November 1987 LIDS-P-1714

ENERGY LIMITED CHANNELS: CODING, MULTIACCESS, AND SPREAD SPECTRUM**

ABSTRACT

Robert G. Gallager *

We consider a class of communication channels modelling situations in which there is an energy constraint but almost an unlimited number of degrees of freedom per unit time. Examples of such channels are broadband additive Gaussian noise channels, fading dispersive channels, and quantum optical channels. We start by restricting such channels to binary inputs and then find the reliability function. In the limit of infinite bandwidth, the reliability function for these channels can be found exactly for all rates if there is a finite capacity in terms of bits per unit energy. Particular attention is paid to how this limiting reliability is approached as bandwidth increases. We then show that the restriction to binary inputs is essentially optimal in the broadband limit. Finally we apply these results to multiaccess use of such channels. Multiaccess coding for these broadband channels provides us with an abstraction of spread spectrum communication.

---

*Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Room 35-206, Cambridge, MA 02139, USA.

**Research was supported by ARO Contract No. DAAL03-86-K-0171 and the National Science Foundation Contract No. NSF-ECS-8310698.

1) INTRODUCTION

There are many communication situations in which the primary constraint on the transmitted sequences arises from power limitations rather than bandwidth limitations. These situations have been widely studied for additive Gaussian noise channels (for example [1-3]) and to a lesser extent for fading dispersive channels (for example [4,5]. There is no cohesive treatment of the class of such channels, and this paper is intended to provide at least the beginning of such a cohesive treatment. For multiaccess communication, in which a large number of transmitters share the same physical medium, it is usually desirable to limit the power of each transmitter so as to limit the interference between the transmitters. In this case again, there is usually more bandwidth than any one transmitter could use effectively, and the primary limit is from the imposed power constraint. For example, in local area networks of the Ethernet type, transmitters use high energy while transmitting, but transmit only a small fraction of time, thus leading to small overall power. In spread spectrum communication, transmitters spread their power over a broad bandwidth, thus resulting in low power per unit bandwidth and not too much interference with other transmitters.

In this paper, we model situations of the above types by a discrete time channel in which the discrete time transmitted and received symbols can be regarded as inputs and outputs over incremental intervals of time and frequency. Initially, we consider point to point communication with a single transmitter and receiver and we restrict the discrete time input to be binary, the symbol 1 representing a non-zero energy input, and the symbol 0 representing a zero energy input. The power constraint is modeled by a limitation on the number of 1's that the transmitter can send in any given block. We derive various bounds on the error probability for coding on such a channel. We then consider the limit as the block length becomes large but the total number of 1's per block remains limited. We then find the reliability function in this limit, i.e., the exponent to achievable error probability per allowable 1 input per block. This reliability function is found exactly for all rates between 0 and capacity, where capacity is in natural units per allowable 1 input. We also find that this capacity is infinite for some channel models and find the conditions under which this occurs.

We next consider a more general discrete time point to point channel with an input alphabet $\{0,1,...,K\}$. The zero input again corresponds to zero energy, and there is some arbitrary energy associated with each of the other input letters. We again impose a constraint on the energy per input symbol . We again derive error bounds for coding and again consider the limit of large block length holding the total energy per block fixed. For the random coding bound, the optimal input distribution becomes binary in this limit. The additve white Gaussian noise channel is somewhat pathological here, since, although a binary input becomes optimal, non-binary inputs are just as good. For many channels in this general class, the random coding bound gives the true reliability function for all rates, whereas for other channels, there is a low rate region where the reliability function can only be bounded. We demonstrate the conditions under which each type of behavior pertains.

Finally, we consider multiaccess channels with the same type of energy constraint on each transmitter's input. We derive a converse to the multiaccess coding theorem for channels with such an energy constraint. This is not an obvious extension of the converse without an energy constraint, and apparently this is new. We then restrict attention to the binary input case in which the channel can be modelled as an <u>or</u> channel followed by an arbitrary binary input memoryless point to point channel. We derive a random coding bound for this class of multiaccess channel for three different cases, one in which the code words from the different transmitters are independently decoded, one in which the code words are decoded in sequence, using the earlier decoded words to help in the decoding, and one in which the decoding of all code words is simultaneous.

2) BINARY INPUT POINT TO POINT CHANNELS

Consider a discrete time memoryless communication channel with binary inputs, $x=\{0,1\}$. For simplicity of notation, assume the outputs are real numbers with transition probability densities $p_0(y)$ and $p_1(y)$. We consider a block code of constraint length N with M code words, $\mathbf{x}_1, \mathbf{x}_2, \dots \mathbf{x}_M$, and rate $R = (\ln M)/N$. We consider the input 0 to be a zero energy input and the input 1 to be a positive energy input; the energy constraint that we impose is then

$$\sum_{n=1}^{N} x_{mn} \leq N\delta \tag{1}$$

for each code word $\mathbf{x}_m = (x_{m1}, x_{m2}, \dots x_{mN})$. For a given code and decoding algorithm, the probability of error is the average over all code words of the probability of incorrect decoding of that code word. Let $P(N,R,\delta)$ be the minimum probability of error for any block code of block length N and rate at least R satisfying the energy constraint in (1). The reliability function $E(R,\delta)$ is then defined as

$$E(R,\delta) = \limsup_N \frac{-\ln P(N,R,\delta)}{N} \tag{2}$$

The coding theorem asserts that $E(R,\delta)$ is positive for R less than the capacity of the channel. Note that the energy of each code word in the codes above is at most $D=N\delta$. What we want to consider in this paper is what happens when N is increased, but the total energy D is held constant. Note that D can also be interpreted as the maximum allowable Hamming distance between each code word and the all zero sequence. For many physical channels, the discrete time elements we are considering here can be viewed as inputs and outputs over incremental intervals of time and frequency, and increasing N can be viewed as increasing the frequency band of the transmission without changing the time duration of the code words and without changing the discrete time channel model.

Note that if N is increased while holding D and the number of code words fixed, then $P(N,R,\delta)$ either decreases or remains constant. The reason for this is that a code for the smaller blocklength could be converted into a code of larger blocklength simply by using 0's for the extra letters of each code word. Thus the limit of $P(N,R,\delta)$ as $N \to \infty$ (with M and $N\delta$ fixed) must exist. It is intuitively almost clear that the optimal code in this limit is a code in which no two code words have 1's in the same positions, thus requiring a block length of M times the energy D. We refer to this code in what follows as the <u>orthogonal</u> code; unfortunately, we have not been able to show the optimality of the orthogonal code, but we shall show that it is essentially optimal. Our major concern is not with the error probability of the orthogonal code, which is very wasteful of block length, but rather with how this limiting error probability is approached with increasing N for a given D.

Given our interest in the limit N->∞, it is desirable to define a normalized rate, $\tilde{R}$ = (ln M)/D = R/δ and a normalized reliability function,

$$\tilde{E}(\tilde{R},\delta) = \lim\sup_N \frac{-\ln P(N,\tilde{R}\,\delta,\delta)}{N\delta} = \frac{E(\tilde{R}\,\delta,\delta)}{\delta} \tag{3}$$

By the argument above, $\tilde{E}(\tilde{R},\delta)$ is nonincreasingwith δ, and thus we must have a limit (perhaps infinite)

$$\tilde{E}(\tilde{R}) = \lim_{\delta\to 0} \tilde{E}(\tilde{R},\delta) \tag{4}$$

The exponent $\tilde{E}(\tilde{R})$ above gives the exponential rate at which error probability can be made to decrease with increasing total energy D = Nδ. There is a subtlety involved here with the order of limits in (3) and (4). In particular, for the orthogonal code discussed above, as one increases D for a given $\tilde{R}$, N must increase and δ must decrease exponentially with D. It is not clear apriori that the exponential rate at which error probability decreases with D for the orthogonal code is the same as $\tilde{E}(\tilde{R})$, since in (3) and (4), one holds δ fixed, but arbitrarily small, while increasing D. We shall find, however, that these exponents are the same, showing that impractically large bandwidths are not required to approach the best exponents.

In the next sub-section, we analyze the random coding bound [1] for these channels, using a fixed value of δ. We next find the reliability function at 0 rate and use this, along with known results about the relationship of lower and upper bounds to error probability, to find upper and lower bounds to the reliability function E(R,δ). We shall find that the upper and lower bounds come together in the limit δ->0, and thus $\tilde{E}(\tilde{R})$ is specified for all $\tilde{R}$. This is rather surprising, since the only other channels for which the reliability function is completely known are a few channels that have a positive zero error capacity, and very noisy channels [1]. Another surprising aspect of these channels is that the capacity (in bits per unit D) can be infinite in the limit δ->0, and we find the conditions under which this can happen.

THE RANDOM CODING BOUND

In applying the random coding bound, we can regard (1) as an energy constraint and then use a shell constraint in the random coding bound as developed in section 7.3 of [1]. An alternative would be to use a fixed composition random coding argument, using Nδ ones in each code word. We take the first alternative here since it is analytically simpler. The result (from theorem 7.3.2 of [1]) is that the reliability function E(R,δ) satisfies

$$E(R,\delta) \geq E_r(R,\delta) \tag{5}$$

where the random coding exponent $E_r(R,\delta)$ is given by

$$E_r(R,\delta) = \max_{0\le\rho\le 1,\ 0\le r} \quad E_0(\rho,r,\delta) - \rho R \tag{6}$$

$$E_0(\rho,r,\delta) = -\ln\int_y \{(1-\delta)p_0(y)^{\frac{1}{1+\rho}}e^{-r\delta} + \delta p_1(y)^{\frac{1}{1+\rho}}e^{r(1-\delta)}\}^{1+\rho}\,dy \tag{7}$$

In this expression, we have used the obvious input distribution, using input 1 with probability δ. It is conceivable that the optimum distribution could use input 1 with a probability less than δ, but we assume in what follows that the constraint δ is always less than the unconstrained optimal probability of input 1. We want to factor out the term $p_0(y)$ from the sum in (7), but observe that there may be a range of y over which $p_0(y) = 0$. To take care of this case, define q as

$$q = \int_{y:p_0(y)=0} p_1(y)dy \tag{8}$$

We shall soon see that whether $q = 0$ or not plays a critical role in whether or not the limiting capacity of the channel, in bits per unit energy, is finite or not. We can now rewrite (7) in the form

$$E_0(\rho,r,\delta) = -\ln\{q\delta^{1+\rho}e^{r(1-\delta)(1+\rho)} + \int p_0(y)[e^{-r\delta}[1-\delta+\delta f(y)]]^{1+\rho}dy\} \tag{9}$$

$$\text{where} \quad f(y) = e^r\left[\frac{p_1(y)}{p_0(y)}\right]^{\frac{1}{1+\rho}} \tag{10}$$

