

On the Combination of Cooperative Diversity and Network Coding for Wireless Uplink Transmissions

Zhiguo Ding, *Member, IEEE*, and Kin K. Leung, *Fellow, IEEE*

Abstract—Cooperative diversity has been recognized as an effective and low-cost technique to combat multipath fading and enhance transmission reliability. Motivated by the fact that many existing cooperative protocols suffer some loss of ergodic capacity, network coding, which is a technique that is well known for its capability to increase system throughput, is proposed in this paper to be combined with cooperative diversity for uplink transmissions. Two kinds of information-theoretic metrics, i.e., outage and ergodic capacities, are developed for the proposed transmission scheme to assist with performance evaluation. The developed analytic results, which are shown to fit well with Monte Carlo simulation, demonstrate that the proposed protocol can achieve better system robustness and larger system throughput simultaneously than comparable schemes. The application of the proposed transmission protocol to scheduling is also studied to achieve better fairness-throughput tradeoff.

Index Terms—Capacity, cooperative communication, multiple input–multiple output, scheduling, system throughput.

I. INTRODUCTION

AS A bandwidth-efficient and low-cost way to provide diversity, cooperative transmission has been shown to be effective in combating multipath fading and is widely used in many communication scenarios, such as wireless sensor networks and cellular systems [1]. Following the initial study in [2]–[4], there has been tremendous research interest in improving the spectral efficiency of cooperative diversity.

For example, a best-relay protocol has been proposed in [5] to avoid the use of relays with deep fading channels, and it is shown that full diversity gain can be achieved by only consuming one extra time slot. Furthermore, to minimize the bandwidth resource consumed by relay transmission, a cooperative protocol was proposed in [6], where the relay forwards its observations at the same time when the source is transmitting new messages. A similar idea was proposed in [7], where the superposition modulation is used to avoid the case that some time slots are entirely used for retransmission. However, many existing cooperative transmission schemes suffer loss of ergodic capacity, which is the price of obtaining a desirable diversity gain.

Manuscript received September 27, 2010; revised December 21, 2010; accepted January 20, 2011. Date of publication February 10, 2011; date of current version May 16, 2011. The work of Z. Ding was supported by the U.K. Engineering and Physical Sciences Research Council under Grant EP/F062079/2. The work of K. K. Leung was supported in part by the U.S. Army Research Laboratory and in part by the U.K. Ministry of Defence under Agreement W911NF-06-3-0001. The review of this paper was coordinated by Prof. B. Hamdaoui.

Z. Ding is with the School of Electrical, Electronic, and Computer Engineering, Newcastle University, NE1 7RU Tyne, U.K.

K. K. Leung is with the Department of Electrical and Electronic Engineering, Imperial College, SW7 2BT London, U.K.

Digital Object Identifier 10.1109/TVT.2011.2112787

Separately studied in the networking layer, network coding has been recognized as a promising technique to increase system throughput [8], [9]. The key feature of network coding is to encourage the relay to forward the mixture of its observations. Because of the broadcasting nature of radio propagation, network coding is ideal to be applied in wireless communications. Since it was devised at a networking layer, it has been assumed that physical layer conditions are perfect, which is not realistic due to multipath fading. The work in [10] is one of the first few to address network coding from the physical layer aspect. As an extension of [10], the combination of network coding and relay selection is studied in [11] to combat fading and explore the dynamic nature of the wireless environment.

Naturally, cooperative diversity and network coding are complementary to each other, and the aim of this paper is to study the combination of network coding and cooperative diversity for the communication scenario where multiple sources are communicating with a common destination. The contribution of this paper is threefold. First, a novel cooperative transmission protocol assisted by network coding is proposed. Multiple sources will simultaneously broadcast, and the signals forwarded by a relay will be a mixture of its observations. Because of the use of network coding, to deliver the same amount of information, the proposed cooperative protocol only needs the same number of time slots as the direct transmission scheme, and therefore, the significant loss of ergodic capacity caused by classical protocols, such as those in [3], can be avoided. To further improve system performance, a relay selection strategy is applied, where only the relay with good connection in both the base station and the sources will be selected. Furthermore, formal proof is provided to show that with two sources involved in the cooperation, the proposed transmission protocol can achieve the largest ergodic sum rate.

Second, the analytic results of two information-theoretic metrics, i.e., outage and ergodic capacities, are developed for the proposed cooperative transmission protocol. It is shown that the proposed protocol can achieve the same diversity order but a constant gain of outage probability compared with many existing cooperative schemes [5], [7]. Furthermore, it is demonstrated that the proposed protocol suffers no loss of ergodic system throughput, unlike other cooperative schemes. A constant performance gain of ergodic capacity can be observed, compared with direct transmission, which is due to the use of network coding and relay selection strategies. Monte Carlo simulation is provided to evaluate the accuracy of the developed analytic results.

Third, the application of the proposed cooperative transmission protocol to scheduling is also discussed in this paper.

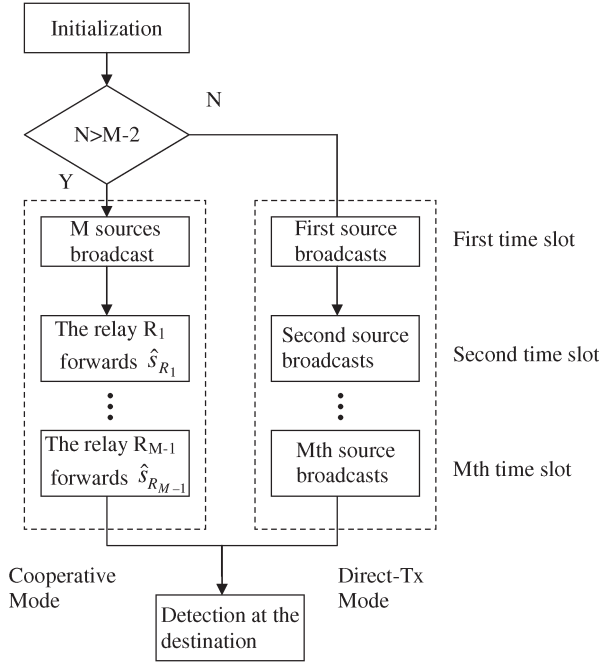


Fig. 1. Diagram for the description of the proposed network-coded cooperative transmission protocol. M denotes the number of sources, and N denotes the number of qualified relays.

Recall that the performance of a scheduler has been fundamentally constrained by the fairness–throughput tradeoff. Since the proposed protocol can improve the users’ channel condition without reducing throughput, it is natural to apply the proposed scheme to scheduling. As shown in this paper, a simple combination of the proposed protocol with a round-robin scheduler can achieve better balance of system throughput and fairness.

II. DESCRIPTION OF TRANSMISSION STRATEGY

Consider a centralized communication scenario where multiple sources are communicating with a common node. Such a communication scenario can be a wireless sensor network, where a data fusion center collects data from multiple sensors, or a traditional cellular mobile system, where multiple users are communicating with a base station. The half duplex mode is considered in this paper, where a node cannot transmit and receive at the same time. Time-division multiplexing access is adopted in this paper, where the developed results can straightforwardly be extended to other multiple-access techniques. Time synchronization is assumed among all nodes, including the source, destination, and relays. Note that for practical systems, such as Global System for Mobile Communications and Universal Mobile Telecommunications System, such an assumption about time synchronization can be realized. Consider that there are M active users who have the data sent to the base station and L inactive users that will act as relays and assist the active users for successful transmission. Fig. 1 shows the transmission strategy of the proposed cooperative protocol.

A. Initialization

It is assumed that the destination has access to the channel state information (CSI) for source–destination and relay–destination channels, and each relay has access to its local

CSI, e.g., its incoming channels from the M sources and its outgoing channel to the destination. During the initialization stage, such an assumption can be realized by using training-based approaches. In particular, each of the M sources will take turns to broadcast training information,¹ where the base station and the L relays will use such training information and accomplish channel estimation. One extra time slot is needed for the base station to broadcast training information, so each relay can estimate its outgoing channel. To obtain relay–destination CSI at the base station, each of the $M - 1$ used relays will transmit training information before their data transmissions.

Based on its local channel information, each relay will know whether it can successfully decode the M users’ information. The criterion for successful decoding will be discussed in detail later. Note that the number of qualified relays, which is denoted as N , has an impact on the decision of which mode can be supported, which implies the following assumption is required by the proposed protocol: Consider that the base station has access to the value of N , i.e., the number of qualified relays, which can be realized by asking each relay to feedback one digit to the base station and indicate whether it is qualified.

If $N \geq (M - 1)$, then the cooperative mode will be used, where the base station needs to broadcast the CSI of the channels between the M sources and the base station to all of the relays. Signaling through the dedicated command channel is assumed to be errorless. If $N < (M - 1)$, then the traditional time-division duplex direct transmission scheme will be used, where the M users take turns to transmit, and there is no cooperation among the users. It is important to point out that the probability of the events with $N \geq (M - 1)$ is much larger than that with $N < (M - 1)$ at high SNR and a fixed data rate, which will be formally proved in the following. In addition, the reason to have the criterion of $N \geq (M - 1)$ will be discussed after the signal model at the base station is introduced.

