CODING AND MULTIACCESS FOR THE ENERGY LIMITED RAYLEIGH FADING CHANNEL

by

İbrahim Emre Telatar
B.S. Middle East Technical University, Ankara
(1986)

SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE
DEGREE OF

MASTER OF SCIENCE
IN ELECTRICAL ENGINEERING AND COMPUTER SCIENCE

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
May 1988

Signature of Author ______________________________
Department of Electrical Engineering and Computer Science
May 16, 1988

Certified by ______________________________
Robert G. Gallager
Thesis Supervisor

Accepted by ______________________________
Arthur C. Smith
Chairman, Departmental Commitee on Graduate Students

CODING AND MULTIACCESS FOR THE ENERGY LIMITED RAYLEIGH FADING CHANNEL

by

İbrahim Emre Telatar

SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE
DEGREE OF
MASTER OF SCIENCE
IN ELECTRICAL ENGINEERING AND COMPUTER SCIENCE
MAY 1988

Abstract

There are many situations when the use of a communication channel is limited by constraints on energy rather than bandwidth. Examples of such channels are broadband additive Gaussian noise channels, fading dispersive channels and quantum optical channels. The above cases can be modeled by a discrete-time, discrete-input, continuous-output, memoryless channel, with constraints on the composition of codes for the channel. First we restrict our attention to a single user channel, and evaluate the channel capacity and reliability function for the case of a Rayleigh fading channel.

We also consider multiaccess use of such channels, and evaluate the possible throughput in the case of equal energy, equal rate users. Also given is a simple test to compare random access with frequency division multiplexing, in the above mentioned situations.

Thesis Supervisor: Robert G. Gallager

Title: Professor of Electrical Engineering

Acknowledgements

I thank my parents, Prof. Dr. Ferzan Telatar and Prof. Dr. Hasan Telatar, for their love and encouragement.

I thank my advisor, Prof. Robert Gallager for his guidance and initiative.

I thank my friends John Spinelli, Rajesh Pankaj, Zhi–Quan Luo, Ying Li, Whay Lee, Diana Dabby and Mohamad Akra for their help and support.

This work was in part supported by a Vinton Hayes Fellowship and an NSF grant NSF–ECS–83/0698. This funding was greatly appreciated.

Contents

List of Figures

Chapter 1

Introduction

In many communication situations the use of a communication channel is limited by constraints on energy rather than bandwidth. Equivalently, the number of degrees of freedom offered by the channel may be too large for a user to utilize effectively. Examples of such channels are broadband additive Gaussian noise channels, fading dispersive channels [Ken 69] and quantum optical channels.

The above cases can be modeled by a discrete-time, discrete-input, continuous-output, memoryless channel, with constraints on the composition of codes for the channel. First we restrict our attention to a single user, binary input case, with symbol 0 representing the zero energy input and symbol 1 representing the positive energy input. The energy constraint is incorporated by restricting the fraction of 1's in any given code-word.

In [Gal 87] the channel capacity for such channels are calculated and bounds are given for the reliability function, both quantities normalized with respect to energy. In the limit of infinite bandwidth, it is shown that the reliability function can be evaluated exactly, and the restriction to binary inputs is shown to be essentially optimal.

In the present work, taking the Rayleigh fading channel as a model, we investigate the dependence of channel capacity and reliability function to energy constraints, and how these quantities approach their limiting behavior as constraints get tight. We also consider multiaccess use of such channels, and evaluate the possible throughput in the case of equal energy, equal rate users. A simple test to compare random access with frequency division multiplexing is developed for the above mentioned situation.

Chapter 2 formally introduces the concepts from information theory which are used in the later chapters. Chapter 3 presents the single user instance of the energy limited communication problem. It is well known that, for a Gaussian channel with infinite bandwidth, the channel capacity in terms of bits/power is independent of the input power. Also, the channel capacity of a Rayleigh fading channel with infinite bandwidth approaches that of the Gaussian channel if the signal to noise ratio is high. We will show that, the Rayleigh fading channel approaches this infinite bandwidth behavior much slower than the Gaussian channel. Chapter 4 investigates the multiuser case, in which we mainly concentrate on the case of a large number of users.

Chapter 2

Background

In this chapter we present some notions from information theory which will be used in the later chapters. Apart from Section 2.5 the chapter is a summary of the first five chapters of [Gal 68]. Section 2.5 on the other hand mainly follows [CsK 81].

2.1 Communication Systems

For the purposes of information theory, a communication system can be modeled as shown

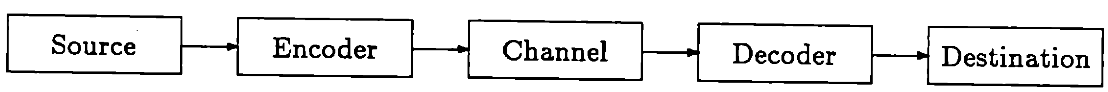

Figure 2.1: Model of a Communication System.

in Figure 2.1. The elements of the model are the following:

Source: The source is modeled by a stochastic process, usually a random sequence of letters drawn from a countable source alphabet.

Encoder: The encoder processes the source output, and prepares it for transmission over the channel.

Channel: The channel represents the transmission medium and will be specified by the triplet of the channel input alphabet, channel output alphabet and a probability measure on the set of output sequences for each input sequence.

Decoder: The task of the decoder is to undo the operations performed by the encoder and to prepare the channel output for the destination.

Destination: The destination is the intended receiver of the message the source generates.

One of the most important results of information theory is that encoding and decoding can be performed in two steps: source encoding and decoding, and channel encoding and decoding. With this result the encoding operation is broken down into two independent steps. In the first the source output is encoded into a string of binary symbols, which serve as a universal medium into which all sources are mapped. The channel encoder then operates on the binary sequences and performs a mapping between the set of binary sequences and the set of channel input sequences.

A particularly simple class of channels is the class of *discrete memoryless channels.* In this channel model, the input and output alphabets of the channel are finite, and each letter in the output sequence is statistically dependent only on the corresponding letter in the input sequence. Hence the probability law relating the output sequence and the input sequence can be described by a conditional probability assignment $P(j \mid k)$, with k and j representing the channel input and output symbols respectively.

2.2 Definitions

A *discrete source* is a sequence $\{u_i\}_{i=1}^{\infty}$ of random variables taking values in a finite set $U_i = U$ called the *source alphabet.* A source is called *stationary* if the joint distribution of $u_{i+1}, \ldots, u_{i+n}$ is independent of i for any n.

Given an ensemble X, i.e., a set X with a probability assignment P on its elements, we say the entropy of X is

$$H(X) = \sum_{x \in X} P(x) \log P(x) \,. \tag{2.1}$$

For a stationary source we define

$$H_L(U) = \frac{1}{L} H(U_1, \ldots, U_L) \,, \tag{2.2}$$

where $H(U_1, \ldots, U_L)$ denotes the entropy of the joint ensemble $U_1 \times \cdots \times U_L$. $H_L(U)$ is nonincreasing with L and we define

$$H_\infty(U) = \lim_{L\to\infty} H_L(U). \tag{2.3}$$

Given two ensembles X and Y we define the *entropy of X conditional on Y* as

$$H(X \mid Y) = \sum_{x\in X}\sum_{y\in Y} P(x,y)\log P(x \mid y). \tag{2.4}$$

We also define the mutual information between these two ensembles as

$$I(X;Y) = \sum_{x\in X}\sum_{y\in Y} P(x,y)\log \frac{P(y \mid x)}{P(y)}. \tag{2.5}$$

Given three ensembles X, Y and Z, the mutual information between X and Y conditional on Z is defined as

$$I(X;Y \mid Z) = \sum_{x\in X}\sum_{y\in Y}\sum_{z\in Z} P(x,y,z)\log \frac{P(y \mid x,z)}{P(y \mid z)}. \tag{2.6}$$

Given finite sets X and Y, a *discrete channel with input alphabet X and output alphabet Y* is a set of conditional probability assignments $P_N\left(y^N \mid x^N\right)$ for each N, $y^N = (y_1, \ldots, y_N) \in Y^N$, and $x^N = (x_1, \ldots, x_N) \in X^N$. A channel is called *memoryless* if

$$P_N\left(y^N \mid x^N\right) = \prod_{n=1}^{N} P(y_n \mid x_n) \tag{2.7}$$

for some P. Hence a discrete memoryless channel is specified by a transition probability matrix $P(j \mid i)$, with i ranging over the channel inputs and j ranging over the channel outputs. The capacity C of a discrete memoryless channel is defined as the maximum mutual information between the channel input and output ensembles,

$$C = \max_{\mathbf{Q}} I(X;Y) = \max_{\mathbf{Q}} \sum_k \sum_j Q(k)\, P(j \mid k) \log \frac{P(j \mid k)}{\sum_i P(j \mid i)\, Q(i)}. \tag{2.8}$$

where $\mathbf{Q}$ ranges over all possible input probability distributions. The capacity of the channel turns out to be an important quantity as the next two sections will show.

2.3 Converse to the Coding Theorem

In this section we will prove the converse to the coding theorem, which states that reliable transmission is not possible over a channel at rates above capacity. The way we prove the theorem is to lower bound the information between the source and destination alphabets, and then to relate the uncertainty left in the input alphabet to the probability of error. The first section, will show that information behaves like an incompressible fluid when passed through a channel.

2.3.1 Data Processing Theorem

Given a communication system with a discrete source emitting a source sequence $u_1, u_2, \ldots$, an encoder which generates $x_1, x_2, \ldots$, a channel with output $y_1, y_2, \ldots$, and a decoder generating $v_1, v_2, \ldots$, as shown in Figure 2.2.

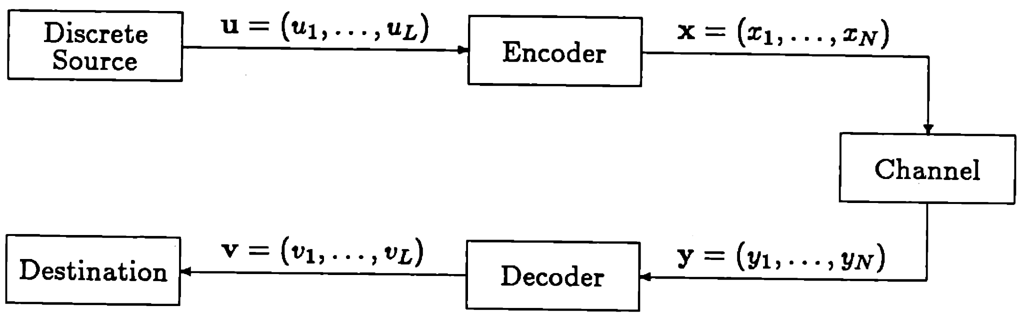

Figure 2.2: A Communication System.

The implication of the model is that the sequences $\mathbf{u}, \mathbf{x}, \mathbf{y}, \mathbf{v}$ form a Markov chain. In this case we have the following theorem:

Theorem 2.1 (Data Processing Theorem) *If the sequences* $\mathbf{u}, \mathbf{x}, \mathbf{y}, \mathbf{v}$ *form a Markov chain, then*

$$I\left(U^L; V^L\right) \leq I\left(X^N; Y^N\right). \tag{2.9}$$

Proof. *Since* $I\left(U^L; Y^N\right) + I\left(X^N; Y^N \mid U^L\right) = I\left(X^N; Y^N\right) + I\left(U^L; Y^N \mid X^N\right)$ *and* $I\left(U^L; Y^N \mid X^N\right) = 0$,

$$I\left(U^L; Y^N\right) \leq I\left(X^N; Y^N\right).$$

Similarly,

$$I\left(U^L; V^L\right) \leq I\left(U^L; Y^N\right),$$

hence completing the proof. □

2.3.2 Fano's Inequality

The following lemma relates conditional uncertainty to the probability of error. The result shows that a lower bound on conditional uncertainty implies a lower bound to the probability of error.

Lemma 2.2 (Fano's Inequality) *Let U, V be a joint ensemble in which the U and V sample spaces contain the same M elements. Let P_e be the probability that the u and v outcomes are different, i.e.,*

$$P_e = \sum_{u \neq v} P(u, v).$$

Then

$$P_e \log(M-1) + \mathcal{H}(P_e) \geq H(U \mid V), \tag{2.10}$$

where $\mathcal{H}(P_e) = -P_e \log P_e - (1 - P_e)\log(1 - P_e)$.

