

## Fundamental Limits and Scaling Behavior of Cooperative Multicasting in Wireless Networks

Ashish Khisti, *Student Member, IEEE*, Uri Erez, *Member, IEEE*, and Gregory W. Wornell, *Fellow, IEEE*

**Abstract**—A framework is developed for analyzing capacity gains from user cooperation in slow-fading wireless networks when the number of nodes (network size) is large. The framework is illustrated for the case of a simple multipath-rich Rayleigh-fading channel model. Both unicasting (one source and one destination) and multicasting (one source and several destinations) scenarios are considered. We introduce a meaningful notion of Shannon capacity for such systems, evaluate this capacity as a function of signal-to-noise ratio (SNR), and develop a simple two-phase cooperative network protocol that achieves it. We observe that the resulting capacity is the same for both unicasting and multicasting, but show that the network size required to achieve any target error probability is smaller for unicasting than for multicasting. Finally, we introduce the notion of a network “scaling exponent” to quantify the rate of decay of error probability with network size as a function of the targeted fraction of the capacity. This exponent provides additional insights to system designers by enabling a finer grain comparison of candidate cooperative transmission protocols in even moderately sized networks.

**Index Terms**—*Ad hoc* networks, cooperative diversity, multicasting, outage capacity, scaling laws, sensor networks, wireless networking.

### I. INTRODUCTION

Cooperative diversity has been proposed as an attractive approach to combatting slow fading in wireless networks [8], [13]. Spatially distributed nodes provide an opportunity to create a distributed virtual antenna array and can provide substantial gains in slow fading environments. There has been a significant interest in studying these gains recently; see, e.g., [1], [2], [9], [10], [12], and the references therein.

A convenient channel model for such problems, as has been widely adopted in the literature, is a quasi-static one in which the parameters are known to the receivers, but not to the transmitters. In such scenarios, the classical Shannon capacity is typically zero due to the positive probability of the channel experiencing an arbitrarily deep fade, so performance is instead quantified in terms of outage capacity, which describes the achievable rate subject to a constraint on the level of outage probability that can be tolerated [11].

Outage analysis applies to a host of multiterminal extensions of such basic channel models as well [14], although the expressions become more cumbersome. To address this, diversity–multiplexing tradeoff analysis provides a suitably coarser scale characterization of such systems by focusing on the high signal-to-noise ratio (SNR) regime and examining how outage probability scales with SNR in this regime for different transmission rates [16].

Manuscript received March 14, 2005; revised November 29, 2005. This work was supported in part by the National Science Foundation under Grant CNS-0434974, by Hewlett-Packard through the MIT/HP Alliance, and by NEC Corporation. The material in this correspondence was presented in part at the 2004 Allerton Conference on Communication, Control and Computing, Monticello, IL, September 2004.

A. Khisti and G. W. Wornell are with the Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA 02139 USA (e-mail: khisti@mit.edu; gww@mit.edu).

U. Erez is with the Department of Electrical Engineering-Systems, Tel-Aviv University, Ramat-Aviv 69978, Israel (e-mail: uri@eng.tau.ac.il).

Communicated by B. Prabhakar, Guest Editor.

Digital Object Identifier 10.1109/TIT.2006.874541

Diversity–multiplexing tradeoff analysis has also proven useful in analyzing a host of simple network problems. For example, [15] extends the analysis to the multiple-access channel, while [8] extends the analysis to the cooperative diversity channel.

While such analysis of cooperative diversity has proven popular, much of the work has been limited to systems in which for a given message there is effectively only a single destination node and a relatively small number of potential relay nodes to participate in the transmission.

In the present correspondence, we develop an alternative framework within which to examine cooperative protocols. First, our emphasis is on the multicasting scenario in which there is one message in the network, but generally multiple destination nodes. We will focus on two extreme special cases of this scenario. One is when *all* nodes in the network are to receive the message, which for convenience we generically refer to as multicasting. The other is when exactly one node in the network is to receive the message, which we refer to as unicasting.

Second, our framework examines the scenario in which the number of nodes in the network is large. This will allow us to examine the associated asymptotic scaling behavior of cooperative networks. As a by-product, we do not need to restrict our attention to high SNR analysis. Indeed, we fix the noise power, normalize the channel statistics, and constrain the total power transmitted in the network. This allows us to parameterize our results in terms of the associated SNR.

Within this framework, we analyze the relationship between transmission rate and the associated error probability. Provided we use codes of sufficiently long block lengths, outage probability dominates the error probability. Specifically, the associated outage event is that not *all* the intended recipients are able to decode the message. Not surprisingly, avoiding outage in multicasting is more difficult than in unicasting.

As our main result, we show that under a multipath-rich Rayleigh-fading network model, a notion of Shannon capacity can be developed. Specifically, there exists a nonzero capacity (dependent on SNR) such that for all rates below capacity, the error probability can be made arbitrarily small provided the network is sufficiently large. Conversely, for all rates above capacity, the error probability is bounded away from zero regardless of network size. Our achievability result is based on a simple two-phase cooperative network protocol we develop. By contrast, when one precludes the possibility of cooperation, the associated capacity is of course zero. Interestingly, our analysis also reveals that despite the fact that multicasting outage behavior is dominated by the *worst* node, the multicasting and unicasting capacities thus defined are identical. Not surprisingly, we also show that for a fixed number of nodes, the probability of outage is still much smaller in unicasting than in multicasting.

We further show that finer scale characterizations of behavior are possible too. In particular, we define a notion of network scaling exponent that characterizes the rate of decay of error probability with network size as a function of the targeted fraction of capacity. Within this analysis, we see, among other insights, that the exponent of our capacity-achieving protocol is quite small for rates that exceed half the network capacity.

While our results are specific to our multipath-rich Rayleigh-fading model, we believe that the associated framework is useful more broadly in the analysis of user cooperation gains in large networks with more realistic—if more complicated—models. Indeed, ultimately our results more generally suggest that just as system analysis asymptotic in block length or SNR has proven useful, so can one that is asymptotic in network size.

The remainder of the correspondence is organized as follows. Section II introduces the network model of interest. The capacity result is

stated in Section III. It is established by providing a converse in Section IV and an achievability argument in Section V. The scaling of the outage probability with the number of nodes is discussed in Section VI, and the network scaling exponent is introduced in Section VII. Finally, Section VIII contains some concluding remarks and directions for future work.

## II. SYSTEM MODEL

We consider a system with  $K$  receiving nodes and one source node. For convenience, we label the source node as node 0, and the receiving nodes as  $\{1, 2, \dots, K\}$ . Of course, in practice, different nodes in the network can act as source nodes over orthogonal time or frequency bands as discussed in [8]. However, for analysis, it suffices to focus on a single configuration.

We assume a narrowband, slow fading channel passband channel model corresponding to a multipath-rich propagation environment. In particular, the channel gains  $h_{jk}$  between arbitrary distinct pairs of nodes  $(j, k)$  are independent and identically distributed (i.i.d.) random variables from a zero-mean, unit-variance circularly symmetric complex Gaussian distribution. In turn, the signal received at node  $k$  at time  $i$  is given by

$$y_k(i) = \sum_{j \in \mathcal{T}(i)} h_{jk} x_j(i) + z_k(i) \quad (1)$$

where  $\mathcal{T}(i)$  is the set of nodes transmitting at time  $i$ , where  $x_j(i)$  is the symbol transmitted by node  $j$  at time  $i$ , and where  $z_k(i)$  denotes circularly symmetric complex i.i.d. Gaussian noise of power  $N_0$ . Furthermore, the noises among the different receivers are mutually independent.

In our model, nodes are subject to a half-duplex constraint, i.e., a node cannot transmit and receive simultaneously. Thus, associated with every valid protocol is a set of binary variables of the form  $D_k(i)$  that specifies at time  $i$  whether node  $k$  is transmitting ( $D_k(i) = 1$ ) or receiving ( $D_k(i) = 0$ ).

