On the Energy Efficiency of Multi-Cell Massive MIMO With Beamforming Training

SI-NIAN JIN\(^1\), DIAN-WU YUE\(^1\), (Senior Member, IEEE), AND HA H. NGUYEN\(^2\), (Senior Member, IEEE)

\(^1\)College of Information Science and Technology, Dalian Maritime University, Dalian 116026, China
\(^2\)Department of Electrical and Computer Engineering, University of Saskatchewan, Saskatoon, SK S7N 5A9, Canada

Corresponding author: Dian-Wu Yue (dwyue@dlmu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61801074 and Grant 61371091.

ABSTRACT
This paper is concerned with a multi-cell downlink (DL) massive multiple-input multiple-output (MIMO) system operating over spatially correlated Rician fading channels. We not only consider estimating channel state information (CSI) at the base station (BS), but also adopt beamforming training (BT) to obtain CSI at the users. With maximum-ratio transmission (MRT) or zero-forcing (ZF) employed at the BS to process the transmit signals, the paper derives closed-form expressions of the sum spectral efficiency for both cases: with and without BT. Based on the obtained closed-form expressions, we investigate the effect of the DL pilot length on the system performance for MRT and ZF precoding and with or without BT. Moreover, when the DL pilot length falls within different ranges, we find out whether using BT leads to better system performance. To address the energy efficiency (EE) maximization problem under the constraints of a given sum spectral efficiency and a maximum total DL transmit power, we transform the problem into a geometric program (GP), which can be solved more efficiently. In particular, we develop one iterative power allocation algorithm for the system with BT scheme. Simulation and numerical results demonstrate that the proposed power allocation algorithm can improve the system EE. Numerical results also show that when the BS uses MRT precoding, the sum spectral efficiency in the high signal-to-noise ratio (SNR) region is much improved with BT.

INDEX TERMS
Massive MIMO, spatially correlated Rician fading, beamforming training, energy efficiency.

I. INTRODUCTION
Multiple-input multiple-output (MIMO) systems equipped with a very large number of antennas, commonly referred to as massive MIMO, or large-scale MIMO, have been recognized as an important technology for current and future generations of wireless communications. The use of a large number of antennas enables energy concentration in a much narrower beam, which leads to much improved received signal strength as well as other advantages, such as improved energy and spectrum efficiencies, and the elimination of thermal noise and fading effects\([1]\)–\([5]\). In massive MIMO, the channel vectors are nearly pairwise orthogonal, and hence, linear signal processing schemes, such as maximum ratio combining/maximum ratio transmission (MRC/MRT), equal-gain combining/equal-gain transmission (EGC/EGT), zero-forcing (ZF), and minimum mean-square error (MMSE) are nearly optimal\([1]\)–\([9]\).

With respect to the uplink (UL) and downlink (DL) signal processing stages of a massive MIMO system, channel state information (CSI) at the base station (BS) and user terminals usually plays a key role in maximizing the network throughput\([10]\). As an alternative to frequency-division duplex (FDD) mode, time-division duplex (TDD) mode is more suitable for large-scale MIMO systems. An important and distinctive feature of a TDD system is channel reciprocity, where the reverse channel can be used as an estimate of the forward channel. For a DL massive MIMO system, the BS usually performs DL precoding based on the CSI obtained by estimating the UL channel\([1]\),\([2]\),\([10]\). Since the length of DL pilots is proportional to the number of antennas at the BS, it would be impractical to acquire CSI for the users in a DL massive MIMO system based on channel reciprocity.
Fortunately, references [11]–[19] come up with a practical solution, called beamforming training (BT), that allows users to obtain DL CSI in a massive MIMO system. In particular, by using BT, the length of DL pilots is determined by the number of users and independent of the number of antennas at the BS. In [11], [12], the rate performance of a single-cell multiuser massive MIMO system with the BT scheme has been studied and [11] proposes a resource allocation strategy to maximize the spectral efficiency for MRT or ZF precoder. In [13], [14], the authors study a multicell multiuser massive MIMO system under Rayleigh fading using BT and derive achievable rate expressions when the BS uses MRT and ZF processing. In addition, [14] designs a method called BT and pilot contamination precoding (BT-PCP) scheme to alleviate the effect of pilot contamination. Simulation results also show that, compared with the conventional PCP method, the proposed BT-PCP scheme can improve the spectral efficiency to the extent that it can closely approach the spectral efficiency under the case of perfect CSI. Considering that Rician fading is a more general fading model than Rayleigh fading, the performance of a multicell multiuser massive MIMO system over Rician fading using BT has been obtained in [15]. For a distributed large-scale MIMO system operating over Rayleigh fading channels, references [16], [17] investigate the benefits of BT scheme and study how the number of antennas and the length of coherence interval affect the spectral efficiency of this system. Based on analytical results, the authors also find that compared to a co-located massive MIMO system, the BT scheme works much more effectively for a distributed massive MIMO system. In [18], [19], the authors mainly focus on the spectral efficiency of a cell-free massive MIMO system with BT scheme, and propose max-min fairness power control and greedy pilot assignment methods to promote spectral efficiency of such a system.

Previous research works on massive MIMO systems consider uncorrelated Rician fading channels [20], [21], which do not take into account the effect of channel correlation commonly present in practical wireless propagation environments. Under spatially uncorrelated Rician fading channels, the rate performance in single-cell and multi-cell UL massive MIMO systems with MRC processing is studied in [20] and [21], respectively. However, in reality, a MIMO channel is generally correlated because the antennas are not sufficiently well separated or the propagation environment does not offer rich enough scattering. Considering the finite number of scattering clusters [22], recent papers study massive MIMO under a more realistic assumption of spatially correlated Rician fading [23]–[26]. Specifically, references [23] and [24] study performance of multi-cell massive MIMO systems and derive the statistical properties of the MMSE, element-wise MMSE (EM-MMSE), and least-square (LS) channel estimates when the BS uses MRC and MRT to perform UL detection and DL precoding, respectively. For other linear processing methods, such as multi-cell MMSE (MMMSE) and linear MMSE (LMMSE) processing, references [25] and [26] derive closed-form asymptotic approximations of the UL spectral efficiency and show how the line-of-sight (LOS) components and pilot contamination affect the overall system performance in multi-cell UL massive MIMO systems.

The above works were mainly concerned with the rate performance of a massive MIMO system. On the other hand, given the need to conserve power consumption and environmental concern, energy efficiency (EE) has become a critical design metric for green communication systems. Therefore, EE optimization is a highly-relevant and important problem for massive MIMO systems, as evidenced in a variety of works on EE optimization research [27]–[31]. For a massive MIMO full-duplex relay system, the authors in [27] propose a power allocation algorithm between the source and relay to maximize the EE, subject to a desired sum spectral efficiency and peak power constraints. For a MIMO two-way relay networks with simultaneous wireless information and power transfer, the authors in [28] present two optimization algorithms to maximize the worst-case EE: an iterative algorithm based on the weighted minimum mean-square error method, and a channel diagonalization algorithm based on the generalized singular value decomposition. For a multi-cell massive MIMO network, an algorithm is proposed in [29] to optimize the number of activated antennas together with power allocation and pilot assignment. Numerical results in [29] show that, compared to the conventional pilot assignment scheme, the proposed algorithm significantly improves both the sum spectral efficiency and EE.

Different from the majority of existing works, [13]–[15] consider the application of BT in a multi-cell massive MIMO system over Rayleigh or Rician fading channels. On the other hand, [23]–[26] examine the rate performance of a multi-cell massive MIMO system under spatially correlated Rician fading channels, but without BT. Given that there is only a limited amount of works on the application of BT for multi-cell massive MIMO systems over spatially correlated Rician fading channels, this paper shall carry out an extensive study on whether BT is beneficial to such systems. While the works in [13]–[15] and [23]–[26] mainly focus on studying the system’s spectral efficiency and do not consider EE, our paper also investigates the EE optimization problem for a multi-cell massive MIMO system over spatially correlated Rician fading channels with BT scheme. In summary, the main contributions of this paper are as follows:

- For the multi-cell DL massive MIMO system operating over spatially correlated Rician fading channels, when the BS uses MRT or ZF precoding to process DL transmission of data and pilots, we derive closed-form expressions of the sum spectral efficiency for both cases, with and without BT.
- Making use of the Bernoulli’s inequality and based on the obtained expressions of the sum spectral efficiency, we find out the relationship between the length of the DL pilot and the system’s performance in both cases of whether or not BT is employed for MRT or ZF precoding. The benefit of whether to implement BT is analyzed and shown to depend on the length of the DL pilot.
For the multi-cell DL massive MIMO system with BT scheme, we develop an optimal power allocation algorithm for assigning DL transmit power in order to maximize the EE for a desired sum spectral efficiency and a given total maximum DL transmit power. By using this algorithm, we can transform the original challenging optimization problem into geometric programs (GPs), which can then be reformulated as a convex problem and solved using standard convex optimization tools. Numerical results demonstrate that, compared to the conventional uniform power allocation, the proposed optimal power allocation algorithm significantly improve the EE of a massive MIMO system.

Notation: Matrices and vectors are represented by boldface upper-case and lower-case letters, respectively. We use trace (·), (·)*, (·)T and (·)H to denote trace, conjugate, transpose and conjugate-transpose operations of matrices, respectively. \( \mathbf{I}_N \) denotes an \( N \times N \) identity matrix and \( [\mathbf{A}]_{mn} \) gives the \((m,n)\) th entry of \( \mathbf{A} \). The smallest element, largest element, real parts, expectation, variance, and covariance operators are denoted by \( \min\{\cdot\}, \max\{\cdot\}, \Re\{\cdot\}, \mathbb{E}\{\cdot\}, \Var\{\cdot\} \) and \( \text{Cov}\{X,Y\} \), respectively. In addition, \( \delta(n,j) \) is the delta function, defined as \( \delta(n,j) = 1 \) if \( n = j \), and \( \delta(n,j) = 0 \) if \( n \neq j \). Finally, we use \( \mathbf{z} \sim \mathcal{CN}(\mathbf{a}, \mathbf{A}) \) to denote a complex Gaussian vector \( \mathbf{z} \) with mean vector \( \mathbf{a} \) and covariance matrix \( \mathbf{A} \).