Proof. *Let E be a binary ensemble with $e = 1$ for $u \neq v$, and $e = 0$ for $u = v$. Then*

$$H(UE \mid V) = H(U \mid V) + H(E \mid UV).$$

since E is completely determined by U and V, $H(E \mid UV) = 0$, hence

$$H(U \mid V) = H(UE \mid V) = H(E \mid V) + H(U \mid VE)$$

$$= H(E \mid V) + P_e H(U \mid V, e = 1) + (1 - P_e) H(U \mid V, e = 0).$$

Since $e = 0$ and V completely determine U, the last term above is zero, also $H(E \mid V) \leq H(E) = \mathcal{H}(P_e)$, and $H(U \mid V, e = 1) \leq \log(M - 1)$, hence

$$H(U \mid V) \leq P_e \log(M-1) + \mathcal{H}(M-1).$$

□

The above result can be extended to apply to sequences of arbitrary length as follows.

Theorem 2.3 *Let U^L, V^L be a joint ensemble of length L sequences, $(u_1, \ldots, u_L)$ and $(v_1, \ldots, v_L)$, where each u_ℓ and v_ℓ belongs to the U and V ensembles respectively. Also assume that U and V ensembles consist of the same M elements. Let $\langle P_e \rangle = \frac{1}{L} \sum_{\ell=1}^{L} P_{e,\ell}$ where $P_{e,\ell}$ is the probability that $u_\ell \neq v_\ell$. Then*

$$\langle P_e \rangle \log (M-1) + \mathcal{H} (\langle P_e \rangle) \geq \frac{1}{L} H \left(U^L \mid V^L \right) \tag{2.11}$$

Proof.

$$\begin{aligned} H \left(U^L \mid V^L \right) &= H \left(U_1 \mid V^L \right) + H \left(U_2 \mid U_1 V^L \right) + \cdots \\ &\quad + H \left(U_L \mid V^L U_1 \cdots U_{L-1} \right) \\ &\leq \sum_{\ell=1}^{L} H \left(U_\ell \mid V_\ell \right). \end{aligned}$$

Using Fano's inequality on each term of the summation,

$$H \left(U^L \mid V^L \right) \leq \sum_{\ell=1}^{L} \left[P_{e,\ell} \log (M-1) + \mathcal{H} (P_{e,\ell}) \right]$$

$$\frac{1}{L} H \left(U^L \mid V^L \right) \leq \langle P_e \rangle \log (M-1) + \frac{1}{L} \sum_{\ell=1}^{L} \mathcal{H} (P_{e,\ell}).$$

Since $\mathcal{H}$ is a concave function, the latter term is less than $\mathcal{H} (\langle P_e \rangle)$, thus completing the proof. □

We can now combine the data processing theorem with the above result and obtain

Theorem 2.4 *Let a discrete stationary source with alphabet size M have entropy $H_\infty (U) = \lim_{L \to \infty} H_L (U)$ where $H_L (U) = (1/L) H \left(U^L \right)$. Let a discrete memoryless channel have a capacity C, and let source sequences of length L be transmitted via N channel uses. Then for any L,*

$$\langle P_e \rangle \log (M-1) + \mathcal{H} (\langle P_e \rangle) \geq H_\infty (U) - \frac{N}{L} C \tag{2.12}$$

Proof. *Using Fano's inequality,*

$$\begin{aligned} \langle P_e \rangle \log (M-1) + \mathcal{H} (\langle P_e \rangle) &\geq \frac{1}{L} H \left(U^L \mid V^L \right) = H_L (U) - \frac{1}{L} I \left(U^L ; V^L \right) \\ &\geq H_L (U) - \frac{1}{L} I \left(X^N ; Y^N \right). \end{aligned}$$

For a discrete memoryless channel, $I \left(X^N ; Y^N \right) \leq NC$, and $H_L (U) \geq H_\infty (U)$, hence the proof is complete. □

2.4 Direct Part to the Coding Theorem

In the preceding section we showed that it is not possible to transmit information over a channel at a rate higher than the capacity of the channel. In this section we will show that it is possible to transmit information reliably at rates lower than the capacity. Notice that to show the direct part it is sufficient to establish the existence of a particular transmission scheme, thus we restrict our attention to particular encoding and decoding procedures.

2.4.1 Block Codes and Maximum Likelihood Decoding

An (N, R) *block encoder* is an encoder which provides a sequence of N channel input symbols to each of M possible messages, where $M \geq e^{NR}$. The sequence of N channel input symbols corresponding to message m is called the *codeword* for message m and is denoted by $\mathbf{x}_m = (x_{m1}, \ldots, x_{mN})$. We will refer to the collection of $\{\mathbf{x}_m\}_{m=1}^{M}$ as the *block code.*

The decoding of a block code can be described by a partitioning of the set of channel output sequences of length N, Y^N into M disjoint subsets $\{Y_m\}_{m=1}^{M}$. The decoder decides on message m if the output sequence belongs to Y_m. Thus, the probability of error, given message m is sent, is

$$P_{e,m} = \sum_{\mathbf{y} \in Y_m^c} P_N(\mathbf{y} \mid \mathbf{x}_m), \tag{2.13}$$

with Y_m^c denoting the complement of the set Y_m. The overall probability of error is then given by

$$P_e = \sum_m \Pr[m] P_{e,m} \tag{2.14}$$

where $\Pr[m]$ is the a-priori probability of message m.

Among possible decoding rules, i.e., choosing the sets Y_m, the *maximum likelihood decoding* will be of particular interest in the following development. This decoding rule assigns the output sequence $\mathbf{y}$ to the decoding set Y_m if

$$\forall m' \neq m \qquad P_N(\mathbf{y} \mid \mathbf{x}_m) \geq P_N(\mathbf{y} \mid \mathbf{x}_{m'}) \tag{2.15}$$

with ties broken arbitrarily. Notice that this decoding rule is optimal if the M messages are equally likely. Throughout the rest of this section we will assume that the decoding scheme is maximum likelihood.

2.4.2 The Random Coding Bound

Our goal in this section is to derive bounds on the error probability of block codes, so as to prove the direct part of the coding theorem. The approach we take is to analyze an ensemble of codes, and to find a bound for the average error probability of the ensemble. To define the ensemble we must define a probability assignment for codes. To this end, let $Q_N(\mathbf{x})$ be a probability assignment on the set of channel input sequences of length N. Consider an ensemble of codes for which each codeword is independently selected from the set of channel input sequences according to the distribution Q_N. Then the probability of a particular code, $\mathbf{x}_1, \ldots, \mathbf{x}_M$, is

$$\prod_{m=1}^{M} Q_N(\mathbf{x}_m).$$

Hence if $\mathcal{C}$ is a code and $P_e(\mathcal{C})$ is the average error probability associated with it then we will bound the expectation $\overline{P}_e$ of $P_e(\mathcal{C})$ over the ensemble of codes. Since the expectation is an upper bound to the minimum, this will also be an upper bound to the average error probability of the best code in the ensemble.

To find the upper bound we will make use of the following result.

Lemma 2.5 *Given a collection of events $\{A_m\}_{m=1}^{M}$, for any $0 \le \rho \le 1$,*

$$\Pr\left[\bigcup_m A_m\right] \le \left(\sum_{m=1}^{M} \Pr[A_m]\right)^{\rho}.$$

Proof. *Since we have*

$$\Pr\left[\bigcup_m A_m\right] \le \min\left\{1, \sum_{m=1}^{M} \Pr[A_m]\right\},$$

and for $0 \le \rho \le 1$

$$\min\left\{1, \sum_{m=1}^{M} \Pr[A_m]\right\} \le \left(\sum_{m=1}^{M} \Pr[A_m]\right)^{\rho},$$

the desired result follows. □

For a fixed $\mathbf{x}_m$, and a fixed received sequence $\mathbf{y}$, under maximum likelihood decoding we will make a decoding error, only if there exists some other codeword $\mathbf{x}_{m'}$ with

$$P_N(\mathbf{y} \mid \mathbf{x}_{m'}) \ge P_N(\mathbf{y} \mid \mathbf{x}_m).$$

Defining A_m as the event that the codeword $\mathbf{x}_{m'}$ is as above,

$$\Pr\left[\text{error} \mid m, \mathbf{x}_m, \mathbf{y}\right] \le \Pr\left[\bigcup_{m' \ne m} A_{m'}\right] \le \left(\sum_{m' \ne m} \Pr\left[A_{m'}\right]\right)^{\rho} \qquad 0 \le \rho \le 1.$$

$$\begin{aligned}\Pr\left[A_{m'}\right] &= \sum_{\mathbf{x}_{m'}: P_N(\mathbf{y}|\mathbf{x}_{m'}) \ge P_N(\mathbf{y}|\mathbf{x}_m)} Q_N\left(\mathbf{x}_{m'}\right) \\ &\le \sum_{\mathbf{x}} Q_N(\mathbf{x}) \left(\frac{P_N(\mathbf{y} \mid \mathbf{x})}{P_N(\mathbf{y} \mid \mathbf{x}_m)}\right)^{s} \qquad \text{for any } s > 0,\end{aligned}$$

yielding

$$\Pr\left[\text{error} \mid m, \mathbf{x}_m, \mathbf{y}\right] \le \left[(M-1) \sum_{\mathbf{x}} Q_N(\mathbf{x}) \left(\frac{P_N(\mathbf{y} \mid \mathbf{x})}{P_N(\mathbf{y} \mid \mathbf{x}_m)}\right)^{s}\right]^{\rho}.$$

Since

$$\overline{P}_{e,m} = \sum_{\mathbf{x}_m} \sum_{\mathbf{y}} Q_N(\mathbf{x}_m) P_N(\mathbf{y} \mid \mathbf{x}_m) \Pr\left[\text{error} \mid m, \mathbf{x}_m, \mathbf{y}\right],$$

we have

$$\overline{P}_{e,m} \le (M-1)^{\rho} \sum_{\mathbf{y}} \left[\sum_{\mathbf{x}_m} Q_N(\mathbf{x}_m) P_N(\mathbf{y} \mid \mathbf{x}_m)^{1-s\rho}\right] \left[\sum_{\mathbf{x}} Q_N(\mathbf{x}) P_N(\mathbf{y} \mid \mathbf{x})^{s}\right]^{\rho}.$$

Choosing $s = 1/(1+\rho)$,

$$\overline{P}_{e,m} \le (M-1)^{\rho} \sum_{\mathbf{y}} \left[\sum_{\mathbf{x}} Q_N(\mathbf{x}) P_N(\mathbf{y} \mid \mathbf{x})^{1/(1+\rho)}\right]^{1+\rho}. \tag{2.16}$$

For a discrete memoryless channel, $P_N(\mathbf{y} \mid \mathbf{x}) = \prod_{n=1}^{N} P(y_n \mid x_n)$. Choosing $Q_N(\mathbf{x})$ as $\prod_{n=1}^{N} Q(x_n)$ for a distribution $Q(k), k = 0, \ldots, K-1$ over the channel input alphabet, equation (2.16) simplifies to

$$\overline{P}_{e,m} \le (M-1)^{\rho} \left\{\sum_{j=0}^{J-1} \left[\sum_{k=0}^{K-1} Q(k) P(j \mid k)^{1/(1+\rho)}\right]^{1+\rho}\right\}^{N}.$$

Considering an (N, R) block code, $M - 1 < e^{NR} \le M$, thus,

$$\overline{P}_{e,m} \le \exp\left\{-N\left[E_0(\rho, \mathbf{Q}) - \rho R\right]\right\}, \tag{2.17}$$

where

$$E_0(\rho, \mathbf{Q}) = -\ln \sum_{j=0}^{J-1} \left[\sum_{k=0}^{K-1} Q(k) P(j \mid k)^{1/(1+\rho)}\right]^{1+\rho}. \tag{2.18}$$

Since the above is valid for every m,

$$\overline{P}_e \leq \sum_{m=1}^{M} \Pr[m] \overline{P}_{e,m} \leq \exp\{-N[E_0(\rho, \mathbf{Q}) - \rho R]\}. \qquad (2.19)$$

Also since ρ and $\mathbf{Q}$ are arbitrary, we can maximize $E_0(\rho, \mathbf{Q}) - \rho R$ over all choices of ρ and $\mathbf{Q}$ to obtain the tightest bound on $\overline{P}_e$. Thus we define the *random coding exponent* as

$$E_r(R) = \max_{0 \leq \rho \leq 1} \max_{\mathbf{Q}} [E_0(\rho, \mathbf{Q}) - \rho R]. \qquad (2.20)$$

Hence we have,

Theorem 2.6 (Random Coding Bound) *Given a discrete memoryless channel with transition probabilities $P(j \mid k)$, any positive integer N and a positive number R, consider the ensemble of (N, R) block codes in which each letter of each codeword is independently selected according to the probability assignment $Q(k)$. Then, for each message m, $1 \leq m \leq \lceil e^{NR} \rceil$, and all $\rho, 0 \leq \rho \leq 1$, the ensemble average probability of error using maximum likelihood decoding satisfies*

$$\overline{P}_{e,m} \leq \exp[-NE_r(R)], \qquad (2.21)$$

moreover

$$\overline{P}_e \leq \exp[-NE_r(R)]. \qquad (2.22)$$

Corollary 2.7 *Given any discrete memoryless channel, any N and any R, there exists an (N, R) block code, with*

$$P_{e,m} \leq 4 \exp[-NE_r(R)] \qquad 1 \leq m \leq M = \lceil e^{NR} \rceil. \qquad (2.23)$$

