CHANNEL STATE TESTING IN INFORMATION DECODING

by

Howard L. Yudkin

B.S.E.E., University of Pennsylvania (1957)

S.M., Massachusetts Institute of Technology (1959)

SUBMITTED IN PARTIAL FULFILLMENT OF THE

REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September, 1964

Signature of Author_____
Department of Electrical Engineering, September 1964

Certified by_____
Thesis Supervisor

Accepted by_____
Chairman, Department Committee on Graduate Students

CHANNEL STATE TESTING IN INFORMATION DECODING

by

HOWARD L. YUDKIN

Submitted to the Department of Electrical Engineering on September 1964 in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

ABSTRACT

A study is made of both block and sequential decoding methods for a class of channels called Discrete Finite State Channels. These channels have the property that the statistical relations between input and output symbols are determined by an underlying Markov chain whose statistics are independent of the input symbols.

A class of (non-maximum likelihood) block decoders is discussed and a particular decoder is analyzed. This decoder has the property that it attempts to probabilistically decode by testing every possible combination of transmitted code word and channel state sequence. An upper bound on error probability for this decoder is found by random coding arguments. The bound obtained decays exponentially with block length for rates smaller than a capacity of the decoding method. The bound is cast in a form so that easy comparison may be made with the corresponding results for the Discrete Memoryless Channel.

A related sequential decoder based on a modification of Fano's decoder is presented and analyzed. It is shown that R_{comp} is equal to the block coding error exponent at zero rate for an appropriate subclass of Discrete Finite State Channels. It is also shown that for this class, the probability of decoding failure for low rates is the probability of error for the block decoding technique presented here.

All results may be specialized to the case of Discrete Memoryless Channels. Some of the results on behavior of the sequential decoding algorithm were not previously available for this case.

Thesis Supervisor: Robert M. Fano
Title: Ford Professor of Engineering

Acknowledgements

It is my pleasure to acknowledge the contributions made to this thesis by my supervisor Professor R.M. Fano and my readers Professors J.M. Wozencraft and R.G. Gallager. A reading of the contents of the thesis will reveal my obvious debt to their ideas and prior work. My association with them has been the most valuable part of my graduate education.

I wish to thank the M.I.T. Lincoln Laboratory for the financial support tendered me under their Staff Associate program. In addition I wish to thank the M.I.T. Research Laboratory of Electronics for the facilities provided to me.

Finally, let me thank my wife, Judith, and my parents for the encouragement and support which they gave me during the extent of my doctoral program.

Table of Contents

List of Figures

Chapter I

Introduction

Most of the results pertaining to the reliability which may be achieved when data are transmitted over a channel have been obtained for the special case of the Discrete Memoryless Channel (DMC). Recent work of Fano[4], Gallager[8], and Shannon, Gallager and Berlekamp[19] has led to an almost complete specification of the smallest probability of error obtainable with maximum likelihood decoding of block codes for the DMC.

Collaterally, the investigation of practical decoding techniques for the DMC has led to the design, construction and testing[22,23] of a sequential decoder based on the sequential decoding technique of Wozencraft[21]. More recently, Fano[5] has presented a new sequential decoder which appears to have great generality of application.

Our purpose in this thesis is to examine decoding techniques for channels that are not of the Discrete Memoryless variety. The channels with which we are concerned are such that at each discrete instant of time one of a finite set of symbols may be transmitted. One of a finite set of output symbols will then be received. The probability that a particular symbol

is received when a particular symbol is transmitted is a function whose value is determined by an underlying finite state stochastic process which is independent of the transmitted symbols. The aspect of memory is introduced by requiring that the probability that the underlying process is in a particular state at a given time is dependent on the sequence of states which the process has occupied in the past. In particular, we will restrict this dependence to be Markovian, which (since we are concerned with finite state processes) is equivalent to allowing the dependence to be over any finite span of previous states! A more careful description of the Channels is presented in Chapter II where appropriate notation is introduced.

A discussion of the broadness of the above model and some of its implications is also presented in Chapter II. We shall call this class of channels, Discrete Finite State Channels (DFSC); sequences of states of the underlying process will be called channel state sequences.

In the following chapters we will examine both block and sequential decoding for the DFSC. The departure in philosophy taken here is that we attempt to decode by probabilistically testing both the

transmitted message and the channel states, rather than the transmitted message alone. Our primary interest is, of course, in the correctness of our decisions on the transmitted messages. The method of testing the compound hypotheses (both message and channel state), however, appears to be natural for sequential decoding. The reason for this statement lies in the fact that the joint statistics of the output, and channel state, given a particular input, are Markovian, while the statistics of the output, alone, are not. By testing both the transmitted message and the channel state we are able to design a sequential decoder which operates in a step-by-step fashion closely related to the operation of such decoders for the DMC. Our ability to achieve such a design is a consequence of the Markovian statistics of the joint event (output and channel state).

A thorough discussion of the particulars of our decoding philosophy is presented in the next chapter. We also discuss, briefly, several alternative approaches to decoding which are suggested by the fact that the DFSC might be described as a time-varying channel. These alternative approaches are those that have arisen when, in engineering

practice, one considers what might be done to improve communication capability of such channels.

To operate in accordance with the above philosophy we must assume that the transmitter and decoder have an explicit probabilistic description of the underlying process. This assumption may be questioned. We observe that this assumption is no worse than the assumption that the probability structure of a given memoryless channel is known. Experience in simulation of the DMC has shown that if the true probabilistic structure of the channel is at all like the assumed structure, then the decoding will behave essentially as predicted theoretically (c.f, Horstein [11]). We should expect the same to be true in the case at hand. In addition, knowledge of the behavior of decoding when the probabilistic description of the channel is known makes available a bound to what might be achieved in practice.

The DFSC fits within the class of channels for which Blackwell, et. al.[1] have investigated capacity. In addition, Kennedy[13] has presented upper and lower bounds to the probability of error achievable with block coding for binary input, binary output DFSC's. Aside from these results and the previously referenced discussions of the DMC, no previous work of relevance to the DFSC appears to be in the literature!

In Chapter II we present a mathematical description of the DFSC and discuss the problem of decoding for this class of channels.

In Chapter III we present various mathematical results which will be applied in the sequel.

In Chapter IV we find an upper bound to the probability of error which can be achieved by block coding for the DFSC when the method of simultaneously testing transmitted information and channel states is employed. A bound which decays exponentially with the block length is found and compared to known results for the DMC.

In Chapter V we examine the behavior of the Fano sequential decoder when used on a DFSC. The results obtained here on maximum information rate for which the first moment of computation is bounded and for various probabilities of error and failure may be specialized to the DMC. Certain of these results for the DMC were previously found by Fano[6]. Certain others have been obtained independently by Stiglitz (unpublished). The results for the DFSC have not been previously obtained.

In Chapter VI we summarize the thesis and suggest and discuss various possible extensions.

Most of the mathematical expressions, equations, and inequalities are numbered in succession in each chapter. For convenience, we will refer to all such expressions as equations. When referencing a previous equation in the same chapter we give its number. When referencing such an equation in a previous chapter we give both the chapter number and the number of the equation. Thus for example, if in Chapter III we wish to refer to equation 2 of that chapter, we call it Equation (2). If, on the other hand, we wish to refer to equation 4 of Chapter II, we call it Equation (2.4).

Chapter II

Introduction to Decoding for the DFSC

A. Description of Channels

We will be concerned with a class of channels where at each discrete instant of time one of a set of K inputs, $x \in X$, $(x=1,2,\ldots,K)$ may be transmitted and one of a set of L outputs $y \in Y$, $(y=1,2,\ldots,L)$ will be received. The probability that output y is received when input x is transmitted is determined as follows:

Suppose we have a B state Markov chain with states $d \in D$, $(d=1,2,\ldots,B)$ and a stationary (i.e., time-invarient) probability matrix $Q = (q_{ij})$ where $q_{ij}(i,j = 1,2,\ldots,B)$ is the probability that when the chain is in state i, the next transition will be to state j. In addition, let there be a set of B^2 probability functions, $p(y/x,d',d)$ defined for all $y \in Y$, $x \in X$ and $d',d \in D$ with the property that:

$$p(y/x,d',d) \geq 0 \quad ; \text{ all } y,x,d',d \qquad (1)$$

and

$$\sum_{Y} p(y/x,d',d) = 1 \quad ; \text{ all } x,d',d \qquad (2)$$

Suppose now that at some time the Markov chain is in state d' and a transition is made to state d, then conditional on this event, the probability that y is received when x is transmitted is $p(y/x,d',d)$. Thus for fixed d', d we may view $p(y/x,d',d)$ as the trans-

ition probability function for a fixed channel.

The aggregate of the Markov chain and the set of functions $p(y/x,d',d)$ will be called a Discrete Finite State Channel (DFSC). We will call the functions $p(y/x,d',d)$ transmission probability functions, and sequences of states from the Markov chain will be called channel state sequences. In this thesis we will restrict ourselves to the case in which the underlying process (i.e., the Markov chain) is irreducible.

Let us pause for a moment and consider the generality of this definition. Although we have defined the transmission probability functions $p(y/x,d',d)$ on the state transitions, we have clearly included the case in which it is desirable to define these functions on the states. To demonstrate this inclusion we need only observe that if we allow $p(y/x,d',d)$ to be independent of d' (or d) our functions are then defined on the states.

Another model which might be considered is the following: Let there be a set of A probability functions $p(y/x,c)$; $(c=1,2,\ldots,A)$. These functions determine the probability of receiving a given output when a given input is transmitted, for the event c occurring. Further, let there be a set of B^2 probability functions $H_{d',d}(c)$; $(d',d=1,2,\ldots,B)$ with

$$H_{d',d}(c) \geq 0 \quad ; \quad \sum_{c=1}^{A} H_{d',d}(c) = 1 \qquad (3)$$

where $H_{d',d}(c)$ is the probability that, when a transition of the Markov chain from state d' to state d takes place, the transmission probability function which determines the input-output statistics is $p(y/x,c)$. The resulting situation may be modelled as a DFSC in either of two ways.

First, each state, d, of the chain may be split into A states, $d_1, d_2, \ldots, d_A$, one for each value of c. For the resulting model we then have:

$$\Pr(d_c / d'_{c'}) = H_{d',d}(c)\, q_{d',d} \qquad (4)$$

and

$$p(y/x,d'_{c'},d_c) = p(y/x,c) \qquad (5)$$

A second alternative is to retain the original description of the chain and take:

$$p(y/x,d',d) = \sum_{c=1}^{A} p(y/x,c)\, H_{d',d}(c) \qquad (6)$$

where we observe that the above equation defines a valid transmission probability function.

We shall find that because of the decoders employed for the DFSC as discusssed in later sections of the chapter, and because of the techniques used to

bound the behavior of these decoders, it is generally desirable to model the channel with as small a number of states in the underlying Markov chain as is possible. For this reason, the second alternative discussed above is adopted when we have such a choice available.

To illustrate further the multiplicity of models which may be used to model a DFSC we consider the special example of a memoryless Binary Symmetric Channel. We may, of course, use the one-state model of the channel as is the usual choice. (Note here that a memoryless channel may always be taken as a DFSC with a single state in the underlying Markov chain). We may also choose a model in which we associate the transmission probability functions with states. We distinguish two types of states, a 0 state and a 1 state with transmission probability functions as shown in Figure 2.1. We may then take any of the models shown in Figure 2.2. Each model clearly is equivalent to a BSC with cross-over probability p. This particular example is of great interest since it allows us to discuss certain deficiencies of our decoders. We will return to this matter in Chapter IV.

To denote sequences of random variables we will use the symbol for the random variable underlined and with a symbol in parentheses indicating the number

p(y/x) for a "0" state

x \ y	1	2
1	1	0
2	0	1

p(y/x) for a "1" state

x \ y	1	2
1	0	1
2	1	0

Figure 2.1 Transmission Probability Functions for "0" States and "1" States

a)

$$Q = \begin{pmatrix} 1-p & p \\ 1-p & p \end{pmatrix} \qquad p < \tfrac{1}{2}$$

"0" state "1" state

A two state model

b)

$$Q = \begin{pmatrix} \frac{1-p}{2} & \frac{1-p}{2} & p/2 & p/2 \\ \frac{1-p}{2} & \frac{1-p}{2} & p/2 & p/2 \\ \frac{1-p}{2} & \frac{1-p}{2} & p/2 & p/2 \\ \frac{1-p}{2} & \frac{1-p}{2} & p/2 & p/2 \end{pmatrix}$$

"0" states "1" states

A four state model

c)

$$Q = \begin{pmatrix} \frac{1-p}{m} & \frac{1-p}{m} & \cdots & \frac{1-p}{m} & \frac{p}{m} & \cdots & p/m & p/m \\ \frac{1-p}{m} & \frac{1-p}{m} & \cdots & \frac{1-p}{m} & \frac{p}{m} & \cdots & p/m & p/m \\ \cdot & \cdot & \cdots & \cdot & \cdot & \cdots & \cdot & \cdot \\ \cdot & \cdot & \cdots & \cdot & \cdot & \cdots & \cdot & \cdot \\ \frac{1-p}{m} & \frac{1-p}{m} & \cdots & \frac{1-p}{m} & \frac{p}{m} & \cdots & p/m & p/m \end{pmatrix}$$

"0" states "1" states

A 2m state model

Figure 2.2 Alternate Models for a BSC

of elements in the sequence. Thus a sequence of n channel inputs will be denoted by $\underline{x}(n)$. The set of all such sequences, which is the n-fold cartesian product of the set of all values of the basic one-dimensional random variable will be denoted by a superscript on the symbol for the one-dimensional set. Thus we speak of $\underline{x}(n) \in X^n$. The position of a particular element of a sequence will be denoted by a subscript. Then, making the obvious analogy of sequences to vectors and elements to components of the vector, we write:

$$\underline{x}(n) = (x_1, x_2, \ldots, x_n) \tag{7}$$

One exception to this rule is that for channel state sequences we will speak of $\underline{d}(n) \in D^n$,

$$\underline{d}(n) = (d_o, d_1, \ldots, d_n) \tag{8}$$

which is actually in D^{n+1}. The reason for this convention is the simplification of notation and this convention should be remembered since it is in continuous use in the sequel. The inclusion of d_o specifies the initial state. As an additional notational convenience we introduce the symbol $\hat{d}_i$,

$$\hat{d}_i = (d_{i-1}, d_i) \tag{9}$$

The notation of equation (7) suggests that we interpret $\underline{x}(n)$ as a row vector. We will use an overbar to denote matrix transposition. Thus $\overline{\underline{x}}(n)$ is a column vector.

The standard introductory reference on Markov chains is Feller[7]. The algebraic treatment of Markov chains in terms of Frobenius's theory of matrices with non-negative elements is given in detail in Gantmacher[9]. An excellent discussion incorporating aspects of both Feller's and Gantmacher's treatments is given by Rosenblatt[17].

B. Block Decoding for the DFSC

We now begin our study of decoding for the DFSC. The situation of block coding is dealt with initially because it is inherently simpler to discuss than sequential decoding.

We wish to transmit one of $M = e^{nR}$ equally likely messages over a DFSC. To do so, we select a set of M channel input sequences $\underline{x}_m(n)$; m=1,2,...,M, and transmit sequence $\underline{x}_m(n)$ to signify that message m occurred at the transmitter.

Upon receipt of the output sequence $\underline{y}(n)$, we attempt to guess which message was transmitted. The best guess, in the sense that it would minimize our probability of error, would be that given by a maximum likelihood decoding scheme. In this case we decide that message k was transmitted if

$$\Pr(\underline{y}(n) \,/\, \underline{x}_k(n)) = \max_m \Pr(\underline{y}(n) \,/\, \underline{x}_m(n)) \quad (10)$$

The probability of error for such a decoding scheme is not readily analyzed for the DFSC, but in principle we may always perform maximum likelihood decoding. The result we expect to obtain when properly chosen block codes are employed on the DFSC is that for rates, R, less than some yet to be determined capacity we are able, by increasing n, the block length, to make the achievable error probability arbitrarily small.

The difficulty that arises, when we attempt to analyze block coding bounds on error probability for maximum likelihood decoding, is that an early step in our derivation of a bound reduces the sharpness of the bound to the point that it is equivalent to a bound on the behavior of the non-maximum likelihood decoder which we ultimately study.

How does one decode for the DFSC? Experience with time-varying channels in general has led various investigators to suggest schemes based on heuristic reasoning. One such scheme may be described as follows: From the received data make an estimate of the channel state sequence. Then, assuming that this estimate is correct, do maximum likelihood decoding as if this assumption were correct. This scheme is embodied physically in such systems as Rake[14] and in systems which utilize techniques of phase estimation and coherent demodulation with the

estimated phase for channels with a time-varying phase shift. This latter scheme is analyzed in some detail by Van Trees[23]. Although these examples apply to continuous channels, the philosophy of approach is clearly applicable to the case of the DFSC. The aspect of these schemes which make them attractive is that for the particular situation for which they are intended, they are readily instrumented in practice while maximum likelihood techniques are not. Both schemes show the following deficiency: The estimate of the channel state is made independently of any hypothesis on the transmitted information. This factor may or may not be bad. Whether it is or not depends on the complex of the rate of transmission, the nature of the particular channel at hand, the choice of modulation, and the interactions among these.

Now consider how such schemes may be applied to the DFSC. We have some rational for deciding that a particular channel state sequence $\underline{d}^*(n)$ has occurred. Then, assuming this decision is correct, we compute: $\Pr(\underline{y}(n) / \underline{x}_k(n), \underline{d}^*(n))$ for each $k = 1,2,\ldots,M$. We then decide that message m was transmitted if:

$$\Pr(\underline{y}(n) / \underline{x}_m(n), \underline{d}^*(n)) = \max_k \Pr(\underline{y}(n) / \underline{x}_k(n), \underline{d}^*(n)) \quad (11)$$

The behavior of such a decoder clearly depends on the method of choosing $\underline{d}^*(n)$. Such methods arise from what amounts to good intuition applied to the particular case at hand. Since we are interested in a broad class of situations, it is unlikely that such intuition could be applied in general. A way out is described below.

Suppose we broaden our approach to include joint estimation of both the channel state sequence which occurs and the transmitted message. We are then not forcing ourselves to decide on the channel state sequence first. Of course, as in the examples discussed above, our primary interest lies in making our decisions on the transmitted message correct. The penalty we pay for being wrong on the channel state sequence is zero if we are right on the transmitted message.

This concept of joint estimation arises in an interpretation of the maximum likelihood recievers for gaussian signals in gaussian noise (see Kailath[12] and Turin[20]). In this case the receivers may be realized in a form in which an estimate is made of the shape of the gaussian signal conditional on the transmitted message having been a particular one. This estimated shape is then used as a reference for a correlation receiver for that particular message. One such estimate and correlation operation is performed for <u>each</u> different transmitted message hypothesis.

