

Capacity of Time-Slotted ALOHA Packetized Multiple-Access Systems Over the AWGN Channel

Muriel Médard, *Senior Member, IEEE*, Jianyi Huang, *Student Member, IEEE*,
 Andrea J. Goldsmith, *Senior Member, IEEE*, Sean P. Meyn, *Fellow, IEEE*, and
 Todd P. Coleman, *Student Member, IEEE*

Abstract—We study different notions of capacity for time-slotted ALOHA systems. In these systems, multiple users synchronously send packets in a bursty manner over a common additive white Gaussian noise (AWGN) channel. The users do not coordinate their transmissions, which may collide at the receiver. For such a system, we define both single-slot capacity and multiple-slot capacity. We then construct a coding and decoding scheme for single-slot capacity that achieves any rate within this capacity region. This coding and decoding scheme for a single time slot combines aspects of multiple access rate splitting and of broadcast codes for degraded AWGN channels. This design allows some bits to be reliably received even when collisions occur and more bits to be reliably received in the absence of collisions. The exact number of bits reliably received under both of these scenarios is part of the code design process, which we optimize to maximize the expected rate in each slot. Next, we examine the behavior of the system asymptotically over multiple slots. We show that there exist coding and decoding strategies such that regardless of the burstiness of the traffic, the system is stable as long as the average rate of the users is within the multiple access capacity region of the channel. In other words, we show that bursty traffic does not decrease the Cover–Wyner capacity region of the multiple access channel. A vast family of codes, which includes the type of codes we introduce for the single-slot transmission, achieve the capacity region, in a sense we define, for multiple-slot transmissions. These codes are stabilizing, using only local information at each of the individual queues. The use of information regarding other queues or the use of scheduling does not improve the multiple-slot capacity region.

Index Terms—ALOHA, broadcast, bursty data, capacity, coding, multiple access, rate splitting.

I. INTRODUCTION

THE flexibility of ALOHA systems, which were first proposed in 1970 by Abramson [1] makes such systems an attractive option for wireless applications, such as data transfer

Manuscript received February 21, 2002; revised September 28, 2002 and November 15, 2002; accepted January 4, 2003. The editor coordinating the review of this paper and approving it for publication is L.-C. Wang. This work was supported in part by the National Science Foundation under grants CCR 99-79381, ITR 00-85929, Career 6891730, and fellowship NSF CCR 0093349, by ONR Young Investigator Award N00014-99-1-0578 and under ONR Grant N00014-99-1-0698, and by the National Reconnaissance Office.

M. Médard and T. P. Coleman are with the Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA 02139 USA (medard@mit.edu).

J. Huang and S. P. Meyn are with the Department of Electrical and Computer Engineering and the Coordinated Sciences Laboratory, University of Illinois, Urbana-Champaign, IL 61801 USA (jhuang7@uiuc.edu).

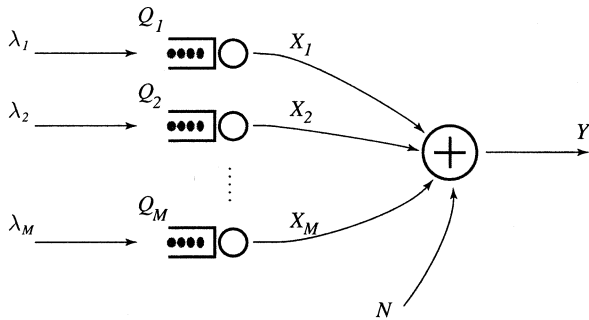
A. J. Goldsmith is with the Department of Electrical Engineering, Stanford University, Stanford, CA 94305 USA (andrea@systems.stanford.edu).

Digital Object Identifier 10.1109/TWC.2003.821175

for nomadic computing. In the original ALOHA system, users transmit packets without any knowledge of other user's current transmissions. If a collision among packets occurs at the receiver, those packets are discarded and the users retransmit. The capacity of ALOHA systems and related collision systems has generally been considered in terms of packet throughput [7], [41], [49]. The stability region of the ALOHA system has been extensively studied. For an infinite number of users, the system is unstable for any input rate [19]. For a finite number of users, there exist bounds and some exact results only for the two-user case or particular arrival processes [6], [51]. Several different control mechanisms have been established to stabilize the operation of the ALOHA system [30], [54] or to perform conflict resolution [19], [27], [32], [34], [37], [40], [45], [48], [65].

In order to avoid total loss of packets to collisions, several coding schemes have been proposed for ALOHA packets [14], [15], [17], [18], [33], [36], [46], [50], [52], [61]. The purpose of such coding is to allow at least part of the data in the packets to be received correctly despite collisions. One example is *spread ALOHA* in which users appear as interference to each other in the event of a collision. However, when careful consideration is given to the dimensions required to spread users, spread ALOHA has been shown in [62] to be detrimental to capacity (as compared to coding without spreading). Bursty multiple-access communications have been considered in [57], but the purpose there is to avoid retransmissions altogether.

The paper is organized as follows. In Section II, we describe a simple time-slotted packetized ALOHA-style model. In Section III, we consider the issue of how to code over a single time slot. We propose a coding scheme that combines multiple-access rate splitting concepts and broadcast codes for degraded additive white Gaussian noise channels. This scheme was presented in part in [42] and [43]. In effect, we create from each user several pseudo-users. Rather than discard all data received during a collision or code sufficiently to receive all the data even in the event of a collision, our scheme enables a tradeoff between the rates obtained in the event of a collision and those obtained when no collision occurs. In the case of no collision, the data from all the pseudo-users corresponding to the single transmitting user is correctly received. In the case of a collision, only a subset of the transmitted data from a user corresponding to the data for a subset of his pseudo-users can be received correctly. The manner in which the energy of each user is split among its corresponding pseudo-users determines the tradeoff between the rates achievable under collisions and

Fig. 1. M -user ALOHA model.

under no collisions. We examine how to optimize the rates allocated to the pseudo-users when we seek to maximize the total expected rate at the receiver for a given traffic arrival distribution in a single slot. We show that the rate benefit of splitting one user into several pseudo-users arises mainly in the case of users with highly asymmetric energies.

In Section IV, we consider the long-term behavior of the system, using a particular definition of capacity for the case of an arbitrarily large number of slots. We show that there exists a family of codes for which the system is stable as long as the average rates of the users are within the multiple access capacity region (the Cover–Wyner rate region [23]). Thus, despite the burstiness of the system, the capacity of the ALOHA system is the same as that of a multiple-access system in which both users transmit continuously. Moreover, the use of information by users or by any other entity, such as a controller or scheduler, of the state of other users' queues does not improve the achievable rate region. We present our conclusions and directions for further research in Section V.

II. MODEL AND BACKGROUND

We consider a multiple-access system, as illustrated in Fig. 1, where M users transmit to a single receiver. The model is time slotted, and each time slot is of length T time units. All the users share a single channel with no multiplicative attenuation but which is corrupted by additive white Gaussian noise (AWGN).

Data arrives at a given user for transmission in the form of fixed-length packets, where different users may have different packet lengths. We define $\Delta_i(j)$ to be the number of bits that arrive at user i in the j th time slot, $j \in \mathbb{Z}_+$. We assume that at most one packet arrives for transmission in each time slot. Thus, $\Delta_i(j)$ equals N_i or zero, where N_i is the number of bits in the i th user's packet. The arrival of packets for transmission at a user is determined by a Bernoulli process in which, at the start of each time slot, a user i receives a new packet with probability p_i and no packets with probability $1 - p_i$. The arrival streams of the various users are mutually independent processes. In particular, the sequence $\{\Delta_i(j) : k \in \mathbb{Z}_+\}$ is i.i.d. with mean $E[\Delta_i(j)] = L_i$, $i = 1, \dots, M$, and the average rate of bits to user i is $\lambda_i := p_i L_i$ (bits per second). The vector $\underline{\lambda}$ is the M -dimensional vector of arrival rates, $\underline{\lambda} = (\lambda_1, \dots, \lambda_M)$.

Moreover, we assume that the moment generating function is finite in a neighborhood of the origin; there exists $\delta > 0$, and $B < \infty$, such that

$$E[\exp(\delta \Delta_i(j))] \leq B \quad (1)$$

for all $j \in \mathbb{Z}_+$, $i = 1, \dots, M$. We discuss in Section VI relaxations of these assumptions.

Let $Q_i(j)$ be the number of bits in buffer i at the beginning of time-slot j , and let $Q(j) = (Q_1(j) \dots Q_M(j))'$ denote the vector of buffer levels. Each user sends data from his buffer. Once a user receives a packet for transmission, the data in that packet is queued at the user and that data is transmitted according to certain policies, which we describe later. Each user i knows the traffic awaiting transmission in its own queue. Moreover, user i may possibly have information $\mathcal{I}_i(j)$ at time j regarding the contents of the queues of other users. A user's queue contains all of its data that has not yet been successfully transmitted, i.e., data that was never transmitted or that was transmitted and received in error due to collisions. Each user has an infinite queue to hold data awaiting transmission.

At the start of a time slot, a user decides whether to transmit over that time slot, and, if he transmits, what data to transmit and how to encode that data. Packets may be transmitted as whole packets or partial packets. Coding is subject to the following constraints.

- 1) For each transmission over a time slot, user i is restricted to using a certain maximum energy σ_i^2 over that time slot.
- 2) User i , if he transmits, must transmit over the whole time slot and use codes of length T .
- 3) At the j th time slot, user i bases his transmission strategy on the contents of his own queue ($Q_i(j)$) and, if known, on information $\mathcal{I}_i(j)$ about other users' queues.
- 4) To ensure synchronization at the receiver, we assume that the receiver has perfect timing information.
- 5) The receiver knows for each user, at each time slot, whether or not that user is transmitting, for instance through identifying tags on the packet transmissions. The absence of a tag for a user indicates the absence of that user. We do not explicitly include such tags or their associated overhead in our model.

Assumption 5) is reasonable if the tags are sufficiently coded, so that they will withstand interference from all users simultaneously. Moreover, our model assumes there is no cost, in terms of time and energy, associated with tag transmission. This is reasonable if the length of a time slot T is very large and the number of users is moderate, so that the bits required for encoding the user identifiers are negligible with respect to the total number of bits in a time slot.

Detailed information regarding queue lengths is typically important in network design [10]. This is particularly important for scheduling, but we do not consider scheduling in this paper. In Section IV, we find that the i th user cannot improve capacity by using information $(Q_k(j-1) : k \neq i)$ about other users' queues.

We do not make here explicit assumptions about what information the senders have about the success of their transmissions. These assumptions are not necessary in Section III, where

a single slot is considered. They become relevant in the case of several slots, as discussed in Section IV. The main issue to consider is whether users are aware of what data was lost to collisions at the receiver. The affect of different assumptions regarding the knowledge of the data lost to collisions will be discussed in Section IV.

We may now present our model for the transmitted and received signals. The receiver receives the sum of the transmissions of the users and the AWGN. The signal transmitted by user i is X_i .

The output of the system at the receiver is Y , given at time unit t by

$$Y[t] = \sum_{i=1}^M X_i[t] + N[t] \quad (2)$$

where the $N[t]$ s are i.i.d. Gaussian and are independent of the processes $\{X_i\}$. The common variance of the $\{N[t]\}$ is denoted σ_N^2 .

At each time unit t the $\{X_i\}$ are mutually independent, conditioned on queue information at the different users. We term collision the event where more than one user transmits in a single time slot.

