SCIENCE CHINA
Information Sciences

• RESEARCH PAPER •

April 2017, Vol. 60 042302:1–042302:13
doi: 10.1007/s11432-016-0069-0

Resource allocation for pilot-assisted massive MIMO transmission

Yun XUE[1], Jun ZHANG[2] & Xiqi GAO[1]*

[1]*School of Information Science and Engineering, Southeast University, Nanjing* 210096*, China;*
[2]*College of Telecommunications and Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing* 210003*, China*

Received January 28, 2016; accepted March 30, 2016; published online November 17, 2016

Abstract This paper is on the resource allocation problem for pilot-assisted multi-user massive multiple-input-multiple-output (MIMO) uplink with linear minimum mean-squared error (MMSE) channel estimation and detection. We utilize the angular domain channel representation for uniform linear antenna arrays, and adopt its equivalent independent and nonidentical distributed channel model. For a given coherence interval and total energy budget, we study the joint optimization of the training length and the training power to maximize the achievable sum-rate. For tractable analysis and low-complexity solution, a tight approximation on the achievable sum-rate is derived first. Then the training length optimization for fixed training power and the training power optimization for fixed training length with respect to the approximate sum-rate maximization are both shown to be concave. An alternative optimization that solves the training length and power iteratively is proposed for the joint resource allocation. In addition, for the special case that the training and data transmission powers are equal, we derive the optimal training lengths for both high and low signal-to-noise-ratio (SNR) regions. Numerical results show the tightness of the derived sum-rate approximation and also the significant performance advantage of the proposed resource allocation.

Keywords massive MIMO, independent nonidentical distribution (i.n.d.), achievable sum-rate, optimal training length, power allocation

Citation Xue Y, Zhang J, Gao X Q. Resource allocation for pilot-assisted massive MIMO transmission. Sci China Inf Sci, 2017, 60(4): 042302, doi: 10.1007/s11432-016-0069-0

1 Introduction

Massive multiple-input-multiple-output (MIMO) configuration is considered to be a key enabler for the 5th generation (5G) wireless systems [1]. By employing a large number (usually hundreds) of antennas at the base station (BS) to simultaneously and jointly serve multiple (i.e., dozens) mobile user terminals (UTs), massive MIMO can greatly improve the performance of cellular networks, including spectral efficiency and energy efficiency. For the evaluation of spectral efficiency, the independent and identically distributed (i.i.d.) Rayleigh fading channels were adopted in the paper by Marzetta [2]. Due to unlimited numbers of antennas at the BS and the asymptotically orthogonality between propagation beam

* Corresponding author (email: xqgao@seu.edu.cn)

vectors of different users, the matched filter (MF) receiver becomes optimal, and the effect of noise and intra-cell interference diminishes. Furthermore, the transmit power can be made arbitrarily low without performance loss. By exploiting the uplink-donwlink reciprocity in time-division duplex (TDD) systems, the required channel state information (CSI) for downlink at the BS can be acquired via uplink training. Nevertheless, because of pilot sequence reuse among cells, pilot contamination persists for multicell systems [3, 4].

In the practical case of BS equipped with a large, but finite number of antennas, pilot contamination has a more severe impact on system performance relative to the effect of additive white noise [5]. The performance of MF and more complex linear receivers has been analyzed in [3, 6]. For finite number of antennas, minimum mean-squared error (MMSE) scheme can provide better performance than MF scheme [6]. However, compared with the physical channel model adopt in [6], the statistical channel models obtained from the physical angular domain analysis are more plausible for the design and performance analysis of practical communication systems [7]. Moreover, the system performance largely depends on the resource allocation between training and data transmission phases and the quality of the estimated CSI.

In [8], the linear MMSE channel estimation with fixed training length and training power was derived and shown to provide near-optimal performance for multicell systems. However, we bring out the channel estimation research based on the background of single-cell system with frequency-division duplex (FDD) operation. FDD system has spurred much recent research interest [1]. Joint Spatial Division and Multiplexing (JSDM) proposed in [9], offers strong potential to achieve massive MIMO link gains. Furthermore, Beam Domain Multiple Access (BDMA) proposed in [10], can even be qualified for medium or high-mobility user applications in single-cell system.

There are very few work on the evaluation of achievable rate with linear MMSE receiver for the statistical channel models obtained from the physical angular domain, as well as, the resource allocation for pilot-assisted massive MIMO communications. Regarding the training length design, an early study for the tradition point-to-point MIMO system showed that if the optimization over the training and data powers is allowed, the optimal training length equals the number of transmit antennas [11]. This conclusion is based on a lower bound on the information-theoretic capacity. For uplink massive MIMO transmission with maximum-ratio combining detection, in a recent work [12], a similar result that the optimal training length equals the number of UTs was obtained based on the nonasymptotic sum-rate maximization under i.i.d. channels. Notice that, some recent work on resource allocation, such as [13], adopts this setting in the training phase.

In this paper, we analyze the achievable sum-rate of pilot-assisted massive MIMO uplink with linear MMSE detection and channel estimation, and derive training resource allocation policies to maximize the achievable sum-rate. We further use the angular domain channel representation for uniform linear antenna arrays, and adopt its equivalent independent and nonidentical distributed (i.n.d.) channel model. The analysis are for asymptotically large number of BS antennas, while the number of UTs are assumed to be fixed and small [3, 9]. We first derive an approximation on the achievable sum-rate of the pilot-assisted massive MIMO uplink. Unlike the sum-rate results in [6] whose evaluation requires iterative numerical procedure, ours is in closed-form. Based on this analytical sum-rate result, we study the joint optimization of the training length and training power. By proving the concavity of the training power optimization for fixed training length and the training length optimization for fixed training power, we propose a low-complexity alternative optimization algorithm for the joint optimization. Simulations are conducted to validate the analytical result and to show the advantage of the proposed resource allocation solution.

2 System model and problem formulation

In this section, the model of the multi-user pilot-assisted massive MIMO system is elaborated, including the channel model and the pilot-assisted uplink transmission scheme. Then the joint training length and power optimization problem is formulated.

2.1 Channel model

We consider a single-cell multi-user massive MIMO system that consists of one BS and K UTs. The BS is equipped with a uniform linear array (ULA) of M antennas spaced with a half wavelength distance. All UTs have single antenna. Denote the $M \times 1$ channel vector from the kth UT to the BS as $\boldsymbol{g}_k$. Using the channel model in [14], we have

$$\boldsymbol{g}_k = \sqrt{\beta_k}\,\boldsymbol{V}_k^{\frac{1}{2}}\boldsymbol{h}_k, \tag{1}$$

where β_k represents the large scale fading, $\boldsymbol{h}_k$ is a complex vector of i.i.d. zero-mean unit variance complex Gaussian elements and $\boldsymbol{V}_k$ is an $M \times M$ diagonal matrix that represents the variance profile of the channels between the kth UT and the BS. The mth diagonal element of $\boldsymbol{V}_k$ (which we denote as v_{km} to help the presentation) is given by

$$v_{km} \triangleq [\boldsymbol{V}_k]_{mm} = M \cdot \boldsymbol{S}_k\left(\arcsin\left(\psi_{m-1}\right)\right)\left[\arcsin\left(\psi_m\right) - \arcsin\left(\psi_{m-1}\right)\right], \text{ for } \ m = 1, \ldots, M, \tag{2}$$

where $\psi_n = n/M$ for $n = 0, 1, \ldots, M$ and $\boldsymbol{S}_k(\theta)$ represents the channel power angle spectrum (PAS) which models the channel power distribution in the angular domain [15]. The channel PAS is normalization as $\int_{-\infty}^{\infty} \boldsymbol{S}_k(\theta)\mathrm{d}\theta = 1$.