II. SYSTEM MODEL
A. CHANNEL MODEL
We consider a massive MIMO system with \( L \) cells where each cell consists of a BS equipped with \( M \) equally-spaced, omnidirectional antennas that serves \( N \) randomly located user equipments. For the user equipments (e.g. micro mobile devices, mobile phones and tablets), they usually can only configure a few antennas due to their own space and size. Similar to only considering mono-antenna user equipment in [32], [33], we also assume that each user equipment is configured with a single antenna in this paper. The matrix \( \mathbf{H}_{il} = [\mathbf{h}_{i1l}, \ldots, \mathbf{h}_{iNl}] \) stands for the MIMO channel matrix between the BS in the \( i \)th cell and the \( N \) users in the \( l \)th cell, where \( \mathbf{h}_{iln} \) denotes the \( nth \) column vector of \( \mathbf{H}_{il} \). The correlated Rician fading channels are modeled such that the channel from the \( nth \) user equipment in the \( l \)th cell to the BS in the \( i \)th cell is represented as [23]–[26]

\[
\mathbf{h}_{iln} = \sqrt{\beta_{iln}} \left( \mathbf{1} \sqrt{1/k_{iln}} \mathbf{g}_{iln} + \sqrt{1/k_{iln}} \tilde{\mathbf{g}}_{iln} \right),
\]

where \( \beta_{iln} \) accounts for the large-scale channel fading effect from the \( nth \) user in the \( l \)th cell to the BS in the \( i \)th cell, whereas the second term in parentheses represents the small-scale channel fading effect. In essence, \( \beta_{iln} \) models the geometric attenuation and shadow fading, which is assumed constant over many coherence time intervals and known \textit{a priori} [1]. In addition, the term accounting for small-scale channel fading effect from the \( nth \) user in the \( l \)th cell to the BS in the \( i \)th cell consists of a Rayleigh-distributed random component \( \mathbf{g}_{iln} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_M) \) to represent the scattered (Non-LOS) signal and a deterministic component \( \tilde{\mathbf{g}}_{iln} \) to capture the specular (LOS) signal. The quantity \( k_{iln} \geq 0 \) is the Rician \( K \)-factor that represents the power ratio of the LOS and Non-LOS components from \( nth \) user in the \( l \)th cell to the BS in the \( i \)th cell. Furthermore, the channel correlation matrix from the \( nth \) user in the \( l \)th cell to the BS in the \( i \)th cell is represented as \( \mathbf{\Delta}_{iln} \). Due to the movement of users, it may cause the effect of phase-shift of the LOS component. But the phase shift will be identical for all BS antennas, and may therefore be accurately tracked in practice. Therefore, we still use the way of [23]–[26] to build the channel model between the BS and the user without considering the effect of phase-shift, (i.e. a LOS path with static phase). Finally, for notational convenience, we define:

\[
\mathbf{R}_{iln} = \beta_{iln} / (1 + k_{iln}) \mathbf{\Delta}_{iln},
\]

\[
\tilde{\mathbf{h}}_{iln} = \sqrt{k_{iln}/(1 + k_{iln})} \tilde{\mathbf{g}}_{iln}.
\]

Therefore, we have \( \mathbf{h}_{iln} \sim \mathcal{CN}(\tilde{\mathbf{h}}_{iln}, \mathbf{R}_{iln}) \), where \( \tilde{\mathbf{h}}_{iln} \) describes the LOS component and \( \mathbf{R}_{iln} \) is a positive semi-definite covariance matrix representing the spatial correlation characteristics of the NLOS component.

B. MMSE CHANNEL ESTIMATION
Each BS needs to acquire CSI for signal processing. Therefore, \( tp \) samples are reserved for performing UL pilot-based channel estimation in each coherence block. Assume that during the channel estimation phase, users of different cells simultaneously transmit the same set of UL orthogonal pilot sequences of \( \tau_p \) symbols, which can be stacked in a \( N \times \tau_p \) matrix \( \sqrt{\tau_p} \mathbf{C} \), which satisfies \( \mathbf{C}^H = \mathbf{I}_N \). The received pilot signal \( \mathbf{Y}_p^l \in \mathbb{C}^M \times \tau_p \) at the BS of the \( i \)th cell is [23]–[25]

\[
\mathbf{Y}_p^l = \sum_{i=1}^{L} \sum_{j=1}^{N} \sqrt{p_t/\tau_p} \mathbf{h}_{ilj} \mathbf{c}_j + \mathbf{N}_p^l,
\]

where \( p_t/\tau_p \) is the transmit power of UL pilot sequences, \( \mathbf{c}_j \) is the \( j \)th row vector of UL orthogonal pilot matrix \( \mathbf{C} \) and the entries of receive noise matrix \( \mathbf{N}_p^l \in \mathbb{C}^M \times \tau_p \) are independent and identically distributed (i.i.d.) \( \mathcal{CN}(0, 1) \) random variables. To estimate the channel \( \mathbf{h}_{iln} \), the received pilot signal \( \mathbf{Y}_p^l \) is multiplied by \( \mathbf{c}_l^H/\sqrt{p_t/\tau_p} \) to produce

\[
\mathbf{y}_p^l = \sum_{i=1}^{L} \mathbf{h}_{iln} + \frac{\mathbf{N}_p^l}{\sqrt{\tau_p/p_t}} \mathbf{c}_n^H.
\]

Since the LOS component changes slowly, the BS can estimate the LOS component very accurately with negligible signal overhead. Therefore, \( \mathbf{h}_{iln} \) is assumed to be known at both the BS and the users. Similarly, \( \mathbf{R}_{iln} \) varies slowly compared to the channel coherence time and is also assumed.
to be perfectly known to the BS and the users. Based on the processed received pilot signal in (5), the BS can apply the MMSE estimation method to obtain an estimate of $h_{iln}$ as follows [23]–[25]:

$$\hat{h}_{iln} = \hat{h}_{iln} + \mathbf{R}_{iln} \Phi_{in} (y_{in}^p - \hat{y}_{in}^p),$$  \hspace{1cm} (6)

where $\hat{y}_{in}^p = \sum_{l=1}^{L} \hat{h}_{iln}$ and

$$\Phi_{in} = \text{Cov} \{ y_{in}^p \}^{-1} = \left( \sum_{l=1}^{L} \mathbf{R}_{iln} + \frac{\mathbf{I}_M}{\tau_p p_p} \right)^{-1}. \hspace{1cm} (7)$$

The true and estimated channel vectors are related as $h_{iln} = \hat{h}_{iln} + \varepsilon_{iln}$, where $\varepsilon_{iln}$ is the estimation error vector. With MMSE estimation, the MMSE estimate $\hat{h}_{iln}$ is statistically independent of the estimation error $\varepsilon_{iln}$ and their distributions are given as [23]–[25]

$$\hat{h}_{iln} \sim CN (\hat{h}_{iln}, \Delta_{iln}), \hspace{1cm} (8)$$

$$\varepsilon_{iln} \sim CN (0, \mathbf{B}_{iln}), \hspace{1cm} (9)$$

where

$$\Delta_{iln} = \mathbf{R}_{iln} \Phi_{in} \mathbf{R}_{iln}, \hspace{1cm} \forall i, l, n, \hspace{1cm} (10)$$

$$\mathbf{B}_{iln} = \mathbf{R}_{iln} - \mathbf{R}_{iln} \Phi_{in} \mathbf{R}_{iln}, \hspace{1cm} \forall i, l, n. \hspace{1cm} (11)$$

For notational convenience, we also define

$$\Delta_{iln} = \mathbf{R}_{iln} \Phi_{in} \mathbf{R}_{iln}, \hspace{1cm} \forall i, l, n. \hspace{1cm} (12)$$

### C. DOWNLINK SPECTRAL EFFICIENCY WITH CONVENTIONAL SCHEME (WITHOUT BEAMFORMING TRAINING)

Section II.B explained how the BS can obtain the UL CSI by processing the pilot sequences from the users. Based on the channel reciprocity of the TDD mode, the BS can use the estimated CSI to precode the DL transmission signal as follows. First, the received signal at the $nth$ user of the $ith$ cell is given as [14]

$$y_{in} = \sum_{l=1}^{L} \hat{h}_{iln}^T \mathbf{F}_l \mathbf{P}_l^\frac{3}{2} \mathbf{x}_l + n_{in}, \hspace{1cm} (13)$$

where the diagonal matrix $\mathbf{P}_l = \text{diag} \{ p_{l1}, \ldots, p_{lN} \} \in \mathbb{C}^{N \times N}$ consists of DL data transmit powers for all users in the $lth$ cell, $\mathbf{x}_l = [x_{l1}, \ldots, x_{lN}]^T$ is the vector of transmit data symbols for users in the $lth$ cell with $\mathbb{E} \{ \mathbf{x}_l \} = \mathbf{0}$ and $\mathbb{E} \{ \mathbf{x}_l \mathbf{x}_l^H \} = \mathbf{I}_N$, $n_{in} \sim CN (0, 1)$ represents the receiver’s noise at the $nth$ user of the $ith$ cell, $\mathbf{h}_{iln}$ denotes the $nth$ column vector of $\mathbf{H}_l$ and $\mathbf{F}_l \in \mathbb{C}^{M \times N}$ is a precoding matrix of the $lth$ cell.

Next, we consider two common linear precoding methods for massive MIMO systems, which are MRT processing $\mathbf{F}_l = \mathbf{F}_l^{\text{MRT}}$ and ZF processing $\mathbf{F}_l = \mathbf{F}_l^{\text{ZF}}$. They are defined, respectively, as [15]

$$\mathbf{F}_l^{\text{MRT}} = \alpha_{l,\text{MRT}} \hat{H}_l^H, \hspace{1cm} (14)$$

$$\mathbf{F}_l^{\text{ZF}} = \alpha_{l,\text{ZF}} \hat{H}_l^H (\hat{H}_l^H \hat{H}_l)^{-1}, \hspace{1cm} (15)$$

where $\hat{H}_l = [\hat{h}_{l1}, \ldots, \hat{h}_{lN}]$, and the values of $\alpha_{l,\text{MRT}}$ and $\alpha_{l,\text{ZF}}$ are chosen to satisfy the power constraint at the BS, i.e., $\mathbb{E} \{ \text{trace} (\mathbf{F}_l \mathbf{F}_l^H) \} = 1$ [15]. Therefore, we have

$$\alpha_{l,\text{MRT}} = \frac{1}{\text{trace} (\mathbf{A}_{iln} \delta (n, j))}, \hspace{1cm} (16)$$

$$\alpha_{l,\text{ZF}} = \frac{1}{\text{trace} (\mathbf{B}_{iln}^{-1} \delta (n, j))}. \hspace{1cm} (17)$$

Therefore, the expression in (13) can be expanded as

$$y_{in} = \sqrt{p_{in}} s_{inn} + \sum_{(j,l) \neq (i,n)} \sqrt{p_{lj}} s_{lnj} x_j + n_{in}, \hspace{1cm} (19)$$

where $s_{lnj} = h_{l1n}^T \mathbf{F}_j^{\text{MRT}}$ and $s_{lnj} = h_{l1n}^T \mathbf{F}_j^{\text{ZF}}$ are the effective CSI for MRT and ZF precoding with $\mathbf{F}_j^{\text{MRT}}$ and $\mathbf{F}_j^{\text{ZF}}$ being the $jth$ column vector of $\mathbf{F}_j^{\text{MRT}}$ and $\mathbf{F}_j^{\text{ZF}}$, respectively.

