Distributed Multiuser Cooperative Network with Heterogenous Relay Clusters

Celal Eşli and Armin Wittneben
ETH Zurich, Communication Technology Laboratory, CH-8092 Zurich, Switzerland
Email: {cesli,wittneben}@nari.ee.ethz.ch

***Abstract*— We propose a multiuser two-hop cooperative network protocol, where several source/destination pairs communicate concurrently over the same physical channel without interfering to each other. A set of half-duplex amplify-and-forward relays assisting them are partitioned into two clusters which have different needs for channel state information (CSI) for calculating the relay gain factors. The cluster composed of relays requiring only local CSI contributes to the received signal power at the destinations, i.e., array gain, but causes multiuser interference. The second cluster needing global CSI within the cluster, cancels the multiuser interference at the destinations in a zero-forcing manner, and can achieve distributed diversity gain through relay selection. It is shown that with drastically reduced CSI exchange overhead, the proposed protocol approaches the rate performances of conventional multiuser relaying systems requiring global network CSI.**

I. Introduction

Cooperative relaying has been considered as an efficient means of improving robustness and spectral efficiency in wireless networks with large number of nodes. In [1], [2] cooperative diversity is introduced by using distributed antennas of different users and generating a virtual array by distributed transmission and signal processing.

The capacity scaling laws of wireless two-hop relay networks are given in [3] for single-user and multiuser (MU) relaying cases. In multiuser relaying, several source/destination (S/D) pairs communicate concurrently over the same physical channel, and the relays process the signals of multiple S/D pairs jointly. Coherent amplify-and-forward (AF) relaying benefits from the distributed beamforming, which let the relayed signals add up coherently at the destinations. When there are infinitely many relays, matching the relay gains appropriately to the channel coefficients, orthogonalizes data streams of different S/D pairs, where the relays only need to know their own backward and forward channel state information (CSI), i.e., local CSI, [3], [4]. For the case of a finite number of relays, a zero-forcing (ZF) scheme based on nullspace projection is proposed in [4]. A minimum number of relays is needed to orthogonalize a given number of S/D pairs. In contrast to [3], this MU-ZF relaying (MUZFR) scheme does not entail multiple antennas at the relays but requires each relay to disseminate its local CSI to all other relays in the network (i.e. global network CSI), which brings an overhead to the system. Having more relays than the necessary minimum number of relays for orthogonalization (i.e. excess relays case), the relay gains can be further optimized [5] to achieve distributed diversity or to maximize sum rate at the expense of increased local CSI (LCSI) exchange load, which can be a crucial overhead in a dense network.

In this paper, we focus on the excess number of relays in the network and propose a heterogeneous relaying protocol (HRP), which achieves the rate performances of MUZFR system with drastically reduced CSI exchange overhead. We partition the relays into two clusters. In the first cluster C_{local}, the relays only need LCSI and cause MU interference at the destinations. The members of the second cluster C_{global} need global CSI (GCSI) within the cluster C_{global} plus the partial channel information of C_{local} and are responsible for orthogonalizing S/D pairs. The minimum number of relays for C_{global} needed to satisfy S/D orthogonalization is found and gain allocation is derived analytically. The cases of excess and lack of relays in C_{global} are also addressed, and the relay gain optimization is modeled as a semidefinite program (SDP). Moreover, assuming a fixed total relay power, the necessary power allocation between cluster has been derived. Diversity gain and improved sum rate are achieved through relay selection for C_{global}.

Notation: We use boldface lowercase and capital letters to indicate vectors and matrices, respectively. The superscripts $(\cdot)^*$, $(\cdot)^T$, $(\cdot)^H$, $(\cdot)^+$ stand for complex conjugate, matrix transpose, complex conjugate transpose, and pseudoinverse, respectively. The operators $\odot$, $\mathrm{E}_{\{x\}}[\cdot]$, $\mathrm{diag}\{\mathbf{x}\}$, $\mathrm{tr}(\mathbf{X})$, $x!$, $\mathrm{Re}\{x\}$, and $\mathrm{Im}\{x\}$ denote the elementwise product, expectation with respect to x, a diagonal matrix with $\mathbf{x}$ on its diagonal, the trace of the matrix $\mathbf{X}$, the factorial of x, the real part of x, and the imaginary part of x, respectively. $x \sim \mathcal{CN}(0, \sigma^2)$ stands for a zero-mean complex normal distribution with variance σ^2.

II. System and Signal Model

We consider a network with N single-antenna S/D pairs and N_r single-antenna AF relays assisting the communication in a half-duplex scheme. According to their need for CSI to determine relay gain factors, the relays are grouped into two clusters C_{local} and C_{global} with $N_{r,1}$ and $N_{r,2}$ relays, respectively. A relay in C_{local} requires only its own LCSI to compute the amplification gain, whereas a C_{global} relay needs the LCSIs of all of the other relays within the cluster in addition to its LCSI, i.e., GCSI within the cluster. It is assumed that there are no direct links between sources and destinations due to shadowing effects or topological conditions. Fig. 1 summarizes the network configuration.

978-1-4244-2110-7/08/$25.00 ©2007 IEEE

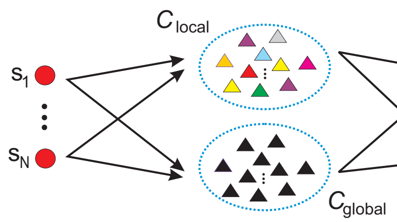

Fig. 1. Heterogenous clusters based two-hop network configuration.

We assume that each relay has its own perfect LCSI knowledge before the communication protocol starts. The communication consists of two phases. In the first phase, called *estimation and feedback phase* (EFP), C_{global} is silent and only C_{local} is active and performs a regular two-hop half-duplex relaying routine. This phase is not performed for data transmission, but for the estimation of the equivalent channel of C_{local}, and then feedbacking this information to C_{global}. Due to our low mobility environment assumption, we assume that the channel stays constant over $2(T_1+T_2)$ time slots. The first $2T_1$ slots are used for the first phase, whereas the last $2T_2$ slots are occupied by the second phase for the actual data transmission. The second phase is called *zero-forcing relaying phase* (ZFP), and both clusters are active. The role of the C_{global} in the ZFP is to optimize its gain factors such that the S/D links are orthogonalized meaning that the interference between different S/D pairs is removed. Note that the relays in C_{local} are unaware of the different phases, but always active and relaying. Moreover, each relay in C_{local} is completely independent from the other relays in the network.

