Multi-Access Communications with Decision Feedback Decoding

by

İbrahim Emre Telatar

S.M. Massachusetts Institute of Technology (1988)
B.S. Middle East Technical University (1986)

Submitted to the Department of Electrical Engineering and Computer Science in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 1992

Signature of Author______________________________
Department of Electrical Engineering and Computer Science
May 1992

Certified by______________________________
Robert G. Gallager
Fujitsu Professor of Electrical Engineering and Computer Science
Thesis Supervisor

Accepted by______________________________
Campbell L. Searle
Professor of Electrical Engineering and Computer Science

Multi-Access Communications with Decision Feedback Decoding

by

İbrahim Emre Telatar

Submitted to the Department of Electrical Engineering and Computer Science
on May 1992, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

There are two fundamental issues of interest in multi-access communications: the stochastic nature of the arrival of messages to be transmitted and the noise and interference that affect the transmission of these messages. The two main schools of research in this field are split by these issues. Each concentrates on one (and only one) of the issues and the classical analysis techniques they use either ignore or trivialize the other aspect of the problem. In this thesis we will present an analysis that takes into account both the random nature of message arrivals and the random nature of the transmission process. This is accomplished by noticing an equivalence between certain aspects of multi-access communications and the classical queueing theoretic problem of processor sharing queues, and by combining this insight with an information theoretic analysis of the multi-access channels.

The latter part of this thesis deals with the information theoretic analysis of errors-and-erasures decoding schemes for single-user discrete memoryless channels. This problem arises naturally from our multi-access communication model, in the sense that such decoding schemes underlie every automatic repeat request system in which the receiver utilizes a feedback line to request retransmissions in case of a detected error (detected errors are also called erasures). Errors-and-erasures systems have been previously analyzed by FORNEY in the late 1960s. Here we will improve upon his results and derive new exponential upper bounds on error and erasure probabilities. A by-product of these results will be a stronger lower bound to the zero-undetected-error capacity of discrete memoryless channels.

Thesis Supervisor: Robert G. Gallager
Title: Fujitsu Professor of Electrical Engineering and Computer Science

Acknowledgements

I thank my advisor, Professor Robert Gallager, for his guidance throughout my graduate study. Without his supervision, insight and contributions this work could not have been possible. To work with him has been a great pleasure and an educational experience in itself.

I thank my thesis committee members, Professors Peter Elias and Sanjoy Mitter, for their helpful comments on my research and presentation.

I thank Professor Pierre Humblet, Professor John Tsitsiklis and Dr. Ofer Zeitouni for their help at various stages of this research. I thank Professors Daniel Stroock and Al Drake for their keen interest in my career.

Over the years I have subjected my officemates and friends at LIDS to endless discussions. I would like to thank the following for their patience and camaraderie that made life as a graduate student an enjoyable experience: Mohamad Akra, Dr. Murat Azizoğlu, Rick Barry, Diana Dabby, Dr. Walid Hamdy, Diane Ho, Dr. Sanjeev Kulkarni, Daniel Chonghwan Lee, Nancy Lee, Dr. Whay Lee, Peter Li, Ying Li, Douglas Mar, Muriel Medard, Dr. Hasan Sait Ölmez, Dr. Cüneyt Özveren, Rajesh Pankaj, Dr. Abhay Parekh, Dr. Tom Richardson, Serap Savari, Jane Simmons, Dr. John Spinelli, Dr. George Stamoulis, Bruno Suard, David Tse, Manos Varvarigos and Murat Veysoğlu. Of these, I especially thank Nancy Lee for her friendship, care, and encouragement. I thank the administrative staff at LIDS, in particular Kathleen O'Sullivan, for providing a friendly atmosphere and making life easy in general. I thank Giampiero Sciutto for keeping the computers running.

I thank my parents, Ferzan and Hasan Telatar, for their love and support.

This work was supported by the U.S. Army Research Office under contract DAAL03-86-K-0171. This funding is greatly appreciated.

Contents

List of Figures

Chapter 1

Introduction

A multi-access communication system consists of a set of transmitters sending information to a single receiver. Each transmitter is fed by an information source generating a sequence of messages; the successive messages arrive for transmission at "random" times. It is usual to assume that the arrival process is POISSON. The signal received at the receiver is related (stochastically) to the signals sent by the transmitters. A common example is the additive GAUSSian noise channel where the received signal consists of the summation of the transmitted signals with an independent GAUSSian noise. We will further assume that the feedback from the receiver is limited; in particular, the possibility of any transmitter observing the received signal is ruled out.

From the description above one sees that there are two issues of interest: (i) the random arrival of messages to the transmitters and (ii) the noise and interference during transmission. The main bodies of research in multi-access communications seem to treat these two issues as if they were separable. The collision resolution approach focuses on the random arrival of messages but ignores noise and trivializes the (mutual) interference of the transmitted signals. The multi-access information theoretic approach, on the other hand, develops accurate models for the transmission process (noise and interference) but the random arrival of messages is totally ignored. In addition, one can say with some oversimplification that the results generated by the two approaches are of different character. The information theoretic results mostly state upper and lower bounds (which sometimes coincide) to the performance of the best possible scheme, whereas collision resolution results mostly analyze the performance of particular algorithms.

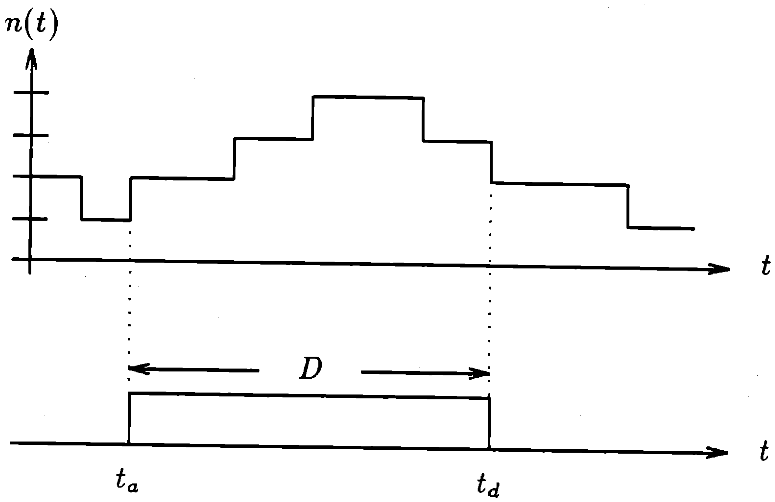
The obstacle faced by the multi-access information theorists in analyzing delay in multi-access systems is twofold. First is the arbitrarily large intervals of time that are required for the coding theorems. Once we have imposed these arbitrarily large time intervals, any delay analysis will be meaningless. This is largely a theoretical problem, and may be worked around. The second is the assumption that each user of the system is active all the time. In principle, by appropriate source coding to remove the randomness in the arrival process, this may be achieved. But, using source coding to remove the burstiness in the arrival of messages will inevitably introduce large delays, thus trivializing the attempts to analyze the delay in the system futile. Note that this second problem is more than a technicality because the delays required to smooth sources is typically orders of magnitude larger than the delays introduced by coding.

In the following few paragraphs we will discuss a somewhat contrived example and will analyze the performance of this multiple-access system by using tools borrowed from both the approaches described above.

Suppose our multi-access environment consists of an additive GAUSSian noise channel. All our transmitters have equal power P, and the noise density is $N_0/2$ over a frequency band of W. Suppose messages arrive at the transmitters in accordance with a POISSON process and the aggregate arrival rate is λ. We will make the simplifying assumption that a new transmitter is created for every message that enters the system. This is known as the infinite user population assumption, and it removes the problem of dealing with queues of messages at each transmitter. As soon as a message arrives, the transmitter will encode it into an infinite duration time signal and begin transmitting it. However, the transmitter will not transmit the whole duration of the signal; it will transmit only until the receiver decodes the message and instructs the transmitter to stop (see Figure 1-1). Thus, if the system is stable, with probability 1 only a finite initial segment of the infinite duration codeword will be transmitted. The decoder will treat each transmitter independently; each message is decoded regarding the other transmissions as noise. If there are n active transmitters at a given time, the signal to noise ratio for a particular transmitter is $P/((n-1)P + N_0W)$. At this point let us assume that the decoder can resolve

$$W \log_2\Bigl(1 + \frac{P}{(n-1)P + N_0W}\Bigr)$$

bits per unit time for each transmitter. Note that this is an unjustified assumption. There is no known coding theorem that guarantees that such a decoder exists. The formula is simply the information theoretic capacity of an additive GAUSSian noise channel with the said signal to noise ratio. With this remark, it is clear that the analysis that follows is *unjustified.* However it provides the intuitive setting in which to understand the essential ideas of a correct analysis presented in Chapter 3. With the assumption, the decoder has



In the figure above, $n(t)$ denotes the number of active transmitters. The lower illustration focuses on a particular transmitter. A message arrives at t_a and is transmitted until t_d at which time it is decoded at the receiver. The duration D of transmission is a random variable which is dependent on the values of $n(t)$ for $t \geq t_a$.

Figure 1-1: Transmission of a packet in the example system

a total information resolving power of

$$nW \log_2\left(1 + \frac{P}{(n-1)P + N_0 W}\right) \quad \text{bits/unit-time}$$

which it shares equally among the active transmitters.

One can liken the situation to that of a processor sharing system where jobs compete for the processor's time. The role of the 'jobs' are taken by the transmitters that are 'served' by the decoder. The more transmitters that are active at a given time, the less 'service' each will get due to interference. We can indeed formulate the problem as a classical processor sharing system in queueing theory, with the following difference: the total processing power of the processor depends on the number of jobs competing for service. The next chapter (Chapter 2) analyzes a processor sharing system with a processor whose power depends on the state of the system. The key observation is that such a system is "quasi-reversible", and this observation allows us to solve for the steady-state probabilities of various events. In particular one can characterize the delay experienced by the incoming messages. The last section in Chapter 2 presents the results obtained when one applies the processor sharing results to our example system.

Note that essential to the working of our example system is the feedback that transmitters receive from the receiver. To understand how these kinds of decision feedback schemes work we will focus on a single user system of this type in Chapter 4. An important result of Chapter 4 will be a new lower bound for the zero-undetected-error capacity of discrete memoryless channels. This lower bound is better than the previously known bounds and we will identify the channels on which our bound is exact. The approach used in these chapters is purely information theoretic. Our efforts to prove a converse to show that the new lower bound is exact for all channels led us to a counter-example: a channel whose zero-undetected-error capacity is larger than the said lower bound. The construction of this example is done in Chapter 5.

Chapter 2

A Queueing Theoretic Result

In the introduction we pointed out the similarity between a particular multi-access scheme and processor sharing queues. The classical queuing theoretic results on processor sharing do not allow the server's rate to depend on the state of the system. In this section we will generalize the classical results to apply to queues in which the server's rate depends on the state of the system. The notion of server rate will be defined during the course of the discussion. The results of the following section can also be found in [Kel79]. However, the derivations we give for these results differ significantly from those in [Kel79] and involve a more natural discretization argument.

2.1 Queues With Processor Sharing

Suppose that customers in a processor sharing system arrive in accordance to a POISSON process of rate λ. Each customer requires a random amount of service, S, distributed according to G:

$$\Pr[S \leq s] = G(s).$$

The service requirements of customers are independent. Given u customers in the system the server can provide service at a rate of $\phi(u) > 0$ units of service per unit time, and divides this service equally among all the customers in the system. That is, whenever there are u customers in the system, each will receive service at a rate of $\phi(u)/u$ per unit time. A customer will depart the system when the service it has received equals its service requirement.

We will assume that $\sup_u \phi(u) < \infty$. In this case we may further assume that $\sup_u \phi(u) \leq 1$ since we can define $S' = S/\sup_u \phi(u)$ and $\phi'(i) = \phi(i)/\sup_u \phi(u)$, and the new system will be identical in performance to the original.

To analyze the system we adopt a discrete time approximation, dividing the time into slots of duration h. Define $\lambda_d = \lambda h$, the probability of an arrival in any time slot, $S_d = h\lceil S/h \rceil$, which has distribution $G_d(k) \stackrel{\text{def}}{=} \Pr[S_d \le kh] = G(hk)$. Further define $g_d(k) \stackrel{\text{def}}{=} \Pr[S_d = hk] = G_d(k) - G_d(k-1)$, and $q(k) = g_d(k+1)/\overline{G}_d(k)$; that is, $q(k)$ is the probability that a customer will depart when it is next served given that it has already received kh units of service. Note that $\overline{G}_d$ denotes $1 - G_d$. Let the state of the system be $\mathbf{s} = (u_0, u_1, \ldots)$, where u_i is the number of customers in the system who have already received ih units of service, and let $u = \sum_i u_i$. We will restrict the state-space to those sequences that are summable and therefore the state-space will be countable. Let $p(\mathbf{s})$ be the limiting state probabilities and let $P(\emptyset)$ be the probability that the system is empty. The server's behavior is described as follows: at the beginning of each time slot, the server

- either, with probability λ_d, accepts a new customer into the system;
- or, with probability $(1 - \lambda_d)$, chooses an existing customer at random, and with probability $\phi(u)$ gives it h units of service.

We will let $\Pr[\mathbf{s} \to \mathbf{s}']$ denote the transition probabilities of the corresponding MARKOV chain. To solve for the limiting state probabilities we will employ the following theorem [Ros83, p. 128]:

> Consider an irreducible MARKOV chain with countable state space E and transition probabilities $\{P_{ij},\, i,j \in E\}$. If one can find non-negative numbers $\{\pi_i,\, i \in E\}$ summing to unity, and a transition probability matrix $\{P^*_{ij},\, i,j \in E\}$ such that
>
> $$\pi_i P_{ij} = \pi_j P^*_{ji}, \quad i,j \in E$$
>
> then $\{\pi_i,\, i \in E\}$ are the stationary probabilities and $\{P^*_{ij},\, i,j \in E\}$ are the transition probabilities of the reversed chain.

Let $\Pr^*[\mathbf{s} \to \mathbf{s}']$ denote the transition probabilities of the reversed chain. We make the following conjecture regarding the reversed process (i.e., a guess at the $\Pr^*$'s):

Conjecture 2.1 *The reverse process is a system of the same type, the state representing the numbers of customers indexed by the residual workload less h. That is $\mathbf{s} = (u_0, u_1, \ldots)$ where u_i is the number of customers who need $(i+1)h$ units of service to depart.*

To analyze the system and verify the conjecture we define $e_i : E \to E$ as

$$e_i(\mathbf{s}) = (u_0, u_{i-1}, \ldots, u_i - 1, u_{i+1}, \ldots)$$

for $\mathbf{s} = (u_0, \ldots, u_{i-1}, u_i, u_{i+1}, \ldots)$, and $u_i > 0$.

With the conjecture we have

in forward: $\mathbf{s} \to e_i(\mathbf{s})$ with probability $(1-\lambda_d)q(i)u_i\phi(u)/u$;
in reverse: $e_i(\mathbf{s}) \to \mathbf{s}$ with (joint) probability $\lambda_d g_d(i+1)$.

Hence if $p(\mathbf{s})$ is the limiting probability, then

$$p(\mathbf{s})(1-\lambda_d)q(i)\frac{u_i\phi(u)}{u} = p(e_i(\mathbf{s}))\lambda_d g_d(i+1),$$

leading to

$$p(\mathbf{s}) = \frac{\lambda_d}{1-\lambda_d}\frac{u}{u_i\phi(u)}\overline{G}_d(i)p(e_i(\mathbf{s})).$$

Solving this iteratively,

$$p(\mathbf{s}) = \frac{1}{\phi_!(u)}\begin{pmatrix} & u & \\ u_0 & u_1 & \ldots \end{pmatrix}\left(\frac{\lambda_d}{1-\lambda_d}\right)^u P(\emptyset)\prod_i \overline{G}_d(i)^{u_i}, \tag{2.1}$$

where $\phi_!(u) = \prod_{i=1}^{u}\phi(i)$. Summing over $u_0, u_1, \ldots$ such that $\sum_i u_i = u$, and making use of the multinomial formula

$$\left(\sum_i a_i\right)^u = \sum_{\{\{u_i\}_i | u_i \geq 0, \sum_i u_i = u\}}\begin{pmatrix} & u & \\ u_0 & u_1 & \ldots \end{pmatrix}\prod_i a_i^{u_i},$$

and noticing that $\sum_{i\geq 0}\overline{G}_d(i) = E[S_d]$, we have

$$\Pr\{u \text{ customers in the system}\} = \frac{1}{\phi_!(u)}\left(\frac{\lambda_d}{1-\lambda_d}E[S_d]\right)^u P(\emptyset), \quad u > 0.$$

Using

$$P(\emptyset) + \sum_{u=1}^{\infty}\Pr\{u \text{ customers in the system}\} = 1$$

we obtain

$$\Pr\{u \text{ customers in the system}\} = \frac{1}{K\phi_!(u)}\left(\frac{\lambda_d}{1-\lambda_d}E[S_d]\right)^u, \quad u \geq 0.$$

where $K = 1 + \sum_{u=1}^{\infty}\left(\frac{\lambda_d}{1-\lambda_d}E[S_d]\right)^u \Big/ \phi_!(u)$.

To verify the conjecture we need to verify that the state probabilities given by equation (2.1) satisfy

$$p(\mathbf{s})\Pr\{\mathbf{s} \to \mathbf{s}'\} = p(\mathbf{s}')\Pr{}^*\{\mathbf{s}' \to \mathbf{s}\} \tag{2.2}$$

for each $\mathbf{s}$ and $\mathbf{s}'$. Both sides of the equation are zero except in the following cases:

1. For some $i \geq 0$, $\mathbf{s}' = e_i(\mathbf{s})$. In forward, this corresponds to the departure of a customer with ih units of service after receiving service one more time. In reverse, it corresponds to the arrival of a customer with service requirement equal to $(i+1)h$. Since $p(\cdot)$ was constructed using (2.2) for this case, the equality is immediate.

2. $\mathbf{s} = (u_0, u_1, \ldots) = e_0(\mathbf{s}')$. In forward this corresponds to the arrival of a customer. In reverse, it corresponds to the departure of a customer after receiving one more unit of service. Thus

$$\Pr\{\mathbf{s} \to \mathbf{s}'\} = \lambda_d, \quad \Pr{}^*\{\mathbf{s}' \to \mathbf{s}\} = (1-\lambda_d)\phi(u+1)\frac{u_0+1}{u+1}.$$

Substituting equation (2.1) and noticing that $\overline{G}_d(0) = 1$, the equality in (2.2) follows.

3. $\mathbf{s} = (u_0, \ldots, u_i, u_{i+1}, \ldots)$ and $\mathbf{s}' = (u_0, \ldots, u_i - 1, u_{i+1}+1, \ldots)$. In both forward and reverse, this corresponds to giving a customer a unit of service and the customer staying in the system. Thus

$$\Pr\{\mathbf{s} \to \mathbf{s}'\} = (1-\lambda_d)\left(\overline{G}_d(i+1)/\overline{G}_d(i)\right) u_i\phi(u)/u,$$

$$\Pr{}^*\{\mathbf{s}' \to \mathbf{s}\} = (1-\lambda_d)(u_{i+1}+1)\,\phi(u)/u.$$

Substituting equation (2.1) and noticing that

$$u_i\begin{pmatrix} & & u & \\ \cdots & u_i & u_{i+1} & \cdots \end{pmatrix} = (u_{i+1}+1)\begin{pmatrix} & & u & \\ \cdots & u_i - 1 & u_{i+1}+1 & \cdots \end{pmatrix},$$

we see that equation (2.2) is satisfied and the conjecture is thus verified.

The conditional probability of the state given u customers in the system is

$$p(\mathbf{s}|u) = p(\mathbf{s})/\Pr\{u \text{ customers in the system}\} = \begin{pmatrix} & u & \\ u_0 & u_1 & \cdots \end{pmatrix} \prod_i \left[\frac{\overline{G}_d(i)}{E[S_d]}\right]^{u_i},$$

which depends only on the equilibrium distribution G_{de} of G_d, defined by $g_{de} = \overline{G}_d/E[S_d]$.

As h tends to 0, the scheme tends to a continuous time processor sharing system; it is easy to see that $\lim_{h\to 0} \lambda_d E[S_d] = \lambda E[S]$. It is intuitively clear that in the limit the randomized service scheme converges to processor sharing. This latter fact can be easily proved as follows: let (s,t) be a time interval during which no arrivals or departures occur. The number of discrete time intervals in this period is $N = (t-s)/h$. Let u be the number of customers in the system during this interval and consider a particular customer in the system. For $1 \le i \le N$, let $X_i = 1$ if this customer receives service at the ith discrete interval and let $X_i = 0$ otherwise. Note that $\{X_i\}_{i=0}^N$ forms an independent and identically distributed sequence of random variables with $\Pr[X_i = 1] = \phi(u)/u$. Thus, the total amount of service this customer receives in this interval is

$$h\sum_{i=1}^{N} X_i = (t-s)N^{-1}\sum_{i=1}^{N} X_i$$

which converges to $(t-s)\phi(u)/u$ with probability 1 as $h \to 0$. Therefore in the limit as $h \to 0$, the service scheme approaches processor sharing. Now we can state a continuous time result.

Theorem 2.2 *For the Processor Sharing Model the number of customers in the system has the distribution*

$$\Pr\{u \textit{ customers in the system}\} = \frac{1}{K\phi_!(u)}(\lambda E[S])^u, \quad u \ge 0,$$

where $K = 1 + \sum_{u=1}^{\infty}(\lambda E[S])^u/\phi_!(u)$. *Given u customers in the system, the completed (or residual) workloads are independent and have distribution G_e defined by $g_e = \overline{G}/E[S]$. The departure process is* POISSON *with rate λ.*

A particularly interesting case is when $\phi(u) = 1$ for all u. Then $K = (1-\lambda E[S])^{-1}$ and

$$\Pr\{u \text{ customers in the system}\} = (1-\lambda E[S])(\lambda E[S])^u, \quad u \ge 0.$$

This last result is the well known processor sharing result [Ros83]. It is also worth noting that one can obtain this processor sharing result by using a round-robin server [Yat90]. In this case the server maintains a list of customers in the system and gives them service by going through the list one by one, starting from the beginning when it reaches the end. The new customers are placed at the current position in the list so that they are served immediately.

2.2 Additive GAUSSian Noise Channel With Deterministic Service

As advertised in the introduction, we now use the results above to analyze the multi-access model considered in the introduction. Let us recall that we are considering an additive GAUSSian noise, multiuser communication system with noise density N_0, bandwidth W and equal power users each with power P. With independent decoding and random coding, given that there are u users in the system,

$$W \log_2 \left(1 + \frac{P}{(u-1)\,P + N_0 W}\right) \quad \text{bits per unit time}$$

of mutual information for each user is generated per unit time. We will assume that the decoder will be able to resolve the same number of bits per unit time. Note that the mutual information model is exact, that is, the formula above is indeed equal to the average mutual information between the transmitted message and the received signal. What is not exact is, whether a message can be decoded with small probability of error when the average mutual information between the message and the received signal is equal to the number of bits in the message. We may attempt to decode the message and let the customer depart when the number of bits in the message is equal to the average mutual information; this will make our analysis exact, but we cannot guarantee that the error probability (the probability that the decoded message is different from the message sent) will be small.

