Training Optimization and Performance of Single Cell Uplink System With Massive-Antennas Base Station

Songtao Lu, *Member, IEEE*, and Zhengdao Wang, *Fellow, IEEE*

***Abstract*— We study the performance of uplink transmission in single-cell wireless systems, where all the transmitters have single antennas and the base station has a large number of antennas. We consider both maximum ratio combining and zero-forcing receivers and both small- and large-scale fading channels. We also characterize the achievable total degrees of freedom (DoF) of such a system without assuming channel state information at the receiver. The system DoF turns out to be the same as that of a single-user multiple-input multiple-output system. However, when the number of users is the same as the number of receive antennas, linear receivers are not sufficient for achieving the maximum total DoF. The amount of energy savings that are possible through the increased number of base-station antennas or increased coherence interval are quantified. Furthermore, the training period and training energy allocation under the average and peak power constraints are optimized jointly to maximize the achievable sum spectral efficiency (SE). The improvement on achievable SE provided by the training duration and energy optimization is verified through multiple numerical simulations.**

***Index Terms*— Massive MIMO, uplink, multiuser, channel estimation, energy allocation, training optimization, degree of freedom (DoF).**

I. Introduction

MASSIVE multiple-input multiple-output (MIMO) systems are a type of cellular communications where the base station is equipped with a large number of antennas. The base station serves multiple mobile stations that are usually equipped with a small number of antennas, typically one. There are several challenges with designing such massive MIMO systems, including e.g., channel state information (CSI) acquisition [3], base station received signal processing [4], downlink precoding with imperfect CSI [5], signal detection algorithm [6], etc. For multi-cell systems, pilot contamination and inter-cell interference also need to be dealt with [7]. There is already a body of results in the literature about the analysis and design of large MIMO systems; see e.g., the overview articles [8]–[11] and references there in. To reveal the potential that is possible with massive MIMO systems, it is important to quantify the achievable performance of such systems in realistic scenarios. For example, it is important to consider practical constraints such as average and peak training power in the channel acquisition process.

A. Scope of This Paper

In this paper, we are interested in the performance of the *uplink* transmission in *single-cell* systems such as stadiums and rural wireless broadband access. However, in practice, the energy spent on sending the wireless signals is limited, while the high quality of the transmission is preferred. There may be several constraints on transmitting the messages, such as power constraints. In particular, we ask what rates can be achieved in the uplink by the mobile users if we assume realistic channel estimation at the base station. Similar analysis has been performed in [12]–[14], but the analysis therein assumes equal power transmission during the channel training phase and the data transmission phase. Also, the effect of the channel coherence interval on the system throughput was discussed in [15] and optimization of the power allocation and training duration for an uplink MIMO system was considered for single-cell and multi-cell systems in [1] and [16] respectively. However, the peak power constraint was not considered. For a fixed training period, to obtain an accurate estimate, the training power needs to be high to enable enough training energy. As a result, peak power constraint, if present, may be violated. The solution is also to optimize the training duration.

If we allow the users to cooperate, then the system can be viewed as a point-to-point MIMO channel. The rates obtained in [17], and the stronger result on non-coherent MIMO channel capacity in [18] can serve as an upper bound for the system sum rate. The question is how much of this sum rate can be achieved without user cooperation and without using elaborate signaling such as signal packing on Grassmannian manifolds. For a system with K mobile users, M base station antennas, and a block fading channel with coherence interval T, we quantify the total degrees of freedom (DoF) and the needed transmission power for achieving a given rate when $M, T \gg 1$, which is a refinement of the corresponding result in [12].

Manuscript received February 21, 2018; revised June 6, 2018 and September 5, 2018; accepted October 7, 2018. Date of publication October 16, 2018; date of current version February 14, 2019. The work in this paper was supported in part by NSF Grant No. 1711922. This paper was presented at the IEEE Global Communications Conference, Austin, TX, USA, December 8–12, 2014 [1], and at the IEEE Wireless Communications and Networking Conference, New Orleans, LA, USA, March 9–12, 2015 [2]. The associate editor coordinating the review of this paper and approving it for publication was R. Zhang. *(Corresponding author: Zhengdao Wang.)*

S. Lu is with the Department of Electrical and Computer Engineering, University of Minnesota Twin Cities, Minneapolis, MN 55455 USA (e-mail: lus@umn.edu).

Z. Wang is with the Department of Electrical and Computer Engineering, Iowa State University, Ames, IA 50011 USA (e-mail: zhengdao@iastate.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCOMM.2018.2876416

Furthermore, the energy allocation and training duration are also optimized for uplink multiuser (MU) MIMO systems in a systematic way. Two linear receivers, maximum ratio combining (MRC) and zero-forcing (ZF), are adopted with imperfect CSI. The average and peak power constraints are both incorporated. We analyze the convexity of this optimization problem, and derive the optimal solution in small-scale fading channels. The solution is in the closed form except in one case where a one-dimensional search of a quasi-concave function is needed. We also develop an iterative algorithm of optimizing the energy allocation in large-scale fading channels. Simulation results are also provided to demonstrate the benefit of optimized training, compared with the equal power allocation considered in the literature, and also illustrate the effect of the peak power constraint on the spectral efficiency (SE) and energy efficiency (EE).

In summary, the main contributions of this paper are:

1) We quantify the total degrees of freedom (DoF) with estimated channels, and the needed transmission power for achieving a given rate when $M, T \gg 1$, which is a refinement of the corresponding result in [12].

2) We provide a complete solution for the optimal training duration and training energy in an uplink MU-MIMO system with both MRC and ZF receivers, under both average and peak power constraints in small-scale fading channels.

3) We also develop an iterative algorithm of balancing the energy expense between the training and data phases with the two receivers in large-scale fading channels under both average and peak power constraints.

B. Related Works

An optimized energy reduction scheme was proposed in [19] for uplink MU-MIMO in a single cell scenario, where both RF transmission power and circuit power consumption were incorporated. In [12] and [20], the achievable rates with perfect or estimated CSI were derived and scaling laws were obtained in terms of the power savings as the number base station antennas was increased. However, the training power and duration were not optimized for rate maximization in channel estimation. In order to take full use of the advantages of the massive MIMO systems, balancing the energy expense between the channel estimation and data transmission takes an important role in improving energy efficiency of the system. The issue of non-ideal hardware and its effect on the achievable rates were investigated in [15] and [21].

A joint pilot and data power control method with a minimum mean-squared error (MMSE) receiver was proposed in [22], which minimized the sum power expense under the signal to interference-plus-noise ratio (SINR) and power constraints of each user. Previous work in [16] maximized the sum SE with respect to power and training duration jointly for the MRC receiver, however every user was assigned the same training power. The sum SE maximization problem was reformulated as a convex problem [23], where the pilot and payload power control for each user were jointly optimized in the single cell massive MIMO systems with MRC and ZF receivers. Unfortunately, the reformulated problem is not equivalent to the original problem. For example, in certain cases the objective value of the reformulated problem could be infinite.

Training design and optimization for uplink massive MIMO systems in a multi-cell setup has been performed in [24], where the problem of insufficient pilots is addressed and non-orthogonal pilots are optimized to maximize the system SE. The problem of optimizing the training pilot duration and update interval was considered in [25], for a massive MIMO system with the MRC receiver. Power allocation for downlink massive MIMO transmission has been considered in [26], where MMSE channel estimation is considered. More recently, the joint power allocation and user association optimization is proposed for multi-cell massive MIMO downlink systems [27], where each user is served by a subset of base stations such that the total transmit power is minimized by optimizing each user's transmit power. Instead of solving a combinatorial assignment problem, a new structure of the pilot signals is proposed by using pilot basis in uplink multi-cell massive MIMO systems [28]. The pilot design problem is further formulated as a max-min fairness problem, where the pilot and data powers of each user are optimized by an iterative locally optimal algorithm.

Notation: We use $\mathbf{A}^{\mathsf{H}}$ to denote the Hermitian transpose of matrix $\mathbf{A}$, $\mathbf{I}_K$ to denote a $K \times K$ identity matrix, $\mathbb{C}$ to denote the complex number set, $\lfloor \cdot \rfloor$ to denote the integer floor operation, *i.i.d.* to denote "independent and identically distributed", and $\mathcal{CN}(0,1)$ to denote circularly symmetric complex Gaussian distribution with zero mean and unit variance.

II. System Model

Consider a single-cell uplink system, where there are K mobile users and one base station. Each user is equipped with one transmit antenna, and the base station is equipped with M receive antennas. The received signal at the base station is expressible as

$$\mathbf{y} = \mathbf{H}\mathbf{P}^{\frac{1}{2}}\mathbf{s} + \mathbf{n}, \tag{1}$$

where $\mathbf{H} \in \mathbb{C}^{M\times K}$ is the channel matrix, matrix $\mathbf{P} = \text{diag}\{p_1, \ldots, p_K\} \in \mathbb{R}^{K\times K}$ is diagonal where each entry models the path gain and shadowing effect between the base station and the kth user, $\mathbf{s} \in \mathbb{C}^{K\times 1}$ is the transmitted signals from all the K users; $\mathbf{n} \in \mathbb{C}^{M\times 1}$ is the additive noise, $\mathbf{y} \in \mathbb{C}^{M\times 1}$ is the received signal. We make the following assumptions:

A1) The channel is block fading such that within a *coherence interval* of T channel uses the channel remains constant. Namely, we assume that the channel coherence interval in seconds is equal to T times the symbol duration T_s in seconds. The entries of $\mathbf{H}$ are *i.i.d.* and taken from $\mathcal{CN}(0,1)$. The CSI is neither available at the transmitters nor at the receiver.

A2) Entries of the noise vector $\mathbf{n}$ are *i.i.d.* and from $\mathcal{CN}(0,1)$.

A3) The average transmit power per user per symbol is ρ. So within a coherence interval the total transmitted energy is ρT.

In summary, the system has four parameters, (M, K, T, ρ). We will allow the system to operate in the ergodic regime,

so coding and decoding can occur over multiple coherent intervals.

A. Channel Estimation

We assume that $K \leq M$ and $K < T$ in this section. To derive the achievable rates for the users, we use a well-known scheme that consists of two phases (see e.g., [17]):
Training Phase. This phase consists of T_τ time intervals. The K users send time-orthogonal signals at power level ρ_τ per user. The training signal transmitted can be represented by a $K \times T_\tau$ matrix $\mathbf{\Phi}$ such that $\mathbf{\Phi}\mathbf{\Phi}^{\mathsf{H}} = E\mathbf{I}_K$, where $E = \rho_\tau T_\tau$ is the total training energy per user per coherent interval. Note that we require $T_\tau \geq K$ to satisfy the time-orthogonality.
Data Transmission Phase. Information-bearing symbols are transmitted by the users in the remaining $T_d = T - T_\tau$ time intervals. The average power per symbol per user is $\rho_d = (\rho T - E)/T_d$.

In the training phase, we will choose $\mathbf{\Phi} = \sqrt{E}\mathbf{I}_K$ for simplicity. Other scaled unitary matrices can also be used without affecting the achievable rate. Note that the transmission power is allowed to vary from the training phase to the data transmission phase. With our choice of $\mathbf{\Phi}$, the received signal $\mathbf{Y}_p \in \mathbb{C}^{M\times T_\tau}$ during the training phase can be written as

$$\mathbf{Y}_p = \mathbf{G}\mathbf{\Phi} + \mathbf{N} = \sqrt{E}\mathbf{G} + \mathbf{N}, \tag{2}$$

where $\mathbf{N} \in \mathbb{C}^{M\times T_\tau}$ is the additive noise and $\mathbf{G} = \mathbf{H}\mathbf{P}^{\frac{1}{2}}$. The equation describes $M \times T_\tau$ independent identities, one for each channel coefficient. The (linear) MMSE estimate for the channel is given by $\widehat{\mathbf{G}} = \frac{1}{\sqrt{E}}\mathbf{Y}_p(\mathbf{P}^{-1}/E + \mathbf{I})^{-1}$ [29], where the kth column of $\widehat{\mathbf{G}}$ is

$$\widehat{\mathbf{G}}_k = \frac{p_k^{\frac{3}{2}}E}{p_k E + 1}\mathbf{h}_k + \frac{p_k\sqrt{E}}{p_k E + 1}\mathbf{n}_k, \tag{3}$$

where $\mathbf{h}_k$ and $\mathbf{n}_k$ are the kth column of $\mathbf{H}$ and $\mathbf{N}$.

The channel estimation error is defined as $\widetilde{\mathbf{G}} = \mathbf{G} - \widehat{\mathbf{G}}$. Thus, we have the kth column of $\widetilde{\mathbf{G}}$, i.e., $\widetilde{\mathbf{G}}_k = \mathbf{G}_k - \widehat{\mathbf{G}}_k = \frac{\sqrt{p_k}}{p_k E+1}\mathbf{h}_k - \frac{p_k\sqrt{E}}{p_k E+1}\mathbf{n}_k$, where $\mathbf{G}_k$ denotes the kth column of $\mathbf{G}$. It is easy to verify that the elements of $\widehat{\mathbf{G}}$ are column-wise *i.i.d.* complex Gaussian with zero mean and variance $\sigma^2_{\widehat{\mathbf{G}}_k} = \frac{p_k^2 E}{p_k E+1}$, and the elements of $\widetilde{\mathbf{G}}$ are column-wise *i.i.d.* complex Gaussian with zero mean and variance $\sigma^2_{\widetilde{\mathbf{G}}_k} = \frac{p_k}{p_k E+1}$. Moreover, $\widehat{\mathbf{G}}$ and $\widetilde{\mathbf{G}}$ are in general uncorrelated as a property of the linear MMSE estimator under the Gaussian assumptions.