A class of decoders may now be thought of immediately. We may for example consider the function $Pr(\underline{y}(n)/\underline{x}_k(n),d(n))$ for all values of both $\underline{d}(n)$ and k. The decoding rule could then be: choose message m as transmitted if

$$\max_{\underline{d}(n)} Pr(\underline{y}(n)/\underline{x}_m(n),\underline{d}(n)) = \max_k \max_{\underline{d}(n)} Pr(\underline{y}(n)/\underline{x}_k(n),\underline{d}(n)) \tag{12}$$

An objection to this decoder which might be raised is that for a particular message which is not the transmitted message, there might be a particular channel state sequence $\underline{d}^*(n)$ such that $Pr(\underline{y}(n)/\underline{x}(n),\underline{d}^*(n))$ is very large.

There are at least two ways of avoiding this unhappy situation. First, by appropriate choice of modulation (i.e., the choice of the $\underline{x}_k(n)$'s) we might be able to avoid the possibility of this occurrence. Again, such a choice is to be found by applying good intuition to the particular case at hand.

A second alternative lies in weighting the probabilities in Equation (12) by a factor which takes into account how probable any sequence $\underline{d}(n)$ is a priori. We may, for example, take a binary weight and assign weight 1 to those channel state sequences whose probability exceeds a given threshold, (say p_o) and weight 0 to the remainder.

Thus if we let D_o be a set such that:

$$D_o = \left\{ \underline{d}(n) \;\middle|\; \Pr(\underline{d}(n)) \geq p_o \right\} \tag{13}$$

and let D_o^c be the complement of this set we might formulate a decoding rule as follows: Pick message m as transmitted if:

$$\max_{\underline{d}(n) \in D_o} \Pr(\underline{y}(n)/\underline{x}_m(n), \underline{d}(n))$$

$$= \max_k \max_{\underline{d}(n) \in D_o} \Pr(\underline{y}(n)/\underline{x}_k(n), \underline{d}(n)) \tag{14}$$

An upper bound on the probability of error for such a decoder can be found, but it is not presented here because it is weaker than the bound for the decoder we do analyze.

The idea of weighting the probabilities in Equation (14) can be extended to the logical conclusion of using as weights the actual a priori probabilities of the state sequences. Thus we are led to the decoder to be employed in this thesis. Our decoding rule is stated as follows:

Choose message m as transmitted if:

$$\max_{\underline{d}(n)} \Pr(\underline{y}(n)/\underline{x}_m(n), \underline{d}(n)) \Pr(\underline{d}(n))$$

$$= \max_k \max_{\underline{d}(n)} \Pr(\underline{y}(n)/\underline{x}_k(n), \underline{d}(n)) \Pr(\underline{d}(n)) \tag{15}$$

Now, we note that:

$$\Pr(\underline{y}(n)/\underline{x}_m(n)\ ) = \sum_{D^n} \Pr(\underline{y}(n)/\underline{x}_m(n),\underline{d}(n))\ \Pr(\underline{d}(n)\ ) \quad (16)$$

It would seem reasonable that, if Equation (15) is true, then with high probability Equation (10) is true. We have not proved the above statement, we have merely suggested its validity. The true relationship between a maximum likelihood decoder and the decoder to be used in this thesis is explored further in Chapter IV.

It is clear that to evaluate the max's in Equation (15) the decoder must test every channel state sequence. This concept of testing both channel state sequences and transmitted messages in order to decode leads to the title of this thesis, "Channel State Testing in Information Decoding". In our decoder we are, in effect, deciding on both the transmitted message and the sequence of channel states. Although we make the latter decision, our primary interest is in the transmitted message and hence in Chapter IV we shall evaluate an upper bound on the probability of decoding error without regard to the probability that the decision on the channel state sequence is correct.

This decoder has the advantages that we are able to obtain an analytical bound on its error probability.

Furthermore, this bound has the desired property (an exponential decay with n) that we would hope to find. Still further, the decoder metric (i.e., $\Pr(\underline{y}(n)/\underline{x}_k(n),\underline{d}(n))\ \Pr(\underline{d}(n))$ may be, with slight modification, used as a metric (see the next section) for a sequential decoder.

That these advantages are obtained should not be construed as meaning that the other decoders discussed above or, in fact, any decoder based on good heuristic reasoning should be precluded. We shall find, for example, that there are many situations in which our decoder is a poor choice. This may be due to the fact that the model chosen for a particular channel is a poor model or that the decoder itself is inherently poor for the case at hand. We can better discuss such matters in Chapter IV.

The point to be emphasized here is that for our decoder we can obtain a bound on error probability whose strengths and weakness in any particular case provide an opportunity to examine the issues at the heart of decoding for the DFSC. In the almost total absence of prior results for channels which are not of the discrete memoryless variety, this opportunity was not previously available.

C. Sequential Decoding for the DFSC

In block decoding we face a dilemma. As we increase n to make the error probability arbitrarily small, while holding the rate, R, constant, the number of messages $M = e^{nR}$ grows exponentially, since for the various alternatives of block decoding discussed above, we must test each possible transmitted sequence. Thus we will, in general, face an exponential amount of computation.

These remarks apply to the DMC as well as the DFSC. In the latter case, for our decoder, the situation is even worse. We must also test every possible channel state sequence. The number of these also grows exponentially with the length, n, of the code.

The most successful technique for avoiding this exponential amount of computation has, in the case of the DMC, been the sequential decoding technique of Wozencraft[21]. Recently, Fano[5] has presented a new sequential decoding algorithm which appears to be somewhat more general. We will use the Fano algorithm with a slight modification to do sequential decoding for the DFSC.

We will restrict the underlying process to have the property that each state may be reached from each other state in a one step transition.

The reason for this restriction will be explained in Chapter V where we discuss its implications.

We assume that the information to be transmitted arrives at the encoder as a stream of equiprobable binary digits which we will call information digits. The encoder is considered to be a finite state device to which are fed $\nu_o \log_2 e$ information digits at a time and whose state at any given time depends on the last $\nu \log_2 e$ information digits which it has accepted. The state may also depend on a particular function of time selected by the designer of the encoder. The encoder output at a given time is then determined uniquely by its state at that time and hence depends on the last $\nu \log_2 e$ information digits fed to it. Such dependence is most readily represented as a tree code in which a particular set of information digits trace a path in the tree along which are listed the channel input symbols generated by the encoder (see Figure 2.3).

The leftmost node of the tree corresponds to the initial state of the encoder which can be assumed to be a state corresponding to a stream of all 0 information digits having been previously fed to the encoder. Each branch corresponds to a particular state of the encoder which is specified by the order number of the branch (i.e., how far into the tree the branch lies) and the last $\nu \log_2 e$ information digits leading to it.

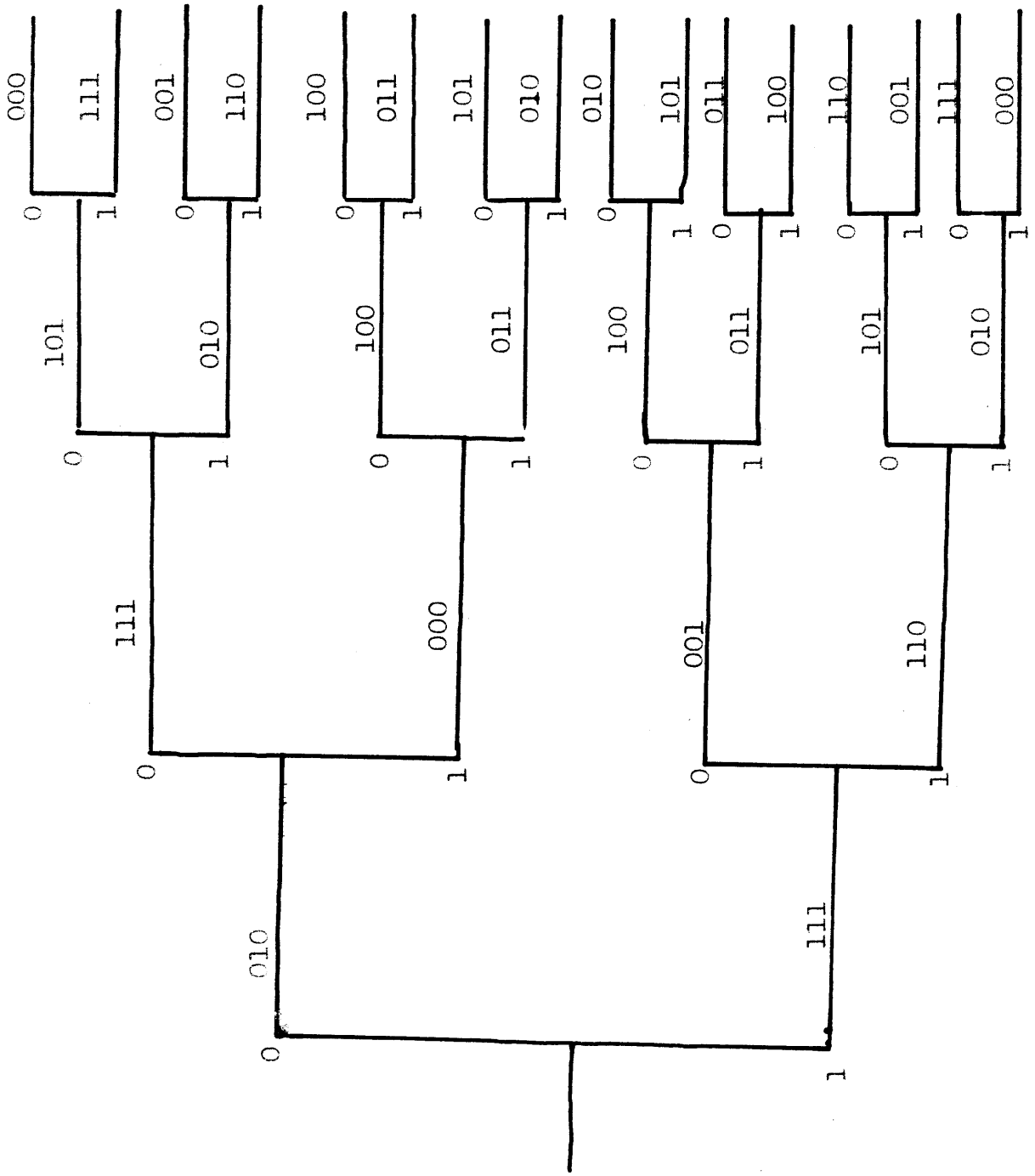


Figure 2.3 A Tree Code with $\nu_o \log_2 e = 1$; $N_o = 3$, $K = 2$

In the figure the information digits are shown just to the left of the branch they generate. The channel symbols corresponding to each branch are shown just above the branch in question.

We assume that the rate, R, is measured in natural units per channel symbol. Thus the number of channel symbols per branch, N_o is given by:

$$N_o = \frac{\nu_o \log_2 e}{R} \tag{17}$$

for each branch.

Now consider two different paths stemming from the same node of the tree. Call this node the reference node. Because the state of the encoder depends on the last $\nu \log_2 e$ information digits fed to it, these two paths must correspond to a sequence of encoder states which are different for at least ν/ν_o branches. Beyond this point corresponding states along the two paths will coincide wherever the sequences of the last $\nu \log_2 e$ information digits along the paths are identical. Two paths stemming from a reference node are called "totally distinct" if the sequences of encoder states along them differ everywhere beyond (i.e., to the right of) the reference node.

The above description of tree codes has been paraphrased from Fano[6]. In Chapter V we will be concerned with an ensemble of such codes. Let us observe at this point that the ensemble (and certainly every member of it) can be generated by an appropriate ensemble of linear feedback shiftregister generators to which are added devices containing stored digits to establish a particular encoder. We will not dwell on the realization of these encoders here, since they have been adequately discussed by Reiffen[16] and Fano[5,6]; but we do state the result that the encoder need have a complexity, as measured in terms of number of elements, that grows only linearly with ν. Note that $\nu \log_2 e$ in this case corresponds to n, the block length, in the case of block coding.

Let us now discuss the method of decoding to be employed. We assume a familiarity with the Fano decoder for the DMC. The decoder computes a metric depending on received and hypothesized transmitted symbols for each branch along a path which is being tested. The running sum of this metric along a path under test is computed. The metric is so chosen that for the actually transmitted path this sum has, with high probability, a monotone increasing (with depth into the tree) lower bound. The decoder is so designed that it searches for and accepts any path having

this property. More precisely, if there are more than one path which have this property, the decoder follows one of them. The decoding procedure is a step-by-step procedure in which each branch is tested individually (rather than long sequences being tested at once as in block decoding). The dependence with depth into the tree arises from the fact that the branches which may be tested at a given time are restricted to those stemming from a tree node which lies along the path accepted up to that time. This reference node is continually updated as the decoding proceeds further and further into the tree. The meaning of this description will become more clear when we examine the details of the decoder for the DFSC.

To adapt the Fano decoder for use on the DFSC we will construct a metric for that case. The viewpoint that we adopt is that we attempt to decode the compound event of transmitted message and channel state sequence which has occurred. Thus, having accepted a path in the tree up to a certain node, the decoder tests all branches stemming from this node, and simultaneously all channel state sequences which are consistent with the state sequence accepted to this node. This concept of jointly testing both message and channel state sequence hypotheses, follows from the discussion of the preceeding section of this

chapter.

Let us now be more precise. Define an arbitrary probability distribution f(y) on the channel output symbols, such that:

$$f(y) > 0 \quad ; \quad y=1,2,\ldots,L$$

$$\sum_{y=1}^{L} f(y) = 1 \tag{18}$$

Now for the branch of order number n, with a particular hypothesis on the transmitted symbols and a particular hypothesis on the channel state sequence, consider the metric

$$g_n = \sum_{j=(n-1)N_o+1}^{nN_o} \ln \frac{p(y_j/x_j,d_j)q_{d_j,d_{j-1}}}{f(y_j)} - U \tag{19}$$

where U is an arbitrary bias.

This metric is the extension to the sequential decoding case of the metric used in the previous section. The significant difference lies in the inclusion of f(y). This function plays the same role here the p(y) plays for the Fano decoder for the DMC. Ideally we would like to include a state

dependent term in the denominator of the argument of the logarithm in Equation (19). We do not do so because we have found such a term to be analytically intractable. The price we pay is that our results for sequential decoding for the DFSC will not, in all cases, bear the same relationship to the results for block coding that is borne in the case of the DMC.

Note that the metric requires knowledge of the present output symbols; the present input symbol hypothesis along the path being followed, the present channel state sequence hypothesis and the most recent channel state hypothesis. Thus, the metric can be computed for each branch in a step-by-step manner which requires only the presence of a tree code generator and a minimal storage of the previous state decision at the decoder.

Now for a particular path in the tree code and a particular sequence of channel states assumed in the decoding define:

$$L_n = \sum_{j=1}^{n-1} g_j \qquad (20)$$

The decoder to be presented below attempts to find a path in the tree and a corresponding sequence of channel states such that along this path the sequence of values L_n has a monotone increasing lower bound.

The operation of the decoder is best explained by examination of a flow chart for it. In Figure 2.4 we present the flow chart.

Here we assume that at each node the branches are numbered in order of the value of the metric along them. Thus $g_{l(n)}$ is the largest value of the metric (consistent with the state assumption on the previous symbol), and $j(n) = 1,2,\ldots,B\ e^{\nu_o}$.

$$\text{Define } \hat{g}_i = \max_{\underline{d}} g_i(\underline{d}) \tag{21}$$

Here $\underline{d}$ is a particular channel state assumption associated with the branch in question. We assume the branches are numbered in order of the value of $\hat{g}$ and $\hat{g}_{l(n)}$ is the largest value of the metric consistent with the state assumption on the previous symbol and $i(n) = 1,2,\ldots,e^{\nu_o}$.

Finally,

$1 \rightarrow F$	stands for:	set F equal to 1
$L_n + \hat{g}_{i(n)} \rightarrow L_{n+1}$	" "	set L_{n+1} equal to $L_n + \hat{g}_{i(n)}$
$L_n + g_{j(n)} \rightarrow L_{n+1}$	" "	" " " " " $L_n + g_{j(n)}$
$n + 1 \rightarrow n$	" "	substitute n+1 for n (increase n by one)
$i(n) + 1 \rightarrow i(n)$	" "	substitute i(n)+1 for i(n)
$T + T_o \rightarrow T$	" "	substitute $T + T_o$ for T
$L_{n+1} : T$	" "	compare L_{n+1} and T; follow path marked $\gtrless$ if $L_{n+1} \gtrless T$.

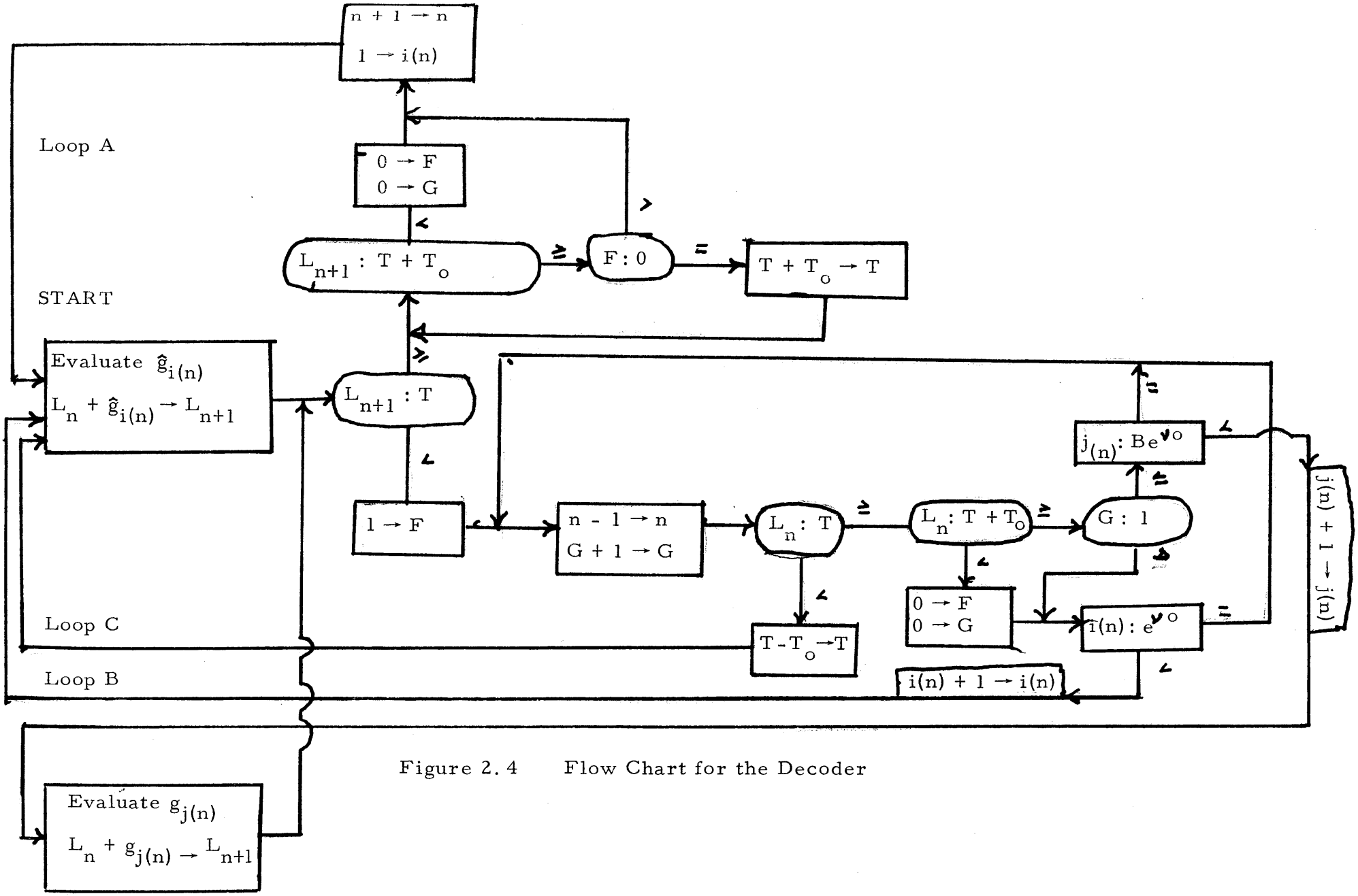


Figure 2.4 Flow Chart for the Decoder

The operation of the decoder is essentially the same as the operation of Fano's decoder in the case of the DMC. The difference lies in the fact that when the decoder is moving forward (i.e., following a path for which L_n is continually increasing) the only state hypotheses utilized are those that maximize the metric for each particular message hypothesis. When the decoder is moving backwards (i.e., following loop B or loop C) for the first time however, we allow the state assumptions to vary over all states consistent with the state decision on the branch preceeding (in order of depth into the tree) the branch presently under investigation. We need never allow this variation for more than one step backwards. This follows from the fact that with Markovian statistics the state sequence can always be forced into any desired state in a one step transition (under the present hypothesis that all states are reachable from all other states in a one step transition). Thus, if a particular path in the tree with a particular channel state sequence hypothesis is one that the decoder can follow successfully, we can always move from this same path with a different state sequence hypothesis to the desired one in a one step transition.