We next upper bound E_0 so as to yield the asymptotic infinite bandwidth result and then later lower bound E_0 so as to discover how the asymptotic result is approached. For the upper bound, we must lower bound the expression inside the log. We first lower bound q by 0, and then, inside the bracketed term in the integral, we use the bound $(1+x)^{(1+\rho)} \ge 1+x(1+\rho)$; this is valid for $x\ge -1$ and $\rho\ge 0$. This yields

$$E_0(\rho,r,\delta) \le r\delta(1+\rho) - \ln\int p_0(y)\{1+\delta(1+\rho)[f(y)-1]\}dy \tag{11}$$

The right hand side of (11) can now be maximized over the free parameter $r \ge 0$, which appears both in the first term and implicitly in f(y). The maximizing r occurs where the integral above is 1; this maximizing r depends on ρ but not on δ and is given by

$$r(\rho) = -\ln \int_y [p_0(y)]^{\frac{\rho}{1+\rho}} [p_1(y)]^{\frac{1}{1+\rho}} dy \tag{12}$$

The integral above exists and is convex $\cup$ in ρ for $\rho>0$ except in the special case in which $p_0(y)p_1(y) = 0$ for all y. This special case is equivalent to a noisless binary symmetric channel and will be treated separately as an example later. We take r(0) to be lim $r(\rho)$ as ρ approaches 0 from above. We assume in what follows (except for that example) that $r(\rho)$ exists. We then define $\tilde{E}_0(\rho)$ as

$$\tilde{E}_0(\rho) = (1+\rho)r(\rho) \tag{13}$$

so that

$$\max_r E_0(\rho,r,\delta) \le \delta\, \tilde{E}_0(\rho) \tag{14}$$

Thus, from (6), we have

$$E_r(R,\delta) \le \max_{0\le\rho\le 1} \left[\delta\tilde{E}_0(\rho) - \rho R\right] \tag{15}$$

On the surface, this is not an entirely satisfactory result, since (15) gives an upper bound to $E_r(R,\delta)$, which is a lower bound to the actual exponent $E(R,\delta)$. We shall show later, when we evaluate the reliability function at R=0, that the right hand side of (15) is also an upper bound to $E(R,\delta)$. Now, however, we evaluate a lower bound to $E_0(\rho,r,\delta)$ and show that the bounds in (14) and (15) are tight in the limit of small δ. What we need is an upper bound on the bracketed term inside the integral of (9). The following lemma gives us the required result.

Lemma 1: For any ρ, $0\le\rho\le 1$, any $x\ge 0$, and any z, $0\le z\le 1$,

$$(1+x-z)^{1+\rho} \le 1 + (1+\rho-\rho z)(x-z) + x^{1+\rho} \tag{16}$$

Proof: The left side of (16) is convex $\cup$ in ρ, so for $0\le\rho\le 1$,

$$(1+x-z)^{1+\rho} \le 1 + (1+\rho)(x-z) + \rho(x-z)^2 \tag{17}$$

To see this, note that the right side of (17) is linear in ρ and (17) is satisfied with equality for ρ equal to 0 and to 1. Next, upper bounding $(x-z)^2$ by $x^2 - z(x-z)$, (17) becomes

$$(1+x-z)^{1+\rho} \le 1 + (1+\rho-\rho z)(x-z) + \rho x^2 \tag{18}$$

For $x\le 1$, $x^2 \le x^{1+\rho}$. Thus, since $\rho\le 1$, the right side of (18) is upper bounded by the right side of (16), establishing the lemma for $x\le 1$. For $x\ge 1$, we have

$$(1+x-z)^{1+\rho} = \left[\frac{1+x-z}{x}\right]^{1+\rho} x^{1+\rho} = \left[1 + \frac{x-z}{x}\right]^{1+\rho} x^{1+\rho} \tag{19}$$

Upper bounding the first term on the right of (19) by the same argument as in (18) and multiplying by $x^{1+\rho}$, we get

$$(1+x-z)^{1+\rho} \le x^{1+\rho} + (1+\rho)(1-z)x^{\rho} + \rho(1-z)^2x^{\rho-1} \tag{20}$$

Since $x\geq1$ here, we can upper bound x^{ρ} by x; also $(1+\rho)(1-z) \le 1+\rho-\rho z$. Thus

$$(1+x-z)^{1+\rho} \le x^{1+\rho} + (1+\rho-\rho z)x + \rho(1-z)^2 = x^{1+\rho} + (1+\rho-\rho z)(x-z) + z-\rho z+\rho$$

Since $0\le z\le1$, we have $z-\rho z+\rho \le 1$, which establishes the lemma for $x\geq1$.

We now apply the lemma to the bracketed expression in the integral of (9), letting x be $\delta f(y)$ and y be δ.

$$E_0(\rho,r,\delta) \ge r\delta(1+\rho) - \ln\{q[\delta e^r]^{1+\rho}+ \int_y p_0(y)\left[1+(1+\rho-\rho\delta)(\delta f(y)-\delta)+[\delta f(y)]^{1+\rho}\right]dy\} \tag{21}$$

Let $F(\rho,r,\delta)$ denote the integral above. We choose $r = r(\rho)$ in evaluating it. Recall from (11) that $\int p_0(y)[f(y)-1]dy = 0$ for this value of r, so that

$$F(\rho,r(\rho),\delta) = 1 + \int p_0(y)[\delta f(y)]^{1+\rho}\, dy \tag{22}$$

We recall that the integral in (9) is over y such that $p_0(y)>0$; f(y) (see (10)) is undefined otherwise. We then have

$$F(\rho,r(\rho),\delta) = 1 + \int_{y:p_0(y)>0} p_0(y)\left[\delta e^r\right]^{1+\rho} \frac{p_1(y)}{p_0(y)}\, dy = 1 + (1-q)\left[\delta e^r\right]^{1+\rho} \tag{23}$$

We have used the definition of q in (8). Substituting (23) into (21),

$$\begin{aligned} E_0(\rho,r(\rho),\delta) &\ge r(\rho)\delta(1+\rho) - \ln\left\{ 1 + \left[\delta e^{r(\rho)}\right]^{1+\rho} \right\} \\ &\ge \delta r(\rho)(1+\rho) - \delta^{1+\rho}\, e^{(1+\rho)r(\rho)} = \delta\left[\tilde{E}_0(\rho)-\delta^{\rho} \exp \tilde{E}_0(\rho)\right] \end{aligned} \tag{24}$$

$$\max_r E_0(\rho,r,\delta) \ge \delta\left[\tilde{E}_0(\rho)-\delta^{\rho} \exp \tilde{E}_0(\rho)\right] \tag{25}$$

The next lemma relates our upper and lower bounds. Recall that we are assuming that $r(\rho)$ exists (i.e., that the channel is not a noiseless BSC).

Lemma 2: For all ρ, $0<\rho\le1$,

$$\lim_{\delta\to 0} \delta^{-1} \max_r E_0(\rho,r,\delta) = \tilde{E}_0(\rho) \tag{26}$$

Proof: From (14), $\delta^{-1} \max_r E_0(\rho,r,\delta) \le \tilde{E}_0(\rho)$. From (25), $\delta^{-1} \max_r E_0(\rho,r,\delta) \ge \tilde{E}_0(\rho) - \delta^{\rho} \exp \tilde{E}_0(\rho)]$. As $\delta \to 0$, δ^{ρ} also approaches 0 and the upper and lower bounds approach equality.

Define

$$\tilde{E}_r(\tilde{R}) = \max_{0\le\rho\le 1} \left[\tilde{E}_0(\rho) - \rho\tilde{R}\right] \tag{27}$$

We show in the next theorem that this is the limiting random coding exponent as δ approaches 0. First, however we discuss the maximization over ρ above. The function $\tilde{E}_0(\rho)$ is nondecreasing and convex $\cap$ in ρ (see the lemma in appendix 5B of [1] for a proof of this). Thus the maximization in (27) is given parametrically by

$$\tilde{E}_r(\tilde{R}) = \tilde{E}_0(1) - \tilde{R} \qquad \text{for } \tilde{R} < \tilde{E}_0'(1)$$

$$\tilde{E}_r(\tilde{R}) = \tilde{E}_0(\rho) - \rho\tilde{E}_0'(\rho) \qquad \text{for } \tilde{R} = \tilde{E}_0'(\rho),\ 0<\rho\le 1$$

$$\tilde{E}_r(\tilde{R}) = \tilde{E}_0(0) \qquad \text{for } \tilde{R} \ge \tilde{E}_0'(0) \tag{28}$$

In the last part of the above equation, $\tilde{E}_0(0)$ and $\tilde{E}_0'(0)$ are taken as limits as $\rho \to 0^+$. What is quite surprising here, however, is that $\tilde{E}_0(0)$ need not be 0. In fact, evaluating the limit from (12) and (13), we find that

$$\tilde{E}_0(0) = -\ln(1-q) \tag{29}$$

This means that if $q>0$, then $\tilde{E}_r(\tilde{R})$ is positive for all $\tilde{R}$; we shall soon see that this implies that the capacity, in bits per unit energy, is infinite. This might seem less surprising if we observe that for the orthogonal code, if message m is sent, correct decoding is assured if, in any position where the m^{th} code word has a one, the output falls in the set of probability q for which input 0 is impossible. Thus the probability of error for the orthogonal code is lower bounded by $(1-q)^D$. In general, $\tilde{E}_r(\tilde{R})$ is decreasing with $\tilde{R}$ out to $\tilde{E}_0'(0)$ and is constant for all larger $\tilde{R}$.