B. Equivalent Channel

Once the channel is estimated, the base station has $\widehat{\mathbf{G}}$ and will decode the users' information using $\widehat{\mathbf{G}}$. We can write the received signal as

$$\mathbf{y} = \widehat{\mathbf{G}}\mathbf{s} + \widetilde{\mathbf{G}}\mathbf{s} + \mathbf{n} \triangleq \widehat{\mathbf{G}}\mathbf{s} + \mathbf{v}, \tag{4}$$

where $\mathbf{v} \triangleq \widetilde{\mathbf{G}}\mathbf{s}+\mathbf{n}$ is the new equivalent noise containing actual noise $\mathbf{n}$ and self interference $\widetilde{\mathbf{G}}\mathbf{s}$ caused by inaccurate channel estimation. Assuming that each element of $\mathbf{s}$ has variance ρ_d during the data transmission phase, and there is no cooperation among the users, the variance of each component of $\mathbf{v}$ is $\sigma_v^2 = \sum_{i=1}^K \frac{\rho_d p_i}{p_i E+1} + 1$.

If we replace $\mathbf{v}$ with a zero-mean complex Gaussian noise with equal variance σ_v^2, but independent of $\mathbf{s}$, then the system described in (4) can be viewed as a MIMO system with perfect CSI at the receiver, and the equivalent signal to noise ratio (SNR) of the kth user is

$$\begin{aligned}\rho_{\text{eff},k} &\triangleq \frac{\rho_d\sigma^2_{\widehat{\mathbf{G}}_k}}{\sigma_v^2} = \frac{\rho_d p_k^2 E}{(p_k E+1)(\sum_{i=1}^K \frac{\rho_d p_i}{p_i E+1}+1)} \\ &= \frac{\rho_d p_k^2}{(p_k + \frac{1}{E})(\sum_{i=1}^K \frac{\rho_d p_i}{p_i E+1}+1)}.\end{aligned} \tag{5}$$

The SNR is the signal power from a single transmitter per receive antenna divided by the noise variance per receive antenna. It is a standard argument that a noise equivalent to $\mathbf{v}$ but assumed independent of $\mathbf{s}$ is "worse" (see e.g., [17]). As a result, the derived rate based on such an assumption is achievable. In the following, for notational brevity, we assume that $\mathbf{v}$ in (4) is independent of $\mathbf{s}$ without introducing a new symbol to represent the equivalent *independent* noise.

Note that the effective SNR $\rho_{\text{eff},k}$ is the actual SNR ρ_d divided by a loss factor $(p_k + \frac{1}{E})(\sum_{i=1}^K \frac{\rho_d^k p_i}{p_i E+1}+1)$. The loss factor can be made small if the energy E used in the training phase is large.

C. Energy Splitting Optimization

The energy in the training phase can be optimized to maximize the effective SNR $\rho_{\text{eff},k}$ in (5) for point-to-point MIMO system, as has been done in [17, Th. 2]. Importantly, with the effective SNR adopted in this paper, the achievable rate with MRC and ZF receivers can be easily optimized in a closed form.

We assume the average transmitted power over one coherence interval T is equal to a given constant ρ, namely $\rho_d T_d + \rho_\tau T_\tau = \rho T$. Let $\alpha \triangleq \rho_\tau T_\tau/(\rho T)$ denote the fraction of the total transmit energy that is devoted to channel training; i.e.,

$$\rho_\tau T_\tau = \alpha\rho T, \quad \rho_d T_d = (1-\alpha)\rho T, \quad 0 \leq \alpha \leq 1. \tag{6}$$

III. Achievable Rates and DoF

A. Rates of Linear Receivers

Given the channel model (4), linear processing can be applied to $\mathbf{y}$ to recover $\mathbf{s}$, as in e.g., [12]. Let $\mathbf{A} \in \mathbb{C}^{K\times M}$ denote the linear processing matrix. The processed signal is

$$\widehat{\mathbf{s}} \triangleq \mathbf{A}\mathbf{y} = \mathbf{A}\widehat{\mathbf{G}}\mathbf{s} + \mathbf{A}\mathbf{v}. \tag{7}$$

The MRC processing is obtained by setting $\mathbf{A} = \widehat{\mathbf{G}}^{\mathsf{H}}$. The ZF processing is obtained by setting $\mathbf{A} = (\widehat{\mathbf{G}}^{\mathsf{H}}\widehat{\mathbf{G}})^{-1}\widehat{\mathbf{G}}^{\mathsf{H}}$.

Based on the equivalent channel model, viewed as a multi-user MIMO systems with perfect receiver CSI and equivalent SNR $\rho_{\text{eff},k}$, the achievable rates for lower bounds derived in [12, Propositions 2 and 3] can then be applied. Also, setting the training period equal to the total number of transmit antennas possesses certain optimality as derived in [17], which means

$T_\tau^* = K$. Specifically, for MRC the following ergodic sum SE is achievable:

$$R^{(\text{MRC})} \triangleq (1 - \frac{K}{T}) \sum_{k=1}^{K} \log_2 \left(1 + \frac{\rho_{\text{eff},k}(M-1)}{\sum_{i=1, i\neq k}^{K} \rho_{\text{eff},i} + 1}\right). \quad (8)$$

For ZF, assuming $M > K$, the following ergodic sum SE is achievable:

$$R^{(\text{ZF})} \triangleq \left(1 - \frac{K}{T}\right) \sum_{k=1}^{K} \log_2 \left(1 + \rho_{\text{eff},k}(M-K)\right). \quad (9)$$

Note that the factor $(1 - \frac{K}{T})$ is due to the fact that during one coherence interval of length T, K time slots have been used for the training purpose. The number of data transmission slots is $T - K$, and the achieved rate needs to be averaged over T channel uses. Also, these rates are actually lower bounds on achievable rates (due to the usage of Jensen's inequality).

We will analyze the DoF in the next section.

B. Degrees of Freedom

We define the DoF of the system as

$$d(M, K, T) \triangleq \sup \lim_{\rho \to \infty} \frac{R^{(\text{total})}(\rho)}{\log_2(\rho)}, \quad (10)$$

where the supremum is taken over the totality of all reliable communication schemes for the system, and $R^{(\text{total})}$ denotes the sum rate of the K users under the power constraint ρ. We may also speak of the (achieved) DoF of one user for a particular achievability scheme, which is the achieved rate of the user normalized by $\log_2(\rho)$ in the limit of $\rho \to \infty$. The DoF measures the multiplexing gain offered by the system when compared to a reference point-to-point single-antenna communication link, in the high SNR regime (see e.g., [30]).

Theorem 1: For an (M, K, T) MIMO uplink system with M receive antennas, K users, and coherence interval T, the total DoF of the system is

$$d(M, K, T) = K^\dagger \left(1 - \frac{K^\dagger}{T}\right), \quad (11)$$

where $K^\dagger \triangleq \min(M, K, \lfloor T/2 \rfloor)$.

Proof: To prove the converse, we observe that if we allow the K transmitters to cooperate, then the system is a point-to-point MIMO system with K transmit antennas, M receive antennas, and with no CSI at the receiver. The DoF of this channel has been quantified in [18], in the same form as in the theorem. Without cooperation, the users can at most achieve a rate as high as in the cooperation case.

To prove the achievability, we first look at the case $K^\dagger < M$. In this case, we note that if we allow only $K^\dagger$ users to transmit, and let the remaining users be silent, then using the achievability scheme describe in Section II-A, each of the $K^\dagger$ users can achieve a rate per user using the zero-forcing receiver given as follows (cf. (16))

$$\left(1 - \frac{K^\dagger}{T}\right) \log_2 \left(1 + \rho_{\text{eff},k}(M - K^\dagger)\right) \quad \forall k. \quad (12)$$

Note that the condition $K^\dagger < M$ is needed. If we choose $E = K^\dagger \rho$ and $\rho_d = \rho$, then the effective SNR in (5) becomes

$$\rho_{\text{eff},k} = \frac{\rho \frac{K^\dagger p_k^2}{K^\dagger p_k + 1/\rho}}{\sum_{i=1}^{K} \frac{p_i}{K^\dagger p_i + 1/\rho} + 1}, \quad \forall k. \quad (13)$$

It can be seen that as $\rho \to \infty$, $\log(\rho_{\text{eff},k})/\log(\rho) \to 1$ and a DoF of the kth user of $(1 - K^\dagger/T)$ is achieved. The total achieved DoF is therefore $K^\dagger(1 - K^\dagger/T)$. Although better energy splitting is possible, as in Section II-C, it will not improve the DoF.

When $K^\dagger = M$, the case is more subtle. In this case the zero-forcing receiver is no longer sufficient. In fact, even the optimal linear processing, which is the MMSE receiver [12, eq. (31)], is not sufficient. The insufficiency can be established by using the results in [31, Sec. IV-C] to show that as $\rho \to \infty$, the effective SNR at the output of MMSE receiver has a limit distribution that is independent of SNR. We skip the details here, since it is not the main concern in this paper.

Instead, we notice that the equivalent channel (4) has an SNR given by (13), which for $K^\dagger \rho > 1$ is greater than $\rho/3$. So, the MIMO system can be viewed as a multiple access channel (MAC) with $K^\dagger$ single-antenna transmitters, and one receiver with M receive antennas. Perfect CSI is known at the receiver, and the SNR between $\rho/3$ and ρ. Using the MAC capacity region result [32, Th. 14.3.1], [33, Sec. 10.2.1], it can be shown that a total DoF of $K^\dagger$ can be achieved over $T - K^\dagger$ the time slots. □

Remark 1: The DoF is the same as that of a point-to-point MIMO channel with K transmit antennas and M receive antennas without transmit- or receive-side CSI [18]. This is not trivial because optimal signaling over non-coherent MIMO channel generally requires cooperation among the transmit antennas. It turns out that as far as DoF is concerned, transmit antenna cooperation is not necessary. However, we do note that user synchronization is needed to prove the result. It is an interesting problem to study the DoF in the asynchronous case.

Remark 2: It can be seen from the achievability proof that for $M > K$, which is generally applicable for massive MIMO systems, ZF at the base station is sufficient for achieving the optimal DoF. However, MRC is not sufficient because ρ shows up both in the numerator and denominator of (15). So as $\rho \to \infty$, the achieved rate is limited. This is due to the interference among the users.

Remark 3: For the case $K^\dagger = M$, non-linear decoding such as successive interference cancellation is needed.

Remark 4: When T is large, a per-user DoF close to 1 is achievable, as long as $K \leq M$.

Remark 5: When M is larger than $K^\dagger$, increasing M further has no effect on the DoF. However, it is clear that more receive antennas is useful because more energy is collected by additional antennas. We will discuss the benefit of energy savings in the next section.

C. Power Savings for Fixed Rate

As more antennas are added to the base station, it is possible that less energy is needed to be transmitted from the mobile stations. Also, in a practical system, the channel statistic information is provided from the downlink, and adaptive power control mechanisms can be adopted for the block fading channel, and thus most of the effect of large-scale fading can be compensated [34]. For this reason and analytical tractability, we will consider the case where is no large-scale fading in the analysis. We note that the derived algorithms are applicable to the case where large-scale fading is present.

Specifically, the effective SNR shown in (13) becomes

$$\rho_{\text{eff}} \triangleq \frac{\rho_d \sigma^2_{\widehat{\mathbf{H}}}}{\sigma_v^2} = \frac{\rho_d E}{K\rho_d + E + 1} = \frac{\rho_d}{1 + \frac{K\rho_d + 1}{E}}. \tag{14}$$

Consequently, for MRC the following ergodic sum SE is achievable:

$$R^{(\text{MRC})} \triangleq K\big(1 - \frac{K}{T}\big) \log_2 \big(1 + \frac{\rho_{\text{eff}}(M-1)}{\rho_{\text{eff}}(K-1)+1}\big). \tag{15}$$

For ZF, assuming $M > K$, the following ergodic sum SE is achievable:

$$R^{(\text{ZF})} \triangleq K\big(1 - \frac{K}{T}\big) \log_2 \left(1 + \rho_{\text{eff}}(M-K)\right). \tag{16}$$

It can be seen from (15) and (16) that when ρ is small, MRC performs better than ZF, which has been previously observed, e.g., [12]. On the other hand, in the low-SNR regime the difference between them is a constant factor $(M-1)/(M-K)$ in the SNR term within the logarithmic functions in (15) and (16). The difference becomes negligible when M is large. Using either result, and the effective SNR in (36), we are able to obtain the following:

If we fix the per-user rate at $R = (1 - K/T)\log_2(1+\rho_0)$, then the required power ρ is

$$\rho = \sqrt{\frac{4\rho_0(T-K)}{MT^2}} + o\left(\frac{1}{\sqrt{M}}\right). \tag{17}$$

This can be proved by setting $\rho M = \rho_0$ in the rate expression for ZF. Since the achievable rate with ZF processing is worse than MRC and MMSE when SNR is very low, the result is still applied for MRC and MMSE processing.

It is interesting to note that increasing T has a similar effect as increasing M on the required transmission power, reducing the power by $1/\sqrt{M}$ or $1/\sqrt{T}$. The reason is the if T is increased, then the energy that can be expended on training is increased, improving the quality of channel estimation, especially in the case where there is a peak power constraint. On the other hand, for (17) to be applicable, we need $M \gg K$.

