This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE ICC 2009 proceedings

Distributed Gradient Based Gain Allocation for Coherent Multiuser AF Relaying Networks

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

***Abstract*—A set of distributed non-cooperating relay antennas is known to be capable of orthogonalizing multiple source-destination pairs in space by coherent amplify-and-forward (AF) relaying techniques. Known relay gain allocation schemes for this setup require either global knowledge of the channel coefficients between all participating nodes or an excess number of relays. The required dissemination of the channel state information (CSI) introduces a major overhead, which in practice may well diminish the spatial multiplexing gain. We introduce a new distributed gradient based gain allocation scheme, which minimizes this overhead. Key to this is the proof, that the gradient of the destination signal-to-interference-plus-noise ratio (SINR) can be calculated in a distributed manner based on local CSI at the relays and very limited feedback from the destinations. In order to minimize the number of iterations in the distributed gradient algorithm we propose two distributed approaches to determine the optimal step size for each iteration. Finally we provide simulation results on the sum rate performance and identify the sequential opening of the spatial channels as a key factor that impacts speed of convergence.**

I. Introduction

The employment of multiple antennas has been identified as the key enabler for high spectral efficiency, since it facilitates multiplexing of several data streams in space rather than in time/frequency. In the meanwhile, there are methods known, e.g., [1], that enables spatial multiplexing (SM), even if both the transmit and the receive antennas are distributed, i.e. not connected to a central encoder and decoder, respectively.

A particularly simple (from a coding perspective) method for realizing distributed SM is based on coherent amplify-and-forward (AF) relaying. The approach has been discussed in [2] and requires distributed relay antennas to assist the communication between terminal nodes without private benefit. The method mimics classical point-to-point MIMO transmit beamforming in a distributed fashion, where the role of the spatial precoder is played by the distributed relay antennas. These impose specific amplifications on their received signal, such that individual pairs of terminals can communicate over interference free effective channels. In the sequel, we refer to this approach as *coherent multiuser relaying*.

For simultaneously multiplexing N single-antenna terminal pairs, the technique described above requires $N^2 - N + 1$ distributed relay antennas [2]. While this requirement appears to be daunting at first glance, it is put into perspective by the remarkably simple structure and low power consumption of the AF relay nodes. In a frequency division duplex implementation each relay just performs an amplification and a frequency translation of the received signal, which can be fully implemented in the analog domain, i.e., they do not require analog-to-digital conversion and digital baseband processing. An equivalent time division duplex (TDD) implementation involves a store and forward (delay) operation, which is typically implemented in digital baseband (RAM). Nevertheless also for TDD, the relays have very low complexity. In essence, the relay nodes in the network have a significantly lower complexity compared to the terminals. This fact allows for the implementation of heterogeneous networks, where a set of terminals coexists with a far larger set of low-cost relay nodes.

The most significant overhead coming with relaying based SM is the requirement for global channel state information (CSI) at each relay. That is to say, the calculation of the relay gain factors requires knowledge of the channel coefficients from all terminal nodes to all relays. Based on the training sequences in the transmit packets of the terminal nodes, each relay can estimate the subset of channel coefficients from all terminals to itself, i.e. its local CSI, without overhead. Whereas, for the gain factor calculation, this local CSI has to be disseminated to all other relays or to a central node. The overhead involved may become prohibitive, in particular if the number of relays is large and/or the channel is fast fading.

The paper at hand addresses this *global CSI dissemination problem*. Specifically, we devise a gradient based method which allows for establishing the spatial coding (complex relay gains) based on a substantially reduced CSI dissemination overhead. In particular, our method renders the amount of CSI to be disseminated over the network *independent* of the number of relays as opposed to a linear scaling in the conventional approach. Our approach is based on the key insight that the derivative of the signal-to-interference-plus-noise ratio (SINR) for a given terminal pair with respect to a specific relay gain coefficient can be determined on the basis of local CSI, if the terminal nodes provide some very limited feedback, which is now independent of the number of relays.

We develop our scheme starting out with the maximization of the SINR for a single source-destination (S-D) pair in the presence of interfering sources. In a second step, we generalize our results to multiple S-D pairs and distributed spatial multiplexing, and consider to maximize the sum rate. It is well-known for gradient based schemes that fast convergence requires a careful choice of step-size and initialization. We derive two protocols for the adaptive distributed calculation of the optimal step-size. The centralized step-size selection protocol is most efficient during the initial phase of the gradient algorithm (acquisition), whereas the decentralized protocol is preferable in the tracking phase. Finally, we provide an experimental analysis of the convergence speed, and elaborate on the impact of initialization on convergence behaviour.

Related Work: The general setting of this paper has been studied in several related publications with different focus. Reference [3] provides the sum-capacity scaling of the network in both the number of S-D pairs and the number of relays. A generalization of the multiuser relaying technique to multiple-antenna nodes has been provided in [4]. References [5], [6] discuss various relay gain allocation strategies aiming to maximize specific achievable rate related performance metrics.

II. System and Signal Model

We consider a wireless network with $2N + N_\mathrm{r}$ single-antenna nodes, where N S-D pairs communicate concurrently over the same physical channel, and N_r AF relay nodes assist the communication in a half-duplex scheme. The relays are non-cooperating, i.e., they do not share their received signals. Motivated by a scenario of range extension through relays, we assume that there is no direct link between S-D pairs.[1]

The matrices describing the frequency flat fading channels between the N sources and all relays, and between all relays and the N destinations are denoted by $\mathbf{H}_\mathrm{sr} \in \mathbb{C}^{N_\mathrm{r}\times N}$ and $\mathbf{H}_\mathrm{rd} \in \mathbb{C}^{N\times N_\mathrm{r}}$. The relays amplify-and-forward their received signals depending on the realization of the two channel matrices. The respective complex multiplications in equivalent baseband are described by an $N_\mathrm{r} \times N_\mathrm{r}$ diagonal matrix $\mathbf{G}$ with the respective gain coefficients on its diagonal. The effective input-output relation from sources to destinations is then written in terms of the spatial transmit vector $\mathbf{s}$, the relay noise vector $\mathbf{v}$ and the destination noise vector $\mathbf{m}$ as [5]

$$\mathbf{d} = \mathbf{H}_\mathrm{rd}\mathbf{G}\mathbf{H}_\mathrm{sr}\mathbf{s} + \mathbf{H}_\mathrm{rd}\mathbf{G}\mathbf{v} + \mathbf{m} = \mathbf{H}_\mathrm{sd}\mathbf{s} + \mathbf{H}_\mathrm{rd}\mathbf{G}\mathbf{v} + \mathbf{m}, \quad (1)$$