Theorem 1: For all $\tilde{R} \ge 0$,

$$\lim_{\delta\to 0} \delta^{-1} E_r(\tilde{R}\delta, \delta) = \tilde{E}_r(\tilde{R}) \tag{30}$$

Proof: From (15), we have

$$\delta^{-1} E_r(\tilde{R}\delta, \delta) \le \tilde{E}_r(\tilde{R})$$

Thus, we need only establish that for any $\varepsilon>0$ we have

$$\delta^{-1} E_r(\tilde{R}\delta, \delta) \geq \tilde{E}_r(\tilde{R}) - \varepsilon \tag{31}$$

for sufficiently small δ. First consider $\tilde{R}$ satisfying $\tilde{R} = \tilde{E}_0'(\rho)$ for some ρ, $0<\rho\leq 1$. Using this value of ρ to lower bound the right side of (6) and combining with (25),

$$E_r(\tilde{R}\delta, \delta) \geq \delta\left[\tilde{E}_0(\rho)-\delta^{\rho} \exp \tilde{E}_0(\rho) - \rho\tilde{R}\right] = \delta\left[\tilde{E}_r(\tilde{R})-\delta^{\rho} \exp \tilde{E}_0(\rho)\right] \tag{32}$$

Dividing both sides by δ, we see that (31) is satisfied for small enough δ. The same argument applies for $\tilde{R} < \tilde{E}_0'(1)$, using 1 for ρ. Finally consider the last case where $R \geq \tilde{E}_0'(0)$. If $E_0(0) = 0$, then $\tilde{E}_r(\tilde{R}) = 0$ and (31) is obviously satisfied for all δ. If $E_0(0) > 0$, then choose ρ to satisfy

$$\tilde{E}_0(\rho) - \rho\tilde{R} = E_0(0) - \varepsilon/2 \tag{33}$$

Using the first half of (32) with this value of ρ, and combining with (33),

$$E_r(\tilde{R}\delta, \delta) \geq \delta\left[\tilde{E}_0(0) - \varepsilon/2 - \delta^{\rho} \exp \tilde{E}_0(\rho)\right] \tag{34}$$

Again (31) is satisfied for small enough δ and the proof is complete.

LOWER BOUNDS TO ERROR PROBABILITY

We next show that $\tilde{E}_r(\tilde{R})$ is also an upper bound on the actual reliability function $\delta^{-1}E(\tilde{R}\delta, \delta)$. Since the random coding exponent is a lower bound to the reliability function, this plus theorem 1 will show that $\tilde{E}_r(\tilde{R})$ is also the limiting reliability function as δ->0. We start by evaluating error probability for a code with two code words. Assume that the code is orthogonal, since for two code words this clearly minimizes error probability. This is a classical binary hypothesis testing problem, made even simpler by the symmetry between the two code words. The error probability (see, for example, section 5.4 of [1]) is given by

$$P_e = O\left(\frac{1}{\sqrt{N}}\right) \exp -D\left[-2 \ln \int_y \sqrt{p_0(y)p_1(y)}\, dy\right] \tag{35}$$

Note that the quantity in brackets is $\tilde{E}_0(1) = \tilde{E}_r(0)$. Also this exponent for two code words is an upper bound to $\tilde{E}(0)$, so we have established that $\tilde{E}_r(0) = \tilde{E}(0)$. It is well known [6] that for any DMC (and thus any value of $\delta>0$) the reliability function is upper bounded by the sphere packing exponent and by the straight line exponent. The sphere packing exponent is given by

$$E_s(R,\delta) = \max_{\rho\geq 0} [\max_{r\geq 0} E_0(\rho,r,\delta) - \rho R] \tag{36}$$

It follows from (14) that

$$E_s(\tilde{R}\delta,\delta) \leq \delta \max_{\rho\geq 0} [\tilde{E}_0(\rho) - \rho\tilde{R}] \tag{37}$$

Letting $\tilde{E}_s(\tilde{R})$ be $\max_{\rho\geq 0}$ $[\tilde{E}_0(\rho)-\rho\tilde{R}]$, we have

$$\tilde{E}_s(\tilde{R}) = \tilde{E}_0(\rho)-\rho\tilde{E}_0'(\rho) \qquad \text{for } \tilde{R} = \tilde{E}_0'(\rho),\ \rho > 0 \tag{38}$$

$$= \tilde{E}_0(0) \qquad \text{for } \tilde{R} \geq \tilde{E}_0'(\rho) \tag{39}$$

The straight line exponent is tangent to the sphere packing exponent (as a function of R) and equals the zero rate exponent, $\tilde{E}_0(1)$, at $R = 0$. The straight line exponent is then upper bounded as the tangent to the upper bound to E_s in (37); i.e. as $\delta[\tilde{E}_0(1)-\tilde{R}]$. Thus combining the upper bound to the sphere packing exponent with this bound to the straight line exponent, we have our earlier upper bound to the random coding exponent; we have thus proven the following theorem:

Theorem 2: For all $\tilde{R} \geq 0$

$$\delta^{-1} E(\tilde{R}\delta,\delta) \leq \tilde{E}_r(\tilde{R}) \tag{40}$$

$$\lim_{\delta\to 0} \delta^{-1}E(\tilde{R}\delta, \delta) = \tilde{E}_r(\tilde{R}) \tag{41}$$

The next issue to be considered is the exponent to error probability with no bandwidth constraint. That is, we want to establish the following theorem.

Theorem 3: For all $\tilde{R} \geq 0$

$$\lim_{D\to\infty} \left[\lim_{N\to\infty} \frac{-\ln P_e(N, \tilde{R}D/N, D/N)}{D}\right] = \tilde{E}_r(\tilde{R}) \tag{42}$$

Comment: Since $P_e(N, \tilde{R}D/N, D/N)$ is non-increasing with N, the limit in N above can be lower bounded by any finite N, say $N = D/\delta$. Thus the left side of (42) is lower bounded by $\tilde{E}_r(\tilde{R})$ and the following proof establishes that the left side of (42) is also upper bounded by $\tilde{E}_r(\tilde{R})$. This requires finding a lower bound to P_e that depends on D and $\tilde{R}$ but not on N.

Proof: For any given block length N and any given code $\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_M$, let Y_m be the m^{th} decoding subset. For ordinary decoding, the sets Y_m are disjoint and for list decoding with a list of L, each $\mathbf{y}$ is in at most L decoding sets Y_m (see [6]; list decoding allows us to combine the sphere packing bound being derived here with the straight line bound). The probability of error given that message m is sent is then

$$P_{em} = \int_{Y_m} \prod_{n=1}^{N} p(y_n | x_{mn}) d\mathbf{y} \tag{43}$$

where $\mathbf{y} = y_1, y_2, \ldots y_N$ and $\mathbf{x}_m = x_{m1}, x_{m2}, \ldots x_{mN}$. Define

$$F(Y_m) = \int_{Y_m} \prod_{n=1}^{N} p_o(y_n) d\mathbf{y} \tag{44}$$

From theorem 5 of [6], for any s, $0 < s < 1$, either

$$P_{em} > \frac{1}{4} \exp[D\,\mu(s) - Ds\,\mu'(s) - s\sqrt{D\mu''(s)}\,] \quad \text{or} \tag{45}$$

$$F(Y_m) > \frac{1}{4} \exp[D\,\mu(s) + D(1-s)\mu'(s) - (1-s)\sqrt{D\mu''(s)}\,] \tag{46}$$

where

$$\mu(s) = \ln \int_{y} p_1(y)^{1-s}\, p_0(y)^s dy \tag{47}$$

Choose s, $0 < s < 1$, so that the right side of (46) equals L/M and assume that L and M are such that this is possible. Since $\sum_m F(Y_m) \le L$, (46) must be violated for some m, so (45) must be satisfied for that m. Since this is true for all codes of M code words, we let $\tilde{R} = [\ln(M/L)]/D$ and have

$$P_e(N,\tilde{R}D/N, D/N) > \exp[D\mu(s) - Ds\mu'(s) - s\sqrt{D\mu''(s)} - \ln 4] \tag{48}$$

for s such that

$$\tilde{R} = \mu(s) - (1-s)\mu'(s) + (1-s)\sqrt{\mu''(s)/D} + (\ln 4)/D \tag{49}$$

This bound is independent of N, so we can pass directly to the limit over D, obtaining

$$\lim_{D\to\infty} [\lim_{N\to\infty} \frac{-\ln P_e(N,\tilde{R}D/N, D/N)}{D}] \leq -\mu(s) + s\mu'(s) \tag{50}$$

for s, $0 < s < 1$, such that

$$\tilde{R} = -\mu(s) - (1-s)\mu'(s) \tag{51}$$

Some algebra reveals that $-\mu(s) + s\mu'(s) = \tilde{E}_0(\rho) - \rho\tilde{E}_0'(\rho)$ and $-\mu(s) - (1-s)\mu'(s) = E_0'(\rho)$ for $s = \rho/(1+\rho)$. Thus, from (38),

$$\lim_{D\to\infty} [\lim_{N\to\infty} \frac{-\ln P_e(N,\tilde{R}D/N, D/N)}{D}] \leq \tilde{E}_s(\tilde{R}) \tag{52}$$

for $\tilde{R} < \tilde{E}_0'(0)$. For $\tilde{R} = \tilde{E}_0'(0)$, the s for which (49) is satisfied approaches 0^+ as $D \to\infty$, so the exponent in (48) approaches $\mu(0) = -\tilde{E}_0(0)$ and (52) is again satisfied. Finally, for $\tilde{R} > \tilde{E}_0'(0)$ and large enough D, (46) is violated (for some m) by all s, $0 < s < 1$, and thus is violated in the limit $s\to 0^+$. Thus (52) is also satisfied for $\tilde{R} > \tilde{E}_0'(0)$ and thus is satisfied for all $\tilde{R} > 0$. Combining this with the straight line exponent as before, the proof of theorem 3 is complete.

We have seen in the above theorems that the reliability function $\tilde{E}_r(\tilde{R})$ in the limit $\delta\to 0$ is similar to the usual reliability functions for discrete memoryless channels with the following two major differences. First the reliability function here is known for <u>all</u> $\tilde{R} > 0$. Second, if $q > 0$, $\tilde{E}_r(\tilde{R})$ is positive for <u>all</u> $\tilde{R} > 0$; that is, the capacity in bits per unit energy of the channel is infinite. A special case of this phenomenon was first observed for a model of quantum optical channel [7,8,9]. It does not suggest that there is anything wrong with the concept of capacity for such channels and in no way gives added importance to the computational cut off rate (see [9] for a more complete discussion); it does suggest that models for which $q > 0$ are not completely realistic, especially for large $\tilde{R}$.

Another peculiar phenomenon that doesn't occur for the DMC is that it is possible for $\lim_{\rho\to 0} \tilde{E}_0'(\rho)$ to be infinite. If q = 0 and this limit is infinite, then the capacity (in bits per unit energy) is infinite, but the exponent approaches 0 as $\tilde{R} \to \infty$.