The source sends one of  $M$  possible messages to the destination node(s) over  $n$  channel uses (i.e.,  $i = 1, 2, \dots, n$ ). The channel gains between all pairs of nodes remains fixed over this duration. In our model, the channel gain  $h_{jk}$  is known to the receiving node  $k$  but not to the transmitting node  $j$ .

We further restrict our attention to protocols in which relay nodes cannot revert to receive mode once they begin transmitting, i.e., if  $D_k(i) = 1$  for some node  $k$  and  $i < n$  then  $D_k(j) = 1$  for all  $i \leq j \leq n$ . This restriction precludes protocols in which transmitting nodes effectively learn and exploit the network channel gains in their encodings.

Finally, for simplicity, we adopt a long-term sum power constraint across the nodes in our model. In particular, with  $X_j(i)$  for  $j \in \mathcal{T}(i)$  denoting the (complex-valued) symbol being transmitted by node  $j$  at time  $i$ , we impose an expected sum power constraint of the form

$$E \left[ \frac{1}{n} \sum_{i=1}^n \sum_{k \in \mathcal{T}(i)} |X_k(i)|^2 \right] \leq P \quad (2)$$

where the expectation is taken over the ensemble of channel realizations and the set  $\mathcal{T}(i)$ , as well as any other randomized aspects of the protocol. Indeed, one will want to consider protocols in which the set  $\mathcal{T}(i)$  depends on the realized channel gains.

The power constraint (2) is a rather natural one for systems in which there are ergodic channel variations but a stringent delay constraint that requires transmission of any particular message within a single coherence interval. Indeed, although we do not satisfy the sum power constraint during the transmission of an individual message, the expectation in (2) ensures that it will be satisfied with high probability over

a sufficiently long sequence of messages. Nevertheless, we remark in advance that the results of the correspondence do not change when we require the sum power constraint to be met with high probability in every coherence interval, and our capacity-achieving protocol can be readily extended to this case. Ultimately, the expected power constraint merely simplifies the exposition.

The preceding discussion characterizes an admissible protocol for our analysis, which we formalize in the following definition.

*Definition 1:* An admissible protocol  $\pi_K$  consists of a set of indicator functions  $\{D_k(i)\} \in \{0, 1\}$ , which determines whether node  $k$  is transmitting or receiving at time  $i$ ; a set of encoding functions  $\{\phi_k(i)\} \in \mathbb{C}$ , which determines the symbol produced by node  $k$  at time  $i$ ; and a set of decoding functions  $\{\psi_k\} \in \{1, 2, \dots, M\}$ , which determines the message decisions produced by node  $k$  at time  $n$ . These functions are further constrained by their usage as described below.

During the initialization phase of the protocol, the source node 0 selects message  $W \in \{1, 2, \dots, M\}$  for transmission. The protocol chooses *a priori* the sequence  $D_0(i)$  for  $i = 1, 2, \dots, n$ . Without loss of generality,  $D_0(1) = 1$  and  $D_k(1) = 0$  for  $k = 1, 2, \dots, K$ . Moreover, the collections of observations  $\mathcal{Y}_k$  at each node  $k$  are initialized:  $\mathcal{Y}_k = \emptyset$ .

At time  $i$ , for  $1 \leq i < n$ , if  $D_0(i) = 1$ , then source node 0 uses encoding function  $\phi_0(i)$  to map  $W$  into a transmitted complex-valued symbol  $x_0(i)$ .

If node  $k \in \{1, 2, \dots, K\}$  is in transmit mode at that time (i.e.,  $D_k(i) = 1$ ), the encoding function  $\phi_k(i)$  at node  $k$  maps  $\mathcal{Y}_k$  and the complex-valued channel gains  $\{h_{jk}, j = 0, 1, \dots, K\}$  into a transmitted complex-valued symbol  $x_k(i)$ , which is transmitted over the channel.

If, instead, the node is in receive mode (i.e.,  $D_k(i) = 0$ ), then it collects the complex-valued measurement  $y_k(i)$  and updates its set of received symbols via  $\mathcal{Y}_k := \mathcal{Y}_k \cup \{y_k(i)\}$ . If  $i = n$ , the decoding function  $\psi_k$  at node  $k$  maps  $\mathcal{Y}_k$  and the complex-valued channel gains  $\{h_{jk}, j = 0, 1, \dots, K\}$  into a decision  $\hat{W}_k$ . Note that without loss of generality,  $D_k(n) = 0$  for at least one value of  $k$ . If  $i < n$ , the node makes a decision whether to switch to transmit mode for the remaining duration. If it decides to switch, it sets  $D_k(j) = 1$  for  $i + 1 \leq j \leq n$ ; otherwise, it sets  $D_k(i + 1) = 0$ .

A cooperation-free protocol is a special case of the preceding definition.

*Definition 2:* A *cooperation-free* admissible protocol is one for which only the source node transmits, i.e.,  $D_k(i) = 0$  for  $1 \leq k \leq K$  and  $1 \leq i \leq n$ .

## III. CODING THEOREMS

We now develop the relationship between transmission rate and error probability for such protocols, in the limit of large network sizes.

We begin with a meaningful definition of capacity.

*Definition 3:* A rate  $R$  is *achievable* for the unicasting (respectively, multicasting) system if for every network size  $K$ , there exists an admissible protocol  $\pi_K$  with  $n_K$  channel uses and  $M_K = 2^{n_K R}$  messages such that the probability that the destination node (respectively, any node) fails to decode the message approaches zero as  $K \rightarrow \infty$ . The (unicasting or multicasting) *capacity*  $C$  is the supremum of all achievable rates.

With this definition, we have the following coding theorem, which is our main result.

*Theorem 1:* The unicasting and multicasting capacities are identical and given by<sup>1</sup>

$$C = \log \left( 1 + \frac{P}{N_0} \right) \quad (3)$$

where  $P$  is as defined in (2) and  $N_0$  is the noise power as defined via (1).

Before presenting our proof of this result, we note that to achieve capacity—indeed, any nonzero rate—requires the use of cooperation. Formally, we have the following result.

*Theorem 2:* The capacity of cooperation-free admissible protocols for both unicasting and multicasting is

$$C_{nc} = 0 \quad (4)$$

whenever  $P$  in (2) is finite and  $N_0$  in (1) is nonzero.

*Proof:* From Definition 2, a destination node must decode directly from the source transmission. Since the channel is a Rayleigh-fading channel with  $P/N_0 < \infty$ , there exists a strictly positive probability of outage—and hence, probability of error—for every positive transmission rate  $R$ . Since these probabilities are independent of  $K$ , (4) follows.  $\square$

We now proceed to the proof of Theorem 1. Since we consider large block lengths, error probability is dominated by outage probability. Specifically, we consider blocks long enough that the probability of error when there is no outage is negligible compared to the outage probability. Thus, in the error analysis in our proof, we restrict our attention to outage probability.

#### IV. PROOF OF CONVERSE PART

We develop a converse via a simple upper bound on the achievable rate  $R$  of Definition 3. In particular, suppose a genie conveys the message  $W$  to nodes  $1, 2, \dots, K-1$  and only destination node  $K$  remains to be served. Thus, nodes  $0, 1, \dots, K-1$  can coordinate to send the message to destination node  $K$ . This is clearly a multiple-input single-output (MISO) antenna system with  $K$  antenna elements and channel knowledge only at the receiver. Thus, for a given rate  $R$ , a lower bound on the outage probability for the MISO channel is a lower bound on that for both unicasting and multicasting systems. We let  $\mathcal{E}_K^{\text{MISO}}$  denote the MISO channel outage event.

To develop such a bound, we first note that it suffices to restrict the input distribution to i.i.d. Gaussian.

*Lemma 1 (Teletar [14]):* The outage capacity of the slow-fading MISO channel with  $K$  transmit antennas in i.i.d. Rayleigh fading with total power constraint  $P$  is achieved by an input distribution with covariance matrix  $\text{diag}(P_1, P_2, \dots, P_K)$  where  $\sum_{j=1}^K P_j = P$ .

