Distributed Gain Allocation in Non-Regenerative Multiuser Multihop MIMO Networks

Raphael Rolny, Jörg Wagner, and Armin Wittneben
Communication Technology Laboratory, ETH Zurich
{rolny, wagner, wittneben}@nari.ee.ethz.ch

***Abstract*—In this work, we study distributed gain allocation to maximize the achievable sum-rate in non-regenerative multihop networks with multiple users and multi-antenna nodes. The source and relay nodes are constrained to forward linear transformations of their input and calculate their transformation matrices based on locally available channel state information and limited feedback from the destination. We relax the optimization problem by imposing a norm constraint instead of a power constraint on each node, and devise a gain allocation scheme that is based on an approximation of the gradient of the sum-rate. By means of computer simulations, we show that the proposed algorithm achieves close to optimal solutions, if a good initialization is provided. This makes the scheme particularly useful for channel tracking in slow fading environments.**

I. Introduction

Future cellular networks have to meet a variety of demands. While the node density in such networks is expected to be further increased, high data rates as well as high reliability, low latency, and low cost are required for networks of the next generation. In order to meet these demands, extensive efforts are made to develop new standards that fulfill the requirements on technologies of the third and fourth generation (3G, 4G) and beyond [1]. Due to the continuously growing node densities in cellular networks, the performance of these communication systems is severely affected by interference of other users. Therefore, sophisticated transmission schemes are required to mitigate these impairments and to allow an efficient use of the available resources. Approaches to enhance the performance in interference limited networks include cooperation between different nodes and/or the use of additional relays. This is foreseen in the upcoming standard LTE-Advanced [2], which is the first mobile telecommunication standard that allows for cooperation and the use of relays.

Different research results have shown that multihop communication via relays can be a cost-effective solution to provide ubiquitous access to high data rate services (see [3], [4] and references therein). Particularly in environments that are affected by strong shadowing, the use of relays can enhance performance considerably. Among other relaying strategies, amplify-and-forward (AF) relaying and, in the MIMO case, generalized AF relaying, where each relay is constrained to forward linear transformations of its input, are thereby of special interest. This non-regenerative approach has substantial advantages, even though it is suboptimal in general. AF relays are easy to implement, avoid coding delays, and are fully transparent to the modulation alphabet used by the terminal nodes. This is particularly interesting in heterogeneous networks with several nodes of different complexity. AF relaying has thus attracted much research interest [5], [6].

While promising gains can be expected by the use of AF relays [7], [8], finding optimal gain coefficients requires global channel state information (CSI) in general. This introduces a large overhead, since all nodes have to disseminate their local CSI to all other nodes or to a central unit. In order to reduce this overhead, a distributed gain allocation scheme that allows each node to compute its own transformation matrix based on locally available CSI is desirable. A gain allocation scheme that allows for optimizing the gain coefficients based on local CSI and very limited feedback from the destination has been proposed in [9], [10] for 2-hop networks without source precoding.

In this paper, we consider a more general framework of multihop networks, where an arbitrary number of source-destination (S-D) pairs communicate with each other over the same physical channel, while the transmission is assisted by several stages of relay nodes (cf. Fig. 1). In our setting, all nodes are equipped with an arbitrary number of antennas and, in contrast to [9], [10], also precoding at the source nodes is included. The extension of the previous work to the general setting considered here fails, because the gradient, which is the key of the distributed algorithm, cannot be computed based on locally available CSI if source precoding and/or more than one relay stage are considered. Therefore, we use an approximation of the gradient that can be computed based on local CSI and limited feedback from the destination. Even though the proposed gain allocation scheme is suboptimal in terms of achievable rate, we show by computer simulations that the performance is close to optimal, if a good initialization for the algorithm is provided.

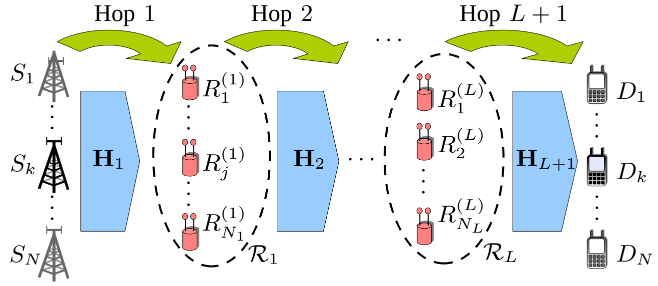

Fig. 1. Structure of the considered multihop network. Destination node D_k is interested in the message sent by source node S_k, while other sources are interfering. The communication is assisted by intermediate stages of relays.

II. Multihop Networks

The multihop networks under consideration consist of N_0 S-D pairs that wish to communicate over the same physical channel. The source nodes as well as the destination nodes are all equipped with an arbitrary number of antennas, and we identify the number of antennas in source node S_k, $k = 1, \ldots, N_0$, with $n_k^{(0)}$ and those of destination node D_k with $n_k^{(L+1)}$. The communication between sources and destinations is assisted by L intermediate relay stages $\mathcal{R}_\ell$, $\ell = 1, \ldots, L$, where stage $\mathcal{R}_\ell$ contains N_ℓ relay nodes. The relays are also equipped with multiple antennas, and we denote the number of antennas in relay $R_j^{(\ell)} \in \mathcal{R}_\ell$ by $n_j^{(\ell)}$.

