for every user. We observe that the denominator is the same for every user k. Therefore the numerator of (14) is the same for all k, which leads to (13). □

We call $w_k p_d^k \gamma_k$ the weighted receive signal power k (SP$_k$). Then we want to find the p_d^k that satisfies Lemma 3 for any given value of $x = w_k p_d^k \gamma_k$, which is provided in the following proposition:

Proposition 1. *For any given value of the weighted SP$_k$ $w_k p_d^k \gamma_k = x$, the optimal p_d^k is given in* (15) *on top of next page. When* (15) *is not real-valued, then such SP$_k$ is not attainable by any feasible power allocation.*

Proof. Making use of Lemma 3 and Theorem 1, we have the following equation:

$$\frac{p_d^k(E_k - (T-K)p_d^k)\beta_k^2}{1 + (E_k - (T-K)p_d^k)\beta_k} = \frac{x}{w_k}. \tag{16}$$

This is equivalent to the quadratic equation

$$\begin{aligned}&(T-K)\beta_k^2(p_d^k)^2 - \beta_k\left(E_k\beta_k + \frac{(T-K)x}{w_k}\right)p_d^k \\ &+ \frac{(E_k\beta_k+1)x}{w_k} = 0.\end{aligned} \tag{17}$$

If the equation has real-valued roots, we observe that sum of roots and products of roots are positive, therefore both roots of the equation are positive. Inspecting (6) we see that smaller p_d^k gives a higher SINR$_k$ when $p_d^k\gamma_k$ is fixed. Therefore we arrive at the result. Moreover when the quadratic equation does not have real-valued roots, then $w_k p_d^k \gamma_k < x$ for all feasible p_d^k and therefore such SP$_k$ is not attainable. □

We now reformulate Problem (11) in terms of SP as presented in the following proposition:

Proposition 2. *Problem* (11) *is reduced to the optimization problem* (18) *with one variable (given on top of next page), where the optimization is done in the domain where the objective function is real, i.e., x is constrained to be achievable for every user k. Finding the optimal x in* (18) *gives the optimal common SP for every user. By using Proposition 1 we can find the optimal p_d^k and p_p^k for every user k to achieve this optimal common SP.*

Proof. We first define $x = w_k p_d^k \gamma_k$ and substitute the results from Proposition 1 into the expression of SINR$_k$. Then the objective function is obtained by changing the maximization of SINR to minimization of 1/SINR and simplifying the expression. □

Finally we present the solution to Problem (18):

Theorem 2. *The common SP that maximizes the minimum weighted SINR is given by $1/y$ where y is the unique optimal solution to an strictly convex optimization problem, and the unique real-valued solution can be found by solving the equation* (19) *on top of next page.*

Proof. First we make the change of variable $y = 1/x$ in (18), then we have the following problem:

$$\begin{aligned}&\underset{y}{\text{minimize}} \quad y + \frac{yE_k}{2(T-K)}\sum_k \beta_k - \\ &\frac{\sum_k \sqrt{E_k^2\beta_k^2 y^2 - 2(T-K)(E_k\beta_k+2)\frac{y}{w_k} + (T-K)^2(\frac{1}{w_k})^2}}{2(T-K)}.\end{aligned} \tag{20}$$

The first term is linear, thus the objective is convex if the last term, which has the form $f(y) = \sqrt{ay^2+by+c}$ is concave. This is verified by taking the second derivative of $f(y)$ which gives

$$\frac{1}{4}\frac{4ac-b^2}{(ax^2+bx+c)^{3/2}}. \tag{21}$$

The second derivative is non-positive when $b^2 - 4ac \geqslant 0$, in such case $f(y)$ is concave.

The kth square root term in (20) satisfies $b^2 - 4ac \geqslant 0$ as

$$\begin{aligned}&\left(2(T-K)(E_k\beta_k+2)\frac{1}{w_k}\right)^2 - 4E_k^2\beta_k^2(T-K)^2\left(\frac{1}{w_k}\right)^2 \\ =&4(T-K)^2\left(\frac{1}{w_k}\right)^2(4E_k\beta_k+4) > 0,\end{aligned} \tag{22}$$

and hence it is strictly concave. The overall function is thus strictly convex. Hence the optimal y can be found by setting the first derivative of the objective to zero and the unique solution is found. □

Since we know that (20) is a strictly convex function in y, hence there will be only one optimal solution and it can be found by line search, such as using bisection method, which makes it easy to implement.

To summarize, we provided a semi-closed form solution to the max-min SE problem with the following procedure:

1) Find the optimal common weighted SP by solving (19) given in Theorem 2, using e.g. bisection.
2) For this SP find all the optimal p_d^k using Proposition 1.
3) Find the optimal p_p^k using Lemma 3.

Finding the optimal power control parameters is reduced to solving an equation with a single variable (or a single-variable convex problem). Therefore the complexity is linear in the number of users being served and independent of the number of antennas, which can be implemented in real-time at the BS.

B. Max-Min for ZF

Similar to the case of the MRC detector, we can write the problem as max-min weighted SINR as follows:

$$\begin{aligned}\underset{\{p_p^k\},\ \{p_d^k\}}{\text{maximize}} \quad & \min_k \frac{w_k(M-K)p_d^k\gamma_k}{1+\sum_{j=1}^K p_d^j(\beta_j-\gamma_j)} \\ \text{subject to} \quad & \tau_p p_p^k + (T-\tau_p)p_d^k \leqslant E_k, \forall k \\ & p_p^k \geqslant 0, p_d^k \geqslant 0, \forall k.\end{aligned} \tag{23}$$

The only difference from (11) is the expressions of the SINRs, which is now taken from (8) by inserting $\tau_p = K$. Due to the negative terms appearing in the denominator of the

$$p_d^k = \frac{E_k\beta_k + (T-K)\frac{x}{w_k} - \sqrt{E_k^2\beta_k^2 - 2(T-K)(E_k\beta_k+2)\frac{x}{w_k} + (T-K)^2(\frac{x}{w_k})^2}}{2(T-K)\beta_k} \tag{15}$$

$$\underset{x\geqslant 0}{\text{minimize}} \quad \frac{1}{x} + \frac{E_k}{2(T-K)x}\sum_k \beta_k - \frac{\sum_k \sqrt{\frac{E_k^2\beta_k^2 - 2(T-K)(E_k\beta_k+2)\frac{x}{w_k} + (T-K)^2(\frac{x}{w_k})^2}{x^2}}}{2(T-K)}, \tag{18}$$

$$\frac{1}{2(T-K)}\sum_k \frac{E_k^2\beta_k^2 y - (T-K)(E_k\beta_k+2)\frac{1}{w_k}}{\sqrt{E_k^2\beta_k^2 y^2 - 2(T-K)(E_k\beta_k+2)\frac{1}{w_k}y + (T-K)^2\frac{1}{w_k^2}}} = 1 + \frac{E_k}{2(T-K)}\sum_k \beta_k. \tag{19}$$

SINR expressions, this problem cannot be directly transformed to a GP problem. Fortunately we observe that the denominators of the SINRs are the same for all users, therefore we can state a similar result as Lemma 4.

Lemma 5. *For the ZF detector, at the optimal point, all* $w_k p_d^k \gamma_k$ *are equal, i.e.,*

$$w_k p_d^k \gamma_k = w_j p_d^j \gamma_j, \ \forall\ j,k = 1,\ldots,K. \tag{24}$$

The proof is similar to that of Lemma 4 and is omitted.

By using Lemma 5 we obtain the following important result:

Theorem 3. *Problem* (23) *can be reformulated as*

$$\begin{aligned} \underset{\{p_p^k\},\ \{p_d^k\}}{\text{maximize}} \quad & \min_k \frac{w_k(M-K)p_d^k\gamma_k}{1+\sum_{j=1}^K p_d^j\beta_j} \\ \text{subject to} \quad & \tau_p p_p^k + (T-\tau_p)p_d^k \leqslant E_k, \forall k \\ & p_p^k \geqslant 0, p_d^k \geqslant 0, \forall k. \end{aligned} \tag{25}$$

This implies that solving problem (25) *gives the same optimal* p_d^k, p_p^k *as solving problem* (23)*, but the objective value is different.*

Proof. Using Lemma 5 we have $w_k p_d^k \gamma_k = w_j p_d^j \gamma_j$ at the optimal point. Moreover the denominator can be written as $1 + \sum_{j=1}^K p_d^j \beta_j - \sum_{j=1}^K p_d^j \gamma_j$ where the last term is equal to $w_k p_d^k \gamma_k \sum_{j=1}^K \frac{1}{w_j}$. Then we can rewrite the weighted SINR as

$$w_k \text{SINR}_k = \frac{M-K}{\frac{1}{w_k p_d^k \gamma_k} + \sum_j \frac{p_d^j \beta_j}{w_k p_d^k \gamma_k} - \sum_j \frac{1}{w_j}} \tag{26}$$

Since $\sum_j \frac{1}{w_j}$ is a constant and the same for every user, the set of parameters that maximizes SINR_k also maximizes the SINR_k if the term $\sum_j \frac{1}{w_j}$ is removed. Therefore both problem are equivalent in the sense that they have the same optimal solutions. □

From Theorem 3 we see that only the constant M is replaced with $M-K$, therefore the power allocation that solves the weighted max-min SE for the MRC also solves the weighted max-min SE for the ZF. The same methods and analytical solutions apply. Therefore we don't need to do a separate optimization for ZF in this case. This implies that the users do not need to know what kind of detector is used at the BS. While the BS can switch between different detectors according to the data traffic requirements or power consumption restrictions.

