How to Scale Up the Spectral Efficiency of Multi-way Massive MIMO Relaying?

How to Scale Up the Spectral Efficiency of Multi-way Massive MIMO Relaying?

Chung Duc Ho, Hien Quoc Ngo, Michail Matthaiou, and Long Dinh Nguyen
Institute of Electronics, Communications and Information Technology (ECIT),
Queen's University Belfast, BT3 9DT, Belfast, U.K.
Email:{choduc01, hien.ngo, m.matthaiou, lnguyen04}@qub.ac.uk

***Abstract*—This paper considers a decode-and-forward (DF) multi-way massive multiple-input multiple-output (MIMO) relay system where many users exchange their data with the aid of a relay station equipped with a massive antenna array. We propose a new transmission protocol which leverages successive cancelation decoding and zero-forcing (ZF) at the users. By using properties of massive MIMO, a tight analytical approximation of the spectral efficiency is derived. We show that our proposed scheme uses only half of the time-slots required in the conventional scheme (in which the number of time-slots is equal to the number of users [1]), to exchange data across different users. As a result, the sum spectral efficiency of our proposed scheme is nearly double the one of the conventional scheme, thereby boosting the performance of multi-way massive MIMO to unprecedented levels. To improve the network energy efficiency, we also propose a power allocation scheme which maximizes the energy efficiency under a given peak power constraint at each user and the relay.**

***Index Terms*—Decode-and-forward, massive MIMO, multi-way relay system, power allocation, spectral and energy efficiency.**

I. Introduction

In the past few years, massive MIMO has attracted significant research attention for its ability to improve the spectral and energy efficiency [2], [3]. In massive MIMO systems, many users can be served by a base station equipped with very large antenna arrays. With very large antenna arrays, the channel vectors between the users and the base station become asymptotically pairwisely orthogonal, and hence, the noise and inter-user interference reduce noticeably without improving the complexity of the system [3]. Furthermore, by using time division duplex (TDD) mode, the channel estimation overhead depends only on the number of active users regardless of the number of base station antennas [4]. This makes massive MIMO scalable and, thus, is one of the key candidates for future wireless communication systems.

On a parallel avenue, multi-way relaying networks have also been investigated to enhance the robustness against the channel variations in distinguished areas, where the direct channels among users are unavailable due to large obstacle and/or heavy path loss in the propagation environment [5]. With the help of the relay station, users that are geographically separated can communicate or exchange their data-bearing symbols much easier. Moreover, a number of papers demonstrate that multi-way relaying networks provide much higher spectral efficiency and communication reliability compared to one-way or two-way relaying systems [6], [7]. For the aforementioned reasons, there is a plethora of potential applications of multi-way relaying networks, including wireless conference or power control in heterogeneous cellular networks.

The combination of multi-way relaying and massive MIMO is very promising since it reaps all benefits of both technologies. Recently, some papers have evaluated the performance of multi-way relaying networks with massive arrays at the relay [8], [9]. In these works, the authors showed that multi-way massive MIMO relay systems can offer huge spectral and energy efficiency. In addition, by using simple linear processing (e.g. ZF and maximum ratio processing), the transmit power of each user can be scaled down proportionally to the number of relay antennas, while maintaining a given quality of service. However, all of aforementioned studies considered a conventional transmission protocol which requires K time-slots to exchange data among K users. This stands out as the main limitation of conventional multi-way relaying schemes.

Different with previous works, in this paper we propose a novel transmission protocol for multi-way massive MIMO relay networks which requires only $\left\lceil \frac{K-1}{2} \right\rceil + 1$ time-slots for the information exchange among the K users. We consider the standard DF technique at the relay station with perfect knowledge of the channel state information (CSI) at the relay and the users. Under these settings, we derive an approximate close-form expression for the spectral efficiency. The approximation is shown to be very tight, especially when the number of relay antennas is large. Furthermore, we also solve a power allocation problem which aims at maximizing the energy efficiency for a given sum spectral efficiency and under a power constraint at the relay and the users.

Notations: Matrices and vectors are expressed as upper and lower case boldface letters, respectively. The superscripts $(\cdot)^T$ and $(\cdot)^H$ stand for the transpose and Hermitian transpose, respectively. We denote by $\mathbf{a}_k$ the k-th column of matrix $\mathbf{A}$. The symbol $\|\cdot\|$ indicates the norm of a vector. The notation $\mathbb{E}\{\cdot\}$ is the expectation operator. The notation $[\mathbf{A}]_{mn}$ or a_{mn} denotes the (m,n)-th element of matrix $\mathbf{A}$, and $\mathbf{I}_K$ is the $K \times K$ identity matrix. Finally, the notation $\mathbf{A} \circ \mathbf{B}$ is defined as the Hadamard product between matrices $\mathbf{A}$ and $\mathbf{B}$.

II. System model

We consider a DF multi-way relay network with a very large antenna array at the relay station. The system includes one relay station equipped with M antennas and K single-antenna users. The bearing-messages from K users are exchanged with the help of the relay station. Each user wants to detect the signals transmitted from $K-1$ other users. We assume that the users and the relay station operate in half-duplex mode and have perfect CSI.[1] Furthermore, we assume that the direct links (user-to-user links) are unavailable due to large path loss and/or severe shadowing.

The channel matrix between the K users and M antennas at the relay is denoted by $\mathbf{G} \in \mathbb{C}^{M \times K}$ and is modeled as

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

where $\mathbf{H} \in \mathbb{C}^{M \times K}$ models small-scale fading with independent $\mathcal{CN}(0,1)$ components, and $\mathbf{D} \in \mathbb{C}^{K \times K}$ is the diagonal matrix of large-scale fading (path loss and log-normal attenuation). Let g_{mk} and h_{mk} be the (m,k)-th element of $\mathbf{G}$ and $\mathbf{H}$, respectively. Thus, we have $g_{mk} = \sqrt{\beta_k} h_{mk}$ where β_k is the k-th diagonal element of $\mathbf{D}$.

In general, the transmission protocol is divided into two phases: multiple-access phase and broadcast phase. In the multiple-access phase, all K users transmit signals to the relay station. In the broadcast phase, the relay station broadcast signals (which are decoded in the multi-access phase) to all the users.

