

# THE CAPACITY OF A CLASS OF CHANNELS<sup>1</sup>

BY DAVID BLACKWELL, LEO BREIMAN, AND A. J. THOMASIAN

*University of California, Berkeley*

**1. Summary.** Shannon's basic theorem on the capacity of a channel is generalized to the case of a class of memoryless channels. A generalized capacity is defined and is shown to be the supremum of attainable transmission rates when the coding and decoding procedure must be satisfactory for every channel in the class.

**2. Definitions and Introduction.** For any positive integer  $n$  and any set  $\mathcal{A}$  we denote by  $\mathcal{A}^{(n)}$  the set of all  $n$ -tuples  $(x_1, \dots, x_n)$  with each  $x_i \in \mathcal{A}$ .

A channel, denoted by  $(\mathcal{A}, \mathcal{B}, P(y|x))$  or by  $P(y|x)$ , consists of two finite sets  $\mathcal{A}, \mathcal{B}$  having  $a \geq 2, b \geq 2$  elements, respectively, and a set of probability distributions  $P(\cdot|x)$  on  $\mathcal{B}$ , one for each  $x \in \mathcal{A}$ .  $P(y|x)$  is interpreted as the probability of receiving  $y \in \mathcal{B}$  given that  $x \in \mathcal{A}$  was transmitted.

The  $n$ -extension of a channel  $(\mathcal{A}, \mathcal{B}, P(y|x))$  is the channel  $(\mathcal{A}^{(n)}, \mathcal{B}^{(n)}, P(v|u))$  where  $v = (y_1, \dots, y_n) \in \mathcal{B}^{(n)}, u = (x_1, \dots, x_n) \in \mathcal{A}^{(n)}$  and  $P(v|u) = \prod_{i=1}^n P(y_i|x_i)$ .

When considering a class of channels,  $(\mathcal{A}, \mathcal{B}, P_\gamma(y|x))$  for  $\gamma \in \mathcal{C}$ , where  $\mathcal{C}$  is an index set, we shall always assume that the  $\mathcal{A}, \mathcal{B}$  sets are the same for each channel in the class. We shall sometimes denote such a class of channels by  $\mathcal{C}$ , the index set.

A  $(G, \epsilon_n, n)$  code for a class  $\mathcal{C}$  of channels for  $G \geq 1, \epsilon_n \geq 0, n$  a positive integer, is a sequence of  $[G]$  distinct elements of  $\mathcal{A}^{(n)}$ ;  $u_1, \dots, u_{[G]}$ ; where  $[G]$  is the largest integer  $\leq G$ , and a sequence of  $[G]$  disjoint subsets of  $\mathcal{B}^{(n)}$ ;  $B_1, \dots, B_{[G]}$ ; such that

$$P_\gamma(B_i^c | u_i) \leq \epsilon_n \quad \text{for } i = 1, \dots, [G] \quad \text{and all } \gamma \in \mathcal{C}.$$

The set  $\{u_1, \dots, u_{[G]}\}$  is called the set of input messages of the code and  $B_i$  is called the decoding set for  $u_i$ . We think of an input letter  $u_i$  of the code as being selected arbitrarily and transmitted over an unknown one of the channels  $P_\gamma, \gamma \in \mathcal{C}$ . The letter  $v$  is received with probability  $P_\gamma(v|u)$  and if  $v \in B_j$  it is decoded as  $u_j$ . Thus, the probability is  $\leq \epsilon_n$  that any input message  $u_i$  will be transmitted so as to be not decoded as  $u_i$ ; regardless of which channel in the class  $\mathcal{C}$  is used.

An  $R \geq 0$  is an attainable transmission rate for a class  $\mathcal{C}$  of channels if there exists a sequence of  $(e^{Rn}, \epsilon_n, n)$  codes for  $\mathcal{C}$  with  $\epsilon_n \rightarrow 0$ . Since  $\mathcal{A}^{(n)}$  has only  $a^n$  points we know that any attainable rate  $R \leq \log a$ . Clearly 0 is an attainable rate for any class of channels. For any class of channels  $\mathcal{C}$  we define  $T = T(\mathcal{C})$  to be the supremum of the set of attainable rates for  $\mathcal{C}$ .

Received February 16, 1959.

<sup>1</sup> This research was supported by the Office of Naval Research under Contract Nonr-222(53).

If  $(\mathcal{A}, \mathcal{B}, P_\gamma(y | x))$  for  $\gamma \in \mathcal{C}$  is a class of channels and  $Q(x)$  is a given probability distribution on  $\mathcal{A}$  then for each  $\gamma \in \mathcal{C}$  we let  $P_\gamma(x, y) = P_\gamma(y | x)Q(x)$  and we define on  $\mathcal{A} \times \mathcal{B}$  the random variable  $J_\gamma$  by

$$J_\gamma(x, y) = \log \frac{P_\gamma(x, y)}{P_\gamma(x)P_\gamma(y)} \quad \text{if } P_\gamma(x, y) > 0$$

$$= 0 \quad \text{if } P_\gamma(x, y) = 0.$$

The dependence of  $P_\gamma$  and  $J_\gamma$  on  $Q$  will usually not be exhibited. Since we will often be interested in expressions of the form  $x \log x$  it is natural to define  $\log 0 = 0$ . We will denote the expectation of a random variable  $X$  with respect to the  $P_\gamma$  distribution by  $E_\gamma X$ . If  $\mathcal{C}$  has only one element we may drop the subscript  $\gamma$ . Finally for any class  $\mathcal{C}$  of channels we define the capacity of the class  $\mathcal{C}$  by

$$C(\mathcal{C}) = C = \sup_{Q(x)} \inf_{\gamma \in \mathcal{C}} E_\gamma J_\gamma$$

where the sup is over all distributions  $Q$  on  $\mathcal{A}$ .

In the case considered by Shannon,  $\mathcal{C}$  has only one element and our formula reduces to  $C = \sup_Q EJ$ , which is the usual formula for the capacity of a memoryless channel. Shannon's theorem then states that  $T = C$ .  $T \geq C$ ,  $T \leq C$  are called the direct and converse halves, respectively. This theorem for a single channel has been proved in various ways and under various conditions by Shannon [12], [13], McMillan [11], Feinstein [6], Khinchin [9], Wolfowitz [14], Blackwell, Breiman, and Thomasian [1]. We will show that within the framework that has been set up

$$T(\mathcal{C}) = C(\mathcal{C})$$

always holds true. This result follows immediately from Theorem 1 which also gives an exponential error bound for any rate  $R < C$ .