V. Joint Pilot and Data Power Control for Weighted Sum SE

In this section we solve the power control problem (9) for the weighted sum SE for MRC and ZF detector. This problem is selected to maximize the total system throughput, and weights are included to provide some fairness between different users. We define the weighted sum SE by choosing $U(R_1,\ldots,R_K) = \sum_{k=1}^K w_k R_k$.

Power control that maximizes sum SE when interference is present is known to be an NP-hard problem in general under perfect channel knowledge [17]. In this part we present a polynomial-time solution to one special case when all sources transmit to the same receiver. When channel estimation errors are present, with the bounding techniques we used for the achievable SE we discover a specific structure that lead to a convex reformulation after a series of transformations. Since optimizing the data power is considered to be a hard problem itself, in the following we first present the case when one only optimizes the data power, then the solution approach is extended to the case of joint optimization of pilot and data power.

A. Weighted Sum SE for MRC

By using Theorem 1, (9) now becomes the following optimization problem:

$$\begin{aligned} \underset{\{p_p^k\},\ \{p_d^k\}}{\text{maximize}} \quad & \sum_k w_k \log_2\left(1 + \frac{M p_d^k \gamma_k}{1+\sum_{j=1}^K \beta_j p_d^j}\right) \\ \text{subject to} \quad & \tau_p p_p^k + (T-\tau_p)p_d^k \leqslant E_k, \forall k, \\ & p_p^k \geqslant 0, p_d^k \geqslant 0, \forall k. \end{aligned} \tag{27}$$

1) Optimizing Data Power: In the case of optimizing data power only, the energy budget constraint reduced to the peak power constraint on the data power given as $P_k = (E_k - \tau_p p_p^k)/(T-\tau_p)$ for user k where p_p^k is now a constant. Therefore we have the following optimization problem:

$$\begin{aligned} \underset{\{p_d^k\}}{\text{maximize}} \quad & \sum_k w_k \log_2\left(1 + \frac{M p_d^k \gamma_k}{1+\sum_{j=1}^K \beta_j p_d^j}\right) \\ \text{subject to} \quad & p_d^k \leqslant P_k, \forall k, \\ & p_d^k \geqslant 0, \forall k. \end{aligned} \tag{28}$$

In this case γ_k are fixed constants and the optimization variables are the data power p_d^k with individual power constraints.

The formulation in (28) is non-convex. However, we use the observation that the denominator of the SINR expression is the same for every user, to obtain a convex reformulation as described in the following theorem:

Theorem 4. *Problem* (28) *can be reformulated into the following convex form:*

$$\begin{aligned} \underset{s,\ \{x_k\}}{\text{maximize}} \quad & \sum_k w_k \log_2(1 + a_k x_k) \\ \text{subject to} \quad & x_k \leqslant \beta_k P_k s, \forall k, \\ & x_k \geqslant 0, \forall k, \\ & \sum_{j=1}^{K} x_j = 1 - s, \end{aligned} \tag{29}$$

where $a_k = M\gamma_k/\beta_k$. The two formulations are equivalent in the sense that they have the same optimal objective value, and the solution to (28) *can be obtained from solution to* (29) *via* $p_d^k = \frac{x_k}{s\beta_k}$.

Proof. First we observe that the denominator of the SINR expression in the objective function of (28) is the same for every user. It is possible for us to apply the following variable substitutions:

1) $x_k = \frac{\beta_k p_d^k}{1+\sum_j \beta_j p_d^j}$;
2) $s = \frac{1}{1+\sum_j \beta_j p_d^j}$, or equivalently, $s = 1 - \sum_j x_j$.

The individual power constraints are changed proportionally. □

Since problem (29) is convex and Slater's condition is always satisfied, standard convex solvers can handle this problem. Moreover we observe that Theorem 4 transforms the problem into a power allocation of virtual parallel channels with individual and sum power constraints. This problem has a water-filling structure when s is fixed. Therefore we investigate the Karush-Kuhn-Tucker (KKT) conditions and obtain the following solution structure which enable us to develop dedicated algorithms that are more efficient than applying standard interior point methods. The results are summarized in the following theorem:

Theorem 5. *The optimal power allocation to the virtual parallel channel* (29) *satisfies the following equations:*

1) $x_k = \min\left(\beta_k P_k s, \max\left(\frac{w_k}{\nu} - \frac{1}{a_k}\right)^+\right),\ \forall k,$
2) $\sum_{j=1}^{K} x_j = 1 - s,$
3) $\nu = \sum_{j=1}^{K} \beta_j P_j \left(\frac{w_j}{\frac{1}{a_j}+x_j} - \nu\right)^+,$

where $(z)^+ = \max(z, 0)$ for any real number z. When s is fixed, the first two conditions are sufficient. Moreover, when

$$\sum_{j=1}^{K} \beta_j P_j \left(\frac{w_j}{\frac{1}{a_j} + \frac{\beta_j P_j}{1+\sum_{j'} \beta_{j'} P_{j'}}}\right) \leqslant \min_k \frac{(1+\sum_j \beta_j P_j) w_k}{\frac{\beta_k P_k}{1+\sum_j \beta_j P_j} + \frac{1}{a_k}}, \tag{30}$$

then it is optimal to let every user use full power.

Proof. The proof can be found in Appendix B. □

With Theorem 5 we develop an efficient algorithm to obtain the optimal power allocation. For fixed s the optimal x_k can be obtained via modified water-filling. Next we apply bisection on s to find the optimal s such that condition (3) in Theorem 5 is satisfied. The use of bisection needs to be justified and is also provided in the appendix. We only need to search for $s \in [\frac{1}{1+\sum_j \beta_j P_j}, 1]$ since this is an implicit constraint from the definition. The s that solves the problem is such that $f(s) \triangleq \sum_{j=1}^{K} \beta_j P_j \left(\frac{1}{\frac{1}{a_j}+x_j} - v\right)^+ - \nu = 0$. As a by-product we also get the condition when it is optimal to for everyone to use full power. The procedure of finding the optimal power control parameters are described in Algorithm 1.

Algorithm 1 Virtual Water-Filling Algorithm for (29)

Initialize $s_l = \frac{1}{1+\sum_j \beta_j P_j}$ and $s_u = 1$. Check if (30) is satisfied, if yes then terminate and output $p_d^k = P_k$. Otherwise compute $s = (s_l + s_u)/2$.
repeat
  solve for x_k and v satisfying conditions (1) and (2) in Theorem 5
  if $f(s) > 0$
    $s_u = s$, s_l remains unchanged
    $s \leftarrow (s_u + s_l)/2$
  else
    $s_l = s$, s_u remains unchanged
    $s \leftarrow (s_u + s_l)/2$
until convergence with $|s_u - s_l| < \epsilon$
return all $p_d^k = \frac{x_k}{\beta_k s}\ \forall k$

2) Joint Pilot and Data Power Optimization: Next we extend the method to the case of joint power control over pilot and data power. The problem can be written as follows:

$$\begin{aligned} \underset{\{p_d^k\},\ \{p_p^k\}}{\text{maximize}} \quad & \sum_k w_k \log_2\left(1 + \frac{M p_d^k \gamma_k}{1 + \sum_{j=1}^{K} \beta_j p_d^j}\right) \\ \text{subject to} \quad & \tau_p p_p^k + (T - \tau_p) p_d^k \leqslant E_k, \forall k, \\ & p_d^k \geqslant 0,\ p_p^k \geqslant 0, \forall k. \end{aligned} \tag{31}$$

Since γ_k depends on p_p^k which is also an optimization variable, the problem is non-convex. However we find out that the tools we developed for the max-min problem help us here as well. More specifically, we make use of Proposition 1 with $w_k = 1\ \forall k$. Define $x_k = p_d^k \gamma_k$ as the SP of user k, then we use Proposition 1 to make a change of variables in (31) and use the same techniques as in the case of optimizing data power only. We obtain the following theorem:

Theorem 6. *Problem* (31) *can be reformulated into the following form:*

$$\begin{aligned} \underset{s,\ \{y_k\}}{\text{maximize}} \quad & \sum_k w_k \log_2(1 + M y_k) \\ \text{subject to} \quad & \sum_{j=1}^{K} \beta_j q(y_j, s) \leqslant 1 - s, \end{aligned} \tag{32}$$

where $q(y_j, s)$ is defined in (33) *on top of next page. The two formulations are equivalent in the sense that they have*

the same optimal objective values, and the solution to (31) *can be obtained from solution to* (32) *via* $p_d^k = q(y_k, s)/s$. *Moreover problem* (32) *is jointly convex in* s *and* y_k.

