IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received November 20, 2019, accepted December 4, 2019, date of publication December 13, 2019,
date of current version December 26, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2959711

Spectral Efficiency Maximization of a Single Cell Massive MU-MIMO Down-Link TDD System by Appropriate Resource Allocation

O. SAATLOU, (Student Member, IEEE), M. OMAIR AHMAD, (Fellow, IEEE), AND M. N. S. SWAMY, (Fellow, IEEE)
Center for Signal Processing and Communications, Department of Electrical and Computer Engineering, Concordia University, Montreal, QC H3G 1M8, Canada
Corresponding author: M. N. S. Swamy (swamy@ece.concordia.ca)

This work was supported in part by the Natural Sciences and Engineering Research Council (NSERC) of Canada, and in part by the Regroupement Strategique en Microelectronique du Quebec (ReSMiQ).

ABSTRACT This paper deals with the problem of maximizing the spectral efficiency in a massive multi-user MIMO downlink system, where a base station is equipped with a very large number of antennas and serves single-antenna users simultaneously in the same frequency band, and the beamforming training scheme is employed in the time-division duplex mode. An optimal resource allocation that jointly selects the training duration on uplink transmission, the training signal power on downlink transmission, the training signal power on uplink transmission, and the data signal power on downlink transmission is proposed in such a way that the spectral efficiency is maximized given the total energy budget. Since the spectral efficiency is the main concern of this work, and its calculation using the lower bound on the achievable rate is computationally very intensive, in this paper, we also derive approximate expressions for the lower bound of achievable downlink rate for the maximum ratio transmission (MRT) and zero-forcing (ZF) precoders. The computational simplicity and accuracy of the approximate expressions for the lower bound of achievable downlink rate are validated through simulations. By employing these approximate expressions, experiments are conducted to obtain the spectral efficiency of the massive MIMO downlink time-division duplexing system with the optimal resource allocation and that of the beamforming training scheme. It is shown that the spectral efficiency of the former system using the optimal resource allocation is superior to that yielded by the latter scheme in the cases of both MRT and ZF precoders.

INDEX TERMS Massive MIMO, power control, resource allocation, spectral efficiency, channel state information acquisition.

I. INTRODUCTION

The use of massive multiple-input multiple-output (MIMO) systems, where a number of users communicate with the base station (BS) with a very large number of antennas, is viable approach for achieving significant improvement in spectral efficiency (SE) [1]–[6]. It has been shown that by employing a very large number of antennas at BS, the interference among the users is canceled, the uncorrelated noise is eliminated and small-scale fading effects are averaged out [7]. In addition, linear detectors such as zero-forcing (ZF) and maximum-ratio combining (MRC) detectors on the uplink (UL) transmission have a near optimal performance along with an acceptable complexity in massive multiuser MIMO (MU-MIMO) systems. In these systems, linear precoders such as ZF and maximum-ratio transmission (MRT) precoders on the downlink (DL) transmission also offer lower complexity along with a near optimal performance [8]. Due to the aforementioned advantages, massive MIMO systems are studied for next generation of cellular networks [9]–[15].

In the cellular networks such as the fifth generation (5G) of mobile communication systems, all users occupy full time-frequency resources both in UL and DL transmissions. In a DL transmission, BS needs to ensure that each user receives only the data intended for it. In a UL transmission, BS requires to recover the individual signals transmitted by the users. In view of this, BS has to perform the huge amount of multiplexing and de-multiplexing signal processing, which is

The associate editor coordinating the review of this manuscript and approving it for publication was Ahmed Mohamed Ahmed Almradi.

feasible using a massive number of antennas and having the channel state information (CSI) at BS [4], [9].

For DL transmission in a massive MIMO system, acquiring the CSI is one of the most challenging topics. BS requires CSI to transmit the precoded signal and the users also require CSI to decode the transmitted signal in DL transmission [16]. This CSI can be estimated by received pilot signals or can be obtained through the feedback from the receiver to the transmitter. In frequency-division duplexing (FDD) operation, the users estimate CSI from DL pilots sent by BS and communicate the estimated CSI back to BS over a feedback channel [17]. This feedback is very costly in massive MU-MIMO, since the number of DL pilots is determined with the massive number of antennas at BS. On the other hand, in time-division duplexing (TDD) operation, BS estimates CSI from UL pilots sent by the users. Due to the channel reciprocity between UL and DL channels in TDD systems, once BS estimates the UL channel, it automatically has a valid estimate of the DL channel. Therefore, the pilot transmission is determined with the number of the users in TDD operation. Typically, the users are smaller than the antennas at BS in massive MIMO systems. As a result, CSI acquisition under TDD mode is more economical and preferable than FDD mode [4], [5].

In DL transmission under TDD operation, the users need to obtain CSI in order to accurately detect the received data symbols. To this end, one simple method is that BS transmits pilot sequences to the users so that each user estimates the DL channel based on the received pilot sequences. This overhead of channel estimation is not effective, since this method depends on the number of antennas at BS. In view of this, it is commonly assumed that the users are aware merely of the statistical properties of the channels and the perfect CSI is not available [18]–[20].

To solve this problem, in [21], the authors have introduced the *beamforming training* (BT) method in such a way that each user efficiently obtains the estimate of CSI in DL massive MU-MIMO transmission. In this method, BS transmits a short pilot sequence to the users and each user estimates the efficient channel gain using the minimum mean-square error (MMSE) channel estimation method. This channel estimation method depends only on the number of users. Thus, the BT scheme is preferable in DL transmission for massive MIMO systems. In [22], the BT scheme has been employed in association with the pilot contamination precoding (PCP) scheme to improve SE in a massive MU-MIMO DL transmission. In [23], a blind algorithm for the estimation of the effective gain is proposed in which no DL pilots are required. It is shown in this paper that the performance of this method is better than that of the BT scheme in [23]. However, the complexity of this method is not negligible since the method is based on the blind channel estimation algorithm. In blind estimation, squares of the absolute values of all the symbols in one data block are required in order to calculate the sample power of the received signal resulting in a high complexity. In particular, the complexity significantly increases when the coherence interval is large. Furthermore, in the blind estimation algorithm, the sum of all the large scale fading coefficients is required in order to estimate the effective channel gain of each user. In addition, some of the parameters, such as the expectation of the channel estimates, are numerically computed using Bayes' rule and Riemann sum. These numerical computations significantly increase the complexity. Hence, in this paper, we employ the BT scheme which requires much less complexity and can be easily implemented.

In addition, some methods have been presented to reduce the dimension of the channel required to be estimated in massive MIMO by applying the spatial basis expansion model [24], [25]. Specifically, in [24], a beam-domain full-duplex scheme has been proposed to make co-time co-frequency uplink and downlink transmissions possible in which a small number of pilot symbols is required. In [25], a method of beam-domain hybrid time switching and power splitting protocol has been proposed in a full-duplex massive MIMO system in order to reduce the pilot symbols.

One of the typical problems in wireless MIMO networks is to study as to how much training is required to estimate CSI. The effect of training sequences on the achievable rate has been investigated in [26] and [27]. In case of multiuser TDD MIMO systems, an attempt has been made in [28] to deal with the problem as to how much time should be spent in training for a given number of transmit antennas, number of receive antennas, and length of the channel coherence time. In addition, it has been shown in [29] that, by varying the transmit powers for the pilot and data sequences, the optimal number of pilot symbols is equal to the number of transmit antennas. In [30] and [31], the performance of channel estimation has been studied for different pilot symbol designs, where a lower bound on the achievable rate has been expressed as a function of the Cramer-Rao bound (CRB). In [32], a hybrid pilot channel estimation technique is proposed for multi-cell multiuser massive MIMO systems, where the pilot duration is optimally selected to maximize SE.

To improve SE of massive MU-MIMO systems, power control among the pilot sequences and payload signals is essential [33]. This power control strategy is employed in UL transmission for different purposes [34]–[36]. For instance, a power allocation scheme has been proposed in [34] in order to optimize the training and data symbols power, in which SE is maximized for a given total energy budget. In [35], a power control scheme based on the channel quality of each user has been proposed in order to maximize the minimum achievable rate of each user, where a MRC detector is used at BS. In [36], a power allocation method over training and data symbols power has been proposed in order to maximize SE, where a ZF receiver is employed. In DL transmission, power among the user is allocated by BS in order to improve the SE of massive MU-MIMO systems [37]–[43]. For instance, a power allocation scheme among the users has been proposed in [37] in order to maximize SE under the total power constraint at BS. In [38], a power control method among the users has been studied in order to maximize SE, where a regularized zero forcing (RZF) precoder is employed. In [39], a power control scheme is proposed to maximize SE in multi-cell massive distributed antenna systems, where the pilot contamination effect is studied. In [40], a power allocation method along with a pilot design is proposed in order to maximize SE in multi-cell massive MIMO systems. A method of power control among the users in conjunction with the BT method has been proposed in [41] based on the waterfilling approach to maximize SE. In [42], a power control scheme between the data and pilot symbols has been proposed in order to improve SE, where the BT scheme is employed. In [43], a method of power allocation among each of the pilot and data symbols of all the users is proposed to maximize SE, where the total power budget per coherence interval for all users is given. In spite of the fact that there are a number of works on the SE maximization in DL transmission, none of these has considered a resource allocation scheme that jointly optimizes the pilot power, data power and duration of training in the BT scheme.

We propose, in this paper, a resource allocation scheme that maximizes SE in DL massive MU-MIMO transmission assuming the pilot and data powers to be different, and in which MRT or ZF precoding is employed. The specific contributions are given below.

- Closed-form approximate expressions for the achievable rates of the aforementioned precoders, where the BT scheme is employed in DL transmission, are derived. To validate the closed-form approximate expressions, the performance of the SE based on the achievable rates is studied by conducting simulation irrespective of the method employed for allocating the resources.
- An optimal resource allocation, which jointly optimizes the training duration in UL and DL transmissions, the training signal power on UL and DL transmissions, and the data signal power on DL transmission for a given total energy budget spent in a coherence interval in order to maximize SE, is proposed.
- It is shown that the performance of the proposed resource allocation scheme is superior to that of other existing schemes in terms of SE.

In spite of the fact that the multi-cell scenario is more practical, there is still much interest in the single-cell scenario, since it can be deployed in specific applications such as in stadiums and rural wireless broadband access [33]. In view of this, in this paper, the single-cell scenario is considered in order to develop the proposed method. This approach developed in the present paper should be helpful in dealing with the multi-cell scenario which will be investigated in the future.

The system model is described in Section II, where the BT method is employed for DL transmission in massive MU-MIMO systems. The achievable rates for MRT and ZF precoders are derived, and the spectral efficiency is defined in Section III. Approximate expressions for the achievable

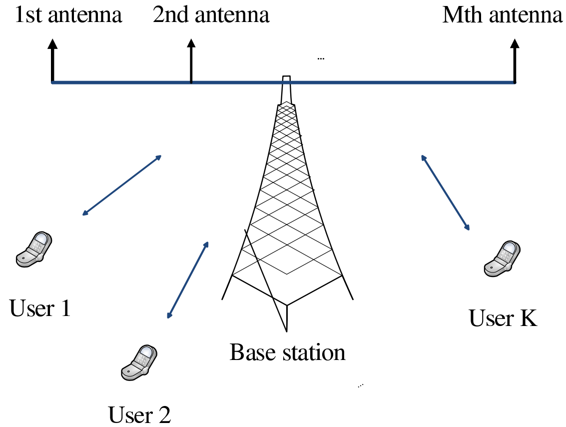

FIGURE 1. Massive MU-MIMO system model.

rates for MRT and ZF precoders are derived in Section IV, and the proposed resource allocation scheme is presented in Section V. Section VI provides numerical results to compare the achievable SE using the proposed resource allocation scheme with that of the BT method. Finally, Section VII concludes the paper by summarizing and highlighting the work of undertakes therein.

