Space-Time Communication Protocols for N-way Relay Networks

Tao Cui and Tracey Ho
Department of Electrical Engineering
California Institute of Technology
Pasadena, CA 91125, USA
Email: {taocui, tho}@caltech.edu

Jörg Kliewer
Klipsch School of Electrical and Computer Engineering
New Mexico State University
Las Cruces, NM 88003, USA
Email: jkliewer@nmsu.edu

Abstract— We address communication protocols for N-way relay networks with M antennas at the relay and a single antenna at the N source terminals. In particular, amplify-and-forward (AF), decode-and-forward (DF), and compress-and-forward (CF) strategies are extended to these networks, and in addition, two new relaying protocols, denoise-and-forward and estimate-and-forward, are proposed. In the first part of the paper, the performance of these schemes is analyzed in terms of the achievable rate region. Also, the optimal diversity-multiplexing tradeoff is derived for both AF and DF. The second part of the paper is devoted to practical space-time transmission strategies. Linear dispersion codes are used, which are optimized by maximizing the sum rate. For AF a diversity order of close to M can be achieved by using a specific space-time code construction.

I. INTRODUCTION

The N-way relay network (NWRN) is a promising wireless network architecture for applications such as wireless teleconferencing, where there are N source terminals in the network and each source terminal needs to exchange information between all other terminals with the help of a relay node. The communication is carried out in two transmission phases, a multiple access channel (MAC) phase where the source terminals send data to the relay in the uplink, and a broadcast phase for the downlink. When $N=2$, the N-way relay network degenerates to the two-way relay channel (TWRC), which has drawn renewed interest [1]–[4] recently. In [1], AF and DF protocols for one-way relay channels in [5] are extended to the half-duplex Gaussian TWRC. Performance analysis and relay function design are discussed in [2], [3] for TWRC with a single relay without memory. In [4], distributed space time coding is proposed in TWRC with multiple relays, where several space time protocols are proposed with different time slots.

In this paper, the achievable rate regions for different protocols such as amplify-and-forward (AF), decode-and-forward (DF), and compress-and-forward (CF) are derived for NWRNs. We also characterize the fundamental diversity-multiplexing tradeoff [6], [7]. Several practical space-time coding protocols are also proposed. Linear dispersion codes [8] are used at the relay, which can be optimized by maximizing the sum capacity. Since each source terminal already knows the signal it has just transmitted, the received signal at the relay can be compressed by network coding to reduce bandwidth usage. A simple space time code is proposed where at any time

slot at most one antenna is active. By analyzing the pairwise error probability (PEP), we find that this code achieves a diversity order $M\left(1-\frac{\log(1+\log P)}{\log P}\right)$, where M is the number of antennas at the relay and P is the total power of the NWRN. By using an Alamouti type space time code with $M=2$, we find that the diversity order two can be attained by AF. Moreover, two new relaying protocols: denoise-and-forward (DNF) and estimate-and-forward (EF) are also proposed.

II. SYSTEM MODEL

We consider an N-way relay network with N source terminals and one relay node. The case of multiple relays is discussed in [9]. There are M antennas on the relay node while each source terminal only has a single antenna. Each source terminal wants to send its data to all the other source terminals. We consider a half duplex system with time division multiplexing. Given time interval T, αT is used for uplink transmission from the source terminals to the relay and $(1-\alpha)T$ is spent for the downlink broadcast from the relay to the source terminals. We further assume that there does not exist direct transmission between 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

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

where h_i^m is the channel fading between source i and the m-th antenna of the relay (we assume reciprocal channels), $w_{r,t}^m$ and $w_{i,t}$ are additive white Gaussian noise (AWGN) realizations with distribution $\mathcal{CN}(0,1)$. The transmitted signals $x_{i,t}$ and $x_{r,t}$ are subject to average power constraints $\frac{1}{T}\sum_{t=1}^{\lfloor \alpha T \rfloor}|x_{i,t}|^2 \leq P_i$, $\forall i=1,...,N$, and $\frac{1}{T}\sum_{t=\lfloor \alpha T \rfloor+1}^{T}|x_{r,t}^m|^2 \leq P_r$. We assume that h_i^m is known perfectly at all the nodes for simplicity and does not change during interval T. The results in this paper also generalize to full duplex transmission by setting $\alpha=1/2$ and doubling the achievable rate regions in Section III.

III. ACHIEVABLE RATE REGIONS

In the following, we assume that the data rate of source terminal i is R_i. The key feature of an N-way relay channel is that each source terminal knows its transmitted signal as side information, which may potentially improve the data rate.