We now establish the following lemma.

*Lemma 2:* Let  $\epsilon > 0$  be arbitrary, and let

$$R = \log \left( 1 + \frac{P}{N_0} \right) + \epsilon$$

for a  $K$ -antenna MISO channel. Then the outage probability  $\Pr \{ \mathcal{E}_K^{\text{MISO}} \}$  is bounded away from zero, i.e.,

$$\inf_K \Pr \{ \mathcal{E}_K^{\text{MISO}} \} > 0. \quad (5)$$

<sup>1</sup>Unless otherwise indicated, all logarithms are base 2.

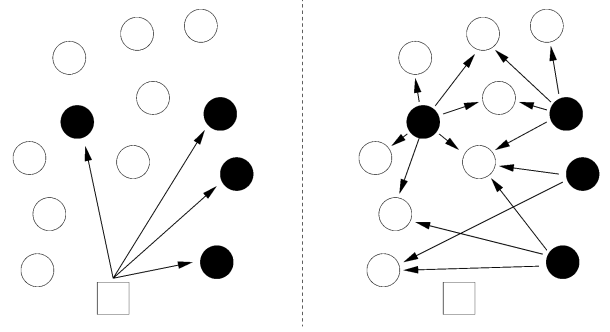


Fig. 1. The two-phase cooperative multicasting protocol. In phase 1 (left), the source node (square) broadcasts at a high rate and only a small fraction (solid) of the many destination nodes (discs) are able to decode. In phase 2 (right), these nodes cooperatively broadcast the message to the remaining nodes using a suitable space-time code.

*Proof:* Let  $P_1^*, P_2^*, \dots, P_K^*$ , be the power allocations that minimize the outage event for the selected rate. The corresponding mutual information is given by

$$I = \log \left( 1 + \frac{1}{N_0} \sum_{i=1}^K P_i^* |h_{iK}|^2 \right). \quad (6)$$

For outage event  $\mathcal{E}_K^{\text{MISO}} = \{R > I\}$ , we have

$$\begin{aligned} \Pr \{ \mathcal{E}_K^{\text{MISO}} \} &= 1 - \Pr \{ I \geq R \} \\ &\geq 1 - \frac{E \left\{ \log \left( 1 + \frac{1}{N_0} \sum_{i=1}^K P_i^* |h_{iK}|^2 \right) \right\}}{R} \end{aligned} \quad (7)$$

$$\geq 1 - \frac{\log \left( 1 + \frac{1}{N_0} \sum_{i=1}^K E[|h_{iK}|^2] P_i^* \right)}{R} \quad (8)$$

$$= 1 - \frac{\log \left( 1 + \frac{1}{N_0} \sum_{i=1}^K P_i^* \right)}{R} \quad (9)$$

$$= 1 - \frac{\log \left( 1 + \frac{P}{N_0} \right)}{R} = \frac{\epsilon}{R} > 0.$$

In the above derivation, (7) follows from the Markov inequality, (8) is a consequence of Jensen's inequality, and (9) follows from  $E[|h_{iK}|^2] = 1$ , for all  $i$  from our fading model. Since the above result holds for all  $K$ , (5) follows.  $\square$

#### V. PROOF OF FORWARD PART

A simple two-phase cooperative protocol can achieve any rate below the capacity (3), as we now develop.

##### A. A Two-Phase Cooperative Protocol

The protocol of interest is depicted in Fig. 1. Specifically, in phase 1, the source node broadcasts the message over  $n_1$  channel uses at a rate  $R_1$  and all nodes attempt to decode the message. Nodes that cannot decode the message discard the source transmission. Then, in phase 2, the nodes that are successful in decoding the message act as relays and form a virtual antenna array, transmitting over the remaining  $n_2 = n - n_1$  channel uses at a rate  $R_2$ . At this point, the intended destination(s) attempt(s) to decode the message and an outage is declared if any of the intended destinations fail.

*Codebook Generation:* Suppose that the source generates  $M$  codewords i.i.d.  $\mathcal{CN}(0, P_1)$  for some  $P_1 > 0$ , each of length  $N_1$ , and all nodes each independently generate  $M$  codewords i.i.d.  $\mathcal{CN}(0, P_2)$  for some  $P_2 > 0$ , each of length  $n_2$ . We describe the main steps of the

protocol for the case of multicasting, but indicate the straightforward modifications for the case of unicasting.

*Phase 1:* The source transmits the codeword corresponding to the intended message from its codebook over  $n_1$  channel uses. We choose the rate in this phase to be (strictly less than, but arbitrarily close to)

$$R_1(\alpha) = \log \left( 1 + G(\alpha) \frac{P_1}{N_0} \right) \quad (10)$$

where

$$G(\alpha) = F^{-1}(1 - \alpha) \quad (11)$$

and  $F(\cdot)$  denotes the cumulative distribution function of an arbitrary channel gain  $|h_{ij}|^2$ , and where we have made the dependency of  $R_1$  on  $\alpha$  explicit.

All nodes attempt to decode the transmission. A node is successful in decoding the message if it finds a codeword in the source codebook that is jointly typical with the received sequence. Let  $K_1$  be the number of nodes that are successful in decoding the message from the source. We label these nodes as  $1, 2, \dots, K_1$ . These nodes participate in Phase 2.

*Phase 2:* Each of the  $K_1$  nodes successful in decoding the Phase 1 transmission next transmits the corresponding codeword from its codebook over  $n_2$  channel uses. The rate in this phase is set to (strictly less than, but arbitrarily close to)

$$R_2(\beta) = \log \left( 1 + \frac{P_2}{N_0} (1 - \beta) \right) \quad (12)$$

where  $0 < \beta < 1$  is design parameter, the dependence of  $R_2$  on which we have made explicit.

Each of the remaining  $K_2 = K - K_1$  nodes attempts to decode the message at the end of the second phase. Node  $k$ , upon receiving its observations<sup>3</sup>  $y_k^{n_2}$  finds a message  $\hat{w}_k$  and a subset of nodes

$$\mathcal{S}_k = \{k_1, k_2, \dots, k_{|\mathcal{S}_k|}\} \subseteq \{1, 2, \dots, K\} \setminus \{k\}$$

such that the corresponding set of codewords

$$\left\{ x_{k_1}^{n_2}(\hat{w}_k), x_{k_2}^{n_2}(\hat{w}_k), \dots, x_{k_{|\mathcal{S}_k|}}^{n_2}(\hat{w}_k) \right\}$$

is jointly typical with  $y_k^{n_2}$ . It declares the message  $\hat{w}_k$  to be the transmitted message if a unique pair  $(\hat{w}_k, \mathcal{S}_k)$  exists and declares a failure otherwise.

In the case of unicasting, if the destination node is successful in decoding the message in Phase 1 then it does not participate in Phase 2. Otherwise, it continues to listen to the transmissions and attempts to decode the Phase 2 transmission. An error occurs if the destination fails to decode the Phase 2 transmission.

### B. Protocol Analysis

First, we analyze code rate. To begin, it is straightforward to verify that  $n_1$  and  $n_2$  are completely determined by the choice of rates. In particular, let the overall rate of our protocol be  $R$ , so that there are  $M = 2^{nR}$  possible messages to send over  $n$  channel uses in the system. Then it follows that

$$n_1 R_1 = n_2 R_2 = \log M. \quad (13)$$

<sup>2</sup>This technicality ensures that the probability of error when not in outage will approach zero uniformly over all channel realizations. A similar technicality applies to the rate in Phase 2.

<sup>3</sup>We use the superscript  $n_2$  to denote the vector formed from the  $n_2$  variables corresponding to time instants  $n_1 + 1, n_1 + 2, \dots, n$ , i.e., Phase 2 of the protocol.