IV. Joint Optimization of Power Allocation and Training Duration

If the peak power, rather than the average power, is limited, then our DoF result still holds because the achievability proof actually uses equal power in the training and data transmission phases. The power savings discussion in the previous subsection still applies, because the system is limited by the total amount of energy available, and not how the energy is expended. In the regime where the SNR is neither very high or very low, the peak power constraint will affect the rate. Also, there is a peak power limit for hardware implementation in practice. We provide a detailed analysis in this section.

a) Energy allocation: We assume that the transmitters are subject to both average power constraint, and peak power constraint:

$$0 \le \rho_d, \quad \rho_\tau \le \rho_{\max}. \tag{18}$$

b) Problem formulation: For an adopted receiver, $\mathcal{A} \in \{\text{MRC}, \text{ZF}\}$, our goal is to maximize the uplink SE subject to the peak and average power constraints. Based on the model in (4), we will consider two linear demodulation schemes: MRC and ZF receivers.

Consider the case where the large-scale fading is compensated. For the MRC receiver, the received SNR for any of the K users' symbols can be obtained by substituting ρ_{eff} into $\rho_{\text{eff}}(M-1)/(\rho_{\text{eff}}(K-1)+1)$ (see [12, eq. (16)]):

$$\mathsf{SNR}^{(\text{MRC})} = \frac{T_\tau \rho_\tau \rho_d (M-1)}{T_\tau \rho_\tau \rho_d (K-1) + K\rho_d + T_\tau \rho_\tau + 1}. \tag{19}$$

For the ZF receiver, the received SNR for any of the K users' symbols can be obtained by substituting ρ_{eff} into $\rho_{\text{eff}}(M-K)$ (see [12, eq. (20)]):

$$\mathsf{SNR}^{(\text{ZF})} = \frac{T_\tau \rho_\tau \rho_d (M-K)}{K\rho_d + T_\tau \rho_\tau + 1}. \tag{20}$$

For either receiver, a lower bound on the sum SE achieved by the K users is given by

$$R^{(\mathcal{A})}(\alpha, T_d) = \frac{T_d}{T} K \log_2(1 + \mathsf{SNR}^{(\mathcal{A})}), \tag{21}$$

where $\mathcal{A} \in \{\text{MRC}, \text{ZF}\}$.

Our optimization problem can be formulated as follows:

$$\begin{aligned}
(\mathbf{OP}) \qquad & \underset{\alpha, T_d}{\text{maximize}} \ R^{(\mathcal{A})}(\alpha, T_d) && \text{(22a)}\\
& \text{subject to } \rho T\alpha + \rho_{\max} T_d \le \rho_{\max} T, && \text{(22b)}\\
& \qquad -\rho T\alpha - \rho_{\max} T_d \le -\rho T, && \text{(22c)}\\
& \qquad 0 \le \alpha \le 1, && \text{(22d)}\\
& \qquad 0 < T_d \le T - K, && \text{(22e)}
\end{aligned}$$

where $R^{(\mathcal{A})}(\alpha, T_d)$ is as given in (21); (22b) and (22c) are from the peak power constraints in the training and data phases, respectively; and the last constraint is from the requirement that $T_\tau \ge K$.

A. SNR Maximization When T_d is Fixed

The feasible set of the problem (OP) is convex, but the convexity of the objective function is not obvious. In this section, we consider the optimization problem when T_d is fixed. In this case, we will prove that $R^{(\mathcal{A})}(\alpha, T_d)$ is concave in α, and derive the optimized α. The result will be useful in the next section where α and T_d are jointly optimized.

For a fixed T_d, from the peak power constraints (22b) and (22c), we have

$$\frac{\rho_{\max} T_\tau}{\rho T} + \big(1 - \frac{\rho_{\max}}{\rho}\big) \le \alpha \le \frac{\rho_{\max} T_\tau}{\rho T}. \tag{23}$$

Combined with (22d), the overall constraints on α are

$$\min\left\{0, \frac{\rho_{\max} T_\tau}{\rho T} + \left(1 - \frac{\rho_{\max}}{\rho}\right)\right\} \le \alpha \le \max\left\{\frac{\rho_{\max} T_\tau}{\rho T}, 1\right\}. \tag{24}$$

In the remaining part of this section, we will first ignore the peak power constraint, and derive the optimal $\alpha \in (0,1)$ for a given T_d. At the end of this section, we will reconsider the effect of the peak power constraint on the optimal α.

1) MRC Case Without Peak Power Constraint: Using (6) we can rewrite (19) as

$$\mathsf{SNR}^{(\mathrm{MRC})}(\alpha) = \frac{M-1}{K-1} \frac{\alpha(\alpha-1)}{\alpha^2 - a_1\alpha - b_1}, \tag{25}$$

where

$$a_1 = 1 + \frac{T_d - K}{\rho T(K-1)}, \quad b_1 = \frac{\rho TK + T_d}{\rho^2 T^2 (K-1)} > 0. \tag{26}$$

It can be verified that $1 - a_1 - b_1 \le 0$.

a) Behavior of the $\mathsf{SNR}^{(MRC)}(\alpha)$ function: Define

$$g(\alpha) := \mathsf{SNR}^{(\mathrm{MRC})} \cdot (K-1)/(M-1). \tag{27}$$

And let $g_d(\alpha) = \alpha^2 - a_1\alpha - b_1$, which is the denominator of $g(\alpha)$.

Lemma 1: The function $g(\alpha)$ is concave in α over $(0,1)$ when $1 - a_1 - b_1 \le 0$ and $b_1 > 0$.

Proof: The proof is elementary but cumbersome, see [2, Lemma 1] for details. □

Lemma 1 gives the convex conditions of the objective function. According to Lemma 1, we know that there is a global maximal point for (25). Taking the derivative of (25) and setting it as 0, we have

$$(1 - a_1)\alpha^2 - 2b_1\alpha + b_1 = 0. \tag{28}$$

Remark 6: It can be observed that when $1 - a_1 - b_1 \le 0$ and $b_1 > 0$, $g_d(\alpha)$ is non-positive at both $\alpha = 0$ and $\alpha = 1$. Since the leading coefficient of $g_d(\alpha)$ is positive, $g_d(\alpha) < 0$ for $\alpha \in (0,1)$, and it has no root in $(0,1)$.

Based on Remark 6, we deduce that $g(\alpha) > 0$ for $\alpha \in (0,1)$. In addition, we have $g(0) = 0$ and $g(1) = 0$. Therefore, there is an optimal α within $(0,1)$ rather than at boundaries.

b) Optimizing α: we discuss the optimal α in three cases, depending on T_d, as compared to K.

- If $T_d = K$, then $1 - a_1 = 0$. Hence, we have $\alpha^* = 1/2$, and
$$\mathsf{SNR}^{(\mathrm{MRC})}\left(\frac{1}{2}\right) = \frac{M-1}{K-1} \frac{1/4}{1/4 + \frac{K(\rho T+1)}{\rho^2 T^2 (K-1)}}. \tag{29}$$
- If $T_d < K$, then $1 - a_1 > 0$. Since $b_1 > 1 - a_1$, $b_1/(1-a_1) > 1$. Between the two roots of (28), the one in between 0 and 1 is
$$\alpha^* = \frac{b_1 - \sqrt{b_1(a_1 + b_1 - 1)}}{1 - a_1}. \tag{30}$$
- If $T_d > K$, then $1 - a_1 < 0$. It can be deduced that in this case α^* in (30) is still between 0 and 1 and therefore is the optimal α.

Substituting (26) into (30), we have

$$\alpha^* = \frac{\sqrt{(\rho TK + T_d)(\rho T T_d + T_d)} - (\rho TK + T_d)}{\rho T(T_d - K)}. \tag{31}$$

We can simplify the expression for the optimal α at both high and low SNR regions:

- At the high SNR region, the optimal α is
$$\alpha^*_{\mathrm{H}} \approx \frac{\sqrt{KT_d} - K}{T_d - K} = \frac{\sqrt{K}}{\sqrt{T_d} + \sqrt{K}}. \tag{32}$$
- Similarly, at the low SNR regime, the optimal α is $\alpha^*_{\mathrm{L}} \approx 1/2$.
As a result, $\mathsf{SNR}^{(\mathrm{MRC})}(\alpha^*_{\mathrm{L}}) = (M-1)/(4T_d(K-1))$. If the SNR is low, the fraction between the training and data is independent on system parameters M, K, ρ_d, ρ_τ, T_τ, and T.

So far we have ignored the peak power constraint. When the peak power is considered, and α^* is not within the feasible set (24), the optimal $\widetilde{\alpha}^*$ with the peak power constraint is the α within the feasible set that is closest to the α^* we derived, which is at one of the two boundaries of the feasible set, due to the concavity of the objective function.

2) ZF Case Without Peak Power Constraint: This optimization problem in the ZF case is similar to that in Section. IV-A.1. Here, we only give the final optimization results.

Using (6) we can rewrite (20) as

$$\mathsf{SNR}^{(\mathrm{ZF})}(\alpha) = \frac{\rho T(M-K)\alpha(1-\alpha)}{(T_d - K)(\gamma + \alpha)}. \tag{33}$$

Define an auxiliary variable when $T_d \neq K$: $\gamma \triangleq \frac{K\rho T + T_d}{\rho T(T_d - K)}$, which is positive if $T_d > K$ and negative if $T_d < K$.

It can be easily verified that in all the three cases, namely $T_d = K$, $T_d > K$, and $T_d < K$, ρ_{eff} is concave in α within $\alpha \in (0,1)$. The optimal value for α that maximizes ρ_{eff} is given as follows:

$$\alpha^* = \begin{cases} -\gamma + \sqrt{\gamma(\gamma+1)}, & T_d > K \\ \dfrac{1}{2}, & T_d = K \\ -\gamma - \sqrt{\gamma(\gamma+1)}, & T_d < K. \end{cases} \tag{34}$$

The maximized effective SNR ρ^*_{eff} is given as

$$\rho^*_{\mathrm{eff}} = \begin{cases} \dfrac{\rho T}{T_d - K}(-2\sqrt{\gamma(\gamma+1)} + (1+2\gamma)), & T_d > K \\ \dfrac{(\rho T)^2}{4K(1+\rho T)}, & T_d = K \\ \dfrac{\rho T}{T_d - K}(2\sqrt{\gamma(\gamma+1)} + (1+2\gamma)), & T_d < K. \end{cases} \tag{35}$$

At the high SNR region ($\rho \gg 1$), we have $\gamma \approx \frac{K}{T_d - K}$, and the optimal values are $\alpha^*_{\mathrm{H}} \approx \sqrt{K}/(\sqrt{T_d} + \sqrt{K})$, $\rho^*_{\mathrm{eff}} \approx \frac{T}{(\sqrt{T_d} + \sqrt{K})^2}\rho$.

At the low SNR region ($\rho \ll 1$), we have $\gamma \approx \frac{T_d}{\rho T(T_d - K)}$, and the optimal values are

$$\alpha^*_{\mathrm{L}} \approx \frac{1}{2}, \quad \rho^*_{\mathrm{eff}} \approx \frac{(\rho T)^2}{4T_d}. \tag{36}$$

The optimized $\mathsf{SNR}^{(\mathrm{ZF})}$ is just given by $(M-K)\rho^*_{\mathrm{eff}}$.

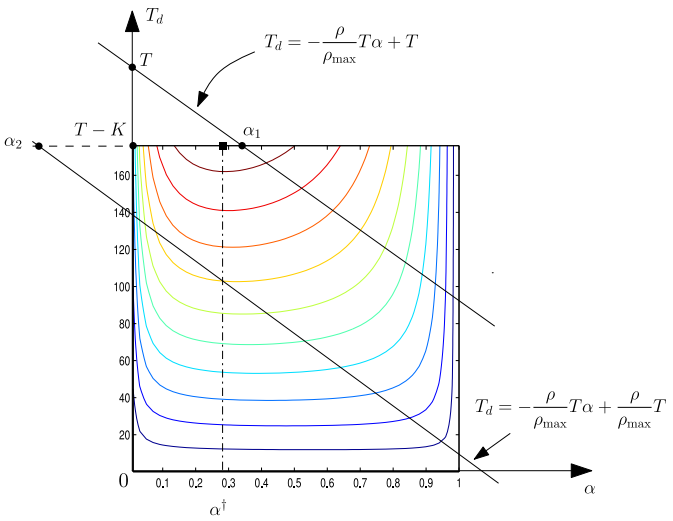


Fig. 1. Feasible region and the contour of the objective function in the MRC case; $T = 196$, $K = 20$ and $M = 50$.

B. Training Optimization With Peak Power Constraint

In this section, α and T_d are jointly optimized for maximizing the achievable rate of the uplink MU-MIMO system as illustrated in (22)–(22e) when both average and peak power constraints are considered.

The feasible set with respect to α and T_d is illustrated in Fig. 1. It can be observed that the feasible region is in between the following two lines

$$T_d = -\rho T\alpha/\rho_{\max} + T, \tag{37}$$

$$T_d = -\rho T\alpha/\rho_{\max} + \rho T/\rho_{\max}, \tag{38}$$

where α and T_d satisfy (22d) and (22e).

We have the following lemma that is useful for describing the shape of our objective function $R^{(\mathcal{A})}(\alpha, T_d)$ when α is fixed.