A. Amplify-and-Forward

In AF, we assume $\alpha=\frac{1}{2}$. The relay precodes the received signals $y_{r,t}^n$ at time t from antenna n, $n=1,...,M$, by multiplying them with a precoding matrix $\Psi=[\psi^{m,n}]$. The transmitted signal from antenna m at time $t+T/2$ is

$$x_{r,t+\frac{T}{2}}^m=\sum_{n=1}^{M}\psi^{m,n}y_{r,t}^n, \quad m=1,...,M, \quad \text{subject to}$$

$$\sum_{i=1}^{N}\sum_{m=1}^{M}\left(2\left|\sum_{n=1}^{M}\psi^{m,n}h_i^n\right|^2 P_i+\sum_{n=1}^{M}|\psi^{m,n}|^2\right)\leq 2P_r. \tag{2}$$

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

$$y_{i,t+\frac{T}{2}}=\sum_{m=1}^{M}h_i^m x_{r,t+\frac{T}{2}}^m+w_{i,t+\frac{T}{2}}$$

$$=\sum_{j=1}^{N}\left(\sum_{m=1}^{M}\sum_{n=1}^{M}\psi^{m,n}h_i^m h_j^n\right)x_{j,t}+\sum_{m=1}^{M}h_i^m\sum_{n=1}^{M}\psi^{m,n}w_{r,t}^n+w_{i,t+\frac{T}{2}}. \tag{3}$$

By canceling the contribution of $x_{i,t}$ from $y_{i,t+\frac{T}{2}}$, we obtain a MAC with $N-1$ users, whose capacity region is given by

$$\sum_{j\in\mathcal{S}}R_j\leq\frac{1}{4}\log\left(1+\frac{2\sum_{j\in\mathcal{S}}\left|\sum_{m=1}^{M}\sum_{n=1}^{M}\psi^{m,n}h_i^m h_j^n\right|^2 P_j}{\sum_{n=1}^{M}\left|\sum_{m=1}^{M}h_i^m\psi^{m,n}\right|^2+1}\right) \tag{4}$$

for all $\mathcal{S}\subseteq\mathcal{S}_i$ where $\mathcal{S}_i=\{1,...,N\}/i$. The achievable rate region is the convex hull of the intersection of the regions (4) for all i. A more general AF can be found in [9], where α is not restricted to $\alpha=\frac{1}{2}$.

B. Decode-and-Forward

In DF, the relay first decodes the signals from the source terminals during time interval αT with a MAC uplink channel. It then re-encodes the resulting signal and broadcasts the encoded signal. To decode the signals at the relay, the source rates should satisfy

$$\sum_{j\in\mathcal{S}}R_j\leq\frac{\alpha}{2}\log\det\left(\mathbf{I}_M+\sum_{j\in\mathcal{S}}\frac{P_j}{\alpha}\mathbf{h}_j\mathbf{h}_j^H\right),\forall\mathcal{S}\subseteq\{1,...,N\}, \tag{5}$$

where $\mathbf{h}_j=[h_j^1,...,h_j^M]^T$. The downlink channel can be considered to be a broadcast channel with side information, where each source knows the signal it just sent. We consider a general scenario where a source needs to communicate messages $W_1,...,W_N$ to N receivers and receiver i already knows W_i. At the relay, W_i is first encoded using codewords $X_i^T=\{X_{i,1},...,X_{i,T}\}$ of length T with rate R_i. The transmitted signal X^T is chosen to be the vector $(X_1,...,X_N)^T$. The received signal at receiver i is denoted as Y_i^T. According to [10, Theorem 1], the set of rates $\{R_i\}$ is achievable iff

$$H(X|W_i)=\sum_{j=1,j\neq i}^{N}R_j\leq I(Y_i;X), \quad i=1,...,N. \tag{6}$$

The signal after re-encoding is precoded with a matrix $\Phi=[\psi^{m,n}]$ before transmitting over different antennas at the relay, i.e., $\sum_{n=1}^{M}\psi^{m,n}x_{r,t}^n$ where $x_{r,t}^n$'s are generated independently with unit average power. Thus, the achievable rate region is

$$\sum_{j\in\mathcal{S}}R_j\leq\frac{\alpha}{2}\log\det\left(\mathbf{I}_M+\sum_{j\in\mathcal{S}}\frac{P_j}{\alpha}\mathbf{h}_j\mathbf{h}_j^H\right), \quad \forall\mathcal{S}\subseteq\{1,...,N\},$$

$$\sum_{j=1,j\neq i}^{N}R_j\leq\frac{1-\alpha}{2}\log\left(1+\sum_{n=1}^{M}\left|\sum_{m=1}^{M}h_i^m\psi^{m,n}\right|^2\right), \quad i=1,...,N, \tag{7}$$

where we require that $\sum_{m=1}^{M}\sum_{n=1}^{M}|\psi^{m,n}|^2=\frac{P_r}{1-\alpha}$.