B. Cooperative Mode

In this paper, we only consider the cooperation between the relays and the sources. At the first time slot, all of the M users will broadcast their data simultaneously, where all of the L relays and the base station listen. Hence, the base station will receive

$$y_{D1} = \sum_{m=1}^M h_{mD}s_m + n_1 \tag{1}$$

where h_{mD} denotes the channel between the m th source and the destination, s_m is the m th user’s message, and n_1 denotes the white Gaussian noise. At the same time, each relay receives

$$y_{r_n} = \sum_{m=1}^M h_{mr_n}s_m + n_{r_n} \quad \forall n \in \{1, \dots, L\}. \tag{2}$$

As discussed before, it is assumed that CSI is available at the receiver. Therefore, each relay will decide whether it can

¹Even for noncooperative schemes, such a stage is still necessary since CSI is required for coherent detection at the base station. The bandwidth resource for training is proportional to the number of active users M .

successfully decode the M users' information based on its local channel information. Define ρ as the SNR, and the criterion used for successful decoding is that its local channel information can satisfy

$$\sum_{m \in \mathcal{S}} \mathcal{R}_m \leq \log \left(1 + \rho \sum_{m \in \mathcal{S}} |h_{mr_n}|^2 \right) \quad \forall \mathcal{S} \subseteq \{1, \dots, M\} \quad (3)$$

which is the capacity region for multiple-access channels [12], [13]. Here, \mathcal{R}_m is denoted as the targeted data rate for the m th user. For simplicity, a symmetrical system is considered here, where R will be used as the common data rate, i.e., $R = \mathcal{R}_m$, $m \in \{1, \dots, M\}$. As long as one relay's incoming channels h_{mr_n} can support all of the inequalities in (3), the relay shall be able to decode all of the source messages. In practice, successive decoding and interference cancellation based approaches can be applied to realize the capacity region of multiple access channels, where the source information with less fading attenuation will be decoded first and then subtracted from the mixture [12], [14]. An alternative to reduce the computational complexity at the relays is to use amplify forward strategies; however, it could cause the difficulty that the CSIs of all the channels, including the source-relay channels, have to be collected at the destination for effective detection [15]. Assume that N of all of the L relay candidates can satisfy the criterion and, hence, be capable of decoding the M sources' information correctly. By ordering the relay-destination channels, denote the N qualified relays as R_1, \dots, R_N , where $|h_{R_n D}|^2 \geq |h_{R_{n+1} D}|^2$. Note that out of the N qualified relays, we only need to use $(M-1)$ best ones.

During the next $M-1$ time slots, the $M-1$ qualified relays with the best relay-destination channel condition (R_1, \dots, R_{M-1}) will take turns to communicate with the base station. The scheduling of transmission can be accomplished in a distributed way by adjusting each relay's backoff time that is inversely proportional to the quality of its relay-destination channel $|h_{R_m D}|^2$ [5], [16]. As a result, a relay with a poorer channel condition has to wait longer, and eventually, the relay with the best channel condition can seize the control of the channel. With the application of network coding, each of the $M-1$ relays transmits a perturbed mixture of the M users' messages

$$\hat{s}_{R_m} = \sum_{m=1}^M \gamma_{R_m, m} s_m \quad (4)$$

where $\gamma_{R_m, m}$ is the weighting factor. The choice of the weighting factors $\gamma_{R_m, m}$ is devised to satisfy

$$\begin{cases} \Gamma_m \perp \Gamma_n, & \forall m \neq n, \quad m, n \in \{0, \dots, M-1\} \\ \Gamma_m^H \Gamma_m = 1, & \forall m \in \{0, \dots, M-1\} \end{cases} \quad (5)$$

where $\Gamma_m = [\gamma_{R_m, 1} \ \dots \ \gamma_{R_m, M}]^T$ for $m \in \{1, \dots, M-1\}$, $\Gamma_0 = (1/p)[h_{1D} \ \dots \ h_{MD}]^T$, and $p = \sum_{m=1}^M |h_{mD}|^2$. Note that the signal combination strategy proposed in (4) can ensure that the transmission power at each relay will be constrained as

$$\mathcal{E} \{ \hat{s}_{R_m} \hat{s}_{R_m}^* \} = \mathcal{E} \{ \Gamma_m^T \mathbf{s} \mathbf{s}^H \Gamma_m^* \} = \Gamma_m^T \Gamma_m^* = 1 \quad (6)$$

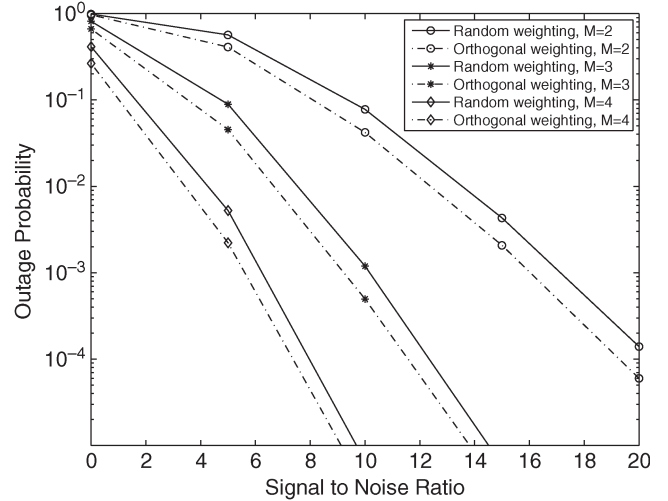


Fig. 2. Use of random weighting factors versus use of orthogonal weighting factors. The data rate is set as $R = 4$ bits/Hz/s.

where $\mathbf{s} = [s_1 \ \dots \ s_M]^T$; the second equality follows from the assumption that the channels are quasi-static, and the last equality follows from (4). In addition, to ensure relay transmission power constrained, the use of the proposed combination strategy can also ensure that the coefficient vectors used by different relays Γ_m are orthogonal to each other. Such an orthogonal property is very useful to simplify the expression of the mutual information, as shown in the following. It is assumed that the weighting factor Γ_m can have an errorless feedback from the base station at the beginning of each relay transmission. Evidently, the feedback of the weighting vector Γ_m will cause an extra system overhead. Recall that the main reason to use such orthogonal factors is to simplify development and obtain explicit expressions of analytic results. In practice, the weighting factor Γ_m can be randomly generated at each relay, so Assumption 4 can be removed, and the feedback of Γ_m is no longer necessary. In Fig. 2, the outage performance of the two schemes using random and orthogonal weighting factors is shown as a function of SNR. As can be seen from the figure, the use of random weighting factors only causes a slight loss of reception reliability. More importantly, the scheme with random weighting factors can still achieve the same diversity gain as that with orthogonal factors since the curves for the two schemes have the same slope.

Hence, during the next $M-1$ time slots, the destination will receive²

$$y_{Dm} = h_{R_m D} \hat{s}_{R_m} + n_m. \quad (7)$$

²At each of these $M-1$ time slots, the M sources can continue transmitting new messages \mathbf{s}_m as in [4] and [6], which can yield a larger sum rate as $\mathcal{I} = ((\log\{(1 + \rho \sum_{m=1}^M |h_{mD}|^2) \prod_{m=1}^{M-1} (1 + \rho \sum_{m=1}^M |h_{mD}|^2 + \rho |h_{R_m D}|^2)\})/M)$. However, such a variance of the proposed scheme requires more sophisticated coding techniques to retrieve the new messages \mathbf{s}_m . Since the key point of this paper is to advocate the combination of cooperative diversity and network coding, we only focus on such orthogonal relay transmission in this paper. All the developed results, including the outage/ergodic capacities and the optimality of M , can be straightforwardly extended to the nonorthogonal scheme.

Stacking over the M time slots, the signal model for the proposed transmission scheme can be shown as

$$\mathbf{y} = \mathbf{D}\mathbf{H}\mathbf{s} + \mathbf{n} \quad (8)$$

where $\mathbf{y} = [y_{D1} \ \cdots \ y_{DM}]^T$, $\mathbf{D} = \text{diag}\{p, h_{R_1D}, \dots, h_{R_{M-1}D}\}$, $\mathbf{H} = [\Gamma_0 \ \cdots \ \Gamma_M]^T$, and $\mathbf{n} = [n_1 \ \cdots \ n_M]^T$. Note that the signal model in the foregoing equation can be viewed as a set of linear equations. There are M observations, which can be used to construct M linear equations. There are M symbols in \mathbf{s} , which can be viewed as unknown variables. Therefore, we have M linear equations to solve M unknown variables, and such linear equations are solvable. This is the reason we need at least $(M - 1)$ relays.

By using the fact that the weighting factors are well defined, i.e., $\mathbf{H}\mathbf{H}^H = \mathbf{I}_M$, the sum rate achieved by the cooperative mode ($N \geq (M - 1)$) can be shown as

$$\mathcal{I}_{M,N}^{co} = \frac{1}{M} \log \left\{ \left(1 + \rho \sum_{m=1}^M |h_{mD}|^2 \right) \prod_{m=1}^{M-1} (1 + \rho |h_{R_mD}|^2) \right\}$$

where the factor $1/M$ is due to the fact that communication happens in M successive time slots.