3) ARBITRARY SET OF INPUTS

Next consider an arbitrary input alphabet {0,1,...,K} and let h(k) be the energy associated with input letter k. Assume that h(0) = 0 and h(k) > 0 for $1 \le k \le K$. Assume further that each code word $\mathbf{x_m} = x_{m1},...x_{mN}$ must satisfy the constraint

$$\sum_{n=1}^{N} h(x_{mn}) \le N\delta. \tag{53}$$

For simplicity we assume that $\delta < h(k)$, $1 \le k \le K$ so that some inputs in each code word must be 0. The random coding bound of (5) to (7) is as before, but now (7) generalizes to

$$E_0(\rho,r,\rho,Q) = -\ln \int_y \{\sum_{k=0}^{K} Q_k\, e^{r[h(k)-\delta]}\, p_k(y)^{\frac{1}{1+\rho}}\}^{1+\rho} \tag{54}$$

where $Q = Q_1,..Q_k$ satisfies $\Sigma_k Q_k h(k) \le \delta$, $Q_0 = 1 - \Sigma_k Q_k$, $Q_k \ge 0$ for $0 \le k \le K$ and $p_k(y)$ is the transition probability density, given input k. We now lower bound the integral in (54), and thus upper bound E_0, by integrating only over y such that $p_0(y) > 0$. We can then factor $p_0(y)$ out of the term in braces and substitute $1 - \Sigma_k Q_k$ for Q_0 to get

$$E_0(\rho,r,\delta,Q) \le r\delta(1+\rho) - \ln \int_y p_0(y)\, \{1 + \sum_{k=1}^{K} Q_k[f_k(y)-1]\}^{1+r}\, dy \tag{55}$$

where

$$f_k(y) = e^{rh(k)}\left[\frac{p_k(y)}{p_0(y)}\right]^{\frac{1}{1+\rho}} \tag{56}$$

Eq. (55) can be further upper bounded by the inequality $(1+x)^{1+\rho} \ge 1 + x(1+\rho)$ for $x \ge -1$, $\rho \ge 0$, yielding

$$E_0(\rho,r,\delta,\mathbf{Q}) \le r\,\delta(1+\rho) - \ln \int_y p_0(y)\{1 + (1+\rho)\sum_{k=1}^{K} Q_k[f_k(y)-1]\}dy \qquad (57)$$

The integral above is a linear function of $\mathbf{Q} = (Q_1,...Q_K)$ subject to the linear constraints $\sum_k Q_k h(k) \le \delta$, $Q_k \ge 0$, and thus the right hand side of (57) is maximized over $\mathbf{Q}$ for any given ρ,r,δ by $Q_k = 0$ for all but one k, $1 \le k \le K$ (i.e. by a binary input). It follows that the right side of (57) can be jointly maximized over r and Q by maximizing over r for each binary choice and then selecting the largest of these K possibilities. Using the result of (11) to (13) for a binary input, then, we have

$$\max_{r,\mathbf{Q}} E_0(\rho,r,\delta,\mathbf{Q}) \le \delta\, \tilde{E}_0(\rho) \qquad (58)$$

where

$$\tilde{E}_0(\rho) = \max_k \frac{-(1+\rho)}{h(k)} \ln \int_y p_0(y)^{\frac{\rho}{1+\rho}} p_k(y)^{\frac{1}{1+\rho}} dy \qquad (59)$$

Note that (59) differs from the earlier binary result in (13) by including the maximization over which non zero input to use and by observing that input k can be used only with a relative frequency $\delta/h(k)$.

We can also obtain a lower bound on $\max_{r,Q} E_0(\rho,r,\delta,\mathbf{Q})$ by restricting Q to have only a single non-zero component, chosen to maximize (59). In this case, the lower bound in (25) applies, so that

$$\max_{r,Q} E_0(\rho,r,\delta,\mathbf{Q}) \ge \delta\Big[\tilde{E}_0(\rho) - \Big(\frac{\delta}{h(k)}\Big)^{\rho} \tilde{E}_0(\rho)\Big] \qquad (60)$$

It is seen that lemma 2 is again valid for this more general case, and theorem 1, which we restate for this case as theorem 4, follows as before.

Theorem 4: for all $\tilde{R} \ge 0$,

$$\lim_{\delta\to 0} \delta^{-1} E_r(\tilde{R}\delta, \delta) = \tilde{E}_r(\tilde{R}) \qquad (61)$$

where

where

$$E_r(R,\delta) = \max_{0\le\rho\le 1, 0\le r, Q} E_0(\rho,r,\delta,Q) - \rho R \tag{62}$$

$$\tilde{E}_r(\tilde{R}) = \max_{0\le\rho\le 1} \tilde{E}_0(\rho) - \rho\tilde{R} \tag{63}$$

The sphere packing exponent $E_s(R,\delta)$ also follows as before with

$$\lim_{\delta\to 0} \delta^{-1} E_s(\tilde{R}\delta,\delta) \le \tilde{E}_s(\tilde{R}) \tag{64}$$

$$\tilde{E}_s(\tilde{R}) = \max_{\rho\ge 0} \tilde{E}_0(\rho) - \rho\tilde{R} \tag{65}$$

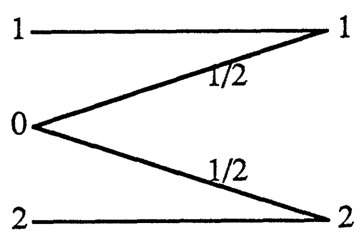
Recall now that in the binary case we showed that $\lim_{\delta\to 0} \delta^{-1}E(\tilde{R}\delta,\delta)$ was equal to $\tilde{E}_r(\tilde{R})$. This was done by evaluating the zero rate exponent, i.e.

$$\tilde{E}(0) = \lim_{\delta\to 0} \left[\lim_{\tilde{R}\to 0} \frac{E(\tilde{R}\delta, \delta)}{\delta} \right]$$

This was shown to be $\tilde{E}_0(1)$ and the desired result followed by combining the straight line exponent with the sphere packing exponent.

In this more general case, we shall find that the zero rate exponent is not <u>always</u> equal to $\tilde{E}_0(1)$. This equality holds for all cases that appear to be of real interest (and thus lim $\delta^{-1}E(\tilde{R}\delta,\delta) = \tilde{E}_r(\tilde{R})$ in these cases) but in cases of less interest, the zero rate exponent exceeds $\tilde{E}_0(1)$.

As an example, consider the DMC in figure 1. From (59), $\tilde{E}_0(\rho) = \rho \ln 2$. Thus $\tilde{E}_r(\tilde{R}) = \ln 2 - \tilde{R}$ for $0 \le \tilde{R} \le \ln 2$ and $\tilde{E}_s(\tilde{R}) = \infty$ for $0 \le R \le \ln 2$. Both are 0 for $\tilde{R} > \ln 2$.



$h(1) = h(2) = 1$

Figure 1

For any blocklength $N \geq D$, this channel can be used as a noiseless binary channel, using inputs 1 or 2 for the first D components of each code word and using input 0 for the final N-D components. Thus the reliability function $\delta^{-1}E(\tilde{R}\delta,\delta)$ is infinite for $0 \leq \tilde{R} \leq \ln 2$ and zero for $\tilde{R} > \ln 2$.

This example can be generalized by adding very small transition probabilities, ε, from input 1 to output 2 and from input 2 to output 1. In this case, there is no zero error capacity and the random coding and sphere packing exponents are equal over a range of rates. For ε small enough, however, the zero rate exponent is achieved by using inputs 1 and 2 exclusively on the first D components of the blocklength and is unequal to $\tilde{E}_0(1)$.

In what follows, we assume that the channel has no zero error capacity. We shall prove the following theorem.

Therem 5: If the zero error capacity is zero, then

$$\tilde{E}(0) = \lim_{\delta\to 0} \frac{E(0,\delta)}{\delta} = \sup_{\mathbf{q}} \frac{\sum_{i,k} -q_i q_k \mu_{ik}(\frac{1}{2})}{\sum_i q_i h(i)} \tag{66}$$

where

$$\mu_{ik}(s) = \ln \int_y p_i(y)^{1-s}\, p_k(s)^s dy \tag{67}$$

and the sup is over probability vectors $\mathbf{q} = q_0,\ldots q_K$ with $q_0 < 1$.

Before proving the theorem, we discuss the supremum in (66) and find the conditions under which it is equal to $\tilde{E}_0(1)$. For any $\mathbf{q}$, define p_i, $1 \leq i \leq k$, by $p_i = q_i/(1-q_0)$. Thus $\mathbf{p}$ is a probability vector over the non-zero inputs. The expression on the right hand side of (66), which we denote $F(\mathbf{q})$, is then

$$F(\mathbf{q}) = \frac{-2q_0 \sum_i p_i\, \mu_{i0}(\frac{1}{2}) - (1-q_0)\sum_{ik} p_i p_k \mu_{ik}(\frac{1}{2})}{\sum_i p_i h(i)} \tag{68}$$

Note that for any given $\mathbf{p}$, $F(\mathbf{q})$ is linear in q_0, and is either maximized at $q_0 = 0$, or as a supremum in the limit where $q_0 \to 1$. Thus the supremum of $F(\mathbf{q})$ can be found by first maximizing the right side of (68) over **p** with $q_0 = 0$ and then with $q_0 = 1$. With $q_0 = 1$, this maximum occurs with some $p_i = 1$, and is given by

$$\max_k -\frac{2\mu_{k0}(\frac{1}{2})}{h(k)} = \tilde{E}_0(1) \tag{69}$$

We thus have proved the following corollary to theorem 5:

Corollary 5.1: If

$$\tilde{E}_0(1) \geq \max_p \frac{\sum_{i=1}^{K}\sum_{k=1}^{K} -p_i p_k \mu_{ik}(\frac{1}{2})}{\sum_{i=1}^{K} p_i h(i)} \tag{70}$$

Then

$$\tilde{E}(0) = \tilde{E}_0(1) \tag{71}$$

This corollary establishes the condition (i.e., (70)) under which the relability function is equal to the random coding exponent for all rates. If (70) is violated, on the other hand, the zero rate exponent is equal to the right side of (70). What the violation of (70) means is that at rates approaching 0, the best codes of a given energy use only non-zero energy inputs for all code words over a block length just long enough to use the available energy. If (70) is satisfied, however, good codes have orthogonal code words and a binary input is sufficient.

