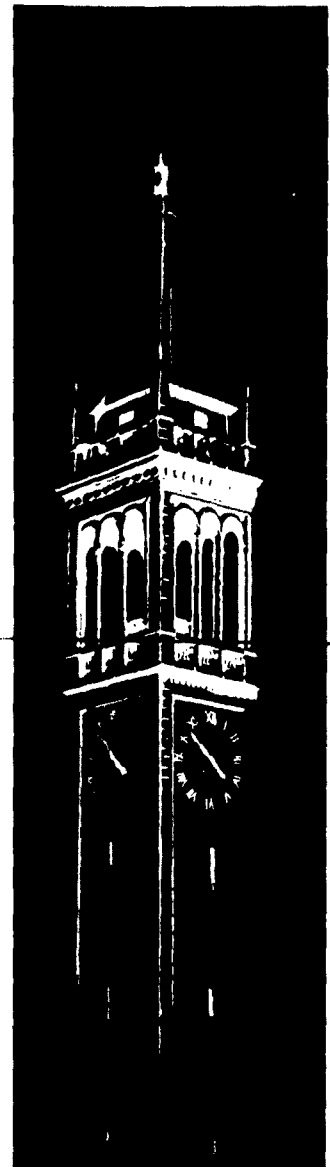

Generalized Coding and Uniform Finite Memory Codes

by

T. J. Wagner

Series No. 60, Issue No. 492
Contract No. Nonr-222(53)
December 26, 1962

ELECTRONICS RESEARCH LABORATORY
UNIVERSITY OF CALIFORNIA
BERKELEY CALIFORNIA

Electronics Research Laboratory
University of California
Berkeley, California

GENERALIZED CODING AND UNIFORM FINITE MEMORY CODES

by

T. J. Wagner

Institute of Engineering Research
Series No. 60, Issue No. 492

Contract No. Nonr-222(53)
December 26, 1962

ACKNOWLEDGMENT

I wish to express my gratitude to Professor Aram Thomasian for his patience and encouragement throughout the period of my graduate study at this university and especially for the benefit of his flawless intuition during the development of the research presented here.

T.J.W.

TABLE OF CONTENTS

SUMMARY

The research presented here, although motivated by the single theme of finding the channel capacity of a discrete memoryless channel for codes other than block codes, is divided into two essentially independent sets of results.

First, in section II a general framework for encoding and decoding is presented which includes block coding. The key concept used with the generalized codes is that of decoding rate. A weak converse is proven using decoding rate which shows that channel capacity for the generalized codes is the same as the usual block coding channel capacity C for a discrete memoryless channel.

Second, in section III uniform finite-memory codes are defined from the general framework after several motivating definitions of properties which seem natural to require of any code. Channel capacity C_u is defined for these codes but what its value is remains an open question. A class of channels is given for which C_u is nonzero for each member of the class. From the converse in section II it is known that $C_u \leq C$.

I. NOTATION AND PRELIMINARY FACTS

For any set A and any positive integer n, A^n denotes the set of all n-tuples of elements from A. $\prod_{i=-\infty}^{r} A_i$ denotes the set of all sequences $(\ldots x_{r-1}, x_r)$ of elements from A while A^I denotes the set of all sequences $\bar{x} = (x_1, x_2, \ldots)$ of elements from A. If $w \in A^n$ then $w(i) = x_i$ for $w = (x_1, \ldots x_n)$ and $1 \le i \le n$. Similarly, $\bar{x}(i) = x_i$ for $\bar{x} \in A^I$. For A a finite set, $|A|$ denotes the number of elements in A and $\mathcal{O}(A^I)$ denotes the $\mathcal{O}$-field of subsets of A^I determined by cylinder sets.