Transmission is divided into $L+1$ time slots and is initiated by the source nodes which transmit their signals simultaneously in the first time slot. Each source node S_k transmits a symbol vector $\mathbf{s}_k \in \mathbb{C}^{n_k^{(0)}}$ which, in equivalent baseband, is linearly precoded by a transformation matrix $\mathbf{G}_k^{(0)} \in \mathbb{C}^{n_k^{(0)} \times n_k^{(0)}}$, i.e. the transmit signal of S_k can be written as $\mathbf{x}_k^{(0)} = \mathbf{G}_k^{(0)} \cdot \mathbf{s}_k$. In time slot ℓ, the relays in stage $\mathcal{R}_\ell$ receive signals from stage $\mathcal{R}_{\ell-1}$ (or from the source stage denoted by $\mathcal{S}$ if $\ell = 1$). Each relay $R_j^{(\ell)}$ performs then a linear transformation of its receive signal with a complex gain matrix before retransmission in time slot $\ell+1$. The transmit signal of $R_j^{(\ell)}$ can thus be written as

$$\mathbf{x}_j^{(\ell)} = \mathbf{G}_j^{(\ell)} \cdot \left(\mathbf{y}_j^{(\ell)} + \mathbf{v}_j^{(\ell)}\right), \tag{1}$$

where $\mathbf{G}_j^{(\ell)} \in \mathbb{C}^{n_j^{(\ell)} \times n_j^{(\ell)}}$ is the transformation matrix, $\mathbf{y}_j^{(\ell)} \in \mathbb{C}^{n_j^{(\ell)}}$ the receive signal, and $\mathbf{v}_j^{(\ell)} \in \mathbb{C}^{n_j^{(\ell)}}$ is the noise induced in $R_j^{(\ell)}$. In time slot $L+1$, the destination stage, denoted by $\mathcal{D}$, receives the transmission of stage $\mathcal{R}_L$.

We assume a slow and frequency flat fading channel with coefficients that are i.i.d. $\mathcal{CN}(0,1)$, if nodes are located in adjacent stages, and zero otherwise. The channel matrix of the ℓ-th hop is denoted by $\mathbf{H}_\ell$, for $\ell = 1, \ldots, L+1$. The overall effective channel from $\mathcal{S}$ to $\mathcal{D}$ is then given by

$$\mathbf{H}_{\text{eff}} = \mathbf{H}_{L+1} \cdot \mathbf{G}_L \cdot \mathbf{H}_L \cdots \mathbf{H}_2 \cdot \mathbf{G}_1 \cdot \mathbf{H}_1 \cdot \mathbf{G}_0, \tag{2}$$

where $\mathbf{G}_\ell$ is the block-diagonal matrix consisting of all transformation matrices of the ℓ-th stage as its diagonal blocks.

Destination node $D_k \in \mathcal{D}$ is only interested in the signal transmitted by source S_k, while the signals from all other sources are interference. The receive signal of D_k can thus be written as

$$\mathbf{d}_k = \mathbf{H}_{\text{eff},k,k}\mathbf{s}_k + \mathbf{H}_{\text{eff},k,-k}\mathbf{s}_{-k} + \sum_{\ell=1}^{L} \mathbf{H}_{\text{rd},k}^{(\ell)} \mathbf{G}_\ell \mathbf{v}_\ell + \mathbf{w}_k, \tag{3}$$

with

- $\mathbf{H}_{\text{eff},k,k}$ the components of the effective channel $\mathbf{H}_{\text{eff}}$ that correspond to S_k and D_k,
- $\mathbf{H}_{\text{eff},k,-k}$ the components of $\mathbf{H}_{\text{eff}}$ that correspond to all other sources S_i, $j \neq k$, and D_k,
- $\mathbf{s}_k$ the symbol vector of S_k intended for D_k,
- $\mathbf{s}_{-k} = \left[\mathbf{s}_1^T, \ldots, \mathbf{s}_{k-1}^T, \mathbf{s}_{k+1}^T, \ldots, \mathbf{s}_{N_0}^T\right]^T$ the vector of all other symbols,
- $\mathbf{H}_{\text{rd},k}^{(\ell)}$ the components of $\mathbf{H}_{L+1}\mathbf{G}_L\mathbf{H}_L \cdots \mathbf{G}_{\ell+1}\mathbf{H}_{\ell+1}$ to D_k, i.e. the (effective) channel from stage $\mathcal{R}_\ell$ to D_k,
- $\mathbf{v}_\ell$ and $\mathbf{w}_k$ the noise induced in $\mathcal{R}_\ell$ and D_k, respectively, with i.i.d. $\mathcal{CN}(0, \sigma_v^2)$ and $\mathcal{CN}(0, \sigma_w^2)$ elements.