It should be pointed out that in this paper the same symbols are used for the derivation of the achievable rates for both the MRT and ZF precoders.

II. MASSIVE SINGLE-CELL SYSTEM MODEL

In this section, the DL transmission is studied in a single-cell massive MU-MIMO system, where a BS with M antenna elements simultaneously communicates with K single antenna users as shown in Fig. 1. It is assumed that $M \gg K$ and BS employs linear precoding technique before the DL transmission to all the users. Thus, BS requires CSI, which is obtained through the UL training. Due to the channel reciprocity between UL and DL channels in TDD operation, BS uses the obtained CSI in order to precode the data for the users.

A. UPLINK CHANNEL ESTIMATION

The orthogonal pilot sequences of length τ_u symbols per coherence interval are simultaneously transmitted by all users in the cell. Since the pilot sequences are orthogonal, $\tau_u \geq K$. The pilot matrix of K users is denoted by $\boldsymbol{\Psi} = [\phi_1, \phi_2, \ldots, \phi_K] \in \mathbb{C}^{\tau_u \times K}$ with the orthogonality property $\boldsymbol{\Psi}^{\dagger}\boldsymbol{\Psi} = \mathbf{I_K}$, where ϕ_k denotes the pilot sequence of kth user and $(.)^{\dagger}$ denotes the Hermitian operation of the associated matrix.

Let $\mathbf{H} \in \mathbb{C}^{M\times K}$ be the channel matrix. In this paper, the elements of $\mathbf{H}$ are independent Gaussian distributed with zero mean and unit variance. Since our proposed resource allocation is not dependent on the large-scale fading coefficient, we neglect the effect of this coefficient for simplicity. Thus, the $M \times \tau_u$ pilot matrix received at BS can be written as [21]

$$\mathbf{Y_u} = \sqrt{\tau_u p_u}\,\mathbf{H}\,\mathbf{\Psi}^\dagger + \mathbf{N_u}, \tag{1}$$

where p_u and $\mathbf{N_u} \in \mathbb{C}^{M\times\tau_u}$ denote respectively, the average pilot transmission power of each user and the received noise matrix at BS. We assume that the elements of $\mathbf{N_u}$ are independent Gaussian distributed with zero mean and unit variance. Using the received pilot matrix given by (1), the minimum mean-square error (MMSE) estimate of $\mathbf{H}$ can be written as [44]

$$\hat{\mathbf{H}} = \frac{\tau_u p_u}{\tau_u p_u + 1}\mathbf{H} + \frac{\sqrt{\tau_u p_u}}{\tau_u p_u + 1}\tilde{\mathbf{N}}_\mathbf{u}, \tag{2}$$

where $\tilde{\mathbf{N}}_u = \mathbf{N_u}\mathbf{\Psi}$ has the same distribution as $\mathbf{N_u}$. In this case, $\mathbf{H}$ is decomposed as

$$\mathbf{H} = \hat{\mathbf{H}} + \varepsilon, \tag{3}$$

where ε denotes the channel estimation error. Since MMSE estimation is employed, ε and $\hat{\mathbf{H}}$ are independent. In addition, ε and $\hat{\mathbf{H}}$ have i.i.d $\mathcal{CN}(0, \frac{1}{\tau_u p_u+1})$ and $\mathcal{CN}(0, \frac{\tau_u p_u}{\tau_u p_u+1})$ elements, respectively [21].

B. TRANSMISSION IN DOWNLINK CHANNEL [21]

BS first uses the channel estimate $\hat{\mathbf{H}}$ obtained in the previous subsection to precode the symbols, and then BS transmits the precoded symbols to the users in DL transmission. In view of this, let s_k be the symbol that BS transmits to the kth user, with $\mathbb{E}\{|s_k|^2\} = 1$ and $\mathbf{W} \in \mathbb{C}^{M\times K}$ be the linear precoding matrix. In this case, the $M \times 1$ transmit signal vector can be written as

$$\mathbf{x} = \sqrt{p_d}\,\mathbf{W}\mathbf{s}, \tag{4}$$

where p_d denotes the average transmit power at BS and $\mathbf{s} \triangleq [s_1, s_2, \ldots, s_K]^T$, where $(.)^T$ denotes the transpose operator. The linear precoding matrix $\mathbf{W}$ should be chosen in such a way that the power constraint at BS is satisfied. Hence, we have $\mathbb{E}\big[||\mathbf{x}||\big] = p_d$, where $||.||$ denotes Euclidean norm, or equivalently, we have

$$\mathbb{E}\big[\mathrm{tr}\big(\mathbf{W}\mathbf{W}^\dagger\big)\big] = 1. \tag{5}$$

In view of this, the received vector at the users can be written as

$$\mathbf{y} = \sqrt{p_d}\,\mathbf{H}^T\mathbf{W}\mathbf{s} + \mathbf{n}, \tag{6}$$

where $\mathbf{n} \sim \mathcal{CN}(0, 1)$ denotes an additive white Gaussian noise. We define $a_{ki} \triangleq \mathbf{h}_k^T\mathbf{w}_i$, where $\mathbf{h}_j$ and $\mathbf{w}_j$ denote the jth columns of $\mathbf{H}$ and $\mathbf{W}$, respectively. In view of this definition, the received signal at the kth user can be decomposed as

$$y_k = \sqrt{p_d}\,a_{kk}s_k + \sqrt{p_d}\sum_{i=1, i\neq k}^{K} a_{ki}s_i + n_k. \tag{7}$$

For accurately detecting the transmitted signal in DL transmission, each user needs to obtain CSI. A conventional method of channel estimation is that BS transmits pilot symbols in such a way that the users estimate the channel using minimum mean-square error (MMSE) estimation. This method is inefficient since the overhead on the aforementioned channel estimation is proportional to M, which tends to infinity in a massive MU-MIMO system. It can be seen from (7) that the user k just needs a_{kk}, which is a scalar value, in order to detect s_k. Thus, each user does not require to obtain the knowledge of H or W. Therefore, the BT scheme is employed to estimate a_{kk} for each user in which just a small amount of the coherence interval is spent [21]. In the BT scheme, the channel estimation is proportional to K, which is much smaller than M. In the next subsection, we explain as to how to estimate a_{kk}.

C. BEAMFORMING TRAINING

In the BT scheme, BS beamforms the pilot sequences in DL transmission after channel estimation in the UL training. Then, the effective channel gain a_{ki} is estimated at each user by the received pilot sequences. We define $\mathbf{S}_p \in \mathbb{C}^{K\times\tau_d}$ to be a pilot matrix in the DL channel, where τ_d denotes the number of symbols for pilot sequences. Using this definition, the pilot matrix is given by

$$\mathbf{S}_p = \sqrt{\tau_d p_p}\;\Phi. \tag{8}$$

where p_p and Φ denote the power of each pilot symbol and the pilot sequence matrix in DL transmission, respectively. Since the pilot sequences are orthogonal, we have $\Phi\Phi^\dagger = \mathbf{I}_K$, which requires that $\tau_d \geqslant K$. In the BT method, using the precoding matrix $\mathbf{W}$, BS beamforms the pilot sequence for the users. In other words, the transmitted pilot matrix is $\mathbf{W}\mathbf{S}_p$. Thus, the received pilot matrix in DL transmission can be expressed as

$$\mathbf{Y}_p^T = \sqrt{\tau_d p_p}\,\mathbf{H}^T\mathbf{W}\Phi + \mathbf{N}_p^T, \tag{9}$$

where $\mathbf{N}_p^T$ denotes the noise matrix in the received signal with i.i.d. $\mathcal{CN}(0, 1)$ entries. To estimate the channel, we use the orthogonality of pilot sequences. In view of this, let $\tilde{\mathbf{Y}}_p^T \triangleq \mathbf{Y}_p^T\Phi^\dagger$. In this case, we have

$$\tilde{\mathbf{Y}}_p^T = \sqrt{\tau_d p_p}\,\mathbf{H}^T\mathbf{W} + \tilde{\mathbf{N}}_p^T, \tag{10}$$

where $\tilde{\mathbf{N}}_p^T \triangleq \mathbf{N}_p^T\Phi^\dagger$ has i.i.d $\mathcal{CN}(0, 1)$ elements. By decomposing $\tilde{\mathbf{Y}}_p^T$ given by (10), we have

$$\tilde{\mathbf{y}}_{p,k}^T = \sqrt{\tau_d p_p}\,\mathbf{h}_k^T\mathbf{W} + \tilde{\mathbf{n}}_{p,k}^T = \sqrt{\tau_d p_p}\,\mathbf{a}_k^T + \tilde{\mathbf{n}}_{p,k}^T, \tag{11}$$

where $\tilde{\mathbf{y}}_{p,k}$ and $\tilde{\mathbf{n}}_{p,k}^T$ represent the kth columns of $\tilde{\mathbf{Y}}_p$ and $\tilde{\mathbf{N}}_p^T$, respectively and $\mathbf{a}_k \triangleq [a_{k1}a_{k2}\ldots a_{kK}]^T$. From (11), kth user estimates $\mathbf{a}_k$. Although the elements of $\mathbf{a}_k$ are correlated and should be jointly estimated, it has been shown in [21] that the performance loss due to independent estimation is negligible. As a result, $a_{k1},\ldots,a_{kK}$ are estimated independently. In view of this, the ith element of $\tilde{\mathbf{y}}_{p,k}$ is employed to estimate a_{ki} using MMSE channel estimation. In this case, the estimation of a_{ki} can be expressed as [44]

$$\hat{a}_{ki} = \mathbb{E}[a_{ki}] + \frac{\sqrt{\tau_d p_p}\mathrm{Var}(a_{ki})}{\tau_d p_p\;\mathrm{Var}(a_{ki}) + 1}\big(\tilde{y}_{p,ki} - \sqrt{\tau_d p_p}\mathbb{E}[a_{ki}]\big) \tag{12}$$

where $\tilde{y}_{p,ki}$ denotes the ith entry of $\tilde{\mathbf{y}}_{p,k}$ and $\mathrm{Var}(a_{ki})$ represents the variance of a_{ki}. This expression looks similar to (10) in [21]. However, in [21] the transmit power for pilot and data symbols are assumed to be the same in DL transmission. It is to be emphasized that in the present paper, we distinguish between the pilot transmit power p_p and the data transmit power p_d in DL transmission.