For subsequent analysis, rewrite (19) as [14]

$$y_{in} = \sqrt{p_{in}} \mathbb{E} \{ s_{inn} \} x_n + \sqrt{p_{in}} (s_{inn} - \mathbb{E} \{ s_{inn} \}) x_n + \sum_{(j,l) \neq (i,n)} \sqrt{p_{lj}} s_{lnj} x_j + n_{in}. \hspace{1cm} (20)$$

For MRT or ZF precoding, when each user detects the received signal based only on the statistical CSI, the achievable rate’s lower bound (without BT) for the $nth$ user of the $ith$ cell is given in (21), as shown at the bottom of this page. Substituting (21) into (22), the DL sum spectral efficiency for MRT and ZF precoding without BT is given as [10]

$$C^{\Theta} = \frac{T - \tau_p}{T} \sum_{i=1}^{L} \sum_{n=1}^{N} R_{in}^{\Theta}, \hspace{1cm} (22)$$

where $\Theta \in \{ \text{MRT}, \text{ZF} \}$ corresponds to MRT or ZF precoding, and $T$ is the coherence time of the channel.

### D. DOWNLINK SPECTRAL EFFICIENCY WITH BEAMFORMING TRAINING

For a massive MIMO system operating in the FDD mode, references [34]–[36] investigate the system’s performance.

$$R_{in}^{\Theta} = \log_2 \left( 1 + \frac{p_{in} \mathbb{E} \{ s_{inn} \}^2}{\sum_{(j,l) \neq (i,n)} p_{lj} \mathbb{E} \{ s_{lnj} \}^2 + p_{in} \text{Var} \{ s_{inn} \} + 1} \right). \hspace{1cm} (21)$$
Based on a beam-domain (BD) DL training approach. Different from [34]–[36], we focus on studying performance of a massive MIMO system operating the TDD mode in this paper.

For a conventional DL massive MIMO system operating the TDD mode, the DL transmission is usually divided into two phases: UL pilot training phase and DL data transmission phase. In the UL pilot training phase, users transmit the UL pilot sequences to the BS, and the BS can estimate the UL CSI by processing the received pilot signal. By exploiting the channel reciprocity in TDD setting, the BS can first use the estimated UL CSI to precode the DL transmit data symbols and then transmit these precoding signals to the users in DL data transmission phase. However, the conventional DL transmission has a disadvantage that the DL instantaneous CSI cannot be obtained at the users and this is not conducive to a better access to useful information for the users. In order to overcome this problem, we consider the use of BT scheme, in which a third phase is allocated to allow the BS transmit DL pilot matrix to the users so that the DL instantaneous CSI can be obtained at the users. The specific implementation is that all BSs of different cells transmit the same set of DL orthogonal pilot sequences synchronously during the BT phase (DL pilot training phase), and each user estimates the effective CSI of $s_{l,n,j}$ based on the received DL pilot matrix. We define $\Psi \in \mathbb{C}^{N \times d}$, $d \geq N$ as DL orthogonal pilot sequences with the property that $\Psi$ is pairwise orthogonal, i.e., $\Psi \Psi^H = I_N$. It follows that the received DL pilot matrix at the users of the $i$th cell can be expressed as [14]

$$\mathbf{y}_i = \sum_{l=1}^{L} \sqrt{\tau_d} \mathbf{H}_l^T \mathbf{F}_l \mathbf{P}^H \mathbf{x}_l + \mathbf{n}_i, \quad (23)$$

where $\mathbf{n}_i \in \mathbb{C}^{N \times d}$ is the receiver noise matrix of the $i$th cell and the entries of $\mathbf{n}_i$ are i.i.d. $CN(0,1)$ random variables. Multiplying the received DL pilot signal at users of the $i$th cell $\mathbf{y}_i$ by $\Psi^H$ gives [37]

$$\tilde{\mathbf{y}}_i = \sum_{l=1}^{L} \sqrt{\tau_d} \mathbf{H}_l^T \mathbf{F}_l \mathbf{P}^H \mathbf{x}_l + \tilde{\mathbf{n}}_i, \quad (24)$$

where $\tilde{\mathbf{n}}_i = \mathbf{n}_i \Psi^H$, and the entries of $\tilde{\mathbf{n}}_i$ are i.i.d. $CN(0,1)$ random variables due to the pairwise orthogonality of DL pilots matrix $\Psi$.

Based on (24), the DL pilot vector at the $n$th user of the $i$th cell can be rewritten as

$$\tilde{\mathbf{y}}_{in} = \sum_{l=1}^{L} \sqrt{\tau_d} \mathbf{s}_{ln} \mathbf{P}^H \mathbf{x}_l + \tilde{\mathbf{n}}_{in}, \quad (25)$$

where $\mathbf{s}_{ln} = [s_{ln,1}, \ldots, s_{ln,N}]$ and $\tilde{\mathbf{n}}_{in}$ is the $n$th row of $\tilde{\mathbf{n}}_i$. According to [38], the entries of $s_{ln}$ can be estimated independently and the linear MMSE estimation of $s_{ln}$ can be given

$$\hat{s}_{ln} = \mathbb{E}\{s_{ln}\} + \frac{\mathbb{Cov}\{s_{ln}, \hat{y}_{in,j}\}}{\mathbb{Cov}\{\hat{y}_{in,j}, \hat{y}_{in,j}\}} (\hat{y}_{in,j} - \mathbb{E}\{\hat{y}_{in,j}\}), \quad (26)$$

where $\hat{y}_{in,j}$ is given by

$$\hat{y}_{in,j} = \sum_{l=1}^{L} \sqrt{\tau_d} \mathbf{P}_l \mathbf{s}_{ln}^H + \tilde{\mathbf{n}}_{in,j}, \quad (27)$$

and $\tilde{\mathbf{n}}_{in,j}$ is the $j$th element of $\tilde{\mathbf{n}}_i$.

From (26), we also obtain the estimation error, $\mu_{lnj} = \hat{s}_{lnj} - s_{lnj}$. Depends on the characteristics of linear MMSE estimation, $\hat{s}_{lnj}$ and $\mu_{lnj}$ are uncorrelated, but not independent. It then follows $s_{lnj} = \hat{s}_{lnj} + \mu_{lnj}$ from that (19) can be rewritten as

$$\mathbf{y}_{in} = \sqrt{P_{in}} \hat{s}_{inin} \mathbf{x}_{in} + \sum_{(l,j)\neq(i,n)} \sqrt{P_{lj}} \hat{s}_{lnj} \mathbf{x}_{lj} + \sum_{l=1}^{L} \sum_{j=1}^{N} \sqrt{P_{lj} \mu_{lnj}} \mathbf{x}_{lj} + \mathbf{n}_{in}, \quad (28)$$

Based on (28), when the BS uses MRT or ZF precoding to process the transmitted signal, the DL achievable rate’s lower bound at the $n$th user of the $i$th cell with the use of BT is given in (29), as shown at the bottom of this page. In (29), $\mathbb{E}\{|\mathbf{\mu}_{lnj}\|^2\}$ is the statistical properties of the interference and noise as worst-case uncorrelated additive noise. This similar methodology have been used in [12], [14] for massive MIMO systems.
Substituting (29) into (30), the DL sum spectral efficiency for MRT and ZF precoding with BT can be given as [14]

\[
\hat{C}^\diamond = \frac{T - \tau_p - \tau_d}{T} \sum_{i=1}^{L} \sum_{n=1}^{N} \hat{R}_{in}^\diamond,
\]

(30)

where \( \hat{C}^\diamond \in \{ \text{MRT, ZF} \} \) corresponds to MRT or ZF precoding, and \( T \) is the coherence time of the channel.

Since (29) is not in closed form, it will cause great difficulties in the theoretical derivation. Driven by these reasons, with the processing technique from Lemma 2 of [14], (29) can be approximated by (31), as shown at the bottom of the previous page. Lemma 2 of [14] is a more accurate approximation method when the number of antennas at the BS is large. Substituting (31) into (32), the sum spectral efficiency for MRT and ZF precoding with BT can be approximated as [14]:

\[
\hat{C}^\diamond = \frac{T - \tau_p - \tau_d}{T} \sum_{i=1}^{L} \sum_{n=1}^{N} \hat{R}_{in}^\diamond.
\]

(32)

III. DOWNLINK SPECTRAL EFFICIENCY OF MRT PRECODING: WITH AND WITHOUT BT

By calculating the expectations of each term in (31) and (21), and plugging the obtained results into (32) and (22), the expressions of the sum spectral efficiency for MRT processing, with and without BT, are given in Theorem 1 and Theorem 2, respectively. Based on Theorem 1 and Theorem 2, we obtain three corollaries concerning the relationship between the length of the DL pilot and the rate performance for a massive MIMO system with and without BT.

Theorem 1: For MRT precoding, the sum spectral efficiency with BT is given as

\[
\hat{C}^{\text{MRT}} = \frac{T - \tau_p - \tau_d}{T} \sum_{i=1}^{L} \sum_{n=1}^{N} \log_2 \left( 1 + \frac{\Gamma_1 + \Gamma_4}{\Gamma_2 + \Gamma_3 - \Gamma_4 + 1} \right),
\]

(33)

where

\[
\Gamma_1 = \frac{p_in^2\alpha_{i\text{,MRT}}^2\alpha_{i\text{,inn}}^2}{N},
\]

(34)

\[
\Gamma_2 = \sum_{l=1}^{L} \sum_{j=1}^{L} p_{lj}^2\alpha_{l\text{,MRT}}^2b_{lj}^2,
\]

(35)

\[
\Gamma_3 = \sum_{(l,j)\neq(i,n)} p_{lj}^2\alpha_{l\text{,MRT}}^2|a_{lj}|^2,
\]

(36)

\[
\Gamma_4 = \frac{\tau_d p_{in}^2\alpha_{i\text{,MRT}}^4b_{i\text{,inn}}^2}{L \sum_{l=1}^{L} p_{in}^2\alpha_{l\text{,MRT}}^2b_{i\text{,inn}}^2} + 1,
\]

(37)

with \( a_{lij} \) and \( \alpha_{l\text{,MRT}} \) being given in (16) and (17), respectively, and

\[
b_{lij} = \text{trace}(R_{in}^A A_{ij}^H) + \tilde{h}_{i\text{,inn}}^H A_{ij} \tilde{h}_{i\text{,inn}} + \tilde{h}_{i\text{,inn}}^H R_{in} \tilde{h}_{i\text{,inn}}.
\]

(38)

Proof: See Appendix A.

Following similar steps in the derivation of \( \hat{C}^{\text{MRT}} \), we can derive an expression for the sum spectral efficiency under the MRT precoding and without BT. Omitting the derivation steps, the result is stated in Theorem 2 and shall be compared with the performance under the MRT precoding and with BT.