Proof. First we introduce a dummy variable t and rewrite (31) as

$$\begin{aligned} \underset{t,\ \{p_d^k\},\ \{p_p^k\}}{\text{maximize}} \quad & \sum_k w_k \log_2\left(1 + \frac{M p_d^k \gamma_k}{t}\right) \\ \text{subject to} \quad & \tau_p p_p^k + (T - \tau_p) p_d^k \leqslant E_k, \forall k, \\ & p_d^k \geqslant 0,\ p_p^k \geqslant 0\ \forall k \\ & 1 + \sum_{j=1}^K \beta_j p_d^j \leqslant t. \end{aligned} \tag{34}$$

The last constraint is relaxed from equality to inequality without changing the solutions to the problem. This is because the objective function is monotonically decreasing in t, thus at the optimal point the last inequality will always be active. Next we apply Proposition 1 with $w_k = 1\ \forall k$. Define $x_k = p_d^k \gamma_k$ as the SP of user k to obtain the following problem

$$\begin{aligned} \underset{t,\ \{x_k\}}{\text{maximize}} \quad & \sum_k w_k \log_2\left(1 + \frac{M x_k}{t}\right) \\ \text{subject to} \quad & 1 + \sum_{j=1}^K \beta_j r(x_j) \leqslant t, \end{aligned} \tag{35}$$

where $r(x_j)$ is defined in (36) on top of next page.

Finally we apply the variable substitution $y_k = x_k/t$ and $s = 1/t$ to obtain (32).

From the proof of Theorem 2 we can deduce that $r(x_j)$ is a convex function in x_j. Next we observe that $q(y_j, s)$ is a perspective transformation of $r(x_j)$ and therefore preserve the convexity [24]. Hence we conclude that (32) is jointly convex in y_js and s. □

Since we have the convex reformulation (32) we can use standard convex solvers to find the optimal solutions efficiently, and the optimal power control parameters can be recovered easily. Here we use the MOSEK solver [25] with CVX [26]

B. Sum SE for ZF

In the case of perfect CSI, maximizing sum SE for ZF is straightforward. This is because the ZF detector completely removes all the interference from other users and creates K parallel channels. However in the case of imperfect CSI, the interference is reduced but still remains, which makes the sum SE problem at least as difficult as with MRC. Fortunately, the techniques we developed for solving the MRC case can be applied here to solve the problem to global optimal. Similarly we will first describe the case of optimizing data power only and then extended to joint pilot and data power optimization.

By using Theorem 1, (9) now becomes the following optimization problem:

$$\begin{aligned} \underset{\{p_p^k\},\ \{p_d^k\}}{\text{maximize}} \quad & \sum_k w_k \log_2\left(1 + \frac{(M-K) p_d^k \gamma_k}{1 + \sum_{j=1}^K p_d^j (\beta_j - \gamma_j)}\right) \\ \text{subject to} \quad & \tau_p p_p^k + (T - \tau_p) p_d^k \leqslant E_k, \forall k \\ & p_p^k \geqslant 0, p_d^k \geqslant 0, \forall k. \end{aligned} \tag{37}$$

1) Optimizing Data Power: In the ZF case, we have the following the following problem:

$$\begin{aligned} \underset{\{p_d^k\}}{\text{maximize}} \quad & \sum_k w_k \log_2\left(1 + \frac{(M-K) p_d^k \gamma_k}{1 + \sum_{j=1}^K (\beta_j - \gamma_j) p_d^j}\right) \\ \text{subject to} \quad & p_d^k \leqslant P_k, \forall k \\ & p_d^k \geqslant 0, \forall k. \end{aligned} \tag{38}$$

We observe that this problem has exactly the same structure as (29) in the MRC case where only the constant β_j changes to $\beta_j - \gamma_j$. Therefore same analysis and algorithm applies here where we substitute all β_j with $\beta_j - \gamma_j$.

2) Joint Pilot and Data Power Optimization: Next we extend this result to the case of joint power control over pilot and data power. The problem is as follows:

$$\begin{aligned} \underset{\{p_d^k\},\ \{p_p^k\}}{\text{maximize}} \quad & \sum_k w_k \log_2\left(1 + \frac{(M-K) p_d^k \gamma_k}{1 + \sum_{j=1}^K (\beta_j - \gamma_j) p_d^j}\right) \\ \text{subject to} \quad & \tau_p p_p^k + (T - \tau_p) p_d^k \leqslant E_k, \forall k \\ & p_d^k \geqslant 0,\ p_p^k \geqslant 0, \forall k. \end{aligned} \tag{39}$$

The transformation we did in the MRC case can be applied here as well as proved by the following corollary:

Corollary 1. *Problem* (39) *can be reformulated into the following form:*

$$\begin{aligned} \underset{s,\ \{y_k\}}{\text{maximize}} \quad & \sum_k w_k \log_2 (1 + (M-K) y_k) \\ \text{subject to} \quad & \sum_{j=1}^K \beta_j q(y_j, s) - \sum_{j=1}^K y_j \leqslant 1 - s, \end{aligned} \tag{40}$$

where $q(y_j, s)$ *is given in* (33) *which is the same as in the MRC case. The two formulations are equivalent in the sense that they have the same optimal objective values, and the solution to* (39) *can be obtained from solution to* (40) *via* $p_d^k = q(y_k, s)/s$. *Moreover problem* (40) *is jointly convex in* s *and* y_k.

Proof. The only difference compared with the case of MRC is that β_j changes to $\beta_j - \gamma_j$ in all expressions. The proof is similar to the case of MRC, and is omitted here for brevity. □

VI. Extension to Correlated Fading Channels

In this section, we extend our results to case of correlated fading channels. We only consider weighted max-min fairness for MRC here, to exemplify how our techniques in the previous sections apply to other channel models. The other cases are left for future work.

For the correlated fading channels, we model $g_k \sim CN(0, \boldsymbol{R}_k)$ where the covariance matrix $\boldsymbol{R}_k$ characterizes the spatial correlation. The large-scale fading is the same for all antennas so all diagonal entries are equal to β_k. The MMSE channel estimation requires the storage of the entire matrix $\boldsymbol{R}_k$ for every user, and the estimation requires the inversion of large matrices – which has a high associated complexity. To avoid this complexity, we adopt the element-wise MMSE estimator proposed in [27]. During the training phase, the BS

$$q(y_j, s) = \frac{E_j\beta_j s + (T-K)y_j - \sqrt{E_j^2\beta_j^2 s^2 - 2(T-K)(E_j\beta_j+2)y_j s + (T-K)^2 y_j^2}}{2(T-K)\beta_j}. \tag{33}$$

$$r(x_j) = \frac{E_j\beta_j + (T-K)x_j - \sqrt{E_j^2\beta_j^2 - 2(T-K)(E_j\beta_j+2)x_j + (T-K)^2 x_j^2}}{2(T-K)\beta_j}. \tag{36}$$

receives the pilot signals, correlates them with pilot sequence of user k and obtains

$$\boldsymbol{y}_k = \sqrt{\tau_p p_p^k}\boldsymbol{g}_k + \boldsymbol{n}_p, \quad k = 1, \ldots, K. \tag{41}$$

The estimate is then

$$\hat{\boldsymbol{g}}_k = \frac{\sqrt{\tau_p p_p^k}\beta_k}{1+\tau_p p_p^k \beta_k}\boldsymbol{y}_k, \quad k = 1, \ldots, K. \tag{42}$$

This estimate, $\hat{\boldsymbol{g}}_k$, is for linear detection of data from user k. With this channel model and estimation method, we obtain the following achievable SE:

Lemma 6. *The capacity of user k with MRC detection under correlated fading and element-wise MMSE estimation is lower bounded by the achievable ergodic SE*

$$R_k^{corr} = \left(1 - \frac{\tau_p}{T}\right)\log_2(1 + \mathrm{SINR}_k^{corr}) \tag{43}$$

where pilot and payload powers are arbitrary,

$$\mathrm{SINR}_k^{corr} = \frac{M p_d^k \gamma_k}{1 + \sum_{j=1}^K \mathrm{tr}(\boldsymbol{R}_j\boldsymbol{R}_k) p_d^j \frac{\gamma_k}{M\beta_k^2} + \sum_{j=1}^K p_d^j \frac{\beta_j}{1+\tau_p p_p^k \beta_k}} \tag{44}$$

and $\gamma_k = \frac{\tau_p p_p^k \beta_k^2}{1+\tau_p p_p^k \beta_k}$.

