Communication Protocols for N-way All-Cast Relay Networks

Tao Cui, Jörg Kliewer*, *Senior Member, IEEE*, and Tracey Ho, *Senior Member, IEEE*

Abstract

We consider communication protocols for N-way all-cast relay networks, which comprise N source terminals such that each source terminal demands messages from all other source terminals with the help of a relay. The derived protocols are characterized by the fact that physical layer network coding is employed at the relay, where each source has side information about the signals it has sent. Amplify-and-forward (AF) and decode-and-forward (DF) protocols are applied to the N-way relay network setting, where the achievable rate regions for those protocols are derived and compared with outer capacity bounds. We propose several practical space-time coding schemes for AF and DF, and introduce two new protocols denoted as denoise-and-forward (DNF) and estimate-and-forward (EF). Further, for AF and DF the fundamental diversity-multiplexing trade-off is characterized.

Index Terms

Relay channels, N-way traffic, cooperation, wireless relay networks.

This work has been supported in part by subcontract #069144 issued by BAE Systems National Security Solutions, Inc. and supported by the Defense Advanced Research Projects Agency (DARPA) and the Space and Naval Warfare System Center (SPAWARSYSCEN), San Diego under Contracts No. N66001-08-C-2013 and W911NF-07-1-0029, by NSF grants CCF-0830666 and CCF-1017632, and by Caltech's Lee Center for Advanced Networking. This paper has been presented in part at the IEEE Global Communications Conference, Dec. 2008, New Orleans, LA, USA.

Tao Cui and Tracey Ho are with the Department of Electrical Engineering, California Institute of Technology, M/C 136-93, Pasadena, CA 91125, USA (Email: {taocui,tho}@caltech.edu).

Jörg Kliewer (* corresponding author) is with the Klipsch School of Electrical and Computer Engineering, New Mexico State University, MSC 3-O, Las Cruces, NM 88003, USA (Email: jkliewer@nmsu.edu, phone: (575) 646-7216, fax: (575) 646-1435).

I. INTRODUCTION

The N-way relay network (NWRN) is a promising wireless network architecture for applications such as wireless teleconferencing, gossip and sensor networks, where there are N source terminals in the network, and each source terminal demands messages from all the other source terminals with the help of a relay node. When $N = 2$, the N-way relay network degenerates to a two-way relay network. Recently, the two-way relay channel (TWRC) has drawn renewed interest from both academic and industrial communities (see, e.g., [1–5]). In [1], AF and DF protocols for one-way relay channels are extended to the half duplex Gaussian TWRC. In [2], network coding [6] is used to increase the sum-rate of two users. With network coding, each node in a network is allowed to perform algebraic operations on received packets instead of only forwarding or replicating received packets. The work in [3], [4] addresses physical layer network coding at the relay in which network coding is generalized from an operation over a finite field to a mapping over the real numbers. In particular, for TWRCs with a single memoryless relay the performance analysis and the design of the relay function are discussed in [3]. Further, in [4], distributed space time coding for TWRC is proposed with multiple relays where amplify-and-forward (AF) decode-and-forward (DF) and denoise and forward protocols are considered. The work in [5] focuses on the broadcast phase of a half duplex TWRC and proves an achievable rate region for a decode and forward scenario at the relay. Further, the capacity region of a broadcast channel where some messages are known *a priori* has been studied in [7]. Another related work is the multiple access relay channel (MARC) [8], [9], where multiple sources communicate with a single destination in the presence of a relay node. Relay strategies for MARC and related capacity theorems are discussed in [8], [9]. On the contrary, all source terminals also serve as destinations in NWRNs. In this sense, we also consider an NWRN as an *all-cast relay network*.

In this paper, we first consider AF and DF protocols, where, in contrast to previous work in [3], [4], achievable rate regions for these schemes are derived. Further, we propose several practical space-time coding protocols for both non-regenerative and regenerative relays, where we address protocols using $K + 1$ time slots. All the source terminals transmit to the relay simultaneously in the first time slot, and after processing the received signal the relay transmits to the source terminals in the remaining K time slots. Different from [4], [10], where coding is performed over symbols received at different time, in the proposed strategies space time coding is carried out over symbols received by different antennas at the relay. We employ linear dispersion codes [11] for coding which can be optimized by maximizing the sum capacity and which provide a general framework that subsumes most of the existing linear space-time codes. Since each source terminal already knows the signal it has transmitted to the relay, the received signal at the relay can be compressed by using physical layer network coding to reduce bandwidth usage and power consumption. A simple space time code is presented where at any one out of the K time slots at most one antenna is active to transmit the signal.

Further, we characterize the fundamental diversity-multiplexing tradeoff (DMT) [12], [13] of AF and DF in the NWRN setting, while DMTs for MIMO, multiple access and cooperative systems are characterized in [12], [13]. In particular, we show that the diversity-multiplexing tradeoff performance can be divided into a lightly loaded and a heavily loaded regime. It is observed that AF performs better than DF in a heavily loaded system while DF is preferable in a lightly loaded system. Similar observations have been made for the MARC in [14], [15]. By maximizing the diversity-multiplexing tradeoff we find that for DF the optimal uplink and downlink time sharing ratio is $1/(N-1)$.

II. System Model

We consider an N-way relay network with N source terminals and a single relay node. There are M antennas on the relay node while each source terminal only has a single antenna. Each source terminal wishes to broadcast its data to all other source terminals and also wishes to decode all other terminals' signals. We consider a half duplex system where nodes cannot simultaneously transmit and receive. We further assume that there does not exist a direct transmission between source terminals, which may be the case e.g. if the source terminals are far away from each other. Thus, the source terminals can only exchange data through the relay node.

We denote by α the fraction of time in which the relay is receiving. Thus, for a transmission scheme operating over a time interval T, period αT is used for uplink transmission from the source terminals to the relay and $(1-\alpha)T$ is spent for downlink transmission from the relay to the source terminals.

Let $x_{i,t}$ and $y_{i,t}$ be the signal sent and received by source terminal i at time t, respectively, and let $x_{r,t}^m$ and $y_{r,t}^m$ be the signal sent and received by the m-th antenna of the relay node at time t, respectively. We thus have the following model:

$$
\begin{aligned}
y_{r,t}^{m} &= \sum_{i=1}^{N} h_i^m x_{i,t} + w_{r,t}^{m}, \quad t = 1, \ldots, \lfloor \alpha T \rfloor, \quad \forall m = 1, \ldots, M, \\
y_{i,t} &= \sum_{m=1}^{M} h_i^m x_{r,t}^{m} + w_{i,t}, \quad t = \lfloor \alpha T \rfloor + 1, \ldots, T, \quad \forall i = 1, \ldots, N,
\end{aligned} \tag{1}
$$

where h_i^m is the channel gain between source i and the m-th antenna of the relay, and $w_{r,t}^m$ and $w_{i,t}$ are complex additive white Gaussian noise (AWGN) samples with distribution $\mathcal{CN}(0,1)$. Note that due to the half duplex assumption, the second equation in (1) does not contain signals from other source terminals. In addition, $x_{i,t}$ and $x_{r,t}$ are subject to the average power constraints

$$
\frac{1}{T} \sum_{t=1}^{\lfloor \alpha T \rfloor} |x_{i,t}|^2 \leq P_i, \ \forall i = 1, \ldots, N, \quad \text{and} \ \frac{1}{T} \sum_{m=1}^{M} \sum_{t=\lfloor \alpha T \rfloor + 1}^{T} |x_{r,t}^{m}|^2 \leq P_r. \tag{2}
$$

We assume reciprocal channels for simplicity. By replacing h_i^m in the second equation in (1) by a different variable g_i^m, most of the results in this paper can be extended to the general case when uplink and downlink channels are different, e.g., when frequency division multiplexing is used. Throughout this paper, we assume that the h_i^m's are independent and are known perfectly at all nodes for simplicity and that they do not change during the interval T. The channel can be estimated via pilot symbols.

III. Basic Protocols and Achievable Rate Regions

In this section, we consider several relay strategies for the N-way relay channel and study their achievable rate regions. In the following, we denote the transmission rate of source terminal i by R_i. The key feature of N-way relay channels is that each source terminal already knows its transmitted signal as side information, which may potentially improve the transmission rate.

A. *Amplify-and-Forward*

The AF strategy is simple and only requires minimal processing at the relay.

i) We first consider the simple case where $\alpha = \frac{1}{2}$. The relay precodes the received signals from all the antennas using a precoding matrix $\Psi = [\psi^{m,n}]$. The transmitted signal from antenna m is

$$
x_{r,t+\frac{T}{2}}^{m} = \sum_{n=1}^{M} \psi^{m,n} y_{r,t}^{n}, \quad m = 1, \ldots, M, \quad \sum_{i=1}^{N} \sum_{m=1}^{M} \left(2 \left| \sum_{n=1}^{M} \psi^{m,n} h_i^n \right|^2 P_i + \sum_{n=1}^{M} |\psi^{m,n}|^2 \right) \leq 2P_r. \tag{3}
$$

The received signal at source node i at time $t+\frac{T}{2}$ is

$$y_{i,t+\frac{T}{2}} = \sum_{m=1}^{M} h_i^m x_{r,t+\frac{T}{2}}^m + w_{i,t+\frac{T}{2}} = \sum_{j=1}^{N} \left(\sum_{m=1}^{M} \sum_{n=1}^{M} \psi^{m,n} h_i^m h_j^n \right) x_{j,t} + \sum_{m=1}^{M} h_i^m \sum_{n=1}^{M} \psi^{m,n} w_{r,t}^n + w_{i,t+\frac{T}{2}}. \quad (4)$$

By canceling the contribution of $x_{i,t}$ from $y_{i,t+\frac{T}{2}}$ we obtain a multiple access channel with $N-1$ users. By applying [16, Theorem 15.3.6] to (4), the achievable rate region is given by the following proposition.

Proposition 1: The achievable rate region for the half duplex N-way relay network by using AF strategy i) is the convex hull of

$$\bigcap_{i=1}^{N} \bigcap_{\mathcal{S} \subseteq \mathcal{S}_i} \left\{ (R_1, \ldots, R_N) \middle| \sum_{j \in \mathcal{S}}^{N} R_j \leq \frac{1}{2} \log \left(1 + \frac{2 \sum_{j \in \mathcal{S}} \left| \sum_{m=1}^{M} \sum_{n=1}^{M} \psi^{m,n} h_i^m h_j^n \right|^2 P_j}{\sum_{n=1}^{M} \left| \sum_{m=1}^{M} h_i^m \psi^{m,n} \right|^2 + 1} \right) \right\}, \quad (5)$$

where $\mathcal{S}_i = \{1, \ldots, N\}/i$ and $/$ denotes set difference.

ii) We next consider the general case where α is an optimization parameter. After receiving signals from the source terminals in the interval αT, the relay transmits on its m-th antenna at any time during the interval $(1-\alpha)T$ the linear combination of all the signals received during time αT, i.e., at time $\tau > \lfloor \alpha T \rfloor$ the transmitted signal through the m-th antenna of the relay is

$$x_{r,\tau}^m = \sum_{t=1}^{\lfloor \alpha T \rfloor} \sum_{n=1}^{N} \psi_{\tau,t}^{m,n} y_{r,t}^n, \; m = 1, \ldots, M, \; \sum_{t=1}^{\lfloor \alpha T \rfloor} \sum_{i=1}^{N} \sum_{m=1}^{M} \left(\frac{\left| \sum_{n=1}^{M} \psi_{\tau,t}^{m,n} h_i^n \right|^2 P_i}{\alpha} + \sum_{n=1}^{M} |\psi_{\tau,t}^{m,n}|^2 \right) \leq \frac{P_r}{1-\alpha}. \quad (6)$$