From (13), it is straightforward to calculate the overall effective rate  $R$  of the system. In particular, since  $\log M = nR$  with  $n_1 + n_2 = n$ , (13) implies that  $R$  satisfies

$$\frac{1}{R} = \frac{1}{R_1(\alpha)} + \frac{1}{R_2(\beta)}. \quad (14)$$

Second, we analyze the power constraint (2). In Phase 1, the transmitted power is  $P_1$ , so provided  $P_1 \leq P$ , our power constraint is met in this phase. To analyze the power used in Phase 2, we begin by noting that on average a fraction  $\alpha$  of the nodes are able to decode the message after Phase 1. Specifically, the number of nodes  $K_1$  successful in Phase 1 has mean

$$E[K_1] = \alpha K \quad (15)$$

since  $K_1$  is a binomial random variable, viz. (cf. (10))

$$K_1 = \sum_{i=1}^K \mathbf{1}_{\{|h_{0i}|^2 > G(\alpha)\}} \quad (16)$$

where  $\mathbf{1}_{\{\cdot\}}$  is an indicator function, which equals 1 if its subscript is true and 0 otherwise, and where we have set  $P_1 = P$ . Hence,

$$E \left[ \sum_{j=1}^{K_1} |X_j|^2 \right] = P_2 E[K_1] = P_2 \alpha K$$

from which we see that the power constraint is satisfied in Phase 2 provided  $P_2 \leq P/\alpha K$ .

Finally, we analyze the outage probability, i.e., the probability of outage of a node that is unable to decode at the end of Phase 2 of the protocol. For convenience, let us exploit symmetry and label this node  $K$ , while the nodes that are successful in Phase 1 we label  $1, 2, \dots, K_1$ . From straightforward MISO system analysis, node  $K$  will fail to decode the message whenever  $K_1 < K$  and

$$R_2(\beta) > \log \left( 1 + G_K(K_1) \frac{P}{N_0} \right) \quad (17)$$

where

$$G_k(k_1) \triangleq \frac{1}{\alpha K} \sum_{j=1}^{k_1} |h_{jk}|^2 \quad (18)$$

is the effective MISO channel gain of node  $k$ , and where we have set  $P_2 = P/(\alpha K)$ . But since  $R_2$  was chosen according to (12) in Phase 2, (17) implies that outage will occur when  $G_K(K_1) \leq 1 - \beta$ . Accordingly, the outage events  $\mathcal{E}_{\alpha, \beta, K}^{\text{uc}}$  and  $\mathcal{E}_{\alpha, \beta, K}^{\text{mc}}$  for unicasting and multicasting, respectively, take the form

$$\mathcal{E}_{\alpha, \beta, K}^{\text{uc}} = \bigcup_{k_1=0}^{K-1} \mathcal{A}_{k_1}^{\text{uc}}$$

$$\mathcal{E}_{\alpha, \beta, K}^{\text{mc}} = \bigcup_{k_1=0}^{K-1} \mathcal{A}_{k_1}^{\text{mc}}$$

where, for  $k_1 \in \{1, 2, \dots, K\}$

$$\mathcal{A}_{k_1}^{\text{uc}} = \{K_1 = k_1, \quad G_K(k_1) \leq 1 - \beta\}$$

$$\mathcal{A}_{k_1}^{\text{mc}} = \left\{ K_1 = k_1, \quad \min_{k: k_1 \leq k \leq K} G_k(k_1) \leq 1 - \beta \right\}.$$

The following lemma provides bound on the conditional outage probability that will be useful in the sequel.

*Lemma 2:* Suppose  $k_1 \geq \alpha K(1 - \beta)$  nodes are successful at the end of Phase 1. Then the conditional probability of outage is given by

$$\Pr \{ \mathcal{E}_{\alpha, \beta, K}^{\text{uc}} \mid K_1 = k_1 \} \leq \left( \frac{\alpha K(1 - \beta)e}{k_1} \right)^{k_1} e^{-\alpha K(1 - \beta)}. \quad (19)$$

*Proof:* To obtain (19), it suffices to bound

$$\Pr\{G_K(K_1) \leq 1 - \beta \mid K_1 = k_1\}$$

since

$$\Pr\{\mathcal{E}_{\alpha,\beta,K}^{\text{uc}} \mid K_1 = k_1\} = \Pr\{G_K(K_1) \leq 1 - \beta \mid K_1 = k_1\}.$$

This can be accomplished by the Chernoff bound, since  $G_K(k_1)$  in (18) can be written as

$$G_K(k_1) = \sum_{j=1}^{k_1} g_j \quad (20)$$

where the

$$g_j = \frac{1}{\alpha K} |h_{jK}|^2 \quad (21)$$

are i.i.d. random variables. Specifically, we obtain, for any  $s > 0$

$$\Pr\{G_K(K_1) \leq 1 - \beta \mid K_1 = k_1\} \quad (22)$$

$$= \Pr\left\{e^{-sG_K(K_1)} \geq e^{-s(1-\beta)} \mid K_1 = k_1\right\} \quad (23)$$

$$\leq e^{s(1-\beta)} E\left[e^{-sG_K(K_1)} \mid K_1 = k_1\right] \quad (24)$$

$$= e^{s(1-\beta)} (E[e^{-sg_i}])^{k_1} \quad (25)$$

$$= \frac{e^{s(1-\beta)}}{(1 + s/(\alpha K))^{k_1}} \quad (26)$$

where (24) follows from the Markov inequality, (25) follows from (20), and (26) follows from evaluating the characteristic function of the exponential random variables (21).

In turn, since (26) holds for all  $s > 0$ , we can choose the particular value

$$s = \frac{k_1}{1 - \beta} - \alpha K. \quad (27)$$

Substituting (27) into (26) yields (19) as desired.  $\square$

We now show that the probability of outage can be made arbitrarily small by selecting  $\alpha K$  appropriately.

*Proposition 1:* The probability of outage in unicasting decreases exponentially with  $\alpha K$  for every  $\beta > 0$ . Specifically, for every  $0 < \epsilon < \beta$

$$\Pr\{\mathcal{E}_{\alpha,\beta,K}^{\text{uc}}\} = E\left[\Pr\{\mathcal{E}_{\alpha,\beta,K}^{\text{uc}} \mid K_1\}\right] \leq \exp(-\alpha K \epsilon^2/4) + \exp\{\alpha K \gamma(\beta, \epsilon)\} \quad (28)$$

where

$$\gamma(\beta, \epsilon) \triangleq \beta - \epsilon + (1 - \epsilon) \ln\left(\frac{1 - \beta}{1 - \epsilon}\right) < 0. \quad (29)$$

*Proof:* To obtain (28) we observe that, for some  $\epsilon \in (0, \beta)$

$$\begin{aligned} & \Pr\{\mathcal{E}_{\alpha,\beta,K}^{\text{uc}}\} \\ &= \sum_{k_1: k_1 < \alpha K(1-\epsilon)} \Pr\{\mathcal{E}_{\alpha,\beta,K}^{\text{uc}} \mid K_1 = k_1\} \Pr\{K_1 = k_1\} \\ & \quad + \sum_{k_1: k_1 \geq \alpha K(1-\epsilon)} \Pr\{\mathcal{E}_{\alpha,\beta,K}^{\text{uc}} \mid K_1 = k_1\} \Pr\{K_1 = k_1\} \quad (30) \\ & \leq \Pr\{K_1 < \alpha K(1-\epsilon)\} + \max_{k_1: k_1 \geq \alpha K(1-\epsilon)} \Pr\{\mathcal{E}_{\alpha,\beta,K}^{\text{uc}} \mid K_1 = k_1\} \\ & = \Pr\{K_1 < \alpha K(1-\epsilon)\} + \Pr\{\mathcal{E}_{\alpha,\beta,K}^{\text{uc}} \mid K_1 = \alpha K(1-\epsilon)\} \quad (31) \end{aligned}$$

where (31) exploits that outage probability is a decreasing function of  $k_1$ . Finally, using the binomial Chernoff bound (see, e.g., [4])

$$\Pr\{K_1 < \alpha K(1-\epsilon)\} \leq e^{-\alpha K \epsilon^2/4}$$

for the first term in (31), and applying Lemma 2 to the second term, yields (28) as desired.  $\square$

In turn, Proposition 1 can be used to bound the corresponding probability of outage in multicasting.

