Spectral Efficient Protocols for Nonregenerative Half-duplex Relaying

Boris Rankov and Armin Wittneben

Communication Technology Laboratory
Swiss Federal Institute of Technology (ETH) Zurich
Sternwartstrasse 7, 8092 Zurich, Switzerland
Email: {rankov, wittneben}@nari.ee.ethz.ch

Abstract

We study two-hop transmission schemes where relays assist the communication between one source/destination pair or multiple source/destination pairs. The relays operate in half-duplex mode, i.e., may not receive and transmit simultaneously at the same time and frequency. This leads to a loss in spectral efficiency due to the pre-log factor 1/2. We propose and study two nonregenerative relaying protocols that avoid the pre-log factor 1/2 but still operate in half-duplex mode. *Firstly*, we consider a relaying protocol where two half-duplex amplify-and-forward (AF) relays alternately forward messages from a source to a destination. The destination employs successive decoding with partial or full cancelation of the inter-relay interference. It is shown that the protocol can recover a significant portion of the half-duplex loss when the inter-relay channel is not too strong. *Secondly*, we propose a relaying protocol where a synchronous bidirectional connection between two nodes is established using one half-duplex relay and show that the sum rate achieves the rate of a full-duplex AF relay channel. We further extend this protocol to realize a bidirectional communication between multiple node pairs assisted by multiple nonregenerative zero-forcing relays.

1 Introduction

The design and analysis of cooperative transmission protocols for wireless networks has recently attracted a lot of interest. Of particular interest are two-hop channels where a relay terminal assists in the communication between a source terminal and a destination terminal. In [1] the authors consider a relay network with one source and one destination both equipped with M antennas and K half-duplex relays each equipped with $N \geq 1$ antennas. In the absence of a direct link between source and destination and the use of amplify-and-forward (AF) relays the authors show that the capacity scales as $\frac{M}{2}\log(\mathrm{SNR})$ for high signal-to-noise ratios (SNR) when $K \to \infty$. The pre-log factor $\frac{1}{2}$ is induced by the half-duplex signaling and causes a substantial loss in spectral efficiency. Further half-duplex relaying protocols with a pre-log factor $\frac{1}{2}$ can be found in [2] and the references therein. One way to avoid the pre-log factor $\frac{1}{2}$ is to use a full-duplex relay that may receive and transmit at the same time and frequency [3], but such a relay is difficult to implement. Large differences in the signal power of the transmitted and the received signal drive the relay's analog amplifiers in the receive chain into saturation and cause problems in the cancelation of the self-interference.

In [4] the authors address the half-duplex loss by proposing a spatial reuse of the relay slot. They consider a base station that transmits K messages to K users and their corresponding relays in K orthogonal time slots. In time slot $K+1$ each relay retransmits its received signal, causing interference to the other users. The capacity of a single connection (base station to user) has then a pre-log factor $\frac{K}{K+1}$ instead of $\frac{1}{2}$. However, the interference in the relay slot results in SNR losses at the users. In [5] the authors propose a signaling scheme where the source transmits new information in every time slot, which leads to collisions at the destination of relayed signals and signals transmitted over the direct link. However, the pre-log factor $\frac{1}{2}$ is avoided with this scheme. Another solution is presented in [6] where the authors propose a transmission scheme with two half-duplex AF relays that alternately forward messages from a source to a destination. In order to decrease the inter-relay interference, one relay performs interference cancelation. This cooperation scheme turns the equivalent channel between source and destination into a frequency-selective channel. A maximum likelihood sequence estimator at the destination is applied to extract the artificially introduced frequency diversity, an idea which is known as delay diversity [7]. The authors do not study the achievable rate of this transmission protocol.

Our Contribution. We propose two half-duplex relaying schemes that mitigate the loss in spectral efficiency due to the half-duplex operation of the relays. *Firstly*, we consider a similar relaying scheme as in [6] but with the difference that both relays are only allowed to amplify-and-forward their received signals (no cancelation of the inter-relay interference at one of the relays as in [6]), whereas the destination employs successive decoding with partial or full cancelation of the inter-relay interference. It is shown that this protocol can recover a significant portion of the half-duplex loss (pre-log factor $\frac{1}{2}$) when the inter-relay channel is not too strong. *Secondly*, we propose a relaying protocol where a synchronous bidirectional connection between two nodes (e.g., two wireless routers) is established using one half-duplex relay. Hereby, the achievable rate in one direction suffers still from the pre-log factor $\frac{1}{2}$ but since two connections are realized in the same physical channel the sum rate achieves the rate of a full-duplex AF relay channel. We then extend this scheme to realize a bidirectional communication between multiple node pairs assisted by multiple nonregenerative zero-forcing relays.

2 Alternating Between Two Relays

We consider transmissions of messages from a source S to a destination D via two amplify-and-forward relays R_1 and R_2, which may not receive and transmit simultaneously. We assume that there is no direct connection between S and D, e.g., due to shadowing or too large separation between S and D. A message is transmitted in two slots. In the first slot the source transmits the message to relay R_1 or R_2 and in the second slot the message is forwarded to the destination. The length of one slot is equal to the length of one codeword (frame) and is NT, where T is the sampling interval and N the number of symbols in each frame. In odd frame slots, $k = 1, 3, 5, \ldots$, relay R_1 receives and R_2 transmits[1], whereas in even frame slots, $k = 2, 4, 6, \ldots$, it is the other way around. This cooperation protocol avoids the pre-log factor $\frac{1}{2}$ since the source transmits a new message in every frame slot and has not to be quiet in each second frame slot. However, since we assume that the relays do not operate in orthogonal channels, there will be interference

[1]Except for $k = 1$, where R_2 does not transmit

between R_1 and R_2 and it is not clear *a priori* whether this inter-relay interference cancels the achieved gain achieved by the increased pre-log factor.