Lemma 2: The function $f(x) = x\ln(1+a/(b+cx))$, when $a, b, c, x > 0$, is concave and monotonically increasing.

Proof: It can be verified by taking the second derivative. See [2, Lemma 2] for details. □

In summary, the objective function has the following two properties:

(**P1**) From Lemma 1, for fixed T_d, $R^{(\mathcal{A})}$ is a concave function with respect to α.

(**P2**) From Lemma 2, for fixed α, $R^{(\mathcal{A})}$ is a concave function and monotonically increasing with respect to T_d.

Since the feasible set is convex, our optimization problem (OP) is a biconvex problem that may include multiple local optimal solutions. However, after studying the convexity of the objective function, we can give the global optimal solutions for both MRC and ZF receivers in the following Theorem and Corollary.

In the remainder of this section, let $\alpha^\dagger$ denote the optimal α when $T_d = T - K$, which is given by Section IV-A.1 and Section IV-A.2 for ZF and MRC processing.

Theorem 2: For the MRC receiver, set $\alpha^\dagger = 1/2$ if $T_d = K$ and otherwise set $\alpha^\dagger$ according to (31) when $T_d = T - K$. Set $\alpha_1 = \rho_{\max}K/\rho T$ and set $\alpha_2 = 1 - \rho_{\max}(T-K)/\rho T$. The solution for the joint optimization of training energy allocation α and the training duration $T_\tau = T - T_d$ is given in three cases. Case 1) If $\alpha_1 < \alpha^\dagger$, then α^* is given by the maximizer of $R^{(\text{MRC})}(\alpha)$ in (69), and $T_d^* = -\rho T\alpha^*/\rho_{\max} + T$; Case 2) If $\alpha_2 > \alpha^\dagger$ then $(\alpha^*, T_d^*) = (\alpha_2, T-K)$; Case 3) If $\alpha_2 < \alpha^\dagger < \alpha_1$, then $(\alpha^*, T_d^*) = (\alpha^\dagger, T-K)$.

Proof: Please see Appendix A □

We also have similar results regarding the optimal energy allocation factor α and training period T_τ for the ZF case. The only difference is that SE $R^{(\text{ZF})}(\alpha)$ should be given by substituting (37) into $R^{(\text{ZF})}(\alpha, T_d)$, which is

$$R^{(\text{ZF})}(\alpha) = \frac{K}{T}\left(-\frac{\rho T}{\rho_{\max}}\alpha + T\right)\log_2(1 + \mathsf{SNR}^{(\text{ZF})}(\alpha)), \tag{39}$$

where

$$\mathsf{SNR}^{(\text{MRC})}(\alpha) = \frac{\alpha(\alpha-1)\rho^2T^2(M-K)}{a_3\alpha^2 - b_3\alpha - c_3}, \tag{40}$$

and $a_3 = \rho^2T^2/\rho_{\max}$, $b_3 = \rho T^2 - \rho TK - \rho T/\rho_{\max}$ and $c_3 = K\rho T + T$. Comparing (69), (70) and (39), (40), we can obtain the results for the ZF receiver straightforwardly as follows.

Corollary 1: For the ZF receiver, set $\alpha^\dagger = 1/2$ if $T_d = K$ and otherwise set $\alpha^\dagger$ according to (34) when $T_d = T - K$. Set $\alpha_1 = \rho_{\max}K/\rho T$ and set $\alpha_2 = 1 - \rho_{\max}(T-K)/\rho T$. The solution for the joint optimization of training energy allocation α and the training duration $T_\tau = T - T_d$ is given in three cases. Case 1) If $\alpha_1 < \alpha^\dagger$, then α^* is given by the maximizer of $R^{(\text{ZF})}(\alpha)$ in (39), and $T_d^* = -\rho T\alpha^*/\rho_{\max} + T$; Case 2) If $\alpha_2 > \alpha^\dagger$ then $(\alpha^*, T_d^*) = (\alpha_2, T-K)$; Case 3) If $\alpha_2 < \alpha^\dagger < \alpha_1$, then $(\alpha^*, T_d^*) = (\alpha^\dagger, T-K)$.

C. Optimized SE When M is Large

When M increases, the transmit power of each user can be reduced proportionally to $1/\sqrt{M}$ for a large M while maintaining a fixed rate as discussed in Section III-C and [12]. Here we discuss the asymptotic achievable SE when $M \to \infty$.

1) Optimized α if T_d is Fixed When $M \to \infty$: If the energy over the training and data phases is allocated differently, we have the following results after optimizing the α for a large M.

Theorem 3: For both ZF and MRC, let $\rho_u \triangleq \sqrt{M}\rho$ be fixed. Then, the maximum achievable SE can be

$$R^{(\mathcal{A})} \to \frac{T_d}{T}K\log_2(1 + \frac{\rho_u^2T^2}{4T_d}), \quad M \to \infty. \tag{41}$$

Proof: According to (25) and (33), when $M \to \infty$, we have

$$\mathsf{SNR}^{(\mathcal{A})}(\alpha) = \frac{\alpha(1-\alpha)\rho_u^2T^2}{T_d}, \tag{42}$$

where the maximum received SNR can be obviously obtained when $\alpha = 1/2$. □

Note, if the peak power constraints are considered, α needs to be within the interval as shown in (24). Otherwise, the optimal solution is located at the boundary of (24).

Remark 7: If the power is allocated equally between the two phases, we have $\alpha = T_\tau/T$ [12], then the difference of

SE between the optimized and the equally allocated power scheme is

$$\Delta R^{(\mathcal{A})}(\alpha) = \frac{T_d}{T}K(\log_2(1+\frac{\rho_u^2 T^2}{4T_d}) - \log_2(1+T_\tau\rho_u^2)) = \frac{T_d}{T}K\log_2\big(\frac{4T_d+\rho_u^2T^2}{4T_d+4T_d(T-T_d)\rho_u^2}\big), \quad (43)$$

where the numerator minus the denominator within the $\log_2(\cdot)$ is equal to $\rho_u^2(T^2-4TT_d+T_d^2) = \rho_u^2(T-T_d)^2 \geq 0$. Therefore, it is clear that the optimized SE is always larger than the unoptimized one. The gain in rate offered by optimizing the energy allocated for training is given by (43).

2) Optimized α and T_d When $M \to \infty$: For both MRC and ZF, under the peak power constraints, the average transmit power of each user is $\rho = \rho_u/\sqrt{M}$, where ρ_u is fixed. Let $\rho/\rho_{\max} \triangleq \xi$. Consequently, the corresponding $\rho_{\max} = \rho_u/(\xi\sqrt{M})$. When $M \to \infty$, applying Theorem. 2 and Corollary. 1, we have the following cases:

- *Case 1: ρ_τ is limited by ρ_{max}*

$$R^{(\mathcal{A})}(\alpha) = K(-\xi\alpha+1)\log_2\big(1+\frac{\alpha(\alpha-1)\rho_u^2 T}{\xi\alpha-1}\big). \quad (44)$$

  Taking the derivative of (44) and setting it to zero, we can obtain the optimal α with one dimension search algorithm [35]. Then, the duration T_d^* can be obtained by (37) directly with substituting α^*.
- *Case 2: ρ_d is limited by ρ_{max}*

$$R^{(\mathcal{A})}(\alpha^*) = K\xi(-\alpha^*+1)\log_2\big(1+\frac{\alpha^*(\alpha^*-1)\rho_u^2 T}{T-K}\big), \quad (45)$$

  where $\alpha^* = 1-(T-K)/(\xi T)$ and $T_d^* = T-K$.
- *Case 3: Neither ρ_d nor ρ_τ is not limited by ρ_{max}*

$$R^{(\mathcal{A})}(\alpha^*) = \frac{T-K}{T}K\log_2\big(1+\frac{\rho_u^2T^2}{4(T-K)}\big), \quad (46)$$

  where $\alpha^* = 1/2$ and $T_d^* = T-K$.

D. SE in Large-Scale Fading Channels

If the adaptive power control is not used, we consider the case where there exits large-scale fading. For the MRC receiver, the received SNR for any of the K users' symbols can be obtained by substituting (5) into $\rho_{\text{eff},k}(M-1)/(\sum_{i=1,i\neq k}^K \rho_{\text{eff},i}+1)$ (see [12, eq. (16)]):

$$\begin{aligned}\mathsf{SNR}_k^{(\text{MRC})} &= \frac{\rho_d p_k^2 E(M-1)}{\rho_d(p_kE+1)\sum_{i=1,i\neq k}^K p_i + \rho_d p_k + p_kE+1} \\ &= \frac{p_k^2T_\tau\rho_\tau\rho_d(M-1)}{T_\tau\rho_\tau\rho_d p_k\sum_{i=1,i\neq k}^K p_i + \rho_d\sum_{i=1}^K p_i + p_kT_\tau\rho_\tau+1} \\ &= \frac{p_k^2\rho^2T^2(M-1)\alpha(\alpha-1)}{\widetilde{a}_1\alpha^2-\widetilde{b}_1\alpha-\widetilde{c}_1}, \end{aligned} \quad (47)$$

where $\widetilde{a}_1 = \rho^2T^2p_k\sum_{i=1,i\neq k}^K p_i$, $\widetilde{b}_1 = T_dp_k\rho T - \rho T\sum_{i=1}^K p_i + \rho^2T^2p_k\sum_{i=1,i\neq k}^K p_i$ and $\widetilde{c}_1 = \rho T\sum_{i=1}^K p_i + T_d$. It can be verified that $1-(\widetilde{b}_1+\widetilde{c}_1)/\widetilde{a}_1 < 1$ and $\widetilde{c}_1/\widetilde{a}_1 > 0$. Applying Lemma 1, we know that $\mathsf{SNR}_k^{(\text{MRC})}$ is a concave function with respect to $\alpha \in (0,1)$.

For the ZF receiver, the received SNR for any of the K users' symbols can be obtained by substituting (5) into $\rho_{\text{eff},k}(M-K)$ (see [12, eq. (20)]):

$$\mathsf{SNR}_k^{(\text{ZF})} = \frac{\rho_d p_k^2 E(M-K)}{\rho_d p_k\sum_{i=1}^K(1+\frac{\Delta_{ki}}{1+p_iE}) + p_kE+1}, \quad (48)$$

where $\Delta_{ki} \triangleq p_i - p_k \ll 1$. Hence, we can get the lower bound of $\mathsf{SNR}_k^{(\text{ZF})}$, which is given by

$$\begin{aligned}\mathsf{SNR}_k^{(\text{ZF})} &= \frac{p_k^2\rho_d T_\tau\rho_\tau(M-K)}{\widetilde{K}_k p_k\rho_d + p_kT_\tau\rho_\tau+1} \\ &= \frac{p_k\rho T(M-K)(1-\alpha)\alpha}{(T_d-\widetilde{K}_k)(\alpha+\gamma')}, \end{aligned} \quad (49)$$

where $\widetilde{K}_k \triangleq \sum_{i=1}^K(1+\Delta_{ki}) \approx K > 1$ and $\gamma' \triangleq \frac{\widetilde{K}_k p_k\rho T+T_d}{p_k\rho T(T_d-\widetilde{K}_k)}$ when $T_d \neq \widetilde{K}_k$. Similar as in (33), it can be verified that $\mathsf{SNR}_k^{(\text{ZF})}, \forall k$ are concave functions with respect to $\alpha \in (0,1)$.

For either receiver, a lower bound on the sum rate achieved by the K users is given by

$$R^{(\mathcal{A})}(\alpha, T_d) = \frac{T_d}{T}\sum_{k=1}^K \log_2(1+\mathsf{SNR}_k^{(\mathcal{A})}). \quad (50)$$

Since the function $\log(1+x)$ is concave and nondecreasing, we know that $\log_2(1+\mathsf{SNR}_k^{(\mathcal{A})}), \forall\mathcal{A}$ are concave, implying $R^{(\mathcal{A})}(\alpha, T_d), \forall\mathcal{A}$ are concave with respect to α.

Applying Theorem. 2 and Corollary. 1, we can obtain the optimal solutions of problem (22) for both MRC and ZF receivers under both average and peak power constraints.

V. SE and Training Optimization With Large-Scale Fading

If the adaptive power control is not applied, there exits large-scale fading in the uplink massive MIMO systems. In this section, we will discuss the energy splitting strategy for training optimization in large-scale fading channels under both average and peak power constraints.

Based on the definition of the equivalent channel in Section II-B, we can obtain the equivalent noise $\sigma_v^2 = \sum_{i=1}^K \frac{\rho_d^i p_i}{p_iE+1} + 1$ and effective SNR of the kth user

$$\begin{aligned}\rho_{\text{eff},k} &\triangleq \frac{\rho_d\sigma_{\widehat{\mathbf{G}}_k}^2}{\sigma_v^2} = \frac{\rho_d^k p_k^2 E}{(p_kE+1)(\sum_{i=1}^K\frac{\rho_d^i p_i}{p_iE+1}+1)} \\ &= \frac{p_k^2\rho_d^k\rho_\tau^k T_\tau}{(p_k\rho_\tau^kT_\tau+1)(\sum_{i=1}^K\frac{\rho_d^ip_i}{p_i\rho_\tau^iT_\tau+1}+1)}, \end{aligned} \quad (51)$$

where the energy splitting strategy is

$$\rho_\tau^kT_\tau = \alpha_k\rho T, \quad \rho_d^kT_d = (1-\alpha_k)\rho T, \quad 0 \leq \alpha_k \leq 1, \quad (52)$$

and ρ_d^k (ρ_τ^k) denotes the power that is allocated to the training (data) phase of the kth user, and α_k denotes the fraction of the total transmit energy of the kth user devoted to channel training.