The received signal at source node i at time τ becomes

$$y_{i,\tau} = \sum_{m=1}^{M} h_i^m x_{r,\tau}^m + w_{i,\tau} = \sum_{t=1}^{\lfloor \alpha T \rfloor} \sum_{j=1}^{N} \left(\sum_{m=1}^{M} \sum_{n=1}^{M} \psi_{\tau,t}^{m,n} h_i^m h_j^n \right) x_{j,t} + \sum_{t=1}^{\lfloor \alpha T \rfloor} \sum_{m=1}^{M} h_i^m \sum_{n=1}^{M} \psi_{\tau,t}^{m,n} w_{r,t}^n + w_{i,\tau}. \quad (7)$$

Therefore, the system is now equivalent to a multiple access MIMO system with $N-1$ users, where each user has $\lfloor \alpha T \rfloor$ transmit antennas and the receiver has $T - \lfloor \alpha T \rfloor$ receive antennas. Let $\mathbf{y}_i = [y_{i,\lfloor \alpha T \rfloor + 1}, \ldots, y_{i,T}]^T$, $\mathbf{x}_j = [x_{j,1}, \ldots, x_{j,\lfloor \alpha T \rfloor}]^T$, $\mathbf{H}_j = \left[\left(\sum_{m=1}^{M} \sum_{n=1}^{M} \psi_{\tau,t}^{m,n} h_i^m h_j^n \right) \right]$, $\mathbf{A}_i = \left[\sum_{m=1}^{M} h_i^m \psi_{\tau,t}^{m,n} \right]$, $\mathbf{w}_r = [w_{r,1}^1, \ldots, w_{r,1}^M, w_{r,2}^1, \ldots, w_{r,2}^M, \ldots, w_{r,\lfloor \alpha T \rfloor}^1, \ldots, w_{r,\lfloor \alpha T \rfloor}^M]^T$, and $\mathbf{w}_i = [w_{i,1}, \ldots, w_{i,\lfloor \alpha T \rfloor}]^T$. We can rewrite (7) in vector form as

$$\mathbf{y}_i = \sum_{j=1}^{N} \mathbf{H}_j \mathbf{x}_j + \mathbf{A}_i \mathbf{w}_r + \mathbf{w}_i. \quad (8)$$

By subtracting $\mathbf{H}_i \mathbf{x}_i$ and applying [16, Theorem 15.3.6], we obtain the following proposition.

Proposition 2: The achievable rate region for the half duplex N-way relay network by using AF strategy ii) is the convex hull of

$$\bigcap_{\mathcal{S}\subseteq\mathcal{S}_i}\left\{(R_1,\ldots,R_N)\left|\sum_{j\in\mathcal{S}}^{N}R_j\le\frac{1}{T}\log\det\left(\mathbf{I}_{(1-\alpha)T}+\sum_{j\in\mathcal{S}}\frac{P_j}{\alpha}\mathbf{H}_j\mathbf{H}_j^H\left(\mathbf{A}_i\mathbf{A}_i^H+\mathbf{I}_{(1-\alpha)T}\right)^{-1}\right)\right.\right\}. \tag{9}$$

We note that the AF strategy in i) is a special case of the one in ii) by choosing $\alpha=\frac{1}{2}$ and $\psi_{\tau,t}^{m,n}=0$, $\forall\tau\neq t+\frac{T}{2}$ and $\psi_{t+\frac{T}{2},t}^{m,n}\neq 0$ in (7), however, the general strategy in ii) provides more freedom for performance improvements.

B. Decode-and-Forward

By using decode-and-forward, the relay first decodes the signals from the source terminals during time interval αT. It then re-encodes the resulting signal and broadcasts the coded signal to all the source terminals in the remaining interval $(1-\alpha)T$. As we show experimentally in Section VI, the DF strategy is useful when the total transmission power is moderate. DF achieves a higher throughput than AF in this regime.

The uplink channel from the source terminals to the relay is a Gaussian multiple access channel. To decode the signals from the source terminals correctly at the relay, by using the capacity region in [16, Section 15.3.6] the source rates should satisfy

$$\sum_{j\in\mathcal{S}}R_j\le\alpha\log\det\left(\mathbf{I}_M+\sum_{j\in\mathcal{S}}\frac{P_j}{\alpha}\mathbf{h}_j\mathbf{h}_j^H\right),\ \forall\mathcal{S}\subseteq\{1,\ldots,N\}, \tag{10}$$

where $\mathbf{h}_j=[h_j^1,\ldots,h_j^M]^T$.

The downlink channel from the relay to the source terminals can be considered to be a broadcast channel with side information, where each source already knows the signal it has sent. We consider a general scenario where a source needs to communicate messages $W_1,\ldots,W_N$ to N receivers and receiver i already knows W_i. The messages $W_1,\ldots,W_N$ are encoded using a channel codebook with $2^{T\sum_{i=1}^{N}R_i}$ codewords according to a Gaussian distribution with zero mean and variance P_r, i.e., $X^{\lfloor(1-\alpha)T\rfloor}=x^{\lfloor(1-\alpha)T\rfloor}(w_1,\ldots,w_N)$, where $X^{\lfloor(1-\alpha)T\rfloor}$ denotes the corresponding vector random variable. Let Y_i denote the received signal at receiver i. Each source terminal i then finds the codeword $X^{\lfloor(1-\alpha)T\rfloor}(\hat{w}_1,\ldots,\hat{w}_{i-1},w_i,\hat{w}_{i+1},\ldots,\hat{w}_N)$ that is jointly typical with the received signal $Y_i^{\lfloor(1-\alpha)T\rfloor}$. At source i, the probability that a random codeword $X^{\lfloor(1-\alpha)T\rfloor}(\tilde{w}_1,\ldots,\tilde{w}_{i-1},w_i,\tilde{w}_{i+1},\ldots,\tilde{w}_N)\neq X^{\lfloor(1-\alpha)T\rfloor}(w_1,\ldots,w_{i-1},w_i,w_{i+1},\ldots,w_N)$ is jointly typical with Y_i is [16]

$$\Pr\left((X^{\lfloor(1-\alpha)T\rfloor}(\tilde{w}_1,\ldots,\tilde{w}_{i-1},w_i,\tilde{w}_{i+1},\ldots,\tilde{w}_N),Y_i)\in\mathcal{A}_\epsilon^{(\lfloor(1-\alpha)T\rfloor)}\right)\leq 2^{-\lfloor(1-\alpha)T\rfloor(I(Y_i;X_r)-3\epsilon)},\tag{11}$$

where $\mathcal{A}_\epsilon^{(\lfloor(1-\alpha)T\rfloor)}$ denotes the set of jointly typical sequences of length $\lfloor(1-\alpha)T\rfloor$. To find the error probability, let $P(E_0)$ denote the probability that the average power constraint is not satisfied, $P(E_s^c)$ denote the probability that $(X^{\lfloor(1-\alpha)T\rfloor}(w_1,\ldots,w_{i-1},w_i,w_{i+1},\ldots,w_N),Y_i)\notin\mathcal{A}_\epsilon^{(\lfloor(1-\alpha)T\rfloor)}$, and $P(E_{v_{-i}})$ denote the probability that $(X^{\lfloor(1-\alpha)T\rfloor}(\tilde{w}_1,\ldots,\tilde{w}_{i-1},w_i,\tilde{w}_{i+1},\ldots,\tilde{w}_N),Y_i)\in\mathcal{A}_\epsilon^{(\lfloor(1-\alpha)T\rfloor)}$ for all $v_{-i}=\{\tilde{w}_1,\ldots,\tilde{w}_{i-1},w_i,\tilde{w}_{i+1},\ldots,\tilde{w}_N\}$ with $\tilde{w}_j\neq w_j$, $j\neq i$. The average probability of error is

$$\begin{aligned}P_e^{(T)}&\leq P(E_0)+P(E_{S_{-i}}^c)+\sum_{v_{-i}}P(E_{v_{-i}})\\&\leq 2\epsilon+\left(2^{\lfloor(1-\alpha)T\rfloor\sum_{j=1,j\neq i}^N R_j}-1\right)2^{-\lfloor(1-\alpha)T\rfloor(I(Y_i;X_r)-3\epsilon)}\\&\leq 2\epsilon+2^{\lfloor(1-\alpha)T\rfloor\left(\sum_{j=1,j\neq i}^N R_j-I(Y_i;X_r)\right)}2^{\lfloor(1-\alpha)T\rfloor 3\epsilon}.\end{aligned}\tag{12}$$

Therefore, if

$$\sum_{j=1,j\neq i}^N R_j<I(Y_i;X),\tag{13}$$

we can choose ϵ and T such that $P_e^{(T)}\leq 3\epsilon$, and the probability of error can be made arbitrarily small as $T\to\infty$. Note that the above approach is identical to generating the codewords $x^{\lfloor(1-\alpha)T\rfloor}(w_1,\ldots,w_N)$ at the relay which can be seen as the generalization of a finite field operation in network coding [6].

Note that similar results for discrete memoryless channels have been obtained in [17] for a general broadcast scenario and in [5] for the broadcast phase of the half duplex TWRC, where in contrast here we consider power constrained continuous memoryless channels. Nevertheless, the Slepian-Wolf coding approach in [17] is similar to a layered scheme consisting of decoding and subsequent network coding at the relay [18].

In NWRN, at the relay, the signal after re-encoding is precoded with a matrix $\Phi=[\psi^{m,n}]$ before transmitting over different antennas, i.e., $\sum_{n=1}^M\psi^{m,n}x_{r,t}^n$ where $\{x_{r,t}^n\}$'s are codewords with unit average power and independent entries. By imposing the average power constraint at the relay, we require that

$$\sum_{m=1}^M\sum_{n=1}^M|\psi^{m,n}|^2=\frac{P_r}{1-\alpha}.\tag{14}$$

At time t, the received signal at the source terminal i is

$$y_{i,t} = \sum_{m=1}^{M} h_i^m \sum_{n=1}^{M} \psi^{m,n} x_{r,t}^n + w_{i,t}, \tag{15}$$

and we can compute

$$I(Y_i; X) = (1-\alpha)\log\left(1 + \sum_{n=1}^{M} \left| \sum_{m=1}^{M} h_i^m \psi^{m,n} \right|^2 \right). \tag{16}$$

Therefore, using (13) the achievable rate region can be stated as follows.

Proposition 3: The achievable rate region for the half duplex N-way relay networks by using DF is the convex hull of the intersection of the following two regions

$$\begin{aligned} \sum_{j\in\mathcal{S}} R_j &\leq \alpha \log\det\left(\mathbf{I}_M + \sum_{j\in\mathcal{S}} \frac{P_j}{\alpha}\mathbf{h}_j\mathbf{h}_j^H\right), \forall \mathcal{S} \subseteq \{1,\ldots,N\}, \\ \sum_{j=1, j\neq i}^{N} R_j &\leq (1-\alpha)\log\left(1 + \sum_{n=1}^{M} \left| \sum_{m=1}^{M} h_i^m \psi^{m,n} \right|^2 \right), i = 1,\ldots,N. \end{aligned} \tag{17}$$

C. Outer Bound

In this section we give two outer bounds on capacity under half duplex operation, which are functions of the fraction of time α in which the relay receives. The first is a bound on half duplex capacity under any coding scheme. The second is a bound on half duplex capacity in the classical restricted case as in Shannon's restricted two way relay channel [19], where the encoding at the terminals does not depend on their received signals. In this case cooperation among source terminals cannot be induced by means of communication via the relay, which is in line with the achievable schemes considered in this paper.

When half duplex networks are considered, a general cut-set outer bound is given by [20, Theorem 1]. In general networks, an exact application of [20, Theorem 1] needs to consider all possible configurations of transmitting and receiving nodes. However, in the N-way relay scenario where the only links are between the terminals and the relay, we only need to consider two configurations, one in which the terminals transmits and the relay receives, and the other in which the relay transmits and the terminals receive.

We have the following proposition on the region of information rates $\{R_i\}$.