III. Conventional Transmission Protocol

In this section, we first summarize a conventional transmission protocol tailored to multi-way massive DF relaying networks. The uplink and downlink spectral efficiencies are then provided in closed-form.

A. Multiple-Access Phase

This phase requires only one time-slot. All the K users transmit their data to the relay in the same time-frequency resource. The $M \times 1$ received vector at the relay is

$$\mathbf{y}_{\text{R}} = \sum_{k=1}^{K} \sqrt{P_{\text{u},k}} \mathbf{g}_k x_k + \mathbf{n}_{\text{R}}, \tag{2}$$

where $\mathbf{x} \triangleq [x_1, x_2 \ldots, x_K]^T$ is the signal vector transmitted from the K users, with $\mathbb{E}\left\{\mathbf{x}\mathbf{x}^H\right\} = \mathbf{I}_K$, $\mathbf{n}_{\text{R}}$ is the noise vector with i.i.d. $\mathcal{CN}(0,1)$ elements, and $P_{\text{u},k}$ is the normalized transmit power of the k-th user.

After receiving the transmitted signals from the K users, the relay employs the maximum ratio combining scheme by multiplying $\mathbf{y}_{\text{R}}$ with $\mathbf{G}^H$ as follows:

$$\mathbf{r} = \mathbf{G}^H \mathbf{y}_{\text{R}}. \tag{3}$$

Then, the k-th element of $\mathbf{r}$, denoted by r_k, is used to decode the signal transmitted from user k. From (3), r_k is given by

$$r_k = \sqrt{P_{\text{u},k}} \|\mathbf{g}_k\|^2 x_k + \sum_{i=1, i\neq k}^{K} \sqrt{P_{\text{u},i}} \mathbf{g}_k^H \mathbf{g}_i x_i + \mathbf{g}_k^H \mathbf{n}_{\text{R}}, \tag{4}$$

[1]This assumption, though idealistic, provides an upper bound on the performance of practical system where the channels are estimated at the users and relay.

where $\mathbf{g}_k$ is the k-th column of $\mathbf{G}$. Therefore, the uplink spectral efficiency of the system in (4) (measured in bit/s/Hz) is given by

$$\text{R}_k^{\text{ul}} = \mathbb{E}\left\{ \log_2 \left(1 + \frac{P_{\text{u},k} \|\mathbf{g}_k\|^4}{\sum\limits_{\substack{i=1 \\ i\neq k}}^{K} P_{\text{u},i} \left|\mathbf{g}_k^H \mathbf{g}_i\right|^2 + \|\mathbf{g}_k\|^2} \right) \right\}. \tag{5}$$

By using Jensen's inequality, a closed-form expression lower bound of the spectral efficiency (5) is given by [3, Eq. (16)]

$$\text{R}_k^{\text{ul}} \geq \tilde{\text{R}}_k^{\text{ul}} = \log_2 \left(1 + \frac{P_{\text{u},k}(M-1)\beta_k}{\sum_{i=1, i\neq k}^{K} P_{\text{u},i}\beta_i + 1} \right). \tag{6}$$

B. Broadcast Phase

In this phase, the relay station transmits all signals decoded in the multiple-access phase to all users in $K-1$ time slots. In the t time-slot, the relay aims to transmit $x_{j(k,t)}$ to user k, $k = 1, \ldots, K$, where

$$j(k,t) \triangleq \begin{cases} (k+t) \text{ modulo } K, & \text{if } \ (k+t) \neq K \\ K, & \text{otherwise.} \end{cases} \tag{7}$$

More precisely, in the t-th time-slot, the relay station transmits

$$\mathbf{s}^{(t)} = \sum_{i=1}^{K} \sqrt{\eta_{j(i,t)}} \mathbf{g}_i x_{j(i,t)}, \tag{8}$$

where $\{\eta_{j(i,t)}\}$ are the power control coefficients at the relay, which are selected to satisfy the power constraint at the relay: $\mathbb{E}\left\{\|\mathbf{s}^{(t)}\|^2\right\} \leqslant P_{\text{r,th}}$. Therefore, we have

$$M \sum_{i=1}^{K} \eta_{j(i,t)} \beta_i \leqslant P_{\text{r,th}}. \tag{9}$$

Then, the received signal at the k-th user is

$$y_k^{(t)} = \mathbf{g}_k^H \mathbf{s}^{(t)} + n_k^{(t)} = \sum_{i=1}^{K} \sqrt{\eta_{j(i,t)}} \ \mathbf{g}_k^H \mathbf{g}_i x_{j(i,t)} + n_k^{(t)}. \tag{10}$$

The k-th user knows its own transmitted signal x_k (or $x_{j(k-t,t)}$), so it can remove the self-interference prior to decoding. Thus, the received signal after self-interference cancelation is

$$\tilde{y}_k^{(t)} = \sqrt{\eta_{j(k,t)}} \|\mathbf{g}_k\|^2 x_{j(k,t)} + \sum_{\substack{i=1 \\ j(i,t)\neq j(k,t), j(k-t,t)}}^{K} \sqrt{\eta_{j(i,t)}} \mathbf{g}_k^H \mathbf{g}_i x_{j(i,t)} + n_k^{(t)}. \tag{11}$$

Then, the corresponding downlink spectral efficiency for the t-th time-slot is

$$\text{R}_k^{\text{dl},(t)} = \mathbb{E}\left\{ \log_2 \left(1 + \frac{\eta_{j(k,t)} \|\mathbf{g}_k\|^4}{\sum\limits_{\substack{i=1 \\ j(i,t)\neq j(k,t), j(k-t,t)}}^{K} \eta_{j(i,t)} \left|\mathbf{g}_k^H \mathbf{g}_i\right|^2 + 1} \right) \right\}. \tag{12}$$

Proposition 1: The spectral efficiency $\mathtt{R}_k^{\text{dl},(t)}$ given by (12) can be lower bounded by

$$\mathtt{R}_k^{\text{dl},(t)} \geq \tilde{\mathtt{R}}_k^{\text{dl},(t)} = \log_2 \left(1 + \frac{\eta_{j(k,t)}(M-1)(M-2)\beta_k^2}{(M-2)\beta_k \sum\limits_{\substack{i=1 \\ j(i,t)\neq j(k,t), j(k-t,t)}}^{K} \eta_{j(i,t)}\beta_i + 1} \right) \quad (13)$$

Proof: The proof is omitted due to space constraints. The details of the proof will be provided in the journal version. ■

IV. Multi-Way Transmission with Successive Cancelation Decoding

In this section, we propose a novel transmission scheme which requires only $\lceil \frac{K-1}{2} \rceil + 1$ time-slots for the information exchange among the K users.

