AD-A071 399 MASSACHUSETTS INST OF TECH CAMBRIDGE LAB FOR INFORMA--ETC F/G 9/4
CODING AND CAPACITY FOR ADDITIVE WHITE GAUSSIAN NOISE MULTI-USE--ETC(U)
MAY 79 L H OZAROW N00014-75-C-1183
UNCLASSIFIED LIDS-TH-917 NL

1 OF 2
AD A071399

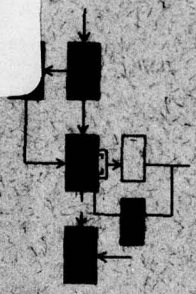
AD A071399

May, 1979

LIDS-TH-917

LEVEL

Research Supported By:
ARPA Contract N00014-75-C-1183
NSF Grant NSF/ENG-77-19971



CODING AND CAPACITY FOR ADDITIVE WHITE GAUSSIAN NOISE MULTI-USER CHANNELS WITH FEEDBACK

DDC FILE COPY

DDC
RECEIVED
JUL 18 1979
A

Lawrence Howard Ozarow

DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

Laboratory for Information and Decision Systems
Formerly
Electronic Systems Laboratory
MASSACHUSETTS INSTITUTE OF TECHNOLOGY, CAMBRIDGE, MASSACHUSETTS 02139

79 07 16 030

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER
2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER

4. TITLE (and Subtitle)
CODING AND CAPACITY FOR ADDITIVE WHITE GAUSSIAN NOISE MULTI-USER CHANNELS WITH FEEDBACK

5. TYPE OF REPORT & PERIOD COVERED
Thesis

6. PERFORMING ORG. REPORT NUMBER
LIDS-TH-917

7. AUTHOR(s)
Lawrence Howard Ozarow

8. CONTRACT OR GRANT NUMBER(s)
ARPA Contract N00014-75-C-1183

9. PERFORMING ORGANIZATION NAME AND ADDRESS
Massachusetts Institute of Technology
Laboratory for Information and Decision Systems
Cambridge, MA 02139

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
Program Code No. 5T10
ONR Identifying No. 049-383

11. CONTROLLING OFFICE NAME AND ADDRESS
Defense Advanced Research Projects Agency
1400 Wilson Boulevard
Arlington, Virginia 22209

12. REPORT DATE
May, 1979

13. NUMBER OF PAGES
147 pages

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)
Office of Naval Research
Information Systems Program
Code 437
Arlington, Virginia 22217

15. SECURITY CLASS. (of this report)
UNCLASSIFIED

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)
Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

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

Deterministic coding schemes are presented for the additive white gaussian noise two user multiple access and broadcast channels with noiseless feedback. The error probabilities for these schemes approach zero at a rate which is doubly exponential in block length.

Outer bounds on the capacity region are also obtained for both channels. The achievable region obtained for the multiple access channel is shown to coincide with the outer bound, yielding a solution of the capacity region for

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-LF-014-6601

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

410 950

this problem. While the achievable region for the broadcast channel does not coincide with the outer bound, for all cases except that in which one channel is a physically degraded version of the other, the achievable region lies outside the set of rates achievable in the absence of feedback. This is the first case in which it has been demonstrated that feedback can enlarge the capacity region of broadcast channels.

(11) May 1979

(14) LIDS-TH-917

(12) 150 p.

(9) Doctoral thesis.

(6)

CODING AND CAPACITY FOR ADDITIVE WHITE GAUSSIAN NOISE MULTI-USER CHANNELS WITH FEEDBACK.

by

(10) Lawrence Howard Ozarow

This report is based on the unaltered thesis of Lawrence Howard Ozarow, submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at the Massachusetts Institute of Technology in May, 1979. The research was conducted at the M.I.T. Laboratory for Information and Decision Systems, with support provided in part by the Advanced Research Project Agency under Contract No. (15) N00014-75-C-1183, and the National Science Foundation under grant (17) NSF-ENG-77-19971.

Laboratory for Information and Decision Systems
Massachusetts Institute of Technology
Cambridge, MA 02139

410 950

CODING AND CAPACITY FOR ADDITIVE WHITE GAUSSIAN NOISE MULTI-USER CHANNELS WITH FEEDBACK

by

LAWRENCE HOWARD OZAROW

B.S., Columbia University
(1970)

S.M., Massachusetts Institute of Technology
(1973)

E.E., Massachusetts Institute of Technology
(1973)

Accession For
NTIS GRA&I
DDC TAB
Unannounced
Justification
By
Distribution/
Availability Codes
Dist | Available/or special
A

SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE
DEGREE OF

DOCTOR OF PHILOSOPHY

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May, 1979

© Lawrence Howard Ozarow 1979

Signature of Author .....
Department of Electrical Engineering and Computer Science,
May 4, 1979

Certified by.....
Thesis Supervisor

Accepted by .....
Chairman, Department Committee

CODING AND CAPACITY FOR ADDITIVE WHITE GAUSSIAN NOISE MULTI-USER CHANNELS WITH FEEDBACK

by

LAWRENCE HOWARD OZAROW

Submitted to the Department of Electrical Engineering and Computer Science on May 4,1979, in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

ABSTRACT

Deterministic coding schemes are presented for the additive white gaussian noise two user multiple access and broadcast channels with noiseless feedback. The error probabilities for these schemes approach zero at a rate which is doubly exponential in block length.

Outer bounds on the capacity region are also obtained for both channels. The achievable region obtained for the multiple access channel is shown to coincide with the outer bound, yielding a solution of the capacity region for this problem. While the achievable region for the broadcast channel does not coincide with the outer bound, for all cases except that in which one channel is a physically degraded version of the other, the achievable region lies outside the set of rates achievable in the absence of feedback. This is the first case in which it has been demonstrated that feedback can enlarge the capacity region of broadcast channels.

THESIS SUPERVISOR: Cyril Leung

Title: Assistant Professor of Electrical Engineering and Computer Science.

ACKNOWLEDGEMENTS

First of all, I would like to thank my thesis supervisor, Professor Cyril Leung for his helpful suggestions and many ideas during this research. I also want to thank him for his careful and thoughtful help during the preparation of the thesis.

The suggestions and interest of my thesis readers, Professors Robert Gallager and Pierre Humblet are greatly appreciated.

I also thank Ms. Leni Gross for her help in typing the manuscript, and Ms. Holly Glaser for preparing the figures.

I want also to thank my supervisors at Lincoln Labs, Dr. Charles Niessen and Dr. Thomas Seay. In connection with my work at Lincoln and my doctoral career in general I want to extend particular thanks to Dr. Joseph De Rosa for his enthusiasm, patience, help, and advice.

Thanks to my parents, Boris and Mollie Ozarow, who have always provided advice, support and encouragement, throughout my life.

Special thanks to Betty for everything.

This research was carried out at the MIT Laboratory for Information and Decision Systems with support provided in part by the Advanced Research Project Agency under Contract No. N00014-75-C-1183, and the National Science Foundation under Grant NSF/ENG-77-19971.

DEDICATED

TO MY

WHOLE FAMILY

TABLE OF CONTENTS

	Page No.
TITLE PAGE	1
ABSTRACT	2
ACKNOWLEDGEMENTS	3
TABLE OF CONTENTS	5
LIST OF FIGURES	7
1. INTRODUCTION	9
1.1 Motivation	9
1.2 Background	13
1.2.1 The Multiple Access Channel	13
1.2.2 The Broadcast Channel	26
1.2.3 The Coding Scheme of Schalkwijk and Kailath	31
2. AWGN MAC WITH FEEDBACK	40
2.1 Achievable Region	42
2.1.1 The Point (R_1^*, R_2^*)	42
2.1.2 Completion of the Achievable Region	52
2.2 Converse	56
2.2.1 Outer Bound for Discrete Memoryless MAC's	58
2.2.2 Outer Bound for AWGN MAC's	68
2.3 The Capacity Region of the AWGN MAC with Feedback	75
2.4 Capacity Region with Correlated Messages and Feedback	82
3. AWGN BROADCAST CHANNEL WITH FEEDBACK	87
3.1 Achievable Region	89
3.2 Outer Bound to the Capacity Region	94
3.2.1 Outer Bound for Discrete Memoryless Channels	94
3.2.2 Outer Bound for the AWGN BC with Feedback	97
3.3 Comparison of Inner and Outer Bounds	101

TABLE OF CONTENTS (continued)

	Page No.
3.3.1 Degraded Channels	101
3.3.2 Channels with Independent Noise	102
3.3.3 Intermediate Cases	112
4. CONCLUSIONS	115
APPENDIX A: ROOTS OF $P(\rho)$	119
APPENDIX B: CONCAVITY OF MUTUAL INFORMATION FOR THE MAC	120
APPENDIX C: CODING FOR THE AUGMENTED AWGN BC WITH FEEDBACK	122
APPENDIX D: AN OUTER BOUND FOR BROADCAST CHANNELS	128
REFERENCES	143
BIOGRAPHICAL NOTE	146

LIST OF FIGURES

Figure No.		Page No.
1.1	Multiple Access Channel with Two Senders.	14
1.2	Capacity Region of the AWGN MAC Without Feedback for $P_1 = P_2 = 10$, $\sigma^2 = 1$.	16
1.3	Achievable Region of Cover and Leung for the AWGN MAC with Feedback for $P_1 = P_2 = 10$, $\sigma^2 = 1$.	21
1.4	Test Encoder for feedback coding Scheme of Cover and Leung.	22
1.5	Broadcast Channel with Two Receivers.	28
1.6	Capacity Region without Feedback and Achievable Region and Outer Bound with Feedback for $P = 10$, $\sigma_1^2 = 1$, $\sigma_2^2 = 2$.	32
1.7	Single-Source Single-Destination Communication with Additive Noise and Feedback.	34
2.1	AWGN MAC with Feedback.	41
2.2	Capacity Region of the AWGN MAC with Feedback for $P_1 = P_2 = 10$, $\sigma^2 = 1$.	57
2.3	Basic Sets for the Outer Bound to the Capacity Region of the AWGN MAC with Feedback.	78
3.1	AWGN Broadcast Channel with Feedback.	88
3.2	Augmented Broadcast Channel.	96
3.3	Achievable Region and Capacity Region for Physically Degraded AWGN BC with Feedback for $P = 10$, $\sigma^2 = 1$, and $\sigma_2^2 = 1$.	103
3.4	Achievable Region and Outer Bound for AWGN BC with Independent Noises and Feedback for $P = 1$, $\sigma_1^2 = \sigma_2^2 = 1$.	105
3.5	Achievable Region and Outer Bound for AWGN BC with Independent Noises and Feedback for $P = 10$, $\sigma_1^2 = \sigma_2^2 = 1$.	106

LIST OF FIGURES (continued)

Figure No.		Page No.
3.6	Achievable Region and Outer Bound for AWGN BC with Independent Noises and Feedback for P = 100, $\sigma_1^2 = \sigma_2^2 = 1$.	107
3.7	Achievable Region and Outer Bound for AWGN BC with Independent Noises and Feedback for P = 1, $\sigma_1^2 = 1$ and $\sigma_2^2 = 2$.	109
3.8	Achievable Region and Outer Bound for AWGN BC with Independent Noises and Feedback for P = 10, $\sigma_1^2 = 1$ and $\sigma_2^2 = 2$.	110
3.9	Achievable Region and Outer Bound for AWGN BC with Independent Noises and Feedback for P = 100, $\sigma_1^2 = 1$ and $\sigma_2^2 = 2$.	111
3.10	Achievable Regions and Outer Bounds for AWGN BC with Feedback. P = 10, $\sigma^2 + \sigma_1^2 = 1$ and $\sigma^2 + \sigma_2^2 = 2$. Common Noise Variance $\sigma^2 = 0,.5,1$.	113
3.11	Achievable Regions and Outer Bounds for AWGN BC with Feedback. P = 10 and $\sigma^2 + \sigma_1^2 = \sigma^2 + \sigma_2^2 = 1$. Common Noise Variance $\sigma^2 = 0,.5,1$.	114
D.1	Basic Sets for Proof of Theorem D.2.	136
D.2	Sato's Bound and Capacity Region of AWGN BC with P = 10, $\sigma_1^2 = 1$ and $\sigma_2^2 = 2$.	140
D.3	Capacity Region of Blackwell's Example.	142

CHAPTER 1

INTRODUCTION

1.1 Motivation

Much of the recent work in information theory has involved multiple user channels. Such configurations seem natural in view of modern communication environments, for example satellite or data networks. In addition the results obtained for multiple user channels are sufficiently different in character from those in single-source single-destination problems to justify interest in the area from a purely theoretical standpoint. A reasonably complete survey of known results in multiple user communication problems is contained in [1].

In a multiple user problem the notion of the capacity of a channel is extended to that of an admissible rate region, or capacity region. In a most general setting, each of M data sources wishes to communicate reliably with each of N data sinks over some given M-input N-output channel. An admissible rate vector is a point $\{R_{ij}\}$ in R_+^{MN} (the positive orthant of real MN space) such that the i^{th} transmitter can communicate to the j^{th} receiver with arbitrarily small probability of error for all i,j at rates R_{ij}. The closure of all such $\{R_{ij}\}$ is the capacity region of the channel. A further generalization sometimes considered is to allow the messages to have some correlation. That is, two or more transmitters might have one common message intended for a receiver, which they can communicate cooperatively, in addition to a set of independent messages. The dual situation is also possible; one trans-

mitter might have a common message intended for all or some subset of the receivers, in addition to private messages for each.

Multi-user problems are rarely considered in such generality. Two canonical problems of both theoretical and practical significance are the two user broadcast channel (BC) in which one transmitter wishes to communicate separate information to two distinct receivers, and the two user multiple access channel (MAC) in which two transmitters wish to communicate separate information to one receiver. Admissible rate regions in these cases are subsets of the positive quadrant of R^2.

In this dissertation we extend these models in that we allow feedback from receivers to transmitters. The results of the use of feedback are interesting since, unlike the single-input single-output channel, where it is known [32] that feedback does not increase the capacity of memoryless channels, it has been shown that in the case of memoryless MAC's, feedback can increase the capacity region ([12],[13]). On the other hand, it has been shown that for at least one broadcast model, that in which one channel is a physically degraded version of the other, feedback does not increase capacity ([8],[16]). Although El-Gamal in [8] conjectures that this result holds true for more general broadcast channels, it will be shown in this dissertation that this conjecture is false.

The results presented in this dissertation will involve deterministic coding schemes for additive white gaussian noise (AWGN) multiple user channels with feedback, whose operation require that the feedback be noiseless. The coding schemes are extensions of the feed-

back coding scheme for single user AWGN channels developed by Schalkwijk and Kailath [2] and Schalkwijk [3], which also required noiseless feedback. The noiseless assumption might be well justified for example, in the case of satellite communication, where the satellite-to-ground link, inherently power limited, is considerably more error prone than the reverse link, over which ground stations can operate at much higher power.

The results obtained here are quite interesting in that the codes are deterministic and reasonably simple to implement, and yield error rates which approach zero with doubly exponential behavior in block length. In addition the data rates achieved both disprove the conjecture mentioned above for the BC, and also exceed previously known achievable rates for the AWGN MAC. Indeed for this MAC we will show that the set of achievable rates coincides with the capacity region.

Before proceeding we summarize two basic multi-user techniques which will be useful in the sequel. The first is time-sharing. The time sharing argument is that if two rate pairs (R_1^1, R_2^1) and (R_1^2, R_2^2) are achievable, any point of the form

$$\begin{aligned} R_1 &= \alpha R_1^1 + (1-\alpha) R_1^2 \\ R_2 &= \alpha R_2^1 + (1-\alpha) R_2^2 \end{aligned} \tag{1.1.1}$$

is achievable for all α between 0 and 1. The intermediate point is achieved by employing a code which achieves (R_1^1, R_2^1) for a fraction α of the time, and a code which achieves (R_1^2, R_2^2) for the remaining $1-\alpha$. One consequence of the time-sharing argument is that capacity regions

are always convex. Another is that in cases for which the capacity region is a polygon, the achievability of the extreme points suffices to demonstrate the achievability of the entire capacity region. Another property derivable from time sharing arguments is that if a point (R_1', R_2') is achievable, then all points (R_1, R_2) which satisfy

$$\begin{aligned} 0 \leq R_1 \leq R_1' \\ 0 \leq R_2 \leq R_2' \end{aligned} \tag{1.1.2}$$

are also achievable. Therefore, in general, achieving points on the boundary (in R^2) of the achievable region suffices to demonstrate the achievability of the entire region.

The other technique which is useful is superposition. The exact form of the superposition argument varies with the specific application, but essentially it is as follows: when two independent codewords are combined by some method appropriate to the structure of the channel, one codeword may be decoded by treating the other as noise (that is, part of the randomization imparted by the channel). Once this codeword is known (at least with high probability) its effect can be removed by the decoder, and the other codeword detected as though the first were known. See [4] for a discussion of the superposition argument in the broadcast context.

In the remainder of this chapter we will discuss the two models to be considered, the nature of previously known results, and the Schalkwijk-Kailath coding scheme. The fundamental information theoretic

quantities used below are as in Gallager [24]. Non-standard or ambiguous terminology will be defined when necessary.

1.2 Background

1.2.1 The Multiple Access Channel

The MAC was first discussed by Ahlswede [9] and Liao [10]. A fairly complete discussion generalized to include the presence of correlated sources was contributed by Slepian and Wolf [11].

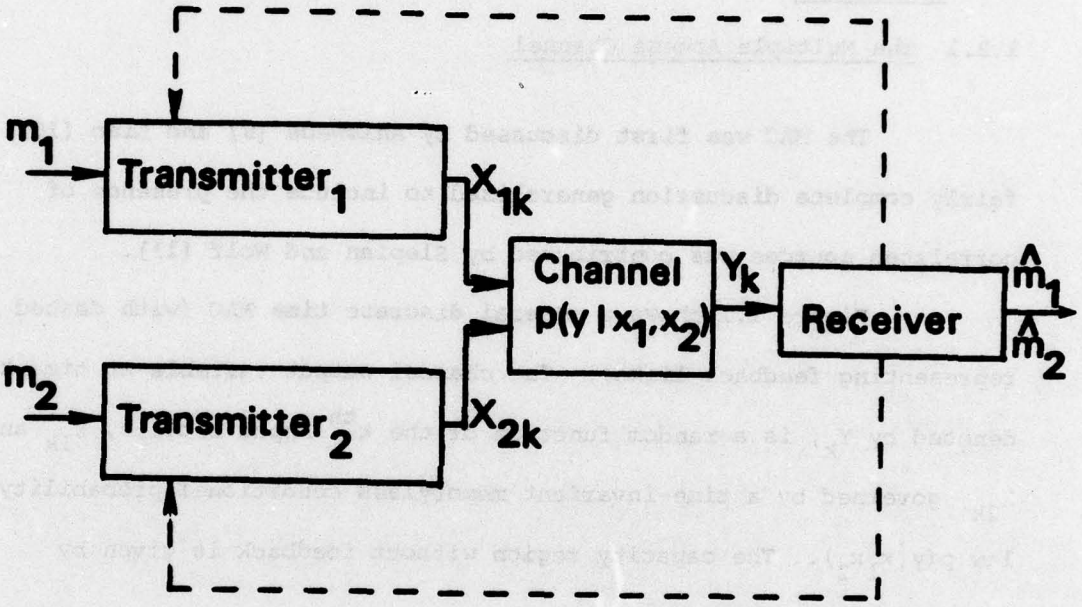
Figure 1.1 shows a general discrete time MAC (with dashed lines representing feedback links). The channel output variable at time k, denoted by Y_k, is a random function of the k^{th} input letters, X_{1k} and X_{2k}, governed by a time-invariant memoryless conditional probability law $p(y|x_1x_2)$. The capacity region without feedback is given by

$$C = \text{co}[\bigcup_{p\in P} K(p)] \tag{1.2.1}$$

where co[ ] denotes closure of the convex hull, $K(p)$ is given by

$$\begin{aligned} K(p) = \{(R_1,R_2):\ & R_1 \le I(X_1;Y|X_2) \\ & R_2 \le I(X_2;Y|X_1) \\ & R_1 + R_2 \le I(X_1X_2;Y)\} \end{aligned} \tag{1.2.2}$$

and P is the set of probability assignments on (X_1,X_2) for which X_1 and X_2 are statistically independent. In the sequel we will omit the customary but redundant imposition of a lower bound of zero for rates.



1.1 Multiple Access Channel with Two Senders.

Analogous expressions hold for some continuous amplitude channels with suitable constraints on the inputs. For the additive white gaussian noise channel, $p(y|x_1x_2) \sim N(x_1+x_2,\sigma^2)$, where $p(\ )$ is in this case a probability density function, resulting from the addition of an independent gaussian noise variable with variance σ^2 to the sum of x_1 and x_2. Under the constraint that $E[X_i^2] \leq P_i$ for $i = 1,2$ (1.2.1) and (1.2.2) reduce to

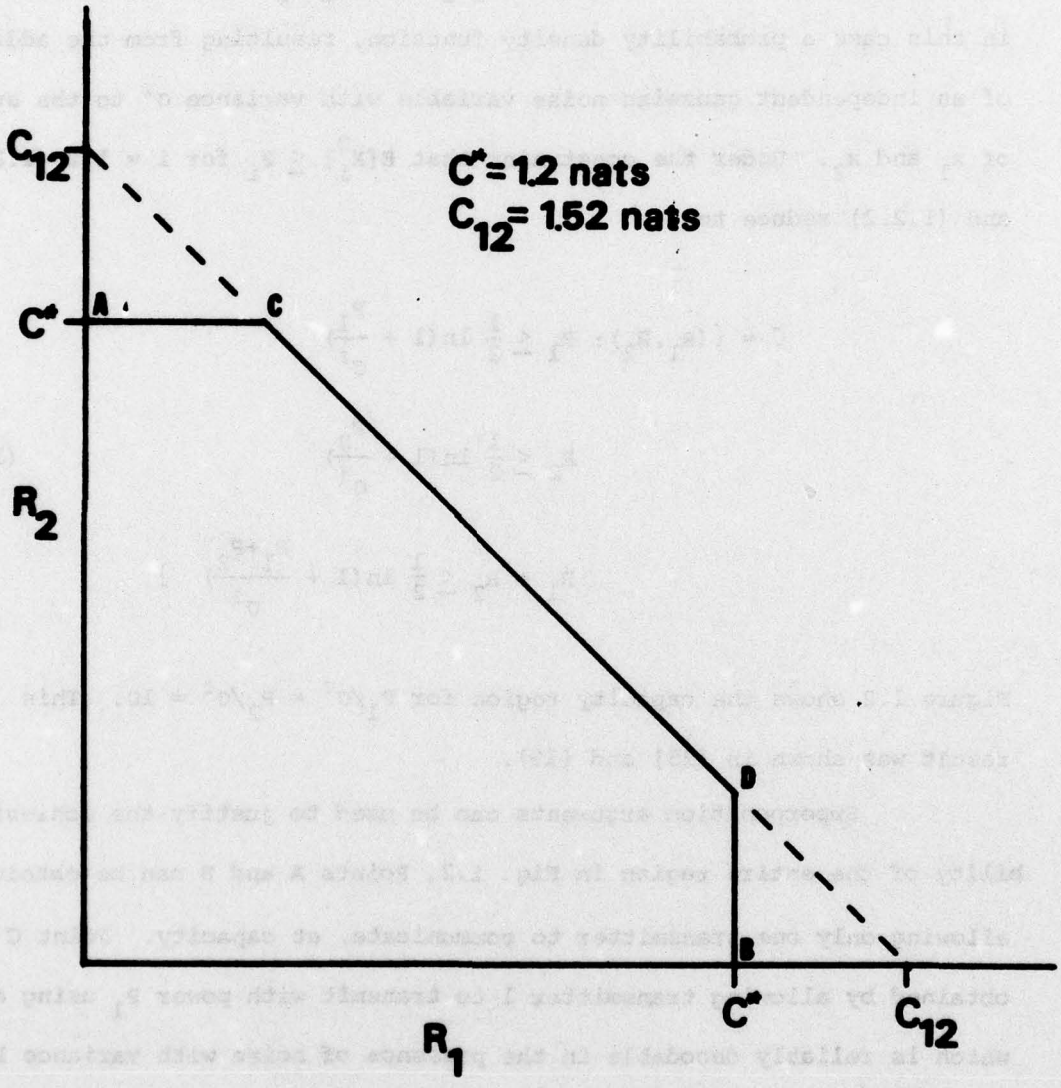
$$C = \{(R_1,R_2): R_1 \leq \frac{1}{2} \ln(1 + \frac{P_1}{\sigma^2})$$

$$R_2 \leq \frac{1}{2} \ln(1 + \frac{P_2}{\sigma^2}) \qquad (1.2.3)$$

$$R_1 + R_2 \leq \frac{1}{2} \ln(1 + \frac{P_1+P_2}{\sigma^2}) \quad \}$$

Figure 1.2 shows the capacity region for $P_1/\sigma^2 = P_2/\sigma^2 = 10$. This result was shown in [25] and [19].

Superposition arguments can be used to justify the achievability of the entire region in Fig. 1.2. Points A and B can be obtained by allowing only one transmitter to communicate, at capacity. Point C is obtained by allowing transmitter 1 to transmit with power P_1 using a code which is reliably decodable in the presence of noise with variance $P_2+\sigma^2$. Then if the code letters $\{x_{ik}\}$ ($k=1,\ldots,N$ and $i=1,2$) are sequences of gaussian random variables with variance P_i, reliable transmission of the message sent by transmitter 1 is possible at rates up to



1.2 Capacity Region of the AWGN MAC without Feedback for $P_1 = P_2 = 10$, $\sigma^2 = 1$.

$$R_1 = H(Y) - H(Y|X_1)$$
$$= \frac{1}{2} \ln 2\pi e(P_1+P_2+\sigma^2) - \frac{1}{2} \ln 2\pi e(P_2+\sigma^2)$$
$$= \frac{1}{2} \ln \left(1 + \frac{P_1}{P_2+\sigma^2} \right) \qquad (1.2.4)$$

Once message 1 is decoded, $\{x_{1k}\}_1^N$ can be subtracted from the received data. Then message 2 may be reliably decoded at rates up to

$$R_2 = H(Y|X_1) - H(Y|X_1X_2)$$
$$= \frac{1}{2} \ln 2\pi e(P_1+\sigma^2) - \frac{1}{2} \ln 2\pi e\sigma^2 \qquad (1.2.5)$$
$$= \frac{1}{2} \ln \left(1 + \frac{P_1}{\sigma^2} \right)$$

This argument may be reversed to obtain point D. The remainder of the capacity region may be achieved by the use of time-sharing.

The addition of feedback, as mentioned above, can enlarge the capacity region. This result was first shown using an ad hoc scheme for the noiseless binary erasure MAC by Gaarder and Wolf [12]. This channel has input alphabets $X_1 = X_2 = \{0,1\}$, output alphabet $Y = \{0,1,e\}$, and a channel probability function given by $p(0|00) = p(1|11) = p(e|01) = p(e|10) = 1$. That is, when the inputs agree, the output equals their common value, and when they disagree an erasure occurs. For this channel, the region specified by (1.2.2) is

$$R_1 \leq 1$$
$$R_2 \leq 1 \qquad (1.2.6)$$

$$R_1+R_2 \leq 1.5 \quad \text{bits/transmission} \tag{1.2.6}$$

Roughly, the approach of [12] is as follows: let the transmitters independently choose sequences of input letters of length N_1, where each letter is equiprobably 0 or 1. They then send these bits over the channel. On the average $N_1/2$ transmissions will result in erasures. For each erasure, specifying either transmitter 1's or transmitter 2's bit will resolve the receiver's uncertainty. By the use of feedback both transmitters know which bits were erased, and what transmitter 1's intended bit was for each transmission (transmitter 1 does since he sent it and transmitter 2 does since it disagreed with his own). The transmitters must now resolve $N_1/2$ bits of uncertainty in N_2 transmissions. With the transmitters sending a common message, the channel is noiseless with ternary output and quaternary input. It may therefore be used at rates up to

$$H(Y) = \log_2 3 \quad \text{bits/transmission} \tag{1.2.7}$$

so that N_1 bits can be conveyed in

$$N_2 = \frac{N_1}{2} \frac{1}{\log_2 3} \quad \text{transmissions.} \tag{1.2.8}$$

Therefore the rate pair achieved is given by

$$R_1 = R_2 = \frac{N_1}{N_1 + N_2} = .76 \quad \text{bits/transmission} \tag{1.2.9}$$

which lies outside the region given by (1.2.6).

Since the appearance of [12], Cover and Leung [13] have found an achievable region for discrete memoryless MAC's with feedback given by

$$A = \text{co}[\bigcup_{p\in P} K(p)] \tag{1.2.10}$$

where $K(p)$ is given by

$$\begin{aligned} K(p) = \{(R_1,R_2): \; & R_1 \le I(X_1;Y|X_2U) \\ & R_2 \le I(X_2;Y|X_1U) \\ & R_1+R_2 \le I(X_1X_2;Y)\} \end{aligned} \tag{1.2.11}$$

and P is the set of joint probability assignments on (U,X_1,X_2) for which

$$p(ux_1x_2y) = p(u)p(x_1|u)p(x_2|u)p(y|x_1x_2) \tag{1.2.12}$$

The region given by (1.2.7) is at least as large as that given by (1.2.2) since we can always choose U to be a degenerate random variable which takes on only one value with probability one, in which case (1.2.7) reduces to (1.2.2).

When applied to the noiseless binary erasure MAC, this region can be shown to include the point (R_1,R_2) given by

$$R_1 = R_2 = .7911 \quad \text{bits/transmission} \tag{1.2.13}$$

An outer bound based on total cooperation between the transmitters can be used to show that the maximum achievable R for which $R_1 = R_2 = R$ is given by

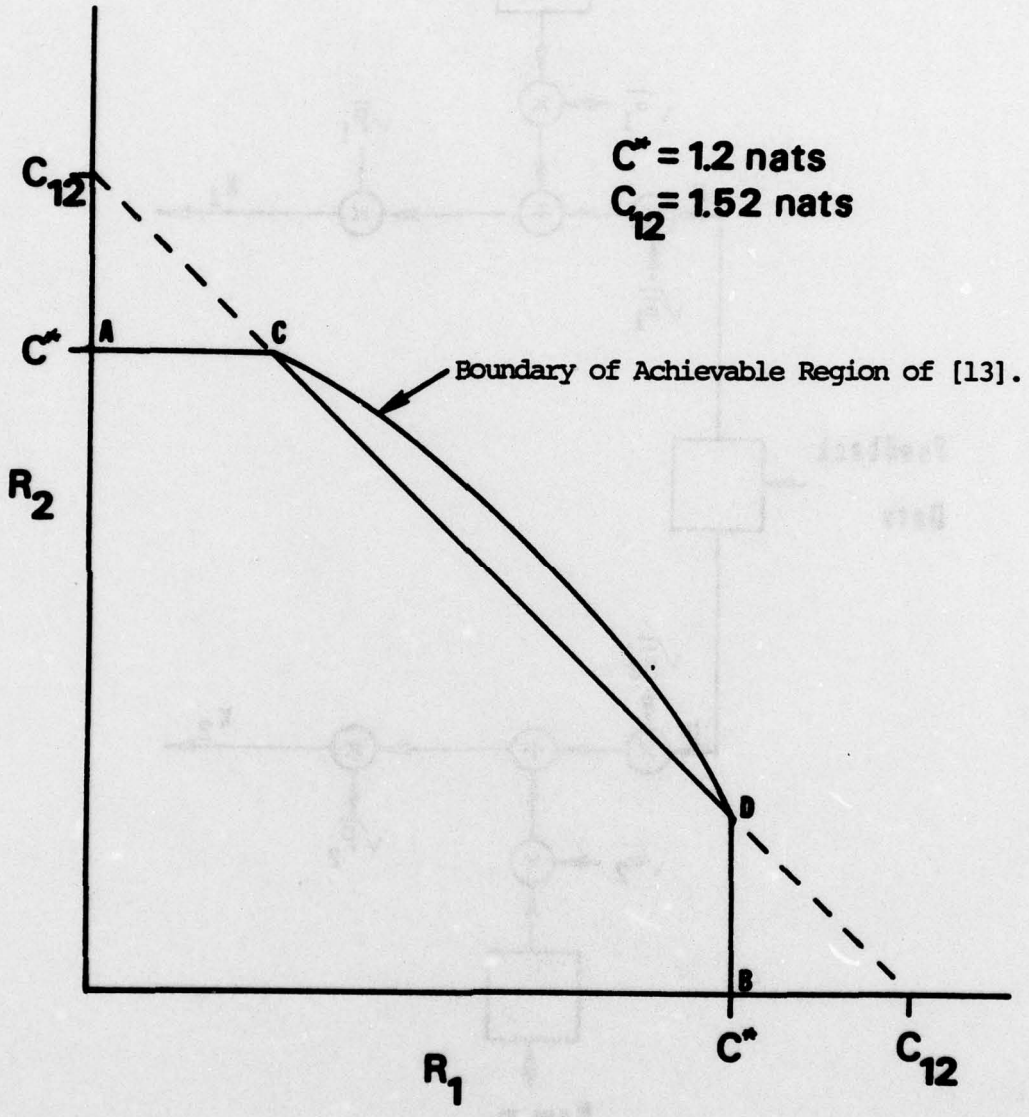
$$R = .7925 \quad \text{bits/transmission} \tag{1.2.14}$$

The result obtained by Cover and Leung is thus very close to optimum for this channel.