For MRC the following sum SE is achievable for the kth user:

$$R_k^{(\text{MRC})}(\boldsymbol{\alpha}, T_d) \triangleq \big(1 - \tfrac{T_\tau}{T}\big) \log_2 \Big(1 + \mathsf{SNR}_k^{(\text{MRC})}(\boldsymbol{\alpha}, T_d)\Big), \tag{53}$$

where $\boldsymbol{\alpha} \triangleq [\alpha_1, \ldots, \alpha_K]$ and

$$\mathsf{SNR}_k^{(\text{MRC})}(\boldsymbol{\alpha}, T_d) \triangleq \frac{(M-1)\rho_{\text{eff},k}}{\sum_{i=1, i\neq k}^{K} \rho_{\text{eff},i} + 1}. \tag{54}$$

For ZF, assuming $M > K$, the following SE of the kth user is achievable:

$$R_k^{(\text{ZF})}(\boldsymbol{\alpha}, T_d) \triangleq \big(1 - \frac{T_\tau}{T}\big) \log_2 \big(1 + \mathsf{SNR}_k^{(\text{ZF})}(\boldsymbol{\alpha}, T_d)\big), \tag{55}$$

where $\mathsf{SNR}_k^{(\text{ZF})}(\boldsymbol{\alpha}, T_d) \triangleq (M-K)\rho_{\text{eff},k}$.

When the large-scale fading is considered, our optimization problem becomes

$$\underset{\{\alpha_k, T_d, \forall k\}}{\text{maximize}} \ \sum_{k=1}^{K} R_k^{(\mathcal{A})}(\boldsymbol{\alpha}, T_d) \tag{56a}$$

$$\text{subject to } \rho T \alpha_k + \rho_{\max} T_d \leq \rho_{\max} T, \quad \forall k \tag{56b}$$

$$-\rho T \alpha_k - \rho_{\max} T_d \leq -\rho T, \quad \forall k \tag{56c}$$

$$0 \leq \alpha_k \leq 1, \quad \forall k \tag{56d}$$

$$0 < T_d \leq T - K. \tag{56e}$$

A. Training Optimization

It is obvious that the kth user's energy splitting will affect the achievable rate of the others such that the optimization problem becomes more complicated in the sense that i) the objective function is not jointly convex with respect to all the variables, and ii) variable T_d in the constraint is coupled with $\alpha_k, \forall k$, resulting that the traditional block coordinate descent (BCD) algorithm that is widely used for solving nonconvex problems does not work. Here, the alternating direction method of multipliers (ADMM) is applied to solve the nonconvex problem. It has been shown in [36]–[38] that ADMM has the provable convergence guarantees to the stationary (locally optimal) points of the nonconvex problem under some mild assumptions which mainly request that the objective function is smooth and the penalty parameter (i.e., ν_k which will be defined later) is sufficiently large (depend on the Lipschitz constant of the objective function).

First, by introducing auxiliary variables $\beta_k, \gamma_k, \forall k$, we can see that problem (56) is equivalent to

$$\underset{\{\alpha_k, T_d, \beta_k, \gamma_k \forall k\}}{\text{minimize}} \ \sum_{k=1}^{K} -R_k^{(\mathcal{A})}(\boldsymbol{\alpha}, T_d) \tag{57a}$$

$$\text{subject to } T_d + \beta_k = T - \frac{\rho T \alpha_k}{\rho_{\max}}, \quad \forall k \tag{57b}$$

$$T_d = \frac{\rho T - \rho T \alpha_k}{\rho_{\max}} + \gamma_k, \quad \forall k \tag{57c}$$

$$0 \leq \alpha_k \leq 1, \quad \forall k \tag{57d}$$

$$0 < T_d \leq T - K, \tag{57e}$$

$$\beta_k, \gamma_k \geq 0 \quad \forall k. \tag{57f}$$

To this end, let us construct the augmented Lagrangian as the following

$$\begin{aligned} &\mathcal{L}(\{\alpha_k\}, \{\beta_k\}, \{\gamma_k\}, T_d; \{\lambda_k\}, \{\mu_k\}) \\ &= \sum_{k=1}^{K} \Bigg(- R_k^{(\mathcal{A})}(\boldsymbol{\alpha}, T_d) + \frac{\nu_k}{2}\Big(T_d - \big(T - \frac{\rho T \alpha_k}{\rho_{\max}} - \beta_k\big) + \frac{\lambda_k}{\nu_k}\Big)^2 \\ &\quad + \frac{\nu_k}{2}\Big(T_d - \big(\frac{\rho T - \rho T \alpha_k}{\rho_{\max}} + \gamma_k\big) + \frac{\mu_k}{\nu_k}\Big)^2 \Bigg), \end{aligned} \tag{58}$$

where $\lambda_k, \mu_k, \forall k$ denote the dual variables associated with equalities (57b) and (57c), and $\nu_k, \forall k$ represent the penalty parameters.

Using ADMM [39], we can obtain the update rules of the primal variables as follows (superscript t denotes the number of iterations):

1) Update of $\alpha_k, \forall k$:

$$\begin{aligned} \alpha_k^{(t+1)} = \arg \min_{0 \leq \alpha_k \leq 1} \Big(&\sum_{k=1}^{K} -R_k^{(\mathcal{A})}(\alpha_k, T_d)\Big) \\ &+ \frac{\nu_k}{2}\Big(T_d - \big(T - \frac{\rho T \alpha_k}{\rho_{\max}} - \beta_k\big) + \frac{\lambda_k}{\nu_k}\Big)^2 \\ &+ \frac{\nu_k}{2}\Big(T_d - \big(\frac{\rho T - \rho T \alpha_k}{\rho_{\max}} + \gamma_k\big) + \frac{\mu_k}{\nu_k}\Big)^2. \end{aligned} \tag{59}$$

2) Update of T_d:

$$\begin{aligned} T_d^{(t+1)} = \arg \min_{0 \leq T_d \leq T-K} \sum_{k=1}^{K} \Bigg(&- R_k^{(\mathcal{A})}(\alpha_k, T_d) \\ &+ \frac{\nu_k}{2}\Big(T_d - \big(T - \frac{\rho T \alpha_k}{\rho_{\max}} - \beta_k\big) + \frac{\lambda_k}{\nu_k}\Big)^2 \\ &+ \frac{\nu_k}{2}\Big(T_d - \big(\frac{\rho T - \rho T \alpha_k}{\rho_{\max}} + \gamma_k\big) + \frac{\mu_k}{\nu_k}\Big)^2 \Bigg). \end{aligned} \tag{60}$$

3) Update of $\beta_k, \forall k$:

$$\begin{aligned} \beta_k^{(t+1)} &= \arg \min_{\beta_k \geq 0} \Big(T_d - \big(T - \frac{\rho T \alpha_k}{\rho_{\max}} - \beta_k\big) + \frac{\lambda_k}{\nu_k}\Big)^2 \\ &= \max\{0, T - T_d - \frac{\rho T \alpha_k}{\rho_{\max}} - \frac{\lambda_k}{\nu_k}\}. \end{aligned} \tag{61}$$

4) Update of $\gamma_k, \forall k$:

$$\gamma_k^{(t+1)} = \arg\min_{\gamma_k \geq 0} \big(T_d - (\frac{\rho T - \rho T \alpha_k}{\rho_{\max}} + \gamma_k) + \frac{\mu_k}{\nu_k}\big)^2 = \max\big\{0, T_d - \frac{\rho T - \rho T \alpha_k}{\rho_{\max}} + \frac{\mu_k}{\nu_k}\big\}. \quad (62)$$

Also, we have the update rules of the dual variables are as follows:

1) Update of dual variables $\lambda_k, \forall k$:

$$\lambda_k^{(t+1)} = \lambda_k^{(t)} + \nu_k \big(T_d^{(t+1)} - (T - \frac{\rho T \alpha_k^{(t+1)}}{\rho_{\max}} - \beta_k^{(t+1)})\big). \quad (63)$$

2) Update of dual variables $\mu^k, \forall k$:

$$\mu_k^{(t+1)} = \mu_k^{(t)} + \nu_k \big(T_d^{(t+1)} - (\frac{\rho T - \rho T \alpha_k^{(t+1)}}{\rho_{\max}} + \gamma_k^{(t+1)})\big). \quad (64)$$

The objective function is Lipschitz continuous and the feasible set of the problem is convex. Applying the theorem in [38, Th. 1], when $\nu_k, \forall k$ are chosen large enough (such that they are larger than the required lower bound of these penalty parameters; see [38, Lemma 9] or [36, Assumption A]), it is guaranteed that the ADMM algorithm can converge to the stationary points of problem (57).

B. Solutions of the Subproblems

The solutions of solving sub-problems (59) and (60) are discussed in this section.

1) ZF Receiver:

- Update of $\alpha_k, \forall k$:
  Lemma 3: For the kth user, when $\alpha_i, \forall i \neq k$ and T_d are fixed, function $\sum_{k=1}^K R_k^{\text{ZF}}(\alpha_k)$ is concave.
  Proof: Please see Appendix C. □
- Update of T_d:
  When $\alpha_k, \forall k$ are fixed, sum rate $\sum_{k=1}^K R_k^{\text{ZF}}(T_d)$ is

$$\sum_{k=1}^K \frac{T_d}{T} \log\big(1 + \frac{(M-K)p_k^2 \rho^2 T^2 (1-\alpha_k)\alpha_k}{(p_k \rho T \alpha_k + 1)(\sum_{i=1}^K \frac{p_i \rho T (1-\alpha_i)}{p_i \rho T \alpha_i + 1} + T_d)}\big).$$

  Applying Lemma 2, we can conclude that $\sum_{k=1}^K R_k^{\text{ZF}}(T_d)$ is also concave with respect to T_d.
  After studying the convexity properties of these objective functions, we know that the each subproblem of updating variables for the ZF receiver is convex. Then, we can take the gradient of the objective function and set it as zero, where the root of the resulting equation is the optimal solution of the subproblem.

2) MRC Receiver:

- Update of α_k:
  When $\alpha_i, i \neq k$ and T_d are fixed, maximizing $\sum_{k=1}^K R_k^{\text{MRC}}(\alpha_k)$ with respect to α_k is not a convex problem, but the optimal solution can be still easily obtained since solving this problem only involves one dimensional search.
- Update of T_d:
  When $\alpha_k, \forall k$ are fixed, we have

$$\begin{aligned}
&\mathsf{SNR}_k^{(\text{MRC})}(T_d) \\
&= \frac{\frac{(M-1)p_k^2 \rho^2 \ T^2 (1-\alpha_k)\alpha_k}{(p_k \rho T \alpha_k + 1)(\sum_{j=1}^K \frac{p_j \rho T (1-\alpha_k)}{p_j \rho T \alpha_j + 1} + T_d)}}{\sum_{i=1, i\neq k}^K \frac{p_i^2 \rho^2 \ T^2 (1-\alpha_i)\alpha_i}{(p_i \rho T \alpha_i + 1)(\sum_{j=1}^K \frac{p_j \rho T (1-\alpha_j)}{p_j \rho T + 1}) + T_d} + 1} \\
&= \frac{\frac{(M-1)p_k^2 \rho^2 T^2 (1-\alpha_k)\alpha_k}{p_k \rho T \alpha_k + 1}}{\sum_{i=1, i\neq k}^K \frac{p_i^2 \rho^2 \ T^2 (1-\alpha_i)\alpha_i}{p_i \rho T \alpha_i + 1} + \sum_{j=1}^K \frac{p_j \rho T (1-\alpha_j)}{p_j \rho T \alpha_j + 1} + T_d}.
\end{aligned}$$

  Applying Lemma 2, we know that $\sum_{k=1}^K R_k^{\text{MRC}}(T_d)$ is concave with respect to T_d, and can be also solved easily.

Algorithm 1 The Training Optimization Algorithm With Both Average and Peak Power Constraints

```
Input: α_k = K/T, ν_k, ∀k and T_d = K.
for t = 1, ..., N do
    for k = 1, ..., K do
        Update primal variables α_k, T_d, β_k, γ_k by (59)–(62).
    end for
    for k = 1, ..., K do
        Update dual variables λ_k, μ_k by (63) and (64).
    end for
end for
```

C. Algorithm Description

By leveraging ADMM, the developed algorithm that splits energy between the training and data phases under both average and peak power constraints is summarized in Algorithm 1, where N denotes the total number of iterations.

When there is no peak power, problem (56) is reduced to

$$\begin{aligned}
&\underset{\{\alpha_k, \forall k\}}{\text{maximize}} \ \sum_{k=1}^K R_k^{(\mathcal{A})}(\boldsymbol{\alpha}) \\
&\text{subject to } 0 \leq \alpha_k \leq 1, \quad \forall k.
\end{aligned} \quad (65)$$

This problem is a special case of (56), where there is no variable coupling in the constraint. Hence, we can simply use the BCD algorithm to solve problem (65). Applying the proposition in [40, Proposition 2.7.1], it is guaranteed that the BCD algorithm converges to the stationary points of problem (65). The algorithm of energy splitting with only the average power constraint is summarized in Algorithm 2.