A. Multiple-Access Phase

The multiple-access phase is the same as the one in the conventional transmission scheme. See Section III-A.

B. Broadcast Phase

Here, we need only $\lceil \frac{K-1}{2} \rceil$ time-slots to transmit all K symbols to all users. The main idea is that: at a given time-slot, the k-th user subtracts all symbols decoded in previous time-slots prior to decoding the desired symbol. Furthermore, after $\lceil \frac{K-1}{2} \rceil$ time-slots, user k receives $\lceil \frac{K-1}{2} \rceil$ signals, and each signal is a linear combination of $K - \lceil \frac{K-1}{2} \rceil - 1$ symbols. So it can detect all $K - \lceil \frac{K-1}{2} \rceil - 1$ symbols without any inter-user interference through the ZF technique. A detailed presentation of the proposed scheme is now provided.

1) First time-slot: The relay intends to send $x_{j(k,1)}$ to the k-th user, for $k = 1, \dots, K$. The signal vector transmitted from the relay is

$$\mathbf{s}^{(1)} = \sum_{i=1}^{K} \sqrt{\eta_{j(i,1)}^{(1)}}\ \mathbf{g}_i x_{j(i,1)}, \quad (14)$$

where $\{\eta_{j(i,1)}^{(1)}\}$, $i = 1, \dots, K$, are the power control coefficients at the relay in the first time-slot which are chosen to satisfy a given power constraint at the relay $\mathbb{E}\left\{\|\mathbf{s}^{(1)}\|^2\right\} \leqslant P_{\text{r,th}}$, or

$$M \sum_{i=1}^{K} \eta_{j(i,1)}^{(1)} \beta_i \leqslant P_{\text{r,th}}. \quad (15)$$

Thus, the received signal at the k-th user is

$$y_k^{(1)} = \mathbf{g}_k^H \mathbf{s}^{(1)} + n_k^{(1)} = \sum_{i=1}^{K} \sqrt{\eta_{j(i,1)}^{(1)}} \mathbf{g}_k^H \mathbf{g}_i x_{j(i,1)} + n_k^{(1)}, \quad (16)$$

where $n_k^{(1)} \sim \mathcal{CN}(0,1)$ is the additive noise at the k-th user in the first time-slot. Since user k knows its transmitted signal x_k (or $x_{j(k-1,1)}$), it can subtract the self-interference before detecting signal $x_{j(k,1)}$. Therefore, the received signal at user k after self-interference cancelation is

$$\tilde{y}_k^{(1)} = \sqrt{\eta_{j(k,1)}^{(1)}} \|\mathbf{g}_k\|^2 x_{j(k,1)} + \sum_{\substack{i=1 \\ j(i,1)\notin \mathcal{V}_{k,1}}}^{K} \sqrt{\eta_{j(i,1)}^{(1)}}\ \mathbf{g}_k^H \mathbf{g}_i x_{j(i,1)} + n_k^{(1)}, \quad (17)$$

where

$$\mathcal{V}_{k,t} \triangleq \{j(k-t,t), j(k-t+1,t), \dots, j(k,t)\}. \quad (18)$$

Then, the corresponding spectral efficiency is written as

$$\mathtt{R}_k^{\text{dl},(1)} = \mathbb{E}\left\{ \log_2 \left(1 + \frac{\eta_{j(k,1)}^{(1)} \|\mathbf{g}_k\|^4}{\sum\limits_{\substack{i=1 \\ j(i,1)\notin \mathcal{V}_{k,1}}}^{K} \eta_{j(i,1)}^{(1)} \left|\mathbf{g}_k^H \mathbf{g}_i\right|^2 + 1} \right) \right\}. \quad (19)$$

2) t-th time-slot: At the t-time-slot, the relay intends to send $x_{j(k,t)}$ to the k-th user, for $k = 1, \dots, K$. The signal vector transmitted from the relay is

$$\mathbf{s}^{(t)} = \sum_{i=1}^{K} \sqrt{\eta_{j(i,t)}^{(t)}}\ \mathbf{g}_i x_{j(i,t)}, \quad (20)$$

where $\{\eta_{j(i,t)}^{(t)}\}$, $i = 1, \dots, K$, are the power control coefficients at the relay in the t-th time-slot chosen to satisfy a given power constraint $P_{\text{r,th}}$ at the relay as:

$$M \sum_{i=1}^{K} \eta_{j(i,t)}^{(t)} \beta_i \leqslant P_{\text{r,th}}. \quad (21)$$

Then, the k-th user sees

$$y_k^{(t)} = \mathbf{g}_k^H \mathbf{s}^{(t)} + n_k^{(t)} = \sum_{i=1}^{K} \sqrt{\eta_{j(i,t)}^{(t)}}\ \mathbf{g}_k^H \mathbf{g}_i x_{j(i,t)} + n_k^{(t)}. \quad (22)$$

The k-th user knows its own transmitted symbols x_k. Furthermore, it also knows its detected symbols in previous time-slots. So, it knows $\{x_{j(k-1,1)}, x_{j(k,1)}, x_{j(k,2)}, \dots, x_{j(k,t-1)}\}$, and, hence, it can remove these symbols to obtain

$$\tilde{y}_k^{(t)} = \sqrt{\eta_{j(k,t)}^{(t)}} \|\mathbf{g}_k\|^2 x_{j(k,t)} + \sum_{\substack{i=1 \\ j(i,t)\notin \mathcal{V}_{k,t}}}^{K} \sqrt{\eta_{j(i,t)}^{(t)}}\ \mathbf{g}_k^H \mathbf{g}_i x_{j(i,t)} + n_k^{(t)}. \quad (23)$$

Thus, the spectral efficiency of the k-th user at the t-th time-slot is

$$\mathtt{R}_k^{\text{dl},(t)} = \mathbb{E}\left\{ \log_2 \left(1 + \frac{\eta_{j(k,t)}^{(t)} \|\mathbf{g}_k\|^4}{\sum\limits_{\substack{i=1 \\ j(i,t)\notin \mathcal{V}_{k,t}}}^{K} \eta_{j(i,t)}^{(t)} \left|\mathbf{g}_k^H \mathbf{g}_i\right|^2 + 1} \right) \right\}. \quad (24)$$

