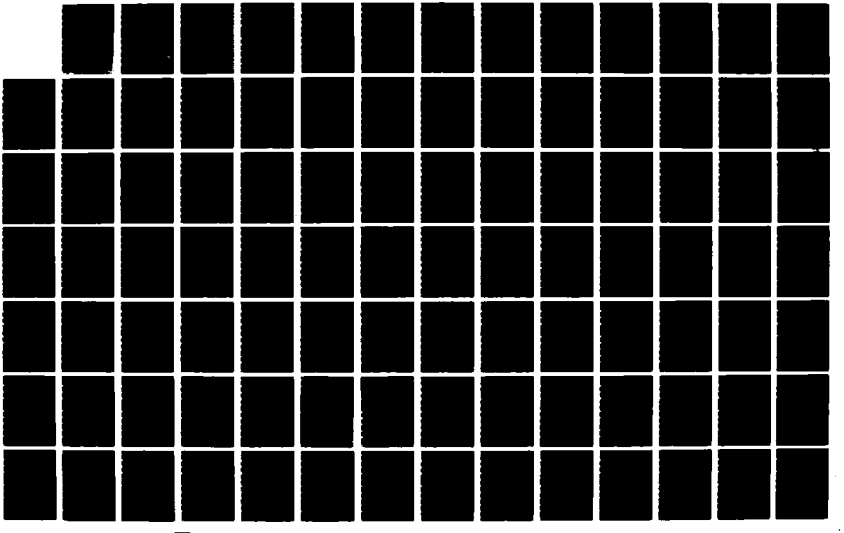
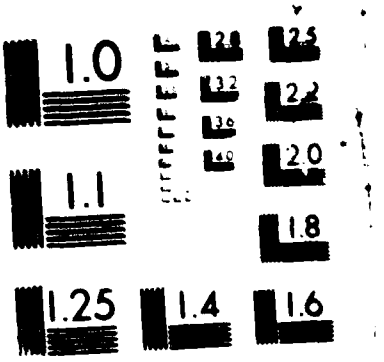
AD-A186 069 PERFORMANCE ANALYSIS OF CODED FREQUENCY-HOPPED SPREAD-SPECTRUM SYSTEMS WI (U) MICHIGAN UNIV ANN ARBOR COMMUNICATIONS AND SIGNAL PROCESSING L M V HEDGE

1/2

UNCLASSIFIED AUG 87 022793-3-T N00014-85-K-0545 F/G 17/4 NL




1.0
1.1
1.25
1.4
1.6
2.0
2.2
2.5

Report 022793-3-T

PERFORMANCE ANALYSIS OF CODED FREQUENCY-HOPPED SPREAD-SPECTRUM SYSTEMS WITH UNKNOWN INTERFERENCE

M. Hegde

COMMUNICATIONS & SIGNAL PROCESSING LABORATORY
Department of Electrical Engineering and Computer Science
The University of Michigan
Ann Arbor, Michigan 48109

August 1987

Technical Report No. 250

Prepared for
OFFICE OF NAVAL RESEARCH
Department of the Navy
Arlington, Virginia 22217

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED	1b. RESTRICTIVE MARKINGS NONE
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION / AVAILABILITY OF REPORT Approved for Public Release; Distribution Unlimited
2b. DECLASSIFICATION / DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) Report 022793-3-T TR 250	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION Communications & Signal Processing Laboratory 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION Office of Naval Research
6c. ADDRESS (City, State, and ZIP Code) The University of Michigan Ann Arbor, Michigan 48109-2122	7b. ADDRESS (City, State, and ZIP Code) 800 North Quincy Street Arlington, Virginia 22217
8a. NAME OF FUNDING / SPONSORING ORGANIZATION 8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER Contract No. N00014-85-K-0545
8c. ADDRESS (City, State, and ZIP Code)	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. / PROJECT NO. / TASK NO. / WORK UNIT ACCESSION NO.

11. TITLE (Include Security Classification)
Performance Analysis of Coded Frequency-Hopped Spread-Spectrum Systems with Unknown Interference

12. PERSONAL AUTHOR(S)
M. V. Hegde

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Tech. Report	FROM ____ TO ____	August 1987	114

16. SUPPLEMENTARY NOTATION

17 COSATI CODES		
FIELD	GROUP	SUB-GROUP

18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
Spread-Spectrum, Game Theory, Frequency Hopping, Jamming, Quantization, Orthogonal Signalling

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

Two classes of problems are considered. In the first class we model the process of communicating in the presence of interference, which is unknown or hostile, as a two-person zero sum game with the communicator and the jammer as the players. The objective functions we consider are mutual information and the channel cutoff rate. The communicator's strategies are distributions on the input alphabet and on a set of quantizers and the jammer's strategies are distributions on the noise power subject to certain constraints. We consider various conditions on the jammer's strategy set and on the communicator's knowledge. For the case with the decoder uninformed of the actual quantizer chosen, we show that, from the communicator's perspective the worst-case jamming strategy is a distribution concentrated at a finite number of points thereby converging a functional optimisation problem into a non-linear programming problem. Moreover, we are able to also characterize the worst-case distributions by means of necessary and sufficient conditions which are easy to verify. For the case with the decoder informed of the actual quantizer chosen we are able to demonstrate the existence of saddle-point strategies. The analysis is also seen to be valid for a number

20. DISTRIBUTION / AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION UNCLASSIFIED
22a. NAME OF RESPONSIBLE INDIVIDUAL Wayne E. Stark	22b TELEPHONE (Include Area Code) (313) 763-0390 22c. OFFICE SYMBOL

DD FORM 1473, 84 MAR 83 APR edition may be used until exhausted. All other editions are obsolete.

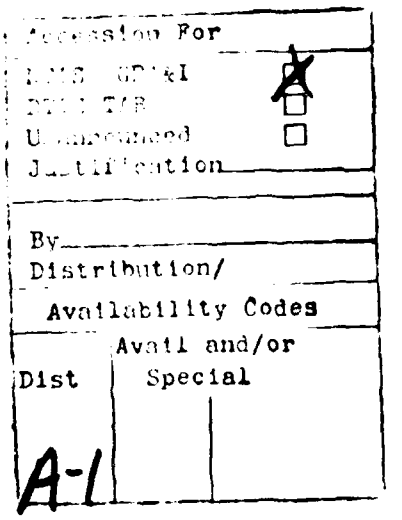
SECURITY CLASSIFICATION OF THIS PAGE
UNCLASSIFIED

19. Abstract (cont.)

of situations where the jammer is adaptive.

The second class of problems we address concerns the performance of orthogonal signalling schemes over channels with unknown partial band interference. The worst case partial band interference is very detrimental to orthogonal signalling and does not allow reliable communication even asymptotically at any signal to noise ratio. We investigate the use of diversity as a simple form of coding in such situations. It is shown that two of the three diversity combining schemes analysed, majority logic combining and linear combining, are unable to overcome the worst case partial band noise, but clipped linear combining is asymptotically able to neutralize the partial band noise, i.e. allow its effect to be no worse than added white Gaussian noise of equivalent noise spectral density

To my father, S.Venkappa Hegde,
to my brothers, Ravi, Harshanna and Sudhir,
and to Daniëlle,
who
C
O
M
P
L
I
C
A
T
E
S
everything.



ACKNOWLEDGEMENTS

I would like to thank my thesis advisor, Dr. Wayne Stark for all the advice, encouragement, support and patience (among other things) he has shown me over the years. He has inspired me by his example and, perhaps unbeknownst to him, is responsible for improving many of my attitudes.

I wish also to express my gratitude to Dr. Demos Teneketzis who has given me unstintingly of his time, knowledge and experience. His generosity is very much treasured.

I am grateful too to Dr. W. Root, Dr. D. Neuhoff and Dr. D. Burns who taught me many of the things I needed to know to write this thesis and furthermore did not abandon me when I requested them to examine what I had done with some of the knowledge they had imparted.

Dr. Paul Shields started me off in this long quest and I am very happy to acknowledge his invaluable influence and assistance.

For their financial support over the years, under contract N00014-85-K0545, I gratefully acknowledge the assistance of the Office of Naval Research.

My friends from the department and my office mates have also contributed in innumerable ways. As they know I will not forget.

Finally I would like to thank Ms. Barb Timm, Ms. Linda Allen and Ms. Beth Olsen for their typing and office support and for their good cheer.

TABLE OF CONTENTS

CHAPTER I

INTRODUCTION

1.1 Unknown Interference and Jamming

Many communication systems have to operate in the presence of interference or noise from other sources. Usually this noise is less familiar to the communicator than the thermal noise originating in the receiver and often very little is known about the particular form of this noise. As a result communication is carried out over a fuzzily-known channel whose parameters depend on the specific kind of noise present. Clearly, if communication is to be carried out reliably, some loss in the rate at which this is performed will have to be suffered to guard against the potentially bad channels that may be encountered. All this is true for a wide variety of situations which arise in practice. For example, multiple access noise (due to transmissions by other users over a common channel) is a case in point. Depending on the message generation rates at the different transmitters the noise may be of unknown intensity, of unknown duration and of unknown timing. Another case which is potentially more detrimental to reliable communication is when there is actually a hostile adversary present, referred to as a jammer, who works at cross-purposes with the communicator. The jammer is both hostile and intelligent in the sense that he makes inimical use of the system parameters that he

knows of. For communication to be possible at all, the jammer must necessarily be constrained. Similarly, the communicator is also subject to constraints in terms of resources such as power or bandwidth. This scenario lends itself quite naturally to a model formulation in terms of two-person non-cooperative game theory with the communicator's strategy set incorporating such parameters as he has control over as, for instance, input signal energy, rate of information symbols, kind of encoding and decoding used, kind of quantization used, and the jammer's strategy set incorporating the noise parameters he introduces into the channel such as, for example, the noise variance or the kinds of distributions of the noise random variables. Various payoff or objective functions have also been studied in the literature such as probability of error, mean square distortion or mutual information.

Information-theoretic analyses which address these situations where the channel parameters at the time of transmission are not completely known have developed well-studied channel models such as the multiple access channel, the interference channel, the compound channel, the arbitrarily varying channel, and various other cases therein. The first two models are suited to the study of interference originating from other users who are competitive but not necessarily hostile in intent. The theme of the analyses in those cases is to achieve or guarantee some kind of cooperation or coordination with limited knowledge of the transmission parameters of the other users through the use of encoders and decoders and to find out the rate region in which reliable communication is possible. The latter three models can be used to model the kind of antagonistic interference originating from a jammer. For instance, the compound channel model may be viewed as one where the jammer chooses one out of a set of possible channels (by choosing one out of a set of possible noise distributions) and the communicator chooses one out of all the possible encoding and decoding strategies with the questions of interest being ones such as: what is the maximum rate at which the communicator can

transmit information reliably irrespective of the particular noise distribution the jammer uses. The solution to this problem [Wolf 78], [Blac 59], which is the capacity of this channel, is $\max_{dP(x)} \min_{C \in \mathcal{C}} I(X;Y)$ (where $dP(x)$ is the distribution of the input symbols, C is a channel selected by the jammer out of the set $\mathcal{C}$, X and Y are the channel input and output random variables respectively and I (. ; .) is the mutual information function). This illustrates the point made earlier of a game theoretic formulation being a natural one for such problems. The solution clearly indicates that the compound channel situation where rate of reliable transmission is the objective and coding and decoding are strategies available to the communicator may be viewed as a two-person game where the communicator's strategy set is the set of all probability distributions on the input and the payoff is the mutual information between the input and the output. The compound channel model is useful in modelling those kinds of jamming where the jamming is not adaptive, i.e. the jammer does not make use of the channel parameters at transmission time to decide his subsequent jamming strategies. If the jammer is sophisticated and adapts his strategy the appropriate model to analyze rate of reliable transmission questions is the arbitrarily "star" varying channel [Csiz 81 pg.233]. Here for each transmission of a channel symbol a new channel (one out of some known set of channels) may be presented to the communicator based on the jammer's knowledge of the previous (and maybe, present) transmitted symbols. Clearly this case includes the compound channel model described before. It also presupposes a jammer who is omniscient in that the channel parameters at the time of every channel transmission are cognizable by him and utilized with malicious intent on the subsequent transmission. Solutions to these problems may be viewed as conservative in the context of less than omniscient real world jammers. The arbitrarily "star" varying channel can also be structured into a two-person zero-sum game theoretic formulation. In Chapter 2 we will utilize the compound channel model for all our

analysis and will explain in the conclusions how the results may, in many cases, be extended to the arbitrarily "star" varying channel.

1.2 Spread Spectrum Counter-measures

The most widely used counter-measure to counteract the effect of a jammer as described above is spread-spectrum modulation. The idea is for the communicator to have potential use of a much larger "space" for transmission than ordinarily needed and use private information (i.e. information not accessible to the jammer) to enable the receiver to determine "where" an individual transmission is located. The use of such a strategy forces the jammer to either expend his finite power over the entire space thereby reducing his effectiveness or to use some other strategy wherein the jammer may allow some transmissions to occur unimpeded but jam the remaining with higher power. In fact, partial-band jamming or pulsed jamming which exploits the latter idea can be shown to be very effective and does seriously degrade the communicator's performance.

Typically in a spread spectrum channel the performance in additive white Gaussian noise is identical to the performance of non-spread systems; namely the bit error probability decreases exponentially with signal-to-noise ratio. However, when subject to worst-case partial-band or pulsed jamming (wherein power is concentrated in time or frequency to affect only a fraction of the symbols transmitted while allowing the remaining to be received "error-free") the bit error probability of a spread-spectrum system decreases only inverse linearly with the signal-to-noise ratio. This is a significant degradation, typically of the order of 30-40 dB for a bit error probability on the order of 10^{-5}.

To remedy this situation most systems use some form of error-correction coding. For example, it can be shown that with a hard decision decoder if the code rate

is small ($< 1/2$) and the jammer is allowed to pulse between several Gaussian distributions then there is no loss in signal-to-noise ratio necessary for reliable communications compared to an additive Gaussian noise channel with the same (average) power. So it can be said that coding (with hard decision demodulation) neutralizes a (power constrained) jammer (i.e., makes the performance the same as additive white Gaussian noise). It can also be shown that the worst case jamming strategy is to pulse between two zero mean Gaussian noise distributions, one of which has zero variance.

As has been well known in the communication field, hard decision decoding loses roughly 2 dB in signal-to-noise ratio compared to soft decision decoding. Thus considerable interest has focused on soft-decision decoding. One problem that has been observed is that if a (soft) decoding algorithm designed for a non-jammed channel is used for a jammed channel then the performance is extremely poor when the jamming strategy is optimized. One method for "overcoming" this difficulty is to assume the jamming noise is one of two distributions (usually one having zero variance called the "off" state and the other called the "on" state) and that the decoder knows perfectly when the jammer is "on" and when the jammer is "off". Using this side information, similar results to the hard decision case have been obtained for the soft decision case (for small rates there is no loss in performance). However assuming this information is available is assuming away the problem. Most systems analyses do not incorporate jamming strategies that affect the reliability of the side information. A jamming strategy not usually considered by such analyses is a strategy that tries to minimize the reliability of the side information.

Motivated by the improvement in performance of soft decisions over hard decisions many researchers have considered decoding algorithms that do not assume side information and do not do hard decision decoding. However, most of these

algorithms still assume the jammer pulses between one of two levels. In Chapter 2 we investigate the case of a decoder that processes symbols from a finite alphabet and where the only constraints on the jammer are average and peak power. We formulate the problem as a game with two players: the jammer whose strategy set consists of distributions modulating the jamming noise, and the communicator, whose strategy set consists of a pair of distributions, one on the input alphabet and one on a set of quantizers. We look for worst-case jamming strategies and investigate when the game admits of a saddle point. We do the analysis using both mutual information (which is equivalent to capacity) and channel cutoff rate as our objective functions.

1.3. Partial-Band Interference

In Chapter 3, we do the detailed analysis for a particular kind of signalling and jamming, i.e. orthogonal signalling in worst-case partial-band jamming. Partial-band jamming is a simple and effective jamming strategy often used against a frequency-hopped spread-spectrum system. It may also be visualized as a worst-case model of partial-band interference, i.e. situations where the frequency-hopped spread spectrum system is subjected to interference in some fraction of the total spread bandwidth the transmitter is using. For example, in a spread-spectrum multiple -access communication system with different users using different hopping patterns such partial-band interference occurs when two users hop to the same frequency-slot at the same time. Another instance is when there is some fading of the transmitted signal in certain frequency bands. The underlying idea in partial-band jamming is this: by concentrating more jamming power on a fraction of the transmitted bits the jammer is willing to allow some transmissions to occur unimpeded but causes high error probabilities for the fraction of bits jammed. As

the uncoded bit error probability varies dramatically with small changes in the bit energy to noise jammer ratio (E_b/N_J) (N_J is the one-sided power spectral density of the jammer's noisewhen spread uniformly across the whole bandwidth) the jammer can, by an appropriate choice of this fraction, cause a significant increase in the average bit error probability. This is the reason for the well known degradation of uncoded systems in worst-case partial-band (or pulsed-time) jamming [Hous 75], [Simon 85]. The worst-case jammer chooses a different fraction for each E_b/N_J and can thereby convert the negative exponential dependence of the bit error probability with E_b/N_J to inverse linear dependence. Thus, for instance, in frequency-hopped M-ary frequency shift keying ($MFSK$), this causes the dramatic degradation of greater than 14 dB at a bit error probability of about 10^{-3} and greater than 30 dB at a bit error probability of about 10^{-5}. For large E_b/N_J typically the worst case fraction jammed is small and thus while the transmissions do not get jammed most of the time those that do are very likely to cause errors. This suggests that some form of coding redundancy that causes data decisions to depend on multiple symbol transmissions can reduce the effectiveness of partial-band jamming. Typically, besides the usual coding gain we have in these situations the addditional gain from the neutralization of the partial-band jammer, i.e. the worst-case fraction with coding is larger. It is known from previous work [Star 82] that, by using codes with rates less than a constant depending on the form of modulation and demodulation used, partial-band jamming can be effectively neutralized.

Orthogonal signals perform well asymptotically in AWGN although they have low rate. Our analysis in Chapter 3 shows that orthogonal signals perform poorly in partial-band jamming even asymptotically . We then investigate the performance of orthogonal signalling with diversity. We do the analysis for diversity using different diversity combining schemes: majority logic combining, linear combining

and clipped linear combining. All the analysis is asymptotic and again we are interested in the performance in the presence of the worst-case jammer. Of the above schemes only clipped linear combining performs well asymptotically.

CHAPTER II

WORST-CASE ANALYSIS OF UNKNOWN INTERFERENCE

2.1 Introduction

We consider a modulator that transmits one out of M signals. This transmitted signal is denoted by the random variable X. The received signal which has been corrupted by the jammer in some fashion is demodulated and quantized into one of L values. In order to disallow the jammer from using knowledge of the quantizer in designing his worst-case strategy, we allow randomization of the quantizer over some given set of quantizers. Clearly such randomization increases the size of the communicator's strategy set. Thus, we view this situation as a game with two players; the jammer and the communicator. The jammer selects the noise in the channel and the communicator chooses the encoder, the decoder and the quantizer. The strategy set for the jammer is the set of all distributions on the power of the jamming noise subject to the given constraints on the peak and average power. The strategy set for the communicator is the set of all distributions on the input alphabet and on the set of quantizers.

For this general set up we show that the worst case jamming strategy from the communicator's perspective is to pulse between a finite number of power levels.

We also consider the case of random decoding strategies where the demodulator output is quantized into a finite number of outputs by a randomized quantizer, i.e. the quantization thresholds are random.

For this case we show that the optimal randomized quantizer can perform better than the nonrandomized quantizer and that from the jammer's point of view the worst-case distribution of the thresholds is also concentrated on a finite number of points. Our basic model can be easily seen to fit a frequency-hop communication system in which the modulation utilizes an M-ary signal set, which in many cases are orthogonal signals. The spread-spectrum bandwith is divided into a large number of frequency slots. Each possible modulated signal is hopped from frequency slot to frequency slot using a pseudo-random hopping pattern. During each hop, one of the M signals is transmitted. There are two important special cases. First, all modulated signals use the same hopping pattern and second, each signal has its own hopping pattern. The demodulator is a coherent or noncoherent matched filter the output of which is then quantized to a finite number of values.

The remainder of the chapter is organized as follows. In Section 2 we define the models we will be considering and give examples for which ou. models apply. In Sections 3 and 4 we derive the above stated results considering the worst case jamming strategy and the optimal quantizer strategy using mutual information as our objective function for the cases with the decoder uninformed about the quantizer and informed about the quantizer respectively. In Section 5 we do the analysis for the case when the quantizer is fixed, i.e. no randomization of the quantizer. We then use the channel cutoff rate as our objective function and derive similar results in Section 6. Finally, in Section 7, we state our conclusions and extend our results.

2.2 Channel Models

In this section we describe the models we use in the subsequent analysis. In all cases we consider a modulator that transmits one out of M signals in D dimensions. This transmitted signal is denoted by the random variable X. The received signal which has been corrupted by the jammer in some fashion is demodulated and quantized into one of L values. The received signal is denoted by the random variable Y. (Y can also be a random vector without changing any of the following analysis).

The general philosophy that we will use is that of game theory with the players being the jammer and the communicator. The jamming strategies are distributions dF on D random variables, $Z_1, Z_2, ..., Z_D$. These random variables represent the power of the jammer in each of the signal dimensions and are modelled as modulating generic noise variables present in the channel. The jammer, however, has an average power constraint and a peak power constraint. More generally the jammer is constrained by

$$\int f(z_1, z_2, ..., z_D) dF(z_1, z_2, ..., z_D) \leq K_J \tag{2.1}$$

and

$$0 \leq Z_j \leq b_j, \quad j = 1, \ldots, D \tag{2.2}$$

where b_j is the peak power constraint and $f(z_1, ..., z_D)$ is some continuous functional of $(z_1, ..., z_D)$. For average power constrained channels with no peak constraint we let b_j become very large.

The output of the demodulator is quantized into one of L values, say $0, 1, ..., L-1$. The output of the quantizer, Y, is also the output of the channel for coding. The strategies for the communicator are to choose a distribution, $dG(\theta)$, on the quantization thresholds and a distribution, $dP(x)$, on the input alphabet. We will

let Θ parametrize the quantizers and assume Θ is some compact subset of R (Θ will be used to denote both the random variable as well as the set of quantizers). We assume Θ and X are independent. For each $(z_1, ..., z_D)$ and θ there is a probability distribution on the output of the channel given the input of the channel:

$$Prob\{Y = y|X = x, \Theta = \theta, Z_1 = z_1, Z_2 = z_2, ..., Z_D = z_D\} = p(y|x, \theta, z_1, z_2, ..., z_D). \tag{2.3}$$

The above model describes the input output relation of the channel for a particular symbol. In addition we model the channel as being memoryless.

In all of our analysis we assume that the jammer and the decoder/quantizer have complete information about the set of strategies possible for each other so that no secret information is considered. As mentioned previously, the performance measure we consider is the largest rate such that reliable communication (in the sense of arbitrarily small error probability) is possible. The type of channels we are considering are known as compound channels with the set of channels (out of which one is chosen) indexed by $dF(z)$. We consider the strategies (distributions) by the jammer to be constant for a whole codeword as opposed to (possibly) changing after each symbol of a codeword which would correspond to an arbitrarily varying channel. For compound channels the capacity with finite input and output is well known to be the maximum of the minimum mutual information. The minimum is over all possible transition probabilities and the maximum is over all probability distributions on the input to the channel. Thus, using the maximum of the minimum mutual information as the performance measure corresponds to the largest rate such that reliable communication is possible no matter what strategy the jammer employs.

We now introduce some notation. Let:

$A = \{0, 1, ..., M-1\}$ be the input alphabet,

$B = \{0, 1, ..., L-1\}$ be the output alphabet,

Θ be the quantizer parameter space (some compact subset of R)

Z be $(Z_1, ..., Z_D)$, $\quad (0 \leq Z_i \leq b_i)$

$p(y|x, \theta, z)$, the transition probability from x to y given θ, z, and

$P_{yx}(\theta, z)$ the corresponding stochastic matrix, $P_{yx}(\theta, z) = [p(y|x, \theta, z)]$.

We assume that

(i) $p(y|x, \theta, z)$ is continuous in z for all θ, x and

(ii) $p(y|x, \theta, z)$ is continuous in θ for all x, z.

Let S denote the set of all probability distributions on the Borel sets of $K \triangleq \{\underline{z} = (z_1, \ldots, z_D) : 0 \leq z_i \leq b_i\}$, and

$$\begin{aligned} I(G, P; F) &= I\left(\int_K \int_\Theta P_{yx}(\theta, z) dG(\theta) dF(z)\right) \\ &= I\left(\int_K P_{yx}^G(z) dF(z)\right) \\ &= I\left(\int_\Theta P_{yx}^F(\theta) dG(\theta)\right) \\ &= I\left(\overline{P}_{yx}(G, F)\right) \end{aligned} \tag{2.4}$$

where $I\left(\overline{P}_{yx}(G, F)\right)$ is the mutual information whenever X and Y are related by the stochastic matrix $\overline{P}_{yx}$.