Theorem 2: Under the MRT precoding and without the use of BT, the sum spectral efficiency of a massive MIMO system is given as [24]

\[
\hat{C}^{\text{MRT}} = \frac{T - \tau_p}{T} \sum_{i=1}^{L} \sum_{n=1}^{N} \log_2 \left( 1 + \frac{\Gamma_1}{\Gamma_2 + \Gamma_3 + 1} \right).
\]

(39)

By comparing Theorem 1 and Theorem 2, we can see that \( \Gamma_4 \) represents the performance improvement brought by BT. But it should be pointed out that using BT incurs a higher pilot cost and results in less DL data transmission. Therefore, when the BS uses the MRT precoding to process the transmitted signal, it is of interest to find out the performance gap between the two cases, with and without the use of BT. Compared with the conventional scheme without BT, the biggest difference is that the scheme with BT can acquire the DL CSI at user terminals by sending the DL pilot. Therefore, the main performance difference between the two schemes is related to the DL pilot length. Below we give three corollaries about the relationship between the DL pilot length and the performance of the two schemes (with and without BT). For example, Corollary 1 and Corollary 2 establish the sufficient (but not necessary) conditions that the length of DL pilot sequences should satisfy so that \( \hat{C}^{\text{MRT}} \geq \hat{C}^{\text{MRT}} \) and \( \hat{C}^{\text{MRT}} \geq \hat{C}^{\text{MRT}} \), respectively. These corollaries are obtained using Bernoulli’s inequality.

Corollary 1: For MRT precoding, the inequality \( \hat{C}^{\text{MRT}} \geq \hat{C}^{\text{MRT}} \) holds if \( \tau_d \) satisfies

\[
N \leq \tau_d \leq \min \left\{ \frac{\min_{l=1}^{L} \left\{ \tau_{\text{in,MRT}}^{\text{lower}} \right\}}{T - \tau_p}, T - \tau_p \right\},
\]

(40)

where

\[
\tau_{\text{in,MRT}}^{\text{lower}} = \frac{p_{in}^2\alpha_{i\text{,MRT}}^4b_{i\text{,inn}}^2}{\sum_{l=1}^{L} p_{in}^2\alpha_{l\text{,MRT}}^2b_{i\text{,inn}}^2} - \frac{\Gamma_1}{\Gamma_1 + \Gamma_5 + \Gamma_6}.
\]

(41)

Proof: See Appendix B.

Thus, compared to the system without BT, Corollary 1 gives the range of \( \tau_d \) when we should adopt BT in DL transmission.

Corollary 2: For MRT precoding, the inequality \( \hat{C}^{\text{MRT}} \geq \hat{C}^{\text{MRT}} \) holds if \( \tau_d \) satisfies

\[
\max \left\{ \frac{\max_{l=1}^{L} \left\{ \tau_{\text{in,MRT}}^{\text{upper}} \right\}}{N}, T - \tau_p \right\} \leq \tau_d \leq T - \tau_p,
\]

(42)

where

\[
\tau_{\text{in,MRT}}^{\text{upper}} = \frac{p_{in}^2\alpha_{i\text{,MRT}}^4b_{i\text{,inn}}^2}{\sum_{l=1}^{L} p_{in}^2\alpha_{l\text{,MRT}}^2b_{i\text{,inn}}^2} \left( T - \tau_p \right) \left( \Gamma_1 + \Gamma_5 + \Gamma_6 \right) \frac{\Gamma_5}{\Gamma_2 + \Gamma_3 + 1},
\]

(43)

\[
\Gamma_5 = \Gamma_2 + \Gamma_3 + 1.
\]

(44)

Proof: See Appendix C.
Similarly, Corollary 2 gives the range of $\tau_d$ when we should choose the conventional scheme without BT in DL transmission. Through Corollary 1 and Corollary 2, we can also find the range of $\tau_d$ that leads to $\bar{C}_{\text{MRT}} = C_{\text{MRT}}$ in some specific cases, which is stated in Corollary 3 below.

**Corollary 3:** For MRT precoding, when the condition $N \leq \tau_{\text{MRT}}^{\text{upper}} \leq \tau_{\text{MRT}}^{\text{lower}} \leq T - \tau_p$ is satisfied, we can find a value of $\tau_d$ within the range of $\tau_{\text{MRT}}^{\text{lower}} \leq \tau_{\text{MRT}} \leq \tau_{\text{MRT}}^{\text{upper}}$ so that $C_{\text{MRT}} = C_{\text{MRT}}$, where $\tau_{\text{MRT}}^{\text{lower}}$ and $\tau_{\text{MRT}}^{\text{upper}}$ are given as

$$
\tau_{\text{MRT}}^{\text{lower}} = \min_{i=1, \ldots, N} \left\{ \tau_{\text{in}, \text{MRT}}^{\text{lower}} \right\},
$$

$$
\tau_{\text{MRT}}^{\text{upper}} = \max_{i=1, \ldots, N} \left\{ \tau_{\text{in}, \text{MRT}}^{\text{upper}} \right\}.
$$

**IV. DOWNLINK SPECTRAL EFFICIENCY OF ZF PRECODING: WITH AND WITHOUT BT**

This section provides the analysis for ZF precoding with and without BT, which parallel the analysis in Section III for MRT precoding. Similar to the case of MRT, based on (22) and (32), we can also obtain expressions for the sum spectral efficiency when ZF precoding is employed with and without BT. The results are stated in Theorem 3 and Theorem 4.

**Theorem 3:** For ZF precoding, the sum spectral efficiency with BT is given as

$$
\bar{C}_{\text{ZF}} = \frac{T - \tau_p - \tau_d}{T} \sum_{i=1}^{L} \sum_{n=1}^{N} \log_2 \left( 1 + \frac{\gamma_1 + \gamma_4}{\gamma_2 + \gamma_3 - \gamma_4 + 1} \right),
$$

where

$$
\gamma_1 = p_{\text{in}, \text{ZF}},
$$

$$
\gamma_2 = \sum_{i=1}^{N} p_{ij} \gamma_{\text{inj}},
$$

$$
\gamma_3 = \sum_{i=1, i \neq j}^{N} p_{ij} \gamma_{\text{inj}},
$$

$$
\gamma_4 = \frac{\tau_d \sum_{l \neq i}^{l} \tilde{h}_{n}^{l} \gamma_{\text{inj}}^{2} + 1}{\sum_{l \neq i}^{l} \tilde{h}_{n}^{l} \gamma_{\text{inj}}^{2}},
$$

with $a_{\text{inj}}$, $b_{\text{inj}}$ and $a_{\text{ZF}}$ being given in (16), (38) and (18), respectively, and

$$
c_{\text{inj}} = \text{trace} \left( \tilde{R}_{\text{in}} A_{\text{inj}} \right) + \tilde{h}_{n}^{l} A_{\text{inj}} \tilde{h}_{n}^{l} + \tilde{h}_{n}^{l} R_{\text{in}} \tilde{h}_{n}^{l} + \left[ \tilde{h}_{n}^{l} \tilde{h}_{n}^{l} \right]^{2} + |\text{trace} (A_{\text{inj}})|^2 \delta (n, j) + 2 \text{Re} \left\{ \text{trace} (A_{\text{inj}}) \tilde{h}_{n}^{l} \tilde{h}_{n}^{l} \delta (n, j) \right\}. $$

**Proof:** See Appendix D.

Following similar steps as in Appendix D, we can also derive the sum spectral efficiency for ZF precoding without BT, which is stated in Theorem 4. Due to space limitation, the proof is omitted.

**Theorem 4:** For ZF precoding, the sum spectral efficiency without BT is given as

$$
C_{\text{ZF}} = \frac{T - \tau_p}{T} \sum_{i=1}^{L} \sum_{n=1}^{N} \log_2 \left( 1 + \frac{\gamma_{\text{inj}}}{\gamma_2 + \gamma_3 + 1} \right).$$

Based on (47) and (56) and following similar derivations in Appendices B and C for MRT precoding, we obtain the following corollaries regarding the relationship between the system performance and the length of DL pilot sequences for ZF precoding with and without BT.

**Corollary 4:** For ZF precoding, it is true that $\bar{C}_{\text{ZF}} \geq C_{\text{ZF}}$ if $\tau_d$ satisfies

$$
N \leq \tau_d \leq \min_{i=1, \ldots, L} \left\{ \min_{n=1, \ldots, N} \left\{ \tau_{\text{in}, \text{ZF}}^{\text{lower}} \right\}, T - \tau_p \right\},
$$

where

$$
\tau_{\text{in}, \text{ZF}}^{\text{lower}} = \frac{p_{\text{in}, \text{ZF}}^2 \gamma_{\text{inj}}^2 (T - \tau_p) - \gamma_1}{\sum_{l \neq i}^{l} \sum_{n=1, \ldots, N} p_{\text{in}, \text{ZF}} \tilde{Y}_{\text{inj}}^{l} + p_{\text{in}, \text{ZF}} \gamma_{\text{inj}}^2},
$$

**Corollary 5:** For ZF precoding, it is true that $C_{\text{ZF}} \geq \bar{C}_{\text{ZF}}$ if $\tau_d$ satisfies

$$
\max_{i=1, \ldots, L} \left\{ \max_{n=1, \ldots, N} \left\{ \tau_{\text{in}, \text{ZF}}^{\text{upper}} \right\}, N \right\} \leq \tau_d \leq T - \tau_p,
$$

where

$$
\tau_{\text{in}, \text{ZF}}^{\text{upper}} = \frac{p_{\text{in}, \text{ZF}}^2 \gamma_{\text{inj}}^2 (T - \tau_p) (\gamma_1 + \gamma_5) - \gamma_1 \gamma_5}{\sum_{l \neq i}^{l} \sum_{n=1, \ldots, N} p_{\text{in}, \text{ZF}} \tilde{Y}_{\text{inj}}^{l} \gamma_1 \gamma_5 + p_{\text{in}, \text{ZF}}^2 \gamma_{\text{inj}}^2 \gamma_5 + p_{\text{in}, \text{ZF}} \gamma_1 \gamma_5 \gamma_{\text{inj}}},
$$

and

$$
\gamma_5 = \gamma_2 + \gamma_3 + 1.
$$

For ZF precoding, Corollary 4 and Corollary 5 show that we should use BT under the condition of (57) and should not use it under the condition of (59). Based on Corollary 4 and Corollary 5, when $\tau_d$ satisfies a certain range, a value of $\tau_d$ can be found so that $\bar{C}_{\text{ZF}} = C_{\text{ZF}}$ and this is stated in Corollary 6.