where $\mathbf{H}_\mathrm{sd}$ is the equivalent two-hop channel matrix. The ith element in a vector corresponds to the ith node in the source, relay or destination stage. The elements of $\mathbf{v}$ and $\mathbf{m}$ are assumed to be independently and identically distributed (i.i.d.) zero-mean circularly symmetric complex Gaussian random variables of variance σ_v^2 and σ_m^2, respectively. For simulations, we assume that the elements of $\mathbf{H}_\mathrm{sr}$ and $\mathbf{H}_\mathrm{rd}$ are i.i.d. zero-mean complex Gaussian variables with unit variance. Furthermore, considering a slow-fading environment, the channel matrices are assumed to stay constant within a transmission cycle duration, and succeeding realizations of the propagation channels are statistically independent. The transmit symbols of the source nodes are assumed to have equal average power P_s. We impose the instantaneous sum transmit power constraint

$$\mathbf{g}^H(P_\mathrm{s}\mathbf{H}_\mathrm{sr}\mathbf{H}_\mathrm{sr}^H \odot \mathbf{I}_{N_\mathrm{r}} + \sigma_v^2\mathbf{I}_{N_\mathrm{r}})\mathbf{g} \;=\; \mathbf{g}^H\mathbf{M}\mathbf{g} \;\leq\; P_\mathrm{r} \quad (2)$$

on the relays, where $\mathbf{G} = \mathrm{diag}\{\mathbf{g}\}$. Moreover, we assume that the relays are phase- and frequency-synchronous. We next define the respective channel vectors from the ith source to all relays $\hat{\mathbf{h}}_{\mathrm{sr},i} \in \mathbb{C}^{N_\mathrm{r}} \triangleq \mathbf{H}_\mathrm{sr}\mathbf{e}_i$; from all sources to the kth relay $\mathbf{h}_{\mathrm{sr},k} \in \mathbb{C}^{N} \triangleq \mathbf{H}_\mathrm{sr}^T\mathbf{e}_k$; from the kth relay to all destinations $\hat{\mathbf{h}}_{\mathrm{rd},k} \in \mathbb{C}^{N} \triangleq \mathbf{H}_\mathrm{rd}\mathbf{e}_k$; and from all relays to the ith destination $\mathbf{h}_{\mathrm{rd},i} \in \mathbb{C}^{N_\mathrm{r}} \triangleq \mathbf{H}_\mathrm{rd}^T\mathbf{e}_i$. Moreover, $\mathbf{h}_{\mathrm{sd},i} \in \mathbb{C}^{N} \triangleq \mathbf{H}_\mathrm{sd}^T\mathbf{e}_i$ denotes the two-hop equivalent channel for the ith S-D pair.

In the sequel, we distinguish between global and local CSI knowledge. The former refers to perfect knowledge of both $\mathbf{H}_\mathrm{sr}$ and $\mathbf{H}_\mathrm{rd}$ at all relays, whereas the latter implies that relay k knows only $\mathbf{h}_{\mathrm{sr},k}$ and $\hat{\mathbf{h}}_{\mathrm{rd},k}$. Note that in order to mitigate interference, the conventional coherent multiuser relaying schemes [2], [4]–[6] require that global CSI is available either at each relay or alternatively at a master node which computes the corresponding gain factors and distributes them to the relays.

III. Distributed Gradient Based Gain Allocation

We introduce a gradient descent based relay gain allocation scheme, where the relays determine their gain factors iteratively in a distributed manner. While our proposal performs exactly same as the global CSI requiring solution, only local CSI plus limited feedback from the destinations are employed. In the following, we focus on the case of a single S-D pair under interference as an interim step towards the multiuser case. Afterwards, we extend these findings to simultaneous communication of multiple S-D pairs.

Assuming that the ith S-D pair is the focus of interest in the presence of $N-1$ independent interferers, the received signal at the ith destination is written as

$$d_i = \underbrace{\mathbf{g}^H\mathbf{D}_{\mathrm{rd},i}\hat{\mathbf{h}}_{\mathrm{sr},i}s_i}_{\text{intended signal}} + \underbrace{\textstyle\sum_{j\neq i}\mathbf{g}^H\mathbf{D}_{\mathrm{rd},i}\hat{\mathbf{h}}_{\mathrm{sr},j}\,s_j}_{\text{interference}} + \underbrace{\mathbf{g}^H\mathbf{D}_{\mathrm{rd},i}\mathbf{v} + m_i}_{\text{noise}},$$

where $\mathbf{D}_{\mathrm{rd},i} \triangleq \mathrm{diag}\{\mathbf{h}_{\mathrm{rd},i}\}$, and $i,j = 1,\ldots,N$. We define a normalized gain vector $\mathbf{w} \triangleq \mathbf{D}_{\mathrm{rd},i}^H\mathbf{g}$, and set $P_\mathrm{s} = 1$. Consequently, the SINR at the ith destination is given by

$$\mathrm{SINR}_i = \frac{\mathbf{w}^H\hat{\mathbf{h}}_{\mathrm{sr},i}\hat{\mathbf{h}}_{\mathrm{sr},i}^H\mathbf{w}}{\mathbf{w}^H\left(\mathbf{H}_{\mathrm{I},i}\mathbf{H}_{\mathrm{I},i}^H + \mathbf{I}_{N_\mathrm{r}}\sigma_v^2\right)\mathbf{w} + \sigma_m^2}, \quad (3)$$

where $\mathbf{H}_{\mathrm{I},i} \triangleq [\hat{\mathbf{h}}_{\mathrm{sr},1}\cdots\hat{\mathbf{h}}_{\mathrm{sr},i-1}\hat{\mathbf{h}}_{\mathrm{sr},i+1}\cdots\hat{\mathbf{h}}_{\mathrm{sr},N}]$. We multiply the left- and right-hand sides of the power constraint (2) with σ_m^2, and express the resultant expression in terms of $\mathbf{w}$ as

$$(\sigma_m^2/P_\mathrm{r})\cdot\mathbf{g}^H\mathbf{M}\mathbf{g} = (\sigma_m^2/P_\mathrm{r})\cdot\mathbf{w}^H\mathbf{D}_{\mathrm{rd},i}^{-1}\mathbf{M}(\mathbf{D}_{\mathrm{rd},i}^{-1})^H\mathbf{w} \leq \sigma_m^2. \quad (4)$$