The flow chart presents an equipment whose complexity is independent of ν. It is intuitively clear that as the parameter ν increases the required speed of operation of this equipment must increase. We

thus evaluate an upper bound on a quantity relating to this required speed in Chapter V under the assumption that $\nu = \infty$.

The quantity which is bounded is the average number of times the decoder follows loop A per node decoded. What we mean by "per node decoded" is the following: We shall find (see the next few paragraphs) that the decoder follows a path which agrees with the transmitted path almost everywhere with overwhelming probability. To ultimately follow this path the decoder may examine a given branch more than once (by being forced back through loop B or C). Once the decoder has examined a given branch on the ultimately accepted path for the last time, we may say that the node(i.e., the information symbols) preceeding this branch has been decoded. It is intuitively clear that most of the time the decoder will follow loop A if it is to ultimately get anywhere. Thus the bound on the average number of times loop A is followed per node decoded gives a reasonable measure of the speed with which the decoder must operate. The result obtained in Chapter V is that for rates of information transmission smaller than a rate R_{comp}, this number of traversals of loop A (i.e., the number of computations) is bounded while for rates exceeding R_{comp} it is not.

An investigation of the decoding algorithm leads to the conclusion that the decoder never makes an irrevocable decision. This follows from the fact that the decoder may move backwards in the tree (i.e., to the left) by following loops B or C. There is no limit to how far back the decoder may move. We may obtain an appreciation for the probability that the decoder ultimately follows the correct path, by inhibiting the ability of the decoder to move backwards indefinitely. If we constrain the backward motion to a fixed number of nodes, which we call a constraint length, we can then determine the probability that the decoder has made an incorrect decision at any node once it moves a constraint length ahead of this node. It is this event which precludes the possibility of the decoder ever moving back to change its incorrect hypothesis. This probability is upper bounded (as in the probability that the decoder is ever required by the algorithm to move back more than a constraint length) under the assumption that $\nu = \infty$. The reason for this assumption will become clear in the next paragraph. It is found that both of these probabilities decay exponentially with the constraint length for rates smaller than R_{comp}. Thus if the rate of information transmission is smaller than R_{comp} we are assured that, except for the errors

to be discussed in the next paragraph, the decoder will eventually follow the correct path if the constraint length is infinite.

There is a class of errors which the decoder can make which we call undetectable errors. These arise in the following fashion. Suppose the decoder follows a path which is correct to a given node, but then is incorrect for the next, say, k information digits, and then is correct once more for the information digits beyond this point. Because of the method of encoding the correct path will differ from this path in $\frac{k}{\nu_o \log_2 e} + (\nu - \nu_o) \log_2 e$ branches, but will agree everywhere else. If the metric on the correct path has a monotone increasing lower bound, then so does this particular incorrect path since the two agree in all but a finite number of branches. Thus there exists circumstances under which the decoder may follow this particular incorrect path and yet never detect that such an event has occurred. The results quoted in the previous paragraph establish that with probability one, the decoder will detect an error that occurs from its following a path which is totally distinct from the correct path beyond a given node. Undetectable errors arise only on paths which are <u>not</u> totally distinct from the correct path.

In Chapter V we will find an upper bound on the average number of undetectable errors made per node decoded. It will be found that this bound decays exponentially with ν and hence all errors may be reduced in probability to arbitrarily small values by increasing ν .

The probability that the ultimately accepted channel state sequence is correct is ignored. We in effect consider all errors in the channel state sequence to be undetectable. It is for this reason that we allow the decoder to change state hypotheses only one step into the past. We justify our viewpoint by observing once more that if the decoder follows the path corresponding to the transmitted information digits, then errors in the channel state sequence are of zero cost.

Chapter III

Mathematical Preliminaries

We interrupt the flow of the thesis at this point to introduce some mathematical results which are required for the following chapters.

A. Convexity and Some Standard Inequalities

We list here some standard inequalities which will prove of use in the sequel. Proofs and discussion of these inequalities may be found in Hardy, et.al.[10]. Throughout, we take λ to be a real number with

$$0 \leq \lambda \leq 1 \tag{1}$$

The Inequality of the Algebraic and Geometric Means:

Let $a,b \geq 0$. Then

$$a^{\lambda} b^{(1-\lambda)} \leq \lambda a + (1-\lambda)b \tag{2}$$

Holder's Inequality:

Suppose $a_i, b_i \geq 0 \quad ; \quad i=1,2,\ldots,N$

Then

$$\sum_{i=1}^{N} a_i b_i \leq \left(\sum_{i=1}^{N} a_i^{\frac{1}{\lambda}} \right)^{\lambda} \left(\sum_{i=1}^{N} b_i^{\frac{1}{1-\lambda}} \right)^{(1-\lambda)} \tag{3}$$

Two additional inequalities of interest are:

$$\left(\sum_{i=1}^{N} a_i\right)^{\lambda} \leq \sum_{i=1}^{N} a_i^{\lambda} \tag{4}$$

and if

$$\sum_{i=1}^{N} b_i = 1 \tag{5}$$

(i.e., $\{b_i\}$ is a probability distribution) then

$$\sum_{i=1}^{N} b_i a_i^{\lambda} \leq \left(\sum_{i=1}^{N} b_i a_i\right)^{\lambda} \tag{6}$$

Minkowski's Inequality:

Suppose $a_{ij} \geq 0$; $i=1,2,\ldots,N$; $j=1,2,\ldots,M$. Then

$$\sum_{i=1}^{N} \left(\sum_{j=1}^{M} a_{ij}^{\lambda}\right)^{\frac{1}{\lambda}} \leq \left\{\sum_{j=1}^{M} \left(\sum_{i=1}^{N} a_{ij}\right)^{\lambda}\right\}^{\frac{1}{\lambda}} \tag{7}$$

We next quote a few results on convexity. A good discussion of these results is given in Blackwell and Girschick[2].

A set, C, of elements c is said to be a convex set if for every $c, c' \in C$ and every λ satisfying equation (1) we have:

$$\lambda c + (1-\lambda)c' \in C \tag{8}$$

The elements may be vectors.

A function, F(c) is said to be convex over the set, C, if

$$F(\lambda c + (1-\lambda)c') \leq \lambda F(c) + (1-\lambda) F(c') \quad (9)$$

If the inequality is reversed the function is said to be concave.

A sufficient condition for convexity of F(c) where c is a real number in some interval and F is twice differentiable is that:

$$\frac{d^2}{dc^2} F(c) \geq 0 \quad (10)$$

This condition is also necessary if F is differentiable, but a non-differentiable function may be convex.

Clearly the set of n-dimensional probability vectors

$$\underline{p} = (p_1, p_2, \ldots, p_n)$$

$$p_i \geq 0 \; ; \; \sum_{i=1}^{M} p_i = 1 \quad (11)$$

is a convex set. In this event we have the following special case of the Theorem of Kuhn and Tucker[14].

Theorem 3.1: A necessary and sufficient condition that $\underline{p}^*$ minimize $F(\underline{p})$ is that: there exists a real number, A, such that:

$$\frac{\partial}{\partial p_i} F(\underline{p}) \Bigg|_{\underline{p} = \underline{p}^*} = A \; ; \; p_i^* > 0 \qquad (12)$$

and

$$\frac{\partial}{\partial p_i} F(\underline{p}) \Bigg|_{\underline{p} = \underline{p}^*} \geq A \; ; \; p_i^* = 0 \qquad (13)$$

Equation (13) allows us to determine if in fact the minimum occurs on the boundary of the set of probability vectors (i.e., for some components of the vector being equal to zero).

B. Bounds on Functions over a Markov Chain

In this section we discuss bounds for functions defined over a finite state Markov chain. The basic results stem from Frobenius's theory of non-negative square matrices (see Gantmacher[9]). The essentials of this theory are given below as Theorem 2.1. We begin with a discussion of irreducible non-negative matrices.

A B x B matrix $Z = (z_{ij})$ is non-negative (i.e., $Z \geq 0$) if

$$z_{ij} \geq 0 \text{ for } i,j = 1,2,\ldots,B \qquad (14)$$

The matrix Z is said to be irreducible if it is impossible by a simultaneous permutation of rows and columns of Z to put it in the form:

$$\begin{pmatrix} Z_1 & , & 0 \\ Z_3 & , & Z_2 \end{pmatrix}$$

where Z_1 and Z_2 are square matrices. Clearly the probability matrix of an irreducible Markov chain is an irreducible matrix.

A vector $\underline{v}_1 = (v_{11}, v_{12}, \ldots, v_{1B})$ is said to be

greater than a vector $\underline{v}_2 = (v_{21}, v_{22}, \ldots, v_{2B})$,

(i.e., $\underline{v}_1 \quad \underline{v}_2$) if $v_{1j} > v_{2j}$; $j=1,2,\ldots,B$ (15)

Frobenius's theorem then states:

Theorem 3.2:

An irreducible non-negative matrix, Z, has a largest positive eigenvalue u which has the following properties:

1) u is a simple root (i.e., of multiplicity one) of the characteristic equation

$$Z - u I = 0 \tag{16}$$

2) If w is any eigenvalue of Z then

$$|w| \leq u \tag{17}$$

3) There exist positive left and right eigenvectors $\underline{y}$ and $\underline{x}$ of Z with eigenvalue u

i.e., $Z \underline{x} = u\underline{x}$; $\underline{x} > 0$

$$\underline{y} Z = u\underline{y} \; ; \; \underline{y} > 0 \tag{18}$$

4) If $\underline{w}$ is a positive eigenvector of Z then $\underline{w}$ has eigenvalue u.

5) Let $\lambda > 0$ and $\underline{w} > 0$ satisfy the equation

$$Z\,\underline{w} \leq \lambda\,\underline{w} \tag{19}$$

$$\text{then } \lambda \geq u \tag{20}$$

where the inequality is strict unless $\underline{w}$ is an eigenvector of Z.

6) u is a monotone function of the matrix elements. That is, if any matrix element is increased, then u is increased.

In the sequel we will be interested in exponential bounds for the powers of the matrix Z (t) where

$$Z(t) = (z_{ij}(t)) \tag{21}$$

and each $z_{ij}(t)$ is positive, twice differentiable, and logarithmically convex in some range of real t, $t_o \leq t \leq t_1$. We say a function, z(t) is logarithmically convex if ln z(t) is convex. This implies that z(t) is convex since for $0 \leq \lambda \leq 1$;

$$z(\lambda t_1 + (1-\lambda)t_2) \leq z(t_1)^{\lambda}\, z(t_2)^{1-\lambda}$$

$$\leq \lambda z(t_1) + (1-\lambda)\, z(t_2) \tag{22}$$

The first inequality above comes from the logarithmic convexity. The second inequality comes from the inequality between algebraic and geometric means (Equation (2)).

Now let the n^{th} power of Z (t) be

$$\left[Z\ (t)\right]^{n} = (z_{ij}^{(n)}(t)\ ) \tag{23}$$

then we have the following theorem:

Theorem 3.3:

Let Z (t) be a B x B non-negative irreducible matrix with elements $z_{ij}(t)$. The elements, $z_{ij}^{(n)}(t)$, of the n^{th} power of Z (t) satisfy the inequality:

$$A_1(t)\ (u(t)\ )^n \leq \sum_{j=1}^{B} z_{ij}^{(n)}(t) \leq A_2(t)\ (u(t)\ )^n \tag{24}$$

Here u(t) is the dominant eigenvalue of Z (t) and $A_1(t)$ and $A_2(t)$ are positive and independent of n. Furthermore, if the $z_{ij}(t)$ are all twice differentiable and logarithmically convex, in a region $t_o \leq t \leq t_1$, then $A_1(t)$ and $A_2(t)$ are twice differentiable and u(t) is twice differentiable and logarithmically convex in the region $t_o \leq t \leq t_1$.

Proof:

Let $\underline{b}(t)$ be a positive right eigenvector of Z (t). Then

$$\sum_{j=1}^{B} z_{ij}(t)\ b_j(t) = u(t)\ b_i(t) \tag{25}$$

and

$$\sum_{j=1}^{B} z_{ij}^{(n)}(t)\ b_j(t) = (u(t))^n\ b_i(t) \qquad (26)$$

Now since $\underline{b}(t)$ is positive it has a smallest component $b^*(t)$ and a largest component $b'(t)$. Thus

$$\frac{b^*(t)}{b'(t)}\ (u(t))^n \leq \frac{b_i(t)}{b'(t)}\ (u(t))^n$$

$$= \frac{1}{b'(t)} \sum_{j=1}^{B} z_{ij}^{(n)}(t)\ b_j(t) \leq \frac{1}{b'(t)} \sum_{j=1}^{B} z_{ij}^{(n)}(t)$$

$$\leq \frac{1}{b^*(t)} \sum_{j=1}^{B} z_{ij}^{(n)}(t)\ b_j(t) = \frac{b_i(t)}{b^*(t)}\ (u(t))^n$$

$$\leq \frac{b'(t)}{b^*(t)}\ (u(t))^n \qquad (27)$$

Here $A_1(t) = \dfrac{b^*(t)}{b'(t)}$ (28)

and $A_2(t) = \dfrac{b'(t)}{b^*(t)}$

are positive and independent of n.

Next observe that u(t) is a solution of the equation:

$$\left| z_{ij}(t) - v\,\delta_{ij} \right| = 0 \qquad (29)$$

The left hand side of this equation is a polynomial of degree B in v, each coefficient of which is a polynomial in the elements $z_{ij}(t)$ of Z(t). Since these elements are twice differentiable, it follows that u(t) is also.

Now since u(t) is a simple root of Equation (29) it follows that the matrix

$$Z(t) - Iu(t) = (z_{ij}(t) - \delta_{ij} u(t)) \qquad (30)$$

has rank B-1. Furthermore, since the matrix Z(t) was irreducible, the vector $\underline{a}$ with components

$$a_1 = 1 \qquad (31)$$

$$a_i = 0 \; ; \; 1 < i \leq B \qquad (32)$$

must be linearly independent of the first row of Z(t) - Iu(t). Thus the B x B matrix Y(t) formed by deleting the first row of Z(t) - Iu(t) and replacing it with $\underline{a}$ must be non-singular. Thus the equation

$$y(t)\, \underline{b}\,(t) = (1,0,0,\ldots,0) \qquad (33)$$

serves to specify $\underline{b}(t)$ which is independent of n. If all coefficients of the $b_i(t)$'s in the above equations are twice differentiable, it follows that $b_i(t)$'s are also. Thus $A_1(t)$ and $A_2(t)$ must be twice differentiable.

Now let $t_o \leq s,\ r \leq t_1$. Then define the vector $\underline{b}$ with components:

$$b_i = (b_i(s))^{\lambda} (b_i(r))^{(1-\lambda)} \quad ; \ i=1,2,\ldots,B \qquad (34)$$

for $0 \leq \lambda \leq 1$. Then we have

$$\sum_{j=1}^{B} z_{ij}(\lambda s + (1-\lambda)r)\, b_j$$

$$\leq \sum_{i=1}^{B} \left[z_{ij}(s)\, b_j(s)\right]^{\lambda} \left[z_{ij}(r)\, b_j(r)\right]^{(1-\lambda)} \qquad (35)$$

by virtue of the logarithmic convexity of the z_{ij}'s. Now applying Holder's inequality (Equation(3)) we have

$$\sum_{j=1}^{B} z_{ij}(\lambda s + (1-\lambda)r)\, b_j$$

$$\leq \left[\sum_{j=1}^{B} z_{ij}(s)\, b_j(s)\right]^{\lambda} \left[\sum_{j=1}^{B} z_{ij}(r)\, b_j(r)\right]^{1-\lambda}$$

$$= u(s)^{\lambda}\, u(r)^{1-\lambda}\, b_i \qquad (36)$$

Thus by Equation (20) of Theorem 2.2 we have:

$$u(\lambda s + (1-\lambda)r) \leq (u(s))^{\lambda} (u(r))^{1-\lambda} \qquad (37)$$

Thus u(t) is logarithmically convex, if the $z_{ij}(t)$'s are. Q.E.D.

The above theorem is essentially the same as that given by Kennedy[13]. Our proof differs somewhat in detail and appears to be simpler. We now prove Theorem 3.4:

Let V(t) be a non-negative, irreducible matrix with elements $z_{ij}(t)^{\frac{1}{t}}$, where the $z_{ij}(t)$'s are logarithmically convex in the region $t_o \leq t \leq t_1$. Let v(t) be the dominant eigenvalue of V(t). Then $(v(t))^t$ is logarithmically convex in the same region of t.