**THEOREM 1:** *Let  $(\mathcal{A}, \mathcal{B}, P_\gamma(y | x))$  for  $\gamma \in \mathcal{C}$  be any class of channels.*

(a) *For any integer  $n$  and any  $R > 0$  such that  $0 \leq C - R \leq 1/2$  there is an  $(e^{Rn}, \epsilon_n, n)$  code for  $\mathcal{C}$  with*

$$\epsilon_n = Ae^{-\frac{(C-R)^2}{B}n}$$

where

$$A = \left[ \frac{2^{10} ab^3}{(C - R)^2} \right]^{2ab} \quad \text{and} \quad B = 2^7 ab.$$

(b) *For any integer  $n$  and  $R > C$  if  $e^{Rn} \geq 2$  then any  $(e^{Rn}, \epsilon_n, n)$  code for  $\mathcal{C}$  must satisfy*

$$\epsilon_n \geq 1 - \frac{C + \frac{\log 2}{n}}{R - \frac{\log 2}{n}}.$$

The sequence of steps used in proving Theorem 1 will be outlined. Theorem 2 presents a basic inequality, for a single channel, which is contained implicitly

in Feinstein [8]. This inequality is of independent interest since it gives the same bound for the maximum probability of error that Shannon [13] gives for the average probability of error. Theorem 2 permits a simple proof of  $T \geq C$  for a single channel. Lemma 2 shows that  $\sup_Q$  in the definition of  $C(\mathcal{C})$  can be replaced by  $\max_Q$ . Theorem 3 gives an exponential bound on the error of a code for one channel, which depends only on  $a, b, (C - R)^2$ . This is convenient in that the particular probabilities  $P(y | x)$  may not be known and, in any case, need not be computed with. Results related to Theorem 3 have been given by Elias [3] and [4], Feinstein [7], Shannon [13], and Wolfowitz [14].

Lemma 3 generalizes the inequality of Theorem 2 to the case when  $\mathcal{C}$  has a finite number of elements, and Theorem 4 generalizes the exponential error bound of Theorem 3 to this case.

Lemma 4 shows that for a given  $\mathcal{A}, \mathcal{B}$  there is a large finite number of channels on  $\mathcal{A}, \mathcal{B}$  such that any channel on  $\mathcal{A}, \mathcal{B}$  is close, in several senses, to one of them. Lemma 5 shows that if a channel has a sequence of codes  $(e^{Rn}, \epsilon_n, n)$  with  $\epsilon_n = e^{-Bn}$  for large  $n$ , with  $B > 0$ , then this same sequence of codes can be used for all channels in a certain neighborhood of the channel. This result justifies some of our attention to exponential error bounds. The technique of Lemma 5 can also be used to get some similar results when the channel probabilities vary from letter to letter.

At this point the direct half of Theorem 1 is demonstrated by approximating the class  $\mathcal{C}$  of channels by a certain finite set of channels  $\mathcal{C}'$  from Lemma 4; obtaining an exponential error bound code for  $\mathcal{C}'$  from Theorem 4; and using Lemma 5 to show that such a code must be satisfactory for  $\mathcal{C}$ .

The converse half of Theorem 1 is then proved.

Before proceeding to the proofs we pause to clear up one point. It is obvious that

$$C(\mathcal{C}) \leq \inf_{\gamma \in \mathcal{C}} \sup_{Q(x)} E_{\gamma} J_{\gamma},$$

i.e.,  $C(\mathcal{C}) \leq$  the capacity of every channel in  $\mathcal{C}$ . We now exhibit an example where  $C(\mathcal{C}) \neq \inf$  of the capacities of channels in  $\mathcal{C}$ . Let  $\mathcal{A} = \mathcal{B} = \{1, 2, 3, 4\}$ ,  $\mathcal{C} = \{1, 2\}$ , and let  $P_1(y | x)$  and  $P_2(y | x)$  be defined by the left and right following matrices, respectively.

$$\begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{pmatrix} \begin{pmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

Let  $Q(x)$  be any distribution on  $\mathcal{A}$  and let  $H_i(Y) = -\sum_y P_i(y) \log P_i(y)$ ,  $H_i(Y | X) = -\sum_x Q(x) \sum_y P_i(y | x) \log P_i(y | x)$ . Using the fact that  $\log x = (\log 2) \log_2 x$  we see that  $(\log 2)^{-1} H_1(Y | X) = Q(1) + Q(2) + 2Q(3) + 2Q(4) = 1 + Q(3) + Q(4)$ . Also from Feinstein [8], p. 15 we have  $(\log 2)^{-1} H_1(Y) \leq 2$  so that  $E_1 J_1 = H_1(Y) - H_1(Y | X) \leq (\log 2)(Q(1) + Q(2))$ . Similarly  $E_2 J_2 \leq (\log 2)(Q(3) + Q(4))$  so that  $C(\mathcal{C}) \leq (1/2) \log 2$ . The case  $Q(i) = 1/4$  for  $i = 1, \dots, 4$  shows that  $C(\mathcal{C}) = (1/2) \log 2$ ; the case

$Q(1) = Q(2) = 1/2$  shows the capacity of channel one to be  $\log 2$ ; the case  $Q(3) = Q(4) = 1/2$  shows the capacity of channel two to be  $\log 2$ . Thus for this example

$$\frac{1}{2} \log 2 = C(\mathcal{C}) < \inf_{\gamma \in \mathcal{C}} \sup_{Q(x)} E_{\gamma} J_{\gamma} = \log 2.$$

**3. A basic inequality.**

**THEOREM 2:** For any channel  $(\mathcal{A}, \mathcal{B}, P(y | x))$ , any distribution  $Q(x)$  on  $\mathcal{A}$ ,  $\alpha > 0, G \geq 1$  there is a  $(G, \epsilon, 1)$  code for the channel with  $\epsilon = Ge^{-\alpha} + P(J \leq \alpha)$ .

