

Capacity Regions for Wireless Ad Hoc Networks

Stavros Toumpis, *Student Member, IEEE* and Andrea J. Goldsmith, *Senior Member, IEEE*

Abstract—We define and study capacity regions for wireless ad hoc networks with an arbitrary number of nodes and topology. These regions describe the set of achievable rate combinations between all source-destination pairs in the network under various transmission strategies, such as variable-rate transmission, single-hop or multihop routing, power control, and successive interference cancellation (SIC). Multihop cellular networks and networks with energy constraints are studied as special cases. With slight modifications, the developed formulation can handle node mobility and time-varying flat-fading channels. Numerical results indicate that multihop routing, the ability for concurrent transmissions, and SIC significantly increase the capacity of ad hoc and multihop cellular networks. On the other hand, gains from power control are significant only when variable-rate transmission is not used. Also, time-varying flat-fading and node mobility actually improve the capacity. Finally, multihop routing greatly improves the performance of energy-constraint networks.

Index Terms—Ad hoc, capacity region, energy constraints, flat fading, mobility, multihop cellular, multihop routing, wireless.

I. INTRODUCTION

WIRELESS networks consist of a number of nodes communicating over a wireless channel. Depending on their architecture, they can be roughly divided in two categories. In those following the *cellular* paradigm, all nodes communicate directly with a base stations that are responsible for controlling all transmissions and forwarding data to the intended users. In those following the *ad hoc* paradigm, all nodes have the same capabilities and responsibilities. Two nodes wishing to communicate can either do so directly, if possible, or route their data through other nodes. Our work deals with this second type of networks.

The nature of the wireless channel, the lack of synchronization, and also the lack of any predetermined topology creates many challenging research topics in the area of ad hoc networks [1], [2]. Traditionally, research has been concentrated on random access [3]–[9], transmission scheduling [10], and routing [11], [12]. Networks with energy constraints are also being studied [13], [14].

Lately, there has also been work on determining the capacity of ad hoc networks. In a recent landmark paper [15], the authors derived lower and upper bounds on the performance of a class of networks in the limit of a large number of nodes, in terms of a single figure of merit, the maximum uniformly achievable communication rate between all nodes and their selected destinations. In this work, we define and investigate capacity regions

for ad hoc networks with any number of nodes. These multidimensional regions are more descriptive, since they contain all achievable combinations of rates between the network nodes under various transmission protocols.

The Shannon capacity region of ad hoc networks remains an open problem, so our capacity regions only define the maximum achievable rates under specific transmission protocols, which may be suboptimal. However, our problem formulation allows us to investigate the impact of different techniques on network performance, including power control, multihop routing, spatial reuse, successive interference cancellation (SIC) and variable-rate transmission.

The remainder of the paper is organized as follows. In Section II, we describe the system model. In Section III, we define rate matrices and capacity regions for ad hoc networks. In Section IV, we define capacity regions for a sequence of five transmission protocols of increasing sophistication, and study them in the context of a random network topology. In Section V, we discuss computational issues. In Section VI, we modify the formulation to study multihop cellular networks and, in Section VII, we extend the formulation to include the effects of time-varying flat-fading and node mobility on the capacity of the network. In Section VIII, we present another modification, suitable for the study of energy-constrained networks. We conclude in Section IX. Throughout the paper, terms being defined are set in **boldface**.

II. SYSTEM MODEL

Consider an ad hoc network with n nodes A_1, A_2, \dots, A_n . Each node has a transmitter, a receiver, and an infinite buffer, and wishes to communicate with some or all of the other nodes, possibly by multihop routing. We assume that nodes cannot transmit and receive at the same time. We also assume that nodes do not wish to multicast information, so every transmission is intended for a single node.

Node A_i transmits at some fixed maximum power P_i and all transmissions occupy the full bandwidth W of the system. We define $P = [P_1 \ P_2 \ \dots \ P_n]^T$ as the **power vector**. When A_i transmits, A_j receives the signal with power $G_{ij}P_i$, where G_{ij} denotes the channel gain between nodes A_i and A_j . We define the **channel gain matrix** to be the $n \times n$ matrix $G = \{G_{ij}\}$. The elements along the diagonal are unimportant and are set to $G_{ii} = 0$. The receiver of each node is subject to thermal noise, background interference from various noise sources such as other networks, and interference from other users, where the interference caused by A_i to A_j is also determined by the link gain G_{ij} . We model thermal noise and background interference jointly, as a single source of additive white Gaussian noise (AWGN), with power spectral density η_i for node A_i . We define the **noise vector** $H = [\eta_1 \ \eta_2 \ \dots \ \eta_n]^T$.

Manuscript received September 26, 2001; revised April 24, 2002 and May 10, 2002; accepted June 3, 2002. The editor coordinating the review of this paper and approving it for publication is W. W. Lu. This work was supported by the Office of Naval Research (ONR) under Grant N00014-99-1-0698.

The authors are with the Department of Electrical Engineering, Stanford University, Stanford, CA 94305-9515 USA (e-mail: stoumpis@stanford.edu).

Digital Object Identifier 10.1109/TWC.2003.814342

Let $\{A_t: t \in \mathcal{J}\}$ be the set of transmitting nodes at a given time, each node A_i transmitting with power P_t . Let us assume that node A_j , $j \notin \mathcal{J}$ is receiving information from node A_i , $i \in \mathcal{J}$. Then the **signal-to-interference and noise ratio (SINR)** at node A_j will be

$$\gamma_{ij} = \frac{G_{ij}P_i}{\eta_j W + \sum_{k \in \mathcal{J}, k \neq i} G_{kj}P_k}. \quad (1)$$

We assume that A_i varies the transmission rate based on γ_{ij} to meet a given performance metric. Specifically, nodes A_i and A_j agree on a transmission rate $r = f(\gamma_{ij})$ where $f(\cdot)$ is a function that reflects the quality of the receiver and the performance metric. For example, based on Shannon capacity, we can set

$$f(\gamma_{ij}) = W \log_2(1 + \gamma_{ij}). \quad (2)$$

Under the Shannon assumption, bits transmitted over the link are received with asymptotically small probability of error as long as (2) holds. Alternatively, $f(\gamma_{ij})$ could be the maximum data rate that satisfies a given BER requirement under a specific modulation scheme such as M -array quadrature amplitude modulation (M -QAM) [16]. Note that in (1), we treat all interference signals from other nodes as noise: this assumption will be relaxed when we consider SIC.

We assume omniscient nodes with perfect knowledge of the channel gain matrix (G) and the noise (H) and power (P) vectors. The transmission protocol for all nodes is agreed to in advance. Thus, no overhead is needed for nodes to determine G , H , P , or the transmission schedule.

III. RATE MATRICES AND CAPACITY REGIONS

In this section, we define the capacity region of a network as a set of rate matrices. Rate matrices provide a mathematical framework for describing the transmission schemes and time-division schedules used by a network.

A. Transmission Schemes and Time-Division Schedules

A **transmission scheme** \mathcal{S} is a complete description of the information flow between different nodes in the network at a given time instant. Therefore, the transmission scheme at a given time consists of all transmit–receive node pairs in operation at that time and, for each of these pairs, the transmission rate and the original source node of the transmitted information. Note that we have assumed that nodes cannot transmit and receive simultaneously, and that the rates used in a transmission scheme are set to $r = f(\gamma_{ij})$, so signals always meet their required performance metric.

As an example, consider the network of Fig. 1 where the node pairs (A_1, A_2) , (A_2, A_3) , (A_3, A_4) , and (A_4, A_1) can all communicate directly but the node pairs (A_1, A_3) and (A_2, A_4) cannot, perhaps because of an obstruction along their line of sight. Therefore, if A_1 wants to communicate with A_3 , it must do so by forwarding data via intermediate nodes, and similarly for traffic between nodes A_2 and A_4 . Two example schemes for this network are \mathcal{S}_1 and \mathcal{S}_2 , depicted in Fig. 1(i) and (ii), respectively. In \mathcal{S}_1 , node A_1 sends its own information to node A_2 and node A_3 sends its own information to node A_4 . As discussed, for

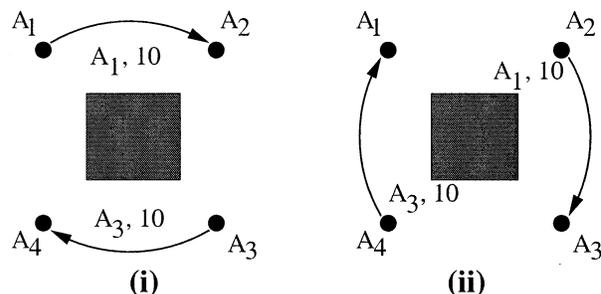


Fig. 1. Two transmission schemes for a network of four nodes. (i) \mathcal{S}_1 . (ii) \mathcal{S}_2 . Transmit–receive node pairs in operation are connected by arrows. The node whose information is being transmitted and the link transmission rate are shown next to the link arrows. The rates are dictated by the SINR of each link and function $f(\cdot)$.

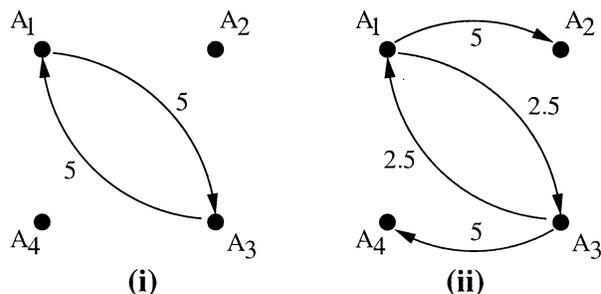


Fig. 2. Information flows for two time-division schedules that use the schemes of Fig. 1. (i) $\mathcal{T}_1 = 0.5\mathcal{S}_1 + 0.5\mathcal{S}_2$. (ii) $\mathcal{T}_2 = 0.75\mathcal{S}_1 + 0.25\mathcal{S}_2$. Arrows signify the end-to-end information flows, and numbers denote the overall communication rates.

each of these transmissions, the bit rate is set to $f(\gamma_{ij})$ where γ_{ij} is the SINR of the link. In the second scheme, A_2 is forwarding A_1 's data to A_3 , and A_4 is forwarding A_3 's data to A_1 (presumably these data were received at a previous time interval.)