Proof of Theorem 5: We first use the expurgated random coding bound to lower bound the reliability $E(0,\delta)$. From [1], the expurgated exponent at rate 0 with constraint δ is given by

$$E_{ex}(0,\delta) = \max_{q:\sum_i q_i h(i) \leq \delta} \sum_{i,k} -q_i q_k \mu_{ik}(\frac{1}{2}) \tag{72}$$

An alternative, given block length N and total energy constraint $N\delta$, is to use a shorter effective constraint length fN ($0<f\leq 1$), using energy δ/f per letter for the first fN letters of the block and using all zeros on the last (1-f)N letters of each code word. Thus, for any f, $0 < f \leq 1$ and any **q** satisfying $\sum_i q_i h(i) \leq \delta/f$, we can achieve an exponent (on a block length of N basis) that is at least

$$-f\sum_{ik} q_i q_k \mu_{ik}(\tfrac{1}{2}) \tag{73}$$

For any given $\mathbf{q}$ such that $q_0<1$, and for any $\delta < \Sigma_i\, q_i h(i)$, we can choose f, $0<f\leq 1$ to satisfy $\Sigma_i\, q_i h(i) = \delta/f$. Substituting this for f in (73) yields

$$\lim_{\delta\to 0} \frac{E(0,\delta)}{\delta} \geq \frac{\sum_{i,k} - q_i q_k \mu_{ik}(\tfrac{1}{2})}{\sum_i q_i\, h(i)} \tag{74}$$

Since this is valid for any $\mathbf{q}$, $q_0 < 1$, the right side of (66) is a lower bound to the left side. In order to establish the opposite inequality, and complete the proof, we use the results in [6 part II]. The results there establish the 0 rate exponent for the unconstrained case, and, rather than repeating the entire (lengthy and difficult) argument for the constrained case, we simply point out the required modifications, using a fixed value δ for the constraint throughout. Theorem 1 of [6 part II] must be modified in a straightforward way (i.e., using Eqs. (1.01) and (1.02) rather than (1.07) and (1.08) of [6 part II]) because of the continuum of the alphabet here. Theorem 2 and lemmas 4.1 and 4.2 of [6 part II] follow without change, but lemma 4.3 must be modified. Lemma 4.3 develops an upper bound on the exponent (there called D_{min}) of an ordered code of 2M code words (the definition of ordered is not relevant to our modification) of block length N. It is shown in (1.36) of [6 part II] that

$$D_{min} \leq \frac{1}{N}\sum_{n=1}\sum_{i,k} q_i^t(n) q_k^b(n)[-\mu_{ik}(\tfrac{1}{2}) - (\tfrac{1}{2}\mu'_{ik}(\tfrac{1}{2})] \tag{75}$$

when $\mu_{ik}(s)$ is defined in (67), $q_i^t(n)$ is the relative frequency of input letter i in position n of the top M code words, and $q_i^b(n)$ is the relative frequency of i in position n of the bottom M code words. Next $q_i(n)$ and $r_i(n)$ are defined by

$$q_i(n) = \frac{1}{2}[q_i^t(n) + q_i^b(n)]$$

$$r_i(n) = \frac{1}{2}[q_i^t(n) - q_i^b(n)]$$

Thus

$$q_i^t(n)q_k^b(n) = q_i(n)q_k(n) + r_i(n)q_k(n) - q_i^t(n)r_k(n) \tag{76}$$

When (76) is substituted into (75), one of the resulting terms is $-(1/N)\sum_n \sum_{i,k} q_i(n)q_k(n)\mu_{ik}(1/2)$. In contrast to Eq. (1.40) of [6 part II], we have, for each n,

$$\sum_{i,k} - q_i(n)q_k(n)\mu_{i,k}(\tfrac{1}{2}) \leq [\sum_i q_i(n)h(i)] \sup_q \frac{\sum_{i,k} - q_i q_k \mu_{ik}(\frac{1}{2})}{\sum_i q_i h(i)} \tag{77}$$

Summing over n and using the energy constraint, we then have

$$\frac{1}{N}\sum_{n=1}^{N}\sum_{i,k} - q_i(n)q_k(n)\mu_{ik}(\tfrac{1}{2}) \leq \delta \sup_q \frac{\sum_{i,k} - q_i q_k \mu_{ik}(\frac{1}{2})}{\sum_i q_i h(i)} \tag{78}$$

For a given δ, the other terms resulting from substituting (76) into (75) go to zero with N as shown in lemmas 4.3, 4.4., and 4.5 of [6 part II]. Thus for any $\delta > 0$,

$$\frac{E(0,\delta)}{\delta} \leq \sup_q \frac{\sum_{i,k} - q_i q_k \mu_{ik}(\frac{1}{2})}{\sum_i q_i h(i)} \tag{79}$$

This completes the proof of theorem 5.

It would be desirable to also establish the result of theorem 3 for this multi-input case, but unfortunately no proof of that has yet been constructed. The results in [6 part II] do not extend in a simple way to the case of no constraint on N for a given energy D.

4) MULTIACCESS COMMUNICATION

Now suppose that, instead of one transmitter sending a code word to a receiver, there are J transmitters, each of which simultaneously sends a code word to a single receiver. Assume that each transmitter has the same channel input alphabet $\{0,1,...,K\}$ and uses a block code of the same constraint length N. Let M_j be the number of code words and $R_j = (\ln M_j)/N$ be the rate of the j^{th} transmitter, $1 \le j \le J$. Let $\{x_m(j), 1 \le m \le M_j\}$ be the code words of the j^{th} code, where $x_m(j) = x_{m1}(j),...,x_{mN}(j)$. Finally, let $h_j(k)$ be the energy associated with transmitter j using letter k and assume that each code word from each transmitter must satisfy the constraint

$$\sum_{n=1}^{N} h_j(x_{mn}(j)) \le N\,\delta_j\,;\ 1 \le m \le M_j,\ 1 \le j \le J \tag{80}$$

We suppose that the transmitters are all synchronized in the sense that the receiver gets a channel output y_n corresponding to the n^{th} channel input from each transmitter. We also assume that the channel is memoryless in the sense that if $x(j) = x_1(j),...,x_N(J)$ is given by

$$p_N(y)|x(1),...x(J)) = \prod_{n=1}^{N} p(y_n|x_n(1),...,x_n(J)) \tag{81}$$

when $p(y|x(1),...x(J))$ is the transition probability density defining the channel.

Given a code for each transmitter, and given a channel transition probability density, we assume that each source j independently produces a message m(j), uniformly distributed from 1 to M_j; the encoder at each transmitter j generates code word $x_{m(j)}(j)$, and the receiver maps the received y into J decoded outputs, $m^*(1),...,m^*(J)$. An error occurs if $m(j) \neq m^*(j)$ for any j, $1 \le j \le J$. There are two different definitions of error probability that are of interest here. The first, block error probability, P_e, is simply the probability that an error occurs over the ensemble of source inputs and channel outputs for a given set of codes. For the second, bit error proabability, we assume that M_j is a power of 2 for each j and that the input to encoder j is a sequence of $\log_2 M_j$ binary digits. The bit error probability for source j, $P_e^*(j)$ is then defined as the expected number of binary digits into encoder j that are decoded in error. The overall bit error probability P_e^* is then defined as $\max_j P_e^*(j)$.

We shall use P_e (as in the previous sections) to discuss the direct part of the coding theorem and exponential error bounds, and use bit error probability, P_e^*, for the converse to the coding theorem. For the converse, we view N as the number of uses of the channel over its lifetime rather than as block length for any given code. Thus, showing that P_e^* is bounded away from 0 for certain transmission rates applies to block codes, convolutional codes, block codes in which the blocks at different transmitters are not synchronized, block codes in which different transmitters have different block lengths, etc. Note that showing that P_e is bounded away from 0, and even showing that P_e approaches 1 with increasing N, does not imply that P_e^* is bounded away from 0.

We now define the achievable region for a multiaccess channel with energy constraints and establish the converge to the coding theorem. The coding theorem and converse for multiaccess channels without an energy constraint was established by Ahlswede [10] and Liao [11] and has been extended in many subsequent papers, for example [12-14]. However, none of the converses use bit error probability and none include general energy constraints. The inclusion of an energy constraint into the converse is non-trivial and forces us to look at some subtleties that appear to have been overlooked in the simpler existing proofs of the converse. We will not bother to prove the forward part of the coding theorem in general since it is a trivial extension of the result without an energy constraint.

Let $Q_j(k)$, $0 \leq k \leq K$ be a probability assignment on the channel input alphabet for transmitter j, $1 \leq j \leq J$, and consider the ensemble $X(1), X(2),...X(J),Y$ in which $X(1),...,X(J)$ are independent, $X(j)$ has the alphabet $\{0,1,...,K\}$ with probability assignment $Q_k(j)$ and Y is described by $p(y|x(1),...,x(J))$.

Let S be an arbitrary non-empty subset of the set $\{1,2,...,J\}$ of transmitters, let $\mathbf{X}(S) = X(i_1),X(i_2),...X(i_j)$ where $S = \{i_1,i_2,...i_j\}$, and let $\mathbf{X}(S^c) = X(i_{j+1})...X(i_J)$ where $\{i_{j+1}, i_{j+2},...,i_J\}$, is the complement of S. Finally let $I_Q(\mathbf{X}(S);Y \mid \mathbf{X}(S^c))$ be the average mutual information between $\mathbf{X}(S)$ and Y conditional on $\mathbf{X}(S^c)$ for the ensemble defined by $\mathbf{Q} = \{Q_k(j), 0 \leq k \leq K, 1 \leq j \leq J\}$. A vector $\mathbf{R} = (R_1,...,R_J)$ of input rates to the J transmitters is called <u>achievable</u> for the joint probability assignment $\mathbf{Q}$ if

$$0 \leq \sum_{i \in s} R_i \leq I_Q(X(S);Y \mid X(S^c)) \quad \text{for all S} \tag{82}$$

Let $\delta = \delta_1,...,\delta_J$ be a vector of allowable energies for the J transmitters. The pair $(\mathbf{R},\delta)$ is defined to be <u>directly achievable</u> if there is some $\mathbf{Q}$ such that $\mathbf{R}$ is achievable for $\mathbf{Q}$ and also such that $\mathbf{Q}$ satisfies the energy constraint, i.e.,

$$\sum_k h_j(k)Q_j(k) \leq \delta_j \quad \text{for } 1 \leq j \leq J \tag{83}$$

Finally the pair $(\mathbf{R},\delta)$ is defined to be <u>achievable</u> if it is in the convex hull of the set of $(\mathbf{R},\delta)$ that are directly achievable. The set of $\mathbf{R}$ such that $(\mathbf{R},\delta)$ is achievable for a given δ is denoted $\mathsf{R}(\delta)$. Note that $\mathsf{R}(\delta)$ is <u>not</u> the same as the convex hull of the set of $\mathbf{R}$ such that $\mathbf{R}$ is achievable for some $\mathbf{Q}$ satisfying (83); we shall shortly see an example of this difference.

Before proceeding, we must understand (82) a little better; for a given $\mathbf{Q}$, (82) is a set of 2^J-1 linear inequalities on the vector $\mathbf{R}$. The set of $\mathbf{R}$ that satisfy (82) is a convex polytope. Let $i_1,i_2,...,i_J$ be a permutation of the integers 1 to J, and let $S_j = \{i_1,i_2,...i_j\}$ for $1 \leq j \leq J$. Then we claim that the solution to

$$\sum_{i \varepsilon S_j} R_i = I_Q(X(S_j);Y \mid X(S_j^c)); \quad 1 \le j \le J \tag{84}$$

is an extreme point of the set of **R** that satisfy (82). It is easy to see that there is a unique **R** satisfying (84) for a given permutation, and it is not hard, by manipulating information inequalities, to see that this satisfies all the other inequalities in (82). This, however, is sufficient to guarantee that **R** is an extreme point. It is also not hard to see that the J! - 1 extreme points (not necessarily distinct) formed in this way constitute all the extreme points other than those with one or more nonzero rates.