Assume that a sequence of K messages is to be transmitted. In frame slot $k \in \{1, 2, \ldots, K\}$ the source S chooses randomly a message (index) $M[k] \in \{1, 2, \ldots, 2^{NR[k]}\}$ according to a uniform distribution with $R[k]$ being the achievable rate in frame k. The message $M[k]$ is then mapped to a codeword $s[k] = (s[k, 1], s[k, 2], \ldots, s[k, N])$ of length N where the symbols $\{s[k, n]\}_{k,n}$ are independent and identically distributed (i.i.d.) according to $\mathcal{CN}(0, P_\mathrm{s})$ with P_s the average transmit power of the source. At the discrete time $[k, n] := (kN + n)\,T$ the receive signal of relay R_p with $p = 2 - \mathrm{mod}\,(k, 2) \in \{1, 2\}$ and $q = 2 - \mathrm{mod}\,(k - i, 2) \in \{1, 2\}$ is given as

$$r[k, n] = h_{0p}[k, n]s[k, n] + n_\mathrm{r}[k, n] + \sum_{i=1}^{k-1} \Big(h_{0q}[k-i, n]s[k-i, n] + n_\mathrm{r}[k-i, n] \Big) f_i[k, n] \quad (1)$$

and

$$f_i[k, n] = \prod_{j=1}^{i} h_{12}[k+1-j, n]g[k-j] \quad (2)$$

denotes the *inter-relay interference factor*. The relay noise samples $\{n_\mathrm{r}[k, n]\}_{k,n}$ are i.i.d. according to $\mathcal{CN}(0, \sigma_\mathrm{r}^2)$. Let S be node 0, R_1 node 1, R_2 node 2 and D node 3. The channel gain between node i and node j at $(kN + n)\,T$ is denoted as $h_{ij}[k, n]$, with $\mathbb{E}|h_{01}|^2 = \mathbb{E}|h_{02}|^2 = \nu_1^2$, $\mathbb{E}|h_{13}|^2 = \mathbb{E}|h_{23}|^2 = \nu_2^2$ and $\mathbb{E}|h_{12}|^2 = \nu_{12}^2$[2]. We assume that $\{h_{ij}[k, n]\}_{k,n}$ are stationary and ergodic processes. The source S knows the *fading distribution* of the channel gains in the network but not the *fading realizations*. The relay nodes R_1 and R_2 do not have any channel knowledge and the destination knows the *fading realizations* of all channel gains in the network. The transmit signal of relay R_p is a scaled version of its received signal: $t[k+1, n] = g[k]r[k, n]$, where $g[k]$ is the scaling coefficient of either relay R_1 or R_2, depending on k.

Destination D observes at $((k + 1)N + n)\,T$ the signal

$$d[k+1, n] = h_{p3}[k+1, n]g[k]r[k, n] + n_\mathrm{d}[k + 1, n] \quad (3)$$

where the noise samples $\{n_\mathrm{d}[k, n]\}_{k,n}$ are i.i.d. according to $\mathcal{CN}(0, \sigma_\mathrm{d}^2)$ and $r[k, n]$ is given in (1). To decode $s[k]$ from (3) the destination receiver first subtracts the previously decoded codewords $s[k-1], \ldots, s[1]$ from the received signal $d[k+1, n]$, because these codewords appear as accumulated inter-relay interference at the destination. However, the influence of the codewords transmitted several frame slots before k may be neglected since they were attenuated several times by the inter-relay channel h_{12}, which acts as forgetting factor for the decoding process, see (1) and (2). After perfect cancelation of m previously decoded codewords $s[k-1], s[k-2], \ldots, s[k-m]$ the destination signal is given by

$$\begin{aligned} d_m[k+1, n] = {} & h_{p3}[k+1, n]g[k]h_{0p}[k, n]s[k, n] + n_\mathrm{d}[k+1, n] \\ & + h_{p3}[k+1, n]g[k] \sum_{i=m+1}^{k-1} h_{0q}[k-i, n]s[k-i, n]f_i[k, n] \\ & + h_{p3}[k+1, n]g[k] \sum_{i=0}^{k-1} n_\mathrm{r}[k-i, n]f_i[k, n] \end{aligned} \quad (4)$$

[2]Due to notational simplicity, we assume equal fading variances from the source to both relays and equal fading variances from both relays to the destination, respectively. The following exposition may be easily extended to the case with unequal fading variances in each link.

where $f_i[k,n] := 1$ for $i = 0\,\forall\, k, n$. For $m = k-1$ all previously transmitted codewords are canceled (*full interference cancelation*). For $m = 0$ all codewords up to $s[k{-}1]$ appear as inter-frame interference when $s[k]$ is decoded. For $0 < m < k-1$ only the last m transmitted codewords are canceled and $s[1], s[2], \ldots, s[k{-}m{-}1]$ remain as interference terms (*partial interference cancelation*). The relay gains are chosen for $k = 2, 3, 4, \ldots$ as

$$g^2[k] = \frac{P_{\mathrm{r}}}{\frac{1}{N}\sum_{n=1}^{N}|r[k,n]|^2} \approx \frac{P_{\mathrm{r}}}{P_{\mathrm{s}}\nu_1^2 + P_{\mathrm{r}}\nu_{12}^2 + \sigma_{\mathrm{r}}^2} := g^2 \tag{5}$$

and in the first frame

$$g[1] = \frac{P_{\mathrm{r}}}{P_{\mathrm{s}}\nu_1^2 + \sigma_{\mathrm{r}}^2} \geq g^2 \tag{6}$$

where P_{r} is the average transmit power of each relay. The approximation in (5) is exact for $N \to \infty$ by the law of large numbers. The inter-relay interference factor (2) may then be written as

$$f_i[k,n] = \prod_{j=1}^{i} h_{12}[k{+}1{-}j,n]g[k-j] = \begin{cases} g^i \prod_{j=1}^{i} h_{12}[k{+}1{-}j,n], & i < k-1; \\ g[1]g^{i-1}\prod_{j=1}^{i} h_{12}[k{+}1{-}j,n], & i = k-1. \end{cases} \tag{7}$$