Proof: Let $\underline{b}(t)$ be a positive right eigenvector of V(t) and let $t_o \leq s,r \leq t_1$. Then define the vector $\underline{b}$ with components

$$b_i = (b_i(s))^{\frac{\lambda s}{\lambda s+(1-\lambda)r}} (b_i(r))^{\frac{(1-\lambda)r}{\lambda s+(1-\lambda)r}} \tag{38}$$

for $0 \leq \lambda \leq 1$. Then we have

$$\sum_{j=1}^{B} z_{ij}(\lambda s + (1-\lambda)r)^{\frac{1}{s + (1-\lambda)r}} b_j$$

$$= \sum_{j=1}^{B} \left[z_{ij}(s)^{\frac{1}{s}} b_i(s) \right]^{\frac{\lambda s}{\lambda s+(1-\lambda)r}} \left[z_{ij}(r)^{\frac{1}{r}} b_i(r) \right]^{\frac{(1-\lambda)r}{\lambda s+(1-\lambda)r}} \tag{39}$$

by virtue of the logarithmic convexity of the $z_{ij}(t)$'s. Now applying Holder's inequality (Equation 3) we have:

$$\sum_{j=1}^{B} z_{ij}(\lambda s+(1-\lambda)r)^{\frac{1}{\lambda s+(1-\lambda)r}} b_j$$

$$\leq \left[\sum_{j=1}^{B}\left[z_{ij}(s)^{\frac{1}{s}} b_i(s)\right]\right]^{\frac{\lambda s}{\lambda s+(1-\lambda)r}} \left[\sum_{j=1}^{B} z_{ij}(r)^{\frac{1}{r}} b_i(r)\right]^{\frac{(1-\lambda)r}{\lambda s+(1-\lambda)r}}$$

$$= (v(s))^{\frac{\lambda s}{\lambda s+(1-\lambda)r}} (v(r))^{\frac{(1-\lambda)r}{\lambda s+(1-\lambda)r}} b_i \tag{40}$$

Thus by Equation (20) of Theorem 2.2 we have:

$$v(\lambda s+(1-\lambda)r) \leq \left[(v(s))^{s}\right]^{\frac{\lambda}{\lambda s+(1-\lambda)r}} \left[(v(r))^{r}\right]^{\frac{(1-\lambda)}{\lambda s+(1-\lambda)r}} \tag{41}$$

It follows then that

$$v(\lambda s+(1-\lambda)r)^{\lambda s+(1-\lambda)r} \leq \left[(v(s))^{s}\right]^{\lambda}\left[(v(r))^{r}\right]^{(1-\lambda)} \tag{42}$$

Q.E.D.

We now prove the following corollary to Theorem 3.3.

Corollary (3.1):

$$\text{Let } V(\underline{d}(n)\ ) = g(d_o) \prod_{k=1}^{n} v(\hat{d}_k) \tag{43}$$

where $v(\hat{d}_k)$ is a non-negative function defined on the state transitions of a Markov chain and $g(d_o)$ is a non-negative function defined on the states. Then

$$\sum_{D^n} V(\underline{d}(n)\ ) \leq A \mu^n \tag{44}$$

where μ is the dominant eigenvalue of the matrix

$$\Lambda = v(i,j)$$

and A is positive and independent of n.

Proof:

$$\text{Let } \Lambda^n = (v^{(n)}(i,j)\ ) \tag{45}$$

Now by Theorem 2.3 we have:

$$\sum_{i=1}^{B} \sum_{j=1}^{B} g(i)\ \ v^{(n)}(i,j) \leq A \mu^n \tag{46}$$

where the constant, A, includes the factor $\sum_{i=1}^{B} g(i)$.

Next observe that by the definition of matrix multiplication

$$\sum_{d_{k-1}=1}^{B} v(\hat{d}_{k-1})\ v\ (\hat{d}_k)$$

defines the element $v^{(2)}(d_{k-2},d_k)$ of the matrix $\mathcal{L}^2$. Thus, upon itterating the sum on D^n in Equation (44) and performing the innermost n-1 sums, we recognize the identity

$$\sum_{D^n} V(\underline{d}(n)\ ) = \sum_{i=1}^{B} \sum_{j=1}^{B} g(i)\ v^{(n)}(i,j) \qquad (47)$$

The Theorem then follows from Equation (46).

In like manner we prove the slightly more complicated

Corollary 3.2:

$$\text{Let } V(\underline{d}(n)\ ) = g(d_o) \prod_{k=1}^{m} v(\hat{d}_k) \prod_{r-m+1}^{n} w(\hat{d}_r) \qquad (48)$$

where $g(d_o)$ and $v(\hat{d}_k)$ are as in Corollary 2.1 and $w(\hat{d}_r)$ is a positive function defined on the state transitions. Then

$$\sum_{D^n} V(\underline{d}(n)\ ) \leq A\mu^m\ w^{n-m} \qquad (49)$$

where w is the dominant eigenvalue of the matrix (w(i,j)).

The proof is simply an elaboration of the preceeding proof!

We close this chapter with the observation that

the lower bound of Equation (24) guarantees that the upper bounds of the preceeding corollaries are exponentially tight.

Chapter IV

Block Decoding for the DFSC

A. Introduction

In this chapter we obtain an upper bound on the block error probability attainable with block codes for the DFSC.

The method of determining a bound on attainable error probability will be to upper bound the average probability of error where the average is with respect to an ensemble of codes in which the various codewords are selected independently by pairs and the letters within each codeword are selected independently from a common distribution given by the probability function $P(x)$. This ensemble of codes is precisely the ensemble used for the same purpose in work on the DMC by Shannon[18], Fano[4], and Gallager[8]. The utility of the resulting bound resides in the theorem

Theorem 4.1: Let $\overline{P}e$ be the average probability for block decoding error over the ensemble of random codes. Then there exists a code in the ensemble with probability of block decoding error less than or equal to $\overline{P}e$. Furthermore, a code selected at random, in accordance with the statistics of the ensemble, will, with probability greater than or equal to $1 - \frac{1}{a}$, have a probability of block decoding error less than or equal to $a\,\overline{P}e$.

The proof is standard and is not repeated here.

We use the symbol Pe,m to refer to the probability of error when message m is transmitted. By the probability of error Pe for a particular code we mean:

$$Pe = \frac{1}{M} \sum_{m=1}^{M} Pe,m \qquad (1)$$

In addition, we denote the average over the ensemble of Pe,m by $\overline{Pe,m}$.

B. Probability of Error Bounds

In this section we will develop an upper bound to the probability of error averaged over the ensemble of random codes. The bound was first developed using generating function arguments. Subsequently, the bound was obtained using arguments based on those given by Gallager[8] in his proof of the random coding bound for the DMC. We present the latter proof here! Details of the random coding argument are omitted since they are covered adequately in the literature!

We first need the

Lemma 4.1

$$Pe,m \leq Pr\left[\begin{array}{l} Pr(\underline{y}(n)/\underline{x}_{m'}(n),\underline{d}^*(n))\ Pr(\underline{d}^*(n)) \\ \geq Pr(\underline{y}(n)/\underline{x}_m(n),\underline{d}(n))\ Pr(\underline{d}(n)) \\ \quad \text{for any } m' \neq m \text{ and any } \underline{d}^*(n) \end{array} \right. \qquad (2)$$

where $\underline{d}(n)$ is the channel state sequence that actually occurs.

Proof:

$$\Pr(\underline{y}(n) / \underline{x}_m(n), \underline{d}(n)) \Pr(\underline{d}(n))$$

$$\leq \max_{\underline{d}'(n)} \Pr(\underline{y}(n) / \underline{x}(n), \underline{d}'(n)) \Pr(\underline{d}'(n)) \quad (3)$$

The Lemma follows from Equation (3) and the decoding rule (Equation(2.15).).

We now prove:

Lemma 4.2

$$\overline{Pe} \leq M^{\rho} \sum_{Y^n} \left\{ \sum_{D^n} \Pr(\underline{d}(n))^{\frac{1}{1+\rho}} \sum_{X^n} \Pr(\underline{x}(n)) \Pr(\underline{y}(n)/\underline{x}(n), \underline{d}(n))^{\frac{1}{1+\rho}} \right\}^{1+\rho}$$

$$0 \leq \rho \leq 1. \quad (4)$$

Proof:

Define the variable, $\theta_m(\underline{y}(n))$

$\theta_m(\underline{y}(n)) = 1$; if event in square brackets in Equation (2) is true. $\theta_m(\underline{y}(n)) = 0$; otherwise.

Then $\theta_m \leq \left\{ \sum_{D^{*n}} \sum_{m' \neq m} \frac{\Pr(\underline{y}^n/\underline{x}_{m'}(n); \underline{d}^*(n))^{\frac{1}{1+\rho}} \Pr(\underline{d}^*(n))^{\frac{1}{1+\rho}}}{\Pr(\underline{y}(n)/\underline{x}_m(n); \underline{d}(n))^{\frac{1}{1+\rho}} \Pr(\underline{d}(n))^{\frac{1}{1+\rho}}} \right\}^{\rho}$

$$\rho \geq 0 \quad (5)$$

Equation (5) follows from the fact that when $\theta_m = 1$ at least one of the terms in the summand exceeds 1, and when $\theta_m = 0$ the right hand side is

positive. Now when message m is transmitted the pair $(\underline{y}(n),\underline{d}(n))$ occur with probability $\Pr(\underline{y}(n)/\underline{x}_m(n),\underline{d}(n))\Pr(\underline{d}(n))$. Thus by virtue of Lemma 4.1 and Equation (5) we have:

$$P_{e,m} = \sum_{Y^n}\sum_{D^n} \Pr(\underline{y}(n)/\underline{x}_m(n),\underline{d}(n))\Pr(\underline{d}(n))\theta_m(\underline{y}(n))$$

$$\leq \sum_{Y^n}\sum_{D^n} \Pr(\underline{d}(n))^{\frac{1}{1+\rho}} \Pr(\underline{y}(n)/\underline{x}_m(n),\underline{d}(n))^{\frac{1}{1+\rho}}$$

$$\cdot\left\{\sum_{m'=m}\sum_{D^{*n}} \Pr(\underline{d}^*(n))^{\frac{1}{1+\rho}} \Pr(\underline{y}(n)/\underline{x}_{m'}(n),d^*(n))^{\frac{1}{1+\rho}}\right\}^{\rho} \qquad (6)$$

Now by assumption on the independence of codeword choices for the random code assignment we have:

$$\overline{P_{e,m}} = \sum_{Y^n}\sum_{D^n} \Pr(\underline{d}(n))^{\frac{1}{1+\rho}} \sum_{X_m^n} \Pr(\underline{x}_m(n))\Pr(\underline{y}/\underline{x}_m(n);\underline{d}(n))^{\frac{1}{1+\rho}}$$

$$\prod_{m'=m}\sum_{X_{m'}^n} \Pr(\underline{x}_{m'}(n))\left\{\sum_{m'=m}\sum_{D^{*n}} \Pr(\underline{d}^*(n))^{\frac{1}{1+\rho}}\right.$$

$$\left.\cdot \Pr(\underline{y}(n)/\underline{x}_{m'}(n),\underline{d}^*(n))^{\frac{1}{1+\rho}}\right\}^{\rho} \qquad (7)$$

We recognize $\prod_{m'=m} \Pr(\underline{x}(n)_{m'})$ to be a probability function. Hence for $\rho \leq 1$; we have from Equation (2.4)

$$\overline{Pe,m} \leq \sum_{Y^n} \sum_{D^n} Pr(\underline{d}(n))^{\frac{1}{1+\rho}} \sum_{X_m^n} Pr(\underline{x}_m(n))\ Pr(y(n)/x_m(n);\underline{d}(n))^{\frac{1}{1+\rho}}$$

$$\left\{ \sum_{m'=m} \sum_{D^{*n}} Pr(\underline{d}^*(n))^{\frac{1}{1+\rho}} \sum_{X_{m'}^n} Pr(\underline{x}_{m'}(n)) Pr(y(n)/x_{m'}(n),\underline{d}^*(n))^{\frac{1}{1+\rho}} \right\}^{\rho} \tag{8}$$

We now recognize that the expression inside the curly brackets of Equation (10) is the same for each m' and is also the same as the corresponding expression outside the curly brackets. Thus $\overline{Pe,m}$ is independent of m and hence by Equation (4) we have:

$$\overline{Pe} = (M-1)^{\rho} \sum_{Y^n} \left\{ \sum_{D^n} Pr(\underline{d}(n))^{\frac{1}{1+\rho}} \sum_{X^n} Pr(\underline{x}(n)) Pr(\underline{y}(n)/\underline{x}(n),\underline{d}(n))^{\frac{1}{1+\rho}} \right\}^{1+\rho} \tag{9}$$

Now bound (M-1) by M. Q.E.D.

This Lemma is basic to all of the results to be obtained below. We now obtain a bound that depends explicitly on the block length, n, and hence establishes a coding theorem! The bound is given for three different cases of DFSC's.

Case I: Channels in which the output specifies the state: We say that a DFSC is a channel in which the output specifies that state, if upon knowing the output symbol we know, without ambiguity, the state occupied by the underlying Markov process. In Figure 4.1

we give an example of a channel in which the output specifies the state. We choose, in this case, to associate the transmission probability functions with the states rather than with the state transitions, because of conventions which will put our bound in the same form as that previously obtained by Gallager (unpublished) for this case. For convenience we write our transmission probability functions as $p(y/x,d)$. We now have:

Theorem 4.2:

The average probability of error for block codes of length n for a DFSC in which the output specifies the state is bounded by

$$\overline{P}e \leq A\, e^{-n\left[E_o(\rho,\underline{p}) - \rho R\right]} \tag{10}$$

where $e^{-E_o(\rho,\underline{p})}$ is the dominant eigenvalue of the matrix $H_1(\rho)$ where

$$H_1(\rho) = \left(q_{ij}\, e^{-E_{oj}(\rho,\underline{p})}\right) \tag{11}$$

and

$$E_{oj}(\rho,\underline{p}) = -\ln \sum_{y=1}^{L} \left[\sum_{x=1}^{K} P(x)p(y/x,j)^{\frac{1}{1+\rho}}\right]^{1+\rho} \tag{12}$$

$$\underline{p} = (p_1, p_2, \ldots, p_k) \tag{13}$$

$$Q = \begin{pmatrix} 1-p & p \\ 1-q & q \end{pmatrix} \qquad 0 \leq p,\ q \leq 1$$

$p\ (y/x,1)$

x \ y	1	2	3	4	5	6
1	1-a	a/2	a/2	0	0	0
2	b/2	1-b	b/2	0	0	0
3	c/2	c/2	1-c	0	0	0

$0 \leq a,b,c \leq 1$

$p\ (y/x,2)$

x \ y	1	2	3	4	5	6
1	0	0	0	1-d	d/2	d/2
2	0	0	0	e/2	1-e	e/2
3	0	0	0	f/2	f/2	1-f

$0 \leq d,e,f \leq 1$

Figure 4.1 A Channel in which the Output Determines the State

and $p_k = P(k) = \Pr(x = k)$ (14)

(recall that there are K input symbols for the channel).

Finally, A is independent of both M and n.

Proof: Observe first that once we have assigned the transmission probability functions to the states we may drop consideration of d_o and take d_1 as the initial state. Then interpreting $\underline{d}(n)$ as $(d_1, d_2, \ldots, d_n)$ we observe that

$$\Pr(\underline{y}(n) \,/\, \underline{x}(n), \underline{d}(n)) = 0$$

unless $\underline{d}(n)$ is the particular state sequence specified by $\underline{y}(n)$. Thus the sum in curly brackets in Equation (4) has only one non-zero term and hence the sum may be removed from the brackets. The resulting bound on error probability is then:

$$\overline{P}e \leq M^{\rho} \sum_{D^n} \Pr(\underline{d}(n)) \sum_{Y^n} \left\{ \sum_{X^n} \Pr(\underline{x}(n)) \Pr(\underline{y}(n)/\underline{x}(n), \underline{d}(n))^{\frac{1}{1+\rho}} \right\}^{1+\rho} \tag{15}$$

Iterate the sums on Y^n and X^n and define

$$v(\hat{d}_i) = \Pr(d_i / d_{i-1}) \sum_{y_i=1}^{L} \left\{ \sum_{x_i=1}^{K} P(x_i) p(y_i / x_i, \hat{d}_i)^{\frac{1}{1+\rho}} \right\}^{1+\rho} \tag{16}$$

Here we use the fact that for random coding

$$\Pr(\underline{x}(n)\ ) = \prod_{i=1}^{n} P(x_i) \tag{17}$$

Then from Corollary (3.1) we obtain the desired result.

Case II: Channels with input rotations

Let $Z_{d',d}$ be a matrix representing the transmission probability function $p(y/x,d',d)$. Thus

$$Z_{d',d} = (p_{k\ell}(d',d)\ ) \tag{18}$$

where

$$p_{k\ell} = p(\ell / k,d',d) \tag{19}$$

We say that a DFSC is a channel with input rotations if for every d' and d the matrix $Z_{d',d}$ may be obtained from the corresponding matrix for every other d' and d by a permutation of rows alone. An example of a channel with input rotations is given in Figure 4.2. Now let P(x), the probability assignment for the random codes, be restricted such that

$$P(x) = P(x') \tag{20}$$

if there exists any two state transitions d',d and c',c such that

$$p_{x',y}(d',d) = p_{x',y}(c',c)\ ;\ \text{all } y = 1,2,\ldots,L \tag{21}$$

$$Q = \begin{pmatrix} 1-a-b & a & b \\ c & 1-c-d & d \\ f & f & 1-e-f \end{pmatrix}$$

$$0 \leq a,b,c,d,e,f \leq 1$$

$p(y/x,1)$

x \ y	1	2	3	4
1	1/8	1/8	1/4	1/2
2	1/3	1/3	1/6	1/6
3	1/16	5/16	7/16	3/16
4	2/5	1/5	1/5	1/5
5	6/11	1/11	3/11	1/11

$p(y/x,2)$

x \ y	1	2	3	4
1	1/3	1/3	1/6	1/6
2	1/8	1/8	1/4	1/2
3	1/16	5/16	7/16	3/16
4	2/5	1/5	1/5	1/5
5	6/11	1/11	3/11	1/11

$p(y/x,3)$

x \ y	1	2	3	4
1	1/3	1/3	1/6	1/6
2	1/8	1/8	1/4	1/2
3	2/5	1/5	1/5	1/5
4	1/16	5/16	7/16	3/16
5	6/11	1/11	3/11	1/11

Figure 4.2 A Channel with Input Rotations

In effect we partition the input symbols into disjoint sets such that all the elements in a given set satisfy Equation (21) for some other element in the set and some pair of state transitions. Then to all the elements in a particular set we assign the same probability in the random code.