We can show that (5) is achievable by using network coding [11] at the physical layer. In general (7) is larger than the rate region by superimposing decoded signals without network coding. Here, network coding acts as a compressor.

C. Compress-and-Forward

CF has been first mentioned in [5] for the relay channel. In our case, a quantization codebook with $2^{T\hat{R}}$ codewords $\{\hat{y}_i^{\lfloor\alpha T\rfloor}\}$ is generated at the relay according to pdf $p_{\hat{Y}}(\hat{y})$. On receiving $y_r^{\lfloor\alpha T\rfloor}=\{y_{r,1},...,y_{r,\lfloor\alpha T\rfloor}\}$, the relay first finds a codeword $\hat{y}_i^{\lfloor\alpha T\rfloor}$ that is jointly typical with $y_r^{\lfloor\alpha T\rfloor}$. The index of $\hat{y}_i^{\lfloor\alpha T\rfloor}$ is then encoded using a codebook $\{x_i^{\lfloor\alpha T\rfloor}\}$. To guarantee that there exists at least one codeword $\hat{y}_i^{\lfloor\alpha T\rfloor}$ that is jointly typical with $y_r^{\lfloor\alpha T\rfloor}$, we require that

$$\hat{R}\geq\alpha I(Y_r;\hat{Y}). \tag{8}$$

To decode $\hat{Y}$ correctly at source terminal i, by using (6), the downlink transmission requires that

$$H(\hat{Y}|X_i)\leq\frac{1-\alpha}{2}\log\left(1+\sum_{n=1}^{M}\left|\sum_{m=1}^{M}h_i^m\psi^{m,n}\right|^2\right), \tag{9}$$

where $\psi^{m,n}$ is defined as in (7). Given $\hat{Y}$, we can decode X_j, $j\neq i$ if and only if

$$\sum_{j\in\mathcal{S}}R_j\leq(1-\alpha)I(\hat{Y};X_{\mathcal{S}}|X_{\mathcal{S}_i/\mathcal{S}},X_i),\mathcal{S}\subseteq\mathcal{S}_i, \tag{10}$$

where $X_{\mathcal{S}}=\{X_i|i\in\mathcal{S}\}$ and $\mathcal{S}_i=\{1,...,N\}/i$.

One way to choose $\hat{Y}$ is by generating $\{\hat{y}_k^{\lfloor\alpha T\rfloor}\}$ using $y_r^{\lfloor\alpha T\rfloor}+\hat{w}_k^{\lfloor\alpha T\rfloor}$, where the entries of $\hat{w}_k^{\lfloor\alpha T\rfloor}$ are i.i.d. AWGN with zero mean and variance $\hat{\sigma}_r^2$ (the entries of $\hat{w}_k^{\lfloor\alpha T\rfloor}$ may have different variances). We thus obtain $H(\hat{Y}|X_i)=\hat{R}-R_i$,

$$I(Y_r;\hat{Y})=\frac{1}{2}\log\det\left(\frac{(1+\hat{\sigma}_r^2)\mathbf{I}_M+\sum_{j=1}^{N}\frac{P_j}{\alpha}\mathbf{h}_j\mathbf{h}_j^H}{\hat{\sigma}_r^2}\right), \tag{11}$$

$$I(\hat{Y};X_{\mathcal{S}}|X_{\mathcal{S}_i/\mathcal{S}},X_i)=\frac{1}{2}\log\det\left(\mathbf{I}_M+\frac{\sum_{j\in\mathcal{S}}\frac{P_j}{\alpha}\mathbf{h}_j\mathbf{h}_j^H}{\hat{\sigma}_r^2}\right),\forall\mathcal{S}\subseteq\mathcal{S}_i. \tag{12}$$

Therefore, the achievable rate region is given by

$$R_i\geq\frac{\alpha}{2}\log\det\left(\frac{(1+\hat{\sigma}_r^2)\mathbf{I}_M+\sum_{j=1}^{N}\frac{P_j}{\alpha}\mathbf{h}_j\mathbf{h}_j^H}{\hat{\sigma}_r^2}\right)$$

$$-\frac{1-\alpha}{2}\log\left(1+\sum_{n=1}^{M}\left|\sum_{m=1}^{M}h_i^m\psi^{m,n}\right|^2\right), \tag{13}$$

$$\sum_{j\in\mathcal{S}}R_j\leq\frac{1-\alpha}{2}\log\det\left(\mathbf{I}_M+\frac{\sum_{j\in\mathcal{S}}\frac{P_j}{\alpha}\mathbf{h}_j\mathbf{h}_j^H}{1+\hat{\sigma}_r^2}\right),\forall\mathcal{S}\subseteq\mathcal{S}_i.$$

In [9], we also show how to choose R_i and how to determine $\psi^{m,n}$ using utility maximization.