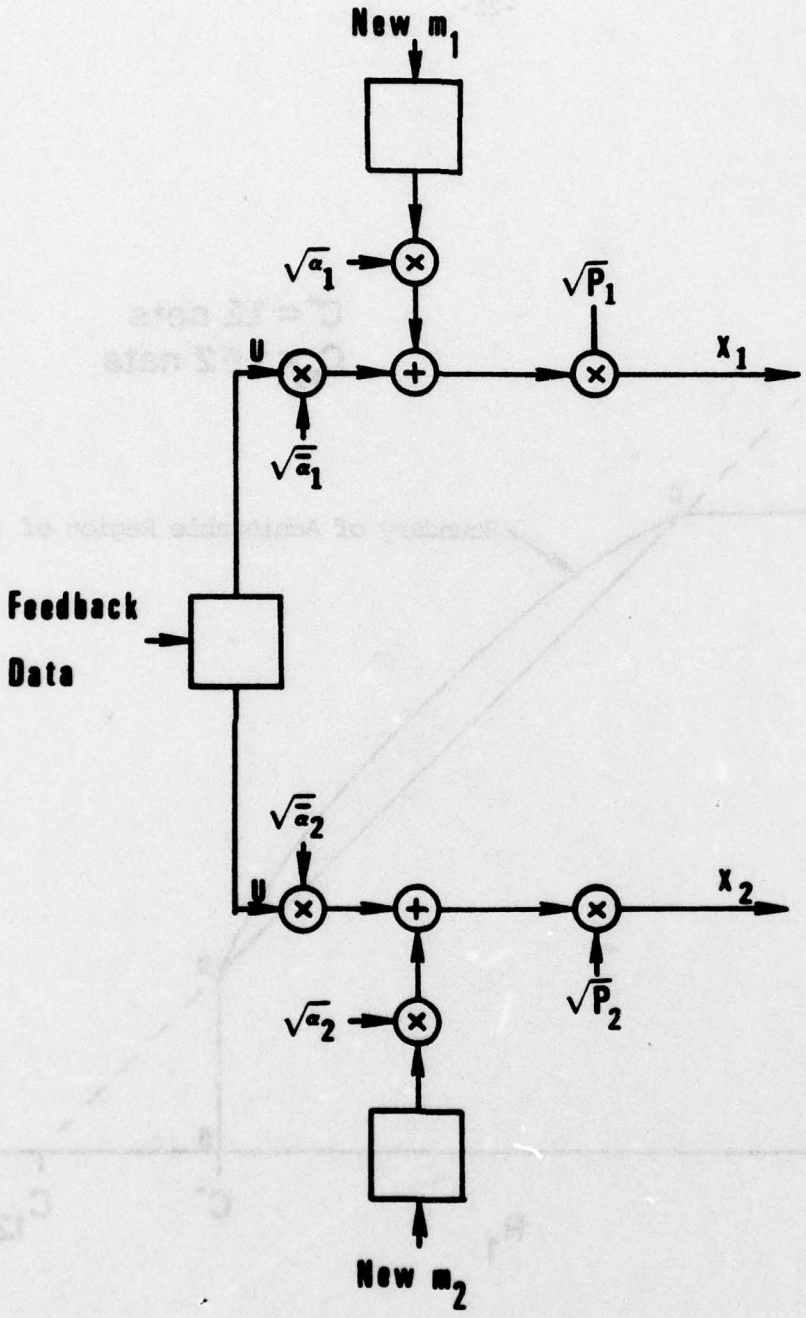
In [13], Cover and Leung found an achievable region for the AWGN MAC with feeedback analogous to that specified by (1.2.7), given by the set of all (R_1,R_2) such that

$$\begin{aligned} R_1 &\leq \frac{1}{2} \ln \left(1 + \frac{\alpha_1 P_1}{\sigma^2} \right) \\ R_2 &\leq \frac{1}{2} \ln \left(1 + \frac{\alpha_2 P_2}{\sigma^2} \right) \quad \text{nats/transmission} \\ R_1 + R_2 &\leq \frac{1}{2} \ln \left(1 + \frac{P_1+P_2+2\sqrt{\bar{\alpha}_1\bar{\alpha}_2 P_1 P_2}}{\sigma^2} \right) \end{aligned} \tag{1.2.15}$$

where $0 \leq \alpha_i \leq 1$ and $\bar{\alpha}_i = 1-\alpha_i$ for $i = 1,2$. In figure 1.3 the region described by (1.2.15) is superimposed on the non-feedback capacity region of figure 1.2. The region given by (1.2.15) is formally identical to that given by (1.2.7) when all the random variables have density functions, power constraints are imposed on the inputs, and a "test encoder" of the form shown in figure 1.4 is employed. This encoder yields the mutual informations appearing in (1.2.15), and the joint probability density function of (U,X_1,X_2,Y) is of the form given in (1.2.12).



1.3 Achievable Region of Cover and Leung for the AWGN MAC with Feedback for $P_1 = P_2 = 10$, $\sigma^2 = 1$.



1.4 Test Encoder for Feedback Coding Scheme of Cover and Leung.

We now present a heuristic discussion of the coding approach of [13] and of the encoder of figure 1.4. The essence of the coding approach is that during a code block transmitter 1 (T_1) uses a fraction α_1 of his power to transmit a "new" message. T_2 independently chooses a new message, transmitted by a fraction α_2 of his power. Each uses the remaining $\bar{\alpha}_i = 1-\alpha_i$ of his power to transmit information known to both transmitters, which is derived from past data. This information is represented by random variable U in fig. 1.4. Using the feedback link, T_2 can both perceive what the receiver has, and decode the new code word sent by T_1. Since U is known by both transmitters and T_2 knows his own transmissions, the rate at which he can reliably decode T_1's new message is simply the mutual information between T_1's input and the receiver's output (which is also T_2's output), given U and X_2. The same reasoning applies with T_1 and T_2 exchanged. Requiring then that each transmitter decode the other's new message, we have that

$$R_i \leq \frac{1}{2} \ln \left(1 + \frac{\alpha_i P_i}{\sigma^2}\right) \qquad (1.2.16)$$

The receiver, however, is rather more confused at this point, as it knows neither U nor either of the new code words a priori. We now define U to be an encoded signal intended to resolve the receiver's uncertainty about the "new" messages of the previous block. Since, by the reasoning of the previous paragraph, each transmitter at the end of a block knows <u>both</u> codewords and the state of the receiver's knowledge, the transmitters can cooperate to coherently resolve the receiver's residual uncertainty about both previous messages. Random variable U denotes this encoded data.

Now this joint codeword is designed to be reliably decodable in the presence of channel noise, and while treating the encoded new data as noise. Therefore U can be decoded first by the receiver and its effect removed, and then the receiver can resolve as much uncertainty about the new messages as possible from the remaining signal.

We can now obtain the last part of equation (1.2.15) (the first part is just (1.2.16)). We note that an upper bound on the amount of uncertainty about a message resolved by a single use of a gaussian channel is given by $\frac{1}{2} \ln(\frac{P_T}{\sigma_e^2})$, where P_T is the total received power, and σ_e^2 is the power of that portion of the received signal which is independent of the message. This is the standard capacity result for gaussian noise channels, with the change that "unwanted" transmitted data are combined with channel noise. We now consider a block of data and the two new messages associated with it. Since that portion of the received signal which corrects old data is decoded and removed, the net signal received is just channel noise plus the two new codewords, with powers $\alpha_1 P_1$ and $\alpha_2 P_2$. Since these variables are independent of each other, the total energy received is just

$$P_T = \alpha_1 P_1 + \alpha_2 P_2 + \sigma^2 \qquad (1.2.17)$$

of which σ^2 is independent of the data. Therefore the uncertainty H_1 removed at the receiver in this stage is bounded by

$$H_1 \leq \frac{1}{2} \ln(1+\frac{\alpha_1 P_1+\alpha_2 P_2}{\sigma^2}) \qquad (1.2.18)$$

Further uncertainty is resolved by the U transmissions in the next block. Since the U data is decoded while treating the superimposed codewords as noise the effective noise for this code is given by

$$\sigma_e^2 = \sigma^2 + \alpha_1 P_1 + \alpha_2 P_2 \tag{1.2.19}$$

The effective total power is just the total received power given by

$$P_T = \alpha_1 P_1 + \alpha_2 P_2 + (\sqrt{\alpha_1}\sqrt{P_1} + \sqrt{\alpha_2}\sqrt{P_2})^2 + \sigma^2 \tag{1.2.20a}$$

$$= P_1 + P_2 + 2\sqrt{\alpha_1\alpha_2}\sqrt{P_1 P_2} + \sigma^2 \tag{1.2.20b}$$

where the first two terms in (1.2.20a) are due to new data, the third term is due to coherent transmission of U, and the fourth term is additive noise. The amount of uncertainty resolved, H_2, is then bounded by the quantity

$$H_2 \leq \frac{1}{2} \ln \frac{P_T}{\sigma_e^2} = \frac{1}{2} \ln \frac{P_1 + P_2 + 2\sqrt{\alpha_1\alpha_2}\sqrt{P_1 P_2} + \sigma^2}{\sigma^2 + \alpha_1 P_1 + \alpha_2 P_2} \tag{1.2.21}$$

The total entropy about the original two messages which can be resolved at the receiver, which bounds the total data rate is then

$$R_1 + R_2 \leq H_1 + H_2 \leq \frac{1}{2} \ln\left(1 + \frac{P_1 + P_2 + 2\sqrt{\alpha_1\alpha_2 P_1 P_2}}{\sigma^2}\right) \tag{1.2.22}$$

which is the last part of(1.2.15)

A rigorous proof of the achievability of this region, based on typical sequence arguments, is provided in [13].

An outer bound to the capacity region for MAC's with feedback can be obtained by generalizing equation (1.2.2) to encompass all joint distributions on X_1, X_2 instead of just statistically independent distributions. Statistical dependence between channel letters is made possible in the feedback case since the transmitters have the common variable represented by the feedback data available, on which to base their new transmitted letters. The authors of [13] exploited this possibility implicitly by having the transmitters partially cooperate in transmitting their joint message. The region obtained, while larger than the non-feedback region, is strictly smaller than the outer bound.

In general, it is not clear how the transmitters can cooperate to achieve the outer bound, or whether it is possible in all cases. In Chapter 2 below, we present a method for the AWGN case which actually achieves the outer bound.

1.2.2 The Broadcast Channel

The broadcast channel was introduced by Cover [4]. Prior to the appearance of [4] communication with more than one receiver was generally considered in the context of time or frequency division multiplexing, in which the transmitter allots part of its total power to each of any number of receivers by dividing its power, in either time or frequency, between signals intended for each receiver. The nature of the results attainable using this approach is discussed at length in [5] where it is shown that the superposition approach introduced in [4] allows simultaneous communication at rates outside the region allowed by various

sharing strategies. An exception to the use of time sharing in the earlier literature is for the case of two users communicating with each other over a common channel, discussed in a quite general setting by Shannon [6].

Figure 1.5a) shows the general two user broadcast channel, with dashed lines representing feedback links. Figure 1.5b) is the special case of physically degraded channels. In this dissertation we treat the problem where the channels are discrete-time additive white gaussian noise channels, with an average power constraint at the transmitter.

In order to obtain a simple characterization of the capacity region of a broadcast channel, it has so far (except for special cases, e.g. [27]) been necessary to be able to compare the channels with outputs Y and Z in some way. Identifying a channel with its output, we say that Y and Z satisfy one of the following relations if the corresponding probabilistic or information theoretic relationship is satisfied. In all cases below, U and X are random variables for which

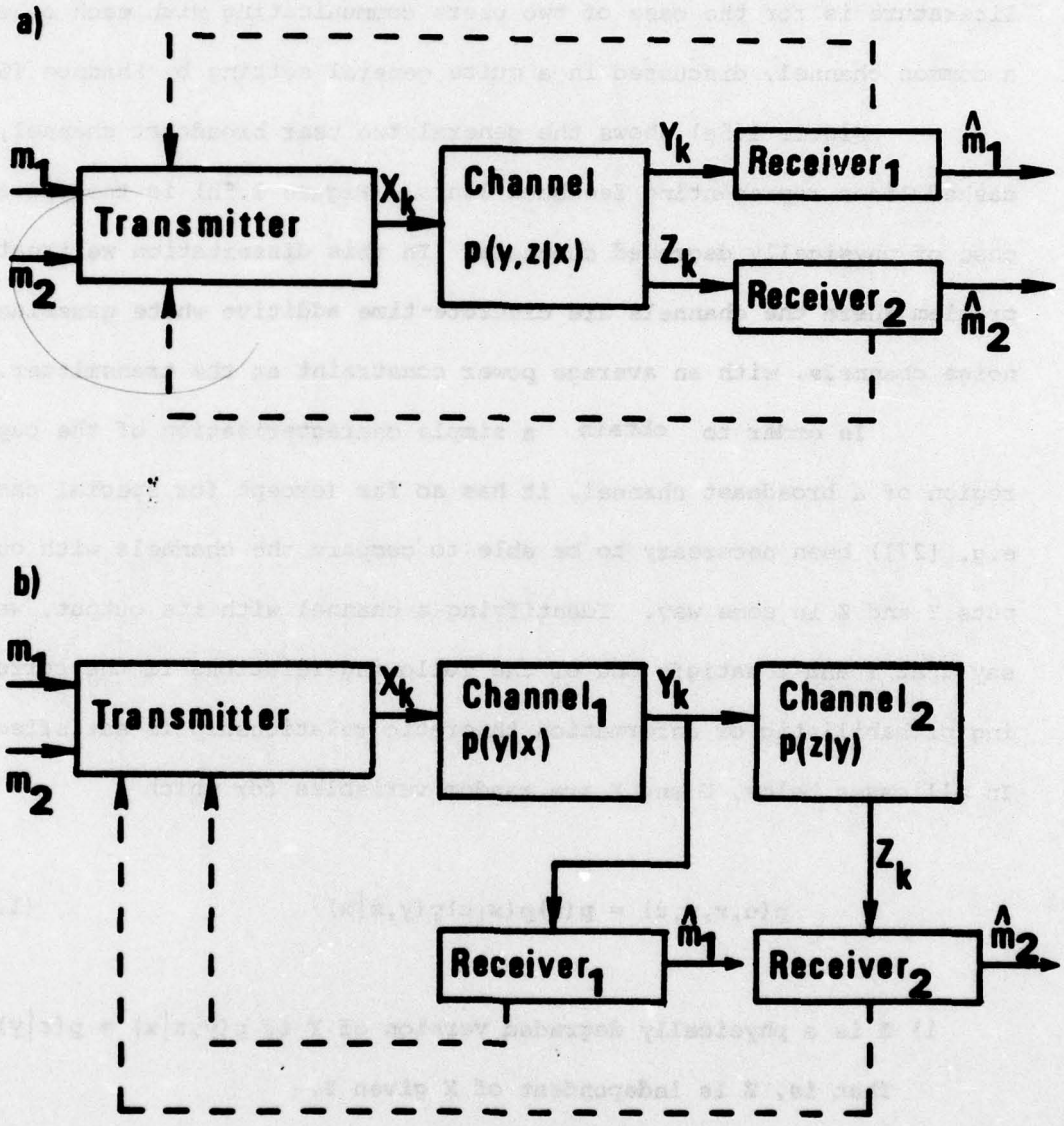
$$p(u,x,y,z) = p(u)p(x|u)p(y,z|x) \tag{1.2.23}$$

i) Z is a physically degraded version of Y if $p(y,z|x) = p(z|y)p(y|x)$. That is, Z is independent of X given Y.

ii) Z is a degraded version of Y if $p(z|x) = \sum_{y} p(z|y)p(y|x)$.

iii) Z is more noisy than Y if $I(U;Z) \leq I(U;Y)$.

iv) Z is less capable than Y if $I(X;Z) \leq I(X;Y)$.



1.5 Broadcast Channel with Two Receivers.
a) General Case
b) Physically Degraded Case

Note that ii) is satisfied if and only if there exists a channel of type i) with the same marginal conditional probabilities $p(y|x)$ and $p(z|x)$.

Each successive ordering can be shown to be strictly weaker than the one preceeding it [1].

The capacity region for each of these orderings has been found in the absence of feedback. For orderings i) and ii) (forward part by Bergmans [28] converse by Gallager [30]) and iii) (Korner and Marton [18]) the capacity region is characterized by the set of all rate triples (R_0,R_1,R_2) where R_0 corresponds to a common message intended for both receivers, R_1 is intended for Y and R_2 is intended for Z, such that

$$\begin{aligned} R_0 + R_2 &\leq I(U;Z) \\ R_1 &\leq I(X;Y|U) \end{aligned} \tag{1.2.24}$$

where $p(u,x,y,z)$ satisfies (1.2.23).

Recently El-Gamal [14] has found the capacity region for ordering iv), given by the set of triples such that

$$\begin{aligned} R_0 + R_1 + R_2 &\leq I(X;Y) \\ R_0 + R_1 + R_2 &\leq I(X;Y|U) + I(U;Z) \\ R_0 + R_2 &\leq I(U;Z) \end{aligned} \tag{1.2.25}$$

The above characterizations exploit the fact that Y is better than Z in some sense and that receiver 1 can use superposition by decoding

the message intended for receiver 2, and then decoding his own message. This approach also leads to the definition of "degraded" message sets (Korner and Marton [33]), for which receiver 2's message is required to be decodable by receiver 1, in which case the capacity region (for a general BC) is given by

$$\begin{aligned} R_1 &\leq I(X;Y|U) \\ R_0 &\leq \min(I(U;Y), I(U;Z)) \end{aligned} \tag{1.2.26}$$

Of course under this restriction R_2 might as well be zero.

The capacity region for the AWGN BC has also been found. Cover [4] found an achievable region given by the set of all rates such that

$$\begin{aligned} R_1 &\leq \frac{1}{2} \ln \left(1 + \frac{\alpha P}{\sigma_1^2}\right) \\ R_2 &\leq \frac{1}{2} \ln \left(1 + \frac{\bar{\alpha} P}{\alpha P + \sigma_2^2}\right) \end{aligned} \quad \text{nats/transmission} \tag{1.2.27}$$

where P is the power constraint, σ_i^2 is the additive noise variance at receiver i ($\sigma_2^2 \geq \sigma_1^2$), α is between 0 and 1, and $\bar{\alpha} = 1 - \alpha$. Bergmans [7] proved the converse. Of course, by extending the notion of degradedness appropriately, using integrals instead of matrix product, it is true that all AWGN BC's are degraded. As in the case of the MAC, the capacity region for the AWGN BC is formally identical to that for discrete BC's.

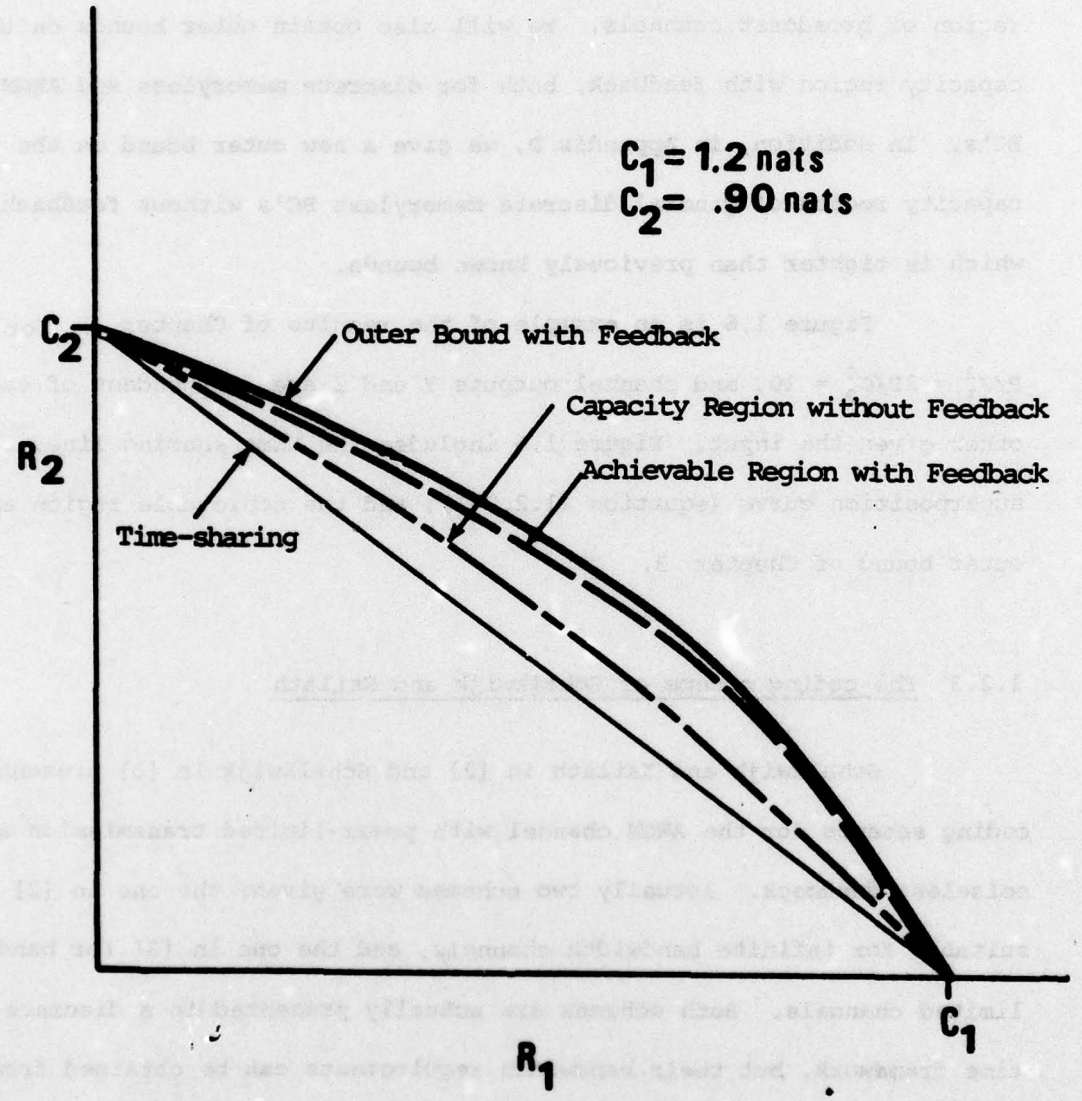
The only previously known feedback result for BC's is El-Gamal's that feedback does not increase the capacity of physically degraded BC's (discrete memoryless [8] and AWGN [16]). In Chapter 3 we present a

constructive coding scheme for AWGN BC's which enlarges the achievable region for cases where the degradedness is non-physical. This is the first demonstration of the fact that feedback can enlarge the capacity region of broadcast channels. We will also obtain outer bounds on the capacity region with feedback, both for discrete memoryless and AWGN BC's. In addition, in Appendix D, we give a new outer bound on the capacity region of general discrete memoryless BC's without feedback which is tighter than previously known bounds.

Figure 1.6 is an example of the results of Chapter 3, for $P/\sigma_1^2 = 2P/\sigma_2^2 = 10$, and channel outputs Y and Z are independent of each other given the input. Figure 1.6 includes the time sharing line, the superposition curve (equation (1.2.27)), and the achievable region and outer bound of Chapter 3.

1.2.3 The coding scheme of Schalkwijk and Kailath

Schalkwijk and Kailath in [2] and Schalkwijk in [3] presented coding schemes for the AWGN channel with power-limited transmission and noiseless feedback. Actually two schemes were given; the one in [2] is suitable for infinite bandwidth channels, and the one in [3] for band-limited channels. Both schemes are actually presented in a discrete-time framework, but their bandwidth requirements can be obtained from sampling theorem considerations. The scheme discussed here is that of [3], since in the infinite bandwidth case, simultaneous communication at capacity can be achieved for the MAC, and time-sharing generates the entire capacity region for the broadcast channel.



1.6 Capacity Region without Feedback and Achievable Region and Outer Bound with Feedback for P = 10, $\sigma_1^2 = 1$ and $\sigma_2^2 = 2$.

We consider the problem of communicating a random variable θ using average-power limited signals over a discrete-time AWGN channel with instantaneous noiseless feedback. This communication model is sketched in Fig. 1.7. Although the noiseless assumption is necessary, feedback with delay can be incorporated with slight modifications. We will communicate θ via a block of N channel transmissions $\{t_k\}^{N-1}$. Assume that after the k^{th} transmission the receiver has an estimate of θ given by

$$\hat{\theta}_k = \theta + \xi_k \tag{1.2.28}$$

where ξ_k is a zero mean gaussian r.v. with variance a_k. Assume further that the transmitter knows ξ_k. Then, at the $k+1^{st}$ time, the transmitter sends

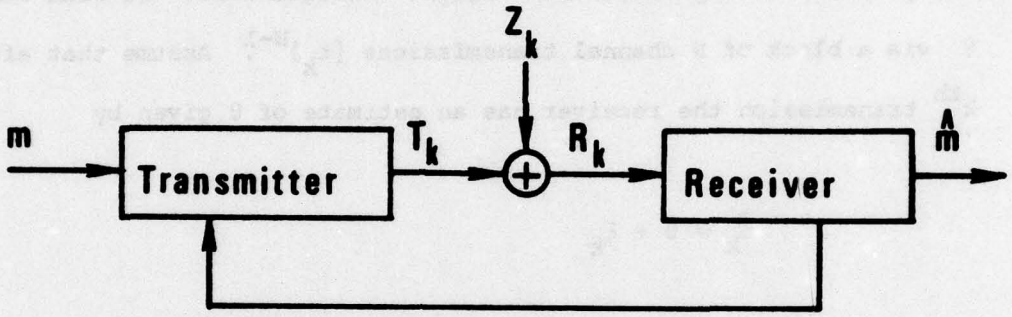
$$t_{k+1} = \sqrt{\frac{P}{a_k}}\,\xi_k \tag{1.2.29}$$

Note that since $\overline{t^2}_{k+1} = P$, the power constraint is satisfied.

The receiver's $k+1^{st}$ channel output is t_{k+1} corrupted by an independent additive zero mean gaussian r.v. with variance σ^2, which we call z_{k+1}. Then denoting the received value by r_{k+1}

$$r_{k+1} = t_{k+1} + z_{k+1} \tag{1.2.30}$$

We have assumed that r_{k+1} is instantaneously available to the transmitter, so that both transmitter and receiver can form the receiver's



1.7 Single-Source Single-Destination Communication with Additive Noise and Feedback.

new estimate, given by

$$\hat{\theta}_{k+1} = \hat{\theta}_k - \frac{\overline{r_{k+1}\xi_k}}{\overline{r_{k+1}^2}} r_{k+1}$$

$$= \theta + \xi_{k+1} \qquad (1.2.31)$$

$\hat{\theta}_{k+1}$ is readily shown to be the maximum likelihood estimate of θ given $\hat{\theta}_k$ and r_{k+1}. From (1.2.28) and (1.2.31) we see that ξ_{k+1} is a linear combination of ξ_k and r_{k+1}, and is available to the transmitter for the next transmission.

To compute the variance of ξ_{k+1}, write

$$a_{k+1} = \overline{\xi_{k+1}^2} = \overline{(\xi_k - \frac{\overline{r_{k+1}\xi_k}}{\overline{r_{k+1}^2}} r_{k+1})^2}$$

$$= a_k - \frac{\overline{r_{k+1}\xi_k}^2}{\overline{r_{k+1}^2}} \qquad (1.2.32)$$

Now

$$r_{k+1} = \sqrt{\frac{P}{a_k}}\, \xi_k + z_{k+1} \qquad (1.2.33)$$

$$\overline{r_{k+1}\xi_k} = \sqrt{\frac{P}{a_k}}\, a_k = \sqrt{P}\sqrt{a_k} \qquad (1.2.34)$$

and

$$\overline{r_{k+1}^2} = \overline{t_{k+1}^2} + \overline{z_{k+1}^2} \qquad (1.2.35)$$

$$= P + \sigma^2$$

Therefore

$$a_{k+1} = a_k - \frac{a_k P}{P+\sigma^2} = a_k \frac{\sigma^2}{P+\sigma^2}$$

$$= a_0 \left(\frac{\sigma^2}{P+\sigma^2} \right)^{k+1} \tag{1.2.36}$$

To compute the performance of this scheme, assume that θ can take on any one of $\|\theta\|$ equally spaced values in the interval $[-\frac{1}{2},\frac{1}{2}]$. Then the spacing between adjacent values of θ is given by

$$\Delta = \frac{1}{\|\theta\| - 1} \tag{1.2.37}$$

After N iterations (a code block), the receiver's estimate is given by

$$\hat{\theta}_{N-1} = \theta + \xi_{N-1} \tag{1.2.38}$$

and the message is decoded as the closest allowable value of θ.

An error is made if ξ_{N-1} is sufficiently large that an incorrect value of θ is closer to the estimate than the true one. This can occur only if

$$|\xi_{N-1}| > \frac{\Delta}{2} \tag{1.2.39}$$

so that the probability of error, P_e, is bounded by

$$P_e \leq \Pr\left[|\xi_{N-1}| > \frac{\Delta}{2} \right] \tag{1.2.40}$$

The inequality comes from the fact that when the true value of θ is at an end point, noise which drives $\hat{\theta}$ out of $[-\frac{1}{2},\frac{1}{2}]$ does not cause an error.

Since ξ_{N-1} is a sum of jointly gaussian r.v.'s, it is gaussian. Therefore

$$P_e \leq 2Q\left(\frac{\Delta}{2\sqrt{\overline{\xi_{N-1}^2}}}\right) \tag{1.2.41}$$

where $Q(x)$ is the tail of the gaussian pdf, equal to $\frac{1}{\sqrt{2\pi}}\int_x^\infty e^{-y^2/2}dy$

Substituting (1.2.36) into (1.2.41) yields

$$P_e = 2Q\left(\frac{1}{2(\|\theta\|-1)\sqrt{a_0}}\left(\frac{P+\sigma^2}{\sigma^2}\right)^{\frac{N-1}{2}}\right) \tag{1.2.42}$$

Since N total iterations are used, if we define R to be the transmission rate in nats/channel use, then

$$\|\theta\| = e^{NR} \tag{1.2.43}$$

In addition the capacity of the channel is given by the well-known result

$$C = \frac{1}{2}\ln\left(\frac{P+\sigma^2}{\sigma^2}\right) \qquad \text{nats/ transmission} \tag{1.2.44}$$

Substituting (1.2.43) in (1.2.42) we obtain

$$P_e \leq 2Q\left(\frac{1}{2\sqrt{a_0}\left(\frac{P+\sigma^2}{\sigma^2}\right)^{1/2}}e^{N(C-R)}\right) \tag{1.2.45}$$

(We have replaced $e^{NR}-1$ by e^{NR} which decreases the argument. Since $Q(\cdot)$ is a monotonic decreasing function, this upper bounds P_e.)