We employ $\mathbb{E}\{\cdot\}$ to denote the expectation operation, and the superscript $(\cdot)^\dagger$ to denote the conjugate-transpose operation, respectively. With the aforementioned channel model, from (1), the covariance matrix of $\boldsymbol{g}_k$ can be calculated as $\mathbb{E}\{\boldsymbol{g}_k\boldsymbol{g}_k^\dagger\} = \beta_k\boldsymbol{V}_k$. We define that $\boldsymbol{G} \triangleq [\boldsymbol{g}_1\ \boldsymbol{g}_2\ \cdots\ \boldsymbol{g}_K]$, which is the $M \times K$ channel matrix between all UTs and the BS. We employ the operator $\mathrm{diag}\{\boldsymbol{a}\}$ to denote the diagonal matrix whose diagonal entries are elements of $\boldsymbol{a}$, and $\mathrm{tr}(\boldsymbol{A})$ to denote the trace of the matrix $\boldsymbol{A}$, respectively. Let $\boldsymbol{R} \triangleq \mathbb{E}\{\boldsymbol{G}^\dagger\boldsymbol{G}\}$. It can be shown straightforwardly that $\boldsymbol{R}$ is diagonal and its kth diagonal element is $[\boldsymbol{R}]_{kk} = \beta_k\mathrm{tr}(\boldsymbol{V}_k) = \beta_k M$. That is $\boldsymbol{R} = M\mathrm{diag}\{\beta_1, \ldots, \beta_K\}$.

2.2 Pilot-assisted uplink transmission scheme

We consider a pilot-assisted uplink transmission scheme, where a coherence interval is composed of two phases: the uplink training phase and the uplink data transmission phase.

2.2.1 *Uplink training phase*

In the uplink training phase, pilots are sent from the UTs to the BS for the BS to learn the channel matrix $\boldsymbol{G}$. Let T_t be the length of the training phase measured in the number of symbol transmissions, p_t be the training power per training symbol for all UTs, and $\sqrt{p_tT_t}\boldsymbol{S}$ be the $K \times T_t$ pilot matrix. Thus the kth UT sends the kth row of $\sqrt{p_tT_t}\boldsymbol{S}$ and all UTs transmit simultaneously. $\boldsymbol{S}$ is assumed to satisfy

$$\boldsymbol{S}\boldsymbol{S}^\dagger = \boldsymbol{I}_K. \tag{3}$$

This implies that the pilot vectors of different UTs are mutually orthogonal and $T_t \geqslant K$.

The $M \times T_t$ received matrix at the BS in the uplink training phase is given by

$$\boldsymbol{Y}_t = \sqrt{p_tT_t}\boldsymbol{G}\boldsymbol{S} + \boldsymbol{N}_t, \tag{4}$$

where $\boldsymbol{N}_t$ is the $M \times T_t$ noise matrix with i.i.d. zero-mean unit variance complex Gaussian elements. Following the results in [16, 17], the linear MMSE estimate of $\boldsymbol{G}$ at the BS is written as

$$\begin{aligned}\hat{\boldsymbol{G}} &= \frac{1}{\sqrt{p_tT_t}}\boldsymbol{Y}_t\left(\boldsymbol{S}^\dagger\boldsymbol{R}\boldsymbol{S} + \frac{M}{p_tT_t}\boldsymbol{I}_{T_t}\right)^{-1}\boldsymbol{S}^\dagger\boldsymbol{R} \\ &\overset{(a)}{=} \left(\boldsymbol{G}\boldsymbol{S} + \frac{1}{\sqrt{p_tT_t}}\boldsymbol{N}_t\right)\left[\frac{p_tT_t}{M}\boldsymbol{I}_{T_t} - \frac{p_t^2T_t^2}{M^2}\boldsymbol{S}^\dagger\left(\boldsymbol{R}^{-1} + \frac{p_tT_t}{M}\boldsymbol{I}_K\right)^{-1}\boldsymbol{S}\right]\boldsymbol{S}^\dagger\boldsymbol{R}\end{aligned}$$

$$= \left(\boldsymbol{G} + \frac{1}{\sqrt{p_t T_t}}\boldsymbol{N}_t\boldsymbol{S}^\dagger\right)\left(\frac{M}{p_t T_t}\boldsymbol{I}_K + \boldsymbol{R}\right)^{-1}\boldsymbol{R}$$
$$= \left(\boldsymbol{G} + \frac{1}{\sqrt{p_t T_t}}\boldsymbol{N}_t\boldsymbol{S}^\dagger\right)\mathrm{diag}\left\{\frac{p_t T_t \beta_1}{1 + p_t T_t \beta_1}, \ldots, \frac{p_t T_t \beta_K}{1 + p_t T_t \beta_K}\right\}. \tag{5}$$

In the above derivations, (a) is obtained by using the matrix inversion identity $(\boldsymbol{A}+\boldsymbol{BCD})^{-1} = \boldsymbol{A}^{-1} - \boldsymbol{A}^{-1}\boldsymbol{B}(\boldsymbol{DA}^{-1}\boldsymbol{B} + \boldsymbol{C}^{-1})^{-1}\boldsymbol{DA}^{-1}$ [18] and the transceiver equation in (4). For any pilot satisfying (3), it can be shown that $\boldsymbol{N}_t\boldsymbol{S}^\dagger$ is an $M \times K$ random matrix with i.i.d. zero-mean unit variance complex Gaussian elements. Define $\hat{\boldsymbol{g}}_k$ as the kth column of $\hat{\boldsymbol{G}}$, and $\boldsymbol{K}_{\hat{\boldsymbol{g}}_k}$ as the covariance matrix of $\hat{\boldsymbol{g}}_k$. Thus, $\hat{\boldsymbol{g}}_k$ satisfies

$$\hat{\boldsymbol{g}}_k \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{K}_{\hat{\boldsymbol{g}}_k}), \tag{6}$$

where $\mathcal{CN}(\boldsymbol{0}, \boldsymbol{K}_{\hat{\boldsymbol{g}}_k})$ represents the circularly symmetric complex Gaussian distribution whose mean vector is $\boldsymbol{0}$ and whose covariance matrix is

$$\boldsymbol{K}_{\hat{\boldsymbol{g}}_k} = \frac{p_t^2 T_t^2 \beta_k^3}{(1 + p_t T_t \beta_k)^2}\boldsymbol{V}_k + \frac{p_t T_t \beta_k^2}{(1 + p_t T_t \beta_k)^2}\boldsymbol{I}_M.$$

Define that $\tilde{\boldsymbol{G}} \triangleq \boldsymbol{G} - \hat{\boldsymbol{G}}$, which is the channel estimation error. The covariance matrix of $\tilde{\boldsymbol{G}}$ can be calculated to be

$$\mathbb{E}\left\{\tilde{\boldsymbol{G}}\tilde{\boldsymbol{G}}^\dagger\right\} = \sum_{k=1}^{K}\left[\frac{(1 + 2p_t T_t \beta_k)\,\beta_k}{(1 + p_t T_t \beta_k)^2}\boldsymbol{V}_k - \frac{p_t T_t \beta_k^2}{(1 + p_t T_t \beta_k)^2}\boldsymbol{I}_M\right]. \tag{7}$$

