THE CAPACITY OF A CLASS OF BROADCAST CHANNELS

BY

ABBAS EL GAMAL

TECHNICAL REPORT NO. 24
JUNE 1, 1977

PREPARED UNDER THE AUSPICES
OF
NATIONAL SCIENCE FOUNDATION
GRANT ENG 76-03684

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
THE CAPACITY OF A CLASS OF BROADCAST CHANNELS

BY

ABBAS EL GAMAL

TECHNICAL REPORT NO. 24
JUNE 1, 1977

PREPARED UNDER THE AUSPICES
OF
NATIONAL SCIENCE FOUNDATION
GRANT ENG 76-03684

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
THE CAPACITY OF A CLASS OF BROADCAST CHANNELS

Abbas El Gamal†
Dept. of Electrical Engineering
Stanford University
Stanford, CA 94305

Abstract

We establish the capacity region for the discrete memoryless broadcast channels \( p(y,z|x) \) for which \( I(X;Y) \geq I(X;Z) \) for all input distributions. We show that the capacity region for this class of channels resembles the capacity region for degraded message sets considered by Körner and Marton [2].

†This work was partially supported by the National Science Foundation Grant ENG 76-03684.
1. **Introduction:**

The broadcast channel was introduced by Cover, who also defined the class of degraded broadcast channels for which one receiver is a stochastically degraded version of the other. Bergmans [4] established a coding theorem for the degraded broadcast channel, and Gallager proved a weak converse for Bergmans' rate region.

In [5], [6] an achievable rate region for the general discrete memoryless broadcast channel was given. No converse has yet been proved except for the case of degraded message sets [2].

As a generalization of the concept of "degradation" in broadcast channels, Körner and Marton introduced (see [1]) the two weaker partial orderings of "less noisy" and "more capable." In [1] the "less noisy" ordering was shown to be strictly weaker than "degradation," and the capacity region of the class of "less noisy" broadcast channels was shown to be similar to the degraded broadcast channel capacity region.

Van der Meulen pointed out in his survey [7] that Ahlswede has shown that there are broadcast channels which fall into the category of "more capable" but that do not satisfy the conditions of the "less noisy" ordering.

In this paper we establish the capacity region of the broadcast channel when one receiver is "more capable" than the other (open problem XXIII in [7]). We prove a coding theorem using superposition codes, and a weak converse to the coding theorem.

The capacity region of this class of broadcast channels will be shown to have a description similar to the capacity region of the broadcast channel with degraded message sets when the private message is sent to the more capable receiver.
2. **Definitions and Statement of the Result:**

The discrete memoryless broadcast channel $(X,P(Y,Z|X),Y×Z)$ consists of three finite sets $X,Y,Z$ and a probability transition matrix $P(y,z|x)$. The $n$th extension for the channel is a broadcast channel $(X,P(y,z|x),Y×Z)$ where

$$p(y,z|x) = \prod_{i=1}^{n} p(y_i,z_i|x_i) \quad (1)$$

An $((M_0,M_1,M_2),n)$ code for a broadcast channel consists of three sets of integers $M_0,M_1,M_2$

$$M_0 = \{1,\ldots,M_0\}$$
$$M_1 = \{1,\ldots,M_1\}$$
$$M_2 = \{1,\ldots,M_2\}$$

an encoding function

$$X : M_0 \times M_1 \times M_2 \to X \quad (3)$$

and two decoding functions

$$g_1 : Y \to M_0 \times M_1 ; g_1(Y) = (\hat{W}_0,\hat{W}_1)$$
$$g_2 : Z \to M_0 \times M_2 ; g_2(Z) = (\hat{W}_0,\hat{W}_2) \quad (4)$$

The set $\{X(w_0,w_1,w_2) : (w_0,w_1,w_2) \in M_0 \times M_1 \times M_2\}$ is called the set of code-words. The integer $w_0$ has the interpretation of the common part of the message, while the integers $w_1,w_2$ are called the independent parts of the message.