At different times, networks may operate under different transmission schemes, for example, in order to provide multihop routing. We assume that the network operation is organized in identical consecutive **frames** of some fixed duration. Within each frame, the network operates using successively schemes $\mathcal{S}_1, \dots, \mathcal{S}_k$, with scheme \mathcal{S}_i operating during a fraction of the frame equal to a_i , where $\sum_{i=1}^k a_i = 1$. We say that the network is using the **time-division schedule** $\mathcal{T} = \sum_{i=1}^k a_i \mathcal{S}_i$, and we refer to the fractions a_i , $i = 1, \dots, k$ as the **weights** of the time-division schedule. For the network of Fig. 1, two possible time-division schedules are $\mathcal{T}_1 = 0.5\mathcal{S}_1 + 0.5\mathcal{S}_2$ and $\mathcal{T}_2 = 0.75\mathcal{S}_1 + 0.25\mathcal{S}_2$. The resulting end-to-end information flows when the network operates under these time-division schedules appear in Fig. 2.

Depending on the ordering of schemes within a time-division schedule, the schedule may imply noncausal routing, so that within a frame a node may forward traffic from another node before that traffic actually arrives. This situation does not pose a problem since we can place before the sequence of frames a time period of finite duration during which some data are backlogged in the intermediate nodes before their final destination. Since the initialization period will have finite duration, the overall performance of the network will not be affected, in the limit of a large number of frames. Therefore, without compromising the generality of our results, we neglect causality in our routing since it significantly complicates the problem and obscures our main results.

B. Rate Matrices

Although transmission schemes are useful for describing the state of the network at a given time, they are not convenient for mathematical manipulation. We will, therefore, use rate matrices to represent transmission schemes. For a network with n nodes, we define the **rate matrix** $R(\mathcal{S})$ of a transmission scheme \mathcal{S} as an $n \times n$ square matrix with entries r_{ij} such that

$$r_{ij} = \begin{cases} r, & \text{if node } A_j \text{ receives information at rate } r \\ & \text{with node } A_i \text{ as the original information} \\ & \text{source} \\ -r, & \text{if node } A_j \text{ transmits information at rate } r \\ & \text{with node } A_i \text{ as the original information} \\ & \text{source} \\ 0, & \text{otherwise.} \end{cases} \quad (3)$$

Each nonzero rate matrix entry denotes a transfer of data. The row index of the entry corresponds to the original source of the data. The column index specifies the receiver or transmitter of the data. Specifically, a negative entry $r_{ij} < 0$ in row i corresponds to the rate at which node A_j transmits information that originated at node A_i . This entry is negative to reflect the fact that the data forwarded cannot be counted in the data that A_j receives from A_i . A positive entry $r_{ij} > 0$ correspond to the rate at which node A_j receives information that originated at node A_i , directly or from another node A_k , in which case $r_{ik} = -r_{ij}$. For example, the rate matrices of schemes \mathcal{S}_1 and \mathcal{S}_2 are, respectively

$$R_1 = \begin{bmatrix} -10 & 10 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -10 & 10 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$R_2 = \begin{bmatrix} 0 & -10 & 10 & 0 \\ 0 & 0 & 0 & 0 \\ 10 & 0 & 0 & -10 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Rate matrices mathematically capture all the information needed to describe the state of the network at a given time: namely, which nodes transmit or receive, at what rate, and from which nodes the data originate. We note that since information must be preserved, i.e., each transmission originates at one node and is received by one node, the elements along any row of a rate matrix must sum to zero.

Up to now, we have associated rate matrices with transmission schemes, which describe the state of the network at a given point in time. However, as time progresses, the network will operate under a time-division schedule that alternates between different schemes. By construction, a time-division schedule of transmission schemes is described by the weighted sum of the corresponding rate matrices with weights equal to the percentage of time that each scheme is in operation. Therefore, if $\sum_{i=1}^N a_i \mathcal{S}_i$ (with $a_i \geq 0$ and $\sum_{i=1}^N a_i = 1$) is a time-division schedule, then its rate matrix will be $R = \sum_{i=1}^N a_i R_i$, where R_1, \dots, R_N are the rate matrices of the schemes $\mathcal{S}_1, \dots, \mathcal{S}_N$.

More succinctly, the following linearity property holds:

$$R \left(\sum_{i=1}^N a_i \mathcal{S}_i \right) = \sum_{i=1}^N a_i R(\mathcal{S}_i). \quad (4)$$

For example, the rate matrices of schedules $\mathcal{T}_1 = 0.5\mathcal{S}_1 + 0.5\mathcal{S}_2$ and $\mathcal{T}_2 = 0.75\mathcal{S}_1 + 0.25\mathcal{S}_2$ will be

$$0.5R_1 + 0.5R_2 = \begin{bmatrix} -5 & 0 & 5 & 0 \\ 0 & 0 & 0 & 0 \\ 5 & 0 & -5 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$0.75R_1 + 0.25R_2 = \begin{bmatrix} -7.5 & 5 & 2.5 & 0 \\ 0 & 0 & 0 & 0 \\ 2.5 & 0 & -7.5 & 5 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

and by comparing them with Fig. 2, we see that they correctly reproduce the information flows of the respective schedules.

C. Ad Hoc Network Capacity Regions and Uniform Capacity

We use the term **transmission protocol** to describe a collection of rules that nodes must satisfy when transmitting. For example, a transmission protocol could be that nodes are only allowed to transmit their own information, must transmit with their maximum power, can transmit simultaneously with other nodes, and treat interfering transmissions as noise (SIC is not allowed). Under a given transmission protocol, a collection of schemes is available to the network. Each of these schemes has a rate matrix. We refer to these as the **basic rate matrices**. Clearly, the less restrictive the transmission protocol, the more schemes are available, and the larger the collection of basic rate matrices.

Since weighted sums of rate matrices describe the net flow of information in the network under a corresponding time-division schedule, we could define the capacity of the network under time-division and a given transmission protocol as the set of weighted sums of all basic rate matrices of this protocol (with the coefficients being positive, and their sum being equal to unity). However, some weighted sums of rate matrices have off-diagonal elements that are negative. Such rate matrices correspond to scenarios where some nodes forward more information from a source than they receive from that source (possibly indirectly, through routing). Clearly, this is not a stable condition, and we, therefore, exclude these sums from the capacity region. All other weighted sums of basic rate matrices can be included in the network capacity region.

Formally, based on the above discussion, we define the **capacity region** of the wireless ad hoc network, under time-division routing and a given transmission protocol, as the convex hull of the basic rate matrices with the restriction that the weighted sums must have nonnegative off diagonal elements. Specifically, if $\{R_1, \dots, R_N\}$ denotes the set of basic rate matrices, the capacity region is

$$C = C(\{R_i\}) \triangleq \text{Co}(\{R_i\}) \cap \mathcal{P}_n$$

$$= \left\{ \sum_{i=1}^N a_i R_i : a_i \geq 0, \sum_{i=1}^N a_i = 1 \right\} \cap \mathcal{P}_n \quad (5)$$

where \mathcal{P}_n is the subset of all $n \times n$ matrices with all their off-diagonal elements nonnegative and $Co\{R_i\}$ denotes the convex hull of the set $\{R_i\}$ of basic matrices.

The shape of the capacity region depends on the pool of basic rate matrices. This pool depends on the network topology and parameters, and also on the transmission protocol. The meaning of the capacity region is the following: Let $R = \{r_{ij}\}$ be a matrix in the capacity region. Then there is a time-division schedule of transmission schemes that are acceptable by the transmission protocol, such that when the network operates under this time division and $i \neq j$, r_{ij} is the rate with which node A_i sends its own information to node A_j , possibly through multiple hops, and $-r_{ii}$ is the total rate with which node A_i is passing information to all other nodes. Since the elements in each row of all matrices in the capacity region must sum to zero, the capacity region is contained in an $n \times (n - 1)$ subspace. This dimensionality is expected, since there are n nodes, each with $(n - 1)$ other nodes with which it may want to communicate. Indeed, each of the $n(n - 1)$ possible communication pairs corresponds to exactly one of the $n(n - 1)$ off-diagonal elements.

To capture the capacity of an ad hoc network with a simple figure of merit, we define the **uniform capacity** C_u of a network under time-division routing and a given transmission protocol as the maximum aggregate communication rate, if all nodes wish to communicate with all other nodes, at a common rate. The uniform capacity is equal to $r_{\max} \times n(n - 1)$, where r_{\max} is the largest r for which the matrix with all its off-diagonal elements equal to r belongs to the capacity region, and $n(n - 1)$ is the total number of source-destination pairs for a network of n nodes.