For the initial transmission, θ is scaled up to the average transmitted power and sent without coding. That is,

$$t_0 = \sqrt{\frac{P}{\overline{\theta^2}}}\,\theta \tag{1.2.46}$$

Therefore

$$r_0 = \sqrt{\frac{P}{\overline{\theta^2}}}\,\theta + z_0 \tag{1.2.47}$$

and

$$\hat{\theta}_0 = \theta + \sqrt{\frac{\overline{\theta^2}}{P}}\; z_0 = \theta + \xi_0 \tag{1.2.48}$$

$$a_0 = E(\xi_0^2) = \frac{\overline{\theta^2}}{P}\sigma^2 \tag{1.2.49}$$

Substituting in (1.2.45) we obtain

$$P_e \leq 2Q\left(\frac{1}{2\sqrt{\overline{\theta^2}}}\sqrt{\frac{P}{P+\sigma^2}}\; e^{N(C-R)}\right) \tag{1.2.50}$$

For any R < C the argument can be made arbitrarily large by increasing N. Thus since Q(x) decreases as x increases, the error probability can be made arbitrarily small. Furthermore, since $Q(x) \sim e^{-x^2/2}$, and $\overline{\theta^2} \simeq \frac{1}{12}$ for N large, then

$$P_e \cong \exp\left(-\frac{3}{2}\,\frac{P}{P+\sigma^2}e^{2N(C-R)}\right) \tag{1.2.51}$$

exhibiting a doubly exponential decrease with block length. This is in contrast with non-feedback codes, whose performance is typically singly exponential in block length.

An additional point worth mentioning is that this scheme works for arbitrary zero mean finite variance noise distributions, as long as the noise is uncorrelated in time, and independent of the data. The equations giving the noise variances are valid regardless of the gaussian assumption, although the expressions for error probability are no longer necessarily true. In [2] it is shown that for the infinite bandwidth channel, the scheme analogous to the one used here does yield an asymptotically gaussian noise term, hence doubly exponential behavior. No such claim is made for the finite bandwidth scheme of [3] (i.e. the one discussed here). Singly exponential decay can be demonstrated as follows.

$$P_e = \Pr[|\xi_{N-1}| > \frac{\Delta}{2}] \tag{1.2.52}$$

By Chebyshev's inequality

$$\begin{aligned} \Pr[|\xi_{N-1}| > \frac{\Delta}{2}] &\leq \frac{4}{\Delta^2}\overline{\xi^2}_{N-1} \\ &\simeq 4\|\theta\|^2 a_0 \left(\frac{\sigma^2}{P+\sigma^2}\right)^{N-1} \\ &\simeq \frac{P+\sigma^2}{3P} e^{-2N(C-R)} \end{aligned} \tag{1.2.53}$$

Therefore, as long as $R < C$, P_e can be made arbitrarily small. Of course, the capacity for non-gaussian channels is in general larger than that for gaussian channels with the same variance, so the set of rates achieved by this scheme is sub-optimum for all but gaussian channels.

CHAPTER 2

AWGN MAC WITH FEEDBACK

In this chapter the capacity region of the additive white gaussian noise MAC with feedback (fig. 2.1) is determined. The forward part of the result will be proved constructively, i.e. a deterministic coding scheme (similar to the one in Section 1.2.3) will be presented in Section 2.1, yielding an achievable region. In Section 2.2 an outer bound to the capacity region is found by means of a weak converse. In Section 2.3 the two regions are shown to be equal.

The result obtained is the following.

Theorem 2.1. Define

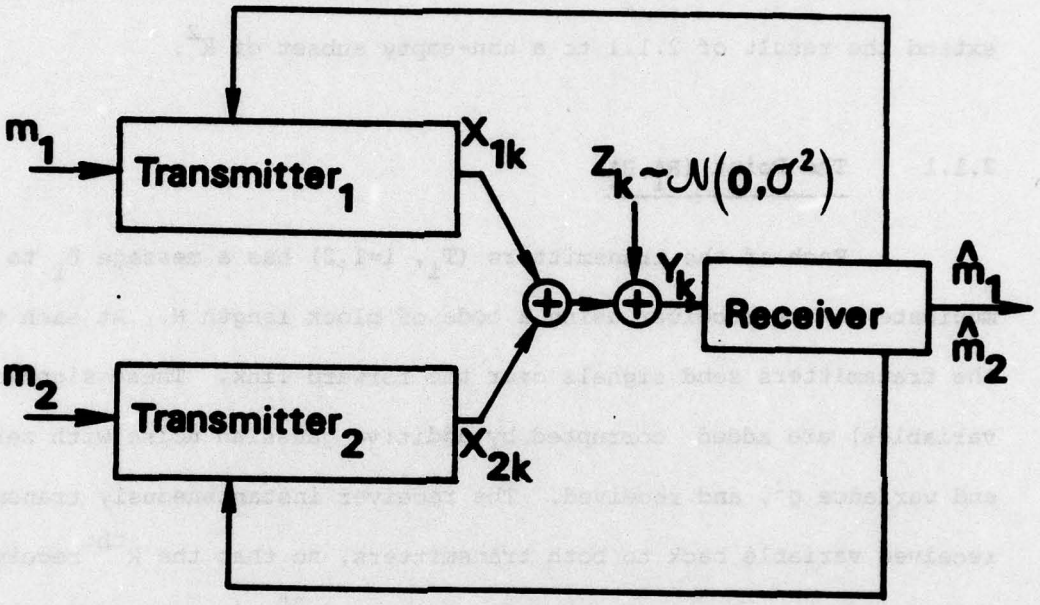
$$C = \bigcup_{0\le\rho\le 1} \{(R_1,R_2): R_1 \le \frac{1}{2}\ln(1+\frac{P_1}{\sigma^2}(1-\rho^2)) \qquad R_2 \le \frac{1}{2}\ln(1+\frac{P_2}{\sigma^2}(1-\rho^2)) \qquad R_1+R_2 \le \frac{1}{2}\ln(1+\frac{P_1+P_2+2\sqrt{P_1P_2}|\rho|}{\sigma^2})\} \tag{2.1.1}$$

C is the capacity region of the AWGN MAC with feedback, where transmitter 1 and 2's signals $\{x_{ik}\}$ $(k = 1,\ldots,N,\ i = 1,2)$ satisfy

$$\sum_{k=1}^{N} \overline{x_{1k}^2} \le NP_1 \tag{2.1.2a}$$

$$\sum_{k=1}^{N} \overline{x_{2k}^2} \le NP_2 \tag{2.1.2b}$$

and the additive noise has variance σ^2.



2.1 AWGN MAC with Feedback.

2.1 Achievable Region

In Section 2.1.1 we demonstrate the achievability of a certain rate point denoted (R_1^*,R_2^*). In 2.1.2, superposition arguments are used to extend the result of 2.1.1 to a non-empty subset of R^2.

2.1.1 The Point (R_1^*,R_2^*)

Each of the transmitters (T_i, i=1,2) has a message θ_i to communicate to the receiver using a code of block length N. At each time k the transmitters send signals over the forward link. These signals (real variables) are added, corrupted by additive gaussian noise with zero mean and variance σ^2, and received. The receiver instantaneously transmits his received variable back to both transmitters, so that the k^{th} received variable y_k may be used in the encoding of the $k+1^{st}$ transmitted data $x_{1,k+1}$ and $x_{2,k+1}$.

We assume that after k-1 channel uses the receiver has estimates of θ_1 and θ_2 of the form

$$\hat{\theta}_1^{k-1} = \theta_1 + \xi_{k-1}$$
$$\hat{\theta}_2^{k-1} = \theta_2 + \eta_{k-1} \qquad (2.1.3)$$

where ξ_{k-1} and η_{k-1} are jointly distributed gaussian random variables, with means zero, variances a_{k-1} and b_{k-1} respectively, and correlation coefficient ρ_{k-1}. Further assume that the estimates are formed from linear

combinations of previously received data. Since both transmitters know all of the previous received data, then T_i knows $\hat{\theta}_i^{k-1}$. Since he also knows θ_i, then T_1 knows ξ_{k-1} and T_2 knows η_{k-1}. Each also knows the covariance parameters, since these are statistical averages.

At the k^{th} step each transmitter would like to amplify his correction term (i.e. ξ_{k-1} for T_1 and η_{k-1} for T_2) and transmit it to the receiver as in the scheme of Section 1.2.3. In general these terms may be negatively correlated, so that when added, they can destructively interfere. However, since all parties involved know ρ_{k-1}, one transmitter can alter the sign of his transmission, so that the transmitted signals are positively correlated and thus constructively interfere. The receiver can then use its knowledge of ρ_{k-1} to compensate for this sign change.

We require then that

$$x_{1k} = \sqrt{\frac{P_I}{a_{k-1}}}\ \xi_{k-1}$$

$$x_{2k} = \sqrt{\frac{P_2}{b_{k-1}}}\ \mathrm{sgn}(\rho_{k-1})\ \eta_{k-1} \qquad (2.1.4)$$

The received data at the k^{th} step is then

$$r_k = x_{1k} + x_{2k} + z_k \qquad (2.1.5)$$

where z_k is the new AWGN term with variance σ^2, independent of all previous data.

Based on the received data and his previous estimates, the receiver forms new estimates of θ_1 and θ_2 given by

$$\begin{aligned}\hat{\theta}_1^k &= \theta_1 + \xi_k \\ &= \hat{\theta}_1^{k-1} - \frac{\overline{r_k \xi_{k-1}}}{\overline{r_k^2}} r_k \\ &= \theta_1 + \xi_{k-1} - \frac{\overline{r_k \xi_{k-1}}}{\overline{r_k^2}} r_k\end{aligned} \tag{2.1.6}$$

Similarly,

$$\hat{\theta}_2^k = \theta_2 + \eta_{k-1} - \frac{\overline{r_k \eta_{k-1}}}{\overline{r_k^2}} r_k \tag{2.1.7}$$

It is easily verified that

$$\begin{aligned}\overline{x_{ik}^2} &= P_i \quad i=1,2 \\ \overline{r_k^2} &= P_1 + P_2 + 2\sqrt{P_1 P_2}|\rho_{k-1}| + \sigma^2 \\ \overline{r_k \xi_{k-1}} &= \sqrt{a_{k-1}}\,(\sqrt{P_1} + \sqrt{P_2}|\rho_{k-1}|) \\ \overline{r_k \eta_{k-1}} &= \sqrt{b_{k-1}}\,\mathrm{sgn}(\rho_{k-1})\,(\sqrt{P_2} + \sqrt{P_1}|\rho_{k-1}|)\end{aligned} \tag{2.1.8}$$

By defining $a_k = \overline{\xi_k^2}$ and $b_k = \overline{\eta_k^2}$ and substituting, we obtain

$$a_k = a_{k-1} \frac{\sigma^2+P_2(1-\rho_{k-1}^2)}{\sigma^2+P_1+P_2+2\sqrt{P_1P_2}|\rho_{k-1}|} \tag{2.1.9a}$$

$$b_k = b_{k-1} \frac{\sigma^2+P_1(1-\rho_{k-1}^2)}{\sigma^2+P_1+P_2+2\sqrt{P_1P_2}|\rho_{k-1}|} \tag{2.1.9b}$$

Similarly, define $c_k = \overline{\xi_k \eta_k}$, and we obtain

$$c_k = c_{k-1} - \frac{\overline{r_k \xi_{k-1}}\ \overline{r_k \eta_{k-1}}}{\overline{r_k^2}} \tag{2.1.9c}$$

By substituting (2.1.8), noting that $\rho_k = \frac{c_k}{\sqrt{a_k b_k}}$, and substituting (2.1.9a) and (2.1.9b) into (2.1.9c) we obtain

$$\rho_k = \frac{\sigma^2\rho_{k-1} - \mathrm{sgn}(\rho_{k-1})\sqrt{P_1P_2}(1-\rho_{k-1}^2)}{\sqrt{\sigma^2+P_1(1-\rho_{k-1}^2)}\ \sqrt{\sigma^2+P_2(1-\rho_{k-1}^2)}} \tag{2.1.10}$$

Let us assume that $\rho_{k-1} = \rho$, where $\rho \in (0,1)$ and ρ solves the quartic equation

$$P(\rho) = \sigma^2(\sigma^2+P_1+P_2+2\sqrt{P_1P_2}\,\rho) - (\sigma^2+P_1(1-\rho^2))(\sigma^2+P_2(1-\rho^2)) \tag{2.1.11}$$

We will show in Section 2.3 that this equation also specifies the joint distribution of X_1X_2 that achieves a point on the boundary of the

outer bound to the capacity region of the AWGN MAC with feedback, and use this fact to determine the capacity region. In Appendix A it is shown that $P(\rho)$ has exactly one root in (0,1). By squaring (2.1.10) and subtracting from 1, we obtain

$$1- \rho_k^2 = (1 - \rho_{k-1}^2) \frac{\sigma^2(\sigma^2+P_1+P_2+2\sqrt{P_1P_2}|\rho_{k-1}|)}{(\sigma^2+P_1(1-\rho_{k-1}^2))(\sigma^2+P_2(1-\rho_{k-1}^2))} \quad (2.1.12)$$

But the rational term equals one by equation (2.1.11). Hence $\rho_k^2 = \rho_{k-1}^2$, and $|\rho_k| = |\rho_{k-1}|$. In (2.1.10) note that if $0<\rho_{k-1}<1$, then ρ_k is strictly less than ρ_{k-1}, and if $-1<\rho_{k-1}<0$, then ρ_k is strictly greater than ρ_{k-1}. Therefore if $|\rho_{k-1}| \in (0,1)$, then

$$\rho_k = -\rho_{k-1} \quad (2.1.13)$$

The rational term in the iterated version of (2.1.12) is now 1, so that (2.1.13) holds inductively.

We now describe an initialization procedure which can achieve any desired initial correlation ($\neq\pm1$). At the -1^{st} epoch, let T_1 transmit

$$x_{1,-1} = \frac{\alpha\sqrt{2}}{\sqrt{\alpha^2+1}} \frac{\sqrt{P_1}}{\sqrt{\theta_1^2}} \theta_1 \quad (2.1.14a)$$

where α is a number to be determined. Let T_2 transmit

$$x_{2,-1} = \frac{\sqrt{2}}{\sqrt{\alpha^2+1}} \frac{\sqrt{P_2}}{\sqrt{\theta_2^2}} \theta_2 \quad (2.1.14b)$$

It is true, then, that

$$\overline{x^2}_{1,-1} = \frac{2\alpha^2}{\alpha^2+1} P_1 \qquad (2.1.15a)$$

$$\overline{x^2}_{2,-1} = \frac{2}{\alpha^2+1} P_2 \qquad (2.1.15b)$$

At the 0^{th} epoch send

$$x_{1,0} = \frac{\sqrt{2}}{\sqrt{\alpha^2+1}} \frac{\sqrt{P_1}}{\sqrt{\theta_1^2}} \theta_1 \qquad (2.1.16a)$$

and

$$x_{2,0} = \frac{\alpha\sqrt{2}}{\sqrt{\alpha^2+1}} \frac{\sqrt{P_2}}{\sqrt{\theta_2^2}} \theta_2 \qquad (2.1.16b)$$

It is easily verified that

$$\overline{x^2}_{1,-1} + \overline{x^2}_{1,0} = 2P_1 \qquad (2.1.17a)$$

and

$$\overline{x^2}_{2,-1} + \overline{x^2}_{2,0} = 2P_2 \qquad (2.1.17b)$$

so that the average power constraints are satisfied.

Now the -1^{st} and 0^{th} received data are

$$r_{-1} = \frac{\sqrt{2}}{\sqrt{\alpha^2+1}} \left[\alpha \frac{\sqrt{P_1}}{\sqrt{\overline{\theta_1^2}}} \theta_1 + \frac{\sqrt{P_2}}{\sqrt{\overline{\theta_2^2}}} \theta_2 \right] + z_{-1} \tag{2.1.18a}$$

$$r_0 = \frac{\sqrt{2}}{\sqrt{\alpha^2+1}} \left[\frac{\sqrt{P_1}}{\sqrt{\overline{\theta_1^2}}} \theta_1 + \alpha \frac{\sqrt{P_2}}{\sqrt{\overline{\theta_2^2}}} \theta_2 \right] + z_0 \tag{2.1.18b}$$

The receiver can weight and combine r_{-1} and r_0 to form the initial estimates

$$\hat{\theta}_1^0 = \left[\frac{(\alpha - \frac{1}{\alpha}) \sqrt{2P_1}}{\sqrt{\overline{\theta_1^2}}} \frac{}{\sqrt{\alpha^2+1}} \right]^{-1} (r_{-1} - \frac{1}{\alpha} r_0)$$

$$= \theta_1 + \frac{\alpha z_{-1} - z_0}{(\alpha^2-1)\sqrt{2P_1}} \sqrt{\alpha^2+1} \sqrt{\overline{\theta_1^2}} \tag{2.1.19a}$$

and

$$\hat{\theta}_2^0 = \theta_2 + \frac{\alpha z_0 - z_{-1}}{(\alpha^2-1)\sqrt{2P_2}} \sqrt{\alpha^2+1} \sqrt{\overline{\theta_2^2}} \tag{2.1.19b}$$

It can easily be shown that

$$a_0 = \frac{(\alpha^2+1)^2\sigma^2}{(\alpha^2-1)^2 2P_1} \overline{\theta_1^2} \tag{2.1.20a}$$

$$b_0 = \frac{(\alpha^2+1)^2\sigma^2}{(\alpha^2-1)^2 2P_2} \overline{\theta_2^2} \tag{2.1.20b}$$

and

$$c_0 = \frac{-\alpha\sigma^2(\alpha^2+1)}{(\alpha^2-1)^2\sqrt{P_1P_2}} \sqrt{\overline{\theta_1^2}\,\overline{\theta_2^2}} \tag{2.1.20c}$$

Therefore,

$$\rho_0 = \frac{c_0}{\sqrt{a_0 b_0}} = \frac{-2\alpha}{\alpha^2+1} \tag{2.1.21}$$

Any value of ρ_0 in (0,1) can be obtained by setting

$$\alpha = \sqrt{\frac{1}{\rho_0^2} - 1} \; - \frac{1}{\rho_0} \tag{2.1.22}$$

Returning to our problem, we choose α so as to obtain

$$\rho_0 = \rho \tag{2.1.23}$$

where ρ solves $P(\rho)=0$.

Equations (2.1.13), (2.1.9) and (2.1.10) then become

$$\rho_k = (-1)^k \rho \tag{2.1.24}$$

$$a_k = a_{k-1} \frac{\sigma^2 + P_2(1-\rho^2)}{\sigma^2+P_1+P_2+2\sqrt{P_1 P_2}|\rho|}$$

$$= a_0 \left[\frac{\sigma^2 + P_2(1-\rho^2)}{\sigma^2+P_1+P_2+2\sqrt{P_1 P_2}|\rho|}\right]^k \tag{2.1.25a}$$

$$b_k = b_0 \left[\frac{\sigma^2 + P_1(1-\rho^2)}{\sigma^2+P_1+P_2+2\sqrt{P_1 P_2}|\rho|}\right]^k \tag{2.1.25b}$$

Define

$$R_1^* = \frac{1}{2} \ln \frac{\sigma^2+P_1+P_2+2\sqrt{P_1P_2}|\rho|}{\sigma^2 + P_2(1-\rho^2)} \tag{2.1.26a}$$

$$R_2^* = \frac{1}{2} \ln \frac{\sigma^2+P_1+P_2+2\sqrt{P_1P_2}|\rho|}{\sigma^2 + P_1(1-\rho^2)} \tag{2.1.26b}$$

Then

$$a_k = a_0 \, e^{-2kR_1^*} \tag{2.1.27a}$$

$$b_k = b_0 \, e^{-2kR_2^*} \tag{2.1.27b}$$

Proceeding as in Section 1.2.3, let θ_i equiprobably take on one of $||\theta_i||$ values distributed uniformly between $-\frac{1}{2}$ and $\frac{1}{2}$. Then for a block of N transmissions (corresponding to N-2 iterations after the initialization)

$$R_i = \frac{1}{N} \ln ||\theta_i|| \text{ nats/transmission} \tag{2.1.28}$$

The separation between adjacent values of θ_i is

$$\Delta_i = \frac{1}{2(||\theta_i|| - 1)} = \frac{1}{2(e^{NR_i} - 1)} > \frac{1}{2e^{NR_i}} \tag{2.1.29}$$

$P_{e,1}$, the error probability for message 1 is given by

$$P_{e,1} \leq 2Q\left(\frac{1}{2e^{NR_1}} \frac{1}{\sqrt{a_0}} e^{(N-2)R_1^*}\right)$$

$$= 2Q\left(\frac{1}{2e^{2R_1^*}\sqrt{a_0}} e^{N(R_1^*-R_1)}\right) \tag{2.1.30}$$

and similarly

$$P_{e,2} \leq 2Q\left(\frac{1}{2e^{2R_2^*}\sqrt{b_0}} e^{N(R_2^*-R_2)}\right) \tag{2.1.31}$$

Both of these can be made arbitrarily small by increasing N, so long as $R_i < R_i^*$ for i=1,2. These error probabilities, like their counterpart for the single user scheme in Section 1.2.3, decay to zero in a doubly exponential manner with increasing block length.

By way of a numerical example, consider $P_1 = P_2 = 10$, and $\sigma^2 = 1$. Then the equation $P(\rho)=0$ is satisfied by $\rho = .71164$, and

$$R_1^* = R_2^* = .8905 \text{ nats/transmission} \tag{2.1.32}$$

For the scheme described in [13], the largest value of R for which $R_1 = R_2 = R$ is achievable is given by

$$R = \frac{1}{2} \ln\left(2\sqrt{\frac{P}{\sigma^2} + 1} - 1\right) = .8643 \text{ nats/transmission} \tag{2.1.33}$$

The largest achievable value without feedback is (from equation (1.2.3))

$$R = \frac{1}{4} \ln 21 = .7611 \text{ nats/transmission} \qquad (2.1.34)$$

2.1.2 Completion of the Achievable Region

We now extend the result of Section 2.1.1. We will describe a superposition approach which allows communication at all points along a curve between (R_1^*, R_2^*) of Section 2.1.1 and point D of Section 1.2 (see Fig. 1.2). By symmetrical arguments, the point (R_1^*, R_2^*) and point C can be connected. By using the fact that any point in R^2 dominated by an achievable point is in turn achievable, we will have demonstrated the achievability of a region of the form shown in Figure 2.2.

We allow T_1 to choose two independent messages, $\theta_1^{(1)}$ and $\theta_1^{(2)}$. T_1 will transmit $\theta_1^{(1)}$ by means of a code of block length N whose letters are drawn independently from a zero-mean normal distribution with variance $\alpha P_1 (0 \leq \alpha \leq 1)$. T_1 uses the remainder of his power, $\bar{\alpha} P_1 (\bar{\alpha} = 1-\alpha)$ to participate in the scheme of Section 2.1.1 while treating the code letters for $\theta_1^{(1)}$ as noise. T_2 uses all of his power as in Section 2.1.1.

Since the code letters for $\theta_1^{(1)}$ are independent identically distributed $N(0, \alpha P_1)$, the results of the previous section apply for the detection of $\theta_1^{(2)}$ and θ_2, except that σ^2 is replaced by $\sigma^2 + \alpha P_1$, and P_1 is replaced by $\bar{\alpha} P_1$ everywhere. Equation (2.1.26) then becomes

$$R_1^*(\alpha) = \frac{1}{2} \ln \left[\frac{\sigma^2 + \alpha P_1 + \bar{\alpha} P_1 + P_2 + 2\sqrt{\bar{\alpha} P_1 P_2}\,|\rho|}{\sigma^2 + \alpha P_1 + P_2(1-\rho^2)} \right] \qquad (2.1.35a)$$

$$= \frac{1}{2} \ln \left[\frac{\sigma^2+P_1+P_2+2\sqrt{\bar{\alpha}P_1P_2}|\rho|}{\sigma^2+\alpha P_1+P_2(1-\rho^2)}\right]$$

$$R_2^*(\alpha) = \frac{1}{2} \ln \left[\frac{\sigma^2+\alpha P_1+\bar{\alpha}P_1+P_2+2\sqrt{\bar{\alpha}P_1P_2}|\rho|}{\sigma^2+\alpha P_1+\bar{\alpha}P_1(1-\rho^2)}\right] \qquad (2.1.35b)$$

$$= \frac{1}{2} \ln \left[\frac{\sigma^2+P_1+P_2+2\sqrt{\bar{\alpha}P_1P_2}|\rho|}{\sigma^2+P_1(1-\bar{\alpha}\rho^2)}\right]$$

The equation satisfied by $|\rho|$ (replacing (2.1.11)) is now

$$P(\alpha,\rho)=(\sigma^2+\alpha P_1)(\sigma^2+P_1+P_2+2\sqrt{\bar{\alpha}P_1P_2}|\rho|)$$

$$-(\sigma^2+P_1(1-\bar{\alpha}\rho^2))(\sigma^2+\alpha P_1+P_2(1-\rho^2))$$

$$= 0 \qquad (2.1.36)$$

At the end of a block, $\theta_1^{(2)}$ and θ_2 are known with high probability at the receiver, so that all of the correction terms are also known. These terms can be subtracted from the received data (which must be stored), leaving just the codeword for $\theta_1^{(1)}$ and the added channel noise. This code may be decoded reliably at rates up to

$$R_1^{**}(\alpha) = \frac{1}{2} \ln\left(\frac{\sigma^2+\alpha P_1}{\sigma^2}\right) \qquad (2.1.37)$$

The net result is that T_1 can communicate at all rates satisfying

$$R_1 < R_1^*(\alpha) + R_1^{**}(\alpha)$$

$$= \frac{1}{2} \ln \left[\frac{\sigma^2+P_1+P_2+2\sqrt{\alpha P_1 P_2}|\rho|}{\sigma^2+\alpha P_1+P_2(1-\rho^2)}\right] + \frac{1}{2} \ln\left(\frac{\sigma^2+\alpha P_1}{\sigma^2}\right) \tag{2.1.38}$$

and T_2 at rates satisfying

$$R_2 < R_2^*(\alpha)$$

$$= \frac{1}{2} \ln \left[\frac{\sigma^2+P_1+P_2+2\ \alpha P_1 P_2|\rho|}{\sigma^2+P_1(1-\bar{\alpha}\rho^2)}\right] \tag{2.1.39}$$

Since ρ satisfies equation (2.1.36), we can solve for the denominator term in $R_1^*(\alpha)$ and express equation (2.1.38) as

$$R_1 < \frac{1}{2} \ln \left[\left(1+\frac{P_1}{\sigma^2}(1-\bar{\alpha}\rho^2)\right)\right] \tag{2.1.40}$$

Note that when $\alpha = 0$, $P(\alpha,\rho)$ reduces to equation (2.1.11). When $\alpha = 1$, $P(\alpha,\rho)$ becomes

$$P(1,\rho) = (\sigma^2+P_1)(\sigma^2+P_1+P_2)-(\sigma^2+P_1)(\sigma^2+P_1+P_2(1-\rho^2)) \tag{2.1.41}$$

which has the unique solution $\rho = 0$. Denoting the solution of (2.1.11) by ρ^*, noting that $P(\alpha,\rho)$ is continuous in both ρ and α, and is guaranteed (by Appendix A) to have a unique root in (0,1), we see that as α varies between 0 and 1, ρ (and $\bar{\alpha}\rho$) varies continuously between ρ^* and 0, so that the right

hand side of equations (2.1.40) and (2.1.39) vary continuously. At $\alpha = 1$, (2.1.40) and (2.1.39) become

$$R_1 < \frac{1}{2} \ln \left[1 + \frac{P_1}{\sigma^2} \right] \tag{2.1.42a}$$

$$R_2 < \frac{1}{2} \ln \left[1 + \frac{P_2}{\sigma^2 + P_1} \right] \tag{2.1.42b}$$

which is just point D in Figure 1.2.

We can estimate (bound) the overall probability of error for this approach as follows: Let $\varepsilon_1, \varepsilon_2$ be the probabilities that $\theta_1^{(2)}$ and θ_2 are decoded incorrectly. By the results of Section 2.1.1 these probabilities can be driven to zero by making N large. Let ε_3 be the probability that $\theta_1^{(1)}$ is decoded incorrectly given that $\theta_1^{(2)}$ and θ_2 are decoded correctly, which also may be made arbitrarily small. Then by the union bound

$$\Pr[\text{error on } \theta_1^{(1)} \text{ or } \theta_1^{(2)} \text{ or } \theta_2] \leq \Pr[\text{error on } \theta_1^{(1)}]$$

$$+ \Pr[\text{error on } \theta_1^{(2)}] + \Pr[\text{error on } \theta_2] \tag{2.1.43}$$

Now,

$$\Pr[\text{error on } \theta_1^{(1)}] = \Pr[\text{error on } \theta_1^{(1)} | \theta_1^{(2)}, \theta_2 \text{ correct}]$$

$$\Pr[\theta_1^{(2)} \theta_2 \text{ correct}] + \Pr[\text{error on } \theta_1^{(1)} | \theta_1^{(2)} \text{ or } \theta_2 \text{ incorrect}] \times$$

$$\Pr[\theta_1^{(2)} \text{ or } \theta_2 \text{ incorrect}]$$

$$\leq \varepsilon_3(1-\Pr[\theta_1^{(2)} \text{ or } \theta_2 \text{ incorrect}]) + \Pr[\theta_1^{(2)} \text{ or } \theta_2 \text{ incorrect}]$$

$$= \varepsilon_3 + \Pr[\theta_1^{(2)} \text{ or } \theta_2 \text{ incorrect}](1-\varepsilon_3)$$

$$\leq \varepsilon_3 + \Pr[\theta_1^{(2)} \text{ or } \theta_2 \text{ incorrect}]$$

$$\leq \varepsilon_1 + \varepsilon_2 + \varepsilon_3 \qquad (2.1.44)$$

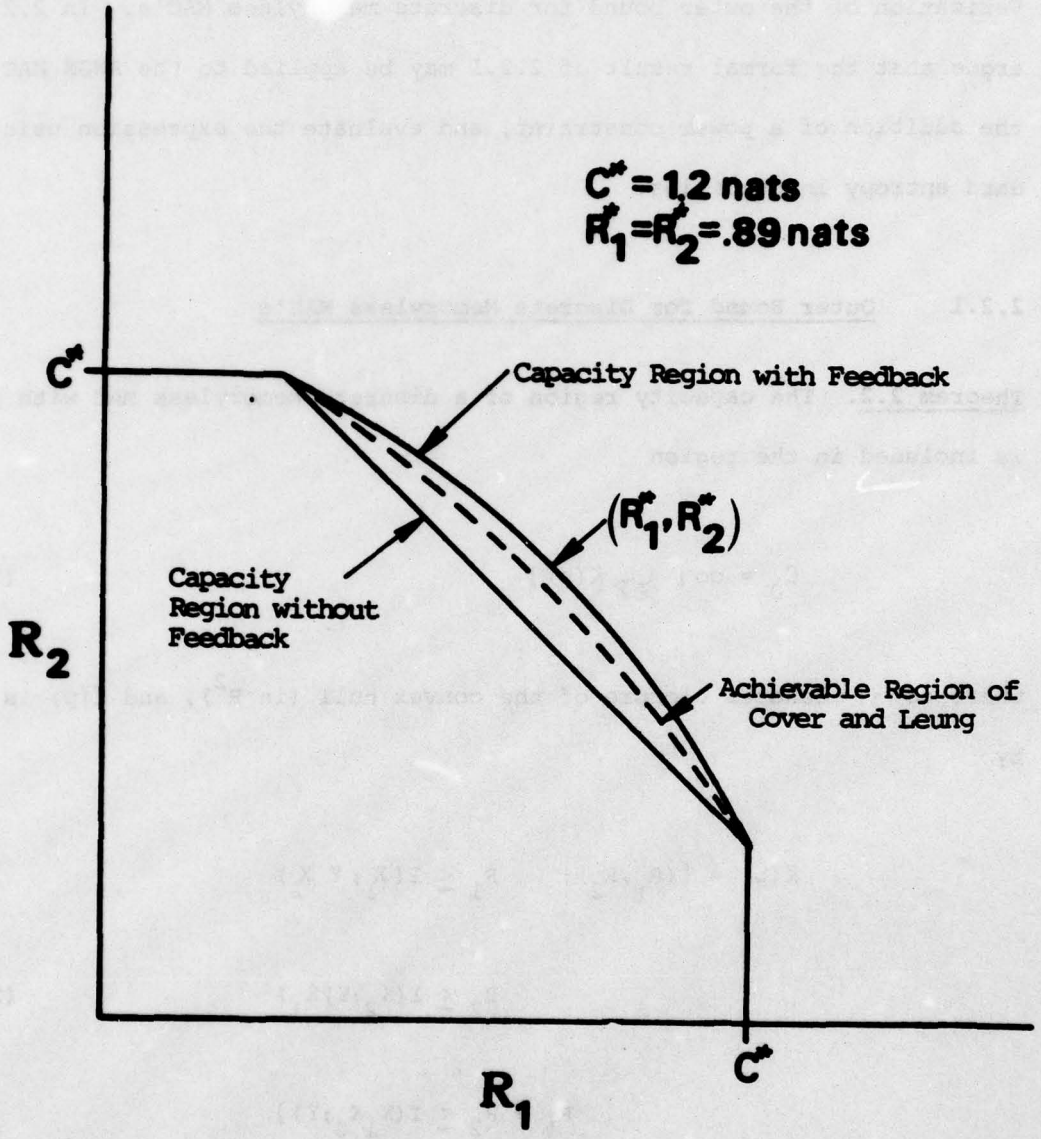
so that (2.1.43) becomes

$$\Pr[\text{error}] < 2\varepsilon_1 + 2\varepsilon_2 + \varepsilon_3 \qquad (2.1.45)$$

We have shown that we can reliably communicate at all rates arbitrarily close to, but dominated by, a curve in R^2 from (R_1^*,R_2^*) given by equations (2.1.26) and (2.1.11) to point D in Figure 1.2. By reversing the roles of T_1 and T_2 on this section we can prove a similar result for a curve between (R_1^*,R_2^*) and point C in Figure 1.2. The entire region of the form displayed in Fig. 2.2 is thus achievable. Doubly exponential error decay is no longer guaranteed, since the error probability is now dominated by the error rate for a non-feedback code, which has singly exponential behavior.