Define

$$p^n_{e_1} = \frac{1}{M_0^{M_1}} \sum_{w_0,w_1 \in M_0 \times M_1} P\{g_1(Y) = (w_0,w_1)|(w_0,w_1)\text{sent}\} \quad (5)$$

$$p^n_{e_2} = \frac{1}{M_0^{M_2}} \sum_{w_0,w_2 \in M_0 \times M_2} P\{g_2(Z) = (w_0,w_2)|(w_0,w_2)\text{sent}\} \quad (6)$$
to be the average probabilities of error of the decoders $g_1$ and $g_2$ respectively.

Also define the rate triple $(R_0,R_1,R_2)$ of an $((M_0,M_1,M_2),n)$ code by

$$R_0 = \frac{1}{n} \log M_0$$
$$R_1 = \frac{1}{n} \log M_1$$
$$R_2 = \frac{1}{n} \log M_2$$

(6)

The rate $(R_0,R_1,R_2)$ is said to be achievable by a broadcast channel if, for any $\epsilon > 0$, for all sufficiently large $n$, there exists an $((M_0,M_1,M_2),n)$ code with

$$M_0 \geq 2^{nR_0}, \quad M_1 \geq 2^{nR_1}, \quad M_2 \geq 2^{nR_2}$$

(7)

such that \(\max \left\{ p^n_{e_1}, p^n_{e_2} \right\} < \epsilon\).

The capacity region $C$ for the broadcast channel is the set of all achievable rates $(R_0,R_1,R_2)$.

Also recall the definitions of the "degradation," "less noisy," and "more capable" orderings on the broadcast channel:

Let $P_1(y|x), P_2(z|x)$ be the two marginals of $P(y,z|x)$, then

1. $Z$ is said to be a degraded form of $Y$ if there exists a probability transition matrix $P_3(z|y)$ such that

$$P_2(z|x) = \sum_{y \in Y} P_1(y|x) P_3(z|y)$$

(8)

2. $Y$ is said to be less noisy than $Z$ if

$$I(U;Z) \leq I(U;Y)$$

(9)
Figure 1. The Capacity Region

\[ C_1 = \max_{p(x)} I(X; Y) \]

\[ C_2 = \max_{p(x)} I(X; Z) \]
for every probability mass function of the form

\[ P(u,x,y,z) = p(u) \ p(x|u) \ p(y,z|x) \]  

(10)

3. \( Y \) is said to be more capable than \( Z \) if

\[ I(X;Z) \leq I(X;Y) \]  

(11)

for all probability distributions on \( X \).

The main result of the paper can now be stated in the following theorem.

Theorem (Capacity region): Let \((X,P(Y,Z|X),Y \times Z)\) be the BC channel defined before, and let \( U \) be an arbitrary random variable with cardinality \( \leq ||X|| + 2 \). If condition (11) holds then the capacity region \( C \) is given by

\[
C = \left\{ (R_0,R_1,R_2) : R_0 + R_1 \leq I(X;Y) \right. \\
\left. R_0 + R_1 + R_2 \leq I(X;Y|U) + I(U;Z) \right. \\
\left. R_0 + R_2 \leq I(U;Z), P \in \mathcal{P} \right\},
\]  

(12)

where \( \mathcal{P} \) is the set of all probability mass functions of the form

\[ p(u,x,y,z) = p(u) \ p(x|u) \ p(y,z|x) . \]  

(13)

Note that

(i) The region is symmetric in \( R_0 \) and \( R_2 \),

(ii) The plane region \((R_1,R_0)\) coincides with the degraded message sets region given in [2],

-5-
(iii) The plane region \((R_0, R_2)\) is defined by

\[ R_0 + R_2 \leq I(X; Z) \]  \hspace{1cm} (14)

and also coincides with the region in [2] when condition (11) is imposed.

(iv) For any fixed \(R_1 = r\) the plane region \((R_0, R_2)\) is a triangle. It is important to note that \(C\) is convex (see Appendix). Thus the usual convexification of the union of information regions is unnecessary.