**Corollary 6:** For ZF precoding, when the inequality $N \leq \tau_{\text{ZF}}^{\text{lower}} < \tau_{\text{ZF}}^{\text{upper}} \leq T - \tau_p$ is satisfied, both schemes, with and without BT, have the same DL sum spectral efficiency for a value of $\tau_d$ in the range $\tau_{\text{ZF}}^{\text{lower}} \leq \tau_d \leq \tau_{\text{ZF}}^{\text{upper}}$, where $\tau_{\text{ZF}}^{\text{lower}}$ and $\tau_{\text{ZF}}^{\text{upper}}$ are given as follows

$$
\tau_{\text{ZF}}^{\text{lower}} = \min_{i=1, \ldots, L} \left\{ \min_{n=1, \ldots, N} \left\{ \tau_{\text{in}, \text{ZF}}^{\text{lower}} \right\} \right\},
$$

$$
\tau_{\text{ZF}}^{\text{upper}} = \max_{i=1, \ldots, L} \left\{ \max_{n=1, \ldots, N} \left\{ \tau_{\text{in}, \text{ZF}}^{\text{upper}} \right\} \right\}.
$$

Remark 1: When the channel correlation matrix $\Delta_{in} = 1_M$ for $\forall i, l, n$, the expression in (33), (39), (47) and (56) are identical to the rate expression for uncorrelated Rician fading channels in reference [15]. Therefore the rate expression given in this paper is a more general result.

V. POWER ALLOCATION

In section III and IV, our main focus is on the achievable rate performance (spectral efficiency). Due to the need to conserve power and serious environmental concern, EE is also an aspect that needs attention for green communication systems. Therefore, we design a power allocation algorithm to optimize EE of the system with BT scheme in this section.

A. POWER ALLOCATION WITH BT

We consider that the DL transmit powers of different users are determined. Given the sum spectral efficiency and the constraint on the maximum DL transmit power, we want to find a power allocation algorithm for determining different DL transmit powers of different users in each cell so that the EE is maximized.

The EE (in bits/J) is defined as the sum spectral efficiency divided by the total power consumption in the network. Based on the linear BS power consumption model, the EE of the system for MRT precoding and ZF precoding with BT is given by [29]

$$EE^\diamond_{BT} = \frac{\tilde{C}^\diamond}{LMp_c + p_0 + \sum_{i=1}^{L} \sum_{n=1}^{N} v p_{in}}.$$  \hspace{1cm} (64)

where $\diamond \in \{\text{MRT}, \text{ZF}\}$ corresponds to MRT or ZF precoding, $\tilde{C}^{\text{MRT}}$ and $\tilde{C}^{\text{ZF}}$ are given in (33) and (47) for MRT precoding and ZF precoding with BT, respectively. The quantity $LMp_c$ represents the total circuit power consumption of all antennas at BS (which is independent of the DL transmit power), where $p_c$ is the power consumed by the circuit at each antenna including the power dissipation of filter and mixer, frequency synthesizer, and digital-to-analog converter; $p_0$ is the basic power consumed at the BS which is independent from the number of antennas; $\nu \geq 1$ is the inefficiency of the power amplifier. Without loss of generality, we assume $\nu = 1$ in this paper.

Based on the numerical results of Section VI, we know that the theoretical results of (33) and (47) are close to the simulation results of (30). Therefore, we can use the expressions of (33) and (47) instead of (30) to represent the sum spectral efficiency with BT for MRT precoding and ZF precoding in (64). So given the sum spectral efficiency and the constraint on the maximum DL transmit power, the optimization problem of maximizing EE can be expressed as [27]

$$\begin{align*}
\text{Maximize} & \quad EE^\diamond_{BT} \\
\text{subject to} & \quad \tilde{C}^\diamond = C^\diamond \\
& \quad 0 \leq \sum_{i=1}^{L} \sum_{n=1}^{N} p_{in} \leq LNp,
\end{align*}$$  \hspace{1cm} (65)

where $\tilde{C}^\diamond$ denotes the given sum spectral efficiency and $LNp$ is the peak power constraint of the total DL transmit power.

Take the system using the MRT precoding and BT as an example. Since $LMp_c$ and $p_0$ are both fixed values, the optimal power allocation problem in (65) can be transformed into

$$\begin{align*}
\text{Minimize} & \quad \sum_{i=1}^{L} \sum_{n=1}^{N} p_{in} \\
\text{subject to} & \quad T - \tau_p - Td \sum_{i=1}^{L} \sum_{n=1}^{N} \log_2 (1 + y_{in}) = \tilde{C}^{\text{MRT}}.
\end{align*}$$  \hspace{1cm} (66)

where $y_{in}$ represents the total circuit power consumption of all antennas at BS (which is independent of the DL transmit power), where $p_c$ is the power consumed by the circuit at each antenna including the power dissipation of filter and mixer, frequency synthesizer, and digital-to-analog converter; $p_0$ is the basic power consumed at the BS which is independent from the number of antennas; $\nu \geq 1$ is the inefficiency of the power amplifier. Without loss of generality, we assume $\nu = 1$ in this paper.
\[
\begin{align*}
\gamma_{in} & \leq \frac{\Gamma_1 + \Gamma_4}{\Gamma_2 + \Gamma_3 - \Gamma_4 + 1}, \quad \forall i, n \\
0 & \leq \sum_{i=1}^{L} \sum_{n=1}^{N} p_{in} \leq LN_p. \quad (66)
\end{align*}
\]

By expanding the inequality in (66), (66) can be rewritten as

\[
\begin{align*}
\text{Minimize} \quad & \sum_{i=1}^{L} \sum_{n=1}^{N} p_{in} \\
\text{subject to} \quad & \prod_{i=1}^{L} \prod_{n=1}^{N} (1 + \gamma_{in}) = 2^{\frac{LN_p}{\rho_d}} \\
0 & \leq \sum_{i=1}^{L} \sum_{n=1}^{N} p_{in} \leq LN_p. \quad (68)
\end{align*}
\]

where (67), as shown at the bottom of the previous page. The objective function in Problem (68) is a posynomial function. However, the equality constraint in (68) is not the monomial function and the inequality constraint in (67) is not the posynomial function either. In fact, Problem (68) is a signomial programming [39]. If the equality constraint and the inequality constraint can be transformed into the form of monomial function and posynomial function, respectively, then Problem (68) becomes a GP which can be reformulated as a convex problem.

In Appendix E, we give an explicit transformation method to approximate the equality constraint and the inequality constraint in the forms of monomial function and posynomial function, respectively. By using a similar technique as in [27] and the approximation presented in Appendix E, we formulate Algorithm 1, a novel successive approximation algorithm, to solve Problem (68) for the case of MRT precoding and BT. Specifically, by using this new iterative algorithm (Algorithm 1), we can successfully convert Problem (68) to a GP and solve it with standard convex optimization tools, such as CVX. Using similar approximations in Appendix E, we can also develop a successive approximation algorithm to find the optimized DL transmit power for the system with ZF precoding and BT, which is also included in Algorithm 1.

In addition, it is pointed out that in Algorithm 1, we choose parameters \( \sigma = 1.5 \) and \( \varphi = 0.01 \) to control the approximation accuracy because these values can maintain a good balance between the approximation accuracy and convergence speed [40]. Due to space constraints, some expressions that will be used in Algorithm 1 are given from (69) to (82), as shown at the bottom of the next page.

\[
\sigma_{\text{MRT}}^{\text{MRT}} = p_{in,t} \tau_d a_2 \alpha_{MRT}^2 b_{inn} \left( f_{\text{MRT}} \right)^{-1} \times \left( b_{inn} \gamma_{in,t} + b_{inn} + a_2^{inn} \right), \quad (72)
\]

\[
\sigma_{\text{MRT}}^{\text{MRT}} = p_{in,t} \tau_d a_2 \alpha_{MRT}^2 b_{inn} \left( f_{\text{MRT}} \right)^{-1}, \quad l \neq i, \quad (73)
\]

\[
\sigma_{\text{MRT}}^{\text{MRT}} = \gamma_{in,t} p_{in,t} \tau_d a_2 \alpha_{MRT}^2 b_{inn} \left( f_{\text{MRT}} \right)^{-1}, \quad (74)
\]

\[
\sigma_{\text{MRT}}^{\text{MRT}} = \left( f_{\text{MRT}} \right)^{-1} \sigma_{\text{MRT}}^{\text{MRT}} p_{in,t} \prod_{l \neq i}^{L} \sigma_{\text{MRT}}^{\text{MRT}} p_{in,t}, \quad (75)
\]

VI. NUMERICAL RESULTS

In this section, we consider a multicell massive MIMO system over correlated Rician fading channels (\( L = 4 \) cells). Each cell has one BS which is equipped with \( M = 50 \) antennas and each BS can provide services to \( N = 5 \) users in the cell. We assume the coherence interval \( T = 1000 \), the UL pilot length is \( t_p = N \), and the UL pilot power is \( p_p = 0 \text{dB} \) in this system. For a massive MIMO system over correlated Rician fading channel, the large-scale channel fading coefficient mainly related to the distance between BS and users. Without losing generality, we set all the direct gains to 1 and all cross gains to 0.1 in the simulations, i.e., for \( i, l = 1, \ldots, L \) and \( n = 1, \ldots, N \), the large-scale channel fading coefficient \( \beta_{ln} = 1 \) if \( i = l \) and \( \beta_{ln} = 0 \) if \( i \neq l \). Moreover, to ensure distinct Rician \( K \)-factor between BS and different users, we assume throughout all the simulations that for all BS, the LOS component \( g_{\text{ln}} \) can be expressed as \( g_{\text{ln}} = e^{-\frac{d}{\lambda}} \sin(\theta_{ln}) \), where \( \lambda \) is the carrier wavelength, \( \theta_{ln} \) is the arrival angle to the \( n \)th user from the BS of the \( i \)th cell, and \( d = \lambda/2 \) is the antenna spacing [41]. In addition, the correlation matrix of the NLOS component is constructed as \( \Lambda_{\text{ln}} = \frac{1}{N} \sum_{m=2}^{M} | r_{\text{ln}} |^2 \), where \( r_{\text{ln}} \) is the coefficient related to antenna correlation between the \( n \)th user equipment in the \( i \)th cell and the BS in the \( i \)th cell [42]. For convenience, we also set \( r_{\text{ln}} = r \) for all BS, \( i, l, n \) in this section.

In Fig. 1, we first illustrate the effect of DL signal-to-noise ratio (SNR), i.e., \( p_{in} \), \( \gamma_{in} \), on the sum spectral efficiency
when using MRT and ZF precoding for the systems with and without BT, respectively, validating the conclusions of Theorems 1, 2, 3 and 4. For the system with BT, the simulation results of the spectral efficiency are obtained based on (30), whereas the theoretical results are obtained by (33) and (47) for MRT and ZF precoding, respectively. For the system without BT, the corresponding expressions for simulation and theoretical results are (22), (39) and (56). We set the channel correlation coefficient \( r = 0.7 \) and the length of the DL pilot \( \tau_p = N \). As expected, our closed-form expressions are almost identical with the simulation results over the entire range of SNR. This figure shows that as the DL SNR increases (for both systems, with and without BT), the sum spectral efficiency keeps increasing and gradually approaches fixed values for MRT and ZF precoding. The sum spectral efficiency obtained with ZF precoding is higher than that obtained with MRT precoding. It is also seen that the system performance with BT gradually exceeds the system performance without BT for both MRT and ZF precoding. Compared to ZF precoding, when the BS uses MRT precoding to process signals, using BT can help to improve the sum spectral efficiency. Overall, compared to the system without
Algorithm 1 (MRT and ZF Precoding With BT)

1: **Initialization**: Firstly, set the initial number of iterations $t = 1$, the tolerance value $\varrho = 0.01$, the maximum number of iterations $t_{\text{max}} = 10$ and the parameter $\omega = 1.5$. For $\forall i, n$, choose some initial values of $p_{in,1}$ and $\gamma_{in,1}$ for MRT or ZF precoding.