We now illustrate the applicability of the above model with an example. Consider a frequency-hopped, binary ($M = 2$) phase shift keyed signal set with data signal $d(t)$:

$$d(t) = \sum_{n=-\infty}^{\infty} X_n \, p_T(t - nT)$$

where $X_n \in \{-1, 1\}$ represents the information bit at time interval nT and

$$p_T(t) \begin{matrix} = 1 & 0 \leq t \leq T \\ = 0 & \text{elsewhere} \end{matrix}.$$

Here $p_T(t)$ is the unit pulse on $[0, T]$ and T is the duration of a data bit. The signal

after modulation can be represented by

$$s(t) = \sqrt{2P}\, d(t) \cos 2\pi f_c t .$$

where P is the power of the transmitted signal. In this example, D, the dimensionality of the signal set, is 1 (we are using antipodal signalling). After frequency hopping we have the signal $s'(t)$ where

$$s'(t) = \sqrt{2P}\, d(t) \cos (2\pi(f_c + f_h(t))t)$$

where

$$f_h(t) = \sum_{n=-\infty}^{\infty} f_n\, p_{T_h}(t - nT_h),$$

$p_{T_h}(t)$ is the unit pulse on $[0,T_h]$, $\{f_n\}_{n=-\infty}^{\infty}$ is an i.i.d. uniform sequence over the set $\{f_1, ..., f_q\}$, and T_h is the hop duration. For simplicity let us assume $T_h = T$. The jamming signal added in the channel may be described as

$$j'(t) = \sum_{i=1}^{q} W_i(t)\hat{j}_i(t)\, \sqrt{2} \cos 2\pi f_i t$$

where

$$W_i(t) = \sum_{l=-\infty}^{\infty} W_{i,l}\, p_T\, (t - lT)$$

and $W_{i,l}^2$ is the jamming power in the i^{th} frequency slot during $l\,T \leq t \leq (l+1)T$ and where $\hat{j}_i(t)$ is some unit normalized noise random process in the i^{th} frequency slot. Also, for the abstract model, Z_1 would be the random variable $W_{i,l}$ in the frequency slot chosen for transmission of the signal during the lth time interval. Given any quantizer θ, the channel transition matrix $p(y|x,\theta,z)$ can be calculated by considering the effect of such a quantizer on the random variables at the output of the matched filter. The received frequency-dehopped signal is

$$r(t) = s(t) + \sum_{i=1}^{q} W_i(t)\hat{j}_i\,(t)\delta(f_h(t), f_i).$$

The nature of $\hat{j}_i\,(t)$ determines the type of jamming that is involved. Thus if $\hat{j}_i\,(t)$ were Gaussian noise, we would be dealing with Gaussian jamming. If,

on the other hand, $\hat{j}_i(t)$ is chosen to be $a_i(t)\cos\left(2\pi f_c t + \hat{\phi}_i(t)\right)$, then we have a model for tone jamming for which

$$a_i(t) = \sum_{n=-\infty}^{\infty} V_n \, p_T(t - nT)$$

where $V_n \in \{-1, +1\}$ and

$$\hat{\phi}_j(t) = \sum_{l=-\infty}^{\infty} \phi_l \, p_T\,(t - lT)$$

where ϕ_l's are i.i.d. random variables uniform on [0,2π].

Consider the Gaussian jamming case with X_n i.i.d. and equally likely to be -1 or 1 and with each transmission equally likely to be in any of the q frequency slots. In terms of our model, since $D = 1$ for the $BPSK$ signal set, Z_1 is a random variable with the same distribution as $W_{i,l}$, $f(z_1) = (z_1^2)$, $K_J = 1$ (say) and b_i's can be any arbitrary constants greater than or equal to 1. For tone jamming the only difference in the model is that N is the random variable $cos\phi$ where ϕ is uniformly distributed over $[0, 2\pi]$.

The normalized output of the demodulator is then of the form

$$U = X + NZ$$

with $X \in +1, -1$, N a generic random variable and Z the jamming strategy chosen by the jammer. We can see that the interference during the lth time interval at the output of the matched filter is a random variable of the form $I = NW_{i,l}$ with probability $\frac{1}{q}$ for $i = 1, \ldots, q$, where N is a standard normal random variable. If the output of the demodulator is quantized by a three level quantizer, for example, then

$$Y = q_3(U)$$

where $q_3(u)$ is given by

$$\begin{aligned} q_3(u) &= 1 \quad u > 0 \\ &= ? \quad -\theta \leq u \leq \theta \\ &= -1 \quad u < -\theta, \end{aligned}$$

$0 \leq \theta \leq 1$. Then $p(y|x,\theta,z)$ is given by (for N Gaussian)

$$\begin{aligned} p(y|x,\theta,z) &= 1 - Q(\frac{1-\theta}{z}) \quad y = x \\ &= Q(\frac{1+\theta}{z}) - Q(\frac{1-\theta}{z}) \quad y =? \\ &= Q(\frac{1+\theta}{z}) \quad y \neq x, y \neq ?. \end{aligned}$$

Returning to our abstract model, the strategies for the jammer are all distributions dF on $Z_1, Z_2, ..., Z_D$ satisfying the given constraints. The strategies for the communicator are all distributions dP on X and all distributions dG on Θ. The performance measures we are interested in is the largest rate such that nearly error-free communication can be achieved, i.e. channel capacity, and R_0, the channel cutoff rate. R_0 is a very useful channel parameter especially for the use of convolutional codes. Many researchers believe R_0 to be a practical limit to the set of rates for which reliable communication is possible.

We consider two different structures for the knowledge of information by the communicator.

I. The decoder is unaware of the actual quantizer chosen but only knows the distribution $dG(\theta)$ on the set of quantizers. The jammer knows only the set of quantizers but not the distribution $dG(\theta)$ chosen by the communicator. He is also aware that the decoder does not know the actual quantizer chosen.

II. The decoder knows the actual quantizer chosen. Again the jammer knows only the the set of quantizers. He also knows that the decoder is aware of the actual quantizer chosen.

Case I is seen to apply to situations where, for reasons of implementation perhaps, the decoding is fixed and not altered with the specific quantizer chosen. It may also be viewed as worst-case in the sense that the decoder's knowledge of the specific quantizer and the utilization of such knowledge can only improve the communicator's performance. When there is no randomization of the quantizer, i.e. the quantizer is fixed, Cases I and II are the same and our results for both cases apply to that situation. Also several special jamming strategies are of interest because of correspondence with physical problems. We will classify the cases as follows.

A. Arbitrary joint distribution on $Z_1, Z_2, ..., Z_D$.

B. $Z_1 = Z_2 = ... = Z_D = Z$.

C. One dimensional jamming, i.e., at most one of the random variables $Z_i \neq 0$.

D. Independent jamming, i.e., $Z_1, Z_2, ..., Z_D$ are independent.

Case B corresponds to the physical situation where the jammer is not able to place different amounts of power in different dimensions of the signal space. Case C corresponds to the case where only one of the dimensions can be jammed at once. Case D corresponds to a frequency-hop communication system with independent hopping for the different symbols. The standard game theoretic description is given below.

Communicator's Perspective

The communicator is interested in the maximum rate at which information can be reliably transmitted no matter what strategy the jammer employs. The communicator designs his system assuming the jammer will somehow find out the strategy he is using and then choose the worst possible distribution on the power levels. In Case I the largest rate for which information can reliably be transmitted is

$$\max_{G,P} \min_{F} I(G,P;F)$$

where $I(G,P;F) \triangleq I(X;Y)$ when (dG,dP) is chosen by the communicator and dF is chosen by the jammer and $I(X;Y) = \sum_{x,y} p(x,y)\log(p(y|x)/p(y))$. That this is the maximum rate of reliable transmission is well known since what we are dealing with is a compound channel with a finite input alphabet and a finite output alphabet and a channel set indexed by the distributions $dF(z)$. [Csiz 81, pgs. 172-173].

Jammer's Perspective

The jammer is interested in the minimum rate such that information can not be reliably transmitted at any higher rate no matter what strategy the communicator employs. The jammer designs his system assuming the communicator will somehow find out the strategy he is using and then design the optimal communication system. In Case I the smallest rate that the jammer can guarantee reliable communication can not be above is

$$\min_{dF} \max_{dG,dP} I(G,P;F).$$

That this is the smallest rate the jammer can guarantee is obvious since for each F the rate above which reliable communication is impossible is $\max_{dG,dP} I(G,P;F)$. In Case II the appropriate mutual information can be written as an expectation of the mutual information for a fixed θ:

$$I(G,P;F) = E_G(I(\theta,P;F))$$

where E_G refers to taking expectations w.r.t. dG and $I(\theta,P;F) \triangleq I(X;Y|\theta)$ where $I(X;Y|\theta) = \sum_{x,y} P(x)p(y|x,\theta)\log(p(y|x,\theta,)/p(y|\theta))$ since X and Θ are independent.

We are now ready to state the results. In brief our results show that when the decoder is informed of the quantization rule then (under a compatibility assumption), there is a saddlepoint in cases A and B, i.e. the jammer's rate and the communicator's rate are equal (Theorem 5). However, when the decoder is not informed of the quantization rule then the jammer's rate and the communicator's rate may differ. However the optimal distributions, F from the communicator's point of view and the G from the jammer's point of view are concentrated on a finite number of points (in all the cases A, B, C and D) (Theorem 1). This converts a functional optimization problem into a finite-dimensional non-linear programming problem.

2.3 Case A1: Decoder Uninformed

The communicator has to determine the distributions $(dG(\theta), dP(x))$ that maximize the amount of information, $I(G, P; F)$, transmitted. The jammer has to find the noise distribution $dF(z)$ to minimize the information received by the decoder. Thus, the quantizer's goal is to achieve

$$\max_{dG(\theta), dP(x)} \min_{dF(z)} I(G, P; F)$$

whereas the jammer wants to achieve

$$\min_{dF(z)} \max_{dG(\theta), dP(x)} I(G, P; F).$$

In this section we show that for any choice of strategy of either player there is a simple characterization of the optimal reaction strategy of his opponent.

Theorem 1: a) The jammer can achieve the minimum in $\max_{dG(\theta),\ dP(x)} \min_{dF(z)} I(G, P; F)$ with a distribution concentrated at at most $M(L-1)+2$ points.

b) The communicator can achieve the maximum in $\min_{dF(z)} \max_{dG(\theta),\ dP(x)} I(G, P; F)$ with a distribution concentrated at at most $M(L-1)+1$ points.

Discussion: Theorem 1(a) says that the communicator in trying to achieve $\max_{dG(\theta),\ dP(x)dF(z)} \min I(G,P;F)$ has to consider only reaction strategies of the jammer that have a finite number of points of support, i.e. for each $(dG(\theta), dP(x))$ chosen by the communicator the worst-case jammer distribution may be assumed to be concentrated at a finite number of points and this number is bounded uniformly (in $(dG(\theta), dP(x))$) by $M(L-1)+2$. It follows that for a fixed quantizer (i.e. no randomization of the quantizer) the worst-case jammer is one who chooses such a finite-dimensional distribution. Similarly Theorem 1(b) says that the jammer may, from his perspective of trying to achieve $\min_{dF(z)} \max_{dG(\theta),\ dP(x)} I(G,P;F)$, consider only finite dimensional reaction strategies on the communicator's part.

To prove these results we use the following facts: (1) the convexity and concavity properties of the mutual information function (it is convex in the channel transition matrix and concave in the input distribution), (2) the equivalence of weak convergence with Levy convergence in our situation (see Appendix B) which we use to show the continuity of our objective function in the strategies as well as compactness of our strategy sets (see Appendix C) (this allows us to conclude that there is a worst case jamming strategy and a best case communicator strategy) and (3) Dubins' Theorem in order to demonstrate that the optimal reaction strategies are described by distributions concentrated on a finite number of points. Dubins' Theorem allows the extreme points of certain convex sets to be written as finite linear combinations of extreme points of larger convex sets.

Proof of Theorem 1:

We prove part (a) in detail. The modifications required to obtain part (b) are obvious. We start by first proving two intermediate results, Lemmas 1 and 2.

Lemma 1: $I(G,P;F)$ is a Levy-continuous functional of $dF(z)$ for any fixed $(dG(\theta), dP(x))$.

Proof of Lemma 1:

First we note that for every $(dG(\theta), dP(x))$, $I(P_{yx})$ is a convex function of P_{yx} [Csiz 81, pg. 50] , i.e.,

$$I(\alpha P^1_{yx} + (1-\alpha)P^2_{yx}) \leq \alpha I(P^1_{yx}) + (1-\alpha)I(P^2_{yx}) \quad 0 \leq \alpha \leq 1$$

and

$$p(y|x,z) = \int_\Theta p(y|x,\theta,z)\, dG(\theta)$$

is a continuous function of z (since $p(y|x,\theta,z)$ is continuous in z and $p(y|x,\theta,z) \leq 1$, this follows from the Dominated Convergence Theorem). Also

$$\begin{aligned} p(y|x) &= \int_K \int_\Theta p(y|x,\theta,z)\, dG(\theta)\, dF(z) \\ &= \int_K p(y|x,z)\, dF(z). \end{aligned}$$

Hence $p(y|x)$ is a Levy-continuous functional of $dF(z)$ and therefore P_{yx} is a Levy-continuous functional of $dF(z)$.

Now $I(G,P;F)$ is a convex function of P_{yx} and hence it is continuous in the interior of the finite-dimensional set $\mathbf{W}$ of all stochastic matrices. (Thus, $I(G,P;F)$ is continuous at any point P_{yx} such that at least one row of P_{yx} is not a one point distribution, i.e. P_{yx} is not deterministic). Hence, $I(G,P;F)$ is a Levy-continuous function of $dF(z)$ for any fixed $(dG(\theta), dP(x))$. □

Let $\mathbf{S} \triangleq$ set of all probability distributions on the Borel subsets of K, and

$$\mathbf{S}^1 \triangleq \{dF(z) \in \mathbf{S} : \int f(z)\, dF(z) = K_J\} \tag{2.5}$$

be a hyperplane in $\mathbf{S}$.

Lemma 2: $I(G,P;F)$ achieves its maximum (minimum) in $\mathbf{S}^1$.

Proof of Lemma 2:

We note that $\mathbf{S}$ is compact in the Levy topology (Appendix C).

Also $\mathbf{S}^1$ is a hyperplane in $\mathbf{S}$ which is closed (since $dF(z) \rightarrow \int_K f(z)\, dF(z)$ is Levy-continuous) in the Levy topology.

Hence $\mathbf{S}^1$ being a closed subset of a compact set is itself (Levy)compact.

Thus Lemma 1 asserts that for fixed $(dG(\theta),\ dP(x))$, $I(G,P;F)$ is a Levy-continuous functional on the compact set $\mathbf{S}^1$. Hence it achieves its minimum (maximum) at some point $dF^*(z) \in \mathbf{S}^1$. □

The above lemmas are used to complete the proof of Theorem 1.

From Lemma 2 we know that $I(G,P;F)$ achieves its minimum in $\mathbf{S}^1$. Let $dF^*(z)$ be a distribution which achieves $\min_{dF(z)} I(G,P;F)$. Denote the corresponding P_{yx} as $P^*_{yx} = [p^*(y|x)]$ i.e.

$$P^*_{yx} = \int_K \int_\Theta p(y|x,\theta,z)\ dG(\theta)\ dF^*(z). \tag{2.6}$$

Now consider the set

$$\begin{aligned} \Lambda &= \{dF(z) \in \mathbf{S}^1 : \int_K \int_\Theta p(y|x,z,\theta)\ dG(\theta)\ dF(z) \\ &= p^*(y|x),\ \ x \in A, y \in B^1\} \end{aligned} \tag{2.7}$$

where $B^1 = \{0,1,\ldots,L-2\}$. The set Λ is the intersection of $\mathbf{S}$ with $M(L-1)+1$ hyperplanes viz. $\mathbf{S}^1$ and the $M(L-1)$ hyperplanes

$$h_{yx} = \{dF(z) \in \mathbf{S}^1 : \int_K \int_\Theta p(y|x,z,\theta)\ dG(\theta)\ dF(z) = p^*(y|x)\}. \tag{2.8}$$

Furthermore:

$\mathbf{S}$ is convex.

$\mathbf{S}$ is linearly bounded ($\mathbf{S}$ being compact in a metric space is bounded and hence its intersection with any line is bounded).

$\mathbf{S}$ being a compact subset of a metric space is closed and any line l in the metric space is closed. Thus $\mathbf{S}$ is also linearly closed.

Hence we have that $\mathbf{S}$ is a convex, linearly closed and linearly bounded set. By Dubins' Theorem [Dubi 62] we can conclude that since Λ is the intersection

of $\mathbf{S}$ with $M(L-1)+1$ hyperplanes, every extreme point of Λ is a convex combination of $M(L-1)+2$ or fewer extreme points of $\mathbf{S}$.

From our construction of Λ we know that $I(G,P;F)$ is constant on Λ. Hence for fixed $(dG(\theta),\ dP(x))$, $I(G,P;F)$ assumes its minimum value at an extreme point of Λ also.

Hence, $I(G,P;F)$ assumes its minimum value at some point $dF(z)$ which is a convex combination of $M(L-1)+2$ or fewer extreme points of $\mathbf{S}$.

Since the extreme points of $\mathbf{S}$ are the one-point distributions, we can finally assert that for each $(dG(\theta),dP(x))$ the jammer can achieve the minimum in

$$\max_{dG(\theta),\ dP(x)} \quad \min_{dF(z)} \quad I(G,P;F)$$

with a distribution concentrated at $M(L-1)+2$ points. This concludes the proof of (a).

For channels which are symmetric for each θ and z, i.e. $p(y|x_1,z,\theta)$ is some permutation of $p(y|x_i,z,\theta)$, we see that the set Λ is actually the intersection of $\mathbf{S}$ with $(L-1)+1$ hyperplanes only, and hence part (a) of the theorem holds with $(L-1)+2=L+1$ instead of $M(L-1)+2$. For M-ary symmetric channels, i.e. channels with M inputs and M outputs and such that for each θ and z, $p(y_i|x_i,z,\theta)=1-\epsilon$ and $p(y_i|x_j,z,\theta)=\dfrac{\epsilon}{M-1}$, $i\neq j$, the bound on the number of points of support reduces to 3.

For (b) we note that the jammer wants to achieve

$$\min_{dF(z)} \quad \max_{dG(\theta),dP(x)} \quad I(G,P;F).$$

This may be written as

$$\min_{dF(z)} \quad \max_{dG(\theta)} \quad C(G,F)$$

where $C(G,F) \triangleq \max\limits_{dP(x)} \quad I(G,P;F)$.

We note that similarly to Lemma 1 for any fixed $dF(z)$, $C(G,F)$ is a continuous functional of $dG(\theta)$. (Simply note that $C(G,F)$ being the maximum of

functions convex in P_{yx} is also convex in P_{yx} and proceed as before). Using our hypothesis that $p(y|x,\theta,z)$ is continuous in θ we can show that

$$\min_{dF(z)} \max_{dG(\theta)} C(G,F)$$

can for any $dF(z)$ be achieved by the decoder/quantizer by a distribution $dG(\theta)$ that is concentrated at at most $M(L-1)+1$ points.

Again for symmetric channels we note that part(b) of the theorem holds with L instead of $M(L-1)+1$. For M-ary symmetric channels this number is 2. The number of points of support is one less than Case A as we have not imposed any constraints on the distributions $dG(\theta)$ chosen by the quantizer. □

2.3.1 Necessary and Sufficient Conditions

We now characterize the finite-dimensional distributions of Section 3.1 by means of necessary and sufficient conditions. We first briefly introduce the necessary definitions and results from optimization theory and then specialize them to our cases.

Let Ω be a convex set and let f be a function from Ω into $\mathbf{R}$. For some fixed x_0 if for all x

$$\lim_{\alpha \downarrow 0} \frac{f((1-\alpha)x_0 + \alpha x) - f(x_0))}{\alpha} \tag{2.9}$$

exists f is said to be weakly differentiable at x_0 and the above limit is denoted as $f'_{x_0}(x)$, the weak derivative at x_0. If f is weakly differentiable in Ω at x_0 for all x_0 in Ω, f is said to be weakly differentiable in Ω. We now state an Optimization Theorem that follows from [Luen 69, pg. 178].

Optimization Theorem: Let f be a continuous, weakly differentiable, convex-cap (concave) map from a compact, convex set to $\mathbf{R}$. Let

$$C \triangleq \sup_{x \in \Omega} f(x). \tag{2.10}$$

Then

1. $C = \max f(x) = f(x_0)$ for some $x_0 \in \Omega$.

2. A necessary and sufficient condition for $f(x_0) = C$ is $f'_{x_0}(x) \leq 0$ for all $x \in \Omega$.

Constrained Optimization Theorem: [Luen 69, pg. 217] Let Ω be a convex subset of a linear vector space and f and g convex-cap functionals on Ω to $\mathbf{R}$. Assume there is an $x_1 \in \Omega$ such that $g(x_1) < 0$ and let

$$C' \triangleq \sup_{\substack{x \in \Omega \\ g(x) \leq 0}} f(x). \tag{2.11}$$

If C' is finite then there exists a constant $\lambda \geq 0$ such that

$$C' = \sup_{x \in \Omega} [f(x) - \lambda g(x)]. \tag{2.12}$$

Furthermore if the supremum in the first equation is achieved by $x_0 \in \Omega$ and $g(x_0) \leq 0$, then this supremum is achieved by x_0 in the second equation and $\lambda g(x_0) = 0$. [Luen 69, pg. 217].

Now given any $(dG(\theta), dP(x))$ and the power constraint

$$\int f(z_1, z_2, ..., z_D) dF(z_1, z_2, ..., z_D) \leq K_J$$

we define

$$U_c(K_J, G) \triangleq \sup_{\substack{F \in \mathbf{S} \\ h_F \leq K_J}} -I(G, P; F) \tag{2.13}$$

where $h_F \triangleq \int_K f(z)\, dF(z)$. To simplify notation we define

$$D: \mathbf{S} \rightarrow \mathbf{R} \text{ by } D(F) = \int_K f(z) dF(z) - K_J. \tag{2.14}$$

Using the Constrained Optimization Theorem we will infer in Theorem 2 that there exists a non-negative constant

$$\lambda = \lambda(G, K_J) \text{ for } D(F) \leq 0 \text{ such that}$$

$$U_c(G, K_J) = \sup_{F \in \mathbf{S}} [-I(G, P; F) - \lambda D(F)]. \tag{2.15}$$

We now formulate necessary and sufficient conditions for the characterization of the optimal distributions of Theorem 1 in the following two theorems.

Theorem 2: $U_c(G, K_J)$ is achieved by a distribution $F_0 \in \mathbf{S}$ satisfying $D(F) \leq 0$ and a necessary and sufficient condition for $U_c(G, K_J) = -I(G, P; F_0)$ is that for some constant $\lambda \geq 0$

$$\int_K [-i(z; G, F_0) - \lambda f(z)] dF(z) \leq -I(G, P; F_0) - \lambda K_J \tag{2.16}$$

for all $F \in \mathbf{S}$

where $i(z; G, F_0) \triangleq \sum_{x,y} P(x)\, p(y|x, z) \log \left(\dfrac{\int p(y|x, z)\, dF_0(z)}{\sum_x P(x) \int p(y|x, z)\, dF_0(z)} \right)$.

Proof of Theorem 2:

$D : \mathbf{S} \rightarrow \mathbf{R}$ is clearly linear, bounded, convex-cap, continuous and weakly differentiable in $\mathbf{S}$ with $D'_{F_1}(F_2) = D(F_2) - D(F_1)$. By choosing F_1 as a distribution with unit mass appropriately we can infer that $D(F_1) < 0$. Next we show that $I(G, P; F)$ is convex in F.

$$\begin{aligned}
I(G, P; \alpha F_1 + (1-\alpha)F_2) &= I(\overline{P}_{yx}(G, \alpha F_1 + (1-\alpha)F_2)) \\
&= I\left(\int_K \int_\Theta p(y|x, \theta, z)\, dG(\theta)\, (\alpha dF_1 + (1-\alpha) dF_2) \right) \\
&= I(\alpha\, \overline{P}_{yx}(G; F_1) + (1-\alpha)\, \overline{P}_{yx}(G; F_2)) \\
&= I(\alpha\, \overline{P}^1_{yx} + (1-\alpha)\, \overline{P}^2_{yx}) \\
&\leq \alpha I(\overline{P}^1_{yx}) + (1-\alpha)\, I(\overline{P}^2_{yx}) \\
&\quad \text{(by the convexity of } I(.) \text{ w. r. t. } P_{yx}) \\
&= \alpha I(G, P; F_1) + (1-\alpha)\, I(G, P; F_2).
\end{aligned} \tag{2.17}$$

Then, since $U_c(G, K_J)$ is finite we can infer from the Constrained Optimization Theorem that there exists some constant $\lambda \geq 0$ such that $U_c = \sup_{F \in \mathbf{S}} [-I(G, P; F) - \lambda D(F)]$.

Now we show that $I(G,P;F)$ is weakly differentiable at all $F \in \mathbf{S}$.

Let $L(\alpha) = I(G,P;\alpha F_1 + (1-\alpha)F_2)$. Since $I(G,P;F)$ is convex in F, $L(\alpha)$ is convex in α. Therefore $\frac{L(\alpha)-L(0)}{\alpha}$ is non-decreasing in α and bounded from below and thus $\lim_{\alpha \downarrow 0} \frac{L(\alpha)-L(0)}{\alpha}$ exists. Furthermore

Lemma 3: $I'_{F_1}(G,P;F_2) = \int i(z;G,F_1)\,dF_2(z) - I(G,P;F_1)$.

Proof of Lemma 3:

See Appendix D.

We now have that $-I(G,P;F)-\lambda D(F)$ is convex-cap, continuous and weakly differentiable in F. Thus, by the Optimization Theorem there is a distribution function $F_o \in \mathbf{S}$ such that $U_c(G,K_J) = -I(G,P;F_o) - \lambda D(F_o)$. The necessary and sufficient condition becomes

$$-I'_{F_0}(G,P;F) - \lambda D'_{F_0}(F) \leq 0 \text{ for all } F \in \mathbf{S} \tag{2.18}$$

or

$$\int_K [-i(z;G,F_0) - \lambda f(z)]dF(z) \leq -I(G,P;F_0) - \lambda h_{F_0}. \tag{2.19}$$

If $h_{F_0} < K_J$ the power constraint is trivial and the constant λ is zero, i.e. $D(F_0) < 0$ but $\lambda D(F_0) = 0$. Thus the necessary and sufficient condition is established. □

From Theorem 1 we know that it is possible to find F_0 from the set of distributions with a finite number of points of support. Finding such an F_0 entails determining the set of points of increase as well as the amounts of increase of F_0 at those points. Let E_0 denote the set of points of increase of F_0. We now show

Theorem 3: Let F_0 be a probability distribution satisfying the power constraint. Then F_0 achieves $U_c(G,K_J)$ iff for some $\lambda \geq 0$

C1) $-i(z;G,F_0) \leq -I(G,P;F_0) + \lambda(f(z) - K_J)$

for all $z \in K$.

C2) $-i(z; G, F_0) = -I(G, P; F_0) + \lambda(f(z) - K_J)$

for all $z \in E_0$.

Proof of Theorem 3:

The sufficiency is clear because if both conditions C1 and C2 the conditions of Theorem 2 hold. We show the necessity.

Assume that F_0 is "optimal" but C1 is not true. Then there must exist some $z_1 \in K$ such that $-i(z_1; G, F_0) > -I(G, P; F_0) + \lambda(f(z_1) - K_J)$. Let $F_1(z)$ be a probability distribution with a unit increase at such a point $z_1 \in K$. Then

$$\int_K [-i(z; G, F_0) - \lambda f(z)]\, dF_1(z) > -I(G, P; F_0) - \lambda K_J \tag{2.20}$$

which contradicts Theorem 2. Hence C1 must be true.