We define a transmission cycle as the duration of the transmission until the local channel conditions are needed to be updated again, i.e., the channel conditions change. A transmission cycle lasts for $2(T_1 + T_2)$ time slots, where, independent from the present phase, the transmitters communicate with relays in the odd indexed slots (i.e., the first hop), and the relays communicate with the destinations in the even indexed slots (i.e., the second hop). The complex scalar transmit symbols at the pth slot are stacked in the vector

$$\mathbf{s}_p = \left[\ s_{p,1}\ s_{p,2}\ \ldots\ s_{p,N}\ \right]^T, \text{ for } p = 1,3,\ldots,2(T_1+T_2)-1.$$

We define $\mathcal{S}_{\text{local}}$ and $\mathcal{S}_{\text{global}}$ as the index sets of relays in C_{local} and C_{global}, respectively. In the following, we present the details of the two communication phases.

A. The Estimation and Feedback Phase

During the first phase slots, i.e., $p = 1,2,\ldots,2T_1$, $\mathbf{s}_p$ is the orthogonal training sequence. In the first hop, $\mathbf{s}_p$ is transmitted only to the C_{local}, since the C_{global} is inactive in the EFP. In this phase, the received signal at the jth relay is

$$r_{p,j} = \sum_{i=1}^{N} h_{j,i} s_{p,i} + w_j, \quad \text{for } p = 1,3,\ldots,2T_1-1 \tag{1}$$

where $j \in \mathcal{S}_{\text{local}}$, $h_{j,i} \sim \mathcal{CN}(0,1)$ is the uplink channel coefficient between the ith source and the jth relay, $w_j \sim \mathcal{CN}(0,\sigma^2_{w_j})$ is the additive white Gaussian noise (AWGN). Upon receiving $r_{p,j}$, each C_{local} relay multiplies its received symbol with the gain factor v_j which is calculated according to *the asymptotic zero-forcing (AZF) gain allocation* proposed in [3], [4]. Hence, the gain factor of the jth relay is

$$v_j = \mathbf{h}_j^H \mathbf{g}_j^*,$$

where $j \in \mathcal{S}_{\text{local}}$, $\mathbf{h}_j = \left[\ h_{j,1}\ h_{j,2}\ \ldots\ h_{j,N}\ \right]^T$ collects the uplink channel coefficients from all sources to the jth relay, $\mathbf{g}_j = \left[\ g_{1,j}\ g_{2,j}\ \ldots\ g_{N,j}\ \right]^T$ collects the downlink channel coefficients from the jth relay to all destinations, and $g_{i,j} \sim \mathcal{CN}(0,1)$ is the channel coefficient between the jth relay and the ith destination. Before forwarding the amplified signals to the destinations, relays scale their signals' power such that each of them transmits with power p_{local}. Hence, the total transmit power of C_{local} is $P_{\text{local}} = N_{r,1} p_{\text{local}}$. We use a per-node power constraint in C_{local} to avoid dependence on other relays in the cluster. Thus, the signal to be transmitted from the jth relay in the second hop is

$$\tilde{r}_{p+1,j} = \sqrt{p_{\text{local}}}\mathbf{h}_j^H \mathbf{g}_j^* r_{p,j} / |\mathbf{h}_j^H \mathbf{g}_j^* r_j|,$$

for $p = 1,3,\ldots,2T_1-1$.

The received signal at the ith destination is

$$\begin{aligned} d_{p+1,i} &= \sum_{j\in\mathcal{S}_{\text{local}}} g_{i,j} u_j h_{j,i} s_{p,i} + \sum_{k=1,\,k\neq i}^{N} \left(\sum_{j\in\mathcal{S}_{\text{local}}} g_{i,j} u_j h_{j,k} \right) s_{p,k} \\ &\quad + \sum_{j\in\mathcal{S}_{\text{local}}} g_{i,j} u_j w_j + n_i \\ &= f^{\text{sd}}_{i,i} s_{p,i} + \sum_{k=1,\,k\neq i}^{N} f^{\text{sd}}_{i,k} s_{p,k} + \tilde{w}_i + n_i \end{aligned} \tag{2}$$

for $p = 1,3,\ldots,2T_1-1$, where $u_j = \sqrt{p_{\text{local}}}\mathbf{h}_j^H \mathbf{g}_j^* / |\mathbf{h}_j^H \mathbf{g}_j^* r_j|$, $f^{\text{sd}}_{i,k}$ represents the equivalent channel between the ith source and the kth destination and $\tilde{w}_i$ is the equivalent amplified relay noise, and $n_i \sim \mathcal{CN}(0,\sigma^2_{n_i})$ is the destination noise term. Using the knowledge of the training sequence and their orthogonality, the ith destination can estimate the equivalent channel for the desired signal $f^{\text{sd}}_{i,i}$, and the equivalent channels for interference $\{f^{\text{sd}}_{i,k}, k = 1,\ldots,N, k \neq i\}$. We define the equivalent channel matrix $\mathbf{F}_{\text{sd}} \in \mathbb{C}^{N\times N}$ of C_{local} as

$$\mathbf{F}_{\text{sd}} = \begin{bmatrix} f^{\text{sd}}_{1,1} & f^{\text{sd}}_{1,2} & \cdots & f^{\text{sd}}_{1,N} \\ f^{\text{sd}}_{2,1} & f^{\text{sd}}_{2,2} & \cdots & f^{\text{sd}}_{2,N} \\ \vdots & \vdots & & \\ f^{\text{sd}}_{N,1} & f^{\text{sd}}_{N,2} & \cdots & f^{\text{sd}}_{N,N} \end{bmatrix}.$$

The last step of EFP is to feedback the $N(N-1)$ off-diagonal elements of the matrix $\mathbf{F}_{\text{sd}}$ to the relays of C_{global} through a secure feedback channel. Here and in the sequel, we assume that EFP is error-free and C_{global} is supplied with the perfect channel state information.