*Proposition 2:* The probability of outage in multicasting decays exponentially with  $\alpha K$ . Specifically

$$\Pr\{\mathcal{E}_{\alpha,\beta,K}^{\text{mc}}\} = E\left[\Pr\{\mathcal{E}_{\alpha,\beta,K}^{\text{mc}} \mid K_1\}\right] \leq K \Pr\{\mathcal{E}_{\alpha,\beta,K}^{\text{uc}}\}. \quad (32)$$

*Proof:* First, we bound the conditional outage probability according to

$$\Pr\{\mathcal{E}_{\alpha,\beta,K}^{\text{mc}} \mid K_1 = k_1\} \quad (33)$$

$$= \Pr\left\{\min_{k:k_1 \leq k \leq K} G_k(K_1) < 1 - \beta \mid K_1 = k_1\right\}$$

$$= \Pr\left\{\bigcup_{i=K_1+1}^K \{G_i(K_1) \leq 1 - \beta\} \mid K_1 = k_1\right\}$$

$$\leq (K - k_1) \Pr\{\mathcal{E}_{\alpha,\beta,K}^{\text{uc}} \mid K_1 = k_1\} \quad (34)$$

$$\leq K \Pr\{\mathcal{E}_{\alpha,\beta,K}^{\text{uc}} \mid K_1 = k_1\} \quad (35)$$

where (34) is a simple application of the union bound. Taking the expectation of both sides of (35) with respect to  $K_1$ , we obtain (32). Finally, since Proposition 1 establishes that the unicasting outage probability decays exponentially, (35) implies that the multicasting outage probability does as well.  $\square$

Propositions 1 and 2 can be used to establish the forward part of the coding theorem for both unicasting and multicasting.

*Proof Theorem 1:* To show that our two-phase protocol can approach the capacity (3) we show that the outage probability can be made arbitrarily small while operating arbitrarily close to the capacity. Suppose that  $\alpha > 0$  and  $\beta > 0$  are arbitrary. Since the outage probability decreases exponentially in  $K$ , we can choose a  $K$  large enough to make the outage probability sufficiently small. Next, note that by choosing  $\alpha$  and  $\beta$  sufficiently small, we can make  $R_1(\alpha)$  sufficiently large and  $R_2(\beta)$  sufficiently close to  $C$ . As particular examples, it suffices to take  $\alpha \sim 1/\log K$  and  $\beta \sim 1/K$  so that  $\alpha, \beta \rightarrow 0$  but  $\alpha K \rightarrow \infty$ . Thus, we can have the effective rate (14) be arbitrarily close to  $C$ , while keeping the outage probability sufficiently small.  $\square$

An intuition behind the achievability result is that in the limit of a large number of nodes, we can find sufficiently many nodes (albeit a small fraction of the population) with very large channel gains and they can be served over a small number of channel uses in the first phase (i.e.,  $n_1$  is a negligible fraction of  $n_2$ ). These nodes then simultaneously cooperate to serve the remaining nodes. Since sufficiently many nodes are transmitting in the second phase, we have enough diversity in the system to drive the outage probability to zero.

### C. Multiple Antenna Generalization

It is possible to generalize our results to the case where the at least some of the nodes in the network have multiple antennas. In particular, suppose that the node  $i$  has  $T_i$  antennas. In the case of unicasting, our two-phase protocol can be straightforwardly extended to obtain the following:

$$C_{\text{ma}}^{\text{uc}} = T_K \log\left(1 + \frac{P}{N_0}\right). \quad (36)$$

In the bound (36) the key quantity of interest is  $T_K$ , the number of antennas at the destination node; the number of antennas at the source and relay nodes does not impact capacity. Note that, in the first phase of the

protocol, we can still communicate to a large number of relay nodes, regardless of the number of antennas at each relay. These nodes then form a virtual antenna to communicate to the destination in Phase 2. This reduces to the case of a multiple-input multiple-output (MIMO) system when the number of transmit antennas is much larger than the number of receive antennas. Using the channel hardening result for such systems—see, e.g., [7]—one can establish that rate  $C_{\text{ma}}^{\text{uc}}$  is achievable. The converse is analogous to the single-antenna case in Section IV.

An analogous argument for multicasting can also be developed:

$$C_{\text{ma}}^{\text{mc}} = \min\{T_1, T_2, \dots, T_K\} \cdot \log\left(1 + \frac{P}{N_0}\right). \quad (37)$$

Evidently, (37) can be much smaller than (36)—the system rate is governed by the node with the fewest antennas in multicasting rather than the destination node.

## VI. OUTAGE SCALING BEHAVIOR

Our capacity result determines the rates for which outage probability goes to zero with increasing network size for multicasting and unicasting. Often, a finer grain analysis is required by system designers. In this section, we develop the manner in which outage probability goes to zero with increasing network size for the two-phase protocol of Section V, which provides several additional insights. For example, while we have shown that multicasting and unicasting share the same capacity, here we show how their respective outage probability curves differ.

### A. Outage Probability Approximations

While (28) and (32) bound the outage probabilities of interest, these bounds are not tight. Nevertheless, good approximations to the actual outage are readily obtained, as we now develop.

The outage probability of a unicasting system under the two-phase protocol can be approximated by

$$\Pr\{\mathcal{E}_{\alpha,\beta,K}^{\text{uc}}\} \approx \frac{1}{\sqrt{K}} \exp\{-\alpha K(1-\beta)\} \times \exp\left\{-K D(\gamma||\alpha) + \gamma K \ln\left(\frac{\alpha(1-\beta)e}{\gamma}\right)\right\} \quad (38)$$

where

$$\gamma = \frac{\sqrt{1+4\mu}-1}{2\mu} \quad (39)$$

with

$$\mu = \frac{\alpha^2(1-\beta)}{1-\alpha}. \quad (40)$$

In turn, the outage probability of the multicasting system can be approximated in terms of this unicasting approximation according to

$$\Pr\{\mathcal{E}_{\alpha,\beta,K}^{\text{mc}}\} = 1 - (1 - \Pr\{\mathcal{E}_{\alpha,\beta,K}^{\text{uc}}\})^K \approx K \Pr\{\mathcal{E}_{\alpha,\beta,K}^{\text{uc}}\}. \quad (41)$$

A derivation of the approximation (38) is provided in the Appendix.

### B. Accuracy of Outage Probability Approximations

In this subsection, we compare our outage probability bounds (28) and (32); and our approximations (38) and (41), to the actual probabilities via Monte Carlo simulations. In particular, we choose a target rate below capacity and evaluate the outage probability as a function of the network size  $K$ . We evaluate the expectations over  $K_1$  in the bounds (28) and (32) by numerical integration.

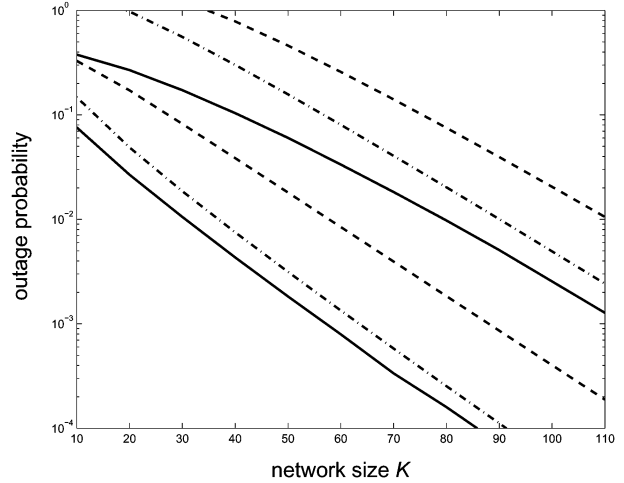


Fig. 2. Outage probability for unicasting and multicasting via our two-phase protocol as a function of network size. The solid curves correspond to our Monte Carlo simulations, the dashed-dotted curves to our analytical approximations, and the dashed curves to our bounds. The top set of curves is for multicasting; the bottom set for unicasting. In the protocol we set  $\beta = 0.5$  and  $R_1(\alpha) = R$ , so that the rate is  $R = (1/2)\log(1 + P/(2N_0)) < C$ .