Proof. The proof is given in Appendix C. □

We observe that Theorem 1 for the optimal training length can be easily extended to cover this case, and therefore the optimization problem we are interested to solve is:

$$\begin{aligned} \underset{\{p_p^k\},\ \{p_u^k\}}{\text{maximize}} \quad & \min_k w_k \mathrm{SINR}_k^{corr} \\ \text{subject to} \quad & \tau_p p_p^k + (T-\tau_p)p_u^k \leqslant E_k, \forall k \\ & p_p^k \geqslant 0, p_u^k \geqslant 0, \forall k. \end{aligned} \tag{45}$$

The epigraph form of (45) is

$$\begin{aligned} \underset{\{p_p^k\},\{p_d^k\},\ \lambda}{\text{maximize}} \quad & \lambda \\ \text{subject to} \quad & w_k M p_d^k p_p^k \beta_k^2 \tau_p \geqslant \\ & \lambda(1 + \tau_p\beta_k p_p^k + \sum_{j=1}^K \beta_j p_d^j + \\ & \tau_p p_p^k \sum_{j=1}^K \mathrm{tr}(\boldsymbol{R}_j\boldsymbol{R}_k) p_d^j \frac{1}{M}), \forall k \\ & \tau_p p_p^k + (T-\tau_p)p_d^k \leqslant E_k, \forall k \\ & p_p^k \geqslant 0, p_d^k \geqslant 0, \forall k. \end{aligned} \tag{46}$$

We recognize (46) as a GP and therefore it can be solved efficiently, using general purpose solvers.

VII. Simulation Results and Discussion

In this section we present simulation results to demonstrate the benefits of our algorithms and compare the performance with the case of no power control (i.e., full equal power) as well as the case of power control on the payload power only (and full power pilots). We consider a scenario with $M = 100$ antennas, $K_0 = 10$ users, and the length of the coherence interval is $T = 200$ (which for example corresponds to a coherence bandwidth of 200 kHz and a coherence time of 1 ms). The users are assumed to be uniformly and randomly distributed in a cell with radius $R = 1000$ m and no user is closer to the BS than 100 m. The path-loss model is chosen as $\beta_k = z_k/r_k^{3.76}$ where r_k is the distance of user k from the BS where z_k represents the independent shadowing effect. Shadowing is chosen to be log-normal distributed with a standard deviation of 8 dB. Due to the long tail behavior of the log-normal distribution there could be some users with very small β_k, therefore in each snapshot the user with the smallest β_k is dropped from service. Therefore the algorithm is run for $K = K_0 - 1 = 9$ users.

The energy budgets $E_k = 10^{-0.5} \times R^{3.76} \times T$ and $E_k = 10^{0.5} \times R^{3.76} \times T$ give a median SNR of -5 dB and 5 dB at the cell edge when using equal power allocation. The weights w_k are set to be equal in all the simulations. The algorithms are run for 1000 Monte-Carlo simulations where in each snapshot the users are dropped randomly in the cell so that the large-scale fading β_k changes.

A. Max-Min SE Results

We compare 4 schemes: 1) the solution to problem (12) (marked as 'Max-min' in the figures); 2) equal power allocation $p_d^k = p_p^k = E_k/T$ (marked as 'Equal Power' in the figures); 3) optimizing only payload power for problem (12) by fixing $p_p^k = E_k/T$ (marked as 'Max-min (data)' in the figures); 4) the scheme that maximizes the sum SE is presented as well for reference (marked as 'sum' in the figures). The same schemes are tested for both MRC and ZF, and low and high SNR scenarios.

In Figure 1 (a) and (b) we plot the cumulative distribution function (CDF) of the minimum SE over different snapshots of user locations for MRC at low and high SNR respectively. We observe that without any power control in almost all of the cases the user with the lowest SNR will get less than 0.5 bit/s/Hz in both low and high SNR scenarios. This is not acceptable if we want to provide decent quality of service to every user being served. With max-min power control for both pilot and data we resolve this problem by guaranteeing the

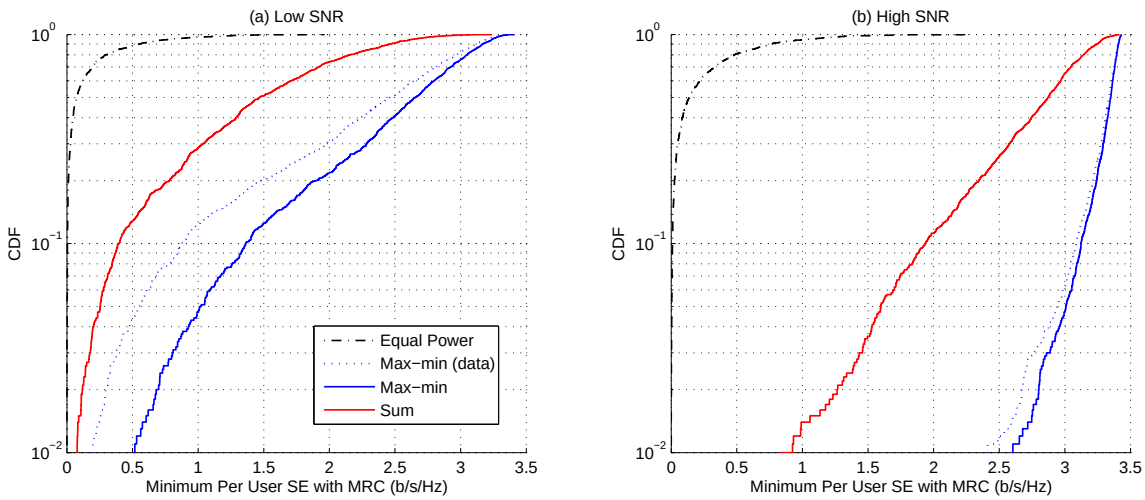

Fig. 1. CDF of the minimum SE with $M = 100$, $K_0 = 10$, $T = 200$, $R = 1000$ m for MRC. Subplots (a) and (b) correspond to low SNR (-5 dB) and high SNR (5 dB) at the cell edge, respectively.

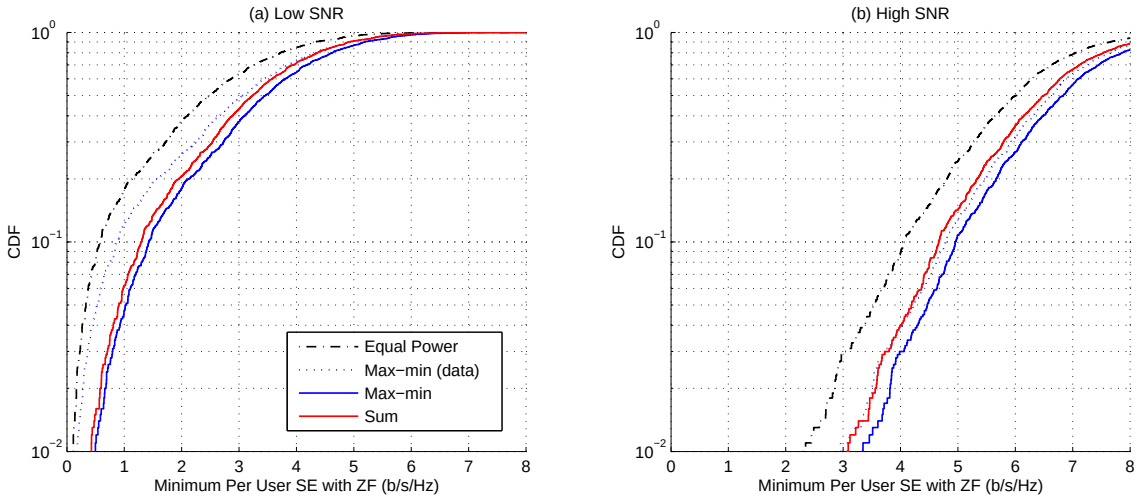

Fig. 2. CDF of the minimum SE with $M = 100$, $K_0 = 10$, $T = 200$, $R = 1000$ m for ZF. Subplots (a) and (b) correspond to low SNR (-5 dB) and high SNR (5 dB) at the cell edge, respectively.

users an SE of more than 1 bit/s/Hz with 0.95 probability and 2.75 bit/s/Hz with 0.5 probability. In low SNR scenarios the joint optimization doubles the 0.95 likely point, from 0.5 to 1 bit/s/Hz, which proves the need of joint pilot and data power optimization at low SNR. In this case with data power control the user with the worst channel would have poor channel estimates that limits the SE, while with joint power control they borrow power from the data part to enhance channel estimation and thereby increase the SE. However in the high SNR scenarios the gain is marginal by the joint optimization, power control over data is enough. This is because the channel estimates are already good enough for linear detection. The performance of the sum SE formulation is not surprising as it is not designed for improving the minimum SE. It boosts the SE of the users with better channels to increase the sum SE, which in turn scarifies the users with worse channels.