Assuming that the power constraint is fulfilled with equality, we substitute the corresponding definition for σ_m^2 from (4) into (3) and obtain an equivalent SINR_i formulation

$$\mathrm{SINR}_i = \frac{\mathbf{w}^H\mathbf{X}_i\mathbf{w}}{\mathbf{w}^H\mathbf{Z}_i\mathbf{w}}, \quad (5)$$

where $\mathbf{Z}_i = \mathbf{H}_{\mathrm{I},i}\mathbf{H}_{\mathrm{I},i}^H + \mathbf{I}_{N_\mathrm{r}}\sigma_v^2 + (\sigma_m^2/P_\mathrm{r})\cdot\mathbf{D}_{\mathrm{rd},i}^{-1}\mathbf{M}(\mathbf{D}_{\mathrm{rd},i}^{-1})^H$, and $\mathbf{X}_i = \hat{\mathbf{h}}_{\mathrm{sr},i}\hat{\mathbf{h}}_{\mathrm{sr},i}^H$. The SINR_i expression in (5) is known as the *generalized Rayleigh quotient* (GRQ). The maximum value of the GRQ, i.e., the maximum achievable SINR, and the corresponding optimal relay gain vector $\mathbf{w}^\star$ can be found through an eigenvalue problem incorporating $\mathbf{X}_i$ and $\mathbf{Z}_i$. In other words, having global CSI knowledge related with the ith S-D pair, i.e., $\mathbf{H}_\mathrm{sr}$ and $\mathbf{h}_{\mathrm{rd},i}$, each relay can compute the maximum SINR_i resulting relay gain efficiently. However, collecting this global knowledge requires the aforementioned

[1] *Notation:* We use boldface lowercase and capital letters to indicate vectors and matrices, respectively. The superscripts $(\cdot)^*$, $(\cdot)^T$, $(\cdot)^H$ stand for complex conjugate, matrix transpose, and complex conjugate transpose, respectively. The operators $\odot$, $\mathrm{E}\{\cdot\}$, $\mathrm{diag}\{\mathbf{x}\}$, $\mathbf{x}[i]$, $\mathbf{X}[i,i]$, $\mathbf{e}_i$, and $\mathbf{I}_N$ denote the element-wise product, expectation, a diagonal matrix with $\mathbf{x}$ on its diagonal, the ith element of $\mathbf{x}$, ith diagonal element of $\mathbf{X}$, all-zero vector with a one in the ith position, and an $N \times N$ identity matrix, respectively.

local CSI dissemination overhead. Addressing this problem, in the following we present a distributed gradient descent based scheme, which does not require global CSI knowledge. Still, the resultant gain vector converges to the optimal relay gain vector obtained from the GRQ.

The gradient descent algorithm updates the gain vector at any iteration by using the vector of the previous iteration, a search direction, i.e., the gradient, and a step size [7]. Applied to our scenario, the update equation for the normalized relay gain vector $\mathbf{w}$ in the nth iteration can be expressed as

$$\mathbf{w}_{n+1} = \mathbf{w}_n + \mu_n \cdot \nabla_{\mathbf{w},n}\mathrm{SINR}_i, \tag{6}$$

where μ_n and $\nabla_{\mathbf{w},n}\mathrm{SINR}_i$ denote the optimal step size and the gradient of SINR_i with respect to $\mathbf{w}$, respectively. Starting from a locally-computable initial gain vector, at each *outer* iteration we are computing the gradient in a distributed manner as will be explained in Section III-A. For each gradient, we optimize the step size iteratively. We refer to the step size optimization cycle as *inner loop* (see Section III-B).

A. Local Gradient Calculation

The gradient of SINR_i with respect to $\mathbf{w}$ is given by

$$\nabla_{\mathbf{w}}\mathrm{SINR}_i = \frac{\mathbf{X}_i\mathbf{w}}{\mathbf{w}^H\mathbf{Z}_i\mathbf{w}} - \frac{\mathbf{Z}_i\mathbf{w}}{\mathbf{w}^H\mathbf{Z}_i\mathbf{w}}\mathrm{SINR}_i, \tag{7}$$

where we drop subscript n in this subsection for the sake of notational simplicity. The following lemma states the information needed at each relay for the local computation of the corresponding element of the gradient (7).

Lemma: *Assume each relay has an a priori knowledge on $\sigma_m^2, \sigma_v^2, P_\mathrm{r}$ and P_s. The sufficient information for the kth relay to "locally" compute the corresponding element $\nabla_{\mathbf{g}}\mathrm{SINR}_i[k]$ of the gradient of SINR_i with respect to $\mathbf{g}$, are local CSI (i.e., $\mathbf{h}_{\mathrm{sr},k}, \hat{\mathbf{h}}_{\mathrm{rd},k}$), the equivalent channel $\mathbf{h}_{\mathrm{sd},i}$, and SINR_i.*

Proof: First, we incorporate the definition of $\mathbf{X}_i$ from (5) and write $\mathbf{X}_i\mathbf{w} = \hat{\mathbf{h}}_{\mathrm{sr},i}\hat{\mathbf{h}}_{\mathrm{sr},i}^H\mathbf{w}$. As the product $\hat{\mathbf{h}}_{\mathrm{sr},i}^H\mathbf{w}$ represents the conjugate of $\mathbf{h}_{\mathrm{sd},i}[i]$, we find that $\mathbf{X}_i\mathbf{w} = \mathbf{h}_{\mathrm{sd},i}^*[i] \cdot \hat{\mathbf{h}}_{\mathrm{sr},i}$. Further, we multiply $\mathbf{X}_i\mathbf{w}$ with $\mathbf{w}^H$ from the left-side, and obtain the received intended signal power at the ith destination, i.e., $\mathbf{w}^H\mathbf{X}_i\mathbf{w} = \mathbf{h}_{\mathrm{sd},i}^*[i] \cdot \mathbf{w}^H\hat{\mathbf{h}}_{\mathrm{sr},i} = \mathbf{h}_{\mathrm{sd},i}^*[i] \cdot \mathbf{h}_{\mathrm{sd},i}[i]$. Further, incorporating the SINR_i definition in (5), we write

$$\mathbf{w}^H\mathbf{Z}_i\mathbf{w} = (\mathbf{w}^H\mathbf{X}_i\mathbf{w})/(\mathrm{SINR}_i) = (|\mathbf{h}_{\mathrm{sd},i}[i]|^2)/(\mathrm{SINR}_i). \tag{8}$$

Next, we define an auxiliary vector

$$\tilde{\mathbf{h}}_{\mathrm{sd},i} \triangleq \left[\mathbf{h}_{\mathrm{sd},i}[1]\cdots\mathbf{h}_{\mathrm{sd},i}[i-1]\;\mathbf{h}_{\mathrm{sd},i}[i+1]\cdots\mathbf{h}_{\mathrm{sd},i}[N]\right]^T,$$

which collects all two-hop equivalent interference channels for the ith destination. Then, we multiply the definition of $\mathbf{Z}_i$ in (5) with $\mathbf{w}$ from the right-side such that we write the resultant expression as

$$\mathbf{Z}_i\mathbf{w} = \mathbf{H}_{\mathrm{I},i}\tilde{\mathbf{h}}_{\mathrm{sd},i}^* + \left(\mathbf{I}_{N_\mathrm{r}}\sigma_v^2 + \frac{\sigma_m^2}{P_\mathrm{r}}(\mathbf{D}_{\mathrm{rd},i}^{-1})\mathbf{M}(\mathbf{D}_{\mathrm{rd},i}^{-1})^H\right)\mathbf{w}. \tag{9}$$