Lemma: Let $\mathbf{Q}^{(1)},...,\mathbf{Q}^{(N)}$ be N joint distributions on the input alphabets, where for $1 \le n \le N$, $\mathbf{Q}^{(n)} = \{Q_j^{(n)}(k);\ 1 \le j \le J,\ 0 \le k \le K\}$ and let $\mathbf{X}^{(n)} = X_1^{(n)},...,X_J^{(n)}$ be the corresponding ensembles. Let $\lambda_n \ge 0$, $\sum_n \lambda_n = 1$. If

$$0 \le \sum_{i \varepsilon S} R_i \le \sum_{n=1}^{N} \lambda_n I(X^{(n)}(S);Y^n \mid X^{(n)}(S^c)) \quad \text{for all } S \tag{85}$$

and
$$\sum_{n=1}^{N} \lambda_n \sum_{k} h_j(k) Q_j^{(n)}(k) \le \delta_j \text{ for } 1 \le j \le J \tag{86}$$

then $(\mathbf{R},\delta)$ is achievable where $\mathbf{R} = (R_1,...R_J)$ and $\delta = (\delta_1,...,\delta_J)$.

Proof: We prove the lemma for N=2; the extension to arbitrary N follows immediately by induction on N. The set of rate vectors that satisfy (85) for a given $\mathbf{Q}^{(1)}$, $\mathbf{Q}^{(2)}$, and λ ($\lambda_1 = \lambda$, $\lambda_2 = 1-\lambda$) forms a convex polytope. As before, each extreme point of this polytope with all $R_i > 0$ satisfies the J equations, $1 \le j \le J$,

$$\sum_{i \varepsilon S_j} R_i = \lambda I_{Q^{(1)}}(X(S_j);Y \mid X(S_j^c)) + (1-\lambda) I_{Q^{(2)}}(X(S_j);Y \mid X(S_j^c)) \tag{87}$$

where $S_j = \{i_1,i_2,...,i_j\}$ for some permutation $i_1,...,i_J$ of the integers 1 to J. For a given permutation (and thus a given extreme point **R**), there is an extreme point $R^{(1)}$ of the set of **R** that satisfies (84) for $\mathbf{Q}^{(1)}$ and an extreme point $\mathbf{R}^{(2)}$ of the set of **R** that satisfies (84) for $\mathbf{Q}^{(2)}$. Clearly $\mathbf{R} = \lambda\mathbf{R}^{(1)} + (1-\lambda)\mathbf{R}^{(2)}$. This same result holds for all the extreme points of (87) involving some $R_i = 0$. Thus every extreme point of (85) (for N=2) is λ times an extreme point of the solutions of (82) for $\mathbf{Q}^{(1)}$ plus $(1-\lambda)$ times an extreme point of the solutions for (82) for $\mathbf{Q}^{(2)}$. Since every point satisfying (85) is a convex combination of the extreme points of (85), it is also equal to λ times a point achievable for $\mathbf{Q}^{(1)}$ plus $(1-\lambda)$ times a point achievable for $\mathbf{Q}^{(2)}$. Thus $(\mathbf{R},\delta)$ is a convex combination of a point $(\mathbf{R}^{(1)},\delta^{(1)})$ directly achievable with $\mathbf{Q}^{(1)}$ and a point $(\mathbf{R}^{(2)},\delta^{(2)})$ achievable with $\mathbf{Q}^{(2)}$.

We are now ready to state the coding theorem and converse; these show that reliable communication is possible if all components of $\mathbf{R}$ are less than those for some $\mathbf{R}^*$ for which $(\mathbf{R}^*,\delta)$ is in the achievable region and that reliable communication is impossible if $(\mathbf{R},\delta)$ is outside of the achievable region.

Theorem 6: Assume that $(\mathbf{R},\delta)$ is achievable for any given memoryless multiaccess channel with J transmitters. Then for any $\varepsilon > 0$, $\mu > 0$, there is an $N(\varepsilon,\mu)$ such that for all block lengths $N \geq N(\varepsilon,\mu)$ there exists a code for each transmitter satisfying the energy constraint δ_j with $M_j \geq \exp[N(R_j-\mu)]$ and there exists a decoding rule such that $P_e \leq \varepsilon$.

We shall not prove this theorem, since, as stated before, it is a minor extension adding energy constraints to the theorem of [10] and [11]. We shall prove the following converse, however, since it is a non-trivial extension. The theorem states that, for a given energy constraint $\delta = (\delta_1,...\delta_J)$, if the bit error probability $P_e{}^*$ is small, then the rate vector $\mathbf{R} = (R_1,...,R_j)$ must be close to an achievable rate vector.

Theorem 7: Consider any given memoryless multiaccess channel with J transmitters and with energy constraint $\delta = (\delta_1,...\delta_J)$. Assume that for $1 \leq j \leq J$ the number of code words M_j for the j^{th} transmitter is a power of 2 and let $R_j = (\ln M_j)/N$ for an arbitrary block length N. Assume that the bit error probability $P_e{}^*$ satisfies $P_e{}^* \leq \varepsilon$. Let $H(\varepsilon) = -\varepsilon \log_2 \varepsilon -(1-\varepsilon) \log_2(1-\varepsilon)$ and let $R^*_j = R_j(1-H(\varepsilon))$. Then $(\mathbf{R}^*,\delta)$ is achievable where $\mathbf{R}^* = (R^*_1,...,R^*_J)$.

Proof: Let $S = \{j_j,j_2,...,j_i\}$ be an arbitrary non-empty subset of the transmitters. The number of binary digits encoded and transmitted by the j^{th} transmitter over the channel block length N is given by $R_jN/\ln 2$. Thus the number of binary digits encoded and transmitted by transmitters in S is given by

$$L = \frac{N}{\ln 2} \sum_{j \varepsilon S} R_j \tag{88}$$

Let U^L be the ensemble of these binary digits (assumed IID and equally likely) and let V^L be the ensemble of the corresponding decoded binary digits. The average error probability over these L binary digits is at most $P_e{}^* \leq \varepsilon$, so from Fano's inequality extended to a sequence of letters (Theorem 4.3.2 in [1]), we have

$$H(\varepsilon) \geq \frac{H(U^L|V^L)}{L \ln 2} \tag{89}$$

where the entropy $H(U^L|V^L)$ is in natural units. Since $H(U^L) = L \ln 2$, this can be rewritten, with the help of (85) as

$$[1 - H(\varepsilon)] \sum_{j \varepsilon S} R_j \leq \frac{I(U^L;V^L)}{N} \tag{90}$$

From the data processing theorem, we have

$$I(U^L;V^L) \leq I(X^N(S);\, Y^N) \tag{91}$$

where $X^N(S) = X^N(j_1)X^N(j_2)...X^N(j_i)$ is the ensemble of transmitted letters from the transmitters in S for the given codes and Y^N is the ensemble of received letters. Using the independence of the code words transmitted by the different transmitters,

$$I(X^N(S)\,;Y^N) \leq I(X^N(S);Y^N|X^N(S^C)) \tag{92}$$

Since the channel is memoryless,

$$I(X^N(S);Y^N|X^N(S^c)) \leq \sum_{n=1}^{N} I(X_n(S);Y_n|X_n(S^c)) \tag{93}$$

Combining (90) to (93) and recalling that $R^*_j = [1 - H(\varepsilon)]R_j$,

$$\sum_{j \varepsilon S} R_j^* \leq \frac{1}{N} \sum_{n=1}^{N} I(X_n(S);Y_n|X_n(S^c)) \tag{94}$$

Since (94) is valid for all non-empty sets S, the previous lemma asserts that (R^*,δ) is achievable.

For the remainder of the paper, we restrict attention to the special case in which each transmitter has a binary alphabet; we assume, as in section 2, that unit energy is used by input 1 and zero energy is used by input 0. Thus an energy constraint δ_j on transmitter j means that each code word used by transmitter j in a code of block length N contains at

$$\sum_{j\varepsilon S} R_j^* \le \frac{1}{N} \sum_{n=1}^{N} I(X_n(S);Y_n|X_n(S^c)) \qquad (94)$$

Since (94) is valid for all non-empty sets S, the previous lemma asserts that (R^*,δ) is achievable.

For the remainder of the paper, we restrict attention to the special case in which each transmitter has a binary alphabet; we assume, as in section 2, that unit energy is used by input 1 and zero energy is used by input 0. Thus an energy constraint δ_j on transmitter j means that each code word used by transmitter j in a code of block length N contains at most $N\delta_j$ ones. For simplicity we often assume that each transmitter has the same energy constraint δ, i.e., $\delta_j = \delta$ for $1 \le j \le J$. We also make the simplifying assumption that the channel output depends statistically only on whether all transmitters send zero or whether one or more transmitters transmit one. Let $p_0(y)$ be the transition probability density given that all transmitters send zero and let $p_1(y)$ be the density given that one or more transmitters send 1.

This multiaccess channel can be viewed as the cascade of two channels, the first of which is a multiaccess or channel for which the output is binary and equal to the logical or of the set of inputs; the second is a point to point channel with binary inputs and transition probability densities $p_1(y)$ and $p_0(y)$. This is not an entirely realistic model since the output of a multiaccess channel usually depends on how many transmitters are simultaneously transmitting a one. Thus a more realistic model would be an adder channel [15] cascaded with a point to point channel with the input alphabet $\{0,1,...,J\}$. We have chosen the simpler model here since our objective is to gain insight into multiaccess coding in the simplest context.