B. The Zero-Forcing Relaying Phase

As the first phase is completed and the partial equivalent channel information of C_{local} is fed back to the C_{global}, the zero-forcing relaying phase takes place. In this phase, the actual data transmission starts and both clusters aid the communication between S/D pairs. Using the global CSI in the cluster and the feedback partial C_{local} CSI, the activated C_{global} relays choose their gain factors such that the interference between different S/D pairs is removed. In other words, the gain factor of each relay in C_{global} is dependent on its own LCSI, LCSIs of the other relays in C_{global}, and partial $\mathbf{F}_{\text{sd}}$.

The received signal model of the jth relay is given by (1), where $j \in \{\mathcal{S}_{\text{global}} \cup \mathcal{S}_{\text{local}}\}$, $p = 2T_1+1, 2T_1+3, \ldots, 2(T_1+T_2)-1$ and $\mathbf{s}_p$ carries the actual messages of the sources to the destinations. The C_{local} relays continue to relay as explained in Section II-A. On the other hand, the relay gain factors, x_l, $l \in \mathcal{S}_{\text{global}}$, of C_{global} relays are chosen such that the S/D links are orthogonalized. The details of the C_{global} relay gain allocation are presented in the next section. Henceforth, the received signal at the ith destination is

$$d_{p+1,i} = \sum_{j \in \mathcal{S}_{\text{local}}} g_{i,j}\Big(u_j \sum_{k=1}^{N} h_{j,k} s_{p,k} + u_j w_j\Big) + \sum_{l \in S_{\text{global}}} g_{i,l}\Big(x_l \sum_{k=1}^{N} h_{l,k} s_{p,k} + x_l v_l\Big) + n_i, \quad (3)$$

for $p = 2T_1+1, 2T_1+3, \ldots, 2(T_1+T_2)-1$, where v_l is the noise term with $\mathcal{CN}(0, \sigma_{v_l}^2)$ of the relay with index l in C_{global}. Defining $\tilde{f}_{i,j}^{\text{sd}}$ as the equivalent channel between the ith source and the jth destination observed only through C_{global} relays, and $\tilde{v}_i$ as the equivalent C_{global} relays amplified noise term, $d_{p+1,i}$ can be expressed as

$$\begin{aligned} d_{p+1,i} &= \big(f_{i,i}^{\text{sd}} + \tilde{f}_{i,i}^{\text{sd}}\big)s_{p,i} + \sum_{k=1,k\neq i}^{N} \big(f_{i,k}^{\text{sd}} + \tilde{f}_{i,k}^{\text{sd}}\big)s_{p,k} + \tilde{v}_i + \tilde{w}_i + n_i \\ &= \big(f_{i,i}^{\text{sd}} + \tilde{f}_{i,i}^{\text{sd}}\big)s_{p,i} + \tilde{v}_i + \tilde{w}_i + n_i, \end{aligned} \quad (4)$$

where the second line follows from the fact that the relay gains $x_l, l \in \mathcal{S}_{global}$, are chosen such that the multi-user interference has been cancelled, i.e., $\sum_{k=1,k\neq i}^{N}\big(f_{i,k}^{\text{sd}} + \tilde{f}_{i,k}^{\text{sd}}\big) = 0$. The total transmission power of C_{global} is P_{global}, which adds to the total power consumption of all relays as $P_{\text{total}} = P_{\text{global}} + P_{\text{local}}$. We impose a fixed total relay power P_{total} to the network to keep the total power consumption constant when utilizing additional relays. In the sequel, it is assumed that $\sigma_{v_l}^2 = \sigma_{w_j}^2 = \sigma_{n_i}^2 = \sigma_n^2, \forall i, j, l$ for the sake of notational simplicity.

III. Zero-Forcing Relay Gain Allocation

In the proposed heterogeneous clusters based MU relaying system, C_{global} is responsible for cancelling the multi-user interference at the destinations caused by C_{global} itself and C_{local}. Each of C_{local} relays is unaware and independent of the other relays in the network. Their gain allocations scheme is dependent only on the LCSI, whereas C_{global} relays exchange CSI within the cluster and receive channel knowledge of C_{local} to decide on the zero-forcing gain allocations.

Let us make the following definitions before proceeding further. The column vectors $\breve{\mathbf{h}}_i$ and $\breve{\mathbf{g}}_i$ collect all the channel coefficients from ith source to all relays in C_{global}, and from all relays in C_{global} to the ith destination, respectively. The compound interference matrix $\mathbf{A} \in \mathbb{C}^{N(N-1) \times N_{r,2}}$ comprises the row vectors $(\breve{\mathbf{h}}_i \odot \breve{\mathbf{g}}_k)^T$ for all $i, k \in \{1, \ldots, N\}$ and $i \neq k$. Then, $\mathbf{Ax}$ delivers a vector containing the equivalent channel coefficients of all multi-user interference terms caused by C_{global}, where $\mathbf{x} = [x_1 \; x_2 \; x_{N_{r,2}}]^T$. Finally, defining the vector $\mathbf{b} \in \mathbb{C}^{N(N-1)}$ as the collection of $f_{i,k}^{\text{sd}}$ with $i, k \in \{1, \ldots, N\}$ and $i \neq k$, the condition for zero-forcing the total multi-user interference is

$$\mathcal{C}1: \quad \mathbf{Ax} + \mathbf{b} = 0. \quad (5)$$

The possible set of solutions of (5) depends on the rank of $\mathbf{A}$. In the following we investigate three different cases for the relation between $N_{r,2}$ and $N(N-1)$. Note that the subsequent relay gain allocations proposed for different $N_{r,2}$ values are independent of $N_{r,1}$.

A. Minimum relay configuration ($N_{r,2} = N(N-1)$)

There is a unique solution for the case $N_{r,2} = N(N-1)$, which is $\mathbf{x}_{\text{zf}} = -\mathbf{A}^{+}\mathbf{b}$. Given that there is a fixed amount of P_{total}, this solution also predefines the total transmission power of each cluster, i.e., P_{global} and P_{local}. The following theorem summarizes this predefined power allocation.