Algorithm 2 The Training Optimization Algorithm With the Average Power Constraint

```
Input: α_k = K/T, ν_k, ∀k and T_d = K.
for t = 1, ..., N do
    for k = 1, ..., K do
        Update α_k by arg max_{0≤α_k≤1} Σ_{k=1}^K R_k^(A)(α_k)
    end for
end for
```

Remark 8: In this case, the recent work in [23, Th. 6] shows that problem (65) can be reformulated into another convex

optimization problem equivalently. Unfortunately, the objective function of the reformulated problem can be infinite in some special case (e.g., when $s = 0$ where s is defined in [23, Th. 6]), resulting that the reformulated problem is not equivalent to problem (65).

D. Complexity of Implementing the Algorithms

The two proposed algorithms are computationally efficient in the sense that each subproblem only involves a one-dimensional optimization problem. Especially, when the subproblem is convex, a simple bisection algorithm can be exploited, which can achieve a small error ϵ in several steps [35]. Therefore, the total complexities of the two algorithms are $\mathcal{O}(NKI)$ where I denotes the number of iteration used in the inner loop of each subproblem. When the subproblem is convex, I is $\mathcal{O}(\log(\frac{1}{\epsilon}))$ and when the subproblem is nonconvex, I is $\mathcal{O}(\frac{1}{\epsilon})$.

E. Sum SE in a Multi-Cell System

In a multi-cell massive MIMO system, the received signal at the base station will be also affected by interference from the neighboring cells. Consider the uplink of the system with L neighboring cells that share the same frequency resource. There is one base station equipped with M antennas in each cell and K users each equipped with one antenna. Let $\mathbf{G}_{ln} = \mathbf{H}_{ln}\mathbf{P}_{ln}^{\frac{1}{2}}$ be the channel matrix between the lth base station and the K users in the nth cell, where $\mathbf{H}_{ln}$ denotes the small-scale fading and the diagonal matrix $\mathbf{P}_{ln} \triangleq \text{diag}\{p_{ln1}, \ldots, p_{lnk}, \ldots, p_{lnK}\}$ represents the large-scale fading, i.e., path loss and shadowing fading. Then, the received signal $\mathbf{Y}_r \in \mathbb{C}^{M\times T_\tau}$ during the training phase can be written as

$$\mathbf{Y}_r = \sqrt{E}\mathbf{G}_{ll} + \underbrace{\sqrt{E}\sum_{i\neq l}^{L}\mathbf{G}_{li}}_{\text{inter-cell interference}} + \mathbf{N}, \tag{66}$$

where $\mathbf{N} \in \mathbb{C}^{M\times T_\tau}$ is the additive noise. Let $\mathbf{h}_{lik}$ and $\mathbf{n}_{llk}$ denote the kth column of $\mathbf{H}_{li}$ and $\mathbf{N}_{ll}$, respectively. The (linear) MMSE estimate for the channel is given by $\widehat{\mathbf{G}}_{ll} = \sqrt{E}\mathbf{Y}_r\mathbf{P}_{ll}(\sum_{i=1}^{L}\mathbf{P}_{li}E + \mathbf{I})^{-1}$, where the kth column of $\widehat{\mathbf{G}}_{ll}$ is

$$\widehat{\mathbf{G}}_{llk} = \frac{p_{llk}^{\frac{3}{2}}E}{\sum_{i=1}^{L}p_{lik}E+1}\mathbf{h}_{llk} + \sum_{i\neq l}^{L}\frac{p_{llk}Ep_{lik}^{\frac{1}{2}}}{\sum_{j=1}^{L}p_{ljk}E+1}\mathbf{h}_{lik} + \frac{p_{llk}\sqrt{E}}{\sum_{i=1}^{L}p_{lik}E+1}\mathbf{n}_{llk}. \tag{67}$$

Let the channel estimation error be $\widetilde{\mathbf{G}}_{ll} = \mathbf{G}_{ll} - \widehat{\mathbf{G}}_{ll}$. Thus, we have the kth column of $\widetilde{\mathbf{G}}_{llk}$, i.e., $\widetilde{\mathbf{G}}_{llk} = \mathbf{G}_{llk} - \widehat{\mathbf{G}}_{llk} = \frac{1}{\sum_{i=1}^{L}p_{lik}E+1}\big(\sqrt{p_{llk}}(\sum_{i\neq l}^{L}Ep_{lik}+1)\mathbf{h}_{llk} - p_{llk}\sum_{i\neq l}^{L}Ep_{lik}^{\frac{1}{2}}\mathbf{h}_{lik} - p_k\sqrt{E}\mathbf{n}_{llk}\big)$, where $\mathbf{G}_{llk}$ denotes the kth column of $\mathbf{G}_{ll}$. It is also easy to verify that the elements of $\widehat{\mathbf{G}}_{ll}$ are column-wise *i.i.d.* complex Gaussian with zero mean and variance $\sigma^2_{\widehat{\mathbf{G}}_{llk}} = \frac{p_{llk}^2E}{\sum_{i=1}^{L}p_{lik}E+1}$, and the elements of $\widetilde{\mathbf{G}}_{llk}$ are column-wise *i.i.d.* complex Gaussian with zero mean and variance $\sigma^2_{\widetilde{\mathbf{G}}_{llk}} = \frac{p_{llk}(\sum_{i\neq l}^{L}p_{lik}E+1)}{\sum_{i=1}^{L}p_{lik}E+1}$. Following the similar derivation steps shown in Section II-B, we can obtain the equivalent SNR of the kth user in the lth cell, i.e.,

$$\rho_{\text{eff},lk} \triangleq \frac{\frac{\rho_d p_{llk}^2E}{\sum_{i=1}^{L}p_{lik}E+1}}{\sum_{n=1}^{L}\sum_{j=1}^{K}\frac{\rho_d p_{lnj}(\sum_{i\neq n}^{L}p_{lij}E+1)}{\sum_{i=1}^{L}p_{lij}E+1}+1},$$

and consequently

$$\begin{aligned}\text{SNR}_{lk}^{(\text{ZF})} &= (M-K)\rho_{\text{eff},lk},\\ \text{SNR}_{lk}^{(\text{MRC})} &= \frac{(M-1)\rho_d E p_{llk}^2}{m_k + \sum_{n=1}^{L}\sum_{i=1}^{K}\rho_d p_{lni} + \sum_{n=1}^{L}p_{lnk}E+1},\end{aligned} \tag{68}$$

where $m_k \triangleq (M-1)\sum_{i\neq l}^{L}p_{lik}^2\rho_d E - \sum_{i=1}^{L}p_{lik}^2\rho_d E + (\sum_{i=1}^{L}\sum_{j=1}^{K}\rho_d p_{lij})\sum_{i=1}^{L}p_{lik}E$, which have been also shown in [12] and [41]. Using (53) and (55), we can get the objective function of (57). Applying the energy splitting

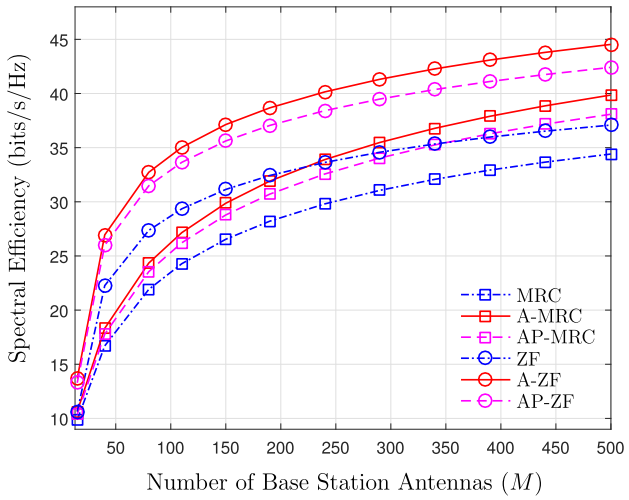


Fig. 2. Comparison between equal and optimized power allocations when the number of base station antennas increases; $\rho_u = 3$dB.

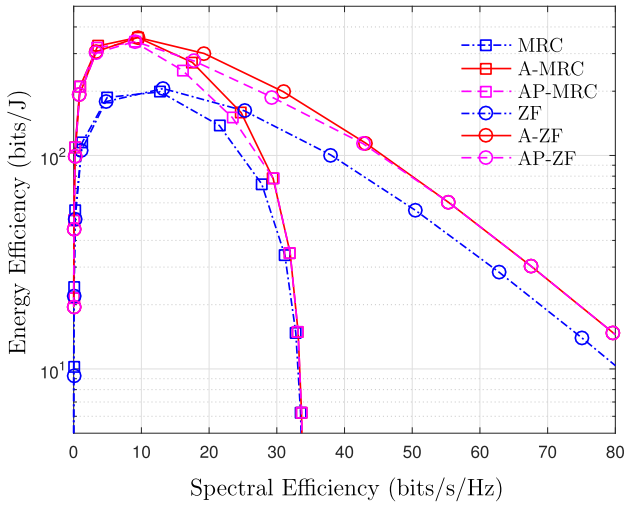


Fig. 3. Comparison of EE versus SE, where $M = 100$.

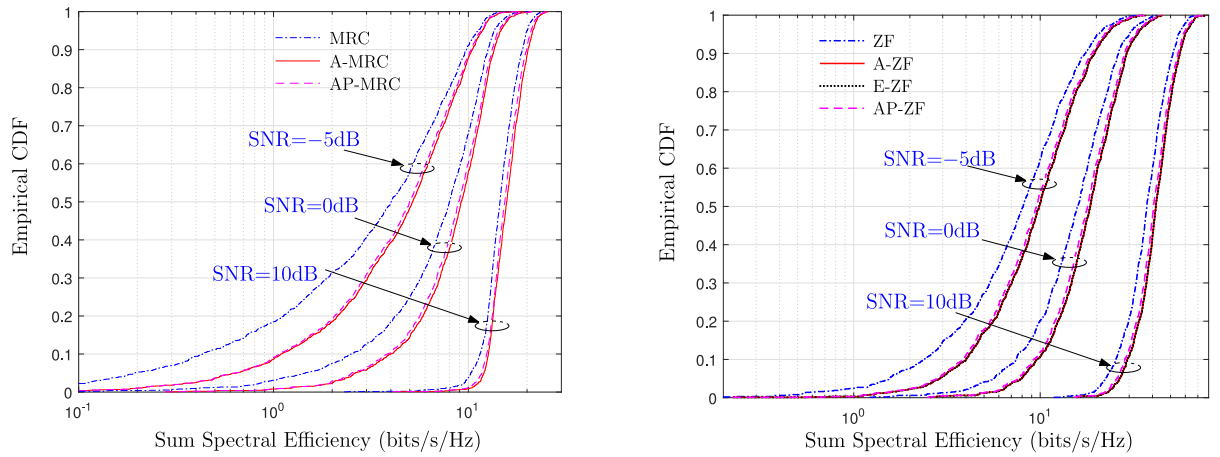


Fig. 4. CDF of the sum SE in large-scale fading channels, where $M = 100$, $K = 10$, $T = 200$. (a) MRC receiver and (b) ZF receiver.

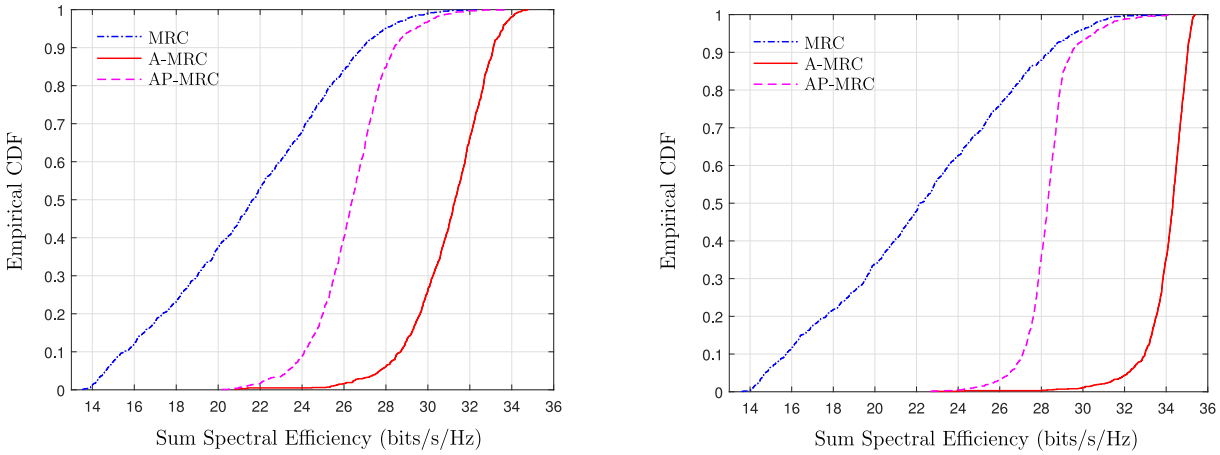


Fig. 5. CDF of the sum SE with the MRC receiver, where $M = 100$, $K = 10$, $T = 200$. (a) SNR at the cell edge is -5dB and (b) SNR at the cell edge is 5dB.

scheme shown in (52), we can exploit Algorithm 1 to solve problem (57) in the large-scale fading case.

VI. Numerical Results

In this section, we compare the SE between the equal power allocation scheme and our optimized one under average and peak power constraints. We consider the following schemes: 1) MRC, which refers to the case where the MRC receiver is used and the same average power is used in both training and data transmission phases [12]. 2) A-MRC, which refers to the case where the MRC receiver is used, the training duration is K, and there is only the average power constraint. 3) AP-MRC, where the MRC receiver is used, and both the training duration and training energy are optimized under both the average and peak power constraints. We will also consider the ZF variants of the above three cases, namely ZF, A-ZF, and AP-ZF. The EE is defined as $\eta^{(\mathcal{A})} \triangleq R^{(\mathcal{A})}(\alpha, T_d)/\rho$, $\mathcal{A} \in \{\text{MRC}, \text{ZF}\}$.