IV. Diversity-Multiplexing Tradeoff

In this section, we characterize the fundamental tradeoff between the diversity and multiplexing gain [6], [7] in N-way relay networks using AF and DF. In the following, we assume that $\sum_{i=1}^{N}P_i+P_r=P$ with the overall power P, $P_1=\cdots=P_N=P_s$ and $NP_s=P_r$, i.e., $P_s=\frac{P}{2N}$ and $P_r=\frac{P}{2}$ and all channel gains are complex Gaussian. As there are N sources in the network, we consider the symmetric case as in [7],

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 transmission strategy. The strategy is said to achieve spatial multiplexing gain r and diversity gain d if

$$\lim_{P \to +\infty} \frac{R^i(P)}{\log P} \geq r, \quad \lim_{P \to +\infty} \frac{P_e^i(P)}{\log P} \leq -d, \quad \forall i = 1,...,N. \quad (14)$$

In AF, by following [6], [7], we obtain the diversity gain

$$d = \begin{cases} M(1-2|\mathcal{S}|r), & \text{if } |\mathcal{S}| > M, \\ |\mathcal{S}|(1-2|\mathcal{S}|r), & \text{if } |\mathcal{S}| \leq M. \end{cases}, \quad (15)$$

where $\mathcal{S} \subseteq \mathcal{S}_i$. Note that the second case is concave in $|\mathcal{S}|$ and its minimum is attained at its boundary $|\mathcal{S}|=1$ or $|\mathcal{S}|=M$. The first case is a decreasing function in $|\mathcal{S}|$. Thus, the worst case d is attained at $|\mathcal{S}|=1$ or $|\mathcal{S}|=N-1$. Finally, we obtain

$$d_{\text{AF}} = \begin{cases} M(1-2(N-1)r), & \text{if } r \geq \frac{M-1}{2(M(N-1)-1)}, \\ 1-2r, & \text{if } r < \frac{M-1}{2(M(N-1)-1)}. \end{cases} \quad (16)$$

From (16), we can see that when the system is lightly loaded $r < \frac{M-1}{2(M(N-1)-1)}$, single user performance is achieved.

In DF, the uplink channel is a MAC, whose diversity multiplexing tradeoff is obtained in [7], i.e.,

$$d^{\text{up}} = \begin{cases} \left(1-\frac{r}{\alpha}\right)\left(M-\frac{r}{\alpha}\right), & \text{if } r \leq \alpha \min\left(1, \frac{M}{N+1}\right), \\ N\left(1-\frac{r}{\alpha}\right)\left(M-\frac{Nr}{\alpha}\right), & \text{if } r > \alpha \min\left(1, \frac{M}{N+1}\right), \end{cases} \quad (17)$$

where $2r$ is due to half duplex. From (7) the downlink channel can be considered to be a MISO channel whose diversity multiplexing tradeoff can be easily obtained as in [6], i.e.,

$$d^{\text{down}} = \left(1-(N-1)\frac{r}{1-\alpha}\right)\left(M-(N-1)\frac{r}{1-\alpha}\right). \quad (18)$$

The diversity gain is the minimum diversity gain of the uplink (17) and the downlink (18), i.e., $d_{\text{DF}} = \min\{d^{\text{up}}, d^{\text{down}}\}$. When $r \leq \alpha \min\left(1, \frac{M}{N+1}\right)$, it is easy to see that the maximum diversity gain is attained at $\frac{\alpha}{1-\alpha} = \frac{1}{N-1}$, or $\alpha = \frac{1}{N}$. When $r > \alpha \min\left(1, \frac{M}{N+1}\right)$, the optimal α depends on M, N, r in a complicated way. We can find the optimal α by numerical analysis. Note that the transmission protocols in Section V correspond to the first case or a lightly loaded system as the rate is fixed by using a specific constellation. Therefore, we should choose $\alpha = \frac{1}{N}$.

V. PRACTICAL SPACE TIME PROTOCOLS

The devised transmission protocols in this section parallel those 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 time slots after processing its received signals, where $\mathcal{Q}$ is a finite constellation set with average power one. We assume that the channel remains static during $K+1$ time slots. At the relay, the received signal at the m-th antenna at the end of the first time slot is

$$y_r^m = \sum_{i=1}^{N} \sqrt{(K+1)P_i} h_i^m s_i + w_r^m, \quad m=1,...,M. \quad (19)$$

The protocols can be classified into two classes: compression based protocols where the relay compresses its received signal by reducing the alphabet size of constellation seen by the relay, and non-compression based protocols. Different from [4]

where coding is performed over symbols received at different time, we apply space time coding over symbols received by different antennas at the relay.

