**QUEEN'S
UNIVERSITY
BELFAST**

Transmission Schemes and Power Allocation for Multiuser Massive MIMO Relaying

Duc Ho, C., Ngo, H-Q., & Matthaiou, M. (2021). Transmission Schemes and Power Allocation for Multiuser Massive MIMO Relaying. *IEEE Transactions on Vehicular Technology*.
<https://doi.org/10.1109/TVT.2021.3098275>

Published in:
IEEE Transactions on Vehicular Technology

Document Version:
Peer reviewed version

Queen's University Belfast - Research Portal:
[Link to publication record in Queen's University Belfast Research Portal](#)

Publisher rights
Copyright 2021, IEEE.
This work is made available online in accordance with the publisher's policies. Please refer to any applicable terms of use of the publisher.

General rights
Copyright for the publications made accessible via the Queen's University Belfast Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The Research Portal is Queen's institutional repository that provides access to Queen's research output. Every effort has been made to ensure that content in the Research Portal does not infringe any person's rights, or applicable UK laws. If you discover content in the Research Portal that you believe breaches copyright or violates any law, please contact openaccess@qub.ac.uk.

Transmission Schemes and Power Allocation for Multiuser Massive MIMO Relaying

Chung Duc Ho, Hien Quoc Ngo, *Senior Member, IEEE* and Michail Matthaiou, *Senior Member, IEEE*

Abstract—This paper proposes and analyzes the performance of two simple transmission protocols for a multiuser massive multiple-input multiple-output relaying system, where K single-antenna users transmit data to a massive-antenna destination through an N -antenna relay node. The proposed transmission protocols take imperfect channel state information (CSI) and power allocation into consideration. In the first transmission protocol, the relay does not need to know the CSI. It just amplifies and forwards the received signals to the destination utilizing the knowledge of the large-scale fading coefficients. In the second protocol, the relay first estimates the channels from all users during the training phase. With these channel estimates, it then uses the classical maximum-ratio combining (MRC) technique to combine all received signals and forwards them to the destination. In both protocols, the destination estimates the channels and employs MRC to decode the signals transmitted from all users. We propose an efficient channel estimation method at the destination in which the destination estimates only the effective channels gains, instead of the whole channels. As a consequence, the channel estimation overhead does not depend on the numbers of relay and destination antennas which makes the system scalable. We derive closed-form expressions for the spectral efficiency of the two transmission protocols. These results allow us to further analyze the performance of these two schemes and to meticulously allocate the transmit powers. Particularly, a max-min power control algorithm is proposed which selects the transmit powers at the relay and users to maximize the lowest spectral efficiency of all users for each large-scale fading realization. We show that, by using the max-min power allocation algorithm in both schemes, the spectral efficiency can be increased significantly, compared to uniform power allocation. Furthermore, a comparison between the two transmission schemes is pursued. If the distance between the users and the relay is large, the first transmission protocol is better and vice versa if this distance is small.

Index Terms—Amplify-and-forward, massive MIMO, power allocation, relaying systems.

Copyright (c) 2015 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.

C. D. Ho is with the Institute of Electrical, Electronics Engineering and Computer Sciences, Thu Dau Mot University, Binh Duong Province 750000, Vietnam (e-mail: hoducchung@tdmu.edu.vn).

H. Q. Ngo and M. Matthaiou are with the Institute of Electronics, Communications and Information Technology (ECIT), Queen's University Belfast, BT3 9DT, Belfast, U.K. (e-mail: {hien.ngo, m.matthaiou}@qub.ac.uk).

The work of C. D. Ho was funded by Thu Dau Mot University under grant number DT.20-030. The work of H. Q. Ngo was supported by the UK Research and Innovation Future Leaders Fellowships under Grant MR/S017666/1. The work of M. Matthaiou was in part supported by a research grant from the Department for the Economy Northern Ireland under the US-Ireland R&D Partnership Programme and by the ESPRC, UK under Grant EP/P000673/1. Parts of this work were presented at the 2019 IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC) [1].

I. INTRODUCTION

Massive multiple-input multiple output (MIMO) has been comprehensively investigated in the past decade and has become one of the core technologies for current and future wireless systems [2]–[6]. It includes a base station equipped with a massive number of active antennas to coherently and simultaneously serve many users. Thanks to the favorable propagation and channel hardening properties, massive MIMO is capable of serving many users at the same time with very high energy efficiency and spectral efficiency. More importantly, these gains are obtained under the use of low complexity linear processing, such as maximum-ratio and zero-forcing.

In colocated massive MIMO, the cell size should be typically large. Otherwise, we need to deploy many massive-antenna base stations, and the effect of pilot contamination will be detrimental. However, with large cells, the users at the cell edges experience poor quality of service due to the exacerbated path loss. To overcome this issue, relaying techniques were proposed to use and have received a lot of research interest [7]–[10]. The combination of massive MIMO and relaying techniques can be classified into three main systems: multipair massive MIMO relaying, multi-way massive MIMO relaying and multiuser massive MIMO relaying systems. The literature review of these categories is presented in the following section.

A. Related Work

Multipair massive MIMO relaying systems were studied in [7], [8]. The work of [7] derived the ergodic rate with maximum-ratio processing and showed that the energy efficiency scales with the number of relay antennas. A new pilot design and power allocation were proposed in [8] to improve the system performance. The performance of multipair massive MIMO relaying systems with full-duplex operation at the relay station was investigated in [11]. The authors showed that by using massive antenna arrays at the relay station together with simple linear processing, the loop interference can be reduced significantly. Full-duplex massive MIMO relaying systems with low-resolution analog-to-digital converters (ADCs) were analyzed in [12], [13]. These papers showed that by applying power allocation optimization, the transmitted power from each user can be reduced dramatically. More specifically, [13] revealed that by utilizing massive antennas at the base station, the performance loss due to the low-resolution ADCs can be compensated. Further investigations on multipair systems have been recently reported in [9], [14]. In [9], the energy efficiency and power scaling laws of multipair

two-way massive MIMO relaying systems with zero-forcing were examined. This study was related with [14], but [14] considered the amplify-and-forward transmission method and imperfect channel state information. To this end, in this work, the sum spectral efficiency and power control are investigated to improve the system performance.

Multi-way massive MIMO was investigated in [15] for single-cell setups and in [16] for multi-cell setups. Both papers show that by combining massive MIMO technology and multi-way relaying channels, we can obtain a huge spectral efficiency and avail of significant power savings. All the above works, however, consider networks where multiple users communicate with each other with the help of a relay equipped with massive antenna arrays.

Multuser massive MIMO relaying systems, where a base station equipped with massive antenna arrays serves a group of users with the help of a relay node, were proposed and studied in [17]. In this paper, the channel state information (CSI) needs to be estimated at the base station and the relay, which increases substantially the channel estimation overhead. In addition, the direct links from the base station to the users were neglected. Furthermore, [18] studied the sum spectral efficiency for uplink and downlink transmission of a multuser massive MIMO system with half duplex amplify-and-forward (AF) relaying. However, in this work, the direct link was again neglected and perfect CSI was assumed to be available at the relay and base station. Recall now that the CSI acquisition is very important in massive MIMO systems and needs to be taken into consideration [4]. Therefore, in our work, we focus on a multuser massive MIMO relaying system taking into account the direct link and imperfect CSI. Simple and scalable transmission protocols with low channel estimation overhead and efficient power control algorithms are proposed and investigated.

B. Contributions

In our work, we propose two simple transmission protocols for multuser massive MIMO relaying systems which rely on how channels are acquired at the relay and also take into account the direct links. The main contributions of this work are summarized as follows:

- 1) We propose two simple transmission schemes including channel estimation protocols and specific signal processing techniques at the relay and the destination. These schemes harness the CSI knowledge at the relay: Transmission Scheme 1 corresponds to the case without any CSI knowledge at the relay, and Transmission scheme 2 corresponds to the case with channel estimation at the relay. In Transmission Scheme 1, all users send their signals to the destination and the relay. The relay just forwards the scaled versions of its received signals to the destination. In Transmission Scheme 2, the relay estimates the channels from all users first. Then, it uses these channel estimates together with maximum-ratio combining (MRC) to combine and forward its received signals to the destination. For both schemes, the destination combines the signals received directly

from the users (direct links) and from the relay to detect all signals transmitted from all users.¹ To do this, the destination needs to have CSI knowledge. A very efficient channel estimation scheme at the destination is proposed where the destination estimates only the effective channel gains. With the proposed channel estimation scheme, the channel estimation overhead does not depend on the numbers of destination and relay antennas; this important property makes our transmission schemes scalable.

- 2) We propose a simple and scalable channel estimation scheme at the destination, where the training phase follows the same transmission protocol as the data transmission phase. With this channel estimation scheme, the destination needs to estimate only the effective channel gains, and hence, the channel estimation overhead does not depend on the numbers of antennas at the relay and destination. It depends only on the number of users.
- 3) We derive an exact closed-form expression for the achievable spectral efficiency for Transmission Scheme 1, and an approximating closed-form expression for the spectral efficiency for Transmission Scheme 2. The approximation is showed to be very tight even when the number of antennas is small.
- 4) From the obtained closed-form expressions for the spectral efficiency, we formulate a max-min power optimization problem which finds the transmit powers of all users and relay to maximize the smallest spectral efficiency of all users, subject to the power constraints of each user and the relay. The power allocation is implemented over large-scale fading time-scales, and improves the system performance significantly, compared to uniform power allocation. We want to stress that the max-min power control methodology is certainly well established, but we are not aware of any work that applies this methodology on our proposed transmission protocols.

Notation: $\mathbf{X}^H$ and $\mathbf{x}^H$ denote the Hermitian of the matrix $\mathbf{X}$ and vector $\mathbf{x}$, respectively. Given a matrix $\mathbf{X}$, we use $\mathbf{x}_k$ to depict its k -th column and $\text{Tr}(\mathbf{X})$ to represent its trace. The notations $\|\cdot\|_F$, $\|\cdot\|$, $\mathbb{E}\{\cdot\}$ and $\text{Var}(\cdot)$ denote the Frobenius norm, norm, expectation and variance operators, respectively. Finally, $\mathbf{I}_K$ denotes the $K \times K$ identity matrix.

II. SYSTEM MODEL

We consider a multuser massive MIMO relaying system as depicted in Fig. 1. The system consists of K users, one relay node, and one destination. All K users send their data to the destination via the assistance of the relay node.² All users have a single-antenna, while the relay and destination

¹In general, it might be possible to design the processing scheme at the relay taking into account the direct links. To do this, the relay needs to know the CSI of the direct links. The destination can send its estimates of the direct links to the relay. However, this technique is inherently impractical because the dimension of the direct link CSI is $M \times K$, where M and K are the number of destination antennas and the number of users, respectively, which causes a huge channel estimation overhead.

²This architecture can be used in practical cellular networks to improve the data rates of users at the cell boundaries or of users who have very weak direct channels to the base station [17], [19].

Fig. 1. Multiuser massive MIMO relaying system.

nodes are equipped with N and M antennas, respectively. We assume that $K \leq N \leq M$, and the relay operates in half-duplex mode with a simple AF scheme. To be more specific, all K users first send data to the relay and the destination on the same time-frequency band. Then, the relay amplifies the received signals and forwards them to destination. Finally, the destination applies the MRC technique to combine the received signals from users (via the direct link) and the signals transmitted from the relay.

Channel notation and model are detailed as follows:

- $\mathbf{G}_{ur} \in \mathbb{C}^{N \times K}$ is the channel from the K users to the relay which is modeled as

$$\mathbf{G}_{ur} = \mathbf{H}_{ur} \mathbf{D}_{ur}^{1/2},$$

where $\mathbf{H}_{ur}$ and $\mathbf{D}_{ur}$ are small-scale fading and large-scale fading matrices, respectively. The small-scale fading matrix is assumed to include independent and identically distributed (i.i.d.) $\mathcal{CN}(0, 1)$ random variables (RVs). The large-scale fading matrix is a diagonal matrix whose k -th diagonal element is denoted by $\beta_{ur,k}$.