Proof. *Choosing a code with $2M$ codewords for which, for equally likely messages,*

$$P_e = \frac{1}{2M} \sum_{i=1}^{2M} P_{e,m} \leq \exp\left[-NE_r\left(\frac{\ln 2M}{N}\right)\right]$$

and removing M codewords including those for which[1]

$$P_{e,m} \geq 2 \exp\left[-NE_r\left(\frac{\ln 2M}{N}\right)\right],$$

[1]Notice that there cannot be more than M codewords with $P_{e,m} \geq 2P_e$.

we are left with a code such that, for each remaining codeword

$$P_{e,m} < 2\exp\left[-NE_r\left(\frac{\ln 2M}{N}\right)\right].$$

Since $E_r\left(\frac{\ln 2M}{N}\right) \geq -\frac{\ln 2}{N} + E_r\left(\frac{\ln M}{N}\right)$, *we have*

$$P_{e,m} < 4\exp\left[-NE_r\left(\frac{\ln M}{N}\right)\right],$$

completing the proof. □

Noticing that $E_0(0, \mathbf{Q}) = 0$, and making use of

$$\left.\frac{\partial E_0(\rho, \mathbf{Q})}{\partial \rho}\right|_{\rho=0} = \sum_{k=0}^{K-1}\sum_{j=0}^{J-1} Q(k)\,P(j \mid k)\ln\frac{P(j \mid k)}{\sum_i Q(i)\,P(j \mid i)} = I(X;Y),$$

and since all functions involved are continuous, for $R < I(X;Y)$, there exists some $0 < \rho \leq 1$ such that

$$E_0(\rho, \mathbf{Q}) - \rho R > 0.$$

Since $I(X;Y)$ can be set equal to C by an appropriate choice of $\mathbf{Q}$ (see equation 2.8 and notice that the space of $\mathbf{Q}$ is compact), for $R < C$, we have $E_r(R) > 0$. Thus for $R < C$, the probability of error can be made arbitrarily small by choosing N large enough.

One also observes that $E_r(R)$ is the supremum of nonincreasing, linear functions. Then it follows that $E_r(R)$ is a convex, nonincreasing function.

Using more complicated arguments one can also derive lower bounds for the error probability of block codes [SGB 67]. Here we state the following result:

Theorem 2.8 *For any* (N, R) *block code for a discrete memoryless channel,*

$$P_e \geq \exp\{-N[E_{sp}(R - o(N)) + o(N)]\}, \tag{2.24}$$

where

$$E_{sp}(R) = \sup_{\rho>0}\left[\max_{\mathbf{Q}} E_0(\rho, \mathbf{Q}) - \rho R\right], \tag{2.25}$$

and $o(N)$ *denotes quantities that approach zero as* N *gets large.*

The above two theorems show the tightness of the random coding bound. If the maximizing ρ in (2.24) is contained in the interval $[0,1]$, which is the case for rates close to capacity, then the upper and lower bounds to the error probability coincide for large N.

The arguments to obtain the random coding bound can be generalized to channels with constrained inputs [Gal 68, Section 7.3]. In this case we constrain our codewords $\mathbf{x}_m = (x_{m1}, \ldots, x_{mN})$ to satisfy

$$\sum_n f(x_{mn}) \leq N\mathcal{E}$$

for a given function $f : X \to \Re$ defined over the channel input alphabet, and a real number $\mathcal{E}$. In this case we have the following theorem [Gal 68, Theorem 7.3.2].

Theorem 2.9 *Given a discrete memoryless channel with transition probabilities $P(j \mid k)$, a positive number R, and an input constraint $\overline{f(x)} \leq \mathcal{E}$. Let $E < E_r(R)$ where*

$$E_r(R) = \max_{\rho\in[0,1], r\geq 0} \max_{\mathbf{Q}} [E_0(\rho, r, \mathbf{Q}) - \rho R] \tag{2.26}$$

with

$$E_0(\rho, r, \mathbf{Q}) = -\ln \sum_j \left[\sum_k Q(k)\, e^{r[f(k)-\mathcal{E}]} P(j \mid k)^{1/(1+\rho)}\right]^{1+\rho} \tag{2.27}$$

and the maximum is taken over all $\mathbf{Q}$ such that

$$\sum_k f(k)\, Q(k) \leq \mathcal{E}.$$

Then for all sufficiently large N, there exists a block code of length N with $M = \lceil e^{NR} \rceil$ codewords $\mathbf{x}_1, \ldots, \mathbf{x}_M$, each satisfying

$$\sum_n f(x_{mn}) \leq N\mathcal{E},$$

and

$$P_{e,m} \leq \exp(-NE). \tag{2.28}$$

2.5 Multiaccess Channels

Given a collection of finite sets $\{X_j\}_{j=1}^J$, Y, a *multiaccess channel with input alphabets $\{X_j\}_{j=1}^J$ and output alphabet Y* is a set of probability distributions

$$P_N\left(y^N \mid x_1^N, \ldots, x_J^N\right)$$

for each N, $y^N \in Y^N$, and $x_j^N \in X_j^N$. A *memoryless multiaccess channel* has the property that

$$P_N\left(y^N \mid x_1^N, \ldots, x_J^N\right) = \prod_{n=1}^{N} P\left(y_n \mid x_{1n}, \ldots, x_{Jn}\right) \tag{2.29}$$

for some P.

An $(N, R_1, \ldots, R_J)$ *multiaccess block encoder* is a collection of (N, R_j) block encoders, $j = 1, \ldots, J$, the jth encoder providing a mapping from a set of M_j messages to X_j^N, with $M_j \geq e^{NR_j}$. Notice that this definition of encoding does not allow joint encoding of the sources.

The decoder for such a code is a mapping from the set of N-length channel output sequences Y^N to $\{1, \ldots, M_1\} \times \cdots \times \{1, \ldots, M_J\}$.

Given a multiaccess channel we say the rate vector $R_1, \ldots, R_J$ is achievable if for every $\epsilon > 0$, and $\delta > 0$, there exists a $(N, R_1', \ldots, R_J')$ multi-access block code with $R_j' \geq R_j - \delta$, and a corresponding decoder with average probability of error less than ϵ. The set of achievable vectors is called the capacity region. We have the following characterization of the capacity region of a multiaccess channel [Ulr 75, see also [CsK 81]].

Theorem 2.10 (Multiaccess Channel Coding Theorem) *The capacity region of a multiaccess channel is the convex hull of the set of rate vectors satisfying*

$$\forall S \subset \{1, \ldots, J\} \quad 0 \leq \sum_{i \in S} R_i \leq I\left(X(S); Y \mid X(S^c)\right) \tag{2.30}$$

where $X(S)$ denoting $\{X_i\}_{i \in S}$, and the convex hull is taken over all possible product form probability distributions over the set $X_1 \times \cdots \times X_J$.

The above form of the theorem, although precise, is not very insightful at the first look. For conceptual simplicity consider the case $J = 2$. Then 2.30 becomes

$$\begin{aligned} R_1 &\leq I\left(X_1; Y \mid X_2\right), \\ R_2 &\leq I\left(X_2; Y \mid X_1\right), \\ R_1 + R_2 &\leq I\left(X_1 X_2; Y\right). \end{aligned}$$

The last equation is simply a restatement of the single user channel coding theorem, the total information transmitted over the channel is bounded by the mutual information. The first equation bounds the rate of the first user if the decoder has the knowledge of the second user's data, the second equation similarly bounds the rate of the second user if the decoder has the knowledge of the first user's data. A typical sketch of the above constraints is shown in Figure 2.3.

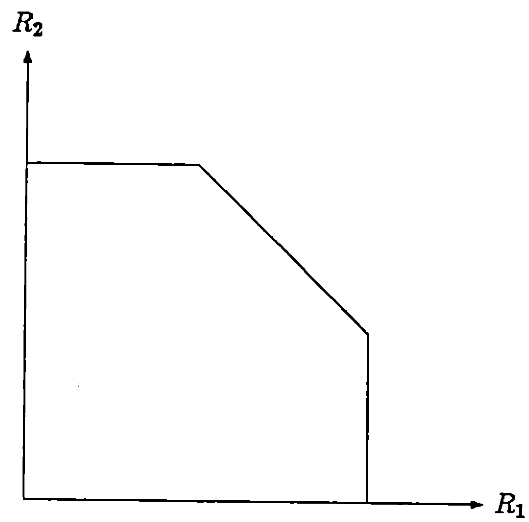

Figure 2.3: Typical sketch of constraints.

The *union* over distributions is analogous to our definition of the capacity as the *maximum* possible mutual information in the single user case. I.e., for each distribution, we can choose codewords randomly as in Section 2.4.2, and satisfy the above constraints; taking the union, we take into account all such coding schemes. Taking the union only over *product form* distributions is to ensure that encoding is done independently, and finally, the convex hull operation corresponds to the time sharing between the coding schemes.

Chapter 3

Point to Point Communication

In this chapter we present results on the single user channel under energy constraints. The channel model we use in this case is a waveform channel and we introduce the concept of *the number of degrees of freedom,* defined as the dimensionality of the space of channel waveforms. As an example, a noiseless channel which has bandwidth W has approximately $2WT$ degrees of freedom,[1] if the duration of channel use is T.

The communication situations we are interested in are such that the number of degrees of freedom offered by the channel is very large. Examples of such channels are the broadband additive Gaussian noise channels, fading dispersive channels, and quantum optical channels. In these cases it may not be possible to utilize all the degrees of freedom effectively because of energy limitations. We describe this situation by imposing a code composition constraint on the codes we use on the channel; this is described in the next section.

3.1 Models and Definitions

Our basic model for communication situations described above is the following [Gal 87]:

- The channel has a binary input alphabet and a continuous output alphabet. The input symbols will be denoted by 0 and 1, and the output space Y, will be the real numbers, $\Re$. The channel is then described by two probability density functions $p_0(y)$ and $p_1(y)$, distribution of the output conditional on the respective input symbol.

[1]The statement immediately follows from the sampling theorem, *a low-pass waveform can be uniquely represented by a sequence of $2WT$ real numbers.*

- The input symbol 0 corresponds to the zero energy input, and the input symbol 1 is assumed to cost an energy $\alpha > 0$ for transmission.

- The channel encoder uses a block code with M codewords of constraint length N. The codewords are denoted by $\mathbf{x}_1, \ldots, \mathbf{x}_M$, and each codeword $\mathbf{x}_i = (x_{i1}, \ldots, x_{iN})$ is constrained to satisfy

$$\frac{1}{N} \sum_j x_{ij} \leq \delta. \tag{3.1}$$

  The memoryless channel assumption implies the independence of the transmissions of symbols in a codeword. Thus the block length N can be interpreted as the number of degrees of freedom of the channel. In this case the user is constrained to use the positive energy symbol only in a δ fraction of the available degrees of freedom.

We define the *rate* of such a code by

$$\widetilde{R} = \frac{\ln M}{\lfloor N\delta \rfloor}. \tag{3.2}$$

Notice that the rate defined as such differs from the traditional rate $\ln M/N$ of Section 2.4 by a factor of δ and measures the amount of information transmitted for each use of symbol 1.

Given a code and a decoder, the probability of error is the average probability of incorrect decoding of a codeword, with the average taken over all codewords. Given a constraint length N and rate $\widetilde{R}$, let $P_e\left(N, \widetilde{R}, \delta\right)$ be the minimum error probability for any block code of constraint length N, rate at least $\widetilde{R}$, and satisfying (3.1). The *reliability function* is then defined by

$$\widetilde{E}\left(\widetilde{R}, \delta\right) = \limsup_{N \to \infty} \frac{-\ln P_e\left(N, \widetilde{R}, \delta\right)}{\lfloor N\delta \rfloor}. \tag{3.3}$$

Similar to $\widetilde{R}$, $\widetilde{E}$ is a measure of the channel reliability with respect to the use of 1's.

Lemma 3.1 *$\widetilde{E}\left(\widetilde{R}, \delta\right)$ is nonincreasing with δ.*

Proof. *Given δ, there exists an increasing sequence $\{N_i\}_{i=1}^{\infty}$ of integers, such that*

$$\lim_{i \to \infty} \frac{-\ln P_e\left(N_i, \widetilde{R}, \delta\right)}{\lfloor N_i\delta \rfloor} = \widetilde{E}\left(\widetilde{R}, \delta\right).$$

Given $\delta' \le \delta$, let $N_i^!$ be such that $\lfloor \delta' N_i^! \rfloor = \lfloor \delta N_i \rfloor$. This can be done since $\delta' < 1$.