2.2.2 *Uplink data transmission phase*

After the training phase in which the BS obtains the channel estimate $\hat{\boldsymbol{G}}$, the uplink data transmission phase follows. Let T (in symbol transmissions) be the length of the coherence interval. Since the length of the training phase is T_t, the length of the data transmission phase is $T - T_t$. Without loss of generality, we consider an arbitrary time slot of the data transmission phase. Let $\boldsymbol{x}$ be the vector containing the information symbols of all users. $\boldsymbol{x}$ is assumed to have zero mean and its covariance matrix is $\mathbb{E}\{\boldsymbol{x}\boldsymbol{x}^\dagger\} = \boldsymbol{I}_K$. Let the transmit power for all UTs be p_{u}. The $M \times 1$ received vector at the BS can be written as

$$\boldsymbol{y} = \sqrt{p_u}\boldsymbol{G}\boldsymbol{x} + \boldsymbol{n} = \sqrt{p_u}\hat{\boldsymbol{G}}\boldsymbol{x} + \sqrt{p_u}\tilde{\boldsymbol{G}}\boldsymbol{x} + \boldsymbol{n}, \tag{8}$$

where $\boldsymbol{n}$ is the $M \times 1$ noise vector following $\mathcal{CN}(0, 1)$.

We define that $\boldsymbol{w} \triangleq \tilde{\boldsymbol{G}}\boldsymbol{x} + \frac{1}{\sqrt{p_u}}\boldsymbol{n}$. As $\boldsymbol{n}$ and $\tilde{\boldsymbol{G}}$ are independent circularly symmetric complex Gaussian random vectors, $\boldsymbol{w}$ is also a circularly symmetric complex Gaussian random vector. It is straightforward to show that its mean is zero and its covariance matrix is

$$\boldsymbol{K}_{\boldsymbol{w}} = \sum_{k=1}^{K}\left[\frac{(1 + 2p_t T_t \beta_k)\,\beta_k}{(1 + p_t T_t \beta_k)^2}\boldsymbol{V}_k - \frac{p_t T_t \beta_k^2}{(1 + p_t T_t \beta_k)^2}\boldsymbol{I}_M\right] + \frac{1}{p_u}\boldsymbol{I}_M$$
$$= \mathrm{diag}\left\{\frac{1}{p_u} + \sum_{k=1}^{K}\frac{(1 + 2p_t T_t \beta_k)\,\beta_k v_{k1} - p_t T_t \beta_k^2}{(1 + p_t T_t \beta_k)^2}, \ldots, \frac{1}{p_u} + \sum_{k=1}^{K}\frac{(1 + 2p_t T_t \beta_k)\,\beta_k v_{kK} - p_t T_t \beta_k^2}{(1 + p_t T_t \beta_k)^2}\right\}. \tag{9}$$

As in [7], we define that

$$\boldsymbol{K}_{\boldsymbol{z},k} \triangleq \sum_{i=1, i\neq k}^{K}\hat{\boldsymbol{g}}_i\hat{\boldsymbol{g}}_i^\dagger + \boldsymbol{K}_{\boldsymbol{w}}. \tag{10}$$

Under linear MMSE detection, the received signal-to-interference-plus-noise-ratio (SINR) for the kth UT is thus given by

$$\rho_k = \hat{\boldsymbol{g}}_k^\dagger \boldsymbol{K}_{\boldsymbol{z},k}^{-1}\hat{\boldsymbol{g}}_k. \tag{11}$$

The average achievable uplink rate for the kth UT is thus

$$R_k \triangleq \left(1 - \frac{T_t}{T}\right) \mathbb{E}\left\{\log_2\left(\rho_k + 1\right)\right\}, \tag{12}$$

where the coefficient $1 - T_t/T$ takes into consideration the length of the training phase. The achievable sum-rate of overall system is

$$R_{\text{sum}} \triangleq \sum_{k=1}^{K} R_k. \tag{13}$$

2.3 Problem formulation

Let P be the total transmit energy constraint for each UT within a coherence interval. Then, we have

$$T_t p_t + (T - T_t)\, p_u = P. \tag{14}$$

Intuitively, with given coherence length and total transmit energy constraint for the coherence interval, the designs of the training length T_t and the training power p_t can significantly affect the system performance. In this work, we consider the joint optimization of the training length and the training power to maximize the total sum-rate. By noticing the equality in (14), the problem can be formulated as follows:

$$\underset{T_t, p_t}{\text{maximize}} \quad R_{\text{sum}}, \quad \text{s.t.} \quad 0 \leqslant p_t \leqslant \tfrac{P}{T_t},\ K \leqslant T_t < T,\ T_t \in \mathbb{N}, \tag{15}$$

where $\mathbb{N}$ denotes the positive integer set. Notice that, we assume that the coherence interval satisfies the condition: $K < T$.

3 Achievable sum-rate analysis and resource allocation

In this section, we solve the resource allocation problem formulated in Subsection 2.3. Due to the complex nature of the sum-rate formula in (10)–(13), to find a solution, we first derive a tractable approximate on the system sum-rate; then conduct the optimization over the approximate sum-rate.

3.1 Achievable sum-rate analysis

In this subsection, we derive an approximate probability density function (PDF) for the SINR following the framework presented in [19]. Then an approximation of the achievable sum-rate for the pilot-aided multi-user massive MIMO uplink is obtained.

By using the matrix inversion identity repetitively, we can rewrite the received SINR for the kth UT given by (11) as

$$\rho_k = \left\{\left[\left(\hat{\boldsymbol{G}}^\dagger \boldsymbol{K}_{\boldsymbol{w}}^{-1} \hat{\boldsymbol{G}} + \boldsymbol{I}_K\right)^{-1}\right]_{kk}\right\}^{-1} - 1. \tag{16}$$

Define $\bar{\boldsymbol{G}} \triangleq M^{\frac{1}{2}} \boldsymbol{K}_{\boldsymbol{w}}^{-\frac{1}{2}} \hat{\boldsymbol{G}}$. Thus the SINR in (16) can be rewritten as

$$\rho_k = \left\{\left[\left(\frac{1}{M}\bar{\boldsymbol{G}}^\dagger \bar{\boldsymbol{G}} + \boldsymbol{I}_K\right)^{-1}\right]_{kk}\right\}^{-1} - 1. \tag{17}$$