- Similarly, $\mathbf{G}_{ud} \in \mathbb{C}^{M \times K}$ and $\mathbf{G}_{rd} \in \mathbb{C}^{M \times N}$ are the channels from the K users and the relay to the destination, respectively. These channels are modeled as

$$\mathbf{G}_{ud} = \mathbf{H}_{ud} \mathbf{D}_{ud}^{1/2} \text{ and } \mathbf{G}_{rd} = \mathbf{H}_{rd} \mathbf{D}_{rd}^{1/2},$$

where $\mathbf{H}_{ud}$ and $\mathbf{H}_{rd}$ are the corresponding small-scale channel matrices, while $\mathbf{D}_{ud} \in \mathbb{C}^{K \times K}$ and $\mathbf{D}_{rd} \in \mathbb{C}^{N \times N}$ are the corresponding large-scale fading matrices, which are diagonal. We assume that the elements of $\mathbf{H}_{ud}$ and $\mathbf{H}_{rd}$ are i.i.d. $\mathcal{CN}(0, 1)$ RVs. We use $\beta_{ud,k}$ and $\beta_{rd,k}$ to denote the k -th diagonal elements of $\mathbf{D}_{ud}$ and $\mathbf{D}_{rd}$, respectively. Furthermore, we can assume that $\beta_{rd,n} = \beta_{ud,n}, \forall n = 1, \dots, N$. This assumption is reasonable since the relay and destination antenna spacing is much smaller than the distance between the relay and the destination.

In our work, we assume that the channels associated with different antennas are uncorrelated. This assumption is reasonable as long as the antennas are sufficiently well separated (i.e. the distance between adjacent antennas is about half a wavelength) and the propagation environment offers rich enough scattering. Note that the physical size of the antenna array can be small even with very large number of antennas [20]. In addition, the practical measurements in [20] demonstrated the

validity of this assumption for large arrays with 128 antennas. Generally, the antenna correlation is more likely to appear in practice, when the antennas are not well separated or the propagation environment does not offer rich enough scattering. Thus, it is interesting to exploit our considered system under correlated channels. We leave this for future work. Depending on the CSI knowledge at the relay, we can have different transmission schemes. In the next sections, we will provide details of two simple transmission protocols.

III. TRANSMISSION SCHEME 1: NO CSI AT THE RELAY

In this transmission protocol, the relay does not have any knowledge of the small-scale fading channels. It just uses the knowledge of large-scale fading to forward an amplified version of the received signal to the destination.

A. Channel Estimation

Since the relay does not need any instantaneous CSI, the channel estimation is done only at the destination. The obtained channel estimates will be used to detect the transmitted data in the data transmission period. To reduce the channel estimation overhead, we propose not to estimate all channel matrices $\mathbf{G}_{ur}$, $\mathbf{G}_{ud}$, and $\mathbf{G}_{rd}$. Instead of this, the destination estimates only the effective channel gains. Thus, the channel estimation and data transmission phases leverage the same transmission protocol.

1) *The First Phase:* In this phase, the K users simultaneously transmit their pilot sequences to the destination and the relay. Denote by $\Phi \in \mathbb{C}^{K \times \tau}$ the pilot matrix assigned to the K users, where τ is the length of the training duration. Then, the destination and relay obtain, respectively, the following $N \times \tau$ and $M \times \tau$ pilot matrices:

$$\mathbf{Y}_{dp}^{(1)} = \sqrt{\tau P_{up}} \mathbf{G}_{ud} \Phi + \mathbf{N}_{dp}^{(1)}, \quad (1)$$

$$\mathbf{Y}_{rp}^{(1)} = \sqrt{\tau P_{up}} \mathbf{G}_{ur} \Phi + \mathbf{N}_{rp}^{(1)}, \quad (2)$$

where P_{up} is the normalized transmit power of each pilot symbol, $\mathbf{N}_{dp}^{(1)} \in \mathbb{C}^{M \times \tau}$ and $\mathbf{N}_{rp}^{(1)} \in \mathbb{C}^{N \times \tau}$ denote the corresponding noise matrices at the destination and the relay including i.i.d. Gaussian RVs with zero-mean and unit variance. We assume that the pilot sequences assigned to all K users are pairwise orthogonal, i.e., $\Phi \Phi^H = \mathbf{I}_K$. To guarantee this we must have $\tau \geq K$.

2) *The Second Phase:* In this phase, the relay forwards a scale version of its received pilot signal, i.e., it forwards $\mathbf{X}_p = \sqrt{\alpha_1} \mathbf{Y}_{rp}^{(1)}$ to the destination, where α_1 represents the normalization factor. This normalization factor is chosen so that the transmit power at the relay is constrained, i.e.,

$$\mathbb{E} \left\{ \|\mathbf{X}_p\|_F^2 \right\} = P_{rp}. \quad (3)$$

From (2) and (3), we obtain α_1 as

$$\begin{aligned} \alpha_1 &= \frac{P_{rp}}{\tau P_{up} \mathbb{E} \left\{ \text{Tr} [\mathbf{G}_{ur} \mathbf{G}_{ur}^H] \right\} + \mathbb{E} \left\{ \text{Tr} \left[\mathbf{N}_{rp}^{(1)} \left(\mathbf{N}_{rp}^{(1)} \right)^H \right] \right\}} \\ &= \frac{P_{rp}}{N \tau \left(P_{up} \sum_{k=1}^K \beta_{ur,k} + 1 \right)}. \end{aligned} \quad (4)$$

Note that, since the dimension of $\mathbf{X}_p$ is $N \times \tau$, the length of this phase is τ (symbols). With $\mathbf{X}_p = \sqrt{\alpha_1} \mathbf{Y}_{rp}^{(1)}$, the destination receives

$$\begin{aligned} \mathbf{Y}_{dp}^{(2)} &= \mathbf{G}_{rd} \mathbf{X}_p + \mathbf{N}_{dp}^{(2)} \\ &= \sqrt{\alpha_1 \tau P_{up}} \mathbf{G}_{rd} \mathbf{G}_{ur} \Phi + \sqrt{\alpha_1} \mathbf{G}_{rd} \mathbf{N}_{rp}^{(1)} + \mathbf{N}_{dp}^{(2)} \\ &= \sqrt{\alpha_1 \tau P_{up}} \mathbf{G}_d \Phi + \mathbf{N}_{dp}^{(2)}, \end{aligned} \quad (5)$$

where $\mathbf{N}_{dp}^{(2)} \in \mathbb{C}^{M \times \tau}$ is the noise matrix at the destination whose entries are i.i.d. $\mathcal{CN}(0, 1)$, and

$$\mathbf{G}_d = \mathbf{G}_{rd} \mathbf{G}_{ur}, \quad (6)$$

$$\mathbf{N}_{dp} = \sqrt{\alpha_1} \mathbf{G}_{rd} \mathbf{N}_{rp}^{(1)} + \mathbf{N}_{dp}^{(2)}. \quad (7)$$

The destination uses $\mathbf{Y}_{dp}^{(2)}$ and the minimum mean-square error (MMSE) method to estimate the effective channel matrices $\mathbf{G}_{ud}$ and $\mathbf{G}_d$ as follows [21]

$$\hat{\mathbf{g}}_{ud,k} = \left(\mathbf{g}_{ud,k} + \frac{\tilde{\mathbf{n}}_{dp,k}^{(1)}}{\sqrt{\tau P_{up}}} \right) \frac{\tau P_{up} \beta_{ud,k}}{\tau P_{up} \beta_{ud,k} + 1}, \quad (8)$$

$$\hat{\mathbf{g}}_{d,k} = \tilde{\sigma}_{d,k} \left(\mathbf{g}_{d,k} + \frac{\tilde{\mathbf{n}}_{dp,k}}{\sqrt{\alpha_1 \tau P_{up}}} \right), \quad (9)$$

where $\tilde{\mathbf{n}}_{dp,k}^{(1)}$, $\tilde{\mathbf{n}}_{dp,k}$ are the k -th columns of $\mathbf{N}_{dp}^{(1)} \Phi^H$ and $\mathbf{N}_{dp} \Phi^H$, respectively, including i.i.d. $\mathcal{CN}(0, 1)$ RVs, and

$$\tilde{\sigma}_{d,k} = \left(1 + \frac{\beta_{ur,k}^{-1} \beta_{rd}^{-1}}{N \alpha_1 \tau P_{up}} (N \alpha_1 \beta_{rd} + 1) \right)^{-1}.$$

With the channel estimates given by (8) and (9), the true channels $\mathbf{g}_{ud,k}$ and $\mathbf{g}_{d,k}$ can be expressed as

$$\mathbf{g}_{ud,k} = \hat{\mathbf{g}}_{ud,k} + \boldsymbol{\epsilon}_{ud,k}, \quad (10)$$

$$\mathbf{g}_{d,k} = \hat{\mathbf{g}}_{d,k} + \boldsymbol{\epsilon}_{d,k}, \quad (11)$$

where $\boldsymbol{\epsilon}_{ud,k}$ and $\boldsymbol{\epsilon}_{d,k}$ are the corresponding channel estimation error vectors.

From the property of the MMSE estimation technique, $\hat{\mathbf{g}}_{ud,k}$ is uncorrelated with $\boldsymbol{\epsilon}_{ud,k}$, and $\hat{\mathbf{g}}_{d,k}$ is uncorrelated with $\boldsymbol{\epsilon}_{d,k}$. Furthermore, since $\mathbf{g}_{ud,k}$ is a Gaussian vector, the elements of $\hat{\mathbf{g}}_{ud,k}$ and $\boldsymbol{\epsilon}_{ud,k}$ are i.i.d. $\mathcal{CN}(0, \sigma_{ud,k})$ and $\mathcal{CN}(0, \epsilon_{ud,k})$, respectively, where $\sigma_{ud,k} = \frac{\tau P_{up} \beta_{ud,k}^2}{\tau P_{up} \beta_{ud,k} + 1}$ and $\epsilon_{ud,k} = \frac{\beta_{ud,k}}{\tau P_{up} \beta_{ud,k} + 1}$. Note that, $\mathbf{g}_{d,k}$ is not Gaussian since it is a product of a Gaussian matrix and a Gaussian vector. Therefore, $\mathbf{g}_{d,k}$ and $\boldsymbol{\epsilon}_{d,k}$ are uncorrelated, but not independent.

B. Data Transmission

We use the same transmission protocol as in Section III-A.

1) *The First Phase:* All K users simultaneously send their symbols to the destination and the relay. Let $\mathbf{x} = [x_1 \dots x_K]^T$, where x_k is the transmitted symbol of the k -th user, with $\mathbb{E}\{\mathbf{x}\mathbf{x}^H\} = \mathbf{I}_K$, and $P_{u,k}$ be the normalized transmit power of user k . Then, the $N \times 1$ and $M \times 1$ signals received at the destination and the relay are, respectively, given by

$$\mathbf{y}_d^{(1)} = \mathbf{G}_{ud} \mathbf{P}_u^{1/2} \mathbf{x} + \mathbf{n}_d^{(1)}, \quad (12)$$

$$\mathbf{y}_r^{(1)} = \mathbf{G}_{ur} \mathbf{P}_u^{1/2} \mathbf{x} + \mathbf{n}_r^{(1)}, \quad (13)$$

where $\mathbf{n}_d^{(1)} \in \mathbb{C}^{M \times 1}$ and $\mathbf{n}_r^{(1)} \in \mathbb{C}^{N \times 1}$ are the noise vectors with i.i.d. $\mathcal{CN}(0, 1)$ components. In addition, $\mathbf{P}_u = \text{diag}\{P_{u,1}, \dots, P_{u,K}\}$.