The ergodic rate in frame $k+1$ measured in b/s/Hz follows as

$$R[k+1] = \begin{cases} \mathbb{E}\log\left(1 + \frac{P_{\mathrm{s}}|h_{13}g[1]h_{01}|^2}{\sigma_{\mathrm{d}}^2 + |h_{13}g[1]|^2\sigma_{\mathrm{r}}^2}\right), & k = 1; \\ \mathbb{E}\log\left(1 + \frac{P_{\mathrm{s}}|h_{p3}gh_{0p}|^2}{\sigma_{\mathrm{d}}^2 + |h_{p3}g|^2\left(P_{\mathrm{s}}r_1[k] + \sigma_{\mathrm{r}}^2 r_2[k]\right)}\right), & k = 2, 3, \ldots, K \end{cases} \tag{8}$$

with $r_1[k] = \sum_{i=m+1}^{k-1}|h_{0q}|^2|f_i[k]|^2$ and $r_2[k] = \sum_{i=0}^{k-1}|f_i[k]|^2$. Note that $f_i[k]$ models the inter-relay interference factor as random variable whereas $f_i[k,n]$ denotes its realization[3]. Clearly, $R[1] = 0$ because after transmission of the first frame no signal is received by the destination yet. The expectations are taken with respect to the statistics of h_{0p}, h_{p3} for $p \in \{1, 2\}$ and h_{12} and depend on the channel model that is used for the fading variables. It can be shown that $R[k+1]$ is a non-increasing sequence when choosing $g[1] = g$. After transmission of a sequence of K messages we get the average rate

$$\overline{R}_K = \frac{1}{K+1}\sum_{k=1}^{K} R[k+1] \geq \frac{K}{K+1}R[K+1] \geq \frac{K}{K+1}\lim_{k\to\infty} R[k] \tag{9}$$

where the pre-log $\frac{K}{K+1} \approx 1$ for large K. The disadvantage of signaling according to (8) is that the source has to adapt the rate for each frame. However, the lower bounds in (9) suggest to use a fixed-rate scheme at the source, either $R[K{+}1]$ or $\lim_{k\to\infty} R[k]$. By using $\lim_{k\to\infty} R[k]$ the rate is independent of the number of messages K to be transmitted. In order to simplify the computation of $\lim_{k\to\infty} R[k]$ we lower bound the rate (8). For $k = 2, 3, \ldots, K$ it is

$$R[k+1] = \mathbb{E}\log\left(\frac{\sigma_{\mathrm{d}}^2 + |h_{p3}g|^2\big(P_{\mathrm{s}}r_1[k] + \sigma_{\mathrm{r}}^2 r_2[k]\big) + P_{\mathrm{s}}|h_{p3}gh_{0p}|^2}{\sigma_{\mathrm{d}}^2 + |h_{p3}g|^2\big(P_{\mathrm{s}}r_1[k] + \sigma_{\mathrm{r}}^2 r_2[k]\big)}\right) \tag{10}$$

$$\geq \mathbb{E}\log\left(\frac{\sigma_{\mathrm{d}}^2 + |h_{p3}g|^2\big(P_{\mathrm{s}}r_{1,k'}[k] + \sigma_{\mathrm{r}}^2 r_{2,k'}[k]\big) + P_{\mathrm{s}}|h_{p3}gh_{0p}|^2}{\sigma_{\mathrm{d}}^2 + |h_{p3}g|^2\big(P_{\mathrm{s}}r_1[k] + \sigma_{\mathrm{r}}^2 r_2[k]\big)}\right) \tag{11}$$

$$\geq \mathbb{E}\log\Big(P_{\mathrm{s}}|h_{p3}gh_{0p}|^2 + \sigma_{\mathrm{d}}^2 + |h_{p3}g|^2\big(P_{\mathrm{s}}r_{1,k'}[k] + \sigma_{\mathrm{r}}^2 r_{2,k'}[k]\big)\Big)$$

$$-\log\Big(\sigma_{\mathrm{d}}^2 + \nu_2^2 g^2\big(P_{\mathrm{s}}\overline{r}_1[k] + \sigma_{\mathrm{r}}^2\overline{r}_2[k]\big)\Big) = R_{\mathrm{low}}[k+1] \tag{12}$$

[3]The same is true for the channels: h_{ij} is the random variable and $h_{ij}[k,n]$ its realization

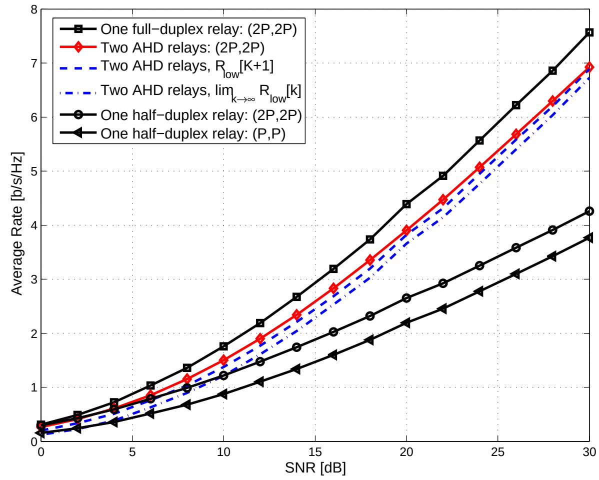
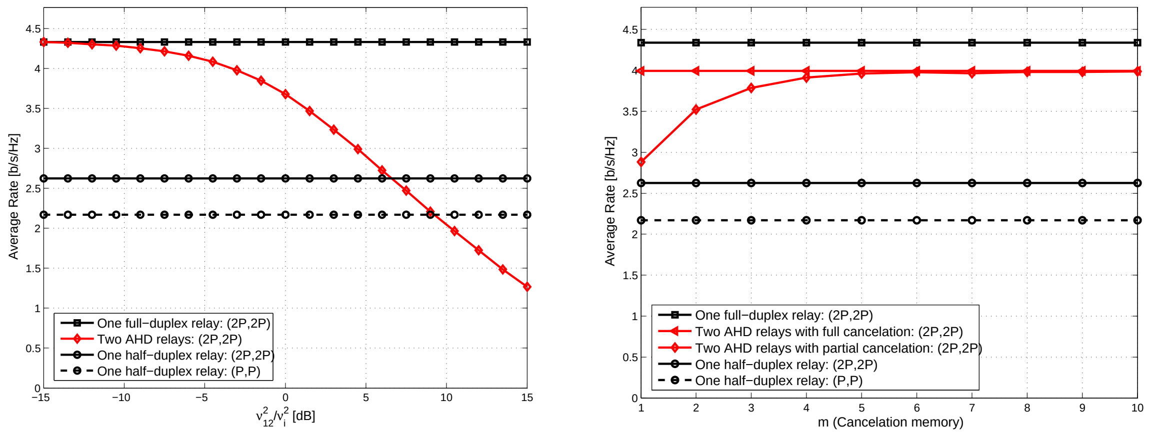