In the example of Figure 4.2. these sets are

$$X_1 = (1,2)$$
$$X_2 = (3,4) \qquad (22)$$
$$X_3 = (5)$$

and the probability assignment is constrained such that

$$P(1) = P(2) = p$$
$$P(3) = P(4) = q \qquad (23)$$
$$P(5) = r$$

where we have

$$2p + 2q + r = 1 \qquad (24)$$

We then have:

Theorem 4.3:

The average probability of error for block codes of length n for a DFSC with input rotations, and P(x) restricted as in Equation (20), is bounded by

$$\overline{P}_e \leq A\, e^{-n\left[E_o(\rho,\underline{p}) - \rho R\right]} \tag{25}$$

where

$$e^{-\frac{E_o(\rho,\underline{p})}{1+\rho}} = e^{-\frac{E_{oij}(\rho,p)}{1+\rho}} \times$$

$$\left[\text{dominant eigenvalue of the matrix } H(\rho)\right]$$

where $H(\rho) = (q_{ij}^{\frac{1}{1+\rho}})$

and $$E_{oij}(\rho,\underline{p}) = -\ln \sum_{y=1}^{L} \left\{ \sum_{x=1}^{K} p(x)p(y/x,i,j)^{\frac{1}{1+\rho}} \right\}^{1+\rho} \tag{26}$$

and is independent of i,j. Furthermore, A is independent of M and n.

Proof: Observe that under the restrictions on P(x)

$$\sum_{x^n} \Pr(\underline{x}(n)) \Pr(\underline{y}(n)/\underline{x}(n),\underline{d}(n))^{\frac{1}{1+\rho}} \text{ is independent}$$

of d(n). This factor may therefore be taken out of the curly brackets in Equation (4). The resulting bound is:

$$\bar{P}e \leq M^{\rho} \sum_{Y^n} \left\{ \sum_{X^n} Pr(\underline{x}(n)\ )\ Pr(\underline{y}(n)/\underline{x}(n),\underline{d}(n)\ )^{\frac{1}{1+\rho}} \right\}^{1+\rho} \cdot$$

$$\left\{ \sum_{D^n} Pr(d(n)\ )^{\frac{1}{1+\rho}} \right\}^{1+\rho} \qquad (27)$$

Iterating the sums on Y^n and X^n we have:

$$\bar{P}e \leq e^{-n\left[E_{oij}(\rho,\underline{p}) - \rho R\right]} \left[\sum_{D^n} Pr(\underline{d}(n)\ )^{\frac{1}{1+\rho}} \right]^{1+\rho} \qquad (28)$$

Then identifying

$$v(\hat{d}_i) = Pr(d_i\ /\ d_{i-1})^{\frac{1}{1+\rho}} \qquad (29)$$

and applying Corollary 3.1 we obtain the result of the theorem.

Case III: General DFSC: In this case we have the:

Theorem 4.4:

The average probability of error for block codes of length n for the DFSC is bounded by the expression

$$Pe \leq A\ e^{-n\left[E_o(\rho,\underline{p}) - \rho R\right]} \qquad (30)$$

where $e^{-\frac{E_o(\rho,p)}{1+\rho}}$ is the maximal eigenvalue of the matrix $H(\rho)$.

$$H(\rho) = \left(q_{ij}^{\frac{1}{1+\rho}} \; e^{-\frac{E_{oij}(\rho,p)}{1+\rho}}\right) \tag{31}$$

and $E_{oij}(\rho,p)$ is given by Equation (26) but is dependent on i and j. Furthermore, A is independent of M and n.

Proof:

We apply Minkowski's inequality (Equation (3.7)) to Equation (4). Identifying:

$$Y^n = i \; ; \; D^n = j \; ; \; \frac{1}{1+\rho} = \lambda \tag{32}$$

$$\Pr(d(n)) \left[\sum_{X^n} \Pr(\underline{x}(n)) \Pr(\underline{y}(n)/\underline{x}(n),\underline{d}(n))^{\frac{1}{1+\rho}}\right]^{1+\rho} = a_{ij}$$

We obtain the result:

$$\bar{P}e \leq M^{\rho}\left\{\sum_{D^n} \Pr(\underline{d}(n))^{\frac{1}{1+\rho}} \cdot\left(\sum_{Y^n}\left[\sum_{X^n} \Pr(\underline{x}(n)) \Pr(\underline{y}(n)/\underline{x}(n),\underline{d}(n))^{\frac{1}{1+\rho}}\right]^{1+\rho}\right)^{\frac{1}{1+\rho}}\right\}^{1+\rho} \tag{33}$$

Next iterate the sums on Y^n and X^n and define:

$$v(\hat{d}_i) = \left[\sum_{y_i=1}^{L} \left\{ \sum_{x_i=1}^{K} P(x_i) p(y_i/x_i, \hat{d}_i)^{\frac{1}{1+\rho}} \right\}^{1+\rho} \right]^{\frac{1}{1+\rho}} \tag{34}$$

Then applying Corollary 3.1 we obtain the desired result.

C. Properties of the Bound

In this section we investigate various properties of the bounds of Theorems 4.2, 4.3, and 4.4. We begin by observing that if a matrix is multiplied by a constant, all eigenvalues of this matrix are multiplied by this same constant. Thus the bound of Theorem 4.3 is identical to the bound of Theorem 4.4. In the proof of Theorem 4.4. we used Minkowski's inequality which was not used in Theorem 4.3. The conclusion we reach is that for channels with input rotations, Minkowski's inequality holds true as an equality if the random code probability assignment is restricted as in Equation (20). We may then treat the results of Theorem 4.3 and 4.4 as the same.

We next observe that the functions $E_{oij}(\rho, p)$ are precisely the functions which appear in Gallager's[8] bound

$$\overline{P}e \leq e^{-n(E_o(\rho, p) - \rho R)} \tag{35}$$

for block coding for the DMC. That is, if we consider $p(y/x,i,j)$ for fixed i,j as the transmission probability

function for a DMC, then the function $E_o(\rho,\underline{p})$ which appears in Equation (35) is just $E_{oij}(\rho,\underline{p})$. Gallager has shown that this function is a twice differentiable, concave function of ρ for a fixed $\underline{p}$. We then have:

Theorem 4.5:

The functions $E_o(\rho,\underline{p})$ of both Theorems 4.2 and 4.4 are twice differentiable, concave functions of for fixed $\underline{p}$.

Proof: From the observations above we have that $q_{ij}e^{-E_{oij}(\rho,\underline{p})}$ is a twice differentiable, logarithmically convex function of ρ for fixed $\underline{p}$. For $E_o(\rho,\underline{p})$ of Theorem 4.2 the desired result then follows from Theorem 3.3. For $E_o(\rho,\underline{p})$ of Theorem 4.4, the result then follows from Theorem 3.4.

We now observe that the matrix H(0) is stochastic (in both cases of interest) and hence has a dominant eigenvalue equal to 1. Thus, $E_o(0,\underline{p}) = 0$. The possible behaviors of $E_o(\rho,\underline{p})$ for fixed $\underline{p}$, are then as is shown in Figure 4.2.

Now define:

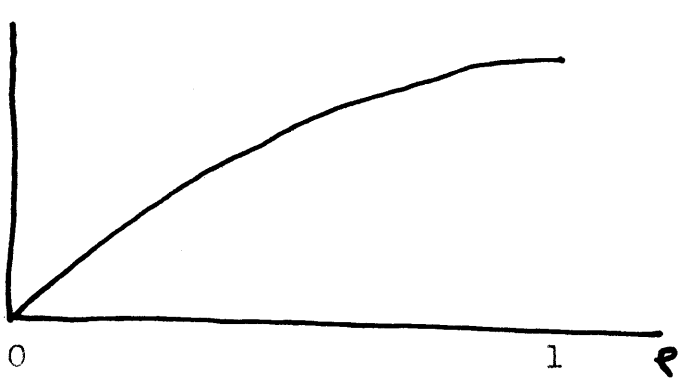
$$E(R,\underline{p}) = \max_{0 \leq \rho \leq 1} E_o(\rho, \underline{p}) - \rho R \qquad (36)$$

We then have, trivially, for all cases of the DFSC:

$$\overline{P}e \leq e^{-n E(R,\underline{p})} \qquad (37)$$

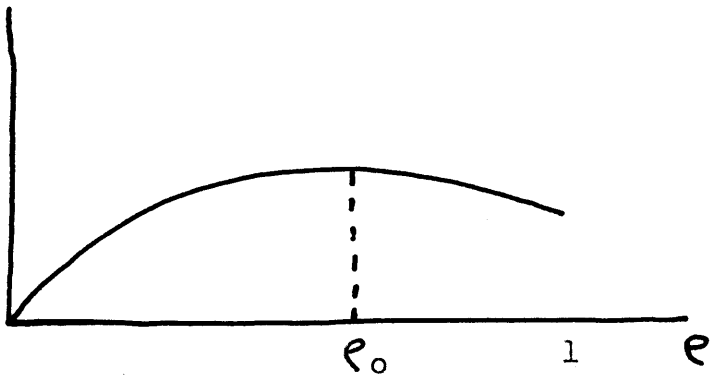
From the concavity of $E_o(\rho)$ we may write:

a)



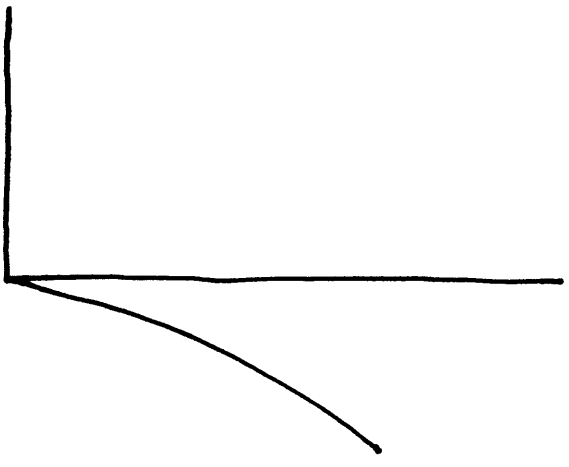
$$\max_{\rho} \; E_o(\rho) = E_o(1)$$

b)



$$\max_{\rho} \; E_o(\rho) = E_o(\rho_o)$$

c)



$$\max_{\rho} \; E_o(\rho) = E_o(0) = 0$$

Figure 4.3 Possible Behaviors of $E_o(\rho, \underline{p})$

Theorem 4.5:

$$\text{If} \quad \frac{\partial E_o(\rho, \underline{p})}{\partial \rho} \Bigg|_{\rho = 0} \leq 0$$

Then $E(R, \underline{p}) = 0$

$$\text{If} \quad \frac{\partial E_o(\rho, \underline{p})}{\partial \rho} \Bigg|_{\rho = 0} \geq 0$$

$$\text{Then } E(R, \underline{p}) = E_o(\rho, \underline{p}) - \rho \frac{\partial E_o(\rho, \underline{p})}{\partial \rho} \tag{38}$$

where ρ is picked such that

$$R = \frac{\partial E_o(\rho, \underline{p})}{\partial \rho} \quad ; 0 \leq \rho \leq 1 \tag{39}$$

If, furthermore,

$$\frac{\partial E_o(\rho, \underline{p})}{\partial \rho} \Bigg|_{\rho = 1} \geq 0$$

$$E(R, \underline{p}) = E_o(1, \underline{p}) - R \tag{40}$$

$$\text{for } R < \frac{\partial E_o(\rho, \underline{p})}{\partial \rho} \Bigg|_{\rho = 1} \tag{41}$$

Finally, for all rates such that Equation (39) may be satisfied we have:

$$\frac{\partial E(R,\underline{p})}{\partial R} = -\rho \tag{42}$$

Proof: All results of the theorem are obvious except Equation (42). For this, set

$$\frac{\partial E(R,\underline{p})}{\partial R} = \frac{\frac{\partial}{\partial \rho}\left[E(R,\underline{p})\right]}{\frac{\partial}{\partial \rho}\left[R\right]} = \frac{\frac{\partial}{\partial \rho}\left\{E_o(\rho,\underline{p}) - \rho\frac{\partial}{\partial \rho}\left[E_o(\rho,\underline{p})\right]\right\}}{\frac{\partial^2}{\partial \rho^2} E_o(\rho,\underline{p})}$$

$$= -\rho \tag{43}$$

The possible behaviors of $E(R,\underline{p})$ are then as is shown in Figure 4.3. The situation in Figure 4.3 is exactly that encountered in the analysis of the DMC.

We may now deduce:

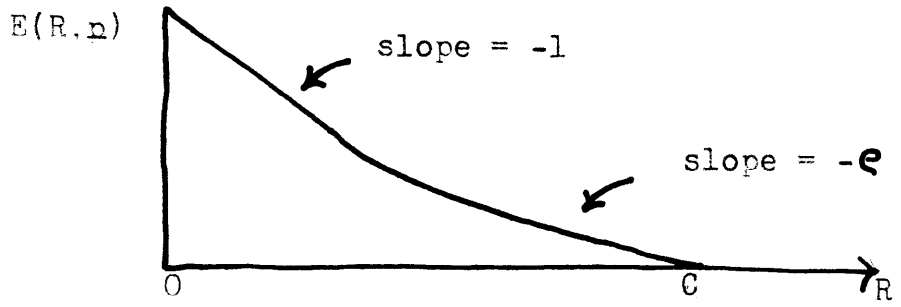
$$\overline{P}_e \leq e^{-n\,E(R)} \tag{44}$$

where $$E(R) = \max_{\underline{p}} E(R,\underline{p}) \tag{45}$$

This maximization cannot be achieved analytically in any cases of generality. Clearly

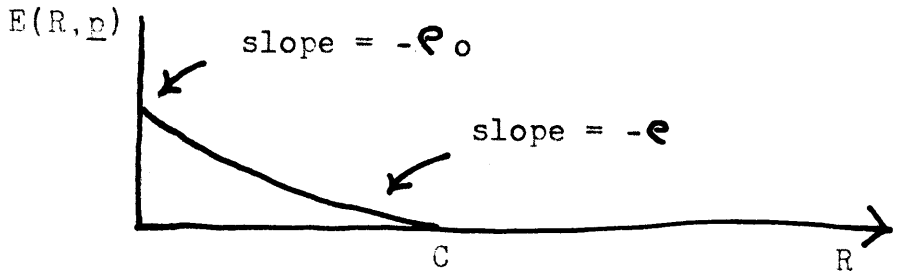
$$E(R) = \max_{\rho} \max_{\underline{p}} E_o(\rho,\underline{p}) - \rho R \tag{46}$$

a)



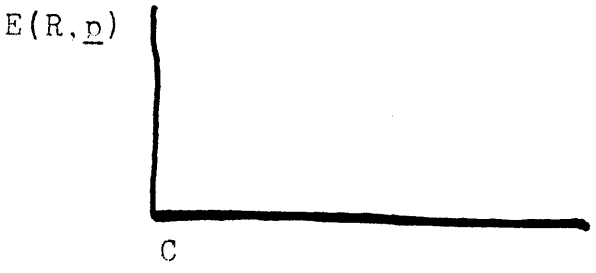
Corresponding to Figure 4.3a

b)



Corresponding to Figure 4.3b

c)



Corresponding to Figure 4.3c

Figure 4.4 Possible Behaviors of E(R,p)

and the max's may be taken in any order. If the max on $\underline{p}$ is evaluated first we may then construct E(R) graphically as shown by Gallager[8]. In general, to perform the max on $\underline{p}$ first we may use the fact from Theorem 3.2, that the dominant eigenvalue of a positive matrix is a monotone function of the matrix elements. Thus if each element of the matrix is decreased or kept fixed, the dominant eigenvalue of the matrix decreases.

For the special case of channels with input rotations, we observe that the constraint on the random code probability assignment defines a convex set of probability vectors $\underline{p}$. Furthermore, in all cases, $E_{oij}(\rho,\underline{p})$ is a concave function of $\underline{p}$. (This follows from Gallager's results on the DMC). Thus, in this special case we may observe that to maximize $E_o(\rho,\underline{p})$ we may maximize $E_{oij}(\rho,\underline{p})$ and conditions on $\underline{p}$ to achieve this maximum are given by the Kuhn-Tucker Theorem (Theorem 3.1).

Finally, we observe that for situations in which there exists a $\underline{p}$ such that

$$\frac{\partial}{\partial} E_o(\rho,\underline{p}) \Big|_{\rho = 0} \neq 0 \qquad (47)$$

it is appropriate to define the capacity, $\mathbb{C}$, of our decoding scheme as:

$$C = \max_{\underline{p}} \frac{\partial}{\partial \rho} E_o(\rho, \underline{p}) \Big|_{\rho = 0} \qquad (48)$$

D. Further Properties of the Bounds

In this section we investigate the "goodness" of our bounds in the sense of their exponential tightness and relation to bounds on maximum likelihood decoders for the DFSC. The discussion is qualitative rather than quantitative for reasons to become obvious below.

Our first observation is that for the special case of a DMC (in which there is only one state in the Markov chain) our bound is equal to that found by Gallager[8] for maximum likelihood decoding. This follows from the fact that in this special case our decoder is, in fact, maximum likelihood.

In more general cases, we may make the following observation: Our Lemma 4.2 may be replaced by the following Lemma for maximum likelihood decoding.

Lemma 4.3:

The average probability of error for maximum likelihood decoding of block codes for the DFSC is bounded by:

$$\overline{P_e} \leq M^{\rho} \sum_{Y^n} \left\{ \sum_{X^n} \Pr(\underline{x}(n)) \Pr(\underline{y}(n)/\underline{x}(n)) ^{\frac{1}{1+\rho}} \right\}^{1+\rho}$$

$$0 \leq \rho \leq 1 \qquad (49)$$

This result appears an an intermediate step in Gallager's derivation of his bound for the DMC. The argument is equally valid for the DFSC. Now observe that:

$$\Pr(\underline{y}(n)/\underline{x}(n))^{\frac{1}{1+\rho}} = \left[\sum_{D^n} \Pr(\underline{y}(n)/\underline{x}(n),\underline{d}(n))\ \Pr(\underline{d}(n))\right]^{\frac{1}{1+\rho}} \tag{50}$$

Then by applying the inequality of Equation (3.4) we have:

$$\Pr(\underline{y}(n)/\underline{x}(n))^{\frac{1}{1+\rho}} \leq \sum_{D^n} \Pr(\underline{d}(n))^{\frac{1}{1+\rho}}\ \Pr(\underline{y}(n)/\underline{x}(n),\underline{d}(n))^{\frac{1}{1+\rho}} \tag{51}$$

By substituting the above bound in Equation (49) we obtain our Equation (4). It then follows that the weakness of our decoder relative to maximum likelihood decoding may be measured in terms of the weakness of the bound in Equation (51).

Next observe that for channels in which the state determines the output, there is only one non-zero term, for fixed $\underline{y}(n)$, in the sum on the right hand side of Equation (50). Thus in this case the equality holds in Equation (51) and our resultant bound (as well as our decoding method) is maximum likelihood.