For our comparison, we set a rate of

$$R = \frac{1}{2} \log\left(1 + \frac{P}{2N_0}\right)$$

which is  $1/2$  of capacity in the high-SNR regime and  $1/4$  of capacity in the low-SNR regime. This rate point is realized by the parameter settings  $G(\alpha) = 1/2$  and  $\beta = 1/2$  in our two-phase protocol, so  $R_1(\alpha) = R_2(\beta)$  in (10) and (12), respectively.

Fig. 2 depicts the results. Several observations are worth emphasizing.

#### Remarks:

- 1) First, the outage curves for both unicasting and multicasting approach zero with our cooperative protocol, which is a consequence of the transmission rate being below capacity. Note that, by contrast, for cooperation-free admissible protocols, the outage curves will not decay with network size.
- 2) Multicasting incurs significant penalty over unicasting in terms of outage probability for a fixed network size  $K$ . In particular, Fig. 2 confirms that the multicasting outage probability is indeed roughly a factor  $K$  larger than the unicasting outage probability.
- 3) The slopes of the outage log-probability curves are asymptotically constant, and the bounds are good predictors of the asymptotic slopes. This is perhaps not surprising since we used Chernoff techniques to derive the bounds. Indeed, in many communication problems the Chernoff exponent is close to the correct exponent. However, the bounds are not particularly close to the the correct outage curves.
- 4) The analytical outage probability approximations are asymptotically quite close to the true curves, converging to within a factor of roughly 3 in probability for large network sizes. In addition, these approximations appear to be actual upper bounds at least in case study depicted, though this is a conjecture.
- 5) The asymptotic slopes of the outage log-probability curves for both unicasting and multicasting are identical. In Section VII, we will develop this slope as the network scaling exponent of the protocol, which we denote using  $E_{\text{ns}}^-$ . For a target outage level, this slope can be used to quantify the asymptotic network size gap between unicasting and multicasting. In particular, suppose that for a fixed choice of  $\alpha$  and  $\beta$  in the protocol,  $K^{\text{uc}}(\epsilon)$

nodes are required to achieve some target outage probability  $\epsilon$  in unicasting. Then the number of nodes required to achieve the same outage probability in multicasting is, asymptotically

$$K^{\text{mc}}(\epsilon) = K^{\text{uc}}(\epsilon) + \frac{1}{E_{\text{ns}}^-} \log K^{\text{uc}}(\epsilon). \quad (42)$$

To verify (42), it suffices to recognize that the vertical distance between the unicasting and multicasting outage probabilities is, in accordance with (41), asymptotically,  $\log K^{\text{uc}}(\epsilon)$ .

## VII. NETWORK SCALING EXPONENT

In this section, we explore, in more detail, the asymptotic rate of decay of the outage probability with network size, which we have termed the network scaling exponent. This exponent captures meaningful information for system designers. For example, at the transmission rates to which Fig. 2 corresponds, outage probabilities for our two-phase protocol decay reasonably quickly in a practical sense—i.e., the network scaling exponent is reasonably large. However, as we will see, at rates close to capacity, it turns out that outage probabilities decay very slowly as a function of network size, corresponding to a small network scaling exponent. This implies that very large network sizes may be needed to achieve practical target error rates.

Before beginning our development, note that the network scaling exponent is the natural counterpart to the classical error exponent for traditional channel codes. In particular, the classical error exponent captures the exponential rate of decay of error probability with block length as a function of the targeted fraction of capacity; see, e.g., [6]. Analogously, the network error exponent captures the exponential rate of decay of error probability in unicasting and multicasting with network size as a function of the targeted fraction of capacity.

Formal definitions follow.

*Definition 4:* The network reliability function with respect to a sequence of admissible protocols  $\pi_K$  in Definition 1 is given by

$$E_{\text{ns}}^-({\pi}_K) = - \lim_{K \rightarrow \infty} \frac{\ln \Pr\{\mathcal{E}_{\pi_K}\}}{K} \quad (43)$$

where  $\mathcal{E}_{\pi_K}$  denotes the outage event for a system with  $K$  nodes under the protocol  $\pi_K$ .

*Definition 5:* The network scaling exponent is the supremum of the network reliability functions of all sequences of admissible protocols with a rate that is at least a fraction  $r$  of the capacity at a given SNR, i.e.,

$$E_{\text{ns}}(r, \text{SNR}) = \sup_{\{\pi_K\} \in \mathcal{P}(r, \text{SNR})} E_{\text{ns}}^-({\pi}_K) \quad (44)$$

where  $\mathcal{P}(r, \text{SNR})$  is a set of sequences of admissible protocols with a rate that is a fraction  $r$  of the capacity.

The following proposition establishes that, as with capacity, unicasting and multicasting are not distinguished by their network scaling exponents.

*Proposition 3:* The network scaling exponent is the same for both unicasting and multicasting.

*Proof:* First, for any sequence of admissible protocols

$$\Pr\{\mathcal{E}_{\pi_K}^{\text{mc}}\} \geq \Pr\{\mathcal{E}_{\pi_K}^{\text{uc}}\}$$

so that

$$E_{\text{ns}}^{\text{mc}}(r, \text{SNR}) \leq E_{\text{ns}}^{\text{uc}}(r, \text{SNR}). \quad (45)$$

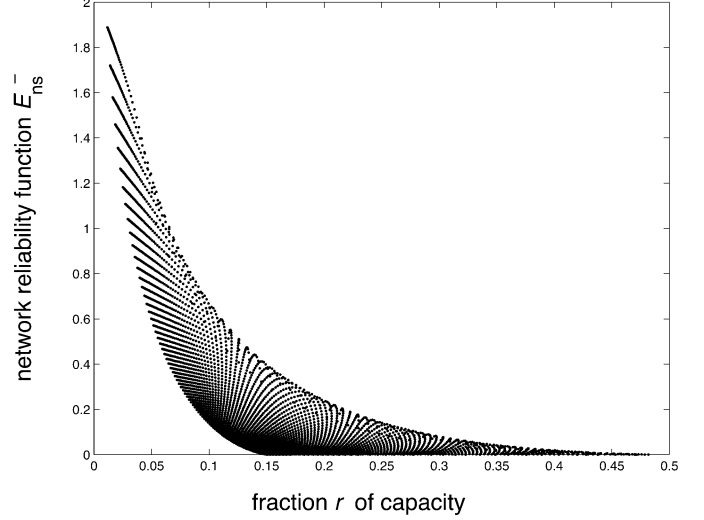


Fig. 3. The upper envelope of the plotted points indicates the network reliability function for the two-phase cooperative protocol as a function of the targeted fraction of capacity  $r$ . Each point corresponds to a particular value of  $\alpha$  and  $\beta$ . In this example, SNR = 0 dB.