IV. CAPACITY REGIONS FOR VARIOUS TRANSMISSION PROTOCOLS

In this section, we define a sequence of capacity regions, each corresponding to a progressively more sophisticated transmission protocol (and consequently a richer pool of basic rate matrices), and then study them in the context of an example ad hoc network. This will illustrate the capacity gains that can be obtained by using these protocols. The example network consists of five nodes, and has a random topology obtained by uniformly and independently distributing five nodes in the box $\{-10 \text{ m} \leq x \leq 10 \text{ m}, -10 \text{ m} \leq y \leq 10 \text{ m}\}$. The power gains between nodes A_i and A_j are given by

$$G_{ij} = K S_{ij} \left(\frac{d_0}{d_{ij}} \right)^\alpha \quad (6)$$

where d_{ij} is the distance between the nodes, K and d_0 are normalization constants set to $K = 10^{-6}$ and $d_0 = 10 \text{ m}$, respectively, the path loss exponent α is set to $\alpha = 4$, and the shadowing factors $S_{ij} = S_{ji}$ are random, independent, and identically generated from a lognormal distribution with a mean of 0 dB and variance $\sigma = 8 \text{ dB}$ (so $S_{ij} = 10^{N_{ij}/10}$ and N_{ij} is Gaussian with expectation $EN_{ij} = 0$ and standard deviation $\sigma_{N_{ij}} = 8$). The power gains are assumed to remain constant for all time; this assumption will be dropped in Section VIII. The transmitter powers are $P_i = 0.1 \text{ W}$, all receivers are subject to AWGN with the same power spectral density $\eta = 10^{-10} \text{ W/Hz}$,

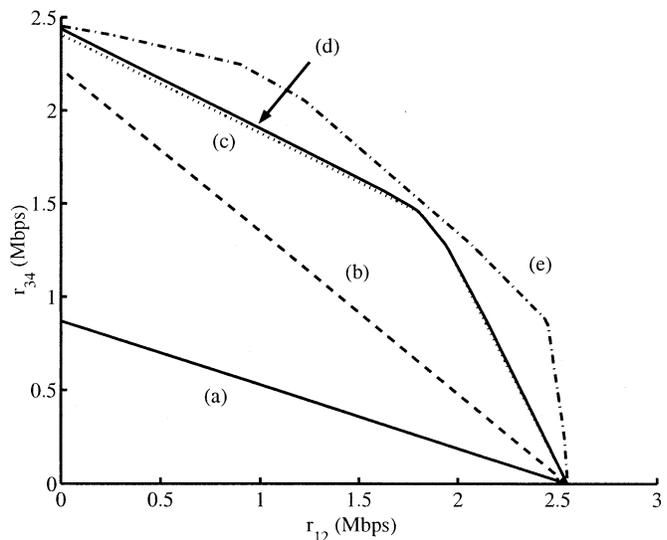


Fig. 3. Capacity region slices of the example ad hoc network along the plane $r_{ij} = 0, (i, j) \neq (1, 2), (3, 4), i \neq j$. (a) Single-hop routing, no spatial reuse. (b) Multihop routing, no spatial reuse. (c) Multihop routing with spatial reuse. (d) Two-level power control added to (c). (e) SIC added to (c).

and the bandwidth is $W = 10^6 \text{ Hz}$. The link data rates are set according to the receiver SINR and the Shannon capacity limit of (2). Note that (2) also reflects the rate that can be achieved using uncoded or coded M -QAM when the transmit power is reduced by an appropriate factor [17].

A. Single-Hop Routing, No Spatial Reuse

We first determine the capacity region when only single-hop routing is allowed (no forwarding) and only one node is transmitting at any time. By only allowing one active node at a time, link data rates are higher since there is no interference, but the network does not take advantage of spatial reuse. Since there are n nodes in the system and each of them has $n - 1$ possible receivers, the network has $N^a = n(n - 1) + 1$ transmission schemes (including the one in which all nodes remain silent), and associated basic rate matrices $R_i^a, i = 1, \dots, N^a$. One of these, the **zero rate matrix**, will correspond to the scheme where all nodes remain silent. Determining the rest of the basic matrices is straightforward using (2), G, P , and H . The capacity region will, therefore, be

$$C^a = Co\{R_i^a, i = 1, \dots, N^a\} \cap \mathcal{P}_n.$$

In Fig. 3(a), we have drawn a two-dimensional slice of C^a along the plane $r_{ij} = 0, (i, j) \neq (1, 2), (3, 4), i \neq j$. This line captures a background rate of zero for node pairs other than $(A_1 \rightarrow A_2)$ and $(A_3 \rightarrow A_4)$. Therefore, only nodes A_1 and A_3 send data. The other nodes never transmit since they do not have data of their own to send and, in single-hop routing, they cannot help in forwarding. Note that the slice is a straight line, as expected, since without spatial reuse at any one time the network supports information transfer between only one source-destination pair. By changing the time percentages that are devoted to each of the two node pairs, different points on the straight line (a) will be achieved. The uniform capacity of the network is $C_u^a = 0.83 \text{ Mb/s}$.

B. Multihop Routing, No Spatial Reuse

Next we consider the case where multihop routing is allowed, but spatial reuse is not, so only one node is transmitting at a given time. Since there are n nodes in the systems, and each has $n - 1$ different possible receivers and n possible nodes to forward data for (including itself and the receiver), there are now $N^b = n^2(n - 1) + 1$ possible transmission schemes (including the one in which all nodes remain silent) and associated rate matrices R_i^b , $i = 1, \dots, N^b$. Determining these matrices is straightforward using (2), G , P , and H . The capacity region under these assumptions will be

$$C^b = Co\{R_i^b, i = 1, \dots, N^b\} \cap \mathcal{P}_n.$$

In Fig. 3(b), we have drawn a slice of C^b along the plane $r_{ij} = 0$, $(i, j) \neq (1, 2), (3, 4)$, $i \neq j$. We note that this slice is again a straight line, as expected, since without spatial reuse at any one time the network supports information transfer between only one source-destination pair. We also note a significant increase in the size of the capacity region as compared with the previous case. This is due to the fact that under multihop routing, the nodes can avoid transmitting directly to their destination over paths with small gains, and instead use multiple hops over channels with much more favorable gains and correspondingly higher rates. Uniform capacity is also increased by 242% to $C_u^b = 2.85$ Mb/s.

C. Multihop Routing With Spatial Reuse

We now allow both multihop routing and spatial reuse. In this case, a network of n nodes will have

$$N^c = \sum_{i=1}^{\lfloor n/2 \rfloor} \frac{n(n-1) \cdots (n-2i+1)}{i!} n^i + 1 \quad (7)$$

distinct transmission schemes. Indeed, the i th term in the above sum is the total number of schemes having i transmit-receive pairs. There are $n(n-1) \cdots (n-2i+1)$ distinct choices for the $2i$ nodes that are involved; however, this number must be divided by $i!$ to account for the fact that pair orderings are unimportant. The total number of pair combinations is multiplied by n^i to account for the different possibilities in the choice of information sources, for each of the pairs. Denoting the basic rate matrices by R_i^c , $i = 1, \dots, N^c$, we can define the capacity region C^c as

$$C^c = Co\{R_i^c, i = 1, \dots, N^c\} \cap \mathcal{P}_n.$$

In Fig. 3(c), we have drawn a slice of C^c along the plane $r_{ij} = 0$, $(i, j) \neq (1, 2), (3, 4)$, $i \neq j$. We note that the slice is no longer a straight line, as the network can now use spatial reuse to maintain multiple active transmissions, and at any time instant more than one stream may be serviced (directly or along a multihop route). The introduction of spatial reuse increases uniform capacity by 26% to $C_u^c = 3.58$ Mb/s, even for this small network of five nodes.

D. Power Control

We have so far assumed that nodes either transmit at their maximum power or remain silent. If we relax this condition and allow each node to transmit at different power levels below the maximum power, then we increase the set of basic rate matrices and thereby, possibly, augment the capacity region. Since there are infinitely many possible power levels, we restrict our attention to power control strategies where node i transmits at one of p possible power levels: $\{(1/p)P_i, (2/p)P_i, \dots, P_i\}$. The network will then have a set of

$$N^d = \sum_{i=1}^{\lfloor n/2 \rfloor} \frac{n(n-1) \cdots (n-2i+1)}{i!} n^i p^i + 1 \quad (8)$$

basic rate matrices. Equation (8) is derived as (7), with the factor p^i being added to account for the different possible power control scenarios, when i pairs are active. Denoting the set of basic rate matrices by R_i^d , $i = 1, \dots, N^d$, the capacity region now becomes

$$C^d = Co\{R_i^d, i = 1, \dots, N^d\} \cap \mathcal{P}_n.$$

In Fig. 3(d), we have drawn a slice of C^d along the plane $r_{ij} = 0$, $(i, j) \neq (1, 2), (3, 4)$, $i \neq j$, for power control with two levels ($p = 2$). We observe that this simple power control strategy does not significantly change this slice of the capacity region. Moreover, the uniform capacity changes less than 1%, to $C_u^d = 3.61$ Mb/s. As expected, additional power levels lead to negligible gains. This result is consistent with other results on variable-rate transmission with power control, which indicate that if the transmission rate is adjusted to the link SINR, additional power control does not significantly improve performance [16]. On the other hand, as we will see later on, power control leads to significant gains when variable-rate transmission is not possible.