2.2 Converse

In this section we obtain an outer bound to the capacity region



2.2 Capacity Region of the AWGN MAC with Feedback for $P_1 = P_2 = 10$ and $\sigma^2 = 1$.

of the AWGN MAC with feedback. In 2.2.1 we give a single letter characterization of the outer bound for discrete memoryless MAC's. In 2.2.2 we argue that the formal result of 2.2.1 may be applied to the AWGN MAC with the addition of a power constraint, and evaluate the expression using standard entropy inequalities.

2.2.1 Outer Bound for Discrete Memoryless MAC's

Theorem 2.2. The capacity region of a discrete memoryless MAC with feedback is included in the region

$$C_0 = \text{co}\left[\bigcup_{p \in P} K(p) \right] \tag{2.2.1}$$

where co[•] denotes closure of the convex hull (in R^2), and $K(p)$ is given by

$$\begin{aligned} K(p) = \{(R_1,R_2): \quad & R_1 \le I(X_1;Y|X_2) \\ & R_2 \le I(X_2;Y|X_1) \\ & R_1 + R_2 \le I(X_1X_2;Y)\} \end{aligned} \tag{2.2.2}$$

and P is the set of all joint probability assignments on X_1, X_2.

The difference between this outer bound and the capacity region of the MAC without feedback is that the union is carried out over all $p(x_1,x_2)$ rather than just product (i.e., independent) distributions.

The convexification implied by (2.2.1) is really unnecessary since $\bigcup_P K(p)$ is already convex. To see this, note that the constraints in (2.2.2) are concave (Appendix B) and consider any two points in $\bigcup_P K(p)$. If they are contained in a single $K(p)$, then any convex combination is also in $K(p)$ since each $K(p)$ is by itself convex. Assume then, that they are in different $K(p)$, say $(R_1^1, R_2^1) \varepsilon K(p_1)$ and $(R_1^2, R_2^2) \varepsilon K(p_2)$. Define $(R_1, R_2) = \alpha(R_1^1, R_2^1) + (1-\alpha)(R_1^2, R_2^2)$ and $p = \alpha p_1 + (1-\alpha)p_2$, where $\alpha \varepsilon [0,1]$. Then

$$R_1 = \alpha R_1^1 + (1-\alpha)R_1^2 \leq \alpha I_{p_1}(X_1;Y|X_2) + (1-\alpha) I_{p_2}(X_1;Y|X_2)$$

$$\leq I_p(X_1;Y|X_2) \qquad (2.2.3a)$$

$$R_2 = \alpha R_2^1 + (1-\alpha)R_2^2 \leq \alpha I_{p_1}(X_2;Y|X_1) + (1-\alpha) I_{p_2}(X_2;Y|X_1)$$

$$\leq I_p(X_2;Y|X_1) \qquad (2.2.3b)$$

$$R_1+R_2 = \alpha(R_1^1+R_2^1) + (1-\alpha)(R_1^2+R_2^2) \leq \alpha I_{p_1}(X_1X_2;Y) + (1-\alpha) I_{p_2}(X_1X_2;Y)$$

$$\leq I_p(X_1X_2;Y) \qquad (2.2.3c)$$

where the subscripting of the mutual informations denotes evaluation at the appropriate probability. The first inequality in each case is by definition of $K(p_i)$ and the second is by the concavity of the mutual informations as

functions of $p(x_1,x_2)$. Equation (2.2.3) shows that $(R_1,R_2) \in K(p)$ C_o, so that C_o is convex. This reasoning does not apply to the MAC without feedback, since a conxex combination of product probabilities is not necessarily a product probability. Therefore convexification is necessary in (1.2.1).

We proceed to prove Theorem 2.1. We define the data sources to independently produce strings of letters drawn from arbitrary discrete alphabets at rates L_s^i (i=1,2) symbols per second. We denote the size of the i^{th} alphabeth by $M_i<\infty$. We further assume that the sources are stationary, that is, the joint statistics of any finite string are independent of time origin. For message $\underline{a}_i$, a string of L letters produced by source i, we define the quantity

$$H_L(a_i) = \frac{1}{L} H(\underline{a}_i) \text{ nats/letter} \qquad (2.2.4)$$

which is the average per-letter entropy of source i, for a string of L letters. In general $H_L(a_i)$ is a bounded monotonically decreasing function of L [24].

We consider the problem of transmitting data produced by sources 1 and 2 over a channel using a block code of length N. Assuming the channel may be used N_c times per second, then

$$L^i = \frac{L_s^i N}{N_c} \qquad (2.2.5)$$

source letters are produced per code block. The entropy of this string is given by

$$H(\underline{a}_i) = L^i H_{L^i}(a_i) \quad \text{nats} \tag{2.2.6}$$

and the amount of information communicated per channel use (rate) is given by

$$R_i \triangleq \frac{1}{N} H(\underline{a}_i) = \frac{L^i}{N} H_{L^i}(a_i) = \frac{L^i_s}{N_c} H_{L^i}(a_i) \text{ nats/transmission} \tag{2.2.7}$$

We now define the encoding and decoding strategies allowed. For transmitter i, we require that

$$X_{i,k} = f_i(\underline{a}_i, \underline{Y}^{k-1}) \quad i=1,2 \quad k=1,N \tag{2.2.8}$$

where $X_{i,k}$ is the i^{th} transmitter's channel symbol at time k, f_i is a deterministic function, and $\underline{Y}^{k-1}$ is the receiver's channel output up to and including time k-1. $\underline{Y}^{k-1}$ is available to both transmitters via the feedback links.

Similarly, define a pair of decoding functions

$$\hat{\underline{a}}_i = g_i(\underline{Y}) \quad i=1,2 \tag{2.2.9}$$

which are functions of $\underline{Y}$, the vector of N channel outputs.

We will use a per-letter form of Fano's inequality ([24], Theorem 4.3.2) which states that

$$\frac{1}{L^i} H(\underline{a}_i | \hat{\underline{a}}_i) \leq \varepsilon_i \ln(M_i - 1) + h(\varepsilon_i) \triangleq \eta_i(\varepsilon_i) \text{ nats/letter} \tag{2.2.10}$$

where
$$\varepsilon_i = \frac{1}{L^i} \sum_{j=1}^{L^i} \Pr[a_{ij} \neq \hat{a}_{ij}] \tag{2.2.11}$$

and

$$h(x) = -x \ln x - (1-x) \ln (1-x) \tag{2.2.12}$$

and a_{ij} (or $\hat{a}_{ij}$) is the j^{th} letter of $\underline{a}_i$ (or $\underline{\hat{a}}_i$).

Now consider source 1. Since $\underline{\hat{a}}_1 = g_1(\underline{Y})$ we can apply the data processing theorem [24] to obtain

$$H(\underline{a}_1|\underline{Y}) \leq H(\underline{a}_1|g_1(\underline{Y})) = H(\underline{a}_1|\underline{\hat{a}}_1) \tag{2.2.13}$$

Also, since conditioning cannot increase entropy,

$$H(\underline{a}_1|\underline{Y}\ \underline{a}_2) \leq H(\underline{a}_1|\underline{Y}) \leq H(\underline{a}_1|\underline{\hat{a}}_1) \tag{2.2.14}$$

Applying (2.2.10) we have

$$H(\underline{a}_1|\underline{Y}\ \underline{a}_2) \leq L^1 \eta_1(\varepsilon_1) \tag{2.2.15}$$

By independence,

$$H(\underline{a}_1|\underline{a}_2) = H(\underline{a}_1) \tag{2.2.16}$$

Subtracting(2.2.15) from (2.2.16) we have

$$I(\underline{a}_1; \underline{Y}|\underline{a}_2) \geq H(\underline{a}_1) - L^1\eta_1(\varepsilon_1) \tag{2.2.17}$$

We bound the left hand side of (2.2.17) as follows:

$$I(\underline{a}_1;\underline{Y}|\underline{a}_2) = H(\underline{Y}|\underline{a}_2) - H(\underline{Y}|\underline{a}_1\underline{a}_2)$$

$$= \sum_{k=1}^{N} [H(Y_k|\underline{Y}^{k-1}\underline{a}_2) - H(Y_k|\underline{Y}^{k-1}\underline{a}_1\underline{a}_2)]$$

a) $$= \sum_{k=1}^{N} [H(Y_k|X_{2k}\underline{Y}^{k-1}\underline{a}_2) - H(Y_k|\underline{Y}^{k-1}\underline{a}_1\underline{a}_2)]$$

b) $$\leq \sum_{k=1}^{N} [H(Y_k|X_{2k}) - H(Y_k|X_{1k}X_{2k}\underline{Y}^{k-1}\underline{a}_1\underline{a}_2)]$$

c) $$= \sum_{k=1}^{N} [H(Y_k|X_{2k}) - H(Y_k|X_{1k}X_{2k})]$$

$$= \sum_{k=1}^{N} I(X_{1k}; Y_k|X_{2k}) \tag{2.2.18}$$

Step a) is true since X_{2k} is a deterministic function of $(\underline{Y}^{k-1}\underline{a}_2)$, step b) is true since conditioning cannot increase entropy, and c) follows from the fact that given the current inputs, the output is independent of the past.

We can combine (2.2.17) and (2.2.18) to obtain

$$\sum_{k=1}^{N} I(X_{1k};Y_k|X_{2k}) \geq H(\underline{a}_1) - L^1\eta_1(\varepsilon_1) \tag{2.2.19}$$

Subtracting both sides from $NR_1 = H(\underline{a}_1)$, we have

$$N[R_1 - \frac{1}{N}\sum_{k=1}^{N} I(X_{1k};Y_k|X_{2k})] \leq L^1\eta_1(\varepsilon_1) \tag{2.2.20}$$

Note that $\frac{L^1}{N} = \frac{L_s^1}{N_c}$ (equation (2.2.5)) which is fixed, so that

$$R_1 - \frac{1}{N}\sum_{k=1}^{N} I(X_{1k};Y_k|X_{2k}) \leq \frac{L_s^1}{N_c}\eta_1(\varepsilon_1) \tag{2.2.21}$$

A similar series of steps, applied to source 2, yields

$$R_2 - \frac{1}{N}\sum_{k=1}^{N} I(X_{2k};Y_k|X_{1k}) \leq \frac{L_s^2}{N_c}\eta_2(\varepsilon_2) \tag{2.2.22}$$

We also consider the joint source $(\underline{a}_1,\underline{a}_2)$. As in equation (2.2.13)

$$\begin{aligned} H(\underline{a}_1\underline{a}_2|\underline{Y}) &\leq H(\underline{a}_1\underline{a}_2|\hat{\underline{a}}_1\hat{\underline{a}}_2) \\ &\leq H(\underline{a}_1|\hat{\underline{a}}_1\hat{\underline{a}}_2) + H(\underline{a}_2|\hat{\underline{a}}_1\hat{\underline{a}}_2) \\ &\leq H(\underline{a}_1|\hat{\underline{a}}_1) + H(\underline{a}_2|\hat{\underline{a}}_2) \end{aligned} \tag{2.2.23}$$

so that

$$I(\underline{a}_1\underline{a}_2;\underline{Y}) \geq H(\underline{a}_1\underline{a}_2) - L^1\eta_1(\varepsilon_1) - L^2\eta_2(\varepsilon_2) \tag{2.2.24}$$

The left hand side is bounded by

$$I(\underline{a}_1\underline{a}_2;\underline{Y}) = H(\underline{Y}) - H(\underline{Y}|\underline{a}_1\underline{a}_2)$$

$$= \sum_{k=1}^{N} [H(Y_k|\underline{Y}^{k-1}) - H(Y_k|\underline{Y}^{k-1}\underline{a}_1\underline{a}_2)]$$

a) $$\leq \sum_{k=1}^{N} [H(Y_k) - H(Y_k|X_{1k}X_{2k}\underline{Y}^{k-1}\underline{a}_1\underline{a}_2)]$$

b) $$= \sum_{k=1}^{N} [H(Y_k) - H(Y_k|X_{1k}X_{2k})]$$

$$= \sum_{k=1}^{N} I(X_{1k}X_{2k};Y_k) \qquad (2.2.25)$$

where a) follows since conditioning cannot increase entropy, and b) follows since given inputs, the current output is independent of the past.

Substituting (2.2.25) into (2.2.24) we have

$$\sum_{1}^{N} I(X_{1k}X_{2k};Y_k) \geq H(\underline{a}_1\underline{a}_2) - L^1\eta_1(\varepsilon_1) - L^2\eta_2(\varepsilon_2) \qquad (2.2.26)$$

Subtracting from $N(R_1+R_2) = H(\underline{a}_1) + H(\underline{a}_2) = H(\underline{a}_1\underline{a}_2)$ (by independence) we get

$$N[R_1 + R_2 - \frac{1}{N}\sum_{k=1}^{N} I(X_{1k}X_{2k};Y_k)] \leq L^1\eta_1(\varepsilon_1) + L^2\eta_2(\varepsilon_2) \qquad (2.2.27)$$

and

$$R_1 + R_2 - \frac{1}{N}\sum_{k=1}^{N} I(X_{1k}X_{2k};Y_k) \leq \frac{L_s^1}{N_c}\eta_1(\varepsilon_1) + \frac{L_s^2}{N_c}\eta_2(\varepsilon_2) \qquad (2.2.28)$$

Now any code of arbitrary block length must yield average (per-transmission) mutual informations which lie inside C_o. To see this, assume a block length of N and a joint probability assignment on $\underline{X}_1, \underline{X}_2$ given by

$$p(\underline{X}_1\underline{X}_2) = \prod_{k=1}^{N} p(X_{1k},X_{2k}|\underline{X}_1^{k-1}\underline{X}_2^{k-1}) \tag{2.2.29}$$

In general, the probabilities at any time will depend on the values at all other times, but the mutual informations of equations (2.2.21) (2.2.22) and (2.2.25) depend only on single-letter probabilities, so this dependence does not affect the average. If we define the "average" probability assignment by

$$p^*(X_1,X_2) = \frac{1}{N}\sum_{k=1}^{N} p_k(X_{1k},X_{2k}) \tag{2.2.30}$$

then we can apply Jensen's inequality to (for example) $I(X_1;Y|X_2)$ to obtain

$$\frac{1}{N}\sum_{1}^{N} I(X_{1k};Y_k|X_{2k}) = \frac{1}{N}\sum I_{p_k}(X_{1k};Y_k|X_{2k})$$

$$\leq I_{p^*}(X_1;Y|X_2) \tag{2.2.31}$$

since $I(X_1;Y|X_2)$ is concave in $p(X_1,X_2)$. All of the mutual informations are concave (Appendix B), so that (2.2.21), (2.2.22) and (2.2.28) become

$$R_1 - I_{p^*}(X_1;Y|X_2) \leq \frac{L_s^1}{N_c}\eta_1(\varepsilon_1) \tag{2.2.32a}$$

$$R_2 - I_{p^*}(X_2;Y|X_1) \leq \frac{L_s^2}{N_c}\eta_2(\varepsilon_2) \tag{2.2.32b}$$

$$R_1 + R_2 - I_{p^*}(X_1,X_2;Y) \leq \frac{L_s^1}{N_c}\eta_1(\varepsilon_1) + \frac{L_s^2}{N_c}\eta_2(\varepsilon_2) \tag{2.2.32c}$$

Assume $(R_1 R_2)$ is not in C_o. Since C_o is a closed convex subset of R_+^2, then there exists a vector (λ_1,λ_2) where $\lambda_1,\lambda_2 \geq 0$ and λ_1,λ_2 are not both zero, for which

$$\lambda_1 R_1 + \lambda_2 R_2 > \sup_{(x_1,x_2)\in C_o} \lambda_1 x_1 + \lambda_2 x_2 \qquad (2.2.33)$$

(see, for example [26]). In particular, there is a positive δ such that for all $(x_1,x_2)\in C_o$

$$\lambda_1 R_1 + \lambda_2 R_2 - \lambda_1 x_1 - \lambda_2 x_2 \geq \delta > 0 \qquad (2.2.34)$$

Combining (2.2.32) and (2.2.34) we have

$$\lambda_1 \frac{L_s^1}{N_c} \eta_1(\varepsilon_1) + \lambda_2 \frac{L_s^2}{N_c} \eta_2(\varepsilon_2) \geq \delta \qquad (2.2.35)$$

Now define

$$0 < \lambda \triangleq \max(\lambda_1\lambda_2) < \infty$$

$$0 < K \triangleq \max\left(\frac{L_s^1}{N_c}, \frac{L_s^2}{N_c}\right) < \infty$$

$$\varepsilon \triangleq \max(\varepsilon_1,\varepsilon_2)$$

$$h'(\varepsilon) \triangleq \begin{cases} h(\varepsilon) & 0 \leq \varepsilon \leq \frac{1}{2} \\ 1 & \frac{1}{2} \leq \varepsilon \leq 1 \end{cases} \qquad (2.2.36)$$

and

$$\eta(\varepsilon) \triangleq \varepsilon \ln(M_1 M_2 - 1) + h'(\varepsilon) \,. \tag{2.2.37}$$

Clearly

$$\eta(\varepsilon) \geq \max(\eta_1(\varepsilon_1), \eta_2(\varepsilon_2)) \tag{2.2.38}$$

and

$$\begin{aligned} \delta &\leq \lambda_1 \frac{L_s^1}{N_c} \eta_1(\varepsilon_1) + \lambda_2 \frac{L_s^2}{N_c} \eta_2(\varepsilon_2) \\ &\leq \lambda \left[\frac{L_s^1}{N_c} \eta_1(\varepsilon_1) + \frac{L_s^2}{N_c} \eta_2(\varepsilon_2)\right] \\ &\leq \lambda K [\lambda_1(\varepsilon_1) + \eta_2(\varepsilon_2)] \\ &\leq 2\lambda K \eta(\varepsilon) \end{aligned} \tag{2.2.39}$$

Therefore

$$\eta(\varepsilon) \geq \frac{\delta}{2\lambda K} > 0 \tag{2.2.40}$$

Since $\eta(0) = 0$ and $\eta(\varepsilon)$ is strictly mononotonically increasing in ε, then (2.2.40) implies that ε is greater than some positive constant, independent of block length. Theorem 2.1 is proven.

2.2.2 Outer Bound for AWGN MAC's

We first note that the formal result of Section 2.2.1 is valid for arbitrary ensembles as long as the various entropies involved can be defined

by either using the conventional "absolute" entropy for discrete random variables, or the "differential" entropy defined by

$$H(x) = -\int_{-\infty}^{\infty} \ln p(x)p(x)dx \tag{2.2.41}$$

where $p(x)$ is a probability density function. We follow the standard notation (e.g. [24]) in using the same symbol for absolute and differential entropy. In going from equation (2.2.17) to (2.2.18) we identified a mutual information defined as a difference between absolute entropies with one defined using differential entropies. This step is justified so long as the differential entropy is defined (i.e. $\underline{Y}$ has a density function) and finite ([24] Section 2.5). $\underline{Y}$ has a density function, since it is a sum of the inputs and gaussian noise. Its entropy is bounded since

$$-\infty < \frac{1}{2} \ln 2\pi e\sigma^2 = H(Y_k|X_{1k}X_{2k}) \leq H(Y_k|X_{ik}) \leq H(Y_k)$$

$$\leq \frac{1}{2} \ln (N(\sqrt{P_1} + \sqrt{P_2})^2 + \sigma^2)$$

$$< \infty \tag{2.2.42}$$

where σ^2 is the noise variance and P_i are the average power constraints. The fourth term above may involve either X_{1k} or X_{2k}. All of the various conditioning properties hold in this case.

We also note that the concavity of the mutual informations in (2.2.2) allows us to reduce the per-block average energy constraint to a

per-letter average energy constraint. That is, for any code satisfying

$$\sum_{k=1}^{N} \overline{x_{ik}^2} \leq NP_i \qquad i=1,2 \tag{2.2.43}$$

a code using the average joint letter probability law (i.e. $P(x_1x_2) = \frac{1}{N}\sum_{k=1}^{N} P(x_{1k},x_{2k})$) yields mutual informations which are no smaller (when averaged over the block), and for which each letter satisfies

$$\overline{x_{ik}^2} \leq P_i \qquad i=1,2,\ k=1,\ldots,N \tag{2.2.44}$$

We proceed by fixing a joint probability assignment on (X_1,X_2) and bounding $K(p)$. Any assignment on (X_1,X_2) has the characteristic parameters

$$\sigma_1^2 = \overline{x_1^2} \leq P_1$$

$$\sigma_2^2 = \overline{x_2^2} \leq P_2$$

$$\rho = \frac{\overline{x_1x_2}}{\sigma_1\sigma_2} \in [-1,1] \tag{2.2.45}$$

We have assumed, without loss of generality, that X_1 and X_2 are zero-mean.

By our definition,

$$\mathrm{Var}(Y) = E[Y^2] = \sigma_1^2 + \sigma_2^2 + 2\sigma_1\sigma_2\rho + \sigma^2 \tag{2.2.46}$$

where σ^2 is the additive noise variance. The gaussian distribution has the

maximum entropy of all continuous distributions subject to a constraint on the variance. Its entropy, for variance s is given by

$$= \frac{1}{2} \ln 2\pi e s \triangleq g(s) \text{ nats} \tag{2.2.47}$$

The entropy of Y is then bounded by

$$H(Y) \leq g(\text{Var}(Y)) = g(\sigma_1^2 + \sigma_2^2 + 2\sigma_1\sigma_2\rho + \sigma^2) \tag{2.2.48}$$

Given X_1 and X_2, the only random component of Y is the additive noise, which is gaussian, so that

$$H(Y|X_1X_2) = g(\sigma^2) \tag{2.2.49}$$

Now consider the conditional entropy of Y given X_2. By definition

$$H(Y|X_2) = E_{X_2}[H(Y|X_2=x_2)] \tag{2.2.50}$$

where the expectation is over X_2. Now given $X_2=x_2$ the entropy of Y may be bounded using (2.2.47), by

$$H(Y|X_2=x_2) \leq g(\text{Var}(Y|X_2=x_2)) \tag{2.2.51}$$

The variance of Y given $X_2=x_2$ is given by

$$\mathrm{Var}(Y|X_2=x_2) = \sigma^2 + \mathrm{Var}(X_1|X_2=x_2) \tag{2.2.52}$$

since channel noise is independent of both X_1 and X_2. The second term on the right hand side of (2.2.52) in turn, is by definition

$$\mathrm{Var}(X_1|X_2=x_2) = E_{X_1}[(x_1 - E_{X_1}[X_1|X_2=x_2])^2|X_2=x_2] \tag{2.2.53}$$

where the expectation is over X_1 given $X_2=x_2$.

Also the conditional expectation of X_1 given $X_2=x_2$ is the unique function of x_2 which minimizes the right hand side. To upper bound the expression, we approximate the conditional mean by the best linear estimate of X_1 given $X_2=x_2$, which is

$$\hat{x}_1 = \frac{\sigma_1}{\sigma_2}\rho x_2 \tag{2.2.54}$$

yielding

$$\mathrm{Var}(X_1|X_2=x_2) \leq E_{X_1}[(x_1 - \frac{\sigma_1}{\sigma_2}\rho\, x_2)^2|x_2] \tag{2.2.55}$$

Averaging over X_2 we have

$$E_{X_2}[\mathrm{Var}(X_1|X_2=x_2)] \triangleq \mathrm{Var}(X_1|X_2)$$

$$= E_{X_2}E_{X_1}[(x_1 - \frac{\sigma_1}{\sigma_2}\rho\, x_2)^2|x_2]$$

$$= E[(x_1 - \frac{\sigma_1}{\sigma_2}\rho\, x_2)^2]$$

$$= \sigma_1^2 + \frac{\sigma_1^2}{\sigma_2^2}\ \rho^2\sigma_2^2 - 2\ \frac{\sigma_1}{\sigma_2}\rho^2\sigma_1\sigma_2$$

$$= \sigma_1^2(1-\rho^2) \qquad (2.2.56)$$

Going back to equation (2.2.50) we note that $g(x)$ is concave in x, so that Jensen's inequality yields

$$H(Y|X_2) \triangleq E[H(Y|X_2=x_2)]$$

$$\leq g(\sigma^2 + E[Var(X_1|X_2=x_2)])$$

$$= g(\sigma^2 + Var(X_1|X_2))$$

$$\leq g(\sigma^2 + \sigma_1^2(1-\rho^2)) \qquad (2.2.57a)$$

since $g(x)$ is monotonic in x.

We can exchange X_1 and X_2 in the above argument to obtain

$$H(Y|X_1) \leq g(\sigma^2 + \sigma_2^2(1-\rho^2)) \qquad (2.2.57b)$$

Substituting (2.2.48, (2.2.49) and (2,2,57) into the definitions of mutual information, and observing that

$$g(x) - g(y) = \frac{1}{2} \ln \frac{x}{y} \qquad (2.2.58)$$

equation (2.2.1)-(2.2.2) becomes

$$C_o = \bigcup_{p \in P} K(p) \tag{2.2.59}$$

where
$$K(p) = \{(R_1,R_2): \; R_1 \leq \frac{1}{2} \ln(1 + \frac{\sigma_1^2}{\sigma^2}(1-\rho^2))$$

$$R_2 \leq \frac{1}{2} \ln(1 + \frac{\sigma_2^2}{\sigma^2}(1-\rho^2))$$

$$R_1 + R_2 \leq \frac{1}{2} \ln(1+ \frac{\sigma_1^2 + \sigma_2^2 + 2\sigma_1\sigma_2\rho}{\sigma^2}) \tag{2.2.60}$$

and

$$P = \{p(x_1,x_2): \; \sigma_1^2 \leq P_1$$

$$\sigma_2^2 \leq P_2$$

$$|\rho| \leq 1\} \tag{2.2.61}$$

We now observe that a region yielded by a negative correlation ρ is included in the corresponding region yielded by $|\rho|$. Also, for positive ρ, the bounds are all maximized when $\sigma_1^2 = P_1$ and $\sigma_2^2 = P_2$. Equation (2.2.59)-(2.2.60) may be re-written then, as

Theorem 2.3. An outerbound on the capacity region of the AWGN MAC with feedback is given by

$$C_o = \bigcup_{0\le\rho\le 1} K'(\rho) \tag{2.2.62}$$

where

$$K'(\rho) = \{(R_1,R_2): \quad R_1 \le \frac{1}{2}\ln(1 + \frac{P_1}{\sigma^2}(1-\rho^2))$$

$$R_2 \le \frac{1}{2}\ln(1 + \frac{P_2}{\sigma^2}(1-\rho^2)) \tag{2.2.63}$$

$$R_1 + R_2 \le \frac{1}{2}\ln(1+\frac{P_1 + P_2 + 2\sqrt{P_1P_2}\rho}{\sigma^2})\} \tag{2.2.63}$$

We have replaced $K(p)$, a function of a probability assignment, by $K'(\rho)$, a function of the correlation coefficient.

Observe that (2.2.63) reduces to (1.2.3), the capacity region of the AWGN MAC without feedback, when $\rho = 0$, which of course for gaussian (x_1,x_2) corresponds to independent transmitted signals.

2.3 The Capacity Region of the AWGN MAC With Feedback

We will now show that the achievable region of Section 2.1 and the outer bound of Theorem 2.3 are identical. To simplify notation in the sequel, we define the following auxiliary quantities:

$$P_T(\rho) = 1 + \frac{P_1 + P_2 + 2\sqrt{P_1P_2}|\rho|}{\sigma^2} \tag{2.3.1a}$$

$$P_A(\rho) = 1 + \frac{P_1}{\sigma^2}(1-\rho^2) \tag{2.3.2b}$$

$$P_B(\rho) = 1 + \frac{P_2}{\sigma^2}(1-\rho^2) \tag{2.3.3c}$$

Equation (2.2.63) then becomes

$$K'(\rho) = \{(R_1,R_2): \quad R_1 \leq \frac{1}{2} \ln P_A(\rho)$$

$$R_2 \leq \frac{1}{2} \ln P_B(\rho)$$

$$R_1 + R_2 \leq \frac{1}{2} \ln P_T(\rho)\} \tag{2.3.4}$$

The achievable region is

$$A = A_1 \cup A_2 \tag{2.3.5}$$

where A_1 is given by (2.1.39) and (2.1.40), which become

$$A_1 = \bigcup_{0\leq\bar{\alpha}\leq 1} \{(R_1,R_2): \quad R_1 \leq \frac{1}{2} \ln P_A(\sqrt{\bar{\alpha}}\rho)$$

$$R_2 \leq \frac{1}{2} \ln P_T(\sqrt{\bar{\alpha}}\rho)/P_A(\sqrt{\bar{\alpha}}\rho)\} \tag{2.3.6}$$

where ρ satisfies (2.1.36), which we repeat:

$$P(\alpha,\rho) = (\sigma^2 + \alpha P_1)P_T(\sqrt{\bar{\alpha}}\rho) - P_A(\sqrt{\bar{\alpha}}\rho)(\sigma^2 + \alpha P_1 + P_2(1-\rho^2))$$

$$= 0 \tag{2.3.7}$$

A_2 is given by an identical expression with P_1 and P_2, R_1 and R_2, and

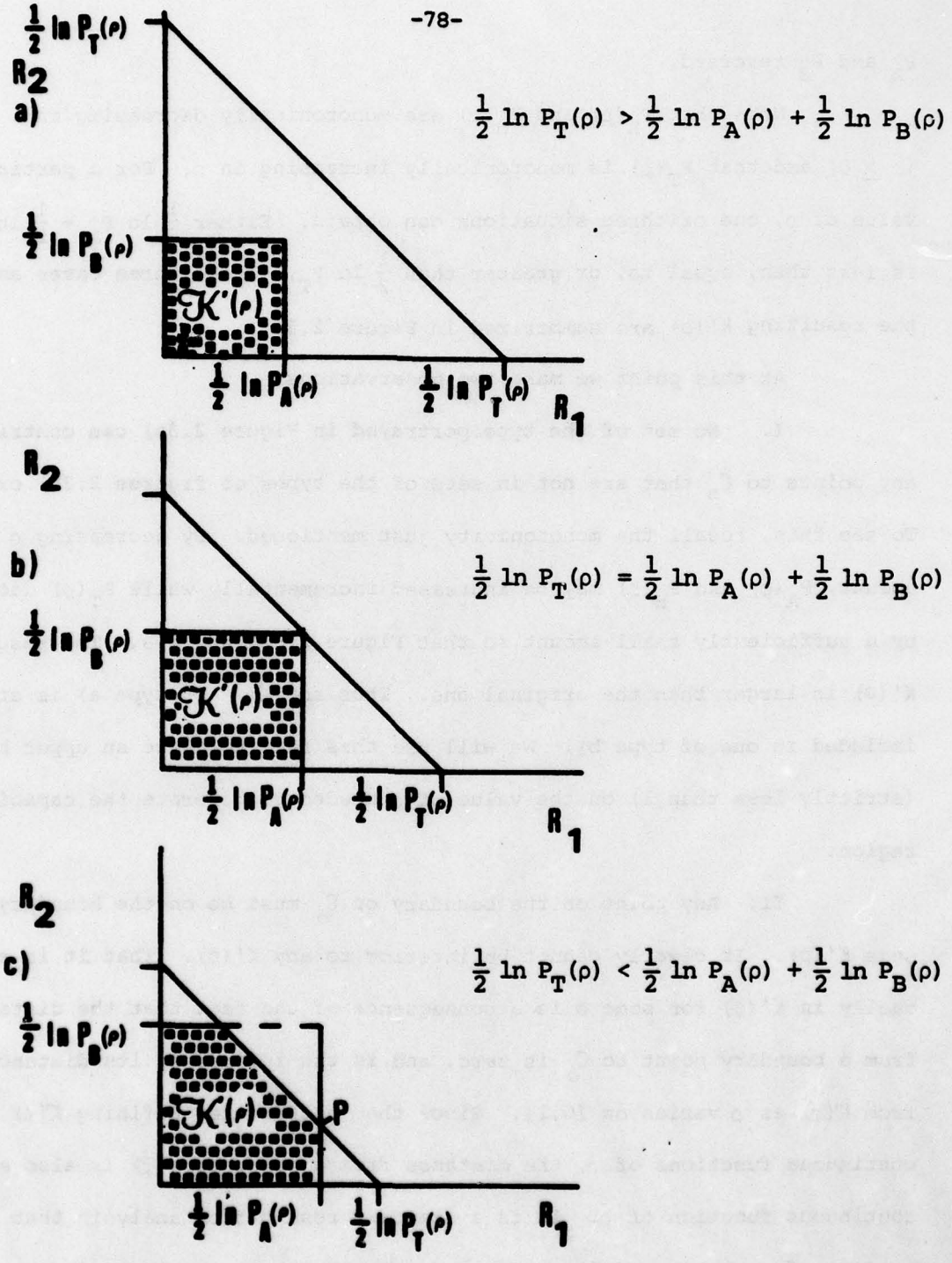
P_A and P_B reversed.