In Figure 2 (a) and (b) we plot the CDF of the minimum SE over different snapshots of user locations for ZF at low and high SNR respectively. We observe that all schemes perform similarly and the gains from joint power control with respect to only power control over data are not as large as in the case of MRC. This is because with ZF most interference is removed by the detector, however in low SNR scenarios joint power control is still necessary as it increases the 0.95 likely point from 0.5 to 1 bit/s/Hz compared to power control over data only. The performance of the sum SE formulation is surprisingly good at both low and high SNR and is even better than the max-min scheme with only data power control. This suggests that with ZF detector we can go for the sum SE formulation and push up the total system throughput without sacrificing much of the worse users' performance.

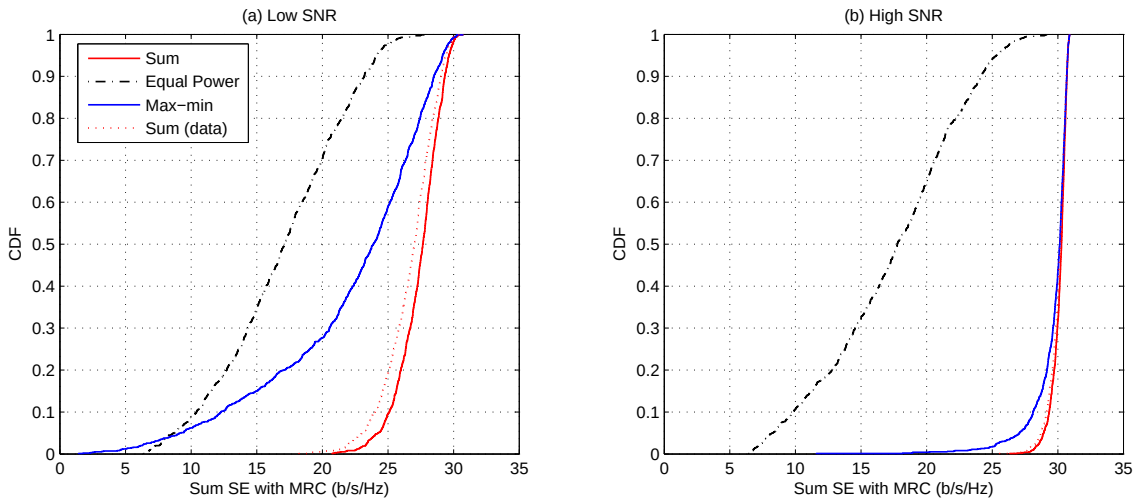

Fig. 3. CDF of the sum SE with $M = 100$, $K_0 = 10$, $T = 200$, $R = 1000$ m for MRC. Subplots (a) and (b) correspond to low SNR (-5 dB) and high SNR (5 dB) at the cell edge, respectively.

B. Sum SE Results

We compare 4 schemes: 1) the scheme that maximizes the sum SE (marked as 'Sum' in the figures); 2) equal power allocation $p_d^k = p_p^k = E_k/T$ (marked as 'Equal Power' in the figures); 3) optimizing the data power only for sum SE by fixing $p_p^k = E_k/T$ (marked as 'Sum (data)' in the figures); 4) the max-min scheme is also presented for reference (marked as 'max-min' in the figures). The same schemes are tested for both MRC and ZF.

In Figure 3 (a) and (b) we plot the CDF of the sum SE for the scenario we described above for MRC at low and high SNR respectively. We observe the optimized power control increases the sum SE significantly. The whole CDF is shifted to the right by almost 15 bit/s/Hz in the low SNR scenario with the proposed power control as compared to equal power allocation. At low SNR the joint power control offers about 10% increase over the case with only data power control. At high SNR the gain is marginal as the SEs of the users have saturated so we are in the $\log$ part of the SE already. The max-min scheme performs well at high SNR due to the saturation of SE, but worse at low SNR. This is because enforcing max-min fairness lead to large sacrifices in sum SE at low SNR. The reason is that with high probability there will be some very disadvantaged user, and everyone else has to cut back significantly to avoid causing near-far interference.

In Figure 4 (a) and (b) we plot the CDF of the sum SE for ZF at low and high SNR respectively. We observe that with ZF when we optimize only the data power the optimal scheme is always using full power. The reason for this is that in single cell systems ZF removes most of the interference, the near-far effects are almost removed by the ZF detector thus creating almost parallel channels. Therefore the scheme with equal power allocation is the same as optimizing data power only. The joint power control offers about 10% improvements over the case with only data power control at low SNR and the gain diminish as the SNR increases. However there will always be a gap between the two schemes, this is because even when the SNR tends to infinity we can use always save power on the pilot and use it for data which increases the SE. The max-min scheme performs poorly in both scenarios, this confirms our suggestion that with ZF we should use the sum SE formulation.

C. Robustness

In this subsection, we present simulation results for the case when the large scale fading parameters are not known perfectly, but obtained through estimation. We assume that the BS collects N processed pilots from each user to perform this estimation. Specifically, denoting each channel realization by $\boldsymbol{g}_k^i$, the processed pilot signals received by the BS for each user can be written as

$$\boldsymbol{y}_k^i = \sqrt{\tau_p p_k}\boldsymbol{g}_k^i + \boldsymbol{w}_k^i, i = 1, \ldots, N, \tag{47}$$

where $\boldsymbol{y}_k^i$ is the processed received signal, τ_p is the length of the pilot, p_k is the signal power and $\boldsymbol{w}_k^i$ is additive noise with variance 1. Then we estimate β_k as follows:

$$\hat{\beta}_k = \frac{\sum_{i=1}^N ||\boldsymbol{y}_k^i||^2 - MN}{MN\tau_p p_k}. \tag{48}$$

This estimate is justified by the fact that

$$\begin{aligned} ||\boldsymbol{y}_k^i||^2 &\approx \tau_p p_k ||\boldsymbol{g}_k^i||^2 + ||\boldsymbol{w}_k^i||^2 \\ &\approx \tau_p p_k \beta_k M + M. \end{aligned} \tag{49}$$

Figure 5 shows the minimum SE achieved by our max-min scheme with the proposed estimator of the large-scale fading parameters. The number of observations is $N = 10$ and the median SNR at the cell edge ranges from -10 dB to 10 dB; all other simulation parameters are the same as in the previous subsection. The estimated βs are treated as the true βs in the optimization (marked as 'Estimated'). The performance is compared with the case when the βs are known perfectly (marked as 'Genie Aided'). We observe that with the simple,

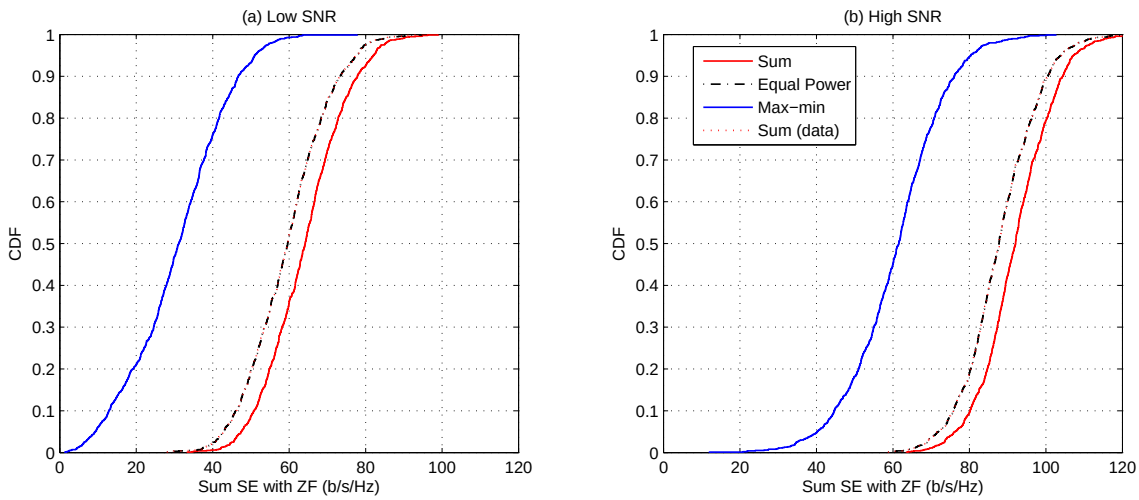

Fig. 4. CDF of the sum SE with $M = 100$, $K_0 = 10$, $T = 200$, $R = 1000$ m for ZF. Subplots (a) and (b) correspond to low SNR (-5 dB) and high SNR (5 dB) at the cell edge, respectively.