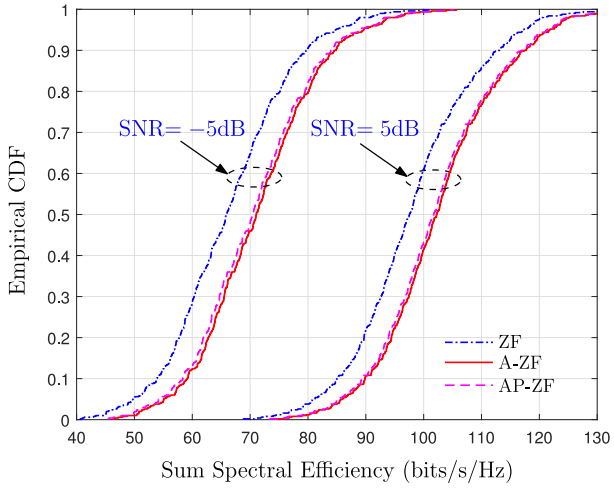


Fig. 6. CDF of the sum SE with the ZF receiver, where $M = 100$, $K = 10$, $T = 200$, and the SNR marker denotes the received SNR at the cell edge.

A. EE and Sum SE With Small-Scale Fading

In our simulations, we set $\rho_{\max} = 2\rho$, $K = 10$, and $T = 200$. In Fig. 2, we show the sum SE of various schemes as the number of antennas increases for $\rho = \rho_u/\sqrt{M}$. It can be seen that SE per user by A-MRC (ZF) is 1.5–4.5 bits/s/Hz and only 1–3.5 bits/s/Hz by MRC (ZF), illustrating that A-MRC (ZF)

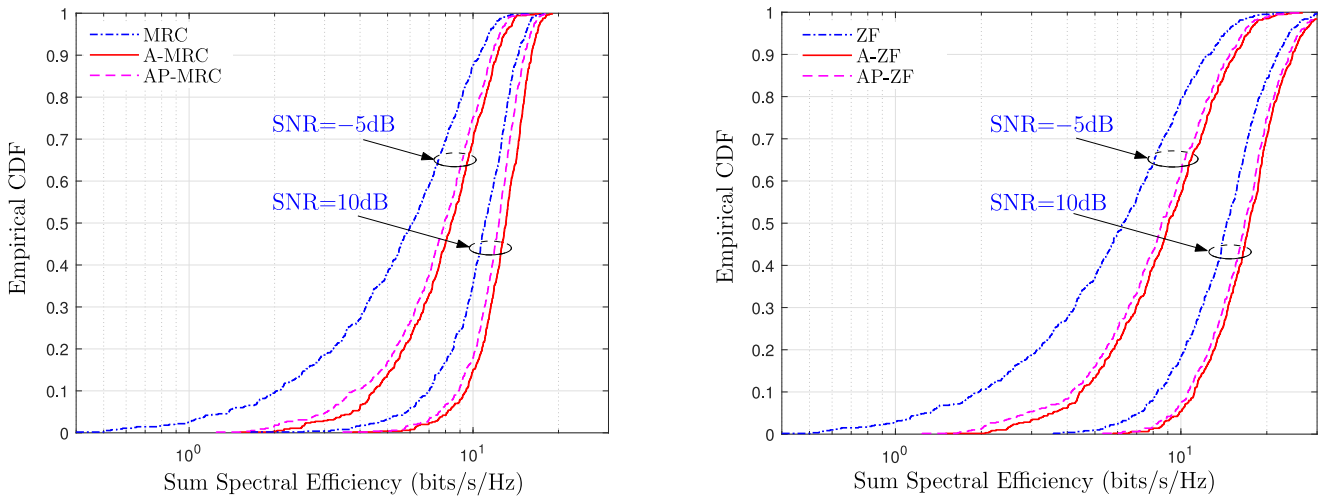


Fig. 7. CDF of the sum SE in the multi-cell massive MIMO system, where $M = 100$, $K = 4$, $T = 200$. (a) MRC receiver and (b) ZF receiver.

performs better than MRC (ZF) as well as AP-MRC (ZF). In Fig. 3, we show EE versus SE. It can also be seen that optimized schemes, e.g., A-MRC (A-ZF) and AP-MRC (AP-ZF), show a significant gain compared with MRC (ZF). In particular, there is an optimal average transmitted power for maximum EE as has been also observed before in [12]. Also, from the simulations, we can see that ZF performs better than MRC at the high SNR region, but worse when the SNR is low.

Moreover, the impact of peak power constraint on SE and EE for both MRC and ZF receivers can be observed through from Fig. 2 to Fig. 3 clearly. They illustrate that when peak power is limited at the training phase, the SE with AP-MRC and AP-ZF cannot be as high as the case with A-MRC and ZF. Although the training period is increased, the time slot is still very precious when the achievable rate needs to be maximized.

B. EE and Sum SE With Large-Scale Fading

We assume that the users are randomly and independently distributed in a single cell with radius $R = 1000$m, where the location of each user follows the uniform distribution and the minimum distance between any user and the base station is 100m. The large-scale fading is modeled as $p_k = z/r_k^3, \forall k$ where r_k is the distance between the kth user and the base station, and z follows a log-normal distribution with zero mean and 8dB standard deviation, representing the shadowing effect. We allocate the energy budgets as $E = \rho T = 10^{-0.5} \times R^3 \times T$, meaning that the SNR at the cell edge is -5dB when equal power allocation is used.

We consider the energy splitting in the following two cases. The first one is maximizing SE with respect to all users (solving problem (22)) and the second one is maximizing SE in terms of each user (solving problem (57)).

1) Joint Power and Training Duration Optimization Over α, T_d: The empirical cumulative distribution function (CDF) of the sum SE over different snapshots of user locations is shown in Fig. 4 with both ZF and MRC receivers. The numerical results are based on 1000 Monte Carlo (MC) trials. We set $\rho_{\max} = 3\rho$. E-ZF refers to the scheme where the exhaustive search of α is used when (48) is adopted for the ZF receiver. In the A-ZF scheme, the lower bound of (48) is used, i.e., (49), in the objective function. It can be observed that the results obtained by E-ZF and A-ZF are very close, illustrating the relaxation is reasonable. Under the different SNRs, it is also shown that when SNR is low, the advantages of using the energy splitting scheme become more obvious, compared with the case without implementing power optimization. Comparing with (a) and (b) in Fig. 4, it can be seen that the improvement obtained by optimizing the sum SE in the MRC case is larger than the ZF case.

2) Joint Power and Training Duration Optimization Over $\alpha_k, \forall k, T_d$: When energy splitting is considered for each user, the achievable SE can be higher than the case where the same energy splitting is used for all users. For AP-MRC and AP-ZF, we set $\nu_k = 10^{-3}, \forall k$ and $\rho_{\max} = 3\rho$. The numerical results are based on 1000 MC trials. The empirical CDF of the sum SE over different snapshots of user locations is shown in Fig. 6 with the ZF receiver and Fig. 5 with the MRC receiver. It can be observed that the optimized power allocation strategy increases the sum SE compared with the equal power allocation scheme. When the peak power constraint is considered, the achievable sum SE is lower than the case where only the average power constraint is applied, illustrating that the results obtained by the A-MRC (A-ZF) method are too optimistic in real applications. AP-MRC (AP-ZF) allows more time slots in the training phase such that SINR can be improved in the data transmission phase, which is the practical strategy for training optimization of the uplink massive MIMO systems. Also, we can see that the peak power constraint affects the sum SE in the case of large-scale fading more significant than the case where there is only small-scale fading, especially for the case shown in Fig. 5 where the MRC receiver is used.

C. Sum SE in Multi-Cell System

Numerical results of maximizing the sum rate in a multi-cell system are depicted in Fig. 7. The system has 7 equal-size hexagonal cells with one at the center and the remaining 6 surrounding it. We focus on analyzing the sum SE of the hexagonal cell in the center. The radius of each cell is 1000m, the path loss model and the distribution of the users are the same as the previous section. Equal power allocation is used in the 6 neighboring cells, and the cases of using MRC and ZF receivers are both considered. It can be observed that the optimized sum SE is higher than the case with equal power allocation. Also, the improvement of the sum SE in the multi-cell system is less than the case of the single-cell system. The reason is that the power allocation scheme cannot deal with the inter-cell interference by optimizing the intra-cell users' power and training duration unless other advanced techniques are adopted, e.g., coordinated multi-point (CoMP) or joint interference management.

VII. Conclusions

In this paper, we considered an uplink multiuser massive MIMO. The channels were assumed to be estimated through training symbols transmitted by the mobile users. We were able to quantify the amount of energy savings that are possible through the increase of either the number of base station antennas, or the coherence interval length. We also quantified the degrees of freedom that is possible in this scenario, which is the same as that of a point-to-point MIMO system. The scheme that achieves the DoF of the system when the number of users is less than the number of base station antennas is linear: zero-forcing is sufficient.

For the case where both average and peak power constraints were considered, we considered the problem of joint training energy and training duration optimization for the MRC and ZF receivers so that the sum SE was maximized. In the small-scale fading channels, we also performed a careful analysis of the convexity of the problem and derived optimal solutions either in closed forms or in one case through a one-dimensional search of a quasi-concave function. For the case where there was large-scale fading, we developed an iterative algorithm that leveraged ADMM and obtained a locally optimal solution. Our results were illustrated and verified through multiple numerical examples.

Appendix

A. Proof of Theorem. 2

Proof: After studying the convexity of the objective function, there are only three possible cases for the optimal solutions, as we discuss below.

1) Case 1 (ρ_τ is Limited by ρ_{max}): Define $\alpha_1 \triangleq \rho_{\max} K/\rho T$, which is the root of $T - K = -\rho T\alpha/\rho_{\max} + T$ in α (see Fig. 1). In the case where $\alpha_1 < \alpha^\dagger$, because of property P2 the optimal (α^*, T_d^*) must be on one of the two lines given by i) $T_d = -\rho T\alpha/\rho_{\max} + T$, $\alpha \in [\alpha_1, 1]$, and ii) $T_d = T - K$, $\alpha \in [0, \alpha_1]$.

On the line $T_d = T - K, \alpha \in [0, \alpha_1]$ the objective function is concave and increasing with α, thanks to property P1. Hence, we only need to consider the line $T_d = -\rho T\alpha/\rho_{\max} + T$, $\alpha \in [\alpha_1, 1]$.

Lemma 4: The objective function $R^{(\mathrm{MRC})}(\alpha, T_d)$ along the line $T_d = -\rho T\alpha/\rho_{\max} + T, \alpha \in [\alpha_1, 1]$ is quasiconcave in α.

Proof: See Appendix B. □

Thanks to Lemma 4, we can find the optimal α by setting the derivative of (69) with respect to α to 0. Efficient one-dimensional searching algorithms such as Newton method or bisection algorithm [35], can be adopted to find out the optimal α.

2) Case 2 (ρ_d is Limited by ρ_{max}): Define $\alpha_2 \triangleq 1 - \rho_{\max}(T - K)/\rho T$, which is the root of $T - K = \rho T\alpha/\rho_{\max} + \rho T/\rho_{\max}$ in α. If $\alpha_2 > \alpha^\dagger$, because of property P2 the optimal (α^*, T_d^*) must be on one of the two lines given by i) $T_d = -\rho T\alpha/\rho_{\max} + T, \alpha \in (\alpha_1, 1)$, $\alpha \in [\alpha_1, 1]$, and ii) $T_d = T - K$, $\alpha \in [\alpha_2, \alpha_1]$. Along the line $T_d = T - K, \alpha \in (\alpha_1, 1)$, the corresponding function is decreasing in α because of property P1. Also considering P2, which implies that the optimal point in this case cannot include $T_d < T - K$, we conclude that the point $(\alpha^*, T_d^*) = (\alpha_2, T - K)$ is the global optimal solution of the problem.

3) Case 3 (Both ρ_d and ρ_τ Are Not Limited by ρ_{max}): If $\alpha_2 < \alpha^\dagger < \alpha_1$, the optimal point is achieved at $(\alpha^*, T_d^*) = (\alpha^\dagger, T - K)$, according to properties P1 and P2. □

B. Proof of Lemma 4

Proof: Consider MRC processing. Substituting (37) into $R^{(\mathrm{MRC})}(\alpha, T_d)$, we have

$$R^{(\mathrm{MRC})}(\alpha) = \frac{K}{T}\left(-\frac{\rho T}{\rho_{\max}}\alpha + T\right)\log_2(1 + \mathsf{SNR}^{(\mathrm{MRC})}(\alpha)), \tag{69}$$

where

$$\mathsf{SNR}^{(\mathrm{MRC})}(\alpha) = \frac{\alpha(\alpha - 1)\rho^2 T^2 (M - 1)}{a_2\alpha^2 - b_2\alpha - c_2}, \tag{70}$$

and $a_2 = \rho^2 T^2(K - 1) + \rho^2 T^2/\rho_{\max}$, $b_2 = \rho^2 T^2(K - 1) + \rho T^2 - \rho TK - \rho T/\rho_{\max}$ and $c_2 = K\rho T + T$. Since $R^{(\mathrm{MRC})}(\alpha) > 0$, in order to prove the quasi-concavity of $R^{(\mathrm{MRC})}(\alpha)$, we need to prove that the super-level set $\mathcal{S}_\beta = \{\alpha | 0 < \alpha < 1, R^{(\mathrm{MRC})}(\alpha) \geq \beta\}$ for each $\beta \in \mathbb{R}^+$ is convex. Equivalently, if we define

$$\phi_\beta(\alpha) = \frac{\beta}{\frac{K}{T}(\frac{\rho T\alpha}{\rho_{\max}} - T)} + \log_2(1 + \mathsf{SNR}^{(\mathrm{MRC})}(\alpha)). \tag{71}$$

we only need to prove that $\mathcal{S}_\phi = \{\alpha | 0 < \alpha < 1, \phi_\beta(\alpha) \geq 0\}$ is a convex set.