Proposition 2: The spectral efficiency $\mathtt{R}_k^{\text{dl},(t)}$ given by (24) can be lower bounded by

$$\mathtt{R}_k^{\text{dl},(t)} \geq \tilde{\mathtt{R}}_k^{\text{dl},(t)} = \log_2\left(1 + \frac{\eta_{j(k,t)}^{(t)}(M-1)(M-2)\beta_k^2}{(M-2)\beta_k \sum\limits_{\substack{i=1 \\ j(i,t)\notin\mathcal{V}_{k,t}}}^{K} \eta_{j(i,t)}^{(t)}\beta_i + 1}\right). \tag{25}$$

Proof: The proof is omitted due to space constraints. ■

3) After $t' = \lceil \frac{K-1}{2} \rceil$ time-slots, the k-th user has received t' signals (the t-th received signal is given by (22)). Furthermore, it has decoded t' symbols. So, it can subtract all t' detected symbols from each received signal to obtain the following results:

$$\begin{cases} \bar{y}_{k,1}^{(t')} = \sum\limits_{\substack{i=1 \\ j(i,t')\notin\mathcal{V}_{k,t'}}}^{K} \sqrt{\eta_{j(i,t')}^{(t')}}\mathbf{g}_k^H \mathbf{g}_{j(k,i-k)} x_{j(i,t')} + n_{k,1}^{(t')}, \\ \bar{y}_{k,2}^{(t')} = \sum\limits_{\substack{i=1 \\ j(i,t')\notin\mathcal{V}_{k,t'}}}^{K} \sqrt{\eta_{j(i,t')}^{(t'-1)}}\mathbf{g}_k^H \mathbf{g}_{j(k,i-k+1)} x_{j(i,t')} + n_{k,2}^{(t')}, \\ \vdots \\ \bar{y}_{k,t'}^{(t')} = \sum\limits_{\substack{i=1 \\ j(i,t')\notin\mathcal{V}_{k,t'}}}^{K} \sqrt{\eta_{j(i,t')}^{(t'-(t'-1))}}\mathbf{g}_k^H \mathbf{g}_{j(k,i-k+t'-1)} x_{j(i,t')} + n_{k,t'}^{(t')}. \end{cases} \tag{26}$$

We can see that we have t' equations, each equation has $(K - t' - 1)$ unknown variables $\{x_{j(i,t')}\}$. Since $t' = \lceil \frac{K-1}{2} \rceil$, the number of equations is greater than or equal to the number of unknown variables. Therefore, the k-th user can detect all remaining $(K - t' - 1)$ symbols $\{x_{j(i,t')}\}$ via the ZF scheme.[2] More precisely, we first rewrite (26) in a matrix-vector form as

$$\bar{\mathbf{y}}_k^{(t')} = \bar{\mathbf{A}}_k^{(t')}\bar{\mathbf{x}} + \bar{\mathbf{n}}_k^{(t')}, \tag{27}$$

where $\bar{\mathbf{x}} \triangleq \begin{bmatrix} x_{j(k,t'+1)} & x_{j(k,t'+2)} & \cdots & x_{j(k,K-1)} \end{bmatrix}^T$,

$$\bar{\mathbf{y}}_k^{(t')} \triangleq \begin{bmatrix} \bar{y}_{k,1}^{(t')} \\ \bar{y}_{k,2}^{(t')} \\ \vdots \\ \bar{y}_{k,t'}^{(t')} \end{bmatrix}, \ \bar{\mathbf{n}}_k^{(t')} \triangleq \begin{bmatrix} n_{k,1}^{(t')} \\ n_{k,2}^{(t')} \\ \vdots \\ n_{k,t'}^{(t')} \end{bmatrix}.$$

and $\bar{\mathbf{A}}_k^{(t')} \in \mathbb{C}^{t' \times (K-t'-1)}$ is defined as $\bar{\mathbf{A}}_k^{(t')} = \mathbf{A}_k^{(t')} \circ \mathbf{E}_\eta^{(t')}$, where

$$\mathbf{A}_k^{(t')} \triangleq \begin{bmatrix} \mathbf{g}_k^H \mathbf{g}_{j(k,1)} & \mathbf{g}_k^H \mathbf{g}_{j(k,2)} & \cdots & \mathbf{g}_k^H \mathbf{g}_{j(k,K-t'-1)} \\ \mathbf{g}_k^H \mathbf{g}_{j(k,2)} & \mathbf{g}_k^H \mathbf{g}_{j(k,3)} & \cdots & \mathbf{g}_k^H \mathbf{g}_{j(k,K-t')} \\ \vdots & \vdots & & \vdots \\ \mathbf{g}_k^H \mathbf{g}_{j(k,t')} & \mathbf{g}_k^H \mathbf{g}_{j(k,t'+1)} & \cdots & \mathbf{g}_k^H \mathbf{g}_{j(k,K-2)} \end{bmatrix},$$

and

$$\mathbf{E}_\eta^{(t')} \triangleq \begin{bmatrix} \sqrt{\eta_{j(k,t'+1)}^{(t')}} & \sqrt{\eta_{j(k,t'+2)}^{(t')}} & \cdots & \sqrt{\eta_{j(k,K-1)}^{(t')}} \\ \sqrt{\eta_{j(k,t'+1)}^{(t'-1)}} & \sqrt{\eta_{j(k,t'+2)}^{(t'-1)}} & \cdots & \sqrt{\eta_{j(k,K-1)}^{(t'-1)}} \\ \vdots & \vdots & & \vdots \\ \sqrt{\eta_{j(k,t'+1)}^{(t'-(t'-1))}} & \sqrt{\eta_{j(k,t'+2)}^{(t'-(t'-1))}} & \cdots & \sqrt{\eta_{j(k,K-1)}^{(t'-(t'-1))}} \end{bmatrix}.$$