3. The Achievability of \(C\):

First notice that because of the symmetry of \(C\) in \(R_0, R_2\) it suffices to show that any \((R_0, R_1, 0)\) or \((0, R_1, R_2)\in C\) is achievable. It follows from (iv) that, by time sharing, any other rate triple in \(C\) can be achieved.

Theorem 2: Any \((R_0, R_1, 0)\in C\) is achievable.

This theorem is identical to Körner and Marton's result on broadcast channels with degraded message sets, although the inequalities are slightly different. We shall sketch the proof for completeness.

Let \(C'\) be the plane region defined by all rate pairs \((R_0, R_1)\) such that \((R_0, R_1, 0)\in C\). Fixing a probability distribution \(p_{X}\) on \(X\), define

\[ C'(p_X) = \left\{ (R_0, R_1) : R_0 + R_1 \leq I(X; Y) \right\} \]

\[ R_0 + R_1 \leq I(X; Y|U) + I(U; Z) \]

\[ R_0 \leq I(U; Z) \]

\[ p(X) = p_X \] \hspace{1cm} (15)

-6-
\[ C' = \bigcup_{p_X} C'(p_X) \]  

(16)

It is sufficient to prove that for an arbitrary \( p_X \) any \( (R_0, R_1) \in C'(p_X) \) is achievable.

Consider the regions

\[
G(p_X) = \left\{ (R_0, R_1) : R_0 + R_1 \leq I(X; Y | U) + I(U; Z) \\
R_0 \leq I(U; Z) \\
p(X) = p_X \right\}
\]  

(17a)

and

\[
L(p_X) = \left\{ (R_0, R_1) : R_0 + R_1 \leq I(X; Y) \\
p(X) = p_X \right\}
\]  

(17b)

Thus,

\[ C'(p_X) = G(p_X) \cap L(p_X) \]  

(18)

Lemma:

Let

\[
K(p_X) = \sup_{p \in \mathcal{P}} \left\{ I(X; Y | U) + I(U; Z) : p(X) = p_X \right\}
\]  

(19)

and

\[
S(p_X) = \left\{ (R_0, R_1) : R_0 \leq I(U; Z) \\
R_1 \leq I(X; Y | U) \\
p(X) = p_X \right\}
\]  

(20)

Then

\[ K(p_X) = I(X; Y) \] and

\[ G(p_X) = \text{convex hull} \left\{ S(p_X) \cup \{0, K(p_X)\} \right\} \]  

(21)

Proof: It suffices to show that \( G(p_X) \) is convex and hence only the extreme points need be considered, namely the two sets of rate pairs

\[
(R_0, R_1) = (I(U; Z), I(X; Y | U))
\]  

(22a)

and

\[
(R_0, R_1) = (0, I(X; Y | U) + I(U; Z)).
\]  

(22b)
Figure (2) shows the two cases

(i) $K(p_X) = I(X;Y)$

(ii) $K(p_X) > I(X;Y)$

Figure 2. The Region $C'(p_X)$

To see that $G(p_X)$ is convex, let $(R_{01}, R_{11})$ and $(R_{02}, R_{12}) \in G(p_X)$ and let $\alpha \in [0,1]$, $\bar{\alpha} = 1 - \alpha$. Then there exists two random variables $(U_1, U_2)$ such that

$$ R_{01} \leq I(U_1;Z) $$

$$ R_{01} + R_{11} \leq I(U_1;Z) + I(X;Y|U_1) $$

and

$$ R_{02} \leq I(U_2;Z) $$

$$ R_{02} + R_{12} \leq I(U_2;Z) + I(X;Y|U_2) $$

Define a random variable $T$ taking values 1,2 with probabilities $\alpha$ and $\bar{\alpha}$ respectively. Then

$$ \alpha R_{01} + \bar{\alpha} R_{02} \leq \alpha I(U_1;Z) + \bar{\alpha} I(U_2;Z) $$