**PROOF:** It is clearly sufficient to construct an  $(M, \epsilon, 1)$  code with the same  $\epsilon$  as in the theorem and with  $M \geq G$ . Let  $A = \{J > \alpha\}$  and for any  $x_0 \in \mathcal{A}$  let  $A_{x_0} = \{(x, y) | (x_0, y) \in A\}$ .  $P(J \leq \alpha) \leq \epsilon$  so that  $P(A) \geq 1 - \epsilon$ , hence there is an  $x_1$  such that  $P(A | x_1) \geq 1 - \epsilon$ . Let  $B_1 = A_{x_1}$ . (Each  $B_k$  will be a cylinder set with base in  $\mathcal{B}$ . The base of  $B_k$  will be the decoding set for  $x_k$ .) At the  $k$ th step select  $x_k$  such that  $P(B_k | x_k) \geq 1 - \epsilon$  where

$$B_k = \bigcup_1^k A_{x_i} - \bigcup_1^{k-1} A_{x_i}.$$

This process will terminate at some  $M \geq 1$ . For every  $x$

$$P\left(A - A \cap \left(\bigcup_1^M A_{x_i}\right) \middle| x\right) < 1 - \epsilon$$

otherwise we could add this  $x$  to  $x_1, \dots, x_M$  contradicting the definition of  $M$ . Thus

$$\begin{aligned} P(A) &= P\left(A \cap \left(\bigcup_1^M A_{x_i}\right)\right) + P\left(A - A \cap \left(\bigcup_1^M A_{x_i}\right)\right) \\ &\leq \sum_1^M P(A_{x_i}) + 1 - \epsilon. \end{aligned}$$

Now if  $(x, y) \in A$  then  $J(x, y) > \alpha$  so that  $P(y | x) > P(y)e^{\alpha}$ . For fixed  $x$  sum both sides of this inequality over all  $y$  such that  $(x, y) \in A$ . Then

$$1 \geq P(A | x) \geq P(A_x)e^{\alpha}.$$

Thus  $P(A_x) \leq e^{-\alpha}$  for any  $x \in \mathcal{A}$  so that  $P(A) \leq Me^{-\alpha} + 1 - \epsilon$ . Since  $P(A) = Ge^{-\alpha} + 1 - \epsilon$ , we have  $M \geq G$ . Clearly the  $B_1, \dots, B_M$  are disjoint and

$$P(B_k | x_k) \geq 1 - \epsilon$$

for  $k = 1, \dots, M$  so the proof is completed.

Consider a single channel  $(\mathcal{A}, \mathcal{B}, P(y | x))$  and let  $Q(x)$  be specified and determine  $P(x, y), J(x, y)$ . Applying Theorem 2 to  $(\mathcal{A}^{(n)}, \mathcal{B}^{(n)}, P(v | u))$  and  $Q(u) = Q(x_1) \cdots Q(x_n)$  with  $\alpha = n(R + EJ)/2, G = e^{Rn}$  we see that for any  $R$  such that  $0 < R < EJ$  there is an  $(e^{Rn}, \epsilon_n, n)$  code for  $(\mathcal{A}, \mathcal{B}, P(y | x))$  with

$$\epsilon_n = e^{-(EJ-R)n/2} + P\left(\frac{1}{n} J' \leq \frac{R + EJ}{2}\right).$$

Now

$$J'(u, v) = \log \frac{P(u, v)}{P(u)P(v)} \quad \text{if } P(u, v) > 0$$

$$= 0 \quad \text{otherwise.}$$

Let  $J''(u, v) = \sum_1^n J_i(x_i, y_i)$  where

$$J_i(x_i, y_i) = \log \frac{P(x_i, y_i)}{P(x_i)P(y_i)} \quad \text{if } P(x_i, y_i) > 0$$

$$= 0 \quad \text{otherwise.}$$

Clearly  $P(J' = J'') = 1$  and  $J''$  is the sum of  $n$  independent random variables each having the distribution of  $J(x, y)$ . Since  $EJ > (R + EJ)/2$  we see that  $\epsilon_n \rightarrow 0$ . Now it is easily seen (and we will shortly prove even more) that for a fixed channel  $EJ$  is a continuous function of  $(Q(x_1), Q(x_2), \dots, Q(x_n))$  and since the domain of the function is a closed bounded subset of Euclidean space the supremum is actually achieved. Thus for any channel  $(\mathfrak{A}, \mathfrak{B}, P(y | x))$  there is a distribution  $Q(x)$  on  $\mathfrak{A}$  such that  $C = EJ$ . Using this  $Q(x)$  in the earlier portions of this paragraph we obtain the direct half of Shannon's theorem for a memoryless channel:  $T \geq C$ .

By introducing a brief epsilon argument in the proof of the direct half of Shannon's theorem we could clearly have ignored the question of whether or not there is a maximizing  $Q(x)$ . Although the fact that there is a maximizing  $Q(x)$  in the general case of a class of channels is not vital in the following work, we will pause to prove this fact now. The proof is based on Lemma 1 which will be needed later.

LEMMA 1: Let  $Q(x), Q'(x)$  be any two distributions on  $\mathfrak{A}$  such that

$$|Q(x) - Q'(x)| \leq \epsilon \leq 1/e \text{ for all } x \in \mathfrak{A}.$$

Then

$$|H(X) - H'(X)| \leq a\epsilon^{1/2}$$

where  $H(X) = -\sum_x Q(x) \log Q(x)$  and  $H'(X) = -\sum_x Q'(x) \log Q'(x)$ .

PROOF: Let

$$f(y) = [-(y + \epsilon) \log (y + \epsilon)] - [-y \log y]$$

where  $0 < \epsilon \leq 1/e$  and  $0 \leq y \leq 1 - \epsilon$ . Then  $f(0) = -\epsilon \log \epsilon > 0$  and  $f(1 - \epsilon) = (1 - \epsilon) \log (1 - \epsilon) < 0$  also

$$f'(y) = -\log (y + \epsilon) - 1 + \log y + 1 = \log \frac{y}{y + \epsilon} < 0$$

so that  $|f(y)| \leq \max \{-\epsilon \log \epsilon, -(1 - \epsilon) \log(1 - \epsilon)\}$ . Now

$$(1 - \epsilon) \log \frac{1}{1 - \epsilon} \leq (1 - \epsilon) \left( \frac{1}{1 - \epsilon} - 1 \right) = \epsilon \leq -\epsilon \log \epsilon$$

since  $\epsilon \leq 1/e$ . Thus

$$|f(y)| \leq -\epsilon \log \epsilon = \frac{\epsilon^{\frac{1}{2} \log \frac{1}{\epsilon}}}{\left(\frac{1}{\epsilon}\right)^{\frac{1}{2}}} \leq \epsilon^{\frac{1}{2}}$$

since  $x^{1/2} - \log x \geq 2 - \log 4 > 0$  for  $x > 0$ . Applying the result  $|f(y)| \leq \epsilon^{1/2}$  to  $y = p$ ,  $\epsilon = q - p$  where  $0 \leq p \leq q \leq 1$  and  $|q - p| \leq 1/e$  we see that

$$|[-p \log p] - [-q \log q]| \leq (|p - q|)^{1/2}$$

which easily gives us the bound on  $|H(X) - H'(X)|$  completing the proof.