Proposition 4: Consider a half duplex N-way relay Gaussian network defined in (1). Let $\{x_i\}$ and $\{y_i\}$ be the transmitted and received signals by the N source terminals, and x and y be the transmitted and received signals by the relay, respectively.
a) If there are no restrictions on the coding scheme, the rate region is contained in

$$\bigcup_{\mathbf{C}=[E\{X_iX_j\}]} \left\{ (R_1,\ldots,R_N) \middle| \sum_{j\in\mathcal{S}} R_j \leq \alpha \log\det\left(\mathbf{I}_M + \sum_{i,j\in\mathcal{S}} C_{i,j}\mathbf{h}_j\mathbf{h}_j^H - \mathbf{W}_{Z,\mathcal{S}^c}\mathbf{C}_{\mathcal{S}^c}^{-1}\mathbf{W}_{Z,\mathcal{S}^c}^T\right), \right.$$
$$\forall\mathcal{S}\subsetneq\{1,\ldots,N\}, \tag{18}$$
$$\left. \sum_{j=1,j\neq i}^{N} R_j \leq (1-\alpha)\log\left(1+\sum_{n=1}^{M}\left|\sum_{m=1}^{M} h_i^m\psi^{m,n}\right|^2\right),\ i=1,\ldots,N \right\},$$

where $\mathbf{C}=[E\{X_iX_j\}]$ is the correlation matrix of $X_1,\ldots,X_N$, $\mathbf{C}_{\mathcal{S}^c}$ contains the entries of $\mathbf{C}$ corresponding to $\mathcal{S}^c$, and

$$\mathbf{W}_{Z,\mathcal{S}^c} = \left[\sum_{i\in\mathcal{S}} \mathbf{h}_i C_{i,j}\right]_{j\in\mathcal{S}^c} \tag{19}$$

with $\mathbf{A}=[\mathbf{a}_j]_{j\mathcal{S}}$ denoting each column of $\mathbf{A}$ is from $\mathbf{a}_j$, $j\in\mathcal{S}$.
b) In the restricted case where encoding at the terminals does not depend on their received signals, the rate region is contained in

$$\left\{ (R_1,\ldots,R_N) \middle| \sum_{j\in\mathcal{S}} R_j \leq \alpha\log\det\left(\mathbf{I}_M + \sum_{j\in\mathcal{S}}\frac{P_j}{\alpha}\mathbf{h}_j\mathbf{h}_j^H\right),\ \forall\mathcal{S}\subsetneq\{1,\ldots,N\}, \right.$$
$$\left. \sum_{j=1,j\neq i}^{N} R_j \leq (1-\alpha)\log\left(1+\sum_{n=1}^{M}\left|\sum_{m=1}^{M} h_i^m\psi^{m,n}\right|^2\right),\ i=1,\ldots,N \right\}, \tag{20}$$

Proof: a) In half duplex transmission, $Y_{\mathcal{S}^c}=0$ when the sources transmit (uplink) and $X_{\mathcal{S}^c}=0$ when the relay transmits (downlink) which means that Y is only a function of $X_1,\ldots,X_N$, and Y_i is only a function of X. Thus, for the α-fraction of time in which the terminals transmit and the relay receives, we have an outer bound given by [16, Theorem 15.10.1]

$$\sum_{j\in\mathcal{S}} R_j \leq \alpha I\left(Y; X_{\mathcal{S}}|X_{\mathcal{S}^c}\right),\ \forall\mathcal{S}\subsetneq\{1,\ldots,N\} \tag{21}$$

for some joint distribution $p(x_1,x_2,\cdots,x_N)p(y|x_1,\ldots,x_N)$. The joint distribution is due to the possible cooperation among source terminals by the reception of signals from the relay [21]. For the downlink, where the relay broadcasts to all receivers for an $(1-\alpha)$-fraction of time, we have according to (13)

$$\sum_{j=1,j\neq i}^{N} R_j \leq (1-\alpha)I(Y_i;X),\ i=1,\ldots,N \tag{22}$$

for some probability distribution $p(x)$.

For Gaussian networks as in (1), $I\left(Y;X_{\mathcal{S}}|X_{\mathcal{S}^c}\right)$ in (21) is maximized when each X_i is Gaussian with zero mean and variance P_i [21], and we can compute (21) as

$$I\left(Y;X_{\mathcal{S}}|X_{\mathcal{S}^c}\right)=h\left(Y|X_{\mathcal{S}^c}\right)-h\left(Y|X_{\mathcal{S}},X_{\mathcal{S}^c}\right), \tag{23}$$

where $h\left(Y|X_{\mathcal{S}},X_{\mathcal{S}^c}\right)=M\log 2\pi e\sigma^2$, with $\sigma^2=1$ be the relay noise variance. To compute $h\left(Y|X_{\mathcal{S}^c}\right)$, let $Z_{\mathcal{S}}=\sum_{i\in\mathcal{S}}\mathbf{h}_iX_i$ and the correlation matrix of X be $\mathbf{R}$, i.e., $[\mathbf{C}]_{i,j}=E\{X_i,X_j\}$. The conditional variance of $Z_{\mathcal{S}}$ given $X_{\mathcal{S}^c}$ is no greater than the variance of $Z_{\mathcal{S}}$ around the linear estimate [22]

$$\hat{Z}_{\mathcal{S}}=V^TX_{\mathcal{S}^c},\ \text{with } V=\mathbf{C}_{\mathcal{S}^c}^{-1}\mathbf{W}_{Z,\mathcal{S}^c}^T, \tag{24}$$

and $\mathbf{C}_{\mathcal{S}^c}$ and $\mathbf{W}_{Z,\mathcal{S}^c}$ are defined in (18). Therefore, we have

$$\begin{aligned}h\left(Y|X_{\mathcal{S}^c}\right)\le&\log(2\pi e)^M\det\left(\mathbf{I}_M+\left(\left(Z_{\mathcal{S}}-E\left(\hat{Z}_{\mathcal{S}}\right)\right)\left(Z_{\mathcal{S}}-\hat{Z}_{\mathcal{S}}\right)^T\right)\right)\\=&\log(2\pi e)^M\det\left(\mathbf{I}_M+\sum_{i,j\in\mathcal{S}}C_{i,j}\mathbf{h}_j\mathbf{h}_j^H-\mathbf{W}_{Z,\mathcal{S}^c}\mathbf{C}_{\mathcal{S}^c}^{-1}\mathbf{W}_{Z,\mathcal{S}^c}^T\right).\end{aligned} \tag{25}$$

Finally, we get

$$I\left(Y;X_{\mathcal{S}}|X_{\mathcal{S}^c}\right)\le\log\det\left(\mathbf{I}_M+\sum_{i,j\in\mathcal{S}}C_{i,j}\mathbf{h}_j\mathbf{h}_j^H-\mathbf{W}_{Z,\mathcal{S}^c}\mathbf{C}_{\mathcal{S}^c}^{-1}\mathbf{W}_{Z,\mathcal{S}^c}^T\right). \tag{26}$$

For (22) we note that

$$\sum_{j\in\mathcal{S}}R_j\le(1-\alpha)I\left(Y_{\mathcal{S}^c};X\right)=(1-\alpha)\left(I(Y_i;X)+I\left(Y_{\mathcal{S}^c\backslash i};X|Y_i\right)\right)\le(1-\alpha)I(Y_i;X),\ \forall i\in\mathcal{S}^c, \tag{27}$$

which is included in

$$\sum_{j=1,j\neq i}^{N}R_j\le(1-\alpha)I(Y_i;X)=(1-\alpha)\log\left(1+\sum_{n=1}^{M}\left|\sum_{m=1}^{M}h_i^m\psi^{m,n}\right|^2\right),\ i=1,\dots,N, \tag{28}$$

where $I(Y_i;X)$ is computed similar to the DF case in (17). Note that X and $X_1,\dots,X_N$ may be correlated. But $I(Y_i;X)$ is maximized when X is Gaussian in (28) regardless of this correlation.
b) In this case, $E\{X_i,X_j\}=0$, $\forall i\neq j$ and $\mathbf{W}_{Z,\mathcal{S}^c}=0$ in (19). The outer bound in this case is thus given by (20). ■

Note that by comparing the restricted half duplex outer bound (20) and the achievable rate of DF (17), we can see that the only difference is that there is no total sum rate constraint at the relay, which is equivalent to $\mathcal{S}\neq\{1,\dots,N\}$ in (20) while DF has this constraint due to the decoding requirement.

IV. Diversity-Multiplexing Tradeoff

In this section, we characterize the fundamental tradeoff between the diversity and multiplexing gain [12], [13] in N-way relay networks using AF and DF, which suggests the scenarios when one strategy is preferred than the other. In the following, we assume that $\sum_{i=1}^{N} P_i + P_r = P$ and that all channel gains are complex Gaussian. As there are N sources in the network, we consider the symmetric case as in [13], where all the sources transmit at the same rate, and they have a common diversity requirement. Let $R^i(P)$ be the achievable rate and $P_e^i(P)$ be the average error probability of source i in a strategy. The strategy is said to achieve spatial multiplexing gain r and diversity gain d if

$$\lim_{P \to +\infty} \frac{R^i(P)}{\log P} \geq r, \quad \lim_{P \to +\infty} \frac{\log P_e^i(P)}{\log P} \leq -d, \quad \forall i = 1, \ldots, N. \tag{29}$$

A. Amplify-and-Forward

Theorem 5: Let $P_1 = \cdots = P_N = P_s$ and $NP_s = P_r$, i.e., $P_s = \frac{P}{2N}$ and $P_r = \frac{P}{2}$. Each antenna's transmit signal at the relay is its received signal scaled by $\sqrt{\frac{P}{M(P+1)}}$. By assuming that all channel gains are i.i.d. complex Gaussian, the diversity gain of AF is

$$d_{\text{AF}} = \begin{cases} M(1 - 2(N-1)r), & \text{if } r \geq \frac{M-1}{2(M(N-1)-1)}, \\ 1 - 2r, & \text{if } r < \frac{M-1}{2(M(N-1)-1)}. \end{cases} \tag{30}$$

Proof: In this case, the achievable rate region of AF can be readily obtained from (5) as

$$\sum_{j \in \mathcal{S}}^{N} R_j \leq \frac{1}{2} \log \left(1 + \frac{\frac{P^2}{N} \sum_{j \in \mathcal{S}} \left| \sum_{m=1}^{M} h_i^m h_j^m \right|^2}{P \sum_{m=1}^{M} |h_i^m|^2 + M(P+1)} \right), \forall \mathcal{S} \subseteq \mathcal{S}_i. \tag{31}$$

Given h_i^m, $\sum_{m=1}^{M} h_i^m h_j^m$ is a complex Gaussian random variable with zero mean and variance $\|\mathbf{h}_i\|^2$. We can thus write $\sum_{m=1}^{M} h_i^m h_j^m = \|\mathbf{h}_i\| x_j$, where x_j is a Gaussian random variable with zero mean and unit variance. Therefore, $\sum_{j \in \mathcal{S}} \left| \sum_{m=1}^{M} h_i^m h_j^m \right|^2 = \|\mathbf{h}_i\|^2 \sum_{j \in \mathcal{S}} |x_j|^2 = g_1 g_2$, where g_1 and g_2 are chi-square random variables with $2M$ and $2|\mathcal{S}|$ degrees of freedom, respectively.