$$ \leq I(U_T;T;Z) $$

(23)
\[(\alpha R_{01} + \alpha R_{02}) + (\alpha R_{11} + \alpha R_{12}) \leq I(U_T, T; Z) + \alpha I(X; Y|U_1) + \alpha I(X; Y|U_2)\]
\[\leq I(U_T, T; Z) + I(X; Y|U_T, T) \quad (24)\]

and \((U_T, T) \to X \to YZ\) form a Markov chain with \(p(X) = p_X\) and \(p(y, z|x)\)
the channel's probability transition matrix. Thus \(G(p_X)\) is convex and the lemma follows directly.

Next, we show that any rate pair \((R_0^*, R_1^*)\) on the upper boundary of \(C'(p_X)\) is achievable. From Fig. (2) notice that the upper boundary of \(C'(p_X)\) consists of two parts:

(i) **The line segment \(BC\)**. Any point \((R_0^*, R_1^*) \in \overline{BC}\) is achievable as in Bergmans [4], since there exists a \(U\) such that
\[
R_1^* = I(X; Y|U), \quad (25)
R_0^* = I(U; Z)
\]
and
\[I(X; Y|U) + I(U; Z) \leq I(X; Y) \quad (26)\]

(ii) **The line segment \(AB\)**. A point \((R_0^*, R_1^*) \in \overline{AB}\) can be achieved by time sharing the two rate pairs
\[
(0, I(X; Y)) \quad \text{and} \quad (R_0^{AB}, R_1^{AB}) \quad (27)
\]
This completes the proof of Theorem 2.

4. **The Converse:**

We now show the optimality of the achievable rate region \(C\) by proving a weak converse.

**Theorem: (Weak converse)** If \((R_0, R_1, R_2) \notin C\), then there exists \(\varepsilon > 0\) such that
\[
\max \left\{ p_{e_1}^n, p_{e_2}^n \right\} \geq \varepsilon \quad \text{for all} \quad n .
\]
Proof: Fano's inequality yields
\[ H(W_0, W_1 | Y) \leq n(R_0 + R_1) p^n_{e,1} + h(p^n_{e,1}) = n\lambda_1n \] \hspace{1cm} (28a)
\[ H(W_0, W_2 | Z) \leq n(R_0 + R_2) p^n_{e,2} + h(p^n_{e,2}) = n\lambda_2n \] \hspace{1cm} (28b)

First consider
\[ n(R_0 + R_1 + R_2) \overset{\Delta}{=} H(W_0, W_1, W_2) = H(W_0) + H(W_1) + H(W_2) \]
\[ = H(W_0, W_1) + H(W_0, W_2) - H(W_0) \]
\[ = I(W_0, W_1; Y) + I(W_0, W_2; Z) - I(W_0; Y) + \]
\[ + H(W_0, W_1 | Y) + H(W_0, W_2 | Z) - H(W_0 | Z) . \]

Substituting from (28), we obtain
\[ n(R_0 + R_1 + R_2) \leq I(W_2; Z | W_0) + I(W_0, W_1; Y) + n(\lambda_1n + \lambda_2n) . \] \hspace{1cm} (29)

Similarly
\[ n(R_0 + R_1 + R_2) \leq I(W_1; Y | W_0) + I(W_0, W_2; Z) + n(\lambda_1n + \lambda_2n) \] \hspace{1cm} (30)

and
\[ n(R_0 + R_2) \overset{\Delta}{=} H(W_0, W_2) \leq I(W_0, W_2; Z) + n\lambda_2n . \] \hspace{1cm} (31)

Next we bound the right hand sides of Equations (29), (30), and (31).