E. SIC

Until now, we have assumed that under transmission schemes with many simultaneous transmissions each node decodes only its intended signal and treats all other signals as noise. However, under a SIC strategy, nodes may decode some signals intended for other nodes first, subtract out this interference, and then decode their own signals. This strategy may cause a node to restrict the transmission rate of an interfering node, since the receiving node must be able to decode the interfering signal. (For example, under Shannon analysis, the interfering signal is assumed to be decoded perfectly as long as the rate at which the interfering signal is set to is less than the capacity of the interfering link.) However, this restriction is balanced by the fact that the receiving node's rate will increase due to the removal of interference. For example, consider the setup of Fig. 1(i). Node A_3 's signal will interfere with node A_2 's reception. A_2 could decode A_1 's signal and treat A_3 's signal as noise, or could first decode and remove the signal from node A_3 and then decode the desired signal from node A_1 . This second decoding strategy will impose an additional constraint on the transmission rate of node A_3 , since its signal will also need to be decoded at A_2 . A similar choice of strategies exists for A_4 . Assuming a network of n

nodes, power control with p levels, and SIC, the total number of transmission schemes, and associated basic rate matrices, will be

$$N^e = \sum_{i=1}^{\lfloor n/2 \rfloor} \frac{n(n-1) \cdots (n-2i+1)}{i!} n^i p^i Z^i(i) + 1 \quad (9)$$

where $Z(i) = 1 * 2 * \cdots * (i-1) + 2 * \cdots * (i-1) + \cdots + (i-1) + 1$. Equation (9) is derived as (8), with the factor $Z(i)^i$ being added to account for all the possible combinations of successive interference strategies. $Z(i)$ represents the total number of SIC strategies available to a specific receiver in the presence of i transmissions; the user may decode its own signal first (1 scenario) or second ($i-1$ scenarios, depending on what signal is decoded first), or third [$(i-1)(i-2)$ scenarios], etc. Denoting the set of basic rate matrices by R_i^e , $i = 1, \dots, N^e$, the capacity region is

$$C^e = \text{Co}\{R_i^e, i = 1, \dots, N^e\} \cap \mathcal{P}_n.$$

Fig. 3(e) shows a slice of C^e for the example ad hoc network, with no power control ($p = 1$) along the plane $r_{ij} = 0$, $(i, j) \neq (1, 2), (3, 4)$, $i \neq j$. This slice indicates that SIC significantly augments the capacity region even without power control. Moreover, the uniform capacity becomes $C_u^e = 4.31$ Mb/s, 19% greater than C_u^c , that corresponds to spatial reuse and multiple hops, but no SIC.

F. Discrete-Rate Transmission

Until now, we have assumed that the transceivers used are capable of variable-rate adaptation as defined by (2), meaning that transmitters automatically adjust the transmission rate to match the SINR at the receiver and achieve the Shannon bound. Such an assumption implies that the available rates are not restricted to a given set of values. We now consider a discrete-rate restriction in our capacity calculations, by use of a step function for $f(\cdot)$ that is bounded above by (2), where each of the steps corresponds to a different possible rate.

In Fig. 4, we plot the uniform capacity of the example ad hoc network of five nodes, for the protocols introduced in the previous subsections, and for four different rate restrictions. Specifically, each of the restrictions consists of L different transmission rates, with $L = 1, 3, 7, 15$. Each of the available rates has a minimum SINR requirement based on the Shannon capacity formula (2). The transmitter chooses the highest rate within the set of possible rates for which the SINR requirement is met. We compare the performance of these different rate restrictions with the performance achieved when there is no rate restriction, as in the previous subsections. As expected, increasing the number of available rates improves capacity. More importantly, power control now leads to significant gains. For example, when only one rate is available, if added to spatial reuse and multihop routing, power control with two power levels raises the uniform capacity by 26%, and outperforms SIC. Power control with three levels raises the uniform capacity by 31%, but using more than three levels leads to negligible gains. This result is consistent with other findings in the literature [16] and shows that power control can be interchanged to some extent with rate adaptation. The performance enhancements achieved by using power con-

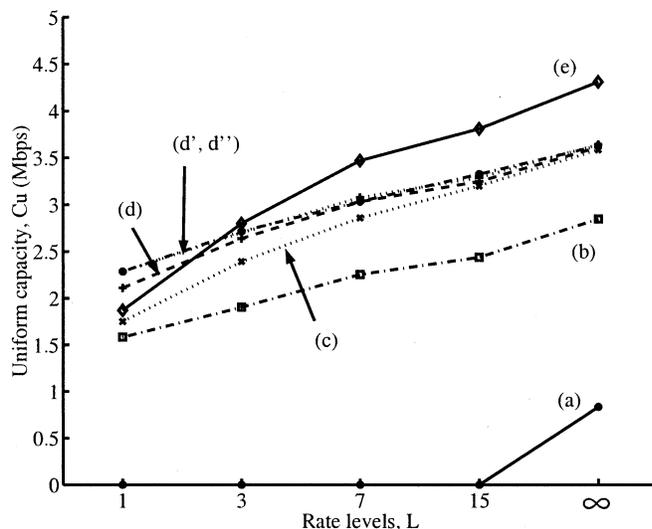


Fig. 4. Uniform capacity versus the number of available rate levels for the example ad hoc network. Each curve corresponds to a different transmission protocol. (a) Single-hop routing, no spatial reuse. (b) Multihop routing, no spatial reuse. (c) Multihop routing with spatial reuse. (d, d', d'') Power control with two, three, and four levels added to (c). (e) SIC added to (c).

rol when time division is not allowed have been studied extensively [18]; this result also shows that power control leads to improved network capacity, even when time division is factored in, provided that variable-rate transmission is not being used.

G. Uniform Capacity of Canonical Topologies

In the previous subsections, we studied the capacity regions of a network with a random topology for a sequence of transmission protocols, and determined that multihop routing, spatial reuse, and SIC all improve the performance significantly. Adding power control yields significant gains only when very limited or no rate adaptation is used. We have determined that these results are general, and hold with little variation for a wide range of random or canonical topologies and modeling parameters.

For example, in Fig. 5 we plot the uniform capacity versus the number of nodes for the five transmission protocols introduced in the previous subsections, and for two canonical topologies: The first, which we define as the **linear topology**, consists of nodes arranged on a linear array, with a spacing of 10 m between them. The second, which we call the **ring topology**, consists of nodes arranged on a circle with a radius of 10 m. The nodes are separated by arches of equal length. In both cases, the bandwidth available to the system is $W = 1$ MHz, all nodes have the same maximum power of $P_i = 0.1$ W, and the thermal noise power spectral density is the same for all receivers and equal to $\eta = 10^{-10}$ W/Hz. We calculate channel gains using the model of (6), with the parameters set to $\alpha = 3$, $d_0 = 10$ m, $K = 10^{-6}$ and $\sigma = 0$ dB (so that the topology, not random shadowing, dominates the performance).

We note that the trend of the curves strongly depends on the topology. In the case of linear networks, the performance of the transmission protocols that do not allow spatial reuse deteriorates fast as the number of nodes increase. The rest of the protocols perform significantly better, but the overall performance

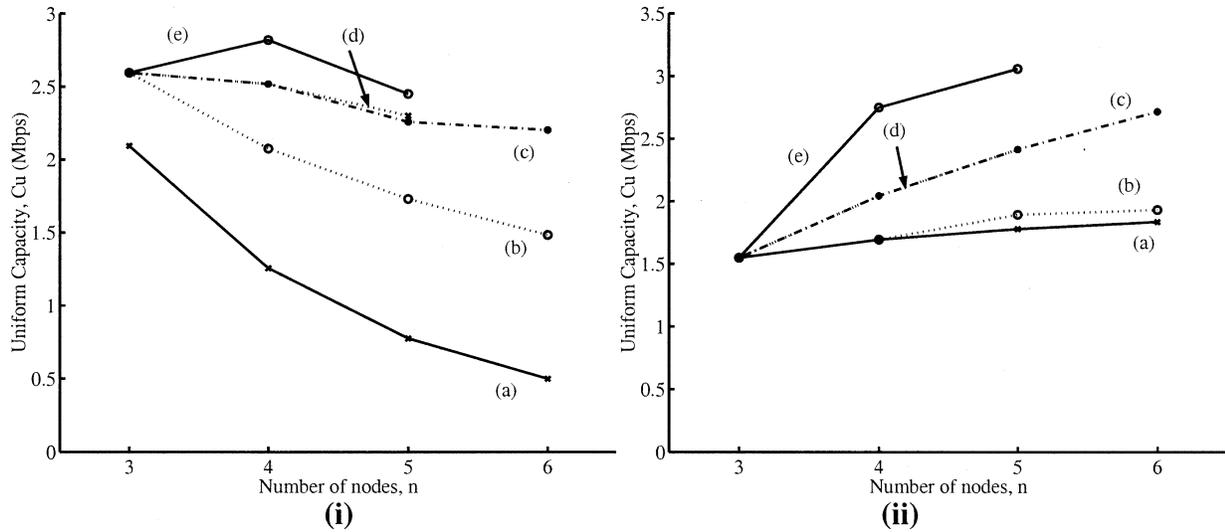


Fig. 5. Uniform capacity of canonical topologies as a function of the number of nodes. (i) Linear networks. (ii) Ring networks. (a) Single-hop routing, no spatial reuse. (b) Multihop routing, no spatial reuse. (c) Multihop routing with spatial reuse. (d) Two-level power control added to (c). (e) SIC added to (c).

decreases with the number of nodes. This means that the loss from the need for multiple hops is greater than the gain from improved spatial separation. In the case of ring networks, comparing Fig. 5(a) and (b) reveals that gains from allowing multiple hops are limited. This is due to the fact that nodes are clustered together in a ring formation. Still, the performance of the protocols that allow concurrent transmissions actually improves as we add more nodes, which implies that networks can take advantage of spatial separation even when more nodes are placed in the same area. This finding is reminiscent of the results appearing in [15].

We have also established that the relative performance shown in the previous figures does not change under a wide range of the transmitter powers, thermal noise powers, and the exponential decay parameter α . SIC was found to be particularly effective in the case of channels with comparable link gains, for example when α is small ($\alpha \sim 2$).