The k-th user applies the ZF scheme to decode the remaining symbols as follows:

$$\tilde{\mathbf{r}}_k^{(t')} = \mathbf{Z}^T \bar{\mathbf{y}}_k^{(t')} = \mathbf{Z}^T \bar{\mathbf{A}}_k^{(t')}\bar{\mathbf{x}} + \mathbf{Z}^T \bar{\mathbf{n}}_k^{(t')}, \tag{28}$$

where

$$\mathbf{Z}^T \triangleq \left(\left(\bar{\mathbf{A}}_k^{(t')}\right)^H \bar{\mathbf{A}}_k^{(t')}\right)^{-1} \left(\bar{\mathbf{A}}_k^{(t')}\right)^H. \tag{29}$$

The n-th element of $\tilde{\mathbf{r}}_k^{(t')}$ will be used to detect $x_{j(k,t'+n)}$. From (28) and the fact that $\mathbf{Z}^T \bar{\mathbf{A}}_k^{(t')} = \mathbf{I}_{K-t'-1}$, the n-th element of $\tilde{\mathbf{r}}_k^{(t')}$ is given by

$$\tilde{\mathbf{r}}_{k,n}^{(t')} = x_{j(k,t'+n)} + \mathbf{z}_n^T \bar{\mathbf{n}}_k^{(t')}. \tag{30}$$

Thus, the corresponding spectral efficiency of the system in (30) is

$$\begin{aligned} \mathtt{R}_k^{\text{dl},(t'+n)} &= \mathbb{E}\left\{\log_2\left(1 + \frac{1}{\|\mathbf{z}_n\|^2}\right)\right\} \\ &= \mathbb{E}\left\{\log_2\left(1 + \frac{1}{\left[\left(\left(\bar{\mathbf{A}}_k^{(t')}\right)^H \bar{\mathbf{A}}_k^{(t')}\right)^{-1}\right]_{nn}}\right)\right\}. \end{aligned} \tag{31}$$

Since (31) has a complicated form that involves a matrix inverse, we cannot obtain an exact closed-form. However, thanks to the trace lemma and law of large numbers (as M goes to infinity) [10], we can obtain the following approximating result.

Proposition 3: As $M \to \infty$, the spectral efficiency $\mathtt{R}_k^{\text{dl},(t'+n)}$ given by (31) converges to

$$\begin{aligned} &\mathtt{R}_k^{\text{dl},(t'+n)} \to \tilde{\mathtt{R}}_k^{\text{dl},(t'+n)} \\ &= \log_2\left(1 + M\beta_k \sum_{i=1}^{t'} \eta_{j(k,t'+n)}^{(t'-i+1)}\beta_{j(k,n+i-1)}\right). \\ &\qquad n = 1, \ldots, K - t' - 1 \end{aligned} \tag{32}$$

Proof: The proof is omitted due to space constraints. ■

V. POWER ALLOCATION

We now turn our attention to the energy efficiency and we apply a power allocation technique for each user and the relay. To this end, we formulate a power allocation scheme to maximize the energy efficiency for a given sum spectral efficiency under peak power constraint. The network energy efficiency (bit/Joule) is defined as the ratio of the sum throughput (in bit/s), which is $B\mathtt{R}_{\text{sum}} =$

[2] In our analysis we consider ZF reception instead of successive interference cancelation to keep the complexity to low levels [4].

$B\frac{1}{t'+1}\sum_{k=1}^{K}\sum_{t=1}^{K-1}\min\left(\mathtt{R}_k^{\text{ul}},\mathtt{R}_k^{\text{dl},(t)}\right)$, to the total power consumption (in Watt) for both downlink and uplink transmission protocol, where B (in Hz) is the transmission bandwidth. Consequently, the energy efficiency is given by [11]

$$\mathtt{EE} = \frac{B\mathtt{R}_{\text{sum}}}{\sum_{k=1}^{K}\frac{1}{\zeta_{\text{u},k}}P_{\text{u},k}+\frac{1}{\zeta_{\text{r}}}\sum_{t=1}^{t'}\sum_{k=1}^{K}M\beta_k\eta_{j(k,t)}^{(t)}+BP_{\text{LD}}\mathtt{R}_{\text{sum}}+P_{\text{LID}}}, \tag{33}$$

where $0<\zeta_{\text{u},k}\leq 1$ and $0<\zeta_{\text{r}}\leq 1$ are the power amplifier efficiencies of the k-th user and the relay, respectively. Furthermore, P_{LD} (in Watt/(bit/s)) represents the load dependent power including power for signal processing (coding and decoding) and the power consumption of the backhaul traffic, and P_{LID} (in Watt) accounts for the load independent power including power required to run the circuit components at the relay/users, and the fixed power consumption for the backhaul link between the relay and the core network. Therefore, the energy efficiency problem is now formulated as

$$\underset{\{P_{\text{u},k},\eta_{j(k,t)}^{(t)}\}}{\text{maximize}}\ \mathtt{EE} \tag{34a}$$

$$\text{s.t. } \mathtt{R}_{\text{sum}}=\mathtt{R}_{\text{th}}, \tag{34b}$$

$$0\leq P_{\text{u},k}\leq P_{\text{u,th}},\ k=1,\ldots,K, \tag{34c}$$

$$0\leq\sum_{k=1}^{K}M\eta_{j(k,t)}^{(t)}\beta_k\leq P_{\text{r,th}},\ t=1,\ldots,t'. \tag{34d}$$

By dividing the numerator and denominator of (33) by $B\mathtt{R}_{\text{sum}}$, the $\mathtt{EE}$ function can be written as

$$\mathtt{EE}=\frac{1}{\frac{\sum_{k=1}^{K}\frac{1}{\zeta_{\text{u},k}}P_{\text{u},k}+\frac{1}{\zeta_{\text{r}}}\sum_{t=1}^{t'}\sum_{k=1}^{K}M\beta_k\eta_{j(k,t)}^{(t)}+P_{\text{LID}}}{B\mathtt{R}_{\text{sum}}}+P_{\text{LD}}}. \tag{35}$$