With the assumption that the decoder can decode as many bits of the message as there is mutual information, we can set the service requirement S of a message to be the number of nats[1] in the message and then the function $\phi(u)$ above is given by

$$\phi(u) = uW \ln\left(1 + \frac{1}{u - 1 + N_0W/P}\right) \quad \text{nats per unit time.}$$

The extra factor u is to take into account that ϕ is a measure of the *total* (not per user) service rate. Since we can compute the statistics of the number of customers in the system, we may use LITTLE's law to compute the average waiting time. For a given value of SNR $\stackrel{\text{def}}{=} P/(N_0W)$, the average number of customers in the system is a function of $\ell \stackrel{\text{def}}{=} \lambda E[S]/W$, which is the loading of the system in terms of nats per second per unit bandwidth: λ is the arrival rate of the messages, thus $\lambda E[S]$ is the nat arrival rate in nats per unit time, and this quantity is further normalized by the available bandwidth. Figure 2-1 shows the dependence of average waiting time to the signal to noise ratio SNR and the loading ℓ. Note that since

$$\lim_{u\to\infty} u \ln\left(1 + \frac{1}{u - 1 + \text{SNR}^{-1}}\right) = 1,$$

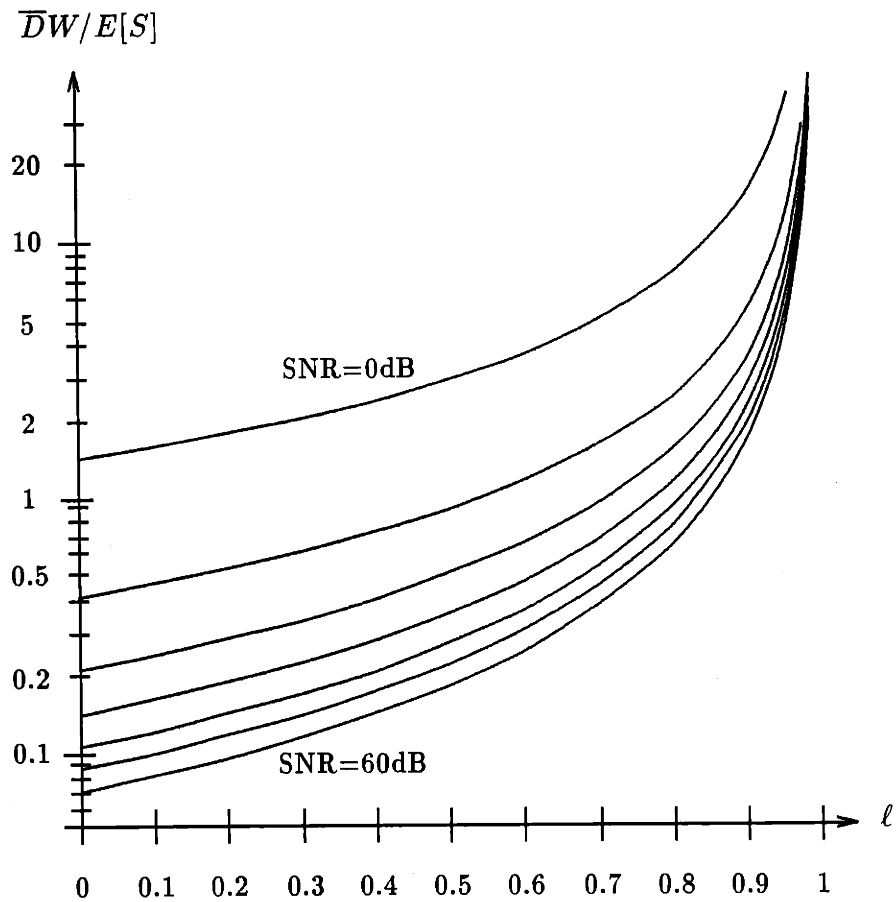
the sum

$$K = 1 + \sum_{u=1}^{\infty} (\lambda E[S])^u / \phi!(u)$$

exists when ℓ is strictly less than unity and diverges for $\ell \geq 1$. Thus the system is stable if and only if $\ell < 1$. In a very real sense then, the throughput of the system is 1 nat per second per Hertz. Note that in the definition of ℓ, the value of $E[S]$ is measured in natural units, not in bits.

It will turn out that the results of this section can be interpreted as the limiting case of the results in the next chapter as the message length gets large.

[1] We will find it convenient to use natural units instead of binary units. Here 1 bit = ln 2 nats.



The average waiting time $\overline{D}$ is shown as a function of the load $\ell \stackrel{\text{def}}{=} \lambda E[S]/W$ and the signal to noise ratio (SNR). The six curves correspond to different SNR values ranging from 0dB to 60dB in increments of 10dB. The delay is normalized by the average service requirement per unit bandwidth $E[S]/W$. Note that the service requirement is measured in natural units, not in bits.

Figure 2-1: Average delay as a function of loading and SNR

Chapter 3

Multi-Access Systems

In this chapter we will present an analysis of a multi-access scheme over an additive white GAUSSian noise channel. In our analysis we will model both the random arrival of messages and the channel noise and interference. The first section presents our model of the channel. We will show that any additive GAUSSian noise channel, whose inputs and outputs consist of waveforms, can be represented as a sequence of scalar GAUSSian noise channels, thereby allowing us to use the tools of information theory. Although this is a long known result, we feel that it is largely ignored outside of the information theory community. A reader already familiar with this representation of waveform channels can skip the following section and continue with Section 3.2 on page 26. That section introduces our multi-access model and shows how, by combining some results from the theory of processor sharing queues and the classical information theoretical analysis technique of random coding, we can get bounds on the performance of this multi-access scheme. Since our model takes into account both arrival and transmission processes in a multi-access system we are able to quantify the tradeoff between the transmission parameters like error probability and arrival parameters like queueing delay and arrival rate.

3.1 Waveform Channels

Most communication problems encountered in practice include a channel whose inputs and outputs are waveforms over some time interval. The usual approach in analyzing such channels from an information theoretic point of view is to assume that the inputs and outputs of the channel belong to some class of functions with a set of properties that allow us to represent each input and output by a countable sequence of coefficients. The channel, at least conceptually, can be specified by describing how the countable sequence that specifies the output relates to the countable sequence that specifies the input. The next section will present an example, namely the case of a bandpass additive white GAUSSian noise channel, before presenting a general treatment of a linear channel with GAUSSian noise. Readers who are not interested in HILBERT spaces may skip the general treatment and take the result on faith. We hope that the example presented by the next section will be convincing enough.

3.1.1 Treatment of the Bandpass Additive GAUSSian Noise Channel

Suppose we are given a channel that is bandlimited to the frequency band $[-f_0 - W/2, -f_0 + W/2] \cup [f_0 - W/2, f_0 + W/2]$ with $f_0 \geq W/2$ and $W > 0$. Assume that the channel response is flat in the band, i.e., the channel transfer function H_c is given by

$$H_c(f) = \begin{cases} 1 & \text{if } \big||f| - f_0\big| \leq W/2 \\ 0 & \text{else.} \end{cases}$$

The input x and output y of the channel are related by

$$y(t) = (h_c * x)(t) + w(t)$$

where h_c is the channel impulse response, $h_c(t) = 2\cos(2\pi f_0 t)\sin(\pi W t)/(\pi t)$, and $w(t)$ is white GAUSSian noise of intensity $N_0/2$, so that, the power spectrum of w is

$$S_w(f) = N_0/2.$$

Since the channel suppresses the component of the input outside the pass-band, we may choose the input signals to be in the pass-band to begin with. Note that any bandlimited signal x can be represented as

$$x(t) = \text{Re}\{\tilde{x}(t)\exp(-j2\pi f_0 t)\}$$

for some $\tilde{x}$ that is bandlimited to $[-W/2, W/2]$. Such a signal, can in turn be represented by uniform samples taken at a rate of W:

$$\tilde{x}(t) = \sum_i (\tilde{x}_i^{(I)} + j\tilde{x}_i^{(Q)})(2W)^{1/2}\frac{\sin(\pi(Wt - i))}{\pi(Wt - i)}.$$

Putting the above together, we see that

$$x(t) = \sum_i x_i^{(I)} \varphi_i^{(I)}(t) + \sum_i x_i^{(Q)} \varphi_i^{(Q)}(t)$$

where

$$\varphi_i^{(I)}(t) = (2W)^{1/2} \frac{\sin(\pi(Wt - i))}{\pi(Wt - i)} \cos 2\pi f_0 t$$

and

$$\varphi_i^{(Q)}(t) = (2W)^{1/2} \frac{\sin(\pi(Wt - i))}{\pi(Wt - i)} \sin 2\pi f_0 t.$$

Note that $\{\varphi_i^{(I)} : i \in \mathbb{Z}\}$ and $\{\varphi_i^{(Q)} : i \in \mathbb{Z}\}$ satisfy the following orthonormality relations:

$$\int \varphi_i^{(I)}(t)\varphi_j^{(I)}(t)\, dt = \int \varphi_i^{(Q)}(t)\varphi_j^{(Q)}(t)\, dt = \delta_{ij} \quad \text{and} \quad \int \varphi_i^{(I)}(t)\varphi_j^{(Q)}(t)\, dt = 0.$$

We can compute $\{x_i^{(I)} : i \in \mathbb{Z}\}$ and $\{x_i^{(Q)} : i \in \mathbb{Z}\}$ from $x(t)$ as follows:

$$x_i^{(I)} = \int x(t)\varphi_i^{(I)}(t)\, dt \quad \text{and} \quad x_i^{(Q)} = \int x(t)\varphi_i^{(Q)}(t)\, dt.$$

At the receiver side, we know that any component of the received signal that is out of the band must be noise, and hence is independent of the transmitted signal, and thus may be ignored. The in-band component of $y(t)$ can be represented in the form that is given for $x(t)$ above:

$$y_{\text{in}}(t) = \sum_i y_i^{(I)} \varphi_i^{(I)}(t) + \sum_i y_i^{(Q)} \varphi_i^{(Q)}(t)$$

where

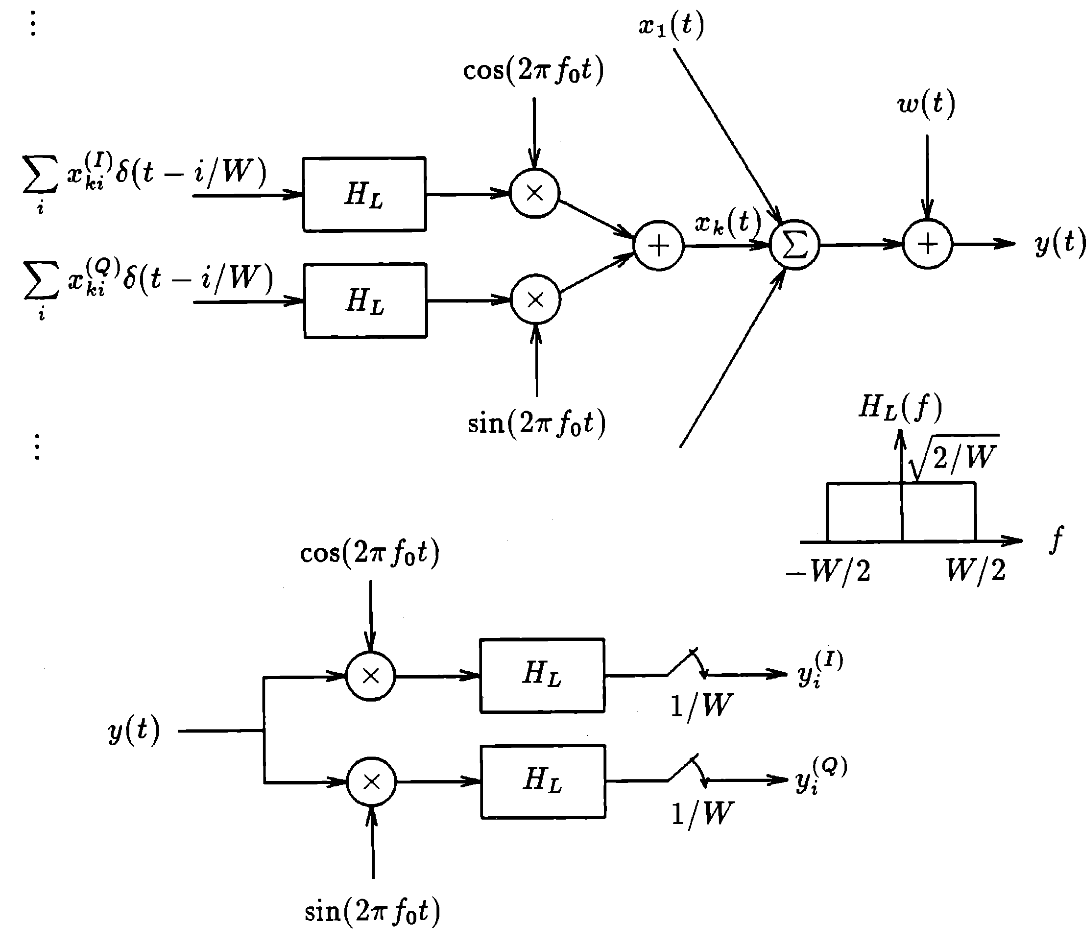
$$y_i^{(I)} = \int y(t)\varphi_i^{(I)}(t)\, dt \quad \text{and} \quad y_i^{(Q)} = \int y(t)\varphi_i^{(Q)}(t)\, dt.$$

If there are multiple users in the channel, each of their signals may be represented as above, and the received signal will be the sum of all their signals and noise. Figure 3-1 shows this case. We then have

$$y_i^{(I)} = \sum_k x_{ki}^{(I)} + n_i^{(I)} \quad \text{and} \quad y_i^{(Q)} = \sum_k x_{ki}^{(Q)} + n_i^{(Q)},$$

where the sum is over the users and

$$n_i^{(I)} = \int w(t)\varphi_i^{(I)}(t)\, dt \quad \text{and} \quad n_i^{(Q)} = \int w(t)\varphi_i^{(Q)}(t)\, dt.$$



The figure shows the model of a multiuser, bandpass, additive GAUSSian noise channel. The model makes use of the fact that any bandpass signal $x(t)$ with spectrum consisting a band of W around f_0 can be written as $x(t) = \mathrm{Re}(\tilde{x}(t) \exp -j2\pi f_0 t)$ where $\tilde{x}$ is a complex lowpass signal occupying the band $[-W/2, W/2]$. Any such low pass signal, in turn, can be represented by its uniform samples taken at a rate of W per second. Since $\tilde{x}$ is complex, these samples have real and imaginary parts, we call them the in-phase (I) and quadrature-phase (Q) components.

Figure 3-1: Model of Bandpass Communications

Recall that $\{\varphi_i^{(I)} : i \in \mathbb{Z}\} \cup \{\varphi_i^{(Q)} : i \in \mathbb{Z}\}$ form an orthonormal sequence, and thus the set of random variables $\{n_i^{(I)} : i \in \mathbb{Z}\} \cup \{n_i^{(Q)} : i \in \mathbb{Z}\}$ is a sequence of independent, identically distributed, GAUSSian, zero mean random variables each with variance $N_0/2$.

We have thus reduced the waveform channel into a sequence of independent *scalar* channels. We can view the waveform channel as follows: for each time t of the form $t = i/W$, the transmitters choose the values of $x_{ki}^{(I)}$ and $x_{ki}^{(Q)}$. This happens W times every unit time interval, and thus each transmitter provides $2W$ samples per unit time. The channel sums the (I) and (Q) values separately, adds noise to each sum and delivers these two values to the receiver as $y_i^{(I)}$ and $y_i^{(Q)}$. Note that the noises added to the (I) and (Q) parts are independent from each other and also independent from the noises in the other time instants.

It can be shown that

$$\lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} x(t)^2\, dt = \lim_{N\to\infty} \frac{W}{2N} \sum_{i=-N}^{N} \left(\left(x_i^{(I)}\right)^2 + \left(x_i^{(Q)}\right)^2 \right)$$

and thus power constraints on x translate to constraints on the sequences $\{x_i^{(I)} : i \in \mathbb{Z}\}$ and $\{x_i^{(Q)} : i \in \mathbb{Z}\}$. In particular, if these sequences are chosen as sequences of independent identically distributed random variables with zero mean and variance $P/(2W)$ then $x(t)$ will have time-average power P with probability 1.

3.1.2 A Mathematical Approach[1]

The channel of the previous section is a special case of linear additive GAUSSian noise channels. Here we present the treatment of this general case. Note that time-invariance is not required. When dealing with linear, additive noise waveform channels, we will find it useful to represent the inputs to the channel (the input waveforms) by the sequence of coefficients of their expansion with respect to a complete orthonormal basis. The implicit assumption is then that the waveforms belong to a separable real HILBERT space X. To recapitulate: given a complete orthonormal basis $\{\varphi_i : i \geq 1\}$ for X, we will represent a waveform $x \in X$ by the sequence $\{x_i \in \mathbb{R} : i \geq 1\}$ where

$$x = \sum_{i\geq 1} x_i \varphi_i,$$

[1] This section is largely based on [Gal68, Chapter 8] ; we have simply translated it into the language of HILBERT spaces. For a treatment of HILBERT spaces the reader is referred to [RSN90] , [Fri82] or [Hal57].

or equivalently $x_i = (x, \varphi_i)$, with $(\cdot,\cdot)$ denoting the inner product in the real HILBERT space. For $x \in X$, we will denote the norm of x by $\|x\| = (x,x)^{1/2}$. Note that, from PARSEVAL's identity, we have

$$\|x\|^2 = \sum_{i\geq 1} |x_i|^2.$$

The channel model will be

$$y = Hx + n$$

with y the output, x the input waveforms, $H : X \to X$ a bounded linear operator, and n "white GAUSSian noise" of intensity $N_0/2$. What we mean by the last statement is the following: for any orthonormal sequence $\{\varphi_i \in X : i \geq 1\}$ the sequence $\{(n, \varphi_i) : i \geq 1\}$ is a sequence of independent GAUSSian random variables, each with mean zero and variance $N_0/2$.

Let $\{\varphi_i \in X : i \geq 1\}$ be the eigenvectors of the operator H^*H, where $H^* : X \to X$ is the adjoint of H. Let $\{\lambda_i : i \geq 1\}$ be the corresponding eigenvalues. Since H^*H is self-adjoint and positive, $\{\lambda_i : i \geq 1\} \subset \mathbb{R}^+$. Without loss of generality we may assume that $\|\varphi_i\| = 1$ for all $i \geq 1$. It then follows that $\{\varphi_i : i \geq 1\}$ form a complete orthonormal set, and also

$$(H\varphi_i, H\varphi_j) = (H^*H\varphi_i, \varphi_j) = \lambda_i(\varphi_i, \varphi_j) = \lambda_i \delta_{ij},$$

and thus $\{\psi_i = \lambda_i^{-1/2} H\varphi_i : i \geq 1, \lambda_i > 0\}$ form an orthonormal set that is complete in the range of H. Let $S = \{i \geq 1 : \lambda_i > 0\}$. Let $P_O : X \to X$ denote the orthogonal projection onto the range of H. Then,

$$P_O z = \sum_{i\in S} (z, \psi_i)\psi_i$$

and it is clear that when we write $y = P_O y + (I - P_O)y$, the second component, consists only of noise, and is independent from the input. Similarly, if P_I denotes the projection onto the space spanned by $\{\varphi_i : i \in S\}$ (that is, the orthonormal complement of the null space of H), then

$$P_I z = \sum_{i\in S} (z, \varphi_i)\varphi_i$$

and if we write $x = P_I x + (I - P_I)x$ we see that only the first component matters in the operation of the channel since the second component produces null output. We can now describe the operation of the channel as

$$y_i = \lambda_i^{1/2} x_i + n_i, \quad i \in S$$

where $y_i = (y, \psi_i)$, $x_i = (x, \varphi_i)$ and $n_i = (n, \psi_i)$. By the definition of n, we see that $\{n_i : i \geq 1\}$ is a sequence of independent, identically distributed, zero mean, GAUSSian random variables, each of variance $N_0/2$. We have now transformed the waveform channel into a sequence of independent *scalar* channels, which we know how to deal with.

As a somewhat more general case, we may consider an additive noise channel where the noise is not white, but is "colored", that is, the noise is of the form Gn where $G : X \to X$ is a bounded linear operator and n is "white" GAUSSian noise. The channel model is then

$$y = Hx + Gn.$$

Let $\{\varphi_{i,g} : i \geq 1\}$ be the normalized eigenvectors of G^*G corresponding to eigenvalues $\{\lambda_{i,g} : i \geq 1\}$. Define $\{\psi_{i,g} = \lambda_{i,g}^{-1/2} G\varphi_{i,g} : i \geq 1, \lambda_{i,g} > 0\}$. Let $S_g = \{i \geq 1 : \lambda_{i,g} > 0\}$. As before, it follows that $\{\varphi_{i,g} : i \geq 1\}$ is a complete orthonormal basis for X and $\{\psi_{i,g} : i \in S_g\}$ is an orthonormal set, complete in the range of G. Let $P_{O,g}$ be the projection onto the range of G:

$$P_{O,g} z = \sum_{i \in S} (z, \psi_{i,g}) \psi_{i,g}.$$

Then, $P_{O,g} G = G$, and

$$P_{O,g} y = P_{O,g} Hx + Gn, \quad y - P_{O,g} y = (I - P_{O,g}) Hx.$$

If $(I - P_{O,g})H \neq 0$, then there exists a noiseless channel from input to the output, and the communication problem in this case is trivial. We therefore assume that $(I - P_{O,g})H = 0$, that is, the range of H is included in the range of G.

Let $K : X \to X$ be given by

$$Kz = \sum_{i \in S_g} \lambda_{i,g}^{-1/2} (z, \psi_{i,g}) \varphi_{i,g}.$$

Note that $KGw = \sum_{i \in S_g} (w, \varphi_{i,g}) \varphi_{i,g}$ and thus $KG = P_{I,g}$, the orthogonal projection onto the orthonormal complement of the null space of of G. Note that K is one-to-one on the range of G (which includes the range of H) and thus knowing KHx one can recover Hx and knowing KGn one may recover Gn. Thus multiplying y by K is an information-lossless operation, and the channel $y = Hx + Gn$ can be replaced by an equivalent channel:

$$z = Ky = KHx + KGn = (KH)x + P_{I,g}n.$$

Now let $\{\varphi_{i,0} : i \geq 1\}$ be the normalized eigenvectors of $(KH)^*KH$, $\{\lambda_{i,0} : i \geq 1\}$ the corresponding eigenvalues, $S_0 = \{i \geq 1 : \lambda_{i,0} > 0\}$, and $\{\psi_{i,0} = \lambda_{i,0}^{-1/2} KH\varphi_{i,0} : i \in S_0\}$. Since the range of G includes the range of H, the range of $KG = P_{I,g}$ includes the range of KH and thus $P_{I,g}\psi_{i,0} = \psi_{i,0}$ for $i \in S_0$, implying:

$$(P_{I,g}n, \psi_{i,0}) = (n, \psi_{i,0}), \quad i \in S_0,$$

and we get an equivalent representation of the channel as

$$z_i = \lambda_{i,0}^{1/2} x_i + n_i, \quad i \in S_0,$$

where $z_i = (z, \psi_{i,0}) = (Ky, \psi_{i,0})$, $x_i = (x, \varphi_{i,0})$ and $n_i = (n, \psi_{i,0})$. As before, the random variables $\{n_i : i \in S_0\}$ form an independent identically distributed zero mean, GAUSSian sequence, and each n_i has variance $N_0/2$.