C. Optimality of the Number of Users for Cooperation

Combining with the case of the non-cooperative mode, the sum rate achieved by the proposed transmission technique can be written as

$$\mathcal{I}_{M,N} = \begin{cases} \mathcal{I}_{M,N}^{co}, & \text{if } N \geq (M - 1) \\ \frac{\log \left\{ \prod_{m=1}^M (1 + \rho |h_{mD}|^2) \right\}}{M}, & \text{if } N < (M - 1). \end{cases} \quad (9)$$

A natural question for the proposed transmission scheme is that how many users can achieve the best performance. Consider a scheme with a fixed number of relays. When the number of users to cooperate is increasing, there will be less qualified relays that can decode all the M users' information. As a result, there will be less freedom to explore the dynamic wireless environments. Hence, the intuition is that the more users participate in cooperation, the worse the system performance will be. The following theorem about the optimality of the number of involved users confirms our intuition.

Theorem 1: For the proposed network-coded cooperative scheme, the largest ergodic sum rate can be achieved where there are only two users participating in cooperation.

Proof: See the Appendix. ■

The benefit of the small number of users for cooperation is to reduce the system complexity. For example, with a small M , the computational complexity to decode the sources' information at each relay will be reduced. The system overhead can also be reduced for a smaller number of M since the destination needs to broadcast less CSI to all of the relays. Therefore, in the following sections, we will only focus on the scenario with two source nodes.

Comparing the expressions of the mutual information achieved by the two modes, it is obvious that the cooperative expression can achieve better performance. The effect of the

proposed cooperative technique is to create M virtual channels, where $|\hat{h}_{1D}|^2 = \sum_{m=1}^M |h_{mD}|^2$ and $|\hat{h}_{mD}|^2 = |h_{R_{m-1}D}|^2$ for $m \in \{2, \dots, M\}$. For a sufficient number of relaying candidates, it can be proved that the quality of the virtual channels is better than the direct transmission channels $|\hat{h}_{mD}|^2 \geq |h_{mD}|^2$, and hence, the mutual information achieved by the proposed cooperative scheme will be larger than the noncooperative scheme. In the following sections, we will provide detailed performance evaluation for the proposed protocol.

III. INFORMATION-THEORETIC METRICS

In this section, two kinds of information-theoretic metrics will be studied, i.e., the outage and ergodic capacities, respectively. The outage capacity measures the robustness performance a system can achieve, and the ergodic capacity focuses on the averaged throughput a system can support. Together, these two information-theoretic metrics will provide a comprehensive performance evaluation for the proposed network-coded cooperative protocol. Note that due to Theorem 1, we will only focus on the scenario with two source nodes.

A. Outage Capacity

First, the definition of the outage capacity is given here.

Definition 1: $\alpha\%$ outage capacity is the maximum data rate R that can be supported with the targeted outage probability $\alpha\%$, i.e.,

$$\max_R P(\mathcal{I} < R) \leq \alpha\%.$$

From its definition, it is necessary to find to the outage probability to obtain the outage capacity.

Rewrite the sum rate supportable for the proposed protocol as

$$\mathcal{I}_{2,N} = \begin{cases} \mathcal{I}_{2,N}^{co}, & \text{if } N \geq 1 \\ \frac{1}{2} \log \left\{ \prod_{m=1}^2 (1 + \rho |h_{mD}|^2) \right\}, & \text{if } N < 1 \end{cases} \quad (10)$$

where $|h_{RD}|^2$ is the relay-destination channel with the largest value among the independent identically distributed (i.i.d.) variables $|h_{R_mD}|^2$ for $m \in \{1, \dots, N\}$.

The outage event can be defined as

$$\mathcal{O} \triangleq \bigcup_{\mathcal{A}} \mathcal{O}_{\mathcal{A}} \quad (11)$$

where the union is taken over all possible subsets $\mathcal{A} \subseteq \{1, 2\}$, and $\mathcal{O}_{\mathcal{A}}$ can be defined as

$$\mathcal{O}_{\mathcal{A}} \triangleq \bigcup_N \mathcal{O}_{N,\mathcal{A}} \quad \forall N \in \{0, \dots, L\} \quad (12)$$

$$\mathcal{O}_{N,\mathcal{A}} \triangleq \left\{ \mathcal{I}(\mathbf{s}_{\mathcal{A}}; \mathbf{y} | \mathbf{s}_{\mathcal{A}^c}, \mathbf{H} = \mathbf{H}, \mathbf{D} = \mathbf{D}, \mathbf{N} = \mathbf{N}) \leq \sum_{i \in \mathcal{A}} R_i \right\} \quad (13)$$

which follows the definition of outage events in [7] and [13]. Furthermore, define $|\mathcal{A}|$ as the number of users in \mathcal{A} . Note that

the symmetric system is of interest in this paper, which means $|\mathcal{A}|R = \sum_{i \in \mathcal{A}} R_i$. Hence, for the case of $N \geq 1$, the mutual information can be written as

$$\begin{aligned} \mathcal{I}_{N, \mathcal{A}_1} &= \log(1 + \rho(|h_{1D}|^2 + \gamma_{21}^2|h_{RD}|^2)) \\ \mathcal{I}_{N, \mathcal{A}_2} &= \log(1 + \rho(|h_{2D}|^2 + \gamma_{22}^2|h_{RD}|^2)) \\ \mathcal{I}_{N, \mathcal{A}_3} &= \frac{1}{2} \log \left\{ \left(1 + \rho \sum_{m=1}^2 |h_{mD}|^2 \right) (1 + \rho|h_{RD}|^2) \right\} \end{aligned}$$

where $\mathcal{I}_{N, \mathcal{A}_n} = \mathcal{I}(\mathbf{s}_{\mathcal{A}_n}; \mathbf{y} | \mathbf{s}_{\mathcal{A}_n^c}, \mathbf{H} = H, \mathbf{D} = D, \mathbf{N} = N)$, $\mathcal{A}_1 = \{1\}$, $\mathcal{A}_2 = \{2\}$, and $\mathcal{A}_3 = \{1, 2\}$, as shown in [13].

The exact expression of the outage probability can be difficult to find, and its upper and lower bounds will be developed in the following:

$$P(\mathcal{O}_{\mathcal{A}_1}) \leq P(\mathcal{O}) = P\left(\bigcup_{\mathcal{A}} \mathcal{O}_{\mathcal{A}}\right) \leq \sum_{\mathcal{A}} P(\mathcal{O}_{\mathcal{A}}) \quad (14)$$

where $P(\mathcal{O}_{\mathcal{A}}) = \sum_N P(\mathcal{O}_{N, \mathcal{A}})P(\mathbf{N} = N)$.

The calculation of the outage probability requires the knowledge of the probability for the event that N of all the L relaying candidates can decode the two users' messages successfully. The following lemma provides an approximation for the criterion that one relay can decode the sources' messages successfully.

Lemma 1: Consider that all the channels are independent and identically Rayleigh faded. Define x and y as the norm of two channels from the two sources to a relay. For high SNR and a fixed data rate, the condition for the relay to successfully decode the sources' messages can be relaxed as follows:

$$\begin{aligned} P\left(x \geq \frac{2^R - 1}{\rho}, y \geq \frac{2^R - 1}{\rho}, x + y \geq \frac{2^{2R} - 1}{\rho}\right) \\ \approx P\left(x \geq \frac{2^R - 1}{\rho}, y \geq \frac{2^R - 1}{\rho}\right). \end{aligned} \quad (15)$$

Proof: See the Appendix. ■

With this approximation, the probability for the number of qualified relays can be obtained from the following theorem.

Theorem 2: Assume that all of the channels are i.i.d. Raleigh fading. Given L relaying candidates, the probability for the number of relays that can decode the two sources' message successfully can be approximated at high SNR and a fixed data rate as

$$P(\mathbf{N} = N) \approx \left(\frac{2^R - 1}{\rho}\right)^{L-N} \sum_{l=N}^L \frac{L!}{N!(L-N)!(L-l)!}. \quad (16)$$

Proof: See the Appendix. ■

One remark for Theorem 2 is that, conditioned on a fixed data rate and high SNR, the event that all of the L relays can decode the M users' information will eventually dominate the other events. Specifically, it can easily be evaluated that at high SNR, we can have $P(\mathbf{N} = L) \rightarrow 1$ and $P(\mathbf{N} = N) \rightarrow 0 \forall 0 \leq N < L$. This also confirms our previous statement that the cooperative mode will be the dominant mode and that the noncooperative mode only has limited effects on the outage

probability. With the help of Theorem 2, the outage probability for the proposed cooperative scheme can be shown as the following theorem.

Theorem 3: Assume that all of the channels are i.i.d. Raleigh fading. With high SNR and a fixed data rate R , the outage probability for the proposed cooperative transmission strategy can be bounded as

$$C_L \left(\frac{2^R - 1}{\rho}\right)^{L+1} \leq P(\mathcal{O}) \leq 4C_L \left(\frac{2^R - 1}{\rho}\right)^{L+1} \quad (17)$$

where $C_L = \sum_{N=0}^L \sum_{l=N}^L (L!/(N+1)!(L-N)!(L-l)!)$

Proof: See the Appendix. ■

As can be seen from Theorem 3, the diversity order that the proposed network coded cooperative scheme can achieve is $L + 1$. A comparison between the proposed scheme and an existing cooperative protocol is provided as follows.