2: **Iteration $t$**: Secondly, compute (69) to (82). Based on these expressions, solve the following GP:

\[
\begin{align*}
\text{Minimize} & \quad \sum_{i=1}^{L} \sum_{n=1}^{N} p_{in} \\
\text{subject to} & \quad \prod_{i=1}^{L} \prod_{n=1}^{N} \sigma_{in,t} \gamma_{in}^{\hat{p}_{in}} = 2^{T_{C}^{\infty}}
\end{align*}
\]

for MRT or (82) for ZF, $\forall i, n$

\[
0 \leq \sum_{i=1}^{L} \sum_{n=1}^{N} p_{in} \leq \text{LNp}
\]

the inequality (81)

\[
0^{-1} p_{in,t} \leq p_{in} \leq \omega p_{in,t}, \quad \forall i, n
\]

\[
\omega^{-1} \gamma_{in,t} \leq \gamma_{in} \leq \omega \gamma_{in,t}, \quad \forall i, n
\]

(83)

For $\forall i, n$, let $p^{*}_{in}$ and $\gamma^{*}_{in}$ denote the solutions of this iteration.

3: Thirdly, when

\[
\max_{i=1,\ldots,L} \left\| \gamma_{in,t} - \gamma_{in}^{*} \right\|^2 \leq \varrho
\]

and

\[
\max_{n=1,\ldots,N} \left\| p_{in,t} - p_{in}^{*} \right\|^2 \leq \varrho
\]

are both established or $t = t_{\text{max}}$, stop the iteration. Otherwise, go to Step 4.

4: Lastly, set $t = t + 1$, $p_{in,t} = p_{in}^{*}$ and $\gamma_{in,t} = \gamma_{in}^{*}$ for $i = 1, \ldots, L; n = 1, \ldots, N$, and go to Step 2.

---

**FIGURE 1.** Effect of SNR on the sum spectral efficiency: $r = 0.7$ and $\tau_{d} = N$.

**FIGURE 2.** Effect of the DL pilot length on the sum spectral efficiency for MRT and ZF precoding: $r = 0.7$ and $p_{in} = 10$ dB. ($\forall i, n$).

**FIGURE 3.** Effect of the DL pilot length on the sum spectral efficiency for ZF precoding: $r = 0.7$ and $p_{in} = 10$ dB. ($\forall i, n$).

BT, using BT together with MRT precoding is very suitable for the multi-cell DL massive MIMO system operating over spatially correlated Rician fading channels. Given the accuracy of the obtained closed-form expressions, only theoretical results are presented in the remaining of this section.

Fig. 2 and Fig. 3 depict the sum spectral efficiency versus the DL pilot length for MRT and ZF precoding, respectively. Here, we assume $r = 0.7$ and $p_{in} = 10$ dB. ($\forall i, n$). It can be seen that the sum spectral efficiency with BT is better than that without BT if $\tau_{d} \leq \tau_{d}^{\text{lower}}$. On the other hand, the opposite is true if $\tau_{d} \geq \tau_{d}^{\text{upper}}$. The performance cross-over point is located in the range of $\tau_{d}^{\text{lower}} \leq \tau_{d} \leq \tau_{d}^{\text{upper}}$, where $\diamond \in \{\text{MRT}, \text{ZF}\}$ corresponds to MRT or ZF precoding. These numerical results verify the correctness of Corollary 1 to Corollary 6 for MRT and ZF precoding. Finally, we can also see that, compared with MRT precoding, ZF precoding reaches the performance cross-over point at a shorter length of the DL pilot.

Fig. 4 shows that the sum spectral efficiency versus the channel correlation coefficient for MRT and ZF precoding with BT scheme. In Fig. 4, we assume $\tau_{d} = N$.
FIGURE 4. Effect of the channel correlation coefficient on the sum spectral efficiency for MRT and ZF precoding with BT scheme: $\tau_d = N$ and $p_{in} = 10\,$dB, $\forall i, n$.

FIGURE 5. Energy efficiency versus sum spectral efficiency for MRT precoding with BT scheme: $p_c = 10\,$dBm, $p_0 = 10\,$dBm, $r = 0.7$ and $\tau_d = N$.

FIGURE 6. Energy efficiency versus sum spectral efficiency for ZF precoding with BT scheme: $p_c = 10\,$dBm, $p_0 = 10\,$dBm, $r = 0.7$ and $\tau_d = N$.

and $p_{in} = 10\,$dB, $\forall i, n$. As expected, as the channel correlation coefficient increase, the sum spectral efficiency gradually decreases. When $r = 0$ (i.e. $[\Delta_{in}] = I_M$, $\forall i, l, n$) and it is equivalent to the system over uncorrelated Rician fading, the sum spectral efficiency can reach its maximum. And when $r$ is less than 0.6, the spectral efficiency attenuation is not large, but after 0.6, the sum spectral efficiency decreases rapidly. The above phenomenons reflect the impact of channel correlation on spectral efficiency.

While Figs. 1 to 4 focus on the system’s performance in terms of the achievable rate, the next two figures present the system’s performance in terms of EE. Specifically, Figs. 5 and 6 plot the EE versus the sum spectral efficiency for MRT or ZF precoding with BT scheme under uniform and optimal power allocations, respectively. The uniform power allocation corresponds to the scenario in which all DL transmit and pilot powers are set at maximum, i.e., $p_{in} = p$, $(i = 1, \cdots, L; n = 1, \cdots, N)$, where $p$ is the maximum power constraint of each user. For the following two cases: MRT and ZF precoding with BT, the optimal power allocations are obtained by Algorithm 1. In Algorithm 1, the initial values of $p_{in,1}$ are chosen as $p_{in,1} = p$, $(\forall i, n)$. After that we can substitute $p_{in} = p$, $(\forall i, n)$ into $\gamma_{in} = \frac{\gamma_1 + \gamma_4}{\gamma_2 + \gamma_3 - \gamma_4 + 1}$ and $\gamma_{in} = \frac{\gamma_1 + \gamma_4}{\gamma_2 + \gamma_3 - \gamma_4 + 1}$ to obtain the initial values of $\gamma_{in,1}$, $(\forall i, n)$ for MRT or ZF precoding with BT. For both examples, we set $p_c = 10\,$dBm, $p_0 = 10\,$dBm, $r = 0.7$ and $\tau_d = N$. From Figs. 5 and 6, we can see that when the BS implements the optimal power allocation, the EE can be improved significantly as compared to the case of exercising the uniform power allocation. These simulation results therefore demonstrate the significant benefit of our proposed power optimization algorithm.

VII. CONCLUSION

In this paper, we have studied and analyzed the sum spectral efficiency and EE of a multi-cell DL massive MIMO system operating over spatially correlated Rician fading and imperfect CSI. Considering that using BT allows the DL pilot cost only dependent on the number of users and independent of the number of BS antennas, the paper examines both cases of using and not using BT for obtaining CSI at users. We derived closed-form expressions of the sum spectral efficiency achieved by MRT and ZF precoding and utilize them to determine the range of the performance cross-over point for the systems using and not using BT. Numerical results were presented to corroborate the analysis. Furthermore, the obtained expressions enabled us to develop one novel iterative power allocation algorithm to improve the system’s EE with BT scheme. Finally, numerical results showed that as the channel correlation coefficient increases, the sum spectral efficiency drops sharply when channel correlation coefficient is large.
APPENDIXES

APPENDIX A

When the BS uses MRT precoding to process the transmitted signal, for the case of \( i, l = 1, \ldots, L \) and \( n, j = 1, \ldots, N \), the effective CSI is given as
\[
s_{lnij} = h_{lnij}^T p_{MRT} = \alpha_{l,MRT} h_{lnij}^T.
\]
(84)

Base on (84), we can analyze the expectations \( \mathbb{E} \{ |s_{lnij}|^2 \} \) and \( \mathbb{E} \{ |\mu_{lnij}|^2 \} \) in (31). In the following, we have \([23],[24]\)
\[
\mathbb{E} \{ s_{lnij} \} = \alpha_{l,MRT}s_{lnij},
\]
(85)
where \( s_{lnij} \) and \( \alpha_{l,MRT} \) are given in (16) and (17), respectively. Based on (85), we obtain
\[
\mathbb{E} \{ \tilde{y}_{ln,j} \} = \sum_{l=1}^{L} \sqrt{\tau_d p_{l} \alpha_{l,MRT} s_{lnij}}.
\]
(86)

From (85) and (86), we use the method in \([24]\) and obtain
\[
\mathbb{E} \{ s_{lnij} \hat{y}_{ln,j}^n \} = \sqrt{\tau_d p_{l} \alpha_{l,MRT}^2 c_{lnij}} \sum_{l=1}^{L} \sqrt{\tau_d p_{l} \alpha_{l,MRT} s_{lnij} a_{lnij}},
\]
(87)
where \( c_{lnij} \) is given in (52). Substituting (85), (86) and (87) into (88), one has
\[
\text{Cov} \{ s_{lnij}, \hat{y}_{ln,j} \} = \sqrt{\tau_d p_{l} \alpha_{l,MRT}^2 b_{lnij}},
\]
(88)
where \( b_{lnij} \) is given in (38). Similarly, we have
\[
\mathbb{E} \{ |\tilde{y}_{ln,j}|^2 \} = \sum_{l=1}^{L} \tau_d p_{l} \alpha_{l,MRT}^2 c_{lnij} + 1
\]
\[
+ \sum_{l=1}^{L} \sum_{l_1 \neq l} \tau_d p_{l} \alpha_{l,MRT} a_{lnij} a_{l_1,nj}^*,
\]
(89)

Substituting (86) and (89) into (90) yields
\[
\text{Cov} \{ \tilde{y}_{ln,j}, \hat{y}_{ln,j} \} = \sum_{l=1}^{L} \tau_d p_{l} \alpha_{l,MRT} b_{lnij} + 1.
\]
(90)

Substituting (85), (86), (88) and (90) into (26) and \( \mu_{lnij} = s_{lnij} - \hat{s}_{lnij} \), we can obtain the expectations \( \mathbb{E} \{ |\tilde{s}_{lnij}|^2 \} \) and \( \mathbb{E} \{ |\mu_{lnij}|^2 \} \) as follows:
\[
\mathbb{E} \{ |\tilde{s}_{lnij}|^2 \} = |\alpha_{l,MRT} s_{lnij}|^2 + \frac{\tau_d p_{l} \alpha_{l,MRT}^2 b_{lnij}}{\sum_{l_1=1}^{L} \tau_d p_{l} \alpha_{l,MRT} a_{lnij} a_{l_1,nj}^* + 1},
\]
(91)
\[
\mathbb{E} \{ |\mu_{lnij}|^2 \} = \alpha_{l,MRT}^2 b_{lnij} - \frac{\tau_d p_{l} \alpha_{l,MRT}^4 b_{lnij}^2}{\sum_{l_1=1}^{L} \tau_d p_{l} \alpha_{l,MRT} a_{lnij} a_{l_1,nj}^* + 1}.
\]
(92)

Finally, substituting (91) and (92) into (32) leads to the result in Theorem 1.