Figure 1: Average rates vs. signal-to-noise ratio for $\nu_1^2 = 1$, $\nu_2^2 = 1$, $\nu_{12}^2 = 0.5$, $K = 50$ and full interference cancelation

with $\overline{r}_1[k] = \sum_{i=m+1}^{k-1} \mathbb{E}|f_i[k]|^2 = \frac{q^{m+1}-q^k}{1-q}$, $\overline{r}_2[k] = \sum_{i=0}^{k-1} \mathbb{E}|f_i[k]|^2 = \frac{1-q^k}{1-q}$ and $q = g^2\nu_{12}^2$. The first inequality (11) follows due to $r_1[k] \geq r_{1,k'}[k] = \sum_{i=m+1}^{k'} |h_{0q}|^2|f_i[k]|^2$ and $r_2[k] \geq r_{2,k'}[k] = \sum_{i=0}^{k'} |f_i[k]|^2$ for $k' < k-1$. The second inequality (12) follows by using Jensen's inequality. For $k \to \infty$ and $q < 1$ we get $\lim_{k\to\infty} \overline{r}_1[k] = \frac{q^{m+1}}{1-q}$ and $\lim_{k\to\infty} \overline{r}_2[k] = \frac{1}{1-q}$ and the lower bound (12) becomes independent of the actual frame number k. Note that for a stationary inter-relay channel h_{12} the statistics of $r_{1,k'}[k]$ and $r_{2,k'}[k]$ become independent of k. Numerical results in the next section show that fixed-rate signaling according to $\lim_{k\to\infty} R_{\text{low}}[k]$ or $R_{\text{low}}[K+1]$ in each frame induces only a small loss compared to variable-rate signaling according to (8), but has the advantage that the source does not have to adapt the rate for each frame. The parameter k' can be used to improve the lower bound: the larger k' the better the lower bound, but the more involved becomes the evaluation of the expectation of the first log-term in (12).

2.1 Numerical Examples

We evaluate the average rates of this relaying scheme by Monte Carlo simulations. The channel gains are distributed i.i.d. (in space and time) complex normal with zero mean and channel variances ν_{12}^2, ν_1^2 and ν_2^2. We choose for all examples $\nu_1^2 = \nu_2^2 = 1$. The additive white Gaussian noise (AWGN) variances are chosen as $\sigma_{\text{r}}^2 = \sigma_{\text{d}}^2 = \sigma^2$ and the transmit powers $P_{\text{s}} = P_{\text{r}} = P$. We simulated 5000 random channels for each SNR value where $\text{SNR} = \frac{P}{\sigma^2}$. In Fig.1 we see that for $\nu_{12}^2 = 0.5$ (inter-relay channel is 3dB weaker than the other channels) the relaying protocol with two alternating half-duplex (AHD) relays and full cancelation of the inter-relay interference achieves an average rate $\overline{R}_K$ that is near to the rate of one full-duplex relay and outperforms clearly the case where only one half-duplex relay is used. The notation $(2P, 2P)$ means that in the first frame slot the total transmit power of the network is $2P$ and in the second frame slot $2P$ (every node transmits with power P). For the lower bound (12) we have chosen $k' = 0$ and $r_1[k] = 0$ (full interference cancelation). We observe from Fig.1 that the performance loss of the fixed-rate schemes based on $\lim_{k\to\infty} R_{\text{low}}[k]$ or $R_{\text{low}}[K+1]$ is small compared to the performance of the variable-rate scheme (8). Fig.2(a) shows the achievable rate (8) for different variances ν_{12}^2 of the inter-relay channel. When the inter-relay channel



(a) Average rate vs. relative strength of the inter-relay channel (with respect to source-relays and relays-destination channel variances $\nu_1^2 = \nu_2^2 = 1$) for SNR=20dB

(b) Full vs. partial interference cancelation with $\nu_{12}^2 = 0.5$ and SNR = 20dB

Figure 2: Average rates of the alternating relays protocol

is not too strong, the AHD scheme performs very well. For inter-relay channels that are considerably stronger than the source-relay and relay-destination channels the AHD scheme does not work well due to the accumulated noise interference of the two relays at the destination. In Fig.2(b) we compare the impact of full and partial interference cancelation on the average rate, cf. (4). We see that after cancelation of about five to six previously transmitted codewords the performance is the same as with full cancelation of the inter-frame interference. The inter-relay channel acts as forgetting factor (2) and lessens the inter-frame interference caused by the previously transmitted codewords.

3 Bidirectional Amplify-and-Forward Relaying

In the relaying scheme described in the previous section we needed two relays to circumvent the pre-log factor $\frac{1}{2}$ in the achievable rate. However, in some cases only one relay may be available to assist the communication between source and destination. Another solution arises, when we assume that two nodes N_1 and N_2 want to establish a synchronous bidirectional connection (for example two wireless routers), i.e., both nodes communicate in both directions through a common half-duplex relay R.