A <u>discrete memoryless channel</u> (DMC) is a triple $(B, \tilde{B}, p)$ where

(i) B is a finite set of elements called inputs,

(ii) $\tilde{B}$ is a finite set of elements called outputs,

(iii) $p = p(\cdot/\cdot)$ is a function on $B \times \tilde{B}$ such that $p(\cdot/y)$ is a probability distribution on $\tilde{B}$ for each $y \in B$, and

(iv) for each positive integer t and for all sequences $(y_1, \ldots y_t) \in B^t$, $(\tilde{y}_1, \ldots \tilde{y}_t) \in \tilde{B}^t$, $p(\tilde{y}_1, \ldots \tilde{y}_t / y_1, \ldots y_t) = \prod_{i=1}^{t} p(\tilde{y}_i / y_i)$.

The <u>n-extension of</u> $(B, \tilde{B}, p)$ is the DMC $(B^n, \tilde{B}^n, q)$ where n is a positive integer and for each $(y_1, \ldots y_n) \in B^n$, $(\tilde{y}_1, \ldots \tilde{y}_n) \in \tilde{B}^n$, $q(\tilde{y}_1, \ldots \tilde{y}_n / y_1, \ldots y_n) = \prod_{i=1}^{n} p(\tilde{y}_i / y_i)$.

A <u>source</u> is a sequence $\{X_i, i = 1,2,\ldots\}$ of finite-valued random variables which are independent, identically distributed with a uniform distribution. In section III a source $\{X_i, -\infty < i < \infty\}$ will be used. Frequently it will be helpful to think of the subscript i as corresponding to time.

Throughout this report all random variables are finite-valued unless otherwise noted. They will be denoted by capital letters X,Y,Z,... and their values by small letters x,y,z,... . The ranges of X and Y will be denoted by A and B respectively, where corresponding affixes are used when necessary. For example, Y ($\tilde{Y}$) will, with a subscript to distinguish order, denote an input (output) random variable to the channel $(B,\tilde{B},p)$. For any random variable Z, $p(z)$ will denote $P[Z=z]$, the probability that $Z=z$.

For finite-valued random variables U and V the numbers

$$I(U) = \sum_{\{u:p(u)>0\}} -p(u)\log p(u)$$

$$I(U/V) = \sum_{\{(u,v):p(u,v)>0\}} -p(u,v)\log\left(p(u,v)/p(v)\right)$$

are called the average uncertainty of U and the average uncertainty of U given V respectively. $J(U,V) = I(U) - I(U/V)$ is called the average mutual uncertainty of U and V. All logarithms used will be to the base 2.

For a probability distribution $p(\cdot)$ on B of $(B,\tilde{B},p)$ with $p(y,\tilde{y}) = p(y)p(\tilde{y}/y)$ and $p(\tilde{y}) = \sum_{y} p(y,\tilde{y})$ for $y \in B$, $\tilde{y} \in \tilde{B}$,

$$C = \sup_{p(\cdot)} J(Y,\tilde{Y}) \tag{1.1}$$

is called the block coding channel capacity of $(B,\tilde{B},p)$. A standard result[1] for the DMC $(B,\tilde{B},p)$ will be used without comment. Let k and n be positive integers and let $X_1,\ldots X_k$ be a sequence of random variables, all with the same range A. Then for any finite collection of functions $\{f_r\} = \{f_r : A^k \to B^n\}$ and any probability distribution

$p(x_1,\cdots x_k,r)$ for the random variables $X_1,\cdots X_k$ and the function f_r (a random variable)

$$J\left((X_1,\cdots X_k),(\tilde{Y}_1,\cdots \tilde{Y}_n)\right) \leq nC \tag{1.2}$$

where $p(x_1,\cdots x_k;\tilde{y}_1,\cdots \tilde{y}_n) = \sum_r p(x_1,\cdots x_k,r) \prod_{i=1}^{n} p(\tilde{y}_i/y_{ri})$ and

$(y_{r1},\cdots y_{rn}) = f_r(x_1,\cdots x_k)$.