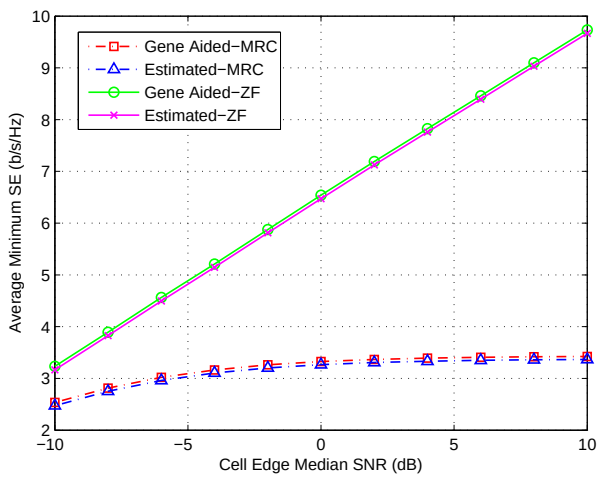

Fig. 5. Average minimum SE with $M = 100$, $K_0 = 10$, $T = 200$, $N = 10$, $R = 1000$ m for estimated large scale fading parameters.

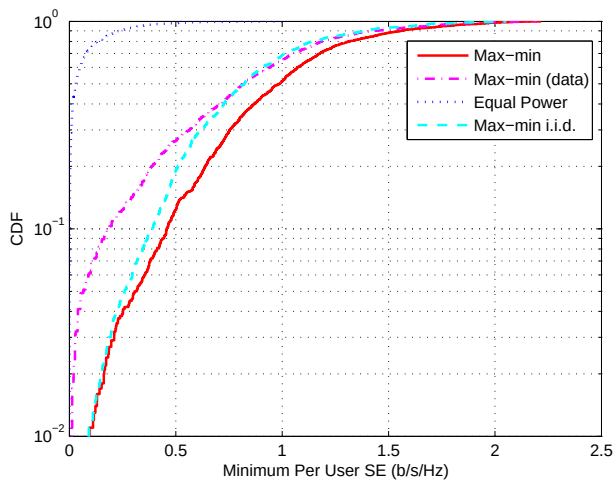

Fig. 6. CDF of the minimum SE with $M = 100$, $K_0 = 10$, $T = 200$, $R = 1000$ m for MRC in correlated fading. The median SNR is -10 dB at the cell edge.

above suboptimal estimator and the small number of training symbols, the performance degradation is almost negligible. We conclude that our scheme shows significant robustness against estimation errors in the large-scale fading parameters.

D. Correlated Fading

In this subsection we look at the performance of joint pilot and data power control in correlated fading channels. We use the one-ring model [28] to model the correlations. An angular spread of 10 degrees is chosen, and the angles of arrival of different users are independent and uniformly distributed between 0 and 180 degrees. The median cell edge SNR is -10 dB and all other parameters are the same as in the above.

Figure 6 shows the CDF of the minimum SE achieved by our scheme with element-wise MMSE channel estimation and MRC. We compare 4 schemes: 1) the solution to problem (46) (marked as 'Max-min'); 2) equal power allocation $p_d^k = p_p^k = E_k/T$ (marked as 'Equal Power'); 3) optimization of only payload power for problem (46), by fixing $p_p^k = E_k/T$ (marked as 'Max-min (data)'); 4) the solution to problem (12) but with application of the power control parameters obtained under the i.i.d. assumption (marked as 'Max-min i.i.d.'). From the plot we see similar behaviors as in the i.i.d. channels, that is, joint pilot and data power control improves the minimum SE substantially. Directly applying the power control parameters obtained under the i.i.d. assumption, neglecting the correlation, yields surprisingly good performance.

Taken together, joint pilot and data power control is highly useful in the low SNR regime also for correlated fading channels. We expect this conclusion to hold also for other channel models, which have to be left for future work.

E. Dependence on SNR, K and T

In this subsection we study the dependence of the gain of joint pilot and data power control on the SNR, number of users and length of coherence interval.

First we investigate the performance of MRC. At low SNR, the noise dominates over interference and we can approximate (6) as

$$\mathrm{SINR}_k \approx M p_d^k \gamma_k \approx M K p_p^k p_d^k \beta_k^2, \tag{50}$$

for $\tau_p = K$. Under the power constraint

$$K p_p^k + (T-K) p_d^k \leqslant E_k, \tag{51}$$

it is straightforward to show that

$$p_p^k = \frac{E_k}{2K} \text{ and } p_d^k = \frac{E_k}{2(T-K)} \tag{52}$$

maximize the approximate SINR. Since $T \gg K$, this means that the user allocates substantially more power to pilots than to data at low SNR. Compared to the case of data power control only, where $p_p^k = p_d^k = E_k/T$, we have

$$\mathrm{SINR}_k^{opt} \approx \frac{T^2}{4K(T-K)} \mathrm{SINR}_k^{data}, \tag{53}$$

where SINR_k^{opt} represents the SINR obtained by optimizing the pilot and data power and SINR_k^{data} represents the SINR obtained by only optimizing the data power.

To conclude, the gain of joint pilot and data power control can be substantial at low SNR, and when K is small relative to T. For ZF, similar results are obtained at low SNR, where the interference can be neglected. Therefore our scheme may be particularly useful for wireless broadband access with stationary terminals, as in that application the coherence interval is usually very long.

At high SNR, when interference dominates over noise, $\gamma_k \approx \beta_k$. Then the impact of the pilot power is negligible for MRC. However for ZF, the interference is cancelled out completely, thus creating parallel channels. More power will be spent on data to boost the SE. However the gain will not be substantial as the spectral efficiency only grows logarithmically with SINR in this regime. Therefore optimizing the data power is most important.

F. Complexity

In this subsection, we characterize the computational complexity of our schemes, and compare it to that of the other digital signal processing that is carried out in massive MIMO systems (in particular, channel estimation and linear detection of the data). We perform the comparison for MRC, as ZF would consume more computational resources.

Since our power control parameters are computed based on the large scale fading, we only have to recompute them at the pace that the large-scale fading changes. The complexity of our algorithm for the max-min problem is of order $O(K)$. The sum SE problem is transformed to a convex problem that can be solved by a general interior point method. Its complexity is $N_{it} O((m+K)^3)$, where N_{it} is the number of Newton iterations required to achieve a predetermined precision, and m is the number of constraints in the problem. The exact number of N_{it} is hard to determine, however in practice, N_{it} is typically in the order of tens [24, Chapter 11]. Therefore 100 should be a good enough bound; in any case, the algorithm may be terminated after 100 iterations. In each Newton iteration we are solving a linear system of equations. Since we have $2K+1$ constraints and $K+1$ variables, the number of operations required for solving this Newton system is about $9K^3$, assuming the use of Cholesky factorization. In the channel estimation phase the number of operations is approximately $2MK^2$, and for MRC detection the number of operations is approximately $2MK$ per data symbol [29]. Therefore the total amount of computations in one coherence interval is approximately $2MK^2 + 2MK(T-K) = 2MKT$. The measurements reported in [30] show that the large-scale fading parameters are constant over a duration that is on the order of 100 times the channel coherence time. Moreover, for the sake of argument, we assume that there are 100 subcarriers in the system. These assumptions result in a relative computational overhead of the proposed sum SE algorithm as $\frac{N_{it} 9K^3}{20000MKT + N_{it} 9K^3}$. We see that even with $N_{it} = 100$ (likely an overestimate) this overhead is on the order of 0.02%.