Theorem 1: *Given that $N_{r,2} = N(N-1)$ and P_{total} is fixed, there is a unique cluster power allocation*

$$\big(P_{\text{local}}, P_{\text{global}}\big) = \left(\frac{P_{\text{total}}}{1+\alpha}, \frac{\alpha P_{\text{total}}}{1+\alpha}\right) \quad (6)$$

where $\mathbf{b}_u = \mathbf{b}/\sqrt{P_{\text{local}}}$, $\tilde{\mathbf{H}} = [\breve{\mathbf{h}}_1, \breve{\mathbf{h}}_2, \cdots, \breve{\mathbf{h}}_N]$, and $\alpha = (\mathbf{A}^{+}\mathbf{b}_u)^H\big(\sigma_n^2 \mathbf{I}_{N_{r,2}} + \sigma_s^2(\tilde{\mathbf{H}}\tilde{\mathbf{H}}^H) \odot \mathbf{I}_{N_{r,2}}\big)(\mathbf{A}^{+}\mathbf{b}_u)$.

Proof: The proof follows immediately from the definition of the total transmission power of C_{global}, which is

$$\begin{aligned} P_{\text{global}} &= \mathbf{x}_{\text{zf}}^H\big(\sigma_n^2 \mathbf{I}_{N_{r,2}} + \sigma_s^2(\tilde{\mathbf{H}}\tilde{\mathbf{H}}^H) \odot \mathbf{I}_{N_{r,2}}\big)\mathbf{x}_{\text{zf}} \\ &= P_{\text{local}}(\mathbf{A}^{+}\mathbf{b}_u)^H\big(\sigma_n^2 \mathbf{I}_{N_{r,2}} + \sigma_s^2(\tilde{\mathbf{H}}\tilde{\mathbf{H}}^H) \odot \mathbf{I}_{N_{r,2}}\big)(\mathbf{A}^{+}\mathbf{b}_u) \\ &= \alpha P_{\text{local}}. \end{aligned} \quad (7)$$

Inserting (7) into $P_{\text{local}} + P_{\text{global}} = P_{\text{total}}$ concludes the proof.□

B. Excess number of relays ($N_{r,2} > N(N-1)$)

Having more C_{global} relays than $N(N-1)$ results in infinitely many solutions for (5). This case implies that the nullspace of $\mathbf{A}$ contains more than the trivial solution, i.e., the all zero vector. Defining $\mathbf{Z}$ = null$\{\mathbf{A}\}$, any vector in the nullspace $\mathbf{Z}$ can be added to any particular solution $\mathbf{x}_o$ of (5), such that

$$\mathcal{C}2: \quad \mathbf{A}(\mathbf{x}_o + \mathbf{Z}\mathbf{x}_n) + \mathbf{b} = 0, \quad (8)$$

where $\mathbf{Z} \in \mathbb{C}^{N_{r,2} \times N_{r,2} - N(N-1)}$ and $\mathbf{x}_n \in \mathbb{C}^{N_{r,2} - N(N-1)}$. Hence, in addition to cancelling the multi-user interference, the gain factors can be further optimized for given figures of merit such as sum rate, fairness or diversity. The vectors $\mathbf{x}_o$, $\mathbf{x}_n$ that are satisfying the $\mathcal{C}2$, enforces a specific set of cluster power allocation pairs, i.e., $(P_{\text{local}}, P_{\text{global}})$.

Theorem 2: *Given that $N_{r,2} > N(N-1)$ and P_{total} is fixed, the feasible set for cluster power allocation pairs is*

$$0 \leq P_{\text{local}} \leq \frac{P_{\text{total}}}{1+\eta} \quad \rightarrow \quad \frac{\eta P_{\text{total}}}{1+\eta} \leq P_{\text{global}} \leq P_{\text{total}}. \qquad (9)$$

Proof: Since $\mathbf{x}_o$ is a particular solution of (5) for given channel conditions and P_{local}, the transmission power of C_{global} is only a function of $\mathbf{x}_n$ as

$$\begin{aligned} f_{P_{\text{global}}}(\mathbf{x}_n) &= (\mathbf{x}_o + \mathbf{Z}\mathbf{x}_n)^H \mathbf{M} (\mathbf{x}_o + \mathbf{Z}\mathbf{x}_n) \\ &= \mathbf{x}_n^H \mathbf{Z}^H \mathbf{M}\mathbf{Z}\mathbf{x}_n + 2\,\mathrm{Re}\{\mathbf{x}_n^H \mathbf{Z}^H \mathbf{M}\mathbf{x}_o\} + \mathbf{x}_o^H \mathbf{M}\mathbf{x}_o, \end{aligned}$$

where $\mathbf{M} = \sigma_n^2 \mathbf{I}_{N_{r,2}} + \sigma_s^2 (\tilde{\mathbf{H}}\tilde{\mathbf{H}}^H) \odot \mathbf{I}_{N_{r,2}}$. The cluster transmission powers are optimization contraints and should be satisfied with equality. The matrix $\mathbf{M}$ is positive-semi definite by definition; hence $f_{P_{\text{global}}}(\mathbf{x}_n)$ is a convex function attaining its minimum at $\mathbf{x}_n^{\min} = -(\mathbf{Z}^H\mathbf{M}\mathbf{Z})^{-1}\mathbf{Z}^H\mathbf{M}\mathbf{x}_o$. Next, it can be shown that

$$f_{P_{\text{global}}}(\mathbf{x}_n^{\min}) = P_{\text{local}} f_{P_{\text{global}}}(\mathbf{x}_n^{\min})\big|_{\mathbf{b}=\mathbf{b}_u}, \qquad (10)$$

where $f_{P_{\text{global}}}(\mathbf{x}_n)\big|_{\mathbf{b}=\mathbf{b}_u}$ denotes that $\mathbf{x}_n^{\min}$ is computed for a particular solution vector $\mathbf{x}_o$ evaluated for $\mathbf{b}_u$, and then fed as input to the function $f_{P_{\text{global}}}(\cdot)$. If $f_{P_{\text{global}}}(\mathbf{x}_n^{\min})$ is smaller than or equal to the design parameter P_{global}, this leads to the conclusion that there is at least one $\mathbf{x}_n$ that satisfies the power constraints. Thus, we have the following constraints to satisfy

$$\eta P_{\text{local}} \leq P_{\text{global}}, \qquad (11)$$

$$P_{\text{global}} + P_{\text{local}} = P_{\text{total}}, \qquad (12)$$

$$P_{\text{global}}, P_{\text{local}} \geq 0, \qquad (13)$$

where $\eta = f_{P_{\text{global}}}(\mathbf{x}_n)\big|_{\mathbf{b}=\mathbf{b}_u}$. Combining and solving the inequalities (11-13), (9) can be written immediately. □

For any given feasible power allocation pair $\big(P_{\text{local}}, P_{\text{global}}\big)$, the additional degrees of freedom can be used to optimize the gain allocations such that the instantaneous minimum link rate is maximized. We prefer to adopt such a *max-min fairness* approach rather than maximizing the sum rate because the latter approach favors the strong links while punishing the weak ones.