Let g be a chi-square random variable with $2n$ degrees of freedom. If v is the exponential order of $1/g$, i.e.,

$$v = -\lim_{P \to +\infty} \frac{\log g}{\log P}, \tag{32}$$

the probability density function of v can be obtained as [23]

$$p_v = \lim_{P \to +\infty} \log P P^{-nv} \exp\left(-P^{-v}\right). \tag{33}$$

For independent random variables $g_1, \ldots, g_L$ where g_i is chi-square distributed with $2n_i$ degrees of freedom, let v_i be the exponential order of g_i. The probability that $\{v_i\}$ belongs to any set $\mathcal{D}$ is dominated by [23]

$$P^{-d^*}, \quad d^* = \inf_{\{v_i\}\in\mathcal{D}} \sum_{i=1}^{L} n_i v_i. \tag{34}$$

By using (34), the outage probability

$$\begin{aligned}&\frac{1}{2}\log\left(1+\frac{\frac{P^2}{N}\sum_{j\in\mathcal{S}}\left|\sum_{m=1}^{M} h_i^m h_j^m\right|^2}{P\sum_{m=1}^{M}|h_i^m|^2+M(P+1)}\right)=\frac{1}{2}\log\left(1+\frac{\frac{P^2}{N}g_1g_2}{Pg_1+M(P+1)}\right)\le|\mathcal{S}|r\log P\\ \Rightarrow&\lim_{P\to+\infty}\frac{1}{\log P}\log\left(1+\frac{\frac{P^2}{N}g_1g_2}{Pg_1+M(P+1)}\right)=1+\lim_{P\to+\infty}\frac{\log(g_1)+\log(g_2)}{\log P}=1-v_1-v_2\le 2|\mathcal{S}|r\end{aligned} \tag{35}$$

is dominated by

$$\max_{v_1,v_2} P^{-(Mv_1+|\mathcal{S}|v_2)}, \quad \text{subject to } 1-v_1-v_2\le 2|\mathcal{S}|r. \tag{36}$$

When $|\mathcal{S}| > M$, $Mv_1 + |\mathcal{S}|v_2$ is minimized by choosing $v_1 = 1-2|\mathcal{S}|r$ and $v_2 = 0$. When $|\mathcal{S}| \le M$, $Mv_1 + |\mathcal{S}|v_2$ is minimized by choosing $v_1 = 0$ and $v_2 = 1-2|\mathcal{S}|r$. Therefore, we obtain the diversity gain d as

$$d = \begin{cases} M(1-2|\mathcal{S}|r), & \text{if } |\mathcal{S}| \ge M, \\ |\mathcal{S}|(1-2|\mathcal{S}|r), & \text{if } |\mathcal{S}| < M. \end{cases} \tag{37}$$

We then need to find the worst case d for all $\mathcal{S}$. Note that the second condition is concave in $|\mathcal{S}|$ and its minimum is attained at its boundary $|\mathcal{S}| = 1$ or $|\mathcal{S}| = M$. The first condition is a decreasing function in $|\mathcal{S}|$. Therefore, the worst case d is attained at $|\mathcal{S}| = 1$ or $|\mathcal{S}| = N-1$ and we obtain (30).

■

From (30) we can see that when the system is lightly loaded, e.g., if $r < \frac{M-1}{2(M(N-1)-1)}$, single user performance is achieved.

B. Decode-and-Forward

Theorem 6: Let $P_1 = \cdots = P_N = P_s$ and $NP_s = P_r$, i.e., $P_s = \frac{P}{2N}$ and $P_r = \frac{P}{2}$. By assuming that all channel gains are i.i.d. complex Gaussian, the diversity gain of DF is

$$d_{\text{DF}} = \begin{cases} (1-Nr)(M-Nr), & \text{if } r < \frac{1}{N}\min\left(1, \frac{M}{N+1}\right), \\ \max_\alpha \min\left(N(1-\frac{r}{\alpha})(M-\frac{Nr}{\alpha}), \left(1-(N-1)\frac{r}{1-\alpha}\right)\left(M-(N-1)\frac{r}{1-\alpha}\right)\right), & \text{if } r \ge \frac{1}{N}\min\left(1, \frac{M}{N+1}\right). \end{cases} \tag{38}$$

Proof: The uplink channel is a multiple access channel, characterized in [13], with a diversity gain of

$$d^{\text{up}}(\alpha) = \begin{cases} \left(1-\frac{r}{\alpha}\right)\left(M-\frac{r}{\alpha}\right), & \text{if } r < \alpha \min\left(1, \frac{M}{N+1}\right), \\ N(1-\frac{r}{\alpha})(M-\frac{Nr}{\alpha}), & \text{if } r \geq \alpha \min\left(1, \frac{M}{N+1}\right), \end{cases} \tag{39}$$

where $\frac{r}{\alpha}$ is due to the half duplex operation. From (17), the downlink channel can be considered to be a MISO channel whose diversity multiplexing tradeoff can be easily obtained as in [12], i.e.,

$$d^{\text{down}}(\alpha) = \left(1-(N-1)\frac{r}{1-\alpha}\right)\left(M-(N-1)\frac{r}{1-\alpha}\right). \tag{40}$$

The diversity gain can now be obtained as $d_{\text{DF}} = \max_\alpha \min\{d^{\text{up}}(\alpha), d^{\text{down}}(\alpha)\}$, leading to (38). When $r \leq \alpha \min\left(1, \frac{M}{N+1}\right)$ the diversity gain d_{DF} is maximized for $\frac{\alpha}{1-\alpha} = \frac{1}{N-1}$, or $\alpha = \frac{1}{N}$. ■

Note that for $r > \alpha \min\left(1, \frac{M}{N+1}\right)$ the optimal α depends on M, N, r in a complicated way, which makes it hard to find a closed form expression for α. However, α can be found via a numerical optimization.

V. Practical Space Time Protocols

The practical protocols in this section are in parallel to those presented in Section III. Specifically, we consider a class of strategies where the source terminals transmit one symbol $s_i \in \mathcal{Q}$ simultaneously in the first time slot and the relay transmits in the following K ($K = (1-\alpha)T$ in Section III) time slots, where $\mathcal{Q}$ is a finite constellation set. This assignment of time slots is due to the fact that each of the N source terminals only has one antenna, whereas the relay typically has $M > 1$ antennas. We assume that the channel remains unchanged during $K+1$ time slots in the following. At the relay, the received signal from the m-th antenna at the end of the first time slot is

$$y_r^m = \sum_{i=1}^{N} \sqrt{(K+1)P_i} h_i^m s_i + w_r^m, \quad m = 1, \ldots, M. \tag{41}$$

The protocols can be classified into two classes: compression based protocols, where the relay compresses its received signal by reducing the alphabet size of the constellation seen by the relay, and non-compression based protocols.

A. Non-Compression Based Protocols

1) Amplify-and-Forward: By using AF, at the relay the transmitted signal over the m-th antenna at time slot $k+1$ is given as

$$x_{r,k}^{m} = \sum_{n=1}^{M} \psi_k^{m,n} y_r^n = \sum_{i=1}^{N} \sqrt{(K+1)P_i} \left(\sum_{n=1}^{M} \psi_k^{m,n} h_i^n \right) s_i + \sum_{n=1}^{M} \psi_k^{m,n} w_r^n, \tag{42}$$

$k = 1, \ldots, K$, $m = 1, \ldots, M$, which represents a special case of a linear dispersion code [11]. In (42), y_r^n can also be replaced by its conjugate complex version $(y_r^n)^*$. Let $y_{i,k}$ be the received signal of source terminal i at time slot $k+1$. We can write $y_{i,k}$ as

$$y_{i,k} = \sum_{m=1}^{M} h_i^m x_{r,k}^m + w_{i,k} = \sum_{j=1}^{N} \sqrt{(K+1)P_j} \left(\sum_{m=1}^{M} \sum_{n=1}^{M} h_i^m \psi_k^{m,n} h_j^n \right) s_j + \sum_{m=1}^{M} \sum_{n=1}^{M} h_i^m \psi_k^{m,n} w_r^n + w_{i,k}, \tag{43}$$

which can be expressed in matrix form as

$$y_{i,k} = \mathbf{h}_i^T \Psi_k \mathbf{H} \mathbf{\Lambda} \mathbf{s} + \mathbf{h}_i^T \Psi_k \mathbf{w}_r + w_{i,k}, \tag{44}$$

where $\mathbf{s} = [s_1, \ldots, s_N]^T$, $\mathbf{h}_i = [h_i^1, \ldots, h_i^M]^T$, $\mathbf{H} = [\mathbf{h}_1, \ldots, \mathbf{h}_N]$, $\Psi_k = [\psi_k^{m,n}]$, $\mathbf{\Lambda} = \text{diag}\{\sqrt{(K+1)P_1}, \cdots, \sqrt{(K+1)P_N}\}$, and $\mathbf{w}_r = [w_r^1, \ldots, w_r^M]^T$. To meet the relay average power constraint, we require

$$\left((K+1) \sum_{j=1}^{N} P_j + 1 \right) \text{tr}\left(\Psi_k \Psi_k^H \right) = \frac{K+1}{K} P_r. \tag{45}$$

Note that (44) can be expressed as

$$\mathbf{y}_i = \mathbf{G}_i \mathbf{H} \mathbf{\Lambda} \mathbf{s} + \mathbf{G} \mathbf{w}_r + \mathbf{w}_i, \tag{46}$$

where $\mathbf{G}_i = [\Psi_1^T \mathbf{h}_i, \cdots, \Psi_K^T \mathbf{h}_i]^T$, $\mathbf{y}_i = [y_{i,1}, \ldots, y_{i,K}]^T$ and $\mathbf{w}_i = [w_{i,1}, \ldots, w_{i,K}]^T$. The ML decoder of s can be easily obtained as

$$\hat{\mathbf{s}} = \underset{\{\tilde{\mathbf{s}} | \tilde{s}_j \in \mathcal{Q},\, j \neq i,\, \tilde{s}_i = s_i\}}{\arg\min} \left\| \left(\mathbf{G}_i \mathbf{G}_i^H + \mathbf{I}_K \right)^{-\frac{1}{2}} \left(\mathbf{y}_i - \mathbf{G}_i \mathbf{H} \mathbf{\Lambda} \mathbf{s} \right) \right\|^2, \tag{47}$$

which can be solved by using sphere decoding [24].

2) Decode-and-Forward: In DF, after receiving the signals from the source terminals, the relay jointly decodes all source terminals' messages as in Section III-B, i.e., $s_1, \ldots, s_N$ are jointly decoded using an ML decoder

$$\{\hat{s}_1, \ldots, \hat{s}_N\} = \underset{s_1, \ldots, s_N}{\arg\min} \sum_{m=1}^{M} \left| y_r^m - \sum_{i=1}^{N} h_i^m s_i \right|^2, \tag{48}$$

which again can be solved by using a sphere decoder [24].

After obtaining $\hat{s}_1, \ldots, \hat{s}_N$ from (48) and assuming that $\hat{s}_1, \ldots, \hat{s}_N$ are decoded correctly, there are two ways to communicate these symbols to the source terminals. We model the downlink channel as a multiple input and single output (MISO) channel and use a linear dispersion code [11] to encode $s_1, \ldots, s_N$. The transmitted signal containing $s_1, \ldots, s_N$ can be written as a K by M matrix given by

$$\mathbf{X}_r = \sum_{i=1}^{N} \left(s_i \mathbf{C}_i + s_i^* \mathbf{D}_i \right), \tag{49}$$

where $\mathbf{C}_i$ and $\mathbf{D}_i$ are linear dispersion coding matrices. The average power constraint requires that $E\{\text{tr}(\mathbf{X}_r \mathbf{X}_r^H)\} = (K+1)P_r$. The received signal at source i can be written as

$$\mathbf{y}_i = \mathbf{X}_r \mathbf{h}_i = \sum_{j=1}^{N} \left(s_j \mathbf{C}_j \mathbf{h}_i + s_j^* \mathbf{D}_j \mathbf{h}_i \right) + \mathbf{w}_i, \tag{50}$$

where $\mathbf{h}_i$ $\mathbf{y}_i$, and $\mathbf{w}_i$ are defined as in (46). Note that s_i is already known at source i and its contribution in $\mathbf{y}_i$ can be subtracted. The ML decoder for s_j, $j \neq i$ in (50) can be readily obtained as in [11] due to the linearity of the code. The design of $\mathbf{C}_i$ and $\mathbf{D}_i$ is similar to that in [11].

3) Denoise-and-Forward: At the relay, $s_1, \ldots, s_N$ are jointly decoded using the ML decoder (48). Different from DF where $\hat{s}_1, \ldots, \hat{s}_N$ are transmitted using a linear dispersion code directly, DNF reconstructs $z_r^m = \sum_{i=1}^N h_i^m \hat{s}_i$. The relay transmits z_r^m because even when the relay cannot decode $x_{1,t}, \ldots, x_{N,t}$ correctly, the signal z_r^m is not far away from the noise free signal $\sum_{i=1}^N h_i^m s_i$. Thus, the proposed approach may remedy noise amplification in AF and hence the name DNF. As in AF, the transmitted signal over the m-th antenna at time slot $k+1$ is

$$x_{r,k}^m = \sum_{n=1}^{M} \psi_k^{m,n} z_r^n, \; k = 1, \ldots, K, \; m = 1, \ldots, M. \tag{51}$$

The ML decoding and precoder design are also similar to AF.