Lemma: Given any probability mass function on \( W_0, W_1, W_2, X, Y, Z \) of the form
\[ p(w_0, w_1, w_2, x, y, z) = p(w_0)p(w_1)p(w_2)p(x | w_0, w_1, w_2) \times \]
\[ \times \prod_{i=1}^{n} p(y_i, z_i | x_i) \] \hspace{1cm} (32)

then
\[ (i) \ I(W_2; Z | W_0) + I(W_0, W_1; Y) \leq \sum_{i=1}^{n} I(X_i; Y_i) \] \hspace{1cm} (33)
\[ (ii) \ I(W_1; Y | W_0) + I(W_0, W_2; Z) \leq \sum_{i=1}^{n} I(X_i; Y_i | U_i) + I(U_i; Z_i) \] \hspace{1cm} (34)
\[(iii) \quad I(W_0, W_2; Z) \leq \sum_{i=1}^{n} I(U_i; Z_i) \quad \text{(35)}\]

where,

\[U_i = (W_0, W_2, Y_{i-1}, Z_i^{i+1}) , \]

\[Y_{i-1} = (Y_1, \ldots, Y_{i-1}) , \]

and

\[Z_i^{i+1} = (Z_{i+1}, \ldots, Z_n) \quad \text{for all} \quad 1 \leq i \leq n \quad . \quad \text{(36)}\]

**Proof:**

First, consider (iii):

\[I(W_0, W_2; Z) = \sum_{i=1}^{n} I(W_0, W_2; Z_i | Z_i^{i+1}) \]

\[\leq \sum_{i=1}^{n} I(W_0, W_2; Z_i^{i+1}; Z_i) \]

\[\leq \sum_{i=1}^{n} I(U_i; Z_i) \quad . \]

Next, using the independence of \( W_0, W_1, W_2 \), note that

\[I(W_1; Y | W_0) \leq I(W_1; Y | W_0, W_2) \]

and

\[I(W_2; Z | W_0) \leq I(W_2; Z | W_0, W_1) \quad . \quad \text{(37)} \]

Now consider (ii):

\[I(W_1; Y | W_0) + I(W_0, W_2; Z) \leq \sum_{i=1}^{n} \left[ I(W_1; Y_i | W_0, W_2, Y_{i-1}) + I(W_0, W_2; Z_i | Z_i^{i+1}) \right] \]

\[\leq \sum_{i=1}^{n} \left[ I(W_1; Y_i | W_0, W_2, Y_{i-1}, Z_i^{i+1}) + I(Z_i^{i+1}; Y_i | W_0, W_2, Y_{i-1}) + I(W_0, W_2, Z_i^{i+1}, Y_{i-1}; Z_i) - I(Y_{i-1}; Z_i | W_0, W_2, Z_i^{i+1}) \right] . \]
It can be shown that (see [8]) a summation by parts yields

\[ \sum_{i=1}^{n} I(Z_{i+1}^{i}; Y_{i} | W_{0}, W_{2}, Y_{i-1}) = \sum_{i=1}^{n} I(Y_{i-1}^{i}; Z_{i} | W_{0}, W_{2}, Z_{i+1}^{i+1}) . \] (38)

Hence two terms cancel in (37), and

\[ I(W_{1}; Y_{i} | W_{0}) + I(W_{0}, W_{2}; Z_{i}) \leq \sum_{i=1}^{n} I(W_{1}; Y_{i} | U_{i}) + I(U_{i}; Z_{i}) \]

\[ \leq \sum_{i=1}^{n} \left[ I(X_{i}; Y_{i} | U_{i}) + I(U_{i}; Z_{i}) \right] , \]

since \( U_{i} \rightarrow X_{i} \rightarrow (Y_{i}, Z_{i}) \) form a Markov chain in this order for all \( 1 \leq i \leq n \).