It can be checked that the first part of $\phi_\beta(\alpha)$, namely $\beta/[\frac{K}{T}(\frac{\rho T\alpha}{\rho_{\max}} - T)]$, is concave for $\alpha \in [0, 1]$. For the other part of $\phi_\beta(\alpha)$, from (70) we know that

$$a_2 - b_2 - c_2 = \rho T(\frac{\rho}{\rho_{\max}} - 1) - T(1 - \alpha\frac{\rho}{\rho_{\max}}) < 0, \tag{72}$$

where $a_2, c_2 > 0$. Applying Lemma 1, we know $\mathsf{SNR}^{(\mathrm{MRC})}(\alpha)$ is concave. Hence, $\log_2(1 + \mathsf{SNR}^{(\mathrm{MRC})}(\alpha))$ is also concave

since function $\log(1+x)$ is concave and non-decreasing [35]. Therefore, its super-level set $\mathcal{S}_\phi$ is convex. It follows that the super-level set $\mathcal{S}_\beta$ of $R^{(\mathrm{MRC})}(\alpha)$ is convex for each $\beta \geq 0$. The objective function is thus quasiconcave. □

C. Proof of Lemma 3

Before proving Lemma 3, we first need the following two lemmas. The proofs are elementary, and not provided here due to space limit.

Lemma 5: The function $f(x) \triangleq \log\left(1+\frac{(1-x)x}{ax+b}\right)$ is concave in x over $(0,1)$ where $a,b>0$ are some constants.

Lemma 6: The function $g(x) \triangleq \log\left(1+\frac{a}{\frac{c}{bx+1}+d}\right)$ is concave in x over $(0,1)$ when $a,b,c,d>0$ are some constants.

The proof of Lemma 3 follows. *Proof:* Substituting (52) into (5), we have

$$\begin{aligned}\rho_{\mathrm{eff},k} &= \frac{p_k^2\rho_d^k\rho_\tau^k T_\tau}{(p_k\rho_\tau^k T_\tau+1)(\sum_{i=1}^K \frac{\rho_d^i p_i}{p_i\rho_\tau^i T_\tau+1}+1)} \\ &= \frac{p_k^2\rho^2T^2(1-\alpha^k)\alpha_k}{\underbrace{(T_d-1+\widetilde{c})p_k\rho T}_{\triangleq a>0}\alpha_k+\underbrace{\rho T p_k+\widetilde{c}+T_d}_{\triangleq b>0}}, \end{aligned} \tag{73}$$

where $\widetilde{c}=\sum_{i=1,i\neq k}^K(1-\alpha_i)\rho T p_i$.

Applying Lemma 5, we know that $R_k^{\mathrm{ZF}}(\alpha_k)$ is concave.

The power of the kth user also generates the interference to the ith user such that the effective SNR at the ith user is

$$\rho_{\mathrm{eff},i}=\frac{p_i^2\rho_d^i\rho_\tau^i T_\tau}{(p_i\rho_\tau^i T_\tau+1)}\frac{T_d}{s},\quad i\neq k$$

where

$$\begin{aligned} s &= T_d\sum_{i=1}^K\frac{\rho_d^i p_i}{p_i\rho_\tau^i T_\tau+1} \\ &= T_d\sum_{i=1,i\neq k}^K\frac{\rho_d^i p_i}{p_i\rho_\tau^i T_\tau+1}+\frac{\rho_d^k p_k}{p_k\rho_\tau^k T_\tau+1}+1 \\ &= \sum_{i=1,i\neq k}^K\frac{T_d\rho_d^i p_i}{p_i\rho_\tau^i T_\tau+1}+\frac{(1-\alpha_k)\rho T p_k}{p_k\rho T\alpha_k+1}+T_d \\ &= \frac{c_k}{b_k\alpha_k+1}+\underbrace{\sum_{i=1,i\neq k}^K\frac{T_d\rho_d^i p_i}{p_i\rho_\tau^i T_\tau+1}+T_d-1}_{\triangleq d>0}, \end{aligned} \tag{74}$$

and $c_k\triangleq 1+\rho T p_k>0, b_k\triangleq p_k\rho T>0$.

Applying Lemma 6, we know that $R_i^{\mathrm{ZF}}(\alpha_k), i\neq k, \forall i$ are concave. Therefore, we can conclude that $\sum_{k=1}^K R_k^{\mathrm{ZF}}(\alpha_k)$ is concave. □

REFERENCES

[1] S. Lu and Z. Wang, "Achievable rates of uplink multiuser massive MIMO systems with estimated channels," in *Proc. Global Commun. Conf. (GLOBECOM)*, Dec. 2014, pp. 3772–3777.

[2] S. Lu and Z. Wang, "Joint optimization of power allocation and training duration for uplink multiuser MIMO communications," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, New Orleans, LA, USA, Mar. 2015, pp. 322–327.

[3] S. Noh, M. D. Zoltowski, Y. Sung, and D. J. Love, "Pilot beam pattern design for channel estimation in massive MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 787–801, Oct. 2014.

[4] M. Gkizeli and G. N. Karystinos, "Maximum-SNR antenna selection among a large number of transmit antennas," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 891–901, Oct. 2014.

[5] S. K. Mohammed and E. G. Larsson, "Per-antenna constant envelope precoding for large multi-user MIMO systems," *EEE Trans. Commun.*, vol. 61, no. 3, pp. 1059–1071, Mar. 2013.

[6] B. Hassibi, M. Hansen, A. Dimakis, H. Alshamary, and W. Xu, "Optimized Markov chain Monte Carlo for signal detection in MIMO systems: An analysis of the stationary distribution and mixing time," *IEEE Trans. Signal Process.*, vol. 62, no. 17, pp. 4436–4450, Sep. 2014.

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

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

[9] A. L. Swindlehurst, E. Ayanoglu, P. Heydari, and F. Capolino, "Millimeter-wave massive MIMO: The next wireless revolution?" *IEEE Commun. Mag.*, vol. 52, no. 9, pp. 56–62, Sep. 2014.

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

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

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

[13] H. Yang and T. L. Marzetta, "Performance of conjugate and zero-forcing beamforming in large-scale antenna systems," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 172–179, Feb. 2013.

[14] Q. Zhang, S. Jin, K.-K. Wong, H. Zhu, and M. Matthaiou, "Power scaling of uplink massive MIMO systems with arbitrary-rank channel means," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 966–981, Oct. 2014.

[15] E. Björnson, M. Matthaiou, and M. Debbah, "Massive MIMO with non-ideal arbitrary arrays: Hardware scaling laws and circuit-aware design," *IEEE Trans. Wireless Commun.*, vol. 14, no. 8, pp. 4353–4368, Aug. 2015.

[16] H. Q. Ngo, M. Matthaiou, and E. G. Larsson, "Massive MIMO with optimal power and training duration allocation," *IEEE Wireless Commun. Lett.*, vol. 3, no. 6, pp. 605–608, Dec. 2014.

[17] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 951–963, Apr. 2003.

[18] L. Zheng and D. N. C. Tse, "Communicating on the Grassmann manifold: A geometric approach to the non-coherent multiple antenna channel," *IEEE Trans. Inf. Theory*, vol. 48, no. 2, pp. 359–383, Feb. 2002.

[19] G. Miao, "Energy-efficient uplink multi-user MIMO," *IEEE Trans. Wireless Commun.*, vol. 12, no. 5, pp. 2302–2313, May 2013.

[20] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "The multicell multiuser MIMO uplink with very large antenna arrays and a finite-dimensional channel," *IEEE Trans. Commun.*, vol. 61, no. 6, pp. 2350–2361, Jun. 2013.

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

[22] 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 Int. Conf. Commun.*, Jun. 2014, pp. 5184–5190.

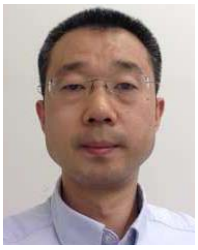
[23] H. V. Cheng, E. Björnson, and E. G. Larsson, "Optimal pilot and payload power control in single-cell massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 65, no. 9, pp. 2363–2378, May 2017.

[24] W. Zhang and W. Zhang, "On optimal training in massive MIMO systems with insufficient pilots," in *Proc. IEEE Int. Conf. Commun.*, May 2017, pp. 1–6.

[25] K. T. Truong, A. Lozano, and R. W. Heath, Jr., "Optimal training in continuous flat-fading massive MIMO systems," in *Proc. 20th Eur. Wireless Conf.*, May 2014, pp. 1–6.

[26] Y. Zhang, W.-P. Zhu, and J. Ouyang,"Energy efficient pilot and data power allocation in multi-cell multi-user massive MIMO communication systems estimation," in *Proc. IEEE 84th Veh. Tech. Conf. (VTC-Fall)*, Sep. 2016, pp. 3571–3575.
[27] T. van 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, Sep. 2016.
[28] T. van Chien, E. Björnson, and E. G. Larsson, "Joint pilot design and uplink power allocation in multi-cell massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 17, no. 3, pp. 2000–2015, Mar. 2018.
[29] S. M. Kay, *Fundamentals of Statistical Signal Processing: Practical Algorithm Development*, vol. 3. London, U.K.: Pearson, 2013.
[30] S. A. Jafar, "Interference alignment: A new look at signal dimensions in a communication network," *Found. Trends Commun. Inf. Theory*, vol. 7, no. 1, pp. 1–134, Jan. 2010.
[31] H. Gao, P. J. Smith, and M. V. Clark, "Theoretical reliability of MMSE linear diversity combining in Rayleigh-fading additive interference channels," *IEEE Trans. Commun.*, vol. 46, no. 5, pp. 666–672, May 1998.
[32] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Hoboken, NJ, USA: Wiley, 1991.
[33] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge, U.K.: Cambridge Univ. Press, 2005.
[34] A. Goldsmith, *Wireless Communication*. Cambridge, U.K.: Cambridge Univ. Press, 2005.
[35] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
[36] M. Hong, Z.-Q. Luo, and M. Razaviyayn, "Convergence analysis of alternating direction method of multipliers for a family of nonconvex problems," *SIAM J. Optim.*, vol. 26, no. 1, pp. 337–364, Jan. 2016.
[37] S. Lu, M. Hong, and Z. Wang, "A nonconvex splitting method for symmetric nonnegative matrix factorization: Convergence analysis and optimality," *IEEE Trans. Signal Process.*, vol. 65, no. 12, pp. 3120–3135, Jun. 2017.
[38] Y. Wang, W. Yin, and J. Zeng, "Global convergence of ADMM in nonconvex nonsmooth optimization," *J. Sci. Comput.*, pp. 1–35, Jun. 2018.
[39] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, Jan. 2011.
[40] D. P. Bertsekas, *Nonlinear Programming*, 2nd ed. Belmont, MA, USA: Athena Scientific, 1999.
[41] K. Guo, Y. Guo, and G. Ascheid, "Energy-efficient uplink power allocation in multi-cell MU-massive-MIMO systems," in *Proc. 21st Eur. Wireless Conf.*, pp. 1–5, May 2015.

Songtao Lu (S'10–M'18) received the Ph.D. degree in electrical and computer engineering from Iowa State University in 2018. He is currently a Post-Doctoral Associate with the Department of Electrical and Computer Engineering, University of Minnesota Twin Cities, Minneapolis. He was a recipient of the Graduate and Professional Student Senate Research Award from Iowa State University in 2015, the Research Excellence Award from the Graduate College of Iowa State University in 2017, and the Student Travel Award from the 20th International Conference on Artificial Intelligence and Statistics in 2017. His primary research interests include wireless communications, optimization, and machine learning.



Zhengdao Wang (S'00–M'02–SM'08–F'16) received the B.S. degree in electronic engineering and information science from the University of Science and Technology of China in 1996, the M.Sc. degree in electrical and computer engineering from the University of Virginia in 1999, and the Ph.D. degree in electrical and Computer engineering from the University of Minnesota in 2002. He is currently with the Department of Electrical and Computer Engineering, Iowa State University. His research interests are in the areas of signal processing, communications, information theory, and machine learning. He was a co-recipient of the *IEEE Signal Processing Magazine* Best Paper Award in 2003, the IEEE Communications Society Marconi Paper Prize Award in 2004, and the *EURASIP Journal on Advances in Signal Processing* Best Paper Award in 2009. He served as an Associate Editor for the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, the IEEE SIGNAL PROCESSING LETTERS, the *IEEE Transactions on Wireless Communications*, *ZTE Communications*, and the IEEE TRANSACTIONS ON SIGNAL PROCESSING, IEEE Signal Processing Society Online Video Library.