The covariance matrix of kth column of $\bar{\boldsymbol{G}}$ can be calculated to be

$$\begin{aligned}\bar{\boldsymbol{V}}_k =& M \boldsymbol{K}_{\boldsymbol{w}}^{-1} \boldsymbol{K}_{\hat{\boldsymbol{g}}_k} \\ =& M\left[\sum_{l=1}^{K}\left(\frac{(1 + 2p_tT_t\beta_l)\,\beta_l}{(1 + p_tT_t\beta_l)^2}\boldsymbol{V}_l - \frac{p_tT_t\beta_l^2}{(1 + p_tT_t\beta_l)^2}\boldsymbol{I}_M\right) + \frac{1}{p_u}\boldsymbol{I}_M\right]^{-1}\end{aligned}$$

$$\times\left[\frac{p_t^2T_t^2\beta_k^3}{(1+p_tT_t\beta_k)^2}\boldsymbol{V}_k+\frac{p_tT_t\beta_k^2}{(1+p_tT_t\beta_k)^2}\boldsymbol{I}_M\right]$$
$$=M\mathrm{diag}\left\{\left[\frac{1}{p_u}+\sum_{l=1}^{K}\frac{(1+2p_tT_t\beta_l)\,\beta_l v_{l1}-p_tT_t\beta_l^2}{(1+p_tT_t\beta_l)^2}\right]^{-1}\frac{p_t^2T_t^2\beta_k^3v_{k1}+p_tT_t\beta_k^2}{(1+p_tT_t\beta_k)^2},\right.$$
$$\left.\dots,\left[\frac{1}{p_u}+\sum_{l=1}^{K}\frac{(1+2p_tT_t\beta_l)\,\beta_l v_{lK}-p_tT_t\beta_l^2}{(1+p_tT_t\beta_l)^2}\right]^{-1}\frac{p_t^2T_t^2\beta_k^3v_{kK}+p_tT_t\beta_k^2}{(1+p_tT_t\beta_k)^2}\right\}. \tag{18}$$

The kth column of $\bar{\boldsymbol{G}}$ can be represented

$$\bar{\boldsymbol{g}}_k=\bar{\boldsymbol{V}}_k^{\frac{1}{2}}\boldsymbol{f}_k, \tag{19}$$

where $\boldsymbol{f}$ is an $M\times 1$ random vector whose entries are i.i.d. $\mathcal{CN}(0,1)$.

Following the work in [19], we can approximate ρ_k by a Gamma distribution with a shape parameter α_k and a scale parameter ζ_k. To have the first two moments of ρ_k match with the corresponding moments of the Gamma distribution, the parametere values can be calculated via the following procedures.

Define $\boldsymbol{T}(z)$ recursively as

$$\boldsymbol{T}(z)\triangleq\left[\sum_{i=1}^{M}\frac{\tilde{\boldsymbol{V}}_{i(-k)}}{1+\mathrm{tr}(\tilde{\boldsymbol{V}}_{i(-k)}\boldsymbol{T}(z))}-z\boldsymbol{I}_{K-1}\right]^{-1}, \tag{20}$$

where $\tilde{\boldsymbol{V}}_i\triangleq\frac{1}{M}\cdot\mathrm{diag}\{[\bar{\boldsymbol{V}}_1]_{ii},\dots,[\bar{\boldsymbol{V}}_K]_{ii}\}$ and $\tilde{\boldsymbol{V}}_{i(-k)}$ is $\tilde{\boldsymbol{V}}_i$ with the kth column removed. Let $\mu_\gamma\triangleq\mathrm{tr}\,(\boldsymbol{T}(-1))/(K-1)$ and $\sigma_\gamma^2\triangleq\mathrm{tr}(\boldsymbol{T}'(-1))/(K-1)$, where $\boldsymbol{T}'(z)$ denotes the first derivative of $\boldsymbol{T}(z)$. It can be solved numerically by

$$\boldsymbol{T}'(z)=\left[\sum_{i=1}^{M}\frac{\tilde{\boldsymbol{V}}_{i(-k)}}{1+\mathrm{tr}(\tilde{\boldsymbol{V}}_{i(-k)}\boldsymbol{T}(z))}-z\boldsymbol{I}_{K-1}\right]^{-2}. \tag{21}$$

The shape parameter and the scale parameter of the Gamma distribution approximation are

$$\alpha_k=\frac{[M-K+1+(K-1)\,\mu_{\gamma,c}]^2}{M-K+1+(K-1)\,\sigma_{\gamma,c}^2}, \tag{22}$$

and

$$\zeta_k=\frac{M-K+1+(K-1)\,\sigma_{\gamma,c}^2}{M-K+1+(K-1)\,\mu_{\gamma,c}}\frac{1}{(1+p_tT_t\beta_k)^2}\left[\frac{1}{M}\sum_{i=1}^{M}\frac{p_t^2T_t^2\beta_k^3v_{ki}+p_tT_t\beta_k^2}{\frac{1}{p_u}+\sum_{l=1}^{K}\frac{(1+2p_tT_t\beta_l)\beta_lv_{li}-p_tT_t\beta_l^2}{(1+p_tT_t\beta_l)^2}}\right]. \tag{23}$$

With this approximation on the distribution of ρ_k, the following approximate result on the system achievable rate is obtained from (12):

$$\begin{aligned}R_k&=\left(1-\frac{T_t}{T}\right)\mathbb{E}\left\{\log_2(\rho_k+1)\right\}\\&\overset{(a)}{\leqslant}\left(1-\frac{T_t}{T}\right)\log_2(\mathbb{E}\{\rho_k\}+1)\\&\approx\left(1-\frac{T_t}{T}\right)\log_2(\alpha_k\zeta_k+1),\end{aligned} \tag{24}$$

where (a) is obtained by using Jensen's inequality $\mathbb{E}\{\log_2(\rho_k+1)\}\leqslant\log_2(\mathbb{E}\{\rho_k\}+1)$ [20, 21]. In accordance with [2, 3, 22], we assume that the number of terminals K remains fixed while the number of BS antennas M grows without bound. In this case, following the deterministic equivalence analysis of the SINR [6], the right-hand-side of (a) is also a tight approximation of R_k and represents its asymptotic behavior for large M. From (24), we obtain an asymptotic approximation of achievable rate of the kth UT as stated in the following proposition.

Proposition 1. When the number of terminals K remains fixed and the number of BS antennas M grows without bound, i.e., $M \to \infty$, the achievable rate for the kth UT can be approximated by

$$\bar{R}_{k,\text{approx}} = \left(1 - \frac{T_t}{T}\right) \log_2 \left[\frac{1}{(1+p_tT_t\beta_k)^2} \sum_{i=1}^{M} \frac{p_t^2T_t^2\beta_k^3 v_{ki} + p_tT_t\beta_k^2}{\frac{1}{p_u} + \sum_{l=1}^{K} \frac{(1+2p_tT_t\beta_l)\beta_l v_{li} - p_tT_t\beta_l^2}{(1+p_tT_t\beta_l)^2}} + 1\right]. \tag{25}$$

An approximation on the achievable sum-rate $\bar{R}_{\text{sum,approx}}$ is thus given by

$$\begin{aligned}\bar{R}_{\text{sum,approx}} &= \sum_{k=1}^{K} \bar{R}_{k,\text{approx}} \\ &= \left(1 - \frac{T_t}{T}\right) \sum_{k=1}^{K} \log_2 \left[\frac{1}{(1+p_tT_t\beta_k)^2} \sum_{i=1}^{M} \frac{p_t^2T_t^2\beta_k^3 v_{ki} + p_tT_t\beta_k^2}{\frac{1}{p_u} + \sum_{l=1}^{K} \frac{(1+2p_tT_t\beta_l)\beta_l v_{li} - p_tT_t\beta_l^2}{(1+p_tT_t\beta_l)^2}} + 1\right]. \end{aligned}\tag{26}$$

Proof. Let $\gamma = (K-1)/M$. From (22) and (23),

$$\alpha_k\zeta_k = (1 - \gamma + \gamma\mu_{\gamma,c}) \frac{1}{(1+p_tT_t\beta_k)^2} \sum_{i=1}^{M} \frac{p_t^2T_t^2\beta_k^3 v_{ki} + p_tT_t\beta_k^2}{\frac{1}{p_u} + \sum_{l=1}^{K} \frac{(1+2p_tT_t\beta_l)\beta_l v_{li} - p_tT_t\beta_l^2}{(1+p_tT_t\beta_l)^2}}. \tag{27}$$

In [19], it has been shown that $0 < \mu_\gamma < 1$. Thus when K is finite and $M \to \infty$, we have $\gamma \to 0$ and

$$\alpha_k\zeta_k \to \frac{1}{(1+p_tT_t\beta_k)^2} \sum_{i=1}^{M} \frac{p_t^2T_t^2\beta_k^3 v_{ki} + p_tT_t\beta_k^2}{\frac{1}{p_u} + \sum_{l=1}^{K} \frac{(1+2p_tT_t\beta_l)\beta_l v_{li} - p_tT_t\beta_l^2}{(1+p_tT_t\beta_l)^2}}. \tag{28}$$

This completes the proof.