A. Non-Compression Based Protocols

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

$$x_{r,k}^m = \sum_{n=1}^{M} \psi_k^{m,n} y_r^n, \quad k=1,...,K, \quad (20)$$

which is actually a special case of linear dispersion code [8]. Let $y_{i,k}$ be the received signal of source i at time slot $k+1$. We can write $y_{i,k}$ as

$$y_{i,k} = \mathbf{h}_i^T \Psi_k \mathbf{H} \Lambda \mathbf{s} + \mathbf{h}_i^T \Psi_k \mathbf{w}_r + w_{i,k}, \quad (21)$$

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

$$\left((K+1)\sum_{j=1}^{N} P_j + 1\right) \text{tr}\left(\Psi_k \Psi_k^H\right) = \frac{K+1}{K} P_r. \quad (22)$$

Note that (21) can be further written as

$$\mathbf{y}_i = \mathbf{G}_i \mathbf{H} \Lambda \mathbf{s} + \mathbf{G} \mathbf{w}_r + \mathbf{w}_i, \quad (23)$$

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

$$\hat{\mathbf{s}} = \underset{\{\tilde{\mathbf{s}} | \tilde{s}_j \in \mathcal{Q}, j \neq i, \tilde{s}_i = s_i\}}{\arg\min} \left\| (\mathbf{G}_i \mathbf{G}_i^H + \mathbf{I}_K)^{-\frac{1}{2}} (\mathbf{y}_i - \mathbf{G}_i \mathbf{H} \Lambda \mathbf{s}) \right\|^2. \quad (24)$$

By using the Chernoff bound, the PEP between $\mathbf{s}_1$ and $\mathbf{s}_2$ at source terminal i can be bounded as

$$\text{PEP} \leq E\left\{ e^{\lambda(\log p(\mathbf{y}_i|\mathbf{s}_1) - \log p(\mathbf{y}_i|\mathbf{s}_2))} \right\}$$
$$= E_{\mathbf{h}_i}\left\{ \det^{-1}\left(\mathbf{I}_M + \frac{\|\Delta\mathbf{s}\|^2}{4} \mathbf{G}_i^H (\mathbf{G}_i \mathbf{G}_i^H + \mathbf{I}_K)^{-1} \mathbf{G}_i\right) \right\}, \quad (25)$$

where $p(\mathbf{y}_i|\mathbf{s})$ is the pdf of $\mathbf{y}_i$ given $\mathbf{s}$, and $\Delta\mathbf{s} = \Lambda(\mathbf{s}_1 - \mathbf{s}_2)$ without the i-th entry.

We consider two special forms of space time coding. i) We choose Ψ_k to be an all zero matrix except the (k,k)-th entry is $\sqrt{\gamma}$, $k=1,...,\min\{K,M\}$. This corresponds to the case that only the k-th antenna is active at time slot k. The relay scales its received signal from antenna k and transmits it over antenna k. We assume that the total power in the network is P, i.e., $\sum_{i=1}^{N} P_i + P_r = P$. To minimize the PEP, we find that $MNP_s = P_r$ or $P_s = \frac{P}{N(M+1)}$. In this case we obtain from (25)

$$\text{PEP} \lesssim \left(\frac{4N(M+1)}{d^2}\right)^M P^{M\left(1-\frac{\log(1+\log P)}{\log P}\right)}, \quad (26)$$

where d is the minimum distance of constellation $\mathcal{Q}$, which achieves a diversity order $M\left(1-\frac{\log(1+\log P)}{\log P}\right)$.

ii) We consider an Alamouti-type code with $M=K=2$. Note that we can also transmit $(y_r^m)^*$ instead of (y_r^m) in (20). By using an Alamouti-type code, we have

$$x_{r,1}^1 = \sqrt{\gamma} y_r^1, x_{r,1}^2 = \sqrt{\gamma} y_r^2, x_{r,2}^1 = \sqrt{\gamma} (y_r^2)^*, x_{r,2}^2 = -\sqrt{\gamma} (y_r^1)^*, \quad (27)$$

where γ satisfies $M^2 \gamma ((K+1)\sum_i^N P_i + 1) = (K+1)P_r$. In this case, the PEP can be obtained by using a similar derivation as in (26). We can show that by using Alamouti code, a diversity order of two can be achieved.