1) *Numerical Results:* Consider an indoor rich-scattering environment that is typically assumed by existing cooperative schemes, as in [3] and [7]. Hence, all of the addressed channels can be treated as identically independent Raleigh distributed. The targeted data rate is $R = 2$ bits/Hz/s, and the number of relays is $L = 2$. The performance of the proposed transmission protocol will be compared with two existing transmission schemes, i.e., the direct transmission scheme and the best relay protocol. The best relay scheme was originally developed in [5] for single-user scenarios, where the source first broadcasts its messages, and at the second time slot, the best relay is selected for cooperation by setting the backoff period of each relay inversely proportional to its channel condition. Note that the source cannot transmit during relay transmissions.

It can be observed from Fig. 3 that the two curves of the proposed protocol and the best relay scheme have the same slope, which confirms our statement that both schemes can achieve the same diversity gain. However, the proposed cooperative scheme can achieve a constant gain of outage probability over the best relay protocol. This is due to the fact that the relay transmission of the proposed protocol is more spectrally efficient than the best relay due to the use of network coding. Furthermore, the two developed upper and lower bounds for the outage probability have also been provided in the figure and shown to fit well with the Monte Carlo simulation results.

B. Ergodic Capacity

Definition 2: The ergodic capacity is the long-term data rate that a system can support, i.e.,

$$\mathcal{C}_e = \int_0^{\infty} \mathcal{I} f_{\mathcal{I}}(\mathcal{I}) d\mathcal{I}$$

where $f_{\mathcal{I}}(\cdot)$ is the probability density function (pdf) of the mutual information \mathcal{I} .

Recall that $P_M(\mathbf{N} = N)$ is a function of the source-relay channel coefficients and that $\mathcal{I}_{M, N}$ is a function of the source-destination and relay-destination channel coefficients. Since all of the channels are assumed to be i.i.d. Raleigh fading, the two variables $\mathcal{I}_{M, N}$ and $P_M(\mathbf{N} = N)$ are independent of

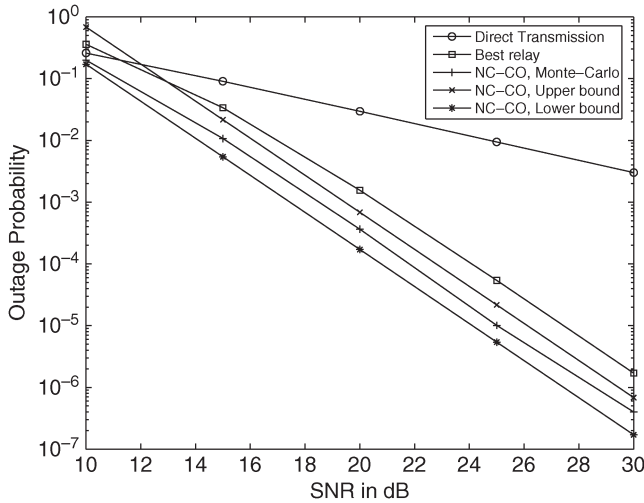


Fig. 3. Outage probability versus SNR. The data rate is set as $R = 2$ bits/Hz/s. Both Monte Carlo simulation and the developed analytic results are shown.

each other. From its definition, the ergodic capacity achieved by the proposed scheme can be calculated as

$$C_e = \sum_{N=0}^L \mathcal{E}\{\mathcal{I}_{M,N}\}P(N = N). \quad (18)$$

Despite its simple form, the expression of the ergodic capacity is very difficult to get since the probability $P(N = N)$ is also a function of the data rate. To obtain explicit analytic results for the ergodic capacity, it is assumed here that there will be a fixed number of relays $N \geq 1$ that can successfully decode the M users' information. The practical meaning for this assumption is that around the active users, there are typically some inactive users that have the line-of-sight connection with the active users. For example, in a large conference room, it is realistic to assume that some inactive users have lossless connection with the active user.

To obtain the expression of the ergodic capacity, the following lemma is needed.

Lemma 2: The following equality holds:

$$\sum_{i=0}^N \binom{N}{i} \frac{(-1)^i}{i+1} = \frac{1}{N+1} \quad (19)$$

for integer N .

Proof: The lemma can be proved by first applying the equation in [17, eq. (0.155.1)] and then some algebraic manipulations. ■

With the help of Lemma 2, the ergodic system throughput can be obtained from the following theorem.

Theorem 4:

$$\mathcal{E}\{\mathcal{I}_N\} \approx (\log \rho - \mathbf{C} \log e) + \frac{1}{2}(\log e + C_N) \quad (20)$$

where \mathbf{C} denotes Euler's constant, and $C_N = N \log e \sum_{i=0}^{N-1} \binom{N-1}{i} (-1)^{i+1} (1/(i+1)) \ln(i+1)$.

Proof: See the Appendix. ■

TABLE I
EVALUATION OF THE VALUE OF THE CONSTANT C_N

N	1	2	3	4	5	6	7	8	9	10
C_N	0	1	1.42	1.66	1.83	1.95	2.05	2.13	2.20	2.26

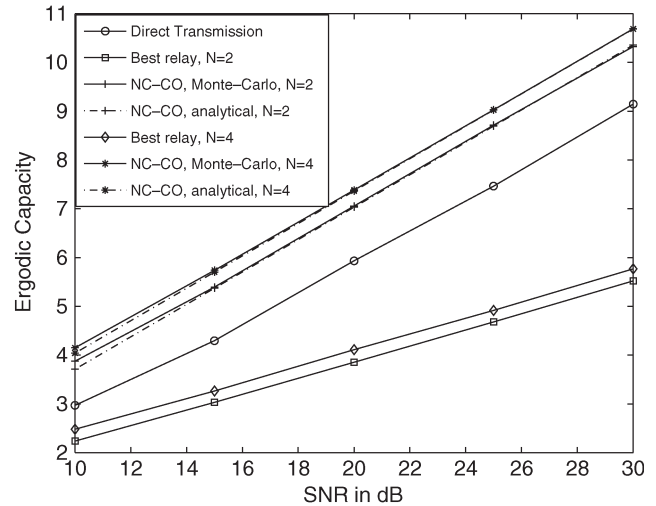


Fig. 4. Ergodic capacity versus SNR. Solid lines show the results obtained by using Monte Carlo simulation, and dashed lines show the developed analytic results.

The value of the constant C_N is only a function of N , and Table I shows its value for different N . As can be seen from the table, the constant C_N is always positive and a monotonically increasing function of N . Recall that the ergodic capacity achieved by direct transmission can be written as

$$\mathcal{E}\{\mathcal{I}_D\} \approx (\log \rho - \mathbf{C} \log e).$$

Hence, the difference of system throughput achieved by the two schemes can be shown as

$$\Delta = \frac{1}{2}(\log e + C_N). \quad (21)$$

Since $C_N \geq 0$ is a monotonically increasing function of N , Δ is always positive, and its value is a proportional function of N . Hence, the proposed cooperative protocol can achieve larger ergodic system throughput than direct transmission, whereas most cooperative diversity schemes suffer some loss for system throughput due to inefficient relay transmission.

1) Numerical Results: The simulation condition is the same as that for the outage probability. The performance of the proposed transmission protocol will be again compared with two existing transmission schemes, i.e., the direct transmission scheme and the best relay protocol. As can be seen from Fig. 4, the best relay protocol achieves the smallest ergodic sum rate among the three transmission schemes, which is due to the extra bandwidth consumption of relay transmission. At high SNR, it is very possible that the source–destination channel is already good enough for reliable communications, where the allocation of only one extra time slot for retransmission can only reduce the system throughput.

Due to the use of network coding and relay selection strategy, relay transmission is more efficiently devised for the proposed

protocol, and hence, a larger ergodic capacity can be achieved compared with the best relay. The performance gain of the proposed scheme over direct transmission is due to the fact that the dynamic nature of radio propagation has been fully utilized and the inactive users/relays have been explored as a new dimension of networking resource. Furthermore, the curves obtained from our developed analytic expressions are very close to the Monte Carlo simulation results, which demonstrate the accuracy of the developed analytic results.

IV. APPLICATION OF THE PROPOSED TRANSMISSION PROTOCOL TO SCHEDULING

Scheduling is fundamental for wireless communications, and its ultimate goal is to optimally accomplish the task that multiple users share resources in a desired way. A fundamental limit of scheduling is the throughput–fairness tradeoff, where the two extreme schedulers, i.e., the max-SNR and round-robin schedulers, are good examples to illustrate the dilemma of scheduling. Consider the wireless communication scenario where M active users are communicating with the same destination. The max-SNR scheduler will always select the user that has the best connection with the base station. Hence, the max-SNR can achieve the largest system throughput, but have poor performance for fairness since a user that has worse connection with the base station will never be served. On the other hand, the round-robin scheduler will ask the multiple user to take turns for transmission. Hence, it can achieve the best fairness but have poor system throughput.