We may now describe the notion of capacity we consider in this paper to analyze the behavior of our system. We assume the time slots to be long enough that rates arbitrarily close to capacity may be achieved over a single time slot. The specific definition of this capacity for our ALOHA channel is given below. The notion of long time slots is the same as for a single user, where rates arbitrarily close to the single-user Shannon capacity can be achieved for codes with a sufficiently long block length (which corresponds to one time slot in our model). User i in time slot j sends one codeword each from a set of K_j^i codebooks $\mathcal{M}_j^{i,\kappa}$, $\kappa = 1, \dots, K_j^i$. We denote the single-slot capacity for user i in slot j , defined below, as λ_j^i , and let $\underline{\lambda}^j = (\lambda_j^1, \dots, \lambda_j^M)$. The ordered set of codebooks $(\mathcal{M}_j^{i,\kappa})_{\kappa=1}^{K_j^i}$ is called the codebook \mathcal{C}_j^i for user i in slot j . We say that the codebook $(\mathcal{C}_j^1, \dots, \mathcal{C}_j^M)$ achieves the single-slot capacity $\underline{\lambda}^j$ in slot j for slot-length T and error probability ξ (is $(T, \xi, \underline{\lambda}^j)$ single-slot capacity achieving) if for some sets $\mathcal{K}_j^1 \subseteq 1, \dots, K_j^1, \dots, \mathcal{K}_j^M \subseteq (1, \dots, K_j^M)$ known to both the transmitter and receiver there exists a decoding policy such that:

- 1) every codeword from a codebook $\mathcal{M}_j^{i,\kappa}$ where $\kappa \in \mathcal{K}_j^i$ is decoded with probability of error ξ or less;
- 2) the rate associated with that codeword transmission equals the single-slot capacity, thus for $i = 1, \dots, M$

$$\sum_{\kappa \in \mathcal{K}_j^i} \frac{\log(|\mathcal{M}_j^{i,\kappa}|)}{T} \geq \lambda_j^i.$$

A codeword that was decoded with probability ξ or less is considered to have been reliably received. We say that a codebook satisfying the conditions outlined above is $(T, \xi, \underline{\lambda}^j)$ single-slot capacity-achieving. Note that this definition differs from the standard capacity definition in that on slot j each user need not send *any* codeword in its codebook \mathcal{C}_j^i with arbitrarily small probability; he need only send a subset of his codewords

with arbitrarily small probability. This subset corresponds to a rate below the maximum associated with the full codebook, to allow for a lower rate to be reliably received in the event of a collision.

We now define multiple-slot capacity based on this single-slot capacity definition. Assume we now transmit over n slots. For a given T and $\xi > 0$, a coding and decoding policy is $(T, \xi, \underline{\lambda})$ capacity-achieving if $\forall i, \forall j, \exists \mathcal{C}_j^i$ that is $(T, \xi, \underline{\lambda}^j)$ single-slot capacity achieving and

$$\lim_{n \rightarrow \infty} \frac{1}{nT} \sum_{j=1}^n \lambda_j^i \geq \lambda_i \quad 1 \leq i \leq M. \quad (3)$$

This notion of capacity is related to other formulations of capacity with a delay constraint or probability of failed transmission, such as delay-limited capacity [31], ϵ -capacity [66], capacity versus outage [16], [47], [59], [60] and expected capacity [24], which itself may be viewed in the context of compound channels [13], [67]. Our own model can be viewed in terms of a compound multiple-access channel. Overviews of these other types of capacities can be found in [10] and [13].

The meaning of capacity-achieving policies as defined above is clear in the context of error exponents. Suppose that for every coding strategy in the policy and every possible arrival pattern in a slot, every user has an error exponent, bounded by $\epsilon > 0$. Error exponents [26] for multiple access channels [39] can be used to find, for large enough T , a lower bound to possible values of ξ of the form $\xi = e^{-\epsilon T}$. Averaging over multiple time slots would yield better error probabilities, similar to those obtained for fading channels, where the coding exponent reflects averaging over several fades.

Under the assumptions spelled out in this section we may obtain strict bounds on the set of achievable rates. Let us suppose that $p_i = 1$ for $i = 1, \dots, M$. We are then in the case where users always have traffic to transmit and each user is aware of whether the other user has traffic to transmit. Then, as long as the vector of arrival rates for the M users $\underline{\lambda} = (\lambda_1 \dots \lambda_M)$ is strictly inside the multiple access achievable rate region [5], [39], we expect the rates to be achievable. This is the basis of the coding schemes considered in the next section.

III. CODING OVER A SINGLE TIME SLOT

In this section, we examine coding over a single time slot. We illustrate the two-user case where each buffer is known to be *empty* at the previous time slot. For the analysis in this section, any traffic to be transmitted at the current time slot can only come from the arrival of new packets in the previous time slot. In the rest of this section, we refer to the rate over a single slot of length T . We assume that each user receives packets/bits to transmit at a rate corresponding to its single-slot capacity. This assumption can readily be relaxed by placing an additional constraint on the capacity of a given user.

We combine concepts from *multiple-access communications* [5], [39]; *broadcast channels* [20]–[23]; and *rate splitting* [28], [53]. The basic idea behind this approach springs from the following observation. In multiple-access channels, capacity is achieved through rate splitting. This involves first constructing *virtual users* that share available power and that

transmit independently. The receiver then decodes the received signals consecutively, so that some users are regarded as noise to other users during decoding. After a user is decoded, the user's contribution to the signal is eliminated, and the noise for the remaining undecoded signal is reduced. A similar approach is taken to achieve capacity in the degraded AWGN broadcast channel. For broadcast AWGN channels, we superimpose two codes, a low-resolution and a high-resolution code. The low-resolution code is decoded by considering the high-resolution code as noise. Once the low-resolution code is decoded, its contribution is eliminated. Hence, there is a similarity between the decoding mechanism used to achieve capacity in multiple-access channels and that used in degraded broadcast channels. In the system, we consider a user code to transmit over two possible channels: a channel with the other user present and a channel without the other user. Thus, our problem bears some traits of both degraded broadcast channels and of multiple-access channels.

For the model considered here, rate splitting is used to map out all points in the single-slot capacity region by splitting the signal of a given user into two signals corresponding to two different pseudo-users. Let us consider a specific example, where user 1 is split into two independent pseudo-users, U_1' and U_1'' , which send sequences of independent Gaussian signals with variance $\beta\sigma_1^2$ and $(1-\beta)\sigma_1^2$, respectively. In this example, we assume that there is no rate splitting for user 2, which maps to a single user U_2 . As in broadcast channels, each of the users we have constructed sends two messages on two separate signals. That is, U_1' sends a low-resolution signal LR_1' and a high-resolution signal HR_1' , which are independent Gaussian i.i.d. sequences with variance $\alpha_1'\beta\sigma_1^2$ and $(1-\alpha_1')\beta\sigma_1^2$, respectively. U_1'' sends low-resolution signal LR_1'' and high-resolution signal HR_1'' , which are independent sequences of i.i.d. Gaussian signals with variance $\alpha_1''(1-\beta)\sigma_1^2$ and $(1-\alpha_1'')(1-\beta)\sigma_1^2$, respectively. U_2 sends low-resolution signal LR_2 and high-resolution signal HR_2 , which are independent sequences of i.i.d. Gaussian signals with variance $\alpha_2\sigma_2^2$ and $(1-\alpha_2)\sigma_2^2$, respectively. Each $\alpha_1, \alpha_2, \beta$ lies in $[0, 1]$. These values can be optimized, which we do in the next section, to maximize average rate. Fig. 2 illustrates this coding scheme, including the rate splitting into pseudo-users and the multiple resolution signaling for each user or pseudo-user. The notations LR and HR are the abbreviations of the low and high resolution, respectively, since we are in effect using a broadcast code within our multiple access scheme.

We decode signals one after the other in the following order:

First LR_1' , then LR_2 , LR_1'' , HR_1'' , HR_2 , and finally HR_1' . (4)

If one of the six signals is not present, the receiver proceeds to the next one. Each signal is decoded so that all signals not yet decoded are considered noise, and signals that have been decoded and reconstructed are cancelled. Here, we assume every low-resolution codeword can be decoded with arbitrarily small probability of error regardless of collisions. The error associated with the high-resolution codewords will depend on whether or not there are collisions.

We may now present the three possible cases that arise and the corresponding decoding rules. Each signal of the LR and HR

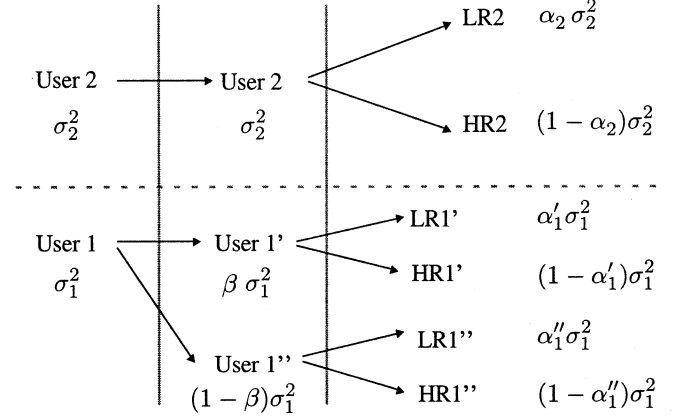


Fig. 2. Representation of the coding scheme.

type has a rate such that it can be decoded within the required probability of error if the SNR is at least

$$\begin{aligned} & \frac{\alpha_1'\beta\sigma_1^2}{\sigma_2^2 + (1-\alpha_1'\beta)\sigma_1^2 + \sigma_N^2} && \text{for } LR_1' \\ & \frac{\alpha_2\sigma_2^2}{\sigma_2^2(1-\alpha_2) + (1-\alpha_1'\beta)\sigma_1^2 + \sigma_N^2} && \text{for } LR_2 \\ & \frac{\alpha_1''(1-\beta)\sigma_1^2}{\sigma_2^2(1-\alpha_2) + (1-\alpha_1'\beta - \alpha_1''(1-\beta))\sigma_1^2 + \sigma_N^2} && \text{for } LR_1'' \\ & \frac{(1-\alpha_1'')(1-\beta)\sigma_1^2}{(1-\alpha_1')\beta\sigma_1^2 + \sigma_N^2} && \text{for } HR_1'' \\ & \frac{(1-\alpha_2)\sigma_2^2}{\sigma_N^2} && \text{for } HR_2 \\ & \frac{\sigma_1^2\beta(1-\alpha_1')}{\sigma_N^2} && \text{for } HR_1'. \end{aligned}$$

Our coding and decoding scheme is defined so that all LR signals above will always have a sufficiently large SNR. These signals are, therefore, always received reliably. For the HRs, they will not have sufficient SNR if user 1 and user 2 send at the same time. If the minimum SNR is not met for any one of the HR signals, that signal is not decoded. We consider the following cases.

Case 1) Only user 2 transmits.

- First, we decode LR_2 , which yields a rate whose maximum is

$$(1/2) \log(1 + (\alpha_2\sigma_2^2/(\sigma_2^2(1-\alpha_2) + (1-\alpha_1'\beta)\sigma_1^2 + \sigma_N^2)))$$

- Next, we decode signal HR_2 , which yields a rate whose maximum is $(1/2) \log(1 + ((1-\alpha_2)\sigma_2^2/(\sigma_N^2)))$.

The total rate is the sum of the above two rates.

Case 2) Only user 1 transmits.

- First, we decode LR_1' , which yields a rate whose maximum is

$$(1/2) \log(1 + (\alpha_1'\beta\sigma_1^2/(\sigma_2^2 + (1-\alpha_1'\beta)\sigma_1^2 + \sigma_N^2))).$$

- Second, we decode signal LR_1'' , which yields a rate whose maximum is

$$\frac{1}{2} \log\left(1 + \frac{\alpha_1''(1-\beta)\sigma_1^2}{\sigma_2^2(1-\alpha_2) + (1-\alpha_1'\beta - \alpha_1''(1-\beta))\sigma_1^2 + \sigma_N^2}\right).$$

- Third, we decode the signal HR_1'' , which yields a rate whose maximum is

$$(1/2) \log \left(1 + \frac{((1 - \alpha_1'')(1 - \beta)\sigma_1^2 / ((1 - \alpha_1')\beta\sigma_1^2 + \sigma_N^2))}{\sigma_1^2} \right).$$

- Finally, we decode HR_1' , yielding a rate whose maximum is $(1/2) \log \left(1 + \frac{\sigma_1^2\beta(1 - \alpha_1')}{\sigma_N^2} \right)$.

The total rate is the sum of the above four rates.

Case 3) User 1 and 2 both transmit.

- First, we decode LR_1' , yielding a rate whose maximum is $(1/2) \log \left(1 + \frac{\alpha_1'\beta\sigma_1^2}{\sigma_2^2 + (1 - \alpha_1')\sigma_1^2 + \sigma_N^2} \right)$.
- Second, we decode LR_2 , which yields a rate whose maximum is

$$(1/2) \log \left(1 + \frac{\alpha_2\sigma_2^2}{\sigma_2^2(1 - \alpha_2) + (1 - \alpha_1')\sigma_1^2 + \sigma_N^2} \right).$$

- Third, we decode signal LR_1'' , which yields a rate whose maximum is

$$\frac{1}{2} \log \left(1 + \frac{\alpha_1''(1 - \beta)\sigma_1^2}{\sigma_2^2(1 - \alpha_2) + (1 - \alpha_1'\beta - \alpha_1''(1 - \beta))\sigma_1^2 + \sigma_N^2} \right).$$

The total rate for user 1 is the sum of the rates of LR_1' and LR_1'' . The total rate for user 2 is the rate LR_2 .

Case 4) Neither user transmits, so the total rate is zero.