Now assume that F_0 is "optimal" but C2 is not true. Then since C1 is true $-i(z; G, F_0) < -I(G, P; F_0) + \lambda(f(z) - K_J)$ for all z in E' where E' is some subset of E_0 with positive measure, i.e.

$$\int_{E'} dF_0(z) = c > 0. \tag{2.21}$$

Since $\int_{E_0 - E'} dF_0(z) = 1 - c$ and on $E_0 - E'$

$$i(z; G, F_0) = I(G, P; F_0) - \lambda(f(z) - K_J) \tag{2.22}$$

and

$$\int_K [i(z; G, F_0) - \lambda f(z)]\, dF_0(z) = \int_{E'} [i(z; G, F_0) - \lambda f(z)]\, dF_0(z)$$
$$+ \int_{E_0 - E'} [i(z; G, F_0) - \lambda f(z)]\, dF_0(z)$$
$$\int_{K - E_0} [i(z; G, F_0) - \lambda f(z)]\, dF_0(z)$$

we have

$$-I(G, P; F_0) - \lambda K_J < -I(G, P; F_0) - \lambda K_J, \tag{2.23}$$

i.e. a contradiction. Hence C2 must be true too. □

Theorems 1 and 3 reduce the calculation of the distributions describing the reaction strategies to finite-dimensional non-linear programming problems. They can be used to simplify the search for conservative strategies which are optimal for either player. In Theorem 4 below we assert the existence of conservative strategies for each player.

Theorem 4: For the game described in Case AI, there exists a conservative strategy $(d\overline{G}(\theta),\ d\overline{P}(x))$ for the communicator and a conservative strategy $d\overline{F}(z)$ for the jammer, i.e. strategies such that

$$\text{i)} \qquad \min_{dF(z)} I(\overline{G},\overline{P};F) = \max_{dP(x)dG(\theta)} \min_{dF(z)} I(G,P;F). \qquad \text{and}$$

$$\text{ii)} \qquad \max_{dP(x),dG(\theta)} I(G,P;\overline{F}) = \min_{dF(z)} \max_{dP(x),dG(\theta)} I(G,P;F) \qquad (2.24)$$

Proof of Theorem 4:

From Lemmas 1 and 2 we note that

a) $I(G,P;F)$ is lower-semicontinuous in $dF(z)$ for each $(dG(\theta),dP(x))$ and

b) There exists $(dG(\theta),dP(x)) \ni I(G,P;F)$ is lower semi-compact in $dF(z)$.

Theorem 4(i) now follows from a fundamental existence theorem [Aubi 82, pg 209, Th. 1]. Theorem 4(ii) follows similarly. □

2.3.2 The Remaining Cases

Case BI: With $F(z)$ now recognized as a one-dimensional distribution Theorems 1 and 2 are easily seen to be true.

Case CI: We redefine **S** as follows: $\mathbf{S} = \bigcup_{i=1}^{M} L_i$ where L_i is the space of product distributions such that

$$Pr(Z_i \geq 0) \geq 0$$

$$Pr(Z_j = 0) = 1,\ j \neq i.$$

By previous arguments each L_i is Levy compact and hence so is $\mathbf{S}$. Now the proofs of Th. 1 and Th. 2 follow as before.

Case DI: We perform the analysis by fixing $D-1$ of the D distributions $dF_1, \ldots, dF_D$. By minor modifications in the proof of Lemma 1 we see that $I(X;Y)$ is a Levy continuous functional of $dF_i(z)$ for each i. Defining $\mathbf{S}$ and $\mathbf{S}^1$ similarly except that now both are spaces of distributions of $dF_i(z_i)$ instead of $dF(z)$ we see that for each $(dG(\theta), dP(x))$ the jammer can achieve the minimum in

$$\max_{(dG(\theta),dP(x))} \min_{dF(z)=dF_1(z_1),dF_2(z_2),\ldots dF_D(z_D)} I(G,P;F) \tag{2.25}$$

with a distribution dF_i concentrated at at most $M(L-1)+2$ points.

Since i is arbitrary we can assert that the jammer can achieve the minimum in (16) with distributions dF_i, $i = 1, \ldots, D$ each of which are concentrated at at most $M(L-1)+2$ points. Part (b) of Theorem 1 and Theorem 2 are easily seen to be true as stated for this case.

2.4 Case AII: Decoder Informed

We have an arbitrary joint distribution on $Z_1, \ldots, Z_D$. The jammer chooses $dF(z)$. The communicator chooses $dG(\theta)$ and further the decoder knows θ. The jammer knows only the the set of quantizers. He also knows that the decoder is aware of the actual quantizer chosen.

In this case we make a "compatibility" assumption, that is, for every θ and $dF(z)$ the capacity-achieving input distribution $dP(x)$ remains the same. While

"compatibility" certainly restricts our model applicability, we show by example that it is often a worst-case assumption. For instance, we know [Dobr 59] that if $M = L$ and if the jammer's strategy set is restricted such that for each distribution $dF(z)$ and quantizer θ, Prob $\{$ error$|x$ $\} \leq \epsilon$ for every x, then the saddle-point strategy for the jammer is to choose a distribution such that

$$p(y|x) = \frac{1}{M} \quad \text{for all } y, x \text{ if } \epsilon > 1 - \tfrac{1}{M}$$

and

$$p(y|x) = \frac{\epsilon}{M-1} \quad y \neq x \text{ if } \epsilon \leq 1 - \tfrac{1}{M}$$
$$= 1 - \epsilon \quad y = x$$

and the saddle-point strategy for the communicator is to choose a uniform distribution on the input alphabet. In our model this corresponds to choosing the canonical noise variables so that $p(y|x, \theta)$ is a symmetric channel for each θ. Such symmetry (and thereby "compatibility") is obtained in a number of other situations as a saddle-point strategy. Under certain conditions, when we have convex constraints in the M noise variables affecting the M inputs of the channel which are invariant under any permutation of the M variables (i.e. a "symmetric" constraint) then the choice of a uniform distribution on the input and the choice of a symmetric channel are saddle-point strategies for the communicator and the jammer respectively (see Appendix E). To describe one more example, if we have M inputs and M outputs,

$$y_i = n_i \quad i = 1, \ldots, M, \quad i \neq j$$
$$y_j = A + n_j, \quad i = j,$$

n_i are $N(0, v_i), i = 1, \ldots, M$ independent random variables and there is further the constraint $\sum_{i=1}^{M} v_i = c$, then by arguments similar to those in Appendix B the saddle point strategy is to choose $v_i = \frac{c}{M}$ and a uniform distribution on the input.

Utilization of the "compatibility" assumption allows us to write the two programs as

$$\min_{dF(z)} \max_{dG(\theta)} E_G(C(\theta, F)).$$

and

$$\max_{dG(\theta)} \min_{dF(z)} E_G(C(\theta, F))$$

where $C(\theta, F) = \max_{dP(x)} I(\theta; F)$ and $I(\theta; F) = I(X; Y|\theta)$.

In this section we prove the existence of a saddlepoint. The main result is stated in the following theorem:

Theorem 5: There exists a pair of distributions $dG^*(\theta), dF^*(z))$ such that

$$E_G(C(\theta, F^*)) \leq E_{G^*}(C(\theta, F^*)) \leq E_{G^*}(C(\theta, F))$$

for all feasible $dG(\theta), dF(z)$, i.e., $(dG^*(\theta), dF^*(z))$ is a saddle point for the game in case AII.

Proof of Theorem 5: The set of all feasible dF 's, i.e.

$$\{dF(z) : \int_K f(z)dF(z) \leq K_J\}, \quad 0 \leq z_i \leq b_i$$

is clearly convex and compact. The set of all dG 's is also convex and compact.

We note that for any fixed $dF(z), C(\theta, F)$ is a continuous function of θ. By our earlier arguments

$$p(y \mid x, \theta) = \int_K p(y|x, \theta, z)dF(z)$$

is a continuous function of θ.

Hence, $P_{yx}(\theta)$ is a continous function of θ. Also $C(\theta, F) = C(P_{yx}(\theta))$ and we know that $C(P_{yx}(\theta))$ is convex in $P_{yx}(\theta)$.

Therefore, for every $\theta \in \Theta \ni P_{yx}(\theta)$ is not deterministic, $C(P_{yx}(\theta))$ is a continous function of $P_{yx}(\theta)$. Hence, for fixed $dF(z), C(\theta, F) = C(P_{yx}(\theta))$ is a continuous function of θ and so

$$E_G(C(\theta, F)) = \int_\Theta C(\theta, F) dG(\theta) \tag{2.26}$$

is a Levy continuous functional of $dG(\theta)$.

Since $E_G(C(\theta, F))$ is linear it is also a concave function of $dG(\theta)$ in $\mathrm{dG}(\theta))$. Next we note that $C(\theta, F)$ is convex in $dF(z)$ for each θ since $C(\theta, F) = C(P_{yx}(\theta))$. Hence

$$C(\theta, \alpha F^1 + (1-\alpha)F^2) < \alpha C(\theta, F^1) + (1-\alpha)C(\theta, F^2) \quad 0 \leq \alpha \leq 1.$$

Taking expectations w.r.t. G

$$\int_\Theta C(\theta, \alpha F^1 + (1-\alpha)F^2) dG(\theta)$$
$$\leq \int_\Theta (\alpha C(\theta, F^1) + (1-\alpha)C(\theta, F^2)) dG(\theta).$$

Thus

$$E_G(C(\theta, \alpha F^1 + (1-\alpha)F^2) \leq \alpha E_G(C(\theta, F^1)) + (1-\alpha)E_G(C(\theta, F^2)).$$

Consequently, $E_G(C(\theta, F))$ is a convex function in $dF(z)$.

Also $E_G(C(\theta, F))$ is Levy-continuous in $dF(z)$. To prove this it suffices to show that for any sequence F_n converging to F in the Levy metric

$$E_G(C(\theta, F_n)) \to E_G(C(\theta, F)).$$

Since convergence in the Levy metric is in our case equivalent to weak convergence (see Appendix B) it suffices to show this for $F_n \xrightarrow{w} F$. However,

$$\begin{aligned}
&\lim_n E_G(C(\theta, F_n)) \\
&\quad = \lim_n \int_\Theta C(\theta, F_n) dG \\
&\quad = \int_\Theta \lim_n C(\theta, F_n) dG \quad \text{(by the Dominated Convergence Theorem)} \\
&\quad = \int_\Theta C(\theta, F) dG \quad (\text{since } C(\theta, F) \text{ is Levy} - \text{continuous in } F) \\
&\quad = E_G(C(\theta, F))
\end{aligned}$$

which proves Levy-continuity in $dF(z)$. From these properties of the objective function and the convexity and compactness of the feasible strategy sets we recognize that the hypotheses of the Sion minmax theorem of game theory are satisfied [Aubi 82, Th 7, pg 218]. This concludes the proof of Theorem 5. □

We note that these saddle-point distributions need not have finite support. However, in this case we have an equlibrium and with no further knowledge of each other's choice of strategy, the jammer and the quantizer should be content utilizing $dG^*(\theta)$ and $dF^*(z)$.

Using the Optimization Theorem and the Constrained Optimization Theorem we can derive necessary and sufficient conditions at these saddle points. Given any $dG(\theta)$ and the power constraint we define

$$\overline{U}_c(K_J, G) \triangleq \sup_{\substack{F \in \mathbf{S} \\ h_F \leq K_J}} - E_G(C(\theta, F)) \tag{2.27}$$

and given any $dF(z)$ we define

$$\overline{V}_c(F) \triangleq \sup_{G \in \mathcal{G}} E_G(C(\theta, F)) \tag{2.28}$$

where $\mathcal{G}$ is the space of distributions on Θ. Then we have

Theorem 6: The saddle-point strategies dF^*, dG^* satisfy to the following inequalities:

$$E_{G^*}(\int(-\bar{i}(z;\theta, F^*) - \lambda f(z))dF(z)) \leq E_{G^*}(-C(\theta, F^*)) - \lambda K_J \tag{2.29}$$

for some $\lambda \geq 0$, for all F where

$$\bar{i}(z;\theta, F) \triangleq \sum_{x,y} P(x)p(y|x, z, \theta) \log \frac{\int p(y|x, z, \theta)dF(z)}{\sum_x P(x) \int p(y|x, z, \theta)dF(z)}$$

Also

$$E_G(C(\theta, F^*)) \leq E_{G^*}(C(\theta, F^*)) \tag{2.30}$$

for all G.

Proof of Theorem 6:

For any F denote the weak derivative of $E_G(C(\theta, F))$ at G_0 as $D_{G_0}(E_G(C(\theta, F))$ and for any G denote the weak derivative of $E_G(C(\theta, F))$ at F_0 as $D_{F_0}(E_G(C(\theta, F))$. Using Lemma 3 and the Dominated Convergence Theorem, we have

$$D_{F_1}(E_G(-C(\theta, F_2)) = E_G(-\int \tilde{i}(z; \theta, F_1) dF_2) + E_G(C(\theta, F_1)) \tag{2.31}$$

for any F_1, F_2.

Also

$$D_{G_1}(E_{G_2}(C(\theta, F))) = E_{G_2}(C(\theta, F)) - E_{G_1}(C(\theta, F)). \tag{2.32}$$

Now letting $F_1 = F^*, G_1 = G^*$ in the first equation we have, using the Constrained Optimization Theorem and the Optimization Theorem and the properties of $E_G(C(\theta, F))$ as in Theorem 2, that a necessary and sufficient condition for F^* to achieve $\overline{U}_c(K_J, G^*)$ is

$$E_{G^*}(-\int (\tilde{i}(z; \theta, F^*) - \lambda f(z)) dF(z)) \leq E_{G^*}(-C(\theta, F^*)) - \lambda K_J \tag{2.33}$$

for some $\lambda \geq 0$, for all F.

Letting $F_1 = F^*, G_1 = G^*$ in the second equation gives us similarly that a necessary and sufficient condition to achieve $\overline{V}_c(F^*)$ is

$$E_G(C(\theta, F^*)) \leq E_{G^*}(C(\theta, F^*)) \tag{2.34}$$

for all G.

Since at a saddle-point $\overline{U}_c(K_J, G^*)$ and $\overline{V}_c(F^*)$ are simultaneously achieved, the theorem follows. □

2.4.1 The Remaining Cases

Case BII: Theorem 5 holds with $F(z)$ as a one-dimensional distribution.

Case CII: Although **S** is compact, it is not convex and so we cannot demonstrate that there is a saddle point strategy.

Case DII: Again we have that $E_G(C(\theta, F))$ is a Levy continuous functional of $dG(\theta)$ and is concave in $dG(\theta)$. Also $E_G(C(\theta, F))$ is Levy continuous in $(dF_1(z), \ldots, dF_D(z))$. However $E_G(C(\theta, F_1, \ldots F_D))$ is not convex in $(F_1, \ldots, F_D)$. Hence we cannot assert the existence of a saddle point in this case.

2.5 Fixed Quantizer

Before concluding this chapter we point out that if we did not have randomized quantization then without "compatibility" the game would have a saddle- point where the jammer's saddle-point distribution need be concentrated at at most $M(L-1)+2$ points. We summarize this in Theorem 7.

Theorem 7: For any quantizer θ, there exists a pair of distributions $dP^*(x), dF^*(z)$ such that

$$I(\theta, P, F^*) \leq I(\theta, P^*, F^*) \leq I(\theta, P^*, F) \tag{2.35}$$

for all feasible dP, dF. Moreover $dF^*(z)$ can be chosen to be concentrated at at most $M(L-1)+2$ points and necessary and sufficient conditions for $dF^*(z)$ and $dP^*(x)$ are for some $\lambda_1, \lambda_2 \geq 0$

$$-i(z; \theta, F^*) \leq -I(\theta, P^*, F^*) + \lambda_1(f(z) - K_J) \tag{2.36}$$

for all $z \in K$ and

$$-i(z; \theta, F^*) = -I(\theta, P^*, F^*) + \lambda_1(f(z) - K_J) \tag{2.37}$$

for all $z \in E_0$ where $i(\cdot; \ldots)$ is as defined in Theorem 2 with G concentrated on θ. Also

$$I_x(\theta, P^*, F^*) = \lambda_2 \tag{2.38}$$

for all $x \ni P^*(x) > 0$ and

$$I_x(\theta, P^*, F^*) \leq \lambda_2 \tag{2.39}$$

for all $x \ni P^*(x) = 0$ where

$$I_x(\theta, P^*, F^*) \triangleq \sum_y p(y|x,\theta) \log \frac{p(y|x,\theta)}{\sum_x P^*(x) p(y|x,\theta)}.$$

Proof of Theorem 7:

From the proof of Theorem 5 we know that all we need to show is that $I(\theta, P, F)$ is (Levy) continuous in $dP(x)$. We show this by considering any sequence $dP_n(x) \xrightarrow{w} dP(x)$ and showing $I(\theta, P_n, F) \to I(\theta, P, F)$. Since x belongs to the finite set A, weak convergence is equivalent to convergence in any finite-dimensional metric.

Now

$$\begin{aligned}
|I(\theta, P_n, F) - I(\theta, P, F)| &= |\sum_{x,y} P_n(x) p(y|x,\theta) \log \frac{p(y|x,\theta)}{\sum_x P_n(x) p(y|x,\theta)} \\
&\quad - \sum_{x,y} P(x) p(y|x,\theta) \log \frac{p(y|x,\theta)}{\sum_x P(x) p(y|x,\theta)}| \\
&\leq |\sum_{x,y} P_n(x) p(y|x,\theta) \log \frac{p(y|x,\theta)}{\sum_x P_n(x) p(y|x,\theta)} \\
&\quad - \sum_{x,y} P_n(x) p(y|x,\theta) \log \frac{p(y|x,\theta)}{\sum_x P(x) p(y|x,\theta)}| \\
&\quad + |\sum_{x,y} P_n(x) p(y|x,\theta) \log \frac{p(y|x,\theta)}{\sum_x P(x) p(y|x,\theta)} \\
&\quad - \sum_{x,y} P(x) p(y|x,\theta) \log \frac{p(y|x,\theta)}{\sum_x P(x) p(y|x,\theta)}| \\
&\leq |\sum_{x,y} P_n(x) p(y|x,\theta)| |\log \frac{\sum_x P_n(x) p(y|x,\theta)}{\sum_x P(x) p(y|x,\theta)}| \\
&\quad + \sum_x D|P_n(x) - P(x)|
\end{aligned} \tag{2.40}$$

where $D = \max_{x,y} p(y|x,\theta) \log \frac{p(y|x,\theta)}{\sum_x p(y|x,\theta)}$

$$\begin{aligned}
&\leq LD|\log \frac{\sum_x P_n(x) p(y|x,\theta)}{\sum_x P(x) p(y|x,\theta)}| \\
&\quad + \sum_x D|P_n(x) - P(x)|.
\end{aligned} \tag{2.41}$$

Again since A is finite we can say that for all $\delta > 0$ $\exists N$ such that for all $n > N$

$$1 - \delta \leq \frac{P_n(x)}{P(x)} \leq 1 + \delta \quad \forall x \in A$$

$$1 - \delta \leq \frac{P_n(x)p(y|x,\theta)}{P(x)p(y|x,\theta)} \leq 1 + \delta \quad \forall x \in A$$

$$1 - \delta \leq \frac{\sum_x P_n(x)p(y|x,\theta)}{\sum_x P(x)p(y|x,\theta)} \leq 1 + \delta \quad \forall x \in A. \tag{2.42}$$

By the continuity of the log function we can say that $\forall \epsilon > 0$ $\exists \delta > 0 \ni$

$$-\epsilon \leq |\log \frac{\sum_x P_n(x)p(y|x,\theta)}{\sum_x P(x)p(y|x,\theta)}| \leq \epsilon.$$

The second term in (2.41) can also clearly be made $\leq \epsilon$ for sufficiently large n. Thus the continuity of $I(\theta, P, F)$ w.r.t. P is confirmed and the first part of the theorem follows. The bound on the number of points of support of dF^* follows from Theorem 1(a). The necessary and sufficient conditions are derived as before from Theorem 3 and well-known results about channel capacity [Gall 68, pg.91]. □

2.6 Channel Cutoff Rate

In this section we show how the results obtained for channel capacity also carry over when the performance measure we choose is R_0, the channel cutoff rate. For a channel given by a transition probability matrix $p(y|x)$, R_0 is defined as

$$R_0 \triangleq \max(-\log E(J(X_1, X_2)))$$

where the maximization is over all distributions $dP(x)$ on the input, X_1 and X_2 are independent random variables with distribution $dP(x)$ and

$$J(X_1, X_2) = \sum_{y \in B} \sqrt{p(y|x_1)p(y|x_2)}.$$

The cutoff rate is the largest number for which there is a linear error exponent viz. a bound of the form $P_e \leq 2^{-n(R_0-R)}$, relating the error probability of the best code of length n and rate R, which is true for all $R < R_0$. It is the rate beyond which sequential decoding of convolutional codes becomes intractable and is widely interpreted to be the largest rate at which "practical" coding systems can be implemented [Vite 79],[Mass 80].

For the compound channel the appropriate R_0 to use is

$$\max_{dP(x)} \min_{p(y|x)\in\mathcal{W}} (-\log E(J(X_1, X_2)))$$

where $\mathcal{W}$ is the channel set. This is the largest number for which there is a linear error exponent for the compound channel. This follows from the random coding exponent function and the sphere packing exponent function for the compound channel [Csiz 81, Lemma 5.4 and Theorem 5.10]. Thus in our game-theoretic formulation the communicator wants to choose $(dP(x), dG(\theta))$ to achieve

$$\max_{dP(x),dG(\theta)} \min_{dF(z)} R_0(G, P, F)$$

where

$$R_0(G, P, F) \triangleq$$

$$-\log(E(\sum_{y\in B} \sqrt{\int\int p(y|x_1, z, \theta) dG(\theta) dF(z)} \sqrt{\int\int p(y|x_2, z, \theta) dG(\theta) dF(z)}))$$

and the jammer chooses $dF(z)$ to attain

$$\min_{dF(z)} \max_{dP(x)} R_0(G, P, F).$$

In case AI we are able to derive results similar to the previous case. Theorem 8 below recovers the same result for $R_0(G, P, F)$ as Theorem 1 did for mutual information.

Theorem 8:a) The jammer can achieve the minimum in $\max_{(dG(\theta),dP(x))} \min_{dF(z)} R_0(G, P, F)$ with a distribution concentrated at at most $M(L-1)+2$ points.

b) The communicator can achieve the maximum in $\min_{dF(z)} \max_{(dG(\theta),dP(x))} R_0(G,P,F)$ with a distribution concentrated at at most $M(L-1)+1$ points.

Proof of Theorem 8:

As in the proof of Lemma 1 we note that P_{yx} is a Levy-continuous function of $dF(z)$. From the functional form of $R_0(G,P,F)$ it is clear that it is a continuous function of P_{yx} (it is easy to see that for all possible P_{yx}, the argument of the log can never be 0). Hence we have that $R_0(G,P,F)$ is a Levy-continuous functional of $dF(z)$ for any $(dG(\theta),dP(x))$. The rest of the proof follows exactly the corresponding steps in the proof of Theorem 1 with $R_0(G,P,F)$ replacing $I(G,P;F)$. For part (b) we point out that $R_0(G,P,F)$ is a Levy-continuous functional of $(dG(\theta),dP(x))$ and proceed as before. □

Furthermore, we can also derive results similar to Theorems 2, 3 and 4 with $R_0(G,P,F)$ as the objective function. We do this in Theorems 9,10 and 11 whose proofs we sketch briefly.

Given any $(dG(\theta),dP(x))$ and the power constraint we define

$$U_R(K_J,G) \triangleq \sup_{\substack{F \in \mathbf{S} \\ h_F \leq K_J}} -R_0(G,P,F) \tag{2.43}$$

and as before

$$D(F) = \int_K f(z)dF(z) - K_J.$$

Theorem 9: $U_R(K_J,G)$ is achieved by a distribution $F_0 \in \mathbf{S}$ satisfying $D(F_0) \leq 0$ and a necessary and sufficient condition for $U_R(K_J,G) = -R_0(G,P,F_0)$ is that for some constant $\lambda \geq 0$

$$\int_K [q(z;F_0,G) - \lambda f(z)]dF(z) \leq T_0(G,P,F_0) - \lambda K_J \tag{2.44}$$

where

$$q(z;F_0,G) \triangleq E\left(\sum_y \left(\frac{\sqrt{\int p(y|x_1,z)dF_1}\,p(y|x_2,z)}{2\sqrt{\int p(y|x_2,z)dF_1}}\right.\right.$$

$$+ \frac{\sqrt{\int p(y|x_2,z)dF_1}p(y|x_1,z)}{2\sqrt{\int p(y|x_1,z)dF_1}}\Bigg)\Bigg).$$

Also

$$T_0(G,P,F) \triangleq \exp(-R_0(G,P,F)).$$

and the expectation is w.r.t. independent random variables X_1, X_2 with common distribution $dP(x)$.

Proof of Theorem 9:

Given (dG, dP) we have from the definition of R_0 that

$$\min R_0(G,P,F) = -\max(-R_0(G,P,F)) = -\max\log(E(J(X_1,X_2)))$$

Maximizing $-R_0(G,P,F)$ is equivalent to maximizing $\exp(-R_0(G,P,F))$ which is the same as maximizing $E(J(X_1,X_2) = E(\sum_y \sqrt{p(y|x_1)p(y|x_2)}$. For notational purposes let us denote $\exp(-R_0(G,P,F))$ as $T_0(G,P,F) = E(\sum_y \sqrt{p(y|x_1)p(y|x_2)})$. We show that $T_0(G,P,F)$ is a convex-cap (concave) functional of F. Let

$$\begin{aligned}
p^1(y|x_1) &\triangleq \int p(y|x_1,z)dF_1(z) \\
p^2(y|x_1) &\triangleq \int p(y|x_1,z)dF_2(z) \\
p^1(y|x_2) &\triangleq \int p(y|x_2,z)dF_1(z) \\
p^2(y|x_2) &\triangleq \int p(y|x_2,z)dF_2(z).
\end{aligned}$$

Then from the inequality $a + b \geq 2\sqrt{ab}$ we have

$$p^1(y|x_1)p^2(y|x_2) + p^1(y|x_2)p^2(y|x_1) \geq 2\sqrt{p^1(y|x_1)p^1(y|x_2)p^2(y|x_1)p^2(y|x_2)}$$

where $p^1(y|x_1)p^2(y|x_2)$ may be chosen as a and $p^1(y|x_2)p^2(y|x_1)$ may be chosen as b.