Since serving a user with poor channel condition is unavoidable to decrease the system throughput, a previous study for scheduling has been focused on how to reduce the resource allocated for such a user and maintain an acceptable fairness. An alternative considered in this paper is to improve the channel condition of such a user and increase its supportable data rate. It is important that such an improvement in channel condition is achieved without causing any loss of ergodic throughput. For this reason, multiple input–multiple output is a good candidate to improve the fairness–throughput tradeoff, but it requires the extra cost of hardware updating for all the users. Many existing cooperative protocols, such as the classical protocols in [3], [5], cannot be used to improve the efficiency of scheduling since the robustness gain, or diversity gain, is obtained by consuming extra bandwidth resources. Naturally, the proposed transmission protocol is a desirable candidate for application to scheduling since it can improve system robustness and throughput at the same time. The key observation is that inactive users, or relays, can be seen as another dimension of bandwidth resource because of the dynamic nature of the radio environment. This observation is important in practice since the number of inactive users is typically far larger than the active users.

A simple application of the proposed protocol to scheduling can first be to randomly group the active users in pair and then apply round-robin scheduling among the groups. This way, the fairness achieved by this simple scheduler is the same as round-robin, and the channel quality of each user can also be improved. Fig. 5 demonstrates the impact of the proposed transmission protocol to scheduling. Assume that the total

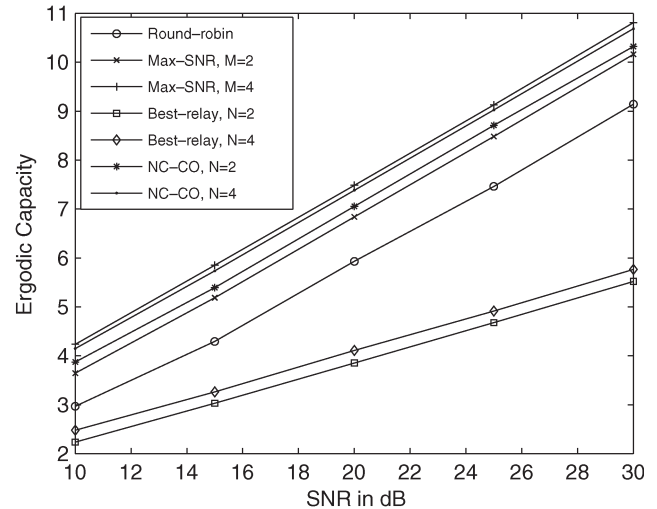


Fig. 5. Ergodic system throughput versus SNR.

number of active users is $M = 2$ or $M = 4$, and the other simulation condition is the same as that in the last section. As can be seen from the figure, the performance of such a simple application can be comparable to the max-SNR scheduler. When there are two active users, the proposed scheduler with $N = 2$ qualified relays can outperform the max-SNR scheduler. When the number of active users is increasing, the proposed scheduler can still achieve similar system throughput to the max-SNR scheduler, as long as there are sufficient qualified relays. By combining the proposed transmission technique with other more advanced schedulers, it is promising that the system throughput can further be increased while maintaining the fairness performance.

V. CONCLUSION

In this paper, we have studied the combination of cooperative diversity and network coding for broadband uplink transmissions. Two kinds of information-theoretic metrics, i.e., outage probability and ergodic capacity, were developed for the proposed transmission scheme for performance evaluation. The application of the proposed transmission protocol to scheduling was also proposed to achieve better fairness–throughput tradeoff. Due to space limitations, the study of practical scheduling protocol design is left as a future direction, where some initial results have been reported in [18].

APPENDIX

Proof of Theorem 1: Recall from (10) that the sum rate supportable for the proposed cooperative scheme is $\mathcal{I}_{M,N}$ conditioned on the fact that there are N relays satisfying (3). Denote the probability of the event that there are N qualified relays as $P_M(\mathbf{N} = N)$, where M is used to point out the fact that the probability is also a function of M . Recall that $P_M(\mathbf{N} = N)$ is only a function of the source–relay channel coefficients, and $\mathcal{I}_{M,N}$ is a function of the source–destination and relay–destination channel coefficients. Since all of the channels are assumed to be i.i.d. Raleigh fading, the two variables $\mathcal{I}_{M,N}$

and $P_M(\mathbf{N} = N)$ are independent of each other. Hence, the ergodic sum rate can be written as

$$\mathcal{C}_M = \sum_{N=0}^L \mathcal{E}\{\mathcal{I}_{M,N}\} P_M(\mathbf{N} = N). \quad (22)$$

To prove the theorem, define the difference between the ergodic sum rate achieved by the two schemes with M and $M - 1$ users to cooperate as

$$\Delta = \mathcal{C}_{M-1} - \mathcal{C}_M.$$

The aim of this proof is to show the difference larger than zero, which means that \mathcal{C}_M is a mono decreasing function of M . For specific \mathbf{N} , the difference can be broken into three parts, i.e.,

$$\Delta = \Delta_1 + \Delta_2 + \Delta_3 \quad (23)$$

where $\Delta_1 = \sum_{N=0}^{M-3} \mathcal{E}\{\mathcal{I}_{M-1,N}\} P_{M-1}(\mathbf{N} = N) - \sum_{N=0}^{M-3} \mathcal{E}\{\mathcal{I}_{M,N}\} P_M(\mathbf{N} = N)$, $\Delta_2 = \mathcal{E}\{\mathcal{I}_{M-1,M-2}\} P_{M-1}(\mathbf{N} = M-2) - \mathcal{E}\{\mathcal{I}_{M,M-2}\} P_M(\mathbf{N} = M-2)$, and $\Delta_3 = \sum_{N=M-1}^L \mathcal{E}\{\mathcal{I}_{M-1,N}\} P_{M-1}(\mathbf{N} = N) - \sum_{N=M-1}^L \mathcal{E}\{\mathcal{I}_{M,N}\} P_M(\mathbf{N} = N)$. From its definition in (23), Δ_1 denotes the difference of the sum rate between two schemes when they all degrade to the direct transmission mode, which implies

$$\Delta_1 = \mathcal{C}_D \left(\sum_{N=0}^{M-3} P_{M-1}(\mathbf{N} = N) - \sum_{N=0}^{M-3} P_M(\mathbf{N} = N) \right) \quad (24)$$

where \mathcal{C}_D denotes the ergodic capacity of direct transmission.

The variable Δ_2 denotes the sum rate difference of the two schemes for the event when the cooperative mode can be operated for the scheme with M sources and when the direct transmission mode is still operated for the scheme with $M - 1$ sources. Hence, Δ_2 can be written as

$$\begin{aligned} \Delta_2 = & \mathcal{E} \left\{ \frac{1}{M-1} \log \left\{ \left(1 + \rho \sum_{m=1}^{M-1} |h_{mD}|^2 \right) \right. \right. \\ & \left. \left. \times \prod_{m=1}^{M-2} (1 + \rho |h_{R_m D}|^2) \right\} \right\} \\ & \times P_{M-1}(\mathbf{N} = M-2) - \mathcal{C}_D P_M(\mathbf{N} = M-2) \\ & \geq \mathcal{C}_D [P_{M-1}(\mathbf{N} = M-2) - P_M(\mathbf{N} = M-2)] \quad (25) \end{aligned}$$

where the inequality follows the assumption that all the channels are i.i.d. Raleigh fading.

The calculation of Δ_3 is more complex, and it is useful to first prove $\Delta'_3 = \mathcal{E}\{\mathcal{I}_{M-1,N} - \mathcal{I}_{M,N}\} \geq 0$. The calculation of this difference can further be broken into two parts, i.e., $\Delta'_3 = \Delta_{31} + \Delta_{32}$, where $\Delta_{31} = \mathcal{E}\{\alpha_{M-1} - \alpha_M\}$, $\Delta_{32} = \mathcal{E}\{\beta_{M-1} - \beta_M\}$, $\alpha_M = (1/M) \log(1 + \rho \sum_{m=1}^M x_{m,M})$, $\beta_M = (1/M) \log\{\prod_{m=1}^{M-1} (1 + \rho y_{m,M})\}$, $x_{m,M} = |h_{mD}|^2$, and $y_{m,M} = |h_{R_m D}|^2$. Without losing generality, it is assumed that two schemes select different sources to participate in cooperation, and hence, $x_{m,M}$ and $x_{m,M-1}$ are i.i.d. variables.