LEMMA 2: For any class of channels  $(\mathfrak{A}, \mathfrak{B}, P_\gamma(y|x))$  for  $\gamma \in \mathfrak{C}$ ,

$$C = \max_{Q(x)} \inf_{\gamma \in \mathfrak{C}} E_\gamma J_\gamma.$$

PROOF: Let  $(\mathfrak{A}, \mathfrak{B}, P(y|x))$  be a channel and  $Q(x)$  a distribution on  $\mathfrak{A}$  determining  $P(x, y) = P(y|x)Q(x)$  and  $J(x, y)$ . Clearly  $EJ = H(X) + H(Y) - H(X, Y)$  where  $H(X) = -\sum_x P(x) \log P(x)$ ,  $H(Y) = -\sum_y P(y) \log P(y)$ ,  $H(X, Y) = -\sum_{x,y} P(x, y) \log P(x, y)$ . Let  $Q'(x)$  be another distribution on  $\mathfrak{A}$  determining  $P'(x, y) = P(y|x)Q'(x)$  and  $J'(x, y)$ , and note that  $E'J' = H'(X) + H'(Y) - H'(X, Y)$  where the primed quantities have analogous definitions. Assume that  $|Q(x) - Q'(x)| \leq \epsilon \leq 1/e$  for all  $x \in \mathfrak{A}$ . Clearly  $|P(x, y) - P'(x, y)| \leq P(y|x) |Q(x) - Q'(x)| \leq \epsilon$  and  $|P(y) - P'(y)| \leq \sum_x |P(x, y) - P'(x, y)| \leq a\epsilon$ . Applying Lemma 1 we get

$$\begin{aligned} |EJ - E'J'| &\leq |H(X) - H'(X)| + |H(Y) - H'(Y)| \\ &\quad + |H(X, Y) - H'(X, Y)| \\ &\leq a\epsilon^{1/2} + b(a\epsilon)^{1/2} + ab\epsilon^{1/2} \leq (a + 2ab)\epsilon^{1/2}. \end{aligned}$$

Thus not only is  $EJ$  continuous in  $Q(x)$  but it is continuous in  $Q(x)$  uniformly in  $Q(x)$  and  $P(y|x)$ . We easily take  $\inf_{\gamma \in \mathfrak{C}}$  on the inequalities

$$E'_\gamma J'_\gamma - (a + 2ab)\epsilon^{1/2} \leq E_\gamma J_\gamma \leq E'_\gamma J'_\gamma + (a + 2ab)\epsilon^{1/2}$$

and see that  $\inf_{\gamma \in \mathfrak{C}} E_\gamma J_\gamma$  is continuous in  $Q(x)$  so that once again there is a maximizing  $Q(x)$  and Lemma 2 is proved.

#### 4. The error bound for one channel.

THEOREM 3: Let  $(\mathfrak{A}, \mathfrak{B}, P(y|x))$  be any channel. For any integer  $n$  and any  $R > 0$  such that  $0 \leq C - R \leq 1/2$ , there is an  $(e^{Rn}, \epsilon_n, n)$  code for the channel with

$$\epsilon_n = 2e^{-\frac{(C-R)^2}{16ab} n}.$$

PROOF: Applying Theorem 2 to  $(\mathfrak{Q}^{(n)}, \mathfrak{B}^{(n)}, P(v|u))$  with  $Q(u) = Q(x_1) \cdots Q(x_n)$ , where  $Q(x)$  is any distribution on  $\mathfrak{A}$ ,  $G = e^{Rn}$ ,  $\alpha = (R + \theta)n$  we see that for any  $R > 0$ ,  $\theta > 0$  there is an  $(e^{Rn}, \epsilon_n, n)$  code for  $(\mathfrak{A}, \mathfrak{B}, P(y|x))$  with

$$\epsilon_n = e^{-n\theta} + P(J'' \leq n(R + \theta))$$

where, as shown in Section 3,  $J''$  is the sum of  $n$  independent random variables, each having the distribution of  $J(x, y)$ . Select  $R > 0$ ,  $0 \leq EJ - R \leq 1/2$  and let  $\theta = (EJ - R)^2$ . Then  $R + \theta \leq R + (EJ - R)/2 = (EJ + R)/2$ .

Thus it remains only to show that

$$P(J'' \leq n(EJ + R)\frac{1}{2}) \leq e^{-\frac{(EJ-R)^2}{16ab}n}$$

(we will need this result later) for we can then choose  $Q$  so that  $C = EJ$ .

A method due to Chernoff [2] will be used to bound the probability in question. Let  $0 \leq t \leq 1$ , then

$$\begin{aligned} P\left(0 \leq \frac{n(EJ + R)}{2} - J''\right) &\leq Ee^{t\left[\frac{n(EJ+R)}{2} - J''\right]} = e^{\frac{tn(EJ+R)}{2}} Ee^{-tJ''} \\ &= \left[e^{\frac{t(EJ+R)}{2}} Ee^{-tJ}\right]^n \end{aligned}$$

so that we need show only that for a proper selection of  $t$ ,

$$e^{\frac{t(EJ+R)}{2}} Ee^{-tJ} \leq e^{-\frac{(EJ-R)^2}{16ab}}$$

Now

$$Ee^{-tJ} = 1 - tEJ + \frac{t^2}{2} EJ^2 e^{-\theta t}, \quad 0 < \theta < 1.$$

We need consider only  $(x, y)$  with  $P(x, y) > 0$ . Terms in  $EJ^2 e^{-\theta t}$  are of the form