On each source node as well as each relay node, we impose an instantaneous transmit power constraint

$$P_j^{(\ell)} \leq P/N_\ell, \qquad \begin{array}{l} \forall j \in \{1, \ldots, N_\ell\}, \\ \forall \ell \in \{0, 1, \ldots, L\}, \end{array} \tag{4}$$

where $P_j^{(\ell)}$ is the transmit power at $R_j^{(\ell)}$ (or S_j if $\ell = 0$). An achievable sum-rate is then given in terms of the covariance matrices of the desired signal, interference, and noise, denoted by $\mathbf{K}_{\text{s}}^{(k)}$, $\mathbf{K}_{\text{i}}^{(k)}$, and $\mathbf{K}_{\text{n}}^{(k)}$, respectively, as

$$r_\Sigma = \sum_{k=1}^{N_0} \Bigg(\log_2 \det \left\{ \mathbf{K}_{\text{s}}^{(k)} + \mathbf{K}_{\text{i}}^{(k)} + \mathbf{K}_{\text{n}}^{(k)} \right\} - \log_2 \det \left\{ \mathbf{K}_{\text{i}}^{(k)} + \mathbf{K}_{\text{n}}^{(k)} \right\} \Bigg) \quad \left[\frac{\text{bit}}{\text{channel use}} \right]. \tag{5}$$

The covariance matrices for user k are given by

$$\mathbf{K}_{\text{s}}^{(k)} = \left[\mathbf{H}_{\text{eff},k,k} \cdot \mathbf{H}_{\text{eff},k,k}^H \right], \tag{6}$$

$$\mathbf{K}_{\text{i}}^{(k)} = \left[\mathbf{H}_{\text{eff},k,-k} \cdot \mathbf{H}_{\text{eff},k,-k}^H \right], \qquad \text{and} \tag{7}$$

$$\mathbf{K}_{\text{n}}^{(k)} = \left[\sigma_v^2 \cdot \sum_{\ell=1}^{L} \mathbf{H}_{\text{rd},k}^{(\ell)} \mathbf{G}_\ell \cdot \mathbf{G}_\ell^H \mathbf{H}_{\text{rd},k}^{(\ell)H} + \sigma_w^2 \cdot \mathbf{I} \right], \tag{8}$$

where we assumed that $\mathbf{E}[\mathbf{s}_k \mathbf{s}_k^H] = \mathbf{I}$, $\forall k$. Note that we dropped the prelog factor in (5) for notational convenience.[1]

III. Distributed Optimization

In the following, we aim to find transformation matrices for the source and relay nodes that maximize the achievable sum-rate (5). However, finding optimal gain coefficients for $\mathbf{G}_0, \ldots, \mathbf{G}_L$ requires global CSI in general, which is not readily available in practice. Therefore, we wish to distribute the gain allocation scheme, such that each node is able to optimize its gain coefficients based on locally available CSI and limited feedback from the destination. To this end, we use a gradient based optimization algorithm similar to the one used in our previous work [10]. However, the extension of this scheme to the general setting of multihop networks with more than two hops and source precoding fails, because several terms that are required to compute the gradient are not locally available in networks with $L \geq 2$. One of the issues is due to the power constraint (4). The transmit power of relay node $R_j^{(\ell)}$ depends on the power of the received signal, which is a function of the gain coefficients of all preceding stages. This implies that $R_j^{(\ell)}$ has to know the gain coefficients of the stages $\mathcal{R}_i$, $i < \ell$, as well as $\mathcal{S}$. Therefore, the different stages have to be decoupled. To this end, we relax the optimization problem by imposing a norm constraint on each node instead of the power constraint:

$$\left\| \mathbf{G}_j^{(\ell)} \right\|_{\text{F}}^2 = \operatorname{Tr}\left\{ \mathbf{G}_j^{(\ell)} \cdot \mathbf{G}_j^{(\ell)H} \right\} \leq \frac{\rho}{N_\ell}, \qquad \forall j, \ell. \tag{9}$$

[1]Due to the use of multiple time slots for the transmission of one symbol vector, the achievable sum-rate needs to be multiplied by a prelog factor. We drop this factor, since its value is not immediately clear without additional assumptions. Depending on the path loss and shadowing effects, new signals can be injected into the network more often than every $L+1$ time slots without causing interference to previously transmitted signals.

Note that (9) is equivalent to (4) for $\ell = 0$. The optimization problem that we wish to solve can now be described as

$$\max_{\{\mathbf{G}_j^{(\ell)}\}_{j,\ell}} r_\Sigma \quad \text{s.t.} \quad \left\|\mathbf{G}_j^{(\ell)}\right\|_\text{F}^2 \leq \frac{\rho}{N_\ell} \quad \forall j, \ell. \tag{10}$$

In first instance, we turn the constrained optimization problem (10) into an unconstrained one by fulfilling the norm constraint with equality through the variable substitution

$$\mathbf{G}_j^{(\ell)} = \sqrt{\frac{\rho}{N_\ell}} \cdot \tilde{\mathbf{G}}_j^{(\ell)} \Big/ \sqrt{\text{Tr}\left\{\tilde{\mathbf{G}}_j^{(\ell)} \cdot \tilde{\mathbf{G}}_j^{(\ell)H}\right\}}. \tag{11}$$

Note that fulfilling the constraint with equality is not optimal in general. The algorithm described later in this section, however, is adapted for optimization without this additional restriction. In the following, the gradient with respect to the gain coefficients in the transformation matrices is computed.