We define ϵ_{ki} to be the channel estimation error. Since MMSE estimation is employed, the estimate $\hat{a}_{ki}$ and the estimation error ϵ_{ki} are uncorrelated. In view of this, the effective channel gain a_{ki} is given by

$$a_{ki} = \hat{a}_{ki} + \epsilon_{ki}. \tag{13}$$

Substituting (13) into (7), we have

$$y_k = \sqrt{p_d}\hat{a}_{kk}s_k + \sqrt{p_d}\sum_{i=1, i\neq k}^{K} \hat{a}_{ki}s_i + \sqrt{p_d}\sum_{i=1}^{K}\epsilon_{ki}s_i + n_k. \tag{14}$$

III. ACHIEVABLE DOWNLINK RATE AND SPECTRAL EFFICIENCY

Employing an approach similar to that used in [45], it can be shown that the achievable DL rate (ADR) for kth user is lower bounded as

$$R_k = \mathbb{E}\left[\log_2\left(1 + \frac{p_d|\hat{a}_{kk}|^2}{p_d\sum_{i=1}^{K}\mathbb{E}\{|\epsilon_{ki}|^2\} + p_d\sum_{i\neq k}^{K}|\hat{a}_{ki}|^2 + 1}\right)\right]. \tag{15}$$

Even though this expression is similar to that obtained in [21], it is noted that the values of $\hat{a}_{ki}$ and $\hat{a}_{kk}$ in (15) are different from those in [21] in view of our distinguishing the pilot transmit power from the data transmit power.

A. MRT PRECODER

When BS uses the MRT precoder in the DL transmission, the precoding matrix is defined as

$$\mathbf{W} = \alpha_{MRT}\,\hat{\mathbf{H}}^*, \tag{16}$$

where $(.)^*$ denotes the conjugate operation of the associated matrix and α_{MRT} is a constant which is employed to satisfy the power constraint given by (5). Thus, we have [21]

$$\alpha_{MRT} = \sqrt{\frac{\tau_u p_u + 1}{MK\tau_u p_u}}. \tag{17}$$

Proposition 1: Using MRT precoding technique, $\hat{a}_{ki}$ and $\mathbb{E}\{|\epsilon_{ki}|^2\}$ are given by

$$\begin{cases} \hat{a}_{ki} = \dfrac{\sqrt{\tau_d p_p}}{\tau_d p_p + K}\tilde{y}_{p,ki} & i \neq k \\ \hat{a}_{kk} = \dfrac{\sqrt{\tau_d p_p}}{\tau_d p_p + K}\tilde{y}_{p,kk} + \dfrac{K}{\tau_d p_p + K}\sqrt{\dfrac{\tau_u p_u M}{K(\tau_u p_u + 1)}} & i = k \\ \mathbb{E}\{|\epsilon_{ki}|^2\} = \dfrac{1}{\tau_d p_p + K} & i\,\forall\,k \end{cases} \tag{18}$$

The proof of this proposition is given in Appendix A.

Substituting $\mathbb{E}\{|\epsilon_{ki}|^2\} = \frac{1}{\tau_d p_p + K}$ into (15), we obtain the lower bound of ADR for kth user as

$$R_k^{MRT} = \mathbb{E}\left[\log_2\left(1 + \frac{p_d|\hat{a}_{kk}|^2}{\frac{K\,p_d}{\tau_d p_p + K} + p_d\sum_{i\neq k}^{K}|\hat{a}_{ki}|^2 + 1}\right)\right]. \tag{19}$$

B. ZF PRECODER

When BS uses the ZF precoder in the DL transmission, the precoding matrix is defined as

$$\mathbf{W} = \alpha_{ZF}\,\hat{\mathbf{H}}^*\left(\hat{\mathbf{H}}^T\hat{\mathbf{H}}^*\right)^{-1}, \tag{20}$$

where α_{ZF} is a normalization constant for satisfying the transmit power constraint given by (5). Thus, we have [21]

$$\alpha_{ZF} = \sqrt{\frac{(M-K)\tau_u p_u}{K(\tau_u p_u + 1)}}. \tag{21}$$

Proposition 2: using ZF precoding technique, $\hat{a}_{ki}$ and $\mathbb{E}\{|\epsilon_{ki}|^2\}$ are given by

$$\begin{cases} \hat{a}_{ki} = \dfrac{\sqrt{\tau_d p_p}}{\tau_d p_p + K(\tau_u p_u + 1)}\tilde{y}_{p,k,i} & i \neq k \\ \hat{a}_{kk} = \dfrac{\sqrt{\tau_d p_p}}{\tau_d p_p + K(\tau_u p_u + 1)}\tilde{y}_{p,k,k} \\ \qquad + \dfrac{\sqrt{M(M-K)\tau_u p_u(\tau_u p_u + 1)}}{\tau_d p_p + K(\tau_u p_u + 1)} & i = k \\ \mathbb{E}\{|\epsilon_{ki}|^2\} = \dfrac{1}{\tau_d p_p + K(\tau_u p_u + 1)} & i\,\forall\,k \end{cases} \tag{22}$$

The proof of this proposition is given in Appendix B.

Substituting $\mathbb{E}\{|\epsilon_{ki}|^2\} = \frac{1}{\tau_d p_p + K(\tau_u p_u + 1)}$ into (15), we obtain the lower bound of ADR for kth user as

$$R_k^{ZF} = \mathbb{E}\left[\log_2\left(1 + \frac{p_d|\hat{a}_{kk}|^2}{\frac{K\,p_d}{\tau_d p_p + K(\tau_u p_u + 1)} + p_d\sum_{i\neq k}^{K}|\hat{a}_{ki}|^2 + 1}\right)\right]. \tag{23}$$

C. SPECTRAL EFFICIENCY

The spectral efficiency S is defined by [21]

$$S = \frac{T - \tau_u - \tau_d}{T}\sum_{k=1}^{K} R_k, \tag{24}$$

where R_k is the lower bound on ADR for the kth user given by (19) and (23) for the MRT and ZF precoders respectively, and T is the length of the coherence interval in DL transmission. The estimate for $\hat{a}_{ki}$ and $\hat{a}_{kk}$ depend on whether the MRT or ZF precoder is used. These estimates are given by (18) and (22) for the MRT and ZF precoders, respectively. In order to obtain $\hat{a}_{ki}$ and $\hat{a}_{kk}$, Monte-Carlo simulations are to performed, wherein the channel and noise matrices are generated. This process requires a large amount of calculations. This motivated us to present a close approximation for the lower bound on ADR, wherein only a low amount of calculations is required.

IV. APPROXIMATION FOR THE LOWER BOUND OF ACHIEVABLE DOWNLINK RATE

To obtain a close approximation of the achievable rate given by (19) and (23) for the MRT and ZF precoders, respectively, we use the following Lemma [37].

Lemma 1 [37]: When X and Y are independent positive random variables,

$$\mathbb{E}\left[\log_2\left(1+\frac{X}{Y}\right)\right] \approx \log_2\left(1+\frac{\mathbb{E}\{X\}}{\mathbb{E}\{Y\}}\right). \tag{25}$$

It has been shown in [37] that when M increases and goes to infinity, the approximation (25) becomes asymptotically exact.

A. MAXIMUM-RATIO TRANSMISSION PRECODING

Employing Lemma 1, a tractable expression for the lower bound of the achievable DL rate for the MRT precoder, given by (19), can be approximated as

$$R_k^{\mathrm{MRT}} \approx \log_2\left(1+\frac{p_d\ \mathbb{E}\{|\hat{a}_{kk}|^2\}}{\frac{K\ p_d}{\tau_d p_p+K}+p_d\sum_{i\neq k}^{K}\ \mathbb{E}\{|\hat{a}_{ki}|^2\}+1}\right), \tag{26}$$

Proposition 3: Substituting (18) into the lower bound of the achievable DL rate given by (26), it can be shown that

$$R_k^{\mathrm{MRT}} \approx \log_2\left(1+\mathrm{SINR}_k^{\mathrm{MRT}}\right) = \tilde{R}_k^{\mathrm{MRT}}, \tag{27}$$

where

$$\mathrm{SINR}_k^{\mathrm{MRT}} = \frac{p_d\ [a\ \tau_d^2 p_p^2 + b\ \tau_d p_p + c]}{\tau_d^2 p_p^2(d\ p_d+1)+\tau_d p_p(e\ p_d+f)+g(p_d+1)}, \tag{28}$$

and

$$\begin{aligned} a &= \alpha_{MRT}^2\left(\frac{\tau_u p_u}{\tau_u p_u+1}\right)^2 M(M+1)+\alpha_{MRT}^2\frac{\tau_u p_u}{(\tau_u p_u+1)^2}M,\\ b &= \frac{2\ \tau_u p_u M}{(\tau_u p_u+1)}+1,\\ c &= \frac{K\ \tau_u p_u M}{(\tau_u p_u+1)},\\ d &= \frac{K-1}{K},\\ e &= 2K-1,\\ f &= 2K,\\ g &= K^2. \end{aligned} \tag{29}$$

The proof of this proposition is given in Appendix C.

B. ZERO-FORCING TRANSMISSION

Employing Lemma 1, a tractable expression for the lower bound of the achievable DL rate for the ZF precoder, given by (23), can be approximated as

$$R_k^{\mathrm{ZF}} \approx \log_2\left(1+\frac{p_d\ \mathbb{E}\{|\hat{a}_{kk}|^2\}}{\frac{K\ p_d}{\tau_d p_p+K(\tau_u p_u+1)}+p_d\sum_{i\neq k}^{K}\ \mathbb{E}\{|\hat{a}_{ki}|^2\}+1}\right), \tag{30}$$

Proposition 4: Substituting (22) into the lower bound of the achievable DL rate given by (30), it can be shown that

$$R_k^{\mathrm{ZF}} \approx \log_2\left(1+\mathrm{SINR}_k^{\mathrm{ZF}}\right) = \tilde{R}_k^{ZF}, \tag{31}$$

where

$$\mathrm{SINR}_k^{\mathrm{ZF}} = \frac{p_d\ [a\ \tau_d^2 p_p^2 + b\ \tau_d p_p + c]}{\tau_d^2 p_p^2(d p_d+1)+\tau_d p_p(e p_d+f)+g p_d+h}, \tag{32}$$

and

$$\begin{aligned} a &= \frac{1}{K(\tau_u p_u+1)}+\alpha_{ZF}^2,\\ b &= 2\sqrt{(M-K)\tau_u p_u}+1,\\ c &= K(M-K)(\tau_u p_u+1)\tau_u p_u,\\ d &= \frac{K-1}{K(\tau_u p_u+1)},\\ e &= (2K-1),\\ f &= 2K(\tau_u p_u+1),\\ g &= K^2(\tau_u p_u+1),\\ h &= K^2(\tau_u p_u+1)^2. \end{aligned} \tag{33}$$

The proof of this proposition is given in Appendix D.

C. SPECTRAL EFFICIENCY

Substituting in (24), the approximate expressions $\tilde{R}_k^{\mathrm{MRT}}$ and $\tilde{R}_k^{\mathrm{ZF}}$ for the achievable rates given by (27) and (31) for the MRT and ZF precoders, respectively, we can obtain approximate values for SE for the two precoders. Thus, SE is obtained directly without requiring any simulation and involves only simple calculations. In Section VI, we show that SE obtained using $\tilde{R}_k^{\mathrm{MRT}}$ ($\tilde{R}_k^{\mathrm{ZF}}$) is very close to that obtained using R_k^{MRT} (R_k^{ZF}).