3.1 One Node Pair Assisted by One Relay

The proposed relaying scheme works as follows: in time slot k both nodes N_1 and N_2 transmit their symbols to relay R in the same time slot and the same bandwidth. The relay scales the received signal in order to meet its average power constraint and retransmits the signal in the next time slot. The received signal at node N_i, $i = 1, 2$, in time slot $k + 1$ is[4]

$$y_i[k{+}1] = h_i[k{+}1]g[k]h_j[k] \cdot x_j[k] + h_i[k{+}1]g[k]h_i[k] \cdot x_i[k] + h_i[k{+}1]g[k] \cdot n_{\mathrm{r}}[k] + n_i[k{+}1] \quad (13)$$

[4] k denotes here discrete symbol time

where $i = 2, j = 1$ for $\mathrm{N}_1 \rightarrow \mathrm{R} \rightarrow \mathrm{N}_2$ and $i = 1, j = 2$ for $\mathrm{N}_1 \leftarrow \mathrm{R} \leftarrow \mathrm{N}_2$[5]. The i.i.d. symbols $x_1[k] \sim \mathcal{CN}(0, P_1)$ and $x_2[k] \sim \mathcal{CN}(0, P_2)$ are the transmit symbols of node N_1 and N_2, respectively. h_1 is the channel gain between N_1 and relay R and h_2 the channel gain between N_2 and relay R[6]. AWGN at the relay is denoted by $n_\mathrm{r} \sim \mathcal{CN}(0, \sigma_\mathrm{r}^2)$ and $n_i \sim \mathcal{CN}(0, \sigma_i^2)$ stands for AWGN at node N_i. The relay gain is chosen as $g[k] = \left(P_\mathrm{r}/\left(P_1|h_1[k]|^2 + P_2|h_2[k]|^2 + \sigma_\mathrm{r}^2\right)\right)^{\frac{1}{2}}$ where P_r is the average transmit power of relay R. Since nodes N_1 and N_2 know their own transmitted signals they can subtract their contribution in (13) prior to decoding, assuming perfect knowledge of the corresponding channel coefficients. The ergodic sum rate is then given by

$$R_\mathrm{sum}^\mathrm{AF} = \frac{1}{2}\mathbb{E}\log\left(1 + \frac{P_1|h_2 g h_1|^2}{\sigma_2^2 + \sigma_\mathrm{r}^2|h_2 g|^2}\right) + \frac{1}{2}\mathbb{E}\log\left(1 + \frac{P_2|h_1 g h_2|^2}{\sigma_1^2 + \sigma_\mathrm{r}^2|h_1 g|^2}\right) \quad (14)$$

i.e., each connection suffers still from the pre-log factor $\frac{1}{2}$. However, the half-duplex constraint can here be exploited to establish a bidirectional connection between two nodes and to increase the sum rate of the network. For the case of equal transmit powers, equal noise variances and equal fading variances the sum rate (14) is equal to the rate of the $\mathrm{N}_1 \rightarrow \mathrm{R} \rightarrow \mathrm{N}_2$ link when R is a full-duplex AF relay. This idea may also be applied to regenerative relaying such as repetition-based decode-and-forward or superposition-based block Markov encoding with backward decoding [8].

3.2 Multiple Node Pairs Assisted by Multiple Relays

We apply the idea described in the previous section to a network with N two-hop communication links and K relays assisting in the communication. Each communication link has data traffic in both directions, whereas relays have no own data to transmit. The network is divided into three sets of nodes: nodes in $\mathcal{M}_1$ want to exchange messages with nodes in $\mathcal{M}_2$ and vice versa and the set $\mathcal{R}$ contains the relay nodes. The input-output relation for the $\mathcal{M}_1 \rightarrow \mathcal{R} \rightarrow \mathcal{M}_2$ communication is

$$\mathbf{y}_2 = \mathbf{H}_2\mathbf{G}\mathbf{H}_1\mathbf{x}_1 + \mathbf{H}_2\mathbf{G}\mathbf{H}_2^\mathrm{T}\mathbf{x}_2 + \mathbf{H}_2\mathbf{G}\mathbf{n}_\mathrm{r} + \mathbf{n}_2 \quad (15)$$

and for the $\mathcal{M}_1 \leftarrow \mathcal{R} \leftarrow \mathcal{M}_2$ communication

$$\mathbf{y}_1 = \mathbf{H}_1^\mathrm{T}\mathbf{G}\mathbf{H}_2^\mathrm{T}\mathbf{x}_2 + \mathbf{H}_1^\mathrm{T}\mathbf{G}\mathbf{H}_1\mathbf{x}_1 + \mathbf{H}_1^\mathrm{T}\mathbf{G}\mathbf{n}_\mathrm{r} + \mathbf{n}_1 \quad (16)$$

where $\mathbf{H}_1$ is the $K \times N$ channel matrix between the nodes in $\mathcal{M}_1$ and $\mathcal{R}$ and $\mathbf{H}_2$ the $N \times K$ channel matrix between the nodes in $\mathcal{R}$ and $\mathcal{M}_2$. The diagonal $K \times K$ matrix $\mathbf{G}$ contains the linear scaling coefficients of all K relays, i.e., the relays do not cooperate. The $N \times 1$ vectors $\mathbf{x}_1$ and $\mathbf{x}_2$ comprise the transmitted symbols from nodes in $\mathcal{M}_1$ and $\mathcal{M}_2$, respectively. Each element (symbol) of $\mathbf{x}_1$ and $\mathbf{x}_2$ is taken from a Gaussian codebook with average transmit power P and is transmitted in the first time slot[7]. The $N \times 1$ vectors $\mathbf{y}_1$ and $\mathbf{y}_2$ comprise the received symbols of nodes in $\mathcal{M}_1$ and $\mathcal{M}_2$ in the second time slot. The $K \times 1$ vector $\mathbf{n}_\mathrm{r}$ and $N \times 1$ vectors $\mathbf{n}_1$, $\mathbf{n}_2$ denote the zero mean AWGN in the relays and the nodes in $\mathcal{M}_1$ and $\mathcal{M}_2$, respectively. The noise variance in each receiving node of the network is assumed to be σ^2. In the following we discuss how

[5] $\mathrm{N}_i \rightarrow \mathrm{N}_j$ indicates information flow from node N_i to node N_j.