The instantaneous rate of the ith S/D link is

$$\mathrm{R}_i = \frac{1}{2}\log_2(1 + \mathrm{SNR}_i), \qquad (14)$$

and the instantaneous signal-to-noise ratio SNR_i is

$$\mathrm{SNR}_i = \frac{\sigma_s^2 \left(\begin{array}{c} (\mathbf{x}_o + \mathbf{Z}\mathbf{x}_n)^H \mathbf{U}_{i,i} (\mathbf{x}_o + \mathbf{Z}\mathbf{x}_n) \\ +2\,\mathrm{Re}\big\{(\mathbf{x}_o + \mathbf{Z}\mathbf{x}_n)^H (\breve{\mathbf{h}}_i \odot \breve{\mathbf{g}}_i)^* f_{i,i}^{\text{sd}}\big\} + |f_{i,i}^{\text{sd}}|^2 \end{array} \right)}{\sigma_{\breve{w}_i}^2 + \sigma_n^2\big(1 + (\mathbf{x}_o + \mathbf{Z}\mathbf{x}_n)^H \mathbf{Q}_i (\mathbf{x}_o + \mathbf{Z}\mathbf{x}_n)\big)}$$

where $\mathbf{U}_{i,k} = (\breve{\mathbf{h}}_k \odot \breve{\mathbf{g}}_i)^* (\breve{\mathbf{h}}_k \odot \breve{\mathbf{g}}_i)^T$, and $\mathbf{Q}_i = \mathrm{diag}\{|\breve{\mathbf{g}}_{\mathrm{i}}|^2\}$. We should remark here that with the previously defined communication protocol the relay amplified noise variance $\sigma_{\breve{w}_i}^2$ is not estimated. Hence, we assume that T_3 time slots should be spent for this estimation on top of $2(T_1 + T_2)$ transmission cycle. The max-min fairness optimization problem follows as

$$\begin{aligned} &\arg\max_{\mathbf{x}_n} \min_{1 \leq i \leq N} \ \mathrm{R}_i \\ &\text{subject to} \ \ (\mathbf{x}_o + \mathbf{Z}\mathbf{x}_n)^H \mathbf{M} (\mathbf{x}_o + \mathbf{Z}\mathbf{x}_n) \leq P_{\text{global}}. \qquad (15) \end{aligned}$$

Using the monotonic equivalence of R_i and SNR_i, and introducing a dummy variable $\gamma > 0$, (15) is modified as

$$\begin{aligned} &\arg\max_{\gamma, \mathbf{x}_n} \ \gamma \\ &\text{subject to} \quad \mathbf{x}_n^H \boldsymbol{\Omega}_i \mathbf{x}_n + 2\,\mathrm{Re}\big\{\mathbf{x}_n^H \boldsymbol{\omega}_i\big\} + \omega_i \geq 0 \ \ \forall i \\ &\qquad\qquad\quad \mathbf{x}_n^H \boldsymbol{\Phi} \mathbf{x}_n + 2\,\mathrm{Re}\big\{\mathbf{x}_n^H \boldsymbol{\phi}\big\} + \phi \leq P_{\text{global}}. \qquad (16) \end{aligned}$$

where $\sigma_s^2 = 1$,

$$\begin{aligned} \boldsymbol{\Omega}_i &= \mathbf{Z}^H (\mathbf{U}_{i,i} - \gamma\sigma_n^2 \mathbf{Q}_i)\mathbf{Z}, \\ \boldsymbol{\omega}_i &= \mathbf{Z}^H \big(\mathbf{U}_{i,i}\mathbf{x}_o + (\breve{\mathbf{h}}_i \odot \breve{\mathbf{g}}_i)^* f_{i,i}^{\text{sd}} - \gamma\sigma_n^2 \mathbf{Q}_i \mathbf{x}_o\big), \\ \omega &= \mathbf{x}_o^H (\mathbf{U}_{i,i} - \gamma\sigma_n^2 \mathbf{Q}_i)\mathbf{x}_o + |f_{i,i}^{\text{sd}}|^2 - \gamma(\sigma_{\breve{w}_i}^2 + \sigma_n^2) \\ &\quad +2\,\mathrm{Re}\{\mathbf{x}_o^H (\breve{\mathbf{h}}_i \odot \breve{\mathbf{g}}_i)^* f_{i,i}^{\text{sd}}\}, \\ \boldsymbol{\Phi} &= \mathbf{Z}^H \mathbf{M}\mathbf{Z}, \boldsymbol{\phi} = \mathbf{Z}^H \mathbf{M}\mathbf{x}_o, \phi = \mathbf{x}_o^H \mathbf{M}\mathbf{x}_o. \end{aligned}$$

We convert (16) to real-valued form to facilitate the problem by introducing

$$\bar{\mathbf{k}} = [\mathrm{Re}\{\mathbf{k}\}^T \ \mathrm{Im}\{\mathbf{k}\}^T]^T, \qquad \bar{\mathbf{K}} = \begin{bmatrix} \mathrm{Re}\{\mathbf{K}\} & -\mathrm{Im}\{\mathbf{K}\} \\ \mathrm{Im}\{\mathbf{K}\} & \mathrm{Re}\{\mathbf{K}\} \end{bmatrix}$$

for vectors and matrices, respectively. Finally, defining $\bar{\mathbf{X}}_n = \bar{\mathbf{x}}_n \bar{\mathbf{x}}_n^T$ and using the relation $\bar{\mathbf{x}}_n^T \bar{\boldsymbol{\Omega}}_i \bar{\mathbf{x}}_n = \mathrm{tr}(\bar{\boldsymbol{\Omega}}_i \bar{\mathbf{x}}_n \bar{\mathbf{x}}_n^T) = \mathrm{tr}(\bar{\boldsymbol{\Omega}}_i \bar{\mathbf{X}}_n)$, (16) becomes