3.2 Random Coding for the Multi-Access Additive GAUSSian Noise Channel with Decision Feedback

We previously analyzed a processor-sharing queueing system. Let us recall the assumptions and the results: apart from the usual assumption of POISSON arrival of customers, we assumed that each customer had a demand which could be characterized by a single real number; moreover, as the customer remained in the system it accumulated service over time with a rate that was a function of the total number of customers currently in the system. When the accumulated service exceeded the demand, the customer departed the system. Under these conditions we showed that the departure process was also POISSON and we found the steady-state distribution of the number of customers in service.

3.2.1 Problem Setting

Here we look at a multi-access system for an Additive White GAUSSian Noise (AWGN) channel with decision feedback from the receiver to the transmitters. The receiver will decode each transmitter regarding others as noise; we may think that there are as many receivers as there are transmitters, each receiver decoding a single transmitter and regarding the other transmitters' signals as interference. In our analysis we will choose the codewords of each transmitter as samples of bandlimited GAUSSian noise. Each receiver will know the codewords of only one transmitter, the signals transmitted by the other transmitters are indistinguishable from those emitted from a GAUSSian noise source. However, we will assume that each receiver is cognizant of the total number of active transmitters at a given time. (This is a sensible assumption: we may imagine that there is a separate channel on which the transmitters announce their start of transmission so that a decoder will be assigned to them.) To be able to use our previous result on processor sharing queues to analyze this multi-access system we must identify the demand of each transmitter and the service rate offered by the receiver to the transmitter. Intuitive candidates for these quantities are the following: the number of bits (i.e., the amount of information) in the transmitters message could constitute the demand, and the average mutual information over the channel could be the service. The intuition behind this hunch is that the number of bits of the transmitters message resolved per unit time should be related to the mutual information over the channel, and thus rate of information flow should constitute the service. The preceding, however, is simply an intuition and needs verification. It will turn out that these intuitive candidates are too simplistic and we will define the demand and service differently in the following section. Nonetheless, we will see that the intuitive candidates can be interpreted as the limiting case of the analysis that follows.

Before proceeding any further let us examine the channel model in more detail. Our transmission model will be that described in Section 3.1.1: the transmitter signals will be constrained to a frequency of width W centered around f_0. The received signal will be the sum of the transmitted signals with white GAUSSian noise. The model is shown in Figure 3-2. In the figure, $x_k(t)$ is the signal of transmitter k and $w(t)$ is the white GAUSSian process, and is independent of the transmitted signals. The power spectrum of $w(t)$ is flat with intensity $N_0/2$:

$$S_{ww}(f) = \tfrac{1}{2}N_0.$$

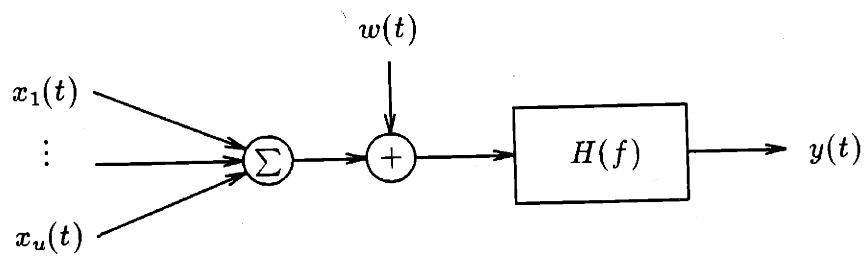


Figure 3-2: Channel Model

In the figure $H(f)$ is an ideal band-pass filter and w is a white GAUSSian process of intensity $N_0/2$.

The filter H is an ideal band pass filter:

$$H(f) = \begin{cases} 1 & \text{if } \left||f| - f_0\right| < W/2 \\ 0 & \text{else.} \end{cases}$$

From the previous section, we see that our continuous time model is equivalent to a discrete time model shown in Figure 3-1 and the relationship of the received signal to the transmitted signals is captured in the equations

$$y_i^{(I)} = \sum_k x_{ki}^{(I)} + n_i^{(I)} \quad \text{and} \quad y_i^{(Q)} = \sum_k x_{ki}^{(Q)} + n_i^{(Q)},$$

where $\{n_i^{(I)} : i \in \mathbb{Z}\}$ and $\{n_i^{(Q)} : i \in \mathbb{Z}\}$ form two independent sequences of independent identically distributed GAUSSian random variables, each with mean zero and variance $N_0/2$. Thus instead of a waveform channel we may think that we have a sequence of *scalar* channels, $\{(C_i^{(I)}, C_i^{(Q)}) : i \in \mathbb{Z}\}$.

The codebooks of the transmitters consist of a set of bandlimited waveforms, so that each waveform x is determined by its samples $\{(x_i^{(I)}, x_i^{(Q)}) : i \in \mathbb{Z}\}$. Let us now use a random coding argument to evaluate the performance of our system. The codebooks of all the transmitters will be chosen according to a probability measure that makes the samples independent and GAUSSian with variance $P/(2W)$ and mean 0. (That is, the codewords will be samples of bandlimited GAUSSian noise of power P.)

One should note that since we are choosing the codewords to be bandlimited, they occupy an infinite duration in time and are also non-causal. (One can see this by examining Figure 3-1 and noticing that the filters are non-causal.) If we admit non-causal filters in our model, then there are no difficulties. However, if one were to implement such a system, then one would use causal filters with finite response times, in which case our model will be only approximate.

Let us focus on a single transmitter–receiver pair, and condition on the process $u(t)$, $t \geq 0$, the number of active transmitter–receiver pairs at time t. Our goal will be to show that a modification of the information based service measure will qualify for the analysis.

The samples of the process $u(t)$ will be integer valued step functions. Define $t_0 \stackrel{\text{def}}{=} 0$ and let t_i, $i \geq 1$ be the times the process changes value and u_i, $i \geq 1$ be the value of the process in the interval $(t_{i-1}, t_i]$.

The noise power for the scalar channels $C_i^{(I)}$ and $C_i^{(Q)}$ is then:

$$\sigma_i^2 = N_0/2 + (u_k - 1)P/(2W), \quad Wt_{k-1} \leq i < Wt_k.$$

This expression indicates that the noise seen by a particular transmitter–receiver pair when there are $u - 1$ other users in the system is GAUSSian with power $N_0/2 + (u - 1)P/(2W)$.

Note that the scalar channels are made available over time at a rate of $2W$ per unit time (during Δt time units we gain $2W\Delta t$ degrees of freedom). Let the number of codewords be M (i.e., the message is $\log_2 M$ bits long). If we use the output of the first d channels to decode the transmitted message, we get the following random coding bound on the error probability [Gal68, pp. 149–150]: for any $0 \leq \rho \leq 1$

$$\bar{P}_e \leq \exp\left[\rho \ln M - \sum_{i=1}^{d} E_0(\rho, \sigma_i)\right].$$

Now, for a GAUSSian channel with independent GAUSSian input ensemble with variance $P/(2W)$ and with noise variance σ^2

$$E_0(\rho, \sigma) = \frac{\rho}{2} \ln\left(1 + \frac{P}{2W\sigma^2(1+\rho)}\right).$$

If we fix a $\rho \in (0, 1]$ and a tolerable error probability P_e, then we can view $-\ln P_e + \rho \ln M$ as the demand, and $E_0(\rho, \sigma)$ as the service (per transmitter–receiver pair per degree of freedom). Note that to cast these parameters in the context of our queueing result, we need to express the service in terms of total service (service to all customers, not just one) per unit time and not in terms of per customer per degree of freedom. This leads to a service rate:

$$2Wu(t)E_0(\rho,\sigma(t)) = u(t)W\rho\ln\left(1+\frac{P}{(1+\rho)(N_0W+(u(t)-1)P)}\right).$$

The factor $2W$ accounts for the degrees of freedom per unit time; to obtain the total service provided by the system we need the additional factor of $u(t)$. Note that scaling the demand and service by the same factor does not change the system; we will do so by factoring out $W\rho$ from both, thus defining the demand as

$$S = W^{-1}[-(\ln P_e)/\rho + \ln M]$$

and the service function (rate of service as a function of total number of customers) as

$$\phi(u) = u\ln\left(1+\frac{P}{(1+\rho)(N_0W+(u-1)P)}\right).$$

In this second form, we see that the hand-wavy analysis we had done after the processor-sharing results were not so far off the mark: there, the demand did not include the $-(\ln P_e)/\rho$ term, and the service function was

$$u\ln\left(1+\frac{P}{N_0W+(u-1)P}\right).$$

In spite of the apparent similarity, differences surface when one tries to find the conditions for the stability of the system. In the model of the previous chapter we had concluded that the system is stable whenever the aggregate arrival rate per unit time per unit bandwidth was less than 1 nat ($\log_2 e$ bits). In this model we have a more realistic result. As we insist on lower error probabilities, the aggregate rate that our system can support decreases. The following analysis makes this relationship clear.

For stability we need:

$$\lim_{u\to\infty}\frac{\lambda E[S]}{\phi(u)} < 1,$$

where λ is the arrival rate. Let us define

$$\ell \stackrel{\text{def}}{=} \lambda E[\ln M]/W$$

as the loading of the queue and

$$E_e \stackrel{\text{def}}{=} -(\ln P_e)/E[\ln M]$$

as the error exponent. The quantity $E[\ln M]$ appearing in the definition of ℓ is the expected length of the messages (in natural units); ℓ is the total nat (as opposed to bit) arrival rate per unit bandwidth (in the units of nats per unit time per unit bandwidth). The error exponent E_e relates the error probability to expected message length. Since $\phi(u) \to 1/(1+\rho)$ as u tends to infinity and

$$E[S] = W^{-1}E[\ln M - (\ln P_e)/\rho] = W^{-1}E[\ln M](1 + E_e/\rho),$$

we get

$$\lim_{u\to\infty} \frac{\lambda E[S]}{\phi(u)} = \ell(1 + E_e/\rho)(1+\rho),$$

and we can rewrite the above stability condition as

$$\ell(1+\rho)(1+E_e/\rho) < 1 \quad \text{for some } \rho \in [0,1]. \tag{3.1}$$

This is equivalent to the statement

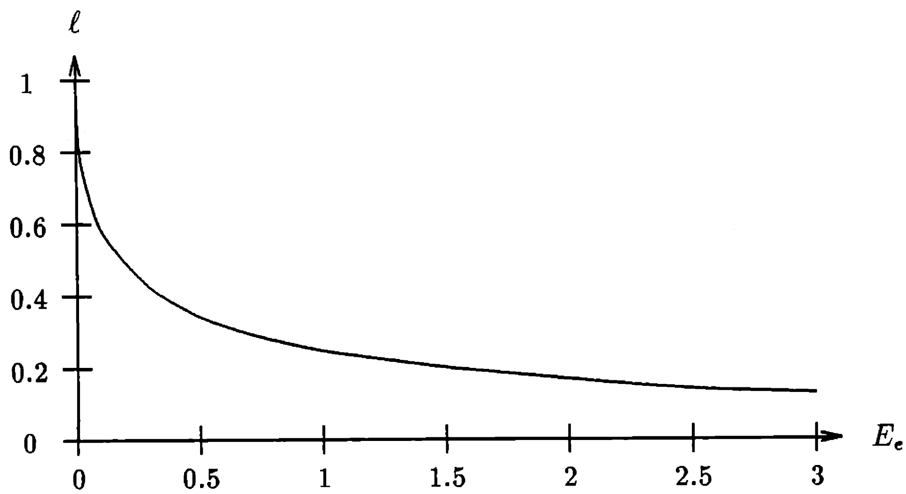
$$\ell \min_{0\le\rho\le1} (1+\rho)(1+E_e/\rho) < 1.$$

This minimization occurs at $\rho = 1$ for $E_e \geq 1$ and at $\rho = \sqrt{E_e}$ for $E_e < 1$. The corresponding value of the minimum is $2(1+E_e)$ and $(1+\sqrt{E_e})^2$ respectively. In sum, the stability region of the processor sharing queue is

$$\{(E_e,\ell) : 0 \le E_e \le 1, 0 \le \ell < \left(1+\sqrt{E_e}\right)^{-2}\} \cup \{(E_e,\ell) : E_e > 1, 0 \le \ell < (2+2E_e)^{-1}\}.$$

Figure 3-3 shows this stability region. The region of interest will depend on the message length. If one has long messages, then one is satisfied with relatively small error exponents, whereas for small message lengths one will be interested in large exponents to achieve small error probability. Note that E_e can be made arbitrarily large if one is willing to sacrifice throughput.

Given a stable (E_e, ℓ) pair and a signal to noise ratio SNR $\stackrel{\text{def}}{=} P/(N_0W)$, our processor



The stability region of the multi-access system is shown. The vertical axis ℓ is the loading of the system, the nat (as opposed to bit) arrival rate per bandwidth. The horizontal axis is the error exponent, the natural logarithm of the error probability divided by the packet length (in nats).

Figure 3-3: Stability region of the multi-access system

sharing results enable us to find the average number $\bar{U}$ of customers in the system:

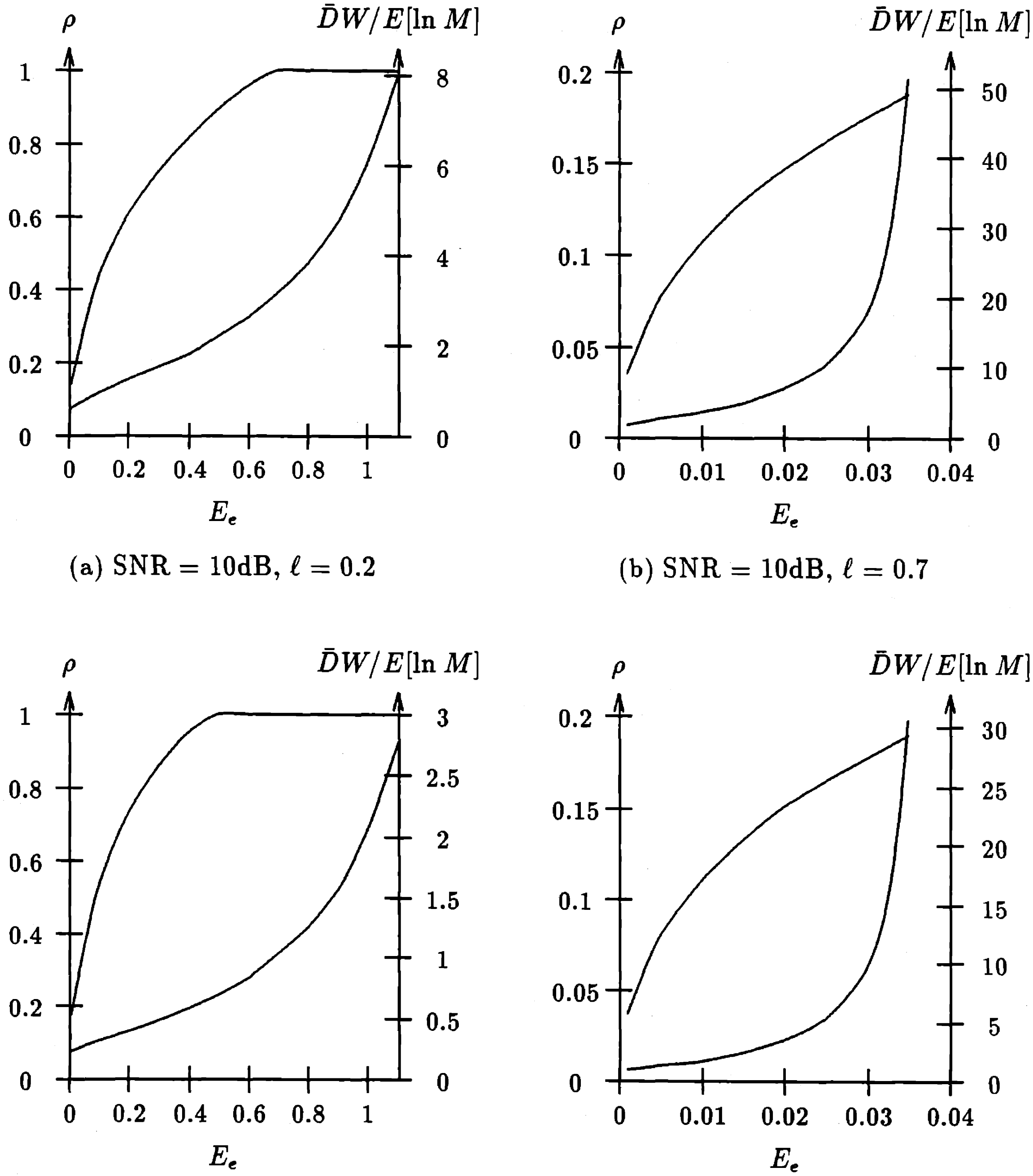
$$\bar{U} = \sum_{u \geq 1} uE[S]^u/\phi_!(u) \Big/ \sum_{u \geq 0} E[S]^u/\phi_!(u)$$

with $\phi_!(u) = \prod_{i=1}^{u} \phi(i)$. Using LITTLE's law we obtain an expression for average delay:

$$\lambda \bar{D} = \bar{U}, \quad W\bar{D}/E[\ln M] = \bar{U}/\ell.$$

Another issue is the choice of ρ. From the discussion above, for any stable (E_e, ℓ) pair, we know that there is a $\rho \in [0,1]$ which makes the system stable. For a stable pair (E_e, ℓ), let Σ be the set of ρ's in $[0,1]$ that satisfy (3.1). Choosing any $\rho \in \Sigma$ will lead to a stable system, but, one should presumably choose ρ to minimize the expected delay. Figure 3-4 shows the optimal ρ and the corresponding delay for various values of ℓ, E_e, and SNR.

Our analysis, although theoretically sound, can be criticized on the following ground: the decoding rule we employ chooses the time of decoding only on the basis of the number of interferers, and ignores what is actually received. The optimal decoder (the one that



(a) SNR = 10dB, $\ell = 0.2$

(b) SNR = 10dB, $\ell = 0.7$

(c) SNR = 30dB, $\ell = 0.2$

(d) SNR = 30dB, $\ell = 0.7$

We see the dependency of expected delay and the optimal ρ as they vary with the error exponent E_e, SNR and ℓ. Each curve in the graph is to be read off from the axis closer to it. Note that the expected delay is normalized with the bandwidth W and the average message length.

Figure 3-4: Expected Delay $\bar{D}$ and optimal ρ vs. error exponent E_e

minimizes the expected delay subject to a probability of error constraint) will make use of all the information available in choosing the decoding time. Let $y_1, \ldots, y_d$ be the outputs of the first d channels and let

$$T(d) = \max_m \frac{\Pr[y_1, \ldots, y_d | m \text{ is transmitted}]}{\sum_{m' \neq m} \Pr[y_1, \ldots, y_d | m' \text{ is transmitted}]}.$$

The optimal rule will decode when $T(d)$ exceeds a pre-determined threshold T, and decode the message m that achieves the maximum in the expression for $T(d)$. This decoding rule does not (as far as I see) lend itself to an additive service characterization, so I don't see how we can analyze this decoding rule. A modified rule would replace the denominator in the $T(d)$ expression by the ensemble probability of the sequence $y_1, \ldots, y_d$. Then, since the channel is memoryless the service function becomes additive, but it also becomes non-deterministic in the sense that knowing the number of interferers does not determine the service rate; it only determines the probability distribution on the rate of service.

3.2.2 An improved decoding rule

We had noted above that the decoding rule used in our analysis determined the decoding time irrespective of the received vector. We will now introduce a slight modification to this decoding rule. In this modified rule, the decoder proceeds in stages. The duration of each stage is determined by the number of interferers during that stage and at the end of each stage the decoder decides whether to decode or to proceed to the next stage. This decision to decode or to proceed is made on the basis of the received signal during the current stage. Note that the transmitter does not need to know about the stages, it just keeps transmitting the signal corresponding to the message. Let P_X be the probability that at the end of a stage the decoder will not decode but proceed with the next stage. Then the number of stages the decoder will take to decode the message is a geometrically distributed random variable with mean $(1 - P_X)^{-1}$. If the expected service requirement of the individual trials is $E[S]$, then the overall service will have expected value $E[S]/(1 - P_X)$. To analyze this decoding rule we will derive a parallel channels result, analogous to the parallel channels result [Gal68, pp. 149–150] on the error exponent. Let us introduce the following quantity:

$$E_0(\rho, s, Q, P) = -\ln \sum_{\mathbf{y}} \left[\sum_{\mathbf{x}} Q(\mathbf{x}) P(\mathbf{y}|\mathbf{x})^{1-s} \right] \left[\sum_{\mathbf{x}'} Q(\mathbf{x}') P(\mathbf{y}|\mathbf{x}')^{s/\rho} \right]^{\rho}.$$

We then have the following [For68]: for each ρ, s satisfying $0 \le s \le \rho \le 1$, and for each $T \ge 1$ there exists a decoding rule with

$$\bar{P}_X \le M^\rho T^s \exp -E_0(\rho, s, Q, P) \quad \text{and} \quad \bar{P}_e \le M^\rho T^{(s-1)} \exp -E_0(\rho, s, Q, P),$$

where the expectation is taken over the ensemble of codes whose codewords are chosen independently according to the distribution Q. Note that the upper bounds on the error and erasure probabilities differ by a factor of T. It is easy to see that if $Q(\mathbf{x}) = Q_1(x_1) \cdots Q_n(x_n)$ and $P(\mathbf{y}|\mathbf{x}) = P_1(y_1|x_1) \cdots P_n(y_n|x_n)$ then,

$$E_0(\rho, s, Q, P) = \sum_{i=1}^{n} E_0(\rho, s, Q_i, P_i)$$

with

$$E_0(\rho, s, Q_i, P_i) = -\ln \sum_y \Big[\sum_x Q_i(x) P_i(y|x)^{1-s}\Big] \Big[\sum_{x'} Q_i(x') P_i(y|x')^{s/\rho}\Big]^\rho .$$

If we are given the values of tolerable error and erasure probabilities P_e and P_X, we may identify their ratio P_X/P_e as T, $-\ln P_X + s \ln T + \rho \ln M$ as the demand and $E_0(\rho, s, Q_i, P_i)$ as service (per degree of freedom). We may now use our queueing results since the service is additive. We may eliminate the parameter T from our definition of demand and write it only in terms of the error and erasure probabilities:

$$S = -(1-s) \ln P_X - s \ln P_e + \rho \ln M.$$

The service thus defined, does not take into account the fact that decoding may consist of several stages. As we have said above, the average total service will exceed $E[S]$ by a factor of $(1 - P_X)^{-1}$. For a given value of P_X, one may still use the curves that we will compute by increasing the loading ℓ by a factor of $(1 - P_X)^{-1}$ and reading the curves at the new loading.

As an aside, note that with $s = \rho/(1+\rho)$ the expression for $E_0(\rho, s, Q, P)$ reduces to the expression for the E_0 of the previous section. With this value of s, the service requirement becomes $S = -\ln P_X^{1/(1+\rho)} P_e^{\rho/(1+\rho)} + \rho \ln M$. In the previous section the service requirement was $-\ln P_e + \rho \ln M$.