Note that $P_A(\rho)$ and $P_B(\rho)$ are monotonically decreasing in $(\rho \geq 0)$ and that $P_T(\rho)$ is monotonically increasing in ρ. For a particular value of ρ, one of three situations can obtain. Either $\frac{1}{2} \ln P_A + \frac{1}{2} \ln P_B$ is less than, equal to, or greater than $\frac{1}{2} \ln P_T$. These three cases and the resulting $K'(\rho)$ are summarized in Figure 2.3.

At this point we make two observations:

I. No set of the type portrayed in Figure 2.3a) can contribute any points to C_o that are not in sets of the types of figures 2.3b) or c). To see this, recall the monotonicity just mentioned. By decreasing ρ a small amount, $P_A(\rho)$ and $P_B(\rho)$ may be increased incrementally while $P_T(\rho)$ decreases by a sufficiently small amount so that Figure 2.3b) applies. The resulting $K'(\rho)$ is larger than the original one. Thus any $K'(\rho)$ of type a) is strictly included in one of type b). We will use this fact to place an upper bound (strictly less than 1) on the value of ρ needed to generate the capacity region.

II. Any point on the boundary of C_o must be on the boundary of some $K'(\rho)$. It clearly cannot be interior to any $K'(\rho)$. That it is actually in $K'(\rho)$ for some ρ is a consequence of the fact that the distance from a boundary point to C_o is zero, and is the infimum of its distance from $K'(\rho)$ as ρ varies on $[0,1]$. Since the coordinates defining $K'(\rho)$ are continuous functions of ρ, the distance from a point to $K'(\rho)$ is also a continuous function of ρ. It is a standard result from analysis that the extrema of a continuous function on a compact set are achieved in the set. As a result of this observation, we see that every point on the boundary of



2.3 **Basic Sets for the Outer Bound to the Capacity Region of the AWGN MAC with Feedback.**

C_o is included in the set of boundary points of the collection of $K'(\rho)$.

We now find the point on the boundary of C_o which maximizes $R_1 + R_2$, and show that it is just the point (R_1^*, R_2^*) of Section 2.1. The $K'(\rho)$ which maximizes $R_1 + R_2$ must be of the type shown in Figure b), since for any $K'(\rho)$:

$$R_1 + R_2 \leq \frac{1}{2} \ln P_A(\rho) + \frac{1}{2} \ln P_B(\rho) \tag{2.3.8a}$$

and

$$R_1 + R_2 \leq \frac{1}{2} \ln P_T(\rho) \tag{2.3.8b}$$

The first bound decreases with ρ; the second increases. The maximum must then occur when they are equal. This equality, however, requires that

$$\frac{1}{2} \ln P_A(\rho) + \frac{1}{2} \ln P_B(\rho) = \frac{1}{2} \ln P_T(\rho) \tag{2.3.9}$$

or

$$P_A(\rho) P_B(\rho) = P_T(\rho) \tag{2.3.10}$$

But this is precisely equation (2.1.11) (or (2.1.36) evaluated at $\alpha = 0$), which defines the correlation coefficient obtained by the scheme of Section 2.1.1. Since the solution to (2.1.11) is unique (Appendix A), our coding scheme yields the appropriate correlation between X_1 and X_2. The point achieved by this scheme is (from (2.3.6) with $\alpha = 0$)

$$R_1^* = \frac{1}{2} \ln P_A(\rho)$$

$$R_2^* = \frac{1}{2} \ln \frac{P_T(\rho)}{P_A(\rho)} \tag{2.3.11}$$

By substituting $P_A(\rho)P_B(\rho)$ for $P_T(\rho)$ we have

$$R_2^* = \frac{1}{2} \ln P_B(\rho) \tag{2.3.12}$$

so that (R_1^*, R_2^*) coincides with the upper right hand corner of $K'(\rho)$ for the ρ (which we will call ρ^*) which yields the maximum (R_1+R_2), and thus is that extremizing point.

Now consider a point on the boundary of C_o for which $R_1 > R_1^*$. By observation II above, any such point must lie on the boundary of $K(\rho')$ for some ρ'. Also $\rho' < \rho^*$ since $P_A(\rho)$ is strictly monotonically decreasing with ρ (observation I). Finding this boundary point for some fixed R_1^1 is equivalent to maximizing R_2 as a function of R_1^1. In this light it is clear that the maximum (and hence the boundary point) occurs at the corner labelled P in Figure 2.3.c), and thus that ρ' must be such that $\frac{1}{2} \ln P_A(\rho') = R_1^1$. If ρ' were larger, then R_1^1 would not be achievable in $K(\rho')$, and if ρ' were smaller, then $P_T(\rho')$ is diminished, so that a smaller R_2 (bounded by $\frac{1}{2} \ln P_T(\rho') - R_1^1$) would result.

These considerations lead us to conclude that the boundary of C_o is generated by the corners of the basic sets $K'(\rho)$ of type c) as ρ varies between 0 and ρ^*. At $\rho = 0$ we have not reached the positive axes, but the time sharing argument applied to the points

$$R_1 = \frac{1}{2} \ln(1 + \frac{P_1}{\sigma^2})$$

$$R_2 = \frac{1}{2} \ln\left(1 + \frac{P_1 + P_2}{\sigma^2}\right) - R_1 \qquad (2.3.13)$$

generated at $\rho = 0$

and $$R_1 = \frac{1}{2} \ln\left(1 + \frac{P_1}{\sigma^2}\right)$$

$$R_2 = 0 \qquad (2.3.14)$$

generated by not using Transmitter 2, suffices to connect with the R_1 axis, and a similar argument connects with the R_2 axis.

Now a point on the boundary of C_o for which $R_1^1 > R_1^*$ is generated by $\rho' < \rho^*$, which is given by

$$R_1^1 = \frac{1}{2} \ln\left(1 + \frac{P_1}{\sigma^2}(1-\rho'^2)\right) \qquad (2.3.15)$$

The value of R_2 (call R_2') corresponding to this value of ρ', at the corner P, is given by

$$R_2^1 = \frac{1}{2} \ln P_T(\rho') - R_1^1$$

$$= \frac{1}{2} \ln \frac{P_T(\rho')}{P_A(\rho')} \qquad (2.3.16)$$

Now consider region A_1. From equation (2.3.6) we see that if a ρ such that $\sqrt{\alpha}\rho = \rho'$ can be obtained as a solution to (2.3.7), then (R_1^1, R_2^1) can be obtained. As noted in Section 2.1.2 at $\alpha = 0$,

$\sqrt{\bar{\alpha}}\rho = \rho = \rho^*$, and at $\alpha = 1$, (obviously) $\sqrt{\bar{\alpha}}\rho = 0$. The function $\sqrt{\bar{\alpha}}\rho$ is continuous in α and ρ. Thus, if ρ (as the solution of (2.3.7)) is a continuous function of α, then all values of $\sqrt{\bar{\alpha}}\rho$ between 0 and ρ^* can be achieved and, in particular any (R_1^1, R_2^1) on the boundary of C_o for which $R_1^1 > R_1^*$ is attained in A_1. The continuity of ρ as a function of α is a straightforward consequence of the continuity of $P(\alpha,\rho)$ in both α and ρ, and the fact that a unique root exists for all α.

The desired result (Theorem 2.1) has been proven, that

$$C_o = A_1 \cup A_2 \tag{2.3.17}$$

and the capacity region is given by C_o. We refer back to Figure 2.2 for an example of the capacity region.

2.4 Capacity Region with Correlated Messages and Feedback

In [11] Slepian and Wolf obtained the capacity region for the discrete memoryless MAC with messages which have a common part. That is, in addition to two private messages $\underline{a}_1$ and $\underline{a}_2$, known only by T_1 and T_2 respectively, there is a common message $\underline{a}_0$, known to both. Slepian and Wolf showed, defining

$$R_i = \frac{1}{N} H(\underline{a}_i) \quad i=0,1,2 \tag{2.4.1}$$

that the region

$$C_c = \text{co}\ \{\bigcup_{p\in P} R(p)\} \tag{2.4.2}$$

where

$$R(p) = \{(R_1,R_2):\quad R_1 \le I(X_1;Y|X_2U)$$

$$R_2 \le I(X_2;Y|X_1U)$$

$$R_1 + R_2 \le I(X_1X_2;Y|U)$$

$$R_0 + R_1 + R_2 \le I(X_1X_2;Y)\ \} \tag{2.4.3}$$

and

$$P = \{p(ux_1x_2y) = p(u)p(x_1|u)p(x_2|u)p(y|x_1x_2)\} \tag{2.4.4}$$

is the capacity region. (Actually they proved the weaker result that outside C_c $\Pr[\underline{a}_i \neq \hat{\underline{a}}_i]$ is bounded away from zero as opposed to a per-letter converse.)

P is the set of joint input-output probabilities where $U \to (X_1,X_2) \to Y$ form a Markov chain in that order, and (X_1X_2) are conditionally independent given u. We will now extend the results of Sections 2.1 and 2.2 appropriately to this case. We will show that if P above is replaced by P' where

$$P' = \{p(ux_1x_2y) = p(u)p(x_1x_2|u)p(y|x_1x_2)\} \tag{2.4.5}$$

that

$$C_c = \text{co}\{\bigcup_{p\in P'} R(p)\} \tag{2.4.6}$$

is an outer bound to the capacity region, and that it is achievable for gaussian channels.

To derive the outer bound we will make the following definition; $\underline{a}_1$ and $\underline{a}_2$ are as in Section 2.3, and $\underline{a}_0$ is the common message. The encoder outputs are then

$$x_{1k} = f_1(\underline{a}_1 \underline{a}_0 \underline{Y}^{k-1}) \tag{2.4.7a}$$

$$x_{2k} = f_2(\underline{a}_2 \underline{a}_0 \underline{Y}^{k-1}) \tag{2.4.7b}$$

Define

$$U_k = (\underline{a}_0, \underline{Y}^{k-1}) \tag{2.4.8}$$

Now, analogous to equation 2.2.17, we have

$$I(\underline{a}_1; \underline{Y} | \underline{a}_0 \underline{a}_2) \geq H(\underline{a}_1) - L^1 \eta_1(\varepsilon_1) \tag{2.4.9}$$

Also
$$H(\underline{Y} | \underline{a}_0 \underline{a}_2) = \sum_{k=1}^{N} H(Y_k | Y^{k-1} \underline{a}_0 \underline{a}_2)$$

a)
$$= \sum_{k=1}^{N} H(Y_k | X_{2k} Y^{k-1} \underline{a}_0 \underline{a}_2)$$

b)
$$= \sum_{k=1}^{N} H(Y_k | X_{2k} U_k \underline{a}_2)$$

c)
$$\leq \sum_{k=1}^{N} H(Y_k | X_{2k} U_k) \tag{2.4.10}$$

where a) is true when x_{2k} is a deterministic function of $(\underline{Y}^{k-1}, \underline{a}_0, \underline{a}_2)$, b) is by definition of U_k, and c) follows from the fact that conditioning does not increase entropy.

Similarly,

$$H(\underline{Y}|\underline{a}_0\underline{a}_2\underline{a}_1) = \sum_{k=1}^{N} H(Y_k|\underline{Y}^{k-1}\underline{a}_0\underline{a}_2\underline{a}_1)$$

a) $$= \sum_{k=1}^{N} H(Y_k|X_{1k}X_{2k}\underline{Y}^{k-1}\underline{a}_0\underline{a}_2\underline{a}_1)$$

b) $$= \sum_{k=1}^{N} H(Y_k|X_{1k}X_{2k})$$

$$= \sum_{k=1}^{N} H(Y_k|X_{1k}X_{2k}U_k) \qquad (2.4.11)$$

where a) follows from the deterministic encoders and b) from the fact that the output is independent of the past, given current inputs.

(2.4.10) and (2.4.11) may be substituted into (2.4.9) to obtain

$$\sum_{k=1}^{N} I(X_{1k};Y_k|X_{2k}U_k) \geq H(\underline{a}_1) - L^1\eta_1(\varepsilon_1) \qquad (2.4.12)$$

We can obtain a similar expression involving $H(\underline{a}_2)$. The bound on $R_1 + R_2$ follows by similarly showing that

$$I(\underline{a}_1\underline{a}_2;\underline{Y}|\underline{a}_0) \leq \sum_{k=1}^{N} I(X_{1k}X_{2k};Y_k|U_k) \qquad (2.4.13)$$

and the bound on $R_0 + R_1 + R_2$ from

$$I(\underline{a}_0\underline{a}_1\underline{a}_2;\underline{Y}) = H(\underline{Y}) - H(\underline{Y}|\underline{a}_0\underline{a}_1\underline{a}_2)$$

$$\leq \sum_{k=1}^{N} I(X_{1k}X_{2k};Y_k) \qquad (2.4.14)$$

All of the subsequent steps of the derivation of Section 3.2.1 are the same. The sums in (2.4.12), (2.4.13) and (2.4.14) can be reduced to single letter quantities by convexity considerations and the fact that $U_k \rightarrow (X_{1k},X_{2k}) \rightarrow Y_k$ is a Markov chain.

To show that the region defined by (2.4.6) is achievable for the AWGN case we use the following superposition: let $\underline{a}_0$ be mapped into a codeword of block length N, whose letters are drawn independent identically distributed from a gaussian distribution $N(0,1)$. Then let T_1 amplify the codeword by $\sqrt{\alpha_1 P_1}$ and T_2 by $\sqrt{\alpha_2 P_2}$ where $0 \leq \alpha_i \leq 1$. The transmitters then use the remainder of their powers ($\bar{\alpha}_i P_i$, i=1,2) to cooperate as in Section 2.1, where the correction terms are computed at the transmitters by assuming that the receiver will correctly decode $\underline{a}_0$. At the end of a block the decoder then subtracts the superimposed codeword and proceeds as in Section 2.1.2. Clearly this scheme can achieve any R_0 in the range

$$0 \leq R_0 \leq \frac{1}{2}\ln\left(1 + \frac{(\sqrt{P_1} + \sqrt{P_2})^2}{\sigma^2}\right) \qquad (2.4.15)$$

and a derivation similar to the one in 2.2.2 and 2.3 shows that the intersection of the achievable region with any plane of the form $R_0 = R_0^*$ coincides with the intersection of the outer bound with that plane.

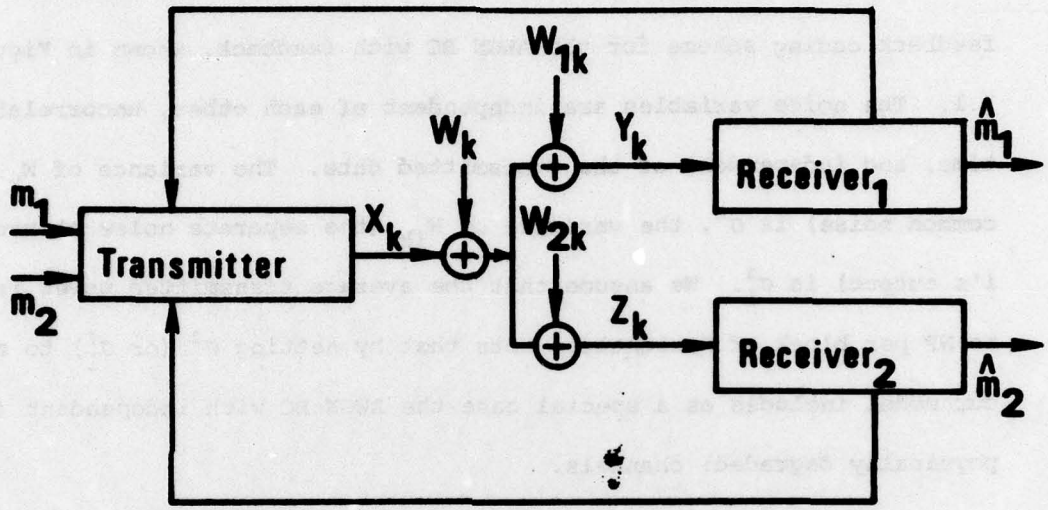
CHAPTER 3

AWGN BROADCAST CHANNEL WITH FEEDBACK

In this chapter we introduce a constructive deterministic feedback coding scheme for the AWGN BC with feedback, shown in Figure 3.1. The noise variables are independent of each other, uncorrelated in time, and independent of the transmitted data. The variance of W_k (the common noise) is σ^2, the variance of W_{ik} (the separate noise at receiver i's output) is σ_i^2. We assume that the average transmitted power is limited to NP per block of N signals. Note that by setting σ^2 (or σ_1^2) to zero, our model includes as a special case the AWGN BC with independent (or physically degraded) channels.

As mentioned in Section 1.2.2 it has been shown that for both the discrete memoryless [8] and AWGN [16] BC, feedback does not enlarge the capacity region if one channel is a physically degraded version of the other. The results of this chapter will show that such is not the case when the channels are not physically degraded, at least for the AWGN case. Since all AWGN broadcast channels are degraded (that is, there exists a cascade channel with the same marginal probability density functions, conditioned on the input), the result of [8] and [16] will have been shown to fail to apply to a class of (continuous) degraded channels.

The contents of this chapter parallel those of Chapter 2. In Section 3.1 we present the feedback coding scheme, and obtain an expression for an achievable region. In Section 3.2 we obtain first



$$W_k \sim \mathcal{N}(0,\sigma^2)$$

$$W_{1k} \sim \mathcal{N}(0,\sigma_1^2)$$

$$W_{2k} \sim \mathcal{N}(0,\sigma_2^2)$$

3.1 AWGN Broadcast Channel with Feedback.

an outer bound for the capacity region of arbitrary discrete memoryless BC's with feedback, and then extend this result to the AWGN case. In Section 3.3 we will compare the achievable regions and outer bounds for some examples.

3.1 Achievable Region

For the problem we are considering, a single transmitter wishes to communicate two messages, θ_1 and θ_2, to two separate receivers using the feedback channel of Figure 3.1. The transmitter will use a single block of N transmissions to communicate both messages. We assume that the feeedback links are noiseless and delayless. In this section the outputs Y_k and Z_k in Figure 3.1 are denoted by r_{1k} and r_{2k} for the sake of notational compactness.

Assume that after k-1 transmissions the receivers have their respective estimates of θ_1 and θ_2 in the form

$$\begin{aligned} \hat{\theta}_1^{k-1} &= \theta_1 + \xi_{k-1} \\ \hat{\theta}_2^{k-1} &= \theta_2 + \eta_{k-1} \end{aligned} \tag{3.1.1}$$

ξ_{k-1} and η_{k-1} are jointly gaussian with zero means, variances a_{k-1} and b_{k-1} and correlation coefficient ρ_k. As in the case of the MAC, we assume that the estimates are deterministic functions of the previously received data, and are therefore available to the transmitter. Since the transmitter also knows the messages, it can form ξ_{k-1} and η_{k-1}.

For the k^{th} transmission, the transmitter forms a linear combination of ξ_{k-1} and η_{k-1}, normalizes it to average power P, and transmits it. Let the transmitted signal x_k be

$$x_k = \sqrt{\frac{P}{\mathcal{D}}} \left[\frac{\xi_{k-1}}{\sqrt{a_{k-1}}} + \frac{\eta_{k-1}}{\sqrt{b_{k-1}}} \; g \; \mathrm{sgn}(\rho_{k-1}) \right] \tag{3.1.2}$$

where $\mathcal{D}$ is the mean squared value of the term in square brackets, given by

$$\mathcal{D} = 1 + g^2 + 2g|\rho_{k-1}| \tag{3.1.3}$$

The parameter $g \in [0,\infty)$ may be varied to allot a varying proportion of the available transmitter power to communication with receiver one vs. receiver two. It is readily seen that with $\mathcal{D}$ chosen as above, the variance of x_k (i.e., the transmitted power) is P.

The receivers then receive noisy versions of x_k, where r_{ik} (i=1,2), which is the i^{th} receiver's channel output at the k^{th} time, is given by

$$r_{ik} = x_k + w_k + w_{ik} \tag{3.1.4}$$

where w_k is the common noise with variance σ^2, and w_{ik} is the separate noise at receiver i, with variance σ_i^2.

The receivers then form their new estimates of the appropriate messages, given by

$$\hat{\theta}_1^k = \theta_1 + \xi_k = \theta_1^{k-1} - \frac{\overline{r_{1k}\xi_{k-1}}}{\overline{r_{1k}^2}} r_{1k}$$

$$\hat{\theta}_2^k = \theta_2^{k-1} - \frac{\overline{r_{2k}\eta_{k-1}}}{\overline{r_{2k}^2}} r_{2k} \qquad (3.1.5)$$

Using the above, it is easy to verify that for i = 1,2

$$\overline{r_{ik}^2} = P + \sigma^2 + \sigma_i^2$$

$$\overline{r_{ik}\xi_{k-1}} = \sqrt{\frac{P}{D}}\sqrt{a_{k-1}}(1 + g|\rho_{k-1}|) \qquad (3.1.6)$$

$$\overline{r_{ik}\eta_{k-1}} = \sqrt{\frac{P}{D}}\sqrt{b_{k-1}}\,\mathrm{sgn}(\rho_{k-1})(g + |\rho_{k-1}|)$$

From equation (3.1.5) we see that

$$\xi_k = \xi_{k-1} - \frac{\overline{r_{1k}\xi_{k-1}}}{\overline{r_{1k}^2}} r_{1k} \qquad (3.1.7a)$$

$$\eta_k = \eta_{k-1} - \frac{\overline{r_{2k}\eta_{k-1}}}{\overline{r_{2k}^2}} r_{2k} \qquad (3.1.7b)$$

These may be squared and averaged to obtain the new variances given by

$$a_k = a_{k-1}\frac{\sigma^2+\sigma_1^2+\frac{P}{D}g^2(1-\rho_{k-1}^2)}{\sigma^2+\sigma_1^2+P}$$

$$b_k = b_{k-1}\frac{\sigma^2+\sigma_2^2+\frac{P}{D}(1-\rho_{k-1}^2)}{\sigma^2+\sigma_2^2+P} \qquad (3.1.8)$$

Similarly, if we define $c_k = \overline{\xi_k \eta_k}$, and $T = (\overline{r_{ik} \xi}_{k-1})(\overline{r_{ik} \eta}_{k-1})$, then we can multiply (3.1.7a) and (3.1.7b), take expected values, and obtain

$$c_k = c_{k-1} - T\left(\frac{1}{r_{1k}^2} + \frac{1}{r_{2k}^2} - \frac{\overline{r_{1k} r_{2k}}}{r_{1k}^2 r_{2k}^2}\right) \tag{3.1.9}$$

Also

$$\overline{r_{1k} r_{2k}} = P + \sigma^2 \tag{3.1.10}$$

By substituting (3.1.6) and (3.1.10) into (3.1.9) and using the definition $c_k = \sqrt{a_k b_k}\, \rho_k$, we obtain the following recursive expression for ρ_k:

$$\rho_k = \frac{(\sigma^2 \Sigma + \sigma_1^2 \sigma_2^2)\rho_{k-1} - \frac{P\Sigma}{D} g\, \mathrm{sgn}(\rho_{k-1})(1-\rho_{k-1}^2)}{\sqrt{\Pi}\sqrt{N_1 N_2}} \tag{3.1.11}$$

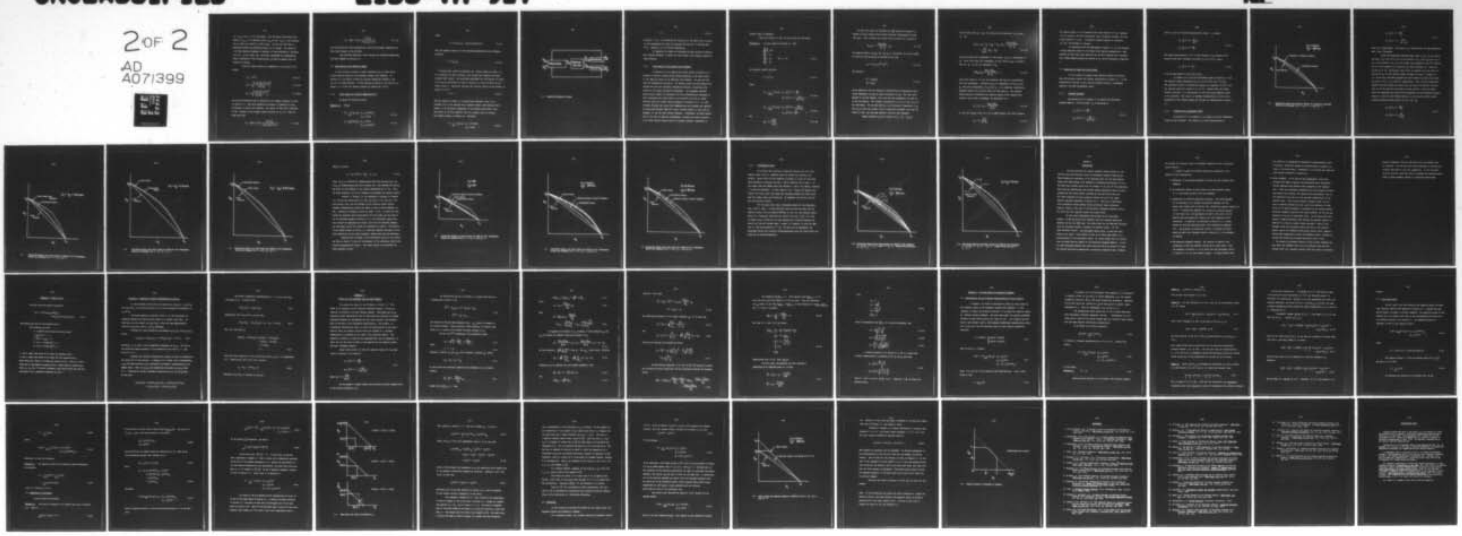
where

$$\begin{aligned} \Sigma &= P+\sigma^2+\sigma_1^2+\sigma_2^2 \\ \Pi &= (P+\sigma^2+\sigma_1^2)(P+\sigma^2+\sigma_2^2) \\ N_1 &= \sigma^2+\sigma_1^2+\frac{P}{D} g^2 (1-\rho_{k-1}^2) \\ N_2 &= \sigma^2+\sigma_2^2+\frac{P}{D}(1-\rho_{k-1}^2) \end{aligned} \tag{3.1.12}$$

As in the case of (2.1.10) for the MAC, equation (3.1.11) may be shown to have a solution in the sense that there exists a $\rho \in [0,1]$ such that if $\rho_{k-1} = \rho$, then $\rho_k = -\rho$. To see this, one can substitute

AD-A071 399 MASSACHUSETTS INST OF TECH CAMBRIDGE LAB FOR INFORMA--ETC F/G 9/4
CODING AND CAPACITY FOR ADDITIVE WHITE GAUSSIAN NOISE MULTI-USE--ETC(U)
MAY 79 L H OZAROW N00014-75-C-1183
UNCLASSIFIED LIDS-TH-917 NL

2 OF 2
AD A071399



END
DATE FILMED
8-79
DDC

$\rho_k = -\rho_{k-1}$ into (3.1.11) and square. The resulting sixth-order polynomial in ρ_{k-1} is of opposite sign at $\rho_{k-1}=0$ and $\rho_{k-1}=1$, and therefore has at least one solution in that range. We have not been able to determine whether the resulting value of ρ is unique. The values of ρ for the numerical examples in Section 3.3 were obtained by iterating (3.1.11). In all cases, the iterations converged to an appropriate value, regardless of the starting point, so that we suspect that the solution is unique.

Given the stable value of ρ, equations (3.1.11) and (3.1.8) become

$$\rho_k = (-1)^k \rho \tag{3.1.13}$$

$$a_k = a_0 \left[\frac{\sigma^2+\sigma_1^2+ \frac{P}{D} g^2(1-\rho^2)}{P+\sigma^2+\sigma_1^2} \right]^k \tag{3.1.14a}$$

$$b_k = b_0 \left[\frac{\sigma^2+\sigma_2^2+ \frac{P}{D} (1-\rho^2)}{P+\sigma^2+\sigma_2^2} \right]^k \tag{3.1.14b}$$

and the initialization may be carried out in a manner analogous to that in Section 2.1. The error probability analysis is identical to that of Section 2.1 and is not repeated. The result is that with ρ selected as the (largest if not unique) stable solution of (3.1.11), then all rates such that

$$R_1 < R_1^*(g) = \frac{1}{2} \ln \frac{P+\sigma^2+\sigma_1^2}{\sigma^2+\sigma_1^2+ \frac{P}{D} g^2(1-\rho^2)} \tag{3.1.15}$$

$$R_2 < R_2^*(g) = \frac{1}{2} \ln \frac{P+\sigma^2+\sigma_2^2}{\sigma^2+\sigma_2^2+\frac{P}{D}(1-\rho^2)} \qquad (3.1.15)$$

are achievable with error probabilities which decay doubly exponentially with block length, as for the MAC.

The solutions obtained in this section are evaluated numerically for some examples in Section 3.3.

3.2 Outer Bound to the Capacity Region

In this section we derive a simple expression for an outer bound to the capacity region of the broadcast channel with feedback. In Section 3.2.1 we derive a bound for discrete memoryless channels, and in 3.2.2 for AWGN channels. In this section we revert to the notation of Figure 3.1, in that the channel outputs are denoted by Y and Z.

3.2.1 Outer Bound for Discrete Memoryless BC's

We prove the following result

Theorem 3.1 Define

$$C_1 = \bigcup_{p \in P} \{(R_1,R_2): R_1 \leq I(X;YZ|U) \quad R_2 \leq I(U;Z) \} \qquad (3.2.1)$$

$$C_2 = \bigcup_{p \in P} \{(R_1,R_2): R_1 \leq I(U;Y) \quad R_2 \leq I(X;YZ|U)\} \qquad (3.2.2)$$

where

$$P = \{p(u,x,y,z) = p(u)p(x|u)p(yz|x)\} \tag{3.2.3}$$

Then the capacity region C of the discrete memoryless BC with feedback satisfies

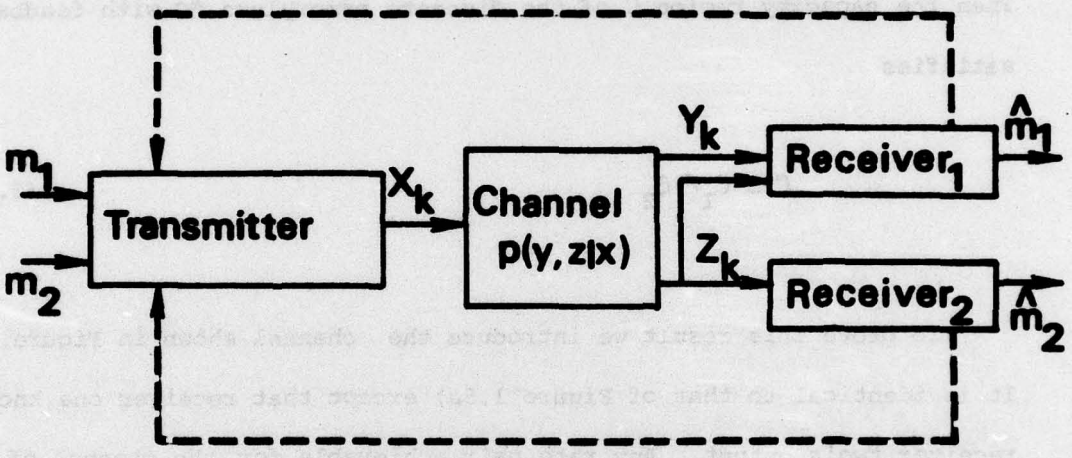
$$C \subset C_1 \cap C_2 \tag{3.2.4}$$

To prove this result we introduce the channel shown in Figure 3.2. It is identical to that of Figure 1.5a) except that receiver one knows receiver two's output. Any rate pair achievable for the channel of Figure 1.5a) must be achievable for that of Figure 3.2, since receiver one can always ignore Z. Therefore, denoting the capacity region of the channel in Figure 3.2 by C',

$$C \subset C' \tag{3.2.5}$$

But the channel of Figure 3.2 is physically degraded, since if we consider (Y,Z) to be receiver one's channel output, then receiver two's output, Z, is trivially independent of the channel input given (Y,Z). The result of [8] then applies, that is, feedback does not enlarge the capacity region in Figure 3.2. Therefore

$$C' = \bigcup_{p \in P} \{(R_1,R_2): R_1 \leq I(X;YZ|U) \quad R_2 \leq I(U;Z)\} \tag{3.2.6}$$



3.2 Augmented Broadcast Channel.

$$= C_1 \tag{3.2.6}$$

Therefore $C \subset C_1$. By reversing the direction of the added link in Figure 3.2 and exchanging the roles of receivers one and two, it follows that $C \subset C_2$. Equation (3.2.4) follows immediately.