Similarly, consider (i)

\[ I(W_{2}; Z_{i} | W_{0}) + I(W_{0}, W_{1}; Y_{i}) \leq \sum_{i=1}^{n} \left[ I(W_{2}; Z_{i}^{i+1} | W_{0}, W_{1}, Z_{i+1}^{i+1}) + I(W_{0}, W_{1}; Y_{i} | Y_{i-1}) \right] , \]

\[ \leq \sum_{i=1}^{n} \left[ I(W_{2}; Z_{i}^{i+1} | W_{0}, W_{1}, Z_{i+1}^{i+1}, Y_{i-1}) + I(Y_{i-1}^{i+1}; Z_{i} | W_{0}, W_{1}, Z_{i+1}^{i+1}) \right. \]

\[ + I(W_{0}, W_{1}; Y_{i}^{i+1} | Y_{i-1}) - I(Z_{i+1}^{i+1}; Y_{i} | W_{0}, W_{1}, Y_{i-1}) \] (39)

Replacing \( W_{2} \) by \( W_{1} \) in (38) and substituting in (39) gives

\[ I(W_{2}; Z_{i} | W_{0}) + I(W_{0}, W_{1}; Y_{i}) \leq \sum_{i=1}^{n} \left[ I(W_{2}; Z_{i}^{i+1} | U_{i}^{i}) + I(U_{i}^{i}; Y_{i}) \right] \]

\[ \leq \sum_{i=1}^{n} \left[ I(X_{i}; Z_{i}^{i+1} | U_{i}^{i}) + I(U_{i}^{i}; Y_{i}) \right] , \]

where \( U_{i}^{i} = (W_{0}, W_{1}, Y_{i-1}, Z_{i+1}^{i+1}) \) and \( U_{i} \rightarrow X_{i} \rightarrow (Y_{i}, Z_{i}) \) form a Markov chain in this order for all \( 1 \leq i \leq n \).

Now, it can be easily shown that (11) implies that

\[ I(X; Z | U) \leq I(X; Y | U) , \] (40)

for all \( U \rightarrow X \rightarrow (Y, Z) \). Thus
\[ I(W_2; Z \mid W_0) + I(W_0; W_1; Y) \leq \sum_{i=1}^{n} \left[ I(X_i; Y_i \mid U_i^i) + I(U_i^i; Y_i) \right] \]

\[ = \sum_{i=1}^{n} I(X_i; Y_i) \]

and the proof of the lemma is completed.

Combining the lemma and Equations (29), (30), and (31) the converse follows easily.

**Acknowledgement:** The author would like to thank Prof. Thomas Cover for his help during the preparation of this paper.
Appendix

\( C \) is Convex: Let \((U_i, X_i, Y_i, Z_i), i = 1, 2\) be two collections of random variables with probability mass functions in \( P \), and let \( T \) be a random variable taking on values 1, 2 with probabilities \( \alpha \) and \( \bar{\alpha} \) respectively. Define
\[
\begin{aligned}
U_T &= U_i \\
X &= X_i \\
Y &= Y_i \\
Z &= Z_i
\end{aligned}
\]

for \( T = i \)

Then \( (T, U_T) \rightarrow X \rightarrow (Y, Z) \) form a Markov chain in this order. Now consider

\[
\alpha I(X_1; Y_1) + \bar{\alpha} I(X_2; Y_2) = \alpha I(X_1; Y_1 | U_1) + \alpha I(U_1; Y_1) + \bar{\alpha} I(X_2; Y_2 | U_2) + \bar{\alpha} I(U_2; Y_2)
\]

\[
= I(U_T; Y | T) + I(X; Y | U_T, T)
\]

\[
\leq I(U_T, T; Y) + I(X; Y | U_T, T)
\]

\[
= I(X; Y)
\]

Next
\[
\alpha I(X_1; Y_1 | U_1) + \alpha I(U_1; Z_1) + \bar{\alpha} I(X_2; Y_2 | U_2) + \bar{\alpha} I(U_2; Z_2) =
\]

\[
= I(X; Y | U_T, T) + I(U_T; Z | T)
\]

\[
\leq I(X; Y | U_T, T) + I(U_T, T; Z)
\]

and
\[
\alpha I(U_1; Z_1) + \bar{\alpha} I(U_2; Z_2) = I(U_T; Z | T) \leq I(U_T, T; Z)
\]
References