Therefore

$$\alpha^2 p^1(y|x_1)p^1(y|x_2) + (1-\alpha)^2 p^2(y|x_1)p^2(y|x_2)$$

$$+ \alpha(1-\alpha)p^1(y|x_1)p^2(y|x_2) + \alpha(1-\alpha)p^2(y|x_1)p^1(y|x_2)$$

$$\geq \alpha^2 p^1(y|x_1)p^1(y|x_2) + (1-\alpha)^2 p^2(y|x_1)p^2(y|x_2)$$

$$+ 2\alpha(1-\alpha)\sqrt{p^1(y|x_1)p^1(y|x_2)p^2(y|x_1)p^2(y|x_2)}$$

$0 \leq \alpha \leq 1$. Taking square-roots we have

$$\sqrt{(\alpha p^1(y|x_1) + (1-\alpha)p^2(y|x_1))(\alpha p^1(y|x_2) + (1-\alpha)p^2(y|x_2)}$$

$$\geq \alpha\sqrt{p^1(y|x_1)p^1(y|x_2)} + (1-\alpha)\sqrt{p^2(y|x_1)p^2(y|x_2)}.$$

Summing over y and taking expectations we have

$$E(\sum_y \sqrt{(\alpha p^1(y|x_1) + (1-\alpha)p^2(y|x_1))(\alpha p^1(y|x_2) + (1-\alpha)p^2(y|x_2)}$$

$$\geq \alpha E(\sum_y \sqrt{p^1(y|x_1)p^1(y|x_2)}) + (1-\alpha)E(\sqrt{p^2(y|x_1)p^2(y|x_2)})$$

thus establishing the desired concavity. Furthermore, denoting the weak derivative of $T_0(G,P,F)$ at F_1 by $D_{F_1}(T_0(G,P,F))$, we need

Lemma 4: $D_{F_1}(T_0(G,P,F)) = \int q(z;F_0,G)dF_2(z) - T_0(G,P,F_1)$.

Proof of Lemma 4:

See Appendix F .

Using the weak derivative and the just derived property of concavity we can proceed as in Theorem 2 to get Theorem 9. □

Theorem 10: Let F_0 be a probability distribution satisfying the power constraint. Then F_0 achieves $U_R(G,K_J)$ iff for some $\lambda \geq 0$

$$q(z;F_0,G) \leq T_0(G,P,F_0) + \lambda(f(z) - K_J) \tag{2.45}$$

for all $z \in K$

$$q(z;F_0,G) = T_0(G,P,F_0) + \lambda(f(z) - K_J) \tag{2.46}$$

for all $z \in E_0$ where E_0 denotes as before the set of points of increase of F_0 (E_0 is finite from Theorem 7).

Proof of Theorem 10:

Follows directly from the proof of Theorem 3. □

Theorem 11: There exists a conservative strategy $(d\overline{G}(\theta), d\overline{P}(x))$ for the communicator and a conservative strategy $d\overline{F}(z)$ for the jammer i.e. strategies such that

$$\text{i)} \max_{dP(x),dG(\theta)} R_0(G,P,\overline{F}) = \min_{dF(z)} \max_{dP(x),dG(\theta)} R_0(G,P,F)$$

and

$$\text{ii)} \min_{dF(z)} R_0(\overline{G},\overline{P},F) = \max_{dP(x)dG(\theta)} \min_{dF(z)} R_0(G,P,F).$$

Proof of Theorem 11:

Again, parallels almost exactly the proof of Theorem 4. □

We also note here that results similar to that derived in the case with mutual information as our objective function hold in cases BI, CI and DI with $R_0(G,P,F)$ as the objective function. However we cannot achieve a saddle-point for the case with side information (with randomized quantizers and "compatibility") because $R_0(G,P,F)$ is not necessarily convex in F and such convexity is essential for any saddle-point to exist. If, on the other hand, we give up randomization of the quantizer (and do not assume "compatibility") then we once again have a saddle-point with the jammer's saddle-point distribution having finite support. This is stated in Theorem 12.

Theorem 12: There exists a pair of distributions $(dP^*(x), dF^*(z))$ such that

$$R_0(G,P,F^*) \le R_0(G^*,P^*,F^*) \le R_0(G^*,P^*,F) \tag{2.47}$$

for all feasible (dP, dG). Moreover, $dF^*(z)$ can be chosen to be concentrated at at most $M(L-1)+2$ points and necessary and sufficient conditions for $dF^*(z)$ and $dP^*(x)$ are that for some $\lambda_1, \lambda_2 \ge 0$

$$q(z;F^*,\theta) \le T_0(\theta,P^*,F^*) + \lambda_1(f(z) - K_J) \tag{2.48}$$

for all $z \in K$. Also

$$q(z;F^*,\theta) = T_0(\theta,P^*,F^*) + \lambda_1(f(z) - K_J) \tag{2.49}$$

for all $z \in E_0$, where E_0 denotes the points of increase of F^* and

$$T_0(\theta, P, F) = \exp(-R_0(\theta, P, F))$$

and

$$\sum_y \sqrt{p(y|x,\theta)}\beta(y, P^*) = \lambda_2 \tag{2.50}$$

for all $x \ni P^*(x) = 0$ where $\beta(y, P) \triangleq \sum_x P(x)p(y|x,\theta)$ and

$$\sum_y \sqrt{p(y|x,\theta)}\beta(y, P^*) \geq \lambda_2 \tag{2.51}$$

for all $x \ni P^*(x) > 0$.

Proof of Theorem 12:

We work with $T_0(\theta, P, F)$ instead of $R_0(\theta, P, F)$. We can do this because

$$R_0(\theta, P, F^*) \leq R_0(\theta, P^*, F^*) \leq R_0(\theta, P^*, F)$$

$$\Leftrightarrow T_0(\theta, P^*, F) \leq T_0(\theta, P^*, F^*) \leq T_0(\theta, P, F^*)$$

and so a saddle-point for R_0 is a saddle-point for T_0 and *viceversa*. Obviously $T_0(\theta, P, F)$ is continuous in P and continuous in F. Moreover from the proof of Theorem 9 we know $T_0(\theta, P, F)$ is convex-cap (concave) in F.

Furthermore $T_0(\theta, P, F)$ is convex in P [Vite 79, pg.140]. Using the Sion minimax theorem as in Theorem 5, the first part of this theorem follows. The bound on the number of points of support follows from Theorem 8. The necessary and sufficient conditions follow from Theorem 10 and well-known properties of the optimizing distributions for the channel cutoff rate [Vite 79, pg.205]. □

2.7 Conclusions

We have constructed fairly general channel models which are capable of representing a number of jamming situations. The jammers we have considered have

all been non-adaptive and using results from the compound channel we are able to give operational significance to our minimax performance measures, i.e. we can assert the existence of encoders and decoders which can perform at arbitrarily low probabilities of error at rates close to our performance measures. Our analysis is also clearly applicable to many restrictions on the jammer's strategy set other than the ones we have considered.

In the case with the decoder uninformed (case I) we have shown that the worst-case jammer strategy (from the communicator's perspective) as well as the worst-case communicator strategy (from the jammer's perspective) needs only be one of the class of distributions with support on a finite number of points. We have a bound on the number of these points of support in terms of the sizes of the input and the output alphabet. Thus we have reduced the computation of the worst case jamming strategies to a finite-dimensional non-linear programming problem. Moreover we can characterize these distributions by necessary and sufficient conditions which are fairly easy to test. All the above has been done for both objective functions: mutual information, which tells us about the fundamental limits to communication in these situations as well as the channel cutoff rate, which tells us about the 'practical' limits to such communication.

In the cases with decoder informed we reduce the communicator's strategy set (either by using the "compatibility" assumption or by fixing a quantizer) . In this case when we have convexity with respect to the jammer's strategy (as in cases AII and BII) we are able to demonstrate the existence of a saddle-point strategy. For the case with non-randomized quantization we are further able to characterize these saddle-point strategies using the earlier theory.

As we have mentioned earlier all the above presupposes non-adaptive jamming. The compound channel model which we use indirectly by our choice of objective function is appropriate to use in this case. We can allow for more sophisticated

jammers if we incorporate the cases where the jammer's strategies are allowed to depend on the previous (and present) channel inputs. The appropriate channel model to use then is that of the arbitrarily "star" varying channel (A^*VC) [Csiz 81, pg.233]. This model generalizes the arbitrarily varying channel (AVC) and includes it as a special case. It is known that the m-capacity (i.e. capacity with maximum probability of error over all the codewords) of the A^*VC is the same as that of the corresponding AVC [Csiz 81, pg.232]. This capacity is known for the case of binary output alphabet (and finite input alphabet) and is known to be equal to $\max_{dP(x)} \min_{W \in \overline{\overline{\mathcal{W}}}} I(X;Y)$ where X and Y are the input and the output respectively, W is any channel chosen from the set of channels $\mathcal{W}$ and $\overline{\overline{\mathcal{W}}}$ is the row-convex closure of $\mathcal{W}$ [Csiz 81]. In our case the jammer's strategy set corresponding is already row-convex closed and hence the appropriate programs would be

a) For the communicator:

$$\max_{(dG(\theta),dP(x))} \min_{dF(z)} I(G,F)$$

b) For the jammer

$$\min_{dF(z)} \max_{(dG(\theta),dP(x))} I(G,F)$$

which is the same objective function as we have used. Similarly, in the case with decoder informed we would obtain the same objective functions. Thus, all the results derived in the previous chapter for the case of mutual information can be extended to the case of the A^*VC channel with binary output. This model may be viewed as a worst-case representation of adaptive jamming. Unfortunately the m-capacity of the AVC is as yet unknown for output sizes greater than 2. On the other hand the a-capacity of the AVC (i.e. the capacity with average probability of error) is known to be either 0 or else $\max_{dP(x)} \min_{W \in \overline{\mathcal{W}}} I(X;Y)$ where $\overline{\mathcal{W}}$ is the convex closure of the set $\mathcal{W}$ to which W belongs [Csiz 81, pg.214]. Since in our model the set of channels is convex as well as row-convex the a-capacity is known to be greater than 0 iff the m-capacity is greater than 0 [Ahls 78]. Thus with average

probability of error whenever the jammer's strategy set is such that he cannot force the capacity to be 0 then all the results of the preceding chapter extend to the case of the A^*VC channel.

CHAPTER III

PERFORMANCE OF ORTHOGONAL SIGNALLING IN UNKNOWN PARTIAL-BAND INTERFERENCE

3.1 Introduction

In this chapter we investigate the performance of simple signalling and demodulation schemes over the partial-band jammed channel. When communicating over the added white Gaussian noise channel, orthogonal signalling suffices asymptotically to achieve capacity, i.e. by choosing M large enough the error probability can be made arbitrarily small for all rates less than capacity or equivalently provided that the ratio of the energy transmitted per information bit and the one-sided power spectral-density E_b/N_0 is greater than ln 2. Conversely no other signals can achieve arbitrarily small error probability when $E_b/N_J < \ln 2$. It is also known [Stark 85a],[Stark 85b], that provided codes of small enough rate are used the capacity of a partial-band jammed channel is the same as that of a white Gaussian noise channel. In the light of this it seems plausible to expect that for the partial-band jammed (PBJ) channel, orthogonal signals with correlation detection which suffice in the white Gaussian case, could be used as a simple scheme to form a reliable communication system. Unfortunately this turns out not to be true and in

Section 2 we demonstrate this by considering the limiting value of the error probability of orthogonal signals in worst case two-level partial-band jamming. For any E_b/N_J the asymptotic probability of error is not zero. The analysis shows that for large E_b/N_J the worst-case jammer ρ is very small. Hence simple redundancy in the form of diversity is next analyzed with majority logic decoding as well as linear combining at the receiving end. Both schemes turn out to be insufficient. The linear combining case indicates that when the outputs of the L diversity transmissions are summed up, small values of ρ, fraction of band jammed, as a function of L, the number of diversity transmissions, play a significant part in degrading the performance when using the sum statistic. This naturally suggests clipped linear combining, wherein we clip the output of each diversity transmission to make a more effective combining technique. The rationale for such a scheme is the expectation that our probability of error will decrease because now the infrequent transmissions with the large noise components will affect our sum much less. To allow for greater generality we allow the clipping level to be a function of L too. Our analysis indicates that for this case we recapture the same threshold phenomenon with orthogonal signals that we had for the AWGN channels. We do this by use of certain Central Limit Theorem approximations. Since such a threshold phenomenon is very sensitive to the kind of approximation used, we need to use a powerful non-uniform "Berry-Esseen" bound due to Michel and a less well-known form of the Central Limit Theorem due to Sirazhdinov and Mamatov. Our analysis shows that provided a certain relation is satisfied by the clipping level, the number of diversity transmissions and the number of signals,then, asymptotically, orthogonal signalling with diversity and clipped linear combining suffices to achieve capacity over the partial-band jammed channel.

3.2 Orthogonal Signaling over the AWGN Channel

We consider the AWGN with one sided noise-power spectral density N_0 watts/Hz and consider a set of M equi-energy orthogonal signals $s_i(t)$, $i = 1, \ldots, M$, limited to time duration T seconds and with average received power S watts. Thus the energy in each signal will be $E = ST$ joules and the orthonormal basis functions can be conveniently chosen as $\phi_i(t) = \frac{1}{\sqrt{E}} s_i(t)$ $(\int_0^T \phi_m(t)\phi_n(t)) = \delta_{mn}$, $m, n = 1, \ldots, M$, where $\delta_{mn} = 1$, $m = n$, and $\delta_{mn} = 0$ $m \neq n$). Using coherent correlation detection the error probability P_e is known to be

$$P_e(\frac{E_b}{N_0}, M) = 1 - \int_{-\infty}^{\infty} (\Phi(u))^{M-1} \frac{1}{\sqrt{2\pi}} \exp\left\{ \frac{-\left(u - \sqrt{\frac{2E_b}{N_0} \log_2 M}\right)^2}{2} \right\} du$$

where $E_b = \frac{E}{\log_2 M}$ = energy per bit and $\Phi(u)$ is the distribution function of a standard normal random variable.

It is well known that [Vite 66]

$$\lim_{M \to \infty} P_e(\frac{E_b}{N_0}, M) = 1 \quad \frac{E_b}{N_0} < \ln 2$$
$$= 0 \quad \frac{E_b}{N_0} > \ln 2$$

Now Shannon's formula for the capacity C of a channel of bandwidth W perturbed by additive Gaussian noise of uniform spectral density N_0 and signal power S is

$$C = W \log(1 + \frac{S}{N_0 W})$$

If we let the bandwidth approach infinity (which is required if we let the number of orthogonal signals increase to infinity) then we have that

$$\lim_{W \to \infty} C = \frac{S}{N_0 \ln 2}$$

the channel coding theorem asserts that for reliable communication over this channel the rate R(bits/sec) must satisfy

$$R < \frac{S}{N_0 \ln 2}$$

Now if T_b denotes the bit duration, i.e. $T_b = \frac{1}{R}$, then this condition is equivalent to $\frac{E_b}{N_0} > \ln 2$. The converse to the above theorem asserts that for all signal sets such that $R > \frac{S}{N_0 \ln 2}$, which is equivalent to $\frac{E_b}{N_0} < \ln 2$, the error probability approaches 1.

However as we show in the following section when we have partial-band jamming, orthogonal signals perform much more poorly.

3.3 Orthogonal Signaling in Partial-band Jamming

We show in this section that the following partial-band jamming strategy causes the error probability to be non-zero asymptotically for any value of $\frac{E_b}{N_J + N_0}$ (N_J is the one-sided power spectral density of the jamming noise). First we describe the model for the signalling and for the frequency hopper and dehopper and the background and jamming noise.

Let $\{\phi_i(t)\}_{i=1}^{M}$ $0 \leq t \leq T$ be a set of orthonormal signals with $s_j(t) = \sqrt{E}\phi_j(t)$ being the signal transmitted corresponding to symbol j. This signal is frequency hopped over q frequencies with one symbol per hop and transmitted over a partial-band jammed channel as $\overline{s}_j(t)$.

The jamming signal $j(t)$ at the receiver is modelled as a weighted sum of band-pass Gaussian processes $j(t) = \sum_{i=1}^{q} Z_i(t) j_i(t)$ where $j_i(t)$ is a Gaussian random process with zero mean and spectral density (one-sided) N_J over a bandwidth W/q Hz where W is the total spread bandwidth of the transmitted signal. In the subsequent analysis W/q is assumed to be much larger than $\frac{1}{T}$. Assume that each M-ary band lies entirely within or without the W/q bandwidth support of some $j_i(t)$ (this is a pessimistic assumption and our probabilities of error are higher than without this assumption). Also assume that the spectral density of $j_i(t)$, $S_i(f)$, is such that $S_i(f)S_j(f) = 0$ for all f and $i \neq k$. Thus $j_i(t)$ and $j_k(t)$ are independent

random processes for $i \neq k$. $Z_i(t)$ is a sequence of non-overlapping pulses of duration T, i.e. $Z_i(t) = Z_{i,m}, \quad mT \leq t \leq (m+1)T$. The jammer has the freedom to choose the distribution of the random variables $Z_{i,m}$ subject to an average power constraint:

$$E(\sum_{i=1}^{q} Z_{i,m}^2) \leq q$$

The partial band jammer chooses the following distribution for $Z_{i,m}$:

$$\Pr(Z_{i,m} = 0) = 1 - \rho \quad 0 \leq \rho \leq 1$$

$$\Pr(Z_{i,m} = \sqrt{\frac{1}{\rho}}) = \rho$$

where ρ is a constant representing the fraction of band which has interference. Thus when the jammer is on $j_i(t)$ has noise spectral density N_J/ρ and when the jammer is off $j_i(t)$ has noise spectral density 0. $Z_{i,m}$ are i.i.d. for each i and $Z_{i,m}$ is independent of $j_i(t)$.

The received signal is thus

$$r(t) = \overline{s}(t) + j(t) + n(t)$$

where $n(t)$ is the thermal noise, which is a white Gaussian process with one sided spectral density $N_0/2$. The signal is dehopped by a frequency dehopper whose output $r_d(t)$ can be written as:

$$r_d(t) = s(t) + \sum_{k=1}^{q} \delta(v_k, f(t)) Z_k(t) \hat{j}_k(t) + n(t)$$

where $\hat{j}_k(t)$ is $j_k(t)$ after frequency translation, $\{v_k\}_{k=1}^{q}$ is the set of frequencies hopped to and $f(t)$ is the hopping pattern i.e.

$$f(t) = f_j, \quad jT \leq t \leq (j+1)T, \quad f_j \in \{v_1, \ldots v_q\}$$

The demodulator processes the received signal by computing the M-dimensional vector

$$y = (y_1, \ldots, y_M)$$

where

$$y_i = \int_0^T r_d(t)\phi_i(t)dt$$

Now assume symbol j was sent, i.e. $s_j(t)$ was sent. Then for $i \neq j$

$$y_i = \int_0^T \sum_{k=1}^{q} \delta(v_k, f(t)) Z_k(t)\hat{j}_k(t)\phi_i(t)dt + \int_0^T n(t)\phi_i(t)dt$$

and for $i = j$

$$y_j = \int_0^T \sum_{k=1}^{q} \delta(v_k, f(t)) Z_k(t)\hat{j}_k(t)\phi_j(t)dt + \sqrt{E} + \int_0^T n(t)\phi_j(t)dt$$

Thus for $i \neq j$

$$y_i = Z_i n_i + N_i$$

and for $i = j$

$$y_j = \sqrt{E} + Z_j n_j + N_j$$

where N_i, $i = 1, \ldots, M$ are i.i.d. Gaussian random variables with mean 0 and variance $N_0/2$ and the n_i, $i = 1, \ldots, M$ are i.i.d. Gaussian random variables with mean 0 and variance $N_J/2$. Thus by conditioning on Z_j and then averaging, the error probability may be written as

$$P_e(\rho, E_b, N_0, N_J, M) = (1-\rho)P_e(\frac{E_b}{N_0}, M) + \rho P_e(\frac{E_b}{N_0 + \frac{N_J}{\rho}}, M)$$

For the worst-case partial-band jammer the error probability is

$$P_e(E_b, N_0, N_J, M) \triangleq \sup_{0 \leq \rho \leq 1} P_e(\rho, E_b, N_0, N_J, M)$$

For $\rho = 0$, $P_e(\rho, E_b, N_0, N_J, M) = P_e(\frac{E_b}{N_0}, M)$.

Now we show that for any signal-to-noise ratio the worst case jammer can ensure that the asymptotic probability of error does not go to zero. This is stated precisely in the following Theorem.

Theorem 1:

i) For $\frac{E_b}{N_0 + N_J} < \ln 2$

$$\lim_{M \to \infty} \max_{0 \le \rho \le 1} \rho P_e \left(\frac{E_b}{\frac{N_J}{\rho} + N_0} , M \right) + (1 - \rho) P_e \left(\frac{E_b}{N_0} , M \right) = 1 \qquad (3.1)$$

ii) For $\frac{E_b}{N_0 + N_J} > \ln 2$

$$\lim_{M \to \infty} \max_{0 \le \rho \le 1} \rho P_e \left(\frac{E_b}{\frac{N_J}{\rho} + N_0} , M \right) + (1 - \rho) P_e \left(\frac{E_b}{N_0} , M \right) = \bar{\rho} \qquad (3.2)$$

where $\bar{\rho}$ is the solution of the equation

$$\frac{E_b}{\frac{N_J}{\rho} + N_0} = \ln 2 \qquad (3.3)$$

i.e. $\bar{\rho} = \frac{\ln 2}{\frac{E_b}{N_J} - \frac{N_0}{N_J} \ln 2}$

Note : $\frac{\ln 2}{\frac{E_b}{N_J}} \le \bar{\rho} = \frac{\ln 2}{\frac{E_b}{N_J} - \frac{N_0}{N_J} \ln 2} \le \frac{\ln 2}{\frac{E_b}{N_0 + N_J}}$

Proof of Theorem 1:

i) $\frac{E_b}{N_0 + N_J} < \ln 2$

$$\lim_{M \to \infty} \max_{0 \le \rho \le 1} \rho P_e \left(\frac{E_b}{\frac{N_J}{\rho} + N_0} , M \right) + (1 - \rho) P_e \left(\frac{E_b}{N_0} , M \right) \qquad (3.4)$$

$$\ge \lim_{M \to \infty} \max_{0 \le \rho \le 1} \rho P_e \left(\frac{E_b}{\frac{N_J}{\rho} + N_0} , M \right) \qquad (3.5)$$

$$\ge \lim_{M \to \infty} \max_{0 \le \rho \le 1} \rho P_e \left(\frac{E_b}{N_0 + N_J} , M \right) \qquad (3.6)$$

(Since $P_e(x, M)$ is a decreasing function of x) (see Appendix G)

$$\geq \lim_{M \to \infty} P_e \left(\frac{E_b}{N_0 + N_J} , M \right) = 1 \tag{3.7}$$

if $\frac{E_b}{N_0 + N_J} > \ln 2$

For any $\epsilon > 0$, let $\rho_1 = \bar{\rho} - \epsilon$ where $\bar{\rho}$ is such that

$$\frac{E_b}{\frac{N_J}{\bar{\rho}} + N_0} = \ln 2 \tag{3.8}$$

Then

$$\frac{E_b}{\frac{N_J}{\rho_1} + N_0} < \ln 2 \tag{3.9}$$

$$\liminf_{M \to \infty} \max_{0 \leq \rho \leq 1} \rho P_e \left(\frac{E_b}{\frac{N_J}{\rho} + N_0} , M \right) + (1 - \rho) P_e \left(\frac{E_b}{N_0} , M \right) \tag{3.10}$$

$$\geq \liminf_{M \to \infty} \max_{0 \leq \rho \leq 1} \rho P_e \left(\frac{E_b}{\frac{N_J}{\rho} + N_0} , M \right) \tag{3.11}$$

$$\geq \liminf_{M \to \infty} \rho_1 P_e \left(\frac{E_b}{\frac{N_J}{\rho_1} + N_0} , M \right) \tag{3.12}$$

$$= \rho_1 . \tag{3.13}$$

Since this is true for any $\epsilon > 0$ we can infer that

$$\liminf_{M \to \infty} \max_{0 \leq \rho \leq 1} \rho P_e \left(\frac{E_b}{\frac{N_J}{\rho} + N_0} , M \right) + (1 - \rho) P_e \left(\frac{E_b}{N_0} , M \right) \geq \bar{\rho} \tag{3.14}$$

Now let ρ_M be the value of ρ which achieves the maximum in

$$\max_{0 \leq \rho \leq 1} \rho P_e \left(\frac{E_b}{\frac{N_J}{\rho} + N_0} , M \right) \tag{3.15}$$

Then for sufficiently large M, $\rho_M \leq \bar{\rho}$. Else for every M there would be a $M_1 > M$ and $\rho_{M_1} > \bar{\rho}$ such that

$$\rho_{M_1} P_e \left(\frac{E_b}{\frac{N_J}{\rho_{M_1}} + N_0} , M_1 \right) \geq \rho P_e \left(\frac{E_b}{\frac{N_J}{\rho} + N_0} , M \right) \tag{3.16}$$

for all ρ, $0 \leq \rho \leq 1$. Now for any $\epsilon > 0$ the right-hand side of (3.16) for sufficiently large M ($M > M_2$ say) can be made greater than $\bar{\rho} - \epsilon$. Since the left-hand side of (3.16) is obviously approaching 0 for $\rho_{M_1} > \bar{\rho}$, we have a contradiction. Therefore

$$\limsup_{M\to\infty} \max_{0\le\rho\le 1} \rho P_e\left(\frac{E_b}{\frac{N_J}{\rho}+N_0},\ M\right) + (1-\rho)\ P_e\left(\frac{E_b}{N_0},\ M\right) \tag{3.17}$$

$$\le\ \overline{\rho}$$

From (3.13) and (3.17) it follows that

$$\lim_{M\to\infty} \max_{0\le\rho\le 1} \rho\ P_e\left(\frac{E_b}{\frac{N_J}{\rho}+N_0},\ M\right) + (1-\rho)\ P_e\left(\frac{E_b}{N_0},\ M\right) \tag{3.18}$$

$$= 1 \quad \text{if} \quad \frac{E_b}{N_0+N_J} < \ln 2$$

$$= \overline{\rho} \quad \text{if} \quad \frac{E_b}{N_0+N_J} > \ln 2$$

The jammer thus is able to thwart reliable communication at any signal-to-noise ratio by choosing ρ small enough (but not too small). Since the choice of a small ρ by the jammer allows a number of transmissions to pass through unscathed but corrupts the rest substantially it seems likely that simple coding such as through diversity may be able to recover some of this loss in rate due to partial-band jamming. We pursue such an investigation in the following section.