Also note that, for the additive GAUSSian noise channel with noise variance σ^2 and for an input distribution that is GAUSSian with variance P,

$$E_0(\rho, s, P/\sigma^2) = \frac{\rho}{2}\ln\left[1 + (s/\rho)\frac{P}{\sigma^2}\right] + \tfrac{1}{2}\ln\left[1 + \frac{(\rho - s - s\rho)(s/\rho)(P/\sigma^2)}{1 + (s/\rho)(P/\sigma^2)}\right]$$

As before, the work function $\phi(u)$ will be given by $2WuE_0$ evaluated at the SNR corresponding to $u-1$ interferers:

$$\begin{aligned}\phi(u) &= \rho u W \ln\left[1 + (s/\rho)\frac{P}{N_0 W + (u-1)P}\right] \\ &\quad + uW\ln\left[1 + \frac{(\rho - s - s\rho)(s/\rho)(P/(N_0W + (u-1)P))}{1 + (s/\rho)(P/(N_0W + (u-1)P))}\right]\end{aligned}$$

The limit of $\phi(u)$ as u approaches infinity is

$$\lim_{u\to\infty}\phi(u) = Ws\left[2 - \frac{1+\rho}{\rho}s\right].$$

If we define $\ell = W^{-1}\lambda E[\ln M]$ as the nat arrival rate per unit bandwidth, $E_e = \dfrac{-\ln P_e}{E[\ln M]}$ as the error exponent and $E_X = \dfrac{-\ln P_X}{E[\ln M]}$ as the erasure exponent we see that

$$E[S] = W\ell[\rho + sE_e + (1-s)E_x],$$

and the stability condition

$$\lim_{u\to\infty}\frac{\lambda E[S]}{\phi(u)} < 1$$

reduces to

$$\ell\frac{\rho + sE_e + (1-s)E_X}{s(2 - s(1+\rho)/\rho)} < 1 \quad \text{for some } 0 \le s \le \rho \le 1.$$

This, in turn, is equivalent to

$$\ell \min_{0\le\rho\le 1}\min_{0\le s\le\rho}\frac{\rho + sE_e + (1-s)E_X}{s(2 - s(1+\rho)/\rho)} < 1.$$

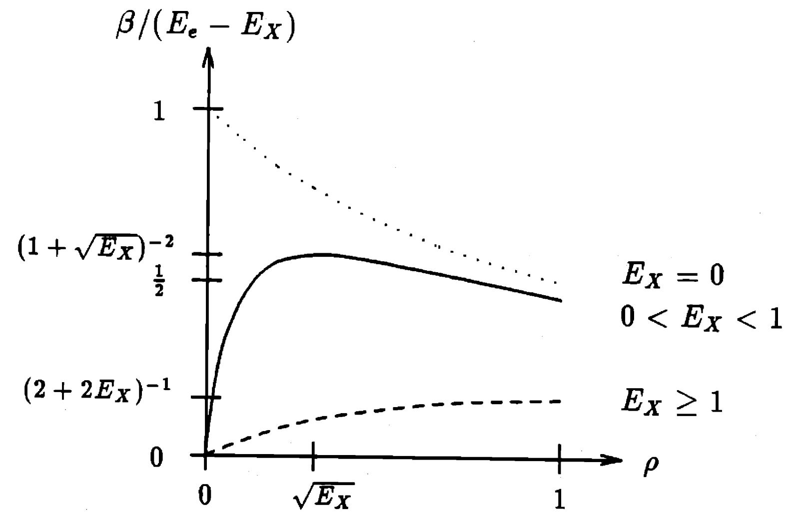
The minimization over s can be done via differentiation and we obtain

$$s = \frac{\rho + E_X}{E_e - E_X}\left(\sqrt{1 + 2\beta} - 1\right)$$

where $\beta = \beta(\rho) \stackrel{\text{def}}{=} \frac{\rho}{1+\rho}\frac{E_e - E_X}{\rho + E_X}$. Using the inequality $\sqrt{1+x} \leq 1 + x/2$, we see that $s \leq \rho/(1+\rho)$ and thus $s \leq \rho$ as desired. Substituting this value of s we get the following condition for stability:

$$\ell \min_{0\leq\rho\leq 1} \tfrac{1}{2}(E_e - E_X)\frac{\sqrt{1+2\beta}}{(\sqrt{1+2\beta}-1)(1+\beta^{-1})-1} < 1 \tag{3.2}$$

where β is defined as above. The quantity to be minimized is a decreasing function of β, and thus the minimizing ρ in equation (3.2) is the one that maximizes $\beta(\rho)$. The



The nature of the mapping $\rho \mapsto \beta(\rho)$ depends crucially on the value of E_X. All the possible cases are shown in the figure above. The maxima and minima of the curves are also indicated.

Figure 3-5: β as a function of ρ for various cases of E_X.

nature of the mapping $\rho \mapsto \beta(\rho)$ depends on the value of E_X. There are three cases as illustrated in Figure 3-5:

1. $E_X = 0$. The range of the mapping $\rho \mapsto \beta$ is $[\frac{1}{2}E_e, E_e]$. The maximum is achieved at $\rho = 0$.

2. $0 < E_X < 1$. The range is $[0, (E_e - E_X)/(1+\sqrt{E_X})^2]$. The maximum is achieved at $\rho = \sqrt{E_X}$.

3. $E_X \geq 1$. The range is $[0, (E_e - E_X)/(2 + 2E_X)]$. The maximum is achieved at $\rho = 1$.

Putting everything together, we have the following stability conditions:

1. If $0 \leq E_X < 1$

$$\tfrac{1}{4}\ell \left(1 + \sqrt{E_X} + \sqrt{\left(1 + \sqrt{E_X}\right)^2 + 2(E_e - E_X)}\right)^2 < 1$$

2. if $E_X \geq 1$

$$\tfrac{1}{2}\ell \left(\sqrt{1 + E_e} + \sqrt{1 + E_X}\right)^2 < 1,$$

Note that at $E_X = 1$, the conditions give identical expressions. Also if $E_X = E_e = E$, we recover the stability conditions of the previous section.

If we are interested in maximizing E_e for a given ℓ irrespective of E_X, we see that this maximum occurs when $E_X = 0$. The condition on ℓ and E_e is then the following:

$$\tfrac{1}{2}\ell \left(1 + E_e + \sqrt{1 + 2E_e}\right) < 1.$$

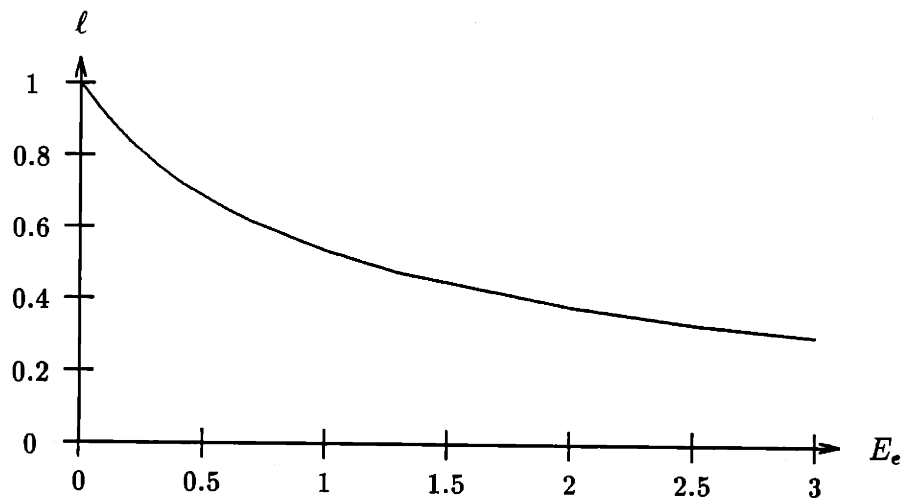
This region is shown in Figure 3-6. For ℓ close to 1, E_e needs to be small, and we can approximate this condition to $E_e < (1 - \ell)$. Thus for large loading ℓ, the system can support error exponents up to $1 - \ell$. Compare that with the decoding rule of the previous section: the largest error exponent the previous system could support for large ℓ is approximately $(1 - \ell)^2/4$.

As in the previous section, we can compute the average delay for any given E_e, E_X, SNR and ℓ. In Figure 3-7 we plot the normalized average delay as a function of E_e for various values of the other parameters. Each point on the curve shows the result of an optimization over ρ and s. A comparison with Figure 3-4 shows that there is a significant improvement in the supportable E_e and the corresponding delay.

We can easily make the following observations:

1. For fixed bandwidth and bit arrival rate, the average delay increases linearly and error probability decreases exponentially with increasing packet length.

2. Increasing the SNR has a more pronounced effect on delay at lower loads than at higher loads. At higher loads, the effective SNR is governed by the interference rather than the intrinsic noise.



Note the enlargement of the region as compared to Figure 3-3. The difference is especially pronounced at ℓ close to 1.

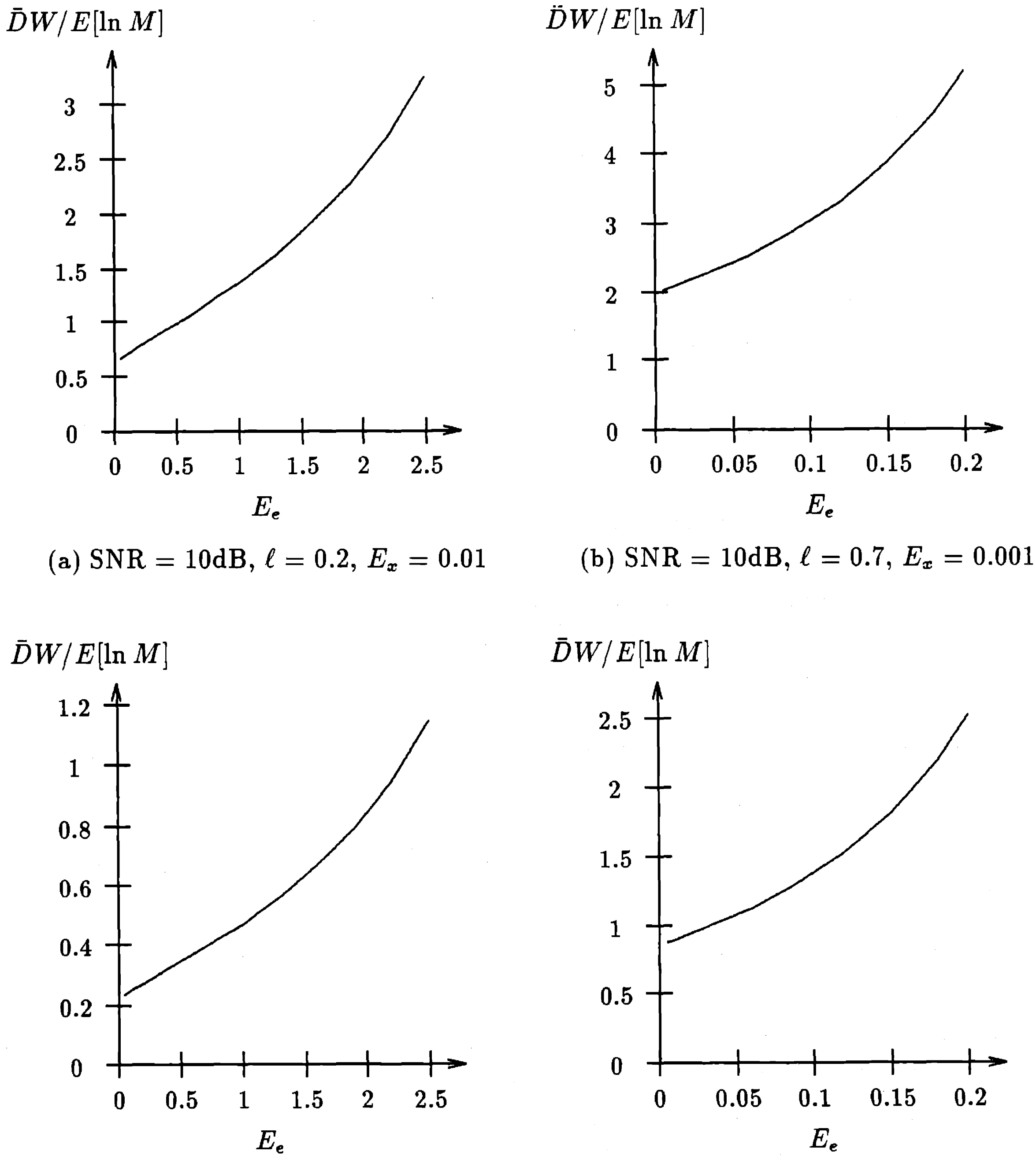
Figure 3-6: The stability region of the multi-access system as $E_X \to 0$.

3. Increasing the bandwidth has a two-fold decreasing effect on delay: (i) it decreases the loading, thereby reducing the normalized delay $W\bar{D}$, and thus (ii) it decreases the actual delay by an extra linear factor.

We can also make the following comments on our system model, and the analysis:

1. The model can be generalized to other waveform channels that have a time-frequency decomposition property. The Rayleigh Fading Channel is an example that comes to mind.

2. The decoding rule the receiver employs fixes the values of the parameters ρ and s for all time. A better decoding rule would optimize the values of ρ and s as more degrees of freedom are observed: when the outputs of the first d channels are observed, the decoder would find the ρ and s that minimize the bound on the error probability and would decode when this upper bound is small enough. However this decoding procedure cannot be analyzed with the machinery we have developed so far: the service is not additive any more. In any event, the analysis we have done with the suboptimal decoding rule provides a lower bound to the performance of this improved decoding rule.



(a) SNR = 10dB, $\ell = 0.2$, $E_x = 0.01$

(b) SNR = 10dB, $\ell = 0.7$, $E_x = 0.001$

(c) SNR = 30dB, $\ell = 0.2$, $E_x = 0.01$

(d) SNR = 30dB, $\ell = 0.7$, $E_x = 0.001$

Figure 3-7: Delay vs. Error Exponent

Chapter 4

Erasures and Errors Decoding for a Single User Channel

In most communication schemes the decoder produces an estimate of the transmitter's intent without indicating its confidence in the estimate. However, such a confidence measure may enhance the performance of systems that make use of the decoder's output. The simplest kind of confidence indication is achieved by allowing the decoder not to decode, i.e., declare an erasure. A communication scheme utilizing such a decoder is called an "errors-and-erasures" scheme. If there is a feedback link available, the decoder can then inform the transmitter of an erasure and request a retransmission. This type of feedback is termed "decision feedback" and is weaker than that in which the transmitter is aware of the actual channel outputs.

The problem of finding the performance of errors-and-erasures and decision feedback schemes has been studied by FORNEY in 1968 [For68]. In his studies, FORNEY describes the optimum decoding rule, and uses a (single-letter) random coding argument to find simultaneous exponential upper bounds to error and erasure probabilities. His results were believed to be exponentially tight at least in the high rate region. CSISZÁR and KÖRNER [CK81] rederived these bounds using a fixed composition argument.

A consequence of FORNEY's studies is the existence of a class of channels for which *zero* error and arbitrarily small erasure probabilities are simultaneously achievable for rates smaller than some "zero error decision feedback" capacity. We will use the term "zero undetected error capacity" to refer to the same quantity, and will denote it by C_{0u}[1].

[1]The reason for the rather awkward symbolism is that other logical choices were taken: C_0 denotes the zero error capacity and C_{0f} denotes the zero error capacity with channel feedback [Sha56]. In general $C_0 \leq C_{0u} \leq C_{0f} \leq C$.

We will write $C_{0u}(P)$ when we are referring to the zero undetected error capacity of a particular channel P. FORNEY also gives the lower bound[2]

$$C_{0u}(P) \geq \max_{Q} \sum_{j} w_j \ln(1/v_j), \tag{4.1}$$

where Q is an input probability distribution, w_j is the corresponding probability of output j and v_j is the collective probability of inputs that lead to output j with positive probability[3]. If one believes that the exponential bounds of FORNEY are tight, one is led to the conclusion that this lower bound is the actual value of C_{0u}.

However, PINSKER and SHEVERDYAEV in 1970 [PS70] proved the following: given a channel P construct the *channel graph*, a bipartite graph obtained by listing the inputs and outputs of the channel and connecting an input-output pair by an edge if the output can be reached from the input with positive probability. If this graph contains no cycles, then $C_{0u}(P)$ is equal to the (usual) capacity $C(P)$. This leads to examples for which (4.1) is not a tight lower bound.

The following example contains the essential ideas in PINSKER and SHEVERDYAEV's proof.

Example 4.1. C_{0u} of the Z-Channel

Consider the channel shown below. Let C denote the capacity of this channel, and let $\mathbf{x}_1, \ldots, \mathbf{x}_M$ be the codewords of a block code of length n, rate $R = (\ln M)/n$ and

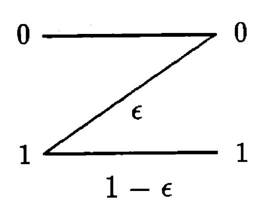


Figure 4-1: Z-Channel

maximum error probability λ. That is, $\Pr[\text{Error}|\mathbf{x}_m] \leq \lambda$ for all $m = 1, \ldots, M$. Classify the codewords according to their weight (the number of 1's they have). Since there are at most $n+1$ classes, the most crowded class has at least $M/(n+1)$ codewords.

[2]We will use natural units for rates and capacities throughout the paper.

[3]Both w_j and v_j are a function of Q as well as P, and should be notated $w_j(Q, P)$ and $v_j(Q, P)$ respectively. However, we will use the shorter form for brevity.

Eliminate all codewords except the members of this most crowded class. The new code has a maximum error probability at most λ (eliminating codewords cannot increase the probability of error) and rate at least

$$\frac{1}{n}\ln\frac{M}{n+1} = R - \frac{\ln(n+1)}{n}.$$

Thus we conclude that if n is large any block code can be converted into a fixed weight code with negligible loss of rate.

Let the common weight of the codewords thus obtained be a, and consider a received sequence $\mathbf{y}$ of weight b. Since the channel cannot change 0's into 1's, $b \leq a$. For a codeword $\mathbf{x}_m$ that leads to $\mathbf{y}$ with positive probability, b of its 1's are received as 1's and $a-b$ of its 1's are received as 0's. Thus

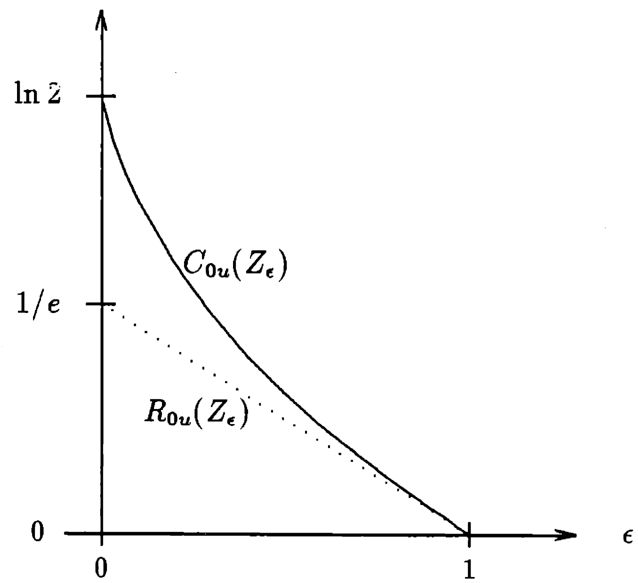
$$\Pr[\mathbf{y}|\mathbf{x}_m] = (1-\epsilon)^b \epsilon^{a-b}.$$

Since the right hand side of the equation is independent of m, we conclude that all m's that are at all possible are equally likely. Therefore, for any $\mathbf{y}$, either there is a single possible $\mathbf{x}_m$, in which case the decoder can be assured of the correctness of its choice, or the decoder has to choose between equally likely candidates. Since in the latter case the decoder will commit to an error with probability at least 1/2 we have

$$\frac{1}{2}\Pr[\text{there is more than one candidate}] \leq \lambda.$$

A decoder that erases whenever there is a chance of making an error will thus have erasure probability at most 2λ. For any R less than C, λ can be made arbitrarily small by increasing n, so we see that $C_{0u} = C$. For this example FORNEY's lower bound gives $(1-\epsilon)/e$ which is strictly smaller than $C = \ln(1 + (1-\epsilon)\epsilon^{\frac{\epsilon}{1-\epsilon}})$, which is illustrated in Figure 4-2.

The above example shows that lower bounding C_{0u} using a single-letter random coding argument does not yield tight answers. In this chapter, we will derive a tighter lower bound for C_{0u}.



The figure compares the lower bound given in (4.1) with the actual undetected error capacity C_{0u} for the Z-channel. The lower bound is denoted by R_{0u}.

Figure 4-2: FORNEY's Lower Bound

4.1 A Tighter Lower Bound to Zero Undetected Error Capacity

To determine our lower bound to C_{0u} we will use a constant composition approach. We will show that[4]

$$C_{0u}(P) \geq \max_{Q} \min_{P' \in \wp(Q,P)} I(Q, P'). \tag{4.2}$$

In this expression P is the transition probability matrix of the given channel, and Q is an input distribution. P_{jk} denotes the conditional probability of output j given input k and Q_k denotes the probability of input k. $\wp(Q, P)$ denotes the set of transition probability matrices P' such that P and P' impose the same output distribution when the input distribution is Q and P' introduces no extra connections to the channel:

$$\wp(Q, P) = \left\{P' \colon P_{jk} = 0 \Rightarrow P'_{jk} = 0, \textstyle\sum_k Q_k P'_{jk} = \sum_k Q_k P_{jk}\right\} \tag{4.3}$$

[4]We will prove that this is a better bound than (4.1) in Appendix 4.A on page 66.

$I(\cdot,\cdot)$ is the mutual information function,

$$I(Q,P') = \sum_{k,j} Q_k P'_{jk} \ln \frac{P'_{jk}}{\sum_i Q_i P'_{ji}}.$$

We will abbreviate the constraint $P_{jk} = 0 \Rightarrow P'_{jk} = 0$ to $P' \ll P$, borrowing this notation from measure theory [Hal74]. One can see that $\wp(Q,P)$ is a convex set and $I(Q,P')$ is convex in P', so that the minimization can be easily done. However, after the minimization over P', the resulting function of Q to be maximized is in general not concave (even though $I(Q,P)$ is), so the maximization may prove to be difficult. This lack of concavity will have important consequences that hinder our search for a converse. We will give an example of this non-concavity after we first prove the lower bound in (4.2).

4.2 The Proof of the Lower Bound

To prove (4.2) one proceeds as follows: Pick a constraint length n and an input distribution Q such that for each input letter k, nQ_k is an integer. An input sequence $\mathbf{x}$ of length n is said to be of composition Q if each input letter k occurs nQ_k times in $\mathbf{x}$. One can easily show that the size of the set of all sequences of composition Q is $\exp nH(Q)$ to the first order of n in the exponent. Here H is the entropy function:

$$H(Q) = -\sum_k Q_k \ln Q_k.$$

Select M codewords $\mathbf{x}_1, \ldots, \mathbf{x}_M$ randomly from this constant composition set according to a uniform probability distribution. The code thus obtained will be a constant composition code.