$$\begin{aligned} P(x, y) \left(\frac{P(x)P(y)}{P(x,y)}\right)^{\theta t} \log^2 \frac{P(x, y)}{P(x)P(y)} &\leq P(x, y) \left(\frac{1}{P(x, y)}\right)^{\theta t} \log^2 \frac{P(x, y)}{P(x)P(y)} \\ &\leq (P(x, y))^{1-t} \log^2 \frac{P(x, y)}{P(x)P(y)} \leq (P(x, y))^{1-t} \log^2 P(x, y) \end{aligned}$$

where the last inequality followed from  $P(x, y) \leq P(x)P(y)/P(x, y) \leq 1/P(x, y)$ . Also

$$\begin{aligned} [(P(x, y))^{\frac{1-t}{2}} \log P(x, y)]^2 &= \left(\frac{2}{1-t}\right)^2 [(P(x, y))^{\frac{1-t}{2}} \log P(x, y)]^{\frac{1-t}{2}}]^2 \\ &\leq \left(\frac{2}{1-t}\right)^2 \frac{1}{e^2} \leq \frac{1}{(1-t)^2}. \end{aligned}$$

Thus

$$Ee^{-tJ} \leq 1 - tEJ + \frac{t^2}{2} \frac{ab}{(1-t)^2} \leq e^{-tEJ + \frac{t^2}{2} \frac{ab}{(1-t)^2}}$$

so that

$$e^{\frac{t(EJ+R)}{2}} Ee^{-tJ} \leq e^{-\frac{1}{2}J(t)}$$

where

$$f(t) = (EJ - R)t - t^2 \frac{ab}{(1 - t)^2}.$$

Let  $t = (EJ - R)/4ab \leq 1/8$  so that  $1/(1 - t)^2 \leq (8/7)^2$ , then

$$f\left(\frac{EJ - R}{4ab}\right) \geq \frac{(EJ - R)^2}{4ab} \left[1 - \left(\frac{8}{7}\right)^2 \frac{1}{4}\right] \geq \frac{(EJ - R)^2}{8ab}$$

completing the proof.

**5. The error bound for a finite set of channels.** Lemma 3 is needed in the proof of Theorem 4.

LEMMA 3: Let  $(\mathcal{A}, \mathcal{B}, P_\gamma(y | x))$  for  $\gamma \in \mathcal{C} = \{1, 2, \dots, L\}$  be a finite class of channels and let  $Q(x)$  be a distribution on  $\mathcal{A}$ , determining  $P_\gamma(x, y), J_\gamma(x, y)$ .

(a) Define a channel  $(\mathcal{A}, \mathcal{B}, P(y | x))$  by  $P(y | x) = (1/L) \sum_{\gamma=1}^L P_\gamma(y | x)$  and let  $Q(x)$  determine  $P(x, y), J(x, y)$ . Then for all  $\alpha, \delta$

$$P(J \leq \alpha) \leq \frac{1}{L} \sum_{\gamma=1}^L P_\gamma(J_\gamma \leq \alpha + \delta) + Le^{-\delta}.$$

(b) For any  $\alpha > 0, G \geq 1, \delta > 0$  there is a  $(G, \epsilon, 1)$  code for  $\mathcal{C}$  with

$$\epsilon = LGe^{-\alpha} + L^2e^{-\delta} + \sum_1^L P_\gamma(J_\gamma \leq \alpha + \delta).$$

PROOF: We first prove part (a).

$$P(J \leq \alpha) = \frac{1}{L} \sum P_\gamma(J \leq \alpha) \leq \frac{1}{L} \sum [P_\gamma(J_\gamma \leq \alpha + \delta) + P_\gamma(J_\gamma > \alpha + \delta; J \leq \alpha)]$$

so that we need only prove that  $P_\gamma(A_\gamma) \leq Le^{-\delta}$  where  $A_\gamma = (J_\gamma \alpha + \delta; J \leq \alpha)$ . For any  $(x, y) \in A_\gamma$  with  $P_\gamma(x, y) > 0$  we have

$$e^\alpha P(y) \geq P(y | x) \geq \frac{1}{L} P_\gamma(y | x) \geq \frac{1}{L} e^{\alpha + \delta} P_\gamma(y)$$

so that  $P_\gamma(y) \leq Le^{-\delta} P(y)$ . Summing this last inequality over all  $y$  such that there is an  $x$  with  $(x, y) \in A_\gamma$  we get  $P_\gamma(A_\gamma) \leq \sum P_\gamma(y) \leq Le^{-\delta}$  which completes the proof of part (a).

Applying Theorem 2 to the channel  $P(y | x)$  defined in part (a) and then using part (a) to bound  $P(J \leq \alpha)$  we find that there is a  $(G, \epsilon_0, 1)$  code for  $P(y | x)$  with

$$\epsilon_0 = Ge^{-\alpha} + P(J \leq \alpha) \leq Ge^{-\alpha} + \frac{1}{L} \sum_\gamma P_\gamma(J_\gamma \leq \alpha + \delta) + Le^{-\delta}.$$

Now  $P_\gamma(y | x) \leq LP(y | x)$  so that if  $x_i$  is an input letter for the  $(G, \epsilon_0, 1)$  code and  $B_i$  is its decoding set, then  $P_\gamma(B_i^c | x_i) \leq L P_\gamma(B_i^c | x_i) \leq L\epsilon_0$ . Thus the  $(G, \epsilon_0, 1)$  code for  $P(y | x)$  is a  $(G, L\epsilon_0, 1)$  code for  $\mathcal{C}$  and the lemma is proved.