2) Decode-and-Forward: In DF, after receiving the signals from the source terminals, the relay jointly decodes the all source terminals' messages using an ML decoder

$$\{\hat{s}_1,...,\hat{s}_N\}=\underset{s_1,...,s_N}{\operatorname{argmin}}\sum_{m=1}^{M}\left|y_r^m-\sum_{i=1}^{N}h_i^m s_i\right|^2, \quad (28)$$

which can be achieved by using a sphere decoder [12]. By (28), the uplink achievable diversity order is M.

After obtaining $\hat{s}_1,...,\hat{s}_N$ from (28), there are two ways to send these symbols. The first approach does not employ network coding. The downlink channel becomes a multiple input and single output (MISO) channel. We use a linear dispersion code [8] to code $s_1,...,s_N$. The transmitted signal from the relay can be written as a K by M matrix

$$\mathbf{X}_r=\sum_{i=1}^{N}(s_i\mathbf{C}_i+s_i^*\mathbf{D}_i). \quad (29)$$

The average power constraint requires that $E\{\operatorname{tr}(\mathbf{X}_r\mathbf{X}_r^H)\}=(K+1)P_r$. The design of $\mathbf{C}_i$ and $\mathbf{D}_i$ is similar to that in [8]. The only difference is that here we need to maximize the sum rate. Let $\mathbf{s}=[s_1,...,s_N]^T$. The ML decoder for s_j, $j\neq i$, at each source terminal can be readily obtained as in [8] and can be implemented by using a sphere decoder [12].

A simple realization of (29) is to transmit $\mathbf{U}_m\mathbf{s}$ over antenna m, $m=1,...,M$, where $\mathbf{U}_m$ is a K by N unitary matrix and $\mathbf{U}_m$ can be chosen randomly to ensure distributed design. Another way to use linear dispersion code is to generate $N-1$ symbols $\tilde{s}_1,...,\tilde{s}_{N-1}$, where $\tilde{s}_i=\sum_{j=1}^{N}\varphi_{i,j}s_j$ is a linear combination of decoded $s_1,...,s_N$. Existing space time codes can be applied on $\tilde{s}_i$. We only need to optimize the linear coefficients $\varphi_{i,j}$ to maximize the sum rate.

The second approach is to use network coding. As source i already knows s_i, it is a waste of bandwidth to transmit a linear combination of s_i. We apply network coding on a finite field $\mathbb{F}$ of size $q=\max\{|\mathcal{Q}|,N\}$ [11]. Let χ_i be the element in $\mathbb{F}$ is corresponding to the index of s_i in $\mathcal{Q}$. We then choose $\xi_{i,j}\in\mathbb{F}$ such that for the given $\chi_{i'}$ and $\eta_i=\sum_{j=1}\xi_{i,j}\chi_j$, $i=1,...,N_1$ we can decode χ_i, $\forall i\neq i'$. For example, we can choose $\xi_{i,j}=1$ if $j=i,i+1 \mod N$ and 0 otherwise. It is easy to verify that by deleting any column from the Toeplitz matrix the resulting matrix is invertible. This Toeplitz matrix also suggests that a modular group is sufficient instead of a finite field, which exists for all $|\mathcal{Q}|$. At the relay we want to transmit $\tilde{x}_i\in\mathcal{Q}_r$ by using a linear dispersion code as (29), $i=1,...,N-1$, where the index of $\tilde{x}_i$ in $\mathcal{Q}_r$ corresponds to η_i. By the construction of η_i, the source terminals can recover the signals correctly. Note that by using network coding, the number of symbols need to transmit is reduced by one and the transmit symbols are superimposed in finite fields rather than in real algebra, which reduces power consumption.

3) Denoise-and-Forward: Denoise-and-forward is similar to DF. At the relay, $s_1,...,s_N$ are jointly decoded using the ML decoder (28). Different from DF where $\hat{s}_1,...,\hat{s}_N$ are transmitted using a linear dispersion code directly, DNF reconstructs $z_r^m=\sum_{i=1}^{N}h_i^m\hat{s}_i$. As in AF, the transmitted signal over the m-th antenna at time slot $k+1$ is

$$x_{r,k}^m=\sum_{n=1}^{M}\psi_k^{m,n}z_r^n, \quad k=1,...,K. \quad (30)$$

Also, the ML decoding and precoder design are similar to AF.