4) Estimate-and-Forward: Different from AF and DF, which are obtained intuitively, we consider using a relay function $g^m(\cdot)$ to process the received signal such that the MSE between the transmitted signal without noise and the processed signal is minimized, i.e.,

$$\min_{g^m} E_{\mathbf{s}} \left\{ \left| \sum_{i=1}^{N} h_i^m s_i - g^m\left(y_r^m\right) \right|^2 \right\}. \tag{52}$$

By expanding the expectation in (52) we obtain

$$E_{\mathbf{s}}\left\{\left|\sum_{i=1}^{N} h_i^m s_i - g^m\left(y_r^m\right)\right|^2\right\} = \frac{1}{2\pi|\mathcal{Q}|^N} \sum_{\mathbf{s}\in\mathcal{Q}^N} \int \left|\sum_{i=1}^{N} h_i^m s_i - g^m\left(y_r^m\right)\right|^2 e^{-\left|y_r^m - \sum_{i=1}^N h_i^m s_i\right|^2} dy_r^m. \tag{53}$$

Minimizing (53) for each $g^m\left(y_r^m\right)$ then yields

$$g^m\left(y_r^m\right) = \frac{\sum_{\mathbf{s}\in\mathcal{Q}^N} \sum_{i=1}^N h_i^m s_i e^{-\left|y_r^m - \sum_{i=1}^N h_i^m s_i\right|^2}}{\sum_{\mathbf{s}\in\mathcal{Q}^N} e^{-\left|y_r^m - \sum_{i=1}^N h_i^m s_i\right|^2}}. \tag{54}$$

Let $z_r^m = \beta g^m\left(y_r^m\right)$, where β is a constant to maintain the average power constraint at the relay. The rest of the protocol follows AF by transmitting $\sum_{n=1}^M \psi_k^{m,n} z_r^n$ over antenna m at time slot $k+1$. Note that when a Gaussian codebook is used, we can rewrite (54) as

$$g^m\left(y_r^m\right) = \frac{\int x e^{-|y_r^m - x|^2} f(x)dx}{\int e^{-|y_r^m - x|^2} f(x)dx} = \frac{(K+1)\sum_{i=1}^N P_i |h_i^m|^2}{(K+1)\sum_{i=1}^N P_i |h_i^m|^2 + 1} y_r^m, \tag{55}$$

where $x = \sum_{i=1}^N h_i^m s_i$ is a Gaussian random variable with zero mean and variance $(K+1)\sum_{i=1}^N P_i |h_i^m|^2$ and $f(x)$ is the pdf of x. From (55), we can see that EF reduces to AF for a Gaussian codebook. However, due to the use of a discrete alphabet $g^m\left(y_r^m\right)$, (54) is not equivalent to AF in the general case.

B. *Compression Based Protocols*

We now consider a compression based DF (CDF) when all terminals and the relay have a single antenna. Two important components of CDF are the quantization codebook and the transmission codebook at the relay. The first codebook determines how to project the received signals to a subspace of the received signal space, and the second one provides the actual transmit symbol by the relay. To use network coding to simplify the mapping design at the relay, the relay can compress its received signals to reduce the power and bandwidth consumption since source i already knows s_i. Intuitively, each terminal only needs $N-1$ signals as a function of the transmitted signals s_i. One straightforward approach is letting the relay transmit $N-1$ superimposed signals directly, i.e, $\sum_{j=1}^N \psi_{i,j} s_j$, $i = 1, \ldots, N-1$, where the $\psi_{i,j}$'s are complex numbers. However, this approach suffers from a power penalty as each terminal only needs linear combinations of $N-1$ signals rather than N signals. When $s_1, \ldots, s_N$ is known at the relay it becomes a multicast problem to transmit $s_1, \ldots, s_N$ to all the sources.

We apply a modular operation to solve this multicast problem. Let $\chi_i \in \{0, 1, \ldots, |\mathcal{Q}|\}$ correspond to the index of s_i in $\mathcal{Q}$. We can choose $\eta_i = (\chi_i + \chi_{i+1}) \mod |\mathcal{Q}|$, $i = 1, \ldots, N-1$. The relay transmits a signal $\tilde{x}_i$ from $\mathcal{Q}$ with index corresponding to η_i by using a suitable mapping (e.g., a natural or a Gray mapping). At any terminal j, on knowing χ_j, we can consecutively recover χ_i, $i \neq j$ using $\{\eta_i\}_{i=1}^{N-1}$. This also shows that $q = |\mathcal{Q}|$ is sufficient and suggests that using a modular group, which exists for all $|\mathcal{Q}|$, is suitable for solving the problem.

The $\tilde{x}_i$'s obtained from the above mapping are encoded using a linear dispersion code and the transmitted signal can be written as a K by M matrix given by

$$\tilde{\mathbf{X}}_r = \sum_{i=1}^{N-1} \left(\tilde{x}_i \tilde{\mathbf{C}}_i + \tilde{x}_i^* \tilde{\mathbf{D}}_i \right). \tag{56}$$

At the source terminals, we first decode $\tilde{x}_i$ in the same way as in the non-compression DF, and we then find the corresponding η_i in $\mathbb{F}$. By the construction of η_i, the source terminals can recover the signals correctly. The design of the precoding matrices $\tilde{\mathbf{C}}_i$ and $\tilde{\mathbf{D}}_i$ also follows that in [11] by maximizing the sum rate. By using network coding, the number of symbols need to be transmitted is reduced by one and the transmit symbols are superimposed in finite fields, which saves both bandwidth and power.

VI. Simulation Results

In this section, experimental results are shown to verify the derived theoretical results. We consider a 3-way relay network in all simulations, i.e., $N = 3$. The total power of the network is P, among which γP is allocated to all the source terminals and $(1-\gamma)P$ is allocated to the relay. Unless otherwise mentioned, all channel fadings have a complex Gaussian distribution with zero mean and unit variance.

We consider maximizing the sum rate $\sum_{i=1}^{N} R_i$, which is given by

$$\begin{aligned} \max_{\Psi} \quad & \frac{N-1}{2N} \sum_{i=1}^{N} \log \left(1 + \frac{2\sum_{j=1, j\neq i}^{N} |\mathbf{h}_i^T \Psi \mathbf{h}_j|^2 P_j}{\|\mathbf{h}_i^T \Psi\|^2 + 1} \right), \\ \text{subject to } & 2\sum_{i=1}^{N} \|\Psi \mathbf{h}_i\|^2 P_i + \operatorname{tr}\left(\Psi \Psi^H\right) \leq 2P_r. \end{aligned} \tag{57}$$

The average achievable rates of AF and DF are compared by averaging over 1000 channel realizations. The AF strategy 1 in (4) and DF using diagonal precoding matrices are denoted as "AF 1 Diagonal" and "DF Diagonal". "AF 1 Opt" and "DF Opt" denote AF and DF using optimal precoding matrices obtained via multi-dimensional search, resp., and "AF 2 Opt" denotes the AF strategy 2 in (7). Due to the complexity of the high dimensional optimization in AF strategy 2, we only consider the case $T = 4$ and $\alpha = 0.5$ and start the multidimensional optimization from the solution for AF strategy 1. We also compare with the sum rate obtained from the outer bound (20) and a full duplex cut-set outer bound denoted as "Restricted Half Duplex Outer Bound" and "Full Duplex Cut-set Bound", respectively.

Fig. 1 compares the achievable sum rate of different protocols as a function of P in a 3-way relay network with $\gamma = 0.2$ and $\gamma = 0.8$, respectively. All channels between the source and the relay have unit average power. The relay has $M = 2$ antennas. We find that in both cases, when P is small, DF achieves a higher sum rate than AF. When $\gamma = 0.2$, the sum rate is limited by the uplink channel. In this case, as (20) does not contain a sum rate constraint, the outer bound is greater than the achievable rate of DF. But when P is large, the downlink channel becomes the bottleneck, and the achievable rate of DF converges to the outer bound (20). When $\gamma = 0.8$, the downlink channel is always the bottleneck and DF performs very close to (20). In addition, the achievable rate difference between AF strategy 1 and AF strategy 2 is very small; as γ increases, their difference diminishes. We also find that the performance loss due to the use of diagonal precoding matrices is small when P is large for both AF strategy 1 and DF, which is advantageous since diagonal matrices are easier to find and to implement in practice.

To consider the impact of network asymmetry, Figs. 2 and 3 compare the achievable sum rate of different protocols when the channels between the sources and the relay have different average power. In particular, in Fig. 2 the channel between source 1 and the relay has unit power gain, while the other two channels have -3 dB power gain. On the other hand, in Fig. 3, the channel between source 1 and the relay has -3 dB power gain, the other two channels have unit power gain. From these figures, we can see that by changing the symmetry of the network the relative performance of different protocols remains the same as in a symmetric network in Fig. 1, though the achievable rates are affected.

Fig. 4 compares the achievable sum rate of different protocols with different γ when the relay has $M = 2$ antennas. We choose $P = 10$. It is seen that DF performs better than AF for all values of γ. Similar observations can be obtained from Fig. 5 where the relay is equipped with $M = 1$ antenna. Here, only the curves for "AF Opt" and "DF Opt" are displayed, the results for diagonal precoding matrices are virtually indistinguishable from the corresponding optimal curves and therefore not shown. From these two figures, we can see that DF achieves a higher sum rate as AF and performs very close to the outer bound (20). Therefore, if we neglect the required processing complexity at the relay, DF is a preferable choice, when the downlink channel is the bottleneck or when the number of users in the network is large, while AF is recommended in low SNR.

Fig. 6 shows the diversity-multiplexing tradeoff for both AF and DF protocols in a 10-way relay network where the relay has $M = 1$ and $M = 3$ antennas, respectively. In Fig. 6(a), the degree of freedom of the whole network is limited by the single antenna at the relay. Therefore, the maximum DMT achievable by both DF and AF is at most one. In Fig. 6(b) the maximum degree of freedom achievable by DF is at most three. Since each source has a single antenna and the relay needs to combine all the signals at the relay, the degree of freedom in AF is limited by the number of antennas at the source, which is one. Fig. 6 suggests that AF is useful only when $M = 1$, while DF is preferable in a lightly loaded system with more than one antenna where the signals can be decoded reliably at the relay.

We now provide simulation results for the practical protocols proposed in Section V. Fig. 7 compares the average symbol error probability of different protocols as a function of P in a 3-way relay network. The relay has $M = 2$ antennas in the simulations. We further choose $K = M$ because the diversity order is determined by $\min(K, M)$. We do not want to choose $K \geq M$ which would imply a smaller rate for the same diversity order, while we do not want to choose $K < M$ to sacrifice diversity order for rate. The power allocation is chosen such that $\gamma = 1$. We employ the AF protocol by using an Alamouti type space time code [25] (denoted as "AF Alamouti"). For comparison purposes we also employ a diagonal space time code at the relay which corresponds to the case that only the k-th antenna at the relay is active in time slot k, $k = 1, 2$ (denoted as "AF Diagonal"). The DF protocol comprises a concatenation of a linear dispersion code with an Alamouti code, where the linear dispersion code is optimized by maximizing the sum rate of all 3 terminals at $P = 20$. After optimization, we find that the coding matrix in (49) is

$$\begin{bmatrix} -0.1314s_1 + 0.7687s_2 + 0.6260s_3 & 0.7880s_1 - 0.1830s_2 + 0.5878s_3 \\ -0.7880s_1^* + 0.1830s_2^* - 0.5878s_3^* & -0.1314s_1^* + 0.7687s_2^* + 0.6260s_3^* \end{bmatrix}, \tag{58}$$

which is denoted as "DF Sum Opt". DF by using the Toeplitz network coding in Section V-A2 with an Alamouti code is denoted as "DF Network Coding". The DNF protocols are defined similarly as their AF counterparts. From Fig. 7, we can see that the performance of EF is very close to DNF. Further, it can be observed that the DF protocols achieve a much better performance than both AF and DNF protocols when P is large. It seems that both DNF and AF attain a smaller diversity order than DF in the observed region. The DNF protocols perform better than their AF counterparts but are still inferior to DF protocols. For the DF protocols we find that by employing network coding a performance gain can be achieved which is due to the power savings of exploiting the inherent compression in network coding. On the other hand, the complexity of AF is much lower than the DF protocols as AF does not need decoding and signal processing at the relay.