Furthermore, if  $\{\pi_K^*\}$  achieves the supremum for the unicasting system, then from a simple application of the union bound it follows that, for each  $K$

$$\Pr\{\mathcal{E}_{\pi_K^*}^{\text{mc}}\} \leq K \Pr\{\mathcal{E}_{\pi_K^*}^{\text{uc}}\}$$

and hence,

$$E_{\text{ns}}^{\text{mc}}(r, \text{SNR}) \geq E_{\text{ns}}^{-, \text{mc}}(\{\pi_K^*\}) = E_{\text{ns}}^{\text{uc}}(r, \text{SNR}). \quad (46)$$

Combining (45) and (46) we obtain  $E_{\text{ns}}^{\text{mc}}(r, \text{SNR}) = E_{\text{ns}}^{\text{uc}}(r, \text{SNR})$  as desired.  $\square$

In the remainder of this section, we analyze a lower bound on the network scaling exponent by optimizing over the class of the two-phase protocols described in Section V. For a fixed choice of  $\alpha$  and  $\beta$ , we can express the fraction of the capacity achieved by the protocol as

$$r(\alpha, \beta, \text{SNR}) = \frac{R(\alpha, \beta, \text{SNR})}{C(\text{SNR})} \quad (47)$$

where we have made the dependency of both  $R$  and  $C$  in (14) and (3), on the parameters of interest explicit. We define the network reliability function of the user cooperation protocol in Section V as

$$E_{\text{ns}}^-(r, \text{SNR}) = \sup_{\alpha, \beta: r(\alpha, \beta, \text{SNR}) \leq r} \left\{ - \lim_{K \rightarrow \infty} \frac{\ln \Pr\{\mathcal{E}_{\alpha, \beta, K}\}}{K} \right\} \quad (48)$$

which constitutes a lower bound on  $E_{\text{ns}}(r, \text{SNR})$  in (44). Note that in the above definition, we have constrained  $\alpha$  and  $\beta$  to be constants independent of  $K$ .

The upper envelope of the points in Fig. 3 indicates the network reliability function of our two-phase protocol. Each point in the plot corresponds to a particular choice of  $\alpha$  and  $\beta$  in the protocol, for which we have numerically evaluated  $E_{\text{ns}}^-$  in (48) for different values of  $r$  at SNR = 0 dB.

Perhaps the most striking observation from Fig. 3 is that the error exponent for the two-phase protocol is quite small when aiming for rates that are more than about half of capacity. This implies that while the protocol is capacity achieving, it may require a prohibitively large number of nodes to achieve rates anywhere close to this capacity. It remains to be determined whether there exist more sophisticated protocols with substantially higher exponents in this regime.

As a final comment, it should also be noted that Fig. 3 effectively characterizes the efficient operating frontier for the protocol. In particular, given a network with  $K$  nodes and an allowable outage probability  $\epsilon$ , one can approximate the reliability function by  $-\ln \epsilon/K$  and determine the corresponding value of  $r$ , which is an estimate of how close one can expect to get to capacity in the system.

### VIII. CONCLUDING REMARKS

Perhaps the main contribution of this correspondence is a framework for analyzing user cooperation protocols in the limit of large network size (number of nodes), which we have illustrated in the case of a multipath-rich Rayleigh-fading environment. Within this framework, we have introduced a meaningful notion of Shannon capacity for this regime and presented a simple two-phase protocol that can achieve rates arbitrarily close to capacity. A finer grain analysis of this two-phase protocol in terms of its network scaling exponent, which characterizes the rate of decay of error probability with network size, shows that it may require prohibitively large number of nodes to achieve rates close to the capacity with this protocol.

One important direction of future work is to study more sophisticated models beyond the Rayleigh-fading model within our framework. One could, for example, incorporate the effects of network geometry and shadowing into the model. More generally, it would be of interest to study a class of channel models for which user cooperation plays a fundamental role in enabling reliable communication in multicasting. The Rayleigh-fading model considered here clearly belongs to this class, but we believe the class may be quite rich and may include many other models of practical importance.

Another important direction is to investigate how system performance changes when the sum power constraint is replaced with individual power constraints. With individual power constraints, the system capacity will increase with the number of nodes—in fact, the MISO upper bound increases according to  $\Theta(\log K)$ . It remains to be determined whether there exist cooperative multicasting protocols that approach this upper bound or whether one can develop tighter upper bounds for this scenario.

Finally, as noted in Section VII, the two-phase protocol may require prohibitively large number of nodes to achieve rates close to the capacity. It remains to investigate whether more sophisticated protocols can improve the network scaling exponent substantially in this regime.

### APPENDIX

#### DERIVATION OF OUTAGE APPROXIMATION (38)

First, we write  $\Pr\{\mathcal{E}_{\alpha,\beta,K}^{\text{uc}}\}$  in the form

$$\begin{aligned} \Pr\{\mathcal{E}_{\alpha,\beta,K}^{\text{uc}}\} &= E\left[\Pr\{\mathcal{E}_{\alpha,\beta,K}^{\text{uc}}|K_1\}\right] \\ &= \sum_{k_1=1}^K \Pr\{K_1 = k_1\} \Pr\{\mathcal{E}_{\alpha,\beta,K}^{\text{uc}} | K_1 = k_1\} \\ &\approx \sum_{k_1=\alpha(1-\beta)K}^K \Pr\{K_1 = k_1\} \Pr\{\mathcal{E}_{\alpha,\beta,K}^{\text{uc}} | K_1 = k_1\}. \end{aligned} \quad (49)$$

Note that we have dropped the contribution of terms with  $k_1 \leq \alpha(1-\beta)K$  in the summation, since we expect their aggregate sum to be small as they deviate significantly from the mean  $E[K_1] = \alpha K$ .

We now approximate each of the two factors in (49). The right factor we approximate by the upper bound (19). The left factor we replace

with via Stirling's approximation for binomial distributions [5, p. 284], yielding

$$\begin{aligned} \Pr\{K_1 = k_1\} &= \binom{K}{k_1} \alpha^{k_1} (1-\alpha)^{(K-k_1)} \\ &\approx \frac{1}{\sqrt{K}} \exp\left\{-KD\left(\frac{k_1}{K} \parallel \alpha\right)\right\} \end{aligned} \quad (50)$$

where  $D(\cdot|\cdot)$  denotes the binary relative entropy function, i.e., for any  $0 < p, q < 1$

$$D(p||q) \triangleq p \ln \frac{p}{q} + (1-p) \ln \frac{1-p}{1-q} \quad (51)$$

and where  $\alpha$  is the parameter of  $K_1$  (cf. (16)).

Thus, substituting (19) and (50) into (49) yields

$$\begin{aligned} \Pr\{\mathcal{E}_{\alpha,\beta,K}^{\text{uc}}\} &\approx \sum_{k_1=\alpha(1-\beta)K}^K \frac{1}{\sqrt{K}} \exp\left\{-KD\left(\frac{k_1}{K} \parallel \alpha\right)\right\} \\ &\times \left(\frac{\alpha K(1-\beta)e}{k_1}\right)^{k_1} \exp\{-\alpha K(1-\beta)\}. \end{aligned} \quad (52)$$

Finally, we approximate (52) by an approximation to the largest single term in the summation, viz.,

$$\begin{aligned} \Pr\{\mathcal{E}_{\alpha,\beta,K}^{\text{uc}}\} &\approx \frac{1}{\sqrt{K}} \exp\{-\alpha K(1-\beta)\} \\ &\times \max_{\gamma \in (\alpha(1-\beta), 1)} \exp\left\{-KD(\gamma||\alpha) + \gamma K \ln\left(\frac{\alpha(1-\beta)e}{\gamma}\right)\right\}. \end{aligned} \quad (53)$$

Since the term in the exponent being minimized in (53) is differentiable and convex in  $\gamma$ , the optimizing  $\gamma$  is the value at which the associated derivative is zero, i.e.,

$$\frac{\gamma^2}{1-\gamma} = \mu \quad (54)$$

where  $\mu$  is as given in (40). Finally, it is straightforward to verify that (54) has a solution in  $(\alpha(1-\beta), 1)$  and it may be solved explicitly, yielding (39).

### ACKNOWLEDGMENT

The authors wish to thank the anonymous reviewers for several observations which helped to improve the quality of the manuscript.