A. Distributed Gradient Computation

The complex gradient of the sum-rate r_Σ is defined as the vector of the partial derivatives with respect to all gain coefficients of all stages

$$\nabla_{\tilde{\mathbf{G}}^*} r_\Sigma = \left[\frac{\partial r_\Sigma}{\partial \tilde{g}_{1,1}^{(1,0)*}}, \frac{\partial r_\Sigma}{\partial \tilde{g}_{1,2}^{(1,0)*}}, \dots \frac{\partial r_\Sigma}{\partial \tilde{g}_{1,n_1^{(0)}}^{(1,0)*}}, \dots \frac{\partial r_\Sigma}{\partial \tilde{g}_{p,q}^{(j,\ell)*}}, \dots\right]^T,$$

where $\tilde{g}_{p,q}^{(j,\ell)*}$ is the complex conjugate of the coefficient in the p-th row and q-th column of $\tilde{\mathbf{G}}_j^{(\ell)}$. The gradient is obtained by the complex chain rule of differentiation

$$\nabla_{\tilde{\mathbf{G}}^*} r_\Sigma = \left((\nabla_{\mathbf{G}^*} r_\Sigma)^T \cdot \hat{\mathbf{J}} + (\nabla_{\mathbf{G}^*} r_\Sigma)^H \cdot \mathbf{J}\right)^T, \tag{12}$$

with $\mathbf{J}$ and $\hat{\mathbf{J}}$ the Jacobians of the variable substitution (11) and its complex conjugate. The elements of $\nabla_{\mathbf{G}^*} r_\Sigma$ are given by

$$\begin{aligned}\frac{\partial r_\Sigma}{\partial g_{p,q}^{(j,\ell)*}} = \frac{1}{\ln(2)} \sum_{k=1}^{N_0} \Bigg(&\text{Tr}\left\{\left(\mathbf{K}_\text{s}^{(k)} + \mathbf{K}_\text{i}^{(k)} + \mathbf{K}_\text{n}^{(k)}\right)^{-1}\right. \\ &\left. \cdot \left(\frac{\partial \mathbf{K}_\text{s}^{(k)}}{\partial g_{p,q}^{(j,\ell)*}} + \frac{\partial \mathbf{K}_\text{i}^{(k)}}{\partial g_{p,q}^{(j,\ell)*}} + \frac{\partial \mathbf{K}_\text{n}^{(k)}}{\partial g_{p,q}^{(j,\ell)*}}\right)\right\} - \\ &\text{Tr}\left\{\left(\mathbf{K}_\text{i}^{(k)} + \mathbf{K}_\text{n}^{(k)}\right)^{-1} \cdot \left(\frac{\partial \mathbf{K}_\text{i}^{(k)}}{\partial g_{p,q}^{(j,\ell)*}} + \frac{\partial \mathbf{K}_\text{n}^{(k)}}{\partial g_{p,q}^{(j,\ell)*}}\right)\right\}\Bigg). \end{aligned} \tag{13}$$

In order to compute the inner derivatives in (13), we factorize the effective channel (2) as

$$\begin{aligned}\mathbf{H}_\text{eff}^{(\ell)} &= \mathbf{H}_\text{eff}^{(\ell)} \cdot \mathbf{I} = \underbrace{\mathbf{H}_{L+1}\mathbf{G}_L \cdots \mathbf{H}_{\ell+1}}_{\triangleq \mathbf{H}_\text{rd}^{(\ell)}} \cdot \mathbf{G}_\ell \cdot \underbrace{\mathbf{H}_\ell \cdots \mathbf{H}_1 \mathbf{G}_0 \mathbf{I}}_{\triangleq \mathbf{H}_\text{sr}^{(\ell)}} \\ &= \mathbf{H}_\text{rd}^{(\ell)} \cdot \mathbf{G}_\ell \cdot \mathbf{H}_\text{sr}^{(\ell)}, \qquad \ell = 0, 1, \dots, L, \end{aligned} \tag{14}$$

and obtain $L+1$ different notations for $\mathbf{H}_\text{eff}$, each with respect to the gain matrix of a particular stage. We further define $\mathbf{H}_{\text{rd},k}^{(\ell)}$ as the components of $\mathbf{H}_\text{rd}^{(\ell)}$ that correspond to destination D_k, $\mathbf{H}_{\text{sr},k}^{(\ell)}$ as the components of $\mathbf{H}_\text{sr}^{(\ell)}$ that correspond to source S_k, and $\mathbf{H}_{\text{sr},-k}^{(\ell)}$ as the components of $\mathbf{H}_\text{rd}^{(\ell)}$ that corresponds to all other source nodes S_j, $j \neq k$. Then, we can write the derivatives of the covariance matrices in (13) as

$$\frac{\partial \mathbf{K}_\text{s}^{(k)}}{\partial g_{p,q}^{(j,\ell)*}} = \left[\mathbf{H}_{\text{eff},k,k} \cdot \mathbf{H}_{\text{sr},k}^{(\ell)H} \mathbf{E}_{p,q}^{(j,\ell)H} \mathbf{H}_{\text{rd},k}^{(\ell)H}\right] \tag{15}$$

$$\frac{\partial \mathbf{K}_\text{i}^{(k)}}{\partial g_{p,q}^{(j,\ell)*}} = \left[\mathbf{H}_{\text{eff},k,-k} \cdot \mathbf{H}_{\text{sr},-k}^{(\ell)H} \mathbf{E}_{p,q}^{(j,\ell)H} \mathbf{H}_{\text{rd},k}^{(\ell)H}\right] \tag{16}$$