For special cases in which $\underline{x}(n)$ and $\underline{d}(n)$ uniquely specify $\underline{y}(n)$, the inequality of Equation (51) may be investigated in terms of the number of sequences $\underline{d}(n)$

which jointly with $\underline{x}(n)$ specify a given $\underline{y}(n)$. If the number is at most algebraic in n, then the bound of Equation (51) can not be exponentially wrong. This means that:

$$\lim_{n\to\infty} \left(\frac{1}{n} \ln \Pr(\underline{y}(n)/\underline{x}(n)) \right)^{\frac{1}{1+\rho}}$$

$$= \lim_{n\to\infty} \frac{1}{n} \ln \sum_{D^n} \Pr(\underline{d}(n))^{\frac{1}{1+\rho}} \Pr(\underline{y}(n)/\underline{x}(n),\underline{d}(n))^{\frac{1}{1+\rho}} \tag{52}$$

Kennedy[13] has called such channels (in the binary input-binary output case) Type I Channels. For such channels, our decoder is asymtotically as good as a maximum likelihood decoder.

The important question of whether any given physical channel can always be modelled as a DFSC with the property of Equation (52) remains unanswered. Such structural questions appear to be far too difficult for ready solution.

At this point we are able to justify our statements regarding choice of models which were made in Section A of Chapter II. The cost of poor modelling, when using our decoder, is easily shown by the example of the memoryless Binary Symmetric Channel. Consider the alternate models shown in Figure 2.2. It may be readily checked that for the two state model our bound (as given by Theorem 4.3) agrees with

the bound which may be obtained by considering the channel as memoryless. However, for the 2m state model the function $E_o(\rho)$ (the inputs are taken equally likely) which is obtained for the two state model is decreased by the quantity $\rho \ln m$. Hence for large enough m, the form of $E_o(\rho)$ can change from that shown in Figure (4.1a) to that shown in Figure (4.1c).

As a consequence, we go from the correct bound to no bound at all.

Finally, let us consider the degredation in our bound of Theorem 4.4 introduced by the use of Minkowski's inequality. We have already observed that for channels with input rotations we suffer no loss at all from this inequality. On the other hand, by comparing Theorems 4.2 and 4.4, we see that for channels in which the output specifies the states, we may, indeed, suffer great loss. The extent of the weakness introduced by this inequality is a subject for future investigation.

E. Final Comments

We have presented a decoding technique for which upper bounds on error probability may be obtained and evaluated. The major virtue of these bounds lies in the fact that we may now replace vague questions, such as "What can be said about decoding for time-varying channels?" by the more specific questions posed in the preceeding section of this chapter.

Although we are unable to answer these questions, at least certain fundamental issues have been brought to light. What we offer is a set of results against which later analyses and results may be compared.

There is one final matter we wish to discuss before closing this chapter. There has long existed the question as to whether or not, by using knowledge of the memory in the channel, it is possible to obtain better error probabilities than by ignoring such memory. To answer this question we must assume that there is a data processing scheme which converts the given channel to a memoryless channel against which coding comparisons can be made.

It is important to emphasize that this comparison must be fair. For example, we may consider the case of a continuous channel with additive gaussian noise and a stochastically varying phase shift. Suppose we signal over this channel with orthogonal signals. We first observe that to fit the channel to our model we may quantize the phase to some desired level. Assuming that, for the resultant model of the channel, Equation (52) is satisfied we may state the following from physical considerations:

1) Our bound is always poorer than what is achievable with coherent detection and known phase.

2) Our bound is always as good as or better than

what is achievable with incoherent detection.

It is clear that it is not fair to compare our results to those for a coherent receiver. On the other hand, it is not clear that incoherent detection represents a good conversion of the channel to a memoryless channel.

This example presents another interesting question. If we quantize the phase to a very fine level, then the resultant $E_{oij}(\rho,\underline{p})$ functions will be relatively large, but the number of states in the model will be large. In some cases, this large number of states may cause a deterioration off-setting the effect of the large $E_{oij}(\rho,\underline{p})$'s on our bound. It may then be true that an optimum level of quantization exists. The investigation of this optimum quantization deserves to be pursued. It is clear that resultant error probability for a system employing such optimum quantization will not be better than true maximum likelihood detection which will perform appropriate integration of the (continuous) distribution on the phase.

To return to the question of utilizing memory in decoding, consider the following example (due to Kennedy[13]).

We have a binary input-binary output channel with two states having the Markov transition probability matrix

$$\begin{pmatrix} 1-p & p \\ p & 1-p \end{pmatrix}; \quad p < \tfrac{1}{2}$$

State 1 is a 0 state and state 2 a 1 state as in Figure (2.1).

There is a temptation in such cases to define for comparison purposes a memoryless binary symmetric channel whose cross-over probability is equal to the stationary probability of state 2. In this example, the resultant channel would have capacity equal to zero. On the other hand an evaluation of the applicable probability of error bound (Theorem 4.3) shows that the $E_o(\rho)$ for this example is equal to that of a memoryless Binary Symmetric Channel with cross-over probability p.

We can in one special case make a reasonably fair estimate of the cost of time variations. Suppose, that for some $\underline{p}$, $E_{oij}(\rho,\underline{p})$ is independent of i and j. When such a situation occurs, we say the channel is state independent for this $\underline{p}$. Thus channels with input rotation are state independent for all $\underline{p}$'s meeting the constraint of Equation (20). Other cases can occur. For example, if the matrices of Equation (18) are such that the matrix for each state transition may be obtained from that for every other state transition by a permutation of columns alone, then a choice of $\underline{p}$ with all components equal

will yield a state-independent channel.

For state independent channels, the bound of Theorem 4.4 may be written as in Theorem 4.3. We then have:

$$E_o(\rho) = E_{oij}(\rho) - (1+\rho) \ln \left[\text{dominant eigenvalue of } (q_{ij}^{\frac{1}{1+\rho}}) \right] \tag{53}$$

It is fair to consider the DMC with transmission probability function equal to that of any particular state transition. For this DMC and the particular $\underline{p}$, Gallager's[8] upper bound on block error probability may be written as:

$$\overline{P}e = e^{-n E_{oij}(\rho, \underline{p}) - \rho R} \tag{54}$$

By comparison with Equations (25) and (53) we see that we may define a loss in reliability due to time variations, and this loss is wholly attributable to the rightmost term of Equation (53). In particular, from Equation (48), the maximum rate at which arbitrary probability of error is guaranteed is reduced due to time variations by an amount

$$\frac{\partial}{\partial \rho} \ln \left[\text{dominant eigenvalue of } (q_{ij}^{\frac{1}{1+\rho}}) \right] \Bigg|_{\rho = 0}$$

If the $\underline{p}$'s for which the channel is state independent include the $\underline{p}$ which yields capacity for the above

defined DMC, then it is fair to call the above term the loss in capacity due to time variations.

Chapter V

Sequential Decoding for the DFSC

A. The Ensemble of Tree Codes

In this chapter we study the decoder described in Section C of Chapter II. We will obtain upper bounds on the three quantities discussed there by random coding methods similar to those used in the preceeding chapter. We begin with a description of the ensemble of codes.

Consider an ensemble of tree codes with the following property. At each time, for each possible state of the encoder, the code symbols generated are selected independently from the common distribution $P(x)$. Furthermore, the symbols generated for any state of the encoder at any given time are selected independently of the symbols generated for any state of the encoder at any other time. Thus the symbols along any path of the tree are selected independently of each other from the common distribution $P(x)$. In addition, the symbols along any totally distinct paths are selected independently of each other from the common distribution $P(x)$ beyond the point in the tree at which they first become distinct. This property does not hold true for paths which are not totally distinct. Wherever

the encoder states along any paths are the same at the same time, the symbols generated at that time must be the same for each of these paths.

Now consider two paths diverging from a reference node in the tree code. Let $\underline{x}(n)$ denote the sequence of n symbols beyond the reference node along the path that the encoder happens to follow. Let $\underline{x}^*(n)$ denote the corresponding sequence of symbols along the other path. Furthermore, let $\underline{y}(n)$ and $\underline{d}(n)$ denote the corresponding sequences of received symbols and channel-state sequence that occurs, respectively. Then, over the ensemble of codes, the 4-tuple of vectors $(\underline{x}(n), \underline{x}^*(n), \underline{y}(n), \underline{d}(n))$ occurs with probability:

$$\Pr(\underline{d}(n)) \Pr(\underline{x}(n)) \Pr(\underline{x}^*(n)) \Pr(\underline{y}(n)/\underline{x}(n),\underline{d}(n))$$

$$= \Pr(d_o) \prod_{i=1}^{n} P(x_i)P(x^*_i)\, q_{d_{i-1},d_i}\, p(y_i/x_i,d_i) \qquad (1)$$

if the paths remain distinct up to length n. If the paths merge at length $k \leq n$, (i.e., have the same sequence of encoder states beyond k) then the 4-tuple occurs with probability:

$$\Pr(\underline{d}(n)) \Pr(\underline{x}(n)) \Pr(\underline{x}^*(k)) \Pr(\underline{y}(n)/\underline{x}(n),\underline{d}(n))$$

$$= \Pr(d_o) \prod_{i=1}^{k} P(x_i)\, P(x^*_i)\, q_{d_{i-1},d_i}\, p(y_i/x_i,d_i) \cdot \prod_{j=k+1}^{n} P(x_j)\, q_{d_{j-1},d_j}\, p(y_j/x_j,d_j) \qquad (2)$$

and $x_i = x^*_i$; i=k+1,k+2,...,n. (3)

The properties of this ensemble will be utilized in the chapter in much the same way that the properties of the ensemble of random block codes were used in the preceeding chapter.

B. Bounds on the Performance of the Decoder-Formulation

We will upper bound the quantities: average number of computations per node decoded, probability of failure, and average number of undetectable errors per node decoded in that order.

We assume, as will be established later, that the decoder will ultimately follow the correct path with some choice of channel state sequence (not necessarily correct).. Now suppose that at some time the decoder first arrives at a particular node on the correct path. We will take this node to be the reference node and take the metric along the correct path at this node to be L_1 for the particular state sequence accepted in first reaching this node. By virtue of the manner in which the decoder chooses its threshold, T, we must have that upon first arriving at this reference node

$$0 \leq L_1 - T < T_o \qquad (4)$$

Now consider the set of all incorrect paths stemming from the reference node. We will upper bound the average number of times the decoder passes

(See Figure 2.4)
through loop A with a hypothesized branch which lies in this set. In addition, we will bound the average number of times that the decoder passes through loop A with the hypothesized branch being the first branch along the correct path stemming from this reference node. If we find these bounds for each node along the correct path taken as a reference node, then we will have considered every possible branch in the tree. Furthermore, if the bounds for a particular reference node are independent of the reference node in question, then a bound on the average number of computations per node decoded will be given by the sum of the bounds computed for any particular reference node.

Now consider the flow chart in Figure 2.4. Any particular branch in the tree code corresponding to a particular message hypothesis can be tested at most B^{N_o+1} times in loop A with a given threshold in effect. This follows from the fact that the node from which the branch stems and the node to which it leads can each be the furthest node into the tree ever accepted along a particular path exactly once. Thus it can be tested with at most B^{N_o} channel state assumptions for the branch in question and B channel state assumptions for the preceeding symbol. We assume here that in choosing channel state sequences for test after first entering a backward mode we

need consider only one sequence for each state assumption for the final symbols on the branch under test. This is so because all state sequences with the same final state will lead to the same behavior of the metric L_n at future nodes. This property is not accounted for in the flow chart of Figure 2.4 because it only complicates the chart. It may be readily incorporated into a physical realization of the decoder so we assume it here. For the flow chart shown in the figure we may replace B^{No+1} by B^{2No}.

Let us clarify the above statements. The first time a given branch is tested in loop A the threshold, T, is increased by incriments of T_o until its value just fails to exceed the value of L_n at the node to which the branch leads. At this time F=0. By the threshold in effect we mean the final value of T reached in this process. Each subsequent time this branch is tested in loop A, we must have F=1 because we are testing this branch as a consequence of having previously followed loop B and reduced the threshold, or else having previously followed loop C. In testing the branch in question these subsequent times, the threshold is not raised in travelling through loop A and hence by the threshold in effect we mean the value of T which is held constant. Each time we test this given branch with a particular assumption on the channel state sequence along this

branch and a particular assumption on the channel state just prior to this branch, the threshold in effect must be different. This follows from the fact that having once tested a branch with a given threshold in effect we will only test this branch again if we are forced, by future events, to reduce the threshold in effect (i.e., we follow loop B). If we are not forced to lower the threshold then there exists a path along which the decoder can move without ever returning to the node from which the branch in question stems.

Now consider a particular node along the correct (actually transmitted) path in the tree. Take this node as a reference node and compute the metrics for all paths stemming from this node as if the metric for the correct path at this node were zero. Also for convenience, of future arguments assign this node order number 1. Let L_n (n=1,2,...) be the sequence of values assumed by the metric along the ultimately accepted path (the correct path) with the ultimately accepted channel state sequence hypothesis. Let

$$L_{min} = \min_n \; L_n \qquad (5)$$

Now consider a branch of order number k in the set of incorrect branches stemming from the reference node and let N(k) be the number of traversals of loop

A made by the decoder using this branch as a hypothesis. Let $L^*_{k+1}(\underline{d}^*(kN_o))$ be the value of the metric at the node in which this branch terminates for the channel state sequence hypothesis $\underline{d}^*(kN_o)$. Finally let A_j be a random variable with

$$A_j = 1 \; ; \text{ if } L^*_{k+1}(\underline{d}^*(kN_o)) \geq L_{min}+(j-1)T_o \text{ for any } \underline{d}^*(kN_o)$$
$$= 0 \; ; \text{ otherwise} \tag{6}$$

then we have:

Theorem 5.1:

$$N(k) \leq B^{N_o+1} \sum_{j=1}^{\infty} A_j \tag{7}$$

Proof: The smallest threshold in effect for which the branch in question is tested satisfies the inequality

$$0 \leq L_{min} - T \leq T_o \tag{8}$$

On the other hand, the largest threshold in effect for which the branch in question is tested, satisfies the inequality

$$0 \leq \max_{\underline{d}^*(kN_o)} L_k(\underline{d}^*(kN_o)) - T \leq T_o \tag{9}$$

Thus, the fact that the branch in question is tested with j different thresholds in effect implies that A_j=1. The theorem then follows from the discussion above.

Now let N(0) be the number of times the first branch on the correct path stemming from the reference node is tested in loop A. Let A_{oj} be defined as in Equation (6) with $L^*_{k+1} = 0$. Then without further comment we have:

Corollary 5.1:

$$N(0) \leq B^{N_o+1} \sum_{j=1}^{\infty} A_{oj} \qquad (10)$$

Now let us switch to a discussion of the probability of failure. Suppose that we impose a finite decoding constraint length, rN_o (recall that we assume $\nu = \infty$). Define $L_{min}(r)$

$$L_{min}(r) = \min_{n=1,2,\ldots,r+1} L_{min} \qquad (11)$$

where we assume that we know the ultimately accepted channel state sequence. We can then observe that the decoder can follow an incorrect path from the reference node from which it stems to a constraint length beyond, when the metric on this path lies above the threshold just below $L_{min}(r)$ at every node. Such an

event constitutes an error, and we denote its probability by Pe(r). It is the fact that we will be able to upper bound Pe(r) by a quantity that becomes arbitrarily small for large r that we can assume that the decoder ultimately accepts the correct path. Of course Pe(r) must be computed conditionally on the assumption that when the decoder first reaches the reference node in question it has not yet made an error. The situation here is analogous to that which arises in the calculation of the probability of error for the Wozencraft[21] sequential decoder.

Now define $L_{m,r+1}(\underline{d}^*(rN_o))$ for the m^{th} of the $\frac{1}{2}e^{rN_oR}$ incorrect paths of length rN_o stemming from a given reference node. Then define the variable $A_j(m)$ as in Equation (6) with L_{min} replaced by $L_{min}(r)$. We then have:

Theorem 5.2:

$$Pe(r) - 1 - \prod_{m=1}^{\frac{1}{2}e^{rN_oR}} (1-A_1(m)) \qquad (12)$$

Proof: The probability of the joint event that an incorrect path has a metric greater than a given value at each of a finite number of points is upper bounded by the probability that the metric exceeds this value at any one of these points. The product in Equation (12) will be zero when any one of the $A_1(m)$'s are equal to 1. Q.E.D.

A failure in the ability of the decoder to operate in accordance with its algorithm will occur if either the events leading to an error occur or if

$$L_{min} - L_{min}(rN_o) \leq 0 \qquad (13)$$

Thus defining $P_f(rN_o)$ as the probability of failure with the decoding constraint length rN_o we have, upon bounding the probability of Equation (13) by $A_{01}(rN_o)$

Corollary 5.2:

$$P_f(rN_o) \leq 1 - \prod_{m=1}^{\frac{1}{2}e^{rN_oR}} (1-A_1(m)) + A_{01}(rN_o) \qquad (14)$$

Finally let us consider the situation of undetectable errors. Let us assume that we know the ultimately accepted channel state sequence that is associated with the ultimately accepted correct path if $\nu = \infty$. Now for ν finite we can follow an incorrect path that diverges from the correct path at a particular reference node and remerges with the correct path (ceases to be distinct) at a node of order number $h + \frac{\nu}{\bar{\nu}_o}$ if the metric along this path lies above the threshold just below $L_{min}(h + \frac{\nu}{\bar{\nu}_o} - 1)$ at every node up to the point at which the correct path and this path remerge. If we follow such a path we make a sequence of $h\nu_o$ undetectable errors

along this path stemming from this particular reference node on the correct path.

There are $\frac{1}{2}e^{hN_oR}$ paths which remerge with the correct path at length (in nodes) $h+\nu/\nu_o$. Now, define $A_j(m)$ as in Equation(6). Then if $P_u(h)$ is the probability that the decoder follows a path stemming from a particular reference node along the correct path and makes h undetectable errors along this path; we have:

Theorem 5.3:

$$P_u(h) \leq 1 - \prod_{m=1}^{\frac{1}{2}e^{hN_oR}} (1-A_1(m)) \qquad (15)$$

The proof is essentially that of Theorem 5.2.

C. Bounds on the Properties of the Decoder - Analytical Results

In this section we obtain analytical bounds for the three quantities of interest. For the purposes of computing these bounds we assume that the path ultimately accepted is the correct path and that the accepted channel state sequence is the one that actually occurs. The significance of this assumption lies in the fact that we can then find bounds using methods similar to those used to find the bounds on Pe in the preceeding chapter.

It will be clear from the operations to be carried out in the remainder of this chapter that the bounds to be computed under this assumption are strictly valid if, beyond the reference node in question, the minimum value of the metric on the correct path with an assumed state sequence that actually occurs is less than or equal to the minimum value of the metric on the path that is ultimately accepted. If this situation does not occur, we note that the separation between these minimum values must always be finite. This statement follows from the fact that any arbitrary channel state sequence may be forced to merge, in a one step transition, with the state sequence that actually occurs. It will be clear in the operations below that such a finite difference introduces no significant alteration in our bounds.