Consider the codes with $\left(N_i, \widetilde{R}, \delta\right)$. The number of codewords in each of these codes is at least $e^{\lfloor N_i \delta \rfloor \widetilde{R}}$, and by appending $N_i^! - N_i$ zeros to each codeword, we obtain a $\left(N_i^!, \widetilde{R}, \delta'\right)$ code. The decoder can simply discard the trailing $N_i^! - N_i$ channel outputs, hence $P_e\left(N_i^!, \widetilde{R}, \delta'\right) \le P_e\left(N_i, \widetilde{R}, \delta\right)$. Then we have,

$$\widetilde{E}\left(\widetilde{R}, \delta'\right) \ge \limsup_{i\to\infty} \frac{-\ln P_e\left(N_i^!, \widetilde{R}, \delta'\right)}{\lfloor N_i^! \delta' \rfloor} \ge \limsup_{i\to\infty} \frac{-\ln P_e\left(N_i, \widetilde{R}, \delta\right)}{\lfloor N_i \delta \rfloor} = \widetilde{E}\left(\widetilde{R}, \delta\right),$$

completing the proof. □

Applying the random coding argument as in Section 2.4.2, we find that

$$\widetilde{E}\left(\widetilde{R}, \delta\right) \ge \widetilde{E}_r\left(\widetilde{R}, \delta\right), \tag{3.4}$$

where $\widetilde{E}_r$, the *random coding exponent,* is given by

$$\widetilde{E}_r\left(\widetilde{R}, \delta\right) = \max_{\rho\in[0,1], r\ge 0} \left[\widetilde{E}_0(\rho, r, \delta) - \rho\widetilde{R}\right] \tag{3.5}$$

$$\widetilde{E}_0(\rho, r, \delta) = -\frac{1}{\delta} \ln \int_Y \left[(1-\delta)\, p_0(y)^{1/(1+\rho)} e^{-r\delta} + \delta p_1(y)^{1/(1+\rho)} e^{r(1-\delta)}\right]^{1+\rho} dy. \tag{3.6}$$

The channel capacity $\widetilde{C}$ is the maximum rate for which reliable communication is possible, i.e., $\widetilde{C}(\delta) = \sup\left\{\widetilde{R} : \widetilde{E}\left(\widetilde{R}, \delta\right) > 0\right\}$, and is measured in nats per use of symbol 1. Equivalently,

$$\widetilde{C} = \frac{1}{\delta} \max_{\mathbf{Q}} I(X;Y),$$

where the maximum is taken over the input distributions that satisfy $\mathbf{Q}(1) \le \delta$. To this end, define $\delta_{\max}$ as the unconstrained optimal value of $\mathbf{Q}(1)$. Then the constrained optimal value of $\mathbf{Q}(1)$ will be

$$\mathbf{Q}(1) = \begin{cases} \delta & \text{for } \delta \le \delta_{\max} \\ \delta_{\max} & \text{else.} \end{cases}$$

Since our main interest lies in the case of small δ, we will assume that the first case holds. Then

$$\widetilde{C}(\delta) = \frac{1}{\delta} \left[\begin{array}{c} (1-\delta) \int_Y p_0(y) \ln p_0(y)\, dy + \delta \int_Y p_1(y) \ln p_1(y)\, dy \\ - \int_Y \left[(1-\delta) p_0(y) + \delta p_1(y)\right] \ln\left[(1-\delta) p_0(y) + \delta p_1(y)\right] dy \end{array} \right]. \tag{3.7}$$

We also define another set of parameters which measure the performance with respect to α, the actual energy used per degree of freedom. In this context we treat α as a variable and we assume that the conditional probability density p_1 depends on α. This assumption is equivalent to assuming a nonbinary channel input alphabet, or again equivalently, assuming that we first choose what energy to use on the channel, and then design our codes using only a binary alphabet. In either case, the next section will show that the restriction to binary inputs is essentially optimal for small δ, and resolve the issue. The quantities we define are then

$$\widehat{R} = \frac{\ln M}{\lfloor N\delta \rfloor \alpha}, \tag{3.8}$$

the rate of the code in nats/energy, and

$$\widehat{E}\left(\widehat{R}, \delta, \alpha\right) = \limsup_{N\to\infty} \frac{-\ln P_e\left(N, \widehat{R}, \delta, \alpha\right)}{\lfloor N\delta \rfloor \alpha}, \tag{3.9}$$

the channel reliability on a per unit energy basis. Here $P_e\left(N, \widehat{R}, \delta, \alpha\right)$ denotes the smallest probability of error for block codes with constraint length N, rate $\widehat{R}$ and satisfying (3.1). As a consequence of lemma 3.1, for any fixed α, $\widehat{E}\left(\widehat{R}, \delta, \alpha\right)$ is nonincreasing with δ. Moreover, being the supremum of nonincreasing functions, $\sup_\alpha \widehat{E}\left(\widehat{R}, \delta, \alpha\right)$ is also nonincreasing with δ.

We also define the channel capacity per unit energy, $\widehat{C}$, as the maximum rate for which reliable communication is possible, i.e, $\widehat{C}(\delta, \alpha) = \sup\left\{\widehat{R} : \widehat{E}\left(\widehat{R}, \delta, \alpha\right) > 0\right\}$. Thus we have $\widehat{C}(\delta, \alpha) = \widetilde{C}(\delta)/\alpha$. Again from lemma 3.1 we can conclude that both $\widetilde{C}$ and $\widehat{C}$ are nonincreasing in δ.

3.2 Tightness of Random Coding and Non-binary Set of Inputs

In Section 2.4.2 we showed that the random coding bound is tight for high rates around capacity. Gallager in [Gal 87, Theorems 1 and 2] shows a stronger result for energy limited binary input channels:

Theorem 3.2 *For all* $\widetilde{R} \geq 0$,

$$\widetilde{E}\left(\widetilde{R}, \delta\right) \leq \widetilde{E}_r\left(\widetilde{R}\right) \tag{3.10}$$

and

$$\lim_{\delta\to 0} \widetilde{E}\left(\widetilde{R},\delta\right) = \widetilde{E}_r\left(\widetilde{R}\right) \tag{3.11}$$

where $\widetilde{E}_r\left(\widetilde{R}\right) = \lim_{\delta\to 0} \widetilde{E}_r\left(\widetilde{R},\delta\right) = \max_{0\le\rho\le 1}\left[-(1+\rho)\ln\int_Y p_0(y)^{\rho/(1+\rho)} p_1(y)^{1/(1+\rho)}\,dy - \rho\widetilde{R}\right].$

The result shows that in the limit of small δ the reliability function can be found exactly for all rates.

Now consider channels with a non-binary input alphabet, say with $X = \{0,\ldots,K\}$. The channel is then described by a set of conditional output probability distributions, $\{p_k(y)\}_{k=0}^{K}$, $p_k(y)$ corresponding to the condition that input symbol is k. Assume that with each input k there is a energy $h(k)$ associated with it, and also assume that $h(0) = 0$. The constraint we now have on each codeword is

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

In [Gal 87] Gallager shows that for a broad class of channels, in the limit of $\delta \to 0$, Theorem 3.2 is essentially valid. More precisely, if

$$\max_k \frac{2\mu_{k0}}{h(k)} \ge \max_{\mathbf{q}} \frac{\sum_{i=1}^{K}\sum_{k=1}^{K} q_i q_k \mu_{ik}}{\sum_{i=1}^{K} q_i h(i)}, \tag{3.12}$$

where $\mu_{ik} = -\ln\int_Y \sqrt{p_i(y)\,p_k(y)}\,dy$ and the second maximum is taken over all probability vectors $\mathbf{q}$, then

$$\widetilde{E}\left(\widetilde{R},\delta\right) \le \widetilde{E}_r\left(\widetilde{R}\right),$$

$$\lim_{\delta\to 0}\widetilde{E}\left(\widetilde{R},\delta\right) = \widetilde{E}_r\left(\widetilde{R}\right),$$

$$\widetilde{E}_r\left(\widetilde{R}\right) = \max_{\rho\in[0,1]}\left[\widetilde{E}_0(\rho) - \rho\widetilde{R}\right],$$

$$\widetilde{E}_0(\rho) = \max_{k>0} -\frac{(1+\rho)}{h(k)}\ln\int_Y p_0(y)^{\rho/(1+\rho)} p_k(y)^{1/(1+\rho)}\,dy.$$

That is, in the limit of $\delta \to 0$, for any given rate $\widetilde{R}$, there is a nonzero symbol k, to be used together with 0, to give an optimum performance over the channel. Thus the restriction to binary inputs is essentially optimal.

3.3 Rayleigh Fading Channel

For each degree of freedom, the Rayleigh fading channel can be described by the conditional probability assignment [Ken 69]

$$p(v \mid u) = \frac{2v}{1+u^2} e^{-v^2/(1+u^2)},$$

where u is the input signal, v is the output signal and under the assumption of a unit amount of noise power. Here we should note the subtlety of assuming that each degree of freedom is independent. This assumption is approximately valid if we do not partition the time and frequency "too finely." That is to say, the the number of degrees of freedom the channel can offer is limited by the *time dispersion* and the *frequency dispersion* of the channel [Ken 69]. If the partitions are such that the time slots are wider than the time dispersion, and frequency slots are wider than the frequency dispersion, then each of the degrees of freedom can be treated as independent.

For ease of analytical manipulations, we model the Rayleigh fading channel in terms of power. In this case, the distribution of the output (power), y, conditional on the input (power), x, is given by

$$p(y \mid x) = \frac{1}{1+x} e^{-y/(1+x)}.$$

Assuming we use power α for symbol 1 and no power for symbol zero, we obtain

$$p_0(y) = \begin{cases} e^{-y} & y \geq 0, \\ 0 & \text{else}, \end{cases} \qquad p_1(y) = \begin{cases} \frac{1}{1+\alpha} e^{-y/(1+\alpha)} & y \geq 0, \\ 0 & \text{else}. \end{cases} \tag{3.13}$$

Notice that each different value of α provides us with an additional input symbol. Thus, (3.12) should be checked:

$$\mu(\alpha, \beta) = -\ln \int_0^\infty \sqrt{\frac{1}{(1+\alpha)(1+\beta)} e^{-y[1/(1+\alpha)+1/(1+\beta)]}}\, dy = \ln \frac{2+\alpha+\beta}{2\sqrt{(1+\alpha)(1+\beta)}},$$

hence $\mu(\alpha, \beta)$ is nonincreasing with α for $\alpha \leq \beta$. Noticing $\mu(\alpha, \beta) = \mu(\beta, \alpha)$, we have

$$\mu(\alpha, \beta) \leq \max\{\mu(0, \alpha), \mu(0, \beta)\} \leq \mu(0, \alpha) + \mu(0, \beta),$$

the last inequality following from the nonnegativity of μ. Also taking into account $h(\alpha) = \alpha$,

$$\frac{\int\int q(\alpha)\, q(\beta)\, \mu(\alpha, \beta)\; d\alpha d\beta}{\int q(\alpha)\, \alpha\, d\alpha} \leq 2 \frac{\int q(\alpha)\, \mu(0, \alpha)\; d\alpha}{\int q(\alpha)\, \alpha\, d\alpha} \leq 2 \sup_\alpha \frac{\mu(0, \alpha)}{\alpha},$$

verifying (3.12).

The next few sections will present the results on the capacity and error exponents for the Rayleigh fading channel, assuming the above distributions.

3.3.1 Capacity Calculations

Making use of (3.7) the capacity $\tilde{C}$ can be calculated as[2]

$$\begin{aligned}
\tilde{C} &= \frac{1}{\delta}\left[\begin{array}{l} (1-\delta)\int_0^\infty e^{-y}\ln e^{-y}\,dy + \delta\int_0^\infty \frac{1}{1+\alpha}e^{-y/(1+\alpha)}\ln\left(\frac{1}{1+\alpha}e^{-y/(1+\alpha)}\right)dy \\ -\int_0^\infty \left[(1-\delta)\,e^{-y}+\delta\frac{1}{1+\alpha}e^{-y/(1+\alpha)}\right]\ln\left[(1-\delta)\,e^{-y}+\delta\frac{1}{1+\alpha}e^{-y/(1+\alpha)}\right]dy \end{array}\right] \\
&= -\ln(1+\alpha)-\frac{1}{\delta}\ln\left(1-\frac{\delta\alpha}{1+\alpha}\right)+\frac{\alpha}{1+\alpha}\int_0^1 \frac{t^{1/\alpha}}{t+\frac{\delta}{(1+\alpha)(1-\delta)}}\,dt \\
&= -\ln(1+\alpha)-\frac{1}{\delta}\ln\left(1-\frac{\delta\alpha}{1+\alpha}\right)+\frac{\alpha}{1+\alpha}T(\alpha,\delta), \qquad (3.14)
\end{aligned}$$

with $T(\alpha,\delta) = \int_0^1 \frac{t^{1/\alpha}}{t+\frac{\delta}{(1+\alpha)(1-\delta)}}\,dt$. Notice that $\frac{\alpha}{1+\alpha}T(\alpha,\delta)$ is Gauss' hypergeometric function [GrR 80],

$$\frac{\alpha}{1+\alpha}T(\alpha,\delta) = {}_2F_1\left(1,1+\frac{1}{\alpha};2+\frac{1}{\alpha};-\frac{\delta}{(1+\alpha)(1-\delta)}\right).$$

An interesting fact about $T(\alpha,\delta)$ is that it is not uniformly continuous in δ at $\delta=0$. For this reason we have $\lim_{\alpha\to\infty}\tilde{C}(\delta=0,\alpha) \neq \lim_{\delta\to 0}\lim_{\alpha\to\infty}\tilde{C}(\delta)$.