[6] Henceforth, we assume reciprocity for all channels.

[7] For simplicity, we omit here the time index in the formulas

to choose the diagonal gain matrix $\mathbf{G}$ such that the transmissions between nodes in $\mathcal{M}_1$ and $\mathcal{M}_2$ become interference-free in both communication directions.

Before continuing we first review the multi-user zero-forcing (ZF) relaying protocol introduced in [9] for unidirectional traffic between nodes in $\mathcal{M}_1$ and $\mathcal{M}_2$. The goal of the multi-user ZF relaying protocol is to choose the linear scaling coefficients (diagonal elements of $\mathbf{G}$) of the AF relays such that the transmissions from nodes in $\mathcal{M}_1$ to nodes in $\mathcal{M}_2$ become orthogonal. For this purpose the input-output relation of the unidirectional $\mathcal{M}_1 \to \mathcal{R} \to \mathcal{M}_2$ communication is rewritten as

$$\mathbf{y}_2 = \mathbf{H}_2\mathbf{G}\mathbf{H}_1\mathbf{x}_1 + \mathbf{H}_2\mathbf{G}\mathbf{n}_\mathrm{r} + \mathbf{n}_2. \tag{17}$$

We define the interference matrix $\mathbf{A}_\mathrm{uni}$ with dimensions $K \times N(N-1)$ to be

$$\mathbf{A}_\mathrm{uni} = \left[\mathbf{h}_1^{(1)} \odot \mathbf{h}_2^{(2)}, \ldots, \mathbf{h}_1^{(1)} \odot \mathbf{h}_q^{(2)}, \mathbf{h}_2^{(1)} \odot \mathbf{h}_1^{(2)}, \mathbf{h}_2^{(1)} \odot \mathbf{h}_3^{(2)}, \ldots, \mathbf{h}_2^{(1)} \odot \mathbf{h}_q^{(2)}, \ldots\right] \tag{18}$$

i.e., the columns of $\mathbf{A}_\mathrm{uni}$ are defined by $\mathbf{h}_p^{(1)} \odot \mathbf{h}_q^{(2)}$ $\forall\, p, q \in \{1, 2, \ldots, N\}$ and $p \neq q$, where $\mathbf{h}_p^{(1)}$ is the vector containing the channel gains from the pth node in $\mathcal{M}_1$ to all relays in $\mathcal{R}$ (i.e., the pth column of $\mathbf{H}_1$) and $\mathbf{h}_q^{(2)}$ is the vector containing the channel gains from the qth node in $\mathcal{M}_2$ to all relays in $\mathcal{R}$ (i.e., the qth column of $\mathbf{H}_2^\mathrm{T}$). The two-hop channel $\mathbf{H}_2\mathbf{G}\mathbf{H}_1$ becomes diagonal if $\mathbf{g} = \mathrm{diag}(\mathbf{G})$ lies in the null space of the interference matrix $\mathbf{A}_\mathrm{uni}$. Let $r = \mathrm{rk}(\mathbf{A}_\mathrm{uni}) = \min\{N^2 - N, K\}$ be the rank of the matrix $\mathbf{A}_\mathrm{uni}$ and define the singular value decomposition $\mathbf{A}_\mathrm{uni} = \mathbf{U}\mathbf{D}\left[\mathbf{V}^{(r)}\ \ \mathbf{V}^{(0)}\right]^\mathrm{H}$, where $\mathbf{V}^{(r)}$ contains the first r right singular vectors of $\mathbf{A}_\mathrm{uni}$ and $\mathbf{V}^{(0)}$ the last $K-r$ right singular vectors. The columns of $\mathbf{V}^{(0)}$ form an orthonormal basis for the null space of $\mathbf{A}_\mathrm{uni}$, i.e., $\mathbf{V}^{(0)} = \mathrm{null}(\mathbf{A}_\mathrm{uni})$. A sufficient condition for the null space to be non-empty is

$$K \geq N(N-1) + 1 \tag{19}$$

and we refer to it as *minimum relay configuration*. The zero-forcing gain vector $\mathbf{g}_\mathrm{ZF}$ is obtained by projecting any gain vector $\mathbf{g}$ onto the null space of $\mathbf{A}_\mathrm{uni}$. In [9] it was shown that for $K \to \infty$ the choice $\mathbf{g}_\infty = \mathrm{diag}\left(\mathbf{H}_2\mathbf{H}_1\right)$ diagonalises the two-hop channel $\mathbf{H}_2\mathbf{G}\mathbf{H}_1$. For finite number of relays $K < \infty$ we choose $\mathbf{g}_\mathrm{ZF} = c\mathbf{Z}\mathbf{Z}^\mathrm{H}\mathbf{g}_\infty$ where $\mathbf{Z} = \mathbf{V}^{(0)}$ and c is chosen such that the average power constraint of the relays is met. Using $\mathbf{g}_\mathrm{ZF}$ the product channel $\mathbf{H}_2\mathbf{G}\mathbf{H}_1$ becomes diagonal and the transmissions between the different source-destination pairs are interference-free. However, the noise terms at the destination nodes in $\mathcal{M}_2$ are still correlated.