Consider transmitting the sequence $\mathbf{x}_m$ and receiving $\mathbf{y}$. Count the number of k to j transformations for this transmit-receive pair. Divide this count by nQ_k (the number of positions where $\mathbf{x}$ was k) to obtain the "observed channel behavior" $\widehat{P}_{jk}$. We will also say that $\mathbf{y}$ is "$\widehat{P}$ generated" from $\mathbf{x}_m$. The conditional probability of $\mathbf{y}$ given $\mathbf{x}_m$ is then

$$P[\mathbf{y}|\mathbf{x}_m] = \prod_{k,j} P_{jk}^{n\widehat{P}_{jk}Q_k} = \left(\prod_{k,j} \widehat{P}_{jk}^{n\widehat{P}_{jk}Q_k}\right) \exp -n \sum_{k,j} Q_k \widehat{P}_{jk} \ln \frac{\widehat{P}_{jk}}{P_{jk}}.$$

Notice that the term in parentheses in the above expression is the conditional probability of $\mathbf{y}$ given $\mathbf{x}$, for a channel with transition probabilities $\widehat{P}$ rather than P. If we define

$$D(\widehat{P}\|P|Q) \stackrel{\text{def}}{=} \sum_{k,j} Q_k \widehat{P}_{jk} \ln \frac{\widehat{P}_{jk}}{P_{jk}},$$

we can rewrite the above as

$$P[\mathbf{y}|\mathbf{x}_m] = \widehat{P}[\mathbf{y}|\mathbf{x}_m] \exp -nD(\widehat{P}\|P|Q).$$

If we sum this expression over all $\mathbf{y}$ that would result in channel behavior $\widehat{P}$ and overbound the right hand side by summing over all $\mathbf{y}$, we get the following bound:

$$\Pr[\text{channel behavior } \widehat{P}] \leq \exp -nD(\widehat{P}\|P|Q).$$

Now consider the event that there is some $\mathbf{x}_{m'}$, $m' \neq m$, for which $P[\mathbf{y}|\mathbf{x}_{m'}] > 0$. Let P' be the observed behavior for the pair $(\mathbf{x}_{m'}, \mathbf{y})$. To satisfy $P[\mathbf{y}|\mathbf{x}_{m'}] > 0$, P' should not contain connections that are not present in P. Also, the fraction of j's in $\mathbf{y}$, w_j, is equal to both $\sum_k Q_k \widehat{P}_{jk}$ and $\sum_k Q_k P'_{jk}$. Thus P' imposes the same output distribution $\widehat{P}$ does under the input distribution Q. Let A denote the set of $\mathbf{x}$'s of composition Q for which the observed behavior of the pair $(\mathbf{x}, \mathbf{y})$ is P'. Out of n positions in $\mathbf{y}$, there are nw_j positions occupied by the letter j. In the corresponding nw_j positions of $\mathbf{x}$ the letter k must occur exactly $nQ_kP'_{jk}$ times. Thus the size of A is given by

$$|A| = \prod_j \frac{(nw_j)!}{\prod_k (nQ_kP'_{jk})!}.$$

From this we can easily show that the fraction of $\mathbf{x}$'s of composition Q for which $(\mathbf{x}, \mathbf{y})$ has observed behavior P' is $\exp -nI(Q, P')$ (again to the first order of n in the exponent).

Combining the above with the union bound,

$$P[\text{Erasure}, P'|\widehat{P}] \leq \min\Big\{1, (M-1)e^{-nI(Q,P')}\Big\} \leq \exp -n(I(Q, P') - R - \epsilon_n)^+,$$

where a^+ denotes the positive part of a, and $\lim_{n\to\infty} \epsilon_n = 0$. Summing over P' and removing the conditioning, we get

$$P[\text{Erasure}] \leq \sum_{\widehat{P}, P'} \exp\Big\{-n\Big[D(\widehat{P}\|P|Q) + \Big(I(Q, P') - R - \epsilon_n\Big)^+\Big]\Big\},$$

where P' is subject to the constraints described above. One can show that there are polynomially many (in n) $\widehat{P}$'s and P''s, so that we can bound the summation by the maximal term with a vanishingly small loss in the exponent:

$$P[\text{Erasure}] \leq \exp\Big\{-n\Big(\min_{\widehat{P},P'}\Big[D(\widehat{P}\|P|Q) + \big(I(Q,P') - R\big)^+\Big] - \epsilon'_n\Big)\Big\}$$

where $\lim_{n\to\infty} \epsilon'_n \to 0$. One now notices that as long as R is less than the right hand side of (4.2) either $D(\widehat{P}\|P|Q)$ or $(I(Q,P') - R)^+$ is positive. (For if $\widehat{P} \neq P$, $D(\widehat{P}\|P|Q) > 0$ else $\widehat{P} = P$ and the constraint on P' is $P' \in \wp(Q,P)$). Since the minimization is done over a compact set and the function to be minimized is continuous we see that the minimum is positive for R less than the right hand side of (4.2).

Note that the bound derived above applies to the ensemble average erasure probability of a class of codes. However, the standard techniques of random coding can easily convert such a result to a bound on the maximal erasure probability of the best code in this class.

It would have been nice to show that the expression in equation (4.2) is in fact C_{0u} rather than being a lower bound to it. The conventional converse proofs for results proven by constant composition code arguments go through the following steps: First find an upper bound to the achievable rate of codes whose codewords have composition Q. Suppose now that this upper bound coincides with the achievability result for such codes. Since any code has a significant constant composition sub-code the converse will follow. Note that this type of approach proves something even stronger: it also finds the performance of constant composition codes. Our achievability result shows that with a code of composition Q one can achieve rates as close to $\min_{P'\in\wp(Q,P)} I(Q,P')$ as desired. For the conventional converse to work we need to show that one cannot do any better with codes of composition Q. This unfortunately is not true as the example below will show. The example demonstrates that $\min_{P'\in\wp(Q,P)} I(Q,P')$ is not a concave function of Q. However, if $I_a(Q)$ denotes the achievable rate with codes of constant composition Q, then the following holds:

Lemma 4.2 *I_a is concave function.*

Proof. By using time sharing between codes of composition Q and Q' one obtain a code of composition $\mu Q + (1-\mu)Q'$ of rate $\mu I_a(Q) + (1-\mu)I_a(Q')$; thus $I_a(\mu Q + (1-\mu)Q') \geq \mu I_a(Q) + (1-\mu)I_a(Q')$. □

We already know that $I_a(Q) \geq \min_{P'\in\wp(Q,P)} I(Q,P')$; we can now improve this lower bound to $I_a(Q) \geq I^{**}(Q)$, where $I^{**}(Q)$ is the smallest concave function not less than

$\min_{P' \in \wp(Q,P)} I(Q,P')$. In Appendix 4.C on page 69 we show a construction to obtain I^{**}. However, as noted in Appendix 4.C the replacement of $\min_{P' \in \wp(Q,P)} I(Q,P')$ by $I^{**}(Q)$ will not change the value of the maxima over Q.

Example 4.3. Non-concavity of $\min_{P' \in \wp(Q,P)} I(Q,P')$ as a function of Q

Consider the following channel (see Figure 4-3 on page 48) with three inputs (0, 1, and 2), three outputs (0, 1, and 2), and transition probabilities

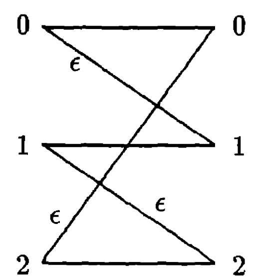


Figure 4-3: Channel for non-concavity example

$$P(j|k) = \begin{cases} 1-\epsilon & i = j \\ \epsilon & j - i = 1 \bmod 3 \\ 0 & \text{else.} \end{cases}$$

For this example only, let us define

$$I_0(Q) = \min_{P' \in \wp(Q,P)} I(Q,P').$$

Consider evaluating I_0 at $Q_0 = Q_1 = Q_2 = 1/3$. The output probabilities for P and hence for $P' \in \wp(Q,P)$ are $W_1 = W_2 = W_3 = 1/3$. The class $\wp(Q,P)$ then consists of all channels P' that are of the same type as P:

$$P'(j|k) = \begin{cases} 1-\epsilon' & i = j \\ \epsilon' & j - i = 1 \bmod 3 \\ 0 & \text{else.} \end{cases}$$

The P' that minimizes $I(Q,P')$ is the one with $\epsilon' = \frac{1}{2}$. The value of $I(Q,P')$ is $\ln(3/2)$, and thus $I_0(\frac{1}{3},\frac{1}{3},\frac{1}{3}) = \ln(3/2)$. Now consider evaluating I_0 for $Q_0 = Q_1 = \frac{1}{2}$, $Q_2 = 0$. Note that since $Q_2 = 0$, the value of $P'(j|2), j = 0,1,2$ is irrelevant in the computation of $I(Q,P')$. The conditions imposed by (4.3) force the remaining transition probabilities to be same as those of the original channel P. So we see that P is a minimizing P' and the value of $I_0(\frac{1}{2},\frac{1}{2},0)$ is $I(Q,P) = \ln 2 - \frac{1}{2}h(\epsilon)$, where $h(x) = -x\ln x - (1-x)\ln(1-x)$ is the binary entropy function. Since the channel has cyclic symmetry, the same value applies to $I_0(0,\frac{1}{2},\frac{1}{2})$ and $I_0(\frac{1}{2},0,\frac{1}{2})$. Now observe that

a. $(\frac{1}{3},\frac{1}{3},\frac{1}{3})$ is a convex combination of $(\frac{1}{2},\frac{1}{2},0)$, $(0,\frac{1}{2},\frac{1}{2})$ and $(\frac{1}{2},0,\frac{1}{2})$:

$$(\tfrac{1}{3},\tfrac{1}{3},\tfrac{1}{3}) = \tfrac{1}{3}(\tfrac{1}{2},\tfrac{1}{2},0) + \tfrac{1}{3}(0,\tfrac{1}{2},\tfrac{1}{2}) + \tfrac{1}{3}(\tfrac{1}{2},0,\tfrac{1}{2}),$$

b. however, for small enough[5] ϵ,

$$I_0(\tfrac{1}{2},\tfrac{1}{2},0) = I_0(0,\tfrac{1}{2},\tfrac{1}{2}) = I_0(\tfrac{1}{2},0,\tfrac{1}{2}) = \ln 2 - \tfrac{1}{2}h(\epsilon) > \ln(3/2) = I_0(\tfrac{1}{3},\tfrac{1}{3},\tfrac{1}{3})$$

that is, $\min_{P'\in\wp(Q,P)} I(Q,P')$ is not concave.

The previous computation implies the following for the channel in the example:

1. Given any $\lambda > 0$ and $\delta > 0$, there is a sufficiently large n for which we can find a code with block length n, and rate greater than $\ln 2 - \frac{1}{2}h(\epsilon) - \delta$ that achieves zero error probability and erasure probability less than λ. Furthermore, we can choose this code such that all the codewords are of composition $(\frac{1}{2},\frac{1}{2},0)$. Let us denote the codewords of this code by $\mathbf{a}_1,\ldots,\mathbf{a}_M$.

2. We can find another code that satisfies the same requirements but whose codewords have composition $(0,\frac{1}{2},\frac{1}{2})$. Indeed, if we let

$$\mathbf{b}_{ij} = \mathbf{a}_{ij} + 1 \bmod 3, \quad i = 1,\ldots,M,\ j = 1,\ldots,n,$$

$\mathbf{b}_1,\ldots,\mathbf{b}_M$ will be such a code. Even another code with codewords $\mathbf{c}_1,\ldots,\mathbf{c}_M$, each with composition $(\frac{1}{2},0,\frac{1}{2})$, can be formed by letting

$$\mathbf{c}_{ij} = \mathbf{b}_{ij} + 1 \bmod 3, \quad i = 1,\ldots,M,\ j = 1,\ldots,n.$$

3. We can form M^3 codewords of length $3n$ by forming the concatenation $\mathbf{a}_i\mathbf{b}_j\mathbf{c}_k$, for each $i,j,k \in \{1,\ldots,M\}$. This new code has rate at least $\ln 2 - \frac{1}{2}h(\epsilon) - \delta$, zero error probability and erasure probability at most 3λ. Note that the composition of each of the codewords of this new code is $(\frac{1}{3},\frac{1}{3},\frac{1}{3})$.

[5] $\epsilon < 0.2622180296581738203814796940489551009 9\ldots$ to be more precise

4. Thus we have constructed a zero error code of composition Q with rate greater than $\min_{P' \in \wp(Q,P)} I(Q, P')$. Hence we see that $\min_{P' \in \wp(Q,P)} I(Q, P')$ is not a tight lower bound to the maximum rate of zero error codes of composition Q. However the supremum of this quantity over Q might be the actual zero undetected error capacity. We will show that this is the case for some but not all classes of channels.

The Result of Pinsker and Sheverdyaev With the lower bound in (4.2) it is easy to obtain the result of Pinsker and Sheverdyaev mentioned earlier: if the graph of the channel as defined on page 42 contains no cycles, then we claim that

$$P' \in \wp(Q, P) \Rightarrow \forall k, j \; Q_k P'(j|k) = Q_k P(j|k),$$

and thus

$$I(Q, P') = I(Q, P).$$

Thus $C_{0u}(P) = \max_Q I(Q, P) = C(P)$, and the lower bound is exact. To prove our claim, let $W_j = \sum_k Q_k P_{jk}$ and note that $P' \in \wp(Q, P)$ implies

$$W_j = \sum_k Q_k P'_{jk} \tag{4.4}$$

and also

$$Q_k = \sum_j Q_k P'_{jk}. \tag{4.5}$$

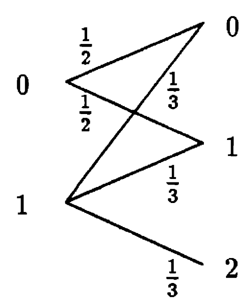
Suppose there exists k_0 and j_0 such that $Q_{k_0} P'_{j_0 k_0} \neq Q_{k_0} P_{j_0 k_0}$. Then, to satisfy (4.4) there must exist $k_1 \neq k_0$ such that $Q_{k_1} P'_{j_0 k_1} \neq Q_{k_1} P_{j_0 k_1}$. To satisfy (4.5) there must exist $j_1 \neq j_0$ such that $Q_{k_1} P'_{j_1 k_1} \neq Q_{k_1} P_{j_1 k_1}$. Continuing in this manner we find a sequence $k_0, j_0, k_1, j_1, \ldots$ such that $k_n \neq k_{n+1}$, $j_n \neq j_{n+1}$, $Q_{k_n} P'_{j_n k_n} \neq Q_{k_n} P_{j_n k_n}$ and $Q_{k_{n+1}} P'_{j_n k_{n+1}} \neq Q_{k_{n+1}} P_{j_n k_{n+1}}$. Furthermore $k_0, k_1, \ldots$ must be all distinct, otherwise, if, say $k_n = k_{n+m}$ then the sequence of nodes $k_n, j_n, \ldots, k_{n+m}$ would form a cycle in the graph. Since the input alphabet is finite, this is a contradiction.

Channels whose graphs do not contain cycles are not the only ones for which $C_{0u}(P) = C(P)$. Channels for which $P(j|k)$ is constant over its support (i.e., the set of positive values of $P(j|k)$ is a singleton) also have this property. We will prove this result by showing that $I(Q, P) \leq I(Q, P')$ for any Q and any $P' \in \wp(Q, P)$.

Another observation is that $C_{0u}(P) > 0$ if and only if there is an output j that can be reached from some but not all inputs: if each output can be reached from every input, then no matter what code we devise every output sequence would be reachable from every codeword, never leading to certainty. On the other hand, if there is one such output then it is easy to see that FORNEY's lower bound is positive.

4.3 Some Special Classes of Channels

Here we note a few classes of channels for which $C_{0u} = C$. One such class, namely, the channels whose graphs are trees was noted above. The channels for which the non-zero values of P_{jk} only depend on k (i.e., the non-zero transitions from any given output k have equal probabilities; see Figure 4-4 for an example) also have the property that



The channel shown above has the property that the non-zero values of P_{jk} depend only on k and not j: all the non-zero transitions from a given input have the same probability; here, the transitions from input 0 have probability $\frac{1}{2}$ and the transitions from 1 have probability $\frac{1}{3}$.

Figure 4-4: Example for $C_{0u} = C$

$C_{0u} = C$. If the non-zero values of P_{jk} are determined by k only, P_{jk} must be of the form

$$P_{jk} = \theta_{jk} c_k$$

where θ_{jk} takes on the values 0 or 1. As proved in Appendix 4.B this means that P is a minimizing P': $I(Q,P) \leq I(Q,P')$ for all $P' \in \wp(Q,P)$. Thus the lower bound in (4.2) is the capacity C of the channel, and thus $C_{0u} = C$.

Similarly channels for which the non-zero values of P_{jk} depend only on j have the same property. The intersection of these classes is the class of channels for which

$$P_{jk} = \theta_{jk} c,$$

that is P_{jk} is either 0 or c.

4.4 Errors-and-Erasures Schemes

Now we turn to the question of improving the exponential bounds derived in [For68]. Recall that the decoder is allowed to declare an erasure and, unlike the previous discussion, is not required to be error free in its estimates. Our aim is to minimize the probability of error for a given bound on the erasure probability. FORNEY [For68] shows that, for a given set of codewords $\{\mathbf{x}_1, \ldots, \mathbf{x}_M\}$, the optimal decoding rule is to decode m if

$$\frac{\Pr[\mathbf{y}|\mathbf{x}_m]}{\sum_{m' \neq m} \Pr[\mathbf{y}|\mathbf{x}_{m'}]} \geq \exp(nT),$$

where n is the length of the block code and T is a parameter chosen to satisfy the requirements on the erasure probability. Note that if $T > 0$ then there can be at most one m satisfying the condition. If no m satisfies the requirement then the decoder declares an erasure. T governs the tradeoff between errors and erasures. Larger values of T impose higher confidence levels, thus decreasing the error probability P_e and increasing the erasure probability P_X. FORNEY shows that for a decoder using this decoding scheme,

$$P_X \leq \exp -nE_1(R,T) \quad \text{and} \quad P_e \leq \exp -nE_2(R,T),$$

where $E_1(R,T) = \max_{Q, 0 \leq s \leq \rho \leq 1} \Big[E_0(s,\rho,Q,P) - \rho R - sT\Big]$, $E_2(R,T) = E_1(R,T) + T$ and

$$E_0(s,\rho,Q,P) = -\ln \sum_j \left[\sum_k Q_k P_{jk}^{1-s}\right] \left[\sum_{k'} Q_{k'} P_{jk'}^{s/\rho}\right]^\rho.$$

There are two natural ways to proceed to improve these bounds. The first is the idea that we can apply the bounds to the channel P^n representing the n-fold use of the original channel P. The second is to rederive the bound independently from a constant composition coding argument. We have pursued the first approach with the input distributions restricted to those that are uniform over a set of constant composition in [TG89]. Here we present the second approach, which seems to be the more natural framework. I believe one can show that the two approaches yield identical results.

Let $\mathbf{x}_1, \ldots, \mathbf{x}_M$ be the codewords, and let $\mathbf{y}$ be the received sequence. Assume that each $\mathbf{x}_m$ is of composition Q, and let P_m be the type of the pair $(\mathbf{y}, \mathbf{x}_m)$. Note that P_mQ

is the composition of $\mathbf{y}$ and thus does not vary with m. We will propose a decoding rule of the following type:

Decode m if

$$P_m \prec P_{m'} \text{ for all } m' \neq m. \tag{4.6}$$

If no m satisfies the criterion (4.6), declare an erasure.

The relation $\prec$ will be left arbitrary; we will only assume that at most one of $V \prec W$ and $W \prec V$ will hold, i.e., that $\prec$ is asymmetric. This will exclude the possibility of more than one m satisfying the decoding criterion. The issue of choosing an appropriate $\prec$ will be discussed later.

Let us try to estimate the erasure and error probabilities for this decoding rule. An erasure will occur if no m satisfies (4.6). In particular, the m that was transmitted should not satisfy (4.6). Thus we may overbound the erasure probability by the probability of this second event[6]. Using the same argument as in Section 4.2,

$$\begin{aligned}
P_X &\leq \Pr[P_m \not\prec P_{m'} \text{ for some } m' \neq m | \mathbf{x}_m \text{ transmitted}] \\
&\leq \sum_{\widehat{P}} \exp(-nD(\widehat{P}\|P|Q)) \sum_{\substack{P':\widehat{P}\not\prec P' \\ P'Q=\widehat{P}Q}} \exp(-n|I(Q,P') - R|^+) \\
&\leq \exp -n[\min_{\substack{\widehat{P},P':\widehat{P}\not\prec P' \\ \widehat{P}Q=P'Q}} D(\widehat{P}\|P|Q) + |I(Q,P') - R|^+ - o(n)].
\end{aligned}$$

Given that m is transmitted, an error will occur if, for some $m' \neq m$, $P_{m'} \prec P_{m''}$ for all $m'' \neq m'$. This event is contained in the event that $P_{m'} \prec P_m$. Thus

$$\begin{aligned}
P_e &\leq \Pr[P_{m'} \prec P_m \text{ for some } m' \neq m | \mathbf{x}_m \text{ transmitted}] \\
&\leq \sum_{\widehat{P}} \exp(-nD(\widehat{P}\|P|Q)) \sum_{\substack{P':P'\prec\widehat{P} \\ P'Q=\widehat{P}Q}} \exp(-n|I(Q,P') - R|^+) \\
&\leq \exp -n[\min_{\substack{\widehat{P},P':P'\prec\widehat{P} \\ \widehat{P}Q=P'Q}} D(\widehat{P}\|P|Q) + |I(Q,P') - R|^+ - o(n)] \\
&= \exp -n[\min_{\substack{\widehat{P},P':\widehat{P}\prec P' \\ \widehat{P}Q=P'Q}} D(P'\|P|Q) + |I(Q,\widehat{P}) - R|^+ - o(n)].
\end{aligned}$$

In the last line, we interchanged the roles of $\widehat{P}$ and P' to illustrate the duality between

[6]Note that this is an overbound to $P_X + P_e$

P_X and P_e. Now define

$$E_X(Q,P,R) = \min_{\substack{\widehat{P},P':\widehat{P}\not\prec P' \\ \widehat{P}Q=P'Q}} D(\widehat{P}\|P|Q) + |I(Q,P') - R|^+, \tag{4.7}$$

$$E_e(Q,P,R) = \min_{\substack{\widehat{P},P':\widehat{P}\prec P' \\ \widehat{P}Q=P'Q}} D(P'\|P|Q) + |I(Q,\widehat{P}) - R|^+. \tag{4.8}$$

We see that E_X and E_e are lower bounds to the erasure and error exponents for this particular decoding rule, and the particular choice of code composition Q.

4.4.1 How to choose $\prec$.

Now we address the issue of choosing a suitable relation $\prec$. Suppose we are given a minimum tolerable value T of the erasure exponent and we want to maximize the value of the error exponent. We would then set

$$\widehat{P} \prec P' \quad \text{whenever } D(\widehat{P}\|P|Q) + |I(Q,P') - R|^+ \leq T.$$

We will then have

$$E_X(R,P,Q) \geq T, \quad E_e(R,P,Q) = \min_{\substack{\widehat{P},P':\widehat{P}Q=P'Q \\ D(\widehat{P}\|P|Q)+|I(Q,P')-R|^+\leq T}} D(P'\|P|Q) + |I(Q,\widehat{P}) - R|^+.$$

Note that if this second minimization yields a value greater than T, the relation $\prec$ thus defined will be asymmetric as required.