3.4 Orthogonal Signalling with Diversity

In this section we try to use time diversity to overcome the partial-band jammer, i.e. make the worst-case partial-band jammer no more deleterious than a broad-band jammer of equivalent power. With diversity L, the energy per bit, E_b, is $\frac{LE}{\log_2 M}$ and we shall use E_b' to denote $\frac{E}{\log_2 M}$.

Assume that symbol j was sent, i.e. $s_j(t)$ was transmitted. Then using the earlier demodulator we have L M-component vector decision variables:

$$y_l = (y_{l,1}, \ldots, y_{l,M}) \quad l = 1, 2, \ldots, L$$

where for $i \neq j$

$$y_{l,i} = Z_{l,i} n_{l,i} + N_{l,i}$$

and for $i = j$

$$y_{l,j} = \sqrt{E} + Z_{l,j}n_{l,j} + N_{l,j}$$

where $n_{l,i}$ is a $N(0, N_J/2)$ Gaussian random variable, $Z_{l,i}$ is 0 with probability 1 - ρ and $\sqrt{\frac{1}{\rho}}$, with probability ρ, $N_{l,i}$ is a $N(0, N_0/2)$ Gaussian random variable and $Z_{l,i}$, $n_{l,i}$ and $N_{l,i}$ are independent. Also, $y_1, \ldots, y_L$ are i.i.d. random vectors.

3.4.1 Majority Logic Combining

We first investigate the following majority logic decoding strategy. The receiver observes the output during each interval of duration T and picks i if the output of the ith correlator is maximum. At the end of L intervals the output of the decoder is the symbol which has been picked a maximum number of times. We do the asymptotic analysis assuming the diversity L to be an increasing function of M for sufficiently large M.

Now in each time interval of length T the probability of error (i.e. the probability that j will not be picked) is

$$\alpha = \rho P_e(\frac{E_b'}{\frac{N_J}{\rho} + N_0}, M) + (1 - \rho)P_e(\frac{E_b'}{N_0}, M)$$

where $P_e(x, y)$ is the probability of error for y orthogonal signals with bit energy to noise ratio being x. Thus the probability of j being picked in each interval is $1 - \alpha$ and the probability of $i \neq j$ being picked in each interval is $\frac{\alpha}{M-1}$. Using the above scheme we denote the probability of error as $P_e^L(\rho_L, E_b', N_J, N_0, M)$ where ρ_L is used to denote the jammer's choice of ρ as a function of L. Since the channel is independent between repetitions and the same input is applied to the channel during each of the L repetitions, the outputs of the channel during the L diversity transmissions are i.i.d. random variables with finite mean. Thus we can utilize the Weak Law of Large Numbers as $L \rightarrow \infty$, i.e. the probability that the proportion

of times we choose j out of the L repetitions differs from 1 - α_L (where $\alpha_L = \rho_L P_e(\frac{E_b'}{\frac{N_J}{\rho_L}+N_0}) + (1-\rho_L)P_e(\frac{E_b'}{N_0}, M)$) by $\epsilon > 0$ goes to zero:

$$\lim_{L\to\infty} \text{Prob}\{|\frac{\text{number of times } j \text{ is picked}}{L} - (1-\alpha_L)| > \epsilon\} \to 0$$

$$\lim_{L\to\infty} \text{Prob}\{|\frac{\text{number of times } i(\neq j) \text{ is picked}}{L} - \frac{\alpha_L}{M-1}| > \epsilon\} \to 0$$

Thus we see that using the above decoding strategy the limiting probability of error will be zero or one according as

$$1 - \alpha_L > \frac{\alpha_L}{M-1} \tag{3.19}$$

or

$$\frac{\alpha_L}{M-1} > 1 - \alpha_L \tag{3.20}$$

and we examine when this is true. Condition (3.19) may be written as

$$\alpha_L < \frac{M-1}{M}. \tag{3.21}$$

Now if $\frac{E_b}{N_J+N_0} > \ln 2$ we know from the previous section that

$$\lim_{M\to\infty} \sup_{0\le\rho\le 1} \rho P_e(\frac{E_b'}{\frac{N_J}{\rho}+N_0}) + (1-\rho)P_e(\frac{E_b'}{N_0}, M) = \frac{\ln 2}{\frac{E_b}{N_J} - \frac{N_0}{N_J}\ln 2} = \overline{\rho}.$$

Hence

$$\rho_L P_e(\frac{E_b'}{\frac{N_J}{\rho_L}+N_0}) + (1-\rho_L)P_e(\frac{E_b'}{N_0}, M) \le \overline{\rho} < 1.$$

Therefore for all sufficiently large M,

$$\alpha_L < \frac{M-1}{M}.$$

Thus we can say that

$$\lim_{M,L\to\infty} \sup_{0\le\rho_L\le 1} P_e^L(\rho_L, E_b', N_J, N_0, M) = 0$$

if $\frac{E_b'}{N_0+N_J} > \ln 2$.

Now assume $\frac{E'_b}{N_0+N_J} < \ln 2$. Condition (3.20) may be written as

$$\rho_L P_e(\frac{E'_b}{\frac{N_J}{\rho_L}+N_0}, M) + (1-\rho_L)P_e(\frac{E'_b}{N_0}, M)$$

$$> \frac{\rho_L P_e(\frac{E'_b}{\frac{N_J}{\rho_L}+N_0}, M) + (1-\rho_L)P_e(\frac{E'_b}{N_0}, M)}{M-1} \tag{3.22}$$

We show that for a particular choice of ρ_L ($\rho_L = 1$) condition (3.20) holds for all sufficiently large M. With $\rho_L = 1$, condition (3.20) becomes

$$1 - P_e(\frac{E'_b}{N_J+N_0}, M) < \frac{P_e(\frac{E'_b}{N_J+N_0}, M)}{M-1} \tag{3.23}$$

for which it clearly suffices if

$$\lim_{M\to\infty}(M-1)(1-P_e(\frac{E'_b}{N_J+N_0}, M)) = 0 \tag{3.24}$$

since we know that

$$\lim_{M\to\infty} P_e(\frac{E'_b}{N_J+N_0}, M) = 1$$

By an easy extension of the derivation in [Vite 64, pp.106-134] (3.24) can be verified. Thus we can say from (3.19),(3.20) and (3.24) that

$$\lim_{M,L\to\infty} \sup_{0\le\rho\le 1} P_e^L(\rho_L, E'_b, N_J, N_0, M) = 1$$

We have established so far that for $\frac{E'_b}{N_0+N_J} > \ln 2$

$$\lim_{M,L\to\infty} \sup_{0\le\rho_L\le 1} P_e^L(\rho_L, E'_b, N_J, N_0, M) = 0 \tag{3.25}$$

and for $\frac{E'_b}{N_0+N_J} < \ln 2$

$$\lim_{M,L\to\infty} \sup_{0\le\rho_L\le 1} P_e^L(\rho_L, E'_b, N_J, N_0, M) = 1 \tag{3.26}$$

Note however that in our diversity signalling scheme the actual bit energy to noise ratio is is not $\frac{E'_b}{N_0+N_J}$ but $\frac{E_b}{N_0+N_J} = \frac{LE'_b}{N_0+N_J}$ and thus in both (3.25) and(3.26) we have allowed very large bit energy to noise ratios. Therefore we can say that

when we use orthogonal signalling with diversity against an intelligent jammer our scheme allows reliable communication when the bit energy to noise ratio ($\frac{E_b}{N_0+N_J}$) increases fast enough with diversity i.e. $\frac{E_b}{N_0+N_J} > L \ln 2$ for sufficiently large L. If $\frac{E_b}{N_0+N_J} < L \ln 2$ for sufficiently large L then the worst-case jammer can frustrate even such high energy to noise ratios.

3.4.2 Linear Combining

Another commonly used method of diversity combining is linear combining. Here we process the output to get the following decision variables:

$$D_i = \sum_{l=1}^{L} y_{l,i} \;\; i = 1, \ldots, M$$

When symbol j is sent then for $i \neq j$

$$D_j = \sum_{l=1}^{L} (Z_{l,j} n_{l,j} + N_{l,j})$$

and for $i = j$

$$D_j = \sum_{l=1}^{L} (\sqrt{E} + Z_{l,j} n_{l,j} + N_{l,j})$$

Again we use $P_e^L(\rho_L, E_b', N_J, N_0, M)$ to denote the probability of error. By conditioning on the number of diversities jammed we can write

$$P_e^L(\rho_L, E_b', N_J, N_0, M) = \sum_{k=0}^{L} \binom{L}{k} \rho_L^k (1-\rho_L)^{L-k} P_e(\frac{LE_b'}{\frac{kN_J}{\rho_L} + LN_0}, M).$$

Consider first the case $\frac{E_b'}{N_0 + N_J} > \ln 2$.

$$P_e^L(\rho_L, E_b', N_J, N_0, M) =$$

$$(1-\rho_L)^L P_e(\frac{E_b'}{N_0}, M) + \sum_{k=1}^{L} \binom{L}{k} \rho_L^k (1-\rho_L)^{L-k} P_e(\frac{\rho_L L E_b'}{kN_J + \rho_L L N_0}, M). \qquad (3.27)$$

Since $\frac{E_b'}{N_0} > \frac{E_b'}{N_0 + N_J} > \ln 2$, the first term goes to 0 with M. Let the jammer choose $\rho_L = \frac{s}{L}$ where s is some number > 0. Let α denote the second term in

(3.27). Then

$$\alpha = \sum_{k=1}^{L} \binom{L}{k} \rho_L^k (1-\rho_L)^{L-k} P_e(\frac{\rho_L L E_b'}{kN_J + \rho_L L N_0}, M)$$

$$= \sum_{k=1}^{L} \binom{L}{k} \rho_L^k (1-\rho_L)^{L-k} P_e(\frac{sE_b'}{kN_J + sN_0}, M)$$

$$= \sum_{k=1}^{L} \binom{L}{k} \rho_L^k (1-\rho_L)^{L-k} P_e(\frac{E_b'}{\frac{kN_J}{s} + LN_0}, M).$$

Since s is arbitrary, let it be chosen so that it satisfies

$$\frac{E_b}{\frac{N_J}{s} + N_0} < \ln 2$$

that is, s is chosen to be less than $\overline{\rho}$ (as in Section 3.1). Then we see that every term $P_e(\frac{E_b'}{\frac{kN_J}{s}+LN_0}, M)$ approaches 1 with increasing L, M. Hence as L, $M \to \infty$, α goes to

$$\lim_{L\to\infty} \sum_{k=1}^{L} \binom{L}{k} \rho_L^k (1-\rho_L)^{L-k}$$

$$= \lim_{L\to\infty} (1 - (1-\rho_L)^L)$$

$$= 1 - e^{-s}.$$

By choosing s to be only slightly smaller than $\overline{\rho}$ we see that

$$\lim_{M,L\to\infty} P_e^L(\rho_L, E_b', N_J, N_0, M) \geq 1 - e^{-\overline{\rho}} > 0. \tag{3.28}$$

Now we consider the case $\frac{E_b'}{N_0 + N_J} < \ln 2$. Suppose that the jammer chooses $\rho_L = 1$ for all L. Then

$$P_e^L(\rho_L, E_b', N_J, N_0, M) = P_e(\frac{E_b'}{N_J + N_0}, M)$$

Since $\frac{E_b'}{N_0 + N_J} < \ln 2$

$$\lim_{L,M\to\infty} P_e(\frac{E_b'}{N_J + N_0}, M) = 1. \tag{3.29}$$

Thus we have established for linear combining that for $\frac{E_b'}{N_0+N_J} > \ln 2$

$$\lim_{M,L\to\infty} \sup_{0\leq\rho_L\leq 1} P_e^L(\rho_L, E_b', N_J, N_0, M) \geq 1 - e^{-\overline{\rho}} \tag{3.30}$$

and for $\frac{E_b'}{N_0+N_J} < \ln 2$

$$\lim_{M,L\to\infty} \sup_{0\le\rho_L\le 1} P_e^L(\rho_L, E_b', N_J, N_0, M) = 1 \tag{3.31}$$

Thus (3.28) indicates that the worst-case jammer (when $\frac{E_b'}{N_0+N_J} > \ln 2$) jams such that $\rho_L \to 0$ and (3.29) shows that if $\frac{E_b'}{N_0+N_J} < \ln 2$ then the worst-case jammer can choose $\rho_L = 1$. In either case the jammer can frustrate very high bit energy to noise ratios ($\frac{E_b}{N_0+N_J} = \frac{LE_b'}{N_0+N_J}$).

We note that although for $\frac{E_b'}{N_0+N_J} > \ln 2$ the worst- case jammer chooses ρ_L such that $\rho_L \to 0$, ρ_L cannot approach 0 too fast for the following reason. Let the jammer choose $\rho_L = \frac{1}{L^\alpha} (\alpha > 1)$. Now

$$P_e^L(\rho_L, E_b', N_J, N_0, M) =$$

$$(1-\rho_L)^L P_e(\frac{E_b'}{N_0}, M) + \sum_{k=1}^{L} \binom{L}{k} \rho_L^k (1-\rho_L)^{L-k} P_e(\frac{\rho_L L E_b'}{kN_J + \rho_L L N_0}, M).$$

By the jammer's choice of ρ_L

$$P_e^L(\frac{1}{L^\alpha}, E_b', N_J, N_0, M) =$$

$$(1-\frac{1}{L^\alpha})^L P_e(\frac{E_b'}{N_0}, M) + \sum_{k=1}^{L} \binom{L}{k} (\frac{1}{L^\alpha})^k (1-\frac{1}{L^\alpha})^{L-k} P_e(\frac{L^{\alpha-1} E_b'}{kN_J + L^{\alpha-1} N_0}, M)$$

Again we point out that the first term goes to 0 with L, M. The second term is

$$\sum_{k=1}^{L} \binom{L}{k} \rho_L^k (1-\rho_L)^{L-k} P_e(\frac{E_b'}{L^{\alpha-1} k N_J + N_0}, M) \tag{3.32}$$

Every term $P_e(\frac{E_b'}{L^{\alpha-1} k N_J + N_0}, M)$ in the summation goes to 1 as L increases and so as $L \to \infty$, (3.32) becomes

$$\lim_{L\to\infty} (1 - (1-\rho_L)^L). \tag{3.33}$$

Next we show that $\lim_{L\to\infty}(1-\rho_L)^L = 1$ and consequently

$$\lim_{M,L\to\infty} P_e^L(\rho_L, E_b', N_J, N_0, M) = 0$$

We now prove that $\lim_{L\to\infty}(1-\rho_L)^L = 1$.

$$(1-\rho_L)^L = (1-\frac{1}{L^\alpha})^L$$

$$= (1-(\frac{1}{L^{\alpha-1}})(\frac{1}{L}))^L$$

We know $\lim_{L\to\infty}(1-\frac{t}{L})^L = e^{-t}$ and so $\forall t \in (0,1)\ \ \exists\, L_0 \ni \forall L \geq L_0$

$$(1-\frac{t}{L})^L \leq (1-(\frac{1}{L^{\alpha-1}})(\frac{1}{L}))^L$$

and so

$$\lim_{L\to\infty}(1-(\frac{1}{L^{\alpha-1}})(\frac{1}{L}))^L \geq \lim_{L\to\infty}(1-\frac{t}{L})^L = e^{-t}$$

Since t is arbitrary we have

$$\lim_{L\to\infty}(1-\rho_L)^L = 1$$

In linear combining we see that since the output of the L diversity transmissions is summed up, small values of ρ_L, while making it less likely that a diversity transmission is jammed, make the probability of error on such a jammed transmission very high because of the low bit energy to noise ratio on such a transmission. The jammer's strategy of choosing ρ_L to be inversely proportional with diversity level L is intuitively explicable. Since L outputs are added he jams such that if he hits one transmission there is enough jamming power to corrupt the sum statistic. On the other hand, trying to put too much jamming power in a single jammed transmission turns out not to be too effective because then the number of good transmissions increases sufficiently enough to overcome the jamming noise. We thus see that in linear combining the few jammed transmissions have a significant effect on the probability of error. This suggests that if we use a form of clipped linear combining wherein we clip the output of each diversity transmission our probability will improve because the infrequent transmissions with the high P_e values will affect our sum much less. Possibly the clipping level can be chosen as a function of L. In the next section we conduct the analysis using this idea.

3.4.3 Clipped Linear Combining

The analysis in the previous sections suggests that the jammer contributes infrequently to the decision statistics but when he does so his contribution is large. This suggests that some form of limiting the jammer's contribution to the decision statistics may be effective. Here we first clip each of the diversity transmission outputs by a symmetric limiter and then combine linearly.

Thus the decision variables we use are

$$D_j' = \sum_{l=1}^{L} C_L\left(Z_{j,l}n_{j,l} + N_{j,l} + \sqrt{E}\right) \quad \text{when } j \text{ is sent} \tag{3.34}$$

$$D_i' = \sum_{l=1}^{L} C_L(Z_{i,l}n_{i,l} + N_{i,l}) \quad i \neq j \tag{3.35}$$

where

$$\begin{aligned} C_L(x) &= x, \quad -\alpha_L \leq x \leq \alpha_L \\ &= \alpha_L, \quad x > \alpha_L \\ &= -\alpha_L, \quad x < -\alpha_L. \end{aligned}$$

The decision rule is to decide that i was sent where $D_i' = \max_i D_i'$. Using this decision rule we calculate the probabilities of being correct.

$$Pr(\text{correct} \mid j \text{ is sent}) = Pr(i = j \mid j \text{ is sent}) \tag{3.36}$$

$$= Pr(D_i' < D_j', \; i \neq j \mid j \text{ is sent}) \tag{3.37}$$

Again we use $P_e^L(\rho_L, E_b', N_J, N_0, M)$ to denote the probability of error. Now let

$$\begin{aligned} D_j = \frac{D_j'}{\sqrt{L}\,\sigma_L'} &= \frac{1}{\sqrt{L}} \sum_{l=1}^{L} \frac{C_L\left(Z_{j,l}n_{j,l} + N_{j,l} + \sqrt{E}\right)}{\sigma_L'} \\ &= \frac{1}{\sqrt{L}} \sum_{l=1}^{L} \frac{Y_{L,l}}{\sigma_L'} \end{aligned} \tag{3.38}$$

where $Y_{L,l} = C_L\left(Z_e\, n_{j,l} + N_{j,l} + \sqrt{E}\right)$ (i.e. the unnormalized decision variable containing the signal) $\sigma_L^j = \sqrt{\text{Var } D_j'}$ and let

$$D_i = \frac{D_i'}{\sqrt{L}\,\sigma_L^j} = \frac{1}{\sqrt{L}} \frac{\sum_{l=1}^{L} C_l\,(Z_L\, n_{i,l} + N_{i,l})}{\sigma_L^j} = \frac{1}{\sqrt{L}} \sum_{l=1}^{L} \frac{X_{L,l}}{\sigma_L^j} \tag{3.39}$$

where $X_{L,l} = C_L(Z_e\, n_{j,l} + N_{j,l}\,)$ (i.e. the unnormalized decision variables with only noise). Note that $Pr(D_i' < D_j', i \neq j | j \text{ is sent}) = Pr(D_i < D_j, i \neq j | j \text{ is sent})$. Using the D_i 's as decision variables we show that we can recapture the asymptotic performance of orthogonal signals over the AWGN channel.

Specifically if the number of orthogonal signals, M, and the diversity level L increase in a certain relation to each other and the clipping level α_L is allowed to increase with L, but not too fast, then the probability of error with clipped linear combining exhibits the same threshhold behavior in worst-case partial-band jamming that orthogonal signalling in AWGN achieves, i.e.

Theorem 2: i) For $\dfrac{E_b}{N_0 + N_J} < \ln 2$

$$\lim_{M,L\to\infty} \sup_{0\le\rho\le 1} P_e^L(\rho_L, E_b', N_J, N_0, M) = 1$$

ii) For $\dfrac{E_b}{N_0 + N_J} > \ln 2$

$$\lim_{M,L\to\infty} \sup_{0\le\rho\le 1} P_e^L(\rho_L, E_b', N_J, N_0, M) = 0$$

Proof of Theorem 2:

We proceed as follows. First we show that the above threshold phenomenon is true if our decision variables were all Gaussian with parameters chosen in a certain way. We do this in Lemma 1. The corresponding probability of correct decoding we denote as $P_{c,a}$. Then by use of some fairly powerful Central Limit approximations we show that the difference between $P_{c,a}$ and the actual probability of error goes

to zero establishing the Theorem. We now proceed with Lemma 1. Let

$$P_{c,a} \triangleq \int_{-\infty}^{\infty} \overline{\phi_J}(x) \left[\overline{\Phi}_J(x)\right]^{M-1} dx$$

where $\overline{\phi_J}(x)$ is the density of a random variable distributed as $\mathcal{N}\left(\sqrt{\frac{L(E+\epsilon_L)}{(\frac{N_0}{2}+\frac{N_J}{2})\beta_L}}, 1\right)$ and $\beta_L \to 1$ and $\epsilon_L \to 0$ and $\overline{\Phi}_J(x)$ is the distribution function of a normal random variable distributed as $\mathcal{N}(0, \tau_L)$ where $\tau_L \to 1$. Next we show that:

Lemma 1:

i) If $\frac{E_b}{N_0 + N_J} \triangleq \frac{LE}{\log_2 M}\left(\frac{1}{N_0 + N_J}\right) < \ln 2$, then $\lim_{M,L\to\infty} P_{c,a} = 0$

ii) If $\frac{E_b}{N_0 + N_J} > \ln 2$, then $\lim_{M,L\to\infty} P_{c,a} = 1$

Proof of Lemma 1:

$$P_{c,a} = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(x - \sqrt{\frac{L(E+\epsilon_L)}{\left(\frac{N_0+N_J}{2}\right)\beta_L}}\right)^2\right]\left[\int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\frac{u^2}{\tau_L}} du\right]^{M-1} dx$$

Let

$$x_1 = x - \sqrt{\frac{L(E+\epsilon_L)}{\left(\frac{N_0+N_J}{2}\right)\beta_L}}, \quad w = \frac{u}{\sqrt{\tau_L}}.$$

Hence

$$P_{c,a} = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x_1^2}\left[\int_{-\infty}^{\frac{x_1}{\sqrt{\tau_L}} + \frac{1}{\sqrt{\tau_L}}\sqrt{\frac{L(E+\epsilon_L)}{\left(\frac{N_0+N_J}{2}\right)\beta_L}}} \frac{\sqrt{\tau_L}}{\sqrt{2\pi}} e^{-\frac{1}{2}w^2} dw\right]^{M-1} dx_1$$

and

$$\lim_{M,L\to\infty} P_{c,a} = \int_{-\infty}^{\infty} \lim_{M,L\to\infty}(\quad) dx_1$$

(from the Dominated Convergence Theorem since the integrand is dominated by $e^{-\frac{1}{2}x_1^2}$ which is integrable). Now consider

$$\gamma \triangleq \lim_{M,L\to\infty} \gamma_{M,L}$$

$$= \lim_{M,L\to\infty} \ln\left(\sqrt{\tau_L}\,\Phi\left(\frac{x_1}{\sqrt{\tau_L}} + \frac{1}{\sqrt{\tau_L}}\sqrt{\frac{2LE + 2L\epsilon_L}{(N_0 + N_J)\beta_L}}\right)\right)^{M-1}$$

Denoting $\frac{E_b}{N_0 + N_J}$ as δ and $\tau_L\beta_L$ as q_L, (3.40) becomes

$$\lim_{M,L\to\infty} \frac{\ln\left(\sqrt{\tau_L}\,\Phi\left(\frac{x_1}{\sqrt{\tau_L}} + \sqrt{\frac{2\delta\,\log_2 M\,\left(1+\frac{\epsilon_L}{E}\right)}{q_L}}\right)\right)}{\frac{1}{M-1}}$$

where $q_L \to 1$. Using L ' Hospital's rule we get

$$\lim_{M,L\to\infty} \frac{\sqrt{\tau_L}}{\Phi\left(\frac{x_1}{\sqrt{\tau_L}} + \sqrt{\frac{2\delta\,\log_2 M\left(1+\frac{\epsilon_L}{E}\right)}{q_L}}\right)}$$

$$\frac{1}{\sqrt{2\pi}}\exp\left[-\frac{1}{2}\left(\frac{x_1^2}{\tau_L} + \frac{2\delta\,\log_2 M\left(1+\frac{\epsilon_L}{E}\right)}{q_L} + \frac{2x_1}{\sqrt{\tau_L}}\sqrt{2\delta\,\log_2 M\,\left(1+\frac{\epsilon_L}{E}\right)}\right)\right]$$

$$\frac{2\delta\left(1+\frac{\epsilon_L}{E}\right)(-(M-1)^2)}{2M\,\ln\,2\sqrt{\frac{2\delta\,\ln M}{\ln\,2}\left(1+\frac{\epsilon_L}{E}\right)}\sqrt{q_L}}$$

$$= \lim_{M,L\to\infty} \frac{1}{\sqrt{2\pi}}\sqrt{\frac{\tau_L\delta\left(1+\frac{\epsilon_L}{E}\right)}{q_L\,\ln\,M\,\ln\,2}}\,\frac{-(M-1)^2}{M}\,\frac{1}{M^{\frac{\delta(1+\epsilon_L)}{\ln\,2\,q_L}}}$$

$$\frac{1}{\Phi\left(\frac{x_1}{\sqrt{\tau_L}} + \sqrt{\frac{2\delta\,\log_2 M\,\left(1+\frac{\epsilon_L}{E}\right)}{q_L}}\right)}$$

$$\exp\left[-\frac{x_1^2}{2\tau_L} - \frac{x_1}{\sqrt{\tau_L}}\sqrt{\frac{2\delta\,\log_2 M\,\left(1+\frac{\epsilon_L}{E}\right)}{q_L}}\right]$$

$\longrightarrow \quad -\infty \quad$ if $\delta < \ln\ 2$

$\longrightarrow\ 0 \quad$ if $\delta > \ln\ 2$

which suffices to make (i) and (ii) true as claimed. Thus $\ln\,\gamma_{M,L} \to -\infty$ if $\delta < \ln 2$ so that $\gamma_{M,L} \to 0$ for $\delta < \ln 2$. For $\delta > \ln 2$, $\ln\,\gamma_{M,L} \to 0$ so that $\gamma_{M,L} \to 1$. □