$$\begin{aligned} &\arg\max_{\gamma, \bar{\mathbf{X}}_n} \ \gamma \\ &\text{subject to} \quad \mathrm{tr}(\bar{\boldsymbol{\Omega}}_i \bar{\mathbf{X}}_n) + 2\bar{\mathbf{x}}_n^T \bar{\boldsymbol{\omega}}_i + \omega_i \geq 0 \ \ \forall i, \\ &\qquad\qquad\quad \mathrm{tr}(\bar{\boldsymbol{\Phi}} \bar{\mathbf{X}}_n) + 2\bar{\mathbf{x}}_n^T \bar{\boldsymbol{\phi}} + \phi \leq P_{\text{global}}, \\ &\qquad\qquad\quad \bar{\mathbf{X}}_n = \bar{\mathbf{x}}_n \bar{\mathbf{x}}_n^T. \qquad (17) \end{aligned}$$

The problem (17) is non-convex because of the rank-1 constraint of $\bar{\mathbf{X}}_n$, but it can be relaxed such that we have the following quasi-convex problem which can be efficiently solved by semidefinite programming,

$$\mathcal{P}_1 := \left\{ \begin{array}{l} \arg\max_{\gamma, \bar{\mathbf{x}}_n, \bar{\mathbf{X}}_n} \quad \gamma \\ \text{subject to} \quad \mathrm{tr}(\bar{\boldsymbol{\Omega}}_i \bar{\mathbf{X}}_n) + 2\bar{\mathbf{x}}_n^T \bar{\boldsymbol{\omega}}_i + \omega_i \geq 0 \ \ \forall i, \\ \qquad\qquad\quad \mathrm{tr}(\bar{\boldsymbol{\Phi}} \bar{\mathbf{X}}_n) + 2\bar{\mathbf{x}}_n^T \bar{\boldsymbol{\phi}} + \phi \leq P_{\text{global}}, \\ \qquad\qquad\quad \begin{bmatrix} \bar{\mathbf{X}}_n & \bar{\mathbf{x}}_n \\ \bar{\mathbf{x}}_n^T & 1 \end{bmatrix} \geq 0, \ \ \bar{\mathbf{X}}_n \succeq 0 \end{array} \right.$$

where the rank-1 constraint is removed and a lower bound to (17) is obtained. The relaxed problem $\mathcal{P}_1$ is equivalent to a SDP feasibility problem when γ is given a priori and can be efficiently solved by SeDuMi [6]. Thus, we use the following bisection algorithm $\mathcal{A}_1$ to solve $\mathcal{P}_1$. While implementing $\mathcal{A}_1$, γ_{min} is set to 0, γ_{max} is chosen large enough according to the operation mean SNR value, and ϵ is a small positive number indicating the precision of the result.

We would like to remark here that so far we assume that the gain allocation is optimized for a given fixed cluster power allocation pair $\big(P_{\text{local}}, P_{\text{global}}\big)$. This fixed power allocation assures that each of C_{local} relays is completely independent from the rest of the network except the first power assignment. Nevertheless, constricting C_{local} relays' independence, an optimization can also be performed over the feasible set of $\big(P_{\text{local}}, P_{\text{global}}\big)$ as defined by *Theorem 2*, which presumably improves the minimum link rate, but is dropped here because of the brevity of the scope of the paper.

Algorithm $\mathcal{A}_1$: Bisection Algorithm for SDP feasibility problem

initiate: $\rightarrow \gamma \in [\gamma_{\min},\ \gamma_{\max}]$

repeat: $\rightarrow \gamma = (\gamma_{\min} + \gamma_{\max})/2$

$\rightarrow$ ***solve the feasibility problem for*** γ :

$$\mathcal{FP}_1 := \left\{ \begin{array}{r} \mathrm{tr}(\bar{\boldsymbol{\Omega}}_i \bar{\mathbf{X}}_n) + 2\bar{\mathbf{x}}_n^T \bar{\boldsymbol{\omega}}_i + \omega_i \geq 0 \ \ \forall i, \\ \mathrm{tr}(\bar{\boldsymbol{\Phi}} \bar{\mathbf{X}}_n) + 2\bar{\mathbf{x}}_n^T \bar{\boldsymbol{\phi}} + \phi \leq P_{\mathrm{global}}, \\ \begin{bmatrix} \bar{\mathbf{X}}_n & \bar{\mathbf{x}}_n \\ \bar{\mathbf{x}}_n^T & 1 \end{bmatrix} \geq 0, \ \ \bar{\mathbf{X}}_n \succeq 0 \end{array} \right.$$

$\rightarrow$ ***if feasible***

$\gamma_{\min} := \gamma$

$\rightarrow$ ***else***

$\gamma_{\max} := \gamma$

until: $\rightarrow \gamma_{\max} - \gamma_{\min} < \epsilon$

C. Lack of relays ($N_{r,2} < N(N-1)$)

Although the condition $N_{r,2} < N(N-1)$ implies that (5) does not have a solution and the multi-user interference can not be cancelled completely, the information about C_{local} can be still used to optimize the gain factors of C_{global} relays. Defining $\mathbf{y} \in \mathbb{C}^{N_{r,2}}$ as the new relay gain amplification vector, the instantaneous signal-to-interference plus noise (SINR) of the ith link can be expressed as

$$\mathrm{SINR}_i = \frac{\left(\mathbf{y}^H \mathbf{U}_{i,i} \mathbf{y} + 2\,\mathrm{Re}\left\{\mathbf{y}^H (\breve{\mathbf{h}}_i \odot \breve{\mathbf{g}}_i)^* f_{i,i}^{\mathrm{sd}}\right\} + |f_{i,i}^{\mathrm{sd}}|^2\right)}{\sum_{k \neq i}\left(\mathbf{y}^H \mathbf{U}_{i,k} \mathbf{y} + 2\,\mathrm{Re}\left\{\mathbf{y}^H (\breve{\mathbf{h}}_k \odot \breve{\mathbf{g}}_i)^* f_{i,k}^{\mathrm{sd}}\right\} + |f_{i,k}^{\mathrm{sd}}|^2\right) + \tilde{\sigma}_n^2},$$