Compared with the sum-rate expression in Subsection 2.3, Proposition 1 provides a tractable closed-form asymptotic approximation for the achievable sum-rate. In addition, Proposition 1 indicates that the derived asymptotic approximation of the achievable sum-rate avoids the computations of μ_γ and σ_γ^2, which have to be solved numerically and need tens of iterations. This reduces the computation complexity in the sum-rate calculations.

3.2 Training length and training power optimization

In this subsection, we solve the optimization of the training length T_t and the training power p_t. For the tractability of the problem and to find low-complexity designs, the asymptotic sum-rate approximation derived in Proposition 1 replaces the exact average sum-rate as the objective function for the optimization problem in (15). Moreover, the problem is a joint optimization with respect to p_t and T_t, and is therefore difficult to tackle. We decompose the problem into two subproblems: (1) for a given p_t, find the optimal T_t that maximizes the achievable sum-rate, and (2) for a given T_t, find p_t that maximizes the achievable sum-rate. The two subproblems are solved in Subsections 3.2.1 and 3.2.2, respectively. Then the overall optimization algorithm is provided in Subsection 3.2.3 via alternative optimization.

3.2.1 *Pilot length optimization*

For an arbitrarily given data transmission power p_u and pilot transmission power p_t, the pilot length optimization problem can be formulated as

$$\begin{aligned} \underset{T_t}{\text{maximize}} \quad & \bar{R}_{\text{sum,approx}}, \\ \text{s.t.} \quad & K \leqslant T_t < T, \\ & T_t \in \mathbb{N}. \end{aligned}\tag{29}$$

Since T_t takes discrete integer value with finite possibilities, it is obvious that exhaustive search can be used to find the optimal solution. But such method is computationally expensive. For a low computation solution we prove the following proposition.

Proposition 2. The asymptotic sum-rate approximation $\bar{R}_{\text{sum,approx}}$ is a concave function of T_t for $T_t \in \mathbb{R}$, with $\mathbb{R}$ being the real number set.

Proof. Define that

$$a_{ki} \triangleq p_t^2 T_t^2 \beta_k^3 v_{ki} + p_t T_t \beta_k^2, \tag{30}$$

$$b_k \triangleq (1 + p_t T_t \beta_k)^2, \tag{31}$$

$$c_{ki} \triangleq \frac{1}{p_u} + \sum_{k=1}^{K} \frac{(1 + 2p_t T_t \beta_k)\, \beta_k v_{ki} - p_t T_t \beta_k^2}{(1 + p_t T_t \beta_k)^2}. \tag{32}$$

Thus,

$$\bar{R}_{k,\text{approx}} = \left(1 - \frac{T_t}{T}\right) \log_2 \left(\sum_{i=1}^{M} \frac{a_{ki}}{b_k c_{ki}} + 1 \right). \tag{33}$$

The second derivative of $\bar{R}_{k,\text{approx}}$ over T_{t} is given by

$$\bar{R}''_{k,\text{approx}} = \frac{[(1 - \frac{T_t}{T}) \sum_{i=1}^{M} (\frac{a_{ki}}{b_k c_{ki}})'' - \frac{2}{T} \sum_{i=1}^{M} (\frac{a_{ki}}{b_k c_{ki}})'](\sum_{i=1}^{M} \frac{a_{ki}}{b_k c_{ki}} + 1) - (1 - \frac{T_t}{T})[\sum_{i=1}^{M} (\frac{a_{ki}}{b_k c_{ki}})']^2}{(\ln 2)(\sum_{i=1}^{M} \frac{a_{ki}}{b_k c_{ki}} + 1)^2}.$$

After tedious but straightforward calculations, we can show that $\bar{R}''_{k,\text{approx}} < 0$. Therefore, $\bar{R}_{k,\text{approx}}$ is a concave function in $K \leqslant T_t < T$. Finally, using the fact that the summation of concave functions is concave, we conclude the proof.

Since the objective function is concave, its optimization can be solved via one-dimensional linear search, such as the golden section search [23]. The ceiling and floor functions of T_t^* are written as $\lceil T_t^* \rceil$ and $\lfloor T_t^* \rfloor$, respectively. As the optimal value for the training length must be an integer, we first find the real value T_t^* that maximizes $\bar{R}_{\text{sum,approx}}$, then compare the sum-rates obtained by using $\lfloor T_t^* \rfloor$ and $\lceil T_t^* \rceil$. The optimal training length is thus the one results in the larger sum-rate.

3.2.2 *Data transmission power optimization*

For an arbitrarily given training length T_t, the data transmission power optimization problem can be formulated as

$$\underset{p_t}{\text{maximize}} \quad \bar{R}_{\text{sum,approx}}, \quad \text{s.t.} \quad 0 \leqslant p_t \leqslant \tfrac{P}{T_t}. \tag{34}$$

The following proposition is proved.

Proposition 3. The asymptotic sum-rate approximation $\bar{R}_{\text{sum,approx}}$ is a concave function of p_t.

Proof. With the definitions in (30)–(32), the asymptotic sum-rate approximation is given in (33). Since the coefficient $1 - T_t/T$ is independent of p_t and the logarithm is also a concave function, it is sufficient to show that $\sum_{i=1}^{M} \frac{a_{ki}}{b_k c_{ki}}$ is concave in p_t for $0 \leqslant p_t \leqslant P/T_t$.

The second derivative of $\bar{R}_{k,\text{approx}}$ over p_t is given by

$$\sum_{i=1}^{M} \left(\frac{a_{ki}}{b_k c_{ki}} \right)'' = -T_t \beta_k^2 b_k^{-2} b_k \sum_{i=1}^{M} \frac{d_{ki}}{c_{ki}^2} + T_t \beta_k^2 b_k^{-1} \sum_{i=1}^{M} \frac{d'_{ki} - 2 d_{ki} c'_{ki} / c_{ki}}{c_{ki}^2},$$

where

$$d_{ki} \triangleq [p_t T_t \beta_k (2v_{ki} - 1) + 1]\, c_{ki} + p_t (p_t T_t \beta_k v_{ki} + 1)(1 + p_t T_t \beta_k)\, c'_{ki}.$$

After tedious but straightforward calculations, we can show that $\sum_{i=1}^{M} (\frac{a_{ki}}{b_k c_{ki}})'' \leqslant 0$. Therefore, $\sum_{i=1}^{M} \frac{a_{ki}}{b_k c_{ki}}$ is a concave function of p_t for $0 \leqslant p_t \leqslant P/T_t$. This concludes the proof.