**THEOREM 4:** *Let  $(\mathfrak{A}, \mathfrak{B}, P_\gamma(y|x))$  for  $\gamma \in \mathfrak{C} = \{1, 2, \dots, L\}$  be a finite class of channels. For any  $R > 0$  such that  $0 \leq C - R \leq 1/2$  there is an  $(e^{Rn}, \epsilon_n, n)$  code with*

$$\epsilon_n = 2L^2 e^{-\frac{(C-R)^2}{16ab}n}.$$

**PROOF.** Applying part (b) of Lemma 3 to the class of channels  $(\mathfrak{A}^{(n)}, \mathfrak{B}^{(n)}, P_\gamma(v|u))$  with  $Q(u) = Q(x_1) \cdots Q(x_n)$  and  $Q(x)$  a distribution for which  $C = \inf_{\gamma \in \mathfrak{C}} E_\gamma J_\gamma$  and  $G = e^{Rn}$ ,  $\alpha = (R + \theta/2)n$ ,  $\delta = \theta n/2$  we see that there is an  $(e^{Rn}, \epsilon_n, n)$  code for  $\mathfrak{C}$  with

$$\epsilon_n = (L + L^2)e^{-\theta n} + \sum_1^L P_\gamma \left( \frac{1}{n} J_\gamma \leq R + \theta \right).$$

Let  $\theta = (C - R)^2$  and note that  $R + (C - R)^2 \leq R + (C - R)/2 \leq R + (E_\gamma J_\gamma - R)/2 = (E_\gamma J_\gamma + R)/2$ . Thus,

$$\epsilon_n \leq (L + L^2)e^{-\frac{(C-R)^2}{16ab}n} + \sum_1^L P_\gamma \left( \frac{1}{n} J_\gamma \leq \frac{1}{2}(R + E_\gamma J_\gamma) \right).$$

Now

$$P_\gamma \left( \frac{1}{n} J_\gamma \leq \frac{1}{2}(R + E_\gamma J_\gamma) \right) \leq P_\gamma \left( \frac{1}{n} J_\gamma \leq \frac{1}{2}(R' + E_\gamma J_\gamma) \right)$$

where  $R' = E_\gamma J_\gamma - (C - R) \geq R$  and  $0 \leq E_\gamma J_\gamma - R' \leq 1/2$ . Therefore, we can apply the result obtained in the proof of Theorem 3 and get

$$P_\gamma \left( \frac{1}{n} J_\gamma \leq \frac{1}{2}(R' + E_\gamma J_\gamma) \right) \leq e^{-\frac{(E_\gamma J_\gamma - R')^2}{16ab}n} = e^{-\frac{(C-R)^2}{16ab}n}.$$

Now  $L \geq 2$  so that  $2L + L^2 = L(L + 2) \leq 2L^2$  and since Theorem 4 reduces to Theorem 3 for  $L = 1$ , the proof is completed.

**6. The direct half of Theorem 1.** Lemmas 4 and 5 are needed for the proof of part (a) of Theorem 1.

**LEMMA 4:** *Let  $\mathfrak{A}, \mathfrak{B}$  be given. For every integer  $M \geq 2b^2$  there is a class of channels  $(\mathfrak{A}, \mathfrak{B}, P_j(y|x))$  with  $\epsilon \in \mathfrak{D}_M$ , where  $\mathfrak{D}_M$  has at most  $(M + 1)^{ab}$  elements, such that for any channel  $(\mathfrak{A}, \mathfrak{B}, P(y|x))$  there is a channel  $(\mathfrak{A}, \mathfrak{B}, P'(y|x))$  in  $\mathfrak{D}_M$  such that:*

- (a)  $|P(y|x) - P'(y|x)| \leq b/M$  for all  $x, y$ .
- (b)  $P(y|x) \leq e^{2b^2/M} P'(y|x)$  for all  $x, y$ .
- (c) For any distribution  $Q(x)$  on  $\mathfrak{A}$  let  $P(x, y) = P(y|x)Q(x)$ ,  $P'(x, y) = P'(y|x)Q(x)$ , then

$$|EJ - E'J'| \leq 2b \left( \frac{b}{M} \right)^{1/2}.$$

**PROOF.** Let  $\mathfrak{D}_M$  be the class of channels  $(\mathfrak{A}, \mathfrak{B}, P(y|x))$  such that for all  $x, y$  we have  $MP(y|x) = \text{an integer}$ . Clearly  $\mathfrak{D}_M$  has at most  $(M + 1)^{ab}$  elements. Given the distributions  $P(y|x)$  we will first construct  $P'(y|x)$  and prove (a),

(b). For this purpose it is enough to carry out the construction for one  $x_0$ . Arrange the “b” numbers  $P(y | x_0)$  in ascending order and designate them by  $p_1 \leq p_2 \leq \dots \leq p_b$ . For  $i = 1, \dots, (b - 1)$  select  $p'_i$  uniquely by  $p_i \leq p'_i < p_i + 1/M$ ,  $Mp'_i =$  an integer.  $p'_i$  will be  $P'(y | x_0)$  with the  $y$  being the one corresponding to  $p_i$ . Clearly

$$p_i \leq e^{\frac{2b^2}{M}} p'_i \quad \text{and} \quad |p_i - p'_i| \leq \frac{b}{M}$$

for  $i = 1, \dots, (b - 1)$ . It remains to show that if  $p'_b = 1 - \sum_{i=1}^{b-1} p'_i$  then  $p'_b \geq 0$  and  $p'_b, p'_b$  satisfy the same relations. Now

$$p'_b \geq 1 - \sum_1^{b-1} \left( p_i + \frac{1}{M} \right) \geq p_b - \frac{b}{M} \geq \frac{1}{b} - \frac{b}{M} \geq \frac{1}{b} - \frac{1}{2b} = \frac{1}{2b}.$$

Thus  $p'_1, \dots, p'_b$  form a distribution and  $p_b \geq p'_b \geq p_b - b/M$  so that

$$|p_b - p'_b| \leq b/M.$$

Also

$$p_b \leq p'_b + \frac{b}{M} \leq p'_b + \frac{2b^2}{M} \frac{1}{2b} \leq p'_b \left( 1 + \frac{2b^2}{M} \right) \leq e^{\frac{2b^2}{M}} p'_b$$

completing the proof of parts (a) and (b).

In the proof of part (c) we will use part (a) and Lemma 1. In order to use Lemma 1 we observe that  $b/M \leq 1/2b \leq 1/4 < 1/e$ . We also note that

$$|P(y) - P'(y)| \leq \sum_x |P(y | x) - P'(y | x)| Q(x) \leq b/M.$$

Now

$$\begin{aligned} |EJ - E'J'| &\leq \left| \left[ -\sum_y P(y) \log P(y) \right] - \left[ -\sum_y P'(y) \log P'(y) \right] \right| \\ &+ \left| \left[ -\sum_{x,y} P(x, y) \log P(x, y) \right] - \left[ -\sum_{x,y} P'(x, y) \log P'(x, y) \right] \right| \leq b \left( \frac{b}{M} \right)^{1/2} \\ &+ \sum_x Q(x) \left| \left[ -\sum_y P(y | x) \log P(y | x) \right] - \left[ -\sum_y P'(y | x) \log P'(y | x) \right] \right| \\ &\leq b \left( \frac{b}{M} \right)^{1/2} + b \left( \frac{b}{M} \right)^{1/2} \end{aligned}$$

and the lemma is proved.