$$\begin{aligned}\frac{\partial \mathbf{K}_\text{n}^{(k)}}{\partial g_{p,q}^{(j,\ell)*}} = \sigma_v^2 \cdot \Bigg(&\mathbf{H}_{\text{rd},k}^{(\ell)} \mathbf{G}_\ell \cdot \mathbf{E}_{p,q}^{(j,\ell)H} \mathbf{H}_{\text{rd},k}^{(\ell)H} + \\ &\sum_{i=1}^{\ell-1} \mathbf{H}_{\text{rd},k}^{(i)} \mathbf{G}_i \cdot \mathbf{G}_i^H \cdot \frac{\partial \mathbf{H}_{\text{rd},k}^{(\ell)H}}{\partial g_{p,q}^{(j,\ell)*}}\Bigg), \end{aligned} \tag{17}$$

where $\mathbf{E}_{p,q}^{(j,\ell)}$ is the single-entry matrix of the same size as $\mathbf{G}_j^{(\ell)}$ with all entries zero except the entry in the p-th row and q-th column, which is one. The terms $\partial \mathbf{H}_{\text{rd},k}^{(\ell)H} \big/ \partial g_{p,q}^{(j,\ell)*}$ are given by $\mathbf{H}_{i+1}^H \mathbf{G}_{i+1}^H \cdots \mathbf{H}_\ell^H \cdot \mathbf{E}_{p,q}^{(j,\ell)H} \cdot \mathbf{H}_{\ell+1}^H \cdots \mathbf{G}_L \mathbf{H}_{L+1}^H$, if $j < \ell$, and $\mathbf{O}$ otherwise.

The Jacobians $\mathbf{J}$ and $\hat{\mathbf{J}}$ contain the partial derivatives of (11) and of its complex conjugate, both with respect to the entries of $\tilde{\mathbf{G}}_j^{(\ell)}$, $\forall \ell \in \{0, \dots, L\}$ and $j \in \{1, \dots, N_\ell\}$. The entries of $\mathbf{J}$ are given by

$$\frac{\partial g_{p,q}^{(j,\ell)}}{\partial \tilde{g}_{p',q'}^{(j',\ell')*}} = \frac{-\frac{1}{2}\sqrt{\frac{\rho}{N_\ell}} \tilde{g}_{p,q}^{(j,\ell)} \left\|\tilde{\mathbf{G}}_j^{(\ell)}\right\|_\text{F}^{-\frac{1}{2}} \text{Tr}\left\{\tilde{\mathbf{G}}_j^{(\ell)} \mathbf{E}_{p',q'}^{(j',\ell')H}\right\}}{\left\|\tilde{\mathbf{G}}_j^{(\ell)}\right\|_\text{F}^2}$$

and those of $\hat{\mathbf{J}}$ by

$$\frac{\partial g_{p,q}^{(j,\ell)*}}{\partial \tilde{g}_{p',q'}^{(j',\ell')*}} = \frac{\xi - \frac{1}{2}\sqrt{\frac{\rho}{N_\ell}} \tilde{g}_{p,q}^{(j,\ell)*} \left\|\tilde{\mathbf{G}}_j^{(\ell)}\right\|_\text{F}^{-\frac{1}{2}} \text{Tr}\left\{\tilde{\mathbf{G}}_j^{(\ell)} \mathbf{E}_{p',q'}^{(j',\ell')H}\right\}}{\left\|\tilde{\mathbf{G}}_j^{(\ell)}\right\|_\text{F}^2},$$

with

$$\xi = \begin{cases} \sqrt{\frac{\rho}{N_\ell}} \cdot \left\|\tilde{\mathbf{G}}_j^{(\ell)}\right\|_\text{F}, & \text{if } j = j', \ell = \ell', p = p', q = q' \\ 0 & \text{otherwise.} \end{cases}$$

Note that the Jacobians are block diagonal matrices, since the variable substitution is applied to each node separately. The overall gradient can, in this case, be decomposed into components that correspond to one relay or source node. These components are given by

$$(\nabla_{\tilde{\mathbf{G}}^*} r_\Sigma)_j^{(\ell)T} = \left((\nabla_{\mathbf{G}^*} r_\Sigma)_j^{(\ell)T} \cdot \hat{\mathbf{J}}_j^{(\ell)} + (\nabla_{\mathbf{G}^*} r_\Sigma)_j^{(\ell)H} \cdot \mathbf{J}_j^{(\ell)}\right)^T,$$

where $(\nabla_{\mathbf{G}^*} r_\Sigma)_j^{(\ell)}$, $\hat{\mathbf{J}}_j^{(\ell)}$, and $\mathbf{J}_j^{(\ell)}$ are the components of the "outer" gradient (13), and the Jacobians, respectively, that correspond to $R_j^{(\ell)}$ (or S_j if $\ell = 0$). Each of these components shall now be computed in each node separately with local CSI and limited feedback from the destination. In the following, we refer to local CSI at $R_j^{(\ell)}$ as the channel coefficients that can be estimated locally at $R_j^{(\ell)}$, i.e. the channel coefficients from $\mathcal{S}$ to $R_j^{(\ell)}$ and the channel coefficients from $R_j^{(\ell)}$ to $\mathcal{D}$.