We define the rate variable r_u^v as follows: u refers to the cases enumerated above, and v refers to the users. Thus, r_2^1 is the rate for user 2 when there is no interference, and r_2^3 is the rate for user 2 when there is interference. We have an achievable rate r_1^2 for user 1 when there is no interference and another achievable rate r_1^3 when there is interference.

Note that our arguments can easily be extended to more than two users. Suppose that we have M users. Every possible set of users is associated with a scenario that corresponds to only those users being present. There are $z(M) = \sum_{j=1}^M \binom{M}{j}$ possible such scenarios. We first split the M users into $2M - 1$ virtual users. Each of these virtual users is then encoded using a $z(M)$ -level coding strategy. For each possible scenario, a particular set of codes will be decoded. If there are m users present, then for each present virtual user $M - m + 1$ codes will be decoded. Our results demonstrate achievability of the proposed coding scheme but not its optimality, which would require showing some fashion of coding converse, possibly similar to the broadcast channel converse [9]. Shamai [60] has applied a broadcast approach, which generalizes [59] to the multiple-access case, to a general class of channels that subsumes our channel. His optimization yields the same results as our optimization. Our results show a particular coding and decoding scheme to achieve the results of [60] and provide an explicit means of achieving a tradeoff for the rates achievable in the absence or presence of a collision. The following section explores that tradeoff in the context of expected rate. An alternate approach, in which broadcast splitting is performed before multiple-access splitting, is given in Appendix A.

A. Maximizing the Total Expected Rate

From the above discussion, we see that determining a coding scheme to achieve capacity is equivalent to determining the values of β , α_1' , α_1'' , and α_2 . Given these parameters, it suffices

to select random codes satisfying WGN-like statistics. In this section, we seek to find β , α_1' , α_1'' , and α_2 to maximize the expected rate, given by

$$E(R) = p_1(1 - p_2)r_1^2 + p_2(1 - p_1)r_2^1 + p_1p_2(r_1^3 + r_2^3). \quad (5)$$

Note that we can optimize $E(R)$ without finding the parameters β , α_1' , α_1'' , and α_2 explicitly [59]. We carry out explicitly the optimization in terms of β , α_1' , α_1'' , and α_2 because our goal is not only to maximize mutual information for a certain coding approach, but also to show explicitly how to design this optimal coding strategy.

We define the following two variables:

$$k_1 = 1 - \alpha_1'\beta - \alpha_1''(1 - \beta), \quad k_2 = 1 - \alpha_2.$$

$k_i = 0$ means we put all the signals of user i in LR, and $k_i = 1$ means we put all the signals in HR. Let us consider all the possible cases for p_1 and p_2 . When $p_1 = p_2 = 0$, we have immediately that $E(R)_{\max} = 0$. We present the case in detail for $p_1 = 0$, $p_2 > 0$ and a similar analysis yields the solution when $p_1 > 0$, $p_2 = 0$. Although these cases reduce to a single user channel, since one of the two users transmits with zero probability, the solutions illustrate the manipulation in the more general case where $p_1 > 0$ and $p_2 > 2$. The solutions for these more general cases, which requires significantly more manipulation, are derived in Appendix B.

Let us examine in detail the boundary point $(p_1, p_2) = (0, p_2)$ with $p_2 > 0$. By the general expression for the total expected rate in (5), we obtain

$$\begin{aligned} E(R) &= p_2 r_2^1 \\ &= p_2 \left[\frac{1}{2} \ln \left(1 + \frac{\alpha_2 \sigma_2^2}{(1 - \alpha_1'\beta)\sigma_1^2 + (1 - \alpha_2)\sigma_2^2 + \sigma_N^2} \right) \right. \\ &\quad \left. + \frac{1}{2} \ln \left(1 + \frac{(1 - \alpha_2)\sigma_2^2}{\sigma_N^2} \right) \right]. \end{aligned}$$

Then we have the following:

- If $\alpha_2 = 0$, then $E(R) = (1/2) \ln \left(1 + \frac{\sigma_2^2}{\sigma_N^2} \right) p_2$.
- If $\alpha_2 \neq 0$, then

$$\begin{aligned} E(R) &\leq p_2 \left[\frac{1}{2} \ln \left(1 + \frac{\alpha_2 \sigma_2^2}{(1 - \alpha_2)\sigma_2^2 + \sigma_N^2} \right) + \frac{1}{2} \ln \left(1 + \frac{(1 - \alpha_2)\sigma_2^2}{\sigma_N^2} \right) \right] \\ &= \frac{1}{2} \ln \left(1 + \frac{\sigma_2^2}{\sigma_N^2} \right) p_2 \end{aligned}$$

and we achieve equality when $(1 - \alpha_1'\beta) = 0$.

Hence, the optimal total expected rate for this case is

$$E(R)_{\max} = \frac{1}{2} \ln \left(1 + \frac{\sigma_2^2}{\sigma_N^2} \right) p_2 \quad (6)$$

where the maximum is achieved when one of the following conditions is satisfied:

- $\alpha_2 = 0$, β , α_1' , α_1'' can be any value in $[0, 1]$;
- $\alpha_2 \neq 0$, $\beta = \alpha_1' = 1$, and α_1'' can take any value in $[0, 1]$.

In other words, all information from user 1 will be decoded first (see Fig. 2). Since user 1 does not send in this case, user 2 codes its message assuming user 1 does not exist. Expression (6) can be seen to be the value we would expect, since it is the capacity for an AWGN channel with noise variance σ_N^2 and signal variance σ_2^2 .

In a similar way, we may verify that, for $p_1 > 0$, $p_2 = 0$, the optimal total expected rate is

$$E(R)_{\max} = \frac{1}{2} \ln \left(1 + \frac{\sigma_1^2}{\sigma_N^2} \right) p_1 \quad (7)$$

where the maximum is satisfied when one of the following conditions is satisfied:

- $\beta \neq 1$, $\alpha_1'' \neq 0$, $\alpha_1' \beta = 0$, $\alpha_2 = 1$;
- $\beta = 1$, $\alpha_1' = 0$, α_1'' , α_2 can be any value in $[0, 1]$;
- $\alpha_1'' = 0$, $\alpha_1' \beta = 0$, α_2 can take any value in $[0, 1]$.

In other words, all information from user 2 will be decoded first (see Fig. 2). Since user 2 does not send in this case, user 1 codes its message assuming user 2 does not exist.

Theorem 1: For any $\xi > 0$, $\exists \tau \in \mathcal{N}$ such that $\forall T > \tau$, there exists a single coding policy for two users which is (T, ξ, λ_j^i) single-slot capacity-achieving for user $i = 1, 2$ over slot j whenever, for any $\alpha, \beta \in [0, 1]^2$

- i) $\lambda_j^1 = 0$ and $\lambda_j^2 \leq LR_2 + HR_2$
or
- ii) $\lambda_j^2 = 0$ and $\lambda_j^1 \leq LR_1' + HR_1' + LR_1'' + HR_1''$
or
- iii) $\lambda_j^1 \leq LR_1' + LR_1''$ and $\lambda_j^2 \leq LR_2$.

The constructive proof of this theorem follows immediately from our discussion. Note that the analysis in this section can readily be extended to the case where the queues are known to have some arbitrary contents at the beginning of the time slot.

B. Numerical Results

We conclude this section with a brief description of some numerical results in which values of k_1 and k_2 were computed which maximize the total expected rate $E(R)$. We consider the special case where $p_1 = p_2 = p$. The optimal values k_1 and k_2 are functions of the probability p .

In each experiment the noise power was fixed, $\sigma_N^2 = 1$, but the signal powers σ_i^2 , $i = 1, 2$ were varied to show the affect of varying SNR. The most interesting cases are found in the highly asymmetric case, with very different SNRs for different users.

Fig. 3 shows results in the symmetric case in which both users have the same transmission power, i.e., $\sigma_1^2 = \sigma_2^2 = \sigma^2$. In each symmetric example considered we found that the best policy is either $k_1 = k_2 = 0$ or $k_1 = k_2 = 1$ for each p . In each example, the coding policy switches at some value of p . Note that, as the SNR increases (σ_i^2 increases from 0.1, 10, 10^4 , 10^7 , and then 10^{10}) this switching point tends to $p = 1/2$. Thus, when the users have equal transmission probabilities and equal transmission powers, no rate splitting between LR and HR is required.

For highly asymmetric SNRs, for some region of p , it is necessary to split at least one user to HR and LR to achieve the maximum total expected rate. Fig. 4 shows five examples with highly asymmetric SNRs to approximately symmetric SNRs. In each of the first three examples there is a region of values of p for which k_1 is strictly between 0 and 1.

IV. STABILITY

Here, we establish stability of the model, as viewed as a discrete-time-controlled stochastic system. We restrict ourselves to

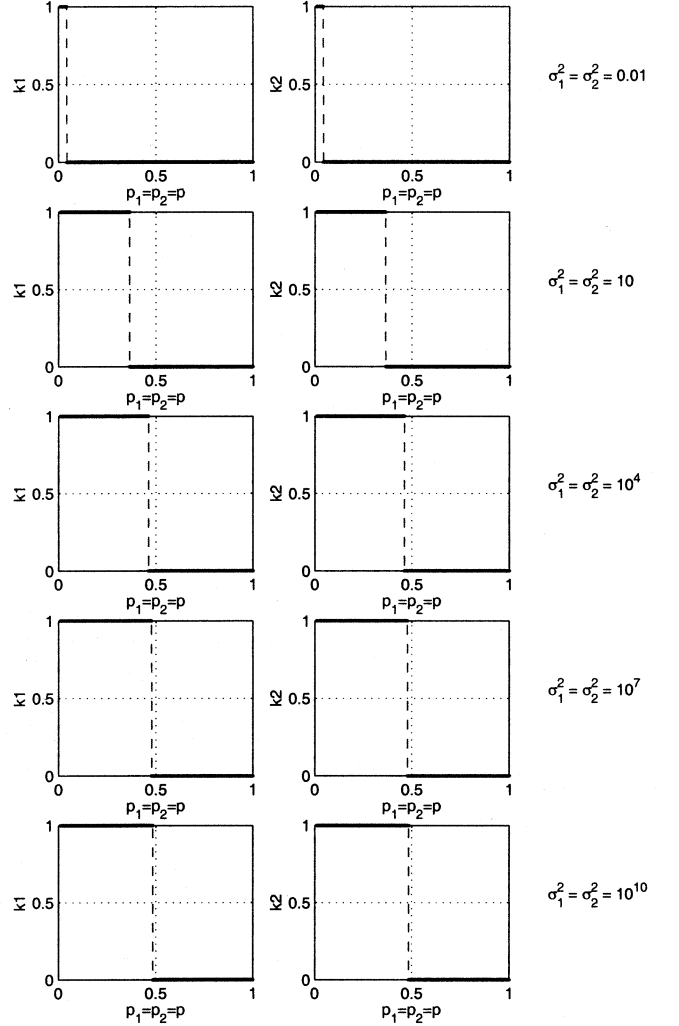


Fig. 3. k_1 and k_2 which maximize $E(R)$ in several symmetric cases: $\sigma_1^2 = \sigma_2^2 = 0.1, 10, 10^4, 10^7$, and 10^{10} , with $\sigma_N^2 = 1$.

the special case of two users, although more general situations can be readily obtained by similar methods.

Since we use coding to allow some bits to be transmitted even in the event of a collision, our analysis of stability is very different than traditional stability analysis of packetized ALOHA systems where collisions entail the loss of all packets involved [8], [25], [29], [30], [38], [55], [56], [58], [63], [64].

In regard to the dynamic model that includes bursty arrival stream to each user, there are two issues to be considered.

A. Impact of Variability

Burstiness is captured by p_i , the probability, at each user, that traffic arrives for transmission in a slot. Simple constructs show that no form of burstiness can improve the total rate. Specifically, consider first a two-user system with no burstiness, i.e., where $p_i = 1$ for all $i = 1, 2$. In that case, the system can be reduced to a system with burstiness if each user holds back transmissions in a probabilistic manner, thus mimicking the affect of bursty arrivals. Hence, burstiness cannot improve the total achievable rate $\sum_{i=1}^2 \lambda_i$, where the achievable rate vector $\underline{\lambda}$ is considered using the definition of Section II.