Fig. 8 compares the average symbol error probability of different protocols when the relay has $M = 1$ antenna. We also include the performance of CDF with physical layer network coding in Section V-B. For each realization of channel fading we choose the best quantization and transmission codebooks according to the procedure in Section V-B. It is observed that optimized CDF attains the best performance in high SNR while AF has a good performance in low SNR, which agrees with the observations in [3]. In high SNR DF outperforms DNF and EF since it uses a linear dispersion code for the downlink, compared to precoding as in DNF and EF. In particular, this can be seen from Fig. 8 where there is a crossover between DF and DNF at an SNR of approximately 17 dB. Below this SNR, DNF (and also EF) perform better than DF since the estimation error for the signal at the m-th receive antenna of the relay, $m = 1, \dots, M$, becomes smaller compared to DF, which offsets the coding gain for linear dispersion coding versus precoding on the downlink in this regime.

Thus, Fig. 8 suggests that when a single antenna system is considered CDF is the preferable choice. Note that in Figs. 7 and 8 no additional forward error correction is employed, and therefore the corresponding rates in the simulations are given as $(1 - \text{symbol error probability}) \cdot (\text{space time code rate})$. Thus, an SER improvement translates into a throughput improvement.

VII. Conclusion

We have studied theoretical and practical properties of several communication protocols for N-way all-cast relay networks. In particular, besides AF and DF we introduce new variants, denoted as estimate-and-forward, denoise-and-forward, and compression-based DF, respectively. As an important theoretical property we have derived the achievable rate regions for these protocols.

Here, we show that network coding can be used to improve the rate regions for the DF-based protocols. Further, outer capacity bounds for the rate regions of the considered protocols were established. As an additional criterion for comparing AF and DF we have also characterized the achievable diversity-multiplexing tradeoff for both schemes. On the practical side, several space-time coding protocols based on AF and DF strategies at the relay were developed. Here, we have focused on linear dispersion space-time codes, which can be optimized by maximizing the sum rate.

In general, AF-based protocols lead to a lower signal processing complexity at the relay compared to DF-based strategies. In contrast, DF-based protocols are advantageous in lightly loaded systems with multiple transmit antennas at the relay and generally in the high-SNR regime. In particular, for a single-antenna relay, CDF is preferable, which uses an implicit compression by physical layer network coding and outperforms DF in high SNR.

REFERENCES

[1] B. Rankov and A. Wittneben, "Spectral efficient signaling for half-duplex relay channels," in *Proc. of Asilomar Conference on Signals, Systems and Computers*, Oct. 2005, pp. 1066–1071.

[2] R. Knopp, "Two-way radio networks with a star topology," in *Proc. of International Zurich Seminar on Communications*, July 2006, pp. 154–157.

[3] T. Cui, T. Ho, and J. Kliewer, "Memoryless relay strategies for two-way relay channels," *IEEE Trans. Commun.*, vol. 57, no. 10, pp. 3132–3143, Oct. 2009.

[4] T. Cui, F. Gao, T. Ho, and A. Nallanathan, "Distributed space-time coding for two-way wireless relay networks," *IEEE Trans. Signal Processing*, vol. 57, no. 2, pp. 658–671, Feb. 2009.

[5] T. Oechtering, C. Schnurr, I. Bjelakovic, and H. Boche, "Broadcast capacity region of two-phase bidirectional relaying," *IEEE Trans. Inform. Theory*, vol. 54, no. 1, pp. 454–458, 2008.

[6] T. Ho, M. Médard, R. Koetter, D. Karger, M. Effros, J. Shi, and B. Leong, "A random linear network coding approach to multicast," *IEEE Trans. Inform. Theory*, vol. 52, no. 10, pp. 4413–4430, Oct. 2006.

[7] Y. Wu, "Broadcasting when receivers know some messages a priori," in *Proc. IEEE Int. Symposium on Inform. Theory*, Nice, France, June 2007, pp. 1141–1145.

[8] G. Kramer, M. Gastpar, and P. Gupta, "Cooperative strategies and capacity theorems for relay networks," *IEEE Trans. Inform. Theory*, vol. 51, no. 9, pp. 3037–3063, Sept. 2005.

[9] K. Gomadam and S. Jafar, "The effect of noise correlation in amplify-and-forward relay networks," *IEEE Trans. Inform. Theory*, vol. 55, no. 2, pp. 3728–3733, Aug. 2008.

[10] Y. Jing and B. Hassibi, "Distributed space-time coding in wireless relay networks," *IEEE Trans. Wireless Commun.*, vol. 5, no. 12, pp. 3524–3536, Dec. 2006.

[11] B. Hassibi and B. Hochwald, "High-rate codes that are linear in space and time," *IEEE Trans. Inform. Theory*, vol. 48, no. 7, pp. 1804–1824, July 2002.

[12] L. Zheng and D. Tse, "Diversity and multiplexing: a fundamental tradeoff in multiple-antenna channels," *IEEE Trans. Inform. Theory*, vol. 49, no. 5, pp. 1073– 1096, May 2003.

[13] D. Tse, P. Viswanath, and L. Zheng, "Diversity-multiplexing tradeoff in multiple-access channels," *IEEE Trans. Inform. Theory*, vol. 50, no. 9, pp. 1859–1874, Sept. 2004.

[14] M. Yuksel and E. Erkip, "Multiple-antenna cooperative wireless sytems: a diversity multiplexing trade-off perspective," *IEEE Trans. Inform. Theory*, vol. 53, no. 10, pp. 3371–3393, Oct. 2007.

[15] D. Chen, K. Azarian, and J. N. Laneman, "A case for amplify-forward relaying in the block-fading multiple-access channel," *IEEE Trans. Inform. Theory*, vol. 54, no. 8, pp. 3728–3733, Aug. 2008.

[16] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed., 2005.

[17] E. Tuncel, "Slepian-wolf coding over broadcast channels," *IEEE Trans. Inform. Theory*, vol. 52, no. 4, pp. 1469–1482, April 2006.

[18] J. Kliewer, T. Dikaliotis, and T. Ho, "On the performance of joint and separate channel and network coding in wireless fading networks," in *Proc. 2007 IEEE Inform. Theory Workshop*, Bergen, Norway, 2007, pp. 1–5.

[19] C. E. Shannon, "Two-way communication channels," in *Proc. 4th Berkeley Symp. Math. Stat. Prob.*, 1961, pp. 611–644.

[20] M. A. Khojastepour, A. Sabharwal, and B. Aazhang, "Bounds on achievable rates for general multi-terminal networks with practical constraints," in *Lecture Notes in Computer Science*, 2003, pp. 146–161.

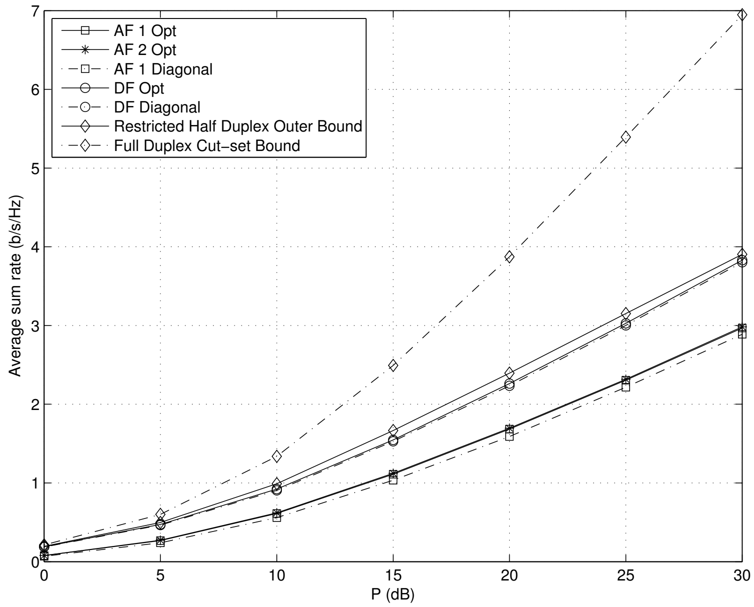
[21] L. Ozarow, "The capacity of the white Gaussian multiple access channel with feedback," *IEEE Trans. Inform. Theory*, vol. 30, no. 4, pp. 623–629, Jul. 1984.

[22] S. Haykin, *Adaptive Filter Theory*, 4th ed. Prentice Hall, 2001.

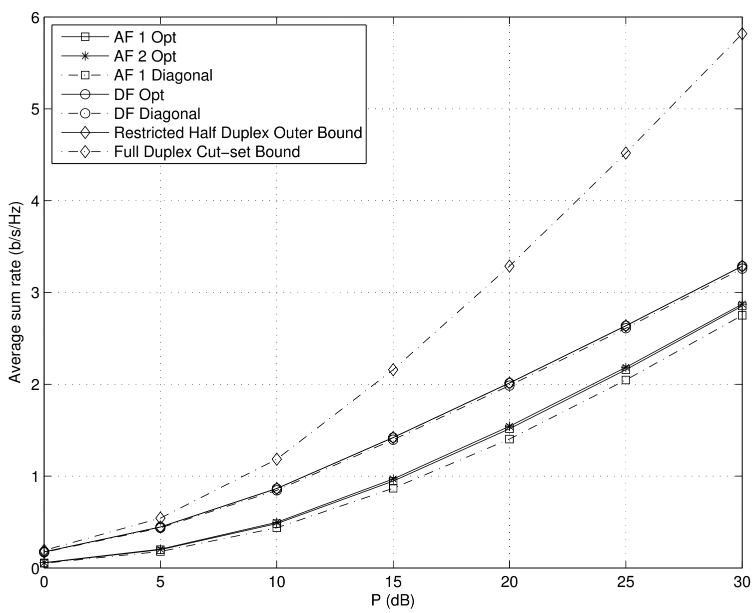
[23] K. Azarian, H. El Gamal, and P. Schniter, "On the achievable diversity-multiplexing tradeoff in half-duplex cooperative channels," *IEEE Trans. Inform. Theory*, vol. 51, no. 12, pp. 4152–4172, Dec. 2005.

[24] B. Hassibi and H. Vikalo, "On the sphere-decoding algorithm I. Expected complexity," *IEEE Trans. Signal Processing*, vol. 53, no. 8, pp. 2806 – 2818, Aug. 2005.

[25] S. M. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE J. Select. Areas Commun.*, vol. 16, no. 8, pp. 1451 – 1458, Oct. 1998.