By applying Jensen's inequality, we can have

$$\begin{aligned} \Delta_{31} & \geq \log \left[\mathcal{E} \left\{ \frac{\left(1 + \rho \sum_{m=1}^{M-1} x_{m,M} \right)^M}{\left(1 + \rho \sum_{m=1}^M x_{m,M} \right)^{M-1}} \right\} \right] \\ & = \log \left[\mathcal{E} \left\{ \frac{(z_{M-1})^M}{(z_M)^{M-1}} \right\} \right]. \quad (26) \end{aligned}$$

Recalling that the Rayleigh fading channels from M source nodes to the relay node are i.i.d., $x_{m,M}$ are independent exponential variables. Hence, $\sum_{m=1}^{M-1} x_{m,M}$ is Chi-square distributed with $2(M-1)$ degrees of freedom, which implies

$$f_{z_M}(z) = \frac{1}{\rho(M-1)!} \left(\frac{z-1}{\rho} \right)^{M-1} e^{-\frac{z-1}{\rho}}. \quad (27)$$

Since it is assumed that $x_{m,M}$ and $x_{m,M-1}$ are i.i.d. variables, we can have

$$\begin{aligned} \mathcal{E} \left\{ \frac{(z_{M-1})^M}{(z_M)^{M-1}} \right\} & = \mathcal{E} \{ (z_{M-1})^M \} \mathcal{E} \left\{ \frac{1}{(z_M)^{M-1}} \right\} \\ & \geq \mathcal{E} \{ (z_{M-1})^M \} \frac{1}{\mathcal{E} \{ (z_M)^{M-1} \}} \quad (28) \end{aligned}$$

where Jensen's inequality is applied since $\psi(x) = 1/x^M$ is a convex function. The variable $\mathcal{E}\{(z_{M-1})^M\}$ can be calculated as

$$\begin{aligned} \mathcal{E} \{ (z_{M-1})^M \} & = \epsilon_1 \int_1^{\infty} z^M (z-1)^{M-2} e^{-\frac{z-1}{\rho}} dz \\ & \geq \epsilon_1 \int_1^{\infty} (z-1)^{2M-2} e^{-\frac{z-1}{\rho}} dz = \frac{(2M-2)!}{(M-2)!} \rho^M \quad (29) \end{aligned}$$

where $\epsilon_1 = (1/\rho^{M-1}(M-2)!)$. Similarly, we can have

$$\begin{aligned} \mathcal{E} \{ (z_M)^{M-1} \} & = \epsilon_2 \int_1^{\infty} z^{M-1} (z-1)^{M-1} e^{-\frac{z-1}{\rho}} dz \\ & \leq e^{-\frac{1}{\rho}} \sum_{k=0}^{2M-2} \frac{(2M-2)!}{k!(M-1)!} \rho^{M-k-1} \end{aligned}$$

where $\epsilon_2 = (1/\rho^M(M-1)!)$.

Now the difference Δ_{31} can be written as

$$\begin{aligned} \Delta_{31} & \geq \log \left[\frac{\frac{(2M-2)!}{(M-2)!} \rho^M}{e^{-\frac{1}{\rho}} \sum_{k=0}^{2M-2} \frac{(2M-2)!}{k!(M-1)!} \rho^{M-k-1}} \right] \\ & \approx (M-1)\rho \geq 0. \quad (30) \end{aligned}$$

On the other hand, we can have

$$\begin{aligned} \Delta_{32} & = \frac{1}{M(M-1)} \mathcal{E} \left\{ (M-1) \sum_{m=1}^{M-2} \log \frac{(1 + \rho y_{m,M-1})}{(1 + \rho y_{m,M})} \right. \\ & \quad \left. + M \sum_{m=1}^{M-2} \log \frac{(1 + \rho y_{m,M-1})}{(1 + \rho y_{M-1,M})} \right\} \end{aligned}$$

which can further be approximated as

$$\Delta_{32} \geq \frac{1}{M} \sum_{m=1}^{M-2} \log \frac{1 + \rho \mathcal{E}\{y_{m,M-1}\}}{1 + \rho \mathcal{E}\{y_{m,M}\}} + \frac{1}{M(M-1)} \sum_{m=1}^{M-2} \log \frac{1 + \rho \mathcal{E}\{y_{m,M-1}\}}{1 + \rho \mathcal{E}\{y_{M-1,M}\}}. \quad (31)$$

With the same number of N qualified relays, recall that the N relays have been ordered as $y_{1,M-1} \geq \dots \geq y_{N,M-1}$ and $y_{1,M} \geq \dots \geq y_{N,M}$. Conditioned on the fact that these relay-destination channels are i.i.d. Rayleigh fading before the ordering, it can easily be shown that $(1 + \rho \mathcal{E}\{y_{m,M-1}\}) / (1 + \rho \mathcal{E}\{y_{m,M}\}) = 1$ and $(1 + \rho \mathcal{E}\{y_{m,M-1}\}) / (1 + \rho \mathcal{E}\{y_{M-1,M}\}) \geq 1$. Hence, $\Delta_{32} \geq 0$. Together with (30), we can conclude that $\mathcal{E}\{\mathcal{I}_{M-1,N} - \mathcal{I}_{M,N}\} \geq 0$. Hence, Δ_3 can be simplified as

$$\Delta_3 \geq \sum_{N=M-1}^L \mathcal{E}\{\mathcal{I}_{M,N}\} [P_{M-1}(\mathbf{N}=N) - P_M(\mathbf{N}=N)] \geq C_D \left(\sum_{N=M-1}^L [P_{M-1}(\mathbf{N}=N) - P_M(\mathbf{N}=N)] \right) \quad (32)$$

where the last inequality follows the fact that the sum rate supported by the cooperative mode is larger than the noncooperative mode. Now, combining (24), (25), and (32), it can be obtained that

$$\Delta \geq C_D \left(\sum_{N=0}^L [P_{M-1}(\mathbf{N}=N) - P_M(\mathbf{N}=N)] \right) \geq 0$$

where the last inequality follows the fact that it is a stronger condition to have N relays to correctly decode M user's information compared with that for the case of $M-1$ users. The theorem is proved. \blacksquare

Proof for Lemma 1: The lemma can be proved since the difference between the two probabilities can be shown as

$$\begin{aligned} \Delta &= P \left(x \geq \frac{2^R - 1}{\rho}, y \geq \frac{2^R - 1}{\rho} \right) \\ &\quad - P \left(x \geq \frac{2^R - 1}{\rho}, y \geq \frac{2^R - 1}{\rho}, x + y \geq \frac{2^{2R} - 1}{\rho} \right) \\ &= P \left(x \geq \frac{2^R - 1}{\rho}, y \geq \frac{2^R - 1}{\rho}, x + y \leq \frac{2^{2R} - 1}{\rho} \right). \quad (33) \end{aligned}$$

Conditioned on the fact that all the channels are i.i.d Rayleigh faded, both x and y are i.i.d exponentially distributed. Therefore, the difference can be upper bounded as

$$\begin{aligned} \Delta &\leq P \left(\frac{2(2^R - 1)}{\rho} \leq x + y \leq \frac{2^{2R} - 1}{\rho} \right) \\ &= \left[1 - e^{-\frac{2^{2R} - 1}{\rho}} \Sigma_1 \right] - \left[1 - e^{-\frac{2(2^R - 1)}{\rho}} \Sigma_2 \right] \\ &\approx \frac{[2^R - 1]^2}{2\rho^2} [(2^R + 1)^2 - 4] \quad (34) \end{aligned}$$

where $\Sigma_1 = \sum_{k=0}^1 ((2^{2R} - 1)/\rho)^k / k!$, $\Sigma_2 = \sum_{k=0}^1 (2(2^R - 1)/\rho)^k / k!$, and the last step follows from the high SNR

atropinization. Therefore, the probability $P(x \geq ((2^R - 1)/\rho), y \geq ((2^R - 1)/\rho), x + y \geq ((2^{2R} - 1)/\rho))$ can be upper bounded by $P(x \geq ((2^R - 1)/\rho), y \geq ((2^R - 1)/\rho))$, and the difference between the two probabilities is decreasing to zero at high SNR. To simplify the development of analytic results, we use $P(x \geq ((2^R - 1)/\rho), y \geq ((2^R - 1)/\rho))$ to approximate $P(x \geq ((2^R - 1)/\rho), y \geq ((2^R - 1)/\rho), x + y \geq ((2^{2R} - 1)/\rho))$. It is important to note that the use of $P(x \geq ((2^R - 1)/\rho), y \geq ((2^R - 1)/\rho))$ yields a lower bound on the performance. Or in the other words, the analytical results obtained by using the proposed approximation are worse than the actual performance achieved by the proposed cooperative protocol. \blacksquare

Proof for Theorem 2: According to Lemma 1, a qualified relay should satisfy

$$|h_{1r_i}|^2 \geq \varepsilon \quad \text{and} \quad |h_{2r_i}|^2 \geq \varepsilon \quad (35)$$

where $\varepsilon = (2^R - 1)/\rho$. Hence, such a qualified relay can be selected as follows: First, select the l of all the L relays that satisfies $|h_{1r_i}|^2 \geq \varepsilon$. Then, within the l candidates, select the N relay that can satisfy $|h_{2r_i}|^2 \geq \varepsilon$. Obviously, it is necessary that $l \geq N$.