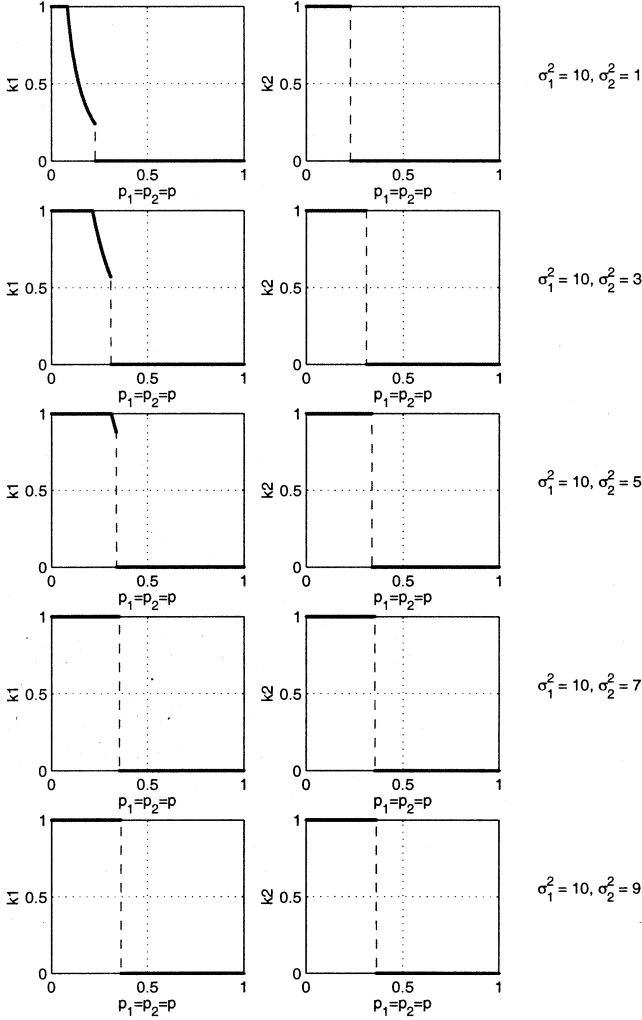


Fig. 4. k_1 and k_2 which maximize $E(R)$ in several asymmetric cases. User 2 has power $\sigma_2^2 = 1, 3, 5, 7,$ and 9 . In all cases, User 1 has power has power $\sigma_1^2 = 10$ and the noise has power $\sigma_N^2 = 1$.

We next find that burstiness cannot *reduce* the achievable rate, since users could average their arrivals over many time slots.

B. Role of Information

Consider a two-user system where, at each time slot, each user has perfect knowledge of the other user's queue state, in terms of bits awaiting transmission, as well as of his own. We obtain a bound on the achievable rate as follows.

Consider the maximum expected total achievable rate over any time slot, subject to the constraint that the probability of error in that time slot is upper bounded by ξ . Standard converse coding theorems establish upper bounds to the attainable rate. If both users transmit, and if $\xi > 0$ is sufficiently small, then this maximal rate is upper bounded by $(1/2) \log(1 + (\sigma_1^2 + \sigma_2^2)/(\sigma_N^2))$. Suppose that only one user, say user 1, transmits in a given time slot. In this case, the maximal achievable rate (for this user, over any individual time slot) is upper bounded by $(1/2) \log(1 + (\sigma_1^2)/(\sigma_N^2))$ for all ξ sufficiently small. Similarly, if only user 2 transmits, the maximal achievable rate (for this user, over any individual time

slot) is upper bounded by $(1/2) \log(1 + (\sigma_2^2)/(\sigma_N^2))$ for all ξ sufficiently small. Thus, for ξ small enough, the single-slot capacity region is upper bounded by the Cover–Wyner multiple-access rate region. Hence, the same converse holds for the multiple-slot capacity region, as stated in Theorem 2.

Theorem 2: There exists $\bar{\xi} > 0$ such that for all $0 < \xi \leq \bar{\xi}$, there exists $\delta > 0$ such that if there exists for some T a coding and decoding policy that is $(T, \xi, \underline{\lambda})$ multiple-slot capacity-achieving, then

$$\sum_{i \in \mathcal{S}} \lambda_i \leq \frac{1}{2} \log \left(1 + \frac{\sum_{i \in \mathcal{S}} \sigma_i^2}{\sigma_N^2} \right) + \delta \quad (8)$$

for all subsets \mathcal{S} of $1, \dots, M$.

In conclusion, we find that neither burstiness nor queue information can make the achievable rate region larger than the multiple-access Cover–Wyner rate region when ξ is small.

In order to prove a coding theorem that shows that the rates strictly in the interior of the Cover–Wyner region are feasible, it is necessary to show that the queue length process is stable whenever $\underline{\lambda}$ lies within the Cover–Wyner region. Theorem 3 below establishes stability.

Theorem 3: Suppose the vector of arrival rate $\underline{\lambda}$ is inside the multiple-access capacity region. Then, for some $\varepsilon > 0$, $B_\varepsilon < \infty$, and any $Q(0) \subseteq \mathbb{R}_+^2$

$$\limsup_{j \rightarrow \infty} E[\exp(\varepsilon \|Q(j)\|)] \leq B_\varepsilon. \quad (9)$$

Proof: We first establish the following version of Foster's criterion (see [44]). Let the Lyapunov function $V : \mathbb{R}_+^2 \rightarrow \mathbb{R}_+$ be defined as $V(x) = \max(x_1, x_2)$, $x \in \mathbb{R}_+^2$. Let $L_0 = \max_{i=1,2} [(1/2) \log(1 + (\sigma_i^2)/(\sigma_N^2))]$. We will show that for some $\varepsilon_0 > 0$, $b_0 < \infty$, and a bounded set $C_0 \subseteq \mathbb{R}_+^2$

$$E[V(j+s) | \mathcal{F}_s] \leq \begin{cases} V(Q(s)) - \varepsilon_0, & Q(s) \notin C_0 \\ b_0, & Q(s) \in C_0 \end{cases} \quad (10)$$

where $C_0 = \{x \in \mathbb{R}_+^2 : x_i \leq L_0, i = 1, 2\}$, and \mathcal{F}_s denotes the information, in terms of past contents of the buffers, about our system up to time slot s .

In order to establish (10), we consider three cases, corresponding to varying backlog at the two queues.

Case 1) $Q(s) \in C_0$. In this case, the bound is obvious for some $\varepsilon_0 < \delta$ under (1) [just take an expectation on both sides of the inequality (10)], regardless of C_0 .

Case 2) $Q_i(s) \gg 1$, for $i = 1, 2$. Given our definition of L_0 , both users have enough bits to send in the next time slot. Because λ is inside the capacity region, (10) holds.

Case 3) $Q(s) \notin C_0$, but only one component, $Q_1(s)$ or $Q_2(s)$, exceeds the threshold L_0 . In these cases, the maximum of $Q_1(s)$ and $Q_2(s)$ decreases even faster than it does in Case 2, so we preserve the drift inequality (10).

We now set $V_\varepsilon(x) = \exp(\varepsilon V(x))$, where $0 < \varepsilon < \delta$. In view of (10) and Lipschitz continuity of V , it can be shown, as in [44, eqn. 16.28], that the following geometric drift holds for some $\beta < 1$, $b < \infty$

$$E[V_\varepsilon(Q(s+1) | \mathcal{F}_s)] \leq \beta V_\varepsilon(Q(s)) + b.$$

This implies that for all s

$$E[V_\varepsilon(Q(s))] \leq \beta^s V_\varepsilon(Q(0)) + \frac{b}{1-\beta}.$$

□

Because $\underline{\lambda}$ lies in the multiple-slot capacity region, the system is stable, so the vector of rates $\underline{\lambda}$ is achievable.

There are two other cases to consider.

- i) If on a boundary of the multiple-access capacity region, we cannot determine the stability of the system.
- ii) If $\underline{\lambda}$ is outside the multiple-access capacity region, then for $\|Q(j)\| \rightarrow \infty$, as $j \rightarrow \infty$, the system is not stable in the strongest sense.

The arguments in this section can readily be extended to the case where we have more than two users. In effect, we have shown that the bursty nature of the data does not affect the reliably received rate, even though the transmitted rate may be affected.

To illustrate the implications of Theorem 3 we consider the following simple policy. When a user, say user 1, empties his buffer, he backs it up very significantly, according to the specific policy. Then, he transmits using a coding scheme that achieves a maximum total rate when both users transmit continuously over all time slots. User 2 follows the same scheme. For any $\xi > 0$, any $\epsilon > 0$, for all large enough T , we may find a policy of this form such that

$$\frac{1}{2} \log \left(1 + \frac{\sigma_1^2 + \sigma_2^2}{\sigma_N^2} \right) - \lambda_1 + \lambda_2 < \epsilon.$$

In effect, if we back up for long enough, the time spent transmitting when users are backed up dominates, and the time spent artificially backing up queues becomes negligible. Thus, this policy is optimal according to our definition of optimality, although it clearly is a poor choice in terms of delay.

We can immediately find that a policy where users transmit at all times as though they were in multiple access mode is optimal and trivially exhibits better delay characteristics than the family of policies mentioned above.

We have not yet discussed the information users need about past collisions. In the family of schemes described above, the information about past collisions is irrelevant, since collisions, which almost always occur, are taken into account in the coding when transmissions occur. On the other hand, for a scheme such as the one presented in Section III, the knowledge of past collisions allows retransmission of the HR components which may have been lost. This is in effect some form of automatic retransmission request (ARQ), albeit very different from the case where complete loss of data occurs in the case of a collision. In particular, let us assume each user knows the type of collisions that have occurred in the past, i.e., each user knows which users transmitted in past slots. The knowledge of past collisions provides not only ARQ but also partial knowledge of the queue, since the users know how many bits were not successfully transmitted. If retransmission of HR components does not take place, then we should consider a capacity definition centered around expected received rate rather than reliably transmitted rate (to within ξ probability of error).

From our discussion, we may state the following coding theorem companion to Theorem 2.

Theorem 4: For any $\xi > 0$ and sufficiently large T , there exists a coding and decoding policy which is $(T, \xi, \underline{\lambda})$ capacity achieving for all $\underline{\lambda}$ in the interior of the Cover–Wyner multiple access capacity region.

V. DELAY ISSUES

Our results indicate that there is a family of transmission strategies that achieve rates arbitrarily close to the multiple access channel capacity region boundary without sharing queue information and without adapting the strategy to the burstiness of the traffic. One cannot, however, interpret our results to mean that it is useless for users to have queue information or to adapt coding to traffic arrival characteristics, such as burstiness. Our analysis has not considered the issue of delay. Recent work has considered power and delay tradeoffs in fading channels [10]. However, in that analysis, users had perfect knowledge of each others' queues. Our results indicate that queue information and adaptive coding are not important from a capacity point of view, but it is reasonable to assume they are useful from a delay point of view. In particular, there exists a tradeoff between delay and energy. This tradeoff is explored in [10] for fading channels, but under the assumption of perfect knowledge of all the queues and centralized control. Such centralized control is not appropriate for our ALOHA-style model, in which transmissions are essentially uncoordinated. The tradeoff between delay and energy in the case of no or very limited queue information in the same system setting as this paper is investigated in [2]. Queue information need not be detailed and coding may not need to be designed to depend as closely on channel and queue knowledge as in Section III. Consider the following family of policies, partially explored in [2]. Each user i transmits a single bit when its queue passes above or below a certain threshold of bits, say sufficient to transmit the maximum single user rate over a single time slot. In a manner similar to that in Section III, the users adapt their coding to the known conditions of the other queues, for instance transmitting in multiple access mode when all users are above the queue threshold. This family of policies is optimal with respect to many slot capacity. For this family of policies, there is a tradeoff between, on the one hand, delay, in terms of time slots, and, on the other hand, energy, rate, queue information, and coding complexity. A topic of further interest when considering delay is determining to what extent placing constraints on average delay, for instance in terms of expected number of slots, affects our achievable rates.

VI. DISCUSSION AND CONCLUSION

We have considered the case of ALOHA systems where some measure of interference cancellation can occur at the receiver. We have shown one such scheme, for the case of a single time slot. This scheme works by combining notions from broadcast codes and rate splitting and adapts coding to the burstiness of the system. This method in effect relies on having several overlaid codes. When we consider many slots, we have seen that the capacity region, in the sense we defined, is the same as the Cover–Wyner multiple-access capacity region. Moreover, we have shown that there is a family of policies that achieve capacity without any queue information being shared among users

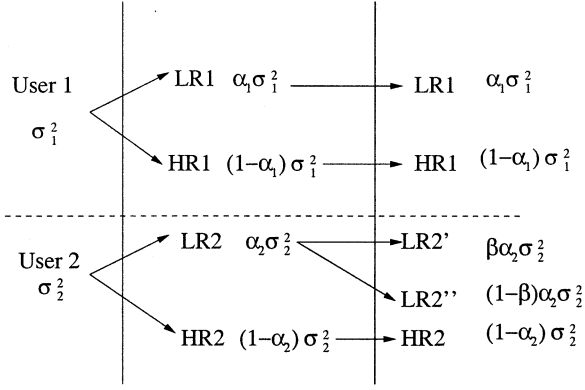


Fig. 5. Representation of the proposed new coding scheme.

and without adapting coding to burstiness of the system. In effect, our results indicate that burstiness does not affect capacity. Interestingly, this insight, without the capacity and coding constructs in this paper, has been present in some of Abramson's work for many years [3], [4].