With the result in Proposition 3, the optimization in (34) can be solved via one-dimensional linear search, such as the golden section search [23].

3.2.3 *Overall optimization algorithm*

The overall joint training length and training power optimization problem with respect to maximizing the asymptotic approximate average sum-rate can be expressed as follows:

$$\underset{T_t,p_t}{\text{maximize}} \quad \bar{R}_{\text{sum,approx}}, \quad \text{s.t.} \quad 0 \leqslant p_t \leqslant \tfrac{P}{T_t},\ K \leqslant T_t < T,\ T_t \in \mathbb{N}. \tag{35}$$

We solve this problem by iteratively and alternatively optimizing T_t and p_t, as explained in Subsections 3.2.1 and 3.2.2. The propose algorithm in provided in Algorithm 1.

Algorithm 1 Iterative optimization algorithm for (35).

1: Initialization: $T_t^{(0)} = K$, $p_u^{(0)} = p_t^{(0)} = P/T$. Calculate $\bar{R}^{(0)}_{\text{sum,approx}}$ based on (26). Let $i = 1$.
2: Given $p_u^{(i-1)}$ and $p_t^{(i-1)}$, find the optimal real value ${T_t}^{*(i)}$ via one-dimensional linear search.
3: Update ${T_t}^{(i)} \leftarrow \lfloor {T_t}^{*(i)} \rfloor$ or ${T_t}^{(i)} \leftarrow \lceil {T_t}^{*(i)} \rceil$, depending on which results in the large approximate sum-rate.
4: Given ${T_t}^{(i)}$, find the optimal $p_t^{(i)}$ via one-dimensional linear search. Calculate $p_u^{(i)}$ using (14).
5: Given $p_u^{(i)}$, $p_t^{(i)}$ and $T_t^{(i)}$, calculate $\bar{R}^{(i)}_{\text{sum,approx}}$ based on (26).
6: If $|\bar{R}^{(i)}_{\text{sum,approx}} - \bar{R}^{(i-1)}_{\text{sum,approx}}|$ is less than the tolerance ϵ, output $T_t^{(i)}$ and $p_t^{(i)}$ as the solution. Otherwise, $i = i + 1$ and go to Step 2.

3.3 Training length design with equal power allocation

Depending on the applications, some communication systems may not have the luxury of adjusting the powers for the training and data transmission phases [11]. Thus in this section, we consider fixed transmission power where the training power p_t and the data transmission power p_u are the same, and study the optimization of the training length. Define that $\rho \triangleq P/T$. Since P is the total transmit energy spent in a coherence interval T and the noise variance is 1, ρ has the interpretation of average transmit signal-to-noise-ratio (SNR) and is therefore dimensionless. For this simpler case, closed-form solutions on the training length for both high SNR and low SNR scenarios are obtained. The results are stated in the following proposition.

Proposition 4. Assume that $p_t = p_u = \rho$. If $\rho \gg 1$, the optimal length of the training interval is $T_t = K$. If $\rho \ll 1$, the optimal length of the training interval is $T_t = T/2$.

Proof. To facilitate the proof, we denote the length of the data transmission phase as T_u. Thus $T_u \triangleq T - T_t$. Then rewriting $\bar{R}_{k,\text{approx}}$ yields

$$\bar{R}_{k,\text{approx}} = \frac{T_u}{T}\log_2\left[\frac{1}{\left[1+\rho\left(T-T_u\right)\beta_k\right]^2}\sum_{i=1}^{M}\frac{\rho^2(T-T_u)^2\beta_k^3 v_{ki} + \rho\left(T-T_u\right)\beta_k^2}{\frac{1}{\rho}+\sum_{l=1}^{K}\frac{[1+2\rho(T-T_u)\beta_l]\beta_l v_{li} - \rho(T-T_u)\beta_l^2}{[1+\rho(T-T_u)\beta_k l]^2}} + 1\right]. \tag{36}$$

Next we show that if $\rho \gg 1$, $\bar{R}_{k,\text{approx}}$ is an increasing function of T_u by proving that $\bar{R}'_{k,\text{approx}} \triangleq \mathrm{d}\bar{R}_{k,\text{approx}}/\mathrm{d}T_u > 0$. Under the aforementioned SNR condition, $\bar{R}_{k,\text{approx}}$ can be approximated as

$$\bar{R}_{k,\text{approx}} \approx \frac{T_u}{T}\log_2\left[\sum_{i=1}^{M}\frac{\rho\left(T-T_u\right)\beta_k v_{ki}}{T-T_u-K+2\sum_{k=1}^{K}v_{ki}} + 1\right]. \tag{37}$$

Differentiating $\bar{R}_{k,\text{approx}}$ in (37) over T_u yields

$$\bar{R}'_{k,\text{approx}} \approx \frac{1}{T}\log_2\left[\rho\sum_{i=1}^{M}\frac{\left(T-T_u\right)\beta_k v_{ki}}{T-T_u-K+2\sum_{k=1}^{K}v_{ki}} + 1\right] - \frac{l_k}{T}, \tag{38}$$

where

$$l_k \triangleq \frac{\log_2\left(\mathrm{e}\right)T_u}{T-T_u}\left(\sum_{i=1}^{M}\frac{v_{ki}}{T-T_u-K+2\sum_{k=1}^{K}v_{ki}}\right)^{-1}\sum_{i=1}^{M}\frac{(2\sum_{k=1}^{K}v_{ki}-K)v_{ki}}{\left(T-T_u-K+2\sum_{k=1}^{K}v_{ki}\right)^2}. \tag{39}$$

We have $\bar{R}'_{k,\mathrm{approx}} > 0$ because

$$\rho > \frac{2^{l_k}-1}{(T-T_u)\beta_k}\left(\sum_{i=1}^{M}\frac{v_{ki}}{T-T_u-K+2\sum_{k=1}^{K}v_{ki}}\right)^{-1}.$$

This shows that $\bar{R}_{k,\mathrm{approx}}$ is an increasing function of T_u. Thus $\bar{R}_{\mathrm{sum,approx}} = \sum_{k=1}^{K}\bar{R}_{k,\mathrm{approx}}$ is also an increasing function of T_u. To maximize $\bar{R}_{\mathrm{sum,approx}}$, T_u takes its maximum value, and equivalently T_t takes its smallest value which is K.

Next, we show that if $\rho \ll 1$, $T_u = T/2$ is the optimal solution. Under this low SNR condition, expanding the sum-rate (36) in a Taylor series for ρ, we have

$$\bar{R}_{k,\mathrm{approx}} \approx \bar{R}_{k,\mathrm{approx}}\big|_{\rho=0} + \frac{\mathrm{d}\bar{R}_{k,\mathrm{approx}}}{\mathrm{d}\rho}\bigg|_{\rho=0}\rho + \frac{1}{2}\frac{\mathrm{d}^2\bar{R}_{k,\mathrm{approx}}}{\mathrm{d}\rho^2}\bigg|_{\rho=0}\rho^2 = \frac{T_u(T-T_u)}{T}2\log_2(\mathrm{e})\,M\beta_k^2. \quad (40)$$

Notice that T_u only appears in the coefficient in front of the log-function. It is straightforward to see that the optimal solution for T_u is $T/2$. This completes the proof.