APPENDIX B

If the inequality \( \bar{C}_{MRT} \geq C_{MRT} \) is satisfied, we first perform the exponential operation and then perform simple algebraic operations on both sides of the inequality. Therefore, the inequality \( \bar{C}_{MRT} \geq C_{MRT} \) can be transformed into the following form
\[
\prod_{i=1}^{L} \prod_{n=1}^{N} \left( 1 + \frac{\Gamma_4}{\Gamma_5 - \Gamma_4} \right) \geq \prod_{i=1}^{L} \prod_{n=1}^{N} \left( 1 + \frac{\Gamma_1 + \Gamma_4}{\Gamma_5 - \Gamma_4} \right),
\]
(93)
where \( \Gamma_1, \Gamma_4 \) and \( \Gamma_5 \) are defined in (34), (37) and (44), respectively.

When the precondition \( N \leq \tau_d \leq T - \tau_p \) is satisfied, we can use Bernoulli’s inequality, i.e., \((1 + x)^r \geq 1 + rx \) for \( x > -1 \) and \( r \geq 1 \) and (93) can be transformed into \([15]\)
\[
\prod_{i=1}^{L} \prod_{n=1}^{N} \left( 1 + \frac{T - \tau_p}{\tau_d} \frac{\Gamma_4}{\Gamma_5 - \Gamma_4} \right) \geq \prod_{i=1}^{L} \prod_{n=1}^{N} \left( 1 + \frac{\Gamma_1 + \Gamma_4}{\Gamma_5 - \Gamma_4} \right).
\]
(94)

Through (94), we know that when the inequality in (95) is satisfied, then the inequality \( \bar{C}_{MRT} \geq C_{MRT} \) must be established. Therefore, (95) as a sufficient (but not necessary) condition of \( \bar{C}_{MRT} \geq C_{MRT} \) can be expressed as
\[
T - \tau_p \geq \frac{\Gamma_4}{\Gamma_5 - \Gamma_4} \geq \frac{\Gamma_1 + \Gamma_4}{\Gamma_5 - \Gamma_4}, \quad \forall i, n.
\]
(95)

Considering the precondition \( N \leq \tau_d \leq T - \tau_p \), we perform a simple algebraic operation on (95) and arrive at the fact that the inequality \( \bar{C}_{MRT} \geq C_{MRT} \) will be true if \( N \leq \tau_d \leq \min \{ \tau_{\text{lower}}_{in,MRT}, T - \tau_p \} \) is satisfied, where \( \tau_{\text{lower}}_{in,MRT} \) is given in (41).

APPENDIX C

For the inequality \( C_{MRT} \geq \bar{C}_{MRT} \), performing the exponential and algebraic operations on both sides of the inequality, we obtain
\[
\prod_{i=1}^{L} \prod_{n=1}^{N} \left( 1 + \frac{\Gamma_1}{\Gamma_5} \right) \frac{T - \tau_p}{\tau_d} \geq \prod_{i=1}^{L} \prod_{n=1}^{N} \left( 1 + \frac{\Gamma_1 + \Gamma_4}{\Gamma_5} \right),
\]
(96)
where \( \Gamma_1, \Gamma_4 \) and \( \Gamma_5 \) are defined in (34), (37) and (44), respectively. By using Bernoulli’s inequality similar to Appendix B, when \( N \leq \tau_d \leq T - \tau_p \), (96) becomes \([15]\)
\[
\prod_{i=1}^{L} \prod_{n=1}^{N} \left( 1 + \frac{T - \tau_p}{\tau_d} \frac{\Gamma_1}{\Gamma_5} \right) \geq \prod_{i=1}^{L} \prod_{n=1}^{N} \left( 1 + \frac{\Gamma_1 + \Gamma_4}{\Gamma_5} \right).
\]
(97)

Based on (97), we know that the inequality (98) is a sufficient (but not necessary) condition of \( C_{MRT} \geq \bar{C}_{MRT} \) and can be given as
\[
\frac{T - \tau_p}{\tau_d} \frac{\Gamma_1}{\Gamma_5} \geq \frac{\Gamma_1 + \Gamma_4}{\Gamma_5}, \quad \forall i, n.
\]
(98)
When (98) and $N \leq \tau_d \leq T - \tau_p$ are simultaneously satisfied, combining the above two conditions and performing simple algebraic operations, the inequality $C^{\text{MRT}} \geq \bar{C}^{\text{MRT}}$ will be established in the case of

$$\max \left\{ \max_{n_1, \ldots, n_N} \left\{ \tau_{\text{in}, \text{MRT}}^{\text{upper}} \right\}, N \right\} \leq \tau_d \leq T - \tau_p,$$

where $\tau_{\text{in}, \text{MRT}}^{\text{upper}}$ is given in (43).

**APPENDIX D**

When the BS uses ZF precoding to process the transmit signal, we define the effective CSI as $s_{\text{inj}} = h_{\text{in}}^T Z F$, $(i, l = 1, \ldots, L; n, j = 1, \ldots, N)$ and also analyze the expectations $E \left\{ |s_{\text{inj}}|^2 \right\}$ and $E \left\{ |s_{\text{inj}}|^2 \right\}$ under the following two cases: Case I: $l = i$ and Case II: $l \neq i$.

Before giving the expectation of each term under Case I and Case II, we firstly use the law of large numbers to give the following approximation [1], [2]:

$$\left( H_{\text{in}}^T H_{\text{in}} \right)^{-1} \approx E_{\text{in}}, \quad M \to \infty, \quad (99)$$

where $E_{\text{in}}$ is a $N \times N$ diagonal matrix whose $(j, j)$th element is $[E_{\text{in}}]_{jj} = \sigma_{\text{inj}}^{-1}$, where $\sigma_{\text{inj}}$ is given in (16). From the approximation of (99), we can obtain [24]

$$\nu_{\text{inj}} = E \left\{ s_{\text{inj}} \right\} = \begin{cases} \alpha_{l, ZF}^l (n, j), & l = i \\ \alpha_{l, ZF}^l a_{l, ZF} (n, j), & l \neq i. \end{cases} \quad (100)$$

Based on (100), one has

$$E \left\{ \tilde{y}_{\text{in}}, j \right\} = \sum_{l=1}^L \sqrt{\tau_d} P_l y_{\text{inj}}, \quad (101)$$

By using the approximation property of (99) and the method of [24], we obtain

$$E \left\{ s_{\text{inj}} \bar{y}_{\text{inj}} \right\} = \begin{cases} \sqrt{\tau_d} P_l y_{\text{inj}}^2 + \bar{y}_{\text{inj}} & l = i \\ \sqrt{\tau_d} P_l \tilde{y}_{\text{inj}} + \sum_{l_1 \neq l}^L \sqrt{\tau_d} P_{l_1} y_{\text{inj}} \bar{y}_{\text{inj}} & l \neq i \end{cases} \quad (102)$$

where $\bar{y}_{\text{inj}}$ and $\tilde{y}_{\text{inj}}$ are given in (53) and (55), respectively. Substituting (100), (101) and (102) into (103), we have

$$\sigma_{\text{inj}} = \text{Cov} \left\{ s_{\text{inj}}, \tilde{y}_{\text{inj}} \right\} = \begin{cases} \sqrt{\tau_d} P_l y_{\text{inj}} & l = i \\ \sqrt{\tau_d} P_l \tilde{y}_{\text{inj}} & l \neq i, \end{cases} \quad (103)$$

where $\tilde{y}_{\text{inj}}$ is given in (54). Similarly, it can be shown that

$$E \left\{ |\tilde{y}_{\text{inj}}|^2 \right\} = \tau_d P_l y_{\text{inj}}^2 + \sum_{l \neq i}^L \tau_d P_l y_{\text{inj}} \tilde{y}_{\text{inj}} + \sum_{l=1}^L \sum_{l_1 \neq l}^L \tau_d \sqrt{P_l P_{l_1} y_{\text{inj}} \bar{y}_{\text{inj}}} + 1. \quad (104)$$

Substituting (101) and (104) into (105), we obtain

$$\eta_{\text{in}, j} = \text{Cov} \left\{ \tilde{y}_{\text{in}, j}, \tilde{y}_{\text{in}, j} \right\} = \tau_d P_l y_{\text{inj}} + \sum_{l \neq i}^L \tau_d P_l \tilde{y}_{\text{inj}} + 1. \quad (105)$$

Substituting (100), (101), (103) and (105) into (26) and

$$\mu_{\text{inj}} = s_{\text{inj}} - \tilde{y}_{\text{inj}},$$

the expectations $E \left\{ |\tilde{y}_{\text{inj}}|^2 \right\}$ and $E \left\{ |\mu_{\text{inj}}|^2 \right\}$ are given as follows:

$$E \left\{ |\tilde{y}_{\text{inj}}|^2 \right\} = \nu_{\text{inj}}^2 + \frac{\sigma_{\text{inj}}^2}{\eta_{\text{in}, j}}. \quad (106)$$

$$E \left\{ |\mu_{\text{inj}}|^2 \right\} = \frac{\sigma_{\text{inj}}^2}{\sqrt{\tau_d P_l} \eta_{\text{in}, j}}. \quad (107)$$

Finally, substituting (106) and (107) into (32) yields the result of Theorem 3.