V. COMPUTATIONAL ISSUES

We have defined the capacity region of a network as the intersection of the set \mathcal{P}_n with the convex hull of the network's basic rate matrices R_1, R_2, \dots, R_N . Therefore, checking if a point is in the network's capacity region is equivalent to checking if the point belongs to \mathcal{P}_n , which is trivial, and checking if it belongs to the convex hull of the network's basic rate matrices. Since rate matrices are isomorphic to vectors of length $n(n-1)$, this is a standard problem in computational geometry and can be solved by a variety of different techniques. We chose to cast it as the following linear program in the $(N-1)$ -dimensional Euclidean space:

$$\begin{aligned} & \text{minimize: } g(x) = \sum_{i=1}^{N-1} x_i \\ & \text{subject to: } 0 \leq x_i \leq 1 \\ & R = \sum_{i=1}^{N-1} x_i R_i \end{aligned} \quad (10)$$

where $\{R_1, \dots, R_N\}$ is the set of all basic rate matrices for the network, and R_N , specifically, is the zero rate matrix (that corresponds to all nodes being silent). If the problem is feasible (i.e., the set of points satisfying the constraints is not empty) with $g(x_{\text{opt}}) \leq 1$, then R belongs to the capacity region, and can be obtained via a time-division schedule of the basic rate matrices with the zero rate matrix R_N being active $100(1 - g(x_{\text{opt}}))$ percent of the time. If the problem is infeasible (i.e., the constraints cannot be satisfied for any point x) or it is feasible with $g(x_{\text{opt}}) > 1$, R does not belong to the capacity region.

By iteratively solving the linear problem (10), we can determine boundary points of the capacity regions. These boundary points correspond to optimal modes of operation for the network. For example, if we wish to maximize rates r_{12} and r_{34} with $r_{12} = r_{34}$ and allow spatial separation, multiple hops, and SIC, but no power control, the "best" time-division schedule is that of Fig. 6 that achieves $r_{12} = r_{34} = 1.6564$ Mb/s. This figure shows that the developed method may be viewed as a solution to the optimal routing problem when links interfere with each other and the communication needs of the network (in terms of the streams that must be serviced and the rates that are required) are arbitrary.

Note that in order to determine the capacity region, the set of all basic rate matrices must first be calculated. As the number of nodes increases, the number of basic rate matrices increases very fast (for example, factorially for the protocol that allows multihop routing and spatial reuse). Since not all matrices will contribute to the capacity region, significant speed gains can be realized by carefully constructing a minimal set of rate matrices that sufficiently describes it. When the transmitters follow the Shannon capacity limit of (2), the MATLAB/C routines we have developed become impractical for networks with more than seven to eight nodes, when SIC is not used, or five to six nodes, when SIC is used. When variable-rate transmission is not used, and the transmitters can only operate with a single fixed rate (provided the SINR is larger than some threshold), the developed routines become impractical for networks with more than around 15 nodes. We are currently developing more

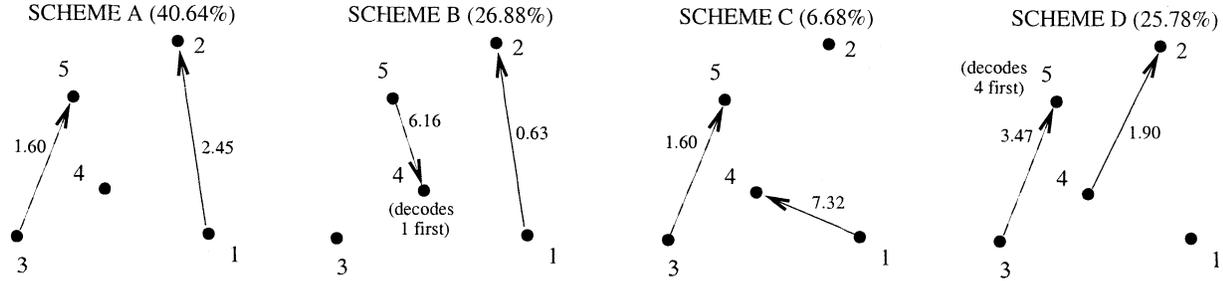


Fig. 6. Time-division schedule that maximizes $r_{12} = r_{34}$, if the rest of the node pairs have no communication requirements, for the example ad hoc network. Arrows signify actual transmissions. Numbers denote link rates (in megabits per second).

sophisticated software that, together with faster hardware, will increase these bounds by three to four nodes in all cases.

VI. MULTIHOP CELLULAR NETWORKS

The developed formulation can be readily applied to the case of multihop cellular networks, in which one node (we assume A_n) acts as a base station, and the rest of the nodes (A_1 to A_{n-1}) can transmit to each other, but are only interested in sending to or receiving from the base station. However, because of the special structure of the network, a simplified formulation can be used to solve for the capacity region. For the case of the uplink communication, instead of defining the rate matrix of a scheme \mathcal{S} , we define the **rate vector** $R(\mathcal{S}) = [r_1 \ r_2 \ \cdots \ r_{n-1}]^T$ in the following manner:

$$r_i = \begin{cases} r, & \text{if node } A_i \text{ transmits information to a node} \\ & \text{(ultimately intended for the base station } A_n) \\ & \text{at rate } r \\ -r, & \text{if node } A_i \text{ receives information from a node} \\ & \text{(ultimately intended for the base station } A_n) \\ & \text{at rate } r \\ 0, & \text{otherwise.} \end{cases} \quad (11)$$

A similar formulation exists for the downlink case. We note that, contrary to the rate matrix formulation, a positive entry signifies transmission of information rather than reception. As in the rate matrix formulation, rate vectors contain all information that is essential for describing the state of the network at a given time.

The rest of the theory can be developed in a manner entirely analogous to the general case. Because of the construction of rate vectors, a weighted time-division schedule of transmission schemes is described by a rate vector equal to the weighted sum of the corresponding rate vectors, with the same weights. Therefore, if a multihop cellular network with n nodes operates under a transmission protocol associated with a set of basic rate vectors $\{R_1, \dots, R_N\}$, the capacity region is

$$\begin{aligned} C &= C(\{R_i\}) \triangleq Co(\{R_i\}) \cap \mathcal{P}_{n-1} \\ &= \left\{ \sum_{i=1}^N a_i R_i : a_i \geq 0, \sum_{i=1}^N a_i = 1 \right\} \cap \mathcal{P}_{n-1} \end{aligned} \quad (12)$$

with \mathcal{P}_{n-1} being the set of all vectors of length $n - 1$ with nonnegative elements. Rate vectors with negative elements are

excluded because they correspond to schedules in which some nodes consistently receive more information than what they transmit, so data ultimately intended for the base station are accumulated to intermediate nodes. The feasibility problem can be solved with a linear program identical to (10). As in the general case, the capacity region depends on the repertoire of basic rate vectors. The set of basic vectors depends on the network topology and parameters, and also on the transmission protocol. The meaning of the capacity region is the following: Let $R = \{r_i\}$ be a vector in the capacity region. Then there is a time division of schemes, allowed by the transmission protocol, such that when the network operates under this time division, r_i is the rate with which node A_i sends its own information to the base station, possibly through multiple hops and time division.

To capture the capacity of a multihop network with a simple figure of merit, we define the **uniform capacity** C_u of a network as the maximum aggregate communication rate, if all nodes wish to communicate with the base station with a common rate. The uniform capacity is equal to $r_{\max} \times (n - 1)$, where r_{\max} is the largest r for which the vector with all its elements equal to r belongs to the capacity region.

As an example, we created a multihop cellular network by placing six nodes randomly, independently, and uniformly in the box $\{-10 \text{ m} \leq x \leq 10 \text{ m}, -10 \text{ m} \leq y \leq 10 \text{ m}\}$ and the base station at the origin. The rest of the parameters were chosen as in the network of Section IV. Slices of the capacity regions for the transmission protocols introduced in Section IV for the general case appear in Fig. 7. The uniform capacities for the same transmission protocols were $C_u^a = 1.52$, $C_u^b = 3.01$, $C_u^c = 3.63$, $C_u^d = 3.67$, and $C_u^e = 4.22$ Mb/s. We have arrived at similar results for various random and canonical topologies and for a wide range of the modeling parameters. So multihop cellular networks behave similarly to ad hoc networks: Multihop routing, spatial reuse, and SIC greatly increase the capacity of the network; improvements from power control are significant only if variable-rate transmission is not used.

VII. FADING AND MOBILITY INCREASE CAPACITY

A. Time-Varying Flat-Fading Channels

With minor modifications, our formulation can handle the case of time-varying flat-fading channels. We return to the general case of ad hoc networks and start by considering a network with a channel that alternates between M different fading states, each being completely described by a gain

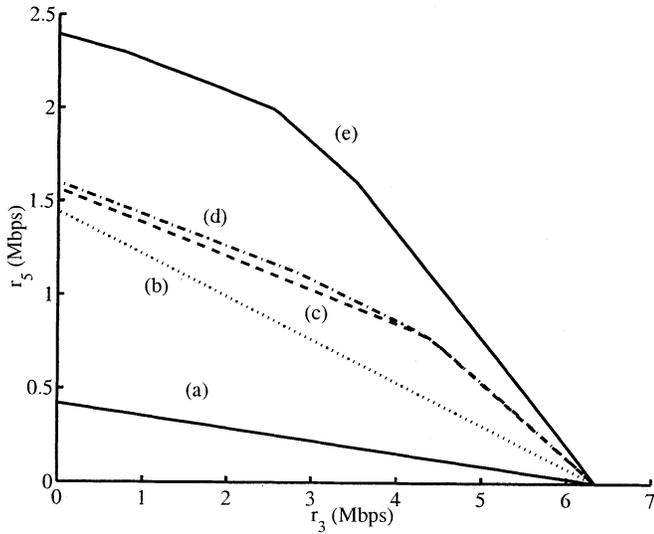


Fig. 7. Capacity region slices of the example multihop cellular network along the plane $r_i = 0, i \neq 3, 5$. (a) Single-hop routing, no spatial reuse. (b) Multihop routing, no spatial reuse. (c) Multihop routing with spatial reuse. (d) Two-level power control added to (c). (e) SIC added to (c).

matrix $G^i, i = 1, \dots, M$. Let $\{R^{ij}, j = 1, \dots, J(i)\}$ be the collection of basic rate matrices for the i th fading state (or a subcollection of them with the same capacity region, as discussed in Section V). Assuming that, over a long period of time, each state will be present for a fraction of time equal to $1/M$, the capacity region of the network can be defined as

$$C = \left\{ \sum_{i=1}^M \sum_{j=1}^{J(i)} a_{ij} R^{ij} : a_{ij} \geq 0, \sum_{j=1}^{J(i)} a_{ij} \leq \frac{1}{M}, \sum_{i=1}^M \sum_{j=1}^{J(i)} a_{ij} = 1 \right\} \cap \mathcal{P}_n. \quad (13)$$

The additional constraint $\sum_{j=1}^{J(i)} a_{ij} \leq 1/M$ simply means that no time-division schedule should use the basic matrices of a given state for longer than that state is available. The capacity region is no longer defined as the convex hull of a collection of matrices, however, we can easily see that it is still convex.