We now extend the multiuser ZF relaying to the bidirectional traffic pattern. Comparing (15) with (17), we see that a receiving node in $\mathcal{M}_2$ suffers additionally from interference caused by its neighbor nodes located in the same set. Each node knows only its own symbol (which was transmitted by the node in the first time slot) and can subtract this contribution from the received signal in the second time slot. The symbols transmitted by the neighbor nodes are unknown and cannot be subtracted. The gain matrix $\mathbf{G}$ has therefore to be chosen such that the channels $\mathbf{H}_2\mathbf{G}\mathbf{H}_2^\mathrm{T}$ and $\mathbf{H}_1^\mathrm{T}\mathbf{G}\mathbf{H}_1$ become diagonal too. For this purpose the interference matrix (18) has to be extended to a $K \times 2N(N-1)$ dimensional matrix $\mathbf{A}_\mathrm{bi} = [\mathbf{A}_\mathrm{uni}, \mathbf{A}_1, \mathbf{A}_2]$ with $\mathbf{A}_1 = \left[\mathbf{h}_1^{(1)} \odot \mathbf{h}_2^{(1)}, \ldots, \mathbf{h}_1^{(1)} \odot \mathbf{h}_N^{(1)}, \mathbf{h}_2^{(1)} \odot \mathbf{h}_3^{(1)}, \ldots, \mathbf{h}_2^{(1)} \odot \mathbf{h}_N^{(1)}, \ldots, \mathbf{h}_{N-1}^{(1)} \odot \mathbf{h}_N^{(1)}\right]$ an $K \times \frac{N(N-1)}{2}$ dimensional matrix and $\mathbf{A}_2$ being defined correspondingly. The minimum relay configuration changes to

$$K \geq 2N(N-1) + 1. \tag{20}$$

The network's sum-rate follows as

$$R^{\mathrm{ZF}}_{\mathrm{sum,sep}} = \frac{1}{2}\mathbb{E}\log\det\left(\mathbf{I}_N + P\mathbf{D}_2^{-1}\mathbf{H}_{12}\mathbf{H}_{12}^{\mathrm{H}}\right) + \frac{1}{2}\mathbb{E}\log\det\left(\mathbf{I}_N + P\mathbf{D}_1^{-1}\mathbf{H}_{21}\mathbf{H}_{21}^{\mathrm{H}}\right) \quad (21)$$

where $\mathbf{H}_{12} = \mathbf{H}_2\mathbf{G}_{\mathrm{ZF}}\mathbf{H}_1$ is the equivalent (diagonal) channel from "left to right" and $\mathbf{H}_{21} = \mathbf{H}_1^{\mathrm{T}}\mathbf{G}_{\mathrm{ZF}}\mathbf{H}_2^{\mathrm{T}}$ the equivalent (diagonal) channel from "right to left". The zero-forcing gain matrix $\mathbf{G}_{\mathrm{ZF}}$ contains the elements of $\mathbf{g}_{\mathrm{ZF}}$, which is the projection of $\mathbf{g}_\infty$ on the nullspace of $\mathbf{A}_{\mathrm{bi}}$. The diagonal matrices $\mathbf{D}_1$ and $\mathbf{D}_2$ contain the diagonal elements of the noise covariance matrices $\mathbf{R}_1 = \sigma^2\left(\mathbf{H}_2\mathbf{G}_{\mathrm{ZF}}\mathbf{G}_{\mathrm{ZF}}^{\mathrm{H}}\mathbf{H}_2^{\mathrm{H}} + \mathbf{I}_N\right)$ and $\mathbf{R}_2 = \sigma^2\left(\mathbf{H}_1^{\mathrm{T}}\mathbf{G}_{\mathrm{ZF}}\mathbf{G}_{\mathrm{ZF}}^{\mathrm{H}}\mathbf{H}_1^{*} + \mathbf{I}_N\right)$, respectively. Since the noise samples at the nodes in $\mathcal{M}_1$ and $\mathcal{M}_2$ are correlated we look also at the case of cooperative nodes in $\mathcal{M}_1$ and $\mathcal{M}_2$ (i.e., two users each with multiple colocated antennas) with joint decoding of data streams. The achievable rate then is

$$R^{\mathrm{ZF}}_{\mathrm{sum,joint}} = \frac{1}{2}\mathbb{E}\log\det\left(\mathbf{I}_N + P\mathbf{R}_2^{-1}\mathbf{H}_{12}\mathbf{H}_{12}^{\mathrm{H}}\right) + \frac{1}{2}\mathbb{E}\log\det\left(\mathbf{I}_N + P\mathbf{R}_1^{-1}\mathbf{H}_{21}\mathbf{H}_{21}^{\mathrm{H}}\right) \quad (22)$$

where $\mathbf{g}_{\mathrm{ZF}}$ is the projection of $\mathbf{g}_\infty$ on the nullspace of $\mathbf{A}_{\mathrm{uni}}$, since the back-propagating self-interference $\mathbf{x}_2$ and $\mathbf{x}_1$ in (15) and (16) may be canceled at the corresponding multi-antenna receivers. Hence, for multiple antenna nodes the minimum relay configuration remains unaltered for the bidirectional ZF relaying scheme. Clearly, $R^{\mathrm{ZF}}_{\mathrm{sum,joint}} \geq R^{\mathrm{ZF}}_{\mathrm{sum,sep}}$, since the noise samples are correlated at the different users, but the rate loss due to separate decoding is small as the numerical examples in the next section show.