Proposition 4 provides useful insights for the high SNR and low SNR scenarios. At high SNR, the smallest training length, K, is enough to have good channel estimation. On the other hand, when the SNR is low, longer training period is needed to improve the channel estimation quality for high spectral efficiency in data transmission. In this case, half of the coherence interval should be devoted to training.

4 Numerical results

In this section, the achievable sum-rate analysis and the resource allocation solution presented in Section 3 are validated through Monte-Carlo simulations. We consider a hexagonal cell with a radius of 1000 m and assume a distance-based path-loss model $\beta_k = z_k\left(r_k/r_h\right)^{\nu}$, where z_k is a log-normal random variable with standard deviation 8 dB, r_k is the distance between the kth UT and the BS, and $\nu = 4$ is the path-loss exponent. We distribute $K = 5$ UTs uniformly in the cell and assume that the BS-UT distance is no smaller than $r_h = 100$ m. The typical outdoor wireless propagation environment is considered, where the channel PAS can be modeled as the truncated Laplacian distribution [15, 24]. The channel PAS is

$$S_k(\theta) = \frac{1}{\sqrt{2}\sigma_k(1-\exp(-\sqrt{2}\pi/\sigma_k))}\cdot\exp\left(\frac{-\sqrt{2}|\theta-\theta_k|}{\sigma_k}\right),\ \text{for } \theta\in[\theta_k-\pi,\theta_k+\pi], \quad (41)$$

where σ_k and θ_k represent the channel angular spread (AS) and the mean angle of arrival (AoA) of the kth UT channel, respectively. We assume that channel ASs are the same for all UTs, so that $\sigma_k = \sigma$. The channel covariance matrices of the UTs are generated according to the model given by (2). We set $\sigma = 0.175$ and the mean AoAs of the UTs from UT 1 to UT 5 are $[-0.7812, -0.4639, 0.0982, 0.6952, 0.9737]$ in radian.

We first validate the tightness of the asymptotic approximate achievable rate and the use of Gamma distribution to approximate the SINR. Figure 1 plots (1) the simulated achievable rate of each user in (12), (2) the analytical approximate achievable rate of each user in (24) where Gamma distribution approximation is used, and (3) the asymptotic closed-form achievable rate in (25). We can see that the derived asymptotic closed-form approximation tightly matches the simulated achievable sum-rate, even when M is as small as 40. The approximation gets tighter as M increases. The figure also validates the use of Gamma distribution to approximate the received SINR in the achievable rate calculation.

To show the advantage of the proposed optimal power allocation, in Figure 2, the cumulative distribution of the sum-rate obtained from 2000 snapshots of large-scale fading are shown for equal power allocation and the proposed optimal power allocation, where the training length is set to be $T_t = K$. As expected, the optimal power allocation improves the sum-rate by about 3 bits/s/Hz. The corresponding ratio of the optimal training power to the optimal data power for $M = 50$ and $M = 100$ is shown in

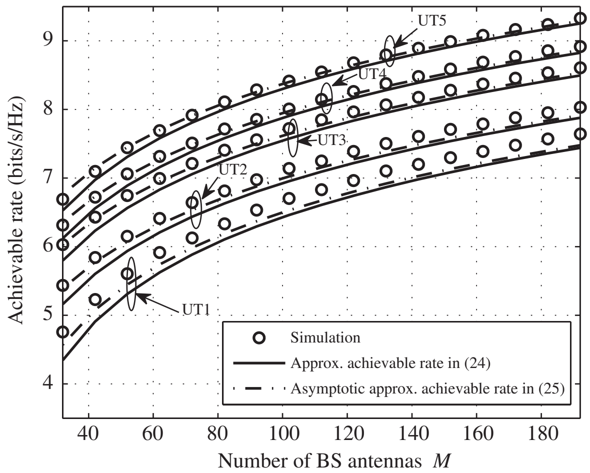

Figure 1 Achievable rates for each UT versus the number of BS antennas M for $T = 200$, $T_t = K = 5$ and $p_t = p_u = 10$ dB.

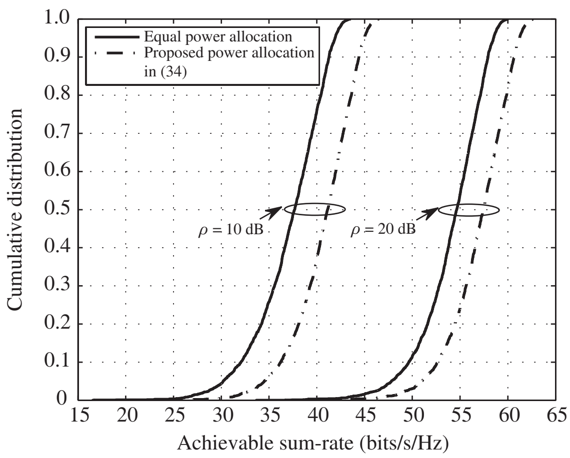

Figure 2 The cumulative distribution function of the achievable sum-rate with equal power allocation and optimizing power allocation for $K = 5$, $M = 128$, $T = 200$ and $T_t = K$.

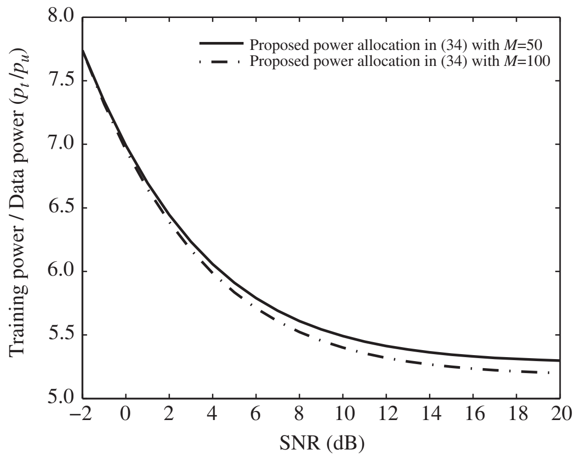

Figure 3 Ratio of the training power to the data power for $K = 5$, $T = 200$ and $T_t = K$.

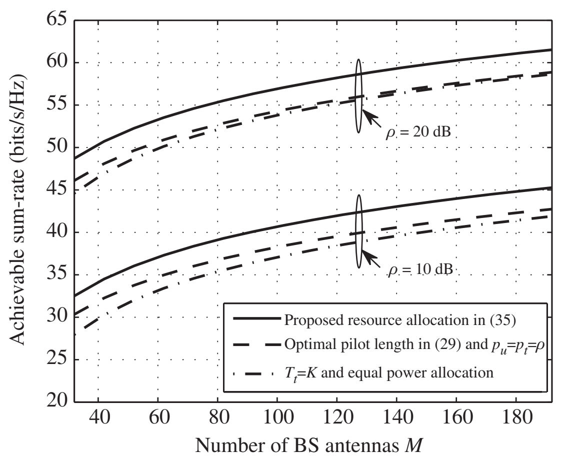

Figure 4 The achievable sum-rate versus the number of antennas M, where $K = 5$ and $T = 200$.