Assuming that $R^{iJ(i)} = R^0, i = 1, \dots, M$ with R^0 being the zero rate matrix, the feasibility problem can be solved by a linear program similar to (10)

$$\begin{aligned} \text{minimize: } g(x) &= \sum_{i=1}^M \sum_{j=1}^{J(i)-1} x_{ij} \\ \text{subject to: } &\begin{cases} R = \sum_{i=1}^M \sum_{j=1}^{J(i)-1} x_{ij} R^{ij} \\ 0 \leq x_i \leq \frac{1}{M}, 1 \leq i \leq M \\ \sum_{j=1}^{J(i)-1} x_{ij} \leq \frac{1}{M}, 1 \leq i \leq M. \end{cases} \end{aligned} \quad (14)$$

If the linear program is feasible (i.e., the set of points satisfying the constraints is not empty) with $g(x_{\text{opt}}) \leq 1$, then R can be

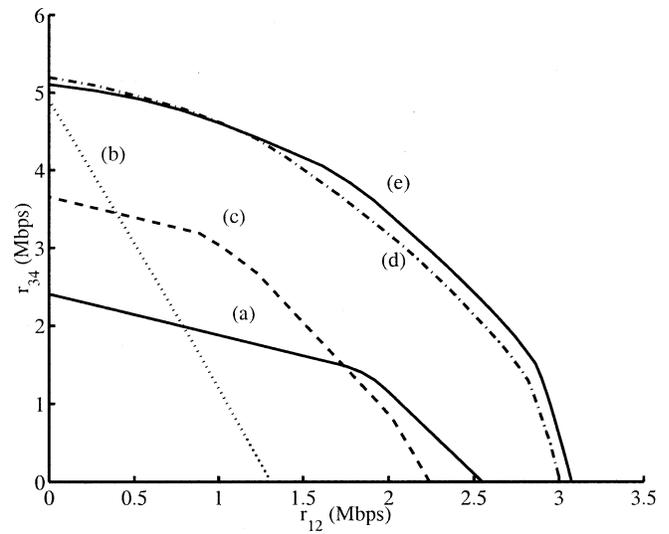


Fig. 8. Slice of the capacity region C^c for the example ad hoc network of Section IV along the plane $r_{ij} = 0, (i, j) \neq (1, 2), (3, 4), i \neq j$ for different combinations of fading states. (a) Fading state $F1$. (b) Fading state $F2$. (c) Fading states $F1, F2$. (d) Fading states $F1, \dots, F10$. (e) Fading states $F1, \dots, F15$.

obtained via a time-division strategy of the basic rate matrices $\{R^{ij}\}$ and the silent rate matrix R^0 that will be “active” $100(1 - g(x_{\text{opt}}))$ percent of the time. Otherwise, R does not belong to the capacity region.

Until now, we assumed that the network operates under a finite number of fading states. This assumption is restrictive, since in most realistic fading models the fading gains can have arbitrary values, so that the total number of fading states is infinite, and the developed formulation becomes inadequate. Furthermore, even if we use a model with a finite number of fading states s , because the gain matrix has $n(n-1)$ nontrivial entries, the resulting number of possible states is $s^{n(n-1)}$, which is intractable for nontrivial values of s and n .

However, the capacity region may be estimated by using a large, but tractable, number of fading states. In this vein, we first create a sequence of independent fading states, according to the statistics of the fading. Then, we use a Monte Carlo approach, and calculate the capacity region as the number of fading states increases, until convergence. Since capacity regions are multi-dimensional, we use convergence of the uniform capacity to indicate capacity region convergence.¹

In the definition of the capacity region, we have intrinsically assumed that nodes schedule their transmissions to take advantage of the different fading states and so are willing to tolerate the random delays associated with waiting for all the required states to come up. Depending on the fading statistics, these delays can be large.

In Fig. 8, we display the convergence of a slice of the capacity region C^c for the network of Section IV as the number of fading states increases. The gain matrices for each of these fading states were created independently, by using the model of (6), with $\sigma = 8$ dB. The uniform capacity converges at the value

¹Since the load of calculations required for the determination of the uniform capacity increases as more states are added, we are practically limited to around 40 states for protocols that allow concurrent transmissions. In all cases, we were able to determine the limit with a tolerance of 5% or better.

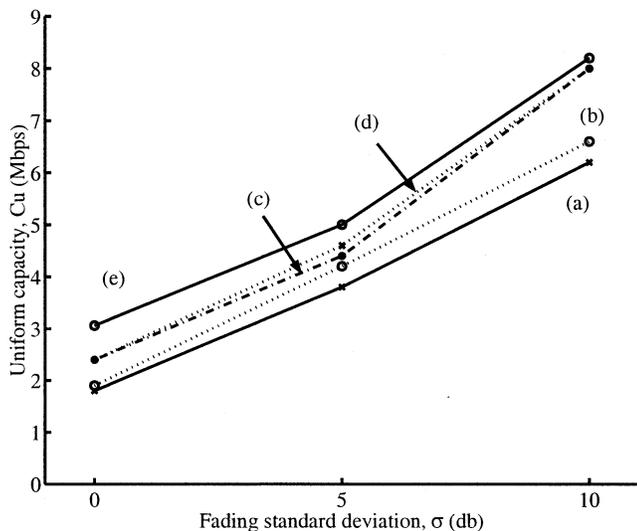


Fig. 9. Estimates of the uniform capacity of various protocols, for a ring ad hoc network with five nodes, as a function of the standard deviation σ of the shadowing component. Estimates are accurate within 0.2 Mb/s. (a) Single-hop routing, no spatial reuse. (b) Multihop routing, no spatial reuse. (c) Multihop routing with spatial reuse. (d) Two-level power control added to (c). (e) SIC added to (c).

$C_u^c = 5.9$ Mb/s which is 120% larger than the uniform capacity with no time-varying flat-fading and $\sigma = 0$ dB, $C_u^c = 2.7$ Mb/s. This value is also 90% larger than the value of the uniform capacity $C_u^c = 3.1$ Mb/s, achieved when there is time-varying flat-fading with $\sigma = 8$ dB, but the nodes do not schedule their transmissions across different fading states, and pick an optimal schedule within each fading state (i.e., as if the current fading state is the only fading state available). The 90% increase is possible because nodes are willing to tolerate large delays.

In Fig. 9, we plot the uniform capacity as a function of the shadowing standard deviation σ , in the case of a ring ad hoc network with five nodes and in the presence of time-varying flat-fading. We note that the uniform capacity increases with the standard deviation, which is a measure of the severity of the fading. This result should actually be expected, since in the presence of fading the network has more degrees of freedom when deciding on the optimal transmission schedule. From another perspective, each fading state will be more favorable for certain combinations of transmissions, and the network will use this state for these combinations.

We note that allowing SIC in addition to spatial reuse yields limited gains when fading is severe ($\sigma = 10$ dB). This is because under severe fading for most of the time, the network has the opportunity to operate using simultaneous transmissions with very little cross interference. In these cases, SIC is ineffective because it imposes very stringent upper bounds on the rates of the interferers.

B. Node Mobility

We can study node mobility using the framework developed for the case of time-varying flat-fading channels. Indeed, the critical element of both settings is that the system alternates between a large number of different channel gain matrices.

Consider an ad hoc network of n nodes A_1, A_2, \dots, A_n where the channel gains are deterministic and depend only on

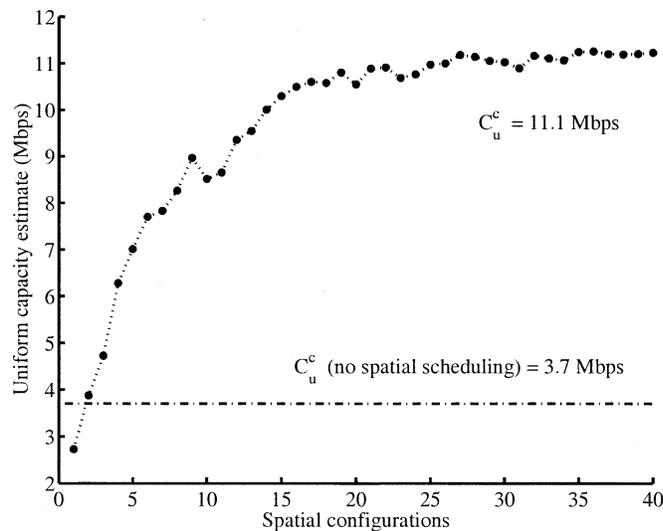


Fig. 10. Convergence of the estimate for the uniform capacity C_u^c with node mobility, for the example ad hoc network of Section IV, with $\sigma = 0$ dB.

distances (for example, path gains are given by (6) with $\sigma = 0$ dB) but the spatial configuration alternates between M different states. Assuming that, after long periods of time, each state is active for a fraction of time equal to $1/M$, the capacity region of the network is again given by (13), where each of the M sets of rate matrices $\{R^{ij}, j = 1, \dots, J(i)\}$ does not describe a different fading state, but a different spatial configuration.