One result used often in the sequel is Fano's Inequality[2]. For two arbitrary random variables X and Y (not necessarily a source output or channel input) the value of X is decided on from the occurrence of Y by a function $g:B\rightarrow A$ (a decoder in the terminology of this report). The probability of error $P\left[g(Y)\neq X\right]$ for any g is related to the average uncertainty of X given Y by Fano's Inequality:

$$I(X/Y) \leq h(P\left[g(Y)\neq X\right]) + P\left[g(Y)\neq X\right]\log|A| \tag{1.3}$$

where $h(x) = -x\log x - (1-x)\log(1-x)$, $0<x<1$, and $h(x)=0$ if $x=0,1$.

II. GENERAL CODING FRAMEWORK

In this section general codes are considered which (i) map infinite source sequences $(x_1,x_2,\cdots)$ into infinite sequences of channel inputs $(y_1,y_2,\cdots)$ and (ii) map infinite channel output sequences $(\tilde{y}_1,\tilde{y}_2,\cdots)$ into infinite decoded source sequences $(\tilde{x}_1,\tilde{x}_2,\cdots)$ in ways other than breaking up the sequences into independent blocks as usually done by block coding. The main result of this section is Theorem 1 which proves that channel capacity for these general codes is the same as the channel capacity for block codes. All channels considered are discrete memoryless channels.

To begin this section the basic facts of block coding are given. A standard way of sending a source output $\{X_i\}$ over a DMC $(B,\tilde{B},p)$ is by block coding. A function $f:A^k \to B^n$ (k and n are positive integers) called the encoder maps (encodes) k-tuples of source outputs into n-tuples of channel inputs as given by

$$\left(Y_{jn+1},\cdots Y_{(j+1)n}\right) = f\left(X_{jk+1},\cdots X_{(j+1)k}\right), \quad j=0,1,\cdots . \tag{2.1}$$

From the channel outputs a function $g:\tilde{B}^n \to A^k$ called the decoder maps (decodes) the channel outputs into source symbols as given by

$$\left(\tilde{X}_{jk+1},\cdots \tilde{X}_{(j+1)k}\right) = g\left(\tilde{Y}_{jn+1},\cdots \tilde{Y}_{(j+1)n}\right), \quad j=0,1,\cdots . \tag{2.2}$$

The goal of the code (f,g) is, of course, to have $\left(\tilde{X}_{jk+1},\cdots \tilde{X}_{(j+1)k}\right) = \left(X_{jk+1},\cdots X_{(j+1)k}\right)$ with high probability, $j=0,1,\cdots$. The diagram of Figure 1 illustrates the block coding relations.

For all $j=1,2,\cdots$ the joint probability distribution of source outputs and channel outputs is defined by

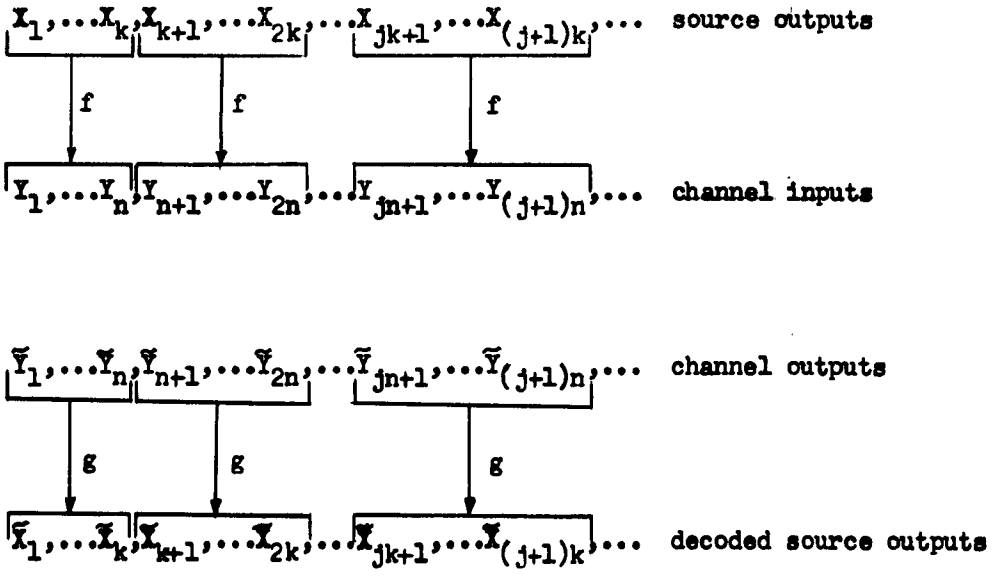

Figure 1. Block Coding Relations for a Block Code (f,g)