Our results make several assumptions regarding data arrival, coding, lack of channel fading, etc. Several extensions are possible. For instance, we may consider different types of data arrival, with some constant streams and some bursty arrivals as in [18]. The coding methods can be readily adapted for different arrival distributions, and the stability results should only depend, for well-behaved distributions, on the average arrival rate of data. We may consider shorter time slots and coding over more than a single slot, either with interleaving or without. The results of Section III would then be changed considerably, but those of Section VI would not be affected, since we could create super slots by amalgamating time slots. Fading, particularly block-fading where fading intervals are an integer number of time slots and the fading parameters are known at the sender and the receiver, may be easily included in the coding method. Stability would then have to take into account the type of time variations of the channel. The values of ξ would also vary with fading and time slots.

APPENDIX A

In this approach, we reverse the order of the rate splitting and do not rate split multiple access for the high-resolution users. As in broadcast channels, each of the users we have constructed sends two messages on two separate signals. That is, U_1 sends signals LR_1 and HR_1 , which are independent WGN signals with variance $\alpha_1\sigma_1^2$ and $(1-\alpha_1)\sigma_1^2$, respectively. U_2 sends signal LR_2 and HR_2 , which are independent WGN signals with variance $\alpha_2\sigma_2^2$ and $(1-\alpha_2)\sigma_2^2$, respectively. The low-resolution component of user 1 is then divided into two independent virtual users for multiple access purposes. Hence, we have $U_{LR1'}$ and $U_{LR1''}$, which send independent WGN signals with variance $\beta_1\alpha_1\sigma_1^2$ and $(1-\beta_1)\alpha_1\sigma_1^2$, respectively. There is no rate splitting for the low-resolution component of user 2. Each $\alpha_1, \alpha_2, \beta_1$ lies in $[0, 1]$.

Fig. 5 illustrates this coding scheme.

The notations LR and HR are the abbreviations of *low resolution* and *high resolution*, respectively, since we are in effect

using a broadcast code within our multiple access scheme. We decode signals one after the other in the following order:

First LR_1' , then LR_2 , LR_1'' , HR_1 , and finally HR_2 .

If one of the five signals is not present, the receiver proceeds to the next one. Each signal is decoded so that all signals not yet decoded are considered noise, and signals that have been decoded and reconstructed are cancelled. Here, we assume the signal can be decoded with over a time slot of length T with probability of error less than or equal to ξ .

Each signal of the LR and HR type has a rate such that it can be decoded within the required probability of error if the SNR is at least

$$\begin{aligned} \text{SNR}'_{\tilde{L}R,1} &= \frac{\tilde{\alpha}_1\tilde{\beta}\sigma_1^2}{\sigma_2^2 + (1-\tilde{\alpha}_1\tilde{\beta})\sigma_1^2 + \sigma_N^2} & \text{for } \tilde{L}R'_1 \\ \text{SNR}_{\tilde{L}R,2} &= \frac{\tilde{\alpha}_2\sigma_2^2}{(1-\tilde{\alpha}_2)\sigma_2^2 + (1-\tilde{\alpha}_1\tilde{\beta})\sigma_1^2 + \sigma_N^2} & \text{for } \tilde{L}R_2 \\ \text{SNR}''_{\tilde{L}R,1} &= \frac{\tilde{\alpha}_1(1-\tilde{\beta})\sigma_1^2}{(1-\tilde{\alpha}_2)\sigma_2^2 + (1-\tilde{\alpha}_1)\sigma_1^2 + \sigma_N^2} & \text{for } \tilde{L}R''_1 \\ \text{SNR}_{\tilde{H}R,1} &= \frac{(1-\tilde{\alpha}_1)\sigma_1^2}{\sigma_N^2} & \text{for } \tilde{H}R_1 \\ \text{SNR}_{\tilde{H}R,2} &= \frac{(1-\tilde{\alpha}_2)\sigma_2^2}{\sigma_N^2} & \text{for } \tilde{H}R_2. \end{aligned}$$

Hence, the low-resolution and high-resolution rates that may be achieved are as follows:

$$\begin{aligned} R_{\tilde{L}R,1} &= \frac{1}{2} \log_2(1 + \text{SNR}'_{\tilde{L}R,1}) + \frac{1}{2} \log_2(1 + \text{SNR}''_{\tilde{L}R,1}) \\ R_{\tilde{H}R,1} &= \frac{1}{2} \log_2(1 + \text{SNR}_{\tilde{H}R,1}) \\ R_{\tilde{L}R,2} &= \frac{1}{2} \log_2(1 + \text{SNR}_{\tilde{L}R,2}) \\ R_{\tilde{H}R,2} &= \frac{1}{2} \log_2(1 + \text{SNR}_{\tilde{H}R,2}). \end{aligned}$$

It can be shown that any rate achieved with the first scheme given in the paper can be achieved with this scheme. Consider a set of allocations of $\alpha'_1, \alpha''_1, \alpha_2, \beta$ for the first scheme. We may choose $\tilde{\alpha}_1, \tilde{\alpha}_2, \tilde{\beta}$ so that

$$\begin{aligned} R_{\tilde{L}R,1} &= R_{LR,1} \\ R_{\tilde{H}R,1} &= R_{HR,1} \\ R_{\tilde{L}R,2} &= R_{LR,2} \\ R_{\tilde{H}R,2} &= R_{HR,2}. \end{aligned}$$

Proof: Let us first set $R_{\tilde{H}R,2} = R_{HR,2}$ by setting

$$\tilde{\alpha}_2 = \alpha_2. \quad (11)$$

Next, we may attempt to set $R_{\tilde{H}R,1} = R_{HR,1}$. Note that from the properties of the AWGN channel capacity function, $C_{\sigma_N^2}(P_1) + C_{\sigma_N^2+P_1}(P_2) = C_{\sigma_N^2}(P_1 + P_2)$, where $C_N(x) = (1/2)\log_2(1 + (x/N))$. Hence

$$\begin{aligned} R_{HR,1} &= R'_{HR,1} + R''_{HR,1} \\ &= C_{\sigma_N^2}((\beta(1-\alpha'_1) + (1-\beta)(1-\alpha''_1))\sigma_1^2) \end{aligned}$$

and we may allow $R_{\tilde{H}R,1} = R_{HR,1}$ by setting $1 - \tilde{\alpha}_1 = \beta(1 - \alpha'_1) + (1 - \beta)(1 - \alpha''_1)$, or equivalently

$$\tilde{\alpha}_1 = (1 - \beta)\alpha''_1 + \beta\alpha'_1. \quad (12)$$

Next, we attempt to set $R_{\tilde{L}\tilde{R},2} = R_{LR,2}$ by choosing our final degree of freedom, $\tilde{\beta}$, accordingly. Note that

$$\begin{aligned} R_{\tilde{L}\tilde{R},2} &= C_{\sigma_N^2 + (1-\tilde{\alpha}_1)\sigma_1^2 + (1-\tilde{\alpha}_2)\sigma_2^2 + \tilde{\alpha}_1(1-\tilde{\beta})\sigma_1^2} (\tilde{\alpha}_2\sigma_2^2) \\ R_{LR,2} &= C_{\sigma_N^2 + (1-\alpha_1)\sigma_1^2 + (1-\alpha_2)\sigma_2^2 + (1-\beta)\alpha_1'\sigma_1^2} (\alpha_2\sigma_2^2). \end{aligned}$$

Hence, we may equate these two rates by setting $\tilde{\alpha}_1(1-\tilde{\beta}) = (1-\beta)\alpha_1'$, or equivalently

$$\tilde{\beta} = \frac{\beta\alpha_1'}{(1-\beta)\alpha_1' + \beta\alpha_1'}. \quad (13)$$

We still must check that with these values of $\tilde{\beta}$, $\tilde{\alpha}_1$, and $\tilde{\alpha}_2$, we in fact arrive at our final set of rates matching: $R_{\tilde{L}\tilde{R},1} = R_{LR,1}$. Let us now verify

$$\begin{aligned} R_{\tilde{L}\tilde{R},1} &= C_{\sigma_N^2 + \sigma_2^2 + (1-\tilde{\alpha}\tilde{\beta})\sigma_1^2} (\tilde{\alpha}_1\tilde{\beta}\sigma_1^2) \\ &+ C_{\sigma_N^2 + (1-\tilde{\alpha}_2)\sigma_2^2 + (1-\tilde{\alpha}_1)\sigma_1^2} (\tilde{\alpha}_1(1-\tilde{\beta})\sigma_1^2) \\ &= C_{\sigma_N^2 + \sigma_2^2 + (\tilde{\alpha}_1(1-\tilde{\beta}) + (1-\tilde{\alpha}_1))\sigma_1^2} (\tilde{\alpha}_1\tilde{\beta}\sigma_1^2) \\ &+ C_{\sigma_N^2 + (1-\alpha_2)\sigma_2^2 + (1-\tilde{\alpha}_1)\sigma_1^2} (\tilde{\alpha}_1(1-\tilde{\beta})\sigma_1^2) \quad (14) \end{aligned}$$

$$= C_{\sigma_N^2 + \sigma_2^2 + (\alpha_1'(1-\beta) + (1-\tilde{\alpha}_1))\sigma_1^2} (\beta\alpha_1'\sigma_1^2) \quad (15)$$

$$+ C_{\sigma_N^2 + (1-\alpha_2)\sigma_2^2 + (1-\tilde{\alpha}_1)\sigma_1^2} (\alpha_1'(1-\beta)\sigma_1^2) \quad (16)$$

$$= C_{\sigma_N^2 + \sigma_2^2 + (\alpha_1'(1-\beta) + \beta(1-\alpha_1') + (1-\beta)(1-\alpha_1'))\sigma_1^2} (\beta\alpha_1'\sigma_1^2) \quad (17)$$

$$+ C_{\sigma_N^2 + (1-\alpha_2)\sigma_2^2 + (\beta(1-\alpha_1') + (1-\beta)(1-\alpha_1'))\sigma_1^2} (\alpha_1'(1-\beta)\sigma_1^2)$$

$$= C_{\sigma_N^2 + \sigma_2^2 + (1-\alpha_1'\beta)\sigma_1^2} (\beta\alpha_1'\sigma_1^2)$$

$$+ C_{\sigma_N^2 + (1-\alpha_2)\sigma_2^2 + (1-\alpha_1'\beta - \alpha_1'(1-\beta))\sigma_1^2} (\alpha_1'(1-\beta)\sigma_1^2)$$

$$= R'_{LR,1} + R''_{LR,1}$$

$$= R_{LR,1} \quad (18)$$

where (14) is due to (11), (15) and (16) are due to (12) and (13), (17) is due to (12), and (18) is due to (12). Note that this collapses the previous $2(2M-1) = 4M-2$ virtual users into $M+2M-1 = 3M-1$ virtual users.

APPENDIX B

In this appendix, we maximize expected capacity for all the remaining cases of p_1 and p_2 .

A. Maximum Total Expected Rate for Boundary Probabilities

Case 1: Only one $p_i = 1$.

We first consider the case where $p_1 = 1$, and $0 < p_2 < 1$. By the general expression for the total expected rate in (5), we obtain (omitting certain algebraic steps)

$$\begin{aligned} E(R) &= (1-p_2)r_1^2 + p_2(r_1^3 + r_2^3) \\ &\leq (1-p_2)\frac{1}{2}\ln\left(\frac{(\sigma_1^2 + \sigma_2^2 + \sigma_N^2)((1-\alpha_1'\beta)\sigma_1^2 + \sigma_N^2)}{((1-\alpha_1'\beta)\sigma_1^2 + \sigma_2^2 + \sigma_N^2)\sigma_N^2}\right) \\ &\quad + \frac{1}{2}\ln\left(\frac{\sigma_1^2 + \sigma_2^2 + \sigma_N^2}{(1-\alpha_1'\beta - \alpha_1'(1-\beta))\sigma_1^2 + \sigma_N^2}\right)p_2 \\ &\leq \frac{1}{2}\ln\left(1 + \frac{\sigma_1^2}{\sigma_N^2}\right) + \frac{1}{2}\ln\left(1 + \frac{\sigma_2^2}{\sigma_1^2 + \sigma_N^2}\right)p_2. \end{aligned}$$

We achieve equality when $\alpha_2 = 1$, $(1 - \alpha_1'\beta) = 1$ and $(1 - \alpha_1'\beta - \alpha_1'(1 - \beta)) = 0$ (i.e., $k_1 = k_2 = 0$). So, the total expected rate for this case is

$$E(R)_{\max} = \frac{1}{2}\ln\left(1 + \frac{\sigma_1^2}{\sigma_N^2}\right) + \frac{1}{2}\ln\left(1 + \frac{\sigma_2^2}{\sigma_1^2 + \sigma_N^2}\right)p_2 \quad (19)$$

where $\beta = 0$, $\alpha_1'' = \alpha_2 = 1$, α_1' can take any value in $[0, 1]$.