Substituting all above derivations into (7), $\nabla_{\mathbf{w}}\mathrm{SINR}_i$ becomes

$$\nabla_{\mathbf{w}}\mathrm{SINR}_i = \Bigg(\hat{\mathbf{h}}_{\mathrm{sr},i}\mathbf{h}_{\mathrm{sd},i}^*[i] - \Big(\mathbf{H}_{\mathrm{I},i}\tilde{\mathbf{h}}_{\mathrm{sd},i}^* + \big(\mathbf{I}_{N_\mathrm{r}}\sigma_v^2 + \frac{\sigma_m^2}{P_\mathrm{r}}(\mathbf{D}_{\mathrm{rd},i}^{-1})\mathbf{M}(\mathbf{D}_{\mathrm{rd},i}^{-1})^H\big)\mathbf{w}\Big)\mathrm{SINR}_i\Bigg)\frac{\mathrm{SINR}_i}{|\mathbf{h}_{\mathrm{sd},i}[i]|^2}. \tag{10}$$

Hence, the kth element of $\nabla_{\mathbf{w}}\mathrm{SINR}_i$, which corresponds to the kth relay's own gain factor $\mathbf{w}[k]$, is given by

$$\nabla_{\mathbf{w}}\mathrm{SINR}_i[k] = \Bigg(\hat{\mathbf{h}}_{\mathrm{sr},i}[k]\mathbf{h}_{\mathrm{sd},i}^*[i] - \Big(\mathbf{h}_{\mathrm{I},i,k}^T\tilde{\mathbf{h}}_{\mathrm{sd},i}^* + \big(\sigma_v^2 + \frac{\sigma_m^2}{P_\mathrm{r}}\frac{(\mathbf{h}_{\mathrm{sr},k}^H\mathbf{h}_{\mathrm{sr},k} + \sigma_v^2)}{|\mathbf{h}_{\mathrm{rd},i}[k]|^2}\big)\mathbf{w}[k]\Big)\mathrm{SINR}_i\Bigg)\frac{\mathrm{SINR}_i}{|\mathbf{h}_{\mathrm{sd},i}[i]|^2},$$

where $\mathbf{h}_{\mathrm{I},i,k}^T$ is the kth row of $\mathbf{H}_{\mathrm{I},i}$. As a last step, the gradient with respect to the physical gain vector $\mathbf{g}$ follows basically from the chain rule as $\nabla_{\mathbf{g}}\mathrm{SINR}_i[k] = \nabla_{\mathbf{w}}\mathrm{SINR}_i[k] \cdot \mathbf{D}_{\mathrm{rd},i}[k,k]$.

In order to compute $\nabla_{\mathbf{g}}\mathrm{SINR}_i[k]$ locally, the kth relay requires the information set

$$\mathcal{I} = \{\hat{\mathbf{h}}_{\mathrm{sr},i}[k], \mathbf{h}_{\mathrm{sr},k}, \mathbf{h}_{\mathrm{rd},i}[k], \mathbf{h}_{\mathrm{sd},i}, \mathbf{h}_{\mathrm{I},i,k}, \mathrm{SINR}_i\}.$$

Incorporating that $\hat{\mathbf{h}}_{\mathrm{sr},i}[k] = \mathbf{h}_{\mathrm{sr},k}[i]$, $\mathbf{h}_{\mathrm{rd},i}[k] = \hat{\mathbf{h}}_{\mathrm{rd},k}[i]$, and $\mathbf{h}_{\mathrm{I},i,k} = \left[\mathbf{h}_{\mathrm{sr},k}[1]\cdots\mathbf{h}_{\mathrm{sr},k}[i-1]\;\mathbf{h}_{\mathrm{sr},k}[i+1]\cdots\mathbf{h}_{\mathrm{sr},k}[N]\right]^T$,we drop $\hat{\mathbf{h}}_{\mathrm{sr},i}[k]$, $\mathbf{h}_{\mathrm{rd},i}[k]$, $\mathbf{h}_{\mathrm{I},i,k}$ from the required information set $\mathcal{I}$ and conclude the proof. ■

The next question to answer is how to gather the necessary information. For the clarity of exposition in the paper, we consider a unidirectional traffic pattern from a set of source nodes to a set of destination nodes. In practice all terminal nodes in the network will transmit and receive packets, leading to a bidirectional traffic pattern. We assume that the packets transmitted by different terminal nodes comprise orthogonal training sequences. Thus, the relays can estimate the local CSI without additional overhead. The destination nodes in turn can use the training sequences to estimate the respective rows of the equivalent channel matrix $\mathbf{H}_{\mathrm{sd}}$ and the receive SINR. This information is subsequently fed back (broadcasted) to all relays. In the sequel, we will refer to the process of training sequence transmission and parameter estimation as *estimate cycle*, and to the feedback of the estimated parameter values as *feedback cycle*. Based on its local channel estimates and the feedback of the destinations each relay can compute the corresponding element of the gradient locally and independently from the other relays in the network.

B. Distributed Step Size Calculation

If global CSI is available, the step size can be efficiently computed with exact or backtracking line searches [7]. In other words, the following unconstrained optimization over μ_n

$$\arg\max_{\mu_n \geq 0} \frac{(\mathbf{g}_n + \mu_n\nabla_{\mathbf{g},n}\mathrm{SINR}_i)^H\tilde{\mathbf{X}}_i(\mathbf{g}_n + \mu_n\nabla_{\mathbf{g},n}\mathrm{SINR}_i)}{(\mathbf{g}_n + \mu_n\nabla_{\mathbf{g},n}\mathrm{SINR}_i)^H\tilde{\mathbf{Z}}_i(\mathbf{g}_n + \mu_n\nabla_{\mathbf{g},n}\mathrm{SINR}_i)} \tag{11}$$

is performed, where $\tilde{\mathbf{X}}_i \triangleq \mathbf{D}_{\mathrm{rd},i}\mathbf{X}_i\mathbf{D}_{\mathrm{rd},i}^H$ and $\tilde{\mathbf{Z}}_i \triangleq \mathbf{D}_{\mathrm{rd},i}\mathbf{Z}_i\mathbf{D}_{\mathrm{rd},i}^H$.