While it is immediately clear that each node can compute its corresponding blocks of the Jacobians (they depend only on the own transformation matrix), the computation of the terms of $(\nabla_{\mathbf{G}^*} r_\Sigma)_j^{(\ell)}$ requires some feedback from $\mathcal{D}$. In order to compute (13), the effective channel $\mathbf{H}_\text{eff}$ as well as the noise covariance matrices $\mathbf{K}_\text{n}^{(k)}$ need do be known. These can be estimated at $\mathcal{D}$ and fed back to the other nodes. In order to compute (15) and (16), local CSI and the knowledge of $\mathbf{H}_{\text{eff}}$ is sufficient, since

$$\mathbf{H}_{\text{sr},k}^{(\ell)H}\mathbf{E}_{p,q}^{(j,\ell)H}\mathbf{H}_{\text{rd},k}^{(\ell)H} = \left(\mathbf{H}_{\text{rd},k}^{(\ell)}[:,p]\cdot\mathbf{H}_{\text{sr},k}^{(\ell)}[q,:]\right)^{H}$$
$$\mathbf{H}_{\text{sr},-k}^{(\ell)H}\mathbf{E}_{p,q}^{(j,\ell)H}\mathbf{H}_{\text{rd},k}^{(\ell)H} = \left(\mathbf{H}_{\text{rd},k}^{(\ell)}[:,p]\cdot\mathbf{H}_{\text{sr},-k}^{(\ell)}[q,:]\right)^{H},$$

where $\mathbf{H}_{\text{rd},k}^{(\ell)}[:,p]$ is the p-th column of $\mathbf{H}_{\text{rd},k}^{(\ell)}$ and $\mathbf{H}_{\text{sr},k}^{(\ell)}[q,:]$ and $\mathbf{H}_{\text{sr},-k}^{(\ell)}[q,:]$ are the q-th row of $\mathbf{H}_{\text{sr},k}^{(\ell)}$ and $\mathbf{H}_{\text{sr},-k}^{(\ell)}$ which can all be estimated locally at $R_j^{(\ell)}$ (or at S_j if $\ell = 0$). For the computation of (17), however, not all terms are locally accessible. We therefore approximate the derivative of the noise covariance matrix (17) as

$$\frac{\partial \mathbf{K}_{\text{n}}^{(k)}}{\partial g_{p,q}^{(j,\ell)*}} \approx \sigma_v^2 \cdot \mathbf{H}_{\text{rd},k}^{(\ell)}\mathbf{G}_\ell \cdot \mathbf{E}_{p,q}^{(j,\ell)H}\mathbf{H}_{\text{rd},k}^{(\ell)H}, \tag{18}$$

for which local CSI is sufficient, since $\mathbf{G}_\ell$ is block-diagonal:

$$\mathbf{H}_{\text{rd},k}^{(\ell)}\mathbf{G}_\ell\mathbf{E}_{p,q}^{(j,\ell)H}\mathbf{H}_{\text{rd},k}^{(\ell)H} = \left(\mathbf{H}_{\text{rd},k}^{(\ell)}\mathbf{E}_{p,q}^{(j\ell)}\mathbf{G}_\ell^H\mathbf{H}_{\text{rd},k}^{(\ell)H}\right)^H$$
$$= \left(\mathbf{H}_{\text{rd},k}^{(\ell)}\mathbf{E}_{p,q}^{(j\ell)} \underbrace{\begin{bmatrix} \mathbf{G}_1^{(\ell)H}\left(\mathbf{H}_{\text{rd},k}[:,\mathcal{K}_1^{(\ell)}]\right)^H \\ \vdots \\ \mathbf{G}_j^{(\ell)H}\left(\mathbf{H}_{\text{rd},k}[:,\mathcal{K}_j^{(\ell)}]\right)^H \\ \vdots \end{bmatrix}}_{\triangleq \mathbf{X}}\right)^H$$
$$= \left(\mathbf{H}_{\text{rd},k}^{(\ell)}[:,p]\cdot\mathbf{X}[q,:]\right)^H,$$

with $\mathcal{K}_j^{(\ell)}$ the index set corresponding to the antennas of $R_j^{(\ell)}$. The single entry matrix $\mathbf{E}_{p,q}^{(j,\ell)}$ selects a single row of $\mathbf{X}$ that contains only channel coefficients that are locally available at $R_j^{(\ell)}$. With the approximation (18), each relay considers only its "own" noise (and the destination noise) for the computation of the derivative (17), while the noise terms induced in the other stages are neglected.

B. Optimization Algorithm

In order to compute the gain coefficients in each node, we apply a gradient based optimization algorithm that updates the transformation matrices according to

$$\tilde{\mathbf{G}}_j^{(\ell)}[m+1] = \mathbf{G}_j^{(\ell)}[m] + \mu[m]\cdot\Delta_j^{(\ell)}[m], \quad \forall j,\ell,$$

where $\mu[m]$ is the step-size and $\Delta_j^{(\ell)}$ is the search direction of the j-th node in the ℓ-th stage in iteration m. If the norm constraint is enforced with equality, the search direction can be chosen as

$$\Delta_j^{(\ell)} = \left(\nabla_{\tilde{\mathbf{G}}^*} r_\Sigma\right)_j^{(\ell)}.$$