For $\delta=0$, $\tilde{C}(\delta=0) = \alpha-\ln(1+\alpha)$. Thus $\hat{C}(\delta=0,\alpha) = 1-\frac{\ln(1+\alpha)}{\alpha}$, and hence $\lim_{\alpha\to\infty}\hat{C}(\delta=0,\alpha)=1$.

For $\delta>0$, we will show that $\lim_{\alpha\to\infty}\hat{C}(\delta,\alpha)=0$. To see this let $\epsilon\in(0,1)$ be arbitrary, and consider

$$\epsilon^{1/\alpha}\int_\epsilon^1 \frac{1}{\left(t+\frac{\delta}{(1+\alpha)(1-\delta)}\right)}\,dt < \int_\epsilon^1 \frac{t^{1/\alpha}}{t+\frac{\delta}{(1+\alpha)(1-\delta)}}\,dt < T(\alpha,\delta) < \int_0^1 \frac{1}{\left(t+\frac{\delta}{(1+\alpha)(1-\delta)}\right)}\,dt$$

$$\epsilon^{1/\alpha}\ln\left(\frac{1+\frac{\delta}{(1+\alpha)(1-\delta)}}{\epsilon+\frac{\delta}{(1+\alpha)(1-\delta)}}\right) < T(\alpha,\delta) < \ln\left(1+\frac{(1+\alpha)(1-\delta)}{\delta}\right).$$

[2]For the intermediate steps see the appendix at the end of this chapter

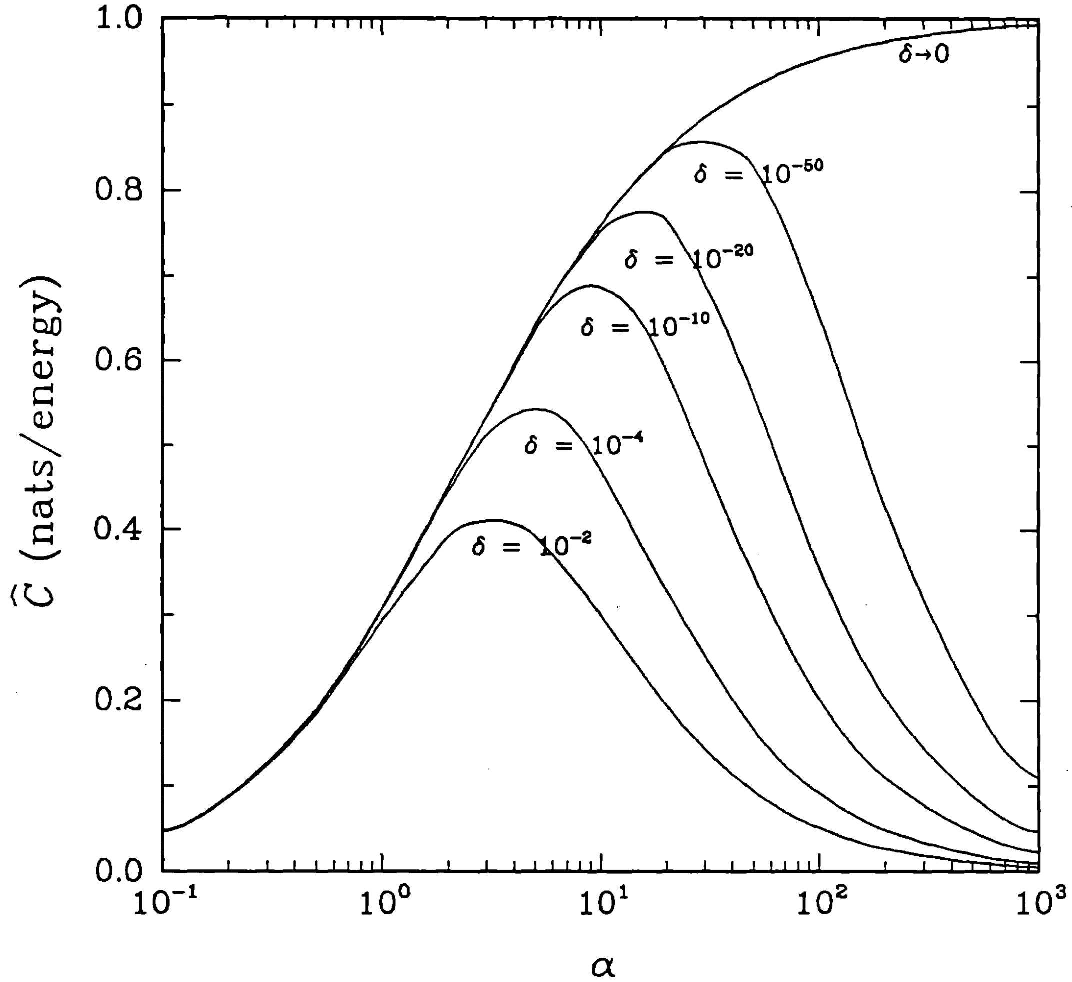

Figure 3.1: $\widehat{C}$ vs. α for the Rayleigh fading channel for different values of δ.

Chosing[3] $\epsilon = \frac{\delta}{\alpha^2} \ln \frac{\alpha}{\delta}$, we have

$$\left(\frac{\delta}{\alpha^2} \ln \frac{\alpha}{\delta}\right)^{1/\alpha} \ln \left(\frac{1 + \frac{\delta}{(1+\alpha)(1-\delta)}}{\frac{\delta}{\alpha^2} \ln \frac{\alpha}{\delta} + \frac{\delta}{(1+\alpha)(1-\delta)}} \right) \; < \; T(\alpha, \delta) \; < \; \ln \left(1 + \frac{(1+\alpha)(1-\delta)}{\delta}\right).$$

Thus we see that $\lim_{\alpha \to \infty} \frac{1}{1+\alpha} T(\alpha, \delta) = 0$, and hence,

$$\lim_{\alpha \to \infty} \widehat{C}(\delta, \alpha) = \lim_{\alpha \to \infty} \left[\frac{\ln(1+\alpha)}{\alpha} - \frac{1}{\alpha\delta} \ln\left(1 - \frac{\alpha\delta}{1+\alpha}\right) + \frac{1}{1+\alpha} T(\alpha, \delta) \right] = 0.$$

[3]This choice of ϵ approximates the value of ϵ that gives the tightest bound.

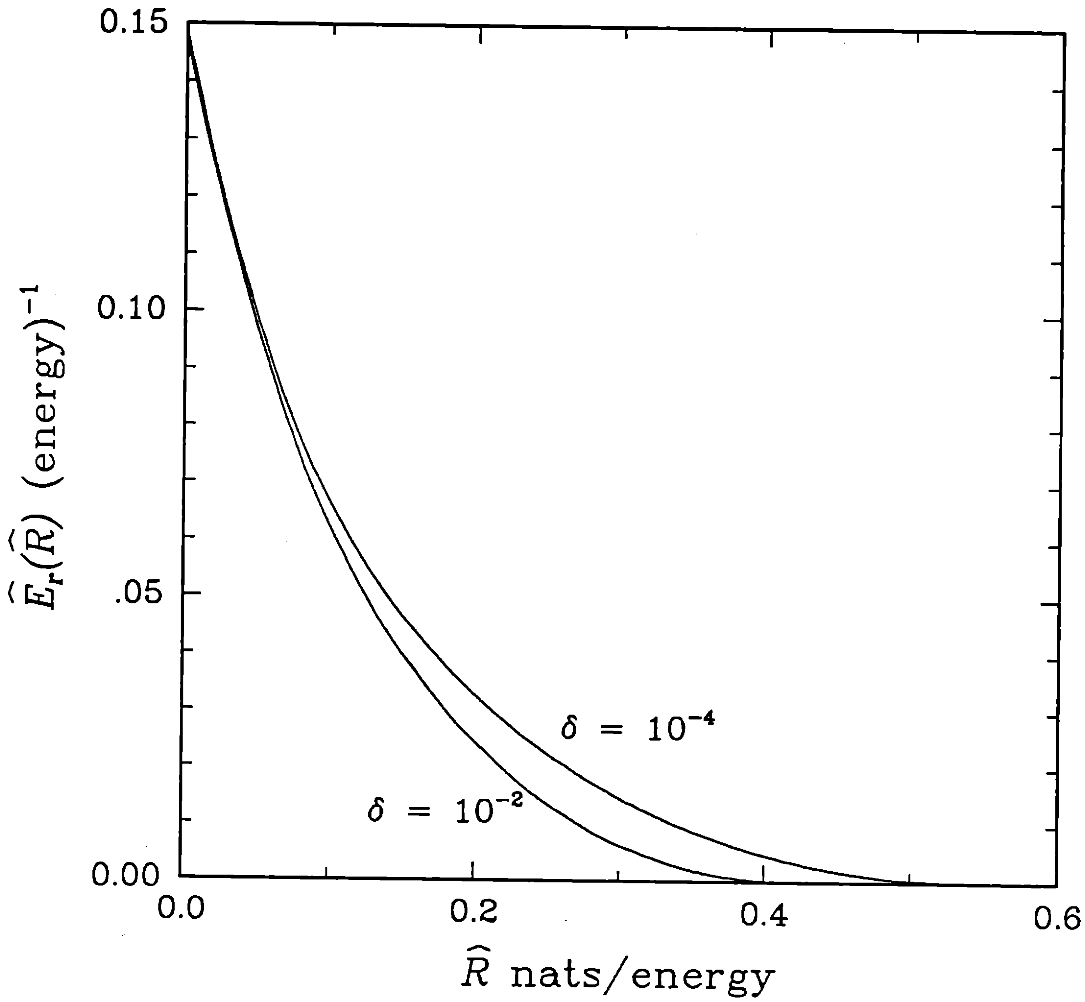

Figure 3.2: Optimized Error Exponent, $\sup_{\alpha} \widehat{E}\left(\widehat{R}, \delta, \alpha\right)$ vs. $\widehat{R}$.

Figure 3.1 displays the capacity of the Rayleigh fading channel as a function of α, for various values of δ. The figure illustrates the point of above discussion, as the peak of the curves shift towards infinity as δ gets smaller. Another conclusion from the figure is that, to get $\widehat{C}$ close to unity, one must choose very small values of δ which is very nonpractical. As an example, the last curve in the figure corresponds to $\delta = 10^{-50}$, and to design such codes, one needs block lengths of at least 10^{50}. Even if one is able to design such codes, the amount of time required to transmit such codes is extremely large, even under most optimistic assumptions about available bandwidth.