2) *The Second Phase:* The relay first amplifies its received signal $\mathbf{y}_r^{(1)}$ as $\mathbf{x}_r = \sqrt{\alpha_2} \mathbf{y}_r^{(1)}$, where α_2 is the normalization factor which is selected so that the normalized transmit power at the relay is constrained by

$$\mathbb{E}\{\|\mathbf{x}_r\|^2\} \leq P_{r,\text{th}}. \quad (14)$$

Then, it broadcasts this amplified version to the destination. Substituting (13) into (14) and using $\mathbf{x}_r = \sqrt{\alpha_2} \mathbf{y}_r^{(1)}$, we obtain

$$\begin{aligned} \alpha_2 &\leq \frac{P_{r,\text{th}}}{\left(\mathbb{E}\left\{ \text{Tr}(\mathbf{G}_{ur} \mathbf{P}_u \mathbf{G}_{ur}^H) \right\} + \mathbb{E}\left\{ \|\mathbf{n}_r^{(1)}\|^2 \right\} \right)} \\ &= \frac{P_{r,\text{th}}}{N \left(\sum_{k=1}^K P_{u,k} \beta_{ur,k} + 1 \right)}. \end{aligned} \quad (15)$$

With the transmit signal vector $\mathbf{x}_r$, the destination receives

$$\begin{aligned} \mathbf{y}_d^{(2)} &= \mathbf{G}_{rd} \mathbf{x}_r + \mathbf{n}_d^{(2)} \\ &= \sqrt{\alpha_2} \mathbf{G}_{rd} \mathbf{G}_{ur} \mathbf{P}_u^{1/2} \mathbf{x} + \sqrt{\alpha_2} \mathbf{G}_{rd} \mathbf{n}_r^{(1)} + \mathbf{n}_d^{(2)} \\ &= \sqrt{\alpha_2} \mathbf{G}_d \mathbf{P}_u^{1/2} \mathbf{x} + \mathbf{n}_d, \end{aligned} \quad (16)$$

where $\mathbf{n}_d^{(2)} \in \mathbb{C}^{M \times 1}$ is the corresponding noise vector whose entries are i.i.d. $\mathcal{CN}(0, 1)$ RVs, and $\mathbf{n}_d = \sqrt{\alpha_2} \mathbf{G}_{rd} \mathbf{n}_r^{(1)} + \mathbf{n}_d^{(2)}$.

By combining the received signal vectors at the destination in the first phase and the second phase given by (12) and (16), respectively, we obtain

$$\begin{aligned} \mathbf{y}_d &= \begin{bmatrix} \mathbf{y}_d^{(1)} \\ \mathbf{y}_d^{(2)} \end{bmatrix} = \begin{bmatrix} \mathbf{G}_{ud} \mathbf{P}_u^{1/2} \mathbf{x} + \mathbf{n}_d^{(1)} \\ \sqrt{\alpha_2} \mathbf{G}_d \mathbf{P}_u^{1/2} \mathbf{x} + \mathbf{n}_d \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{G}_{ud} \mathbf{P}_u^{1/2} \\ \sqrt{\alpha_2} \mathbf{G}_d \mathbf{P}_u^{1/2} \end{bmatrix} \mathbf{x} + \begin{bmatrix} \mathbf{n}_d^{(1)} \\ \mathbf{n}_d \end{bmatrix} = \mathbf{G} \mathbf{x} + \tilde{\mathbf{n}}, \end{aligned} \quad (17)$$

where $\mathbf{G} = \begin{bmatrix} \mathbf{G}_{ud} \mathbf{P}_u^{1/2} \\ \sqrt{\alpha_2} \mathbf{G}_d \mathbf{P}_u^{1/2} \end{bmatrix}$, and $\tilde{\mathbf{n}} = \begin{bmatrix} \mathbf{n}_d^{(1)} \\ \mathbf{n}_d \end{bmatrix}$.

3) *Data Detection:* From the received signal vector $\mathbf{y}_d$ in (17), the destination will detect the desired signal $\mathbf{x}$. To do this, the MRC scheme is implemented using the channel estimates acquired in the channel estimation period (see Section III-A). The $K \times 1$ combined signal vector after using MRC is

$$\tilde{\mathbf{y}}_d = \hat{\mathbf{G}}^H \mathbf{G} \mathbf{x} + \hat{\mathbf{G}}^H \tilde{\mathbf{n}}, \quad (18)$$

where $\hat{\mathbf{G}} = \begin{bmatrix} \hat{\mathbf{G}}_{ud} \mathbf{P}_u^{1/2} \\ \sqrt{\alpha_2} \hat{\mathbf{G}}_d \mathbf{P}_u^{1/2} \end{bmatrix}$. To decode x_k , the k -th entry of $\tilde{\mathbf{y}}_d$ is used.

IV. TRANSMISSION SCHEME 2: CSI AT THE RELAY

In this transmission scheme, more advanced signal processing can be performed at the relay to improve the system performance. To do this, the relay needs to know the channels. The channels from the K users can be estimated at the relay through the training phase, and the channel estimation overhead depends only on the number of users. The channels

from the destination to the relay can be also estimated. But the corresponding channel estimation overhead depends on the number of destination antennas which is very large. Therefore, in this transmission scheme, we consider the case where the relay estimates only the channels from the users.

A. Channel Estimation

1) *The First Phase:* Similarly to Transmission Scheme 1, all K users will send orthogonal pilot sequences to the relay. But in this case, the relay will estimate the channel from the K users. The received pilot matrix at the relay is $\mathbf{Y}_{rp}^{(1)}$ given by (2). After projecting $\mathbf{Y}_{rp}^{(1)}$ onto $\mathbf{Y}_{rp}^{(1)}$, we obtain the MMSE estimate of $\mathbf{g}_{ur,k}$ as

$$\hat{\mathbf{g}}_{ur,k} = \left(\sqrt{\tau P_{up}} \mathbf{g}_{ur,k} + \tilde{\mathbf{n}}_{rp,k}^{(1)} \right) \left(\frac{\sqrt{\tau P_{up}} \beta_{ur,k}}{\tau P_{up} \beta_{ur,k} + 1} \right), \quad (19)$$

where $\tilde{\mathbf{n}}_{rp,k}^{(1)} = \left[\mathbf{N}_{rp}^{(1)} \Phi^H \right]_k$ with i.i.d. $\mathcal{CN}(0, 1)$ RVs. We denote by $\epsilon_{ur,k}$ the estimation error vector of $\mathbf{g}_{ur}$, i.e.,

$$\mathbf{g}_{ur,k} = \hat{\mathbf{g}}_{ur,k} + \epsilon_{ur,k}. \quad (20)$$

Then, $\hat{\mathbf{g}}_{ur,k}$ and $\epsilon_{ur,k}$ are independent. Furthermore, the elements of $\hat{\mathbf{g}}_{ur,k}$ are i.i.d. $\mathcal{CN}(0, \sigma_{ur,k})$, and the elements of $\epsilon_{ur,k}$ are i.i.d. $\mathcal{CN}(0, \epsilon_{ur,k})$, where $\sigma_{ur,k} = \frac{\tau P_{up} \beta_{ur,k}^2}{\tau P_{up} \beta_{ur,k} + 1}$ and $\epsilon_{ur,k} = \frac{\beta_{ur,k}}{\tau P_{up} \beta_{ur,k} + 1}$.

2) *The Second Phase:* Since the relay knows the channel estimate of $\mathbf{G}_{ur}$, it can process the received pilot signal before forwarding it to the destination. Here, we assume that the relay uses the simple MRC scheme to combine the received signals. MRC scheme is used because it has low computational complexity and can be implemented in a distributed manner [22]. In addition, its performance is remarkably good when the number of antennas is large. To improve the system performance, more complicated signal processing techniques, e.g. zero-forcing processing, can be used. However, they require more computing power than MRC does. A comprehensive study of the tradeoff between the system performance and the implementation complexity of these processing techniques will be part of our future work. With MRC, the amplified version of the pilot signal received at the relay is

$$\mathbf{X}_p = \sqrt{\alpha_3} \hat{\mathbf{G}}_{ur}^H \mathbf{Y}_{rp}^{(1)},$$

where α_3 is the normalization factor, which is chosen to satisfy the power constraint at the relay:

$$\mathbb{E} \left\{ \|\mathbf{X}_p\|_F^2 \right\} = P_{rp}. \quad (21)$$

Thus, we can obtain the normalization factor α_3 in closed-form as in (22) shown at the top of the page.

The relay does not know the channel to the destination.³ Thus, it just forwards the MRC pilot signal $\mathbf{X}_p$ towards the destination without applying any type of precoding. Since the dimension of the combined pilot signal $\mathbf{X}_p$ is $K \times \tau$, the relay

³At the relay, to acquire the channel between the relay and the destination $\mathbf{G}_{rd}$, the channel estimation overhead will be proportional to the number of destination antennas which is prohibitively large in the considered massive MIMO topology.

needs only K antennas (out of N antennas) to forward $\mathbf{X}_p$. The antenna selection is out of scope of this work. Here, we just consider a simple method: the relay randomly chooses K antennas from N antennas. Note that we do not need to repeat this step for each transmission. Denote by $\bar{\mathbf{G}}_{rd} \in \mathbb{C}^{M \times K}$ the channel from the K selected antennas at the relay to the destination. Then, the $M \times \tau$ received pilot signal at the destination can be expressed as:

$$\begin{aligned} \mathbf{Y}_{dp}^{(2)} &= \bar{\mathbf{G}}_{rd} \mathbf{X}_p + \mathbf{N}_{dp}^{(2)} \\ &= \sqrt{\alpha_3 \tau P_{up}} \bar{\mathbf{G}}_{rd} \hat{\mathbf{G}}_{ur}^H \mathbf{G}_{ur} \Phi + \sqrt{\alpha_3} \bar{\mathbf{G}}_{rd} \hat{\mathbf{G}}_{ur}^H \mathbf{N}_{rp}^{(1)} + \mathbf{N}_{pd}^{(2)} \\ &= \sqrt{\alpha_3 \tau P_{up}} \bar{\mathbf{G}}_d \Phi + \bar{\mathbf{N}}_{pd}, \end{aligned} \quad (23)$$

where

$$\bar{\mathbf{G}}_d = \bar{\mathbf{G}}_{rd} \hat{\mathbf{G}}_{ur}^H \mathbf{G}_{ur}, \quad (24)$$

$$\bar{\mathbf{N}}_{pd} = \sqrt{\alpha_3} \bar{\mathbf{G}}_{rd} \hat{\mathbf{G}}_{ur}^H \mathbf{N}_{rp}^{(1)} + \mathbf{N}_{pd}^{(2)}. \quad (25)$$

From (23), by using the same methodology in Section IV-A1, we obtain the MMSE estimate of $\bar{\mathbf{g}}_{d,k}$ at the destination as

$$\hat{\bar{\mathbf{g}}}_{d,k} = \left(\bar{\mathbf{g}}_{d,k} + \frac{\tilde{\mathbf{n}}_{dp,k}}{\sqrt{\alpha_2 \tau P_{up}}} \right) \tilde{\sigma}_{d,k}, \quad (26)$$

where

$$\tilde{\sigma}_{d,k} = \left(1 + \frac{N \alpha_3 \beta_{rd} \sum_{k=1}^K \sigma_{ur,k} + 1}{\alpha_3 \tau P_{up} \text{Var}(\bar{\mathbf{g}}_{d,k})} \right)^{-1}, \quad (27)$$

with

$$\text{Var}(\bar{\mathbf{g}}_{d,k}) = N \beta_{rd} (\sigma_{ur,k} + \epsilon_{ur,k}) \sum_{k'=1}^K \sigma_{ur,k'}. \quad (28)$$

B. Data Transmission

We use the same transmission protocol as in Section IV-A.

1) *The First Phase:* This phase is the same as that of Transmission Scheme 2, see Section III-B1. Thus, the $N \times 1$ and $M \times 1$ received signals at the relay and destination are given by (13), and (12), respectively.

2) *The Second Phase:* The relay combines the received signal $\mathbf{y}_r^{(1)}$ using the channel estimate $\hat{\mathbf{G}}_{ur}^H$ and forwards the processed signal $\mathbf{x}_r = \sqrt{\alpha_4} \hat{\mathbf{G}}_{ur}^H \mathbf{y}_r^{(1)}$ to the destination. Here, α_4 is the power normalization factor at the relay, which is chosen to satisfy

$$\mathbb{E} \left\{ \|\mathbf{x}_r\|^2 \right\} \leq P_{r,\text{th}}. \quad (29)$$

From (29), α_4 is given by (30) shown at the top of the page, where $\mathbf{E}_{ur}$ is the estimation error matrix associated with $\mathbf{G}_{ur}$.