We can see that, to maximize $\mathtt{EE}$, we can minimize the first term of denominator in (35). As a consequence, the power allocation problem (34) is equivalent to

$$\underset{P_{\text{u},k},\eta_{j(k,t)}^{(t)}}{\text{minimize}}\ \sum_{k=1}^{K}\frac{1}{\zeta_{\text{u},k}}P_{\text{u},k}+\frac{M}{\zeta_{\text{r}}}\sum_{t=1}^{t'}\sum_{k=1}^{K}\beta_k\eta_{j(k,t)}^{(t)} \tag{36a}$$

$$\text{s.t. } \frac{1}{t'+1}\sum_{k=1}^{K}\sum_{t=1}^{K-1}\log_2\left(1+\min\left(\gamma_{j(k,t)}^{\text{ul}},\gamma_k^{\text{dl},(t)}\right)\right)=\mathtt{R}_{\text{th}}, \tag{36b}$$

$$(34c),(34d), \tag{36c}$$

where

$$\gamma_{j(k,t)}^{\text{ul}}\triangleq\frac{e_{k,t}P_{\text{u},j(k,t)}}{\sum_{i=1,i\neq k}^{K}\beta_iP_{\text{u},j(i,t)}+1}, \tag{37}$$

$$\gamma_k^{\text{dl},(t)}\triangleq\begin{cases}\frac{f_{k,t}\eta_{j(k,t)}^{(t)}}{q_{k,t}\sum_{i=1,\,j(i,t)\notin\mathcal{V}_{k,t}}^{K}\beta_i\eta_{j(i,t)}^{(t)}+1} & \text{if } t=1,\ldots,t'\\ M\beta_k\sum_{i=1}^{t'}\eta_{j(k,t)}^{(t'-i+1)}\beta_{j(k,t-t'+i-1)} & \text{if } t=t'+1,\ldots,K-1\end{cases}, \tag{38}$$

and where $e_{k,t}=(M-1)\beta_{j(k,t)},\ f_{k,t}=(M-1)(M-2)\beta_k^2,\ q_{k,t}=(M-2)\beta_k$.

The problem (36) can be rewritten as

$$\underset{P_{\text{u},k},\eta_{j(k,t)}^{(t)},\gamma_{k,t}}{\text{minimize}}\quad(36a)\quad\text{s.t. }(34c),(34d) \tag{39a}$$

$$\frac{1}{t'+1}\sum_{k=1}^{K}\sum_{t=1}^{K-1}\log_2(1+\gamma_{k,t})=\mathtt{R}_{\text{th}}, \tag{39b}$$

$$\gamma_{k,t}\leq\gamma_{j(k,t)}^{\text{ul}},\ k=1,\ldots,K,\ t=1,\ldots,K-1, \tag{39c}$$

$$\gamma_{k,t}\leq\gamma_k^{\text{dl},(t)},\ k=1,\ldots,K,\ t=1,\ldots,t', \tag{39d}$$

$$\gamma_{k,t}\leq\gamma_k^{\text{dl},(t)},\ k=1,\ldots,K,\ t=t'+1,\ldots,K-1. \tag{39e}$$

Since $e_{k,t},\ f_{k,t}$, and $q_{k,t}$ are positive numbers, (39) can be equivalently written as

$$\underset{P_{\text{u},k},\eta_{j(k,t)}^{(t)},\gamma_{k,t}}{\text{minimize}}\quad(36a)\quad\text{s.t. }(34c),(34d) \tag{40a}$$

$$\prod_{k=1}^{K}\prod_{t=1}^{K-1}(1+\gamma_{k,t})=2^{(t'+1)\mathtt{R}_{\text{th}}}, \tag{40b}$$

$$\frac{\gamma_{k,t}}{e_{k,t}}\sum_{i=1,i\neq k}^{K}\beta_iP_{\text{u},j(i,t)}\left(P_{\text{u},j(k,t)}\right)^{-1}+\frac{\gamma_{k,t}}{e_{k,t}}\left(P_{\text{u},j(k,t)}\right)^{-1}\leq 1,$$
$$k=1,\ldots,K,\ t=1,\ldots,K-1, \tag{40c}$$

$$\frac{\gamma_{k,t}q_{k,t}}{f_{k,t}}\left(\eta_{j(k,t)}^{(t)}\right)^{-1}\sum_{i=1,\,j(i,t)\notin\mathcal{V}_{k,t}}^{K}\beta_i\eta_{j(i,t)}^{(t)}+\frac{\gamma_{k,t}}{f_{k,t}}\left(\eta_{j(k,t)}^{(t)}\right)^{-1}\leq 1,$$
$$k=1,\ldots,K,\ t=1,\ldots,t', \tag{40d}$$

$$\frac{\gamma_{k,t}}{t'M\beta_k}\prod_{i=1}^{t'}\left(\eta_{j(k,t)}^{(t'-i+1)}\beta_{j(k,t-t'+i-1)}\right)^{-1/t'}\leq 1,$$
$$k=1,\ldots,K,\ t=t'+1,\ldots,K-1. \tag{40e}$$

The objective function and all the inequality constraints in (40) are posynomial functions. However (40b) is not a monomial function, and hence, problem (40) is nonconvex problem. In order to solve the problem we use the technique in [12, Lemma 1] by finding an approximate solution of (40) through solving a sequence of geometric programming (GPs). More precisely, we can approximate $(1+\gamma_{k,t})$ by $\kappa_{k,t}\gamma_{k,t}^{\xi_{k,t}}$ near a point $\hat{\gamma}_{k,t}>0$, where $\xi_{k,t}\triangleq\frac{\hat{\gamma}_{k,t}}{1+\hat{\gamma}_{k,t}}$ and $\kappa_{k,t}\triangleq\hat{\gamma}_{k,t}^{-\xi_{k,t}}(1+\hat{\gamma}_{k,t})$. Consequently, near a point $\hat{\gamma}_{k,t}>0$, we have

$$\prod_{k=1}^{K}\prod_{t=1}^{K-1}(1+\gamma_{k,t})\approx\prod_{k=1}^{K}\prod_{t=1}^{K-1}\kappa_{k,t}\gamma_{k,t}^{\xi_{k,t}}, \tag{41}$$

which is a monomial function. Thus, by using the local approximation given by (41), the optimization problem (40) can be approximated by a GP which can be solved effectively by using CVX [13]. Furthermore, by using a similar technique as in [12], each iteration of the GP is obtained by replacing the posynomial objective function with its best local monomial approximation near the solution obtained at the previous iteration. The successive approximation algorithm to solve (40) is summarized in Algorithm 1.