In Appendix D we apply the technique of this section to general BC's without feedback, to obtain an outer bound to the capacity region of those channels.

3.2.2 Outer Bound for the AWGN BC with Feedback

In Section 2.2.2 we applied the formal result of Section 2.2.1 directly to obtain a single letter characterization of the outer bound on the capacity region of the AWGN MAC with feedback. We used the fact that the differential entropy of the channel output was well-defined and bounded and that the resulting information theoretic constraints were concave in the input probability assignment. For broadcast channels where neither channel is noiseless, the use of differential entropies is still justified, but since $I(U;Z)$ is not concave in $p(u,x)$ we cannot directly apply the single letter expression of Section 3.2.1. In [16], El-Gamal obtained the result that feedback does not increase the capacity of physically degraded AWGN BC's using a method similar to that used by Bergmans [7] for the case without feedback. Furthermore, El-Gamal showed that in the case of physical degradedness, allowing the better receiver to see both channel outputs does not increase capacity, regardless of

whether there is feedback.

Using the results of [16], we will prove the following:

Theorem 3.2 For the channel of Figure 3.1, with

$$\begin{aligned} &\sum_{k=1}^{N} \overline{x_k^2} \le NP \\ &\overline{w_k^2} = \sigma^2 \\ &\overline{w_{1k}^2} = \sigma_1^2 \qquad (k = 1,\ldots N) \\ &\overline{w_{2k}^2} = \sigma_2^2 \end{aligned} \tag{3.2.7}$$

the capacity region satisfies

$$C \subset C_1 \cap C_2 \tag{3.2.8}$$

where

$$C_1 = \bigcup_{0\le\alpha\le 1} \{(R_1,R_2): R_1 \le \frac{1}{2} \ln\left(1 + \frac{\alpha P}{\sigma_e^2}\right) \quad R_2 \le \frac{1}{2} \ln\left(1 + \frac{\bar{\alpha} P}{\alpha P+\sigma^2+\sigma_2^2}\right) \} \tag{3.2.9a}$$

$$C_2 = \bigcup_{0\le\alpha\le 1} \{(R_1,R_2): R_1 \le \frac{1}{2} \ln\left(1 + \frac{\bar{\alpha} P}{\alpha P+\sigma^2+\sigma_1^2}\right) \quad R_2 \le \frac{1}{2} \ln\left(1 + \frac{\alpha P}{\sigma_e^2}\right) \} \tag{3.2.9b}$$

and

$$\sigma_e^2 = \sigma^2 + \frac{\alpha^2\sigma_2^2}{\sigma_1^2+\sigma_2^2} \tag{3.2.10}$$

To show this result we consider the AWGN version of Figure 3.2. Assume a coding strategy which allows reliable communication at some rate pair. Then consider the problem where receiver one's output is

$$(S_k, Z_k) = \left(\frac{\sigma_1^2 Z_k + \sigma_2^2 Y_k}{\sigma_1^2 + \sigma_2^2}, Z_k \right) \tag{3.2.11}$$

The mapping between (S_k, Z_k) and (Y_k, Z_k) is invertible (as long as $\sigma_2^2 \neq 0$), so that any code involving an encoder of the form

$$X_k = f(\theta_1, \theta_2, \underline{Y}^{k-1}, \underline{Z}^{k-1}) \tag{3.2.12}$$

and decoders

$$\begin{aligned} \hat{\theta}_1 &= g_1(\underline{Y}, \underline{Z}) \\ \hat{\theta}_2 &= g_2(\underline{Z}) \end{aligned} \tag{3.2.13}$$

can be applied to the new channel by substituting the appropriate functional dependence on $\underline{S}, \underline{Z}$ for every occurrence of $\underline{Y}, \underline{Z}$. The code, when applied to the new channel, will have the same probability of error as on the old channel. The channels represented by $((Y,Z),Z)$ and $((S,Z),Z)$ are equivalent. We now show that Z_k is statistically independent of X_k given S_k so that the channel (S,Z) is physically degraded, and from the result in [16], has the same capacity $((S,Z),Z)$ with feedback.

Random variable Z_k may be written as $Z_k = S_k + (Z_k - S_k)$.

We need show that $(Z_k - S_k)$ is statistically independent of X_k given S_k. But

$$Z_k - S_k = X_k + W_k + W_{2k} - \left(X_k + W_k + \frac{\sigma_1^2 W_{2k} + \sigma_2^2 W_{1k}}{\sigma_1^2 + \sigma_2^2}\right)$$
$$= \frac{\sigma_2^2}{\sigma_1^2+\sigma_2^2}(W_{2k} - W_{1k}) \qquad (3.2.14)$$

Since the noises are independent of the signal, $Z_k - S_k$ is independent of X_k. Since the noises are independent of each other, $Z_k - S_k$ is independent of W_k. As for the remainder of S_k,

$$E\left[(W_{2k} - W_{1k})\left(\frac{\sigma_2^2 W_{1k} + \sigma_1^2 W_{2k}}{\sigma_1^2 + \sigma_2^2}\right)\right] = 0 \qquad (3.2.15)$$

Since both terms in (3.2.15) are gaussian, and they are uncorrelated, they are independent. Therefore $Z_k - S_k$ is independent of both X_k and S_k, and hence independent of X_k given S_k. Z is therefore a physically degraded version of S and the result of [16] applies. The capacity region desired is therefore the capacity region of the channel with outputs S and Z with no feedback. By definition of S,

$$S_k = X_k + W_k + \frac{\sigma_1^2 W_{2k} + \sigma_2^2 W_{1k}}{\sigma_1^2 + \sigma_2^2} \qquad (3.2.16)$$

so that the channel from X to S is an AWGN channel with noise variance

$$\sigma_e^2 = \sigma^2 + \frac{\sigma_1^2 \sigma_2^2}{\sigma_1^2 + \sigma_2^2} \qquad (3.2.17)$$

The channel from X to Z is gaussian with noise variance $\sigma^2 + \sigma_2^2$. Equation (1.2.27) applies, with the appropriate noise variances inserted, and this yields equation (3.2.9a). A symmetric argument applies in obtaining (3.2.9b), and Theorem 3.2 follows.

In connection with the AWGN model of Figure 3.2, we have obtained a coding approach suited to this model which employs superposition, made possible in this case since receiver one can form receiver two's estimate. This coding scheme achieves all points in C_1, and is discussed in Appendix C.

3.3 Comparison of Inner and Outer Bounds

In this section we present some numerical results for achievable and converse regions for various realizations of Figure 3.1. In the subsections to follow, we will discuss degraded channels, independent channels, and some intermediate cases.

3.3.1. Degraded Channels

By setting $\sigma_1^2 = 0$ in Figure 3.1 we obtain the physically degraded AWGN BC. For this model, C_1 is specified by

$$R_1 \leq \frac{1}{2} \ln \left(1 + \frac{\alpha P}{\sigma^2}\right)$$

$$R_2 \leq \frac{1}{2} \ln \left(1 + \frac{\bar{\alpha} P}{\alpha P + \sigma^2 + \sigma_2^2}\right) \qquad (3.3.1)$$

which is just the standard superposition region. C_2 becomes

$$R_1 \leq \frac{1}{2} \ln \left(1 + \frac{\bar{\alpha}P}{\alpha P+\sigma^2}\right)$$

$$R_2 \leq \frac{1}{2} \ln \left(1 + \frac{\alpha P}{\sigma^2}\right) \qquad (3.3.2)$$

The region described by (3.3.2) is the triangle in R_2 bounded by the positive axes and a straight line from $(C_1,0)$ to $(0,C_1)$, where

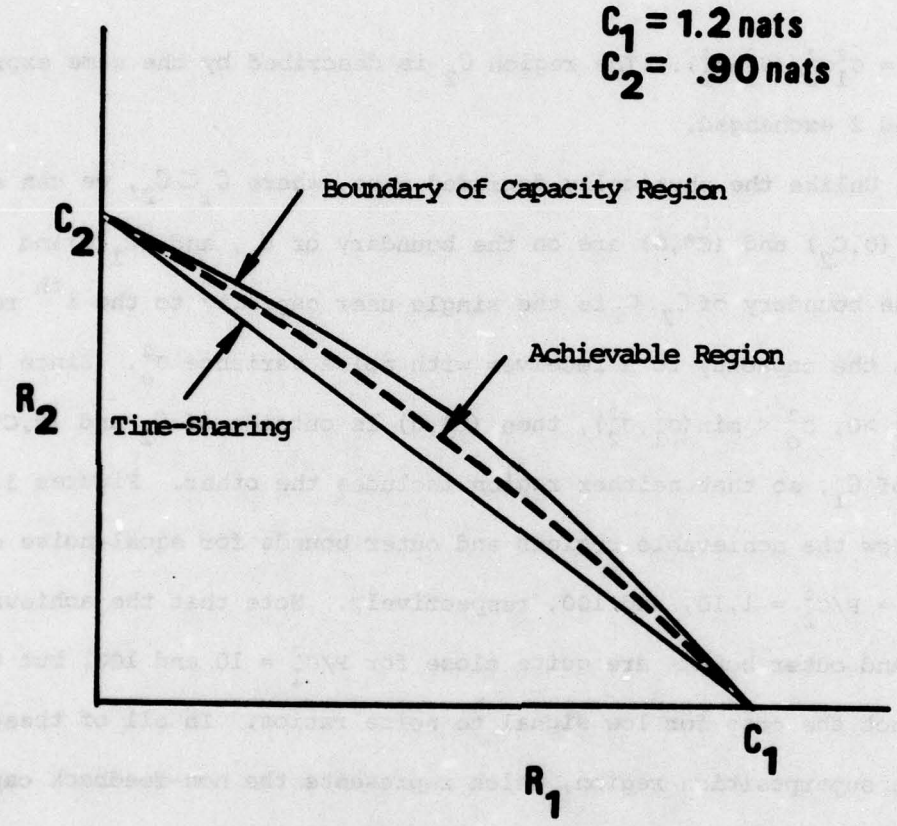
$$C_1 = \frac{1}{2} \ln \left(1 + \frac{P}{\sigma^2}\right) \qquad (3.3.3)$$

It is an easy matter to show that $C_1 \subset C_2$.

In Figure 3.3 we plot the achievable region of Section 3.1 and the capacity region represented by (3.3.1) for P=10, and $\sigma^2 = \sigma_2^2 = 1$. The achievable region is obtained by varying g in equation (3.1.15), and the capacity region by varying α in (3.3.2). Observe that the coding scheme of Section 3.1 is sub-optimal in that the entire capacity region is not achieved. This result is somewhat disconcerting in view of the optimality of the related scheme for the MAC, as demonstrated in Section 2.3.

3.3.2 Channels with Independent Noise

By setting $\sigma^2 = 0$ in Figure 3.1, we obtain a BC with independent noises at the receivers. The region C_1 is then characterized by



3.3 Achievable Region and Capacity Region for Physically Degraded AWGN BC with Feedback for $P = 10$, $\sigma^2 = 1$ and $\sigma_2^2 = 1$.

$$R_1 \leq \frac{1}{2} \ln \left(1 + \frac{\alpha P}{\sigma_e^2}\right)$$

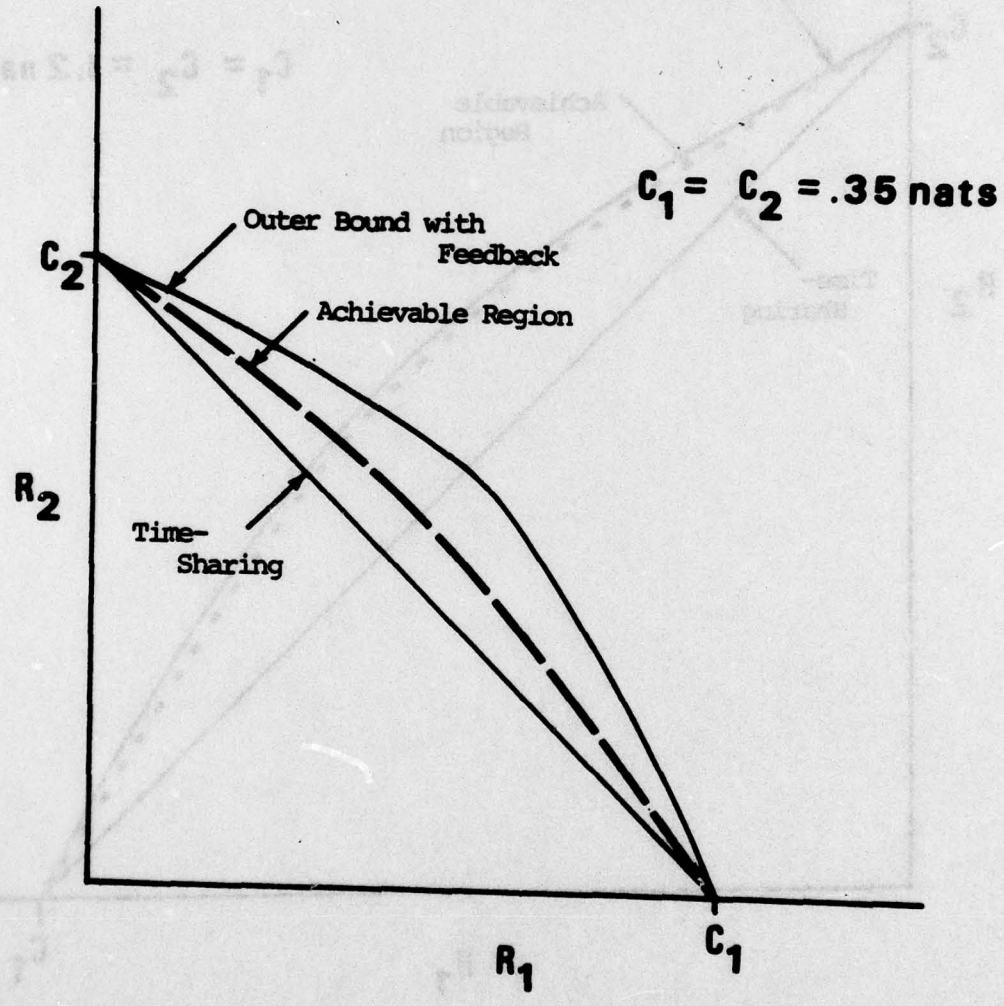
$$R_2 \leq \frac{1}{2} \ln \left(1 + \frac{\bar{\alpha} P}{\alpha P + \sigma_2^2}\right) \qquad (3.3.4)$$

where $\sigma_e^2 = \sigma_1^2\sigma_2^2/(\sigma_1^2+\sigma_2^2)$. The region C_2 is described by the same expression with 1 and 2 exchanged.

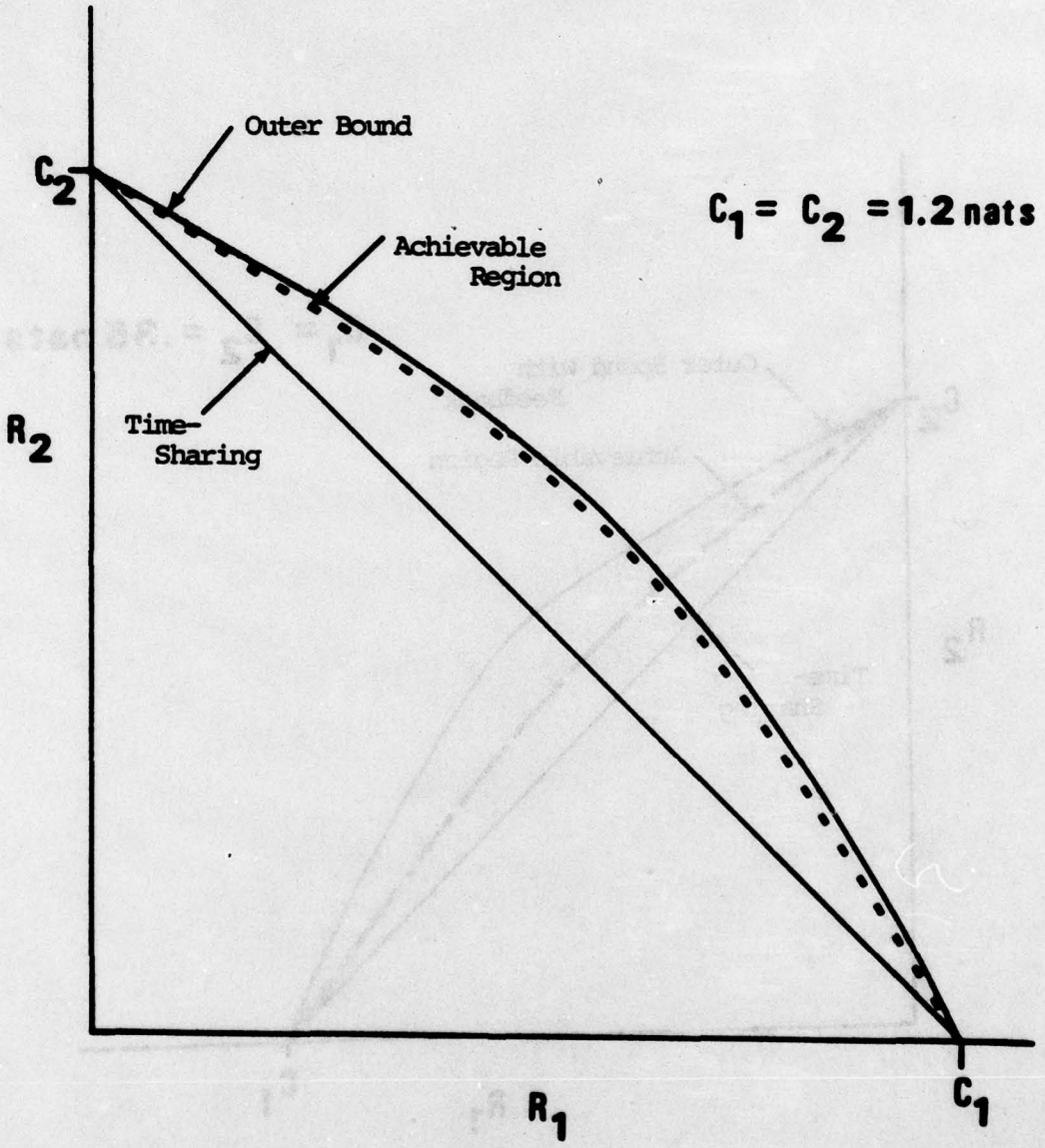
Unlike the physically degraded case, where $C_1 \subset C_2$, we can easily see that $(0,C_2)$ and $(C^*,0)$ are on the boundary of C_1, and $(C_1,0)$ and $(0,C^*)$ are on the boundary of C_2. C_i is the single user capacity to the i^{th} receiver, and C^* is the capacity to a receiver with noise variance σ_e^2. Since for σ_1^2 and $\sigma_2^2 > 0$, $\sigma_e^2 < \min(\sigma_1^2,\sigma_2^2)$, then $(C^*,0)$ is outside of C_2 and $(0,C^*)$ is outside of C_1, so that neither region includes the other. Figures 3.4 to 3.6 show the achievable regions and outer bounds for equal noise cases, for $P/\sigma_1^2 = P/\sigma_2^2 = 1, 10$, and 100, respectively. Note that the achievable regions and outer bounds are quite close for $P/\sigma_1^2 = 10$ and 100, but that this is not the case for low signal to noise ratios. In all of these cases the superposition region, which represents the non-feedback capacity region, is just the time sharing region, as can be seen from the superposition equations which become, for $\sigma_1^2 = \sigma_2^2 = \sigma_c^2$,

$$R_1 \leq \frac{1}{2} \ln \left(1 + \frac{\alpha P}{\sigma_c^2}\right)$$

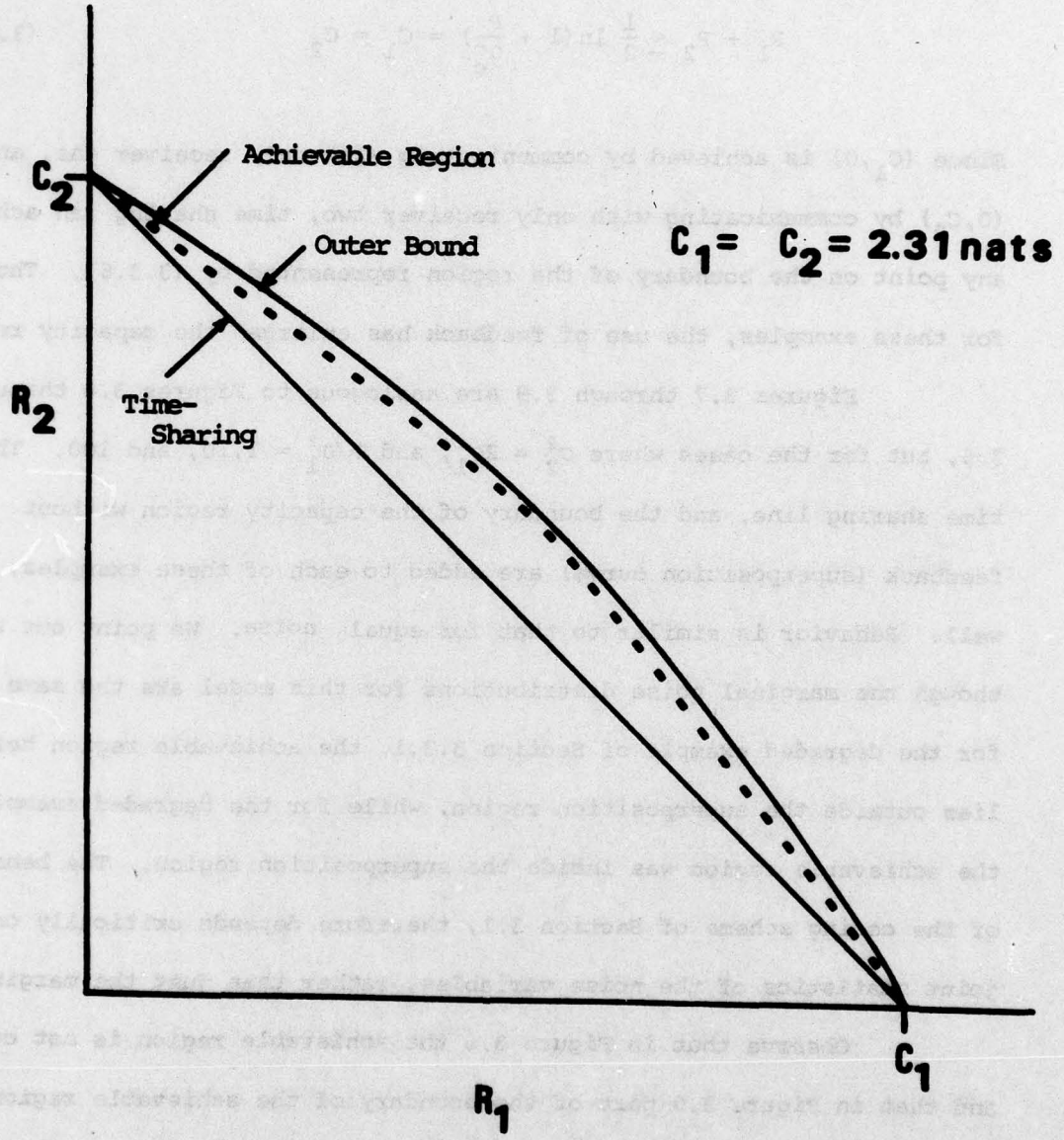
$$R_2 \leq \frac{1}{2} \ln \left(1 + \frac{\bar{\alpha} P}{\alpha P + \sigma_c^2}\right) \qquad (3.3.5)$$



3.4 Achievable Region and Outer Bound for AWGN BC with Independent Noises and Feedback for $P = 1$, $\sigma_1^2 = \sigma_2^2 = 1$.



3.5 Achievable Region and Outer Bound for AWGN BC with Independent Noises and Feedback for P = 10, $\sigma_1^2 = \sigma_2^2 = 1$.



3.6 Achievable Region and Outer Bound for AWGN BC with Independent Noises and Feedback for P = 100, $\sigma_1^2 = \sigma_2^2 = 1$.

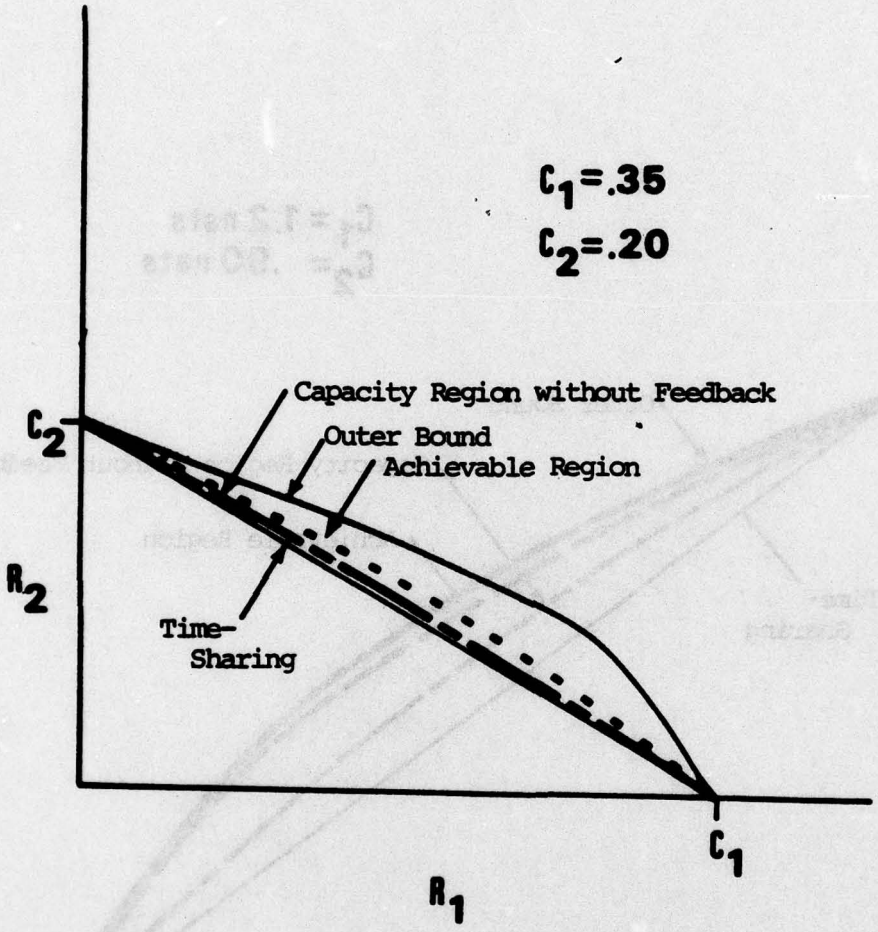
Adding, we obtain

$$R_1 + R_2 \leq \frac{1}{2} \ln(1 + \frac{P}{\sigma_c^2}) = C_1 = C_2 \qquad (3.3.6)$$

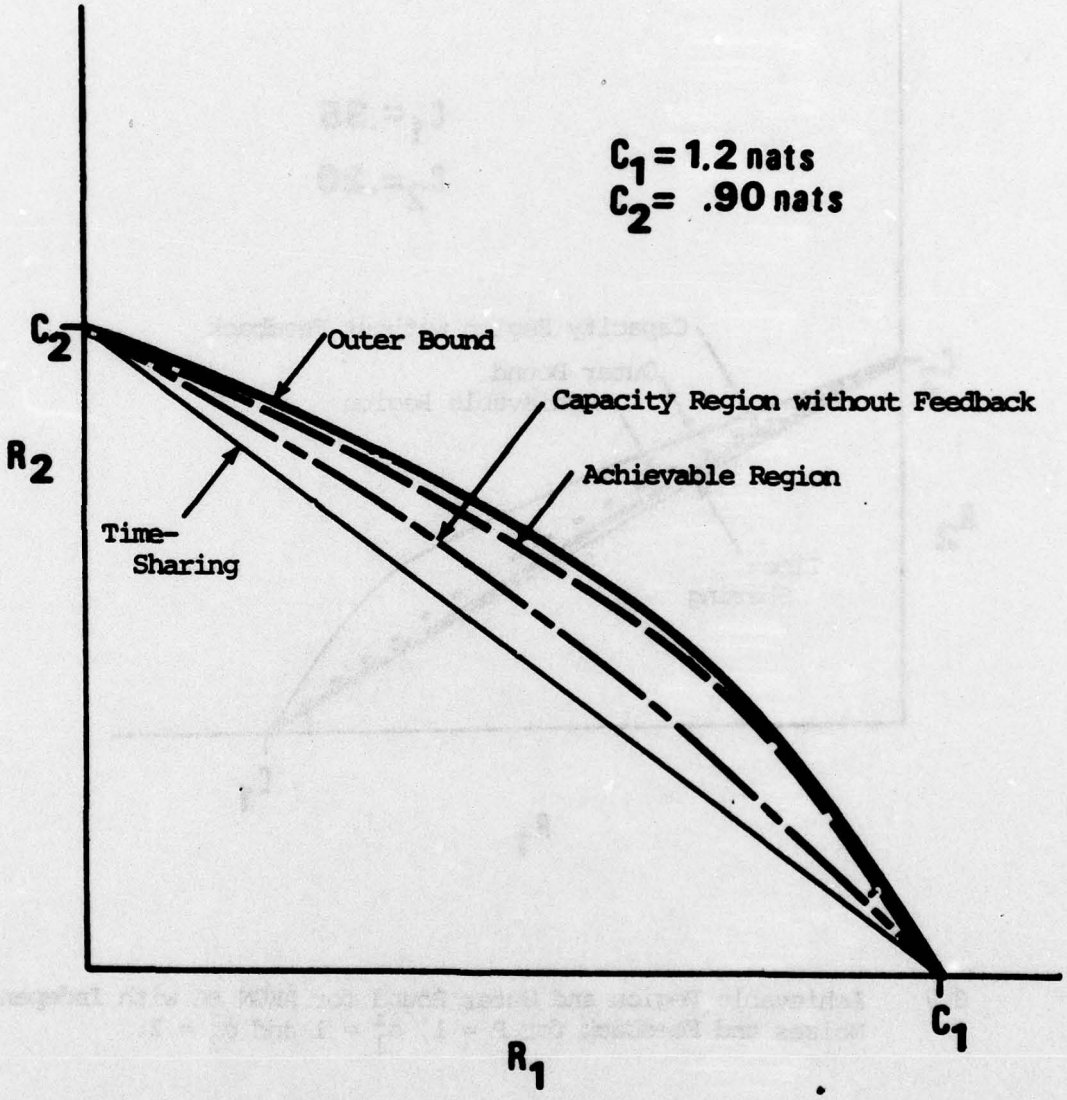
Since $(C_1,0)$ is achieved by communicating with only receiver one, and $(0,C_2)$ by communicating with only receiver two, time sharing can achieve any point on the boundary of the region represented by (3.3.6). Thus, for these examples, the use of feedback has enlarged the capacity region.