Note that, as in the channel estimation period, to forward $\mathbf{x}_r$ to the destination, the relay uses only K antennas. Thus, the $M \times 1$ received signal vector at the destination is given by

$$\begin{aligned} \mathbf{y}_d^{(2)} &= \bar{\mathbf{G}}_{rd} \mathbf{x}_r + \mathbf{n}_d^{(2)} \\ &= \sqrt{\alpha_4} \bar{\mathbf{G}}_{rd} \hat{\mathbf{G}}_{ur}^H \mathbf{G}_{ur} \mathbf{P}_u^{1/2} \mathbf{x} + \sqrt{\alpha_4} \bar{\mathbf{G}}_{rd} \hat{\mathbf{G}}_{ur}^H \mathbf{n}_r^{(1)} + \mathbf{n}_d^{(2)} \\ &= \sqrt{\alpha_4} \bar{\mathbf{G}}_d \mathbf{P}_u^{1/2} \mathbf{x} + \bar{\mathbf{n}}_d, \end{aligned} \quad (31)$$

where $\bar{\mathbf{n}}_d = \sqrt{\alpha_4} \bar{\mathbf{G}}_{rd} \hat{\mathbf{G}}_{ur}^H \mathbf{n}_r^{(1)} + \mathbf{n}_d^{(2)}$.

$$\begin{aligned}
\alpha_3 &= \frac{P_{rp}}{\tau P_{up} \mathbb{E} \left\{ \text{Tr} \left(\hat{\mathbf{G}}_{ur}^H \mathbf{G}_{ur} \mathbf{G}_{ur}^H \hat{\mathbf{G}}_{ur} \right) \right\} + \mathbb{E} \left\{ \text{Tr} \left(\hat{\mathbf{G}}_{ur}^H \mathbf{N}_{rp}^{(1)} \left(\mathbf{N}_{rp}^{(1)} \right)^H \hat{\mathbf{G}}_{ur} \right) \right\}} \\
&= \frac{P_{rp}}{\tau P_{up} \left(\sum_{k=1}^K \sum_{\substack{i=1 \\ i \neq k}}^K \left(N(N+1) \sigma_{ur,k}^2 + N \sigma_{ur,k} \sigma_{ur,i} \right) + N \sum_{k=1}^K \sigma_{ur,k} \sum_{k'=1}^K \epsilon_{ur,k'} \right) + N \tau \sum_{k=1}^K \sigma_{ur,k}}. \quad (22)
\end{aligned}$$

$$\begin{aligned}
\alpha_4 &\leq \frac{P_{r,th}}{\mathbb{E} \left\{ \text{Tr} \left(\hat{\mathbf{G}}_{ur}^H \hat{\mathbf{G}}_{ur} \mathbf{P}_u \hat{\mathbf{G}}_{ur}^H \hat{\mathbf{G}}_{ur} \right) \right\} + \mathbb{E} \left\{ \text{Tr} \left(\hat{\mathbf{G}}_{ur}^H \mathbf{E}_{ur} \mathbf{P}_u \mathbf{E}_{ur}^H \hat{\mathbf{G}}_{ur} \right) \right\} + \mathbb{E} \left\{ \text{Tr} \left(\hat{\mathbf{G}}_{ur}^H \hat{\mathbf{G}}_{ur} \right) \right\}} \\
&= \frac{P_{r,th}}{N \sum_{k=1}^K \sum_{\substack{i=1 \\ i \neq k}}^K P_{u,k} \left((N+1) \sigma_{ur,k}^2 + \sigma_{ur,i} \sigma_{ur,k} + \sigma_{ur,k} \epsilon_{ur,k} + \sigma_{ur,i} \epsilon_{ur,k} \right) + N \sum_{k=1}^K \sigma_{ur,k}}, \quad (30)
\end{aligned}$$

From (12) and (31), the received signal vector at the destination after two phases can be collected as

$$\begin{aligned}
\mathbf{y}_d &= \begin{bmatrix} \mathbf{y}_d^{(1)} \\ \mathbf{y}_d^{(2)} \end{bmatrix} = \begin{bmatrix} \mathbf{G}_{ud} \mathbf{P}_u^{1/2} \mathbf{x} + \mathbf{n}_d^{(1)} \\ \sqrt{\alpha_4} \hat{\mathbf{G}}_d \mathbf{P}_u^{1/2} \mathbf{x} + \tilde{\mathbf{n}}_d \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{G}_{ud} \mathbf{P}_u^{1/2} \\ \sqrt{\alpha_4} \hat{\mathbf{G}}_d \mathbf{P}_u^{1/2} \end{bmatrix} \mathbf{x} + \begin{bmatrix} \mathbf{n}_d^{(1)} \\ \tilde{\mathbf{n}}_d \end{bmatrix} = \tilde{\mathbf{G}} \mathbf{x} + \tilde{\mathbf{n}}, \quad (32)
\end{aligned}$$

where $\tilde{\mathbf{G}} = \begin{bmatrix} \mathbf{G}_{ud} \mathbf{P}_u^{1/2} \\ \sqrt{\alpha_4} \hat{\mathbf{G}}_d \mathbf{P}_u^{1/2} \end{bmatrix}$ and $\tilde{\mathbf{n}} = \begin{bmatrix} \mathbf{n}_d^{(1)} \\ \tilde{\mathbf{n}}_d \end{bmatrix}$.

3) *Data Detection*: Similarly, since the destination knows the estimate of $\tilde{\mathbf{G}}$, it can use MRC to detect all signals transmitted from all K users. The $K \times 1$ combined signal vector after using MRC is

$$\bar{\mathbf{y}}_d = \hat{\mathbf{G}}^H \tilde{\mathbf{G}} \mathbf{x} + \hat{\mathbf{G}}^H \tilde{\mathbf{n}}, \quad (33)$$

where $\hat{\mathbf{G}} = \begin{bmatrix} \hat{\mathbf{G}}_{ud} \mathbf{P}_u^{1/2} \\ \sqrt{\alpha_4} \hat{\hat{\mathbf{G}}}_d \mathbf{P}_u^{1/2} \end{bmatrix}$. The k -th stream (the k -th element) of $\bar{\mathbf{y}}_d$ is used to decode x_k .

V. ACHIEVABLE SPECTRAL EFFICIENCY

In this section, we derive closed-form expressions for the spectral efficiency of Transmission Schemes I and II, using the use-and-then-forget bounding technique in [4].

A. Transmission Scheme 1

From (18), the k -th element of $\bar{\mathbf{y}}_d$ used to decode x_k is

$$\begin{aligned}
\tilde{y}_{d,k} &= \hat{\mathbf{g}}_k^H \mathbf{g}_k x_k + \sum_{\substack{i=1 \\ i \neq k}}^K \hat{\mathbf{g}}_k^H \mathbf{g}_i x_i + \hat{\mathbf{g}}_k^H \tilde{\mathbf{n}} \\
&= \mathbb{E} \left\{ \hat{\mathbf{g}}_k^H \mathbf{g}_k \right\} x_k + \tilde{n}_{d,k}, \quad (34)
\end{aligned}$$

where

$$\tilde{n}_{d,k} = (\hat{\mathbf{g}}_k^H \mathbf{g}_k - \mathbb{E} \left\{ \hat{\mathbf{g}}_k^H \mathbf{g}_k \right\}) x_k + \sum_{\substack{i=1 \\ i \neq k}}^K \hat{\mathbf{g}}_k^H \mathbf{g}_i x_i + \hat{\mathbf{g}}_k^H \tilde{\mathbf{n}}. \quad (35)$$

We treat the first term of the RHS of (34) as the desired signal part, and the second term as the effective noise. Clearly, the desired signal and the effective noise are uncorrelated. Therefore, by using the use-and-then-forget bounding method in [4], we can obtain the following achievable spectral efficiency for user k :

$$R_{d,k} = \frac{T-2\tau}{2T} \log_2 (1 + \gamma_k) \quad (\text{bit/s/Hz}), \quad (36)$$

where T is the length of the coherence interval, and the signal-to-interference-plus-noise ratio (SINR) γ_k is given by

$$\gamma_k = \frac{|\mathbb{E} \left\{ \hat{\mathbf{g}}_k^H \mathbf{g}_k \right\}|^2}{\text{Var}(\hat{\mathbf{g}}_k^H \mathbf{g}_k) + \text{IU}_k + \text{AN}_k}, \quad (37)$$

with

$$\text{IU}_k = \sum_{\substack{i=1 \\ i \neq k}}^K \mathbb{E} \left\{ |\hat{\mathbf{g}}_k^H \mathbf{g}_i|^2 \right\} \quad \text{and} \quad \text{AN}_k = \mathbb{E} \left\{ |\hat{\mathbf{g}}_k^H \tilde{\mathbf{n}}|^2 \right\}. \quad (38)$$

Note that the pre-log factor $\frac{T-2\tau}{2T}$ in the spectral efficiency expression (36) reflects the overhead due to the use of the half-duplex AF relaying technique and the channel estimation.

Remark 1: The achievable spectral efficiency given by (36) is obtained via the use-and-then-forget bounding technique which relies on the channel statistics only. This technique is widely used in massive MIMO literature (see [4] and references herein) because it yields a simple and insightful closed-form solution which facilitates system design, resource allocation to name but a few. In addition, because of the channel hardening property of massive MIMO, detection using only the channel statistics is nearly optimal. To see this more quantitatively, we consider a genie-aided destination case, where the destination is assumed to know perfectly the instantaneous channel gain. Under a genie-aided destination, we can obtain the following achievable spectral efficiency:

$$R_{d,k}^g = \frac{T-2\tau}{2T} \mathbb{E} \left\{ \log_2 \left(1 + \frac{|\hat{\mathbf{g}}_k^H \mathbf{g}_k|^2}{\sum_{\substack{i=1 \\ i \neq k}}^K |\hat{\mathbf{g}}_k^H \mathbf{g}_i|^2 + \mathbb{E} \left\{ |\hat{\mathbf{g}}_k^H \tilde{\mathbf{n}}|^2 \right\}} \right) \right\}, \quad (39)$$

Fig. 2. Sum spectral efficiency versus the number of destination antennas for Transmission Scheme 1, generated by one large-scale fading snapshot. Here, we choose $P_{u,k} = P_r = 0.1$ W for all k , $P_{rp} = P_{u,k}$ and $P_{up} = P_{u,k}$. The details of the system setup are discussed in Section VII.

where the expectation inside the log function is taken over the noise term. Figure 2 shows a comparison between the achievable spectral efficiency (36) and the genie-aided spectral efficiency (39), for one snapshot of large-scale fading. We can see that the gap is remarkably small. This verifies the effectiveness of the use-and-then-forget bounding technique.

Proposition 1: With Transmission Scheme I, the achievable spectral efficiency of user k is $R_{d,k} = \frac{T-2\tau}{2T} \log_2(1 + \gamma_k)$, where γ_k can be expressed in closed-form as

$$\gamma_k = \frac{M^2 P_{u,k}^2 (\sigma_{ud,k} + N \alpha_2 \tilde{\sigma}_{d,k} \beta_{ur,k} \beta_{rd})^2}{\text{Var}(\hat{\mathbf{g}}_k^H \mathbf{g}_k) + \text{IU}_k + \text{AN}_k}, \quad (40)$$

with $\text{Var}(\hat{\mathbf{g}}_k^H \mathbf{g}_k)$, IU_k , and AN_k are, respectively, given by (40), (41), and (42), shown at the top of the next page.

Proof: See Appendix A. ■

We want to stress that the closed-form expression for the achievable spectral efficiency provided in Proposition 1 is an exact result. Thus, it is valid for any finite number of users, relay and destination antennas.

Remark 2: The result in Proposition 1 enables us to obtain some basic insights into the performance in the massive MIMO regime (i.e., $M \rightarrow \infty$). We assume that $\frac{M}{N} = \kappa$ is fixed, and the relay transmits with full power, i.e., $\alpha_2 = \frac{P_{r,\text{th}}}{N(\sum_{k=1}^K P_{u,k} \beta_{ur,k} + 1)}$.