The assumption that there is a finite number M of configurations is rather artificial, and a more realistic model would be that nodes move continuously. This would in turn imply that the number of possible states (and, hence, the number of rate matrices) is infinite, and as a consequence, the developed formulation is insufficient. However, the capacity region can be estimated by using a large, but finite, number of spatial configurations, in the same way as a finite number of fading states was used in the case of time-varying flat-fading. Specifically, let us assume that the vector stochastic process describing the node movement is stationary and ergodic (for example, two-dimensional independent Brownian walks constrained to a box). Then, instead of using all states, we can use a large number of independent realizations of the node positions to estimate the capacity region.

In Fig. 10, we plot the estimate for the uniform capacity versus the number of spatial configurations used for an ad hoc network of five nodes. All parameters are set as in the example ad hoc network of Section IV with $\sigma = 0$ dB (and no time-varying flat-fading). The transmission protocol allows multiple hop routing and spatial reuse, but no power control or SIC. The node movements are modeled as independent two-dimensional standard Brownian motions constrained to the box $\{-10 \text{ m} \leq x \leq 10 \text{ m}, -10 \text{ m} \leq y \leq 10 \text{ m}\}$, so we can create a realization of the process at a fixed time by distributing the nodes randomly, uniformly, and independently in the box. The uniform capacity converges to the value $C_u^c = 11.1$ Mb/s, after including around 40 spatial configurations. In the figure, we also plot the uniform capacity that is achieved by the network if scheduling across different spatial configurations is

not used, so that at any time instant nodes are using only the current configuration in the most efficient way to communicate, and are not exploiting node mobility. In this case, the capacity is $C_u^c = 3.7$ Mb/s, and the 200% increase achieved by using scheduling across spatial configurations is only possible if nodes are willing to tolerate large delays. This exchange of delay with throughput, under node mobility, is reminiscent of the results in [19].

VIII. ENERGY-CONSTRAINED NETWORKS

We now apply the methodology of the previous sections to develop a formulation suitable for the study of energy-constrained networks. In such networks, nodes have a finite amount of energy that they can use for transmitting or receiving [13]. As a consequence, continuous communication with a constant nonzero rate is not possible, but rather the goal is to maximize the total number of bits communicated.

A. System Model

We consider a network of n nodes A_1, A_2, \dots, A_n , described by the system model of Section II. For simplicity, we assume that A_n acts as a base station, and the rest of the nodes wish to send information to it (modifying the formulation to treat the downlink case or the more general ad hoc case is straightforward). In addition, each node $A_i, i = 1, \dots, n-1$ has some finite amount of energy E_i to use for its communication needs. Let $E = [E_1 \ E_2 \ \dots \ E_{n-1}]^T$ be the **energy vector** of the network.

We make the assumption that, under a given transmission scheme \mathcal{T} , when node A_i transmits to node A_j with power P_i for t seconds (so that the dissipated energy is $E_i = P_i t$), the number of bits transmitted is $B = f(\gamma_{ij})t$, where

$$\gamma_{ij} = \frac{G_{ij}P_i}{\eta_j W + \sum_{k \in \mathcal{T}, k \neq i} G_{kj}P_k}$$

is the link SINR and the function $f(\cdot)$ defines the achievable data rate (in bits per second) for the link SINR. We assume that the relation $B = f(\gamma_{ij})t$ holds true for any duration t , which implies that the achievable data rate is independent of the number of bits transmitted. Our formulation is not consistent with previous information theoretical approaches [20], [21]; however, it is well justified for any practical modulation and coding strategy, as long as the number of bits transmitted is much greater than the bits per symbol (for uncoded modulation), the block length (for block-coded modulation), or the memory of the convolutional coder (for convolutional coding).

Note that nodes expend energy not only for transmitting, but also for receiving. We make the assumption that the energy spent to decode a sequence of bits is proportional to its length.

Our transmission strategy aims to maximize the number of bits communicated between the nodes and does not consider delay. Thus, since any transmitting node will eventually expend its energy and become silent, there is no gain by allowing concurrent transmissions: If there are nonzero cross-channel gains, some transmissions will experience interference and

fewer than the maximum possible bits will be transmitted. If all cross-channel gains are zero, spatial reuse will not hurt, but it will not produce any gains either. Therefore, for our capacity analysis, we assume that no concurrent transmissions take place in the network. On the other hand, using multiple hops still makes sense and can augment the capacity region significantly.

B. Capacity Region Formulation

We still define transmission schemes as in Section III, and describe them in terms of rate vectors, as defined in Section VI. In addition, if $q_i(\mathcal{S})$ is the power consumed by node A_i during scheme \mathcal{S} , then we define the **power dissipation vector** of the scheme \mathcal{S} to be the vector $Q(\mathcal{S}) = [q_1(\mathcal{S}) \ q_2(\mathcal{S}) \ \dots \ q_{n-1}(\mathcal{S})]^T$. Nodes will typically need most of their power for transmission; however, the above definition allows us to factor in the power required to receive signals.

We assume that the network operates under a transmission protocol that allows schemes $\mathcal{S}_1, \mathcal{S}_2, \dots, \mathcal{S}_N$, corresponding to the rate vectors R_1, R_2, \dots, R_N and power dissipation vectors Q_1, Q_2, \dots, Q_N . Based on the discussion of the previous subsection, we assume that the transmission protocol does not permit concurrent transmissions. Depending on the particular communication requirements of the nodes, the network will operate under a succession of some of the N different schemes, each being active for a specific amount of time. With no loss of generality, we assume that each scheme will appear at most once in this succession (separate occurrences can be lumped together). Any succession of schemes is acceptable, as long as the energy dissipation associated with this succession satisfies the requirements. This leads to the following definition of the **capacity region** of the network:

$$C = C(\{R_i\}, \{Q_i\}) \triangleq \left\{ \sum_{i=1}^N t_i R_i : t_i \geq 0, \sum_{i=1}^N t_i Q_i \leq E \right\} \cap \mathcal{P}_{n-1}. \quad (15)$$

Contrary to the previous cases, the capacity region is not a convex hull, but it is trivial to check that it is still convex. Also, the coefficients $\{t_i\}$ do not represent time percentages, but actual times. The meaning of the capacity region is the following: If $B = \{b_i\} \in C$, with $B = \sum_{i=1}^N t_i R_i$, then there is a strategy under which each of the nodes A_j will send b_j bits of data to the base station. The strategy consists of using the schemes corresponding to the R_i s, each for exactly t_i seconds. Since in each scheme only one node will be active, we are guaranteed that the schemes can be temporally placed so that nodes used for forwarding data will receive the data before they are required to forward them.

As in the previous cases, the feasibility problem can be formulated as a linear program in the N -dimensional Euclidean space. Moreover, the linear program has no objective function. Point B belongs to the capacity region if and only if the following system has a solution $\{t_1, \dots, t_N\}$:

$$B = \sum_{i=1}^N t_i R_i, \quad E \geq \sum_{i=1}^N t_i Q_i. \quad (16)$$

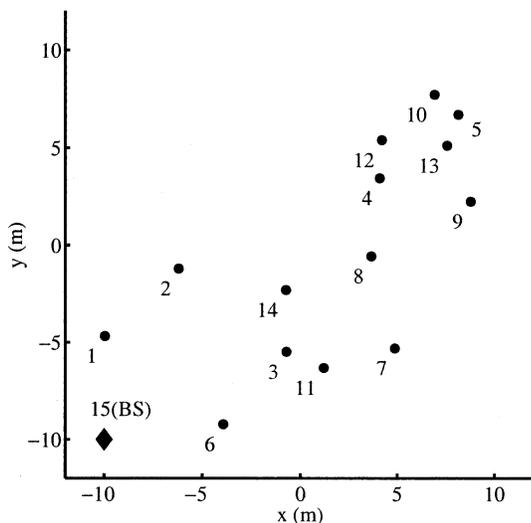


Fig. 11. Example energy-constrained network with 15 nodes.

C. A Numerical Example

In Fig. 11, we display an energy-constrained network with 15 nodes and a random topology, created as the example multihop cellular network of Section VI with the exception that now the base station is placed on the point $\{x = -10 \text{ m}, y = -10 \text{ m}\}$. All system parameters are chosen as in the network of Section VI. All nodes transmit with power $P_i = 0.1 \text{ W}$, but waste no power when they receive. The energy available to each node is $E_i = 10 \text{ J}$. We assume that transmitters use variable-rate M -QAM, with the requirement that the probability of symbol error is always less than 10^{-4} , and trellis coding with a coding gain of 3.5 dB is used. Under these assumptions, the achievable rate $f(\gamma_{ij})$ is well approximated [17] by

$$f(\gamma_{ij}) = W \log_2 \left(1 + \frac{\gamma_{ij}}{\Gamma} \right)$$

where the gap Γ is equal to $\Gamma = 3 \text{ dB}$. Γ is equal to the gap between uncoded M -QAM and capacity (6.5 dB), minus the coding gain (3.5 dB).