4) Estimate-and-Forward: At each received antenna, we first use a 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.,

$$\min_{g^m}E_{\mathbf{s}}\left\{\left|\sum_{i=1}^{N}h_i^m s_i-g^m(y_r^m)\right|^2\right\}, \quad (31)$$

which has the solution

$$g^m(y_r^m)=\frac{\sum_{\mathbf{s}\in\mathcal{Q}^N}\sum_{i=1}^{N}h_i^m s_i e^{-\left|y_r^m-\sum_{i=1}^{N}h_i^m s_i\right|^2}}{\sum_{\mathbf{s}\in\mathcal{Q}^N}e^{-\left|y_r^m-\sum_{i=1}^{N}h_i^m s_i\right|^2}}. \quad (32)$$

Let $z_r^m=\beta g^m(y_r^m)$, where β is a constant to maintain the average power 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 $k+1$. We have considered minimizing the estimation error at each antenna. Given $\psi_k^{m,n}$, we can also minimize the MSE of symbols at source terminals. When applying EF to Gaussian codebooks, we find that EF reduces to AF and that the achievable rate region is identical to that of AF.

B. Compression Based Protocols

We mainly consider compression based DF (CDF), where other protocols can be extended similarly. To motivate the design principle, we consider a two-way relay network with a single antenna at the relay, i.e., $N=2$ and $M=1$. The received signal at the relay is $y_r=h_1x_1+h_2x_2+w_r$. Let $\mathcal{Q}_q$ and $\mathcal{Q}_t$ be the quantization and transmission constellations, respectively. There exists a mapping that maps $x_q\in\mathcal{Q}_q$ to $x_t\in\mathcal{Q}_t$, i.e., $x_t=\mathcal{M}(x_q)$. On receiving y_r, the relay first finds that

$$x_q=\underset{x\in\mathcal{Q}_q}{\operatorname{argmin}}|y_r-x|^2. \quad (33)$$

The relay then transmits $x_t=\mathcal{M}(x_q)$. If we choose $\mathcal{Q}_q=\{h_1x_1+h_2x_2|x_1,x_2\in\mathcal{Q}\}$, CF is similar to DNF by decoding $h_1x_1+h_2x_2$ first. But different from DNF which transmits a scaled version of x_q, CF transmits x_t instead. If $\mathcal{M}(x)=\beta x$, then CF reduces to DNF but clearly this is not the optimal choice of $\mathcal{Q}_q$ and $\mathcal{Q}_t$. In the following, we always choose $\mathcal{Q}_q=\{h_1x_1+h_2x_2|x_1,x_2\in\mathcal{Q}\}$. The problem now becomes how to choose $\mathcal{M}(\cdot)$ and $\mathcal{Q}_t$. Note that an important difference between CDF in this subsection and the traditional CF in one-way relay channel [5] is that the size of $\mathcal{Q}_t$ can be smaller than that of $\mathcal{Q}_q$ due to the availability of side information at the receivers. This approach can be interpreted as network coding at the physical layer, where the size of the constellation is reduced due to the inherent compression in network coding.

To ensure correct decoding, we need to make sure that given x_t and x_1, there exists a unique x_2 such that $x_t=\mathcal{M}(h_1x_1+h_2x_2)$. Therefore, we can construct a graph $\mathcal{G}$, where each vertex in $\mathcal{G}$ corresponds to a point in $\mathcal{Q}_q$. There is an edge between $h_1x_1+h_2x_2$ and $h_1x_1+h_2x_2'$ and an edge between $h_1x_1+h_2x_2$ and $h_1x_1'+h_2x_2$, $\forall x_1\neq x_1', x_2\neq x_2'$. Hence, a valid

mapping $\mathcal{M}(\cdot)$ corresponds to a coloring of $\mathcal{G}$. To find the optimal constellations, we can exhaust all possible colorings of $\mathcal{G}$ and for each coloring we can find $\mathcal{Q}_t$ to maximize $\min\{|h_1|^2 d_{2,t}^2, |h_2|^2 d_{1,t}^2\}$ under the average power constraint at the relay, where $d_{i,t}$ is the minimum distance between any two points in $\mathcal{Q}_t$ given x_i. The optimal $\mathcal{M}(\cdot)$ is the one that achieves the minimum average end-to-end error probability. Note that the above approach can be extended to the general case where the relay has multiple antennas and to more than two source terminals.

After obtaining $\mathcal{Q}_q$ and $\mathcal{Q}_t$, CDF can be naturally extended to AF and EF. After finding x_q from (33), compression based AF transmits $\beta(y_r - x_q + \mathcal{M}(x_q))$, where β is a scalar to satisfy the average power constraint. Likewise, assuming that x_q is received at the relay without noise, compression based EF chooses a relay function f to minimize $E\{|\mathcal{M}(x_q) - f(y_r)|^2\}$.