Figures 3.7 through 3.9 are analogous to Figures 3.4 through 3.6, but for the cases where $\sigma_2^2 = 2\sigma_1^2$, and $P/\sigma_1^2 = 1,10$, and 100. The time sharing line, and the boundary of the capacity region without feedback (superposition curve) are added to each of these examples, as well. Behavior is similar to that for equal noise. We point out that though the marginal noise distributions for this model are the same as for the degraded example of Section 3.3.1, the achievable region here lies outside the superposition region, while for the degraded example, the achievable region was inside the superposition region. The behavior of the coding scheme of Section 3.1, therefore depends critically on the joint statistics of the noise variables, rather than just the marginals.

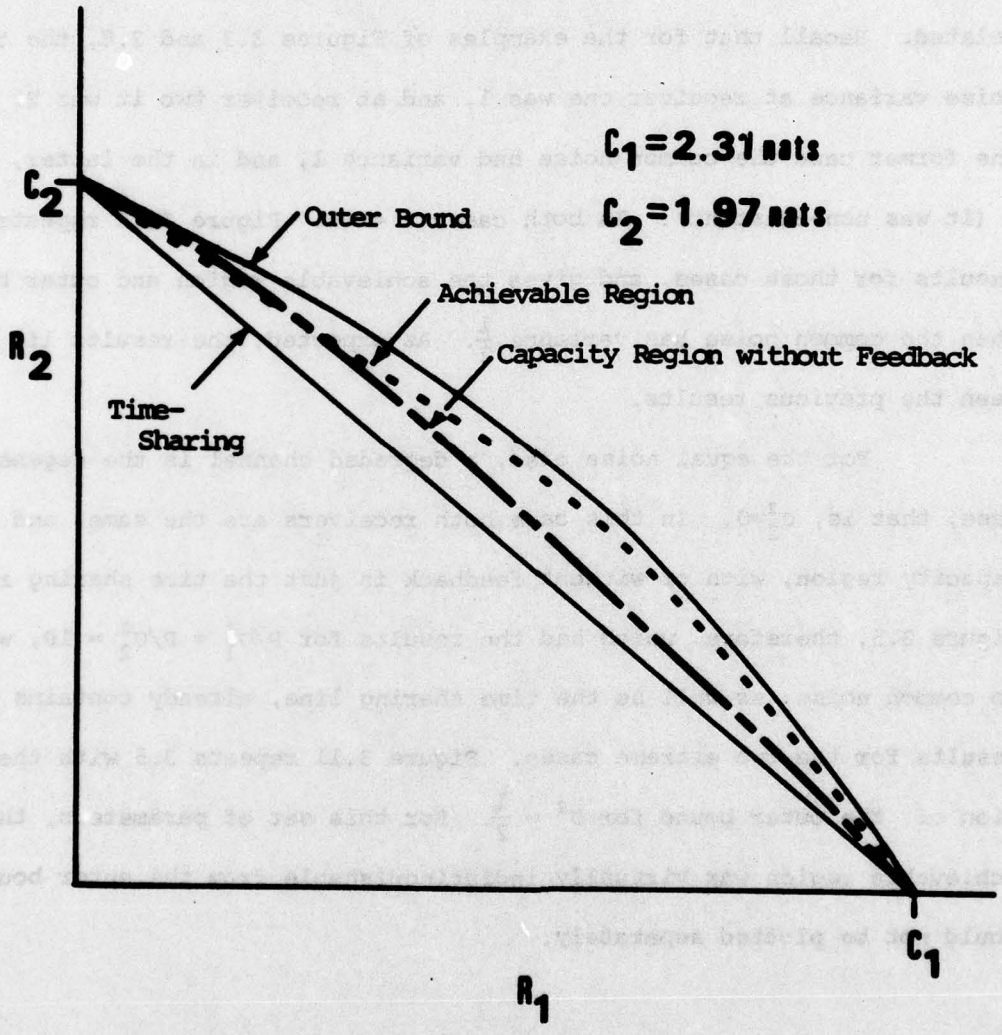
Observe that in Figure 3.6 the achievable region is not convex, and that in Figure 3.9 part of the boundary of the achievable region lies inside the superposition region. The coding scheme is sub-optimal for these examples as well.



3.7 Achievable Region and Outer Bound for AWGN BC with Independent Noises and Feedback for P = 1, $\sigma_1^2 = 1$ and $\sigma_2^2 = 2$.



3.8 Achievable Region and Outer Bound for AWGN BC with Independent Noises and Feedback for P = 10, $\sigma_1^2 = 1$ and $\sigma_2^2 = 2$.

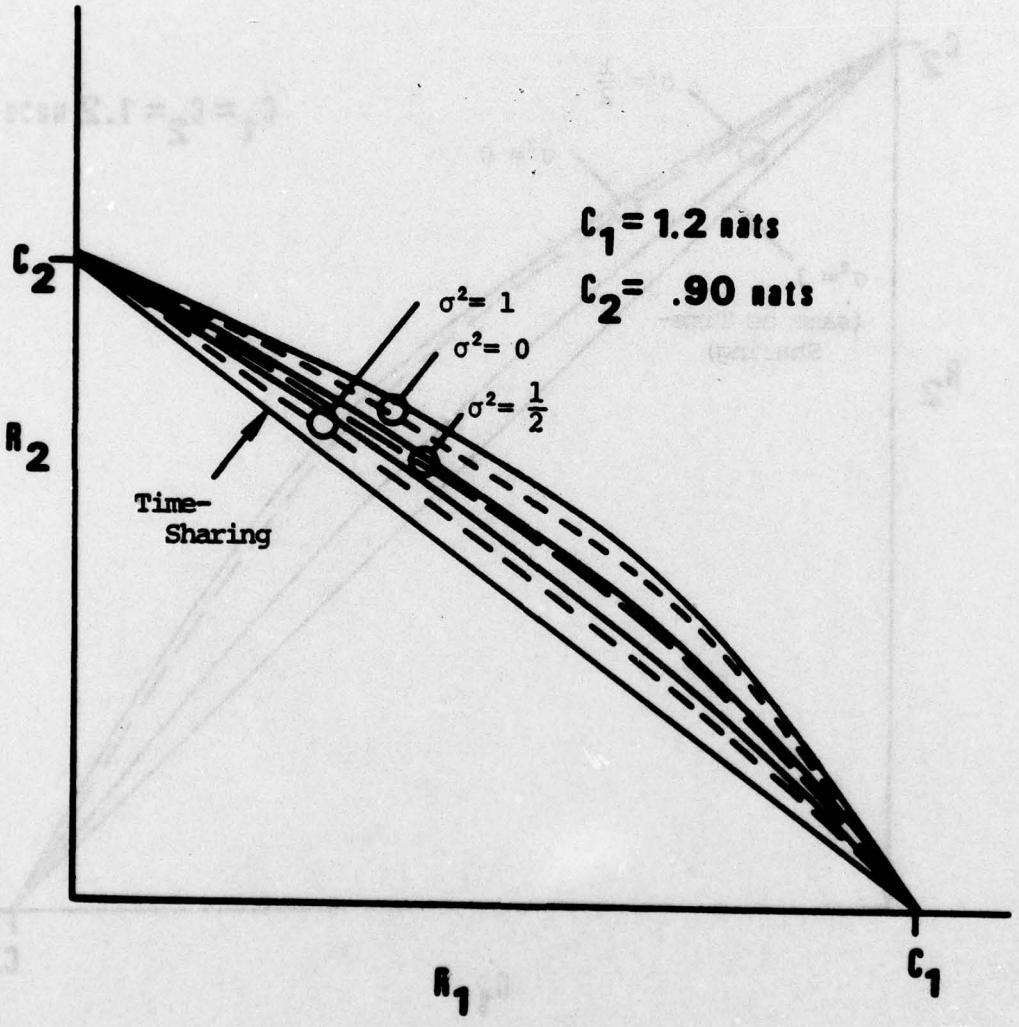


3.9 Achievable Region and Outer Bound for AWGN BC with Independent Noises and Feedback for P = 100, $\sigma_1^2 = 1$ and $\sigma_2^2 = 2$.

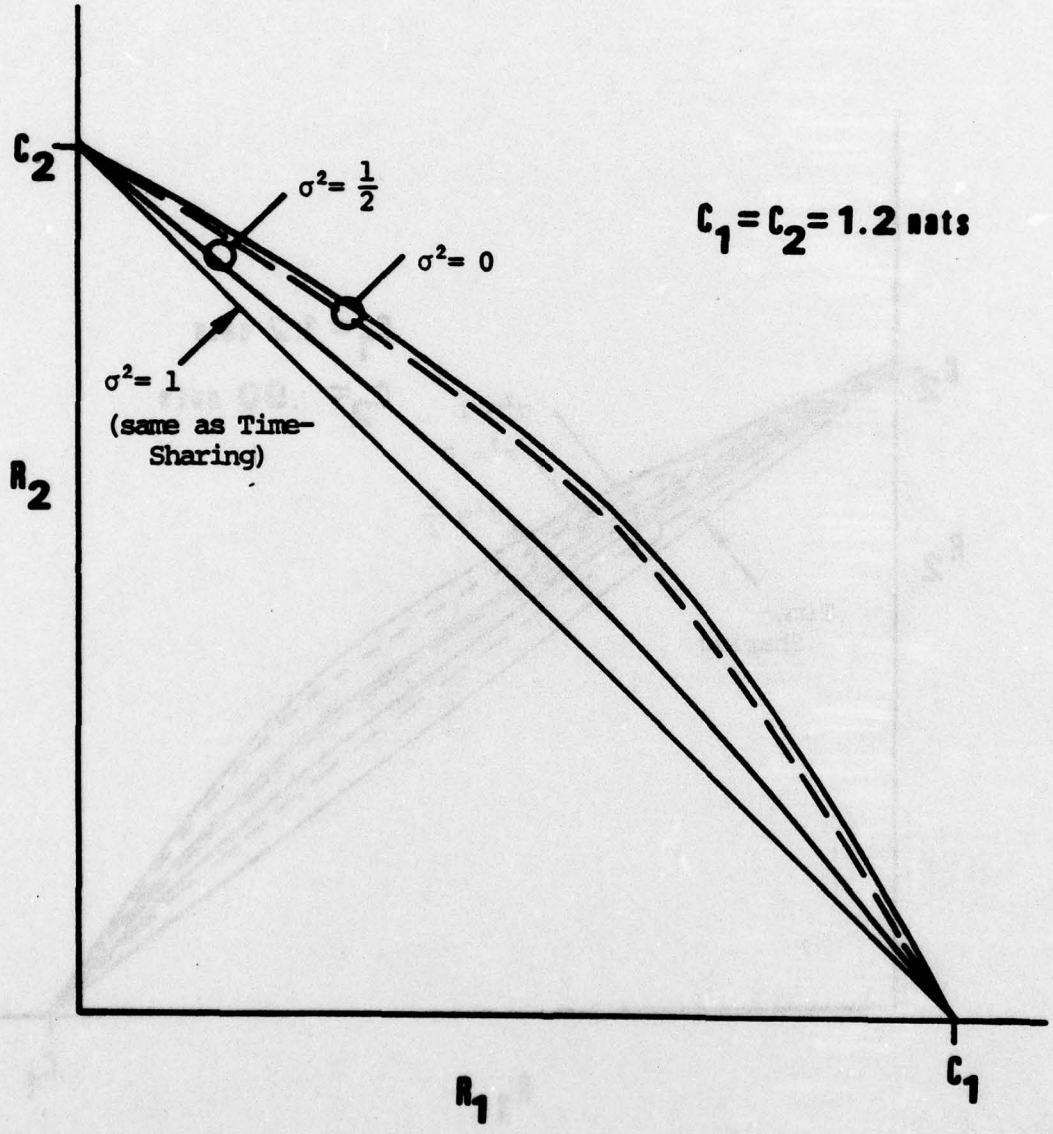
3.3.3 Intermediate Cases

We conclude this section on numerical results with two intermediate cases, that is, examples where the noises are partially correlated. Recall that for the examples of Figures 3.3 and 3.8, the total noise variance at receiver one was 1, and at receiver two it was 2. In the former case the common noise had variance 1, and in the latter, variance 0 (it was non-existent). In both cases P = 10. Figure 3.10 repeats the results for those cases, and gives the achievable region and outer bound when the common noise has variance $\frac{1}{2}$. As expected, the results lie between the previous results.

For the equal noise case, a degraded channel is the degenerate case, that is, $\sigma_2^2=0$. In this case both receivers are the same, and the capacity region, with or without feedback is just the time sharing region. Figure 3.5, therefore, which had the results for $P/\sigma_1^2 = P/\sigma_2^2 = 10$, with no common noise, as well as the time sharing line, already contains the results for the two extreme cases. Figure 3.11 repeats 3.5 with the addition of the outer bound for $\sigma^2 = \frac{1}{2}$. For this set of parameters, the achievable region was virtually indistinguishable from the outer bound, and could not be plotted separately.



3.10 Achievable Regions and Outer Bounds for AWGN BC with Feedback. $P = 10$ and $\sigma^2 + \sigma_1^2 = 1$ and $\sigma^2 + \sigma_2^2 = 2$. Common Noise Variance $\sigma^2 = 0, .5, 1$.



3.11 Achievable Regions and Outer Bounds for AWGN BC with Feedback. $P = 10$ and $\sigma^2 + \sigma_1^2 = \sigma^2 + \sigma_2^2 = 1$. Common Noise Variance $\sigma^2 = 0, .5, 1$.

CHAPTER 4

CONCLUSIONS

We have presented two similar feedback coding schemes for the discrete-time AWGN multiple access and broadcast channels respectively. These schemes are extensions of the approach given for the band-limited single user AWGN channel with feedback by Schalkwijk and Kailath [2],[3]. The multi-user schemes share with the schemes of [2] and [3] the properties that they are deterministic and achieve doubly exponential decay of error probability with block length, although for the MAC this holds only for rates dominated by those achieved without the use of the superposition approach described in Section 2.1.2. The sets of achievable rates afforded by these schemes exceed previously known achievable regions, except in the case of the physically degraded broadcast channel, for which the true capacity region was already known.

We have also established outer bounds for the achievable regions of the MAC and BC with feedback, both for discrete memoryless channels and for the AWGN case. The outer bound for the AWGN MAC coincides with the achievable region, yielding the capacity region. For the AWGN broadcast channel the achievable region fails to reach the outer bound in all cases. This failure is due, to at least some extent, to a sub-optimality of the achievable region; the coding scheme fails to achieve even the known capacity region of the physically degraded AWGN BC. It has not been determined whether the outer bound for the BC is itself too loose. The results obtained do demonstrate a previously unexpected fact: feedback

can enlarge the capacity region of broadcast channels as well as multiple access channels.

A number of areas for further research are suggested by the results of this dissertation:

1) Extensions of the general approach to other two user channels with feedback.

2) An achievable region or outer bound for M user channels, where M > 2, for either the MAC or BC with feedback.

3) Application to discrete memoryless channels. This work depended on the existence of an optimal constructive approach for the single user channel and on the fact that transmitted signals combine linearly. A constructive approach for single user erasure channels is well known [24], and approaches for BSC's ([20],[29]) and for general discrete memoryless channels [21] have appeared in the literature. In addition, as mentioned in Section 2.1, a coding scheme for discrete memoryless MAC's with feedback has appeared [13]. One question of particular interest is whether the outer bound for MAC's with feedback found in Section 2.2.1 is achievable in general.

4) The gaussian broadcast channel. The results of Chapter 3 are incomplete in that the capacity region has not been found. From the examples of Section 3.3 it is clear that the achievable region of Section 3.1 is not the capacity region. It seems doubtful that

the addition of superposition followed by convexification would be optimal, though the regions so obtained might in general lie close to the outer bound. Conversely, it is unclear how tight the outer bound of Section 3.2 really is.

5) Noisy feedback. In [2] and [3] and subsequently ([22],[23]) attempts were made to analyze and optimize the basic Schalkwijk-Kailath approach when additive noise appeared in the feedback link. Under the reasonable assumption that both forward and feedback powers were limited, the results were discouraging: none of the approaches were able to yield reliable communication at any positive rate. From a purely Shannon theoretic point of view this is somewhat beside the point, since the existence of codes which do not employ feedback is guaranteed by the coding theorem. Noiseless feedback simplified the coding problem, but did not add anything in the way of achievable rates. In the multi-user case, however, feedback generally does enlarge the capacity region, in addition to allowing a simple coding procedure. Noise in the feedback links must therefore affect the size of the capacity region, and as the feedback links become totally noisy, capacity regions must degenerate to their non-feedback values. Results in neither the forward nor converse directions exist as yet.

Two classes of problems related to that of noisy feedback are that where the feedback links are of a different type than the forward links (e.g. gaussian forward links with noisy or noiseless

discrete feedback), and the case where only one feedback link is available. For the MAC this would correspond to allowing the receiver feed back to only one transmitter. In the BC either only one receiver could feed back, or perhaps the receivers would share a single feedback channel in a multiple access mode.

APPENDIX A: ROOTS OF P(ρ)

We prove that the quartic polynomial

$$P(\rho) = \sigma^2(\sigma^2+P_1+P_2+2\sqrt{P_1P_2}\rho) - (\sigma^2+P_1(1-\rho^2))(\sigma^2+P_2(1-\rho^2)) \qquad (A.1)$$

has exactly one root in the interval $0<\rho<1$.

The following are true:

1. A quartic has at most four real roots.
2. $\lim_{\rho\to\infty} P(\rho) = -\infty$
3. $P(0) = -P_1P_2 < 0.$
4. $P(1) = \sigma^2(\sqrt{P_1}+\sqrt{P_2})^2 > 0.$
5. $P(-1) = \sigma^2(\sqrt{P_1}-\sqrt{P_2})^2 \geq 0.$

3. and 5. imply that there is at least one negative root.

2. and 4. imply that there is at least one root greater than 1.

Hence there are (from 1.) at most two roots in (0,1). But by 3. and 4. there are an odd number of roots in (0,1). Hence there is exactly one. Since P_1, P_2, and σ^2 are all arbitrary, this result holds also for the polynomial $P(\alpha,\rho)$ defined by equation (2.1.36).

APPENDIX B: CONCAVITY OF MUTUAL INFORMATION FOR THE MAC

In this appendix we show that the quantities $I(X_1X_2;Y)$, $I(X_1;Y|X_2)$, and $I(X_2;Y|X_1)$ are concave functions of the joint probability assignment of X_1 and X_2.

The first quantity is concave, since it is just the mutual information between the input and the output of a channel (see [24]). We need only show the result for $I(X_1;Y|X_2)$, since the same demonstration holds for $I(X_2;Y|X_1)$ with X_1 and X_2 exchanged.

Consider two input probability assignments $p_1(x_1,x_2)$ and $p_2(x_1,x_2)$. Define

$$p_3(x_1,x_2) = \alpha p_1(x_1,x_2) + (1-\alpha)p_2(x_1,x_2) \quad \alpha \in [0,1] \qquad (B.1)$$

Obviously, p_3 is also a joint probability assignment on X_1,X_2. Now define the auxiliary random variable Φ, with probability law $Pr[\Phi = 1] = \alpha$, and $Pr[\Phi = 2] = 1-\alpha$.

Consider the following communication scheme: we wish to communicate the value of Φ to the receiver. A sample of Φ is drawn, and a corresponding X_1,X_2 are drawn from the joint probability assignment corresponding to the sample value. That is (X_1,X_2) has probability assignment $p_i(x_1,x_2)$ when $\Phi = i$. Denoting the mutual information resulting from p_i by $I_i(X_1;Y|X_2)$, we have that

$$\begin{aligned} I(X_1;Y|X_2\Phi) &= Pr[\Phi=1]I_1(X_1;Y|X_2) + Pr[\Phi=2]I_2(X_1;Y|X_2) \\ &= \alpha I_1(X_1;Y|X_2) + (1-\alpha)I_2(X_1;Y|X_2) \end{aligned}$$

The mutual information unconditioned on Φ is just $I(X_1;Y|X_2)$ evaluated at p_3. In other words,

$$I(X_1;Y|X_2) = I_3(X_1;Y|X_2) \tag{B.3}$$

Subtracting (B.2) from (B.3), we have that

$$I(X_1;Y|X_2) - I(X_1;Y|X_2\Phi) = I_3 - \alpha I_1 - (1-\alpha)I_2 \tag{B.4}$$

The left hand side is

$$\begin{aligned}[H(Y|X_2) - H(Y|X_1X_2)]-[H(Y|X_2\Phi) - H(Y|X_1X_2\Phi)] \\ &= H(Y|X_2) - H(Y|X_2\Phi) \\ &= I(Y;\Phi|X_2) \\ &\geq 0 \end{aligned} \tag{B.5}$$

where the first equality in (B.5) holds since given X_1, X_2, Y is independent of Φ. Substituting (B.5) into (B.4), we have

$$I_3 - \alpha I_1 - (1-\alpha)I_2 \geq 0 \tag{B.6}$$

Therefore $I(X_1;Y|X_2)$ is concave in $p(x_1,x_2)$.

APPENDIX C:

CODING FOR THE AUGMENTED AWGN BC WITH FEEDBACK

We consider the model of the BC shown in Figure 3.2. This model is an extension of the canonical broadcast channel, in that one receiver is allowed to see both channel outputs. This model may be appropriate in some applications, say in cases where one receiver is located physically between the transmitter and the other receiver, and received data is fed back to the transmitter directionally. The example is interesting theoretically since it allows the exploitation of the superposition idea, and obtains results which are optimum (i.e. reliable communication is possible at all rates in the capacity region). The approach of Section 3.1 does not use superposition, and is suboptimal, at least for the only case in which it was applied and the capacity region was known (degraded channel).

Recall from Section 3.2 that the capacity region for the AWGN version of Figure 3.2 is given by

$$R_1 \leq \frac{1}{2} \ln \left(1 + \frac{\alpha P}{\sigma_e^2}\right)$$

$$R_2 \leq \frac{1}{2} \ln \left(1 + \frac{\bar{\alpha} P}{\alpha P + \sigma^2 + \sigma_2^2}\right) \qquad (C.1)$$

where $\sigma_e^2 = \sigma^2 + \dfrac{\sigma_1^2 \sigma_2^2}{\sigma_1^2 + \sigma_2^2}$.

We now present a coding scheme which achieves reliable communication at all points satisfying (C.1).

All definitions are as in Section 3.1, except that after k-1 transmissions receiver 1 has

$$\hat{\theta}_1^{k-1} + d_{k-1}\theta_2 = \theta_1 + d_{k-1}\theta_2 + \xi_{k-1}$$

$$\hat{\theta}_2^{k-1} = \theta_2 + \eta_{k-1} \tag{C.2}$$

and receiver 2 has the same estimate of θ_2. d_{k-1} above is assumed to be a known constant. Since receiver 1 knows receiver 2's channel data, there is no problem in his forming the same estimate of θ_2.

Now for the k^{th} transmission, the transmitter sends

$$x_k = \sqrt{\frac{P}{D}} \left(\frac{\xi_{k-1}}{\sqrt{a_{k-1}}} + g \frac{\eta_{k-1}}{\sqrt{b_{k-1}}} \right) \tag{C.3}$$

where

$$D = 1 + g^2 \tag{C.4}$$

Receiver 1 obtains r_{1k} and r_{2k}, while receiver 2 obtains r_{2k}, where

$$r_{ik} = x_k + w_k + w_{ik} \tag{C.5}$$

At this point both receivers upgrade their estimates of θ_2 as before, forming

$$\hat{\theta}_2^k = \hat{\theta}_2^{k-1} - \frac{\overline{r_{2k}\eta_{k-1}}}{\overline{r_{2k}^2}} r_{2k} \tag{C.6}$$

Assume that $\overline{\xi_{k-1}\eta_{k-1}} = 0$. Then

$$\overline{r_{1k}\eta_{k-1}} = \overline{r_{2k}\eta_{k-1}} = \sqrt{\frac{P}{D}}\, g\, \sqrt{b_{k-1}} \tag{C.7}$$

Since

$$\eta_k = \eta_{k-1} - \frac{\overline{r_{2k}\eta_{k-1}}}{\overline{r_{2k}^2}}\, r_{2k} \tag{C.8}$$

then

$$b_k = \overline{\eta_{k-1}^2} = b_{k-1} - \frac{\overline{r_{2k}\eta_{k-1}}^2}{\overline{r_{2k}^2}}$$

$$= b_{k-1} - \frac{\frac{P}{D}\, g^2 b_{k-1}}{P+\sigma^2+\sigma_2^2} = b_{k-1}\, \frac{\sigma^2+\sigma_2^2+\frac{P}{1+g^2}}{P+\sigma^2+\sigma_2^2} \tag{C.9}$$

To upgrade his estimate of θ_1, receiver 1 first combines r_{1k} and r_{2k} by forming the maximum likelihood estimate of x_k

$$r_k' = \frac{\sigma_2^2 r_{1k}+\sigma_1^2 r_{2k}}{\sigma_1^2+\sigma_2^2} = x_k + w_k + \frac{\sigma_2^2 w_{1k}+\sigma_1^2 w_{2k}}{\sigma_1^2+\sigma_2^2} = x_k + w_k + w_k' \tag{C.10}$$

He then subtracts $\sqrt{\frac{P}{D}}\, \frac{g}{\sqrt{b_{k-1}}}\, (\hat{\theta}_2^{k-1})$ from r_k'. The η_{k-1} terms cancel, yielding

$$r_{3k} = \sqrt{\frac{P}{D}}\, \frac{\xi_{k-1}}{\sqrt{a_{k-1}}} - \sqrt{\frac{P}{D}}\, \frac{g}{\sqrt{b_{k-1}}}\, \theta_2 + w_k + w_k' \tag{C.11}$$

Treating θ_2 as an unknown (but non-random) parameter, then

$$\overline{r_{3k}^2} = \frac{P}{D} + \sigma^2 + \overline{w_k'^2} = \frac{P}{D} + \sigma_e^2 \tag{C.12}$$

and

$$\overline{r_{3k}\xi_{k-1}} = \sqrt{\frac{P}{D}}\, \sqrt{a_{k-1}} \tag{C.13}$$

Receiver 1 now forms

$$\hat{\theta}_1^k + d_k\theta_2 = \hat{\theta}_1^{k-1} + d_{k-1}\theta_2 - \frac{\overline{r_{3k}\xi_{k-1}}}{\overline{r_{3k}^2}}\, r_{3k}$$

$$= \theta_1 + d_k\theta_2 + \xi_k \qquad \text{(C.14)}$$

By substituting equation (C.11) above, defining r_{3k}, it is seen that

$$d_k = d_{k-1} + \sqrt{\frac{a_{k-1}}{b_{k-1}}}\ \frac{\frac{P}{D}\, g}{\frac{P}{D} + \sigma_e^2} \qquad \text{(C.15)}$$

and

$$\xi_k = \xi_{k-1}\left(1 - \frac{\frac{P}{D}}{\frac{P}{D} + \sigma_e^2}\right) + \sqrt{\frac{P}{D}}\ \frac{\sqrt{a_{k-1}}}{\frac{P}{D} + \sigma_e^2}\,(w_k + w_k') \qquad \text{(C.16)}$$

which may be squared and averaged to obtain

$$a_k = \overline{\xi_k^2} = a_{k-1}\left(\frac{\sigma_e^2}{\frac{P}{D} + \sigma_e^2}\right)^2 + \frac{P}{D}\,\frac{a_{k-1}\sigma_e^2}{\left(\frac{P}{D} + \sigma_e^2\right)^2}$$

$$= a_{k-1}\frac{\sigma_e^2}{\frac{P}{D} + \sigma_e^2} \qquad \text{(C.17)}$$

We can multiply equations (C.8) and (C.16) and average to obtain the following recursive expression for the correlation between the estimate noises:

$$\overline{\xi_k\eta_k} = \overline{\xi_{k-1}\eta_{k-1}} + \frac{\overline{r_{2k}\eta_{k-1}}}{\overline{r_{2k}^2}}\,\frac{\overline{r_{3k}\xi_{k-1}}}{\overline{r_{3k}^2}}\,\overline{r_{2k}r_{3k}} - \frac{\overline{r_{3k}\xi_{k-1}}\ \overline{r_{3k}\eta_{k-1}}}{\overline{r_{3k}^2}}$$

$$- \frac{\overline{r_{2k}\eta_{k-1}}\,\overline{r_{2k}\xi_{k-1}}}{\overline{r_{2k}^2}} \qquad \text{(C.18)}$$

By assumption $\overline{\xi_{k-1}\eta_{k-1}} = 0$. This implies that $\overline{r_{3k}\eta_{k-1}} = 0$, so that the first and third terms of (C.18) are zero. From the definition of r_{3k} above, we see that $\overline{r_{3k}\xi_{k-1}} = \overline{r_{2k}\xi_{k-1}}$, so that defining $T = \overline{r_{ik}\xi_{k-1}}\overline{r_{2k}\eta_{k-1}}$ for i = 2 and 3, we obtain

$$\overline{\xi_k\eta}_k = \frac{T}{\overline{r_{2k}^2}} \left(\frac{\overline{r_{2k}r_{3k}}}{\overline{r_{3k}^2}} - 1 \right) \tag{C.19}$$

But from (C. 5) and (C.11) we obtain

$$\begin{aligned} \overline{r_{2k}r_{3k}} &= \frac{P}{D} + \overline{(w_k + w_k')(w_k + w_{2k})} \\ &= \frac{P}{D} + \sigma^2 + \overline{w_k'w}_{2k} \\ &= \frac{P}{D} + \sigma^2 + \frac{\sigma_1^2\sigma_2^2}{\sigma_1^2 + \sigma^2_2} \\ &= \frac{P}{D} + \sigma_e^2 \\ &= \overline{r_{3k}^2} \end{aligned} \tag{C.20}$$

Substituting into (C.19), then, $\overline{\xi_k\eta}_k = 0$.

To initialize, the messages are sent separately (amplified up to expected power P), so that

$$\begin{aligned} \hat{\theta}_1^0 + d_0\theta_2 &= \theta_1 + \frac{w_{-1} + w'_{-1}}{\sqrt{12P}} \\ \hat{\theta}_2^0 &= \theta_2 + \frac{w_0 + w_{2,0}}{\sqrt{12P}} \end{aligned} \tag{C.21}$$

Then

$$\begin{aligned} d_0 &= 0 \\ a_0 &= \frac{\sigma_e^2}{12P} \\ b_0 &= \frac{\sigma^2+\sigma_2^2}{12P} \\ \overline{\xi_0 \eta_0} &= 0 \end{aligned} \tag{C.22}$$

Thus the hypothesis that $\overline{\xi_k \eta_k} = 0$ is true by induction, and

$$a_k = a_0\left[\frac{\sigma_e^2}{\frac{P}{\mathcal{D}}+\sigma_e^2}\right]^k \tag{C.23a}$$

$$b_k = b_0\left[\frac{\frac{P}{\mathcal{D}}+\sigma^2+\sigma_2^2}{P+\sigma^2+\sigma_2^2}\right]^k \tag{C.23b}$$

$$d_k = \frac{\frac{P}{\mathcal{D}}g}{\frac{P}{\mathcal{D}}+\sigma_e^2}\sum_{j=1}^{k}\sqrt{\frac{a_{j-1}}{b_{j-1}}} \tag{C.23c}$$

A similar analysis as in Sections 2.1 and 3.1 yields that reliable communication is achieved at all R_1 and R_2 such that

$$\begin{aligned} R_1 &\le \frac{1}{2}\ln\left(\frac{\frac{P}{\mathcal{D}}+\sigma_e^2}{\sigma_e^2}\right) \\ R_2 &\le \frac{1}{2}\ln\left(\frac{P+\sigma^2+\sigma_2^2}{\frac{P}{\mathcal{D}}+\sigma^2+\sigma_2^2}\right) \end{aligned} \tag{C.24}$$

Since $\mathcal{D} = 1+g^2$, $\mathcal{D} \in [1,\infty]$, and $\frac{1}{\mathcal{D}} \in (0,1]$. Defining $\alpha = \frac{1}{\mathcal{D}}$, we obtain the desired result.