Here, we aim at designing schemes which perform (11) in a distributed manner without requiring global CSI. To this end, we propose a centralized and a decentralized protocol, which both employ only local information at the relays and efficiently benefit from several additional *estimate* and *feedback* cycles.

1) The Decentralized Protocol: The decentralized protocol uses an *inner loop*, which consists of several estimate cycles and a 1-bit feedback cycle to determine the optimal step size. Initially we define a minimum value $\mu_n^{\min}$, and a maximum value $\mu_n^{\max}$ for the step size. Starting from $\mu_n^{\min}$ and iterating until $\mu_n^{\max}$ with an appropriate inner increment δ, the network performs multiple estimate cycles between relays and the destination. In the qth cycle the step size $\mu_{n,q} \triangleq \mu_n^{\min}+(q-1)\delta$ is used. The destination measures the corresponding SINR at its antenna for each step-size choice $\mu_{n,q}$. As soon as it detects that SINR is reduced with respect to the previous choice, it feeds back a stop bit through a dedicated control channel. This terminates the inner loop and the relays use the step size of the previous cycle as optimal step size for the present gradient.

The gain vector in this inner loop does not necessarily satisfy the instantaneous relay sum power constraint (2). Because as the vector is updated by each choice of μ_n, its length increases monotonically on average. Hence, this may introduce a bias in the SINR estimation. In order to cope with this, we need to normalize the updated gain vector at each inner loop iteration such that the relay sum transmit power, i.e. $\mathbf{g}_{n+1}^H \mathbf{M} \mathbf{g}_{n+1}$, is independent of the choice of μ_n. For a given $\mu_{n,q}$, the sum transmit power without gain vector normalization is given by

$$\begin{aligned}\mathbf{g}_{n+1}^H \mathbf{M} \mathbf{g}_{n+1} = \|\mathbf{M}^{\frac{1}{2}} \mathbf{g}_n\|^2 + \mu_{n,q}^2 \|\mathbf{M}^{\frac{1}{2}} \cdot \nabla_{\mathbf{g},n} \mathrm{SINR}_i\|^2 \\ +2\mu_{n,q} \mathrm{Re}\big\{(\nabla_{\mathbf{g},n} \mathrm{SINR}_i)^H \mathbf{M} \mathbf{g}_n\big\}. \qquad (12)\end{aligned}$$

If the three summands in (12) are known at the relays, the necessary scaling can also be determined locally. Hence, we propose to measure these quantities at the destination and fed them back to the relays. Each measurement requires one channel use. The key idea is to use *over-the-air addition* through the relay-to-destination channel to perform the required matrix multiplications in (12). During these *measurement cycles*, the relays do not act as forwarders but rather transmit a specific locally computable value each. Let us now explain *over-the-air addition* for measuring $\|\mathbf{M}^{\frac{1}{2}} \mathbf{g}_n\|^2$ at the intended destination. With the available local CSI, the kth relay computes a transmit symbol $\alpha_k \in \mathbb{C}$, which is defined as $\alpha_k \triangleq (|\mathbf{g}_n[k]|^2 \cdot (\mathbf{h}_{\mathrm{sr},k}^H \mathbf{h}_{\mathrm{sr},k} + \sigma_v^2))/\hat{\mathbf{h}}_{\mathrm{rd},k}[i]$, where the square root of the numerator represents the absolute value of the kth element of $\mathbf{M}^{\frac{1}{2}} \mathbf{g}_n$ (see (2)), and the denominator compensates the forward channel. As all N_{r} relays concurrently transmit the respective α_k, the received signal at the ith destination becomes $d_i = \sum_{k=1}^{N_{\mathrm{r}}} \hat{\mathbf{h}}_{\mathrm{rd},k}[i] \alpha_k = \sum_{k=1}^{N_{\mathrm{r}}} |\mathbf{g}_n[k]|^2 \cdot (\mathbf{h}_{\mathrm{sr},k}^H \mathbf{h}_{\mathrm{sr},k} + \sigma_v^2)$, which perfectly computes $\|\mathbf{M}^{\frac{1}{2}} \mathbf{g}_n\|^2$. Thus, the destination measures this received signal amplitude, and feeds it to the relays. Likewise, to measure the second and the third summands, the kth relay transmits $|\nabla_{\mathbf{g},n} \mathrm{SINR}_i[k]|^2 (\mathbf{h}_{\mathrm{sr},k}^H \mathbf{h}_{\mathrm{sr},k} + \sigma_v^2)/\hat{\mathbf{h}}_{\mathrm{rd},k}[i]$ and $\nabla_{\mathbf{g},n} \mathrm{SINR}_i[k] \cdot \mathbf{g}_n^*[k] (\mathbf{h}_{\mathrm{sr},k}^H \mathbf{h}_{\mathrm{sr},k} + \sigma_v^2)/\hat{\mathbf{h}}_{\mathrm{rd},k}[i]$, respectively. Note that these cycles are performed only once at the start of line search.

2) The Centralized Protocol: Here the ith destination computes the optimal μ_n through (11) locally, and then broadcasts it back to all relays in the network. To this end, the destination needs to know the following six terms: $\mathbf{g}_n^H \tilde{\mathbf{X}}_i \mathbf{g}_n$, $\mathbf{g}_n^H \tilde{\mathbf{Z}}_i \mathbf{g}_n$, $(\nabla_{\mathbf{g},n} \mathrm{SINR}_i)^H \tilde{\mathbf{X}}_i (\nabla_{\mathbf{g},n} \mathrm{SINR}_i)$, $(\nabla_{\mathbf{g},n} \mathrm{SINR}_i)^H \tilde{\mathbf{Z}}_i (\nabla_{\mathbf{g},n} \mathrm{SINR}_i)$, $\mathrm{Re}\big\{\mathbf{g}_n^H \tilde{\mathbf{X}}_i (\nabla_{\mathbf{g},n} \mathrm{SINR}_i)\big\}$ and $\mathrm{Re}\big\{\mathbf{g}_n^H \tilde{\mathbf{Z}}_i (\nabla_{\mathbf{g},n} \mathrm{SINR}_i)\big\}$ (see (11)). Since the destination, as well as all relays, has already the knowledge of SINR_i and $\mathbf{h}_{\mathrm{sd},i}$, the first and the second terms are locally available, i.e., $\mathbf{g}_n^H \tilde{\mathbf{X}}_i \mathbf{g}_n = |\mathbf{h}_{\mathrm{sd},i}[i]|^2$, and $\mathbf{g}_n^H \tilde{\mathbf{Z}}_i \mathbf{g}_n = |\mathbf{h}_{\mathrm{sd},i}[i]|^2/\mathrm{SINR}_i$.