For the case where $0 < p_1 < 1$, $p_2 = 1$, we obtain in a similar fashion that

$$E(R)_{\max} = \frac{1}{2}\ln\left(1 + \frac{\sigma_2^2}{\sigma_N^2}\right) + \frac{1}{2}\ln\left(1 + \frac{\sigma_1^2}{\sigma_2^2 + \sigma_N^2}\right)p_1$$

where $\beta = \alpha_1' = \alpha_2 = 1$, α_1'' can take any value in $[0, 1]$.

Case 2: $p_1 = 1$ and $p_2 = 1$.

The total expected rate can be shown to be the same as the maximum total rate for the multiple access channel, thus satisfying our conditions for stability discussed in Section II. Hence

$$E(R)_{\max} = \frac{1}{2}\ln\left(1 + \frac{\sigma_1^2 + \sigma_2^2}{\sigma_N^2}\right) \quad (20)$$

where $(1 - \alpha_1'\beta - \alpha_1''(1 - \beta)) = 0$, $(1 - \alpha_2) = 0$.

B. Maximum Total Expected Rate for Interior Probabilities

We now consider the case where the vector p lies in the interior of the simplex: $0 < p_1 < 1$ and $0 < p_2 < 1$. By the general expression for the total expected rate in (5) (omitting algebraic steps), we have the equation shown at the bottom of the page.

For simplicity, we consider in detail only the special case when $p_1 = p_2 = p$. For the case $p_1 \neq p_2$, see [35]. We note that the conclusion in either case is the same. The optimum (k_1, k_2) occurs on the boundary of its domain.

For the special case $0 < p_1 = p_2 = p < 1$, the total expected rate can be written as

$$\begin{aligned} E(R) &= p(r_1^3 + r_2^3) + p(1-p)(r_1^2 - r_1^3 + r_2^2 - r_2^3) \\ &= \frac{1}{2}\ln\left(\frac{\sigma_1^2 + \sigma_2^2 + \sigma_N^2}{(1-\alpha_1'\beta - \alpha_1''(1-\beta))\sigma_1^2 + (1-\alpha_2)\sigma_2^2 + \sigma_N^2}\right) \\ &\quad + p(1-p)\left[\frac{1}{2}\ln\left(1 + \frac{(1-\alpha_1'\beta - \alpha_1''(1-\beta))\sigma_1^2}{\sigma_N^2}\right)\right. \\ &\quad \left. + \frac{1}{2}\ln\left(1 + \frac{(1-\alpha_2)\sigma_2^2}{\sigma_N^2}\right)\right]. \end{aligned}$$

Given the definition in Section III-A, i.e., $k_1 = 1 - \alpha_1'\beta - \alpha_1''(1 - \beta)$, $k_2 = 1 - \alpha_2$, we obtain

$$\begin{aligned} E(R) &= p\frac{1}{2}\ln\left(\frac{\sigma_1^2 + \sigma_2^2 + \sigma_N^2}{k_1\sigma_1^2 + k_2\sigma_2^2 + \sigma_N^2}\right) \\ &\quad + p(1-p)\left[\frac{1}{2}\ln\left(1 + \frac{k_1\sigma_1^2}{\sigma_N^2}\right) + \frac{1}{2}\ln\left(1 + \frac{k_2\sigma_2^2}{\sigma_N^2}\right)\right]. \quad (21) \end{aligned}$$

According to their definitions, k_1 and k_2 are independent and may take values in $[0, 1]$, so we get a maximization problem of a two-variable function over a closed region. The maximum points are either on the boundary or the relative maxima inside the region. Next, we prove that there are no relative maxima inside the region.

Take the derivatives of $E(R)$ in (21) over k_1 and k_2 , we get

$$\frac{\partial E(R)}{\partial k_1} = -\frac{1}{2}p\frac{\sigma_1^2}{k_1\sigma_1^2 + k_2\sigma_2^2 + \sigma_N^2} + \frac{1}{2}p(1-p)\frac{\sigma_1^2}{k_1\sigma_1^2 + \sigma_N^2}$$

$$\frac{\partial E(R)}{\partial k_2} = -\frac{1}{2}p \frac{\sigma_2^2}{k_1 \sigma_1^2 + k_2 \sigma_2^2 + \sigma_N^2} + \frac{1}{2}p(1-p) \frac{\sigma_2^2}{k_2 \sigma_2^2 + \sigma_N^2}. \quad (22)$$

$$(23)$$

By letting $\partial E(R)/\partial k_1 = 0$ and $\partial E(R)/\partial k_2 = 0$, we determine an extreme point: $k_{10} = p\sigma_N^2/(1-2p)\sigma_1^2$, $k_{20} = p\sigma_N^2/(1-2p)\sigma_2^2$. Taking the second order derivatives of $E(R)$ at this point yields

$$\begin{aligned} \frac{\partial^2 E(R)}{\partial k_1^2} \Big|_{(k_{10}, k_{20})} &= \frac{1}{2}p \frac{\sigma_1^4}{(k_{10}\sigma_1^2 + k_{20}\sigma_2^2 + \sigma_N^2)^2} \\ &\quad - \frac{1}{2}p(1-p) \frac{\sigma_1^4}{(k_{10}\sigma_1^2 + \sigma_N^2)^2} \\ &= \frac{1}{2}p \frac{\sigma_1^4}{\left(\frac{p\sigma_N^2}{1-2p} + \frac{p\sigma_N^2}{1-2p} + \sigma_N^2\right)^2} \\ &\quad - \frac{1}{2}p(1-p) \frac{\sigma_1^4}{\left(\frac{p\sigma_N^2}{1-2p} + \sigma_N^2\right)^2} \\ &= \frac{1}{2}p(1-2p)^2 \frac{\sigma_1^4}{\sigma_N^4} \left(1 - \frac{1}{1-p}\right) < 0 \quad (24) \end{aligned}$$

$$\frac{\partial^2 E(R)}{\partial k_2^2} \Big|_{(k_{10}, k_{20})} = \frac{1}{2}p(1-2p)^2 \frac{\sigma_2^4}{\sigma_N^4} \left(1 - \frac{1}{1-p}\right) < 0 \quad (25)$$

$$\begin{aligned} \frac{\partial^2 E(R)}{\partial k_1 \partial k_2} \Big|_{(k_{10}, k_{20})} &= \frac{1}{2}p \frac{\sigma_1^2 \sigma_2^2}{(k_{10}\sigma_1^2 + k_{20}\sigma_2^2 + \sigma_N^2)^2} \\ &= \frac{1}{2}p \frac{(1-2p)^2 \sigma_1^2 \sigma_2^2}{\sigma_N^4} \quad (26) \end{aligned}$$

$$\frac{\partial^2 E(R)}{\partial k_2 \partial k_1} \Big|_{(k_{10}, k_{20})} = \frac{1}{2}p \frac{(1-2p)^2 \sigma_1^2 \sigma_2^2}{\sigma_N^4}. \quad (27)$$

From

$$k_{10} = \left(\frac{p\sigma_N^2}{(1-2p)\sigma_1^2}\right) > 0 \Rightarrow p < 0.5 \Rightarrow \left(1 - \left(\frac{1}{1-p}\right)\right)^2 < 1$$

we have

$$\begin{aligned} \frac{\partial^2 E(R)}{\partial k_1^2} \Big|_{(k_{10}, k_{20})} &\quad \frac{\partial^2 E(R)}{\partial k_2^2} \Big|_{(k_{10}, k_{20})} \\ &\quad - \frac{\partial^2 E(R)}{\partial k_1 \partial k_2} \Big|_{(k_{10}, k_{20})} \frac{\partial^2 E(R)}{\partial k_2 \partial k_1} \Big|_{(k_{10}, k_{20})} < 0. \quad (28) \end{aligned}$$

The results in (24)–(28) imply (k_{10}, k_{20}) is not a relative extremum, and the maximum points are on the boundary, i.e., $k_1 = 0$, or $k_2 = 0$, or $k_1 = 1$, or $k_2 = 1$. We consider four cases.

Case 1) If $k_1 = 0$, then from (23) we know that, for this case

$$\frac{\partial E(R)}{\partial k_2} = -\frac{1}{2}p \frac{\sigma_2^2}{k_2 \sigma_2^2 + \sigma_N^2} + \frac{1}{2}p(1-p) \frac{\sigma_2^2}{k_2 \sigma_2^2 + \sigma_N^2} < 0$$

then $E(R)$ is a decreasing function of k_2 .

So, when $k_2 = 0$, we achieve the relative maximum of $E(R)$ on line $k_1 = 0$.

Case 2) If $k_2 = 0$, then from (22) we know that, for this case

$$\frac{\partial E(R)}{\partial k_1} = -\frac{1}{2}p \frac{\sigma_1^2}{k_1 \sigma_1^2 + \sigma_N^2} + \frac{1}{2}p(1-p) \frac{\sigma_1^2}{k_1 \sigma_1^2 + \sigma_N^2} < 0.$$

So, when $k_1 = 0$, we achieve the relative maximum of $E(R)$ on line $k_2 = 0$. From Cases 1 and 2, the relative maximum on the boundary $k_1 = 0$ or $k_2 = 0$ is

$$E(R)^* = \frac{1}{2}p \ln \left(1 + \frac{\sigma_1^2 + \sigma_2^2}{\sigma_N^2}\right). \quad (29)$$

Case 3) If $k_1 = 1$, then from (23) we know that, for this case

$$\frac{\partial E(R)}{\partial k_2} = -\frac{1}{2}p \frac{\sigma_2^2}{\sigma_1^2 + k_2 \sigma_2^2 + \sigma_N^2} + \frac{1}{2}p(1-p) \frac{\sigma_2^2}{k_2 \sigma_2^2 + \sigma_N^2}. \quad (30)$$

For different p in $[0, 1]$, we have the following three possibilities.

1) $p \leq \sigma_1^2/(\sigma_1^2 + \sigma_2^2 + \sigma_N^2)$
 $\partial E(R)/\partial k_2 \geq 0$ implies that, when $k_2 = 1$, we get the relative maximum

$$E(R)^* = p(1-p) \left[\frac{1}{2} \ln \left(1 + \frac{\sigma_1^2}{\sigma_N^2}\right) + \frac{1}{2} \ln \left(1 + \frac{\sigma_2^2}{\sigma_N^2}\right) \right]. \quad (31)$$

2) $\sigma_1^2/(\sigma_1^2 + \sigma_2^2 + \sigma_N^2) \leq p \leq \sigma_1^2/(\sigma_1^2 + \sigma_N^2)$
 $\partial E(R)/\partial k_2 = 0$ implies that, when $k_2 = ((1-p)\sigma_1^2 - p\sigma_N^2)/p\sigma_2^2$, we get the relative maximum

$$\begin{aligned} E(R)^* &= \frac{1}{2}p \ln \left(\frac{p(\sigma_1^2 + \sigma_2^2 + \sigma_N^2)}{\sigma_1^2}\right) \\ &\quad + \frac{1}{2}p(1-p) \left[\ln \left(1 + \frac{\sigma_1^2}{\sigma_N^2}\right) + \ln \left(\frac{(1-p)\sigma_1^2}{p\sigma_N^2}\right) \right]. \quad (32) \end{aligned}$$