Now consider the variable A_j for an incorrect path of kN_o symbols stemming from a reference node on the correct path. Let L_{n+1} be the value of the metric on the correct path (with the state sequence that actually occurs) and define the variable $A_{j,n}$ by

$$A_{j,n} = 1 \; ; \text{ if } L_{k+1}(\underline{d}^*(kN_o)) \geq L_{n+1} + (j-2)T_o$$

$$\text{for some } (\underline{d}^*(kN_o)) \qquad (16)$$

$$= 0 \; ; \text{ otherwise}$$

Then we have:

Lemma 5.1:

$$A_j \leq \sum_{n=1}^{\infty} A_{j,n} \tag{17}$$

Proof: For some n we must have

$$L_{n+1} = L_{min} \tag{18}$$

Thus, if $A_{j,n} = 0$ for all n we must have $A_j = 0$. If $A_{j,n} = 1$ for any n the right hand side of Equation (17) is greater than or equal to 1 and hence overbounds A_j.

Now consider the variable $A_j(m)$ for the m^{th} of M incorrect paths of kN_o symbols stemming from a reference node on the correct path. Define the variable $A_{j,n}(m)$ by

$$A_{j,n}(m) = 1 \text{ ; if } L_{m,k+1}(\underline{d}^*(kN_o) \geq L_{n+1} + (j-2)T_o$$

$$\text{for some } \underline{d}^*(kN_o) \tag{19}$$

$$A_{j,n}(m) = 0 \text{ ; otherwise}$$

Then we have

Lemma 5.2:

$$1 - \prod_{m=1}^{M} (1-A_j(m)) \leq \sum_{n=1}^{k} \left[1 - \prod_{m=1}^{M} (1-A_{j,n}(m)) \right] \tag{20}$$

The proof is simply an elaboration of the preceeding proof and is omitted.

At this point we introduce some additional notation: Define the sequences $\underline{x}(n-k) \in X^{n-k}$ having components

$$\underline{x}(n-k) = (x_{k+1}, x_{k+2}, \ldots . x_n) \tag{21}$$

In like manner define the sequences $\underline{y}(n-k)$ and $\underline{d}(n-k)$. Now let the sequence of transmitted symbols corresponding to the m^{th} of the incorrect paths be $\underline{x}_m(n)$ and let $\underline{x}(n)$ be the sequence of transmitted symbols corresponding to the correct path. In addition, recalling the metric defined in Chapter II, Section C, we let

$$F(\underline{y}(k)) = \prod_{i=1}^{k} f(y_i) \; ; \; F(\underline{y}(n-k)) = \prod_{i=k+1}^{n} f(y_i) \tag{22}$$

Then we have:

Theorem 5.4:

If $n \geq k$

$$\overline{1 - \prod_{m=1}^{M} (1 - A_{j,n}(m))} \leq JM^{\rho} e^{\frac{-(j-2)\rho T_o}{1+\rho}} e^{-KE_o(\rho, \underline{p})} \cdot e^{-(N-K)\left[\frac{E_o(\rho, \underline{p}) - \rho U}{1+\rho}\right]} \; ; \; 0 \leq \rho \leq 1 \tag{23}$$

If $k \geq n$

$$1 - \prod_{m=1}^{M} (1 - A_{j,n}(m)) \leq JM^{\sigma} e^{-\frac{(j-2)\sigma T_o}{1+\sigma}} e^{-NE_o(\sigma, \underline{p})} \cdot e^{-(K-N)\left[E_1(\sigma, \underline{p}) + \frac{\sigma}{1+\sigma}(E_o(\sigma, \underline{p}) + U)\right]}$$

$$; \ 0 \leq \sigma \leq 1 \qquad (24)$$

Here $K = kN_o$, $N = nN_o$ (25)

$E_o(\rho, \underline{p})$ is given by Equation (4.26) and

$e^{-E_1(\sigma, \underline{p})}$ = dominant eigenvalue of the matrix $G(\sigma)$ (26)

$$G(\sigma) = (q_{ij} F_{ij}(\sigma)) \qquad (27)$$

where $$F_{ij}(\sigma) = \left[\sum_Y f(y) \left\{\sum_X P(x) P(y/x,i,j)\right\}^{-\sigma}\right]^{\frac{1+\sigma}{1+\sigma}} \qquad (28)$$

The constant J is independent of M, N, and K.

The proof of the above Theorem is given in the Appendix.

This Theorem plays the role here played by Lemma 4.2 and Theorem 4.4 in the previous chapter. We may use it to derive the following important theorems:

Theorem 5.5:

The maximum rate at which the average number of computations per node decoded converges, for $\nu = \infty$, is bounded below by $R_{comp}(U)$. Here $R_{comp}(U)$ is given by the smaller of the following:

a) $E_o(\rho^*, p)$ where ρ^* is the largest value of $\rho \leq \rho_o$ for which

$$U \leq \frac{E_o(\rho, p)}{\rho} \tag{29}$$

(Recall from Figure 4.3 that $E_o(\rho_o, p) = \max_{0 < \rho \leq 1} E_o(\rho, p)$)

b) $$\max_{0 < \sigma \leq 1} \min \left\{ E_1(\sigma, p) + \frac{\sigma}{1+\sigma} (E_o(\sigma, p) + U), E_o(\sigma, p) \right\}$$

Proof: Let $\overline{N}(k)$ be the average over the ensemble of codes, channel outputs and channel state sequences of $N(k)$. We may obtain an upper bound on $\overline{N}(k)$ from Theorem 5.1 and Lemma 5.1 by using the bounds of

Theorem 5.4 with M = 1. This upper bound is independent of which particular incorrect branch of order number k we discuss. There are $\frac{1}{2} e^{kN_oR}$ such branches and hence, the total average computation will be bounded by:

$$\sum_{k=0}^{\infty} \sum_{n=1}^{\infty} \sum_{i=1}^{\infty} e^{kN_oR} \overline{A}_{j,n}$$

We introduce the bounds of Theorem 5.4 and sum on j first. Observing that

$$\sum_{j=1}^{\infty} e^{-j\frac{\rho}{1+\rho} T_o}$$

is finite for all $\rho > 0$ we may eliminate this sum from further consideration.

Next we split the sums on k and n as follows:

$$\sum_{k=0}^{\infty} \sum_{n=1}^{\infty} e^{kN_oR} \overline{A}_{j,n} = \sum_{k=0}^{\infty} \sum_{n-k=0}^{\infty} e^{kN_oR} \overline{A}_{j,n}$$

$$+ \sum_{n=1}^{\infty} \sum_{k-n=0}^{\infty} e^{kN_oR} \overline{A}_{j,n} \quad (30)$$

Using the bound of Equation (23) in the first sum on the right hand side of Equation (30) we see that this sum will converge if both

$$E_o(\rho, \underline{p}) - R > 0$$

and

$$E_o(\rho, \underline{p}) - \rho U > 0 \qquad (31)$$

Thus condition a) of the theorem is established upon recalling the properties of $E_o(\rho, \underline{p})$ from Chapter IV.

In like manner, we introduce Equation (24) into the second sum on the right hand side of Equation (30) and derive the following conditions for convergence:

$$E_o(\sigma, \underline{p}) - R > 0$$

$$E_1(\sigma, \underline{p}) + \frac{\sigma}{1+\sigma} \left(E_o(\sigma, \underline{p}) + U\right) - R > 0 \qquad (32)$$

Q.E.D.

We now prove:

Theorem 5.6:

The ensemble average $\overline{Pe}(r)$ of $Pe(r)$ is bounded by

$$\overline{Pe}(r) \leq J_1 \, e^{-rN_o \left[E_o(\sigma, \underline{p}) - \sigma R\right]} \qquad (33)$$

where J_1 is a constant, and

$$E_1(\sigma, \underline{p}) + \frac{\sigma U - E_o(\sigma, \underline{p})}{1+\sigma} \geq 0 \qquad (34)$$

If Equation (34) is not true, then

$$\overline{Pe}(r) \leq J_2\, e^{-rN_o\left[E_1(\sigma,p)+\frac{\sigma(E_o(\sigma,p)+U)}{1+\sigma}\ -\sigma R\right]} \tag{35}$$

where J_2 is a constant.

Proof: Substitute the bound of Lemma 5.2 into Theorem 5.2.

Then apply the bound of Equation (24) with $M=e^{rN_oR}$ and $k=r$. Next sum over k-n from 0 to r-1. If the condition of Equation (34) is met then the sum is not exponential in r and just contributes to the constant J_1. If Equation (34) is not true then the sum contributes an exponential factor to yield Equation (35). This factor is determined by the identity

$$\sum_{i=0}^{n-1} e^{ix} = \frac{e^{nx}-1}{e^{x}-1} \leq \frac{e^{nx}}{e^{x}-1} \tag{36}$$

Note that a similar bounding of $A_{o1}(rN_o)$ does not contribute an exponentially poorer term and thus $P_f(rN_o)$ has the same bound (except for a different constant) as $Pe(rN_o)$.

Finally we prove

Theorem 5.7: Let $\overline{Nu}$ be the average number of undetectable errors made per node decoded. Then $\overline{Nu}$ is bounded by

$$\overline{Nu} \leq J\, e^{-\nu/\nu_o\ N_oR_{comp}(U)} \tag{37}$$

where J is a constant.

Proof: Upon setting

$M=\frac{1}{2}\, e^{-\nu/\nu_o N_o R}\, e^{(h+\nu/\nu_o)N_o R}$ in Equations (23) and (25) we use them to bound the right hand side of Equation (15) via Lemma 5.2. Thus, ignoring constants, we have

$$\overline{Pu}(h+\nu/\nu_o) \leq e^{-\nu/\nu_o N_o R\sigma}\, e^{-(h+\nu/\nu_o)N_o\left[E_o(\sigma,\underline{p})-\sigma R\right]}$$

$$\sum_{i=0}^{(h+\nu/\nu_o-1)N_o} e^{-i\left[E_1(\rho,\underline{p})+\frac{\sigma U-E_o(\sigma,\underline{p})}{1+\sigma}\right.}$$

$$+\, e^{-\nu/\nu_o N_o R\rho}\, e^{-(h+\nu/\nu_o)N_o\left[E_o(\rho,\underline{p})-\rho R\right]}$$

$$\cdot \sum_{i=1}^{\infty} e^{-i\left[\frac{E_o(\rho,\underline{p})-\rho U}{1+\rho}\right]} \qquad (38)$$

The last sum will converge only if

$$U \leq \frac{E_o(\rho,\underline{p})}{\rho}$$

and have the bound

$$e^{-(\nu/\nu_o)N_o E_o(\rho,\underline{p})}\; e^{-hN_o\left[E_o(\rho,\underline{p})-\rho R\right]}$$

We may treat the first term on the right hand side exactly in accordance with the proof of the previous theorem. Thus if Equation (34) is true, the term in

question is bounded by $e^{-\nu/\nu_o N_o E_o(\sigma,\underline{p})}$ x [right hand side of Equation (33) with r=h].

If Equation (34) is not true, it is bounded by

$$e^{-\nu/\nu_o N_o \left[E_1(\sigma,\underline{p}) + \sigma \frac{(E_o(\sigma,\underline{p})+U}{1+\sigma}\right]} \quad x$$

[right hand side of Equation (35) with r=h].

Now form

$$\overline{Nu} = \sum_{h=0}^{\infty} h\, \overline{Pu}(h+ \nu/\nu_o) \tag{38}$$

Observe that the conditions for convergence of this sum are <u>less</u> stringent than the requirement

$$R < R_{comp}(U) \tag{39}$$

The conditions on the exponent which dominates the resultant expression are precisely those that determine $R_{comp}(U)$. Q.E.D.

D. <u>Discussion</u>

The bounds developed here are unfortunately left in terms of the function $E_1(\rho,\underline{p})$ which depends on f(y). Note, however, that if f(y) can be chosen such that

$$E_1(\rho,\underline{p}) \geq 0 \tag{40}$$

we may draw the conclusion (in the non-trivial case $C \geq 0$):

Theorem 5.8:

If $E_1(\rho,\underline{p}) \geq 0$ then

$$R_{comp} = \max_{U} R_{comp}(U) = E_o(\rho_o,\underline{p}) \tag{41}$$

Furthermore, for

$$\frac{E_o(\rho_o,\underline{p})}{\rho_o} \leq U \leq \frac{\partial}{\partial\rho} E_o(\rho,\underline{p}) \Big|_{\rho = 0} \tag{42}$$

$$R_{comp}(U) = E_o(\rho^*,\underline{p}) \tag{43}$$

where $$U = \frac{E_o(\rho^*,\underline{p})}{\rho^*} \tag{44}$$

Proof:

For ρ^* chosen as in Equation (44) we have:

$$E_1(\rho^*,\underline{p}) + \frac{\rho^*}{1+\rho^*}\,(E_o(\rho^*,\underline{p})+U) = E_1(\rho^*,\underline{p})+E_o(\rho^*,\underline{p})-E_o(\rho^*,\underline{p}) \tag{45}$$

Now $$\frac{\partial}{\partial\rho}\,\frac{E_o(\rho,\underline{p})}{\rho} = \frac{\rho E_o'(\rho)-E_o(\rho)}{\rho^2} \leq 0 \;;\; 0 \leq \rho \leq \rho_o \tag{46}$$

This follows from the properties of the block coding exponents of Chapter IV. The numerator of the above expression is just the negative of our block coding exponents.

Thus ρ^* in Equation (44) must be the largest value of ρ for which

$$U \leq \frac{E_o(\rho)}{\rho} \tag{47}$$

The conditions on R_{comp} are then satisfied by the choice of the theorem statement.

Next let us consider the bound on $\overline{Pe}(r)$. We have Theorem 5.9:

Suppose $E_1(\sigma, p) \geq 0$. Let σ^* be such that

$$E_o(\sigma^*, p) - \sigma^* R = \max_{\sigma} E_o(\sigma, p) - \sigma R = E(R, p) \tag{48}$$

Then for all U such that

$$E_1(\sigma^*, p) + \frac{\sigma^* U - E_o(\sigma^*, p)}{1+\sigma^*} \geq 0 \tag{49}$$

We have

$$\overline{Pe}(r) \leq K\, e^{-rN_o E(R,p)} \tag{50}$$

For all U such that Equation (49) is not correct, we have

$$\overline{Pe}(r) \leq K\, e^{-rN_o(R_{comp}(U) - \sigma_o R)} \tag{51}$$

where σ_o is such that the minimum in condition b) for R_{comp} is achieved!

Proof: The theorem follows directly from inspection of Theorem 5.6 and the conditions for $R_{comp}(U)$. The only question that arises is whether for any (R,U) pair Equation (49) can ever be satisfied. Observe that for $\sigma^* = \rho_o$, Equation (49) will be satisfied for all

$$U \geq \frac{E_o(\rho_o)}{\rho_o} \tag{52}$$

This result is non-trivial for the case $\rho_o = 1$.

It is now appropriate to show that there exists channels for which f(y) may be adjusted such that $E_1(\rho,p) \geq 0$. For channels with input rotations we have that $\sum_{x=1}^{K} P(x)\ P(y/x,i,j)$ is independent of i and j. Thus for these channels we may pick

$$f(y) = \sum_{x=1}^{K} P(x)\ P(y/x,i,j) \tag{53}$$

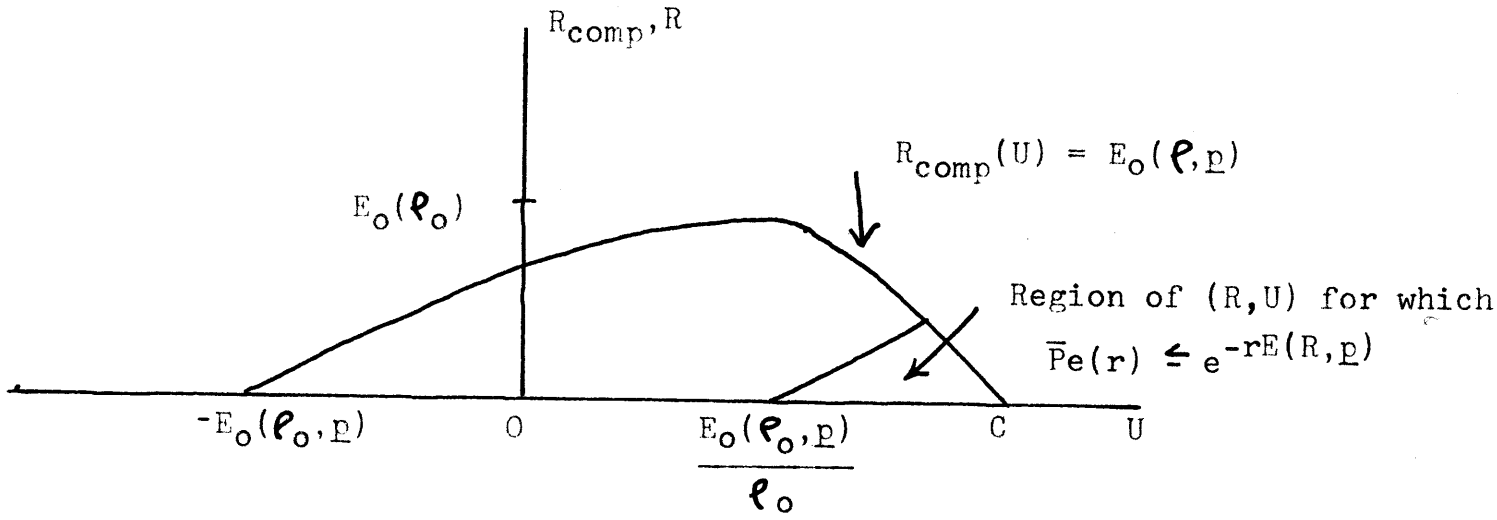
This choice insures the desired property of $E_1(\rho,p) = 0$ as may be checked from the definition in Theorem 5.4. We note here that a DMC is a channel with input rotations, so that the results obtained in this chapter may be specialized to that case. An analysis of the decoder's performance on the DMC has been carried out independently by Stiglitz (unpublished). The results in this case may be made somewhat stronger than those contained here because it is not necessary

to make some of the approximations carried out in the proof of Theorem 5.4. The resultant value of R_{comp}, however, is unchanged.