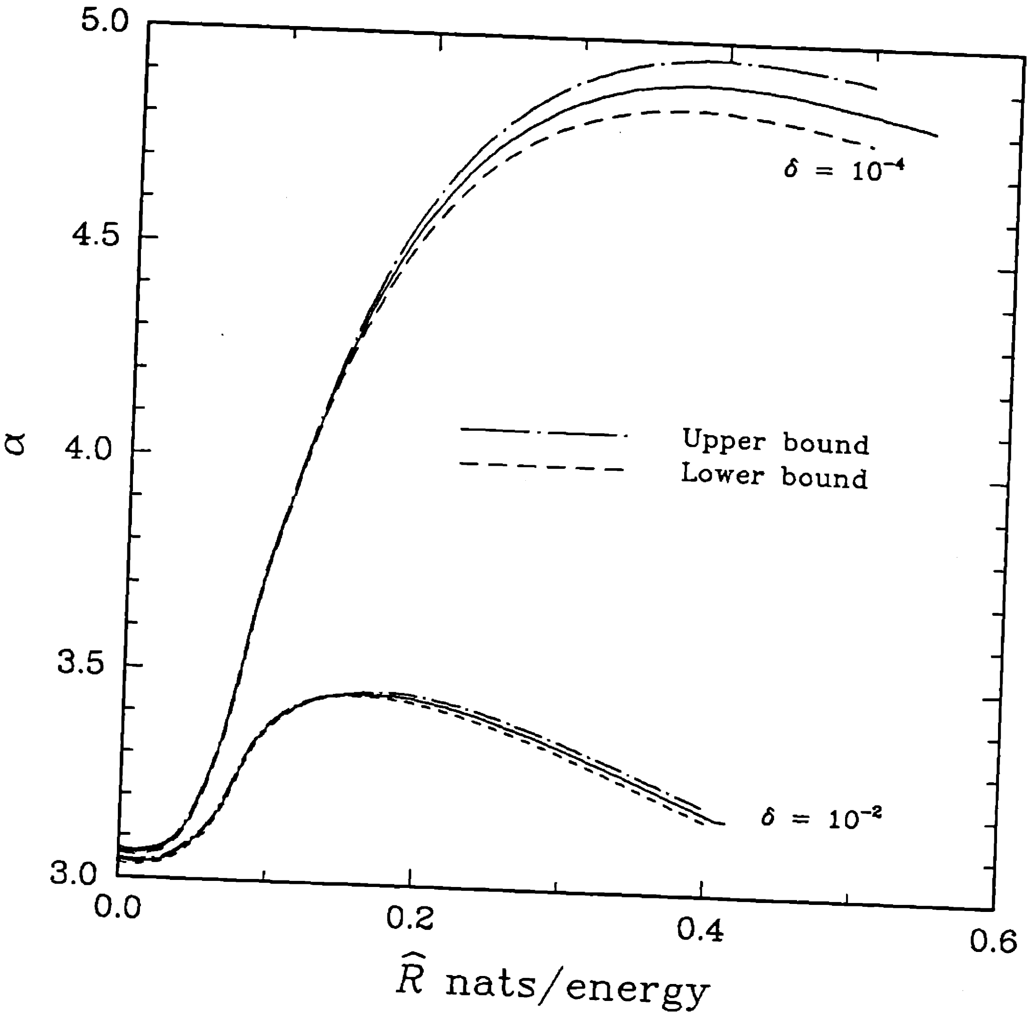

Figure 3.3: Optimizing α, $\arg\sup_{\alpha} \widehat{E}\left(\widehat{R}, \delta, \alpha\right)$ vs. $\widehat{R}$.

In the limit of infinite bandwidth, i.e., $\delta \to 0$, and for large SNR, i.e., α, we see that the Rayleigh fading channel has a capacity $\widehat{C} = 1$. For an additive Gaussian noise channel in the limit of infinite bandwidth, $\widehat{C} = 1$ regardless of the value of SNR. Thus the above figure also indicates that, although the additive Gaussian noise channel and the Rayleigh fading channel have the same capacity in the limiting case of infinite bandwidth and large signal to noise ratio, the Rayleigh fading channel approaches this limit much slower than the Gaussian channel.

3.3.2 Exponent Calculations and Optimum SNR

Using (3.5), we can calculate the random coding exponent $\widetilde{E}_r\left(\widetilde{R},\delta\right)$, for any $\widetilde{R}$, providing a lower bound to the actual error exponent $\widetilde{E}\left(\widetilde{R},\delta\right)$. From the results in Section 2.4.2 this lower bound will be tight for rates close to capacity. Also from Section 3.2 we know that $\lim_{\delta\to 0}\widetilde{E}_r\left(\widetilde{R},\delta\right)$ provides an upper bound to the actual error exponent.

The results of Section 3.2 can be carried over to $\widehat{E}$ and $\widehat{E}_r$. Thus we have

$$\widehat{E}\left(\widehat{R},\delta,\alpha\right)\geq\widehat{E}_r\left(\widehat{R},\delta,\alpha\right), \tag{3.15}$$

where

$$\widehat{E}_r\left(\widehat{R},\delta,\alpha\right)=\max_{\rho\in[0,1],r\geq 0}\left[\widehat{E}_0\left(\rho,r,\delta,\alpha\right)-\rho\widehat{R}\right] \tag{3.16}$$

$$\widehat{E}_0\left(\rho,r,\delta,\alpha\right)=-\frac{1}{\delta\alpha}\ln\int_Y\left[\left(1-\delta\right)p_0\left(y\right)^{1/(1+\rho)}e^{-r\delta}+\delta p_1\left(y\right)^{1/(1+\rho)}e^{r(1-\delta)}\right]^{1+\rho}dy. \tag{3.17}$$

The resulting error exponent can then be optimized over α to find the optimal SNR and the error exponent. Figure 3.2 shows the dependence of the optimized error exponent to the rate for different values of δ. The companion Figure 3.3 shows the optimizing value of α, (SNR), vs. rate, $\widehat{R}$. The maximization over α is a very costly process due to the difficulty of the computation of the objective function. Another complication is that, the maximum occurs in a broad range of α, hence it becomes even more difficult to find the maximizing α. For this reason in Figure 3.3 we provide upper and lower bounds for the optimal SNR.

3.A Appendix: Derivation of $\widetilde{C}$ for the Rayleigh Fading Channel

In this section we show the steps in obtaining (3.14) from (3.7). With

$$p_0(y) = \begin{cases} e^{-y} & y \geq 0, \\ 0 & \text{else,} \end{cases} \quad \text{and} \quad p_1(y) = \begin{cases} \frac{1}{\beta} e^{-y/\beta} & y \geq 0, \\ 0 & \text{else,} \end{cases}$$

where $\beta \stackrel{\text{def}}{=} 1 + \alpha$, equation (3.7) becomes

$$\widetilde{C} = \frac{1}{\delta} \left[\begin{array}{c} (1-\delta) \int_0^\infty e^{-y} \ln e^{-y} \, dy + \delta \int_0^\infty \frac{1}{\beta} e^{-y/\beta} \ln \left(\frac{1}{\beta} e^{-y/\beta} \right) dy \\ - \int_0^\infty \left[(1-\delta) e^{-y} + \delta \frac{1}{\beta} e^{-y/\beta} \right] \ln \left[(1-\delta) e^{-y} + \delta \frac{1}{\beta} e^{-y/\beta} \right] dy \end{array} \right].$$

Noticing that for $a > 0$, $\int_0^\infty \frac{1}{a} e^{-u/a} \ln \left(\frac{1}{a} e^{-u/a} \right) du = -1 - \ln a$, the above can be written as

$$\widetilde{C} = \frac{1}{\delta} \left[-1 - \delta \ln \beta - \int_0^\infty \left[(1-\delta) e^{-y} + \delta \frac{1}{\beta} e^{-y/\beta} \right] \ln \left[(1-\delta) e^{-y} + \delta \frac{1}{\beta} e^{-y/\beta} \right] dy \right].$$

Defining $x = e^{-y/\beta}$, $dy = \beta dx/x$, the above integral becomes

$$\int_0^1 \left[(1-\delta) \beta x^{\beta-1} + \delta \right] \ln \left[(1-\delta) x^\beta + \frac{\delta}{\beta} x \right] dx.$$

The new integral can be decomposed by writing δ as $\frac{\delta}{\beta} + \delta \left(1 - \frac{1}{\beta} \right)$:

$$\int_0^1 \left[(1-\delta) \beta x^{\beta-1} + \frac{\delta}{\beta} \right] \ln \left[(1-\delta) x^\beta + \frac{\delta}{\beta} x \right] dx + \delta \left(1 - \frac{1}{\beta} \right) \int_0^1 \ln \left[(1-\delta) x^\beta + \frac{\delta}{\beta} x \right] dx.$$

The first integral can be evaluated by defining $u = \left[(1-\delta) x^\beta + \frac{\delta}{\beta} x \right]$:

$$\int_0^{1-\delta\left(1-\frac{1}{\beta}\right)} \ln u \, du = \left[1 - \delta \left(1 - \frac{1}{\beta} \right) \right] \ln \left[1 - \delta \left(1 - \frac{1}{\beta} \right) \right] - \left[1 - \delta \left(1 - \frac{1}{\beta} \right) \right].$$

The integrand of the second integral can be written as $\ln x + \ln \left[(1-\delta) x^{\beta-1} + \frac{\delta}{\beta} \right]$, and noticing that $\int_0^1 \ln x \, dx = -1$, we can write

$$\widetilde{C} = \frac{1}{\delta} \left[\begin{array}{c} -\delta \ln \beta - \left[1 - \delta \left(1 - \frac{1}{\beta} \right) \right] \ln \left[1 - \delta \left(1 - \frac{1}{\beta} \right) \right] \\ -\delta \left(1 - \frac{1}{\beta} \right) \int_0^1 \ln \left[(1-\delta) x^{\beta-1} + \frac{\delta}{\beta} \right] dx \end{array} \right].$$

Defining $t = x^{\beta-1}$, the last term in the above expression becomes

$$\frac{\delta}{\beta}\int_0^1 \ln\left[\frac{\delta}{\beta} + (1-\delta)\,t\right] t^{-1+1/(\beta-1)}\,dt.$$

Integrating by parts, we obtain:

$$\delta\left(1-\frac{1}{\beta}\right)\left[\ln\left[1-\delta\left(1-\frac{1}{\beta}\right)\right] - (1-\delta)\int_0^1 \frac{t^{1/(\beta-1)}}{\frac{\delta}{\beta}+(1-\delta)\,t}\,dt\right],$$

hence

$$\tilde{C} = -\ln\beta - \frac{1}{\delta}\ln\left[1-\delta\left(1-\frac{1}{\beta}\right)\right] + \left(1-\frac{1}{\beta}\right)\int_0^1 \frac{t^{1/(\beta-1)}}{t+\frac{\delta}{\beta(1-\delta)}}\,dt,$$

which is equal to (3.14) with $\beta = \alpha + 1$.

Chapter 4

Multiaccess Communication

In this chapter, we consider the use of a channel with a very large number of degrees of freedom by more than one user, as opposed to the single user of the previous chapter. Again we will assume that each user has a limited amount of energy, and will find the maximum possible user rate in the special case of equal rate, equal energy users.

4.1 Models and Definitions

The basic model we choose for multiaccess use is a cascade of a combiner and a point to point channel as shown in Fig. 4.1.

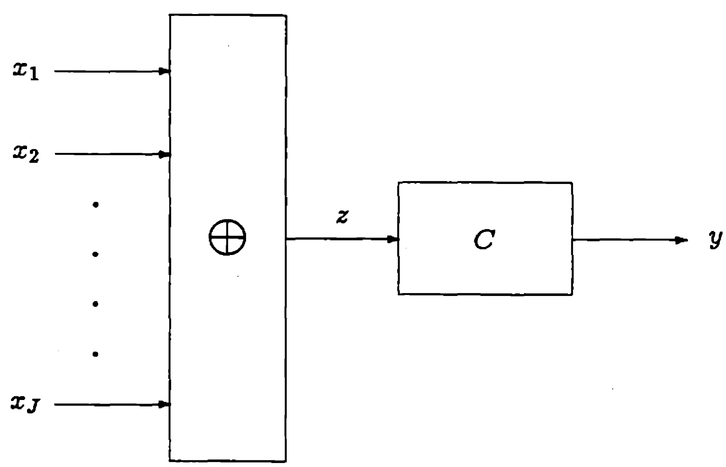

Figure 4.1: Multiaccess Communication Model

In this case a set of J users access a point to point channel C via a combiner $\bigoplus$: $X_1 \times \cdots \times X_J \rightarrow Z$. Usual choices for the combiner are the *OR*-channel and the *ADD*-channel, which at their outputs respectively do not and do provide information about the number of users utilizing a particular degree of freedom. That is, for the *OR*-channel, the output of the combiner is the logical OR of the set of inputs, and for the *ADD*-channel, the output of the combiner is the arithmetic sum of the set of inputs. The above model with a noiseless channel has been studied by many researchers, e.g. [ChW 81]. It is also important to realize that the above model is the channel model per degree of freedom. Also notice that the *ADD*-channel is the most general model for a symmetrical set of users.

We will further assume that each of the input alphabets X_i, $i = 1, \ldots, J$ is binary, and similar to Section 3.1, we constrain the codewords of user j to contain no more than a δ_j fraction of 1's. Thus associated with a multiple access block code, we have a rate vector $\boldsymbol{R}$ and a constraint vector $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_J)$.

The multiaccess coding theorem of Section 2.5 can be extended [Gal 87] to include the energy limitations. The main difference between the unconstrained multiaccess theorem and the constrained multiaccess theorem lies in the timesharing (convex hull) operation. The twist in the constrained case is that we may obtain a code which meets the constraint by timesharing between two codes, both of which violate the constraint individually.