- 1) From (40), we have $\gamma_k \rightarrow \infty$, as $M \rightarrow \infty$. We can see that when the number of destination antennas goes to infinity, the inter-user interference and noise can be cancelled out. As a result, the spectral efficiency increases without bound.
- 2) Let $P_{u,k} = \frac{E_{u,k}}{M}$, where $E_{u,k}$ is fixed (independent of M). As $M \rightarrow \infty$, we have

$$\gamma_k \rightarrow \frac{E_{u,k} (\sigma_{ud,k} + P_{r,\text{th}} \tilde{\sigma}_{d,k} \beta_{ur,k} \beta_{rd})^2}{\text{AN}_k^\infty}, \quad (43)$$

where

$$\begin{aligned} \text{AN}_k^\infty &= \tilde{\sigma}_{d,k}^2 P_{r,\text{th}} (P_{r,\text{th}}(\kappa + 1) + 1) \left(\beta_{ur,k} + \frac{1}{\tau P_{up}} \right) \\ &+ \sigma_{ud,k} + \frac{P_{r,\text{th}}^2 \tilde{\sigma}_{d,k}^2 \beta_{rd}}{\varsigma_1 \tau P_{up}} + \frac{P_{r,\text{th}} \tilde{\sigma}_{d,k}^2}{\varsigma_1 \tau P_{up}}, \end{aligned} \quad (44)$$

and where $\varsigma_1 = \alpha_1/N$.

We can see that the spectral efficiency converges to a constant as $M \rightarrow \infty$. It means that by using a large number of antennas, the transmit power of each user can be cut down proportionally to $1/M$ without performance degradation.

B. Transmission Scheme 2

From (33), the k -th element of $\bar{\mathbf{y}}_d$ is

$$\begin{aligned} \bar{y}_{d,k} &= \hat{\mathbf{g}}_k^H \bar{\mathbf{g}}_k x_k + \sum_{\substack{i=1 \\ i \neq k}}^K \hat{\mathbf{g}}_k^H \bar{\mathbf{g}}_i x_i + \hat{\mathbf{g}}_k^H \bar{\mathbf{n}} \\ &= \mathbb{E} \{ \hat{\mathbf{g}}_k^H \bar{\mathbf{g}}_k \} x_k + \bar{n}_{d,k}, \end{aligned} \quad (45)$$

where

$$\bar{n}_{d,k} = (\hat{\mathbf{g}}_k^H \bar{\mathbf{g}}_k - \mathbb{E} \{ \hat{\mathbf{g}}_k^H \bar{\mathbf{g}}_k \}) x_k + \sum_{\substack{i=1 \\ i \neq k}}^K \hat{\mathbf{g}}_k^H \bar{\mathbf{g}}_i x_i + \hat{\mathbf{g}}_k^H \bar{\mathbf{n}}. \quad (46)$$

Thus, the corresponding achievable spectral efficiency of the k -th user is

$$\bar{R}_{d,k} = \frac{T-2\tau}{2T} \log_2(1 + \bar{\gamma}_k) \quad (\text{bit/s/Hz}), \quad (47)$$

where $\bar{\gamma}_k$ is given by

$$\bar{\gamma}_k = \frac{|\mathbb{E} \{ \hat{\mathbf{g}}_k^H \bar{\mathbf{g}}_k \}|^2}{\text{Var}(\hat{\mathbf{g}}_k^H \bar{\mathbf{g}}_k) + \bar{\text{IU}}_k + \bar{\text{AN}}_k}, \quad (48)$$

where

$$\bar{\text{IU}}_k = \sum_{\substack{i=1 \\ i \neq k}}^K \mathbb{E} \{ |\hat{\mathbf{g}}_k^H \bar{\mathbf{g}}_i|^2 \}, \quad (49)$$

$$\bar{\text{AN}}_k = \mathbb{E} \{ |\hat{\mathbf{g}}_k^H \bar{\mathbf{n}}|^2 \}. \quad (50)$$

Proposition 2: With Transmission Scheme II, the achievable spectral efficiency of user k is $\bar{R}_{d,k} = \frac{T-2\tau}{2T} \log_2(1 + \bar{\gamma}_k)$, where $\bar{\gamma}_k$ can be expressed in closed-form as

$$\bar{\gamma}_k = \frac{\bar{\text{N}}_k}{\text{Var}(\hat{\mathbf{g}}_k^H \bar{\mathbf{g}}_k) + \bar{\text{IU}}_k + \bar{\text{AN}}_k}, \quad (51)$$

where $\bar{\text{N}}_k$, $\text{Var}(\hat{\mathbf{g}}_k^H \bar{\mathbf{g}}_k)$, $\bar{\text{IU}}_k$, and $\bar{\text{AN}}_k$ are approximately given by (52), (53), (54), and (55) on the next page. Fur-

$$\begin{aligned} \text{Var}(\hat{\mathbf{g}}_k^H \mathbf{g}_k) &= P_{u,k}^2 \left(M\sigma_{ud,k}^2 + MN\alpha_2^2 \tilde{\sigma}_{d,k}^2 (M+N+1) \beta_{ur,k}^2 \beta_{rd}^2 + M\epsilon_{ud,k} \sigma_{ud,k} \right. \\ &\quad \left. + \frac{\alpha_2^2 \tilde{\sigma}_{d,k}^2}{\tau P_{up}} MN \beta_{ur,k} \beta_{rd}^2 (M+N) + \frac{\alpha_2^2 \tilde{\sigma}_{d,k}^2}{\alpha_1 \tau P_{up}} MN \beta_{ur,k} \beta_{rd} \right), \end{aligned} \quad (40)$$

$$\begin{aligned} \text{IU}_k &= \sum_{\substack{i=1 \\ i \neq k}}^K P_{u,k} P_{u,i} \left(M\sigma_{ud,k} \sigma_{ud,i} + M\sigma_{ud,k} \epsilon_{ud,i} + MN(M+N) \alpha_2^2 \tilde{\sigma}_{d,k}^2 \beta_{rd}^2 \left(\beta_{ur,k} + \frac{1}{\tau P_{up}} \right) \beta_{ur,i} \right. \\ &\quad \left. + \frac{\alpha_2^2 \tilde{\sigma}_{d,k}^2}{\alpha_1 \tau P_{up}} MN \beta_{rd} \beta_{ur,i} \right), \end{aligned} \quad (41)$$

$$\text{AN}_k = P_{u,k} \left(M\sigma_{ud,k} + MN \tilde{\sigma}_{d,k}^2 \alpha_2 \beta_{rd} (\alpha_2 (M+N) + 1) \left(\beta_{ur,k} + \frac{1}{\tau P_{up}} \right) + \frac{\alpha_2^2 \tilde{\sigma}_{d,k}^2}{\alpha_1 \tau P_{up}} MN \beta_{rd} + M \frac{\alpha_2 \tilde{\sigma}_{d,k}^2}{\alpha_1 \tau P_{up}} \right). \quad (42)$$

thermore, $\mathcal{I}_k$, $\mathcal{G}_k$, $\mathcal{J}_k$, $\mathcal{M}_k$, $\mathcal{N}_k$, and $\mathcal{O}_k$ in (53), (54) and (55) are

$$\mathcal{I}_k = M^2 \beta_{rd}^2 \alpha_4^2 \tilde{\sigma}_{d,k}^2 N(N+1)(N+2) \sigma_{ur,k}^3,$$

$$\mathcal{G}_k = M^2 \beta_{rd}^2 \alpha_4^2 \tilde{\sigma}_{d,k}^2 N(N+1) \sum_{\substack{i=1 \\ i \neq k}}^K \sigma_{ur,k} \sigma_{ur,i}$$

$$\mathcal{J}_k = M^2 \beta_{rd}^2 \alpha_4^2 \tilde{\sigma}_{d,k}^2 N \sum_{\substack{i=1 \\ i \neq k}}^K \sum_{\substack{j=1 \\ j \neq k,i}}^K \sigma_{ur,k} \sigma_{ur,i} \sigma_{ur,j},$$

$$\mathcal{M}_k = M^2 \beta_{rd}^2 \alpha_4^2 \tilde{\sigma}_{d,k}^2 N(N+1) \sum_{i=1}^K \sigma_{ur,i} \epsilon_{ur,k}$$

$$\mathcal{N}_k = M \beta_{rd} \alpha_4^2 \tilde{\sigma}_{d,k}^2 N(N+1) \sum_{k=1}^K \sigma_{ur,k}^2,$$

respectively,

$$\mathcal{O}_k = M \beta_{rd} \alpha_4^2 \tilde{\sigma}_{d,k}^2 \sum_{k=1}^K \sum_{\substack{i=1 \\ i \neq k}}^K \sigma_{ur,k} \sigma_{ur,i}.$$

Proof: We first use $M\beta_{rd}\mathbf{I}_N$ to approximate $\mathbf{G}_{rd}^H \mathbf{G}_{rd}$, and then follow a similar methodology as in Section 1.⁴ The detailed derivation of (52) is provided in Section VIII-B. Other expressions follow the same method. ■

The closed-form expression provided in Proposition 2 is an approximating result derived under the assumption that the number of destination antennas M is large. From a theoretical point of view, this closed-form result is valid for large M . However, from the numerical results in Section VII, we will observe that this approximation is very tight (and hence, valid) even when M is small (e.g., $M = 10$).

VI. MAX-MIN POWER ALLOCATION

In this section, we consider a max-min power optimization problem. With max-min power allocation, we optimally select the transmit powers at the users and the relay to maximize

⁴The approximation comes from the law of large numbers, i.e. $\frac{\mathbf{G}_{rd}^H \mathbf{G}_{rd}}{M} \rightarrow \beta_{rd} \mathbf{I}_N$, as $M \rightarrow \infty$.

the lowest spectral efficiency of all users, and hence, improve the network connectivity. This technique has been widely investigated in the massive MIMO space [4]. In addition, our max-min power control method can be directly extended to a max-min weighted spectral efficiency problem: $\max \min\{w_k R_k\}$, where $w_k > 0$ is the weighting factor associated with user k . A user with lower priority will be assigned a larger weighting factor. The power allocation can be performed at a central processing unit or at the destination. The resulting optimal power values can be sent to the users and the relay via dedicated channels. Note that this is done for each large-scale fading realization which changes very slowly with time.

A. Transmission Scheme 1

For Transmission Scheme 1, the max-min power allocation is formulated as follows:

$$\max_{P_{u,k}, \alpha_2} \min_{k=1, \dots, K} R_{d,k} \quad (56a)$$

$$\text{s.t. } 0 \leq P_{u,k} \leq P_{u,\text{th}}, \quad k = 1, \dots, K, \quad (56b)$$

$$N\alpha_2 \left(\sum_{k=1}^K P_{u,k} \beta_{ur,k} + 1 \right) \leq P_{r,\text{th}}, \quad (56c)$$

where $R_{d,k}$ is spectral efficiency of user k given in (36). The constraints (56b) and (56c) represents the power constraints at the user and the relay, respectively, with $P_{u,\text{th}}$ and $P_{r,\text{th}}$ being the corresponding maximum powers. Since $\log(1+x)$ is an increasing function with respect to x , problem (56) is equivalent to

$$\max_{P_{u,k}, \alpha_2} \min_{k=1, \dots, K} \gamma_k \quad (57a)$$

$$\text{s.t. } (56b), (56c). \quad (57b)$$

We can see that problem (57) is not a convex problem (not jointly convex with respect to α_2 and $P_{u,k}$). To tackle this issue, we propose to use an efficient and simple iterative method. More precisely, we decouple problem (57) into two sub-problems: solving $P_{u,k}$ with fixed α_2 , and solving α_2 with fixed $P_{u,k}$. As a consequence, a solution for (57) is obtained by alternatively solving these sub-problems.