Assume that the J transmitters are numbered so that the rate vector $\mathbf{R} = (R_1,R_2,...R_J)$ satisfies $R_1 \ge R_2 \ge ... \ge R_J \ge 0$. Let $\mathbf{Q}$ be the input probability assignment in which each transmitter uses 1 with probability δ and 0 with probability $1-\delta$. The set of rate vectors $\mathbf{R}$, subject to the above ordering, that are feasible for $\mathbf{Q}$ is given by the following inequalities, $1 \le j \le J$:

$$0 \le \sum_{i\varepsilon S_j} R_i \le I_Q(X(S_j);Y|X(S_j^{\,c})) \qquad (95)$$

where $S_j = \{1,...,j\}$. From the symmetry between the transmitters, it can be seen that all the inequalities in (82) are satisfied if these inequalities are satisfied. Let $I(\varepsilon)$ be the mutual information on the point to point part of the multiaccess channel given that input 1 is used with probability ε, i.e.,

$$I(\varepsilon) = \int_y dy[\varepsilon p_1(y)\ln \frac{p_1(y)}{\varepsilon p_1(y)+(1-\varepsilon)p_0(y)} + (1-\varepsilon)p_0(y)\ln \frac{p_0(y)}{\varepsilon p_1(y)+(1-\varepsilon)p_0(y)}] \quad (96)$$

Then we can evaluate the right hand side of (95) to yield

$$0 \le \sum_{i\varepsilon S_j} R_i \le (1-\delta)^{J-j}\, I(1-(1-\delta)^j) \quad (97)$$

In the important special case where all the transmitter rates are equal to some fixed R, the most stringent inequality above is j = J, where we have

$$JR \le I(1-(1-\delta)^J) \quad (98)$$

Let ε_{max} be the value of ε that maximizes $I(\varepsilon)$; i.e. $I(\varepsilon_{max})$ is the capacity of the point to point channel. If $1 - (1-\delta)^J > \varepsilon_{max}$, then the energy constraint on each channel should be reduced to the point where $1 - (1-\delta)^J = \varepsilon_{max}$. In this case, the sum of the rates from the individual transmitters is limited only by the capacity of the point to point channel. This remains true for unequal rates so long as each inequality in (97) is satisfied.

For situations in which some rates are much larger than others, the directly achievable region can be enlarged by using input assignments other than the one above. In general, for an energy constraint δ on each transmitter, let transmitter j use input 1 with probability $\delta_j \le \delta$. For such a **Q**, the average mutual informations in (95) are evaluated to yield

$$0 \le \sum_{i\varepsilon S_j} R_i \le [\prod_{i=j+1}^{J} (1-\delta_j)]I[1 - \prod_{i=1}^{j} (1-\delta_j)] \quad (99)$$

There appears to be no simple characterization of the set of directly achievable rates, although it is not hard to verify whether or not a given rate vector is directly achievable. One starts with (99) for j = 1 and $\delta_1 = \delta$; this places a lower bound on the product of $(1-\delta_j)$ from 2 to J. Next one considers (99) for j = 2 and one chooses δ_2 to provide the smallest possible lower bound on the product of $(1-\delta_i)$ from 3 to J. One continues with larger j until either all constraints are met or some constraint can not be met.

The directly achievable region above is typically smaller than the entire achievable region. Consider a convex combination of J rate vector / energy vector pairs in which the j^{th}, say

$(\mathbf{R}^{(j)}, \delta^{(j)})$ has $R_i^{(j)} = \delta_i^{(j)} = 0$ for all $i \neq j$. This corresponds to frequency or time division multiplexing between the transmitters, as will be discussed later. The achievability inequalities of (82) reduce to the single inequality

$$0 \leq R_j^{(j)} \leq I_{Q^{(j)}}(X_j;Y) = I(\delta_j^{(j)}) \tag{100}$$

Letting $\lambda_1,\ldots,\lambda_J$ be non-negative numbers summing to 1, consider the convex combination $\mathbf{R} = \sum_j \lambda_j \mathbf{R}^{(j)}$, $\delta = \sum_j \lambda_j\, \delta_j^{(j)}$. This means that $R_j = \lambda_j R_j^{(j)}$ and $\delta_j = \lambda_j \delta_j^{(j)}$. From (100), then, $(\mathbf{R},\delta)$ is achievable if

$$0 \leq R_j \leq \lambda_j I(\delta_j/\lambda_j) \tag{101}$$

We shall refer to the rate region defined by (101 as the frequency division multplexing (FDM) region. In the special case where all rates are equal to some fixed R, all energy constraints are equal to some fixed δ, and $\lambda_j = 1/J$ for $1 \leq j \leq J$, this reduces to

$$0 \leq JR \leq I(J\delta) \tag{102}$$

Note that if $J\delta \leq \varepsilon_{max}$, then the rates achievable via these convex combinations are strictly larger than the directly achievable rates.

Note that $NJ\delta$ is an upper bound on the number of ones that can enter the point to point channel so that $I(J\delta)$ is an upper bound (for $J\delta \leq \varepsilon_{max}$) on the mutual information per letter on the point to point channel. Thus $JR = I(J\delta)$ is the outer limit of the achievable region for equal rates. Figure 2 illustrates the achievable rate region for $J = 2$. The extreme points of (97) are shown in the figure as lying on the boundary of the feasible region. It is conjectured that this is true in general.

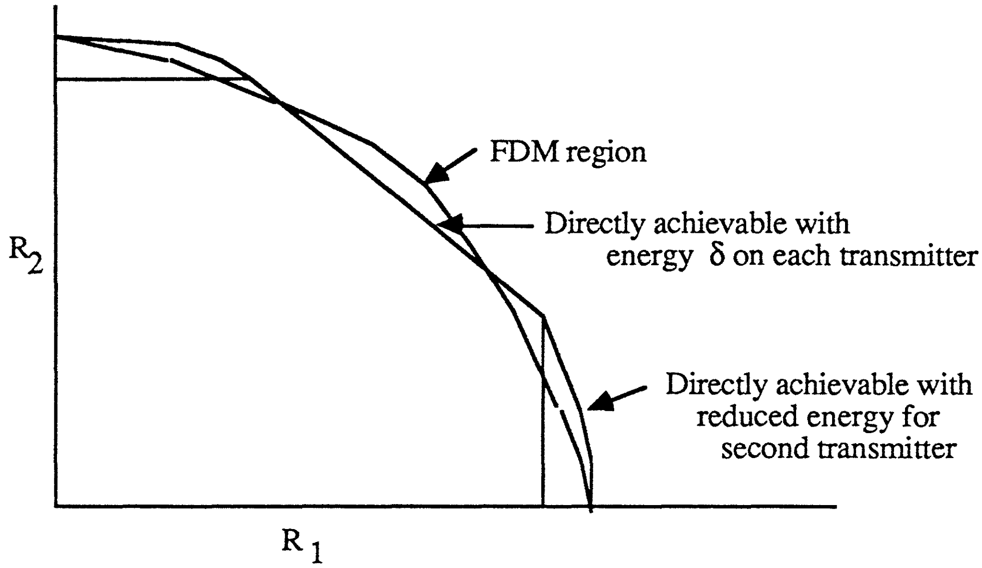


Figure 2

Achievable region for two transmitter case. Not shown is the convex hull between the FDM region and the directly achievable region.

Next let us investigate the use of coding to achieve rates in the interior of the feasible region. First suppose that the desired set of rates lie in the interior of the FDM region. Then for an overall block length N, each transmitter can be allocated a disjoint set of positions in the block, allocating $\lambda_j N$ positions (which can be viewed as a frequency band) to the j^{th} transmitter. The j^{th} transmitter than uses a code within its sub-block of $\lambda_j N$ positions using at most δN ones in each code word, or δ/λ_j ones per allocated position. Since there is no interference between transmitters, the decoder can decode each transmitter's code words independently. Thus the system reduces to J independent point to point channels with energy constraints; conventinal coding techniques can be used and the error probability results of section 2 are directly applicable.

The FDM approach is often undesirable for data transmission. The major reason for this is not so much that it doesn't achieve the full achievable rate region but rather that data souces tend to be bursty. A transmitter must be allocated enough bandwidth to send its messages with tolerable delay, but this leads to the bandwidth being wasted when there is nothing to send. One solution to this problem is to allocate bandwidth to a transmitter only when it has data to send. This, however, introduces scheduling delay and distributed scheduling algorthms. While this approch is often appropriate, it is worthwhile to explore other approaches. Another disadvantage of FDM is that it is susceptible to jamming, either intentional or unintentional.

Next, suppose that the desired set of rates lie in the interior of the directly achievable region for some input probability assignment Q in which transmitter j uses input 1 with probability

$\delta_j \leq \delta$, $1 \leq j \leq J$. Consider a randomly chosen ensemble of codes in which each transmitter j independently chooses each component of each code word to be 1 with probability δ_j subject to a shell constraint. This can be viewed as an abstraction of spread spectrum coding (see, for example [16,17]. In practice, a spread spectrum transmitter pseudo-randomly chooses a subset of the available N components and then uses algebraic or convolutional coding to select code words within that subset; all code words use input zero (i.e. don't transmit) for all components outside that subset.

One advantage of this approach is that if there are a very large set of transmitters, but only J of them are transmitting in a block (i.e., the others are sending a constant stream of zeros), then the error probability is the same as if only J transmitters existed. Thus no scheduling is required and there is no loss of efficiency due to bursty sources (other than that J becomes a random variable and the error probability becomes large as J becomes too large).

Decoding is a major problem with codes of this type. We discuss three possibilities. The first, and simplest, is to decode each transmitter's code words independently, averaging over the random coding ensembles of the other transmitters. Consider a given transmitter j. Given that j transmits a 1, the output y has the conditional density $p_1(y)$. Given that j transmits a 0, however, the output y depends on whether one or more other transmitters send a 1. The probability that one or more of the other transmitters send a 1, averaged over their ensembles of codes, is $\varepsilon_j = 1 \prod_{i\neq j}(1-\delta_i)$. thus, given that transmitter j transmits 0, the output has the conditional density $p^*_0(y) = \varepsilon_j p_1(y) + (1-\varepsilon_j)p_0(y)$. The reliability function for transmitter j, using independent decoding is then given directly by the results of section 2, using $p^*_0(y)$ in place of $p_0(y)$. For δ_j small, the asymptotic results of section 2 apply. The fact that the other transmitters use a shell constraint complicates this result somewhat but can be shown to not effect the reliability function. The reliability function for transmitter j goes to 0 at $R_j = I(X_j;Y)$. This maximum rate is given by

$$I(X_j;Y) = I(X_1,X_2,\ldots,X_J;Y) - I(X_1,\ldots,X_{j-1}X_{j+1},\ldots,X_J;Y|X_j)$$

Using (99), this is

$$I(X_j;Y) = I\big[1 - (1-\delta_j)(1-\varepsilon_j)\big] - (1-\delta_j)I\big[\varepsilon_j\big] \tag{103}$$

For the special case in which J is large, $\delta_i = \delta$ for all i, and $1 - (1-\delta)^J = \varepsilon_{max}$ (i.e., δ is set to limit multiaccess interference rather than by intrinsic energy constraints), this becomes

$$R_j \leq I(X_i;Y) \approx \frac{-\ln(1-\varepsilon_{max})I(\varepsilon_{max})}{J} \tag{104}$$

Comparing this with (98), we see that the use of independent decoding reduces the achievable rates by a factor of $-\ln(1-\varepsilon_{max})$. For cases where δJ is small relative to 1, this loss in achievable rate is less severe.