As an example of this consider a set of non-negative numbers $\xi_1, \ldots, \xi_J$, that sum to unity, and consider the time sharing of J coding schemes, the jth scheme used ξ_j fraction of the time. Suppose the jth scheme has a rate $\boldsymbol{R}_j = R_j \boldsymbol{e}_j / \xi_j$ and meets the constraint $\boldsymbol{\delta}_j = \delta_j \boldsymbol{e}_j / \xi_j$, where $\boldsymbol{e}_j$ denotes the jth unit vector. Then the overall code has a rate $\boldsymbol{R} = (R_1, \ldots, R_J)$ and meets the constraint $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_J)$. This particular coding scheme will be referred to as the *FDM scheme*.

To take into account the above fact, we define the following:

- The pair $(\boldsymbol{R}, \boldsymbol{\delta})$ is *directly achievable* if, for some joint product input distribution $\mathbf{Q}$ satisfying $\forall j \; Q_j(1) \leq \delta_j$,

$$\forall S \subset \{1, \ldots, J\} \quad 0 \leq \sum_{i \in S} R_i \leq I\left(X(S); Y \mid X(S^c)\right). \tag{4.1}$$

The above is just a generalization of Theorem 2.10.

- The pair $(\boldsymbol{R}, \delta)$ is ***achievable*** if, it is in the convex hull of the directly achievable region. Notice that the convex hull is taken over both the elements of the pair, not just $\boldsymbol{R}$.

With the above definitions we have the following Theorem [Gal 87].

Theorem 4.1 *The capacity region of a multiaccess channel with energy constraint δ is the set of rates $\boldsymbol{R}$, for which $(\boldsymbol{R}, \delta)$ is achievable.*

4.2 Rayleigh Fading Channel

Consider a set of J users accessing a Rayleigh Fading channel. Suppose the transmitter of each user has the same power α (per degree of freedom), and suppose the noise power (per degree of freedom) in the channel is unity. If the transmitters of the users are non-coherent, i.e, they have random phase with respect to each other, then we conclude that the power in a particular degree of freedom is the sum of the powers of individual users utilizing that particular degree of freedom. Thus the appropriate combiner model is the *ADD*-combiner. Then the distribution of the output of the Rayleigh fading channel conditional on k users using the channel is given by

$$p_k(y) = \begin{cases} \frac{1}{\alpha_k} e^{-y/\alpha_k} & \text{for } y \geq 0, \\ 0 & \text{else,} \end{cases} \tag{4.2}$$

where $\alpha_k = 1 + k\alpha$.

Consider a case in which the constraint for each user is the same, i.e., $\delta_j = \delta$. If δ is small enough, the directly achievable region will be obtained for an input distribution such that the users are independent and each user has a probability δ of using a 1. Then the unconditional output distribution becomes

$$p_Y(y) = \sum_{j=0}^{J} \binom{J}{j} \delta^j (1-\delta)^{J-j} p_j(y), \tag{4.3}$$

and the directly achievable region is characterized by

$$\forall S \subset \{1, \dots, J\} \quad 0 \leq \sum_{i \in S} R_i \leq$$

$$\sum_{i=0}^{s} \sum_{j=0}^{J-s} \binom{s}{i} \binom{J-s}{j} \delta^{i+j} (1-\delta)^{J-(i+j)} \int_0^\infty p_{i+j}(y) \ln \frac{p_{i+j}(y)}{\sum_{k=1}^{s} \binom{s}{k} \delta^k (1-\delta)^{s-k} p_{j+k}(y)} dy,$$

where s is the cardinality of S.

Due to the computational intractability of the above set of equations, we turn our attention to a simpler case of a large number of users with equal rates.

4.2.1 Infinite User Case

Suppose now that J is very large and $\delta_j = \lambda/J$ for each user j. With equal rate users $R_j = R$, $j = 1, \ldots, J$, Section 4.3, will show that the FDM region will not be larger than the directly achievable region. Although in general the capacity region may be larger than both the FDM and directly achievable regions, usually the FDM and directly achievable regions are of more interest. We also note that the directly achievable region is also referred as the random access region. Thus, the maximum possible rate (for random access) will be derived from 4.1. However, the binding constraint in 4.1 is the constraint corresponding to $S = \{1, \ldots, J\}$,

$$JR \leq I(X_1 \cdots X_J; Y),$$
$$\tilde{R} \leq \frac{1}{\lambda} I(X_1 \cdots X_J; Y). \tag{4.4}$$

By symmetry, the value of the bound is maximized by a distribution over the users for which each user uses input 1 with probability λ/J. The output distribution then becomes

$$p_Y(y) = \sum_{j=0}^{J} \binom{J}{j} \left(\frac{\lambda}{J}\right)^j \left(1 - \frac{\lambda}{J}\right)^{J-j} p_j(y).$$

Noticing that for large J,

$$\binom{J}{j} \left(\frac{\lambda}{J}\right)^j \left(1 - \frac{\lambda}{J}\right)^{J-j} \rightarrow e^{-\lambda} \frac{\lambda^j}{j!},$$

we can write

$$p_Y(y) = \sum_{j=0}^{\infty} \mu_j p_j(y). \tag{4.5}$$

where $\mu_j = e^{-\lambda} \dfrac{\lambda^j}{j!}$.

Equation 4.4 then becomes

$$\tilde{R} \leq \frac{1}{\lambda} \left[\sum_{k=0}^{\infty} \mu_k \int_0^{\infty} p_k(y) \ln p_k(y)\, dy - \int_0^{\infty} p_Y(y) \ln p_Y(y)\, dy \right].$$

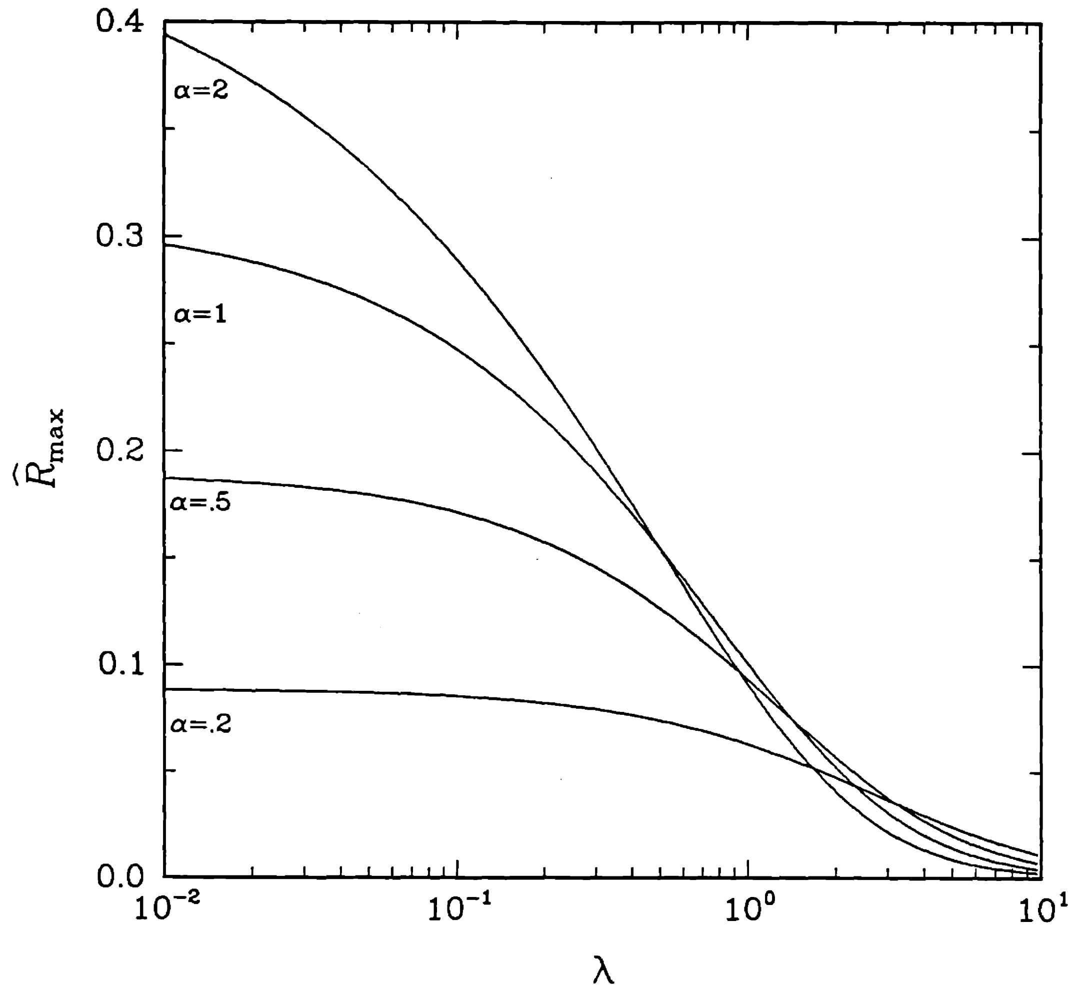

Figure 4.2: $\widehat{R}_{\max}$ vs. λ for different values of α.

The terms appearing in the first summation easily evaluate to

$$\int_0^\infty p_k(y) \ln p_k(y)\, dy = -\left(1 + \ln \alpha_k\right),$$

but the second integral does not lend itself to any simplifications. See the appendix at the end of this chapter for a result on the behavior of $p_Y(y)$. As in the previous chapter, we define the rate of a user with respect to energy as $\widehat{R} = \widetilde{R}/\alpha$. Figure 4.2 shows the value of the rate, $\widehat{R}$, achievable by random access as a function of λ, the expected number of users utilizing one degree of freedom. The figure clearly shows that throughput (in terms of nats/energy) decreases as λ increases, this is due to the increased interference of the users.

On the other hand we see that an increase in α does not necessarily increase the throughput (again in terms of nats/energy). Figure 4.3 on page 39 illustrates this more clearly. Thus for any given λ, there is an optimum α for which throughput is maximized.

4.3 Comparison of FDM and Directly Achievable Region for the Binary Input Adder Channel

With FDM, the highest possible rate of each user is given by

$$\tilde{C}_{\mathrm{FDM}}(\lambda) = \frac{1}{\lambda}\left[\begin{array}{c}(1-\lambda)\int_Y p_0(y)\ln p_0(y)\,dy + \lambda\int_Y p_1(y)\ln p_1(y)\,dy \\ -\int_Y \left[(1-\lambda)p_0(y)+\lambda p_1(y)\right]\ln\left[(1-\lambda)p_0(y)+\lambda p_1(y)\right]dy\end{array}\right]. \quad (4.6)$$

Notice that the above is the same as 3.7 with λ replacing δ.

With random access, however, the rate of each user is given by

$$\tilde{C}_{\mathrm{RA}}(\lambda) = \frac{1}{\lambda}\left[\sum_k \mu_k \int_Y p_k(y)\ln p_k(y)\,dy - \int_Y\left[\sum_k \mu_k p_k(y)\right]\ln\left[\sum_k \mu_k p_k(y)\right]dy\right]. \quad (4.7)$$

We notice that for all well behaved functions $p_k(y)$,

$$\lim_{\lambda\to 0}\tilde{C}_{\mathrm{FDM}}(\lambda) = \lim_{\lambda\to 0}\tilde{C}_{\mathrm{RA}}(\lambda).$$

On the other hand,

$$\frac{\partial}{\partial\lambda}\left[\tilde{C}_{\mathrm{FDM}}(\lambda) - \tilde{C}_{\mathrm{RA}}(\lambda)\right]\Big|_{\lambda=0} = \int_Y p_1(y)\ln\frac{p_1(y)}{p_0(y)}\,dy - \frac{1}{2}\int_Y p_2(y)\ln\frac{p_2(y)}{p_0(y)}\,dy.$$

Hence the RA scheme will outperform the FDM scheme for small λ, if,

$$\int_Y p_2(y)\ln\frac{p_2(y)}{p_0(y)}\,dy - 2\int_Y p_1(y)\ln\frac{p_1(y)}{p_0(y)}\,dy > 0. \quad (4.8)$$

Conversely if

$$\int_Y p_2(y)\ln\frac{p_2(y)}{p_0(y)}\,dy - 2\int_Y p_1(y)\ln\frac{p_1(y)}{p_0(y)}\,dy < 0,$$

then FDM performs better than random access. In case of equality, we need to look at higher order derivatives to reach a conclusion.