But as already mentioned, enforcing the constraint with equality is not optimal in general. A relay node that receives a weak signal due to small channel coefficients from the previous stage, retransmits mainly noise and little desired signal. Forcing this relay to transmit with high power can, in this case, possibly reduce the achievable rate. In order to drop this additional restriction, the variable substitution (11) can be modified to

$$\mathbf{G}_j^{(\ell)} = \begin{cases} \sqrt{\frac{\rho}{N_\ell}}\cdot\tilde{\mathbf{G}}_j^{(\ell)}/\left\|\tilde{\mathbf{G}}_j^{(\ell)}\right\|_{\text{F}}, & \text{if } \left\|\tilde{\mathbf{G}}_j^{(\ell)}\right\|_{\text{F}} > \sqrt{\frac{\rho}{N_\ell}} \\ \tilde{\mathbf{G}}_j^{(\ell)}, & \text{if } \left\|\tilde{\mathbf{G}}_j^{(\ell)}\right\|_{\text{F}} \le \sqrt{\frac{\rho}{N_\ell}}, \end{cases}$$

where the transformation matrix is only normalized, if the constraint is violated. If a specific $\tilde{\mathbf{G}}_j^{(\ell)}$ does not violate the constraint, $\mathbf{G}_j^{(\ell)} = \tilde{\mathbf{G}}_j^{(\ell)}$ and the search direction can be chosen as

$$\Delta_j^{(\ell)} = \left(\nabla_{\mathbf{G}^*} r_\Sigma\right)_j^{(\ell)},$$

i.e., only the "outer" gradient is considered and the Jacobians are dropped. In the other case, i.e. if $\mathbf{G}_j^{(\ell)}$ is a normalized version of $\tilde{\mathbf{G}}_j^{(\ell)}$, two different search directions have to be considered:

- If an update with $\Delta_j^{(\ell)} = (\nabla_{\mathbf{G}^*} r_\Sigma)_j^{(\ell)}$ with a small step-size μ_{min} leads to a transformation matrix that does *not* violate the constraint, this search direction is used,
- if this update leads to a transformation matrix that *does* violate the constraint, the norm constraint is enforced with equality and the total gradient including the Jacobians has to be considered, i.e., the search direction is in this case given by $\Delta_j^{(\ell)} = (\nabla_{\tilde{\mathbf{G}}^*} r_\Sigma)_j^{(\ell)}$.

After the search direction for iteration step m is obtained, we optimize the step-size $\mu[m]$ by a line search method that is designed to achieve low complexity. To this end, we start with some small initial step-size $\mu_0 = \mu_{\text{min}}$ and increase the step-size according to $\mu_{t+1} = \alpha\cdot\mu_t$, for some $\alpha > 1$, as long as the sum-rate with instantaneous choice of μ_{t+1} is larger as the sum-rate with the previous choice. As soon as μ_{t+1} leads to a smaller sum-rate than with the previous step-size, the line search terminates and $\mu[m] = \mu_t$ is chosen for the step-size of iteration step m.

In order to achieve a faster convergence behavior than with the simple gradient search method outlined above, we can refine the algorithm by utilizing conjugate gradients (CG). In this case, however, inner products of the current and previous search direction need to be computed. These can be obtained by *over-the-air additions* (cf. [9], [10]).

C. Overhead

The knowledge of $\mathbf{H}_{\text{eff}}$ as well as $\mathbf{K}_{\text{n}}^{(k)}$ is required by all source and relay nodes for the computation of the search directions. Estimates of these matrices can be obtained at $\mathcal{D}$ by the use of orthogonal pilot sequences transmitted simultaneously by the sources. The matrices can then be fed back through the network to the relay and source nodes. However, the dimensions the these matrices are not dependent on the number of relays or antennas therein and remain fixed, even if N_ℓ grows large. The pilots transmitted by $\mathcal{S}$ and additional sequences transmitted simultaneously by $\mathcal{D}$ in the feedback cycle can be used for estimates of the local CSI in the relays. The overhead due to such an estimation and feedback cycle does thus not depend on the relays per stage. More relay stages, however, imply a larger overhead, since all signals have to traverse the entire network.

An additional overhead is introduced by the line search as well as for the computation of the inner products for the CG algorithm, if this refinement is used. But also this overhead does not scale with the number of relays per stage. Moreover, numerical results show that also the number of iterations required to to achieve close to optimal sum-rates does not scale at all, or if so, very slowly with the number of relays per stage. Accordingly, our gain allocation scheme proves to be particularly useful for networks with a large number of relays that are grouped into few stages.

IV. Simulation Results & Discussion

We assess the performance of the proposed gain allocation scheme by means of computer simulations. We choose networks with $L = N_0 = N_1 = N_2 = 2$ and two antennas in each node. In the simulations, we set the noise variances to $\sigma_v^2 = \sigma_w^2 = 1$ and compare the results of the proposed scheme with the achievable sum-rates obtained by an optimization algorithm that has access to global CSI and optimizes with respect to the initially imposed power constraint. For a fair comparison, we scale the resulting transformation matrices after optimization with the norm constraint such that the power constraint is fulfilled. Two different scaling methods are used. In the first method, each node scales its gain coefficients individually, if the power constraint is violated. In the second method, the scaling is performed stage-wise, where all transformation matrices of the same stage are normalized by the same factor. Note that this scaling method requires an additional over-the-air addition cycle.