$$\bar{\mathbf{N}}\mathbf{U}_k \approx M^2 P_{u,k}^2 \left[\sigma_{ud,k} + \alpha_4 \tilde{\sigma}_{d,k} \beta_{rd} \left(N(N+1) \sigma_{ur,k}^2 + N \sum_{\substack{k=1 \\ k' \neq k}}^K \sigma_{ur,k} \sigma_{ur,k'} + N \sum_{k'=1}^K \epsilon_{ur,k} \sigma_{ur,k'} \right) \right]^2, \quad (52)$$

$$\begin{aligned} \text{Var}(\hat{\mathbf{g}}_k^H \hat{\mathbf{g}}_k) &\approx P_{u,k}^2 \left(M(M+1) \sigma_{ud,k}^2 + M \epsilon_{ud,k} \sigma_{ud,k} + \mathcal{I}_k \left((N+3) \sigma_{ur,k} + \sum_{\substack{i=1 \\ i \neq k}}^K \sigma_{ur,i} + \sum_{\substack{j=1 \\ j \neq k}}^K \sigma_{ur,j} \right. \right. \\ &\quad \left. \left. + 2 \epsilon_{ur,k} + \frac{1}{\tau P_{\text{up}}} \right) + \mathcal{G}_k \left(\sigma_{ur,k} \sigma_{ur,i} + \sum_{\substack{j=1 \\ j \neq k, i}}^K \sigma_{ur,k} \sigma_{ur,j} + 2 \epsilon_{ur,k} \sigma_{ur,i} + 2 \epsilon_{ur,k} \sigma_{ur,k} + 2 \sigma_{ur,i} \right. \right. \\ &\quad \left. \left. + \frac{1}{\tau P_{\text{up}}} (\sigma_{ur,i} + \sigma_{ur,k}) + \sigma_{ur,i} \right) + \mathcal{J}_k \left(\epsilon_{ur,k} + 1 + \frac{1}{\tau P_{\text{up}}} \right) + \mathcal{M}_k (1 + \sigma_{ur,j} \epsilon_{ur,k}) \right. \\ &\quad \left. + \frac{\mathcal{N}_k}{\tau P_{\text{up}}} \left(M \beta_{rd} \epsilon_{ud,k} + \frac{1}{\alpha_3} \right) + \frac{\mathcal{O}_k}{\tau P_{\text{up}}} \left(M \beta_{rd} \epsilon_{ud,k} + \frac{1}{\alpha_3} \right) + 2 \alpha_4 \tilde{\sigma}_{d,k} M^2 \sigma_{ud,k} \beta_{rd} \left(N(N+1) \sigma_{ur,k}^2 \right. \right. \\ &\quad \left. \left. + \sum_{\substack{i=1 \\ i \neq k}}^K \sigma_{ur,k} \sigma_{ur,i} \right) + 2 \alpha_4 \tilde{\sigma}_{d,k} M^2 N \sigma_{ud,k} \beta_{rd} \epsilon_{ud,k} \sum_{\substack{k'=1 \\ k' \neq k}}^K \sigma_{ur,k'} \right) - \bar{\mathbf{N}}\mathbf{U}_k, \end{aligned} \quad (53)$$

$$\begin{aligned} \mathbf{I}\mathbf{U}_k &\approx \sum_{\substack{i=1 \\ i \neq k}}^K P_{u,k} P_{u,i} \left(M \sigma_{ud,k} \sigma_{ud,i} + M \sigma_{ud,k} \epsilon_{ud,i} \mathcal{I}_k (\sigma_{ur,i} + \epsilon_{ur,k} + 2) + \mathcal{G}_k (\sigma_{ur,k} \sigma_{ur,i} + \sigma_{ur,k} \sigma_{ur,j} \right. \\ &\quad \left. + 3 \sigma_{ur,j} + 3 \sigma_{ur,i} + 2 \sigma_{ur,k}) + \mathcal{J}_k \left(\sigma_{ur,i} + 2 + \epsilon_{ur,k} + \frac{(N+1) \epsilon_{ur,i} \epsilon_{ur,j}}{\tau P_{\text{up}}} \right) + \epsilon_{ur,k} \epsilon_{ur,i} (MN(N+1) \right. \\ &\quad \left. \times \sigma_{ur,j} \mathcal{O}_k + M^2 N \alpha_4^2 \beta_{rd}^2 \tilde{\sigma}_{d,k}^2 \sigma_{ur,j} \sigma_{ur,k'}) + \frac{M^2 N \alpha_4^2 \beta_{rd}^2 \tilde{\sigma}_{d,k}^2 \sigma_{ur,j} \sigma_{ur,k'}}{\tau P_{\text{up}}} + \frac{1}{\alpha_3 \tau P_{\text{up}}} \left(\mathcal{N}_k \right. \right. \\ &\quad \left. \left. + MN \beta_{rd} \alpha_4^2 \tilde{\sigma}_{d,k}^2 \sum_{\substack{j=1 \\ j \neq k}}^K \sigma_{ur,i} \sigma_{ur,j} \right) + \frac{MN \beta_{rd} \alpha_4^2 \tilde{\sigma}_{d,k}^2}{\alpha_3 \tau P_{\text{up}}} \sum_{j=1}^K \sigma_{ur,j} \epsilon_{ur,i} \right), \end{aligned} \quad (54)$$

$$\begin{aligned} \mathbf{A}\mathbf{N}_k &\approx P_{u,k} \left(M \beta_{ud,k} + \mathcal{I}_k + \mathcal{G}_k (2 \sigma_{ur,k} + \sigma_{ur,i}) + (N+1) \mathcal{J}_k + M \beta_{rd} \alpha_4^2 \tilde{\sigma}_{d,k}^2 \left(2N(N+1) \sigma_{ur,k}^2 \right. \right. \\ &\quad \left. \left. + 2N \sum_{\substack{i=1 \\ i \neq k}}^K \sigma_{ur,k} \sigma_{ur,i} \right) + M \beta_{rd} \epsilon_{ur,k} (\mathcal{O}_k \sigma_{ur,i} + N \sigma_{ur,i} \sigma_{ur,j}) + MN \beta_{rd} \epsilon_{ur,k} \tilde{\sigma}_{d,k}^2 \alpha_4^2 \sum_{i=1}^K \sigma_{ur,i} \right. \\ &\quad \left. + \frac{M \beta_{rd}}{\tau P_{\text{up}}} (\mathcal{O}_k \sigma_{ur,i} + N \sigma_{ur,i} \sigma_{ur,j}) + \frac{M(N+1) \mathcal{M}_k \beta_{rd}}{\epsilon_{ur,k} \tau P_{\text{up}}} + \frac{M \beta_{rd} \mathcal{M}_k}{\epsilon_{ur,k} \alpha_3 \tau P_{\text{up}}} + \frac{M \alpha_4^2 \tilde{\sigma}_{d,k}^2}{\alpha_3 \tau P_{\text{up}}} \right). \end{aligned} \quad (55)$$

1) Solve $P_{u,k}$ with Fixed α_2 : When α_2 is fixed, to solve $P_{u,k}$, we introduce a new variable t , and obtain the following problem from (57):

$$\begin{aligned} \max_{\{P_{u,k}\}, t} \quad & t & (58a) \\ \text{s.t.} \quad & \gamma_k \geq t, \quad \forall k & (58b) \end{aligned}$$

$$\begin{aligned} & (56b), (56c). & (58c) \\ & \max_{\{P_{u,k}\}, t} \quad t & (59a) \\ & \text{s.t.} \quad (56b), (56c), & (59b) \end{aligned}$$

$$a_k f_k + b_k f_k \sum_{\substack{i=1 \\ i \neq k}}^K P_{u,k}^{-1} P_{u,i} + c_k f_k P_{u,k}^{-1} < \frac{1}{t}, \quad \forall k, \quad (59c)$$

We can see that the constraint in (58b) is not a posynomial function. However, we can reformulate it to a form of posynomial function as: $\frac{1}{\gamma_k} \leq \frac{1}{t}$, $\forall k$. To this end, problem (58)

becomes where $f_k = (\hat{\sigma}_{ud,k} + N \alpha_2 \tilde{\sigma}_{d,k} \beta_{ur,k} \beta_{rd})^2 M^2$, a_k , b_k , and

c_k are given, respectively, by

$$a_k = M\hat{\sigma}_{ud,k}^2 + MN\alpha_2^2\tilde{\sigma}_{d,k}^2(M+N+1)\beta_{ur,k}^2\beta_{rd}^2 + M\sigma_{eud,k}\hat{\sigma}_{ud,k} + \frac{\alpha_2^2\tilde{\sigma}_{d,k}^2}{\tau P_{up}}MN\beta_{ur,k}\beta_{rd}^2(M+N) + \frac{\alpha_2^2\tilde{\sigma}_{d,k}^2}{\alpha_1\tau P_{up}}MN\beta_{ur,k}\beta_{rd}, \quad (60)$$

$$b_{k,i} = \sum_{i=1, i \neq k}^K M \left(N(M+N)\alpha_2^2\tilde{\sigma}_{d,k}^2\beta_{rd}^2 \left(\beta_{ur,k} + \frac{1}{\tau P_{up}} \right) \beta_{ur,i} + \hat{\sigma}_{ud,k}\hat{\sigma}_{ud,i} + \hat{\sigma}_{ud,k}\sigma_{eud,i} + \frac{\alpha_2^2\tilde{\sigma}_{d,k}^2}{\alpha_1\tau P_{up}}N\beta_{rd}\beta_{ur,i} \right), \quad (61)$$

$$c_k = M\hat{\sigma}_{ud,k} + \frac{\alpha_2^2\tilde{\sigma}_{d,k}^2}{\alpha_1\tau P_{up}}MN\beta_{rd} + M\frac{\alpha_2\tilde{\sigma}_{d,k}^2}{\alpha_1\tau P_{up}} + MN\tilde{\sigma}_{d,k}^2\alpha_2\beta_{rd} \left(\alpha_2(M+N+1) \right) \left(\beta_{ur,k} + \frac{1}{\tau P_{up}} \right). \quad (62)$$

Clearly, problem (59) is a geometric program (GP). Therefore, it can be solved efficiently by using convex optimization tools.

2) *Solve α_2 with Fixed $P_{u,k}$* : From (48), we can see that the SINR γ_k is an increasing function with respect to α_2 . Thus, if $P_{u,k}$ is fixed, we obtain the following optimal value for α_2 :

$$\alpha_2 = \frac{P_{r,th}}{N \left(\sum_{k=1}^K P_{u,k}\beta_{ur,k} + 1 \right)}. \quad (63)$$

Based on the discussions in Sections VI-A1 and VI-A2, we propose an iterative algorithm to solve problem (57) as in Algorithm 1.

Algorithm 1: Iterative algorithm to solve (57)

Initialize: choose the power thresholds $P_{u,th}$, $P_{r,th}$, and the number of iterations N_I . The initial value of $P_{u,k}$ is set to equal $P_{u,th}$, for all k , and set $n = 0$.

Repeat

- Set $n = n + 1$
- Set $\alpha_2 = \frac{P_{r,th}}{N \left(\sum_{k=1}^K P_{u,k}\beta_{ur,k} + 1 \right)}$.
- Solve the GP problem (59). Let $P_{u,k}^*$ is the optimal solution.
- Update $P_{u,k} = P_{u,k}^*$

Until $n = N_I$

B. Transmission Scheme 2

Since the SINR expression for Transmission Scheme 2 has the same form as the one for Transmission Scheme 1, we can follow the same method as in previous section (i.e. Section VI-A) to find the optimal $P_{u,k}$ and α_4 , for $k = 1, \dots, K$.

VII. NUMERICAL RESULTS AND DISCUSSIONS

This section presents numerical results to verify the correctness our analytical results, to analyze the performance of our proposed relaying transmission schemes, and to evaluate the benefits of proposed power allocation. The K users are located uniformly at random inside a 1000×1000 m² square with the

Fig. 3. Sum spectral efficiency versus the number of destination antennas for Transmission Scheme 1 and Scheme 2, generated by one large-scale fading snapshot. Here, we choose $\bar{P}_{u,k} = \bar{P}_r = 0.1$ W for all k , $\bar{P}_{up} = \bar{P}_{u,k}$ and $\bar{P}_{up} = \bar{P}_{u,k}$.

Fig. 4. Sum spectral efficiency versus the number of relay antennas for Transmission Scheme 1 and scheme 2. We choose $M/N = 10$ and the other parameters are the same as in Fig. 3.

center being at $(0, 0)$. The destination and relay locations are $(2000, 0)$ and $(1000, 0)$, respectively.