For the second possibility with respect to decoding, consider a situation in which the rates are ordered, $R_1 \geq R_2 \ldots \geq R_J$, and, for some **Q**

$$R_j \leq I_Q(X_j;Y \mid X(S_j^c)) \; ; \; 1 \leq j \leq J \tag{105}$$

where $S_j = \{1,...,j\}$. By comparison with (84), it can be seen that the rate vector **R** for which those inequalities are satisfied with equality is an extreme point of the achievable region for **Q**.

If all the inequalities of (105) are satisfied with strict inequality, then one can select a randomly chosen code for each transmitter. One can decode such a set of codes by first decoding for transmitter J. Conditioning on the decoded word for J, one next decodes the code word for J-1, and so forth down to the code word for transmitter 1. Viterbi [18] has investigated this decoding strategy for an additive white Gaussian noise channel. He and Verdu [19] consider also the problems of synchronization for this channel. We shall consider the reliability function for this type of decoding shortly, but first treat the final possibility for decoding.

Suppose that $\mathbf{R} = (R_1,...R_J)$ is achievable for input distribution **Q**. Suppose that the code words from all decoders are decoded simultaneously, using, for example, maximum likelihood decoding on the set of J codes. No computationally attractive decoding techniques are known for simultaneous decoding, with the possible exception of sequential decoding as developed by Arikan [20]. The random coding bound to error probability for this case was derived by Slepian and Wolf [21] and was extended to shell constraints in [22]. We now develop this bound using a shell constraint, for the multiaccess channels under consideration here.

The probability of error, P_e, for one or more of the J code words can be expressed as

$$P_e = \Sigma_S P_e(S) \tag{106}$$

where $S = \{i_1,i_2,...,i_j\}$ is a non-empty set of transmitters and $P_e(S)$ is the probability that the code words for $i_1,i_2,...,i_j$ are all decoded incorrectly and that the code words for S^c are all decoded correctly. Since the number of sets S is 2^J-1, independent of the block length N, the reliability function is simply the minimum of the reliability functions over S.

Consider a particular set S and let

$$\varepsilon = 1 - \prod_{i \varepsilon S} (1-\delta_i); \; \sigma = 1 - \prod_{i \varepsilon S^c} (1-\delta_i) \tag{107}$$

Thus ε is the probability that one or more transmitters in S send a 1 in a particular component of the block and σ is the probability that one or more transmitters in S^c send a 1. $P_e(S)$, averaged over the ensemble of codes, can be upper bounded by the probability that the set of transmitted code words in S is less likely, given y, than some other set of transmitted words in S, conditional on, and then averaged over, some set of code words in S^c. Since transmitters in S collectively use input 1 to the point to point part of the channel with probability ε and the other transmitters collectively use input 1 with probability σ, the random coding bound [22] is given by

$$P_e(S) \le A \exp[-N E_r(\sum_{i\varepsilon S} R_i)] \tag{108}$$

$$E_r(\sum_{i\varepsilon S} R_i) = \max_{0\le\rho\le 1, r\ge 0} \{E_0(\rho,r) - \rho \sum_{i\varepsilon S} R_i] \tag{109}$$

$$E_0(\rho,r) = -\ln \Big\{ \sigma \int_y [\varepsilon\, e^{r(1-\varepsilon)} p_1(y)^{\frac{1}{1+\rho}} + (1-\varepsilon) e^{-r\varepsilon} p_1(y)^{\frac{1}{1+\rho}}]^{1+\rho} dy$$

$$+ (1-\sigma) \int_y [\varepsilon\, e^{r(1-\varepsilon)} p_1(y)^{\frac{1}{1+\rho}} + (1-\varepsilon) e^{-r\varepsilon} p_0(y)^{\frac{1}{1+\rho}}]^{1+\rho} dy \Big\}$$

$$= -r\varepsilon(1+\rho) - \ln\Big\{ \sigma[\varepsilon e^r + (1-\varepsilon)]$$

$$+ (1-\sigma) \int_y [\varepsilon e^r p_1(y)^{\frac{1}{1+\rho}} + (1-\varepsilon) p_0(y)^{\frac{1}{1+\rho}}]^{1+\rho} dy \Big\} \tag{110}$$

The parameter A in (108) can be upper bounded as in [1], but it is not exponential in N, so that $E_r(\Sigma_{i\varepsilon S} R_i)$ is a lower bound to the reliability function for S and $\min_S E_r(\Sigma_{i\varepsilon S} R_i)$ is a lower bound to the overall reliability function. It is positive if all the inequalities in (105) are satisfied with strict inequality. $E_r(\Sigma_{i\varepsilon S} R_i)$ can be approximated by the techniques of section 2, but the approximations are not very close if ε is an appreciable fraction of 1.

A similar analysis of error probability can be used if the rates are ordered and satisfy (105) with strict inequality. Since the code word for the j^{th} transmitter is now decoded regarding the outputs of the first j-1 transmitters as noise, the probability of error for the j^{th} transmitter, conditional on decoding transmitters j+1 to J correctly, is given by (108) and (109), with (110) replaced by

$$E_0(\rho,r) = -r\delta_j(1+\rho) - \ln \Big\{ \sigma[\delta_j e^r + (1-\delta_j] + (1-\sigma) \int_y [\delta_j e^r p_1(y)^{\frac{1}{1+\rho}} + (1-\delta_j)p_0^*(y)^{\frac{1}{1+\rho}} dy \Big\} \tag{111}$$

where $\sigma = 1 - \Sigma_{i>j}(1-\delta_i)$ and

$$p_0^*(y) = \prod_{i<j} (1-\delta_i)p_0(y) + [1-\prod_{i<j} (1-\delta_i)]p_1(y) \tag{112}$$

These exponents are positive if (105) is satisfied with strict inequality, but the exponents are somewhat smaller than those for simultaneous decoding.

These results remain valid if the blocks used by the different transmitters are not synchronized (see Hui and Humblet [23] for a discussion of the unsynchronized case), but again we note that our model is not sufficient to discuss the problem of lack of component synchronization. For rates that are achievable only as convex combinations of directly achievable rates, one can time share between codes that are directly achievable. This of course requires block synchronization. As shown in [23], only directly achievable (R,δ) pairs are achievable with no block synchronization between the transmitters. As shown in [24], however, if there is a bounded uncertainity between the components of the different transmitters, time sharing can still be used by using block lengths large relative to the maximum uncertainity.

REFERENCES

[1] R. G. Gallager, *Information Theory and Reliable Communication,* New York: Wiley, 1968

[2] A. J. Viterbi and J. K. Omura, *Principles of Digital Communication and Coding*, New York: McGraw Hill, 1979.

[3] J. M. Wozercraft and I. M. Jacobs, *Principles of Communication Engineering*, New York: Wiley, 1965.

[4] R. S. Kennedy, *Fading Dispersive Channels*, New York: Wiley, 1969.

[5] J. S. Richters, "Communication over Fading Dispersive Channels," M.I.T. RLE Tech Report 464, 1967.

[6] C. E. Shannon, R. G. Gallager, and E. R. Berlekamp, "Lower Bounds to Error Probability for Coding on Discrete Memoryless Channels," Information and Control 10, pp. 65-103 (part I), pp. 522-552 (part II).

[7] R. J. McEliece, "Practical Codes for Photon Communication," IEEE Trans. IT, Vol. IT-27,#4, July 1981.

[8] J. L. Massey, "Capacity, Cutoff Rate, and Coding for a Direct Detection Optical Channel," IEEE Trans. Comm., VolCOM-29, #11, pp 1615-1621, Nov. 1981.

[9] P. A. Humblet, "Error Exponents for a Direct Detection Optical Channel," M.I.T. LIDS Report LIDS-P-1337, Oct. 1983; also Proc. 23rd Allerton Conference on Communication, Control and Computing, Monticello, Ill., Oct. 1985.

[10] R. Ahlswede, "Multi-way Communication Channels," Proc. 2nd Inter. Symp. I.T., Tsahkadsor, Armenian S.S.R., 1971.

[11] H. Liao, "A Coding Theorem for Multiple Access Communications," in Proc. Int. Symp. I.T., Asilomar, Calif, 1972; also "Multiple Access Channels," Ph.D. thesis,, Univ. of Hawii, 1972.

[12] A. El Gamal and T. M. Cover, "Multiple user Information Theory," Proc. IEEE, vol 68, pp. 1466-1483, De. 1980

[13] E. C. van der Meulen, "A Survey of Multi-way Channels in Information Theory 1961-1976," IEEE Trans. I.T., vol IT-23, Jan 1977.

[14] I. Csiszar and J. Korner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, New York: Academic Press 1981.

[15] T. Kasami, S. Lin, V. K. Wei, and S. Yamamura, "Graph Theoretic Approaches to the Code Construction for the Two-user Multiple-access Binary Adder Channel," IEEE Trans. I.T., Vol IT-20, pp. 114-130, Jan. 1983.

[16] M. B. Pursley, "Frequency-Hop Transmission for Satellite Packet Switching and Terrestrial Packet Radio Networks," IEEE Trans. I.T., Vol IT32, pp 652-668, Sept. 1986.

[17] Special Issue on Spread Spectrum: IEEE Trans. Comm., May 1982.

[18] A. J. Viterbi, "Spread Spectrum Multiple Access with Binary Modulation can Approach Shannon Capacity for the Aggregate Gaussian Channel" Manuscript.

[19] S. Verdu, "Maximum Likelihood Sequence Detection for Intersymbol Interference Channels: A New Upper Bound on Error Probability," IEEE Trans. I.T., vol IT-33, pp. 62-69, Jan. 1987.

[20] E. Arikan, "Sequential Decoding for Multiple Access Channels," Ph. D. Thesis, Dept. of EECS, M.I.T., Nov. 1985.

[21] D. Slepian and J. Wolf, "A Coding Theorem for multiple access channels with correlated sources," BSTJ, vol. 52, pp. 1037-1076, Sept. 1973.

[22] R. G. Gallager, "A Perspective on Multiaccess Channels," IEEE Trans. I.T., vol IT-31, pp. 124-143, Mar. 1985.

[23] J. Y. N. Hui and P. A. Humblet, "The Capacity region of the Totally Asynchronous Multiple-access Channel," IEEE Trans. I.T., vol IT-31, pp. 207-217, Mar. 1985.

[24] T. M. Cover, R. J. McEliece, and E. C. Posner, "Asynchronous Multi-access channel Capacity," IEEE Trans. I.T., vol IT-26,pp. 291-298, Nov. 1980.