### REFERENCES

- [1] K. Azarian, H. El 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.
- [2] A. Bletsas, A. Khisti, D. Reed, and A. Lippman, "A simple cooperative diversity method based on network path selection," *IEEE J. Sel. Areas Commun.*, to be published.
- [3] H. Boche and E. Jorswieck, "Outage probability in multiple antenna systems," *Europ. Trans. Telecommun.*, to be published.
- [4] J. Canny. Combinatorics and Discrete Probability (CS174) Course Notes. Univ. Calif., Berkeley. [Online]. Available: <http://www.cs.berkeley.edu/jfc/cs174lects/lec9/lec9.html>
- [5] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.
- [6] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.
- [7] B. M. Hochwald, T. L. Marzetta, and V. Tarokh, "Multiple-antenna channel hardening and its implications for rate feedback and scheduling," *IEEE Trans. Inf. Theory*, vol. 50, no. 9, pp. 1893–1909, Sep. 2004.



- [8] 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. 3062–80, Dec. 2004.
- [9] J. N. Laneman and G. W. Wornell, "Distributed space-time coded protocols for exploiting cooperative diversity in wireless networks," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2415–2525, Oct. 2003.
- [10] R. U. Nabar, H. Bölcskei, and F. W. Kneubühler, "Fading relay channels: Performance limits and space-time signal design," *IEEE J. Sel. Areas Commun.*, vol. 22, no. 6, pp. 1099–1109, Aug. 2004.
- [11] L. Ozarow, S. Shamai (Shitz), and A. Wyner, "Information theoretic considerations for cellular mobile radio," *IEEE Trans. Veh. Technol.*, vol. 43, no. 2, pp. 359–738, May 1994.
- [12] N. Prasad and M. K. Varanasi, "Diversity and multiplexing tradeoff bounds for cooperative diversity protocols," in *Proc. Int. Symp. Information Theory*, Chicago, IL, Jun./Jul. 2004, p. 268.
- [13] A. Sendonaris, E. Erkip, and B. Aazhang, "User cooperation diversity—Part I: System description," *IEEE Trans. Commun.*, vol. 51, no. 11, pp. 1927–38, Nov. 2003.
- [14] Í. E. Teletar, "Capacity of multi-antenna Gaussian channels," *Europ. Trans. Telecommun.*, vol. 10, pp. 585–596, Nov./Dec. 1999.
- [15] D. N. C. Tse, P. Viswanath, and L. Zheng, "Diversity–multiplexing tradeoff in multiple access channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 9, pp. 1859–1874, Sep. 2004.
- [16] L. Zheng and D. N. C. Tse, "Diversity and multiplexing: A fundamental tradeoff in multiple antenna channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 5, pp. 1073–1096, May 2003.

## On Outer Bounds to the Capacity Region of Wireless Networks

Sahand Haji Ali Ahmad, Aleksandar Jovičić, *Student Member, IEEE*, and Pramod Viswanath, *Member, IEEE*

**Abstract**—In this correspondence, we study the capacity region of a general wireless network by deriving fundamental upper bounds on a class of linear functionals of the rate tuples at which joint reliable communication can take place. The widely studied transport capacity is a *specific* linear functional: the coefficient of the rate between a pair of nodes is equal to the Euclidean distance between them. The upper bound on the linear functionals of the capacity region is used to derive upper bounds to scaling laws for *generalized* transport capacity: the coefficient of the rate between a pair of nodes is equal to some arbitrary *function* of the Euclidean distance between them, for a class of minimum distance networks. This upper bound to the scaling law meets that achievable by multihop communication over these networks for a wide class of channel conditions; this shows the optimality, in the scaling-law sense, of multihop communication when studying generalized transport capacity of wireless networks.

**Index Terms**—Ad hoc wireless networks, capacity region, cut-set bounds, isometric embedding, multihop, transport capacity.

### I. INTRODUCTION

A characterization of the capacity region of wireless networks is one of the long standing open problems in information theory. Progress in

Manuscript received February 12, 2005; revised April 25, 2005. This work was supported in part by the National Science Foundation by Grant CCR-0312413 and by a Grant from Motorola, Inc., as part of the Motorola Center for Communication.

The authors are with the Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801 USA (e-mail: alihaji@uiuc.edu; jovicicuiuc.edu; pramodv@uiuc.edu).

Communicated by R. W. Yeung, Guest Editor.

Digital Object Identifier 10.1109/TIT.2006.874533

this direction has recently come by relaxing the questions asked: focus has been on the *transport capacity* [3], a specific linear functional of rate tuples at which jointly reliable communication is possible. The coefficient of the rate between a pair of nodes in this linear functional is equal to the Euclidean distance between the two nodes. Further, much emphasis is on the scaling law of the transport capacity as a function of the number of nodes in the network. Recent results have shown that multihop communication is scaling-law optimal for a class of minimum distance networks provided the wireless channel decays fast enough [4], [7], [8]. A study of the sum capacity of the wireless network appears in [5].

In an effort to understand the entire capacity region of the wireless network, we shift the focus from transport or sum capacity to studying arbitrary linear functionals of the rate tuples in the capacity region. The capacity region is a compact convex set. Thus, understanding the extremum of linear functionals of the rate tuples in the capacity region is an alternative but equivalent characterization (the *convex dual* in the language of convex analysis; see [6]) of the capacity region. In this correspondence, we derive simple and robust upper bounds to a class of linear functionals: the upper bound simply depends on the distance between the nodes, channel attenuation conditions and the coefficients of the linear functional. The key technique is a combination of the results of isometric embeddability of an arbitrary metric space into the  $l_1$  metric space and the familiar information-theoretic cut-set upper bounds to the capacity region.

We use this simple upper bound to exactly characterize the scaling law behavior over arbitrary linear and planar networks of *generalized* transport capacity

$$C_f \stackrel{\text{def}}{=} \max_{(R_{ij}) \text{ in the capacity region}} \sum_{i,j=1}^n R_{ij} f(r_{ij}) \quad (1)$$

where  $R_{ij}$  and  $r_{ij}$  denote, respectively, the rate of reliable communication and the Euclidean distance between a pair of nodes  $(i, j)$ . When  $f(\cdot)$  is the identity function, i.e.,  $f(x) = x$ , we get back the transport capacity. The scaling law of generalized transport capacity depends crucially on the long range behavior of the function  $f(\cdot)$ ; of particular interest is the parameter

$$\alpha_f \stackrel{\text{def}}{=} \overline{\lim}_{x \rightarrow \infty} \left( \frac{\log f(x)}{\log x} \right). \quad (2)$$

To be able to state the scaling law precisely, we briefly describe our wireless channel model adopted from [4]: the received signal at node  $j$  at time  $m$  is

$$y_j[m] = \sum_{i \neq j} \frac{h_{ij}^s[m]}{(1+r_{ij})^\delta} x_i[m] + z_j[m]. \quad (3)$$

Here  $x_i[m]$  is the signal transmitted by node  $i$  at time  $m$  and  $z_j[m]$  is i.i.d. white Gaussian noise. Each node  $j$  has an average transmit power constraint (denoted by  $P_j$ ). The fading channel  $\{h_{ij}^s[m]\}_m$  is a stationary and ergodic zero-mean stochastic process that is independent for each pair of nodes  $(i, j)$  and models the small scale fluctuations of frequency flat fading. For simplicity, we assume that  $\mathbb{E}[|h_{ij}^s[m]|^2] = 1$  for all  $i, j, m$ . The large scale variations are modeled explicitly<sup>1</sup> through the decay of signal level: a factor of  $\frac{1}{(1+r_{ij})^\delta}$  from node  $i$  to node  $j$ . In free space  $\delta = 1$ , with a single reflected path along with the

<sup>1</sup>The *far field* signal decay is usually denoted by  $r_{ij}^{-\delta}$ . Here, we have written  $(1+r_{ij})^{-\delta}$  to ensure that our model makes sense when nodes get close; i.e., the average received power is not more than the average transmit power.