The large-scale fading coefficients depend on the path loss due to the locations of users, relay and destination as well as the shadowing effect. They are modeled as [11], [23]

$$\beta_k = \frac{L_k}{1 + \left(\frac{z_k}{z_0} \right)^\nu}, \quad (64)$$

where $\beta_k \in \{\beta_{ur,k}, \beta_{ud,k}, \beta_{rd}\}$, L_k presents the shadowing which is a log-normal random variable with standard deviations of σ_L dB, ν is the path loss exponent, z_k is the distance from the k -th user to the relay/destination or the distance from the relay to destination, and z_0 denotes a reference distance.

In all examples, we choose $z_0 = 100$ m, $\sigma_L = 8$ dB, $\nu = 4$, $T = 200$ and $\tau = K$. Furthermore, we use $\bar{P}_{u,k}$, $\bar{P}_r$ to denote

Fig. 5. Sum spectral efficiency versus the large-scale fading effect of the direct link (c).

Fig. 6. Cumulative distribution of the per-user spectral efficiency for Transmission Scheme 1. We select $M = 100$, $\bar{P}_{u,k} = \bar{P}_r = 0.1$ W for all k , $\bar{P}_p = \bar{P}_r$ and $\bar{P}_{up} = \bar{P}_{u,k}$.

the transmit powers of user k and the relay, respectively; $\bar{P}_{u,th}$ and $\bar{P}_{r,th}$ to indicate the maximum powers of the users and the relay, respectively; and $\bar{P}_{up}$, and $\bar{P}_{rp}$ to illustrate the pilot powers sent from the users and the relay, respectively. The corresponding normalized transmit SNRs $P_{u,k}$, P_r , $P_{u,th}$, $P_{r,th}$, P_{up} , and P_{rp} can be calculated by dividing these powers by the noise power [22].

Firstly, we verify the accuracy of our analytical results for Transmission Schemes 1 and 2. Fig. 3 shows the sum spectral efficiency of Transmission Schemes 1 and 2 versus the number of destination antennas M for different values of N and K for one large-scale fading snapshot.⁵ The “Analysis” curves in Fig. 3 is obtained from the analytical results in Proposition 1

⁵Note that we consider one snapshot of large-scale fading because we want to make sure that the comparison is made under the same setup. It is not reasonable to compare the average values because even if the average values are close (or the same), we still cannot guarantee that the instantaneous values (corresponding to each large-scale fading realization) are close (or the same).

Fig. 7. Cumulative distribution of the per-user spectral efficiency for Transmission Scheme 2. The parameters are the same as in Fig. 6.

Fig. 8. Cumulative distribution of the transmit power (in Watt) for Transmission Scheme 1 and Scheme 2. We choose $M = 100$, $\bar{P}_{u,k} = \bar{P}_r = 0.1$ Watt for all k , $\bar{P}_{rp} = \bar{P}_r$ and $\bar{P}_{up} = \bar{P}_{u,k}$.

and Proposition 2. The “simulation” curves are generated via the Monte-Carlo method using (36) for Transmission Scheme 1 and (47) for Transmission Scheme 2. It is clear from both figures that our analytical results match perfectly with simulation results for both cases. Note that the analytical result in Proposition 2 is an approximating result. We can see from Fig. 3 that this approximation is very tight even when $M = 10$. Note further that, with Transmission Scheme 1, the relay just amplifies and forwards its received signals to the destination. During this process, the relay may amplify the inter-user interference and the noise too, and hence, using more relay antennas does not imply a better performance. This is why we can see from Fig. 3 that there is a crossing point of curve ($K = 7, N = 7$) and curve ($K = 7, N = 15$).

Secondly, we show in Fig. 4 the behavior of the sum spectral efficiency with different number of users K when the number of relay antennas increases for Transmission Schemes 1 and 2. In this example, we fix the ratio of $M/N = 10$. It is clear from the figure that when the number of relay antennas increases,

(a) Transmission Scheme 1. We choose the same parameters as in Fig. 6.

(b) Transmission scheme 2. We choose the same parameters as in Fig. 7.

Fig. 9. Sum spectral efficiency versus the number of iterations with power allocation from Algorithm 1.

the sum spectral efficiency significantly increases for both schemes. This is in line with the discussion in Remark 2.

We next compare the performance of our proposed massive MIMO relaying systems with the conventional massive MIMO systems without the help of a relay. Here, we choose Transmission Scheme 1 for the comparison. The comparison between Transmission Schemes 1 and 2 will be studied later in Fig. 10. In this example, we take one large-scale fading snapshot. To investigate the effect of the direct link, we first scale the large-scale coefficients of the direct link by c and then vary c . The factor c reflects the effect of large-scale fading of the direct link. If the distance between the destination and the users is short, then c may be large and vice versa. Fig. 5 depicts the sum spectral efficiency versus c with different K, N , and M . We use the results in [4, Table 3.1] to generate the curves for the case of conventional massive MIMO (without using relay). As expected, when c is small (and hence, the large-scale fading is small), we need to use a relay node to reduce the effect of path loss. As a result, our proposed relaying system outperforms the conventional massive MIMO when c is small

Fig. 10. Average spectral efficiencies of Schemes 1 and 2 versus the distance d for different K and N , with power allocation. The parameters setup are the same as in Fig. 6 and Fig. 7.

and vice versa when c is large.

Figs. 6 and 7 compare the per-user spectral efficiencies of the cases with and without power control for Transmission Scheme 1 and 2, respectively. For uniform power allocation, we choose $\bar{P}_{u,k} = 0.1$ W for all $k = 1, \dots, K$. For our proposed power allocations (Section VI-A for Scheme 1 and Section VI-B for Scheme 2), we choose the initial values as $\bar{P}_{u,th} = \bar{P}_{r,th} = 0.1$ W, $\bar{P}_{up} = \bar{P}_{u,th}$ and $\bar{P}_{rp} = \tau \bar{P}_{u,th}$. The cumulative distributions of the spectral efficiency in our examples are generated as follows: (i) 100 random realizations of user locations and shadowing fading profiles are generated; (ii) the corresponding spectral efficiencies of these 100 realizations are computed; and (iii) a cumulative distribution is generated over the so-obtained spectral efficiencies. As we can see from these figures, the proposed power allocation schemes offer much more 95%-likely spectral efficiency than the uniform power allocation does. For example, the proposed power allocation scheme can improve the 95%-likely spectral efficiency by a factor of 5.8 for Scheme 1 and a factor of 5.7 for Scheme 2, compared to uniform power allocation scheme, at $K = 5$ and $N = 50$. Furthermore, by comparing Fig. 6 with 7, we can see that Scheme 2 is better than Scheme 1 in this setup.

Fig. 7 shows the cumulative distribution of the transmit power for our two transmission schemes with $M = 100$. It is clear from Fig. 7 that the proposed power allocation utilizes much less transmit power than the uniform power control does.

Then, we investigate the convergence of the Algorithm 1 for Transmission Schemes 1 and 2 as shown in Fig. 9. For both transmission schemes, the proposed algorithm converges quickly, just after 4 – 10 iterations.

Finally, we study the impact of the distance from users to the relay and the distance from relay to destination on the performance of Transmission Schemes 1 and 2. Here, the relay and the destination are located at $(d, 0)$ and $(2d, 0)$ (in meters), and we vary d to evaluate the system performance. Fig. 10 shows the average spectral efficiency (which is taken over large-scale realizations) versus the distance d . In this figure, our proposed power allocation is implemented. We

can see from the figure that at low and medium distance d , Transmission Scheme 2 provides much better average spectral efficiency compared to Transmission Scheme 1. Nevertheless, when the distance is large, the effective noise at the relay is amplified in Scheme 2, and hence, Scheme 1 is better.

VIII. CONCLUSION

In this paper, we exploited a multiuser massive MIMO relaying system with two simple transmission schemes. In the first transmission scheme, no channel state information was needed at the relay node and the users. The relay just amplifies its received signals (transmitted from all users) and sends these scaled versions to the destination. In the second transmission scheme, the relay estimates all channel information from all users. Then, it uses these channel estimates and the MRC technique to combine the received signals before forwarding them to the destination. In both schemes, we proposed a method to estimate the channels at the destination for signal detection. With our proposed scheme, the destination did not need to estimate all channels. It just needed to estimate only the effective channel gains, and hence, the channel estimation overhead is independent of the numbers of antennas at the relay and the destination. Our transmission schemes are very simple, yet, they can improve the system performance noticeably, especially when the direct transmission links from the users to the destination experience high path loss. In addition, an efficient and simple max-min power allocation algorithm which chooses the transmit powers at the relay and the user to maximize the minimum spectral efficiency was proposed. Our proposed transmission protocols can be extended to multiple-relay setups because the MRC scheme can be implemented in a distributed manner. However the performance analysis as well as power control require a comprehensive study. Thus, we leave this interesting multiple-relay system for our future work.

APPENDIX

A. Proof of Proposition 1

To derive the closed-form expression for the spectral efficiency, we need to compute $|\mathbb{E}\{\hat{\mathbf{g}}_k^H \mathbf{g}_k\}|^2$, $\text{Var}(\hat{\mathbf{g}}_k^H \mathbf{g}_k)$, IU_k , and AN_k in closed forms.

- Compute $|\mathbb{E}\{\hat{\mathbf{g}}_k^H \mathbf{g}_k\}|^2$:

From the definitions of $\mathbf{G}$ and $\hat{\mathbf{G}}$ in (17) and (18), we have

$$\begin{aligned} \mathbb{E}\{\hat{\mathbf{g}}_k^H \mathbf{g}_k\} &= \mathbb{E}\{P_{u,k} \hat{\mathbf{g}}_{ud,k}^H \mathbf{g}_{ud,k} + \alpha_2 P_{u,k} \hat{\mathbf{g}}_{d,k}^H \mathbf{g}_{d,k}\} \\ &= P_{u,k} \mathbb{E}\{\hat{\mathbf{g}}_{ud,k}^H \mathbf{g}_{ud,k}\} + \alpha_2 P_{u,k} \mathbb{E}\{\hat{\mathbf{g}}_{d,k}^H \mathbf{g}_{d,k}\}. \end{aligned} \quad (65)$$

Substituting (6) and (9) into (65), together using the fact that the channel estimate $\hat{\mathbf{g}}_{ud,k}$ and its corresponding channel estimation error are independent, we obtain

$$\begin{aligned} \mathbb{E}\{\hat{\mathbf{g}}_k^H \mathbf{g}_k\} &= P_{u,k} \mathbb{E}\{\|\hat{\mathbf{g}}_{ud,k}\|^2\} + \alpha_2 P_{u,k} \mathbb{E}\{\mathbf{g}_{ur,k}^H \mathbf{G}_{rd}^H \mathbf{G}_{rd} \mathbf{g}_{ur,k}\} \tilde{\sigma}_{d,k} \\ &= M P_{u,k} (\sigma_{ud,k} + N \alpha_2 \tilde{\sigma}_{d,k} \beta_{ur,k} \beta_{rd}), \end{aligned} \quad (66)$$

which yields the numerator of (40).

- Compute $\text{Var}(\hat{\mathbf{g}}_k^H \mathbf{g}_k)$: We have

$$\begin{aligned} \text{Var}(\hat{\mathbf{g}}_k^H \mathbf{g}_k) &= \mathbb{E}\{|\hat{\mathbf{g}}_k^H \mathbf{g}_k|^2\} - |\mathbb{E}\{\hat{\mathbf{g}}_k^H \mathbf{g}_k\}|^2 \\ &= \mathbb{E}\{|\hat{\mathbf{g}}_k^H \mathbf{g}_k|^2\} - M^2 P_{u,k}^2 (\sigma_{ud,k} + N \alpha_2 \tilde{\sigma}_{d,k} \beta_{ur,k} \beta_{rd})^2. \end{aligned} \quad (67)$$

- We now calculate $\mathbb{E}\{|\hat{\mathbf{g}}_k^H \mathbf{g}_k|^2\}$:

Following the same method as in (65) and (66), we have

$$\begin{aligned} \mathbb{E}\{|\hat{\mathbf{g}}_k^H \mathbf{g}_k|^2\} &= M^2 P_{u,k}^2 (\sigma_{ud,k} + N \alpha_2 \tilde{\sigma}_{d,k} \beta_{ur,k} \beta_{rd})^2 \\ &\quad + P_{u,k}^2 \left(M \sigma_{ud,k}^2 + M N \alpha_2^2 \tilde{\sigma}_{d,k}^2 (M + N + 1) \beta_{ur,k}^2 \beta_{rd}^2 \right. \\ &\quad \left. + M \epsilon_{ud,k} \sigma_{ud,k} + \frac{\alpha_2^2 \tilde{\sigma}_{d,k}^2}{\tau P_{up}} M N \beta_{ur,k} \beta_{rd}^2 (M + N) \right. \\ &\quad \left. + \frac{\alpha_2^2 \tilde{\sigma}_{d,k}^2}{\alpha_1 \tau P_{up}} M N \beta_{ur,k} \beta_{rd} \right). \end{aligned} \quad (68)$$

Substituting (68) into (67), we obtain (40). With the same methodology as in computing (40), we can deduce (41) and (42), and hence, we can arrive at the desired result as in Proposition 1.