We will first examine a special case in detail: Suppose we are interested in very small but positive values of T. Then the demand on erasures is that their probability decays to zero as the blocklength increases, but no restriction is made on the rate of such decay. The limiting behavior as the values of T get ever smaller can be examined by analyzing the case when $T = 0$. We shall call the maximum achievable error exponent the "feedback exponent" and denote it by $E_f(R,P,Q)$. In this case

$$D(\widehat{P}\|P|Q) + |I(Q,P') - R|^+ \leq 0 \quad \text{implies} \quad \widehat{P}_{jk} = P_{jk} \text{ for } Q_k > 0, \text{ and, } I(Q,P') \leq R,$$

and thus

$$E_f(R,P,Q) = E_e(R,P,Q) = \min_{\substack{P':I(Q,P')\le R\\ PQ=P'Q}} D(P'\|P|Q) + |I(Q,P)-R|^+.$$

If we define

$$E_{fsp}(R,P,Q) = \min_{\substack{V:I(Q,V)\le R\\ PQ=VQ}} D(V\|P|Q)$$

then we can express the above as

$$E_f(R,P,Q) = E_{fsp}(R,P,Q) + |I(Q,P)-R|^+.$$

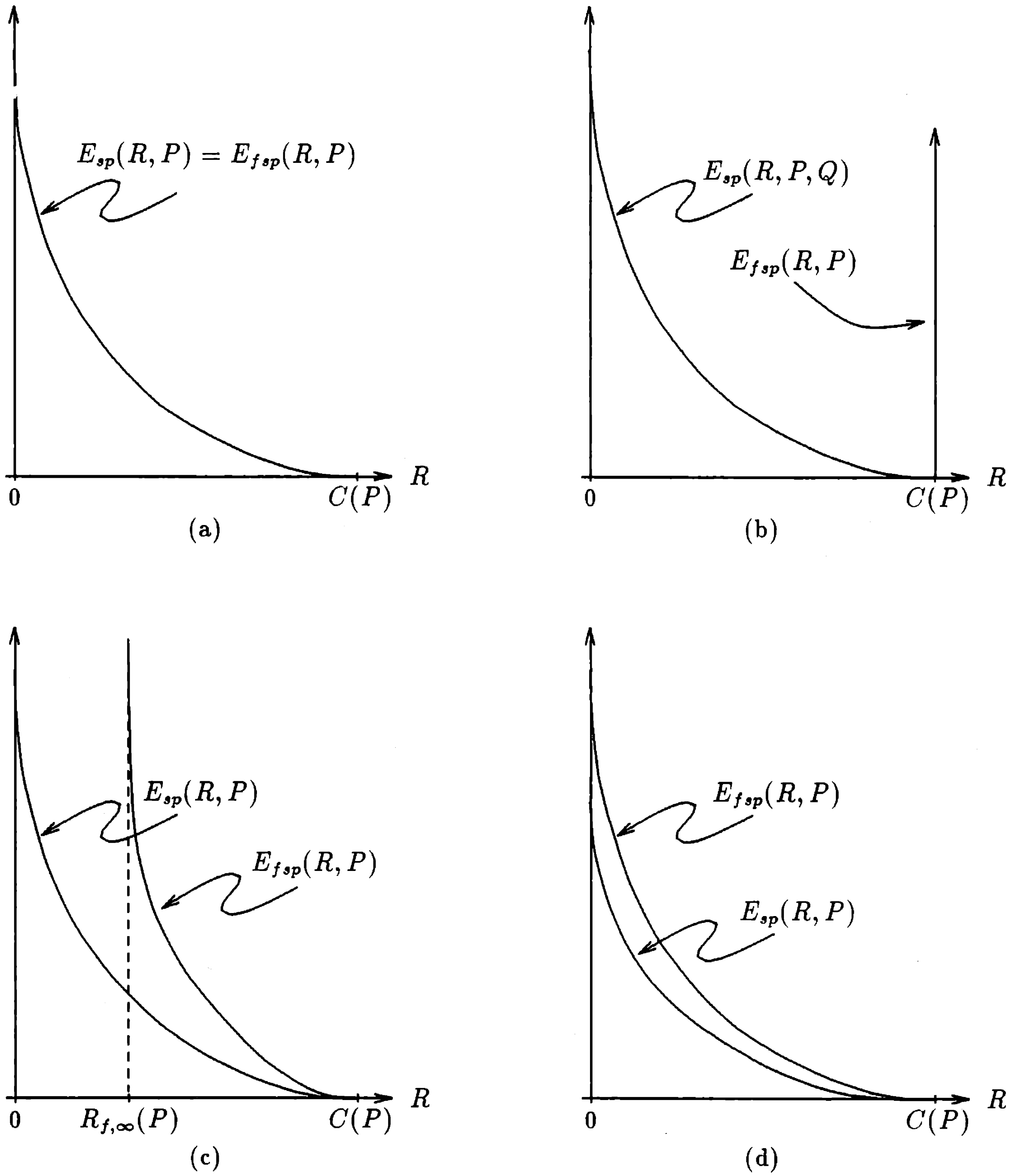
Note that $E_{fsp}(R,P,Q) \ge E_{sp}(R,P,Q) \stackrel{\text{def}}{=} \min_{V:I(Q,V)\le R} D(V\|P|Q)$. In particular, for tree channels, we know that $R < I(Q,P)$ implies $E_{fsp}(R,P,Q) = +\infty$. For totally symmetric channels, with a Q achieving the maxima in $E_{sp}(R,P) \stackrel{\text{def}}{=} \max_Q E_{sp}(R,P,Q)$, we have $E_{fsp}(R,P,Q) = E_{sp}(R,P)$, and, $E_e(R,P,Q) = E_{sp}(R,P) + C(P) - R$. Also note that for a given P and Q, E_{fsp} is a convex, non-increasing function of R. The second of these assertions is trivial and the first one follows from the convexity of $D(V\|P|Q)$ and $I(Q,V)$ with respect to V. We may also define $E_{fsp}(R,P) \stackrel{\text{def}}{=} \max_Q E_{fsp}(R,P,Q)$, which will inherit the convexity and monotonicity (in R) of $E_{fsp}(R,P,Q)$. Figure 4-5 illustrates the relationship between $E_{fsp}(R,P)$ and $E_{sp}(R,P)$ for various classes of channels. Figure 4-5(a) illustrates the case of a binary symmetric channel; for this case, E_{fsp} and E_{sp} coincide. In Figure 4-5(b) we see the case of the Z-channel: $E_{fsp}(R)$ is infinite for $R < C$. Figure 4-5(c) shows the case for which $0 < \max_Q \min_{P'\in\wp(Q,P)} I(Q,P') < C$, and Figure 4-5(d) shows the case for which $C_{0u} = 0$. Unshown is the possibility of $E_{sp}(R)$ becoming infinite.

We can obtain a curious result if we evaluate the feedback exponent E_f at $R=0$:

$$\begin{aligned} E_f(0,P,Q) &= E_{fsp}(0,P,Q) + I(Q,P) \\ &= \min_{\substack{V:I(Q,V)\le 0\\ PQ=VQ}} D(V\|P|Q) + I(Q,P) \end{aligned}$$

Now, $I(Q,V)\le 0$ and $VQ=PQ$ imply $V_{jk} = W_j \stackrel{\text{def}}{=} \sum_k Q_k P_{jk}$, and we get

$$E_f(0,P,Q) \;=\; \sum_{k,j} Q_k W_j \ln \frac{W_j}{P_{jk}} + \sum_{k,j} Q_k P_{jk} \ln \frac{P_{jk}}{W_j}$$



The figure shows the relationship between E_{sp} and E_{fsp} for various channels. In (a) we see the case of a totally symmetric channel, (b) shows the case for which $R_{f,\infty}(P) \stackrel{\text{def}}{=} \max_Q \min_{P' \in \wp(Q,P)} I(Q,P') = C_{0u} = C$, (c) illustrates the case for which $0 < R_{f,\infty} < C$, and in (d) we see the case for which $C_{0u} = 0$ but $E_{sp} \neq E_{fsp}$.

Figure 4-5: Relationship between E_{sp} and E_{fsp}

$$
\begin{aligned}
&= \sum_{k,j} Q_k(W_j - P_{jk}) \ln \frac{W_j}{P_{jk}} \\
&= \sum_{k,j} Q_k(W_j - P_{jk}) \ln \frac{1}{P_{jk}} \\
&= \sum_{i,k,j} Q_i Q_k P_{jk} \ln \frac{P_{jk}}{P_{ji}},
\end{aligned}
$$

and

$$E_f(0,P) \stackrel{\text{def}}{=} \max_Q E_f(0,P,Q) = \max_Q \sum_{i,k,j} Q_i Q_k P_{jk} \ln \frac{P_{jk}}{P_{ji}}.$$

We will prove the following theorem:

Theorem 4.4 *For any given channel P, and any input distribution Q, an erasure exponent of 0 and an error exponent of $E_f(0,P,Q)$ is achievable. Moreover, for any sequence of block codes $C_1, C_2, \ldots$ of increasing block length and with $P_e(C_i) \to 0$, $P_X(C_i) \to 0$ and $M(C_i) \to \infty$,*

$$P_e(C_i) \geq \exp -n_i[E_f(0,P) + o(n_i)].$$

Here n_i is the blocklength of the code C_i and $M(C_i)$ denotes the number of codewords in the code C_i.

Proof. The first part of the theorem is already proved. We need to prove the converse. Let $\mathbf{x}_1, \ldots, \mathbf{x}_M$ be the codewords and let $D_1, \ldots, D_M$ be the decoding regions corresponding to these codewords. We will use the fact that processing does not increase informational divergence, i.e., for any two distributions V and V' and for any set S

$$V(S) \ln \frac{V(S)}{V'(S)} + (1 - V(S)) \ln \frac{1 - V(S)}{1 - V'(S)} \leq D(V \| V').$$

By rearranging the terms we get

$$V(S) \ln \frac{1}{V'(S)} + (1 - V(S)) \ln \frac{1}{1 - V'(S)} \leq D(V \| V') + h(V(S))$$

where h is the binary entropy function. Further lower bounding the right hand side by eliminating the second term, and rearranging, we get

$$V'(S) \geq \exp - \frac{D(V \| V') + h(V(S))}{V(S)}.$$

Let us take $V = P^n(\cdot|\mathbf{x}_m)$, $V' = P^n(\cdot|\mathbf{x}_{m'})$ and $S = D_m$. We see that

$$P^n(D_m|\mathbf{x}_{m'}) \geq \exp - \frac{D(P^n(\cdot|\mathbf{x}_m)\|P^n(\cdot|\mathbf{x}_{m'})) + h(P^n(D_m|\mathbf{x}_m))}{P^n(D_m|\mathbf{x}_m)}.$$

Now let $E_{m,m'} \stackrel{\text{def}}{=} D(P^n(\cdot|\mathbf{x}_m)\|P^n(\cdot|\mathbf{x}_{m'}))$. Then

$$\begin{aligned}
E_{m,m'} &= D(P^n(\cdot|\mathbf{x}_m)\|P^n(\cdot|\mathbf{x}_{m'})) \\
&= \sum_{\mathbf{y}} P^n(\mathbf{y}|\mathbf{x}_m) \ln \frac{P^n(\mathbf{y}|\mathbf{x}_m)}{P^n(\mathbf{y}|\mathbf{x}_{m'})} \\
&= \sum_{\ell=1}^{n} \sum_{y_\ell} P(y_\ell|x_{m\ell}) \ln \frac{P(y_\ell|x_{m\ell})}{P(y_\ell|x_{m'\ell})} \\
&= \sum_{\ell=1}^{n} \sum_{i,k} \chi_{i,k}^{\ell}(m,m') \sum_{j} P_{ji} \ln \frac{P_{ji}}{P_{jk}},
\end{aligned}$$

where

$$\chi_{i,k}^{\ell}(m,m') = \begin{cases} 1 & \text{if } x_{m\ell} = i, x_{m'\ell} = k, \\ 0 & \text{else.} \end{cases}$$

Now let us sum over all $m \neq m'$ and get

$$\begin{aligned}
\sum_{m \neq m'} E_{m,m'} &= \sum_{m,m'} E_{m,m'} \\
&= \sum_{\ell=1}^{n} \sum_{i,j,k} P_{ji} \ln \frac{P_{ji}}{P_{jk}} \sum_{m,m'} \chi_{i,k}^{\ell}(m,m') \\
&= M^2 \sum_{\ell=1}^{n} \sum_{i,j,k} Q_i^{\ell} Q_k^{\ell} P_{ji} \ln \frac{P_{ji}}{P_{jk}},
\end{aligned}$$

where $Q_k^{\ell} = |\{m : x_{m\ell} = k\}|/M$. The first step follows from $E_{m,m} = 0$, and the last step follows from

$$i \neq k \Rightarrow |\{(m,m') : x_{m\ell} = i, x_{m'\ell} = k\}| = |\{m : x_{m\ell} = i\}| \cdot |\{m' : x_{m'\ell} = k\}|,$$

and for $i = k$, $\ln(P_{ji}/P_{jk}) = 0$. Thus,

$$M(M-1) \min_{m \neq m'} E_{m,m'} \leq \sum_{m \neq m'} E_{m,m'} \leq nM^2 \max_{Q} \sum_{i,j,k} Q_i Q_k P_{ji} \ln \frac{P_{ji}}{P_{jk}} = nM^2 E_f(0,P)$$

and hence

$$\min_{m \neq m'} E_{m,m'} \leq (M/(M-1))nE_f(0,P).$$

It is now clear that for m and m' achieving the minimum above,

$$P_e \geq P^n(D_m|\mathbf{x}_{m'}) \geq \exp -n\left[\frac{M}{M-1}\frac{1}{P^n(D_m|\mathbf{x}_m)}[E_f(0,P)+1/n]\right]$$

where we have upper bounded the binary entropy term by 1. Since P_e and P_X are both approaching zero, $P^n(D_m|\mathbf{x}_m)$ approaches 1. Similarly $M/(M-1) \to 1$ as n gets large. From these observations the proof follows. □

We see that the feedback exponent E_f is tight at zero rate. This is somewhat unexpected, since one usually has to go through expurgation techniques to obtain tight bounds at low rates. One should also note that E_f is only a lower bound to the actual feedback exponent. The fact that it is not equal to the actual feedback exponent at rates greater than zero will be implied by the result of the next section. Namely, if E_f were equal to the actual feedback exponent, our lower bound to C_{0u} would be tight. The next section establishes that the latter fact is not true, and hence, E_f is in general not equal to the feedback exponent.

For values of T other than zero, we may lower bound E_e by removing the constraint $\widehat{P}Q = P'Q$. We then get

$$\begin{aligned}
E_e(R,P,Q) &\geq \min_{\widehat{P},P':D(\widehat{P}\|P|Q)+|I(Q,P')-R|^+\leq T} D(P'\|P|Q)+|I(Q,\widehat{P})-R|^+ \\
&= \min_{0\leq a\leq T}\ \min_{D(\widehat{P}\|P|Q)\leq a}\ \min_{I(Q,P')\leq T+R-a} D(P'\|P|Q)+|I(Q,\widehat{P})-R|^+ \\
&= \min_{0\leq a\leq T}[E_{sp}(R+T-a,P,Q)+|\min_{D(\widehat{P}\|P|Q)\leq a} I(Q,\widehat{P})-R|^+] \\
&= \min_{0\leq a\leq T}[E_{sp}(R+T-a,P,Q)+|E_{sp}^{-1}(a,P,Q)-R|^+],
\end{aligned}$$

where we define

$$E_{sp}^{-1}(a,P,Q) = \inf\{R : E_{sp}(R,P,Q) \leq a\},$$

as a customary definition of an inverse function for decreasing functions. For $R \geq R_{\text{crit}}$ and $T \leq E_{sp}(R,P,Q)$, it can be shown that the minimization occurs at $a = T$, and we obtain

$$E_e(R,P,Q) \geq E_{sp}(R,P,Q) + E_{sp}^{-1}(T,P,Q) - R$$

If we set $\tilde{R} \stackrel{\text{def}}{=} E_{sp}^{-1}(T, P, Q) \geq R$, we see that

$$E_X(R, P, Q) \geq E_{sp}(\tilde{R}, P, Q), \quad E_e(R, P, Q) \geq E_{sp}(R, P, Q) + \tilde{R} - R,$$

are simultaneously achievable for any $\tilde{R} \geq R$. This same result will be rederived in the next two sections by a different approach.

4.4.2 A Parametric Approach

Instead of fixing a particular erasure exponent and trying to find the largest possible error exponent, we may parametrize the tradeoff between these two exponents. To this end, consider for $\lambda \geq 0$

$$\min_{\hat{P}, P' : \hat{P}Q = P'Q} [D(\hat{P} \| P | Q) + |I(Q, P') - R|^+ + \lambda(D(P' \| P | Q) + |I(Q, \hat{P}) - R|^+)] \quad (4.9)$$

Since the minimization is done on a compact set, and since the function is continuous, the minima is achieved. Let $\hat{P}(\lambda)$ and $P'(\lambda)$ be the values of $\hat{P}$ and P' that achieve the minimum. We see that for any $\hat{P}$ and P' such that $\hat{P}Q = P'Q$, at most one of

$$D(\hat{P} \| P | Q) + |I(Q, P') - R|^+ < D(\hat{P}(\lambda) \| P | Q) + |I(Q, P'(\lambda)) - R|^+ \quad (4.10)$$

$$D(P' \| P | Q) + |I(Q, \hat{P}) - R|^+ < D(P'(\lambda) \| P | Q) + |I(Q, \hat{P}(\lambda)) - R|^+ \quad (4.11)$$

can be satisfied. We will put $\hat{P} \prec P'$ whenever (4.10) is satisfied. Note that this will ensure that $\prec$ is asymmetric. Now using equations (4.7) and (4.8) we can easily see that

$$E_X(Q, P, R) \geq \tilde{E}_X(Q, P, R, \lambda) \stackrel{\text{def}}{=} D(\hat{P}(\lambda) \| P | Q) + |I(Q, P'(\lambda)) - R|^+$$

and

$$E_e(Q, P, R) \geq \tilde{E}_e(Q, P, R, \lambda) \stackrel{\text{def}}{=} D(P'(\lambda) \| P | Q) + |I(Q, \hat{P}(\lambda)) - R|^+.$$

It is clear that λ governs the tradeoff between the error and erasure exponents. However, the range of values of $D(\hat{P}(\lambda) \| P | Q) + |I(Q, P'(\lambda)) - R|^+$ as λ changes may not cover the range between 0 and $E_{sp}(R, P, Q)$. For example, even for $\lambda = 0$, the lower bound to erasure exponent may be strictly positive, inhibiting the analysis of the feedback exponent.

4.4.3 Comparison with Csiszár and Körner's results

Here we will compare our results on the error and erasure exponents with those derived in [CK81], which are in turn essentially the same as those in [For68]. One difference between [CK81] and [For68] is that the sphere packing exponent used in [CK81] is tighter for compositions that are not optimal; this difference disappears once one maximizes over the input distribution.

First consider a relaxed version of the minimization done in (4.9):

$$\min_{\hat{P},P'}[D(\hat{P}\|P|Q) + |I(Q,P') - R|^+ + \lambda(D(P'\|P|Q) + |I(Q,\hat{P}) - R|^+)] \tag{4.12}$$

The only difference between (4.9) and (4.12) is in the removal of the constraint $\hat{P}Q = P'Q$. Relaxing our constraints can only decrease the values of the exponents we obtain. One notices that this new minimization can be done in separate steps:

$$\min_{\hat{P}} D(\hat{P}\|P|Q) + \lambda|I(Q,\hat{P}) - R|^+ + \lambda[\min_{P'} D(P'\|P|Q) + \lambda^{-1}|I(Q,P') - R|^+]. \tag{4.13}$$

Using the notation of [CK81, p.174] we recognize the above as

$$E_{r,\lambda}(R,P,Q) + \lambda E_{r,1/\lambda}(R,P,Q),$$

where $E_{r,\lambda}(R,Q,P) = \min_V D(V\|P|Q) + \lambda|I(Q,V) - R|^+$. It can be shown ([CK81, p. 174]) that

$$E_{r,\lambda} = \begin{cases} E_{sp}(R,Q,P) & \text{if } R \geq R_\lambda \\ E_{sp}(R_\lambda,Q,P) + \lambda(R_\lambda - R) & \text{if } 0 \leq R \leq R_\lambda, \end{cases}$$

where $E_{sp}(R,P,Q) = \min_{V:I(Q,V)\leq R} D(V\|P|Q)$ is the sphere-packing exponent and R_λ is the smallest R at which the convex curve $E_{sp}(R,P,Q)$ meets its supporting line of slope $-\lambda$. For $R \leq I(Q,P)$, the minimization for $E_{sp}(R,P,Q)$ always occurs at a V for which $I(Q,V) = R$, so we can write $E_{sp}(R,P,Q) = \min_{V:I(Q,V)=R} D(V\|P|Q)$ in this range. Note that the minimizing V in $E_{r,\lambda}$, V^*, satisfies

$$\begin{array}{lll} D(V^*\|P|Q) = E_{sp}(R,P,Q) & I(Q,V^*) = R & \text{if } R \geq R_\lambda, \\ D(V^*\|P|Q) = E_{sp}(R_\lambda,P,Q) & I(Q,V^*) = R_\lambda & \text{if } R \leq R_\lambda. \end{array}$$

We can now express a lower bound to erasure and error exponents in terms of the sphere packing exponent E_{sp}. For a given R, choose $0 \leq \lambda \leq 1$. There are three possibilities:

1. $R \leq R_{1/\lambda} \leq R_\lambda$. Denote the minimizing $\widehat{P}$ and P' in (4.13) by $\widehat{P}(\lambda)$ and $P'(\lambda)$[7]. We see that $I(Q,\widehat{P}(\lambda)) = R_\lambda$, $D(\widehat{P}(\lambda)\|P|Q) = E_{sp}(R_\lambda, P, Q)$, $I(Q,P'(\lambda)) = R_{1/\lambda}$, $D(P'(\lambda)\|P|Q) = E_{sp}(R_{1/\lambda}, P, Q)$, and

$$D(\widehat{P}(\lambda)\|P|Q) + |I(Q,P'(\lambda)) - R|^+ = E_{sp}(R_\lambda, P, Q) + R_{1/\lambda} - R,$$

$$D(P'\|P|Q) + |I(Q,\widehat{P}(\lambda)) - R|^+ = E_{sp}(R_{1/\lambda}, P, Q) + R_\lambda - R,$$

and that these are achievable error and erasure exponents. Figure 4-6 gives a geometric construction for these quantities.

2. $R_{1/\lambda} \leq R \leq R_\lambda$. We see that $I(Q,\widehat{P}(\lambda)) = R$, $D(\widehat{P}(\lambda)\|P|Q) = E_{sp}(R,P,Q)$, $I(Q,P'(\lambda)) = R_\lambda$ and $D(P'(\lambda)\|P|Q) = E_{sp}(R_\lambda, P, Q)$. Then

$$D(\widehat{P}(\lambda)\|P|Q) + |I(Q,P'(\lambda)) - R|^+ = E_{sp}(R_\lambda, P, Q),$$

$$D(P'(\lambda)\|P|Q) + |I(Q,\widehat{P}(\lambda)) - R|^+ = E_{sp}(R, P, Q) + R_\lambda - R,$$

and $E_{sp}(R,P,Q) + R_\lambda - R$ and $E_{sp}(R_\lambda, P, Q)$ are achievable error and erasure exponents. We can also give a geometrical construction to obtain these lower bounds. The procedure is illustrated in Figure 4-7.