where $\sigma_s^2 = 1$, $\tilde{\sigma}_n^2 = \sigma_{\bar{w}_i}^2 + \sigma_n^2\left(1 + \mathbf{y}^H \mathbf{Q}_i \mathbf{y}\right)$. Adopting a *maxmin* SINR fairness approach and following similar steps (15)-(17), we obtain

$$\mathcal{P}_2 := \left\{ \begin{array}{ll} \arg\max_{\gamma, \bar{\mathbf{y}}, \bar{\mathbf{Y}}} & \gamma \\ \text{subject to} & \mathrm{tr}(\bar{\boldsymbol{\Pi}}_i \bar{\mathbf{Y}}) + 2\bar{\mathbf{y}}^T \bar{\boldsymbol{\pi}}_i + \pi_i \geq 0 \ \ \forall i, \\ & \mathrm{tr}(\bar{\mathbf{M}} \bar{\mathbf{Y}}) \leq P_{\mathrm{global}}, \\ & \begin{bmatrix} \bar{\mathbf{Y}} & \bar{\mathbf{y}} \\ \bar{\mathbf{y}}^T & 1 \end{bmatrix} \geq 0, \ \ \bar{\mathbf{Y}} \succeq 0, \end{array} \right.$$

where

$$\begin{aligned} \boldsymbol{\Pi}_i &= \mathbf{U}_{i,i} - \gamma \sum_{k=1,k\neq i}^{N} \mathbf{U}_{i,k} - \gamma \sigma_n^2 \mathbf{Q}_i, \\ \boldsymbol{\pi}_i &= (\breve{\mathbf{h}}_i \odot \breve{\mathbf{g}}_i)^* f_{i,i}^{\mathrm{sd}} - \gamma \sum_{k=1,k\neq i}^{N} (\breve{\mathbf{h}}_k \odot \breve{\mathbf{g}}_i)^* f_{i,k}^{\mathrm{sd}}, \\ \pi_i &= |f_{i,i}^{\mathrm{sd}}|^2 - \gamma \sum_{k=1,k\neq i}^{N} |f_{i,k}^{\mathrm{sd}}|^2 - \gamma(\sigma_{\bar{w}_i}^2 + \sigma_n^2). \end{aligned}$$

$\mathcal{P}_2$ is solved with a similar bisection algorithm to $\mathcal{A}_1$.

IV. Relay Selection for Clusters

The assignment of relays to clusters can be inherently dependent on geographical conditions of the communication area or topological conditions of different relays with respect to each other. Hence, assigning relays which are in the close proximity of each other to the same cluster would be an efficient and realistic scenario in terms of minimizing CSI exchange overhead. On the other hand, the estimated instantaneous channel conditions can be incorporated to decide on the best cluster structures to maximize the efficiency of the chosen figure of merit, e.g., outage rate, sum rate, etc. In the following, we propose a simple and efficient protocol for the assignment of the relays to clusters, which exploits the trade-off between CSI exchange load and the rate performance.

We partition the relays into sub-clusters of $N_m = N(N-1)$ relays, i.e., the minimum relay configuration. This leads to $\lfloor N_r/N_m \rfloor$ sub-clusters if we would like to have different relays in different clusters, or if we drop this constraint, then we have a total combination of $N_r!/((N_r - N_m)!N_m!)$ sub-clusters. The members of each sub-cluster disseminate their LCSI to the other members of the same sub-cluster, which results GCSI within each sub-cluster.

A slightly modified version of the communication protocol mentioned in Section II is used for relay selection. We assume that all N_r relays in the network are members of the C_{local} cluster and the destinations estimate the equivalent channel, $\hat{\mathbf{F}}_{\mathrm{sd}}$, of the whole relay network. Then, this knowledge is partially fed back to all sub-clusters in the network. Let us now focus on a single sub-cluster, say the qth one, C_q. Since the members of C_q have already exchanged the LCSI of each other, they can subtract this information $\hat{\mathbf{F}}_{\mathrm{sd},q}$ from $\hat{\mathbf{F}}_{\mathrm{sd}}$ such that they obtain the equivalent channel of the relays in the rest of the network, $\hat{\mathbf{F}}'_{\mathrm{sd},q}$. Using the GCSI within the sub-cluster and $\hat{\mathbf{F}}'_{\mathrm{sd},q}$, the qth sub-cluster can calculate the necessary gain allocation vector $\mathbf{x}_{\mathrm{zf},q}$ through Section III-A. Since the sub-clusters only know their effect on amplified noise at the destinations but not the others', they can only approximate the SNR at each destination. Using these approximations, they compute the link rates and the corresponding sum rate that they can supply if they are chosen as the C_{global}. The calculated approximate rates from all sub-clusters are sent back to a master node (e.g., a previously selected source node), which then based on the largest supplied rate, decides which sub-cluster is chosen as the C_{global}.

The number of the sub-clusters defines the trade-off between the performance and the LCSI exchange overhead. Having $N_r!/((N_r - N_m)!N_m!)$ sub-clusters is the extreme case and corresponds to the GCSI throughout the network, which is not preferable when compared with MUZFR with GCSI. On the other hand, defining only one sub-cluster is equivalent to choose relays clusters randomly.

V. Simulation Results

In this section, we present Monte-Carlo simulation results. We use two reference systems, where all relays in the network perform either AZF relaying with local CSI [4] or MUZFR with global network CSI [5]. Unless otherwise stated, $\sigma_s^2 = 1$, and the signal-to-noise ratio $\mathrm{SNR} = \sigma_s^2/\sigma_n^2 = 20$ dB.

Fig. 2 shows that using only the minimum number of relays necessary to orthogonalize S/D links (e.g., 2 relays for $N = 2$), the heterogeneous relaying protocol (HRP) without relay selection (RS) halves the gap between the two reference systems with almost no CSI exchange w.r.t. MUZFR. Adapting the assignment of relays clusters to the estimated channel coefficients, we perform an exhaustive search over all relays

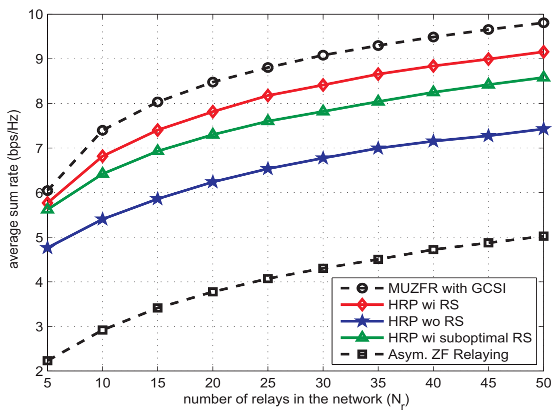

Fig. 2. Average sum rates vs. N_r for $N = 2$, $N_{r,2} = 2$ (i.e., minimum relay configuration for C_{global}), SNR = 20 dB.