Therefore, the probability of the number N can be found as

$$P(\mathbf{N} = N) = \sum_{l=N}^L P(\mathbf{N} = N | l) P(l = l) \quad (36)$$

where $P(l = l)$ denotes the probability that l out of L relays can decode the messages from the first source correctly, and $P(\mathbf{N} = N | l = l)$ denotes the probability that N out of the l relays can decode the second source message correctly. To calculate $P(l = l)$, assume that $x_i = |h_{1r_i}|^2$ have been ordered as

$$x_{(1)} \leq \dots \leq x_{(L)}.$$

The probability that there are l relays capable to decode the first source's information can be obtained as

$$P(l = l) = P(x_{(L-l)} \leq \varepsilon, x_{(L-l+1)} \geq \varepsilon). \quad (37)$$

Applying order statistics [19], the joint pdf of $x_{(L-l)}$ and $x_{(L-l+1)}$ can explicitly be obtained, and the probability of $P(l = l)$ can be calculated as

$$\begin{aligned} P(l = l) &= c_f \int_0^\varepsilon f(x_{(L-l)}) [F(x_{(L-l)})]^{L-l-1} dx_{(L-l)} \\ &\quad \times \int_\varepsilon^\infty f(x_{(L-l+1)}) [1 - F(x_{(L-l+1)})]^{l-1} dx_{(L-l+1)} \\ &= \frac{L!}{(L-l)! l!} [1 - e^{-\varepsilon}]^{L-l} e^{-l\varepsilon} \quad (38) \end{aligned}$$

where $c_f = L! / (L-l-1)! (l-1)!$. For the two extreme cases, i.e., $l = 0$ and $l = L$, it can easily be proved that $P(l = 0) = [1 - e^{-\varepsilon}]^L$ and $P(l = L) = [e^{-\varepsilon}]^L$. It is interesting to point out that these two extreme cases are perfectly fit with the expression given in (38).

Following similar steps, we have

$$P(\mathbf{N} = N | \mathbf{l} = l) = \frac{l!}{(l - N)!N!} [1 - e^{-\varepsilon}]^{l-N} e^{-N\varepsilon}$$

for $0 \leq N \leq l$. Combined with (38), the probability to have N qualified relays can be approximated as

$$P(\mathbf{N}) = \sum_{l=N}^L \frac{L!}{(L-l)!(l-N)!N!} [1 - e^{-\varepsilon}]^{L-N} e^{-(N+l)\varepsilon} \\ \approx \sum_{l=N}^L \frac{L!}{(L-l)!(l-N)!N!} \varepsilon^{L-N}$$

where the last equation follows the exponential expansion and the assumption of $\varepsilon \ll 1$. ■

Proof for Theorem 3: For the case of $N \geq 1$, the outage probability for the event $P(\mathcal{O}_{N,A_i})$ ($i = 1, 2$) can be written as³

$$P(\mathcal{I}_{N,A_1} \leq R) = P(|h_{1R}|^2 + \gamma_{21}^2 z \leq \beta). \quad (39)$$

For the case of $M = 2$, the weighting coefficient should be equal to $\gamma_{21}^2 = |h_{2R}|^2 / \sum_{m=1}^2 |h_{iR}|^2$. By defining $x = |h_{1R}|^2$ and $y = |h_{2R}|^2$, the outage capacity can be written as

$$P(\mathcal{I}_{N,A_1} \leq R) = \int \int \int_{x+\frac{y}{x+y}z \leq \beta} f_x(x)f_y(y)f_z(z) dx dy dz. \quad (40)$$

The exact expression of this outage probability will be difficult to get directly because it is difficult to obtain the integral region. Therefore, in the following, the upper and lower bounds will be developed to approximate the outage probability:

$$P(\mathcal{I}_{N,A_1} \leq R) = P\left(x + \frac{y}{x+y}z \leq \beta\right) \\ \geq P(x + z \leq \beta). \quad (41)$$

Note that $f_x(x) = e^{-x}$ and $f_z(z) = N f_x(z) f_x(z)^{N-1}$ from order statistics [19]. Hence, the lower bound of the outage probability can be easy to obtain as

$$P(\mathcal{I}_{N,A_1} \leq R) \geq \int_0^\beta [1 - e^{-(\beta-z)}] N e^{-z} (1 - e^{-z})^{N-1} dz \\ \approx \int_0^\beta [\beta - z] N z^{N-1} dz = \beta^{N+1} \frac{1}{(N+1)} \quad (42)$$

where the approximation follows the exponential expansion. To find the upper bound, we can have

$$P(\mathcal{I}_{N,A_1} \leq R) = P\left(x + \frac{1}{\frac{x}{y} + 1}z \leq \beta | y \geq 1\right) P(y \geq 1) \\ + P\left(x + \frac{y}{x+y}z \leq \beta | y < 1\right) P(y < 1)$$

³Without losing generality, the probability for $P(\mathcal{O}_{N,A_1})$ is calculated here.

$$\leq P\left(\frac{x}{y} + \frac{1}{\frac{x}{y} + 1}z \leq \beta\right) \\ + P\left(x + \frac{1}{x+1}z \leq \beta\right). \quad (43)$$

Define $w = x/y$, and the density functions of the new variable w can be obtained as $F_w(w) = w/(w+1)$ and $f_w(w) = 1/(w+1)^2$. Now, the probability can be written as

$$P\left(w + \frac{1}{w+1}z \leq \beta\right) = \int_0^\beta \int_0^{c_\beta} f_z(z) dz f_w(w) dw \\ \approx \int_0^\beta (\beta - w)^N dw = \frac{\beta^{N+1}}{N+1} \quad (44)$$

where $c_\beta = (\beta - w)(w + 1)$. Following similar steps, we can have

$$P\left(x + \frac{1}{x+1}z \leq \beta\right) \approx \frac{\beta^{N+1}}{N+1}. \quad (45)$$

It is interesting to point out that the two probabilities in (44) and (45) are exactly same. Combining these two equations, the upper bound can be obtained as

$$P(\mathcal{I}_{N,A_1} \leq R) \leq \frac{2\beta^{N+1}}{N+1}. \quad (46)$$

Comparing (46) with (42), the outage probability of $P(\mathcal{I}_{N,A_i} \leq R)$ can be bounded as

$$\frac{\beta^{N+1}}{N+1} \leq P(\mathcal{I}_{N,A_i} \leq R) \leq \frac{2\beta^{N+1}}{N+1} \quad (47)$$

$\forall i \in \{1, 2\}, \forall N \in \{1, \dots, L\}$.

For the extreme case of $N = 0$, the outage probability can be shown as $P(\mathcal{I}_{0,A_i} \leq R) = P(|h_{iR}|^2 \leq \beta)$. It can easily be found that the expression of the outage probability in (47) can also be applied for the case of $N = 0$.

By using Theorem 2, the outage probability $P(\mathcal{O}_{A_1})$ can be bounded as

$$\left(\frac{2^R - 1}{\rho}\right)^{L+1} \sum_{N=0}^L \sum_{l=N}^L \frac{L!}{(N+1)!(L-N)!(L-l)!} \leq P(\mathcal{O}_{A_1}) \\ \leq \left(\frac{2^R - 1}{\rho}\right)^{L+1} \sum_{N=0}^L \sum_{l=N}^L \frac{2L!}{(N+1)!(L-N)!(L-l)!}.$$

For the case of $N \geq 1$, the outage probability $P_{\mathcal{O}}(\mathcal{A}_3)$ can be calculated as follows:

$$P(\mathcal{I}_{N,A_3} \leq 2R) = P(\log \{z_1(1 + \rho y)\} \leq 2R)$$

where $z_1 = 1 + \rho \sum_{m=1}^2 |h_{mD}|^2$, and it is a Chi-square distribution with 4 degrees of freedom, i.e.,

$$q_{z_1}(z_1) = \frac{1}{\rho} \frac{z_1 - 1}{\rho} e^{-\frac{z_1 - 1}{\rho}}. \quad (48)$$

Define $z_2 = 1 + \rho y$, whose cumulative distribution function is $F_{z_2}(z_2) = [(z_2 - 1)/\rho]^N$. Therefore, we can have

$$P(\mathcal{I}_{N,A_3} \leq 2R) = \int_1^{2^{2R}} F_{z_2} \left(\frac{2^{2R}}{z_1} \right) f_{z_1}(z_1) dz_1 = \frac{C_R}{\rho^{N+2}} \quad (49)$$

$\forall N \in \{1, \dots, L\}$, where $C_R = \int_1^{2^{2R}} [(2^{2R} - x)/x]^N (x - 1) dx$ is a constant not related with SNR. For the extreme case of $N = 0$, the outage probability can be degraded as $P(\mathcal{I}_{0,A_3} \leq 2R) = P(\log\{(1 + \rho \sum_{m=1}^2 |h_{mD}|^2)\} \leq 2R) = (2^{2R} - 1)^2 / 2\rho^2$. Interestingly, this extreme case can perfectly be fit into the general expression of the outage probability in (49). Again, using Theorem 2, the outage probability $P(\mathcal{O}_{A_3})$ can be written as

$$P(\mathcal{O}_{A_3}) = \sum_{N=0}^L P(\mathcal{O}_{A_3,N}) P(\mathbf{N} = N) \approx \left(\frac{1}{\rho}\right)^{L+2} \sum_{N=0}^L \sum_{l=N}^L \frac{C_R (2^R - 1)^{L-N} L!}{(N)! (L-N)! (L-l)!} \quad (50)$$