3) $p \geq \sigma_1^2/(\sigma_1^2 + \sigma_N^2)$
 $(\partial E(R)/\partial k_2) \leq 0$ implies that, when $k_2 = k_1 = 0$, we get the relative maximum

$$E(R)^* = \frac{1}{2}p \ln \left(1 + \frac{\sigma_1^2 + \sigma_2^2}{\sigma_N^2}\right). \quad (33)$$

$$\begin{aligned} E(R) &= p_1(1-p_2)r_1^2 + p_2(1-p_1)r_2^2 + p_1p_2(r_1^3 + r_2^3) \\ &= p_1(1-p_2)(r_1^2 - r_1^3 + r_1^3) + p_2(1-p_1)(r_2^2 - r_2^3 + r_2^3) + p_1p_2(r_1^3 + r_2^3) \\ &= p_1r_1^3 + p_2r_2^3 + p_1(1-p_2)(r_1^2 - r_1^3) + p_2(1-p_1)(r_2^2 - r_2^3) \\ &= p_1 \frac{1}{2} \ln \left(\frac{(\sigma_1^2 + \sigma_2^2 + \sigma_N^2)((1-\alpha'_1\beta)\sigma_1^2 + (1-\alpha_2)\sigma_2^2 + \sigma_N^2)}{((1-\alpha'_1\beta)\sigma_1^2 + \sigma_2^2 + \sigma_N^2)(1-\alpha'_1\beta - \alpha''_1(1-\beta))\sigma_1^2 + (1-\alpha_2)\sigma_2^2 + \sigma_N^2} \right) \\ &\quad + p_2 \frac{1}{2} \ln \left(1 + \frac{\alpha_2\sigma_2^2}{(1-\alpha'_1\beta)\sigma_1^2 + (1-\alpha_2)\sigma_2^2 + \sigma_N^2} \right) \\ &\quad + p_1(1-p_2) \frac{1}{2} \ln \left(1 + \frac{(1-\alpha'_1\beta - \alpha''_1(1-\beta))\sigma_1^2}{\sigma_N^2} \right) + p_2(1-p_1) \frac{1}{2} \ln \left(1 + \frac{(1-\alpha_2)\sigma_2^2}{\sigma_N^2} \right) \end{aligned}$$

Case 4) If $k_2 = 1$, then from (22) we know that, for this case

$$\frac{\partial E(R)}{\partial k_1} = -\frac{1}{2}p \frac{\sigma_1^2}{k_1 \sigma_1^2 + \sigma_2^2 + \sigma_N^2} + \frac{1}{2}p(1-p) \frac{\sigma_1^2}{k_1 \sigma_1^2 + \sigma_N^2}. \quad (34)$$

For different p in $[0, 1]$, we have the following three possibilities.

- 1) $p \leq \sigma_2^2/(\sigma_1^2 + \sigma_2^2 + \sigma_N^2)$
 $\partial E(R)/\partial k_1 \geq 0$ implies that, when $k_1 = 1$, we get the relative maximum

$$E(R)^* = p(1-p) \left[\frac{1}{2} \ln \left(1 + \frac{\sigma_1^2}{\sigma_N^2} \right) + \frac{1}{2} \ln \left(1 + \frac{\sigma_2^2}{\sigma_N^2} \right) \right]. \quad (35)$$

- 2) $\sigma_2^2/(\sigma_1^2 + \sigma_2^2 + \sigma_N^2) \leq p \leq \sigma_2^2/(\sigma_2^2 + \sigma_N^2)$
 $\partial E(R)/\partial k_2 = 0$, hence when $k_1 = ((1-p)\sigma_2^2 - p\sigma_N^2)/p\sigma_1^2$, we get the relative maximum

$$E(R)^* = \frac{1}{2}p \ln \left(\frac{p(\sigma_1^2 + \sigma_2^2 + \sigma_N^2)}{\sigma_2^2} \right) + \frac{1}{2}p(1-p) \left[\ln \left(1 + \frac{\sigma_2^2}{\sigma_N^2} \right) + \ln \left(\frac{(1-p)\sigma_2^2}{p\sigma_N^2} \right) \right]. \quad (36)$$

- 3) $p \geq \sigma_2^2/(\sigma_2^2 + \sigma_N^2)$
 $\partial E(R)/\partial k_1 \leq 0$, hence when $k_1 = k_2 = 0$, we get the relative maximum

$$E(R)^* = \frac{1}{2}p \ln \left(1 + \frac{\sigma_1^2 + \sigma_2^2}{\sigma_N^2} \right). \quad (37)$$

Based on the discussions above, the maximum expected rate for the case $p_1 = p_2 = p$ is

Case 1) $\sigma_1^2 = \sigma_2^2 = \sigma^2$

- A) $p \leq \sigma^2/(2\sigma^2 + \sigma_N^2)$, $E(R)_{\max} = \max((29), (31))$, where (29) and (31) are the equation numbers.
 B) $\sigma^2/(2\sigma^2 + \sigma_N^2) < p < \sigma^2/(\sigma^2 + \sigma_N^2)$, $E(R)_{\max} = \max((29), (32)) = (36)$.
 C) $p > \sigma^2/(\sigma^2 + \sigma_N^2)$, $E(R)_{\max} = (29)$.

Case 2) $\sigma_1^2 > \sigma_2^2$

- A) $\sigma_1^2/(\sigma_1^2 + \sigma_2^2 + \sigma_N^2) \leq \sigma_2^2/(\sigma_2^2 + \sigma_N^2)$
 a) $p \leq \sigma_2^2/(\sigma_1^2 + \sigma_2^2 + \sigma_N^2)$, $E(R)_{\max} = \max((29), (31))$.
 b) $\frac{\sigma_2^2/(\sigma_1^2 + \sigma_2^2 + \sigma_N^2)}{\sigma_1^2/(\sigma_1^2 + \sigma_2^2 + \sigma_N^2)} < p < \frac{\sigma_2^2/(\sigma_2^2 + \sigma_N^2)}{\sigma_1^2/(\sigma_1^2 + \sigma_2^2 + \sigma_N^2)}$, $E(R)_{\max} = \max((29), (31), (36))$.
 c) $\sigma_1^2/(\sigma_1^2 + \sigma_2^2 + \sigma_N^2) < p < \sigma_2^2/(\sigma_2^2 + \sigma_N^2)$, $E(R)_{\max} = \max((29), (32), (36))$.
 d) $\frac{\sigma_2^2/(\sigma_2^2 + \sigma_N^2)}{\sigma_1^2/(\sigma_1^2 + \sigma_2^2 + \sigma_N^2)} \leq p < \sigma_1^2/(\sigma_1^2 + \sigma_2^2 + \sigma_N^2)$, $E(R)_{\max} = \max((29), (31))$.
 e) $p > \sigma_1^2/(\sigma_1^2 + \sigma_2^2 + \sigma_N^2)$, $E(R)_{\max} = (29)$.
 B) $\sigma_1^2/(\sigma_1^2 + \sigma_2^2 + \sigma_N^2) \geq \sigma_2^2/(\sigma_2^2 + \sigma_N^2)$
 a) $p \leq \sigma_2^2/(\sigma_1^2 + \sigma_2^2 + \sigma_N^2)$, $E(R)_{\max} = \max((29), (31))$.
 b) $\frac{\sigma_2^2/(\sigma_1^2 + \sigma_2^2 + \sigma_N^2)}{\sigma_1^2/(\sigma_1^2 + \sigma_2^2 + \sigma_N^2)} < p < \frac{\sigma_2^2/(\sigma_2^2 + \sigma_N^2)}{\sigma_1^2/(\sigma_1^2 + \sigma_2^2 + \sigma_N^2)}$, $E(R)_{\max} = \max((29), (31), (36))$.
 c) $\sigma_2^2/(\sigma_2^2 + \sigma_N^2) < p < \sigma_1^2/(\sigma_1^2 + \sigma_2^2 + \sigma_N^2)$, $E(R)_{\max} = \max((29), (31))$.
 d) $\sigma_1^2/(\sigma_1^2 + \sigma_2^2 + \sigma_N^2) \leq p < \sigma_1^2/(\sigma_1^2 + \sigma_2^2 + \sigma_N^2)$, $E(R)_{\max} = \max((29), (32))$.
 e) $p > \sigma_1^2/(\sigma_1^2 + \sigma_2^2 + \sigma_N^2)$, $E(R)_{\max} = (29)$.

Case 3) $\sigma_1^2 < \sigma_2^2$, and our conclusions are similar to those of 2).

We can see, for the extreme case where $p_1 = p_2 = 1 > \sigma_1^2/(\sigma_1^2 + \sigma_N^2)$ and $\sigma_2^2/(\sigma_2^2 + \sigma_N^2)$

$$E(R)_{\max} = (29) = \frac{1}{2} \ln \left(1 + \frac{\sigma_1^2 + \sigma_2^2}{\sigma_N^2} \right)$$

which is the capacity of the multiple-access channel.

ACKNOWLEDGMENT

The authors would like to acknowledge useful discussions with D. P. Taylor, K. Wasserman, and S. Shamai.

REFERENCES

- [1] N. Abramson, "The aloha system—Another alternative for computer communications," in *Proc. Fall Joint Computer Conf.*, 1970.
- [2] T. P. Coleman and M. Médard, "Trade-off between energy and delay in wireless packetized systems," in *Proc. 39th Annual Allerton Conf. Communications, Control and Computing*, 2001.
- [3] N. Abramson, "The throughput of packet broadcasting channels," *IEEE Trans. Commun.*, vol. COM-25, Jan. 1977.
- [4] —, "Spread ALOHA for VSAT's," *The ALOHA System*, Univ. Hawaii, B86-4a, 1986.
- [5] R. Ahlswede, "Multi-way communication channels," *Proc. ISIT*, pp. 23–52, 1971.
- [6] V. Anantharam, "The stability region of the finite-user, slotted ALOHA system protocol," *IEEE Trans. Inform. Theory*, vol. 37, pp. 535–540, 1991.
- [7] E. Arkan, "Multi-Access in packet radio networks," M.S., MIT, 1982.
- [8] Banh-Tri-An and E. Gelenbe, "Etude et simulation de réseaux de diffusion sous contrôle," *RAIRO Informatique/Computer Science*, vol. 11, no. 3, pp. 301–311, 1977.
- [9] P. Bergmans, "A simple converse for broadcast channels with additive white Gaussian noise," *IEEE Trans. Inform. Theory*, vol. 20, pp. 279–280, Mar. 1974.
- [10] R. A. Berry, "Power and delay trade-offs in fading channels," Ph.D. dissertation, MIT, 2000.
- [11] I. Bettesh and S. Shamai, "A low delay algorithm for the multiple access channel with Rayleigh fading," in *Proc. PIMRC 98*, 1998, pp. 1367–1372.
- [12] —, "Outage analysis for multiple access channel with Rayleigh fading," in *Proc. ISIT*, 2000, p. 303.
- [13] E. Biglieri, J. Proakis, and S. Shamai, "Fading channels: Information-theoretic and communication aspects," *IEEE Trans. Inform. Theory*, vol. 44, pp. 2619–2692, Oct. 1998.
- [14] A. M. Y. Bigloo, T. A. Gulliver, and V. K. Bhargava, "Maximum-likelihood decoding and code combining for DS/SSMA slotted ALOHA," *IEEE Trans. Commun.*, vol. 45, pp. 1602–1612, 1997.
- [15] D. Brady and L. Merakos, "Throughput performance of multiuser detection in unslopped contention channels," *Proc. IEEE INFOCOM*, 1994.
- [16] G. Caire, G. Taricco, and E. Biglieri, "Minimum outage probability for slowly-fading channels," in *Proc. ISIT*, 1998, p. 7.
- [17] G. Caire, E. Leonardi, and E. Viterbo, "Concatenated coding for packet transmission over the collision channel," in *Proc. ISIT*, 1998, p. 288.
- [18] G. Caire and D. Tunitetti, "ARQ protocols for the Gaussian collision channel," in *Proc. ISIT*, 2000, p. 406.
- [19] J. I. Capetanakis, "The multiple access broadcast channel: Protocol and capacity considerations," *IEEE Trans. Inform. Theory*, vol. IT-35, 1979.
- [20] T. M. Cover, "Broadcast channels," *IEEE Trans. Inform. Theory*, vol. IT-18, pp. 2–14, Jan. 1972.
- [21] —, "An achievable rate region for the broadcast channel," *IEEE Trans. Inform. Theory*, vol. IT-21, pp. 399–404, July 1975.
- [22] —, "Comments on broadcast channels," *IEEE Trans. Inform. Theory*, vol. 44, pp. 2524–2530, Oct. 1998.
- [23] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.
- [24] M. Effros and A. Goldsmith, "Capacity definitions and coding strategies for general channels with receiver side information," in *Proc. ISIT*, 1998, p. 39.