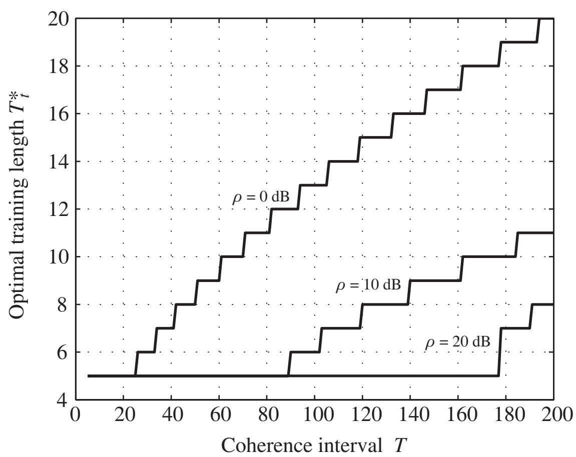

Figure 5 The optimal training length as a function of the length of coherence interval T, where $p_t = p_u = \rho$, $K = 5$, and $M = 128$.

Figure 3. We can see that at low SNR, we spend more power during the training phase, and vice versa at high SNR.

In Figure 4, the system sum-rates are shown for the following cases: (1) the optimal solution of problem (35), (2) the optimal pilot length in (29) and equal power allocation, and (3) the pilot length is K and equal power allocation. We can see that the joint optimization of training length and training power can achieve considerably higher sum-rate than the uplink transmission with equal power and $T_t = K$, and improves the sum-rate by about 4 bits/s/Hz.

Figure 5 displays the optimal training length that maximizes the asymptotic approximate sum-rate for different lengths of coherence interval. The training power and the data transmission power are set to be the same, i.e., $p_t = p_u = \rho$. The number of BS antennas is set as 128. We see that as the SNR ρ decreases, the optimal training length increases. When the SNR is set to be 20 dB, $T_t = K = 5$ is the optimal training length as the coherence interval changes from 5 to 178.

5 Conclusion

This paper addressed the optimal resource allocation problem for the pilot-aided single-cell multi-user massive MIMO system uplink to jointly select the training length and the training power for given coherence interval and total energy budget during the coherence interval. Under angular domain representation of the channels, a tight asymptotic achievable sum-rate was derived in closed-form. Then the separate training interval and training power optimization problems that maximize the derived sum-rate were shown to be concave, based on which a low-complexity alternative optimization algorithm for the joint design was proposed. For the special case that the training power and data transmission power are the same, the optimal training lengths were show to be the same as the UT number for asymptotically high SNR and equals half the coherence interval for asymptotically low SNR. Numerical results shown appreciable benefits of the proposed optimal resource allocation.

Acknowledgements This work was supported by National High-tech R&D Program of China (863) (Grant Nos. 2015AA011305, 2014AA01A704), National Natural Science Foundation of China (Grant Nos. 61320106003, 61521061), National Science and Technology Major Project of China (Grant No. 2015ZX03001035-002), Program for Jiangsu Innovation Team, Natural Science Foundation through the Jiangsu Higher Education Institutions of China (Grant No. 15KJB510025), and Natural Science Foundation Program through Jiangsu Province of China (Grant No. BK20150852). We would like to sincerely thank Dr. JING Yindi at the Department of Electrical and Computer Engineering of the University of Alberta for helpful discussion and suggestion.

Conflict of interest The authors declare that they have no conflict of interest.

References

1 You X H, Pan Z W, Gao X Q, et al. The 5G mobile communication: the development trends and its emerging key techniques (in Chinese). Sci Sin Inform, 2014, 44: 551–563
2 Marzetta T L. Noncooperative cellular wireless with unlimited numbers of base station antennas. IEEE Trans Wirel Commun, 2010, 9: 3590–3600
3 Ngo H Q, Marzetta T L, Larsson E G. Energy and spectral efficiency of very large multiuser MIMO systems. IEEE Trans Commun, 2013, 61: 1436–1449
4 Ashikhmin A, Marzetta T L. Pilot contamination precoding in multi-cell large scale antenna systems. In: Proceedings of IEEE International Symposium on Information Theory, Cambridge, 2012. 1137–1141
5 Gopalakrishnan B, Jindal N. An analysis of pilot contamination on multi-user MIMO cellular systems with many antennas. In: Proceedings of IEEE 12th International Workshop on Signal Processing Advances in Wireless Communications, San Francisco, 2011. 381–385
6 Hoydis J, Brink S, Debbah M. Massive MIMO in the UL/DL of cellular networks: how many antennas do we need? IEEE J Sel Areas Commun, 2013, 31: 160–171
7 Tse D, Viswanath P. Fundamentals of Wireless Communication. London: Cambridge University Press, 2009
8 Yin H F, Gesbert D, Filippou M, et al. A coordinated approach to channel estimation in large-scale multiple-antenna systems. IEEE J Sel Areas Commun, 2013, 31: 264–273

9 Adhikary A, Nam J, Ahn J Y, et al. Joint spatial division and multiplexing-the large-scale array regime. IEEE Trans Inf Theory, 2013, 59: 6441–6463
10 Sun C, Gao X Q, Jin S, et al. Beam division multiple access transmission for massive MIMO communications. IEEE Trans Commun, 2015, 63: 2170–2184
11 Hassibi B, Hochwald B M. How much training is needed in multiple-antenna wireless links? IEEE Trans Inf Theory, 2003, 49: 951–963
12 Ngo H Q, Marzetta T L, Larsson E G. Massive MIMO with optimal power and training duration allocation. IEEE Wirel Commun Lett, 2014, 3: 605–608
13 Guo K F, Guo Y, Fodor G, et al. Uplink power control with MMSE receiver in multi-cell MU-Massive-MIMO systems. In: Proceedings of IEEE International Conference on Communications, Sydney, 2014. 5184–5190
14 You L, Gao X Q, Xia X G, et al. Pilot reuse for massive MIMO transmission over spatially correlated Rayleigh fading channels. IEEE Trans Wirel Commun, 2015, 14: 3352–3366
15 Pedersen K I, Mogensen P E, Fleury B H. A stochastic model of the temporal and azimuthal dispersion seen at the base station in outdoor propagation environments. IEEE Trans Wirel Commun, 2000, 49: 437–447
16 Biguesh M, Gershman A B. Training-based MIMO channel estimation: a study of estimator tradeoffs and optimal training signals. IEEE Trans Signal Process, 2006, 54: 884–893
17 Zhang X L, Matthaiou M, Coldrey M, et al. Impact of residual transmit RF impairments on training-based MIMO systems. IEEE Trans Commun, 2015, 63: 2899–2911
18 Moon T K, Stirling W C. Mathematical Methods and Algorithms for Signal Processing. Upper Saddle River: Prentice Hall, 2000
19 Li P, Paul D, Narasimhan R, et al. On the distribution of SINR for the MMSE MIMO receiver and performance analysis. IEEE Trans Inf Theory, 2006, 52: 271–286
20 Boyd S, Vandenberghe L. Convex Optimization. London: Cambridge University Press, 2004
21 Gao F F, Zhang R, Liang Y C, et al. Design of learning based MIMO cognitive radio systems. IEEE Trans Veh Technol, 2010, 59: 1707–1720
22 Ngo H Q, Marzetta T L, Larsson E G. Analysis of the pilot contamination effect in very large multicell multiuser MIMO systems for physical channel models. In: Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Prague, 2011. 3464–3467
23 Chong E K P, Zak S H. An Introduction to Optimization. 3rd ed. Hoboken: Wiley-Interscience, 2008
24 Cho Y S, Kim J, Yang W Y, et al. MIMO-OFDM Wireless Communications with MATLAB. Singapore: John Wiley & Sons (Asia) Pte Ltd., 2010