According to (14), the outage probability can be bounded as

$$P(\mathcal{O}_{A_1}) \leq P(\mathcal{O}) = P\left(\bigcup_A \mathcal{O}_A\right) \leq \sum_A P(\mathcal{O}_A) \quad (51)$$

which can be written as

$$\epsilon_L \left(\frac{2^R - 1}{\rho}\right)^{L+1} \leq P(\mathcal{O}_{A_1}) \leq P(\mathcal{O}) \leq 4\epsilon_L \left(\frac{2^R - 1}{\rho}\right)^{L+1} + \left(\frac{1}{\rho}\right)^{L+2} \sum_{N=0}^L \sum_{l=N}^L \frac{C_R (2^R - 1)^{L-N} L!}{(N)! (L-N)! (L-l)!}$$

where $\epsilon_L = \sum_{N=0}^L \sum_{l=N}^L (L! / (N+1)! (L-N)! (L-l)!)$. With the assumption of high SNR and fixed data rate, it can be found that $P(\mathcal{O}_{A_3})$ will be dominated by the probability $P(\mathcal{O}_{A_i})$ ($i = 1, 2$). The theorem is proved. ■

Proof for Theorem 4: Rewrite the mutual information as

$$\mathcal{I}_N = \frac{1}{2} \log \left\{ \left(1 + \rho \sum_{m=1}^2 |h_{mD}|^2 \right) (1 + \rho |h_{RD}|^2) \right\}$$

$\forall N \in \{1, \dots, L\}$. Hence, the ergodic capacity can be shown as

$$\mathcal{E}\{\mathcal{I}_N\} = \mathcal{E} \left\{ \frac{1}{2} \log \left(1 + \rho \sum_{m=1}^2 |h_{mD}|^2 \right) \right\} + \mathcal{E} \left\{ \frac{1}{2} \log (1 + \rho |h_{RD}|^2) \right\}.$$

The calculation of the ergodic capacity can be accomplished by two steps. First, define $z = 1 + \rho \sum_{m=1}^2 |h_{mD}|^2$, which has the pdf as $f_z(z) = (1/\rho)((z-1)/\rho)e^{-((z-1)/\rho)}$. The first factor in the previous equation can be shown as

$$\mathcal{E}\{\log z\} \approx \log e + \log \rho - \mathbf{C} \log e. \quad (52)$$

For the section step, define $z_N = 1 + \rho |h_{RD}|^2$ for the case in which there are N qualified relays. The latter part of (52) can be written as

$$\mathcal{E}\{\log z_N\} = \frac{N \log e}{\rho} \int_1^\infty \ln z e^{-\frac{z-1}{\rho}} \left[1 - e^{-\frac{z-1}{\rho}} \right]^{N-1} dz. \quad (53)$$

Applying binomial expansion, we can have

$$\mathcal{E}\{\log z_N\} = \frac{N \log e}{\rho} \sum_{i=0}^{N-1} \binom{N-1}{i} (-1)^{i+1} \frac{e^{\frac{i+1}{\rho}} \rho}{i+1} \mathbf{Ei} \left(-\frac{i+1}{\rho} \right).$$

For large SNR, we can have the following approximation:

$$\mathcal{E}\{\log z_N\} \approx N \log e \sum_{i=0}^{N-1} (-1)^{i+1} \frac{1}{i} \binom{N-1}{i} [\ln \rho - \mathbf{C}] + N \log e \sum_{i=0}^{N-1} \binom{N-1}{i} (-1)^{i+1} \frac{1}{i+1} \ln(i+1).$$

Applying Lemma 2, we can have

$$\mathcal{E}\{\log z_N\} \approx \log \rho - \mathbf{C} \log e + C_N \quad (54)$$

where $C_N = N \log e \sum_{i=0}^{N-1} \binom{N-1}{i} (-1)^{i+1} (1/i+1) \ln(i+1)$.

In summary, the ergodic sum rate that the proposed scheme can achieve is written as

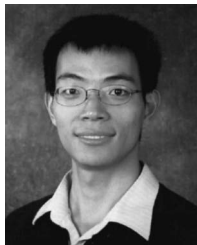
$$\mathcal{E}\{\mathcal{I}_N\} \approx \frac{1}{2} (\log e + \log \rho - \mathbf{C} \log e + [\log \rho - \mathbf{C} \log e + C_N]) = (\log \rho - \mathbf{C} \log e) + \frac{1}{2} (\log e + C_N).$$

The theorem is proved. ■

REFERENCES

- [1] T. S. Rappaport, *Wireless Communications: Principles and Practice*. Englewood Cliffs, NJ: Prentice-Hall, 1998.
- [2] A. Sendonaris, E. Erkip, and B. Aazhang, "User cooperation diversity—Part I: System description," *IEEE Trans. Commun.*, vol. 51, no. 11, pp. 1927–1938, Nov. 2003.
- [3] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3062–3080, Dec. 2004.
- [4] R. U. Nabar, H. Bolcskei, and F. W. Kneubuhler, "Fading relay channels: Performance limits and space-time signal design," *IEEE J. Sel. Areas Commun.*, vol. 22, no. 6, pp. 1099–1109, Aug. 2004.
- [5] A. Bletsas, A. Khisti, D. P. Reed, and A. Lippman, "A simple cooperative diversity method based on network path selection," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 3, pp. 659–672, Mar. 2006.
- [6] K. Azarian, H. E. Gamal, and P. Schniter, "On the achievable diversity-multiplexing tradeoff in half-duplex cooperative channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4152–4172, Dec. 2005.
- [7] Z. Ding, T. Ratnarajah, and C. Cowan, "On the diversity-multiplexing tradeoff for wireless cooperative multiple access systems," *IEEE Trans. Signal Process.*, vol. 55, no. 9, pp. 4627–4638, Sep. 2007.
- [8] R. Ahlswede, N. Cai, S. R. Li, and R. W. Yeung, "Network information flow," *IEEE Trans. Inf. Theory*, vol. 46, no. 4, pp. 1204–1216, Jul. 2000.
- [9] S. Katti, H. Rahul, W. Hu, D. Katabi, M. Medard, and J. Crowcroft, "Xors in the air: Practical wireless network coding," in *Proc. ACM SIGCOMM*, Sep. 2006, pp. 243–254.
- [10] S. Zhang, S. Liew, and P. Lam, "Physical layer network coding," in *Proc. 12th Annu. Int. Conf. Mobile Comput. Netw.*, Sep. 2006, pp. 63–68.
- [11] Z. Ding, K. K. Leung, D. L. Goeckel, and D. Towsley, "On the study of network coding with diversity," *IEEE Trans. Wireless Commun.*, vol. 8, no. 3, pp. 1247–1259, Mar. 2009.

- [12] T. Cover and J. Thomas, *Elements of Information Theory*, 6th ed. New York: Wiley, 1991.
- [13] D. N. C. Tse, P. Viswanath, and L. Zheng, "Diversity-multiplexing trade-off in multiple-access channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 9, pp. 1859–1874, Sep. 2004.
- [14] E. Biglieri, R. Calderbank, A. Constantinides, A. Goldsmith, A. Paulraj, and H. V. Poor, *MIMO Wireless Communications*. Cambridge, U.K.: Cambridge Univ. Press, 2007.
- [15] Z. Ding, T. Ratnarajah, and K. K. Leung, "On the study of network coded AF transmission protocol for wireless multiple access channels," *IEEE Trans. Wireless Commun.*, vol. 8, no. 1, pp. 118–123, Jan. 2009.
- [16] S. Zhu and K. K. Leung, "Distributed cooperative routing for UWB ad-hoc networks," in *Proc. IEEE ICC*, Jun. 2007, pp. 3339–3344.
- [17] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series and Products*, 6th ed. New York: Academic, 2000.
- [18] Z. Ding, K. K. Leung, and M. Peng, "On the combination of network coding with diversity and opportunistic scheduling," in *Proc. IET Commun. Conf. Wireless, Mobile Comput.*, Dec. 2009, pp. 1–5.
- [19] H. A. David and H. N. Nagaraja, *Order Statistics*, 3rd ed. New York: Wiley, 2003.



Zhiguo Ding (S'03–M'05) received the B.Eng. degree in electrical engineering from the Beijing University of Posts and Telecommunications, Beijing, China, in 2000 and the Ph.D. degree in electrical engineering from Imperial College, London, U.K., in 2005.

From July 2005 to June 2010, he was with Queen's University Belfast, Belfast, U.K.; Imperial College; and Lancaster University, Lancaster, U.K. Since June 2010, he has been a Lecturer with Newcastle University, Tyne, U.K. His research interests are cross-

layer optimization, cooperative diversity, statistical signal processing, and information theory.



Kin K. Leung (F'01) received the B.S. degree (with first-class honors) from the Chinese University of Hong Kong, Shatin, Hong Kong, in 1980 and the M.S. and Ph.D. degrees in computer science from the University of California, Los Angeles, in 1982 and 1985, respectively.

He started his career with AT&T Bell Labs, Murray Hill, NJ, in 1986 and worked at its successor companies AT&T Labs and Bell Labs of Lucent Technologies, Holmdel, NJ, until 2004. Since 2004, he has been the Tanaka Chair Professor in Internet

Technology with Imperial College, London, U.K. His research interests include network resource allocation, medium access control protocols, Transmission Control Protocol/Internet Protocol, distributed optimization algorithms, mobility management, network architecture, real-time applications and teletraffic issues for broadband wireless networks, wireless sensors, and ad hoc networks. He is also interested in a wide variety of wireless technologies, including IEEE 802.11, 802.16, and third-generation, as well as future generation cellular networks.