In Fig. 12, we plot a slice of two capacity regions along the plane $b_i = 0, i \neq 5, 10$. The capacity regions correspond to a single-hop and a multiple hop transmission protocol, with no spatial reuse, SIC, or power control.

In the single-hop case, the capacity region is rectangular, since in this case no forwarding is allowed, and the best (and only) strategy is for the two nodes to transmit sequentially to the base station. As expected, multihop routing greatly improves performance. As an example, the point on the boundary of the multihop capacity region for which $b_5 = b_{10}$ is achieved by a transmission strategy consisting of 23 transmissions that actually depletes the energy reserves of all nodes. This result is typical and shows that optimum strategies aggressively take advantage of all available resources to service all data streams. With this transmission strategy, both nodes send $b_5 = b_{10} = 5.60 \text{ Mb/s}$ to the base station, which is only slightly less than the number of bits either node would transmit if it were the only one using the network (7.62 and 7.45 Mb/s, respectively), although the two nodes are placed close to each other. This result can be explained by noting that there are

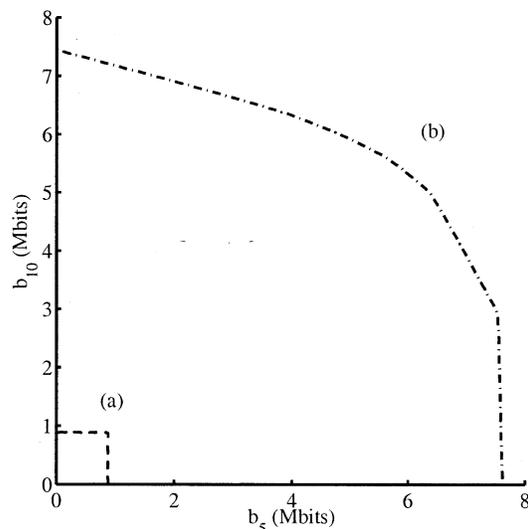


Fig. 12. Capacity region slices of the network of Fig. 11 along the plane $b_i = 0, i \neq 5, 10$. (a) Single-hop routing. (b) Multihop routing.

enough nodes between A_5, A_{10} , and the base station, such that two roughly independent paths to the base station can exist. We have arrived at similar results for various transceiver models, and for a wide range of modeling parameters.

IX. CONCLUSION

We have developed a mathematical framework for finding the capacity region of an ad hoc or multihop cellular wireless network under time-division routing and a given transmission protocol, possibly in the presence of time-varying flat-fading or node mobility. We use this framework to determine the network performance that can be obtained using various transmission protocols for a number of different network topologies. We show that multihop routing, spatial reuse, and SIC all lead to significant gains, but gains from power control are significant only if very limited or no rate adaptation is used. We also determine that fading and node mobility can actually improve network capacity. Finally, we introduce a formulation suitable for the study of energy-constrained networks.

A major limitation of our capacity formulation is the very fast increase of basic rate matrices as the number of nodes increases. As was discussed, the capacity region can be completely reproduced by a much smaller set of rate matrices. Based on this observation, a possible future research direction could be toward developing methods for determining a “good enough” subset of rate matrices, that will be manageable and also reproduce most of the capacity region. These methods may be deterministic (for example, requiring all such rate matrices to have a certain set of characteristics) or stochastic (for example, genetic algorithms). Such techniques will allow us to study capacity regions of larger networks, and if simple enough could also be integrated to the study of medium access control and routing protocols in ad hoc networks.

REFERENCES

- [1] *Special Issue on Advances in Mobile Ad Hoc Networking, IEEE Pers. Commun.*, vol. 8, Feb. 2001.

- [2] *Special Issue on Wireless Ad Hoc Networks, IEEE J. Select. Areas Commun.*, vol. 17, Aug. 1999.
- [3] F. A. Tobagi, "Modeling and performance analysis of multihop packet radio networks," *Proc. IEEE*, vol. 75, pp. 135–155, Jan. 1987.
- [4] P. Karn, "MACA: A new channel access method for packet radio," in *Proc. 9th Computer Networking Conf.*, London, ON, Canada, Sept. 1990, pp. 134–140.
- [5] V. Bharghavan, A. Demers, S. Shenkar, and L. Zhang, "MACAW: A media access protocol for wireless LAN's," in *Proc. ACM SIGCOMM*, London, UK, Aug. 1994, pp. 212–225.
- [6] *IEEE ComSoc LAN MAN Standards Committee Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications*, IEEE Standard 802.11, 1997.
- [7] C. L. Fullmer and J. J. Garcia-Luna-Aceves, "Floor acquisition multiple access (FAMA) for packet-radio networks," in *Proc. ACM SIGCOMM*, Cambridge, MA, Aug. 1995, pp. 262–273.
- [8] S. Xu and T. Saadawi, "Does the IEEE802.11 MAC protocol work well in multihop wireless ad hoc networks?," *IEEE Commun. Mag.*, vol. 39, pp. 130–137, June 2001.
- [9] C.-K. Toh, V. Vassiliou, G. Guichal, and C.-H. Shih, "MARCH: A medium access control protocol for multihop wireless ad hoc networks," in *Proc. IEEE MILCOM*, Los Angeles, CA, Oct. 2000, pp. 512–516.
- [10] A. Ephremides and T. V. Truong, "Scheduling broadcasts in multihop radio networks," *IEEE Trans. Commun.*, vol. 38, pp. 456–460, Apr. 1990.
- [11] E. M. Royer and C.-K. Toh, "A review of current routing protocols for ad-hoc mobile wireless networks," *IEEE Pers. Commun.*, vol. 6, pp. 46–55, Apr. 1999.
- [12] J. Broch, D. A. Maltz, D. B. Johnson, Y.-C. Hu, and J. Jetcheva, "A performance comparison of multi-hop wireless ad hoc network routing protocols," in *Proc. ACM/IEEE MOBICOM*, Dallas, TX, Oct. 1998, pp. 85–97.
- [13] G. J. Pottie, "Wireless sensor networks," in *Proc. IEEE Information Theory Workshop*, Killarney, Kerry, Ireland, June 1998, pp. 139–140.
- [14] A. Michail and A. Ephremides, "Energy efficient routing for connection-oriented traffic in ad-hoc wireless networks," in *Proc. 11th IEEE PIMRC*, vol. 2, London, UK, Sept. 2000, pp. 762–766.
- [15] P. Gupta and P. R. Kumar, "The capacity of wireless networks," *IEEE Trans. Inform. Theory*, vol. 46, pp. 388–404, Mar. 2000.
- [16] A. Goldsmith and S. Chua, "Variable-rate variable-power M -QAM for fading channels," *IEEE Trans. Commun.*, vol. 45, pp. 1218–1230, Oct. 1997.
- [17] M. V. Eyuboğlu and G. D. Forney, Jr., "Trellis precoding: Combined coding, precoding and shaping for intersymbol interference channels," *IEEE Trans. Inform. Theory*, vol. 38, pp. 301–314, Mar. 1992.
- [18] N. Bambos, "Toward power-sensitive network architectures in wireless communications: Concepts, issues, and design aspects," *IEEE Pers. Commun.*, vol. 5, pp. 50–59, June 1998.
- [19] M. Grossglauser and D. Tse, "Mobility increases the capacity of wireless ad hoc networks," in *Proc. INFOCOM*, vol. 3, Anchorage, AK, Apr. 2001, pp. 1360–1369.

- [20] R. Gallager, "Energy limited channels: Coding, multiaccess and spread spectrum," in *Proc. Conf. Information Science Systems*, Princeton, NJ, Mar. 1988, p. 372.
- [21] S. Verdú, "On channel capacity per unit cost," *IEEE Trans. Inform. Theory*, vol. 36, pp. 1019–1030, Sept. 1990.



Stavros Toumpis (S'98) received the Diploma in electrical and computer engineering from the National Technical University of Athens, Athens, Greece, in 1997, and the M.S. degrees in electrical engineering and in mathematics from Stanford University, Stanford, CA, in 1999 and 2002, respectively. He is currently working toward the Ph.D. degree at Stanford University.

From 1998 to 1999, he worked as a Research Assistant for the Mars Global Surveyor Radio Science Team, providing operational support. Since 2000, he is a Member of the Wireless Systems Laboratory, Stanford University, Stanford, CA. His research includes work in the capacity of wireless networks, medium access control, and channel modeling.



Andrea J. Goldsmith (S'90–M'93–S'93–M'95 – SM'99) received the B.S., M.S., and Ph.D. degrees in electrical engineering from the University of California, Berkeley, in 1986, 1991, and 1994, respectively.

From 1986 to 1990, she was affiliated with Maxim Technologies, where she worked on packet radio and satellite communications systems, and from 1991 to 1992, she was affiliated with AT&T Bell Laboratories, where she worked on microcell modeling and channel estimation. She was an Assistant Professor of Electrical Engineering at the California Institute of Technology, Pasadena, from 1994 to 1998, and is currently an Assistant Professor of Electrical Engineering at Stanford University, Stanford, CA. Her research includes work in capacity of wireless channels and networks, wireless communication theory, adaptive modulation and coding, multiantenna systems, joint source and channel coding, communications and control, and adaptive resource allocation for cellular systems and ad hoc wireless networks.

Dr. Goldsmith is a Terman Faculty Fellow at Stanford and a recipient of the Alfred P. Sloan Fellowship, a National Science Foundation CAREER Development Award, the Office of Naval Research Young Investigator Award, a National Semiconductor Faculty Development Award, an Okawa Foundation Award, and the David Griep Memorial Prize from the University of California, Berkeley. She is an editor for the IEEE TRANSACTIONS ON COMMUNICATIONS and the *IEEE Personal Communications Magazine*.