LEMMA 5: Let  $(\mathcal{A}, \mathcal{B}, P'(y | x))$ ,  $(\mathcal{A}, \mathcal{B}, P(y | x))$  be two channels and  $A$  a non-negative number such that  $P(y | x) \leq e^A P'(y | x)$  for all  $x, y$ . Any  $(e^{Rn}, \epsilon_n, n)$  code for  $(\mathcal{A}, \mathcal{B}, P'(y | x))$  is an  $(e^{Rn}, \epsilon_n e^{An}, n)$  code for  $(\mathcal{A}, \mathcal{B}, P(y | x))$ .

PROOF: Let  $u = (x_1, \dots, x_n) \in \mathcal{A}^{(n)}$ ,  $v = (y_1, \dots, y_n) \in \mathcal{B}^{(n)}$ . Then

$$P(v | u) = \prod_1^n P(y_i | x_i) \leq e^{An} \prod_1^n P'(y_i | x_i) = e^{An} P'(v | u).$$

Thus for any subset  $D$  of  $\mathcal{B}^{(n)}$  and any  $u \in \mathcal{A}^{(n)}$  we have

$$P(D | u) \leq e^{An} P'(D | u).$$

Let  $u_i \in \mathcal{A}^{(n)}$  be an input message and  $B_i$  the corresponding decoding set of an  $(e^{Rn}, \epsilon_n, n)$  code for  $(\mathcal{A}, \mathcal{B}, P'(y|x))$ . Then

$$P(B_i^c | u_i) \leq e^{An} P'(B_i^c | u_i) \leq e^{An} \epsilon_n$$

and the proof is completed.

We turn now to the proof of part (a) of Theorem 1. For each  $P(y|x) \in \mathcal{C}$  select a  $P'(y|x) \in \mathcal{D}_M$  according to Lemma 4 and let  $\mathcal{C}'$  denote this set of channels. Let  $C' = C(\mathcal{C}')$ . Since  $\mathcal{C}'$  has at most  $(M+1)^{ab}$  elements we know from Theorem 4 that if  $R' > 0, 0 \leq C' - R' \leq 1/2$  then there is an  $(e^{R'n}, \epsilon'_n, n)$  code for  $\mathcal{C}'$  with

$$\epsilon'_n = 2(M+1)^{2ab} e^{-\frac{(C'-R')^2}{16ab}n}$$

For each  $P(y|x) \in \mathcal{C}$  there is a  $P'(y|x) \in \mathcal{C}'$  such that

$$P(y|x) \leq e^{\frac{2b^2}{M}} P'(y|x)$$

so that from Lemma 5 the code which we have for  $\mathcal{C}'$  is an  $(e^{R'n}, \epsilon_n, n)$  code for  $\mathcal{C}$  with

$$\epsilon_n = 2(M+1)^{2ab} \exp \left\{ -\left[ \frac{(C'-R')^2}{16ab} - \frac{2b^2}{M} \right] n \right\}$$

Let  $C = C(\mathcal{C})$  and let  $Q(x)$  be a maximizing distribution for  $\mathcal{C}$ . We wish to show that  $C'$  cannot be very much smaller than  $C$ . For every  $P'(y|x) \in \mathcal{C}'$  there is a  $P(y|x) \in \mathcal{C}$  such that  $EJ \leq E'J' + 2b(b/M)^{1/2}$  where we use  $Q(x)$  in both cases. Thus for every  $P'(y|x) \in \mathcal{C}'$

$$C = \inf_{\mathcal{C}} EJ \leq E'J' + 2b \left( \frac{b}{M} \right)^{1/2}$$

so that

$$C \leq \inf_{\mathcal{C}'} E'J' + 2b \left( \frac{b}{M} \right)^{1/2} \leq C' + 2b \left( \frac{b}{M} \right)^{1/2}$$

Let  $R > 0$  be given such that  $0 < C - R \leq 1/2$ . We must show how to select  $R'$  and  $M$  to get our result into the final form.

We select an integer  $M$  such that

$$\frac{2^8 ab^3}{(C-R)^2} \leq M \quad \text{and} \quad (M+1) \leq \frac{2^9 ab^3}{(C-R)^2}$$

so that

$$2b \left( \frac{b}{M} \right)^{1/2} \leq \frac{C-R}{2} \quad \text{and} \quad \frac{2b^2}{M} \leq \frac{(C-R)^2}{2^7 ab}$$

We define  $R'$  by

$$C' - R' = C - R - 2b \left( \frac{b}{M} \right)^{1/2} \geq \frac{C-R}{2} > 0.$$

Clearly  $C' - R' \leq 1/2$  so that we have an  $(e^{R'n}, \epsilon_n, n)$  code for  $\mathfrak{C}$  with

$$\begin{aligned} \epsilon_n &\leq 2(M + 1)^{2ab} \exp - \left\{ \frac{(C - R)^2}{4(16ab)} - \frac{(C - R)^2}{2^7 ab} \right\} \\ &\leq 2 \left[ \frac{2^9 ab^3}{(C - R)^2} \right]^{2ab} \exp - \left\{ \frac{(C - R)^2}{2^7 ab} \right\}. \end{aligned}$$

The inequality  $C \leq C' + 2b(b/M)^{1/2}$  shows that  $R' \geq R$  and an  $(e^{R'n}, \epsilon_n, n)$  code for  $\mathfrak{C}$  can easily be reduced to an  $(e^{Rn}, \epsilon_n, n)$  code for  $\mathfrak{C}$  so that part (a) of Theorem 1 is proved.