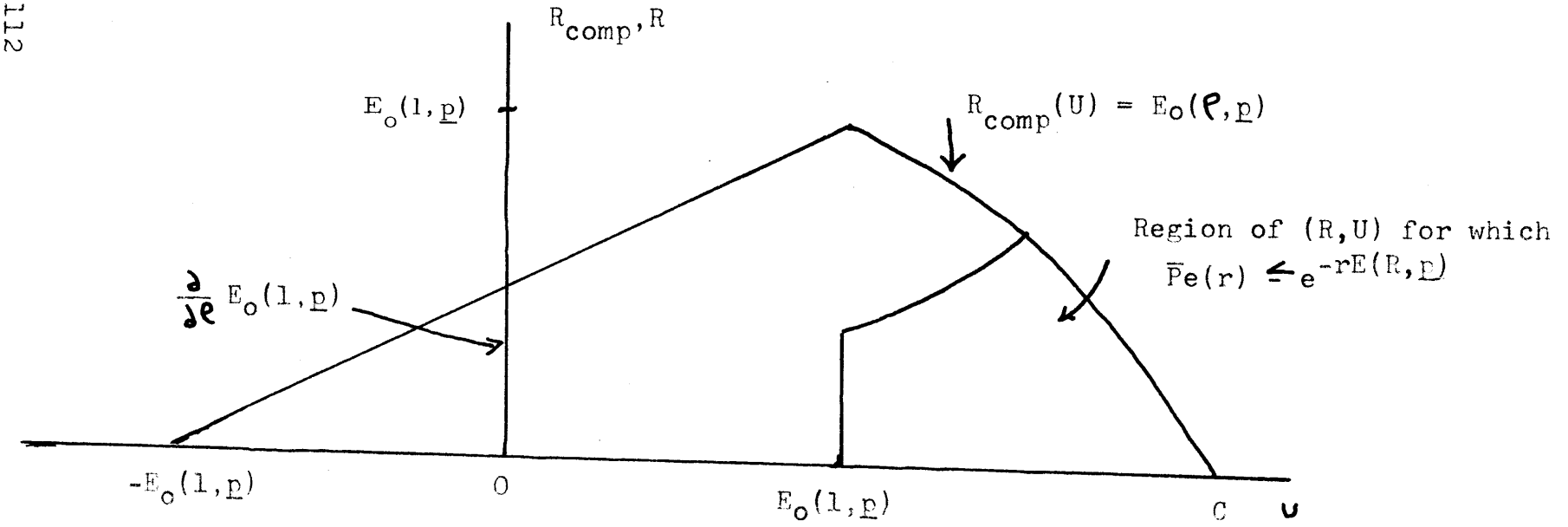
In Figure 5.1 we show the results in a graphical form.

E. Final Comments

In this chapter we have analyzed a sequential decoder for the general DFSC. The assumption that every state was reachable from every other state was used only in Theorem 5.1. It is clear that although this assumption was not unphysical in the first place, it may be removed in certain special cases (for example the unphysical channels in which the output determines the state). The primary reason for not discarding this assumption is that the resultant decoder (Figure 2.4) gains efficiency in not having to examine large numbers of unnecessary state sequences. On this note we close the chapter.



a) $\rho_o < 1$



b) $\rho_o = 1$

Figure 5.1 $R_{comp}(U)$ versus U for $E_1(\rho, p) = 0$

Chapter VI

Concluding Remarks

We have presented an analysis of a class of channels which might be alternately described as channels with memory or time-varying channels. This analysis has been tied to a particular approach to decoding for these channels. As has been pointed out earlier in the thesis, our results can be viewed as a starting point for further analyses aimed at removing deficiencies left and answering questions posed here.

Beyond the particular suggestions for further research made in the thesis, we might add:

1) An attempt to find lower bounds to the probability of error attainable with block codes for the DFSC.

2) A study of higher order moments of computation for sequential decoding for the DFSC.

3) A study of channels in which the statistics of the underlying Markov chain are dependent on the input symbol.

4) An extension of the results here to continuous channels.

Finally, let us note that theory in the absence of experimentation is just theory.

Appendix

Proof of Theorem 5.4

Let $\underline{x}(nN_o)$ be the sequence of nN_o transmitted symbols along the correct path stemming from the reference node. Let $\underline{x}_m(kN_o)$ be the corresponding sequence of kN_o symbols along the m^{th} incorrect path. Finally let $\underline{d}(n)$ be the state sequence that actually occurs. Then, we have by definition:

$$\left(\neq - \prod_{m=1}^{M} (1-A_{j,n}(m)) = 1 \; ; \text{ if } e^{L_{m,k+1}(\underline{d}^*(K))-L_{min}(n)-(j-2)T_o} \geq 1\right.$$

$$\text{for any } m \leq M \text{ and any } \underline{d}^*(K)$$

$$= 0 \; ; \text{ otherwise} \qquad (1)$$

Here $N = nN_o$ and $K = kN_o$

Then we have:

$$1-\prod_{m=1}^{M}(1-A_{j,n}(m)) \leq \left(\sum_{m=1}^{M} \sum_{D^*K} e^{\frac{1}{1+\rho}\left[L_{m,k+1}(\underline{d}^*(K))-L_{min\ n}-(j-2)T_o\right]}\right)^{\rho}$$

$$; 0 \leq \rho \qquad (2)$$

This follows from the fact that when $1-\prod_{m=1}^{M}(1-A_{j,n}(m))=1$ at least one term in the sum on the right hand side of Equation (2) must be greater than or equal to 1.

On the other hand, the right hand side is always positive. Then recalling the definitions of the L's from Equation (2.19) and (2.20) we have:

$$e^{L_{m,k+1}(\underline{d}^*(K))} = \frac{Pr(\underline{y}(K)/\underline{x}_m(K),\underline{d}(K))Pr(\underline{d}(K))}{F(\underline{y}(K))} \; e^{-KU} \tag{3}$$

and a similar expression for $L_{min}^{(n)}$. Thus

$$1-\prod_{m=1}^{M}(1-A_{j,n}(m))$$

$$\leq e^{(N-K)\frac{\rho}{1+\rho}U - (j-2)\frac{\rho}{1+\rho}T_o} \cdot \left[\frac{F(\underline{y}(K))}{F(\underline{y}N)}\right]^{\frac{\rho}{1+\rho}}$$

$$\cdot\left\{\sum_{m=1}^{M}\sum_{D^*K}\frac{Pr(\underline{y}(K)/\underline{x}_m(K),\underline{d}^*(K))^{\frac{1}{1+\rho}}\; Pr(\underline{d}^*(K))^{\frac{1}{1+\rho}}}{Pr(\underline{y}(N)/\underline{x}(N),\underline{d}(N))\; Pr(\underline{d}(N))^{1/1+\rho}}\right\}^{\rho} \tag{4}$$

We now average the right hand side of Equation (4). We must consider the separate cases $N \geq K$ and $K \geq N$.

Case I: $N \geq K$

In this case the ensemble statistics are:

$$\Pr(\underline{y}(N)/\underline{x}(N),\underline{d}(N))\ \Pr(\underline{d}(N))\ \prod_{m=1}^{M} \Pr(\underline{x}_m(K))$$

Introduce the symbols $\underline{x}(N-K)$, etc. as in Chapter V and observe:

$$1-\prod_{m=1}^{M}(1-A_{j,n}(m)) \leq e^{(N-K)\frac{\rho}{1+\rho}}\ U\ e^{-(j-2)\frac{\rho}{1+\rho}}\ T_o^{\ M\rho}$$

$$\cdot\left\{\sum_{Y^K}\left(\sum_{D^K}\Pr(\underline{d}(K)^{\frac{1}{1+\rho}}\sum_{X^K}\Pr(\underline{x}(K))\Pr(\underline{y}(K)/\underline{x}(K),\underline{d}(K))^{\frac{1}{1+\rho}}\right)\right.$$

$$\left(\sum_{D^*K}\Pr(\underline{d}^*(K)^{\frac{1}{1+\rho}}\sum_{X^*K}\Pr(\underline{x}^*(K)\Pr(\underline{y}(K)/\underline{x}^*(K),\underline{d}^*(K))^{\frac{1}{1+\rho}}\right)^{\rho}$$

$$\left[\sum_{D^{N-K}}\Pr(\underline{d}(N-K)/d_K)^{\frac{1}{1+\rho}}\sum_{Y^{N-K}}F(\underline{y}^{N-K}))\sum_{X^{N-K}}\Pr(\underline{x}(N-K))\right.$$

$$\left.\left.\left\{\frac{\Pr(\underline{y}(N-K)/\underline{x}(N-K),\underline{d}(N-K))}{F(\underline{y}(N-K))}\right\}^{\frac{1}{1+\rho}}\right]\right\} \qquad (5)$$

Here we have used the factorization properties

$$\Pr(\underline{y}(N)/\underline{x}(N),\underline{d}(N)) = \Pr(\underline{y}(K)/\underline{x}(K),\underline{d}(K)\ )$$

$$\Pr(\underline{y}(N-K)/\underline{x}(N-K),\underline{d}(N-K),d_K) \quad (6)$$

etc., and averaged as in the case of the proof of Lemma 4.2.

Next recognize that $F(\underline{y}(N-K)\ )$ is a probability distribution and use Equation (3.6) with $\lambda = \frac{1}{1+\rho}$ to establish

$$\sum_{D^{N-K}} \Pr(\underline{d}(N-K))^{\frac{1}{1+\rho}} \sum_{Y^{N-K}} F(\underline{y}(N-K)) \left[\sum_{X^{N-K}} \Pr(\underline{x}(N-K)) \cdot \left\{ \frac{\Pr(\underline{y}(N-K)/\underline{x}(N-K),\underline{d}(N-K),d_K)}{F(\underline{y}(N-K))} \right\}^{\frac{1}{1+\rho}} \right]^{\frac{1+\rho}{1+\rho}}$$

$$\leq \sum_{D^{N-K}} \Pr(\underline{d}(N-K))^{\frac{1}{1+\rho}} \left[\sum_{Y^{N-K}} \left\{ \sum_{X^{N-K}} \Pr(\underline{x}(N-K)) \right. \right.$$

$$\left. \left. \Pr(\underline{y}(N-K)/\underline{x}(N-K),\underline{d}(N-K),d_K)^{\frac{1}{1+\rho}} \right\}^{1+\rho} \right]^{\frac{1}{1+\rho}} \quad (7)$$

By comparison with Equation (4.33) and with the aid of Corollary 2.1, we may bound the right hand side of Equation (6) by $e^{-(N-K)\frac{E_o(\rho,\underline{p})}{1+\rho}}$. The remainder of the right hand side of Equation (5) may be compared with Equation (4.4). With the help of Theorem 4.4 we may bound this term to yield the desired result (Equation (5.23)).

Case II: $K \geq N$

Here the ensemble statistics are:

$$\Pr(\underline{y}(K)/\underline{x}(K),\underline{d}(K))\ \Pr(\underline{d}(K))\ \Pr(\underline{x}(K)) \prod_{m=1}^{M} \Pr(\underline{x}_m(K))$$

We may handle this case by interchanging the roles of N and K in Equation (5) and replacing the factor that appears on the left hand side of Equation (6) by the factor:

$$\sum_{D^{K-N}} \Pr(\underline{d}(K-N)) \sum_{Y^{N-K}} \sum_{X^{K-N}} \Pr(\underline{x}(K-N))\Pr(\underline{y}(K-N)/\underline{x}(K-N),\underline{d}(K-N),d_K)$$

$$\left\{\sum_{D^{*K-N}} \Pr(\underline{d}^*(K-N))^{\frac{1}{1+\rho}} \sum_{X^{*K-N}} \Pr(\underline{x}^*(K-N)\ )\right.$$

$$\left.\left[\frac{\Pr(\underline{y}(K-N)/\underline{x}^*(K-N),\underline{d}^*(K-N),d^*_K)}{F(\underline{y}(K-N))}\right]^{\frac{1}{1+\rho}}\right\}^{\rho}$$

We now apply Holder's inequality (Equation (3.3)) with $\lambda = \frac{1}{1+\rho}$ to upper bound this factor by:

$$\sum_{D^{K-N}} \Pr(\underline{d}(K-N)) \left\{ \sum_{Y^{K-N}} F(\underline{y}(K-N))^{-\rho} \left[\sum_{X^{K-N}} \Pr(\underline{x}(K-N)) \Pr(\underline{y}(K-N)/\underline{x}(K-N),\underline{d}(K-N),d_K) \right]^{1+\rho} \right\}^{\frac{1}{1+\rho}}$$

$$\left\{ \sum_{Y^{K-N}} \left[\sum_{D^*{}^{K-N}} \Pr(\underline{d}^*(K-N))^{\frac{1}{1+\rho}} \left(\sum_{X^*{}^{K-N}} \Pr(\underline{x}(K-N)) \right. \right. \right.$$

$$\left. \left. \left. \Pr(\underline{y}(K-N)/\underline{x}^*(K-N),\underline{d}^*(K-N),d^*_K)^{\frac{1}{1+\rho}} \right) \right] ^{1+\rho} \right\}^{\frac{\rho}{1+\rho}}$$

Comparing the rightmost factor of the above with Equation (4.4) and then applying the results of Theorem 4.4, we may bound this factor by $e^{-(K-N)E_0(\rho,\underline{p})\frac{\rho}{1+\rho}}$. The remaining factor of the above may be handled by Corollary (2.1) to yield the bound $e^{-(K-N)E_1(\rho,\underline{p})}$. The rest of the proof proceeds as in Case I to obtain the desired result (Equation (5.25)).

Q.E.D.

The bounding methods presented here and in Lemma 4.2 are closely related to the technique of generating function bounds which have been widely used in the past. Discussion of this latter technique is given by Fano[4]. The sharpness of such techniques may be proven by means of the "Central Limit Theorem with Large Deviations" due to Shannon (unpublished). A good presentation of an independent derivation of this theorem is given by Blackwell and Hodges[3].

Bibliography

1) Blackwell, D., L. Breiman and A.J. Thomasian: "Proof of Shannon's transmission theorem for finite-state indecomposable channels", Annals of Math. Stat., 29, No. 4, pp. 1209-1220 (1958).

2) Blackwell, D. and M.A. Girshick, "Theory of Games and Statistical Decisions, John Wiley and Sons, New York, 1954, Chapt.2.

3) Blackwell, D. and J.L.Hodges, "The Probability in the Extremal Tail of a Convolution", Annals of Mathematical Statistics, 30, pp. 1113-1120, 1959.

4) Fano, R.M., Transmission of Information, John Wiley and Sons, Inc., and Technology Press, Inc., New York, 1961.

5) Fano, R.M., "A Heuristic Discussion of Probabilistic Decoding", IEEE Trans. on Information Theory, IT-9, pp. 64-74, 1963.

6) Fano, R.M., MIT course 6.575-6.626 notes, 1963 (unpublished).

7) Feller, W., An Introduction to Probability Theory and its Application, 2nd Ed., John Wiley and Sons, Inc., New York, 1957, Chapt. 15.

8) Gallager, R.G., "A Simple Derivation of the Coding Theorem and Some Applications", to be published in IEEE Trans. on Information Theory.

9) Gantmacher, F.R., Applications of the Theory of Matrices, Interscience Publishers, New York, 1959, Chapt. 3.

10) Hardy, G.H., J.E. Littlewood and G. Polya, Inequalities, Cambridge, Univ. Press, 1962.

11) Horstein, M. "An Experimental Study of Sequential Decoding for the Binary Symmetric Channel", MIT Lincoln Laboratory Group Report, 34-74, Nov. 20, 1958.

12) Kailath, T., "Correlation Detection of Signals Perturbed by a Random Channel", IRE Trans. on Information Theory, IT-6, pp. 361-367, 1960.

13) Kennedy, R.S., "Finite-State Binary Symmetric Channels", Sc.D. Thesis, MIT, Dept. of Electrical Engineering, Jan. 1963.

14) Kuhn, H.W. and A.W. Tucker, "Nonlinear Programming", Second Berkeley Symposium on Mathematical Statistics and Probability, J. Neyman (Ed.), p. 181, 1951.

15) Price, R. and P.E. Green, "A Communication Technique for Multipath Channels", IRE Proceeding, 46, pp. 555-569, 1958.

16) Reiffen, B., "Sequential Encoding and Decoding for the Discrete Memoryless Channel", MIT Research Laboratory of Electronics, Tech. Rept. No. 374. Aug. 1960.

17) Rosenblatt, M., Random Processes, Oxford, Univ. Press, New York, 1962.

18) Shannon, C.E., "Certain Results in Coding Theory for Noisy Channels", Information and Control, 1, pp. 6-25, 1951.

19) Shannon, C.E., Gallager, R.G. and Berlekamp, E., "A Lower Bound to the Probability of Error for Block Codes for the Discrete Memoryless Channel", to be published.

20) Turin, G.L., "On Optimal Diversity Reception", IRE Trans. on Information Theory, IT-7, pp. 154-166, 1961.

21) Wozencraft, J.M. and Reiffen, B., Sequential Decoding, Technology Press, and John Wiley and Sons, New York, 1961.

22) Wozencraft, J. M. and K.M. Perry, "Seco: A self-regulating error correcting coder-decoder", IRE Trans. on Information Theory, IT-8, pp. S128-135.

23) Wozencraft, J.M. et.al., "Application of Sequential Decoding to High-Rate Data Communication on a Telephone Line", IEEE Trans. on Information Theory, IT-9, pp. 124-126, 1963.

24) Van Trees, H.L., "Optimum Power Division in Coherent Communication Systems", MIT Lincoln Laboratory, Tech. Rept. No. 301, Feb. 1963.

BIOGRAPHICAL NOTE

Howard Louis Yudkin was born on June 22, 1936 in Philadelphia, Pennsylvania. Upon graduation from Central High School, Philadelphia, in 1953, he entered the Moore School of Electrical Engineering at the University of Pennsylvania where he received his B.S.E.E. degree in 1957. Mr. Yudkin then entered the Massachusetts Institute of Technology as an MIT Lincoln Laboratory Staff Associate and received his S.M. degree in 1959. From 1959 until 1962, Mr. Yudkin was a staff member of the MIT Lincoln Laboratory. During this time, he was primarily engaged with the study and design of communications systems.

In 1962 Mr. Yudkin reentered MIT as a Lincoln Laboratory Staff Associate. His previous experience includes work at the Frankford Arsenal in fire control systems, at the Moore School in radar warning systems, and at IBM in thin magnetic films.

Mr. Yudkin now resides in Framingham, Massachusetts with his wife and two daughters.

Publications of the Author

1) "The A Posteriori Probability Computing Channel" Lincoln Lab. Report 2G-25-7, Sept. 12, 1958 with B. Reiffen.

2) "Position Location with Multi-Terminal Antenna Systems", Master's Thesis, MIT, May 25, 1959.

3) "Impulse Noise on an H-44 Telephone Circuit", Lincoln Lab. Report 25G-0012, April 1960 with R. Pfeiffer.

4) "Some Results in the Measurement of Impulse Noise on Several Telephone Circuits", Proc. National Electronics Conference, 1960.

5) "The Design of an 'Error Free' Data Transmission System for Telephone Circuits", AIEE Trans. Communications and Electronics, July 1961, with B. Reiffen and W. Schmidt.

6) "Signal Selection", Lincoln Lab. Div. 2 Quarterly Progress Report June 1961.

7) "Experiments in the Improvement of the Impulse Response of Telephone Circuits", Lincoln Lab. Report 25G-4 Nov. 27, 1961.

8) "Infrared Pulsed and CW Range Radar", Lincoln Lab. Apollo Report AP-14, Sept. 25, 1962 with B. Goldstein, M. Macrakis and N. Sullivan.

9) "A Note on Signalling Over Spread Channels", Lincoln Lab. Memo 65L-007, Sept. 63 with B. Reiffen.

10) "An Error Bound for Gaussian Signals in Gaussian Noise", MIT Research Laboratory of Electronics, Quarterly Progress Report No. 73, pp. 149-155, April, 1964.