Next we show that

$$\lim_{M\to\infty}\left|P_{e,a} - P_e^L(\rho_L, E_b', N_J, N_0, M)\right| = 0 \tag{3.41}$$

where $P_c^L(\rho_L, E_b', N_J, N_0, M) = 1 - P_e^L(\rho_L, E_b', N_J, N_0, M)$. For ease of notation we denote $P_c^L(\rho_L, E_b', N_J, N_0, M)$ simply as P_c. Now

$$| P_{c,a} - P_c | = | P_{c,a} - \int_{-\infty}^{\infty} \overline{\phi}_j(x) [F_{D_i}(x)]^{M-1} dx + \int_{-\infty}^{\infty} \overline{\phi_j}(x) [F_{D_i}(x)]^{M-1} dx - P_c |$$

where $\overline{\phi_j}(x)$ is the density of a random variable distributed as a $\mathcal{N}\left(\frac{L(E+\epsilon_L)}{\beta_L \frac{(N_0+N_J)}{2}}\right)$ with $\epsilon_L \to 0$ and $\beta_L \to 1$ and $F_{D_i}(x)$ is the distribution function of the random variable D_i. By the triangle inequality

$$| P_{c,a} - P_c | \leq \int_{-\infty}^{\infty} \overline{\phi}_j(x) | F_{D_i}^{M-1}(x) - \overline{\Phi}_i^{M-1}(x) | dx + \int_{-\infty}^{\infty} | \overline{\phi}_j(x) - F_{D_j}'(x) | | F_{D_i}(x)|^{M-1} dx + C[Pr\ |Y_{Li}| \geq \alpha_L]^L$$

where $F_{D_j}'(x)$ is the density of the absolutely continuous part of $F_{D_j}(x)$ and C is some constant. In general if $0 \leq a, b \leq 1$

$$| a^M - b^M | = | (a-b)(a^{M-1} + a^{M-2} b + a^{M-3} b^2 + \ldots\ldots + b^{M-1}) | \leq |a-b| M.$$

Let

$$\Delta_1 = \int_{-\infty}^{\infty} \overline{\phi}_j(x) | F_{D_i}^{M-1}(x) - \overline{\Phi}_i^{M-1}(x) | dx$$

and

$$\Delta_2 = \int_{-\infty}^{\infty} | \overline{\phi}_j(x) - F_{D_j}'(x) | | F_{D_i}(x)|^{M-1} dx + C[Pr\ |Y_{Li}| \geq \alpha_L]^L$$

Then

$$\Delta_1 \leq \int_{-\infty}^{\infty} M \overline{\phi}_j(x) | F_{D_i}(x) - \overline{\Phi}_i(x) | dx$$

and

$$\Delta_2 = \int_{-\infty}^{\infty} | \overline{\phi}_j(x) - F_{D_j}'(x) | | F_{D_i}(x)|^{M-1} dx + C[Pr\ |Y_{Li}| \geq \alpha_L]^L$$

Now

$$\begin{aligned} F_{D_i}(x) &= Pr\left(\frac{1}{\sqrt{L}}\sum_{l=1}^{L}\frac{X_{Ll}}{\sigma_L^l} \leq x\right) \\ &= Pr\left(\frac{1}{\sqrt{L}}\sum_{l=1}^{L}\frac{X_{Ll}}{\sigma_L^l} \leq \frac{x}{\frac{\sigma_L^l}{\sigma_L^j}}\right) \\ &= F_{\overline{D}_i}\left(\frac{x}{\frac{\sigma_L^l}{\sigma_L^j}}\right) \text{ (say)} \end{aligned}$$

Let $\frac{\sigma_L^l}{\sigma_L^j} = \beta_L$. Then $\beta_L \rightarrow 1$ (see Appendix J). Hence

$$\Delta_1 \leq \int_{-\infty}^{\infty} M\,\overline{\phi_j}(x) \mid F_{\overline{D}_i}\left(\frac{x}{\beta_L}\right) - \Phi_i\left(\frac{x}{\beta_l}\right) \mid dx$$

$$= \int_{-\infty}^{\infty} \frac{M}{\beta_l}\,\overline{\phi_j}\,(\beta_L x) \mid F_{\overline{D}_i(x)} - \overline{\Phi}_i(x) \mid dx.$$

Now we use Michel's version of the non-uniform Berry-Esseen Theorem [Mich 81], i.e.

Theorem : If $X_1, \ldots, X_n$ is a sequence of i.i.d. random variables with $E(X_1) = 0$, $E(X_1^2) = 1$ and $F_n(x)$ denotes the distribution of $\frac{X_1 + \ldots + X_n}{\sqrt{n}}$ then

$$\mid F_n(x) - \Phi(x) \mid \leq \frac{C_0}{\sqrt{n}}\,\frac{E \mid X_1 \mid^3}{(1 + |x|^3)}$$

(Furthermore $C_0 \leq 30.54$)

Since $\overline{D}_i = \frac{1}{\sqrt{L}}\sum_{l=1}^{L}\frac{X_{Ll}}{\sigma_L^l} = \frac{1}{\sqrt{L}}\sum_{l=1}^{L} V_{Ll}$ (say) where V_{Ll} are i.i.d. with $E(V_{Ll}) = 0$, $E(V_{Ll}^2) = 1$ and $E(|V_{Ll}|^3) \leq \alpha_L^3$ we have

$$\Delta_1 \leq \int_{-\infty}^{\infty} \frac{c\,M\,\alpha_L^3}{\sqrt{L}(1 + |x|^3)}\,dx$$

$$< \bar{c}\,\frac{M\alpha_L^3}{\sqrt{L}}$$

for some constants c and $\bar{c}$. Therefore if M, α_L and L are chosen so that

$$\frac{M\alpha_L^3}{\sqrt{L}} \rightarrow 0 \quad \text{then } \Delta_1 \rightarrow 0$$

Now we address the second term Δ_2.

$$\Delta_2 = \int_{-\infty}^{\infty} \left| \overline{\phi_j}(x) - F'_{D_j}(x) \right| \, |F_{D_i}(x)|^{M-1} \, dx$$

$$+ C \, [Pr \, |Y_{L,l}| \geq \alpha_L]^L$$

It follows that

$$\nu_L \triangleq C \, [P\{|Y_{L,l}| \geq \alpha_L\}]^L \leq C \, [P\{|Y_{L,l} - \mu_L| \geq \alpha_L - \mu_L]^L$$

where $\mu_L = \mu_{Y_{L,l}}$. Now

$$\nu_L \leq C \left(\frac{E[(Y_{L,l} - \mu_L)^2]}{(\alpha_L - \mu_L)^2} \right)^L$$

$$\leq C \left(\frac{N_0 + N_J}{(\alpha_L - \mu_L)^2} \right)^L$$

where the last step follows from Appendix I. Since $\alpha_L \to \infty$ and μ_L is bounded, we have

$$\lim_{L \to \infty} \nu_L \leq 0.$$

It follows that

$$\lim_{L \to \infty} \nu_L = 0.$$

Next we note that $F'_{D_j}(x)$ is the density of the absolutely continuous component of the normalized sum of L i.i.d. random variables, $\frac{Y_{L,l}}{\sigma_L^j}$, each of which has variance 1 and mean $\mu_L = \frac{\sqrt{E}}{\sigma_L^j} + \epsilon_L$ where $\epsilon_L \to 0$ and $\sigma_L^j \to \frac{N_0}{2} + \frac{N_J}{2}$ (see Appendix J). Since $\overline{\phi}_j(x)$ has been chosen to be the density of a normal random variable with the same mean and variance i.e. of a $N\left(\frac{\sqrt{L(E + \epsilon_L)}}{\beta_L \left(\frac{N_0}{2} + \frac{N_J}{2} \right)}, 1 \right)$ random variable where $\epsilon_L \to 0$, $\beta_L \to 1$ we have that the first term of Δ_2 is

$$\leq \int_{-\infty}^{\infty} \left| \overline{\phi}_j(x) - F'_{D_j}(x) \right| dx$$

Using a transformation of variables $p = x - \mu_L$ we have that Δ_2 is

$$= \int_{-\infty}^{\infty} \left| \overline{\phi}_j(p) - F'_{D_j}(p) \right| dp$$

where now $\phi_j(p)$ is the density of a standard normal random variable and $\hat{F}'_{D_j}(p)$ is the density of the absolutely continuous part of the sum of L i.i.d. random variables each of which has variance 1, mean 0 and bounded absolute third moment. Hence by the result in Appendix H

$$\Delta_2 \leq \int_{-\infty}^{\infty} \frac{\alpha_L^3}{6\sqrt{L}} (x^3 - 3x)\, \phi(x)dx + \frac{C_1}{\sqrt{L}}$$

$$\leq \frac{1}{3\sqrt{2\pi}} (1 - 4e^{-\frac{3}{2}}) \frac{\alpha_L^3}{\sqrt{L}} + \frac{C_1}{\sqrt{L}}$$

$$\leq \frac{C_2\, \alpha_L^3}{\sqrt{L}}$$

Thus if M, L and α_L are chosen so that $\frac{M\, \alpha_L^3}{\sqrt{L}} \rightarrow 0$ we have that

$$\lim_{M\rightarrow\infty} P_C^a = \lim_{M\rightarrow\infty} P_c^L(\rho_L, E'_b, N_J, N_0, M)$$

Thus we can conclude that asymptotically our clipped linear combining receiver exhibits the same threshold behavior demonstrated for $P_{c,a}$ in Lemma 1. Thus

i) If $\frac{E_b}{N_0 + N_J} < \ln 2$ $P_c^L(\rho_L, E'_b, N_J, N_0, M) \rightarrow 0$

ii) If $\frac{E_b}{N_0 + N_J} > \ln 2$ $P_c^L(\rho_L, E'_b, N_J, N_0, M) \rightarrow 1$

As we have imposed no restrictions on ρ_L we can say that

i) for $\frac{E_b}{N_0 + N_J} < \ln 2$

$$\lim_{M,L\rightarrow\infty} \sup_{0\leq\rho\leq 1} P_e^L(\rho_L, E'_b, N_J, N_0, M) = 1$$

ii) and for $\frac{E_b}{N_0 + N_J} > \ln 2$

$$\lim_{M,L\rightarrow\infty} \sup_{0\leq\rho\leq 1} P_e^L(\rho_L, E'_b, N_J, N_0, M) = 0$$

3.5 Conclusions

We have investigated the asymptotic performance of orthogonal signals over channels with both thermal noise as well as unknown partial-band interference. Knowing that such signalling suffices asymptotically to communicate over the AWGN at the limits prescribed by the channel capacity theorem we tried to recover from the effects of the worst-case unknown partial-band interference on the performance of such signalling. The worst-case partial-band jammer does degrade the asymptotic performance severely but he needs to optimize his strategy for each value of the bit energy to noise ratio ($E_b/(N_J + N_0)$) chosen by the communicator. Our analysis reveals that for bit energy to noise ratios above a constant ($\ln 2$) the jammer can be most effective if he jams only a fraction (ρ) of the band. This is because the probability of error near the values of $E_b/(N_J + N_0)$ around $\ln 2$ rises dramatically with a small decrease in $E_b/(N_J + N_0)$. The fraction ρ jammed gets smaller as $E_b/(N_J + N_0)$ gets larger (observe however that it does not get too small). This indicates that the jammer is wilfully reducing the probability of affecting a transmission in order that he may cause more serious damage when he does affect a transmission. This observation suggests that simple coding such as diversity in such a case may be very effective.

Diversity over the partial-band interference channel was next investigated using, at first, majority logic decoding. In this scheme, since the jammer is willing to accept a small probability of affecting transmissions, we expect that the majority of the received diversity transmissions would be received error free. However, the worst-case jammer optimizes his ρ with respect to the diversity level and he is able to ensure that even for very large $E_b/(N_J + N_0)$ the asymptotic symbol error probability is 1. Reliable communication in this case is possible only if $E_b/(N_J + N_0)$ increases with diversity level faster than a certain rate.

Our next diversity scheme was linear combining wherein we simply added the outputs of each diversity transmission. The hope is that the few diversity transmissions that are jammed are nullified in the sum statistic by the many good receptions. In this case the jammer can, by an appropriate choice of ρ_L ensure that for any $E_b/(N_J + N_0)$ the asymptotic error probability is non-zero. The jammer's choice of ρ_L is inversely proportional to the diversity level and thus for large L the jammer is jamming a very small fraction but with a large power. The effect hoped for i.e. the swamping out of the few bad receptions by the many good ones does take place but only if the jammer's choice of ρ_L goes to zero much faster.

All this suggests that in our diversity combining we must find a way of limiting the contribution of any individual diversity transmission to the overall decision statistic. We therefore proceed with clipped linear combining wherein we first clip the output of each diversity transmission and then simply add. The output statistics are thus the sums of many i.i.d. random variables which suggests the use of some form of Central Limit Theorem Approximations. To ensure the threshold behaviour that we are looking for in the error probability we need to use powerful versions of the Central Limit Theorem for which we need the clipping level, the diversity level and the number of signals to satisfy a certain relation. Doing so we can show that the worst-case jammer can be neutralized asymptotically, i.e. he is seen to be no more detrimental to reliable communication than the AWGN channel of equivalent noise spectral density.

CHAPTER IV

CONCLUSIONS

In Chapter 2 we have constructed fairly general channel models which are capable of representing a number of jamming situations. All our analysis was done in a game-theoretic framework. We view the entire transmitted sequence as one play of a zero-sum two person game. In the case with no side information (Case I) we have characterized the worst-case jammer strategy by the number of points of support of the worst-case distribution as well as by necessary and sufficient conditions at these points. This allows us to formulate the search for the worst case jammer strategy as a finite dimensional nonlinear programming problem. Although the necessary and sufficient conditions are not easy to solve for, they are fairly easy to test. Given the convexities of the objective functions this suggests that it would be possible to develop efficient steepest ascent (or descent) computational algorithms for these optimization problems. Much the same held true for both our performance measures, mutual information and channel cutoff rate.

In the cases with the decoder informed we reduce the communicator's strategy set (either by using the "compatibility" assumption or by fixing a quantizer). In this case when we have convexity with respect to the jammer's strategy (as in cases AII and BII) we are able to demonstrate the existence of a saddle-point strategy. For the case with non-randomized quantization we are further able to characterize these saddle-point strategies using the earlier theory. Part of the reason we get

saddle point strategies in Case II and not in Case I is that the randomization of the quantizers in Case I does not average out the objective function. In Case II where the decoder is informed of the actual quantizer chosen the objective function does get averaged out and thus we do actually use randomized strategies in this case.

Although our analysis was done mainly for non-adaptive jammers we find that a number of our results hold true for the case of adaptive jamming as well. An appropriate model to use in this case was seen to be the arbitrary "star" varying channel. Despite the increase in the jammer's strategy set we find that in a number of cases he is able to cause no more loss than if he were non-adaptive. Other generalizations of this game are possible, i.e.if we allow the communicator to change his strategy after every transmission based on feedback from the previous transmission or based on observing the jamming noise. These would be sequential games with exchange of information and it is not clear what kind of objective function would have operational significance in this case. While we have an upper bound on the number of points of support of the worst case distributions it is possible that the number actually needed is less. Also it is possible that the performance is robust with regard to the number of jamming levels and that a jammer with a few levels is able to do fairly well. Such questions can best be answered by numerical analysis.

In Chapter 3 we have investigated the asymptotic performance of orthogonal signals over channels with both thermal noise as well as unknown partial band interference. The worst-case partial-band jammer does degrade the asymptotic performance severely but he needs to optimize his strategy for each value of the bit energy to noise ratio $E_s/(N_J + N_0)$ chosen by the communicator. Our analysis reveals that for bit energy to noise ratios above a constant [illegible] he will be most effective if he jams only a fraction [illegible] of the band. For [illegible] $E_s/(N_J + N_0)$ the worst case jamming is no more effective than [illegible] equivalent noise spectral density. The maximum [illegible]

band jamming: getting the communication system to perform around the sharp rise in the probability of error curve. To do this however, the jammer must allow a high probability of not jamming a particular transmission. This suggests coding by way of diversity as a means of overcoming the jammer.

The worst-case jammer tries to counteract such coding by making a few transmissions affect the resultant decision statistic significantly. Both majority logic decoding and linear combining do not perform well against such a jammer. However, by limiting the effect of a single transmission on the decision statistic, clipped linear combining is able to asymptotically neutralize partial-band jamming.

Our analysis in Chapter 3 was entirely asymptotic. Other interesting questions that could be asked are; how should diversity be chosen as a function of the number of signals to achieve a given probability of error against the worst- case jammer ? If the jammer has a peak power constraint then what values of diversity and signal set size will achieve a given error probability against the worst-case jammer ?

Although we did our analysis using coherent detection our results are valid for noncoherent detection in all cases except for the clipped linear combining case. This is because the properties we use of the probability of error of M orthogonal signals with coherent detection remain valid even in the noncoherent case. These properties are the monotone decreasing nature of the probability of error with the bit energy to noise ratio and the asymptotic threshold behaviour of the probability of error. However for the clipped linear combining case, the Gaussian approximations we used in the coherent case, do not necessarily work. It would be interesting to determine if there is a diversity combining scheme in the noncoherent case which would neutralize the partial-band jammer.

APPENDICES

APPENDIX A

Consider, the following metric on the space of D-dimensional distributions on K.

$$\begin{aligned} d(F,G) &= \inf\{h : F(x_1-h, x_2-h, \ldots, x_D-h) - h \le G(x_1,\ldots,x_D) \\ &\le F(x_1+h,\ldots,x_D+h)+h \qquad \text{forall}(x_1,\ldots,x_D)\}. \end{aligned}$$

We check that $d(F,G)$ satisfies the properties of a metric:

1. $d(F,G) \ge 0$ and $= 0$ iff $F = G$.
2. $d(F,G) = d(G,F)$.
3. $d(F,H) \le d(F,G) + d(G,H)$.

1. Clearly, $d(F,G) \ge 0$. If $d(F,G) = 0$ we consider a sequence $h_n \downarrow 0$ and from the right-continuity of distribution functions and the definition of d we get

   $G(x_1,\ldots,x_D) \le F(x_1,\ldots,x_D)$.

   Similarly

   $F(x_1,\ldots,x_D) \le G(x_1,\ldots,x_D)$

   $\therefore F = G$

2. Let $d(F,G) = d$.

Then for all $h \ge d$ and $\forall(x_1,\ldots,x_D)$

$F(x_1-h,\ldots,x_D-h) - h \le G(x_1,\ldots,x_D)$.

$\therefore F(x_1,\ldots,x_D) \leq G(x_1+h,\ldots,x_D+h)+h.$

Similarly $F(x_1,\ldots,x_D) \geq G(x_1-h,\ldots,x_D-h)-h.$

$\therefore d(F,G) = d(G,F)$.

3) Let $d(F,G)=d_1, d(G,H)=d_2, d(F,H)=d_3.$

Then for $h_1 \geq d_1, h_2 \geq d_2$ and $\forall(x_1,\ldots,x_D)$

$F(x_1-h_1,\ldots,x_D-h_1)-h_1 \leq G(x_1,\ldots,x_D) \leq F(x_1+h_1,\ldots,x_D+h_1)+h_1$

and

$G(x_1-h_2,\ldots,x_D-h_2)-h_2 \leq H(x_1,\ldots,x_D) \leq G(x_1+h_2,\ldots,x_D+h_2)+h_2$

$F(x_1-h_1-h_2,\ldots,x_D-h_1-h_2)-h_1-h_2 \leq H(x_1,\ldots,x_D)$

$\leq F(x_1+h_1+h_2,\ldots,x_D+h_1+h_2)+h_1+h_2$.

$\therefore d_3 \leq d_1+d_2.$

$\therefore d(F,G)$ is a metric.

A sequence of distribution functions F_n on $\mathbf{R}^D$ is said to converge weakly to F iff for any bounded continuous function $f(\underline{x})$ defined on $\mathbf{R}^D$ (where $\underline{x}$ is $(x_1,\ldots,x_D)$

$$\int_{\mathbf{R}^D} f(\underline{x})dF_n(\underline{x}) \longrightarrow \int_{\mathbf{R}^D} f(\underline{x})dF.(\underline{x})$$

APPENDIX B

Lemma: With F , F_n denoting the distribution functions of random variables $\underline{X}(= (X_1, X_2, \ldots, X_D))$ such that $t_i \leq X_i \leq u_i$ the following are equivalent:

1. $F_n \longrightarrow F$ at every point $\underline{x}$ which is a continuity point of the distribution $F(\underline{x})$.

2. $d(F_n, F) \longrightarrow 0$.

3. $F_n \xrightarrow{w} F$.

Proof: The proof is accomplished by showing the following: $1 \Rightarrow 2, 1 \Rightarrow 3, 2 \Rightarrow 1, \ 3 \Rightarrow 1$.

i) $1 \Rightarrow 2$. Let C denote the set of continuity points of $F(\underline{x})$. Clearly C is dense in $\mathbf{R}^D$. Choose $a_1, a_2, \ldots, a_D$ $(\in \ C)$ such that $a_i \leq t_i, \quad i = 1, \ldots, D$ and $b_1, b_2, \ldots, b_D (\in \ C)$ such that $b_i = u_i + 1$. Subdivide each $[a_i, b_i]$ by points $a_i = a_{i0} < a_{i,1} < a_{i,2} < \ldots < a_{i,s} = b_i \quad a_{i,k} \in C$ such that $a_{i,k} - a_{i,k-1} < \epsilon$. Let $E = \{(x_1, \ldots, x_m) : a_i \leq x_i \leq b_i\}$. Clearly $Pr(E) = 1$.

Let L denote the lattice of points with i^{th} coordinate equal to a_{ik} $0 \leq k \leq s$. L has $(s+1)^D$ points. Denote a generic point of L by l. Given any $\epsilon \geq 0$ choose N large enough so that for $n \geq N$ at all points $l \in L$ the following inequality is satisfied:

$$\mid F_n(l) - F(l) \mid \leq \frac{\epsilon}{2}.$$

Now we prove that for every $\underline{x}$ and for $n \geq N$

$$F(x_1 - \epsilon, \ldots, x_D - \epsilon) - \epsilon \leq F_n(x_1, \ldots x_D) \leq F(x_1 + \epsilon, \ldots x_D + \epsilon) + \epsilon$$

a) Consider $\underline{x} \in E$.

Then $\underline{x}$ lies in one of the lattice cells. Let $\bar{l}$ denote the closest lattice point such that $l \geq \underline{x}$ and let $\underline{l}$ denote the closest lattice point such that $l \leq \underline{x}$. Clearly $\bar{l} \leq (x_1 + \epsilon, \ldots, x_D + \epsilon)$ and $\underline{l} \geq (x_1 - \epsilon, \ldots, x_D - \epsilon)$.

Now

$$F_n(\underline{x}) \leq F_n(\bar{l}) \leq F(\bar{l}) + \frac{\epsilon}{2} \leq F(x_1 + \epsilon_1 \ldots, x_D + \epsilon) + \frac{\epsilon}{2}.e$$

Also

$$F_n(\underline{x}) \geq F_n(\underline{l}) \geq F(\underline{l}) - \frac{\epsilon}{2} \geq F(x_1 - \epsilon, \ldots, x_D - \epsilon) - \frac{\epsilon}{2}. \qquad \text{(B.2)}$$

Hence

$$F(x_1 - \epsilon, \ldots, x_D - \epsilon) - \frac{\epsilon}{2} \leq F_n(x) \leq F(x_1 + \epsilon, \ldots, x_D + \epsilon) + \frac{\epsilon}{2}.$$

$$\therefore F(x_1 - \epsilon, \ldots, x_D - \epsilon) - \epsilon \leq F_n(x) \leq F(x_1 + \epsilon, \ldots, x_D + \epsilon) + \epsilon.$$

b) Consider now $\underline{x} \notin E$.

We examine the two cases.

i) $\underline{x} \ni$ either $x_i \geq a_i$ for all i or $x_i \leq b_i$ for all i. Call the set of all such $\underline{x}$, W.

ii) $\underline{x} \notin W$. By our selection of E for such $\underline{x}$, $F(\underline{x}) = 0$.

Case i) When $\underline{x} \in W$ and $x_i \leq b_i$ for all i then by our selection of E, $F(\underline{x}) = 0$, $F_n(\underline{x}) = 0$. Hence

$$F_n(\underline{x}) \leq F(\underline{x}) + \epsilon$$

$$F_n(\underline{x}) \geq F(\underline{x}) - \epsilon$$

$$\therefore F(x_1 - \epsilon, \ldots, x_D - \epsilon) - \epsilon \leq F_n(\underline{x}) \leq F(x_1 + \epsilon, \ldots, x_D + \epsilon) + \epsilon.$$

When $\underline{x} \in W$ and $x_i \geq a_i$ for all i define the following sets:

$$E^1 \triangleq \{\underline{x}: a_i \le x_i \le b_i, \ i = 1, \ldots, D\}$$

$$E^2 \triangleq E - E^1.$$

Define l^* as follows: consider all the components of $\underline{x}$, $x_{i_1}, \ldots x_{i_q}$ which are greater than $b_{i_1}, \ldots, b_{i_q}$ respectively. Call the set of such indices Q. Keeping the components with indices $\notin Q$ as before, reduce the components with indices in Q to $x^*_{i_1}, \ldots x^*_{i_q}$ so that $b_{i_k} < x^*_{i_k} < u_{i_k}$, $b_{i_k} < x^*_{i_k} - \epsilon < u_{i_k}$ and and $b_{i_k} < x^*_{i_k} + \epsilon < u_{i_k}$ with $i_k \in Q$. This vector we call l^* (it is clear that such a l^* can always be found). By construction $F(\underline{x}) = F(\underline{l}^*)$, $F_n(\underline{x}) = F_n(\underline{l}^*)$, $F(x_1 - \epsilon, \ldots, x_D - \epsilon) = F(l^*_1 - \epsilon, \ldots, l^*_D - \epsilon)$ and $F(x_1 + \epsilon, \ldots, x_D + \epsilon) = F(l^*_1 + \epsilon, \ldots, l^*_D + \epsilon)$. Also, $\underline{l}^*$ clearly belongs to E. Therefore this case is reduced to the case when $\underline{x} \in E$. Hence by the argument in part(a)

$$F(x_1 - \epsilon, \ldots, x_D - \epsilon) - \epsilon \le F_n(x) \le F(x_1 + \epsilon, \ldots, x_D + \epsilon) + \epsilon.$$

Case ii) When $\underline{x} \ni W$ let l denote the lattice point closest to $\underline{x}$.