Algorithm 1: Successive approximation algorithm for (40)

1. *Initialization*: Define a tolerance ϵ, the maximum number of iterations L, and parameter α. Set $n := 1$, choose the initial values of $\gamma_{k,t}$ as $\hat{\gamma}_{k,t}^{(1)}$.
2. *Iteration* n: Compute $\xi_{k,t}^{(n)} \triangleq \frac{\hat{\gamma}_{k,t}^{(n)}}{1+\hat{\gamma}_{k,t}^{(n)}}$ and $\kappa_{k,t}^{(n)} \triangleq \left(\hat{\gamma}_{k,t}^{(n)}\right)^{-\xi_{k,t}^{(n)}} \left(1+\hat{\gamma}_{k,t}^{(n)}\right)$.
$$\underset{P_{\mathrm{u},j(k,t)},\eta_{j(k,t)}^{(t)},\gamma_{k,t}}{\text{minimize}} \quad (36a) \quad \text{s.t.}$$
$$\prod_{k=1}^{K}\prod_{t=1}^{K-1} \kappa_{k,t}^{(n)} \left(\gamma_{k,t}^{(n)}\right)^{\xi_{k,t}^{(n)}} = 2^{(t'+1)\mathtt{R}_{\text{th}}},$$
$$(34c), (34d), (40c), (40d), (40e),$$
$$\alpha^{-1}\hat{\gamma}_{k,t}^{(n)} \le \gamma_{k,t} \le \alpha\hat{\gamma}_{k,t}^{(n)}.$$
Denote the optimal solutions by $\gamma_{k,t}^{*}$.
3. *Stopping criterion*: If $\max_k = |\hat{\gamma}_{k,t}^{(n)} - \gamma_{k,t}^{*}| < \epsilon$ or $n = L \rightarrow$ stop. Otherwise, go to step 4.
4. *Update initial values*: Set $n := n+1$, $\hat{\gamma}_{k,t}^{(n)} = \gamma_{k,t}^{*}$, go to step 2.

VI. Numerical Results

In this section, we provide numerical results to evaluate the performance of our proposed transmission and power allocation schemes for the multi-way massive MIMO relaying.

First, Fig. 1 shows the sum spectral efficiency $\mathtt{R}_{\text{sum}}$ versus the number of relay antennas with $K = 10$, $\beta_k = 1$, $\forall k$. Here, uniform power control is considered: $P_{\mathrm{u},k} = P_{\mathrm{u}} = 0$ dB, $\eta_{j(i,t)} = \alpha^{(t)}$ given by [14, Eq. (8)] for conventional schemes; and $P_{\mathrm{u},k} = P_{\mathrm{u}} = 0$ dB, $\eta_{j(i,t)}^{(t)} = P_{\mathrm{r,th}}/(M\sum_{i=1}^{K}\beta_i)$ for the proposed scheme. The "proposed scheme analysis" curve represents our analytical results obtained by using the lower bounds (6), (25), and the asymptotic result (32). The "proposed scheme simulation" curve is generated from the outputs of a Monte-Carlo simulator using (5), (24), and (31). We can see that the proposed approximation is very tight, even with small number of antennas. Figure 1 also compares the performance of our proposed scheme with the one of the conventional DF scheme (Section III-B) and the conventional AF scheme in [14]. We can see that our proposed scheme significantly outperforms other schemes. The sum spectral efficiency of our proposed scheme improves by factors of nearly 2 and 3 compared with the conventional DF scheme and the conventional AF scheme, respectively. This is due to the fact that with the conventional schemes, we need in total K time-slots to exchange the information among the K users, while with our proposed scheme, we need only $\lceil\frac{K-1}{2}\rceil + 1$ time-slots.

Next, we consider a more practical scenario where the large-scale fading β_k changes depending on the locations of the users and the shadow fading. To generate the large-scale

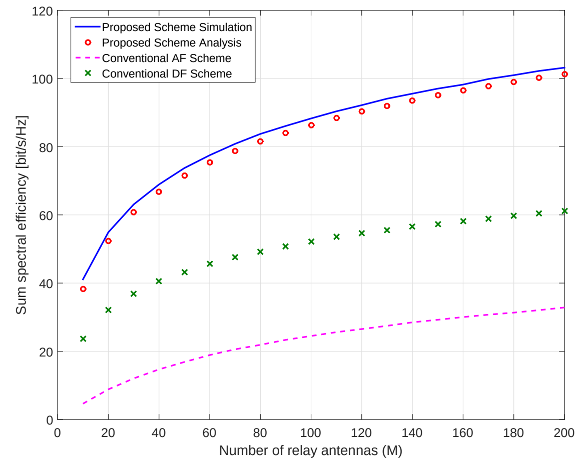

Fig. 1: The comparison of the sum spectral efficiency with different schemes versus the number of relay antennas. We choose $P_{\mathrm{u}} = 0$ dB, $P_{\mathrm{r,th}} = 10$ dB, $K = 10$, $\beta_k = 1$.

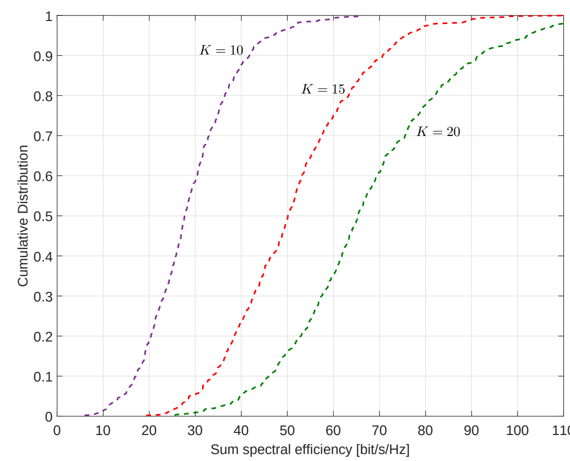

Fig. 2: Cumulative distribution of the sum spectral efficiency for different K. We choose $M = 100$.

fading, we use the same model as in [1]. Figure 2 illustrates the cumulative distribution of the sum spectral efficiency of our proposed scheme for $K = 10, 15$, and 20. The values of $P_{\mathrm{u},k}$ and $P_{\mathrm{r,th}}$ are the same as Fig. 1. As expected, the sum spectral efficiency increases when K increases. The 95%-likely sum spectral efficiency for $K = 20$ is about 40.4 bit/s/Hz which is much higher than that for $K = 10$ and $K = 15$.