V. OPTIMAL RESOURCE ALLOCATION

In this section, we present the proposed resource allocation scheme in order to maximize SE. It has been shown in [33] that allocating optimal powers for the training symbols and data symbols increases SE, where SE is a function of the energy per bit (EPB) defined as

$$\eta \triangleq \frac{\frac{\tau_d}{T}p_p+\frac{\tau_u}{T}p_u+(1-\frac{\tau_d+\tau_u}{T})p_d}{S}. \tag{34}$$

It can be observed from (34) that when $p_p = p_u = p_d$ and $\tau_d = \tau_u$, we have $\eta = \frac{p_d}{S}$. Moreover, it can be observed from (24) that when the transmit power is reduced below a certain threshold, the bit energy increases. Hence, the minimum bit energy is obtained at a non-zero SE [34]. Operating below this SE is evidently inefficient. However, this regime can be operated by increasing the transmit power for training and reducing the transmit power for data. Motivated by these observations, we propose an optimal resource allocation to jointly select the training duration on UL transmission (τ_u), the training duration on DL transmission (τ_d), the training signal power on DL transmission (p_p), the training signal power on UL transmission (p_u), and the data signal power on DL transmission (p_d) in order to maximize SE for a given total energy budget spent in a coherence interval. In view of this, let the total transmit energy constraint at BS and each user be E_{td} and E_{tu}, respectively. Thus, we have

$$\tau_d p_p + (T - \tau_d - \tau_u) p_d \leq E_{td} \ , \tag{35}$$

and

$$\tau_u p_u \leq E_{tu}. \tag{36}$$

From (35), the channel estimate is degraded when $\tau_d p_p$ decreases, but the energy for the data transmission phase $(T-\tau_d-\tau_u)p_d$ is increased under the total energy constraint at BS. Hence, SE may improve. Moreover, the accuracy of the channel estimate is improved by allocating more energy to the training transmission phase. However, less energy should be allocated to the data transmission phase to satisfy (35). Hence, SE may again improve. In addition, from (36), it is straightforward that total energy constraint at each user is allocated to the UL training transmission phase in order to improve SE. In view of this, there are optimal values of τ_u, τ_d, p_p, p_u, and p_d which maximize SE for given E_{td}, E_{tu}, and T. Mathematically speaking, we have

$$\max_{p_u, p_d, p_p, \tau_u, \tau_d} \ S$$
$$\text{s.t.} \begin{cases} \tau_d p_p + (T - \tau_d - \tau_u) p_d \leq E_{td} \\ \tau_u p_u \leq E_{tu} \\ p_p \geq 0, \quad p_d \geq 0, \ p_u \geq 0 \\ \tau_u \geq K, \quad \tau_d \geq K \\ \tau_u + \tau_d \leq T \end{cases} \tag{37}$$

Lemma 2: The energy constraint given by (35) is satisfied with equality at the optimal solution.

Proof: Since the expressions for SINR given by (28) and (32) are monotonically increasing with p_p for a given p_d and vice versa, it can be observed from (24) that S is an increasing function of p_p when p_d is given. In addition, S is an increasing function of p_d when p_p is given. Hence, S is maximized when BS uses all the energy budget in one coherence interval, i.e., $\tau_d p_p + (T - \tau_d - \tau_d) p_d = E_{td}$. ■

Lemma 3: The energy constraint given by (36) is satisfied with equality at the optimal solution.

Proof: Since SINRs given by (28) and (32) are monotonically increasing with p_u, it can be observed from (24) that S is an increasing function of p_u. Hence, S is maximized when each user employs all energy budget in one coherence interval, i.e., $\tau_u p_u = E_{tu}$. ■

Remark 1: It has been shown in [29] that when the transmit powers for pilot and data sequences are allowed to vary, the optimal number of training symbols is equal to the number of transmit antennas M. On the other hand, if the training and data powers are to be made equal, the optimal number of training symbols can be larger than the number of transmit antennas M. In massive MIMO systems, M is very large. Thus, it is ineffective that we optimally choose $\tau_d = M$. On the other hand, the BT scheme is employed to efficiently estimate the channel gain for each user in order to reduce the number of training symbols in DL transmission. In view of this, in this paper, following the work in [21], we relax τ_d in the optimization problem given by (37).

According to Lemma 1, Lemma 2, and Remark 1, the optimization problem given by (37) can be rewritten as

$$\max_{p_d, p_p, p_u, \tau_u} \ S$$
$$\text{s.t.} \begin{cases} \tau_d p_p + (T - \tau_d - \tau_u) p_d = E_{td} \\ \tau_u p_u = E_{tu} \\ p_p \geq 0, \quad p_d \geq 0, \ p_u \geq 0 \\ K \leq \tau_u \leq T - \tau_d \end{cases} \tag{38}$$

There are optimal values of τ_u, p_u, p_p and p_d which maximize SE. In the next subsections, we intend to find the optimal values of τ_u and p_u, and simplify the optimization problem given by (38) for two linear precoders, namely, MRT and ZF. To this end, we first introduce the following theorem.

Theorem 1: The function

$$g(x) = x \ \log_2 \left(1 + \frac{\beta_k}{\varsigma_k + \mu_k \ x}\right) \tag{39}$$

is a strictly increasing function in $x \in (0, \infty)$, where $\beta_k > 0$, $\varsigma_k > 0$, and $\mu_k > 0$.

Proof: When β_k, ς_k, and μ_k are positive, we have

$$g'(x) = \frac{-1}{\ln 2}\left(\frac{\beta_k \mu_k x}{(\varsigma_k + \mu_k x)(\beta_k + \varsigma_k + \mu_k x)} - \ln\left(1 + \frac{\beta_k}{\varsigma_k + \mu_k x}\right)\right), \tag{40}$$

and

$$g''(x) = \frac{1}{\ln 2} \frac{-\beta_k \mu_k^2 (\beta_k + 2\varsigma_k) x - 2\beta_k \varsigma_k \mu_k (\beta_k + \varsigma_k)}{(\varsigma_k + \mu_k x)^2 (\beta_k + \varsigma_k + \mu_k x)^2} < 0, \tag{41}$$

where $g'(x)$ and $g''(x)$ are first and second derivatives of $g(x)$, respectively. From (41), we conclude that $g'(x)$ is a strictly deceasing function in x since $g''(x) < 0$. As a result, $g'(x) > g'(\infty) = 0$ which implies that $g(x)$ is a strictly increasing function in $x \in (0, \infty)$. ■

A. MAXIMUM-RATIO TRANSMISSION

To satisfy the first constraint of the optimization problem given by (38), we have $p_d = \frac{E_{td} - \tau_d p_p}{T - \tau_d - \tau_u}$. Using MRT precoder and substituting $p_d = \frac{E_{td} - \tau_d p_p}{T - \tau_d - \tau_u}$ into (27) and then, (27) into (24), we have

$$S(p_u, \tau_u, p_p, p_d) = \sum_{k=1}^{K} g_k(p_u, \tau_u, p_p, p_d) \tag{42}$$

where

$$g_k(p_u, \tau_u, p_p, p_d) = \left(1 - \frac{\tau_d + \tau_u}{T}\right) \log_2 \left(1 + \frac{\beta_k}{\varsigma_k + \mu_k (1 - \frac{\tau_d + \tau_u}{T})}\right), \tag{43}$$

and

$$\beta_k = \frac{(E_{td} - \tau_d p_p)}{T}\left[a(\tau_d p_p)^2 + b(\tau_d p_p) + c\right],$$
$$\varsigma_k = \frac{(E_{td} - \tau_d p_p)}{T}\left[d\,(\tau_d p_p)^2 + e\,(\tau_d p_p) + g\right],$$
$$\mu_k = \left((\tau_d p_p)^2 + f(\tau_d p_p) + g\right). \tag{44}$$

Proposition 4: The optimal value of τ_u given by the optimization problem (38) is K.

Proof: Assume that τ_u^*, $p_u^* = \frac{E_{tu}}{\tau_u^*}$, p_p^*, and $p_d^* = \frac{E_{td} - \tau_d p_p^*}{T - \tau_d - K}$ are the optimal solution of the optimization problem given by (38) which satisfy the constraints where $\tau_u^* > K$. Then, we choose $\overline{\tau_u} = K$, $\overline{p_u} = \tau_u^* p_p^*/K$, $\overline{p_p} = p_p^*$, and $\overline{p_d} = \frac{E_{td} - \tau_d p_p^*}{T - \tau_d - K}$ satisfying the constraints of the optimization problem given by (38). Note that with this choice, we have $\overline{\tau_u p_u} = \tau_u^* p_u^* = E_{tu}$. Substituting $\overline{p_u}$, $\overline{\tau_u}$,$\overline{p_p}$, and $\overline{p_d}$ into (43) yields

$$g_k(\overline{p_u}, \overline{\tau_u}, \overline{p_p}, \overline{p_d}) = \left(1 - \frac{\tau_d + K}{T}\right)\log_2 \left(1 + \frac{\overline{\beta_k}}{\overline{\varsigma_k} + \overline{\mu_k}(1 - \frac{\tau_d + K}{T})}\right) \tag{45}$$

where

$$\overline{\beta_k} = \frac{(E_{td} - \tau_d p_p^*)}{T}\left[a(\tau_d p_p^*)^2 + b(\tau_d p_p^*) + c\right],$$
$$\overline{\varsigma_k} = \frac{(E_{td} - \tau_d p_p^*)}{T}\left[d\,(\tau_d p_p^*)^2 + e\,(\tau_d p_p^*) + g\right],$$
$$\overline{\mu_k} = \left((\tau_d p_p^*)^2 + f(\tau_d p_p^*) + g\right). \tag{46}$$

Knowing $\overline{\tau_u p_u} = \tau_u^* p_u^*$ and using Theorem 1 and the fact $\tau_u^* > K$, we have

$$g_k(\overline{p_u}, \overline{\tau_u}, \overline{p_p}, \overline{p_d}) > \left(1 - \frac{\tau_d + \tau_u^*}{T}\right)\log_2\left(1 + \frac{\overline{\beta_k}}{\overline{\varsigma_k} + \overline{\mu_k}(1 - \frac{\tau_d + \tau_u^*}{T})}\right) = g_k(p_u^*, \tau_u^*, p_p^*, p_d^*). \tag{47}$$

Thus, from (42) and (47), we have

$$S(\overline{p_u}, \overline{\tau_u}, \overline{p_p}, \overline{p_d}) > S(p_u^*, \tau_u^*, p_p^*, p_d^*). \tag{48}$$

There is a contradiction between the equation given by (48) and the assumption. Thus, we have $\tau_u^* \le K$. Moreover, since the pilot sequences are orthogonal, we also have $\tau_u^* \ge K$. As a result, $\tau_u^* = K$, i.e., the optimal value of τ_u given by the optimization problem (38) is K. ■

At this stage, we find the optimal value of p_u with help of the following lemma.

Lemma 4: The optimal value of p_u in the optimization problem given by (38) is $p_u^* = E_{tu}/K$.

Proof: To satisfy the second constraint of the optimization problem given by (38), we have $\tau_u^* p_u^* = E_{tu}$. Since $\tau_u^* = K$, thus, $p_u^* = E_{tu}/K$. ■

According to Lemma 4 and Proposition 4, the optimization problem given by (38) can be rewritten as [42]

$$\max_{p_d} S\big|_{p_p = \frac{E_{td}}{\tau_d} - \left(\frac{T - \tau_d - K}{\tau_d}\right)p_d}$$
$$\text{such that } \left\{ 0 \le p_d \le \frac{E_{td}}{T - \tau_d - K}. \right. \tag{49}$$

Lemma 5: The objective function of the optimization problem given by (49) is concave with respect to p_d.