Then

$$F_n(\underline{x}) = 0, F(\underline{x}) = 0, F(l) = 0.$$

$$F_n(\underline{x}) \le F(\underline{x}) + \epsilon$$

$$F_n(\underline{x}) \ge F(\underline{x}) \ge \epsilon.$$

Hence

$$F(x_1 - \epsilon, \ldots, x_D - \epsilon) - \epsilon \le F_n(\underline{x}) \le F(x_1 + \epsilon, \ldots, x_D + \epsilon) + \epsilon.$$

Thus $1 \Rightarrow 2$.

ii) $1 \Rightarrow 3$.

Take any function $f(\underline{x})$ bounded and continuous on E. Since E is compact f is uniformly continuous on E. Denote by U an upper bound of $|f(\underline{x})|$ and choose points

$$a_i = a_{i,0} < a_{i,1} < \quad < a_{i,s} = b_i \qquad \begin{matrix} a_{i,k} \in C \\ i = 1, \ldots D \end{matrix}$$

so that we have a lattice L on E such that $|f(l_1) - f(l_2)| < \epsilon$ where l_1 and l_2 are points belonging to the same lattice cell. Construct the function f_ϵ which is constant on the lattice cells as follows:

$$f_\epsilon(\underline{x}) = f(\underline{y}^j) \quad \underline{x} \in E \text{ and } \underline{y}^j \text{ is some interior}$$

point of the lattice cell j to which $\underline{x}$ belongs

$$= 0 \qquad x \notin E$$

Obviously for any distribution function $G(\underline{x})$

$$\int f_\epsilon(\underline{x})dG(\underline{x}) = \sum_{j=1}^{|L|} f(\underline{y}^j)\Delta_{a^j_{1,k_1},a^j_{1,k_1+1}} \quad \ldots \Delta_{a^j_{D,k_D},a^j_{D,k_D+1}} G(x_1,\ldots,x_D)$$

where a^j_{l,k_l}, a^j_{l,k_l+1} are the l-coordinates of $\underline{y}^j$ and where

$$\Delta_{b_1,a_1},\ldots,\Delta_{b_D,a_D}F(x_1,\ldots x_D) = F_0 - F_1 + F_2,\ldots + (-1)^D F_D$$

F_1 is the sum of all $\binom{D}{1}$ terms of the form $F(c_1,\ldots,c_D)$ with $c_k = a_k$ for exactly i integers in $\{1,\ldots,D\}$ and $c_k = b_k$ for the remaining $D-i$ integers.

Since $F_n(\underline{x}) \to F(\underline{x})$ at the lattice points

$$\int f_\epsilon(\underline{x})dF_n(x) \to \int f_\epsilon(\underline{x})dF(\underline{x}).$$

Also

$$\begin{aligned}\int |f(\underline{x}) \;-\; f_\epsilon(\underline{x})|dF(\underline{x}) &= \int_{E^C} |f(\underline{x}) - f_\epsilon(\underline{x})|dF(\underline{x}) + \int_E |f(\underline{x}) - f_\epsilon(\underline{x})\;|dF(\underline{x}) \\ &\leq\; 0 + \int_E \in dF(\underline{x}) = \epsilon.\end{aligned}$$

Similarly

$$\int |f(\underline{x}) - f_\epsilon(\underline{x})|dF_n(\underline{x}) \leq \epsilon$$

Hence

$$|\int f(\underline{x})dF_n(\underline{x}) - \int f(\underline{x})dF(\underline{x})| \;\leq 3\,\epsilon$$

for sufficiently large n. Since ϵ was arbitrary we have $1 \Rightarrow 3$

iii) $2 \Rightarrow 1$

Let $\underline{x}_0$ be a continuity point of $F(\underline{x})$. Then for every $\epsilon > 0$ there exists $\delta > 0$ such that

$$| F(\underline{x}) - F(\underline{x}_0) | < \epsilon$$

if $|\underline{x} - \underline{x}_0| \leq \delta$ where $||\underline{x}|| = \sqrt{x_1^2 + \ldots + x_D^2}$. Let $h = \min(\epsilon, \sqrt{\frac{\delta}{D}})$ and let n be sufficiently large so that $d(F_n, F) < h$. Then

$$\begin{aligned} F_n(\underline{x}_0) &\geq F(x_{0,1} - h, \ldots, x_{0,D} - h) - h \geq F(\underline{x}_0) - 2\epsilon \\ F_n(\underline{x}_0) &\leq F(x_{0,1} + h, \ldots, x_{0,D} + h) + h \leq F(\underline{x}_0) + 2\epsilon. \end{aligned}$$

Since ϵ is arbitrary $2 \Rightarrow 1$.

iv) $3 \Rightarrow 1$.

Let $\underline{x}_0$ be a continuity point of $F(\underline{x})$ and let $F_n \xrightarrow{w} F$. Take $\delta > 0$ such that for $||\underline{x} - \underline{x}_0|| < \sqrt{D}\delta \quad |F(\underline{x}) - F(\underline{x}_0)| < \epsilon$. Define the following sets:

$$J \triangleq \{\underline{x} : \ x_i \leq x_{0,i}, \ \ i = 1, \ldots, D\}.$$

$$J_\delta \triangleq \{\underline{x} : \ x_i \leq x_{0,i} - \delta, \ \ i = 1, \ldots, D\}.$$

$$J^\delta \triangleq \{\underline{x} : \ x_i \leq x_{0,i} + \delta, \ \ i = 1, \ldots, D\}.$$

$$J_1 \triangleq J - J_\delta.$$

$$J_2 \triangleq J^\delta - J.$$

Construct the functions:

$$\begin{aligned} f_1(\underline{x}) &= 1 \qquad \underline{x} \in J_\delta \\ &= \frac{1}{D\delta} \sum_1^D x_{0,i} - \delta - \max(x_{0,i} - \delta, x_i) \qquad x \in J_1 \\ &= 0 \qquad \text{elsewhere.} \end{aligned}$$

$$\begin{aligned} f_2(\underline{x}) &= 1 \qquad \underline{x} \in J \\ &= \frac{1}{D\delta}\sum_{1}^{D} x_{0,i} - \max(x_{0,i}, x_i) \qquad x \in J_2 \\ &= 0 \qquad \text{elsewhere.} \end{aligned}$$

$f_1(x)$ and $f_2(x)$ are both continuous functions ranging between 0 and 1. On J_1, $f_1(x) \leq 1$ and on J_2, $f_2(x) \leq 1$.

Then

$$\int f_1(\underline{x})dF(\underline{x}) \geq \int_{J_\delta} 1dF(\underline{x}) = F(x_{0,1} - \delta, \ldots, x_{0,D} - \delta) \geq F(\underline{x}_0) - \epsilon \qquad \text{(B.3)}$$

$$\int f_2(\underline{x})dF(\underline{x}) \leq \int_{J^\delta} 1dF(\underline{x}) = F(x_{0,1} + \delta, \ldots, x_{0,D} + \delta) \leq F(\underline{x}_0) + \epsilon \qquad \text{(B.4)}$$

$$\int f_1(\underline{x})dF_n(\underline{x}) \leq \int_J 1dF_n(\underline{x}) = F_n(\underline{x}_0) \qquad \text{(B.5)}$$

$$\int f_2(\underline{x})dF_n(\underline{x}) \geq \int_J 1\ dF_n(\underline{x}) = F_n(\underline{x}_0). \qquad \text{(B.6)}$$

From the fact that $F_n \xrightarrow{w} F$ for sufficiently large n

$$|\int f_1(\underline{x})dF_n(\underline{x}) - \int f_1(\underline{x})\ dF(\underline{x})| < \epsilon \qquad \text{(B.7)}$$

$$|\int f_2(\underline{x})dF_n(\underline{x}) - \int f_2(\underline{x})\ dF(\underline{x})| < \epsilon. \qquad \text{(B.8)}$$

From (B.3),(B.4),(B.5),(B.6), (B.7) and (B.8)

$$F(\underline{x}_0) - 2\epsilon \leq F_n(\underline{x}_0) \leq F(\underline{x}_0) + 2\epsilon$$

Since ϵ was arbitrary $3 \Rightarrow 1$. This completes the proof of the proposition

APPENDIX C

Lemma: The set **S** of distribution functions of random variables $\underline{x} = (x_1, \ldots, x_D)$ such that $0 \leq x_i \leq b_i$ is compact in the space of distribution functions on $\underline{x}$.

Proof: Let a sequence of distribution functions $F_n(\underline{x})$ be given. Pick a countable set C everywhere dense on the set $\mathbf{R}^D$, $C = \{\underline{x}_1, \ldots, \underline{x}_s, \ldots\})$. By Helly's Weak Compactness Theorem [Loeve 77, pg. 181] there exists a subsequence $F_{n_1}(\underline{x}), \ldots, F_{nk}(\underline{x}), \ldots$ which converges at every point $\underline{x} = \underline{x}_s$. Let

$$v(\underline{x}_s) = \lim_{k \to \infty} F_{n_k}(\underline{x}_s)$$

and set

$$F(x) = \sup_{\underline{x}_s \leq \underline{x}} v(\underline{x}_s).$$

The function $F(\underline{x})$ is defined everywhere on $\mathbf{R}^D$ and is obviously non-decreasing and right-continuous. Clearly $F(\underline{x}) = 1$ for $x_i \geq b_i$ and $F(x_1, \ldots, c, \ldots, x_D) = 0$ if $c < 0$. Thus $F(\underline{x})$ is a distribution function on $\mathbf{R}^D$ and $F \in \mathbf{S}$.

Also, it is easy to see that $F_{n_k}(\underline{x})$ converges to $F(\underline{x})$ at every continuity point of $F(\underline{x})$. This is equivalent to weak convergence of $F_{n_k}$to$F$ [Ash 72, Th 4.5.1] which by Lemma 1 is equivalent to convergence in the Levy metric. Hence $d(F_{n_k}, F) \to 0$.

Hence any sequence of points belonging to **S** has a convergent subsequence in **S**. Since the space of all distribution functions is a metric space (with the Levy metric) **S** is compact.

APPENDIX D

Lemma 3 : $I'_{F_1}(G;F_2) = \int i(z;G,F_1)dF_2(z) - I(G;F_1)$

where $i(z;G,F_1) = \sum_{x,y} p(x)p(y|x,z)\log\left(\frac{\int p(y|x,z)dF_1}{\sum_x p(x)\int p(y|x,z)dF_1}\right)$.

Proof of Lemma 3 :

$$I'_{F_1}(G;F_2) = \lim_{\alpha \downarrow 0}\frac{1}{\alpha}\left\{\sum_{x,y} p(x)\left(\int\int p(y|x,z,\theta)[(1-\alpha)dF_1 + \alpha dF_2]dG(\theta)\right)\right.\cdot$$

$$\log\frac{(\int\int p(y|x,z,\theta)[(1-\alpha)dF_1+\alpha dF_2]dG(\theta))}{\sum_x p(x)(\int\int p(y|x,z,\theta)[(1-\alpha)dF_1+\alpha dF_2]dG(\theta)}$$

$$-\sum_{x,y} p(x)(\int\int p(y|x,z,\theta)dF_1 dG(\theta)).$$

$$\left.\log\frac{(\int\int p(y|x,z,\theta)dF_1dG(\theta))}{\sum_x p(x)(\int\int p(y|x,z,\theta)dF_1dG(\theta)}\right\}. \qquad \text{(D.1)}$$

Denoting $\int p(y|x,z,\theta)dG(\theta)$ as $p(y|x,z)$

$$I'_{F_1}(G;F_2) = \lim_{\alpha\downarrow 0}\frac{1}{\alpha}\left\{\sum_{x,y} p(x)\left[\int p(y|x,z)[(1-\alpha)dF_1 + \alpha dF_2\right]\right.$$

$$\log\frac{\int p(y|x,z)[(1-\alpha)dF_1+\alpha\, dF_2]}{\sum_x p(x)\int p(y|x,z)[(1-\alpha)dF_1+\ \alpha dF_2]}$$

$$\left.-\int p(y|x,z)dF_1\log\frac{\int p(y|x,z)dF_1}{\sum_x p(x)\int p(y|x,z)dF_1}\right\}$$

$$= \lim_{\alpha\downarrow 0}\frac{1}{\alpha}\left\{\sum_{x,y} p(x)\left[\int p(y|x,z)\alpha dF_2\log\left(\frac{\int p(y|x,z)[(1-\alpha)dF_1+\alpha dF_2]}{\sum_x p(x)\int p(y|x,z)[(1-\alpha)dF_1+\alpha\, dF_2]}\right)\right.\right.$$

$$- \int p(y|x,z)\alpha dF_1 \log\left(\frac{\int p(y|x,z)[(1-\alpha)dF_1+\alpha dF_2]}{\sum_x p(x)\int p(y|x,z)[(1-\alpha)dF_1+\alpha dF_2]}\right)\Bigg]\Bigg\}$$

$$+\lim_{\alpha\downarrow 0}\frac{1}{\alpha}\sum_{x,y}p(x)\left[\int p(y|x,z)dF_1\log\left(\frac{\int p(y|x,z)[(1-\alpha)dF_1+\alpha dF_2]}{\sum_x p(x)\int p(y|x,z)[(1-\alpha)dF_1+\alpha dF_2]}\right)\right.$$

$$\left.- \int p(y|x,z)dF_1\log\left(\frac{\int p(y|x,z)dF_1}{\sum_x p(x)\int p(y|x,z)dF_1}\right)\right]$$

$= a + b(\text{say}).$

By choosing a sequence $\alpha_n \downarrow 0$ and using weak convergence of $(1-\alpha_n)dF_1 + \alpha_n dF_2$ to dF_1

$$a = \int i(z;G,F_1)dF_2 - I(G;F_1)$$

$$b = \frac{d}{d\alpha}\left[\sum_{x,y}p(x)\int p(y|x,z)dF_1\log\left(\frac{\int p(y|x,z)[(1-\alpha)dF_1+\alpha dF_2]}{\sum_x p(x)\int p(y|x,z)[(1-\alpha)dF_1+\alpha dF_2}\right)\right]_{\alpha=0}.$$

Taking the derivative

$$b = \sum_{x',y}p(x')\int p(y|x',z)dF_1\left\{\frac{\sum_x p(x)\int p(y|x',z)[(1-\alpha)dF_1+\alpha dF_2]}{\int p(y|x,z)[(1-\alpha)dF_1+\alpha dF_2]}\right.$$

$$\cdot\frac{1}{d^2}\left[\left(\sum_x p(x)\int p(y|x,z)[(1-\alpha)dF_1+\alpha dF_2]\right)\int p(y|x',z)(dF_2-dF_1)\right.$$

$$\left.\left.-\int p(y|x',z)[(1-\alpha)dF_1+\alpha dF_2]\left(\sum_x p(x)\int p(y|x,z)(dF_2-dF_1)\right)\right]\right\}$$

where $d \triangleq \sum_x p(x)\int p(y|x,z)[(1-\alpha)dF_1+\alpha dF_2$.

After some algebraic manipulation it can be shown that $b \to 0$ as $\alpha \downarrow 0$.

APPENDIX E

Here we consider a communication game [illegible] an input distribution r on the M [illegible] the $M \times L$ transition probability matrix [illegible] output random variables respectively and [illegible] random variable associated with the conditional [illegible] feasible $\underline{n}$'s $(= (n_1, \ldots, n_M)$ be compact. The [illegible] $(n_1, \ldots, n_M))$. Assume this function is linear and that for [illegible] $1, \ldots, M$ the channel chosen is symmetric. Let $I(r, \underline{n}) \triangleq I(\cdot, \cdot)$ [illegible] is r and B's choice is $\underline{n}$. Let $n_1, \ldots, n_M$ be constrained by [illegible] $i = 1, \ldots, c$ where f_i is a convex, symmetric function of $n_1, \ldots, n_M$ [illegible] invariant under any permutation of $n_1, \ldots, n_M$. Then a saddle point strategy exists for both players and for player A it is to chooose a uniform distribution on the input and for player B it is to choose all the components of $\underline{n}$ equal. [illegible] exists $\underline{n}^*$ with all its components equal such that

$$I(r, \underline{n}^*) \leq I(r^*, \underline{n}^*) \leq I(r^*, \underline{n})$$

where r^* corresponds to the uniform input distribution.

Proof: Step 1: $I(r, \underline{n}^*) \leq I(r^*, \underline{n}^*)$

This follows from the fact that the mutual information between the input and the output of a symmetric channnel is maximized by the uniform distribution.

Step 2: $I(r^*, \underline{n}^*) \leq I(r^*, \underline{n})$

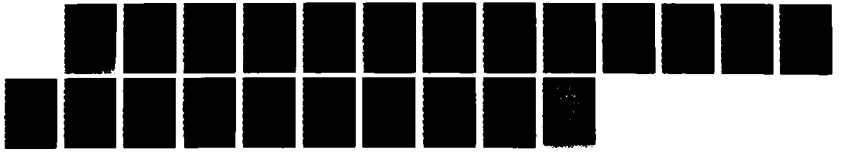
... which is linear in $\underline{n}$, $I(r, \underline{n})$ is convex ... the constraints the set of feasible $\underline{n}$'s is a convex ...

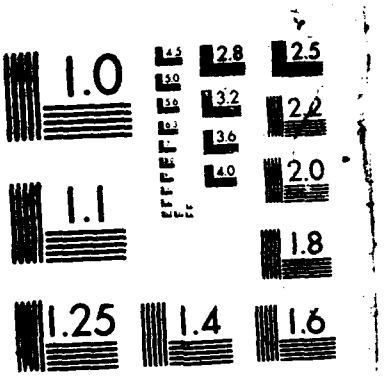
... $I(r, \underline{n}^*) = \ldots$ be achieved at some $\underline{n}_1 \neq \underline{n}^*$. Then we ... $\underline{n}_1$ ... that the minimum is also achieved at $\underline{n}^*$. The ... the input and the symmetry of the constraints ... permutation of $\underline{n}_1$, say, we have a new channel $p^*(y|x)$ ... a rearranging of the inputs of the original channel. The mutual ... $I(r, \underline{n}_1)$. Now consider all the $M!$ permutations ... permutations are not distinct but it does not matter). ... $\frac{1}{M!}\sum \underline{n}_1' = \underline{n}_2$, say. Every component of $\underline{n}_2$ is ... Now from the convexity of $I(r, \underline{n})$ w.r.t. $\underline{n}$ we know that

$$I(r, \underline{n}_2) = I\left(r, \frac{1}{M!}\sum \underline{n}_1'\right) \leq \frac{1}{M!}\sum I(r, \underline{n}_1')$$
$$= I(r, \underline{n}_1)$$

$$I(r, \underline{n}_2) = I(r, \underline{n}_1)$$

... $I(r, \underline{n})$ is achieved at $\underline{n}_2$ too. The result then follows from the ... that $I(r, \underline{n})$ is convex in r.





APPENDIX F

Lemma 4 : $D_{F_1}(T_0(G,P,F_2)) = E\left(\sum_y \left(\frac{\sqrt{\int p(y|x_1,z)dF_1}\int p(y|x_2,z)dF_2}{2\sqrt{p(y|x_2,z)dF_1}} + \right.\right.$

$$\left.\left.\frac{\sqrt{\int p(y|x_2,z)dF_1}\int p(y|x_1,z)dF_2}{2\sqrt{\int p(y|x_1,z)dF_1}}\right)\right) - T_0(G,F_1).$$

Proof of Lemma 4:

$$D_{F_1}(T_0(G,P,F_2)) = \lim_{\alpha \downarrow 0}\frac{1}{\alpha}\left[E\left(\sum_y \sqrt{\int p(y|x_1,z)(\alpha dF_2 + (1-\alpha)dF_1)}\right.\right.$$

$$\left.\sqrt{\int p(y|x_2,z)(\alpha dF_2 + (1-\alpha)dF_1)}\right) - E\left(\sum_y \sqrt{\int p(y|x_1,z)dF_1 \int p(y|x_2,z)dF_1}\right)\Bigg]$$

$$= \frac{d}{d\alpha}\left(E\left(\sum_y \sqrt{\int p(y|x_1,z)(\alpha dF_2 + (1-\alpha)dF_1)}\sqrt{\int p(y|x_2,z)(\alpha dF_2 + (1-\alpha)dF_1)}\right)\right)$$

at $\alpha = 0$.

Using the Dominated Convergence Theorem we have

$$E\left(\frac{d}{d\alpha}\left(\sum_y \cdots\right)\right)$$

$$= E\left(\sum_y \left(\frac{\sqrt{\int p(y|x_1,z)\ (\alpha dF_2 + (1-\alpha)dF_1)}\int p(y|x_2,z)(dF_2 - dF_1)}{2\sqrt{\int p(y|x_2,z)(\alpha dF_2 + (1-\alpha)dF_1)}}\right.\right.$$

$$\left.\left. + \frac{\sqrt{\int p(y|x_2,z)(\alpha dF_2 + (1-\alpha)dF_1)}\int p(y|x_1,z)(dF_2 - dF_1)}{2\sqrt{\int p(y|x_1,z)(\alpha dF_2 + \ (1-\alpha)dF_1)}}\right)\right)$$

at $\alpha = 0$

$$= E\left(\sum_y \left(\frac{\sqrt{\int p(y|x_1,z)dF_1}\int p(y|x_2,z)(dF_2-dF_1)}{2\sqrt{\int p(y|x_2,z)dF_1}}\right.\right.$$

$$\left.\left.+\frac{\sqrt{\int p(y|x_2,z)dF_1}\int p(y|x_1,z)(dF_2-dF_1)}{2\sqrt{\int p|y|x_1,z)dF_1}}\right)\right)$$

$$= E\left(\sum_y \left(\frac{\sqrt{\int p(y|x_1,z)dF_1}\int p(y|x_2,z)dF_2}{2\sqrt{p(y|x_2,z)dF_1}}\right.\right.$$

$$\left.\left.+\frac{\sqrt{\int p(y|x_2,z)dF_1}\int p(y|x_1,z)dF_2}{2\sqrt{\int\ p(y|x_1,z)dF_1}}\right)\right) - T_0(G,P,F_1).$$

APPENDIX G

$P_e\,(x, M)$ is a monotone decreasing function of x.

i) Coherent detection:

As $x \uparrow$, $Q\left(z + \sqrt{2x \log M}\right) \downarrow$

$$\therefore\ K = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{z^2}{2}} \left[1 - Q(z + \sqrt{2x \log M}\right]^{M-1} dz \uparrow$$

$$\therefore\ P_e\,(x, M) = 1 - K \downarrow$$

ii) Non-coherent detection:

As $x \uparrow$, $I_0\,(\sqrt{x}z)\, e^{-\frac{z^2}{2}} \downarrow$

$$\therefore\ P_e(x, M) = \int_0^{\infty} z I_0(xz) \exp\left(-\frac{z^2 + x^2}{2}\right) \left[1 - \prod_{\substack{i=1 \\ i \neq M}}^{M} (1 - \exp\left(-\frac{z^2}{2}\right)\right] dz \downarrow$$

APPENDIX H

We modify a result due to S. Kh. Sirazhdinov and M. Mamatov [Sira 62].

Let $X_1, \ldots, X_n$ be independent random variables with the common distribution $F(x)$. Let $E(X_i) = 0$, $E(X_i^2) = 1$. Let $F_n(x)$ be the distribution of $\frac{X_1 + \quad + X_n}{\sqrt{n}}$. It is well-known that $F_n(x)$ can be represented uniquely as a sum of the form

$$F_n(x) = \lambda_n C_n(x) + (1 - \lambda_n) S_n(x) \quad 0 \leq \lambda_n \leq 1 \tag{H.1}$$

where $C_n(x)$ represents the absolutely continuous part and $S_n(x)$ represents the singular and step-function parts of $F_n(x)$

Let $p_n(x)$ be the density of the absolutely continuous part.

Theorem 1: If $\exists\, n_0$ such that $\lambda_{n_0} > 0$ and if $\alpha_3 = E(X_i^3) < \infty$ then for sufficiently large n

$$\int_{-\infty}^{\infty} \mid p_n(x) - \phi(x) \mid dx = \int_{-\infty}^{\infty} \frac{\mid \alpha_3 \mid}{6\sqrt{n}} (x^3 - 3x)\, \phi(x) dx + \frac{C_1}{\sqrt{n}} \tag{H.2}$$

where

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \tag{H.3}$$

i.e. the density of the standard normal distribution. and C_1 is some universal constant

Proof of Theorem 1:

Let

$$g_n(t) = e^{-\frac{t^2}{2}} + \frac{\alpha_3}{6\sqrt{n}} (it)^3 e^{-\frac{t^2}{2}} \tag{H.4}$$

where $i = \sqrt{-1}$.

and let $f(t)$ and $f_n(t) = f^n(\frac{t}{\sqrt{n}})$ (where $f^n(.)$ indicates the n-fold convolution) be the characteristic functions of the distributions $F(x)$ and $F_n(x)$ respectively. We need the following Lemmas.

Lemma 1: If $\beta_3 = E \mid X_i \mid^3 < \infty$ and $\mid t \mid \leq \frac{\sqrt{n}}{24\beta_3}$ then for

i) $|f_n(t) - g_n(t)| \leq \frac{C_2}{\sqrt{n}} (\mid t \mid^3 + \mid t \mid^6) \exp\left(-\frac{t^2}{4}\right)$

ii) $|f_n'(t) - g_n'(t)| \leq \frac{C_3}{\sqrt{n}} (\mid t \mid^2 + \mid t \mid^7) \exp\left(-\frac{t^2}{4}\right)$

Proof of Lemma 1:

i) is proved in [Gned 54].

ii) is proved in [Esse 58].

Lemma 2: For n_0 as in the hypotheses of Theorem 1,

$$F_{n_0}(x) = p\, H_1(x) + q\, H_2(x) \tag{H.5}$$

where $p > 0$, $q \geq 0$, $p + q = 1$ and $H_1(.)$ and $H_2(.)$ are distribution functions such that if $h_1(t)$ is the characteristic function corresponding to $H_1(x)$ then:

1) $\int_{-\infty}^{\infty} \mid h_1(t) \mid^2 dt < \infty$

2) Given any $\epsilon > 0$ there exists a $C_4 > 0$ such that the inequality $\mid h_1(t) \mid \leq \exp(-C_4)$ holds for $\mid t \mid \geq \epsilon$.

Proof of Lemma 2:

See [Prok 52].