3. $R_{1/\lambda} \leq R_\lambda \leq R$. We see that $\widehat{P}(\lambda)$ and $P'(\lambda)$ satisfy

$$D(\widehat{P}(\lambda)\|P|Q) = D(P'(\lambda)\|P|Q) = E_{sp}(R,P,Q)$$

and

$$I(Q,\widehat{P}(\lambda)) = I(Q,P'(\lambda)) = R.$$

Thus

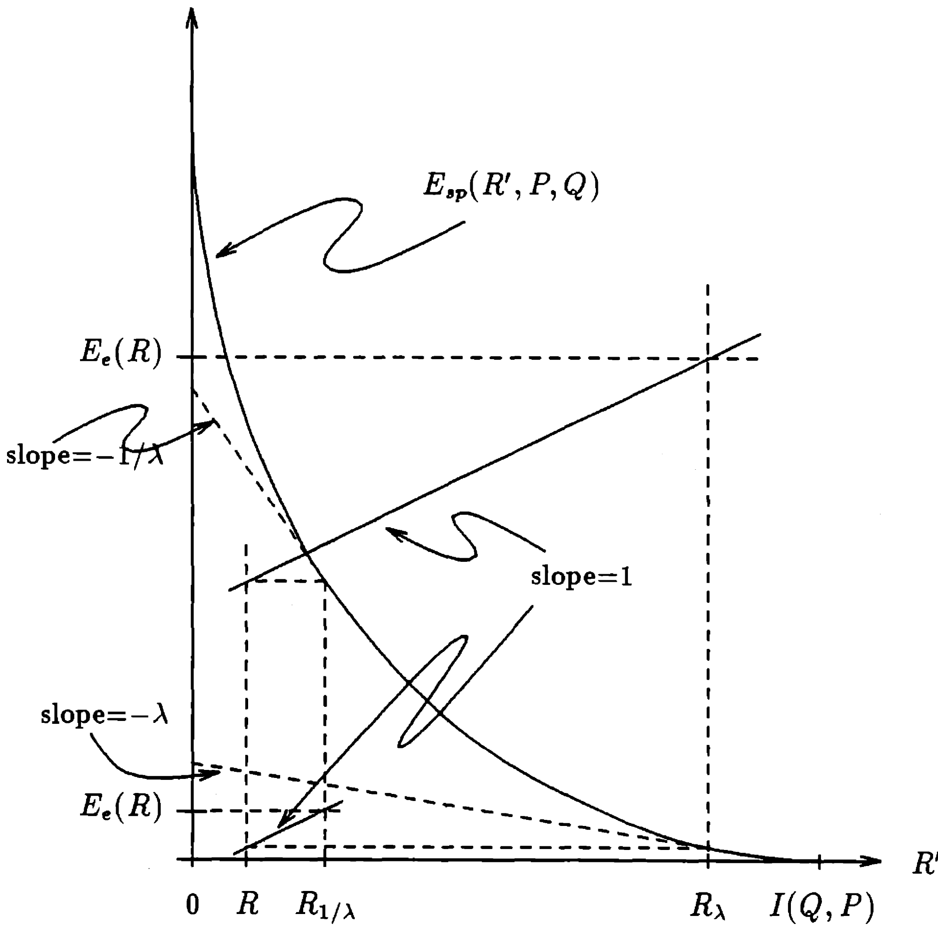
$$D(\widehat{P}(\lambda)\|P|Q) + |I(Q,P'(\lambda)) - R|^+ = E_{sp}(R,P,Q),$$

$$D(P'(\lambda)\|P|Q) + |I(Q,\widehat{P}(\lambda)) - R|^+ = E_{sp}(R,P,Q),$$

and thus the bound does not improve on the previous case.

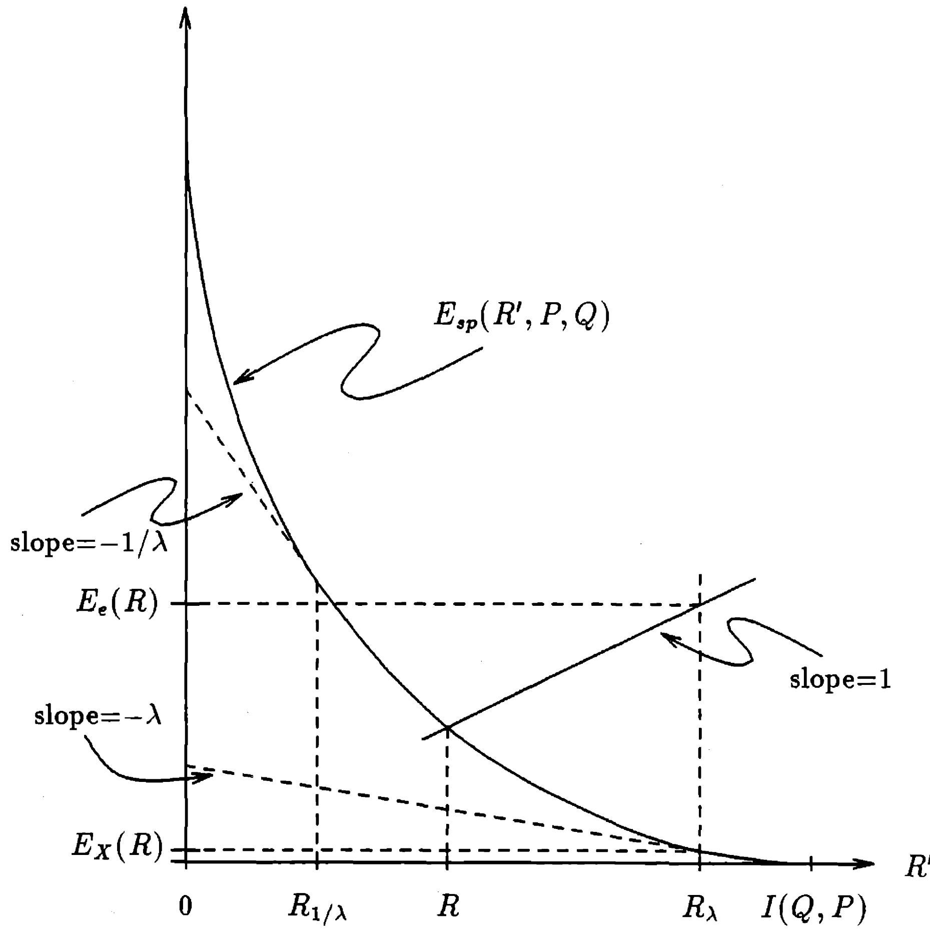
For $R \geq R_{\text{crit}} \stackrel{\text{def}}{=} R_1$, $R_\lambda \geq R$ implies $R_{1/\lambda} \leq R$. In this case, by varying λ we see that $E_{sp}(R) - R + \tilde{R}$ and $E_{sp}(\tilde{R})$ are achievable error and erasure exponents for any $\tilde{R} \geq R$. Csiszár and Körner give similar lower bounds: For every $\tilde{R} \geq R$, $E_{r,\lambda}(R,P,Q) - R + \tilde{R}$

[7]We could have written $\widehat{P}(1/\lambda)$ in place of $P'(\lambda)$



With $R \leq R_{1/\lambda} \leq R_\lambda$ we can achieve error and erasure exponents of $E_{sp}(R_{1/\lambda}) + R_{1/\lambda} - R$ and $E_{sp}(R_\lambda) + R_\lambda - R$. The figure gives a geometric construction for these quantities.

Figure 4-6: Construction of error and erasure exponents for small R



For any given value of $R_{1/\lambda} \leq R \leq R_\lambda$, we can achieve an error exponent of $E_{sp}(R) - R + R_\lambda$ and an erasure exponent of $E_{sp}(R_\lambda)$. The figure also shows $E_{r,\lambda}(R)$ and $E_{r,1/\lambda}(R)$ of Csiszár and Körner. In the figure R_λ denotes the value of R at which $E_{sp}(R)$ meets a slope of $-\lambda$. Similarly $R_{1/\lambda}$ denotes the value of R at which $E_{sp}(R)$ meets a slope of $-1/\lambda$.

Figure 4-7: Construction of error and erasure exponents for high R.

for error and $E_{r,1/\lambda}(\tilde{R}, P, Q)$ for erasure exponents. Since E_{sp} dominates both $E_{r,\lambda}$ and $E_{r,1/\lambda}$, we see that our construction gives at least as good bounds in this region.

4.A Comparison of Equation (4.2) and (4.1)

We had claimed that the lower bound in (4.2) is better than the one given in (4.1). Here we prove that it is indeed so.

Lemma 4.5 $-\sum_j w_j \ln v_j \leq I(Q,P)$ *where* $w_j = \sum_k Q_k P_{jk}$ *and* $v_j = \sum_{k:P_{jk}>0} Q_k$.

Proof. On noting

$$\begin{aligned} -\sum_j w_j \ln v_j - I(Q,P) &= \sum_j \sum_k Q_k P_{jk} \ln \frac{w_j}{v_j P_{jk}} \\ &= \sum_j \sum_{k:P_{jk}>0} Q_k P_{jk} \ln \frac{w_j}{v_j P_{jk}} \\ &\leq \sum_j \sum_{k:P_{jk}>0} Q_k P_{jk} \left[\frac{w_j}{v_j P_{jk}} - 1\right] \\ &= \sum_j w_j - 1 = 0, \end{aligned}$$

the proof follows. □

Lemma 4.6 *If* Q*,* P*, and* P' *satisfy* $P' \ll P$ *and* $w_j = \sum_k Q_k P'_{jk} = \sum_k Q_k P_{jk} = w'_j$*, then*

$$-\sum_j w_j \ln v_j \leq -\sum_j w'_j \ln v'_j,$$

$$v_j = \sum_{k:P_{jk}>0} Q_k \quad \textit{and} \quad v'_j = \sum_{k:P'_{jk}>0} Q_k.$$

Proof. Simply note that $P' \ll P$ implies

$$v_j = \sum_{k:P_{jk}>0} Q_k \geq \sum_{k:P'_{jk}>0} Q_k = v'_j,$$

and that ln is an increasing function. □

Theorem 4.7 *Equation (4.2) is a better lower bound than (4.1). That is,*

$$\max_Q -\sum_j w_j \ln v_j \leq \max_Q \min_{P' \in \wp(Q,P)} I(Q,P').$$

Proof. By the two previous lemmas, for all $P' \in \wp(Q, P)$

$$-\sum_j w_j \ln v_j \leq -\sum_j w'_j \ln v'_j \leq I(Q, P')$$

and thus

$$-\sum_j w_j \ln v_j \leq \min_{P' \in \wp(Q,P)} I(Q, P').$$

Now take maximums over both sides to complete the proof. $\square$

4.B Some Technical Aspects of $\min_{P' \in \wp(Q,P)} I(Q, P')$

Here we state some of the properties of the P' that minimizes

$$I(Q, P') = \sum_{k,j} Q_k P'_{jk} \ln \frac{P'_{jk}}{\sum_i Q_i P'_{ji}}$$

while being constrained to the set

$$\wp(Q, P) = \{P_{jk} = 0 \Rightarrow P'_{jk} = 0, \sum_k Q_k P'_{jk} = \sum_k Q_k P_{jk}\}.$$

To this end we form the Lagrangian

$$F_Q(P', \lambda, \mu) = I(Q, P') + \sum_j \lambda_j \sum_k Q_k P'_{jk} + \sum_k \mu_k \sum_j P'_{jk}.$$

Note that the first additional term corresponds to the constraint that the probability of the letters symbols should be unaltered when P is replaced by P' and the second additional term is to ensure that P' is a probability distribution. Taking the derivative of F_Q with respect to those P'_{jk} that are not bound by the $P' \ll P$ constraint and setting these derivatives equal to zero we obtain

$$Q_k \ln \frac{P'_{jk}}{\sum_i Q_i P_{ji}} + \lambda_j Q_k + \mu_k = 0,$$

which yields (again for those P'_{jk} not bound by $P' \ll P$)

$$P'_{jk} = c_k d_j.$$

If we let

$$\theta_{jk} = \begin{cases} 1 & P_{jk} > 0 \\ 0 & \text{else,} \end{cases}$$

we can write

$$P'_{jk} = \theta_{jk} c_k d_j$$

The choice of c_k and d_j must satisfy the conditions

$$\sum_j \theta_{jk} c_k d_j = 1 \quad \text{and} \quad \sum_k Q_k \theta_{jk} c_k d_j = \sum_k Q_k P_{jk}.$$

4.C Minimal Concave Functions

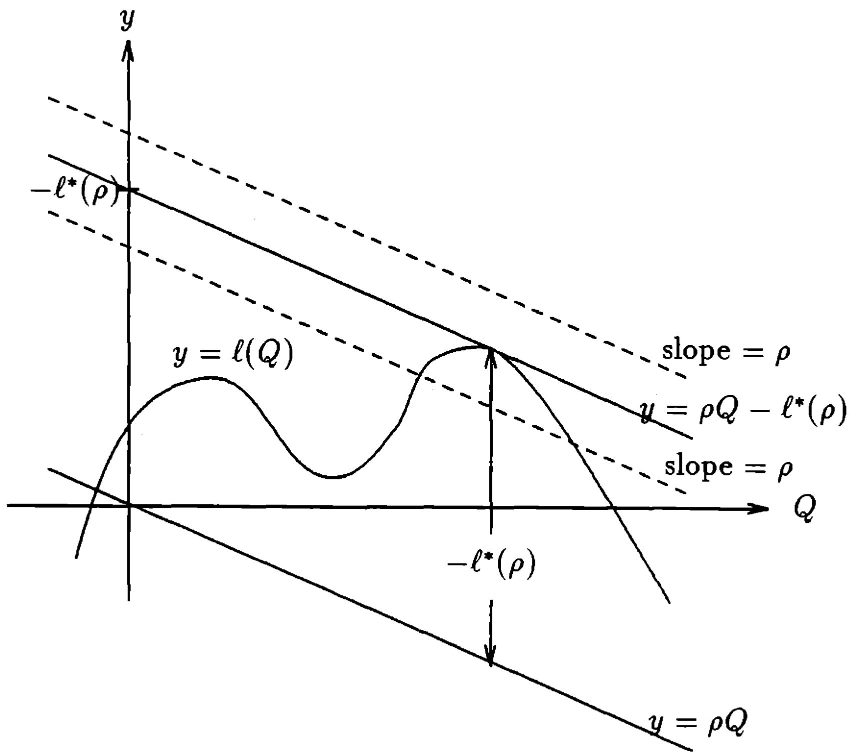
In Section 4 we claimed the following: Let $\ell : \mathbb{R}^K \to \mathbb{R}$. Define $\ell^* : \mathbb{R}^K \to \overline{\mathbb{R}}$ as[8]

$$\ell^*(\rho) = \inf_Q \rho \cdot Q - \ell(Q),$$

and $\ell^{**} : \mathbb{R}^K \to \overline{\mathbb{R}}$ as

$$\ell^{**}(Q) = \inf_\rho \rho \cdot Q - \ell^*(\rho);$$

then ℓ^{**} is the minimal concave function that dominates ℓ. In the preceding $\overline{\mathbb{R}}$ denotes the



$-\ell^*(\rho)$ is the minimal y-intercept of all lines of slope ρ that lie completely above ℓ. Some other lines of slope ρ are shown in the figure in dotted lines.

Figure 4-8: Illustration of $\inf_Q \rho Q - \ell(Q)$.

extended reals, i.e., $\overline{\mathbb{R}} = \mathbb{R} \cup \{\pm\infty\}$, and $\cdot$ denotes scalar product, i.e., $\rho \cdot Q = \sum_k \rho_k Q_k$. The

[8] ℓ^* is called the LEGENDRE transform of ℓ.

concept of minimal concave function is well defined since the class of concave functions is closed with respect to taking minimums. For brevity we will prove the claim only for $K = 1$, the proof is verbatim for other K.

By definition $-\ell^*(\rho) = \sup_Q \ell(Q) - \rho Q$ is the amount by which the line ρQ should be shifted so that it lies above ℓ (see Figure 4-8). Thus $\rho Q - \ell^*(\rho)$ is the line with minimum y-intercept among the lines of slope ρ that are above ℓ. The function ℓ^{**} is the lower envelope of these lines. Being a minimum of lines it is a concave function, and since all the lines were above ℓ to begin with so is ℓ^{**}.

Now suppose f is any other concave function that is above ℓ. Since f is concave, for each Q_0, there exists a ρ_0 such that for all Q

$$\ell(Q) \leq f(Q) \leq \rho_0(Q - Q_0) + f(Q_0).$$

Rearranging the terms we see that, for all Q

$$\rho_0 Q_0 - f(Q_0) \leq \rho_0 Q - \ell(Q);$$

thus $\rho_0 Q_0 - f(Q_0) \leq \ell^*(\rho_0)$, and hence

$$\ell^{**}(Q_0) \leq \rho_0 Q_0 - \ell^*(\rho_0) \leq f(Q_0),$$

proving that ℓ^{**} is indeed the minimal concave function that dominates ℓ.

As a further note, the supremum of ℓ^{**} and ℓ coincide:

$$\sup_Q \ell(Q) = \sup_Q \ell^{**}(Q).$$

This is rather easy to see since the constant function $\sup_Q \ell(Q)$ is trivially concave and dominates ℓ.

Chapter 5

Notes on the Converse

In this section we will show, by means of an example, that the inequality in (4.2) is, in general, not an equality. That is, we will find a channel P for which

$$C_{0u}(P) > \max_{Q} \min_{P' \in \wp(Q,P)} I(Q,P).$$

Upon GALLAGER's suggestion [Gal91a] we will search for an example within the class of "almost noiseless" channels. The class consists of channels that have the same number K of outputs as they have inputs, and have transition probabilities

$$P_{jk} = \begin{cases} 1 - \epsilon_{kk} & k = j \\ \epsilon_{jk} & \text{else.} \end{cases}$$

We will assume that for a set E of ordered input–output pairs, ϵ_{jk}, for $j \neq k$ satisfies

$$\begin{aligned} 0 < \epsilon_{jk} < \epsilon, &\quad \text{for } (k,j) \in E, \\ \epsilon_{jk} = 0, &\quad \text{else.} \end{aligned}$$

By convention we will include 'diagonal' elements of the form (k,k) in this set E, i.e., E will the set of (k,j) pairs such that

$$E = \{(k,j) : P_{jk} > 0\}.$$

We will be interested in computing C_{0u} in the limiting case as ϵ approaches 0 (and hence the name "almost noiseless"). In this limit $\sum_k Q_k P_{jk} = Q_j$, i.e, the output probabilities are the same as the input probabilities. Note that this limiting case is completely characterized by the directed graph $G = (V, E)$, where $V = \{0, 1, \ldots, K-1\}$ is the set of inputs (identical to the set of outputs) and E is the set of pairs described above. Figure 5-1 shows a few examples of channels and their associated graphs. In drawing these graphs, we will omit the self loops.

Suppose now that we are given $U \subset V$ such that the induced graph (U, F) contains no cycles. This means that if we restrict the input set to U, the aforementioned theorem of PINSKER and SHEVERDYAEV [PS70] applies, and as $\epsilon \downarrow 0$ we get

$$C_{0u} \geq \ln |U|.$$

Maximizing the cardinality of U, we can write

$$C_{0u} \geq \ln \max_{\substack{U \subset V: (U, E(U)) \\ \text{is acyclic}}} |U|,$$

where $(U, E(U))$ denotes the induced graph, i.e., $E(U) = E \cap (U \times U)$. We can sharpen the bound by applying it to a channel that denotes the n-fold use of the original one. Let

$$E^n = \{(\mathbf{x}, \mathbf{y}) : \mathbf{x}, \mathbf{y} \in V^n,\ P(\mathbf{y}|\mathbf{x}) > 0\} = \{(\mathbf{x}, \mathbf{y}) : \forall i\ (x_i, y_i) \in E\}.$$

Also define $E^n(U) = E^n \cap (U \times U)$. Then we have

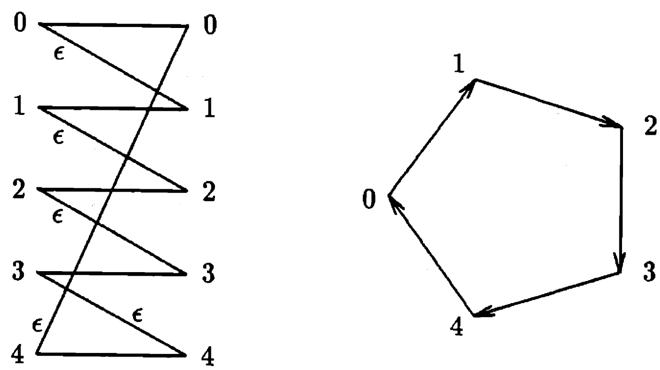
$$C_{0u} \geq \frac{1}{n} \ln \max_{\substack{U \subset V^n: (U, E^n(U)) \\ \text{is acyclic}}} |U|. \tag{5.1}$$

Furthermore, C_{0u} is the limit superior[1] of the right hand side as $n \to \infty$. To see that, suppose $\{\mathbf{x}_1, \ldots, \mathbf{x}_M\}$ is the set of codewords of a zero-error block code with block length n. Viewed as a subgraph of V^n, this set must not only be acyclic, but each node must be isolated from the others. If not, assume there is a directed link $\mathbf{x}_l \to \mathbf{x}_m$. Whenever $\mathbf{x}_m$ is transmitted, the received sequence will be $\mathbf{x}_m$ with very high probability (since the channel is almost noiseless). But the decoder will not be able to decode since the possibility of $\mathbf{x}_l$ being the transmitted codeword cannot be eliminated.

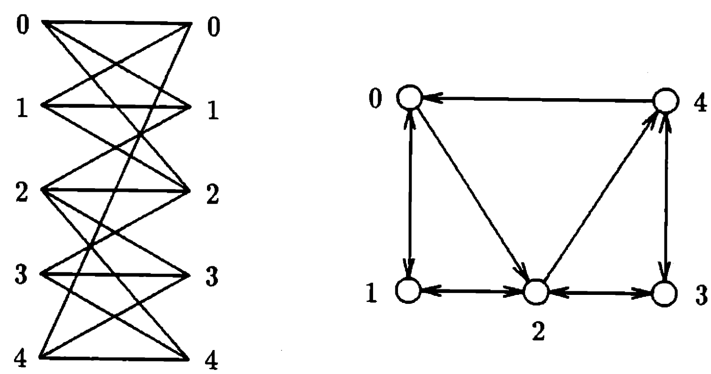
One hopes that the expression in (5.1) would be independent of n; in that case considering $n = 1$ would suffice to determine C_{0u}. However this is not the case as the following example[2] shows.

[1] In appendix 5.A on page 78 we will prove that the right hand side has a limit. This fact is not necessary for our current discussion.