Proof: First, we substitute $p_p = \frac{E_{td}}{\tau_d} - (\frac{T - \tau_d - K}{\tau_d})p_d$ into (27) which yields a concave function with respect to p_d in the range $0 \le p_d \le \frac{E_{td}}{T - \tau_d - K}$. Knowing that $\log_2(1 + x)$ is a concave function, we conclude that $\log_2(1 + \text{SINR}_k^{MRT})$ is also concave. Moreover, since the summation of concave functions is also concave, we conclude the proof of Lemma 5. ■

In view of this, there is a global maximum point for the optimization problem given by (49). To obtain a globally optimal solution, any convex optimization scheme can be employed. We have employed the FMINCON function in MATLAB's optimization toolbox to derive the optimal solution of the optimization problem given by (49). It can be seen from (34) that when SE is maximized, EPB is minimized for a given E_{td}. As a result, this solution also provides the minimum value of EPB.

B. ZERO-FORCING

To satisfy the first constraint of the optimization problem given by (38), we have $p_d = \frac{E_{td} - \tau_d p_p}{T - \tau_d - \tau_u}$. Using ZF precoder and substituting $p_d = \frac{E_{td} - \tau_d p_p}{T - \tau_d - \tau_u}$ into (31) and then, (31) into (24), we have

$$S(p_u, \tau_u, p_p, p_d) = \sum_{k=1}^{K} g_k(p_u, \tau_u, p_p, p_d) \tag{50}$$

where

$$g_k(\tau_u, p_u, p_p, p_d) = \left(1 - \frac{\tau_d + \tau_u}{T}\right) \times \log_2\left(1 + \frac{\beta_k}{\varsigma_k + \mu_k(1 - \frac{\tau_d + \tau_u}{T})}\right), \tag{51}$$

and

$$\beta_k = \frac{(E_{td} - \tau_d p_p)}{T}\left[a(\tau_d p_p)^2 + b(\tau_d p_p) + c\right],$$
$$\varsigma_k = \frac{(E_{td} - \tau_d p_p)}{T}\left[d\,(\tau_d p_p)^2 + e\,(\tau_d p_p) + g\right],$$
$$\mu_k = \left((\tau_d p_p)^2 + f(\tau_d p_p) + h\right). \tag{52}$$

Proposition 5: The optimal value of τ_u given by the optimization problem (38) is K.

Proof: Assume that τ_u^*, $p_u^* = \frac{E_{tu}}{\tau_u^*}$, τ_u^*, p_p^*, and $p_d^* = \frac{E_{td} - \tau_d p_p^*}{T - \tau_d - K}$ are the optimal solution of the optimization problem given by (38) which satisfy the constraints where $\tau_u^* > K$. Then, we choose $\overline{\tau_u} = K$, $\overline{p_u} = \tau_u^* p_p^*/K$, $\overline{p_p} = p_p^*$, and $\overline{p_d} = \frac{E_{td} - \tau_d p_p^*}{T - \tau_d - K}$ satisfying the constraints of the optimization problem given by (38). Note that with this choice, we have

$\overline{\tau_u p_u} = \tau_u^* p_u^* = E_{tu}$. Substituting $\overline{p_u}$, $\overline{\tau_u}$,$\overline{p_p}$, and $\overline{p_d}$ into (51) yields

$$g_k(\overline{p_u}, \overline{\tau_u}, \overline{p_p}, \overline{p_d}) = \left(1 - \frac{\tau_d + K}{T}\right) \times \log_2\left(1+\frac{\overline{\beta_k}}{\overline{\varsigma_k}+\overline{\mu_k}(1-\frac{\tau_d+K}{T})}\right), \quad (53)$$

where

$$\overline{\beta_k} = \frac{(E_{td} - \tau_d p_p^*)}{T}\left[a(\tau_d p_p^*)^2 + b(\tau_d p_p^*) + c\right],$$
$$\overline{\varsigma_k} = \frac{(E_{td} - \tau_d p_p^*)}{T}\left[d\,(\tau_d p_p^*)^2 + e\,(\tau_d p_p^*) + g\right],$$
$$\overline{\mu_k} = \left((\tau_d p_p^*)^2 + f(\tau_d p_p^*) + h\right). \quad (54)$$

Knowing $\overline{\tau_u p_u} = \tau_u^* p_u^*$ and using Theorem 1 and the fact $\tau_d^* > K$, we have

$$\begin{aligned} &g_k(\overline{p_u}, \overline{\tau_u}, \overline{p_p}, \overline{p_d}) \\ &> \left(1 - \frac{\tau_d + \tau_u^*}{T}\right)\log_2\left(1 + \frac{\overline{\beta_k}}{\overline{\varsigma_k} + \overline{\mu_k}(1 - \frac{\tau_d+\tau_u^*}{T})}\right) \\ &= g_k(p_u^*, \tau_u^*, p_p^*, p_d^*). \end{aligned} \quad (55)$$

Thus, from (50) and (55), we have

$$S(\overline{p_u}, \overline{\tau_u}, \overline{p_p}, \overline{p_d}) > S(p_u^*, \tau_u^*, p_p^*, p_d^*). \quad (56)$$

There is a contradiction between the equation given by (56) and the assumption. Thus, we have $\tau_u^* \leq K$. Moreover, since the pilot sequences are orthogonal, we also have $\tau_u^* \geq K$. As a result, $\tau_u^* = K$, i.e., the optimal value of τ_u given by the optimization problem (38) is K. ■

At this stage, we find the optimal value of p_u with help of the following lemma.

Lemma 6: The optimal value of p_u is $p_u^* = E_{tu}/K$.

Proof: To satisfy the second constraint of the optimization problem given by (38), we have $\tau_u^* p_u^* = E_{tu}$. Since $\tau_u^* = K$, thus, $p_u^* = E_{tu}/K$. ■

According to Lemma 6 and Proposition 5, the optimization problem given by (38) can be rewritten as

$$\begin{aligned} &\max_{p_d} S|_{p_p=\frac{E_{td}}{\tau_d}-\left(\frac{T-\tau_d-K}{\tau_d}\right)p_d} \\ &\text{such that } \left\{ 0 \leq p_d \leq \frac{E_{td}}{T-\tau_d-K}. \right. \end{aligned} \quad (57)$$

Lemma 7: The objective function of the optimization problem given by (57) is concave with respect to p_d.

Proof: First, we substitute $p_p = \frac{E_{td}}{\tau_d} - (\frac{T-\tau_d-K}{\tau_d})p_d$ into (31) which yields a concave function with respect to p_d in the range $0 \leq p_d \leq \frac{E_{td}}{T-\tau_d-K}$. Knowing the fact that $\log_2(1+x)$ is a concave function, we conclude that $\log_2(1+\text{SINR}_k^{ZF})$ is also concave. Moreover, since the summation of concave functions is also concave, we conclude the proof of Lemma 7. ■

In view of this, there is a global maximum point for the optimization problem given by (57). To obtain a globally optimal solution, any convex optimization scheme can be employed. We have employed the FMINCON function in MATLAB's optimization toolbox to derive the optimal solution of the optimization problem given by (57). It can be seen from (34) that when SE is maximized, EPB is minimized for a given E_{td}. As a result, this solution also provides the minimum value of EPB.

VI. EXPERIMENTAL RESULTS

We first validate the approximate expressions $\tilde{R}_k^{\text{MRT}}$ and $\tilde{R}_k^{\text{ZF}}$ for the lower bound of the achievable rate given by (27) and (31) for the MRT and ZF precoders, respectively. We then utilize these approximate expressions to study the SE performance of the proposed optimal power allocation and compare the results with that of equal power allocation. In all the experiments conducted, we define $SNR \triangleq \frac{E_{td}}{T}$. Since E_{td} is the total transmit energy spent in a coherence interval T and the noise variance is 1, SNR has the interpretation of average transmit signal-to-noise ratio (SNR). Thus, SNR is dimensionless. We also choose $\tau_d = K$, and $T = 200$ (corresponding to a coherence bandwidth of 200 KHz and a coherence time of 1 ms) in all examples.

A. VALIDATION OF THE APPROXIMATE EXPRESSION FOR SE

In order to obtain SE for the MRT and ZF precoders, we first substitute the estimates of a_{ki} given by (18) and (22) in the expression for R_k given by (19) and (23), respectively. Then, we substitute (19) and (23) in the expression for SE given by (24). In order to obtain a_{ki}, 1000 Monte-Carlo simulations are carried out, where in the channel and noise matrices are generated for each snapshot. This process needs a large amount of calculation for a given SNR (for example, the run time for obtaining the SE for a given SNR is 21.76 seconds, using MATLAB software and a PC with Intel(R) Core(TM) i5 @ 2.7 GHz processor and 4 GB installed memory (RAM)). In order to validate the approximate expressions for SE for the MRT and ZF precoders, we substitute the lower bound of ADR given by (27) and (31), respectively, in the expression for SE given by (24). In this case, a small amount of calculation without any Monte-Carlo simulation is required (for example, the run time for obtaining the SE for a given SNR is 0.01 seconds, using MATLAB software and a PC with Intel(R) Core(TM) i5 @ 2.7 GHz processor and 4 GB installed memory (RAM)).

Figs. 2 and 3 show SE versus SNR when $M = 10$ and $M = 50$ employing the MRT and ZF precoders, respectively. It can be seen from Fig. 2 that SE obtained using the approximate expressions for $\tilde{R}_k^{\text{MRT}}$ is very close to that obtained using the actual one, for equal power allocation as well as for optimal power allocation when the MRT precoder is used. It can also be seen from Fig. 3 that SE obtained using the approximate expressions for $\tilde{R}_k^{\text{ZF}}$ is very close to that obtained using the actual one, for equal power allocation as well as for optimal power allocation when the ZF precoder is used. Hence, we can simply employ the approximated expressions $\tilde{R}_k^{\text{MRT}}$ and $\tilde{R}_k^{\text{ZF}}$ given by (27) and (31) in order to obtain SE for the MRT and ZF precoders, respectively, rather than

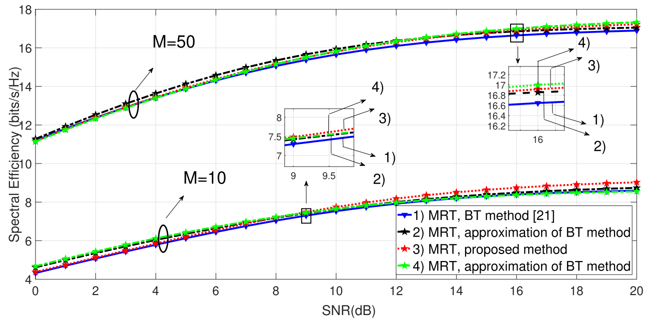

FIGURE 2. SE versus *SNR* when $E_{tu} = 6.9dB$ and $K = 5$.