APPENDIX D: AN OUTER BOUND FOR BROADCAST CHANNELS

D.1 Preliminaries and an Alternate Representation of Prior Results

In Chapter 3 we found it desirable to obtain an outer bound on the capacity region of the broadcast channel with feedback. In this appendix, we apply the methods of Section 3.2 to bound the capacity region of channels without feedback. The best such bound for general broadcast channels is that found by Sato [31] (also in [1]). For a general BC with input X and outputs Y and Z, with marginal conditional probabilities $p'(y|x)$ and $p'(z|x)$, let $\mathcal{P}$ be the following class of joint channel probability functions

$$\mathcal{P} = \{p(yz|x): \sum_{y} p(yz|x) = p'(z|x) \quad \sum_{z} p(yz|x) = p'(y|x)\} \tag{D.1}$$

Then for each $p \in \mathcal{P}$ define

$$R(p) = \text{co}[\bigcup_{q\in\mathcal{Q}} \{(R_1,R_2): R_1 \le I_q(X;Y) \quad R_2 \le I_q(X;Z) \quad R_1 + R_2 \le I_q(X;YZ)|_p\}] \tag{D.2}$$

where $\mathcal{Q}$ is the set of all possible input distributions. Sato's outer bound is then

$$C' = \bigcap_{p\in\mathcal{P}} R(p) \tag{D.3}$$

In equation (D.2) and throughout this appendix it is convenient to express $I(X;YZ)$ as $I_q(X;YZ)_{|p}$ to denote dependence on p, the channel probability function, and q, the input probability assignment. Similarly $I(X;Y)$ and $I(X;Z)$ are written as $I_q(X;Y)$ and $I_q(X;Z)$ to denote their dependence on q. They are independent of p for all $p \in P$.

The inequalities which define $\mathcal{R}(p)$ in (D.2) follow from standard information theoretic arguments, and the intersection in (D.3) holds since a code for the actual channel must be a code for every channel with the same marginal conditional probabilities.

If we define the following quantity:

$$J_q(X;YZ) = \inf_{p \in P} I_q(X;YZ)_{|p} \tag{D.4}$$

we obtain an alternate representation of (D.2)-(D.3). Defining the region

$$\begin{aligned} C^* = \text{co}\Big[\bigcup_{q \in Q} \{(R_1,R_2):\ & R_1 \leq I_q(X;Y) \\ & R_2 \leq I_q(X;Z) \\ & R_1 + R_2 \leq J_q(X;YZ) \ \}\Big] \end{aligned} \tag{D.5}$$

we will prove

<u>Theorem D.1</u> $\quad C^* = C'$ (D.6)

Before proving Theorem D.1 we introduce the following lemmata:

Lemma D.1 $J_q(X;YZ)$ is concave in q.

This follows from Theorem 5.5 of [26].

Lemma D.2 For $R(p)$ defined as in (D.2), and for any non-negative vector $\underline{\lambda}$ in R^3, define

$$M(\underline{\lambda}) = \sup_q [\lambda_1 I_q(X;Y) + \lambda_2 I_q(X;Z) + \lambda_3 I_q(X;YZ)\,|_p] \qquad (D.7)$$

Then a point (R_1^*, R_2^*) is in $R(p)$ if and only if for all $\underline{\lambda} \geq 0$,

$$\lambda_1 R_1^* + \lambda_2 R_2^* + \lambda_3 (R_1^* + R_2^*) \leq M(\underline{\lambda}) \qquad (D.8)$$

The above applies to the set $\mathcal{C}^*$ with J_q subsitituted for $I_q(X;YZ)\,|_p$ in (D.7).

This lemma follows from an application of the separating hyperplane theorem (Theorem 11.3 of [26]). We note here that the convexification in (D.2) and (D.5) is redundant, since both underlying unions are already convex (since all of the constraints are concave and Q is convex).

Lemma D.3 Since $I_q(X;YZ)\,|_p$ is bounded and continuous in p and q, convex in p and concave in q, and P and Q are convex and compact, then

$$\inf_p \sup_q I_q(X;YZ)\,|_p = \sup_q \inf_p I_q(X;YZ)\,|_p \qquad (D.9)$$

This is Lemma 37.3.2 of [26]. Note that the continuity and compactness hypotheses imply that sequential limits of informations are actually achieved.

We now prove Theorem D.1. If (R_1^*,R_2^*) is in C^*, then there is some q for which the three inequalities in (D.5) are satisfied (again, by continuity and compactness, (R_1^*,R_2^*) is not only approached as a limit, but actually achieved), but since for all p, $I_q(X;YZ)_{|p} \geq J_q(X;YZ)$, the same q yields informations which satisfy the three inequalities in (D.2) for all p. Therefore $C^* \subset C'$.

Conversely, assume (R_1^*,R_2^*) is in C'. From Lemma D.2, it is true that for all p ϵ , and for all $\underline{\lambda} \geq 0$,

$$\lambda_1 R^*_1 + \lambda_2 R_2^* + \lambda_3 (R_1^*+R_2^*) \leq \sup_q [\lambda_1 I_q(X;Y) + \lambda_2 I_q(X;Z) + \lambda_3 I_q(X;YZ)_{|p}] \quad (D.10)$$

Since (D.10) holds for all $p \in P$, we can take the infimum of the right hand side over p, and apply Lemma D.3 to obtain

$$\lambda_1 R_1^* + \lambda_2 R_2^* + \lambda_3 (R_1^*+R_2^*) \leq \sup_q \inf_p [\lambda_1 I_q(X;Y) + \lambda_2 I_q(X;Z) + \lambda_3 I_q(X;YZ)_{|p}] \quad (D.11)$$

Only the last term in (D.11) depends on p, and its infimum is J_q. Therefore

$$\lambda_1 R_1^* + \lambda_2 R_2^* + \lambda_3 (R_1^*+R_2^*) \leq \sup_q [\lambda_1 I_q(X;Y) + \lambda_2 I_q(X;Z) + \lambda_3 J_q(X;YZ)] \quad (D.12)$$

and by Lemma D.2, (R_1^*,R_2^*) is in C^*. Therefore $C' \subset C^*$, and Theorem D.1 is

proved.

D.2 A New Outer Bound

We now give a new outer bound on the capacity region of broadcast channels, based on the approach of Section 3.2. Consider the augmented channel of Figure 3.2 without feedback. The capacity region of this channel can be no smaller than that of the corresponding BC without the link between receiver 2 and receiver 1. The augmented channel is degraded, so that it has capacity region

$$C_1(p) = \bigcup_{q \in \mathcal{Q}'} \{(R_1, R_2): R_1 \leq I_q(X;YZ|U)|_p, \; R_2 \leq I_q(U;Z)\} \tag{D.13}$$

where

$$\mathcal{Q}' = \{q(u,x,y,z) = q(u)q(x|u)p(yz|x)\} \tag{D.14}$$

The capacity region C of the true channel must lie in $C_1(p)$ for all $p \in \mathcal{P}$, so that

$$C \subset \bigcap_{p \in \mathcal{P}} C_1(p) \tag{D.15}$$

By reversing the direction of the added link, we get

$$C \subset \bigcap_{p \in \mathcal{P}} C_2(p) \tag{D.16}$$

where

$$C_2(p) = \bigcup_{q \in \mathcal{Q}'} \{(R_1, R_2): R_1 \leq I_q(U;Y) \quad R_2 \leq I_q(X;YZ|U)_{|p}\} \tag{D.17}$$

Therefore, we have the following

Theorem D.2 The capacity region of the general discrete memoryless BC satisfies

$$C \subset [\bigcap_{p \in \mathcal{P}} C_1(p)] \cap [\bigcap_{p \in \mathcal{P}} C_2(p)]$$

$$= \bigcap_{p \in \mathcal{P}} [C_1(p) \cap C_2(p)] \tag{D.18}$$

where $\mathcal{P}$ is defined by (D.1).

D.3 Comparison of the Bounds

We will prove the following

Theorem D.3 The bound of Theorem D.2 is tighter than that of Theorem D.1. That is

$$\bigcap_{p \in \mathcal{P}} [C_1(p) \cap C_2(p)] \subset C^* \tag{D.19}$$

To show this, we first find an outer bound for $\bigcap_{p\in P} C_1(p)$. For any p, if $(R_1,R_2) \in C_1(p)$, then there exists a q for which

$$\begin{aligned} R_1 &\leq I_q(X;YZ|U)|_p \\ R_2 &\leq I_q(U;Z) \end{aligned} \tag{D.20}$$

Since $U \to X \to (Y,Z)$ is a markov chain (by definition of Q'), then by the data processing theorem ([24] Theorem 4.3.3)

$$R_2 \leq I_q(U;Z) \leq I_q(X;Z) \tag{D.21}$$

Also,

$$\begin{aligned} R_1 + R_2 &\leq I_q(X;YZ|U)|_p + I_q(U;Z) \\ &\leq I_q(X;YZ|U)|_p + I_q(U;YZ)|_p \\ &= I_q(UX;YZ)|_p \\ &= I_q(X;YZ)|_p \end{aligned} \tag{D.22}$$

Therefore,

$$C_1(p) \subset \bigcup_{q\in Q} \{(R_1,R_2): R_1 + R_2 \leq I_q(X;YZ)|_p \quad R_2 \leq I_q(X;Z)\} \tag{D.23}$$

Using an approach similar to that which proved Theorem D.1, we can show that

$$\bigcap_{p\in P} C_1(p) \subset C_1^o = \bigcup_{q\in Q} \{(R_1,R_2): R_1 + R_2 \leq J_q(X;YZ) \quad R_2 \leq I_q(X;Z) \} \tag{D.24}$$

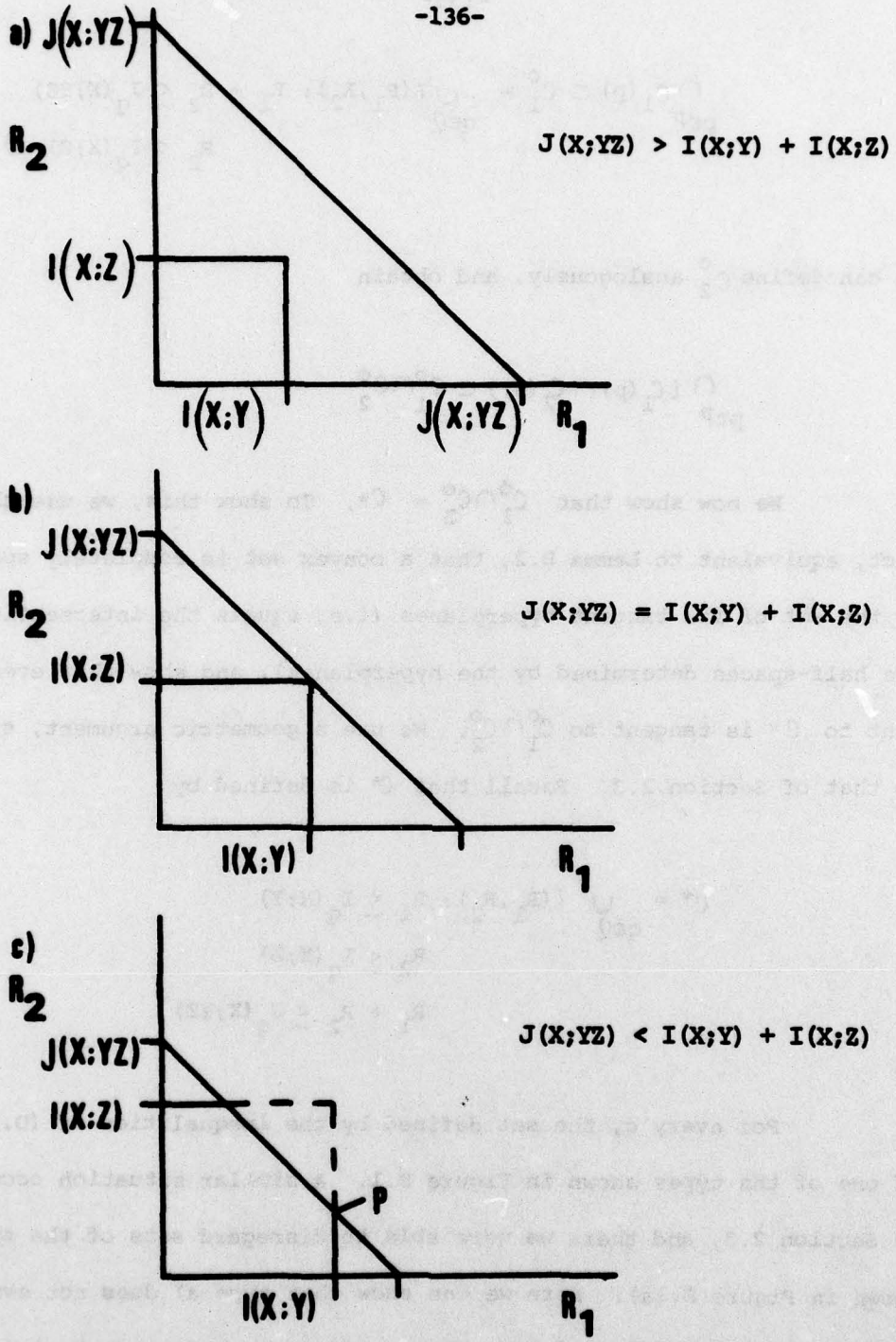
We can define C_2^o analogously, and obtain

$$\bigcap_{p\in P} [C_1(p) \cap C_2(p)] \subset C_1^o \cap C_2^o \tag{D.25}$$

We now show that $C_1^o \cap C_2^o = C^*$. To show this, we use the fact, equivalent to Lemma D.2, that a convex set is completely specified by the set of its tangent hyperplanes (i.e. equals the intersection of the half-spaces determined by the hyperplanes), and show that every tangent to C^* is tangent to $C_1^o \cap C_2^o$. We use a geometric argument, similar to that of Section 2.3. Recall that C^* is defined by

$$C^* = \bigcup_{q\in Q} \{(R_1,R_2): R_1 \leq I_q(X;Y) \quad R_2 \leq I_q(X;Z) \quad R_1 + R_2 \leq J_q(X;YZ)\} \tag{D.26}$$

For every q, the set defined by the inequalities of (D.26) is of one of the types shown in Figure D.1. A similar situation occurred in Section 2.3, and there we were able to disregard sets of the type shown in Figure D.1a). Here we can show that type a) does not even occur. Consider that channel in P for which Y and Z are independent given X.



D.1 Basic Sets for Proof of Theorem D.2.

This channel is always in $\mathcal{P}$. Call this channel p_0. For any q,

$$J_q(X;YZ) = \inf_{p} I_q(X;YZ)\big|_p \leq I_q(X;YZ)\big|_{p_0} \tag{D.27}$$

Since, for p_0, Y and Z are independent given X, it is true that

$$\begin{aligned} I_q(X;YZ)\big|_{p_0} &= H_q(YZ)\big|_{p_0} - H_q(Y|X) - H_q(Z|X) \\ &\leq H_q(Y) + H_q(Z) - H_q(Y|X) - H_q(Z|X) \\ &= I_q(X;Y) + I_q(X;Z) \end{aligned} \tag{D.28}$$

where we have dropped the dependence on p_0 for quantities which depend only on the marginal conditional probability functions. Combining (D.27) and (D.28) we have that

$$J_q(X;YZ) \leq I_q(X;Y) + I_q(X;Z) \tag{D.29}$$

Therefore sets of the type depicted in Figure D.1a), which correspond to the reverse (strict) inequality, do not occur.

Now consider a tangent to C^*. From continuity and compactness, it must be tangent to one of the sets of Figure D.1. Assume the tangent has equation $R_1 + \lambda R_2 = k_1(\lambda)$, where $0 \leq \lambda \leq 1$. The point of tangency must be the point marked P in Figure D.1, since the line has a slope less than -1. Now assume that this line is not tangent to C_2^o. The basic sets of C_2^o are the same as those of Figure D.1, except that the horizontal

line corresponding to the constraint on R_1 is absent. If the tangent we are considering is not tangent to C_2^o, assume that there is a tangent with the same slope and a larger constant, say $k_2(\lambda) > k_1(\lambda)$. The case of a smaller constant cannot occur, since $C^* \subset C_2^o$. This new line $R_1 + \lambda R_2 = k_2(\lambda)$ is tangent to a basic set of C_2^o for some other q at the point corresponding to P. We can evaluate the basic set of C^* for the new q. The new line is tangent to this set at point P, since the imposition of a constraint on R_2 will not delete that point, because of inequality (D.29). Therefore, there is a point of C^* lying above the assumed tangent, leading to a contradiction. Hence, all tangents of the form $R_1 + \lambda R_2 = k(\lambda)$, for $0 \leq \lambda \leq 1$ are tangent to C_2^o.

In a similar fashion, tangents of the form $\lambda R_1 + R_2 = k(\lambda)$ for $0 \leq \lambda \leq 1$ can be shown to be tangent to C_1^o.

From these two facts, it is clear that C^* is no smaller than $C_1^o \cap C_2^o$. Since each of the latter sets includes C^*, C^* is no larger than the intersection. Therefore $C_1^o \cap C_2^o = C^*$, and Theorem D.3 is proved.

Since $C^* \subset R^2$, but is defined by three inequalities, the fact that it can be generated by intersecting sets formed by pairwise combinations of the constraints is intuitively satisfying.

D.4 Discussion

In this section we evaluate the bounds for two simple cases, the degraded channel and Blackwell's example.

For a degraded channel, the cascade (physically degraded) channel

is in P. Since in general $I_q(X;YZ) \geq I_q(X;Y)$ with equality for cascade channels, then the cascade channel achieves the infimum in (D.4) and

$$J_q(X;YZ) = I_q(X;Y) \tag{D.30}$$

C* then becomes

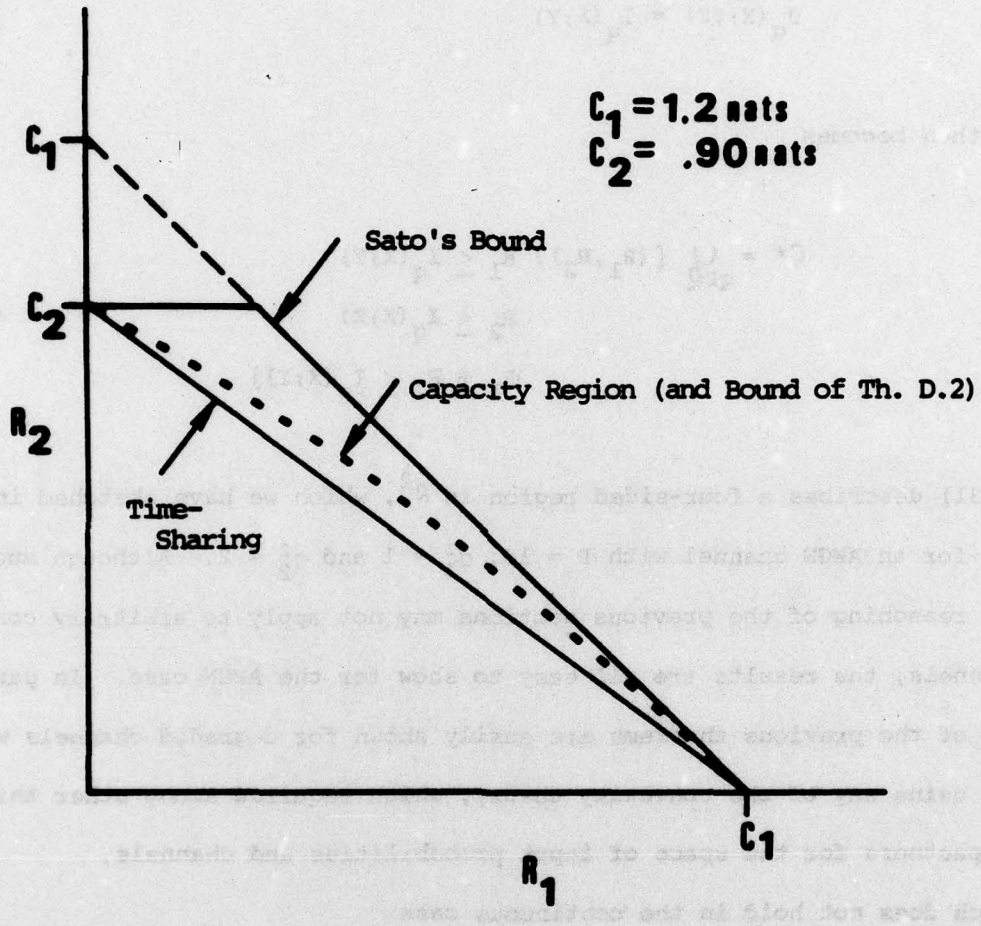
$$\begin{aligned} C^* = \bigcup_{q\in Q} \{(R_1,R_2): \; & R_1 \leq I_q(X;Y) \\ & R_2 \leq I_q(X;Z) \\ & R_1 + R_2 \leq I_q(X;Y)\} \end{aligned} \tag{D.31}$$

(D.31) describes a four-sided region in R^2, which we have sketched in Figure D.2 for an AWGN channel with $P = 10$, $\sigma_1^2 = 1$ and $\sigma_2^2 = 2$. Although much of the reasoning of the previous sections may not apply to arbitrary continuous channels, the results are all easy to show for the AWGN case. In particular all of the previous theorems are easily shown for degraded channels without using any of the convexity theory, which required among other things compactness for the space of input probabilities and channels, which does not hold in the continuous case.

The region $C_1(p)$ described by equation (D.13) becomes for the cascade channel

$$\begin{aligned} C_1(p) = \bigcup_{q\in Q'} \{(R_1,R_2): \; & R_1 \leq I(X;Y|U) \\ & R_2 \leq I(U;Z) \quad \} \end{aligned} \tag{D.32}$$

which is the true capacity region. This region is also sketched in Figure



D.2 Sato's Bound and Capacity Region of AWGN BC with P = 10, $\sigma_1^2 = 1$ and $\sigma_2^2 = 2$.

D.2. Clearly, in this case the bound of Theorem D.2 is strictly tighter than that of Theorem D.1, and indeed is exact.

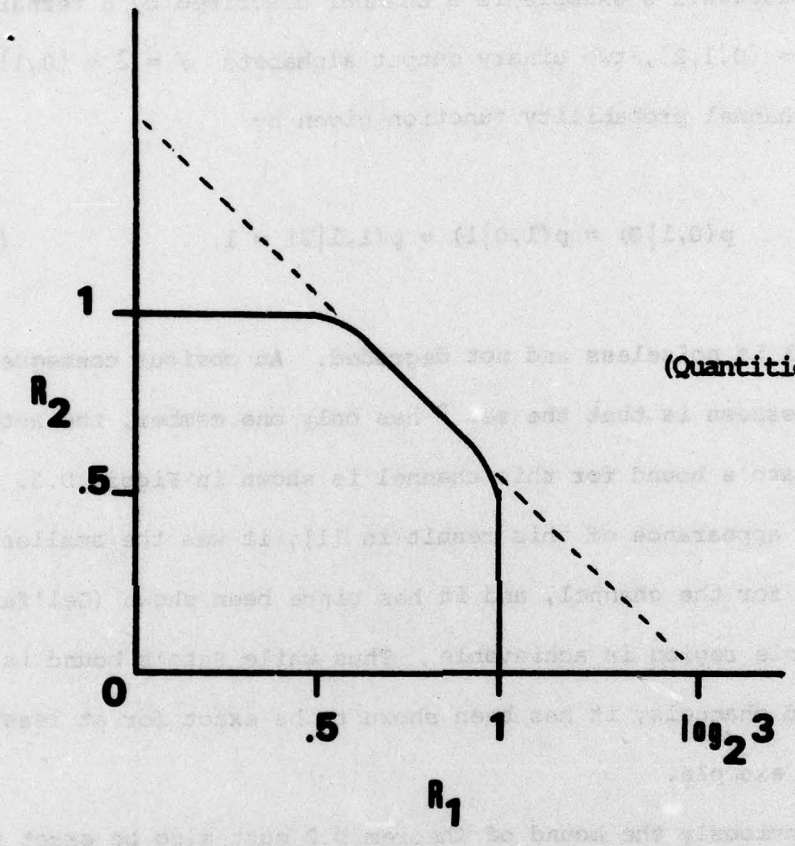
Blackwell's example is a channel described by a ternary input alphabet $X = \{0,1,2\}$, two binary output alphabets $Y = Z = \{0,1\}$, and the joint channel probability function given by

$$p(0,1|0) = p(1,0|1) = p(1,1|2) = 1 \qquad \text{(D.33)}$$

This channel is noiseless and not degraded. An obvious consequence of the noiselessness is that the set P has only one member, the actual channel. Sato's bound for this channel is shown in Figure D.3. At the time of the appearance of this result in [1], it was the smallest known outer bound for the channel, and it has since been shown (Gel'fand [27]) that the whole region is achievable. Thus while Sato's bound is loose for degraded channels, it has been shown to be exact for at least one non-trivial example.

Obviously the bound of Theorem D.2 must also be exact for this example.

NOTE: As this manuscript was going into final preparation, a paper appeared in which a new outer bound on the capacity region of discrete memoryless BC's was given (Marton [34]). We have not had time to compare the bound of [34] and Theorem D.2.



(Quantities are in bits)

D.3 Capacity Region of Blackwell's Example.

REFERENCES

1. Van der Meulen, E.C. "A Survey of Multi-way Channels in Information Theory: 1961-1976," IEEE Trans. on Inf. Th., Vol. IT-23, pp. 1-37, Jan. 1977.

2. Schalkwijk, J.P.M. and Kailath, T. "A Coding Scheme for Additive Noise Channels with Feedback - Part I: No Bandwidth Constraint," IEEE Trans. on Inf. Th., Vol. IT-12, pp. 172-182, Apr. 1966.

3. Schalkwijk, J.P.M. "A Coding Scheme for Additive Noise Channels with Feedback - Part II: Band-Limited Signals," IEEE Trans. on Inf. Th., Vol. IT-12, pp. 183-189, Apr. 1966.

4. Cover, T.M. "Broadcast Channels," IEEE Trans. on Inf. Th., Vol. IT-18, pp. 2-14, Jan. 1972.

5. Bergmans, P.P. and Cover, T.M. "Cooperative Broadcasting," IEEE Trans. on Inf. Th., Vol. IT-20, pp. 317-324, May 1974.

6. Shannon, C.E. "Two-way Communication Channels," Proc. 4th Berkeley Symp. Math. Stat. and Prob., Vol. 1, pp. 611-644, 1961.

7. Bergmans, P.P. "A Simple Converse for Broadcast Channels with Additive White Gaussian Noise," IEEE Trans. Inf. Th., Vol. IT-19, pp. 772-777, Nov. 1973.

8. El-Gamal, A. "The Feedback Capacity of Degraded Broadcast Channels," IEEE Trans. on Inf. Th., Vol. IT-24, pp. 379-381, May 1978.

9. Ahlswede, R. "On Two-way Communication Channels and a Problem by Zarankiewicz," Trans. 6th Prague Conf. on Inf. Th., Stat. Dec. Functions, Random Processes, pp. 23-37, 1971.

10. Liao, H. Multiple Access Channels, Ph.D. dissertation, Dept. of E.E., Univ. Hawaii, 1972.

11. Slepian, D. and Wolf, J.K. "A Coding Theorem for Multiple Access Channels with Correlated Sources," Bell System Technical Journal, Vol. 52, pp. 1037-1076, Sep. 1973.

12. Gaarder, N.T. and Wolf, J.K. "The Capacity Region of a Multiple Access Discrete Memoryless Channel can Increase with Feedback," IEEE Trans. on Inf. Th., Vol. IT-21, pp. 100-102, Jan. 1975.

13. Cover, T.M. and Leung-Yan-Cheong, S.K. "A Rate Region for the Multiple Access Channel with Feedback," Stanford Stat. Tech. Report 417, Sept. 1976.

14. El-Gamal, A. "The Capacity of a Class of Broadcast Channels," IEEE Trans. on Inf. Th., Vol. IT-25, pp. 166-169, Mar. 1979.

15. Shannon, C.E. "A Mathematical Theory of Communication," Bell System Technical Journal, Vol. 27, pp. 379-423,623-656, July and Oct. 1948

16. El-Gamal, A. "The Capacity of the Gaussian Broadcast Channel with Feedback," Stanford Dept. Stat. Tech. Report #25, Jun. 1977.

17. Cover, T. "Open Problems in Information Theory," Proc. 1975 IEEE-USSR Joint workshop on Inf. Th., pp. 35-37, 1976.

18. Korner, J. and Marton, K. "The Comparison of Two Noisy Channels," Trans. Coll. on Inf. Th., Keszthely, Hungary 1975, pp. 411-423.

19. Cover, T. "Some Advances in Broadcast Channels," Advances in Communication Systems, Vol. 4, A. Viterbi, ed., San Francisco: Academic Press, 1975.

20. Schalkwijk, J.P.M. "A Class of Simple and Optimal Strategies for Block Coding on the Binary Symmetric Channel with Noiseless Feedback," IEEE Trans. on Inf. Th., Vol. IT-17, pp. 283-287, May 1971.

21. Ahlswede, R. "A Constructive Proof of the Coding Theorem for Discrete Memoryless Channels with Feedback," Trans. 6th Prague Conf. on Inf. Th., Stat. Decision Functions, Random Processes, pp.39,50, 1971.

22. Kashyap, R.L. "Feedback Coding Schemes for an Additive Noise Channel with a Noisy Feedback Link," IEEE Trans. Inf. Th.,Vol. IT-14, pp. 471-480, May 1968.

23. Butman, S. "A general Formulation of Linear Feedback Systems with Solutions," IEEE Trans. Inf. Th., Vol. IT-15, pp.392-400, May 1969.

24. Gallager, R.G. Information Theory and Reliable Communication, New York: Wiley, 1968.

25. Wyner, A.D. "Recent Results in the Shannon Theory," IEEE Trans. Inf. Th., Vol. IT-20, pp. 2-10, Jan. 1974.

26. Rockafellar, R.T. Convex Analysis, Princeton: Princeton U. 1970.

27. Gel'fand, S.I. "Capacity of one Broadcast Channel," Problemy Peredachi Informatsii, Vol. 13, pp. 106-108, Jul.-Sep. 1977.

28. Bergmans, P.P. "Random Coding Theorems for Broadcast Channels with Degraded Components," IEEE Trans. Inf. Th., Vol. IT-19, pp. 197-207, Mar. 1973.

29. Berlekamp, E.R. "Block Coding for the Binary Symmetric Channel with Noiseless, Delayless, Feedback," in Error Correcting Codes, pp. 61-85, New York: Wiley, 1968.

30. Gallager, R.G. "Capacity and Coding for Degraded Broadcast Channels," Problemy Peredachi Informatsii, Vol. 10, pp. 3-14, Jul.Sep. 1974.

31. Sato, H. "An Outer Bound to the Capacity Region of Broadcast Channels, " IEEE Trans. on Inf. Th., Vol. IT-24, pp. 374-377, May 1978.

32. Shannon, C.E. "The Zero Error Capacity of a Noisy Channel," IRE Trans. on Inf. Th., Vol. IT-2, pp. 8-19, Sept. 1956.

33. Korner, J. and Marton, K. "General Broadcast Channels with Degraded Message Sets," IEEE Trans. on Inf. Th., Vol. IT-23, pp. 60-64, Jan. 1977.

34. Marton, K. "A Coding Theorem for the Discrete Memoryless Broadcast Channel," IEEE Trans. on Inf. Th., Vol IT-25, pp. 306-311, May 1979.

BIOGRAPHICAL NOTE

Lawrence Ozarow was born in the Bronx, N.Y. on October 31, 1949. He received the B.S. degree in electrical engineering from Columbia University in June 1970. From September 1970 to June 1973 he was enrolled at M.I.T., where he received the S.M. and E.E. degrees in electrical engineering in June, 1973. His master's research was in the area of optical communication.

From March 1973 to September 1976 he was employed at CNR, Inc. in Needham, Ma., working in the area of design and analysis of coding and modulation techniques for communication over fading channels.

In September 1976 he returned to M.I.T. to resume his graduate studies. From September 1976 to June 1978 he was a Research Assistant at the M.I.T. Lincoln Laboratories, Lexington, Ma., working in the area of satellite communication. From June 1978 t June 1979 we was a recipient of a Vinton Hayes Fellowship.

Mr. Ozarow's publications to date are "Packet Switching in a Processing Satellite," Proceedings of the IEEE, Jan. 1978 (with J.K. DeRosa), "Efficient Packet Satellite Communication," accepted for publication, IEEE Transacations on Communication Theory (with J.K. De Rosa and L. Weiner), "Capacity Limits for Binary Codes in the Presence of Interference," IEEE Transactions on Communication Theory, Feb. 1979 (with D.Chase), and "Combined Voice and Packet Communication Via Satellite," invited paper, I.C.C. '79, June, 1979 (with J.K. De Rosa).

Mr. Ozarow is a member of Tau Beta Pi and Eta Kappa Nu.