In order to measure the third, the fourth and the fifth terms, the network performs an *estimate and feedback* cycle with $\nabla_{\mathbf{g},n} \mathrm{SINR}_i$ as the relay gain vector instead of $\mathbf{g}_n$, and the corresponding SINR_i^{∇} and $\mathbf{h}_{\mathrm{sd},i}^{\nabla}$ are measured/estimated at the destination. Consequently, the related terms can be computed as $(\nabla_{\mathbf{g},n} \mathrm{SINR}_i)^H \tilde{\mathbf{X}}_i (\nabla_{\mathbf{g},n} \mathrm{SINR}_i) = |\mathbf{h}_{\mathrm{sd},i}^{\nabla}[i]|^2$, $(\nabla_{\mathbf{g},n} \mathrm{SINR}_i)^H \tilde{\mathbf{Z}}_i (\nabla_{\mathbf{g},n} \mathrm{SINR}_i) = |\mathbf{h}_{\mathrm{sd},i}^{\nabla}[i]|^2/\mathrm{SINR}_i^{\nabla}$, and $\mathrm{Re}\big\{\mathbf{g}_n^H \tilde{\mathbf{X}}_i (\nabla_{\mathbf{g},n} \mathrm{SINR}_i)\big\} = \mathrm{Re}\big\{\mathbf{h}_{\mathrm{sd},i}^*[i] \mathbf{h}_{\mathrm{sd},i}^{\nabla}[i]\big\}$.

The last term can only be partially computed with the available information at the destination, i.e., SINR_i, SINR_i^{∇}, $\mathbf{h}_{\mathrm{sd},i}$ and $\mathbf{h}_{\mathrm{sd},i}^{\nabla}$. Taking a closer look, we have

$$\begin{aligned}\mathrm{Re}\big\{\mathbf{g}_n^H \tilde{\mathbf{Z}}_i (\nabla_{\mathbf{g},n} \mathrm{SINR}_i)\big\} = \mathrm{Re}\big\{\sigma_v^2 \mathbf{g}_n^H \mathbf{D}_{\mathrm{rd},i} \mathbf{D}_{\mathrm{rd},i}^H (\nabla_{\mathbf{g},n} \mathrm{SINR}_i) \\ + \mathbf{g}_n^H \mathbf{D}_{\mathrm{rd},i} \mathbf{H}_{\mathrm{I},i} \mathbf{H}_{\mathrm{I},i}^H \mathbf{D}_{\mathrm{rd},i}^H (\nabla_{\mathbf{g},n} \mathrm{SINR}_i) + \frac{\sigma_m^2}{P_{\mathrm{r}}} \mathbf{g}_n^H \mathbf{M} \nabla_{\mathbf{g},n} \mathrm{SINR}_i\big\} \qquad (13)\end{aligned}$$

where the second summand can be computed using $\mathbf{h}_{\mathrm{sd},i}$ and $\mathbf{h}_{\mathrm{sd},i}^{\nabla}$ jointly. However, we need some other *measurement* cycles to obtain the first and the third summands in (13). For the $\sigma_v^2 \mathbf{g}_n^H \mathbf{D}_{\mathrm{rd},i} \mathbf{D}_{\mathrm{rd},i}^H (\nabla_{\mathbf{g},n} \mathrm{SINR}_i)$ term, a concurrent transmission is performed from all relays to the destination, where the kth relay transmits $(\sigma_v^2 \mathbf{g}_n^*[k] |\hat{\mathbf{h}}_{\mathrm{rd},k}[i]|^2 \cdot \nabla_{\mathbf{g},n} \mathrm{SINR}_i[k])/\hat{\mathbf{h}}_{\mathrm{rd},k}[i]$. Lastly, the third term in (13) is similarly obtained as explained in the decentralized option. In conclusion, the destination is capable to perform the unconstrained maximization (11) with the available information gathered after several *measurement* cycles. As the optimal $\mu_n^{\star}$ is found, it is broadcasted to all relays. Note that in order to cope with the monotonic increase of SINR with respect to μ_n, this scheme also requires an appropriate scaling as introduced in the previous subsection.

3) Discussion: Comparing the two protocols we observe that both require some measurement cycles, where specific relay transmit symbols are used to efficiently evaluate matrix expressions. The decentralized scheme furthermore uses multiple estimate cycles, i.e., training sequences from all sources, to iteratively determine the optimum step size. In contrast, the centralized option requires estimate cycles to determine some constants in (11) at the destination. In general the decentralized protocol is preferable in terms of overhead, if few inner iterations are necessary. For this reason, we use the centralized option in the acquisition phase, and the decentralized option in the tracking phase of the gradient scheme.

C. Conjugate Gradient and Convergence Issues

The conjugate gradient (CG) algorithm, where the search direction is chosen to be orthogonal to the previously searched directions, is a commonly preferred scheme due to its fast convergence properties with respect to conventional gradient algorithms. With conjugate gradients, the update equation and the search direction are modified to $\mathbf{g}_{n+1} = \mathbf{g}_n + \mu_n \boldsymbol{\phi}_n$ and $\boldsymbol{\phi}_n = \nabla_{\mathbf{g},n} \mathrm{SINR}_i + \beta_n \cdot \nabla_{\mathbf{g},n-1} \mathrm{SINR}_i$. There are several different formulas to compute the constant $\beta_n \in \mathbb{R}^+$ [8]. For further details of CG, we refer the interested reader to [8].

The CG methods for the solution of *(generalized) Rayleigh quotient* problems have been well-studied in the literature (see [9] and references therein), where several different algorithms with various complexity have been proposed and the corresponding convergence analyses have been performed. Henceforth, we employ the CG scheme in order to accelerate the convergence speed of our distributed scheme.

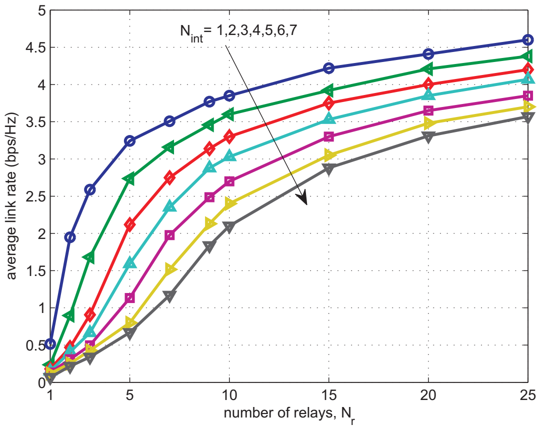

Figure 1. Single source-destination pair in the presence of N_{int} interference sources: Average achievable link rates versus number of relays N_{r} with optimal relay gain allocation (SNR = 20 dB).