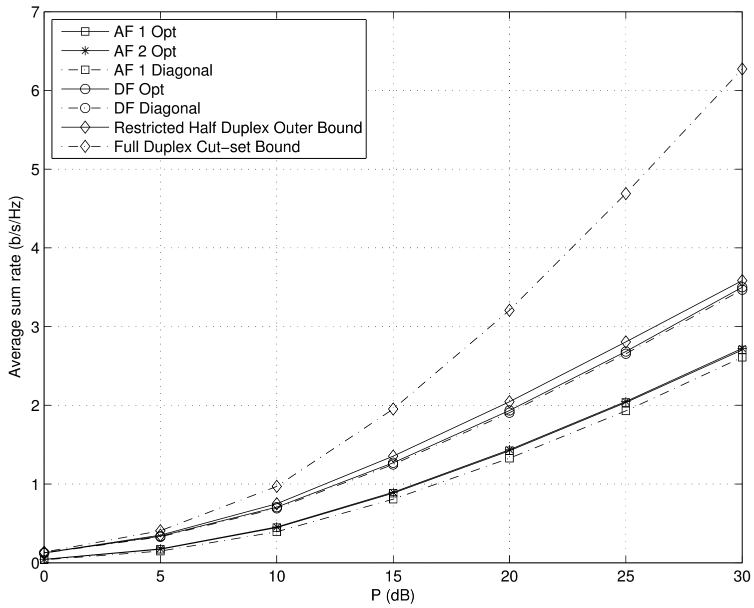


(a) $\gamma = 0.2$

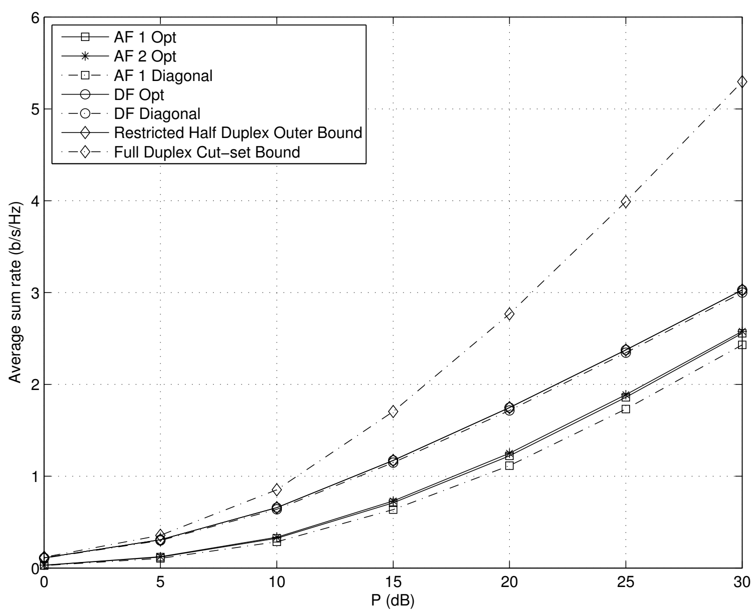


(b) $\gamma = 0.8$

Fig. 1. Achievable sum rate comparison of different protocols as a function of P in a 3-way symmetric relay network, where all channels have unit variance, γP denotes the overall source power and $(1-\gamma)P$ the relay power, respectively. The relay has $M = 2$ antennas, and all channel gains are complex Gaussian with zero mean and unit variance.

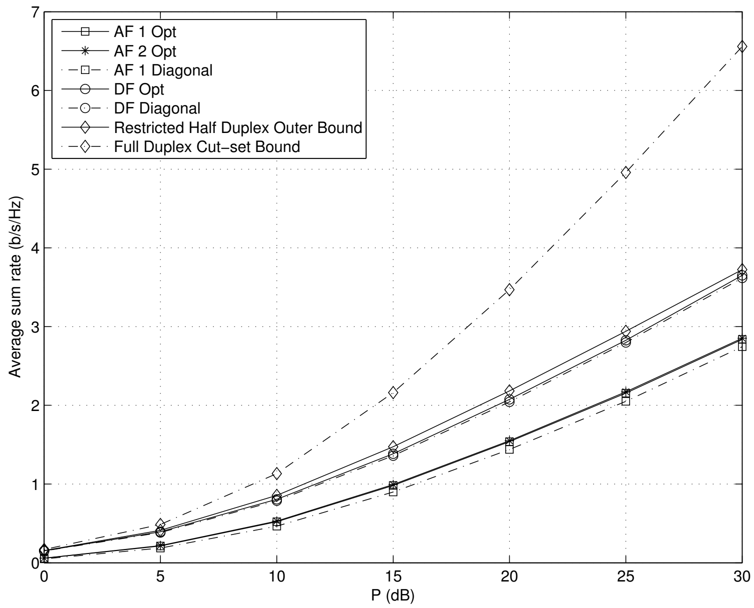


(a) $\gamma = 0.2$

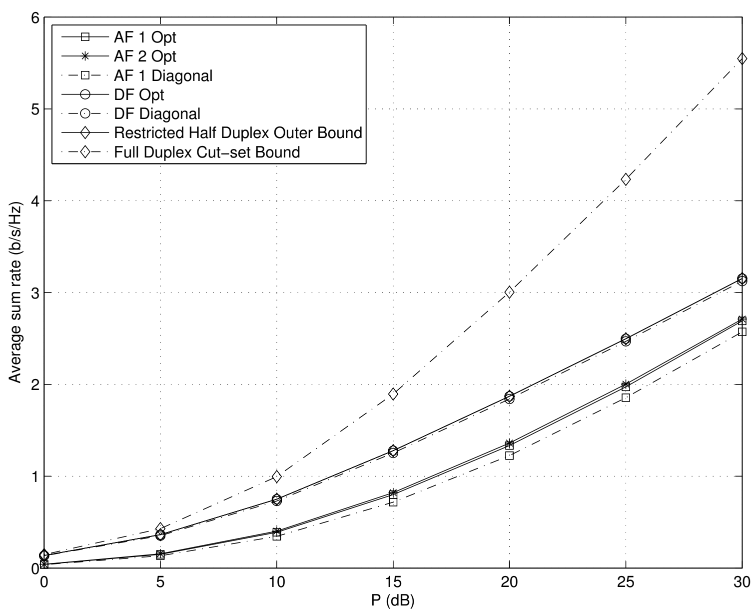


(b) $\gamma = 0.8$

Fig. 2. Achievable sum rate comparison of different protocols as a function of P in a 3-way symmetric relay network, where the channel between source 1 and the relay has unit power gain, the other two channels have -3 dB power gain, γP denotes the overall source power and $(1-\gamma)P$ the relay power, respectively. The relay has $M = 2$ antennas, and all channel gains are complex Gaussian with zero mean and unit variance.



(a) $\gamma = 0.2$



(b) $\gamma = 0.8$

Fig. 3. Achievable sum rate comparison of different protocols as a function of P in a 3-way symmetric relay network, where the channel between source 1 and the relay has -3 dB power gain, the other two channels have unit power gain, γP denotes the overall source power and $(1-\gamma)P$ the relay power, respectively. The relay has $M = 2$ antennas, and all channel gains are complex Gaussian with zero mean and unit variance.

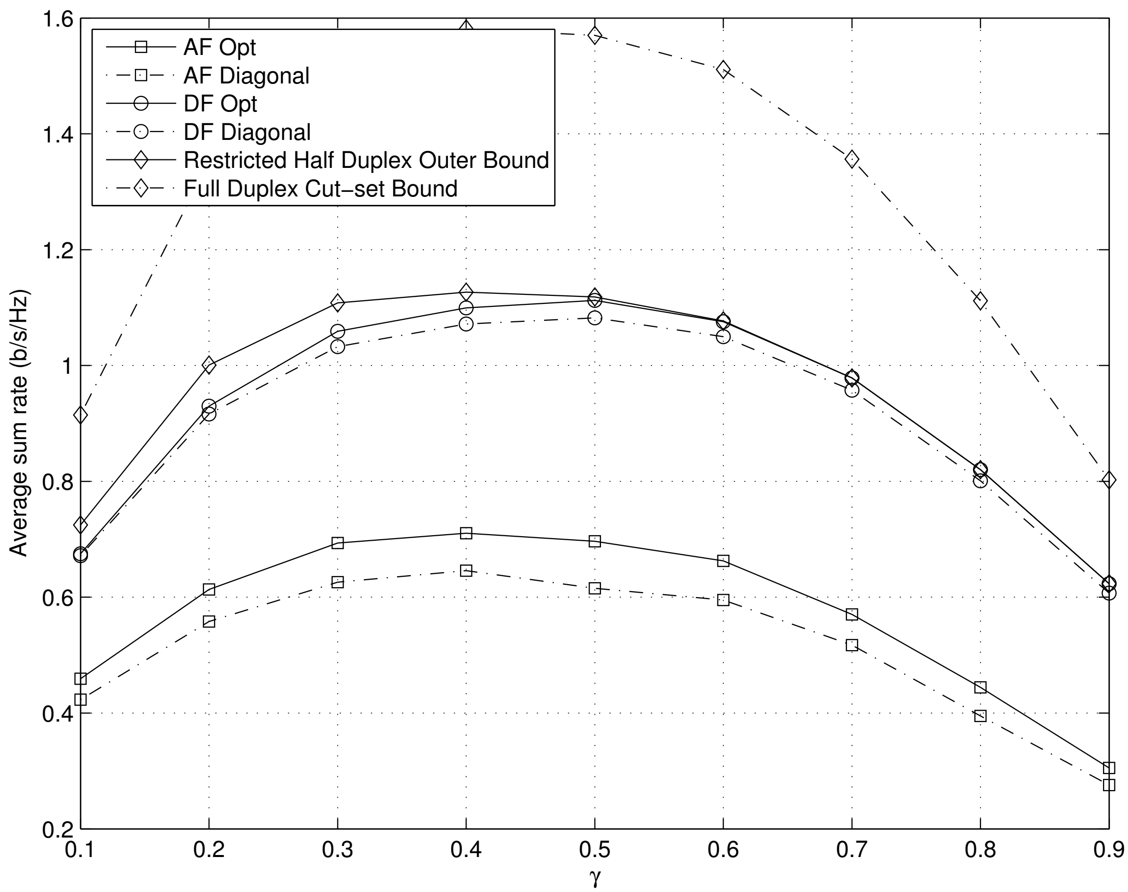


Fig. 4. Achievable sum rate comparison for different protocols as a function of γ in a 3-way symmetric relay network with $P = 10$ and all channels have unit variance. The relay has $M = 2$ antennas, and all channel gains are complex Gaussian with zero mean and unit variance.