Finally, we evaluate the energy efficiency performance given in Section V by solving the power minimization problem for both the users and the relay. The simulation parameters for power allocation are tabulated in Table I [11].

Furthermore, the large-scale fading matrix is chosen by taking one snapshot of the practical setup in Fig. 2, as $\mathbf{D} = \text{diag}[0.0361 \quad 0.0368 \quad 0.0794 \quad 0.01398 \quad 0.0315 \quad 0.0523$

TABLE I:

SYSTEM PARAMETERS FOR THE SIMULATION IN FIG. 3

Parameters	Values
Transmission bandwidth B	20 MHz
The load dependent power P_{LD}	1.15 Watt/(Gbit/s)
The load independent power P_{LID}	5 Watt
The power amplify efficiency of each user $\zeta_{\text{u},k}$	0.2
The power amplify efficiency at the relay ζ_{r}	0.29

0.0235 0.0457 0.0654 0.0776 0.0533 0.0335 0.0657 0.0554 0.0736]. Figure 3 illustrates the energy efficiency versus the sum spectral efficiency under two power schemes namely uniform and proposed power allocation with different number of users K. In this example, we consider the cases with $K = 5$, and 15. The dashed curves correspond to the uniform power allocation where all K users transmit the same power, i.e., $P_{\text{u},k} = P_{\text{u,th}}, \forall k = 1, \dots, K$, and the relay uses their maximum power $P_{\text{r,th}}$. The solid curves correspond to the proposed power allocation by evaluating Algorithm 1. The initial values setup of Algorithm 1 are given as follows: $\epsilon = 10^{-2}$, $\alpha = 1.1$, $L = 5$, and $\hat{\gamma}_{k,t}^{(1)}$ is obtained from the uniform power allocation.

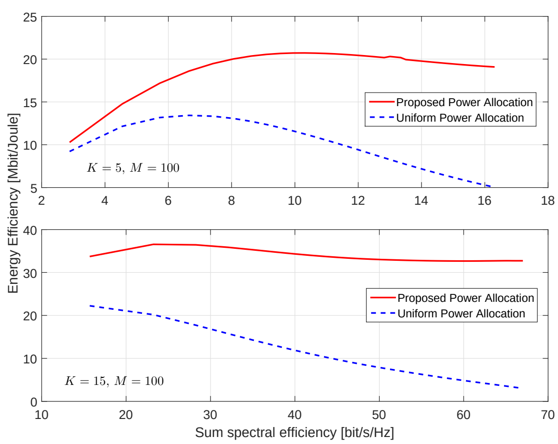

Fig. 3: The energy efficiency versus sum spectral efficiency with uniform and the proposed power allocation schemes. We choose $P_{\text{r,th}} = KP_{\text{u,th}}$, $M = 100$.

We can see that our proposed power control method offers significant better performance compared to using the conventional uniform power allocation method. For instance, with $K = 5$, in order to obtain the sum spectral efficiency of 10 bit/s/Hz, the energy efficiency of the proposed power allocation is nearly double compared to that for the uniform power control scheme. The gap between the two schemes increases significantly as the sum spectral efficiency increases. The reason is that large sum spectral efficiency corresponds to the high power regime in which our proposed power allocation scheme is more beneficial.

VII. Conclusion

We proposed a novel and useful transmission scheme for multi-way massive MIMO relay systems with decode-and-forward protocol at the relay. While the conventional scheme needs K time-slots to exchange all data among K users, our proposed scheme, which is based on successive cancelation decoding, needs only $\left\lceil \frac{K-1}{2} \right\rceil + 1$ time-slots. Thus, the sum spectral efficiency of our proposed scheme is nearly double the sum spectral efficiency of the conventional scheme. In addition, we proposed a power allocation scheme which aims at maximizing the total energy efficiency for a given sum spectral efficiency. The proposed power allocation was iteratively solved via a sequence of GPs, and increased the energy efficiency nearly two times compared to uniform power allocation.

Acknowledgement

This work was supported by project no. 3811/QD-UBND, Binh Duong government, Vietnam. The work of M. Matthaiou was supported in part by the EPSRC under grant EP/P000673/1.

References

[1] 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. Letters*, vol. 21, no. 4, pp. 849–852, Apr. 2017.

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

[3] 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.

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

[5] D. Gündüz, A. Yener, A. Goldsmith, and H. V. Poor, "The multiway relay channel," *IEEE Trans. Inf. Theory*, vol. 59, no. 1, pp. 51–63, Jan. 2013.

[6] Y. Tian and A. Yener, "Degrees of freedom for the MIMO multi-way relay channel," *IEEE Trans. Inf. Theory*, vol. 60, no. 5, pp. 2495–2511, May 2014.

[7] A. Amah and A. Klein, "Non-regenerative multi-way relaying with linear beamforming." in *Proc. IEEE PIMRC*, Sep. 2009, pp. 1843–1847.

[8] G. Amarasuriya, E. G. Larsson, and H. V. Poor, "Wireless information and power transfer in multi-way massive MIMO relay networks," *IEEE Trans. Wireless Commun.*, vol. 15, no. 6, pp. 3837–3855, June 2015.

[9] G. Amarasuriya and H. V. Poor, "Multi-way amplify-and-forward relay networks with massive MIMO," in *Proc. IEEE PIMRC*, Sep. 2014, pp. 595–600.

[10] S. Wagner, R. Couillet, M. Debbah, and D. T. Slock, "Large system analysis of linear precoding in MISO broadcast channels with limited feedback," *IEEE Trans. Inf. Theory*, vol. 58, no. 7, pp. 4509–4537, July 2012.

[11] 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, Jun. 2015.

[12] P. C. Weeraddana, M. Codreanu, M. Latva-aho, and A. Ephremides, "Resource allocation for cross-layer utility maximization in wireless networks," *IEEE Trans Veh. Technol.*, vol. 60, no. 6, pp. 2790–2809, July 2011.

[13] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, UK: Cambridge University Press, 2004.

[14] C. Ho, H. Q. Ngo, M. Matthaiou, and T. Q. Duong, "Multi-way massive MIMO relay networks with maximum-ratio processing," in *Proc. IEEE SigTelCom*, Jan. 2017, pp. 124–128.