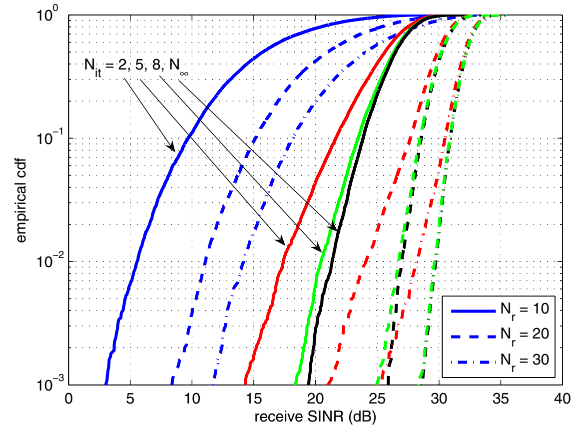

Figure 2. Single source-destination pair in the presence of two interference sources: Empirical CDFs of receive SINR for different numbers of relays N_{r} and outer iterations N_{it} in the gradient algorithm (SNR = 25 dB).

D. Extension to Multiple Source/Destination Pairs

In the case of multiple S-D pairs, the relays can employ the local gradient calculation scheme in Section III-A to determine the gradient of the SINR for each destination. However, the important decision to make in the multiuser case is how to employ the gradients, i.e., what is the objective function?

In this paper, we choose the sum rate of the N S-D pairs as the objective function to be maximized. Assuming that i.i.d. Gaussian symbols are transmitted, the sum rate is given by

$$\mathrm{R}_{\text{sum}} = \textstyle\sum_{i=1}^{N} \log_2(1+\mathrm{SINR}_i),$$

and its gradient with respect to the relay gain vector $\mathbf{g}$ is

$$\nabla_{\mathbf{g}} \mathrm{R}_{\text{sum}} = \frac{1}{\log_e(2)} \cdot \sum_{i=1}^{N} \frac{1}{1+\mathrm{SINR}_i} \nabla_{\mathbf{g}} \mathrm{SINR}_i. \qquad (14)$$

As each relay can compute its elements of the gradients $\nabla_{\mathbf{g}}\mathrm{SINR}_i$, and has acquired knowledge of all SINR_i in this process, it can as well compute its contribution to $\nabla_{\mathbf{g}}\mathrm{R}_{\text{sum}}$.

Dropping further details for the sake of the brevity, an algorithm similar to the SINR maximization is used to maximize R_{sum} by incorporating (14) in the update equation. Both protocols for step size calculation are applicable with some additional *estimate and feedback* cycles due to the increased number of cross terms of the signal, interference and noise.

IV. Simulation Results and Discussions

In this section we provide insight into the speed of convergence of the considered conjugate gradient method by means of computer experiments. We fix $P_{\mathrm{s}} = P_{\mathrm{r}} = 1$ and $\sigma_m^2 = \sigma_v^2$. We define the average signal-to-noise ratio as $\mathrm{SNR} = P_{\mathrm{s}}/\sigma_m^2$.

a) Single S-D Pair under Interference: Figures 1 and 2 illustrate the performance of the coherent AF relaying scheme with single S-D pair in the presence of multiple interference sources. For the results in Fig. 1, the gain vector has been determined by directly solving the generalized Rayleigh quotient problem (5). As such Fig. 1 provides an upper bound on the performance of any gradient based scheme. Specifically, it depicts the average achievable link rate versus N_{r} for various numbers of interference sources N_{int}. We observe a strictly increasing achievable rate in N_{r}. Up to the point $N_{\mathrm{r}} = N_{\text{int}}$, this increase is explained by the capability of suppressing more and more interfering streams. Moreover, for $N_{\mathrm{r}} > N_{\text{int}}$, the relay nodes are capable of further improving the SINR by the realization of a coherent superposition of the desired signal at the destination node (distributed array gain).

Fig. 2 shows the performance of the distributed gradient based gain allocation scheme for $N_{\text{int}} = 2$. It depicts the empirical cumulative distribution function (CDF) of the receive SINR in three sets of curves, which corresponds to $N_{\mathrm{r}} = 10, 20$, and 30 relays. The sets are identified by line style (solid, dashed, dashed-dotted). For each set, the CDF after $N_{\text{it}} = 2, 5, 8$ and N_∞ outer iterations is shown, where N_∞ corresponds to the optimal result obtained by the generalized Rayleigh quotient problem (5). The number of iterations is identified by the line color (blue, red, green, black). In all cases $N_{\text{it}} = 8$ iterations (green) suffice to achieve almost optimal performance (N_∞, black). For a given number of iterations, the performance improves dramatically with N_{r}. After $N_{\text{it}} = 5$ iterations (red), for example the $N_{\mathrm{r}} = 10$ configuration (solid) achieves an outage SINR of 17 dB at 1% outage probability, whereas with $N_{\mathrm{r}} = 20$ relays the outage SINR improves by 7 dB to 24 dB. Adding 10 more relays ($N_{\mathrm{r}} = 30$) further improves the performance by 4 dB.

The starting point has considerable impact on the convergence speed of gradient searches. In the experiments of this subsection we initialize the gradient algorithm with a local CSI based gain vector, whose kth element is given by

$$\mathbf{g}_0[k] = \left(\mathbf{h}_{\mathrm{sr},k}[i] \cdot \hat{\mathbf{h}}_{\mathrm{rd},k}[i]\right) / \left(\sigma_v^2 |\hat{\mathbf{h}}_{\mathrm{rd},k}[i]|^2 + (\sigma_m^2/P_{\mathrm{r}}) \mathbf{h}_{\mathrm{sr},k}^H \mathbf{h}_{\mathrm{sr},k}\right).$$

Up to a constant factor, this gain vector is optimal in the absence of interference, since it implements a maximum ratio combining of the desired signal contributions at the destination. This initialization yields a huge improvement of the convergence speed as compared to a random initialization.

b) Multiple Source-Destination Pairs: Next, we discuss the performance of the distributed gradient based gain allocation for multiple S-D pairs. Objective function is the sum rate of the N links. In contrast to the single S-D pair case here no analytical solution is available. Fig. 3 illustrates the convergence behaviour (outer iterations) of the conjugate gradient

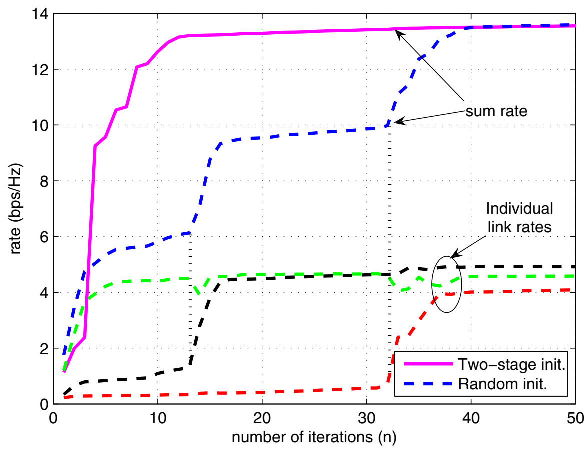

Figure 3. Three source-destination pairs: Sample trajectory of achievable rates at each outer iteration for a typical channel realization with $N_{\mathrm{r}} = 30$.