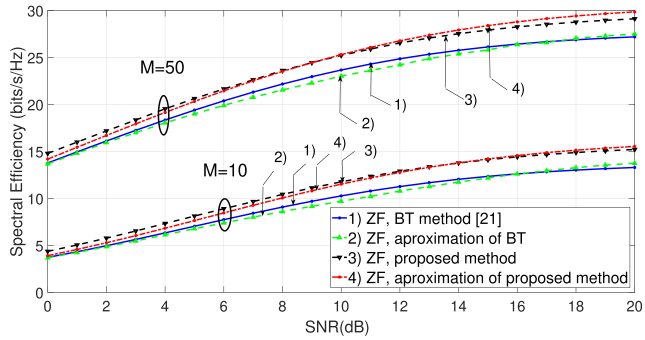

FIGURE 3. SE versus *SNR* when $E_{tu} = 6.9dB$ and $K = 5$.

using (15) to obtain SE with a high complexity. We expect that for other methods of power allocation, the approximate expressions can be used to obtain SE.

B. PERFORMANCE COMPARISON OF THE PROPOSED SCHEME AND THAT OF EQUAL POWER ALLOCATION

It can be seen from Figs 2 and 3 that the proposed method outperforms the BT method of [21] in terms of SE using both precoders. This superiority in the performance is attributed to the optimal transmitted power p_d^* and p_p^* and the optimal training duration τ_u which have been obtained in order to maximize SE. Moreover, it can be also seen from Fig 3 that the superiority of the proposed resource allocation scheme is more outstanding at high *SNR*. The reason for this is that the proposed resource allocation method outperforms when more energy budget is allocated to the users.

Furthermore, Fig. 4 shows the variation of the ratio of the optimal pilot power p_p to the optimal transmitted data power p_d for the MRT and ZF precoders. We can see that in order to maximize SE, more power should be allocated to the data symbols at high *SNR* and less power at low *SNR*. It is also seen that the approximately half of the total energy budget is employed for DL training and the other half is employed for DL data transmission at low *SNR*.

Besides, we compare SE of the proposed method with a genie receiver where the channel estimation error given by (13) is zero. It is assumed that kth user can perfectly estimate a_k and the power is equally allocated to p_p and p_d in the BT

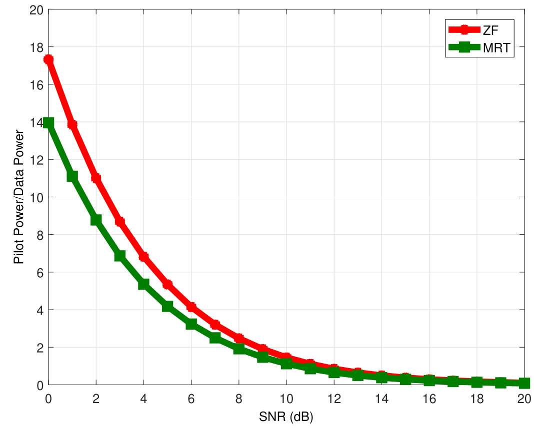

FIGURE 4. Ratio of the optimal pilot power to the optimal data power when $E_{tu} = 6.9dB$ and $K = 5$.

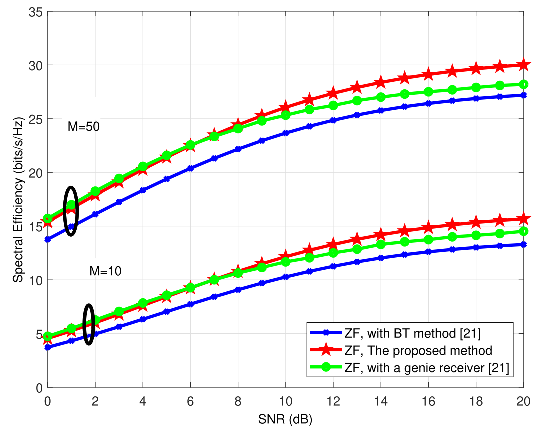

FIGURE 5. SE versus *SNR* when $E_{tu} = 6.9dB$ and $K = 5$ for the ZF precoder.

scheme. In this case, the SE is given by [21]

$$S_G = \frac{T - \tau_d - \tau_u}{T} \sum_{k=1}^{K} \mathbb{E}\Bigg[\log_2\Bigg(1 + \frac{p_d |\hat{a}_{kk}|^2}{p_d \sum_{i \neq k}^{K} |\hat{a}_{ki}|^2 + 1}\Bigg)\Bigg]. \tag{58}$$

Fig. 5 compares SE of the proposed method given by (37) with that of provided by (58). Here, we choose $K = 5$. It can be seen from Fig. 5 that the proposed method outperforms the method presented in [21] with a genie receiver at high SNR in terms of SE. In addition, at low SNR the difference between the performance of the proposed method and that provided with a genie receiver is negligible in terms of SE. As a result, the proposed method performs better than a genie receiver.

Finally, Figs. 6 and 7 show SE versus the number of antennas at BS (M) for the MRT and ZF precoders, respectively. From these figures, it can be seen that the performance of the proposed method is better than that provided in [21] in terms of the higher SE. In addition, when the number of antennas M is increased at BS, the difference between SE of the proposed method and that provided in [21] increases. As a result, the proposed method is an appropriate scheme for massive MIMO systems, where M goes to infinity. Moreover, it can also be seen from these figures that when the number

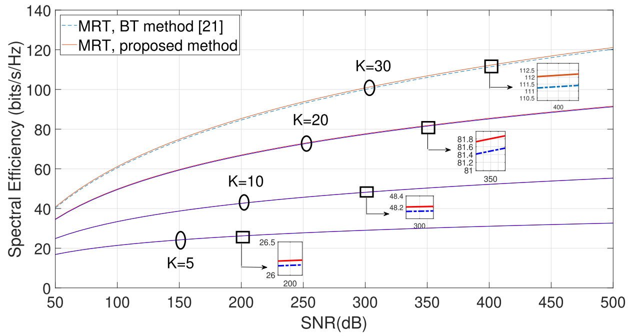

FIGURE 6. SE versus *SNR* when $E_{tu} = 6.9dB$ for the MRT precoder.

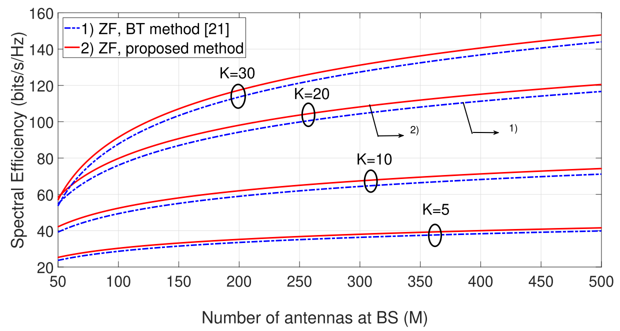

FIGURE 7. SE versus *SNR* when $E_{tu} = 6.9dB$ for the ZF precoder.

of users increases, the performance of the proposed resource allocation scheme significantly improves in terms of SE for both the MRT and ZF precoders. Hence, the proposed method is also an appropriate scheme for MU-MIMO systems, where multi-users are served in a cell.

C. COMPLEXITY

The optimization problems given by (57) and (49) are convex and are solved by FMINCON function in MATLAB's optimization toolbox using the interior point method. This method requires $\log(n_c)$ number of iterations, where n_c is the total number of constraints [46]. Each iteration requires $\mathcal{O}(n_c n_v^2)$ operations, where n_v is the number of variables. As a result, the total computational complexity for solving each problem is [46]

$$C = \mathcal{O}((n_c n_v^2)\log(n_c)). \quad (59)$$

In the optimization problems given by (49) and (57), we have $n_c = K$ and $n_v = 1$. Thus, $C = \mathcal{O}(K\log(K))$. In addition, the run times for solving the optimization problems given by (49) and (57) are 0.074471 and 0.098622 seconds, respectively, (using MATLAB software and a PC with Intel(R) Core(TM) i5 @ 2.7 GHz processor and 4 GB installed memory (RAM)). This indicates that the complexity of the proposed method is very low.

VII. CONCLUSION

This paper has investigated the downlink transmission in a multi-user massive MIMO system under time-division duplexing operation via a beamforming training method. We have derived an approximate expression for the achievable downlink rate and the accuracy of this approximation has been verified by obtaining numerical results. We have also proposed a resource allocation method in order to maximize the spectral efficiency and evaluated the performance of the proposed scheme conducting simulations. We have found that the number of pilot sequences for the channel estimation in the uplink channel should be equal to the number of users in order to maximize the spectral efficiency. We have also found that the spectral efficiency is remarkably improved at high *SNR* by allocating more power to data symbols for a given total power budget. In addition, we have shown that the performance of the proposed method is superior to that of the beamforming training method of [21] in terms of the spectral efficiency.

In a future work, we intend to extent these results to the multi-cell scenario.

APPENDIXES

APPENDIX A

PROOF OF PROPOSITION 1

Using the MRT precoder, we have $a_{ki} = \alpha_{MRT}\mathbf{h}_k^T\mathbf{h}_i^*$ [21]. Thus, $\mathbb{E}\{a_{ki}\}$ and $\mathrm{Var}(a_{ki})$ are given by [21]

$$\begin{cases} \mathbb{E}\{a_{ki}\} = 0 \quad i \neq k \\ \mathbb{E}\{a_{kk}\} = \dfrac{\sqrt{\tau_u p_u M}}{K(\tau_u p_u + 1)} \quad i = k \\ \mathrm{Var}(a_{ki}) = \dfrac{1}{K} \quad i\,\forall\, k \end{cases} \quad (60)$$

Substituting (60) in (12), we can obtain $\hat{a}_{ki}$ and $\hat{a}_{kk}$ given by (18). To prove Proposition 1, we also need to calculate $\mathbb{E}\{|\epsilon_{ki}|^2\}$.

For $i = k$, using (13) and (18), $\mathbb{E}\{|\epsilon_{kk}|^2\}$ can be written as

$$\begin{aligned}\mathbb{E}\{|\epsilon_{kk}|^2\} &= \mathbb{E}\left\{\left|\frac{K}{\tau_d p_p + K}a_{kk} - \frac{\sqrt{\tau_d p_p}}{\tau_d p_p + K}\tilde{n}_{p,kk}\right.\right. \\ &\quad \left.\left. - \frac{K}{\tau_d p_p + K}\sqrt{\frac{\tau_u p_u M}{K(\tau_u p_u + 1)}}\right|^2\right\} \\ &= \Big(\frac{K}{\tau_d p_p + K}\Big)^2\mathbb{E}\{|a_{kk}|^2\} + \frac{\tau_d p_p}{(\tau_d p_p + K)^2}. \quad (61)\end{aligned}$$

We also know that $\mathbb{E}\{|a_{kk}|^2\} = \alpha_{MRT}^2\big(\frac{\tau_u p_u}{\tau_u p_u + 1}\big)^2 M(M+1) + \alpha_{MRT}^2\frac{\tau_u p_u}{(\tau_u p_u+1)^2}M$ [21]. Substituting $\mathbb{E}\{|a_{kk}|^2\}$ into (61), we have

$$\mathbb{E}\{|\epsilon_{kk}|^2\} = \frac{1}{\tau_d p_p + K}. \quad (62)$$

For $i \neq k$, using (13) and (18), $\mathbb{E}\{|\epsilon_{ki}|^2\}$ can be written as

$$\begin{aligned}\mathbb{E}\{|\epsilon_{ki}|^2\} &= \mathbb{E}\{|\frac{K}{\tau_d p_p + K}a_{ki} - \frac{\sqrt{\tau_d p_p}}{\tau_d p_p + K}\tilde{n}_{p,ki}|^2\} \\ &= \Big(\frac{K}{\tau_d p_p + K}\Big)^2\mathbb{E}\{|a_{ki}|^2\} + \frac{\tau_d p_p}{(\tau_d p_p + K)^2}. \quad (63)\end{aligned}$$

We also know that $\mathbb{E}\{|a_{ki}|^2\} = \frac{1}{K}$ [21]. Substituting $\mathbb{E}\{|a_{ki}|^2\}$ into (63), we have

$$\mathbb{E}\{|\epsilon_{ki}|^2\} = \frac{1}{\tau_d p_p + K}. \tag{64}$$

Substituting (60) in (12) and having (62) and (64), we conclude the proof of Proposition 1.