VI. SIMULATION RESULTS

Exemplarily we consider a three-way relay network where the relay has $M=2$ antennas. Fig. 1 compares the average symbol error probability of different protocols as a function of P. The power is allocated according to the considerations in Section V. The AF protocols by using the Alamouti type space time code and the diagonal space time code in Section V-A.1 are denoted as "AF Alamouti" and "AF Diagonal", respectively. The AF scheme obtained by using random unitary precoding matrix is denoted as "AF Unitary". The DF protocol by concatenating a linear code with an Alamouti code is optimized by maximizing the sum rate of all three terminals at $P=20$. After optimization, we find the coding matrix is

$$\begin{bmatrix} -0.1314x_1 + 0.7687x_2 + 0.6260x_3 & 0.7880x_1 - 0.1830x_2 + 0.5878x_3 \\ -0.7880x_1^* + 0.1830x_2^* - 0.5878x_3^* & -0.1314x_1^* + 0.7687x_2^* + 0.6260x_3^* \end{bmatrix},$$
(34)

which is denoted as "DF Sum Opt". The DF strategy with Toeplitz matrix-based network coding and an Alamouti code in Section V-A.2 is denoted as "DF Network Coding". Further, the DF scheme with random unitary precoding matrices is denoted as "DF Unitary". The DNF protocols are defined similarly as their AF counterparts. We do not show the results of EF because its performance is very close to DNF. From Fig. 1, we can see that 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 still are inferior to the DF protocols. For these protocols, we find that by using network coding there is a performance gain over those schemes which are not employing network coding. Furthermore, in all AF, DF and DNF, we find that using random unitary matrices incurs a performance loss even though it may be suitable for distributed scenarios as random linear network coding [11]. More results can be found in [9]. In particular, we find that if P is small, DF outperforms both CF and AF strategies in terms of the average sum rate, where CF is superior for larger values of P. Also, DF is seen to outperform all other strategies in terms of the average symbol error probability.
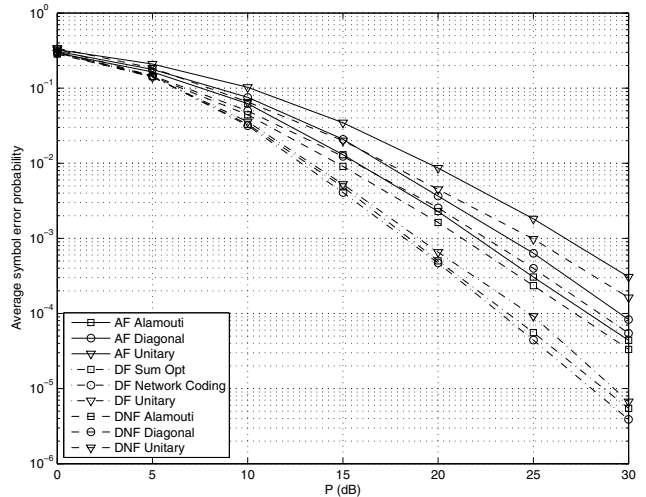

Fig. 1. Average symbol error probability comparison of different protocols as a function of P in a 3-way relay network. The relay has $M=2$ antennas.

VII. CONCLUSION

We have analyzed several space-time communication protocols for N-way relay networks in terms of the achievable rate region. In particular, for CF and DF strategies we have seen that applying network coding at the relay improves the rate region. Moreover, for AF and DF we also determined the optimal diversity-multiplexing tradeoff. The second part of the paper was devoted to practical transmission strategies. We demonstrated that for AF with M antennas at the relay and an overall power of P a diversity order slightly smaller than M can be achieved by employing a specific space-time code construction.

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] T. Cui, T. Ho, and J. Kliewer, "Some results on relay strategies for memoryless two-way relay channels," in *Proc. of Information Theory and Applications Workshop*, Jan. 2008.

[3] ——, "Relay strategies for memoryless two-way relay channels: Performance analysis and optimization," in *Proc. of IEEE ICC*, May 2008.

[4] T. Cui, F. Gao, T. Ho, and A. Nallanathan, "Distributed space-time coding for two-way wireless relay networks," in *Proc. of IEEE ICC*, May 2008.

[5] T. Cover and A. Gamal, "Capacity theorems for the relay channel," *IEEE Trans. Inform. Theory*, vol. 25, no. 5, pp. 572–584, Sept. 1979.

[6] 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.

[7] 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.

[8] 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.

[9] T. Cui and T. Ho, "Communication protocols for n-way relay networks," Caltech, Tech. Rep., Mar. 2008.

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

[11] 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.

[12] T. Cui and C. Tellambura, "Generalized feedback detection for spatial multiplexing multi-antenna systems," *IEEE Trans. Wireless Commun.*, vol. 7, no. 2, pp. 594–603, Feb. 2008.