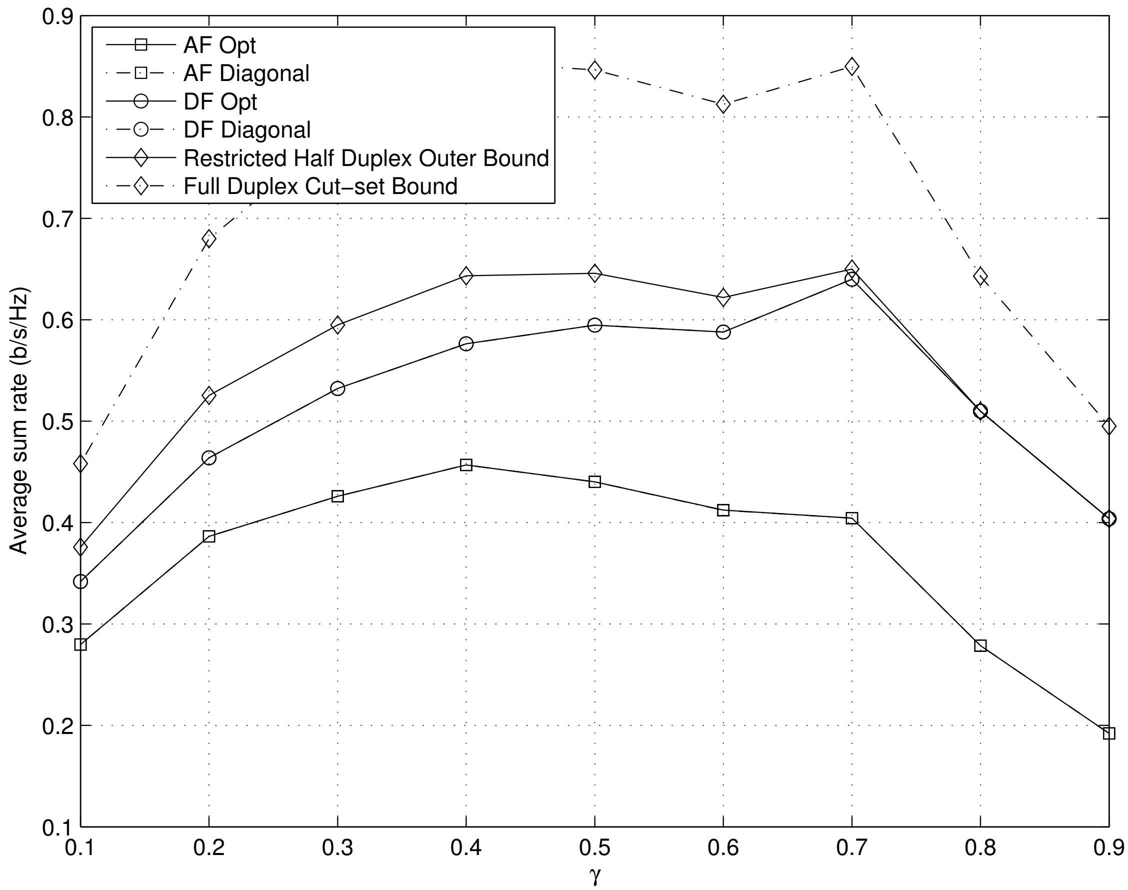
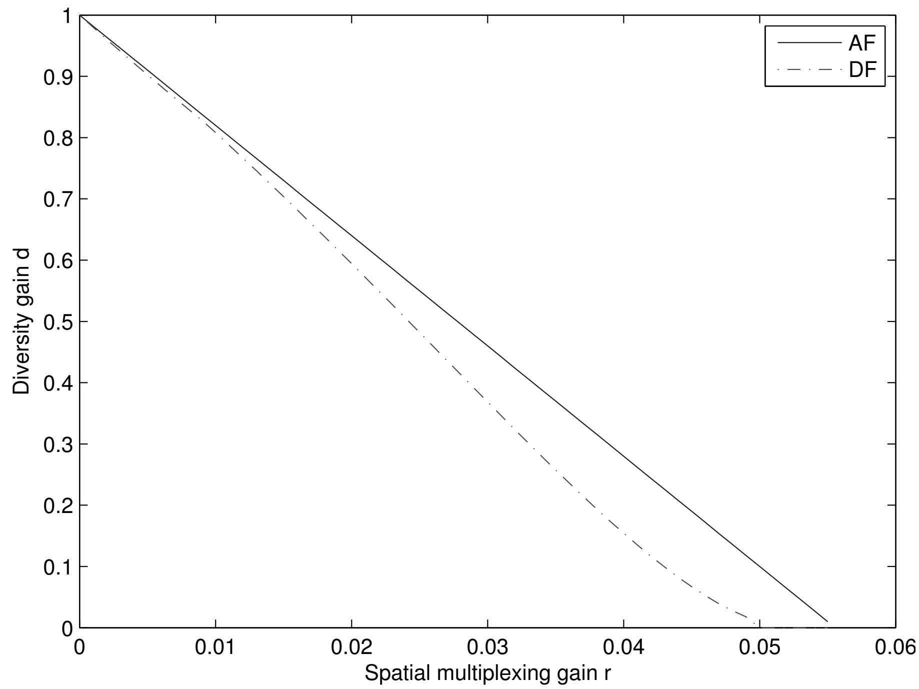
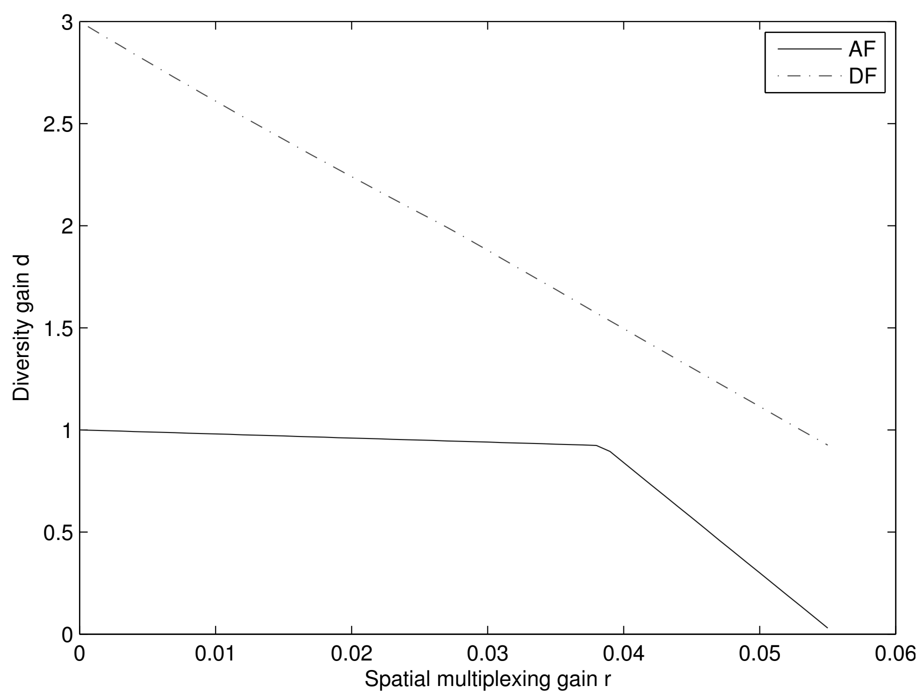


Fig. 5. Achievable sum rate comparison for different protocols as a function of γ in a 3-way relay network with $P = 10$. The relay has $M = 1$ antennas, and all channel gains are complex Gaussian with zero mean and unit variance.



(a) $M = 1$



(b) $M = 3$

Fig. 6. Diversity-multiplexing tradeoff comparison between AF and DF protocols in a 10-way relay network for different number of antennas M.

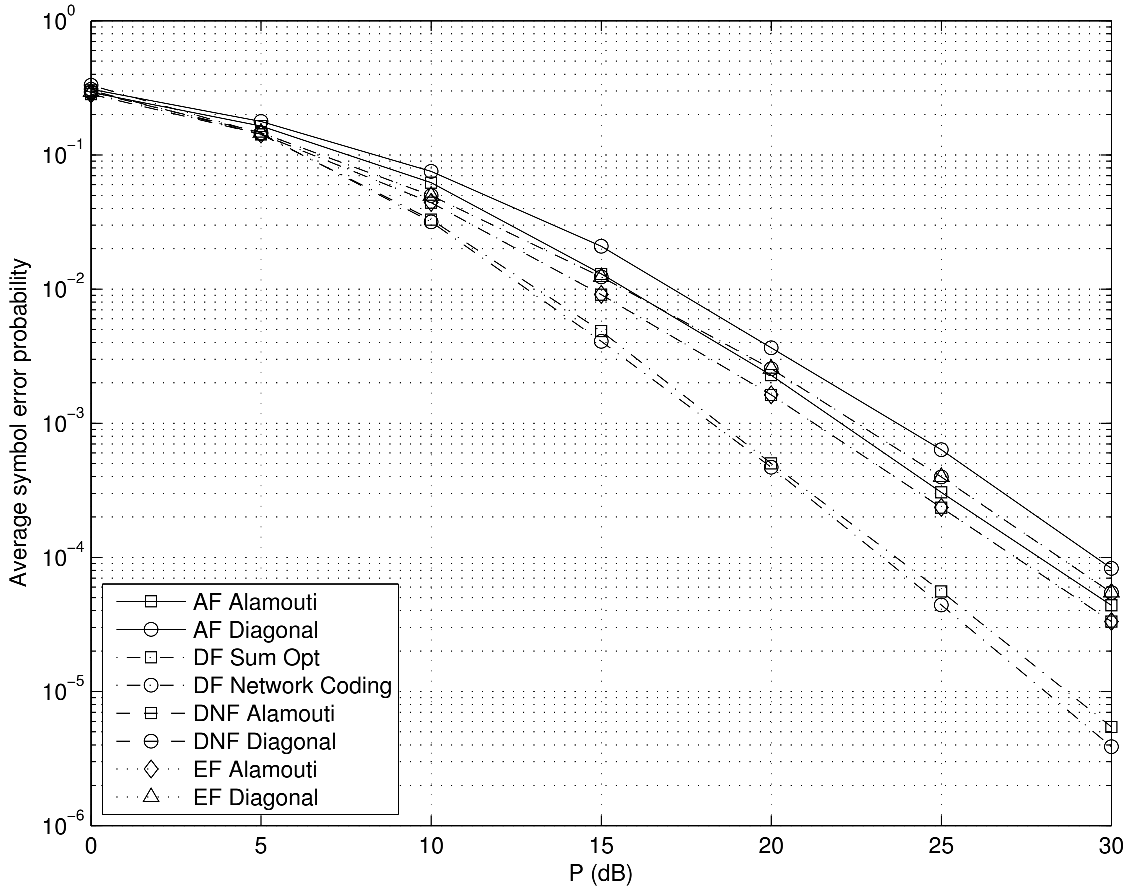


Fig. 7. Average symbol error probability comparison for different protocols as a function of P in a 3-way relay network. The relay has $M = 2$ antennas. All channel gains are complex Gaussian with zero mean and unit variance.

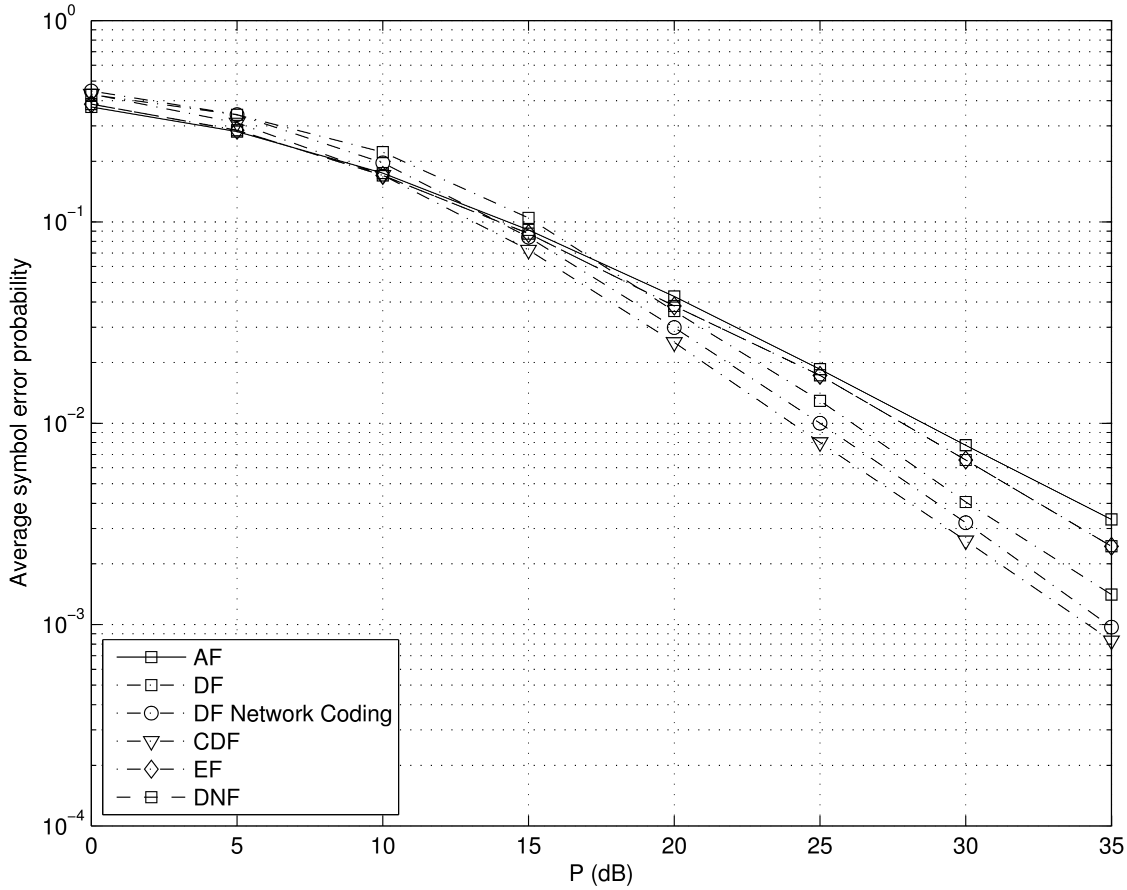


Fig. 8. Average symbol error probability comparison for different protocols as a function of P in a 3-way relay network. The relay has $M = 1$ antennas. All channel gains are complex Gaussian with zero mean and unit variance.