Here we should note that the result is valid only if the users are restricted to binary inputs. If the users were allowed to use a continuous alphabet, then by FDM each user can

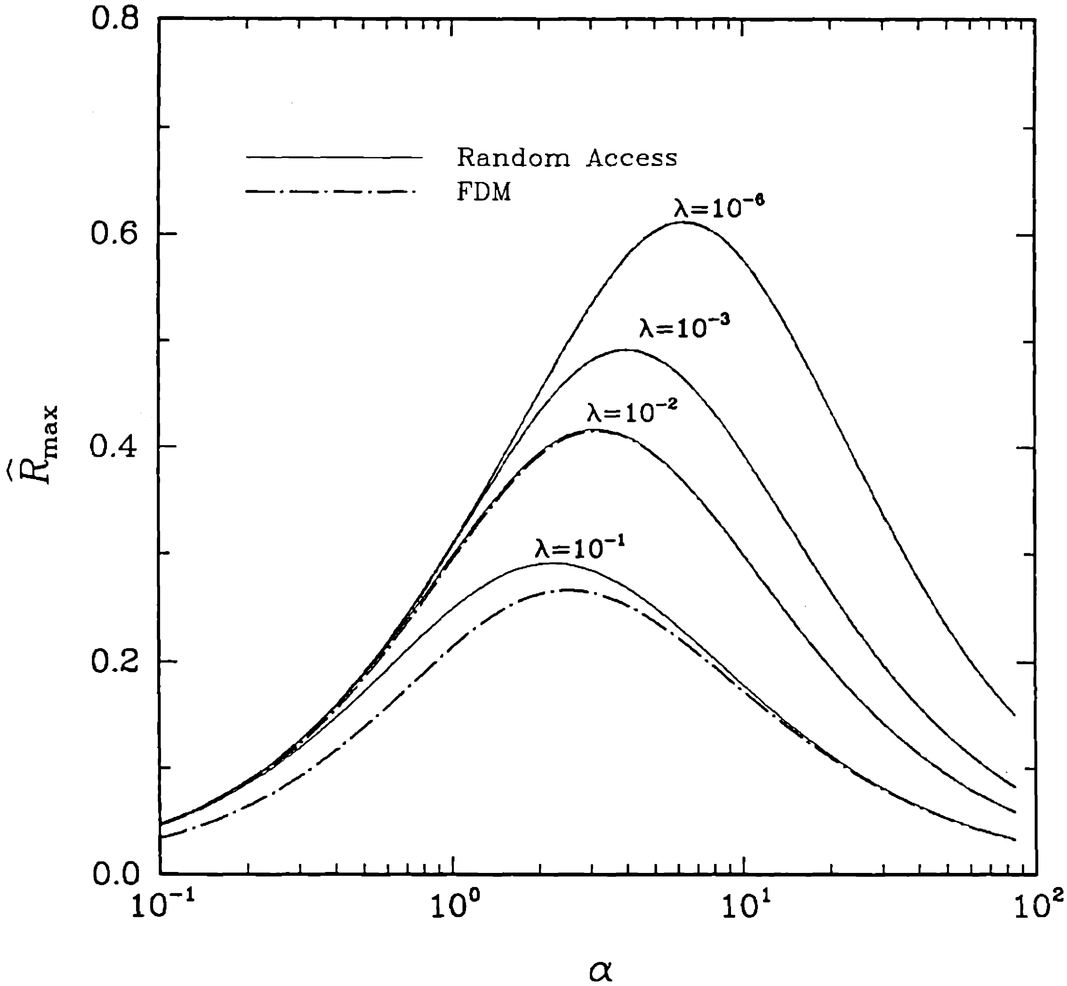

Figure 4.3: Comparison of the FDM and Random Access schemes.

use the point to point channel to its full capacity, which is not possible with random access. Thus, for this case FDM would perform better than random access.

For the Rayleigh fading channel with the *ADD*-combiner, the test quantity in (4.8) evaluates to

$$\int_0^\infty \frac{1}{\alpha_2} e^{-y/\alpha_2} \ln \frac{e^{y/\alpha_2}}{\alpha_2 e^{-y}}\, dy - 2\int_0^\infty \frac{1}{\alpha_1} e^{-y/\alpha_1} \ln \frac{e^{-y/\alpha_1}}{\alpha_1 e^{-y}}\, dy = \ln \frac{\alpha_1^2}{\alpha_2} + \alpha_2 + 1 - 2\alpha_1,$$

using $\alpha_k = 1 + k\alpha$, the above becomes

$$\ln \frac{\alpha_1^2}{\alpha_2} = \ln \frac{1 + 2\alpha + \alpha^2}{1 + 2\alpha} > 0.$$

Hence for the Rayleigh fading channel with the *ADD*-combiner there is always a region of λ for which random access outperforms FDM. Figure 4.3 shows the comparison of the FDM and random access for the Rayleigh fading channel with the *ADD*-combiner. The figure suggests that random access is superior to FDM not only for small λ but for all λ.

It is interesting to note that, if the combiner were an *OR*-combiner, i.e.,

$$\forall k > 0 \quad p_k(y) = p_1(y),$$

then we would have

$$\int_Y p_2(y) \ln \frac{p_2(y)}{p_0(y)}\, dy - 2\int_Y p_1(y) \ln \frac{p_1(y)}{p_0(y)}\, dy =$$
$$\int_Y p_1(y) \ln \frac{p_0(y)}{p_1(y)}\, dy \leq \int_Y p_1(y) - p_0(y)\, dy = 0.$$

where the last inequality follows from $\ln x \leq x - 1$, with equality, if and only if, $p_0(y) = p_1(y)$ almost everywhere. Thus for the *OR*-combiner, regardless of the nature of the channel FDM will perform at least as well as random access, for some region of λ.

4.A Appendix: The Limiting Behavior of $p_Y(y)$

In Section 4.2.1 we had derived that, for a Rayleigh fading channel used by an infinite number of users each accessing the channel with a vanishingly small probability, the channel output distribution, $p_Y(y)$, is given by

$$p_Y(y) = \sum_{k=0}^{\infty} \mu_k p_k(y), \tag{4.9}$$

where

$$\mu_k = e^{-\lambda}\frac{\lambda^k}{k!}, \tag{4.10}$$

$$p_k(y) = \frac{1}{\alpha_k}e^{-y/\alpha_k}, \tag{4.11}$$

$$\alpha_k = 1 + k\alpha. \tag{4.12}$$

In this section we will investigate the behavior of $p_Y(y)$ for large y. We begin by considering the characteristic function $\Psi_Y(s) = E[\exp sY]$ of the random variable Y. Then

$$\Psi_Y(s) = \sum_{k=0}^{\infty} \mu_k \Psi_k(s),$$

where

$$\begin{aligned}\Psi_k(s) &= \int_0^{\infty} e^{sy}\frac{1}{\alpha_k}e^{-y/\alpha_k}\,dy \\ &= \frac{1}{1-\alpha_k s} \quad \text{for } s < 1/\alpha_k.\end{aligned}$$

Hence,

$$\Psi_Y(s) = \sum_{k=0}^{\infty} \frac{\mu_k}{1-\alpha_k s} \quad \text{for } s \leq 0.$$

The above characteristic function exhibits a rare behavior of *not* existing on any open interval containing the origin. The conclusion is that, $p_Y(y)$ decays slower than any exponential. It is also interesting to note that all the finite moments of Y are finite, hence $p_Y(y)$ decays faster than any inverse power.

We state the following result on the behavior of $p_Y(y)$:

Lemma 4.2

$$\lim_{y\to\infty} e^{-\beta\sqrt{y\ln y}}p_Y(y) = \begin{cases} 0 & \text{for } \beta > -\sqrt{2/\alpha}, \\ \infty & \text{for } \beta < -\sqrt{2/\alpha}. \end{cases} \tag{4.13}$$

Proof. *Let* $a_k(y) = e^{-\beta\sqrt{y \ln y}} \mu_k p_k(y)$. *Then for large* k,

$$a_k(y) = \frac{e^{-\lambda}}{\alpha\sqrt{2\pi}} \exp\left[-k \ln k + k - \frac{3}{2}\ln k - \frac{y}{\alpha k} - \beta\sqrt{y \ln y} + o(k)\right],$$

where $o(k)$ *denotes terms which approach zero as* $k \to \infty$, *independent of* y. *Now consider a sequence* $y_k = \alpha k^2 \ln k$. *Then* $y_k \to \infty$ *as* $k \to \infty$, *and*

$$a_k(y_k) = \frac{e^{-\lambda}}{\alpha\sqrt{2\pi}} \exp\left[-\beta\sqrt{2\alpha} k \ln k \sqrt{1 + \frac{\ln(\alpha \ln k)}{2 \ln k}} - 2k \ln k - \frac{3}{2}\ln k + o(k)\right].$$

For $\beta < -\sqrt{2/\alpha}$, *the above term becomes arbitrarily large as* $k \to \infty$. *Since*

$$e^{-\beta\sqrt{y_k \ln y_k}} p_Y(y_k) = \sum_{\ell=0}^{\infty} a_\ell(y_\ell),$$

and all terms are non-negative, $e^{-\beta\sqrt{y_k \ln y_k}} p_Y(y_k)$ *becomes arbitrarily large as* $k \to \infty$, *resulting in the second half of the lemma.*

Now consider $\beta > -\sqrt{2/\alpha}$. *Then we claim that* $a_k(y) \to 0$ *uniformly in* k *as* $y \to \infty$. *To prove the claim we proceed as follows:*

It suffices to show that the largest of the terms $\{a_k(y)\}_{k=0}^{\infty}$ *approaches zero as* y *approaches* ∞. *To this end we observe that the maximal term for large* y *satisfies* $\alpha k_{\max}^2 \ln k_{\max} = y$, $k_{\max}$ *is an increasing function of* y, *and* $k_{\max} \to \infty$ *as* $y \to \infty$. *In this case*

$$a_{k_{\max}}(y) = \frac{e^{-\lambda}}{\alpha\sqrt{2\pi}} \exp\left[\begin{array}{c} -\beta\sqrt{2\alpha} k_{\max} \ln k_{\max} \sqrt{1 + \frac{\ln(\alpha \ln k_{\max})}{2 \ln k_{\max}}} \\ -2k_{\max} \ln k_{\max} - \frac{3}{2}\ln k_{\max} + o(k_{\max}) \end{array}\right].$$

With $\beta > -\sqrt{2/\alpha}$, *the above term approaches 0, hence the proof of the claim.*

Also since $\sum_k a_k(y)$ *converges for every* y, *and* $a_k(y)$ *are non-negative, we have*

$$\lim_{y\to\infty} \sum_{k=0}^{\infty} a_k(y) = \sum_{k=0}^{\infty} \lim_{y\to\infty} a_k(y) = 0,$$

completing the proof. □

Chapter 5

Conclusion

The performance of a class of channels under energy limitations is studied with particular interest in the Rayleigh fading channel. It is shown that to achieve high capacities on a Rayleigh fading channel, in terms of bits/energy, very complex encoding and decoding schemes are necessary. For the case of single user communication, the channel capacity $\widehat{C}$, and the random coding exponent $\widehat{E}_r$, are evaluated and the dependence of these quantities on the energy constraint δ and the channel SNR α is investigated.

The multiaccess performance of the Rayleigh fading channel is evaluated for the special case of a large number of equal rate, equal energy users, and a comparison of FDM and random access schemes is presented. It is shown that the random access scheme performs better than FDM for at least small access rates.

A few problems remain open, however:

- In Section 3.2 we stated conditions under which the binary use of a non-binary channel is essentially optimal. These conditions seem to hold for any well behaved channel. A better understanding of these conditions seem essential to the further development of the area.

- In Section 4.2.1 we just considered the FDM and random access regions, and assumed that the directly achievable region will not expand in the direction of the main diagonal (equal rates) if we take the convex hull. This statement is not true in general, and is difficult to verify even in simple cases. A simple test to guarantee the statement would be very useful.

References

[ChW 81] Shih-Chun Chang and Jack K. Wolf, "On the T-User M-Frequency Noiseless Multiple-Access Channel with and Without Intensity Information," *IEEE Transactions on Information Theory,* Vol. IT-27, no. 1, January 1981.

[CsK 81] I. Csizár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems.* New York: Academic Press, 1981.

[Gal 68] R. G. Gallager, *Information Theory and Reliable Communication.* New York: Wiley, 1968.

[Gal 87] R. G. Gallager, "Energy Limited Channels: Coding, Multiaccess and Spread Spectrum," M.I.T. LIDS Report LIDS–P–1714, November 1987.

[GrR 80] I. S. Gradshteyn and I. M. Ryzhik, *Table of Series, Integrals and Products,* Corrected and Enlarged Edition. New York: Academic Press, 1980.

[Ken 69] R. S. Kennedy, *Fading Dispersive Channels.* New York: Wiley, 1969.

[SGB 67] C. E. Shannon, R. G. Gallager and E. R. Berlekamp, "Lower Bounds to Error Probability for Coding on Discrete Memoryless Channels. Part I," *Information and Control,* Vol. 10, No. 1, February, pp. 65–103, 1967.

[Ulr 75] M. L. Ulrey, "The Capacity Region of a Channel with s senders and r receivers," *Information and Control,* Vol. 29, pp. 185–203, 1975.