APPENDIX B
PROOF OF PROPOSITION 2

Using the ZF precoder, we have $a_{ki} = \mathbf{h}_k^T \mathbf{w}_i^*$ [21], where w_i is the ith column of matrix W given by (20). Thus, $\mathbb{E}\{a_{ki}\}$ and $\text{Var}(a_{ki})$ are given by [21]

$$\begin{cases} \mathbb{E}\{a_{ki}\} = 0 & i \neq k \\ \mathbb{E}\{a_{kk}\} = \alpha_{ZF} & i = k \\ \text{Var} a_{ki} = \dfrac{1}{K(\tau_u p_u + 1)} & i \,\forall\, k \end{cases} \tag{65}$$

Substituting (65) in (12), we can obtain $\hat{a}_{ki}$ and $\hat{a}_{kk}$ given by (22). To prove Proposition 2, we also need to calculate $\mathbb{E}\{|\epsilon_{ki}|^2\}$.

For $i = k$, using (13) and (22), $\mathbb{E}\{|\epsilon_{ki}|^2\}$ can be written as

$$\begin{aligned} &\mathbb{E}\{|\epsilon_{kk}|^2\} \\ &= \mathbb{E}\Bigg\{\Bigg|\frac{K(\tau_u p_u + 1)}{\tau_d p_p + K(\tau_u p_u + 1)} a_{kk} - \frac{\sqrt{\tau_d p_p}}{\tau_d p_p + K(\tau_u p_u + 1)} \tilde{n}_{p,kk} \\ &\quad - \frac{\Delta}{\tau_d p_p + K(\tau_u p_u + 1)}\Bigg|^2\Bigg\} \\ &= \frac{\big(K(\tau_u p_u + 1)\big)^2 \mathbb{E}\{|a_{kk}|^2\}}{\big(\tau_d p_p + K(\tau_u p_u + 1)\big)^2} + \frac{\tau_d p_p}{(\tau_d p_p + K(\tau_u p_u + 1))^2} \\ &\quad - \frac{2\Delta K(\tau_u p_u + 1)\mathbb{E}\{a_{kk}\}}{\big(\tau_d p_p + K(\tau_u p_u + 1)\big)^2} + \frac{\Delta^2}{\big(\tau_d p_p + K(\tau_u p_u + 1)\big)^2}. \end{aligned} \tag{66}$$

where $\Delta = \sqrt{K(M-K)\tau_u p_u(\tau_u p_u + 1)}$. We also know that $\mathbb{E}\{|a_{kk}|^2\} = \frac{1}{K(\tau_u p_u + 1)} + \alpha_{zf}^2$ and $\mathbb{E}\{a_{kk}\} = \alpha_{zf}$ [21]. Substituting $\mathbb{E}\{|a_{kk}|^2\}$ and $\mathbb{E}\{a_{kk}\}$ into (66), we have

$$\mathbb{E}\{|\epsilon_{kk}|^2\} = \frac{1}{\tau_d p_p + K(\tau_u p_u + 1)} \tag{67}$$

For $i \neq k$, using (13) and (22), $\mathbb{E}\{|\epsilon_{ki}|^2\}$ can be written as

$$\begin{aligned} \mathbb{E}\{|\epsilon_{ki}|^2\} &= \mathbb{E}\{|\frac{K}{\tau_d p_p + K} a_{ki} - \frac{\sqrt{\tau_d p_p}}{\tau_d p_p + K} \tilde{n}_{p,ki}|^2\} \\ &= (\frac{K}{\tau_d p_p + K})^2 \mathbb{E}\{|a_{ki}|^2\} + \frac{\tau_d p_p}{(\tau_d p_p + K)^2}. \end{aligned} \tag{68}$$

We also know that $\mathbb{E}\{|a_{ki}|^2\} = \frac{1}{K(\tau_u p_u + 1)}$ [21]. Substituting $\mathbb{E}\{|a_{ki}|^2\}$ into (68), we have

$$\mathbb{E}\{|\epsilon_{ki}|^2\} = \frac{1}{\tau_d p_p + K(\tau_u p_u + 1)} \tag{69}$$

Substituting (65) in (12) and having (67) and (69), we conclude the proof of Proposition 2.

APPENDIX C
PROOF OF PROPOSITION 3

To prove *Proposition* 3, first, we calculate $\mathbb{E}\{|\hat{a}_{ki}|^2\}$ for MRT precoding.

$\mathbb{E}\{|\hat{a}_{ki}|^2\}$:

For $i = k$, from (18) and using the fact that $\mathbb{E}\{|a_{kk}|^2\} = \alpha_{MRT}^2 \big(\frac{\tau_u p_u}{\tau_u p_u + 1}\big)^2 M(M+1) + \alpha_{MRT}^2 \frac{\tau_u p_u}{(\tau_u p_u + 1)^2} M$ and $\mathbb{E}\{a_{kk}\} = \sqrt{\frac{\tau_u p_u M}{K(\tau_u p_u + 1)}}$ [21], we have

$$\begin{aligned} &\mathbb{E}\{|\hat{a}_{kk}|^2\} \\ &= \frac{1}{(\tau_d p_p + K)^2} \mathbb{E}\Big[\big|\tau_d p_p a_{kk} + \sqrt{\tau_d p_p} \tilde{n}_{pkk}^T + K\mathbb{E}\{a_{kk}\}\big|^2\Big] \\ &= \frac{1}{(\tau_d p_p + K)^2}\Big[\mathbb{E}\{|a_{kk}|^2\}(\tau_d p_p)^2 + 2K\big(\mathbb{E}\{a_{kk}\}\big)^2(\tau_d p_p) \\ &\quad + (K\mathbb{E}\{a_{kk}\})^2\Big] = \frac{1}{(\tau_d p_p + K)^2}\Big[\Big(\alpha_{MRT}^2 \frac{\tau_u p_u}{(\tau_u p_u + 1)^2} M \\ &\quad + \alpha_{MRT}^2 \big(\frac{\tau_u p_u}{\tau_u p_u + 1}\big)^2 M(M+1)\Big)(\tau_d p_p)^2 \\ &\quad + \big(\frac{2\tau_u p_u M}{\tau_u p_u + 1} + 1\big)\tau_d p_p + \frac{\tau_u p_u M K}{(\tau_u p_u + 1)}\Big] \end{aligned} \tag{70}$$

For $i \neq k$, from (18) and using the fact that $\mathbb{E}\{|a_{ki}|^2\} = \frac{1}{K}$ and $\mathbb{E}\{a_{ki}\} = 0$ [21], we have

$$\mathbb{E}\{|\hat{a}_{ki}|^2\} = \frac{\tau_d p_p}{(\tau_d p_p + K)^2}\Big[\frac{\tau_d p_p}{K} + 1\Big] \tag{71}$$

Then, by substituting (70) and (71) into (26), the proof of *Proposition* 3 is completed.

APPENDIX D
PROOF OF PROPOSITION 4

To prove *Proposition* 4, first, we calculate $\mathbb{E}\{|\hat{a}_{ki}|^2\}$ for ZF precoding.

First, $\mathbb{E}\{|\hat{a}_{ki}|^2\}$:

For $i = k$, from (22) and using the fact that $\mathbb{E}\{|a_{kk}|^2\} = \frac{1}{K(\tau_u p_u + 1)} + \alpha_{zf}^2$ and $\mathbb{E}\{a_{kk}\} = \alpha_{zf}$ [21], we have

$$\begin{aligned} &\mathbb{E}\{|\hat{a}_{kk}|^2\} \\ &= \frac{1}{(\tau_d p_p + K(\tau_u p_u + 1))^2} \mathbb{E}\Big[\big|\tau_d p_p a_{kk} + \sqrt{\tau_d p_p} \tilde{n}_{p,kk}^T + \Delta\big|^2\Big] \\ &= \frac{1}{(\tau_d p_p + K(\tau_u p_u + 1))^2}\Big[\mathbb{E}\{|a_{kk}|^2\}(\tau_d p_p)^2 \\ &\quad + \big(2\Delta\mathbb{E}\{a_{kk}\} + 1\big)\tau_d p_p + \Delta^2\Big] \\ &= \frac{1}{(\tau_d p_p + K(\tau_u p_u + 1))^2}\Big[\big(\frac{1}{K(\tau_u p_u + 1)} + \alpha_{zf}^2\big)(\tau_d p_p)^2 \\ &\quad + \big(2\Delta\alpha_{zf} + 1\big)\tau_d p_p + \Delta^2\Big] \end{aligned} \tag{72}$$

where $\Delta = \sqrt{K(M-K)\tau_u p_u(\tau_u p_u + 1)}$.

For $i \neq k$, from (13) and (22), we have

$$\mathbb{E}\{|\hat{a}_{ki}|^2\} = \frac{\tau_d p_p}{(\tau_d p_p + K(\tau_u p_u + 1))^2}\Big[\frac{\tau_d p_p}{K(\tau_u p_u + 1)} + 1\Big] \tag{73}$$

Then, by substituting (72) and (73) into (30), the proof of *Proposition* 4 is completed.

REFERENCES

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

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

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

[4] T. L. Marzetta, E. G. Larsson, H. Yang, and H. Q. Ngo, *Fundamentals of Massive MIMO*. Cambridge, U.K.: Cambridge Univ. Press, 2016.

[5] A. Chockalingam and B. S. Rajan, *Large MIMO Systems*. Cambridge, U.K.: Cambridge Univ. Press, 2014.

[6] O. Saatlou, M. O. Ahmad, and M. N. S. Swamy, "Spectral efficiency maximization of multiuser massive MIMO systems with finite-dimensional channel via control of users power," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 65, no. 7, pp. 883–887, Jul. 2018.

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

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

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

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

[11] C. Xu, Y. Hu, C. Liang, J. Ma, and L. Ping, "Massive MIMO, non-orthogonal multiple access and interleave division multiple access," *IEEE Access*, vol. 5, pp. 14728–14748, 2017.

[12] E. L. Bengtsson, F. Rusek, S. Malkowsky, F. Tufvesson, P. C. Karlsson, and O. Edfors, "A simulation framework for multiple-antenna terminals in 5G massive MIMO systems," *IEEE Access*, vol. 5, pp. 26819–26831, 2017.

[13] J. Li, D. Wang, P. Zhu, and X. You, "Energy efficiency optimization of distributed massive MIMO systems under Ergodic QoS and Per-RAU power constraints," *IEEE Access*, vol. 7, pp. 5001–5013, 2019.

[14] W. Cheng, H. Zhang, L. Liang, H. Jing, and Z. Li, "Orbital-angular-momentum embedded massive MIMO: Achieving multiplicative spectrum-efficiency for mmWave communications," *IEEE Access*, vol. 6, pp. 2732–2745, Dec. 2018.