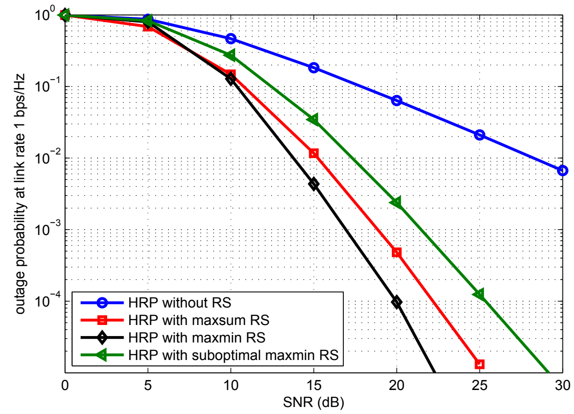

Fig. 3. The outage probability vs. SNR for $N = 2$, $N_r = 6$, $N_{r,2} = 2$ (i.e., minimum relay configuration for C_{global}).

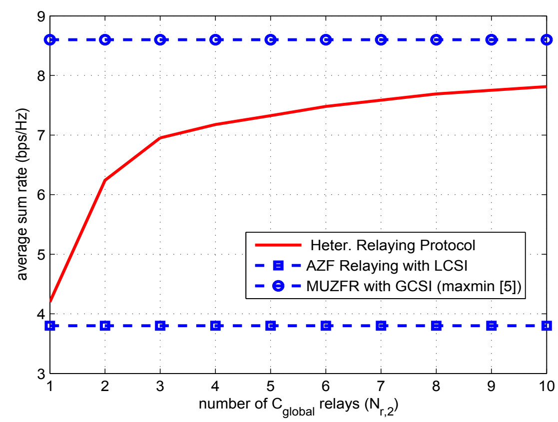

Fig. 4. Average sum rates vs $N_{r,2}$ for the excess relays case in C_{global} and $N = 2$, $N_r = 20$, SNR = 20 dB.

in the network to find the relays to assign to the C_{global} that maximizes the instantaneous sum rate. As depicted in **Fig. 2** with "*HRP wi RS*", the sum rate performance approaches the MUZFR considerably. Nevertheless, such an exhaustive search suffers from LCSI exchange load. Next, we plot the performance of our proposed simple relay selection scheme (i.e., "*HRP wi suboptimal RS*"), where we used only 5 sub-clusters through all N_r values. The proposed scheme performs efficiently without causing too much LCSI exchange load. For instance, for the extreme case of $N_r = 50$, even with 5 sub-clusters the simple relay selection improves the sum rate significantly, where there is a total possible of 50!/(48!2!) = 1225 sub-cluster structures.

For the minimum relay configuration case, adding relays to C_{local} contributes to array gain. However, it is also possible to achieve distributed diversity gain with RS. As depicted in **Fig. 3**, searching for the best C_{global} relays maximizing the minimum link rate achieves diversity (i.e., *maxmin RS*) [5]. Another interesting result in **Fig. 3** is the diversity achieved by *maxsum RS* while selecting the best relays to maximize sum rate.

The cases of excess and lack of relays in C_{global} are evaluated in **Fig. 4**. All possible relay distributions to clusters are investigated for $N_r = 20$. Having one additional relay on top of the minimum relay configuration in C_{global} (i.e., $N_{r,2} = 3$), the HRP doubles the sum rate of AZF. As $N_{r,2}$ approaches to $N_r = 20$, the HRP reaches the performance of MUZFR. Moreover, it is realized that sum rate is not very sensitive to excess number of relays in C_{global} after about $N_{r,2} = 6$. Hence, we can benefit from the trade-off between LCSI exchange load and sum rate performance as we increase $N_{r,2}$. It should also be noted that excess relays in C_{global} leads to a distributed diversity gain and fairness inbetween the links due to adopted maxmin S(I)NR approach [5].

VI. Conclusions

We have proposed a coherent multiuser AF relaying protocol consisting of heterogeneous relay clusters. The corresponding relay gain factor and cluster transmit power allocations have been investigated and the cases of excess and lack of relays are formulated as semidefinite programs. The CSI exchange load, which is the drawback of coherent multiuser relaying, has been drastically reduced with the proposed clusterings. The C_{global} relays cancel the multiuser interference, whereas the C_{local} relays provide array gain to the system. The distributed diversity gain and/or improved sum rate can be achieved through the proposed simple and efficient relay selection for clusters.

References

[1] J. N. Laneman and G. W. Wornell, "Distributed Space-Time-Coded Protocols for Exploiting Cooperative Diversity in Wireless Networks," *IEEE Trans. on Inf. Theory*, vol. 49, no. 10, pp. 2415–2425, Oct. 2003.

[2] A. Sendonaris, E. Erkip, and B. Aazhang, "User Cooperation Diversity - Part I and II," *IEEE Trans. on Commun.*, Nov. 2003.

[3] R. U. Nabar, Ö. Oyman, H. Bölcskei, and A. J. Paulraj, "Capacity scaling laws in MIMO wireless networks," in *Proc. Allerton Conf. on Commun., Control, and Computing, Monticello, IL*, Oct. 2003, pp. 378–389.

[4] A. Wittneben and B. Rankov, "Distributed Antenna Systems and Linear Relaying for Gigabit MIMO Wireless," in *IEEE VTC Fall*, Sep. 2004.

[5] C. Eşli, Stefan Berger, and A. Wittneben, "Optimizing zero-forcing based gain allocation for wireless multiuser networks," *IEEE ICC'07*, accepted.

[6] J. F. Sturm, "Using SeDuMi 1.02, a Matlab toolbox for optimization over symmetric cones," *Opt. Meth. and Soft.*, Vol. 11-12, 1999, pp. 625-653.