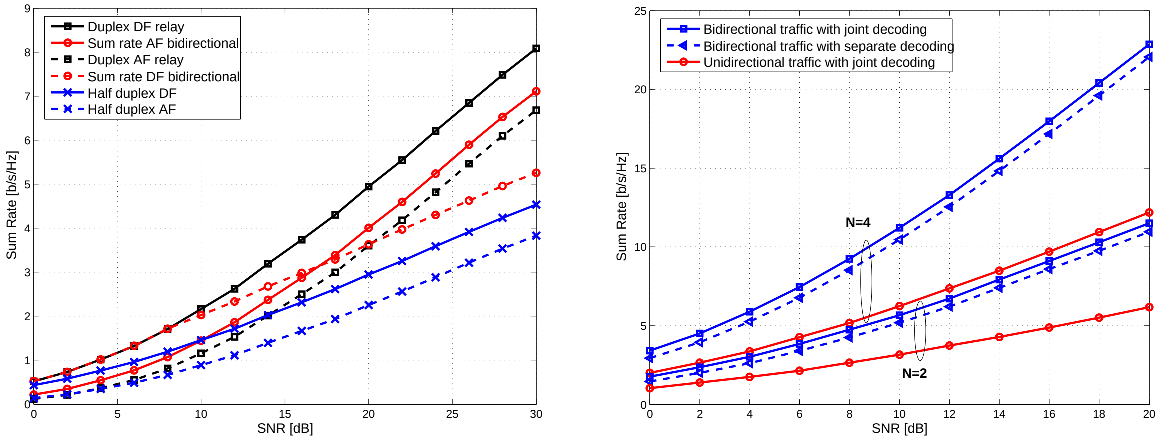
3.3 Numerical Examples

We evaluate the sum rates of this relaying scheme by Monte Carlo simulations. For the numerical examples all channel gains are distributed i.i.d. $\mathcal{CN}(0,1)$ (in space and time). Each node (relays as well as sources) consumes a transmit energy $P \cdot 2T$ over two time slots and AWGN is distributed i.i.d. $\mathcal{CN}(0,\sigma^2)$. We simulated 5000 random channels for each SNR value where SNR $= \frac{P}{\sigma^2}$. **One Node Pair:** In Fig.3(a) we compare the ergodic sum rate of the bidirectional AF relaying protocol with a bidirectional repetition-based decode-and-forward (DF) relaying protocol where the sum rate is given by[8]

$$R^{\mathrm{DF}}_{\mathrm{sum}} = \frac{1}{2}\min\left\{\mathbb{E}\log\left(1+P\left(|h_1|^2+|h_2|^2\right)\right), \mathbb{E}\log\left(1+\frac{P|h_1|^2}{2}\right)+\mathbb{E}\log\left(1+\frac{P|h_2|^2}{2}\right)\right\}. \quad (23)$$

We see that the AF scheme outperforms the DF scheme in the high SNR region and is inferior in the low and mid SNR region. The reason is that the sum rate for the multiple access channel in the first time slot of the DF scheme dominates the minimum in (23) for high SNR. We compare these schemes with the rates achievable for unidirectional relaying with all combinations of half-duplex/duplex and AF/DF. We observe that the AF bidirectional protocol outperforms the rate of the unidirectional duplex AF scheme. This is due to a higher total network power consumed by the bidirectional scheme. When normalizing to the same network power, teh rates of both schemes coincide. **Multiple Node Pairs:** In Fig.3(b) we compare the sum rates of unidirectional and bidirectional ZF relaying. We see that the loss due to separate decoding is small, since the transmissions are orthogonalized and neglecting the noise correlations at the receiving nodes does not degrade the performance significantly. We observe that bidirectional ZF relaying with $N = 2$ node pairs and $M = 5$ relays (minimum relay configuration) achieves almost the same sum rate as unidirectional relaying with $N = 4$ node pairs and $M = 13$ relays.

[8] Bidirectional protocols for regenerative relaying are elaborated in [8] in more detail.



(a) Bidirectional relaying for one node pair assisted by one AF or DF relay

(b) Bidirectional relaying for multiple node pairs assisted by several AF relays

Figure 3: Sum rates for bidirectional relaying

4 Conclusions

We studied two half-duplex nonregenerative relaying protocols with increased spectral efficiency compared to conventional half-duplex relay protocols. For the first scheme (two alternating relays) it was shown that when the inter-relay channel is not to strong a large portion of the half-duplex loss can be recovered. In the second relaying protocol we established a bidirectional connection between two nodes using one half-duplex relay. It was shown that the sum rate of the AF scheme achieves the rate of the duplex AF relay channel. This scheme was then extended to the case of multiple node pairs assisted by multiple zero-forcing AF relays.

References

[1] H. Bölcskei, R. U. Nabar, O. Oyman, and A. J. Paulraj, “Capacity scaling laws in MIMO relay networks,” *IEEE Trans. Wireless Commun.*, Apr. 2004. Submitted.

[2] J. N. Laneman, D. N. Tse, and G. W. Wornell, “Cooperative diversity in wireless networks: Efficient protocols and outage behavior,” *IEEE Trans. Inform. Theory*, vol. 50, pp. 3062–3080, Dec. 2004.

[3] T. M. Cover and A. A. El Gamal, “Capacity theorems for the relay channel,” *IEEE Trans. Inform. Theory*, vol. 25, pp. 572–584, Sept. 1979.

[4] O. Munoz, A. Augustin, and J. Vidal, “Cellular capacity gains of cooperative MIMO transmission in the downlink,” in *Proc. IZS*, (Zurich, Switzerland), pp. 22–26, Feb. 2004.

[5] A. Wittneben and B. Rankov, “Impact of cooperative relays on the capacity of rank-deficient MIMO channels,” in *Proc. Mobile and Wireless Communications Summit (IST)*, (Aveiro (Portugal)), pp. 421–425, 2003.

[6] T. J. Oechtering and A. Sezgin, “A new cooperative transmission scheme using the space-time delay code,” in *Proc. ITG Workshop on Smart Antennas*, (Munich), Mar. 2004.

[7] A. Wittneben, “A new bandwidth efficient transmit antenna modulation diversity scheme for linear digital modulation,” in *Proc. ICC*, vol. 3, pp. 1630–1634, 1993.

[8] B. Rankov and A. Wittneben, “Spectral efficient signaling for half-duplex relay channels,” in *Proc. Asilomar Conf. Signals, Syst., Comput.*, (Pacific Grove, CA), Nov. 2005.

[9] A. Wittneben and B. Rankov, “Distributed antenna systems and linear relaying for gigabit MIMO wireless,” in *IEEE Veh. Tech. Conf.*, (Los Angeles, LA), pp. 3624–3630, Sept. 2004.