We conclude that while the complexity calculation given here represents a first-order estimate only, the extra efforts for solving the joint optimization problem is negligible in representative cases.

VIII. Conclusion

We considered the optimal joint pilot and data power allocation problems in single cell uplink massive MIMO systems with MRC or ZF detection. It was first proved that the optimal length of the training interval equals the number of users. Using the SE as performance metric and setting a total energy budget, the power control was formulated as optimization problems for two different objective functions: the weighted minimum SE and the weighted sum SE. The optimal power control policy was found for the case of maximizing the weighted minimum SE by a semi-closed form solution to a single variable equation with unique solution. The optimal power control parameters were shown to be the same for MRC and ZF. For maximizing the sum SE a convex reformulation was found and efficient solution algorithms were developed. The methods have also been extended to handle the case of correlated fading, although a complete treatment of all aspects of that case is left for future work.

Simulation results demonstrated the advantage of joint optimization over both pilot and data power, and how the two objectives behave at low and high cell-edge SNRs. With MRC we have a clear choice to make between max-min and sum SE, which is dependent on the system requirements. With ZF we can maximize sum SE without sacrificing much in min SE. The need of joint pilot and data power control is particularly important at low SNR, while at high SNR optimizing only data power seems to be good enough. Since multi-cell systems are interference-limited, we predict that we will get results similar to the low SNR results, particularly if a large pilot reuse factor is used to get single-cell-like estimation quality. The numerical

results were also justified by a theoretical analysis in the low and high SNR regime. This analysis showed that the gain is more substantial when the number of users, K, is small compared to the length of the coherence interval, T.

Future work includes extension of the methodologies to multi-cell systems and more sophisticated system models.

Appendix

A. Proof of Theorem 1

Before proving Theorem 1 we state and prove two lemmas.

Lemma 7. *For any $x \geqslant 0$, we have $\ln(x) \geqslant \frac{x-1}{x}$ with equality if and only if $x = 1$.*

Proof. Write $f(x) = \ln(x) - \frac{x-1}{x}$, then we have $f'(x) = \frac{1}{x} - \frac{1}{x^2}$. Observing that $f'(x) \leqslant 0, \forall x \in (0,1]$ and $f'(x) \geqslant 0, \forall x > 1$, we can conclude that $x = 1$ is the minimum point of $f(x)$ at which $f(x) = 0$. Thus we have $f(x) \geqslant 0, \forall x > 0$, which proves the lemma. □

Lemma 8. *For any positive constants a, b and c, $g(x) = x\log_2\left(1 + \frac{a}{bx+c}\right)$ is a strictly monotonic increasing function in x for all $x > 0$.*

Proof. Taking the first derivative we have

$$\begin{aligned} g'(x) &= \frac{1}{\ln(2)}\ln\left(1 + \frac{a}{bx+c}\right) \\ &\quad + \frac{x}{\ln(2)} \cdot \frac{1}{1 + a/(bx+c)} \cdot \frac{-a}{(bx+c)^2 \cdot b} \\ &= \frac{1}{\ln(2)}\left(\ln\left(1 + \frac{a}{bx+c}\right) - \frac{abx}{(bx+c)(bx+c+a)}\right) \\ &> \frac{1}{\ln(2)}\left(\ln\left(1 + \frac{a}{bx+c}\right) - \frac{a}{bx+c+a}\right) \geqslant 0. \end{aligned} \tag{54}$$

The first inequality comes from the fact that $bx/(bx+c) < 1$ for any strictly positive b and c. The last inequality follows from putting $1 + a/(bx+c)$ in Lemma 7. □

Next, we prove Theorem 1 by contradiction. Assume that $\tau_p^* > K$, p_p^{k*} and p_d^{k*} is the optimal solution to problem (9). From Lemma 3 we know that

$$\tau_p^* p_p^{k*} + (T - \tau_p^*)p_d^{k*} = E_k, \ k = 1, \ldots, K. \tag{55}$$

We will now construct a new feasible point that gives a higher objective function. Choose $\tau_p^{'} = K$, $p_p^{k'} = \tau_p^* p_p^{k*}/K$, $p_d^{k'} = (E_k - \tau_p^* p_p^{k*})/(T-K)$ for every user k, then $\gamma_k^{'} = \gamma_k^*$ as $\tau_p^{'} p_p^{k'} = \tau_p^* p_p^{k*}$. We compare the value of $R_k(\tau_p, p_p^k, p_d^k)$ for these two sets of parameters. The achievable SE for user k with our new construction is

$$R_k(K, p_p^{k'}, p_d^{k'}) = \left(1 - \frac{K}{T}\right)\log_2\left(1 + \frac{a_k}{T - K + c_k}\right) \tag{56}$$

where $a_k = M\gamma_k^{'}(E_k - \tau_p^* p_p^{k*})$, $c_k = \sum_{j=1}^K \beta_j(E_j - \tau_p^* p_p^{j*})$ for the MRC, and $a_k = (M-K)\gamma_k^{'}(E_k - \tau_p^* p_p^{k*})$, $c_k = \sum_{j=1}^K(\beta_j - \gamma_j^{'})(E_j - \tau_p^* p_p^{j*})$ for the ZF. Then we observe that

$$R_k(\tau_p^*, p_p^{k*}, p_d^{k*}) = \left(1 - \frac{\tau_p^*}{T}\right)\log_2\left(1 + \frac{a_k}{T - \tau_p^* + c_k}\right). \tag{57}$$

Next we apply Lemma 8 with $x = T - \tau_p$ we can know that TR_k is a strictly monotonic increasing function in $T - \tau_p$. Therefore $R_k(K, p_p^{k'}, p_d^{k'}) > R_k(\tau_p^*, p_p^{k*}, p_d^{k*})$ and the maximum is achieved at $\tau_p = K$ due to the constraint $\tau_p \geqslant K$. This is a contradiction to the assumption that $\tau_p^* > K$, hence $\tau_p^* = K$. Since this holds for every k, we have proved the theorem for any monotonic increasing function $U(R_1, \ldots, R_k)$.

B. Proof of Theorem 5

We first state the Lagrangian function of problem (29):

$$\begin{aligned} L(s, \{x_k\}, \{\lambda_k\}, \{\mu_k\}, \nu) &= \sum_k w_k \log_2(1 + a_k x_k) \\ &- \sum_k \lambda_k(x_k - \beta_k P_k s) + \sum_k \mu_k x_k - \nu(\sum_k x_k + s - 1). \end{aligned} \tag{58}$$

Then we can write the KKT conditions [24] for problem (29):

$$\begin{aligned} \frac{w_k}{\frac{1}{a_k} + x_k} - \lambda_k + \mu_k - \nu &= 0, \forall k, \\ \lambda_k \geqslant 0, \ x_k \leqslant \beta_k P_k s, \ \lambda_k(x_k - \beta_k P_k s) &= 0, \forall k, \\ \mu_k \geqslant 0, \ x_k \geqslant 0, \ \mu_k x_k &= 0, \forall k, \\ \sum_k x_k &= 1 - s, \\ \sum_k \lambda_k \beta_k P_k - \nu &= 0. \end{aligned} \tag{59}$$

We construct a set of solution to the above KKT conditions as follows:

$$x_k = \min\left(\beta_k P_k s, \left(\frac{w_k}{\nu} - \frac{1}{a_k}\right)^+\right), \forall k, \tag{60}$$

$$\lambda_k = \left(\frac{w_k}{\frac{1}{a_k} + x_k} - \nu\right)^+, \forall k, \tag{61}$$

$$\mu_k = (\nu - a_k)^+, \forall k. \tag{62}$$

We can easily verify this set of solutions together with condition (59) and (59) satisfies the overall KKT conditions. When s is considered to be a constant, the last condition of (59) is not necessary as it corresponds to $\frac{\partial L}{\partial s} = 0$. This set of solutions is a function of ν and we are looking for ν such that (59) is satisfied. For a given s, finding the optimal x_ks and ν can be done using algorithms in [31] and [32]. Then we perform bisection on s to find the optimal s that satisfies (59). Using bisection we are looking for the zero crossing point of a univariate function, and this requires the function to have different signs on each end of the interval. To justify that we can use bisection, we need to check the sign of $f(s) \triangleq \sum_{j=1}^K \beta_j P_j \left(\frac{w_j}{\frac{1}{a_j} + x_j} - \nu\right)^+ - \nu$ on the boundaries, which corresponds to checking $s = \frac{1}{1+\sum_j \beta_j P_j}$ and $s = 1$.