**7. Converse half of Theorem 1.** The proof is based on Lemma 6.

LEMMA 6: Let  $G$  be an integer,  $\mathfrak{A}$  a finite set and let  $u_1, \dots, u_G$  be distinct elements of  $\mathfrak{A}^{(n)}$ . Define  $Q(x)$  on  $\mathfrak{A}$  by

$$Q(x) = \frac{1}{nG} \sum_{j=1}^G (\text{the number of times that } x \text{ appears in } u_j).$$

Then any  $(G, \epsilon, n)$  code, for a channel  $(\mathfrak{A}, \mathfrak{B}, P(y | x))$  which uses these  $u_1, \dots, u_G$  for inputs must satisfy

$$(1 - \epsilon) \log G - \log 2 \leq nEJ$$

where  $Q(x)$  is used to define  $P(x, y)$  and  $J(x, y)$ .

PROOF: Define a distribution  $\nu(u)$  on  $\mathfrak{A}^{(n)}$  by  $\nu(u) = 1/G$  if  $u$  is one of  $u_1, \dots, u_G$  and  $\nu(u) = 0$  otherwise. Define a distribution  $P(u, v)$  on  $\mathfrak{A}^{(n)} \times \mathfrak{B}^{(n)}$  by  $P(u, v) = P(v | u)\nu(u)$  where  $P(v | u)$  is obtained from the  $n$ -extension of  $(\mathfrak{A}, \mathfrak{B}, P(y | x))$ . Now define  $n$  distributions on  $\mathfrak{A} \times \mathfrak{B}$  by

$$P^{(i)}(x, y) = P(y | x)\nu^{(i)}(x)$$

for  $i = 1, \dots, n$  where

$$\nu^{(i)}(x) = \sum_{x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n} \nu(x_1, \dots, x_{i-1}, x, x_{i+1}, \dots, x_n)$$

and observe that  $Q(x) = (1/n) \sum_{i=1}^n \nu^{(i)}(x)$ . Thus, the lemma will be proved if the following chain of inequalities is proved.

$$\begin{aligned} (1 - \epsilon) \log G - \log 2 &\leq \sum_{u,v} P(u, v) \log \frac{P(u, v)}{P(u)P(v)} \\ &\leq \sum_{i=1}^n \sum_{x,y} P^{(i)}(x, y) \log \frac{P^{(i)}(x, y)}{P^{(i)}(x)P^{(i)}(y)} \leq n \sum_{x,y} P(x, y) \log \frac{P(x, y)}{P(x)P(y)}. \end{aligned}$$

Using  $\log x = (\log 2) \log_2 x$  to convert a result from Feinstein [8], pp. 29, 39, 44; which is due to Fano [5]; we obtain the first inequality. The second inequality follows from page 30 of Feinstein [8]. We proceed to prove the third inequality. Now

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n \sum_{x,y} P^{(i)}(x, y) [\log P(y | x) - \log P^{(i)}(y)] \\ = \sum_{x,y} P(x, y) \log P(y | x) - \frac{1}{n} \sum_{i=1}^n \sum_y P^{(i)}(y) \log P^{(i)}(y) \end{aligned}$$

but

$$\begin{aligned} -\frac{1}{n} \sum_{i=1}^n \sum_y P^{(i)}(y) \log P^{(i)}(y) &\leq -\sum_y \left( \frac{1}{n} \sum_{i=1}^n P^{(i)}(y) \right) \log \left( \frac{1}{n} \sum_{i=1}^n P^{(i)}(y) \right) \\ &= -\sum_y P(y) \log P(y) = -\sum_{x,y} P(x, y) \log P(y) \end{aligned}$$

where this last inequality follows from Lemma 4 on page 16 of Feinstein [8]. Combining the above, we complete the proof of the third inequality and hence of the lemma.

From Lemma 6 we immediately obtain that if  $G$  is an integer then for any  $(G, \epsilon, n)$  code for a class  $\mathcal{C}$  of channels there is a  $Q(x)$  on  $\mathcal{A}$  such that

$$(1 - \epsilon) \log G - \log 2 \leq n \inf_{\gamma \in \mathcal{C}} E_\gamma J_\gamma \leq nC.$$

Now  $e^{Rn}$  may not be an integer but

$$\log [e^{Rn}] \geq \log (e^{Rn} - 1) \geq nR + \log(1 - e^{-Rn}) \geq nR - \log 2$$

so that

$$(1 - \epsilon) (nR - \log 2) \leq nC + \log 2$$

which completes the proof of part (b) of Theorem 1.

#### REFERENCES

- [1] DAVID BLACKWELL, LEO BREIMAN, A. J. THOMASIAN, "Proof of Shannon's transmission theorem for finite-state indecomposable channels," *Ann. Math. Stat.*, Vol. 29 (1958), pp. 1209-1220.
- [2] HERMAN CHERNOFF, "A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 493-507.
- [3] PETER ELIAS, "Coding for noisy channels," *I.R.E. Convention Record* (1955), part 4, pp. 37-44.
- [4] PETER ELIAS, "Coding for two noisy channels," *Proceedings of the London Symposium on Information Theory*, Butterworth Scientific Publications, London, 1955.
- [5] R. M. FANO, "Statistical theory of communication," notes on a course given at the Massachusetts Institute of Technology, 1952, 1954.
- [6] AMIEL FEINSTEIN, "A new basic theorem of information theory," *I.R.E. Trans. P.G.I.T.*, September, 1954, pp. 2-22.
- [7] AMIEL FEINSTEIN, "Error bounds in noisy channels without memory," *I.R.E. Trans. P.G.I.T.*, September, 1955, pp. 13-14.
- [8] AMIEL FEINSTEIN, *Foundations of Information Theory*, McGraw-Hill, New York, 1958.
- [9] A. I. KHINCHIN, "On the fundamental theorems of information theory," *Uspekhi Matematicheskikh Nauk.*, Vol. 21 (1956), pp. 17-75.
- [10] A. I. KHINCHIN, *Mathematical Foundations of Information Theory*, Dover Publications, Inc., 1957.
- [11] BROCKWAY McMILLAN, "The basic theorems of information theory," *Ann. Math. Stat.*, Vol. 24(1953), pp. 196-219.
- [12] C. E. SHANNON, "A mathematical theory of communication," *Bell System Technical Journal*, Vol. 27 (1948), pp. 379-423, and 623-656.
- [13] CLAUDE E. SHANNON, "Certain results in coding theory for noisy channels," *Information and Control*, Vol. 1 (1957), pp. 6-25.
- [14] J. WOLFOWITZ, "The coding of messages subject to chance errors," *Illinois J. Math.*, Vol. 1 (1957), pp. 591-606.