[15] J. Zhang, X. Xue, E. Björnson, B. Ai, and S. Jin, "Performance analysis and power control of cell-free massive MIMO systems with hardware impairments," *IEEE Access*, vol. 6, pp. 55302–55314, Sep. 2018.

[16] 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. 4849–4864, Dec. 2015.

[17] Y. Wang, C. Li, Y. Huang, D. Wang, T. Ban, and L. Yang, "Energy-efficient optimization for downlink massive MIMO FDD systems with transmit-side channel correlation," *IEEE Trans. Veh. Technol.*, vol. 65, no. 9, pp. 7228–7243, Sep. 2016.

[18] J. Hoydis, S. ten Brink, and M. Debbah, "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.

[19] 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.

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

[21] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Massive MU-MIMO downlink TDD systems with linear precoding and downlink pilots," in *Proc. Allerton Conf. Commun., Control, Comput.*, Oct. 2013, pp. 293–298.

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

[23] H. Q. Ngo and E. G. Larsson, "No downlink pilots are needed in TDD massive MIMO," *IEEE Trans. Wireless Commun.*, vol. 16, no. 5, pp. 2921–2935, May 2017.

[24] X. Xia, K. Xu, D. Zhang, Y. Xu, and Y. Wang, "Beam-domain full-duplex massive MIMO: Realizing co-time co-frequency uplink and downlink transmission in the cellular system," *IEEE Trans. Veh. Technol.*, vol. 66, no. 10, pp. 8845–8862, Oct. 2017.

[25] K. Xu, Z. Shen, Y. Wang, X. Xia, and D. Zhang, "Hybrid time-switching and power splitting swipt for full-duplex massive MIMO systems: A beam-domain approach," *IEEE Trans. Veh. Technol.*, vol. 67, no. 8, pp. 7257–7274, Aug. 2018.

[26] J. Baltersee, G. Fock, and H. Meyr, "Achievable rate of MIMO channels with data-aided channel estimation and perfect interleaving," *IEEE J. Sel. Areas Commun.*, vol. 19, no. 12, pp. 2358–2368, Dec. 2001.

[27] X. Ma, L. Yang, and G. B. Giannakis, "Optimal training for MIMO frequency-selective fading channels," *IEEE Trans. Wireless Commun.*, vol. 4, no. 2, pp. 453–466, Mar. 2005.

[28] T. L. Marzetta, "How much training is required for multiuser MIMO?" in *Proc. Asilomar Conf. Signals, Syst. Comput. (ACSSC)*, Oct. 2006, pp. 359–363.

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

[30] M. Dong and L. Tong, "Optimal design and placement of pilot symbols for channel estimation," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Proc. (ICASSP)*, May 2001, pp. 2109–2112.

[31] L. Berriche, K. Abed-Meraim, and J. C. Belfiore, "Cramer-Rao bounds for MIMO channel estimation," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Montreal, QC, Canada, May 2004, pp. 397–400.

[32] J. Li, C. Yuen, D. Li, X. Wu, and H. Zhang, "On hybrid pilot for channel estimation in massive MIMO uplink," *IEEE Trans. Veh. Technol.*, vol. 68, no. 7, pp. 6670–6685, Jul. 2019.

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

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

[35] H. V. Cheng, E. Björnson, and E. G. Larsson, "Uplink pilot and data power control for single cell massive MIMO systems with MRC," in *Proc. Int. Symp. Wireless Commun. Syst. (ISWCS)*, Brussels, Belgium, Aug. 2015, pp. 396–400.

[36] O. Saatlou, M. O. Ahmad, and M. N. S. Swamy, "Spectral efficiency maximization of single cell massive multiuser MIMO systems via optimal power control with ZF receiver," in *Proc. IEEE Annu. Int. Symp. Pers., Indoor, Mobile Radio Commun. (PIMRC)*, Montreal, QC, Canada, Oct. 2017, pp. 1–5.

[37] Q. Zhang, S. Jin, M. McKay, D. Morales-Jimenez, and H. Zhu, "Power allocation schemes for multicell massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 14, no. 11, pp. 5941–5955, Nov. 2015.

[38] S. Zarei, W. Gerstacker, R. R. Müller, and R. Schober, "Low-complexity linear precoding for downlink large-scale MIMO systems," in *Proc. IEEE Annu. Int. Symp. Pers., Indoor, Mobile Radio Commun. (PIMRC)*, London, U.K., Sep. 2013, pp. 1119–1124.

[39] J. Zuo, J. Zhang, C. Yuen, W. Jiang, and W. Luo, "Energy efficient downlink transmission for multi-cell massive DAS with pilot contamination," *IEEE Trans. Veh. Technol.*, vol. 66, Feb. 2017, pp. 1209–1221, Feb. 2017.

[40] T. Van Chien, E. Bjornson, and E. G. Larsson, "Joint pilot design and uplink power allocation in multi-cell massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 17, no. 3, pp. 2000–2015, 2018.

[41] O. Saatlou, M. O. Ahmad, and M. N. S. Swamy, "Spectral efficiency maximization for massive multiuser MIMO downlink TDD systems via data power allocation with MRT precoding," in *Proc. IEEE Veh. Technol. Conf. (VTC-Fall)*, Toronto, ON, Canada, Sep. 2017, pp. 1–5.

[42] O. Saatlou, M. O. Ahmad, and M. N. S. Swamy, "Optimal power allocation for massive MU-MIMO downlink TDD systems," in *Proc. IEEE Can. Conf. Elect. Comput. Eng. (CCECE)*, Vancouver, BC, Canada, May 2016, pp. 1–4.

[43] O. Saatlou, M. O. Ahmad, and M. N. S. Swamy, "Joint data and pilot power allocation for massive MU-MIMO downlink TDD systems," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 66, no. 3, pp. 512–516, Mar. 2019.

[44] S. M. Kay, *Fundamentals of Statistical Signal Processing: Detection Theory*. Englewood Cliffs, NJ, USA: Prentice-Hall, 2006.

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

[46] A. D. Shoaei, M. Derakhshani, and T. Le-Ngoc, "Efficient LTE/Wi-Fi coexistence in unlicensed spectrum using virtual network entity: Optimization and performance analysis," *IEEE Trans. Commun.*, vol. 66, no. 6, pp. 2617–2629, Jun. 2018.

O. SAATLOU received the B.Sc. degree in telecommunications engineering from Azad University, Tehran, Iran, in 2009, and the M.Sc. degree in telecommunications engineering from Shahed University, Tehran, in 2013. He is currently pursuing the Ph.D. degree in electrical and computer engineering with Concordia University, Montreal, QC, Canada, under supervision of Dr. M. O. Ahmad and Dr. M. N. S. Swamy. His research interests include the areas of massive MIMO, resource allocation, and relay networks. He is the Co-Chair of Event Coordinator Team, the IEEE Montreal IRC, and the Co-Organizer.

M. OMAIR AHMAD (S'69–M'78–SM'83–F'01) received the B.Eng. degree in electrical engineering from Sir George Williams University, Montreal, QC, Canada, and the Ph.D. degree in electrical engineering from Concordia University, Montreal. From 1978 to 1979, he was a Faculty Member with the New York University College, Buffalo, NY, USA. In 1979, he joined the faculty of Concordia University as an Assistant Professor of computer science. He was a Founding Researcher with Micronet Company, Ltd., Sapporo, Japan, as a Canadian Network of Centers of Excellence, from 1990 to 2004. Previously, he was an Examiner of the order of Engineers of Quebec. He joined the Department of Electrical and Computer Engineering, Concordia University, where he was the Chair, from 2002 to 2005, and is currently a Professor. He is the Concordia University Research Chair (Tier I) in multimedia signal processing. He has authored in the area of signal processing and holds four patents. His current research interests include the areas of multidimensional filter design, speech, image and video processing, nonlinear signal processing, communication DSP, artificial neural networks, and VLSI circuits for signal processing. In 1988, he was a member of the Admission and Advancement Committee of the IEEE. He was a recipient of numerous honors and awards, including the Wighton Fellowship from the Sandford Fleming Foundation, an induction to Provosts Circle of Distinction for Career Achievements, and the Award of Excellence in Doctoral Supervision from the Faculty of Engineering and Computer Science, Concordia University. He has served as the Program Co-Chair for the 1995 IEEE International Conference on Neural Networks and Signal Processing, the 2003 IEEE International Conference on Neural Networks and Signal Processing, and the 2004 IEEE International Midwest Symposium on Circuits and Systems. He was a General Co-Chair of the 2008 IEEE International Conference on Neural Networks and Signal Processing. He is the Chair of the Montreal Chapter IEEE Circuits and Systems Society. He was the Local Arrangements Chairman of the 1984 IEEE International Symposium on Circuits and Systems. He was an Associate Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS PART—I: FUNDAMENTAL THEORY AND APPLICATIONS, from 1999 to 2001.

M. N. S. SWAMY received the B.Sc. degree (Hons.) in mathematics from Mysore University, India, the Diploma degree in electrical communication engineering from the Indian Institute of Science, Bangalore, and the M.Sc. and Ph.D. degrees in electrical engineering from the University of Saskatchewan, Saskatoon, Canada. He was conferred in 2009 the title of Honorary Professor at National Chiao Tong University, Taiwan. He served as the Chair of the Department of Electrical Engineering, Concordia University, Montreal, QC, Canada, from 1970 to 1977. He was the Dean of Engineering and Computer Science, Concordia University, from 1977 to 1993, where he is currently a Research Professor with the Department of Electrical and Computer Engineering. During that time, he developed the Faculty into a research-oriented one from what was primarily an undergraduate Faculty. Since July 2001, he holds the Concordia Chair (Tier I) in signal processing. He has also taught in the Electrical Engineering Department, Technical University of Nova Scotia, Halifax, and the University of Calgary, Calgary, and the Department of Mathematics, University of Saskatchewan. He has published extensively in the areas of number theory, circuits, and systems and signal processing. He holds five patents. He is the coauthor of nine books and several book chapters. Dr. Swamy was a Founding Member of Micronet from its inception in 1990 as a Canadian Network of the Centers of Excellence until its expiration in 2004, and also its coordinator for Concordia University. He is a Fellow of the Institute of Electrical Engineers, U.K.; the Engineering Institute of Canada; the Institution of Engineers, India; and the Institution of Electronic and Telecommunication Engineers, India. He was inducted in 2009 to the Provosts Circle of Distinction for career achievements. He was a recipient of many IEEE-CAS Society awards, including the Education Award in 2000, Golden Jubilee Medal in 2000, and the 1986 Guillemin-Cauer Best Paper Award. He has served the IEEE in various capacities, such as the President-Elect in 2003, the President in 2004, the Past-President in 2005, the Vice President (Publications) from 2001 to 2002, the Vice-President in 1976, the Editor-in-Chief of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS—I, from June 1999 to December 2001, an Associate Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS, from June 1985 to May 1987, a Program Chair for the 1973 IEEE CAS Symposium, a General Chair for the 1984 IEEE CAS Symposium, the Vice-Chair for the 1999 IEEE Circuits and Systems (CAS) Symposium, and a member of the Board of Governors of the CAS Society. He has been the Editor-in-Chief of the journal *Circuits, Systems and Signal Processing* (CSSP), since 1999. Recently CSSP has instituted a Best Paper Award in his name.