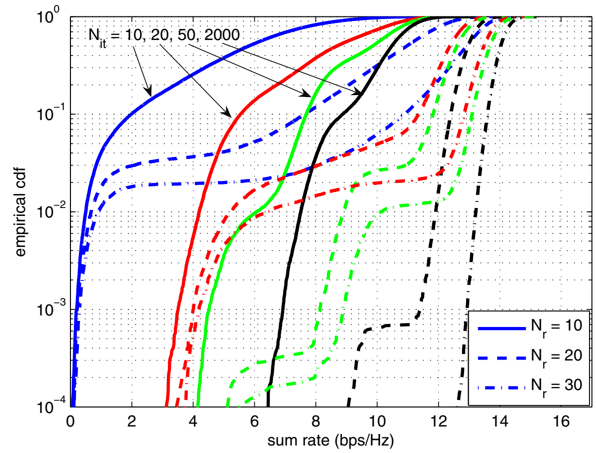

Figure 4. Three source-destination pairs: Empirical CDFs of the achievable sum-rate for various numbers of relays N_{r} and outer iterations N_{it} in the gradient algorithm (SNR = 25 dB).

algorithm for a typical channel realization, $N = 3$ S-D pairs and $N_{\mathrm{r}} = 30$ relays. The dashed lines refer to a random gain vector initialization. According to the trajectories of the individual link rates after 35 iterations all three links support a rate of about 4.5 bps/Hz each. The convergence behaviour of the three trajectories is dramatically different however. The links (spatial subchannels) open up one after the other with intermediate periods without much improvement. This leads to the plateau structure of the sum rate trajectory (dashed-blue). While this basic behaviour can be observed for most channel realizations, the length of the plateaus between the transitions varies substantially.

In order to capture this statistical influence, we consider the empirical CDF of the sum rate in Fig. 4. The results are based on 100000 channel realizations, and we employ the heuristic two-stage initialization. Three sets of curves are shown, which refer to $N_{\mathrm{r}} = 10$ (solid), 20 (dashed) and 30 (dashed-dotted) relays. Parameter within each set is the number of iterations N_{it} (line color). A comparison of the curves with the same color (same N_{it}) shows that with increasing number of relays, the CDF tends to develop one or several plateaus. We attribute this to the different number of subchannels, which are open after N_{it} iterations. This effect becomes more pronounced as the number of relays grows. Let us consider the $N_{\mathrm{r}} = 30$ case in more detail. After $N_{\mathrm{it}} = 10$ iterations (dashed-dotted-blue) either all three subchannels are open, which leads to a sum rate of about 13bps/Hz or no subchannel is open. The latter happens with a probability of about 1%. After 20 iterations (dashed-dotted-red) either three or one subchannel (sum rate 4bps/Hz) are open in our experiment. We conjecture the existence of another plateau, which corresponds to no open subchannels. This case occurs with such a low probability however, that it is not observable in the figure. After 50 iterations (dashed-dotted-green), we observe either three (13bps/Hz), two (9bps/Hz) or one (4-5bps/Hz) open subchannels. Finally after 2000 iterations (dashed-dotted-black), with probability 99.99% all subchannels are open (no plateau within the depicted probability range).

The conjectured multi-plateau structure of the CDF is nicely supported by curves of $N_{\mathrm{it}} = 50$ iterations (green), which exhibit two plateaus each. A comparison in between these curves reveals, that the probability of difficult channel matrices, which require a large number of iterations to open up at least two subchannels, drops with an increasing number of relays, e.g., the plateaus between three and two open subchannels occur respectively at decreasing probabilities of 3×10^{-1}, 3×10^{-2}, 1×10^{-2}. This is a result of the increasing diversity in the system. Thus, the gradient scheme benefits from the number of relays.

The speed of convergence is substantially affected by the initial gain vector. For multiple S-D pairs we cannot apply the initial vector from the previous subsection, because it is optimized for one specific S-D pair. We have found a heuristic initialization method however, which hugely improves convergence. The solid magenta line in Fig. 3 shows achieved the sum rate trajectory for the same channel as the dashed curves. Note the dramatic improvement. For the sake of brevity in this paper we cannot elaborate on this scheme. Nevertheless, in Fig. 4 we used this optimized initialization scheme in order to provide realistic performance figures. For a random initialization basically the same CDFs would be obtained, if we increase the number of iterations by an order of magnitude.

References

[1] V. R. Cadambe and S. A. Jafar, "Interference alignment and the degrees of freedom for the K user interference channel," ArXiv, vol. 707, Jul. 2007.

[2] A. Wittneben and B. Rankov, "Distributed antenna systems and linear relaying for gigabit MIMO wireless," in *IEEE VTC Fall*, Sep. 2004.

[3] V. I. Morgenshtern and Helmut Bölcskei, "Crystallization in Large Wireless Networks", *IEEE Trans. Inf. Theory*, vol. 53, no. 10, Oct. 2007.

[4] T. Abe, H. Shi, T. Asai, and H. Yoshino, "Relay techniques for MIMO wireless networks with multiple source and destination pairs," *EURASIP J. on Wireless Commun. and Networking*, vol. 2006, pp. 19, April 2006.

[5] C. Eşli, Stefan Berger, and A. Wittneben, "Optimizing zero-forcing based gain allocation for wireless multiuser networks," in *Proc. of IEEE Inter. Conf. on Commun. 2007*, June 2007, pp. 5825 - 5830.

[6] S. Berger and A. Wittneben "Cooperative distributed multiuser MMSE relaying in wireless ad-hoc networks ," in *Proc. of IEEE Asilomar Conf. on Sig., Sys. and Comp., 2005*, California, Nov. 2005, pp. 1072-1076.

[7] S. Boyd and L. Vandenberghe, "Convex Optimization," Cambridge University Press, 2004.

[8] J. R. Shewchuk, "An introduction to the conjugate gradient method without the agonizing pain,", School of Computer Science, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA, $1\frac{1}{4}$. edition, 1994.

[9] H. Yang, "Conjugate gradient methods for the Rayleigh quotient minimization of generalized eigenvalue problems,", Computing, 51 (1993).