Since the optimization at hand is highly non-convex, the optimization algorithm can converge to different local optima, depending on the initialization. While some of the solutions achieve full multiplexing gain, other initializations can converge to local optima that do not suppress the interference of all users and hence achieve only a fraction of the achievable multiplexing gain. Starting the gain allocation scheme with a "good" initial guess is therefore crucial for achieving good performance. The proposed algorithm is therefore particularly useful for channel tracking in slow fading environments, where a global algorithm initiates the communication between $\mathcal{S}$ and $\mathcal{D}$ and the distributed scheme tracks the changes of the channel.

In Fig. 2, the empirical cumulative distribution function (CDF) of the sum-rates are plotted for different values of SNR $= P/N_\ell, \forall \ell$, where the distributed algorithm is initialized with the best solution of the global optimizer. In this case, we observe that the distributed scheme achieves close to optimal solutions for all SNR values, if the scaling according to the power constraint is performed per stage. For lower SNR's however, the per node scaling achieves better results.

In Fig. 3, we show the tracking performance of the distributed algorithm for SNR $= 20$ dB with a time varying channel $\mathbf{H}_\ell[m+1] = \vartheta \mathbf{H}_\ell[m] + \sqrt{1-\vartheta^2}\mathbf{H}_{\text{i.i.d.}}$, $\vartheta \in [0,1]$, that is fixed until the optimization has terminated and changes by an additive i.i.d. component afterwards. The parameter ϑ is a measure for the correlation of the channel compared to the previous realization. In each optimization (for each channel realization), the distributed algorithm is initialized

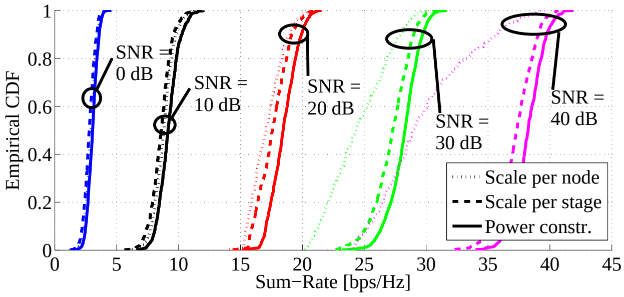

Fig. 2. CDF of achievable sum-rates after optimization for different SNR values. For each channel realization, only the best solution is considered.

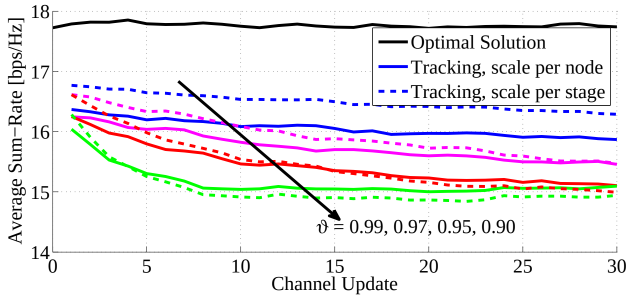

Fig. 3. Average sum-rate after channel tracking in a slow fading environment.

with the preceding solution (or with the global optimum at the beginning). If the channel increment is small ($\vartheta = 0.99$), the distributed scheme shows similarly good performance results as before. When the correlation of the actual channel realization with the initial one decreases (due to a smaller ϑ or increasing m), the performance decreases to a constant value. However, we can conclude that the proposed algorithm is an efficient scheme that achieves close to optimal solutions, if the channel changes sufficiently slowly, while the introduced overhead is reduced considerably with respect to a global optimization scheme.

References

[1] ITU-R, "Requirements related to technical performance for IMT-Advanced radio interface(s)," ITU, Tech. Rep. ITU-R M.2134, Nov. 2008.

[2] 3GPP, Technical Specification Group Radio Access Network, "Further advancements for E-UTRA: Physical layer aspects (Release 9)," Tech. Rep. 3GPP TR 36.814 V0.4.1 (2009-02), Feb. 2009.

[3] M. Cao, X. Wang, S.-J. Kim, and M. Madihian, "Multi-hop wireless backhaul networks: A cross-layer design paradigm," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 4, pp. 738 –748, May 2007.

[4] L. Le and E. Hossain, "Multihop cellular networks: Potential gains, research challenges, and a resource allocation framework," *IEEE Communications Magazine*, vol. 45, no. 9, pp. 66 –73, 2007.

[5] Y. Rong and Y. Hua, "Optimality of diagonalization of multi-hop MIMO relays," *IEEE Trans. Wireless Commun.*, vol. 8, no. 12, pp. 6068 –6077, 2009.

[6] S.-W. Jeon, S.-Y. Chung, and S. A. Jafar, "Degrees of freedom of multi-source relay networks," in *Proc. 47th Annual Allerton Conf.*, Sept. 2009.

[7] S. Berger and A. Wittneben, "Cooperative distributed multiuser MMSE relaying in wireless ad-hoc networks," in *Asilomar Conference on Signals, Systems, and Computers 2005*, Pacific Grove, CA, Nov. 2005.

[8] O. Munoz, J. Vidal, and A. Agustin, "Non-regenerative MIMO relaying with channel state information," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, March 2005, pp. 361–364.

[9] C. Esli, J. Wagner, and A. Wittneben, "Distributed gradient based gain allocation for coherent multiuser AF relaying networks," *ICC*, 2009.

[10] R. Rolny, J. Wagner, C. Esli, and A. Wittneben, "Distributed gain matrix optimization in non-regenerative MIMO relay networks," in *Asilomar Conf. on Signals, Systems, Computers*, Pacific Grove, CA, Nov. 2009.