When $s = \frac{1}{1+\sum_j \beta_j P_j}$, then to satisfy (59) $x_k = \beta_k P_k s$, and thus $\lambda_k \geqslant 0, \ \forall k$ and $\mu_k = 0, \ \forall k$. This is equivalent to

$$\nu \leqslant \frac{w_k}{\frac{1}{a_k} + \frac{\beta_k P_k}{1+\sum_j \beta_j P_j}}, \ \forall k. \tag{63}$$

On the other hand from the last condition of (59) we have

$$\nu = \frac{1}{1+\sum_j \beta_j P_j} \sum_{j=1}^{K} \beta_j P_j \left(\frac{w_j}{\frac{1}{a_j} + \frac{\beta_j P_j}{1+\sum_{j'} \beta'_j P'_j}} \right). \quad (64)$$

Therefore for both (63) and (64) to hold the condition is

$$\frac{1}{1+\sum_j \beta_j P_j} \sum_{j=1}^{K} \beta_j P_j \left(\frac{w_j}{\frac{1}{a_j} + \frac{\beta_j P_j}{1+\sum_{j'} \beta_{j'} P_{j'}}} \right) \leqslant \min_k \frac{w_k}{\frac{1}{a_k} + x_k}. \quad (65)$$

In such case we can always find the ν that satisfies the KKT conditions, which means that it is optimal to let every user use full power. On the other hand if (65) does not hold, there is no ν that can satisfy all conditions simultaneously, and we can easily check that $f(s) > 0$.

When $s = 1$, $\lambda_k = 0$, $\forall k$ and therefore $f(s) = -\nu \leqslant 0$. Hence we have verified that $f(s)$ have different signs on the boundaries. Moreover (29) is a convex problem and Slater's condition is satisfied. Therefore the KKT conditions are sufficient and necessary for optimality, there will be one s such that $f(s) = 0$ within the boundaries. The optimal s can therefore be found via bisection.

C. Proof of Lemma 6

We apply the bounding techniques used in [33] to develop an achievable SE as

$$\left(1 - \frac{\tau}{T}\right) \log_2 \left(1 + \mathrm{SINR}_k^{corr}\right), \quad (66)$$

where

$$\mathrm{SINR}_k^{corr} = \frac{|E[\hat{\boldsymbol{g}}_k^H \boldsymbol{g}_k]|^2 p_d^k}{\sum_j E[|\hat{\boldsymbol{g}}_k^H \boldsymbol{g}_j|^2] p_d^j - |E[\hat{\boldsymbol{g}}_k^H \boldsymbol{g}_k]|^2 p_d^k + E[||\hat{\boldsymbol{g}}_k||^2]} \quad (67)$$

The results follow from calculating the terms in the above expressions, using standard results.

References

[1] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Processing Magazine*, vol. 30, no. 1, pp. 40–60, Jan 2013.

[2] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Transactions on Wireless Communications*, vol. 9, no. 11, pp. 3590–3600, November 2010.

[3] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang, "An overview of massive MIMO: Benefits and challenges," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 742–758, Oct 2014.

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

[5] H. Yang and T. L. Marzetta, "Performance of conjugate and zero-forcing beamforming in large-scale antenna systems," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 2, pp. 172–179, February 2013.

[6] ——, "Capacity performance of multicell large-scale antenna systems," in *2013 51st Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, Oct 2013, pp. 668–675.

[7] J. Hoydis, S. ten Brink, and M. Debbah, "Massive MIMO in the UL/DL of cellular networks: How many antennas do we need?" *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 2, pp. 160–171, February 2013.

[8] H. Yang and T. L. Marzetta, "A macro cellular wireless network with uniformly high user throughputs," in *2014 IEEE 80th Vehicular Technology Conference (VTC2014-Fall)*, Sept 2014, pp. 1–5.

[9] K. Guo, Y. Guo, G. Fodor, and G. Ascheid, "Uplink power control with mmse receiver in multi-cell mu-massive-MIMO systems," in *2014 IEEE International Conference on Communications (ICC)*, June 2014, pp. 5184–5190.

[10] H. Q. Ngo, M. Matthaiou, and E. G. Larsson, "Massive MIMO with optimal power and training duration allocation," *IEEE Wireless Communications Letters*, vol. 3, no. 6, pp. 605–608, Dec 2014. (Correction published in vol. 4, pp. 225, Apr. 2015.)

[11] J. Choi, "Massive MIMO with joint power control," *IEEE Wireless Communications Letters*, vol. 3, no. 4, pp. 329–332, Aug 2014.

[12] H. V. Cheng, E. Björnson, and E. G. Larsson, "Uplink pilot and data power control for single cell massive MIMO systems with mrc," in *2015 International Symposium on Wireless Communication Systems (ISWCS)*, Aug 2015, pp. 396–400.

[13] X. Li, E. Björnson, E. G. Larsson, S. Zhou, and J. Wang, "Massive MIMO with multi-cell MMSE processing: Exploiting all pilots for interference suppression," *CoRR*, vol. abs/1505.03682, 2015. [Online]. Available: http://arxiv.org/abs/1505.03682

[14] K. Guo, Y. Guo, and G. Ascheid, "Energy-efficient uplink power allocation in multi-cell mu-massive-MIMO systems," in *Proceedings of European Wireless 2015; 21th European Wireless Conference*, May 2015, pp. 1–5.

[15] Y. Zhang and W. P. Zhu, "Energy-efficient pilot and data power allocation in massive MIMO communication systems based on MMSE channel estimation," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2016, pp. 3571–3575.

[16] S. Lu and Z. Wang, "Joint optimization of power allocation and training duration for uplink multiuser MIMO communications," in *2015 IEEE Wireless Communications and Networking Conference (WCNC)*, March 2015, pp. 322–327. (extended version [Online] available: http://arxiv.org/abs/1410.7675)

[17] Z. Q. Luo and S. Zhang, "Dynamic spectrum management: Complexity and duality," *IEEE Journal of Selected Topics in Signal Processing*, vol. 2, no. 1, pp. 57–73, Feb 2008.

[18] M. Chiang, C. W. Tan, D. Palomar, D. O'Neil, and D. Julian, "Power control by geometric programming," *IEEE Transaction on Wireless Communications*, vol. 6, no. 7, pp. 2640–2651, July 2007.

[19] S. Shi, M. Schubert, and H. Boche, "Rate optimization for multiuser MIMO systems with linear processing," *IEEE Transactions on Signal Processing*, vol. 56, no. 8, pp. 4020–4030, Aug 2008.

[20] 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.

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

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

[23] E. Björnson, E. A. Jorswieck, M. Debbah, and B. Ottersten, "Multiobjective signal processing optimization: The way to balance conflicting metrics in 5g systems," *IEEE Signal Processing Magazine*, vol. 31, no. 6, pp. 14–23, Nov 2014.

[24] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge, U.K.: Cambridge University Press, 2004.

[25] M. ApS, *MOSEK Optimization Suite Release 8.0.0.42*, 2016. [Online]. Available: http://docs.mosek.com/8.0/intro.pdf

[26] M. Grant and S. Boyd. (2013, Sept) Cvx: Matlab software for disciplined convex programming, version 2.0 beta. [Online]. Available: http://cvxr.com/cvx

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

[28] A. Adhikary, J. Nam, J. Y. Ahn, and G. Caire, "Joint spatial division and multiplexing - the large-scale array regime," *IEEE Transactions on Information Theory*, vol. 59, no. 10, pp. 6441–6463, Oct 2013.

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

[30] I. Viering, H. Hofstetter, and W. Utschick, "Spatial long-term variations in urban, rural and indoor environments," in *Proceedings of the 5th COST*, vol. 273. Citeseer, 2002.

[31] D. P. Palomar and J. R. Fonollosa, “Practical algorithms for a family of waterfilling solutions,” *IEEE Transactions on Signal Processing*, vol. 53, no. 2, pp. 686–695, Feb 2005.

[32] P. He, L. Zhao, S. Zhou, and Z. Niu, “Water-filling: A geometric approach and its application to solve generalized radio resource allocation problems,” *IEEE Transactions on Wireless Communications*, vol. 12, no. 7, pp. 3637–3647, July 2013.

[33] J. Jose, A. Ashikhmin, T. L. Marzetta, and S. Vishwanath, “Pilot contamination and precoding in multi-cell TDD systems,” *IEEE Transactions on Wireless Communications*, vol. 10, no. 8, pp. 2640–2651, August 2011.