**APPENDIX E**

In order to transform Problem (68) into a GP problem, we first consider converting the equality constraint in (68) from a posynomial function to a monomial function. More precisely, by using the technique in (40) of reference [39], we can use $\sigma_{\text{inj}}^{\gamma_{\text{inj}}} \approx 1 + \gamma_{\text{inj}}$ near the point $\bar{y}_{\text{inj}}$, where $\delta_{\text{inj}} = \bar{y}_{\text{inj}} (1 + \gamma_{\text{inj}})^{-1}$ and $\sigma_{\text{inj}} = \bar{y}_{\text{inj}} (1 + \gamma_{\text{inj}})$. Therefore, near the point $\bar{y}_{\text{inj}}$, the left hand side of the equality constraint in (68) can be approximated as

$$\prod_{i=1}^L \prod_{n=1}^N (1 + \gamma_{\text{inj}}) \approx \prod_{i=1}^L \prod_{n=1}^N \sigma_{\text{inj}}^{\gamma_{\text{inj}}} \approx 1 + \gamma_{\text{inj}}. \quad (108)$$

Thus, the equality constraint in Problem (68) becomes a monomial function. For the inequality constraint in (67), the left and right hand sides of the inequality are both posynomial functions. Thus, we consider converting the right hand side of the inequality to a monomial function by using the approximation method in (40) of reference [39], which is as follows:

$$a_{\text{inj}}^2 + \gamma_{\text{inj}} P_{\text{inj}}\tau_d a_{\text{inj}}^2 b_{\text{inj}}^2 + \sum_{l=1}^L P_{\text{inj}}\tau_d a_{\text{inj}}^2 b_{l\text{inj}}^2 + \sum_{l=1}^L \sum_{l \neq i}^L P_{\text{inj}}\tau_d a_{\text{inj}}^2 b_{l\text{inj}} = \tilde{a}_{\text{inj}}^2 + \gamma_{\text{inj}} P_{\text{inj}} \sigma_{\text{inj}} \sum_{l=1}^L P_{\text{inj}} \tilde{a}_{\text{inj}}, \quad (109)$$

where

$$f_{\text{in}} = \tilde{P}_{\text{inj}}\tau_d a_{\text{inj}}^2 b_{\text{inj}} \left( b_{\text{inj}} \tilde{y}_{\text{inj}} + b_{\text{inj}} + a_{\text{inj}}^2 \right) + \sum_{l=1}^L \tilde{P}_{\text{inj}}\tau_d a_{\text{inj}}^2 b_{l\text{inj}} + a_{\text{inj}}^2, \quad (110)$$

$$a_{\text{inj}}^2 + \gamma_{\text{inj}} P_{\text{inj}}\tau_d a_{\text{inj}}^2 b_{\text{inj}}^2 + \sum_{l=1}^L \tilde{P}_{\text{inj}}\tau_d a_{\text{inj}}^2 b_{l\text{inj}} + a_{\text{inj}}^2, \quad (111)$$

$$a_{\text{inj}}^2 + \gamma_{\text{inj}} P_{\text{inj}}\tau_d a_{\text{inj}}^2 b_{l\text{inj}}^2 + \sum_{l=1}^L \tilde{P}_{\text{inj}}\tau_d a_{\text{inj}}^2 b_{l\text{inj}} + a_{\text{inj}}^2, \quad (112)$$

$$a_{\text{inj}}^2 + \gamma_{\text{inj}} P_{\text{inj}}\tau_d a_{\text{inj}}^2 b_{l\text{inj}}^2 + \sum_{l=1}^L \tilde{P}_{\text{inj}}\tau_d a_{\text{inj}}^2 b_{l\text{inj}} + a_{\text{inj}}^2, \quad (113)$$
\[ \sigma_{\text{in}} = \frac{1}{\ln \gamma_{\text{in}}} \left( \prod_{l \neq i} \gamma_{\text{in}} \right) \left( \frac{1}{\gamma_{\text{in}}} \right). \]  

(114)

Therefore, when \( \bar{p}_{\text{in}} \) and \( \gamma_{\text{in}} \) are close to \( \tilde{p}_{\text{in}} \) and \( \tilde{\gamma}_{\text{in}} \), respectively, for \( i = 1, \ldots, L \) and \( n = 1, \ldots, N \), (109) is established. By dividing the left hand side of the inequality in (67) by this approximate monomial function in (109), we can convert (67) to a polynomial function in (81), as shown at the bottom of page 10. Through the two approximations in (108) and (109), Problem (68) becomes a GP, which can be solved by CVX.

REFERENCES

[1] H. Quoc 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.

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

[3] J. Hoydis, S. ten Brink, and S. Shahabuddin, “Massive MIMO in the UL/DL of cellular networks: How many antennas do we need?” *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 160–171, Feb. 2013.

[4] H. Quoc Ngo, T. L. Marzetta, and E. G. Larsson, “Analysis of the pilot contamination effect in very large multicell MIMO systems for physical channel models,” in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Prague, Czech Republic, May 2011, pp. 3464–3467.

[5] M. A. Albreem, M. Juntti, and S. Shahabuddin, “Massive MIMO detection techniques: A survey,” *IEEE Commun. Survey Tuts.*, vol. 21, no. 4, pp. 3109–3132, 4th Quart., 2019.

[6] S.-N. Jin, D.-W. Yue, and H. H. Nguyen, “Power scaling laws of massive MIMO full-duplex relaying with hardware impairments,” *IEEE Access*, vol. 6, pp. 40860–40882, 2018.

[7] S.-N. Jin, D.-W. Yue, and H. H. Nguyen, “Equal-gain transmission in massive MIMO systems under Ricean fading,” *IEEE Trans. Veh. Technol.*, vol. 67, no. 10, pp. 9656–9668, Oct. 2018.

[8] M. Fozouomi, H. Q. Ngo, M. Matthaiou, S. Jin, and G. C. Alexandropoulou, “Hybrid processing design for multipair massive MIMO relaying with channel spatial correlation,” *IEEE Trans. Commun.*, vol. 67, no. 1, pp. 107–123, Jan. 2019.

[9] X. Wang, J. Liu, and C. Zhai, “Wireless power transfer-based multi-pair two-way relaying with massive antennas,” *IEEE Trans. Wireless Commun.*, vol. 16, no. 11, pp. 7672–7684, Nov. 2017.

[10] 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.

[11] O. Saatli, M. O. Ahmad, and M. N. S. Swamy, “Spectral efficiency maximization of a single cell massive MU-MIMO down-link TDD system by appropriate resource allocation,” *IEEE Access*, vol. 7, pp. 182758–182771, 2019.

[12] H. Quoc Ngo, E. G. Larsson, and T. L. Marzetta, “Massive MU-MIMO downlink TDD systems with linear precoding and downlink pilots,” in *Proc. 51st Annu. Allerton Conf. Commun., Control, Comput.*, Monticello, IL, USA, Oct. 2013, pp. 293–298.

[13] A. Khansefid and H. Minn, “Achievable downlink rates of MRC and ZF precoders in massive MIMO with uplink and downlink pilot contamination,” *IEEE Trans. Commun.*, vol. 63, no. 12, pp. 4894–4896, Dec. 2015.

[14] J. Zuo, J. Zhang, C. Yuan, W. Jiang, and W. Luo, “Multicell multuser massive MIMO transmission with downlink training and pilot contamination precoding,” *IEEE Trans. Veh. Technol.*, vol. 65, no. 8, pp. 6301–6314, Aug. 2016.

[15] S.-N. Jin, D.-W. Yue, and H. H. Nguyen, “Multicell massive MIMO: Downlink rate analysis with linear processing under Ricean fading,” *IEEE Trans. Veh. Technol.*, vol. 68, no. 4, pp. 3777–3791, Apr. 2019.

[16] J. Li, D. Wang, P. Zhu, J. Wang, and X. You, “Downlink spectral efficiency of distributed massive MIMO systems with linear beamforming under pilot contamination,” *IEEE Trans. Veh. Technol.*, vol. 67, no. 2, pp. 1130–1145, Feb. 2018.

[17] J. Li, D. Wang, P. Zhu, and X. You, “Benefits of beamforming training scheme in distributed large-scale MIMO systems,” *IEEE Access*, vol. 6, pp. 7432–7444, 2018.
[40] P. C. Weeraddana, M. Codreanu, M. Latva-aho, and A. Ephremides, “Resource allocation for cross-layer utility maximization in wireless networks,” IEEE Trans. Veh. Technol., vol. 60, no. 6, pp. 2790–2809, Jul. 2011.
[41] N. Ravindran, N. Jindal, and H. C. Huang, “Beamforming with finite rate feedback for LOS MIMO downlink channels,” in Proc. IEEE Global Telecommun. Conf., Washington, DC, USA, Nov. 2007, pp. 4200–4204.
[42] G. Taricco and E. Riegler, “On the ergodic capacity of correlated Rician fading MIMO channels with interference,” IEEE Trans. Inf. Theory, vol. 57, no. 7, pp. 4123–4137, Jul. 2011.

**SI-NIAN JIN** received the B.S. degree in communication engineering from Shandong University, Weihai, China, in 2014, and the M.S. degree in information and communication engineering from Dalian Maritime University, Dalian, China, in 2017, where he is currently pursuing the Ph.D. degree in information and communication engineering. His research interests include massive MIMO systems and cooperative communications.

**DIAN-WU YUE** (Senior Member, IEEE) received the B.S. and M.S. degrees in mathematics from Nankai University, Tianjin, China, in 1986 and 1989, respectively, and the Ph.D. degree in communications and information engineering from the Beijing University of Posts and Telecommunications, Beijing, China, in 1996. From 1989 to 1993, he was a Research Assistant of applied mathematics with the Dalian University of Technology, Dalian, China. From 1996 to 2003, he was an Associate Professor of communications and information engineering with the Nanjing University of Posts and Telecommunications, Nanjing, China. From 2000 to 2001, he was a Visiting Scholar with the University of Manitoba, Winnipeg, MB, Canada. From 2001 to 2002, he was a Postdoctoral Fellow of the University of Waterloo, Waterloo, ON, Canada. In November 2017, May 2018, and January 2019, he was a Visiting Professor with the University of Saskatchewan, Saskatoon, SK, Canada. Since December 2003, he has been a Full Professor of communications and information engineering with Dalian Maritime University, Dalian. His current research interests include massive MIMO systems, millimeter wave MIMO communications, and cooperative relaying communications.

**HA H. NGUYEN** (Senior Member, IEEE) received the B.Eng. degree from the Hanoi University of Technology (HUT), Hanoi, Vietnam, in 1995, the M.Eng. degree from the Asian Institute of Technology (AIT), Bangkok, Thailand, in 1997, and the Ph.D. degree from the University of Manitoba, Winnipeg, MB, Canada, in 2001, all in electrical engineering. He joined the Department of Electrical and Computer Engineering, University of Saskatchewan, Saskatoon, SK, Canada, in 2001, and became a Full Professor, in 2007. He currently holds the position of NSERC/Cisco Industrial Research Chair of low-power wireless access for sensor networks. He has coauthored, with Ed. Shwedyk, the textbook *A First Course in Digital Communications* (Cambridge University Press). His research interests fall into the broad areas of communication theory, wireless communications, and statistical signal processing. He is a Fellow of the Engineering Institute of Canada (EIC) and a Registered Member of the Association of Professional Engineers and Geoscientists of Saskatchewan (APEGS). He was an Associate Editor of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, from 2007 to 2011, and the IEEE WIRELESS COMMUNICATIONS LETTERS, from 2011 to 2016. He currently serves as an Associate Editor for the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY.