Hierarchical Wireless Networks: Capacity Bounds using the Constrained Multiple-Access Relay Channel Model

Invited Paper

Lalitha Sankaranarayanan*
WINLAB, Department of ECE
Rutgers University
Piscataway, NJ 08854
Email: lalitha@winlab.rutgers.edu

Gerhard Kramer†
Bell Laboratories
Lucent Technologies
Murray Hill, NJ 07974
Email: gkr@bell-labs.com

Narayan B. Mandayam*
WINLAB, Department of ECE
Rutgers University
Piscataway, NJ 08854
Email: narayan@winlab.rutgers.edu

***Abstract*— A wireless ad hoc network is modeled as a three-tier hierarchical network comprised of low-power source nodes that communicate with a wired access-point via an intermediate relay node. Such a model is appropriate for hybrid ad hoc networks where cooperation between the sources may not be possible or desirable. It is shown that constraining the relay to transmit and receive in either different time slots or frequency bands for a fixed total time and bandwidth yields the same capacity bounds. The rates achieved by cooperative strategies are compared with the rates achieved by traditional multi-hopping and are shown to be substantially better than for certain channels and geometries.**

I. Introduction

The demand for ubiquitous communications is driving the development of a variety of wireless devices and technologies that facilitate ad hoc communications. Such devices, under different size and processing constraints, can form a network of sensor nodes that monitor events and collect data or share bandwidth and energy resources to facilitate communication with each other or a backbone network. The challenge lies in designing such networks to ensure an efficient use of the limited bandwidth and power resources.

In their seminal work on the throughput of wireless ad hoc networks [1], Gupta and Kumar showed that for a network of n homogeneous nodes that cooperate to forward data, the throughput per node falls asymptotically with increasing number of nodes. This decrease in throughput is a direct result of interference and bandwidth restrictions where each node allocates some of its throughput to forwarding packets for neighboring nodes. They also showed that introducing wireless relay nodes does not change the scaling properties. In [2], Liu et al limit cooperation between source nodes by introducing a regular network of base stations (access points) connected by a high-bandwidth wired network within the ad hoc network of n source nodes. They show that in such a *hybrid* ad hoc network where each source node transmits at W bits/sec, a scaling in network throughput capacity as $\Theta(mW)$ can be achieved when the number of base stations, m, scales faster than $\sqrt{n}$ thus requiring a significant investment in the wired infrastructure.

Consider a three-tier hierarchical network that results from the introduction of wireless relay nodes serving exclusively as forwarders in a hybrid wireless network of sources and base stations described above. For a network where the source nodes have a one-hop link to the nearest relay, and forwarding (cooperation) is limited to the relays, determining the relay density and bandwidth allocation that minimizes the number of access points while preserving throughput is an open problem. While relays may not reverse the scaling behavior, it is possible that they reduce the required number of wired access points and also lower the power consumption of the source nodes, both valuable resources in an ad hoc network.

In addition to these general and theoretic networking issues, there are operational advantages to hierarchical heterogeneous layering that cannot be achieved with a “flat” homogeneous network. For example, dedicating select nodes with appropriate power and processing capabilities as relays preserves the limited battery resources of source nodes and facilitates scalable routing [3]. The relay layer also helps eliminate complex economic or social incentives needed to encourage cooperation in general ad hoc networks. Thus, for a variety of applications, a relatively small number of higher-level network elements with access to more power and better processing capabilities could greatly improve the performance of the overall system in terms of reliability, longevity, and flexibility.

In [4], capacity bounds and cooperative strategies for a *simple* hierarchical ad hoc network formed by a cluster of nodes that communicate with an access point via a relay node are presented. The network is modeled as a multiple-access relay channel (MARC) with Gaussian noise and fading. Two modes of relay operation that result from placing constraints on its

*Supported in part by the NSF under grant no. ITR-0205362.
†Supported in part by the NSF under grant no. CCR-0325673.

simultaneous transmit-receive capabilities are considered. In this paper, we show that the case of constraining the relay to transmit and receive in different time slots is equivalent to using orthogonal frequency bands under identical symbol time and bandwidth constraints. We also compare the performance of traditional multi-hopping with the cooperative strategies presented in [4]. The paper is organized as follows. In Section II, we briefly discuss the model and the cooperative strategies. In Section III, the rate bounds resulting from constraining the relay in time and frequency are compared. In Section IV, we consider two example geometries for a two-source network and compare the performance of different communication strategies.

II. Gaussian MARC: Model and Strategies

A. Model

A model for an M-source Gaussian MARC consists of $M+1$ inputs signals X_{ki}, $k=1,2,\ldots,M+1$ from the sources and the relay node, and two output signals $Y_{M+1,i}$ and $Y_{M+2,i}$ at the relay and destination, respectively, where i is a time index [4]. The channel is used n times and the received signals at terminals $M+1$ and $M+2$ are

$$Y_{M+1,i} = \left(\sum_{k=1}^{M} h_{M+1,ki} X_{ki} \right) + Z_{M+1,i} \tag{1}$$

$$Y_{M+2,i} = \left(\sum_{k=1}^{M+1} h_{M+2,ki} X_{ki} \right) + Z_{M+2,i} \tag{2}$$

where Z_{ji}, $j \in \{M+1, M+2\}$ is proper (circularly symmetric) complex Gaussian noise with zero-mean and unit variance. The transmitted signals from the k^{th} source and the relay are constrained in power as

$$\sum_{i=1}^{n} E(|X_{ki}|^2) \Big/ n \leq P_k \quad k=1,2,\ldots,M+1 \tag{3}$$

The parameter h_{jki} is the fading experienced by the signal from the k^{th} transmitter at the j^{th} receiver in the i^{th} symbol and is assumed known only at the j^{th} receiver. In this analysis, analogous to [4], we consider two kinds of fading channels:

1) constant no fading $h_{jki} = 1 \Big/ \sqrt{d_{jk}^{\gamma}}$ for all $i \in [1, n]$ where d_{jk} is the distance between the j^{th} receiver and the k^{th} source and γ is the path-loss exponent.
2) ergodic phase-fading with $h_{jki} = e^{j\theta_{jki}} \Big/ \sqrt{d_{jk}^{\gamma}}$ where θ_{jki} is a uniformly distributed random variable between $[-\pi, \pi]$.

The analysis for these models generalizes to other types of fading such as Rayleigh fading [5].

The above model permits the relay to transmit and receive simultaneously. In general, however, physical and practical constraints limit the relay to either transmit or receive, thereby resulting in a *constrained-MARC* (C-MARC) (see also [6]). In [4], the C-MARC was defined as the MARC of (1)−(3) with the constraints

$$Y_{M+1} = 0 \ \text{ if } X_{M+1} \neq 0 \tag{4}$$

and that $X_{M+1} = 0$ for a fraction α of the total time. The advantage of this approach is that one can apply the theory developed for the MARC directly to the C-MARC. This definition, however, constrains the relay to employ time division duplexing (TDD) between its transmit and receive states.

Alternatively, the relay can transmit and receive at the same time but in non-overlapping frequency bands. Thus, the relay employs frequency division duplexing (FDD) by receiving in a fraction α of the total bandwidth W and transmitting in the fraction $(1-\alpha)W$. The sources and destination, in general, use both bands to transmit and receive respectively. Defining $X_{mi}^{(\alpha)}$ and $X_{mi}^{(1-\alpha)}$ as the m^{th} transmitted signals in the bands of bandwidth αW and $(1-\alpha)\,W$ respectively, we write $X_{mi} = (X_{mi}^{(\alpha)}, X_{mi}^{(1-\alpha)})$ for all i and $m \in [1, M+1]$ with $X_{M+1,i}^{(\alpha)} = 0$. We call the former frequency band the α band and the latter the $(1-\alpha)$ band. The power constraint (3) is applied to the FDD model in the same way as the TDD model. The received signals are

$$Y_{M+1,i} = (Y_{M+1,i}^{(\alpha)}, 0) \tag{5}$$

$$Y_{M+2,i} = (Y_{M+2,i}^{(\alpha)}, Y_{M+2,i}^{(1-\alpha)}) \tag{6}$$

where at the j^{th} receiver, $j \in [M+1, M+2]$, $Y_{ji}^{(\alpha)}$ and $Y_{ji}^{(1-\alpha)}$ in the α and $(1-\alpha)$ bands are

$$Y_{ji}^{(\alpha)} = \left(\sum_{k=1}^{j-1} h_{M+1,ki} X_{ki}^{(\alpha)} \right) + Z_{ji}^{(\alpha)} \tag{7}$$

$$Y_{ji}^{(1-\alpha)} = \left(\sum_{k=1}^{j-1} h_{M+1,ki} X_{ki}^{(1-\alpha)} \right) + Z_{ji}^{(1-\alpha)} \tag{8}$$

The variables $Z_{ji}^{(\alpha)}$ and $Z_{ji}^{(1-\alpha)}$ are zero-mean proper complex Gaussian noise variables in the corresponding bands with variance α and $1-\alpha$ respectively.

We remark that, in general, any node in a hierarchical network can sleep, transmit or receive for some fraction of the total time with constraints on the average power in each mode and over all modes as in (3) [7]. Thus, the capacity analysis of constrained relay channels should, in general, take into account the positive probabilities (fraction of time) of each node being in one of three modes, *sleep*, *listen*, or *talk* (*SLoT*), at any time [7]. For the C-MARC considered here, we assume that the probability that the sources listen or the destination transmits is zero. In the next section we show that a relay network with time-duplexed SLoT constraints is equivalent to a relay network with frequency duplexed nodes and the same SLoT constraints for a fixed total bandwidth, symbol time, and average power.

B. Bounds and Cooperative Strategies

A rate-tuple $(R_1, R_2, \ldots, R_M)$ is said to be *achievable* if there are encoders and decoders such that the probability that the destination node makes an error in decoding any of the M messages is less than ϵ for all positive ϵ. The *capacity region* is the closure of the set of achievable rate tuples. Outer bounds on the capacity region of the MARC and C-MARC can be obtained in a manner similar to the well-known cut-set bounds for networks and are presented in [4]. Inner bounds are obtained by constructing codes and computing their achievable rates.

In [4], coding strategies for the classic single-source relay channel [8] are extended to obtain various strategies for the MARC and C-MARC. The cooperative strategy of [8, theorem 1] is generalized as the *decode-and-forward* (DF) strategy where the relay decodes the source messages before forwarding them to the destination. For the C-MARC, an additional strategy of *partial decode-and-forward* (P-DF) results when the relay is limited to decoding only one of the two message streams from each source while the destination decodes both over both fractions. The *compress-and-forward* (CF) strategy extends the strategy of [8, theorem 6] where the relay facilitates reliable detection at the destination by forwarding a quantized version of its received signal to the destination while the *amplify-and-forward* (AF) strategy considers a simple relay that forwards an amplified version of its received signal to the destination.

III. Time vs. Frequency Duplexing

For a Gaussian C-MARC, outer bounds and achievable strategies are presented in [4] for a TDD relay. We now show that for the strategies considered in [4] and the assumption that the channel state information (CSI) is unknown at the transmitters, the same rate bounds are obtained if the relay employs FDD. We fix the bandwidth W and symbol time T and assume that the sources and destination use all available time and bandwidth to transmit and receive, respectively. Thus, for TDD the sources and relay share the same bandwidth while for FDD the sources and relay transmit at the same time. The transmitter nodes are subject to the same power constraints in both cases.

Consider the DF strategy. For a fixed W and T, the rate bounds at the relay in the fraction (time or frequency) α are

$$\sum_{m\in G} R_m \leq \alpha I(\mathbf{X}_{(G)}; Y_{M+1}|\mathbf{X}_{(G^c)}) \tag{9}$$

where $X_{M+1}^{(\alpha)} = 0$ and $G \subseteq \{1, 2, \ldots, M\}$ such that $\mathbf{X}_{(G)} = \{X_m : m \in G\}$. The destination decodes the source messages using the received symbols from the sources in both fractions and from the relay in the $1-\alpha$ fraction for both TDD and FDD. The resulting bounds are

$$\sum_{m\in G} R_m \leq \begin{pmatrix} \alpha I(\mathbf{X}_{(G)}; Y_{M+2}|\mathbf{X}_{(G^c)}) + \\ (1-\alpha)\, I(\mathbf{X}_{(G)}, X_{M+1}; Y_{M+2}|\mathbf{X}_{(G^c)}) \end{pmatrix} \tag{10}$$

For the case where the source and relay operating modes are known at all transmitters and receivers, Gaussian signalling maximizes the rates in (9) and (10) for Gaussian channels [9].

For the C-MARC, we assume that the m^{th} source, $m \in [1, M]$, allocates $P_m^{(\alpha)}$ and $P_m^{(1-\alpha)}$ as the average power in the fractions (time or frequency) α and $1-\alpha$ respectively, such that $\alpha P_m^{(\alpha)} + (1-\alpha)\, P_m^{(1-\alpha)} \leq P_m$. Then, the received SNR at the relay for any $G \subseteq S$ for the TDD case is $\sum_{m\in G} |h_{M+1,m}|^2 P_m^{(\alpha)}$ where, as defined in Section II, the noise power over the bandwidth W is unity. For the FDD case, the same SNR is achieved at the relay since now both the signal and noise power in the α band are scaled by α thus yielding the same rate bounds at the relay in both cases. We can similarly show that the source signals have the same SNR at the destination in each fraction for both TDD and FDD.

A time-duplexed relay can transmit at most $P_{M+1}/(1-\alpha)$ in its transmit fraction $1-\alpha$ subject to (3) resulting in a received SNR at the destination of $|h_{M+1,M+2}|^2 P_{M+1}/(1-\alpha)$. On the other hand, a frequency-duplexed relay transmits over all n time symbols, with average power P_{M+1} in each symbol, achieving a receive SNR $|h_{M+1,M+2}|^2 P_{M+1}/(1-\alpha)$ at the destination in the band $(1-\alpha)W$ where $(1-\alpha)$ is the noise power in that band. Thus, the SNR at the destination for the TDD and FDD case are the same for all nodes, resulting in the same achievable rates for the DF strategy for both cases.

The above argument also holds for the P-DF strategy and the CF strategy. In both cases, the constraint on bandwidth or time translates to scaling the mutual information by a factor α or $1-\alpha$ while cooperation, if any, between the sources and relay is achieved in the $1-\alpha$ fraction at the destination. We do not consider the AF strategy for a FDD relay when the sources also transmit in the $1-\alpha$ fraction as it does not lend itself to the same interpretation of a multiple-access inter-symbol interference (ISI) channel at the destination that it does for TDD. Finally, we remark that the analysis can be generalized to allow the source and destination nodes to have a positive probability of being in sleep mode. It can then similarly be shown that the C-MARC rate bounds obtained when all the nodes time-duplex between their three SLoT modes are the same as those obtained if all nodes used orthogonal frequency bands for the three modes.

IV. Wireless Examples

In [4], two example geometries are considered to illustrate the gains achieved by using a relay relative to direct transmission. We consider the same geometries here as shown in Fig.

1. While the two geometries chosen clearly illustrate capacity achieving strategies, they are also reflective of the typical performance achieved by the various strategies considered here for an arbitrary placement of source and relay nodes. Case 1 is a geometry with a symmetric positioning of the sources with respect to the relay and destination while case 2 is a collinear geometry with both sources at the origin and the destination a unit distance away from the origin. In both cases, the relay moves along the line connecting the destination with the origin. We compare the cooperative strategies described in Section II with traditional multi-hop routing, where messages from the sources are forwarded to the destination in two hops, the first from the source to the relay in the fraction α and the second from the relay to destination in the fraction $1-\alpha$. Without any loss of generality, we assume that all transmit nodes employ TDD for this analysis.

In multi-hop routing, the sources do not transmit when the relay forwards thereby conserving power. To make a fair comparison, we apply the same restriction on the C-MARC by imposing a sleep state on the source nodes in the $1-\alpha$ time fraction to obtain a Gaussian orthogonal multiple-access relay channel. We also make appropriate modifications to the resulting rate bounds for the DF, P-DF, CF, and AF rate bounds in [4]. A direct consequence of this restriction is that the DF and P-DF strategies simplify to the same strategy that we shall henceforth refer to as DF. Further, as is common in multi-hop routing, we half-duplex the transmitters by setting $\alpha = 1/2$. We remark that for this SLoT choice the sources and relay transmit in orthogonal channels, and there now exists an equivalent frequency-duplexed representation for the AF strategy with the same rates achieved by both TDD and FDD under the constraints discussed in Section III.

We present and analyze the results separately for the two fading models we consider, namely, the constant non-fading and the ergodic fast fading channel. For the following analysis the path loss exponent γ is chosen as 4 and all logarithms are evaluated with respect to base 2 so that the resulting rates are in units of bits per channel use.

A. No Fading

The sum-rate for the DF, AF, CF, and multi-hop (MH) strategies, in addition to the outer bounds (OB) are plotted as a function of the relay's position in Figs. 2 and 3 for cases 1 and 2 respectively and $P_1 = P_2 = P_3 = 2$. The rate bounds for the DF, CF, AF, and OB are obtained by setting $X_m^{(1-\alpha)} = 0$ for $m = 1,2$ in the C-MARC bounds presented in [4]. The rate bounds for the MH strategy are given by the set of rate pairs (R_1, R_2) that satisfy, for all $G \subseteq \{1,2\}$,

$$\sum_{m\in G} R_m \leq \min\left\{\alpha I(\mathbf{X}_{(G)};Y_3|\mathbf{X}_{(G^c)}),(1-\alpha)\,I(X_3;Y_4)\right\} \tag{11}$$

for an input distribution $\alpha p(x_1)p(x_2) + (1-\alpha)p(x_3)$ that is chosen Gaussian to maximize rates over a white Gaussian fading channel. Finally, the disadvantage of half-duplexing is

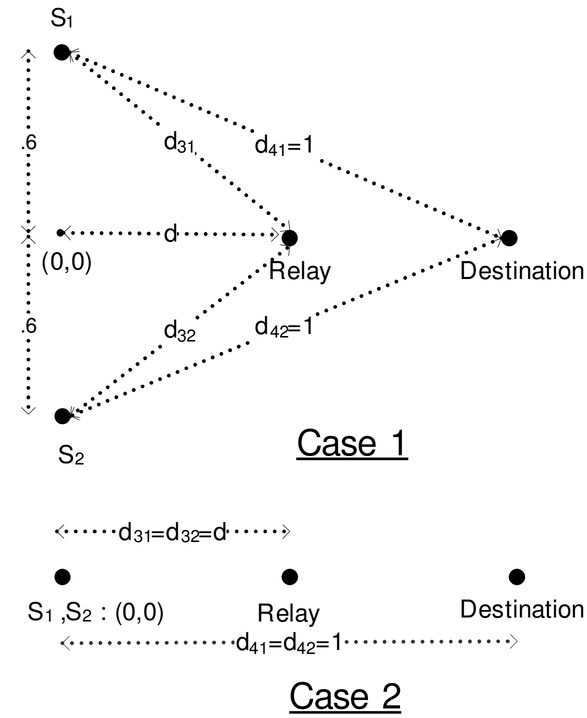

Fig. 1. Two geometries for a two-sensor MARC

clearly revealed by plotting the maximum achievable rates over all α for DF and MH as well as OB.

As shown in Fig. 2 and Fig. 3, the DF strategy meets the outer bounds when the relay is closer to the sources than it is to the destination. This *clustering* of sources and relay yields a high rate channel between the sources and relay resulting in the smaller rates at the destination, same for both DF and OB, determining the bounds. Clearly from the figures, the DF strategy demonstrates the usefulness of a relay that is not physically located between the sources and destination, as is possible in a network of randomly deployed nodes, provided the destination uses the received signals from both the sources and relay to decode. The poor performance of the MH strategy relative to the other strategies for relay positions not close to the destination is a direct consequence of the destination not using the received information from both fractions to decode. MH achieves the same rates as DF when the relay approaches the destination.

For relay positions closer to the destination, the CF and AF strategies approach the outer bounds. The CF and AF strategies result in a multiple-access system with two receive antennas at the destination, one receiving the direct signal and the other, a delayed noisy version of the signal received at the relay. The choice between AF and CF when the relay is close to the destination can be made based on energy and processing limitations of the relay and overall system latency requirements. The rates for the CF strategy result from

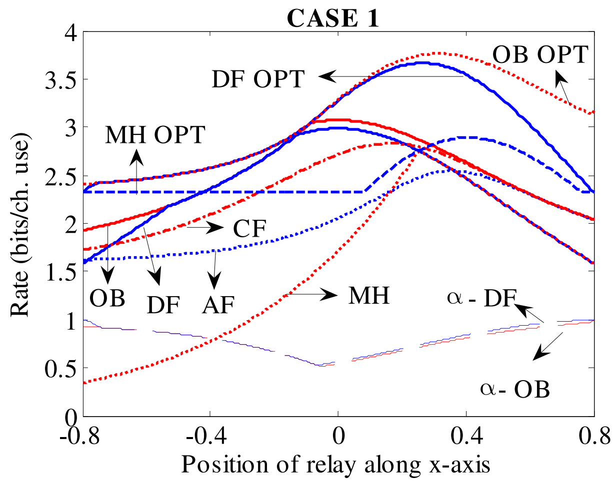

Fig. 2. Upper bound and achievable sum-rates for Case 1

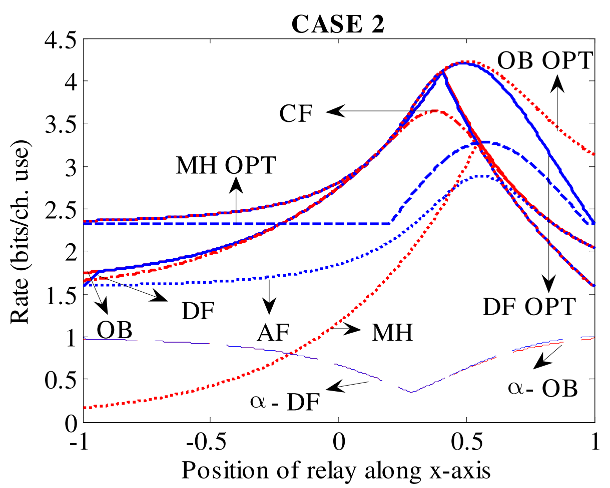

Fig. 3. Upper bound and achievable sum-rates for Case 2

the destination exploiting the correlation between its received signal, Y_4, and that at the relay, Y_3, in addition to decoding the relay's signal X_3 ([8, theorem 6]). If the destination did not exploit this correlation in decoding Y_3, it can be shown that for $\alpha = 1/2$ and half-duplexed sources, CF simplifies to the AF strategy.

We remark that the fraction α that maximizes the rates achievable for each strategy is, in general, not $1/2$ for any arbitrary network geometry. In [10], for $M = 1$, it is shown that capacity is achieved when α is chosen as the fraction α^*_{DF} that maximizes the DF strategy. In general, however, for any choice of source and relay positions, $\alpha^*_{DF} \neq \alpha^*_{OB}$ where α^*_{OB} maximizes the outer bounds. However, DF maximized over α, shown in Figs. 2 and 3 as DF OPT, achieves capacity when the sources and relay form a *cluster* such that $\alpha^*_{DF} = \alpha^*_{OB}$. The optimal OB curve is shown as OB OPT. Maximizing MH rates over α in (11) results in the sources transmitting full-duplex (FD) ($\alpha = 1$) when they are closer than the relay is to the destination. The advantage of multi-hopping relative to direct FD for relay positions between the sources and destination is revealed by the plot MH OPT. These plots also reveal the region where half-duplexing the sources and relay is advantageous relative to FD direct. Finally, allowing the source nodes to transmit cooperatively with the relay in the $1 - \alpha$ fraction for the same average power will, in general, increase the outer bounds and achievable rates for all strategies.

B. Ergodic Phase Fading

An ergodic phase fading model is appropriate for hierarchical ad hoc networks deployed over fast-changing terrain or in high-mobility environments. Similar to the no fading case, the DF and P-DF strategies achieve capacity for the ergodic channel when the sources and relay form a cluster [4]. For the present analysis where the sources sleep in the $1-\alpha$ fraction, the rates obtained for the ergodic phase-fading channel for the DF and MH strategies simplify to those obtained for the no fading case. The capacity achieving behavior of DF and DF OPT over the relay positions [-.525, -.13] and [-.3,.4] is evident in Figs. 2 and 3 for case 1 and 2 respectively.

Thus, given a choice in the placement of the relay, for the appropriate channel and transmitter power constraints, a cooperative strategy can be chosen to achieve the best rate from the strategies considered here.

References

[1] P. Gupta and P. R. Kumar, "The capacity of wireless networks," *IEEE Trans. Inform. Theory*, vol. 46, no. 2, pp. 388–404, Mar. 2000.
[2] B. Liu, Z. Liu, and D. Towsley, "On the capacity of hybrid wireless networks," in *Proc. 2003 IEEE Infocom*, San Francisco, CA, Apr. 2003.
[3] E. M. Belding-Royer, "Multi-level hierarchies for scalable ad hoc routing," *Wireless Networks*, vol. 9, no. 5, pp. 461–478, Sept. 2003.
[4] L. Sankaranarayanan, G. Kramer, and N. B. Mandayam, "Hierarchical sensor networks: Capacity theorems and cooperative strategies using the multiple-access relay channel model," in *First IEEE Conference on Sensor and Ad Hoc Communications and Networks*, Santa Clara, CA, Oct. 2004.
[5] G. Kramer, M. Gastpar, and P. Gupta, "Information-theoretic multi-hopping for relay networks," in *International Zurich Seminar on Communications*, ETH Zurich, Switzerland, Feb. 2004, pp. 192–195.
[6] M. A. Khojastepour, A. Sabharwal, and B. Aazhang, "On the capacity of 'cheap' relay networks," in *37th Annual Conf. Information Sciences and Systems*, Baltimore, MD, Mar. 2003.
[7] G. Kramer, "Models and theory for relay channels with receive constraints," in *42nd Annual Allerton Conf. on Commun., Control, and Computing*, Allerton, IL, Sept. 2004.
[8] T. Cover and A. E. Gamal, "Capacity theorems for the relay channel," *IEEE Trans. Inform. Theory*, vol. 25, no. 5, pp. 572–584, Sept. 1979.
[9] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.
[10] Y. Liang and V. Veeravalli, "Gaussian orthogonal relay channel: Optimal resource allocation and capacity," Aug. 2004, preprint.