- [25] M. J. Ferguson, "On the control, stability, and waiting time in a slotted ALOHA random-access system," *IEEE Trans. Commun.*, pp. 1306–1311, Nov. 1975.
- [26] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.
- [27] —, "Conflict resolution in random access broadcast networks," in *AFOSR Workshop*, 1978, pp. 74–76.
- [28] A. Grant, B. Rimoldi, R. Urbanke, and P. Whiting, "Single user coding for the discrete memoryless multiple access channel," *Proc. ISIT*, p. 448, 1995.
- [29] B. Hajek, "Stochastic approximation methods for decentralized control of multiaccess communications," *IEEE Trans. Inform. Theory*, vol. IT-31, pp. 176–184, Mar. 1985.
- [30] B. Hajek and T. van Loon, "Decentralized dynamic control of a multiaccess broadcast channels," *IEEE Trans. Automat. Contr.*, vol. AC-27, pp. 559–569, June 1982.
- [31] S. Hanly and D. Tse, "Multiaccess fading channels: Part II: Delay-limited capacities," *IEEE Trans. Inform. Theory*, vol. 44, pp. 2816–2831, Nov. 1998.
- [32] J. Hayes, "An adaptive technique for local distribution," *IEEE Trans. Commun.*, vol. COM-26, pp. 1178–1186, 1978.
- [33] A. Hottinen and R. Wichman, "Utilization of spread ALOHA with multiuser detection," *ISIT*, p. 123, 1998.
- [34] J.-C. Huang and T. Berger, "Delay analysis of interval searching contention resolution algorithms," *IEEE Trans. Inform. Theory*, vol. IT-31, pp. 264–273, Mar. 1985.
- [35] J. Huang, "Capacity of time-slotted ALOHA systems," Master, University of Illinois at Urbana-Champaign, 2000.
- [36] S. W. Kim, "Stabilization of slotted ALOHA spread-spectrum communication networks," *IEEE J. Select. Areas Commun.*, vol. 8, pp. 555–561, May 1990.
- [37] J. Komlós and A. G. Greenberg, "An asymptotically fast nonadaptive algorithm for conflict resolution in multiple-access channels," *IEEE Trans. Inform. Theory*, vol. IT-31, pp. 302–306, Mar. 1985.
- [38] S. S. Lam and L. Kleinrock, "Packet switching in a multiaccess broadcast channel: Dynamic control procedure," *IEEE Trans. Commun.*, vol. COM-23, pp. 891–904, Sept. 1975.
- [39] H. Liao, "Multiple-Access channels," Ph.D. dissertation, Univ. Hawaii, 1972.
- [40] J. L. Massey, "Collision resolution algorithms and random-access communications," in *Multi User Communication Systems*, G. Longa, Ed., 1981, vol. 2, pp. 73–137.
- [41] J. L. Massey and P. Mathys, "The collision channel without feedback," *IEEE Trans. Inform. Theory*, vol. 31, pp. 192–204, 1985.
- [42] M. Médard and A. J. Goldsmith, "Capacity of time-slotted ALOHA systems," *WCNC*, vol. 1, pp. 490–494, 1999.
- [43] M. Médard, S. P. Meyn, J. Huang, and A. J. Goldsmith, "Capacity of time-slotted ALOHA systems," *Proc. ISIT*, p. 407, 2000.
- [44] S. Meyn and R. Tweedie, *Markov Chains and Stochastic Stability*. New York: Springer-Verlag, 1993.
- [45] J. Moseley and P. Humblet, "A class of efficient contention resolution algorithms for multiple access channels," *IEEE Trans. Commun.*, vol. COM-33, pp. 145–151, 1985.
- [46] S.-J. Oh and K. M. Wasserman, "Dynamic spreading gain control in multiservice CDMA networks," *IEEE J. Select. Areas Commun.*, vol. 17, pp. 918–927, 1999.
- [47] L. H. Ozarow, S. Shamai, and A. D. Wyner, "Information theoretic considerations for cellular mobile radio," *IEEE Trans. Vehic. Technol.*, vol. 43, pp. 359–378, May 1994.
- [48] S. S. Panwar, D. F. Towsley, and J. K. Wolf, "On the throughput of degenerate intersection and first-come first-served collision resolution algorithms," *IEEE Trans. Inform. Theory*, vol. IT-31, pp. 274–279, Mar. 1985.
- [49] N. Pippenger, "Bounds on the performance of protocols for a multiple access broadcast channel," *IEEE Trans. Inform. Theory*, vol. IT-27, pp. 45–151, Mar. 1981.
- [50] A. Polydoros and J. Silvester, "Slotted random access spread-spectrum networks: An analytical framework," *IEEE J. Select. Areas Commun.*, vol. SAC-5, pp. 989–1002, 1987.
- [51] R. Rao and A. Ephremides, "On the stability of a multiple-access system," *IEEE Trans. Inform. Theory*, vol. 34, pp. 918–930, 1988.
- [52] D. Raychaudhuri, "Performance analysis of random access packet-switched code-division multiple access systems," *IEEE Trans. Commun.*, vol. COM-29, 1981.
- [53] B. Rimoldi and R. Urbanke, "A rate splitting approach to the Gaussian multiple-access channel," *IEEE Trans. Inform. Theory*, vol. 42, pp. 364–375, 1996.
- [54] R. Rivest, "Network control by Bayesian broadcast," *IEEE Trans. Inform. Theory*, vol. IT-33, pp. 323–328, 1987.
- [55] W. A. Rosenkrantz and D. Towsley, "On the instability of the slotted ALOHA multiaccess algorithm," *IEEE Trans. Automat. Contr.*, vol. AC-28, pp. 994–996, Oct. 1983.
- [56] T. N. Saadawi and A. Ephremides, "Analysis, stability, and optimization of slotted ALOHA with a finite number of buffered users," *IEEE Trans. Automat. Contr.*, vol. AC-26, pp. 680–689, June 1981.
- [57] J. W. Sennott and L. I. Sennott, "A queueing model for analysis of a bursty multiple-access communication channel," *IEEE Trans. Inform. Theory*, vol. IT-27, pp. 317–321, May 1981.
- [58] D. Seret and C. Macchi, "Contrôle adaptif d'un réseau de transmission de paquets," in *Flow Control in Computer Networks*. Amsterdam, The Netherlands: North-Holland, 1979.
- [59] S. Shamai, "A broadcast strategy for the Gaussian slowly fading channel," *Proc. ISIT*, p. 150, 1997.
- [60] —, "A broadcast approach for the multiple-access slow fading channel," in *Int. Symp. Information Theory*, 2000, p. 128.
- [61] T. Shinomiya and H. Suzuki, "Slotted ALOHA mobile packet communication systems with multiuser detection in a base station," *IEEE Trans. Vehic. Technol.*, vol. 49, pp. 948–955, May 2000.
- [62] F. Tarköy, "Information-theoretic aspects of spread ALOHA communications," in *Proc. PIMRC*, 1995, pp. 1318–1320.
- [63] B. S. Tsybakov, "Survey of USSR contributions to random multiple-access communications," *IEEE Trans. Inform. Theory*, vol. IT-31, pp. 143–165, Mar. 1985.
- [64] B. S. Tsybakov and V. L. Bakirov, "Stability of asynchronous ALOHA system," *Problemy Peredachi Informatsii*, vol. 20, pp. 82–94, 1984.
- [65] B. S. Tsybakov and V. A. Mikhailov, "Free synchronous packet access in a broadcast channel with feedback," *Problemy Peredachi Informatsii*, vol. 14, pp. 32–59, 1978.
- [66] S. Verdú and T. S. Han, "A general formula for channel capacity," *IEEE Trans. Inform. Theory*, vol. 40, pp. 1147–1157, July 1994.
- [67] J. Wolfowitz, *Coding Theorems of Information Theory*. Berlin, Germany: Springer-Verlag, 1978.

Muriel Médard (S'91–M'95–SM'00) received the B.S. degrees in electrical engineering and computer science as well as mathematics in 1989, the B.S. degree in humanities in 1990, the M.S. degree in electrical engineering in 1991, and the Sc.D. degree in electrical engineering in 1995, all from the Massachusetts Institute of Technology, (MIT) Cambridge.

She is an Esther and Harold E. Edgerton Associate Professor in the Department of Electrical Engineering and Computer Science, MIT, and a member of the Laboratory for Information and Decision Systems. She was previously an Assistant Professor at the Electrical and Computer Engineering Department and a member of the Coordinated Science Laboratory at the University of Illinois, Urbana-Champaign. From 1995 to 1998, she was a Staff Member at MIT Lincoln Laboratory in the Optical Communications and the Advanced Networking Groups. Her research interests include the areas of reliable communications, with particular emphasis on networks and information theory.

Dr. Médard serves as an Associate Editor for the *Optical Communications and Networking Series* of the *IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS* and as an Associate Editor for Communications for the *IEEE TRANSACTIONS ON INFORMATION THEORY*. She has served as a Guest Editor for the *IEEE JOURNAL OF LIGHTWAVE TECHNOLOGY* and was an Associate Editor for the *OSA Journal of Optical Networking*. She is the recipient of an NSF Career Award. She was awarded the IEEE Leon K. Kirchmayer Prize Paper Award, 2002, ("The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel," *IEEE Trans. Inform. Theory*, vol. 46, pp. 935–946, May 2000). She was co-awarded the Best Paper Award (G. Weichenberg, V. Chan, M. Médard, "Reliable architectures for networks under stress," in *4th Int. Workshop Design Reliable Communication Networks (DRCN 2003)*, Banff, Alberta, Canada, Oct. 2003).

Jianyi Huang (S'00) received the B.A. degree from the Department of Automation, Tsinghua University, Beijing, China, in 1996, and the M.S. degree in electrical engineering from the University of Illinois, Urbana-Champaign, in 2000. He is now working toward the Ph.D. degree (with Prof. S. Meyn) in electrical engineering at the University of Illinois, Urbana-Champaign.

His research interests include channel capacity and channel coding.

Andrea J. Goldsmith (S'90–M'93–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 communication 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 then joined the Electrical Engineering Department, Stanford University, Stanford, CA, where she is now an Associate Professor and holds the Bredt Faculty Development Scholar Chair. Her research interests include work in the capacity of wireless channels and networks, wireless communication theory, adaptive resource allocation, multiantenna channels, communications and control, energy-constrained communications, and cross-layer design for cellular systems and *ad hoc* wireless networks.

Dr. Goldsmith is a Recipient of the National Academy of Engineering Gilbreth Lectureship, an Alfred P. Sloan Fellowship, a Stanford Terman Faculty Fellowship, the 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 was an Editor for the *IEEE TRANSACTIONS ON COMMUNICATIONS* from 1995 to 2001 and has been an Editor for the *IEEE Wireless Communications Magazine* since 1995. She is also an elected member of Stanford's Faculty Senate.

Sean P. Meyn (S'85–M'87–SM'95–F'02) received the B.A. degree in mathematics (*summa cum laude*) from the University of California, Los Angeles, in 1982 and the Ph.D. degree in electrical engineering from McGill University in 1987.

He completed a two-year postdoctoral fellowship at the Australian National University, Canberra, Australia. He is now a Professor in the Department of Electrical and Computer Engineering, and a Research Professor in the Coordinated Science Laboratory at the University of Illinois, Urbana-Champaign. He is coauthor (with R. Tweedie) of the monograph *Markov Chains and Stochastic Stability*, (London, U.K.: Springer-Verlag, 1993). He has held visiting positions at the Australian National University; the Indian Institute of Science, Bangalore; the Technion, Haifa, Israel; and at the University of Versailles, France. His research interests include stochastic processes, optimization, complex networks, and information theory.

Dr. Meyn has served on the editorial boards of several journals in the systems and control and applied probability areas. He was a University of Illinois Vice Chancellor's Teaching Scholar in 1994 and received jointly with Tweedie the 1994 ORSA/TIMS Best Publication In Applied Probability Award. During 1997–1998, he was a Fulbright Research Scholar in India.

Todd P. Coleman (S'99) received B.S. degrees in electrical engineering and in computer engineering (*summa cum laude*) from the University of Michigan, Ann Arbor, in 2000. He received the M.S. degree in electrical engineering from the Massachusetts Institute of Technology (MIT), Cambridge, in 2002. He is currently working toward the Ph.D. degree in electrical engineering at MIT.

His research interests include multiple-user information theory, distributed source compression, and queueing theory.

Mr. Coleman has received an NSF Graduate Research Fellowship, the EECS Morris J. Levin Award for Best Masterworks Oral Thesis Presentation, the Harrison Award for Excellence in Location Estimation, MIT 6.432 (Fall 2000), the University of Michigan Hugh Rumler Senior Class Prize, a Michigan Association of Governing Boards Award, the University of Michigan EECS Senior Scholar Award, the University of Michigan NAACP Academic Achievement Award, and the University of Michigan Spirit of Martin Luther King, Jr. Award.