Lemma 3: If $p > 0$, $q \geq 0$, $p + q = 1$ then for sufficiently large n,

$$\sum_{m - np < -\sqrt{n} \log n} \binom{n}{m} p^m q^n \leq \frac{C_5}{n^2} \tag{H.6}$$

Proof of Lemma 3:

Let W be a Binomial (n,p) random variable. Then L.H.S. of (H.6) is $Pr(W < np - \sqrt{n}\log n)$. Hence (H.6) is true iff $Pr(W \geq np - \sqrt{n}\log n) \leq 1 - \frac{C_5}{n^2}$

Now from the generalized Cheybyshev inequality we know that

$$Pr(w \geq np - \sqrt{n}\log n) \leq \frac{E\,(f(w))}{f\,(np - \sqrt{n}\log n)} \tag{H.7}$$

where f is any increasing, non-negative function on its range.

Choosing $f\,(W) = W^3$ we get

$$Pr\,(W \geq np - \sqrt{n}\log n) \neq \frac{E\,(W^3)}{(np - \sqrt{n}\log n)^3}$$

$$\leq \frac{n^3\,p^3 - (3p^3 - p^2)n^2 + (2p^3 - p^2 + p)^n}{n^3\,p^3 + 3n^2\,p^2\,\sqrt{n}\log n - 3n^2\,p\,(\log n)^2 - n\,\sqrt{n}\,(\log n)^3}$$

$$= 1 + \frac{n^2(-3\,p^3 + p^2 - 3p^2\,\sqrt{n}\log n + 3p(\log n)^2) + n(2p^3 - p^2 + p + \sqrt{n}\,(\log n)^3)}{n^3p^3 + 3n^2p^2\,\sqrt{n}\log n - 3n^2\,p(\log n)^2 - n\,\sqrt{n}\,(\log n)^3}$$

$$\geq 1 - \frac{C_5}{n^2} \tag{H.8}$$

Now we proceed with the proof of Theorem 1.

Let $n = n_0 m + r \quad 0 \leq r < n_0$. Then according to Lemma 2,

$$F_n(x) = \sum_{j=0}^{m} \binom{m}{j}\, p^j\, q^{m-j}\, H_1^j\left(x\sqrt{\frac{n}{n_0}}\right) * H_2^{m-j}\left(x\sqrt{\frac{n}{n_0}}\right) * F^r\left(x\sqrt{n}\right) \tag{H.9}$$

Here * denotes convolution and the exponents of the distribution functions H and F denote the corresponding numbers of convolutions.

We now divide the sum into two parts. By $\sum_1$ we will mean the sum over those values of j for which $j - mp > -\sqrt{m}\log m$ and by $\sum_2$ we will mean the sum over those values of j for which $j - mp < -\sqrt{m}\log m$, and by $\sum$ we will mean the sum over all values of j so that

$$F_n(x) = \sum_1 \binom{m}{j} p^j q^{m-j} H_i^j * H_2^{m-j} * F^r + \sum_2 \binom{m}{j} p^j q^{m-j} H_1^j * H_2^{m-j} * F^r \qquad \text{(H.10)}$$

The distribution $H_1^j * H_2^{m-j} * F^r$ has the characteristic function

$$h_1^j \left(t\sqrt{\frac{n_0}{n}}\right) h_2^{m-j} \left(t\sqrt{\frac{n_0}{n}}\right) f^r \left(\frac{t}{\sqrt{n}}\right)$$

and for $j \geq 2$ it has a square integrable density which we denote by $p_{mj}(x)$

Let

$$\phi_n(x) = \phi(x)\left[1 + \frac{\alpha_3}{6\sqrt{n}}(x^3 - 3x)\right]$$

Then

$$\int_{-\infty}^{\infty} | p_n(x) - \phi_n(x) | dx \leq \int_{-\infty}^{\infty} | \sum_1 \binom{m}{j} p^j q^{m-j} p_{mj}(x) - \phi_n(x) | dx + \sum_2 \binom{m}{j} p^j q^{m-j}$$

and therefore by Lemma 3

$$\int_{-\infty}^{\infty} | p_n(x) - \phi_n(x) | dx \leq \int_{-\infty}^{\infty} | \sum_1 - \phi_n(x) | dx + \frac{C_5}{n^2}$$

By the Cauchy-Schwarz inequality we have

$$\left(\int_{-\infty}^{\infty} | \sum_1 - \phi_n(x) | dx\right)^2 \leq \int_{-\infty}^{\infty} | \sum_1 - \phi_n(x) |^2 (1+x^2)dx \int_{-\infty}^{\infty} \frac{dx}{1+x^2} \qquad \text{(H.11)}$$

$$= \pi\left[\int_{-\infty}^{\infty} | \sum_1 - \phi_n(x) |^2 dx + \int_{-\infty}^{\infty} x^2 | \sum_1 - \phi_n(x) |^2 dx\right] \qquad \text{(H.12)}$$

$$= \pi [I + I_1] \text{ say} \qquad \text{(H.13)}$$

We evaluate I and I_1 separately. Since $\sum_1$ is square-integrable we have from Parseval's Theorem.

$$I = \int_{-\infty}^{\infty} | \sum_1 - \phi_n(x) |^2 \, dx$$

$$= \int_{-\infty}^{\infty} | \sum_1 \binom{m}{j} p^j q^{m-j} h_1^j \left(t\sqrt{\tfrac{n_0}{n}}\right) h_2^{m-j} \left(t\sqrt{\tfrac{n_0}{n}}\right) f^r \left(\frac{t}{\sqrt{n}}\right) - g_n(t) |^2 \, dt$$

$$\leq \int_{t \leq \Delta\sqrt{n}} | \sum_1 - g_n(t) |^2 \, dr + 2 \int_{t > \Delta\sqrt{n}} | \sum_1 |^2 \, dt$$
$$+ 2 \int_{t > \Delta\sqrt{n}} | g_n(t) |^2 \, dt \qquad \text{(H.14)}$$

where $\Delta \leq \dfrac{1}{24\beta_3}$

From Lemmas 1 and 3

$$\int_{|t| \leq \Delta\sqrt{n}} | \sum_1 - g_n(t) |^2 \, dt = \int_{|t| \leq \Delta\sqrt{n}} | \sum - g_n(t) - \sum_2 |^2 \, dt$$

$$\leq 2 \int_{|t| \leq \Delta\sqrt{n}} | f_n(t) - g_n(t) |^2 \, dt + 2 \int_{|t| \leq \Delta\sqrt{n}} | \sum_2 |^2 \, dt$$

$$\leq \frac{C_7}{n} \qquad \text{(H.15)}$$

Also

$$\int_{|t| > \Delta\sqrt{n}} | \sum_1 |^2 \, dt$$

$$\leq \int_{|t| > \Delta\sqrt{n}} | \sum_1 \binom{m}{j} p^j q^{m-j} h_2^{m-j} f^r |^2 | h_1 |^{mp} \, dt$$

$$\leq \int_{|t| > \Delta\sqrt{n}} (\sum_1 \binom{m}{j} p^j q^{m-j})^2 | h_1 \left(t\sqrt{\tfrac{n_0}{n}}\right) |^{mp} \, dt$$

$$\leq \int_{|t| > \Delta\sqrt{n}} | h_1 \left(t\sqrt{\tfrac{n_0}{n}}\right) |^{mp} \, dt$$

$$\leq \sqrt{\frac{n_0}{n}} \int_{|z| > \Delta\sqrt{n_1}} | h_1(z) |^{mp} \, dz$$

$$\leq \sqrt{\frac{n}{n_0}} \, e^{-C_4(mp-2)} \int_{-\infty}^{\infty} | h_1(z) |^2 \, dz \leq \frac{C_9}{n} \qquad \text{(H.16)}$$

since $j > \dfrac{mp}{2} > 1$ in the sum $\sum_1$ (for sufficiently large values of m)

Also

$$\int_{|t| > \Delta\sqrt{n}} | g_n(t) |^2 \, dt \leq \frac{C_{10}}{n} \qquad \text{(H.17)}$$

It follows from (H.15), (H.16) and (H.17) that

$$I \leq \frac{C_{11}}{n} \tag{H.18}$$

We now estimate I_1.

$$I_1 = \int_{-\infty}^{\infty} x^2 \mid \sum_1 - \phi_n(x) \mid^2 dx$$

$$= \int_{-\infty}^{\infty} \mid \sum_1 \binom{m}{j} p^j q^{m-j} h_1^j h_2^{m-j} f^r)' - g_n'(t) \mid^2 dt$$

$$= I_1' + I_2' + I_3' \tag{H.19}$$

and

$$I_1' = \int_{|t| \leq \Delta\sqrt{n}} \mid (\sum_1)' - g_n'(t) + (\sum_2)' - (\sum_2)' \mid^2 dt$$

$$\leq 2 \int_{|t| \leq \Delta\sqrt{n}} \mid f_n'(t) - g_n'(t) \mid^2 dt + 2 \int_{|t| \leq \Delta\sqrt{n}} \mid (\sum_2)' \mid^2 dt$$

$$\leq \frac{C_{12}}{n} \tag{H.20}$$

from Lemmas 1 and 3.

From Lemma 2 we have

$$I_2' = 2 \int_{|t| > \Delta\sqrt{n}} \mid (\sum_1)' \mid^2 dt \leq \frac{C_{13}}{n} \tag{H.21}$$

Also

$$I_3' = \int_{|t| > \Delta\sqrt{n}} \mid g_n'(t) \mid^2 dt \leq \frac{C_{14}}{n} \tag{H.22}$$

Thus from (H.18) and (H.23) we get

$$I_1 \leq \frac{C_{15}}{n} . \tag{H.23}$$

From (H.20) , (H.21) and (H.22) we get

$$\int_{-\infty}^{\infty} \mid p_n(x) - \phi_n(x) \mid dx \leq \frac{C}{\sqrt{n}} \tag{H.24}$$

Hence

$$\int_{-\infty}^{\infty} \mid p_n(x) - \phi(x) \mid \; lx \tag{H.25}$$

$$\leq \int_{-\infty}^{\infty} \mid \phi_n(x) - \phi(x) \mid \; dx \; + \int_{-\infty}^{\infty} \mid p_n(x) - \phi_n(x) \mid \; dx \tag{H.26}$$

$$\leq \int_{-\infty}^{\infty} \frac{\mid \alpha_3 \mid}{6\sqrt{n}} \, (x^3 - 3x) \, dx \; + \; \frac{C_1}{\sqrt{n}} \tag{H.27}$$

APPENDIX I

$$\text{Var } (Y_{L,l}) \leq \frac{N_0}{2} + \frac{N_J}{2}$$

Proof:

Denote Var $(Y_{L,l})$ as $V(Y_{L,l})$

Then

$$\begin{aligned}
V(Y_{L,l}) = {} & V(Y_{L,l}|Z_{j,l}=0)\ Pr(Z_{j,l}=0) \\
& + V(Y_{L,l}|Z_{j,l}\neq 0)\ Pr(Z_{j,l}\neq 0) \\
= {} & (1-\rho_L) V(C_L(N_{j,l}+\sqrt{E})) \\
& + \rho_L\ V\left(\left(C_L\left(\frac{1}{\sqrt{\rho_L}}\ n_{j,l} + N_{j,l}\sqrt{E}\right)\right)\right. \\
\leq {} & (1-\rho_L)\ V\ \left(N_{j,l}+\sqrt{E}\right) \\
& + \rho_L\ V\ \left(\frac{1}{\sqrt{\rho_L}}\ n_{j,l}\ +\ N_{j,l}\ +\ \sqrt{E}\ \right) \\
= {} & (1-\rho_L)\ V\ (N_{j,l})\ +\ \rho_L\ V\ \left(\frac{1}{\sqrt{\rho_L}}\ n_{j,l}\ +\ N_{j,l}\ \right) \\
= {} & (1-\rho_L)\ \frac{N_0}{2}\ +\ \frac{N_J}{2}\ +\ \rho_L\ \frac{N_0}{2} \\
= {} & \frac{N_0}{2}\ +\ \frac{N_J}{2}
\end{aligned}$$

APPENDIX J

We show here that $\beta_L \triangleq \frac{\sigma_I^i}{\sigma_L^j} \rightarrow 1$ by showing that $\beta_L^2 \rightarrow 1$. Now $\beta_L^2 = \frac{Var(X_{L,l})}{VarY_{L,l})}$. Hence it suffices to show that for the worst case jammer $Var(X_{L,l}) \rightarrow N_J/2 + N_0/2$ and $Var(Y_{L,l} \rightarrow N_J/2 + N_0/2$. From Appendix I we know that $Var(X_{L,l}) \leq N_J/2 + N_0/2$ and $Var(Y_{L,l}) \leq N_J/2 + N_0/2$. Now by simply choosing the sequence $\{\rho_L = 1\}_1^\infty$ it is clear from Appendix I that the jammer can achieve $Var(X_{L,l}) \rightarrow N_J/2 + N_0/2$ and $Var(Y_{L,l}) \rightarrow N_J/2 + N_0/2$. Thus clearly the worst-case jammer can achieve $Var(X_{L,l}) \rightarrow N_J/2 + N_0/2$ and $Var(Y_{L,l}) \rightarrow N_J/2 + N_0/2$

.

We also show here that $E(Y_{L,l}) = \sqrt{E} + \epsilon_L$ where $\epsilon_L \rightarrow 0$.

$$E(Y_{L,l}) = E(Y_{L,l}|Z_{j,l} = 0)Pr(Z_{j,l} = 0) + E(Y_{L,l}|Z_{j,l} \neq 0)Pr(Z_{j,l} \neq 0)$$

$$= (1 - \rho_L)(\sqrt{E} - \delta_L) + \rho_L(\sqrt{E} - \zeta_L)$$

$$\sqrt{E} + (1 - \rho_L)\delta_L + \rho_L\zeta_L.$$

Consider first the term $(1 - \rho_L)\delta_L$

$$(1 - \rho_L)\delta_L = \int_{-\alpha_L}^{-\alpha_L + 2\sqrt{E}} \frac{2x}{\sqrt{2\pi}N_0} e^{\frac{-(x - \sqrt{E})^2}{N_0}} dx$$

$$= (1 - \rho_L) \int_{-\alpha_L - \sqrt{E}}^{-\alpha_L + \sqrt{E}} \frac{2(y + \sqrt{E})}{\sqrt{2\pi N_0}} e^{\frac{-y^2}{N_0}} dy$$

$$(1 - \rho_L)\frac{1}{\sqrt{2\pi}}(e^{\frac{-(-\alpha_L - \sqrt{E})^2}{N_0}} - e^{\frac{-(-\alpha_L + \sqrt{E})^2}{N_0}})$$

$$+(1-\rho_L)\int_{-\alpha_L-\sqrt{E}}^{-\alpha_L+\sqrt{E}} \frac{2\sqrt{E}}{N_0} e^{\frac{-y^2}{N_0}}\, dy$$

Now as $L \rightarrow \infty$ the first term clearly goes to zero. Using the Mean Value Theorem [Bart 76, pg. 230] from calculus, the second term goes to zero too.

Now consider the second term $\rho_L \zeta_L$.

$$\rho_L \zeta_L = \rho_L \int_{-\alpha_L-\sqrt{E}}^{-\alpha_L+\sqrt{E}} \frac{2(y+\sqrt{E})}{\sqrt{2\pi}(N_0+\frac{N_J}{\rho_L})} e^{\frac{-y^2}{N_0+\frac{N_J}{\rho_L}}}\, dy$$

$$= \rho_L \frac{1}{\sqrt{2\pi}} (e^{\frac{-(-\alpha_L-\sqrt{E})^2}{N_0+\frac{N_J}{\rho_L}}} - e^{\frac{-(-\alpha_L+\sqrt{E})^2}{N_0+\frac{N_J}{\rho_L}}})$$

$$+\rho_L \int_{-\alpha_L-\sqrt{E}}^{-\alpha_L+\sqrt{E}} \frac{2\sqrt{E}}{\rho_L N_0 + N_J} e^{\frac{-y^2}{N_0+\frac{N_J}{\rho_L}}}\, dy$$

Now if $\inf_L \rho_L$ were greater than zero then clearly each term goes to zero with increasing L. If $\inf_L \rho_L = 0$ then each term is of the form $\rho_L c$ where $c \leq 1$ and hence each of the above terms goes to zero with increasing L.

BIBLIOGRAPHY

BIBLIOGRAPHY

[1] [Ahls 78] R. Ahlswede, "Elimination of correlation in random codes for arbitrarily varying channels", *Zeitschrift fur Wahrscheinlichkeitstheorie*, no. 33, pp. 159-175, 1978

[2] [Arau 80] A.Araujo, E.Gine, *The Central Limit Theorem for Real and Banach Valued Random Variables*, Wiley, !980

[3] [Ash 65] R.Ash, *Information Theory*, Interscience Publishers, 1965

[4] [Ash 72] R. Ash, *Real Analysis and Probability*, Academic Press,Inc., 1972

[5] [Aubi 82] J.P. Aubin, *Mathematical Methods of Game and Economic Theory*, North-Holland, 1982

[6] [Bart 76] R.G. Bartle, *The Elements of Real Analysis*, Wiley, 1976

[7] [Bhat 76] R.N.Bhattacharya, R.Ranga Rao, *Normal Approximation and Asymptotic Expansions*, Wiley, 1976

[8] [Blac 57] N.M. Blachman,"Communication as a game", *Wescon 1957 Conference Record*, 1957

[9] [Blac 54] D. Blackwell, M.A. Girshick, *Theory of Games and Statistical Decisions*, Dover Publications, INc., 1954

[10] [Blac 59] D. Blackwell, L. Breiman, A.J. Thomasian,"The capacity of a class of channels",*Annals of Mathematical Statistics*, no.30, pp.1229-1241

[11] [Blac 60] D. Blackwell, L. Breiman, A.J. Thomasian,"The capacities of certain channel classes under random coding", *Annals of Mathematical Statistics*, no.31, pp.558-567

[12] [Bord 85] J.M. Borden, D.J.Mason, R.J.McEliece,"Some information theoretic saddlepoints", *SIAM Journal on Control and Optimization*, vol. 23, no. 1, Jan 1985

[13] [Chan 85] L.F. Chang, An information-theoretic study of ratio-threshold antijam techniques, Ph.D. Thesis, University of Illinois at Urbana-Champaign, 1985

[14] [Crep 85] P.N. Crepeau, D.N. McGregor,"Anomalous union bound behaviour for MFSK signalling on inverse linear channels", *IEEE Transactions on Communications*, vol. COM-33, no. 7, pp. 1153-1159, Nov. 1985

[15] [Crep 87] P.N. Crepeau,"Performance of coded FH/MFSK with a quantizer-limiter in a worst-case partial-band gaussian interference Channel", *Proceedings of the Military Communications Conference, 1987*, 12.2.1-12.2.6, 1987

[16] [Csiz 81] I. Csiszar and J. Korner, *Information Theory : Coding Theory for Discrete Memoryless Systems*, Academic Press, 1981

[17] [Dobr 59] R.L. Dobrushin,"Optimum information transmission through a channel with unknown parameters", *Radio Engineering Electronics*, vol. 4, no. 12, 1959

[18] [Dubi 62] L.E. Dubins,"One extreme points of convex sets", Journal of Mathematical Analysis and Applications, pp. 237-244, 1962

[19] [Eric 85] T. Ericson,"The arbitrarily varying channel and the jamming problem", *Internal Report LiTH-ISY-I-0772, Department of Electrical Engineering, Linkoping University, Sweden*, 1985

[20] [Esse 58] C.G. Esseen,"On mean central limit theorems", *Kungl. Tekn. Hogsk. Handl., Stockholm*, no. 121, 1958

[21] [Fell 68] W. Feller,"On the Berry-Esseen theorem", *Zeitschrift fur Wahrscheinlichkeitstheorie*, no.10, pp.261-268, 1968

[22] [Frie 86] K.J. Friedrichs," Error analysis for noncoherent M-ary orthogonal communication systems in the presence of arbitrary Gaussian interference", *IEEE Transactions on Communications* vol. 34, pp. 817-821, Aug. 1986

[23] [Gera 82] E.A. Geraniotis, M.B. Pursley," Error probabilities for slow frequency-hopped spread-spectrum multiple access communications over fading channels", *IEEE Transactions on Communications* vol. 31, pp.996-1009, May 1982

[24] [Gall 68] R.G. Gallagher, *Information Theory and Reliable Communication*, Wiley, 1968

[25] [Gned 54] B.V. Gnedenko, A.N. Kolmogorov, *Limit Distributions for Sums of Independent Random Variables*, Addison-Wesley, 1954

[26] [Hall 82] P. Hall, *Rates of Convergence in the Central Limit Theorem*, Pitman Publishing Inc., 1982

[27] [Hous 75] S.W. Houston,"Modulation techniques for communication, Part1: Tone and noise jamming performance of spread spectrum M-ary FSK and 2,4-ary DPSK waveforms", *Proc. IEEE National Aerospace and Electronics Conference*, pp. 51-58, June 1975

[28] [Karl 59] S. Karlin, *Mathematical Methods and Theory in Games, Programming and Economics, vols 1 and 2* Addison- Wesley, 1959

[29] [Kell 85] C.M. Keller, Diversity combining for frequency-hop spread-spectrum communications with partial-band interference and fading, Ph.D. Thesis, University of Illinois at Urbana-Champaign, Sept. 1985

[30] [Ky 51] Ky Fan,"Fixed-point and minimax theorems in locally convex toplogical linear spaces", *Proceedings of the National Academy of Science*, vol. 38, pp. 119- 124, 1951

[31] [Lee 84] J.S. Lee , R.H.French, L.E. Miller," Probability of error analyses of a BFSK frequency-hopping system with diversity under partial-band jamming interference-Part 1: Performance of a square-law linear combining soft decision receiver", *IEEE Transactions on Communications*, vol. 32,,pp.645-653, June 1984

[32] [Loev 77] M. Loeve, *Probability Theory I*, Springer-Verlag, 1977

[33] [Lamp 66] J. Lamperti, *Probability*, W.A. Benjamin Inc., 1966

[34] [Luen 69] D.G. Luenberger *Optimization by Vector Space Methods*, Wiley, 1969

[35] [Mass 81] J.L. Massey,"Coding and modulation in Digital Communications", *Proceedings of the International Zurich Seminar on Digital Communications*, Mar. 1974

[36] [McEl 77] R.J. McEliece, *The Theory of Information and Coding*, Addison-Wesley, 1977

[37] [McEl 81] R.J. McEliece, W.E. Stark,"An information-theoretic study of communication in the presence of jamming", *IEEE International Con-*

ference on Communication, Conference Record, pp.45.3.1-45.3.5, June 1981

[38] [McEl 84] R.J. McEliece, W.E. Stark,"Channels with block interference", *IEEE Transactions on Information Theory*, vol. 30, pp.44-53, Jan. 1984

[39] [McEl 83] R.J. McEliece, E.R. Rodemich,"A study of optimal abstract jamming strategies vs. noncoherent MFSK", *Milcom Record1983*, pp. 1.1.1 -1.1.6, 1983

[40] [McEl 83] R.J. McEliece,"Communication in the presence of jamming- an information theoretic approach", in *Secure Digital Communications*, Springer- Verlag, pp. 127-166 1983

[41] [McEl 82] R.J. McEliece, W.E. Stark"The optimal code rate vs. a partial band jammer", *Military Communications Conferencw Record 1982* , pp. 45.3.1 - 45.3.5, 1982

[42] [Mich 81] R. Michel,"On the constant in the nonuniform version of the Berry-Esseen Theorem", *Zeitschrift fur Wahrscheinlichkeitstheorie*, no. 55, pp.109-117, 1981

[43] [Peng 86] W.C. Peng, Some communication jamming games, Ph.D. Thesis, University of Southern California, Jan 1986

[44] [Petr 75] V.V. Petrov, *Sums of Independent Random Variables*, Springer-Verlag, 1975

[45] [Proa 83] J.G. Proakis, *Digital Communications*, McGraw-Hill, 1983

[46] [Prok 52] Y.V. Prokhorov,"A local limit theorem for densities", *Dokl. Akad. Nauk SSSR* , pp.797-800, 1952

[47] [Root 61] W.L. Root,"Communication through unspecified additive noise", *Information and Control*, vol. 4, pp. 15-29, 1961

[48] [Scha 68] H. Schubert, *Topology*, Macdonald and co. Ltd., 1968

[49] [Simo 85] M.K. Simon, J.K. Omura, R.A. Scholz, B.K. Levitt *Spread Spectrum Communications, vols 1,2 and 3*, Computer Science Press, 1985

[50] [Sira 62] S.K. Sirazhdinov, M. Mamatov,"On convergence in the mean for densities", *Theory of Probability and its Applications*,no. 4, pp.425-428, 1962

[51] [Smit 71] J.G. Smith,"The information capacity of amplitude- and variance-constrained scalar gaussian channels", *Information and Control*, vol. 18, pp.203-219, 1971

[52] [Star 82] W.E. Stark, Coding for frequency-hopped spread-spectrum channels with partial-band interference, Ph.D. Thesis, University of Illinois at Urbana-Champaign, 1982

[53] [Star 85a] W.E. Stark,"Coding for frequency-hopped spread-spectrum communication with partial-Band interference-Part 1: capacity and cut-off rate", *IEEE Transactions on Communications* vol. 33, no. 10, Oct. 1986

[54] [Star 85a] W.E. Stark,"Coding for frequency-hopped spread-spectrum communication with partial-band interference-Part 2: coded performance", *IEEE Transactions on Communications* vol. 33, no. 10, Oct. 1986

[55] [Star 86] W.E. Stark,D. Teneketzis, S.K. Park,"Worst-case analysis of

partial-band interference", *Proceedings of the 1986 Conference on Information Sciences and Systems* , 1986

[56] [Stig 66] I.G. Stiglitz,"Coding for a class of unknown channels", *IEEE Transactions on Information Theory*, vol. 12, pp.189-195, 1966

[57] [Vite 64] A.J. Viterbi,"Phase coherent communication over the continuous Gaussian Channel"in *Digital Communications with Space Applications* edited by S. Golomb.

[58] [Vite 66] A.J. Viterbi, *Principles of Coherent Communication*, Mcgraw-hill, 1966

[59] [Vite 79] A.J. Viterbi, J.K. Omura, *Principles of Digital Communication and Coding*, McGraw-hill, 1979

[60] [Wits 80] H.S. Witsenhausen," Some aspects of convexity useful in information theory', *IEEE Transactions on Information Theory*, vol. 26, pp.265-271, May 1980

[61] [Wolf 78] J.Wolfowitz, *Coding Theorems of Information Theory*, Springer-Verlag, 1978

[62] [Woze 65] J.M. Wozencraft, I.M. Jacobs, *Principles of Communication Engineering*, Wiley, 1965

END

DATE

FILMED

JAN

1988