B. Derivation of (52)

We first compute:

$$\begin{aligned} \mathbb{E}\{\hat{\mathbf{g}}_k^H \bar{\mathbf{g}}_k\} &= \mathbb{E}\left\{ \left[\sqrt{P_{u,k}} \hat{\mathbf{g}}_{ud,k}^H \quad \sqrt{\alpha_4 P_{u,k}} \hat{\mathbf{g}}_{d,k}^H \right] \begin{bmatrix} \sqrt{P_{u,k}} \mathbf{g}_{ud,k} \\ \sqrt{\alpha_4 P_{u,k}} \bar{\mathbf{g}}_{d,k} \end{bmatrix} \right\} \\ &= P_{u,k} \mathbb{E}\{\hat{\mathbf{g}}_{ud,k}^H \mathbf{g}_{ud,k}\} + \alpha_4 P_{u,k} \mathbb{E}\{\hat{\mathbf{g}}_{d,k}^H \bar{\mathbf{G}}_{rd} \hat{\mathbf{G}}_{ur}^H \mathbf{g}_{ur,k}\} \\ &= \alpha_4 P_{u,k} \mathbb{E}\{\mathbf{g}_{ur,k}^H \hat{\mathbf{G}}_{ur} \bar{\mathbf{G}}_{rd}^H \bar{\mathbf{G}}_{rd} \hat{\mathbf{G}}_{ur}^H \mathbf{g}_{ur,k}\} \tilde{\sigma}_{d,k} \\ &\quad + P_{u,k} \mathbb{E}\{\|\hat{\mathbf{g}}_{ud,k}\|^2\}. \end{aligned} \quad (69)$$

We next use the approximation $\frac{\mathbf{G}_{rd}^H \mathbf{G}_{rd}}{M} \approx \beta_{rd} \mathbf{I}_N$ to obtain

$$\begin{aligned} \mathbb{E}\{\hat{\mathbf{g}}_k^H \bar{\mathbf{g}}_k\} &\approx \alpha_4 P_{u,k} \tilde{\sigma}_{d,k} M \beta_{rd} \mathbb{E}\{\hat{\mathbf{g}}_{ur,k}^H \hat{\mathbf{G}}_{ur} \hat{\mathbf{G}}_{ur}^H \hat{\mathbf{g}}_{ur,k}\} \\ &\quad + M P_{u,k} \sigma_{ud,k} + \alpha_4 P_{u,k} \tilde{\sigma}_{d,k} M \beta_{rd} \mathbb{E}\{\epsilon_{ur,k}^H \hat{\mathbf{G}}_{ur} \hat{\mathbf{G}}_{ur}^H \epsilon_{ur,k}\} \\ &\approx M P_{u,k} [\sigma_{ud,k} + \alpha_4 \tilde{\sigma}_{d,k} \beta_{rd} (N(N+1) \sigma_{ur,k}^2 \\ &\quad + N \sum_{\substack{k=1 \\ k' \neq k}}^K \sigma_{ur,k} \sigma_{ur,k'} + N \sum_{k'=1}^K \epsilon_{ur,k} \sigma_{ur,k'})] \end{aligned} \quad (70)$$

Finally, (70) can be obtained by using $\bar{\mathbf{N}}_k = |\mathbb{E}\{\hat{\mathbf{g}}_k^H \bar{\mathbf{g}}_k\}|^2$.

REFERENCES

- [1] C. D. Ho, H. Q. Ngo, and M. Matthaiou, "Massive MIMO AF relaying with channel estimation and power control techniques," in *Proc. IEEE SPAWC*, July 2019.
- [2] 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.
- [3] M. Ke, Z. Gao, Y. Wu, X. Gao, and R. Schober, "Compressive sensing-based adaptive active user detection and channel estimation: Massive access meets massive MIMO," *IEEE Trans. Signal Process.*, vol. 68, pp. 764–779, Jan. 2020.

- [4] T. L. Marzetta, E. G. Larsson, H. Yang, and H. Q. Ngo, *Fundamentals of Massive MIMO*. Cambridge University Press, 2016.
- [5] J. Zhang, E. Björnson, M. Matthaiou, D. W. K. Ng, H. Yang, and D. J. Love, "Prospective multiple antenna technologies for beyond 5G," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 8, pp. 1637–1660, Aug. 2020.
- [6] M. Vikalp, S. Ekant, and B. Rohit, "Energy-efficient massive MIMO multi-relay NOMA systems with CSI errors," *IEEE Trans. Commun.*, vol. 68, no. 12, pp. 7410–7428, Aug. 2020.
- [7] S. Jin, X. Liang, K.-K. Wong, X. Gao, and Q. Zhu, "Ergodic rate analysis for multipair massive MIMO two-way relay networks," *IEEE Trans. Wireless Commun.*, vol. 14, no. 3, pp. 1480–1491, Mar. 2015.
- [8] L. Pan, Y. Dai, W. Xu, and X. Dong, "Multipair massive MIMO relaying with pilot-data transmission overlay," *IEEE Trans. Wireless Commun.*, vol. 16, no. 6, pp. 3448–3460, Jun. 2017.
- [9] J. Qian, C. Masouros, and M. Matthaiou, "Multi-pair two-way massive MIMO relaying with zero forcing: Energy efficiency and power scaling laws," *IEEE Trans. Commun.*, vol. 68, no. 3, pp. 1417–1431, Mar. 2020.
- [10] B. Dutta, R. Budhiraja, N. Seshadri, and R. D. Koilpillai, "Large-system analysis of AF full-duplex massive MIMO two-way MRC/MRT relaying," *IEEE Trans. Commun.*, vol. 68, no. 4, pp. 2061–2077, Apr. 2020.
- [11] H. Q. Ngo, H. A. Suraweera, M. Matthaiou, and E. G. Larsson, "Multipair full-duplex relaying with massive arrays and linear processing," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 9, pp. 1721–1737, Sept. 2014.
- [12] C. Kong, C. Zhong, S. Jin, S. Yang, H. Lin, and Z. Zhang, "Full-duplex massive MIMO relaying systems with low-resolution ADCs," *IEEE Trans. Wireless Commun.*, vol. 16, no. 8, pp. 5033–5047, Aug. 2017.
- [13] Y. Xiong, Z. Zhang, and N. Wei, "Achievable rates for massive MIMO relaying systems with variable-bit ADCs/DACs," *IEEE Commun. Lett.*, vol. 68, no. 7, pp. 991–994, Feb. 2020.
- [14] C. Kong, C. Zhong, M. Matthaiou, E. Björnson, and Z. Zhang, "Spectral efficiency of multipair massive MIMO two-way relaying with imperfect CSI," *IEEE Trans. Veh. Tech.*, vol. 68, no. 7, pp. 6593–6607, July 2019.
- [15] C. D. Ho, H. Q. Ngo, M. Matthaiou, and T. Q. Duong, "On the performance of zero-forcing processing in multi-way massive MIMO relay networks," *IEEE Commun. Lett.*, vol. 21, no. 4, pp. 849–852, Apr. 2017.
- [16] D. Kudathanthirige and G. Amarasureiya, "Multicell multi-way massive MIMO relay networks," *IEEE Trans. Veh. Technol.*, vol. 66, no. 8, pp. 6831–6848, Jan. 2017.
- [17] —, "Sum rate of massive MIMO downlink with simultaneous relayed and direct transmissions," in *Proc. IEEE ICC*, June 2017.
- [18] X. Yang, X. Liang, T. Song, W. Xu, and S. Jin, "Multiuser massive MIMO AF relaying: Spectral efficiency and power allocation," *IEEE Access*, vol. 6, pp. 18 894–18 906, Mar. 2018.
- [19] D. Kudathanthirige and G. Amarasureiya, "Massive MIMO configurations for multi-cell multi-user relay networks," *IEEE Trans. Wireless Commun.*, vol. 17, no. 3, pp. 1849–1868, Mar. 2018.
- [20] X. Gao, O. Edfors, F. Rusek, and F. Tufvesson, "Linear pre-coding performance in measured very-large MIMO channels," in *Proc. IEEE VTC*, Sept. 2011.
- [21] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ: Prentice Hall, 1993.
- [22] H. Q. Ngo, A. Ashikhmin, H. Yang, E. G. Larsson, and T. L. Marzetta, "Cell-free massive MIMO versus small cells," *IEEE Trans. Wireless Commun.*, vol. 16, no. 3, pp. 1834–1850, Mar. 2017.
- [23] C. Ho, H. Q. Ngo, M. Matthaiou, and T. Q. Duong, "Multi-way massive MIMO with maximum-ratio processing and imperfect CSI," in *Proc. IEEE EUSIPCO*, Oct. 2017.

nam. His research interests include MU-MIMO, massive MIMO, multi-way relay networks with massive MIMO and relaying system.

Hien Quoc Ngo is currently a Lecturer with Queen's University Belfast, U.K. He has coauthored many research articles in wireless communications and coauthored the Cambridge University Press textbook *Fundamentals of Massive MIMO* (2016). His main research interests include massive (large-scale) MIMO systems, cell-free massive MIMO, physical layer security, and cooperative communications. He received the IEEE ComSoc Stephen O. Rice Prize in Communications Theory in 2015, the IEEE ComSoc Leonard G. Abraham Prize in 2017, and the Best Ph.D. Award from EURASIP in 2018. He also received the IEEE Sweden VT-COM-IT Joint Chapter Best Student Journal Paper Award in 2015. He was awarded the UKRI Future Leaders Fellowship in 2019. He serves as the Editor for the IEEE Transactions on Wireless Communications, IEEE Wireless Communications Letters, the Digital Signal Processing, and the Physical Communication (Elsevier). He was a Guest Editor of IET Communications, and a Guest Editor of IEEE ACCESS in 2017.

Michail Matthaiou (S'05–M'08–SM'13) was born in Thessaloniki, Greece in 1981. He obtained the Diploma degree (5 years) in Electrical and Computer Engineering from the Aristotle University of Thessaloniki, Greece in 2004. He then received the M.Sc. (with distinction) in Communication Systems and Signal Processing from the University of Bristol, U.K. and Ph.D. degrees from the University of Edinburgh, U.K. in 2005 and 2008, respectively. From September 2008 through May 2010, he was with the Institute for Circuit Theory and Signal Processing, Munich University of Technology (TUM), Germany working as a Postdoctoral Research Associate. He is currently a Professor of Communications Engineering and Signal Processing and Deputy Director of the Centre for Wireless Innovation (CWI) at Queen's University Belfast, U.K. after holding an Assistant Professor position at Chalmers University of Technology, Sweden.

Dr. Matthaiou and his coauthors received the IEEE Communications Society (ComSoc) Leonard G. Abraham Prize in 2017. He currently holds the ERC Consolidator Grant BEATRICE (2021–2026) focused on the interface between information and electromagnetic theories. He was awarded the prestigious 2018/2019 Royal Academy of Engineering/The Leverhulme Trust Senior Research Fellowship and also received the 2019 EURASIP Early Career Award. His team was also the Grand Winner of the 2019 Mobile World Congress Challenge. He is currently the Editor-in-Chief of Elsevier Physical Communication, a Senior Editor for IEEE WIRELESS COMMUNICATIONS LETTERS and an Associate Editor for the IEEE JSAC SERIES ON MACHINE LEARNING FOR COMMUNICATIONS AND NETWORKS.