[2] The example we have here is a simplified version of one due to GALLAGER [Gal91a]. His example is a graph with 7 nodes. It is not difficult to prove that 5 is the least number of nodes necessary to



(a) A 'cyclic' channel and its graph



(b) A more complicated channel and its graph

This figure shows two channels and the graphs associated with them. A directed link (k, j) appears in the graph if input k leads to output j. The first example (a) shows a 5×5 channel with input k leading to output $k + 1 \bmod 5$, which results in a cyclic graph. The second example (b) shows a more complicated channel and its associated graph. We will refer to this channel in Example 5.1 on page 74. The transition probabilities are not marked so as not to clutter the figure; it is understood that they obey the convention given in the text. Note that some of the links in the associated graph are bi-directional while some others are not.

Figure 5-1: Examples of "almost noiseless" channels and their associated graphs

Example 5.1.

Take the graph shown in Figure 5-1b. One can see that any set of three nodes has to include a cycle: if node 1 is in the set, the set cannot include nodes 0 and 2; thus it has to include nodes 3 and 4. But 3 and 4 form a cycle of length 2; thus node 1 cannot be in our acyclic set. A similar argument shows node 3 cannot be in the set either. That leaves us with nodes 0, 2 and 4, but these form a cycle too. However, the square of this graph has an acyclic component of size 5: $\{01, 14, 22, 40, 43\}$[3] (See Figure 5-2).

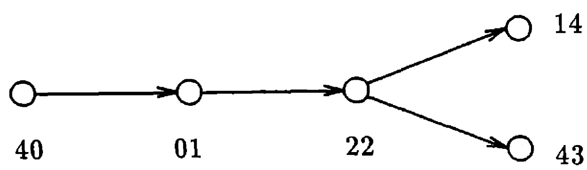


Figure 5-2: A five node acyclic subgraph of the square of the graph shown in Figure 5-1b

How does all this relate to finding an example for which

$$C_{0u} > \max_{Q} \min_{P' \in \wp(Q,P)} I(Q, P')?$$

The key to this is to relate $\max_Q \min_{P' \in \wp(Q,P)} I(Q, P')$ for an "almost noiseless" channel P and $\ln \max_{U \subset V, \text{acyclic}} |U|$ of its associated graph. If one looks at 'cyclic' channels as in Figure 5-1a, one sees that the two quantities are equal (see Appendix 5.B on page 80), and in general, the argument given in the previous paragraph establishes that

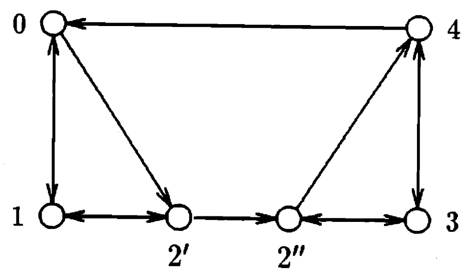
$$\max_{Q} \min_{P' \in \wp(Q,P)} I(Q, P') \geq \ln \max_{U \subset V, \text{acyclic}} |U|.$$

It is not necessary to prove the equality of the two terms in general to obtain an example; it suffices to show equality for the channel shown in Figure 5-1b. The following, due to GALLAGER [Gal91b], will do precisely that.

First consider a related channel, the graph of which is shown in Figure 5-3. Note that we have split node 2 of the original channel into two distinct nodes $2'$ and $2''$. We will refer to this channel as $\hat{P}$. First, we will show that for any input distribution Q for the

construct such an example.

[3]This is one of 24 such subsets, all of which were enumerated by a small C program. There are no acyclic components of size 6. It is interesting to note that the problem of finding the largest acyclic component of a graph is NP-complete [GJ79, p. 195].



Note that node 2 of the graph in Figure 5-1b is broken into two separate nodes, $2'$ and $2''$, and a directed between these two nodes is added.

Figure 5-3: A channel related to the one in Figure 5-1b

original channel P, there exists an input distribution $\widehat{Q}$ for this related channel $\widehat{P}$ and a $\widehat{P}' \in \wp(\widehat{Q}, \widehat{P})$ such that

$$\widehat{Q}_k = Q_k, \quad k = 0,1,3,4, \quad \widehat{Q}_{2'} + \widehat{Q}_{2''} = Q_2, \quad \text{and} \quad I(\widehat{Q}, \widehat{P}') \leq \ln 2.$$

To this end, choose $\widehat{P}''$ such that

$$\widehat{P}''_{jk} = \widehat{Q}_j/(\widehat{Q}_0 + \widehat{Q}_1 + \widehat{Q}_{2'}), \quad j,k = 0,1,2',$$

$$\widehat{P}''_{jk} = \widehat{Q}_j/(\widehat{Q}_0 + \widehat{Q}_1 + \widehat{Q}_{2''}), \quad j,k = 2'',3,4,$$

and $\widehat{P}''_{jk} = 0$ for all other j and k. We see that $I(\widehat{Q}, \widehat{P}'') = h(\widehat{Q}_0 + \widehat{Q}_1 + \widehat{Q}_{2'}) \leq \ln 2$, $\widehat{P}''\widehat{Q} = \widehat{P}\widehat{Q}$, but $\widehat{P}'' \not\ll \widehat{P}$. We will remedy that in our next step. First choose $\widehat{Q}_{2'}$ and $\widehat{Q}_{2''}$ such that $Q_2 = \widehat{Q}_{2'} + \widehat{Q}_{2''}$ and $\widehat{Q}_{2'}\widehat{P}''_{02'} = \widehat{Q}_{2''}\widehat{P}''_{42''}$, i.e.,

$$\frac{\widehat{Q}_0\widehat{Q}_{2'}}{\widehat{Q}_0 + \widehat{Q}_1 + \widehat{Q}_{2'}} = \frac{\widehat{Q}_4\widehat{Q}_{2''}}{\widehat{Q}_{2''} + \widehat{Q}_3 + \widehat{Q}_4}.$$

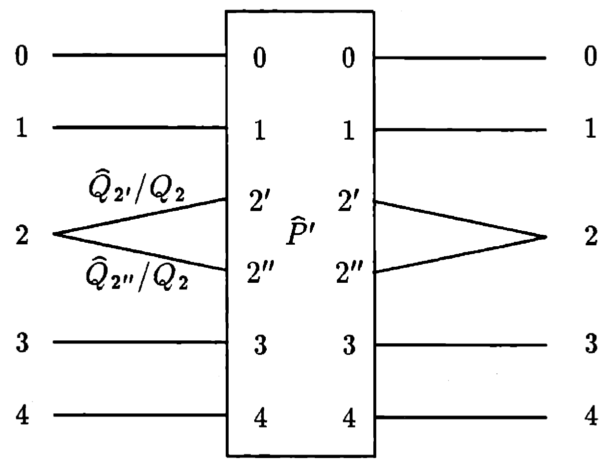
Note that a solution always exists, since the left hand side is an increasing continuous function of $\widehat{Q}_{2'}$ with the value 0 when $\widehat{Q}_{2'} = 0$, whereas the right hand side is an increasing continuous function of $\widehat{Q}_{2''}$ (decreasing in $\widehat{Q}_{2'}$ with the value 0 when $\widehat{Q}_{2''} = 0$. Now define $\widehat{P}'$ by setting

$$\widehat{P}'_{02'} = \widehat{P}'_{2''4} = 0, \quad \widehat{Q}_{2'}\widehat{P}'_{2''2'} = \widehat{Q}_4\widehat{P}'_{04} = \widehat{Q}_0\widehat{P}''_{2'0},$$

and $\hat{P}'_{jk} = \hat{P}''_{jk}$ for all other j and k. Note that $\hat{P}' \in \wp(\hat{Q}, \hat{P})$ and

$$\begin{aligned}
I(\hat{Q}, \hat{P}') &= I(\hat{Q}, \hat{P}'') + \hat{Q}_{2'}\hat{P}'_{2'2''} \ln \frac{\hat{Q}_{2'}\hat{P}'_{2'2''}}{\hat{Q}_{2'}\hat{Q}_{2''}} + \hat{Q}_4\hat{P}'_{04} \ln \frac{\hat{Q}_4\hat{P}'_{04}}{\hat{Q}_4\hat{Q}_0} \\
&\quad - \hat{Q}_{2'}\hat{P}''_{02'} \ln \frac{\hat{Q}_{2'}\hat{P}''_{02'}}{\hat{Q}_{2'}\hat{Q}_0} - \hat{Q}_4\hat{P}''_{2''4} \ln \frac{\hat{Q}_4\hat{P}''_{2''4}}{\hat{Q}_{2''}\hat{Q}_4} \\
&= I(\hat{Q}, \hat{P}'') \\
&\leq \ln 2.
\end{aligned}$$

Finally consider the following arrangement shown in Figure 5-4. The end-to-end channel,



Here we cascade 3 channels: the first splits the input 2 into 2′ and 2″, the second is the channel $\hat{P}'$ described above, and the third makes the outputs 2′ and 2″ indistinguishable. Note that if the input distribution is Q, the distribution at the input of $\hat{P}'$ is $\hat{Q}$.

Figure 5-4: A cascade of channels, with $\hat{P}'$ in the middle

P' satisfies $P' \in \wp(Q, P)$ and by the data processing theorem

$$I(Q, P') \leq I(\hat{Q}, \hat{P}') \leq \ln 2.$$

Thus for this channel, we see that

$$\max_{Q} \min_{P' \in \wp(Q,P)} I(Q, P') = \ln 2,$$

whereas, $C_{0u}(P) \geq \frac{1}{2} \ln 5 > \ln 2$.

5.A Right Hand Side of (5.1) Has a Limit

Here we prove that the right hand side of equation (5.1) has a limit. We will first prove a result of independent interest, that the product of two acyclic graphs is acyclic. Let us first define what we mean by a cycle:

Definition 5.2 *Given a graph (V, E), a sequence of nodes $v_1, \ldots, v_n$ form a cycle if:*

1. *they are all distinct, ($i \neq j$ implies $v_i \neq v_j$),*
2. *n is at least 2,*
3. *for all $1 \leq k < n$ $(v_k, v_{k+1}) \in E$, and $(v_n, v_1) \in E$.*

Theorem 5.3 *Suppose $G = (V, E)$ and $H = (W, F)$ are two acyclic graphs. Then the graph $(V \times W, E \times F)$ is also acyclic. By $E \times F$ we mean*

$$E \times F = \{((v, w), (x, y)) : (v, x) \in E,\ (w, y) \in F\}.$$

Proof. Suppose $(V \times W, E \times F)$ contains a cycle, i.e., a sequence of distinct pairs (v_k, w_k), $k = 1, \ldots, n$ such that for $1 \leq k < n$, $(v_k, v_{k+1}) \in E$, $(w_k, w_{k+1}) \in F$, $(v_n, v_1) \in E$ and $(w_n, w_1) \in F$. Consider the projection of this sequence on its first coordinate, $v_1, \ldots, v_n$. If $v_1 = v_2 = \cdots = v_n$, then the sequence $w_1, \ldots, w_n$ forms a cycle in (W, F) and we arrive at a contradiction. Thus suppose that not all v_k's are equal. However, the sequence $v_1, \ldots, v_n$ may contain runs: l and k for which $v_l = v_{l+1} = \cdots = v_{l+k}$. Remove these runs by retaining only one term of each run. Let the resulting sequence be $u_1, \ldots, u_m$[4]. We see that $m > 1$; for $1 \leq k < m$, $(u_k, u_{k+1}) \in E$, and $(u_m, u_1) \in E$. If $u_1, \ldots, u_m$ are all distinct, then they form a cycle in (V, E); otherwise let

$$L = \sup\{k : u_1, \ldots, u_k \text{ are all distinct,}\},$$

and

$$l = \inf\{k : u_k = u_{L+1}\}.$$

[4]For the purist, define a function $r : \mathbb{Z}^+ \to \mathbb{Z}^+$ recursively as

$$\begin{aligned} r(1) &= 1, \\ \text{and for } k > 1, \quad r(k) &= \begin{cases} \inf\{l : r(k-1) < l \leq n,\ v_l \neq v_{l-1}\} & \text{if the set is not empty} \\ n+1 & \text{else.} \end{cases} \end{aligned}$$

Let $m = \sup\{k : r(k) \leq n\}$, and let $u_k = v_{r(k)}$ for $k = 1, \ldots, m$.

Note that $1 < L < m$ and $l \leq L$. Furthermore, $l \neq L$, since $u_L \neq u_{L+1}$ by construction. Thus $u_l, \ldots, u_L$ are all distinct, and form a cycle in (V, E) leading to a contradiction. □

Now let

$$a_n = \frac{1}{n} \ln \max_{\substack{U \subset V^n : (U, E^n(U)) \\ \text{is acyclic}}} |U|.$$

We see that $0 \leq a_n \leq \ln |V|$, and, by the theorem above, for any n and m,

$$(n+m)a_{n+m} \geq n a_n + m a_m, \tag{5.2}$$

for if $U_n \subset V^n$ and $U_m \subset V^m$ are maximal acyclic subgraphs, then $U_n \times U_m$ is an acyclic subgraph of V^{n+m}.

Theorem 5.4 *The sequence a_n, $n = 1, 2, \ldots$, has a limit.*

Proof. By equation (5.2) we see that the subsequence $a_{2^k}, k = 0, 1, \ldots$, is non-decreasing and bounded, and thus has a limit a, with $0 \leq a \leq \ln |V|$.

Given $0 < \epsilon < 1$, choose M such that $a_{2^k} \geq a - \epsilon/(1+a)$ for all $k \geq M$.

Note that any N can be written as $N = \sum_{k \geq 0} 2^k b_k$ with $b_k \in \{0, 1\}$. Furthermore only finitely many terms of this sum will be non-zero. For $N > N(\epsilon) = 2^M (1+a)/\epsilon$ we see that

$$N a_N \geq \sum_{k \geq 0} 2^k b_k a_{2^k} \geq \sum_{k \geq M} 2^k b_k a_{2^k} \geq (N - 2^M) a_{2^M} \geq (N - 2^M)(a - \epsilon/(1+a)),$$

thus

$$a_N \geq (1 - \epsilon/(1+a))(a - \epsilon/(1+a)) \geq a - \epsilon.$$

To prove an inequality in the opposite direction, choose K such that $2N \leq 2^K \leq 4N$. Then

$$2^K a \geq 2^K a_{2^K} \geq N a_N + (2^K - N) a_{2^K - N} \geq N a_N + (2^K - N)(a - \epsilon),$$

$$2^K \epsilon + N a \geq N a_N,$$

$$a + 4\epsilon \geq a_N.$$

Thus, we see that a is the limit of the sequence a_n, $n = 1, 2, \ldots$. □

5.B Case of the 'Cyclic' Channels

Now consider a channel with $K > 1$ inputs and K outputs with transition probabilities

$$P_{kj} = \begin{cases} 1-\epsilon & k = j \\ \epsilon & j - k = 1 \bmod K \\ 0 & \text{else.} \end{cases}$$

As ϵ approaches zero from above, its value shows up only in the $\widehat{P}' \ll P$ constraint but does not change the output probabilities, i.e., the probability of output j will be equal to the probability of input j.

If the input probabilities are given by Q_k, $k = 0, \ldots, K-1$, the constraint on $\widehat{P}'$ will be

$$\begin{aligned} Q_0\widehat{P}'_{00} + Q_{K-1}(1 - \widehat{P}'_{K-1,K-1}) &= Q_0 \\ Q_1\widehat{P}'_{11} + Q_0(1 - \widehat{P}'_{00}) &= Q_1 \\ &\vdots \\ Q_k\widehat{P}'_{kk} + Q_{k-1}(1 - \widehat{P}'_{k-1,k-1}) &= Q_k \\ &\vdots \\ Q_{K-1}\widehat{P}'_{K-1,K-1} + Q_{K-2}(1 - \widehat{P}'_{K-2,K-2}) &= Q_{K-1} \end{aligned}$$

Leading to

$$Q_0(1 - P_{00}) = \cdots = Q_k(1 - \widehat{P}'_{kk}) = \cdots = Q_{K-1}(1 - \widehat{P}'_{K-1,K-1}) = c,$$

where c satisfies $0 \leq c \leq \min_k Q_k$. Thus we see that

$$\widehat{P}'_{kk} = 1 - c/Q_k$$

and

$$I(Q, \widehat{P}') = \sum_k \left[2Q_k \ln \frac{1}{Q_k} + (Q_k - c)\ln(Q_k - c) + c \ln c \right].$$

Taking the derivative of this expression with respect to c and setting equal to 0 yields

$$\sum_k \ln\left(\frac{Q_k}{c} - 1\right) = 0, \tag{5.3}$$

or equivalently

$$\sum_k \ln(Q_k - c) = \sum_k \ln c.$$

Substituting this equivalent form in the expression for $I(Q, \widehat{P}')$, we get

$$I = \min_c I(Q, \widehat{P}') = \sum_k Q_k \ln \frac{Q_k - c}{Q_k^2}$$

with c satisfying (5.3). Now consider

$$I - \ln(K-1) = \sum_k Q_k \ln \frac{Q_k - c}{Q_k^2(K-1)} \leq \sum_k \frac{1 - c/Q_k}{K-1} - 1.$$

If we can show that $\sum_k c/Q_k \geq 1$ for c satisfying (5.3), then the last quantity in the equation above is less than zero, and hence

$$I \leq \ln(K-1).$$

This last step is equivalent to showing that

$$\sum_k \log \alpha_k = 0 \Rightarrow \sum_k \frac{1}{1 + \alpha_k} \geq 1,$$

with $\alpha_k = (Q_k/c - 1)$. To this end, let us minimize $\sum_k [1 + \alpha_k]^{-1}$ subject to $\sum_k \ln \alpha_k = 0$, i.e, $\prod_k \alpha_k = 1$. The minimum occurs when one of the α_k's approaches zero while the others approach infinity. The value of the infimum is 1, and hence the result follows.

Chapter 6

Conclusion

In the first part of this thesis (Chapters 2 and 3) we described a multi-access communication system and presented a method of analysis that accurately models both the message arrival process and the transmission, and shows the tradeoff between such quantities as error probability and queueing delay. However, one should note that, we are not making any claims that the multi-access scheme proposed is in any way superior to the current multi-access schemes. In fact, it is clear from Figure 2-1 and the discussion following it, that our multi-access scheme is limited to 1 nat per second per Hertz of throughput regardless of the signal to noise ratio. Thus, schemes that are derivatives of ALOHA, employing a scheduling of transmissions, will exceed the performance of our system at high signal to noise ratios. This, however, is not because of the unscheduled nature of the transmissions in our multi-access system, rather, it is because of the independent decoding of the messages. A true multi-access decoder would employ joint decoding of the transmitters and perform better than the system we have analyzed. Indeed, an analysis akin to that given at the end of Chapter 2 indicates that regardless of the signal to noise ratio, the throughput of the system is unbounded. However, the analysis that leads to this result is based on the same heuristic as in Chapter 2 and its predictions should be taken with a grain of salt.

We note the following as problems to research further in connection with this part of the thesis.

- Recall that our analysis assumes that all the transmitters have equal power. To relax that assumption and still make use of the techniques we have developed entails the solution to the processor sharing problem with multiple classes of customers. Our attempts to tackle this problem have so far failed.

- We believe that the analysis can be applied to transmission models other than additive white GAUSSian noise channels. As long as one can discretize the time as we have done in Section 3.1.1 the only modification in the analysis is to the quantity E_0.

- The reduction of the waveform channel to a sequence of scalar channels is done in a somewhat clumsy way, since it involves non-causal and infinite duration signals. At this point we do not see any clean way of getting around this problem, since it is inherent in bandlimited systems.

In the second part of the thesis (Chapters 4 and 5) we focused on an information theoretic problem, namely, the performance of errors and erasures decoding schemes over discrete memoryless channels. We derived new and stronger upper bounds to error and erasure probabilities, and a new and stronger lower bound to the zero undetected error capacity. We also showed that the bounds are not equal to the actual quantities. Again there are several open problems, all seemingly difficult:

- What is the zero undetected error capacity of discrete memoryless channels? Even though we have shown that for some channels our bound is exact, there is no known closed form expression to compute this quantity. It is also clear that one gets closer and closer to the capacity by considering the bounds for the n-fold channel derived from the original. One question is whether one achieves the capacity at some finite n.

- One can describe a multi-access analog of the errors and erasures schemes described in Chapter 4. This will differ from the multi-access model described in the first chapters of the thesis, because the new model will not consider random message arrivals, and the channel will be a general discrete memoryless channel. One can calculate upper bounds to error and erasure probabilities using a single-letter random coding argument. However the results so obtained would not be tight, since they would be the analogs of FORNEY's bounds. If one can generalize the fixed composition type arguments to this multi-access case, then one can derive stronger upper bounds to error and erasure probabilities. At present we do not know how to do this generalization.

Bibliography

[CK81] Imre Csiszár and Janos Körner. *Information Theory: Coding Theorems for Discrete Memoryless Systems.* Academic Press, New York, 1981.

[For68] G. David Forney. Exponential error bounds for erasure, list and decision feedback schemes. *IEEE Transactions on Information Theory*, IT-14(2):206–220, March 1968.

[Fri82] Avner Friedman. *Foundations of Modern Analysis.* Dover Publications, Inc., New York, 1982.

[Gal68] Robert G. Gallager. *Information Theory and Reliable Communication.* John Wiley & Sons, New York, 1968.

[Gal85] Robert G. Gallager. A perspective on multiaccess channels. *IEEE Transactions on Information Theory*, IT-31(2):124–142, March 1985.

[Gal91a] Robert G. Gallager. Private Communication, August 1991.

[Gal91b] Robert G. Gallager. Private Communication, September 1991.

[GJ79] Michael R. Garey and David S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness.* W. H. Freeman and Company, New York, 1979.

[Hal57] Paul Richard Halmos. *Introduction to Hilbert Space and the Theory of Spectral Multiplicity.* Chelsea Publishing Company, New York, second edition, 1957.

[Hal74] Paul Richard Halmos. *Measure Theory.* Springer-Verlag, New York, 1974.

[Kel79] Frank P. Kelly. *Reversibility and Stochastic Networks.* Wiley series in probability and mathematical statistics. John Wiley & Sons, New York, 1979.

[PS70] Mark S. Pinsker and A. Yu. Sheverdyaev. Transmission capacity with zero error and erasure. *Problems in Information Transmission (Problemy Peredachi Informatsii)*, 6(1):20–24, 1970.

[Ros83] Sheldon M. Ross. *Stochastic Processes.* John Wiley & Sons, New York, 1983.

[RSN90] Frigyes Riesz and Béla Szökefalvi-Nagy. *Functional Analysis.* Dover Publications, Inc., New York, 1990. Translation of Leçons d'analyse fonctionelle by Leo F. Boron. An unabridged publication of the work first published by Frederick Ungar Publishing Co., New York, in 1955.

[Sha56] Claude E. Shannon. The zero error capacity of a noisy channel. *IRE Transactions on Information Theory*, IT-2:8–19, September 1956.

[TG89] İ. Emre Telatar and Robert G. Gallager. Zero error decision feedback capacity of discrete memoryless channels. Technical Report LIDS–P–1927, Massachusetts Institute of Technology, Laboratory for Information and Decision Systems, November 1989.

[Yat90] Roy D. Yates. *High Speed Round Robin Queueing Networks.* PhD thesis, Massachusetts Institute of Technology, Department of Electrical Engineering and Computer Science, Cambridge, MA, 1990.

Index