$$p(x_1,\cdots x_{jk};\tilde{y}_1,\cdots\tilde{y}_{jn}) = \frac{1}{|A|^{jk}} \prod_{i=1}^{jn} p(\tilde{y}_i/y_i) \quad \text{where} \tag{2}$$

$(y_{rn+1},\cdots y_{(r+1)n}) = f(x_{rk+1},\cdots x_{(r+1)k})$, $r=0,\cdots j-1$. Thus blocks of k-tuples from the source are encoded and decoded independently of other blocks so that $P\left[(\tilde{X}_{jk+1},\cdots\tilde{X}_{(j+1)k}) \neq (X_{jk+1},\cdots X_{(j+1)k})\right]$ is independent of j. $e(f,g) = P\left[(\tilde{X}_1,\cdots\tilde{X}_k) \neq (X_1,\cdots X_k)\right]$ is called the probability of error for the block code (f,g).

$R = \frac{k}{n} \log|A|$ is called the rate of the block code (f,g). As the average number of bits per channel input, it measures the density of source outputs per channel input. From the decoding viewpoint (the viewpoint which is useful for general codes used later) R is the average number of bits decoded per channel output.

The block coding channel capacity C of a DMC $(B,\tilde{B},p)$ has the following property

$$C = \sup\left\{R' : \inf_{(f,g)}\left\{e(f,g) : \frac{k}{n}\log|A| \leq R'\right\} = 0\right\} \tag{2}$$

which is just a result of the usual coding theorem for a DMC $(B,\tilde{B},p)$.[3]

From the observation that there are many conceivable ways other than block coding to send a source output $(x_1,x_2\cdots)$ over a channel, the question naturally arises as to whether (i) a code rate can be defined for a general class of codes (for which channel capacity is defined for some suitable error criterion) and (ii), if so, do block codes attain the channel capacity of the general codes? To begin answering (i) and (ii) a general class of encoders and decoders is first defined.

One of the properties that any practical encoder should have and which a block encoder does have is that the n^{th} channel input depends on only a finite number of source outputs. Similarly, a decoder should decide on the k^{th} source symbol after some finite number of channel outputs. Therefore the following definitions are made.

An <u>encoder</u> for $\{X_i\}$ and $(B,\tilde{B},p)$ is a sequence of functions $\{f_n: A^{K(n)} \to B\}$ where $\{K(n)\}$ is a sequence of positive integers and f_n determines the n^{th} channel input from the first K(n) source outputs. Since there are many different sequences $\{K(n)\}$ which could be used for the representation of the same encoder it is assumed that K(n) is, for each n, the smallest positive integer r such that the first n channel inputs depend on at most the first r source outputs. With this assumption, if the source subscripts of $\{X_i\}$ correspond to time in seconds, then K(n) is the earliest time in seconds that the n^{th} channel input could be sent where, of course, the n^{th} channel input is not sent before the $(n-1)^{th}$ even though it may be determined by earlier source outputs.

The block encoder $f: A^{k'} \to B^{n'}$ is obviously a special case of the encoder defined above. With the general encoder, however, the n^{th} channel input can depend on the whole source output $x_1, \cdots x_{K(n)}$ while the n^{th} channel input of the block encoder can depend on at most the source outputs $x_{jk'+1}, \cdots x_{(j+1)k'}$ where j satisfys $jn' < n \leq (j+1)n'$. In addition, the functions f_n which correspond to the block encoder $f: A^{k'} \to B^{n'}$ are periodic with period n'.

A decoder for $\{X_i\}$ and $(B,\tilde{B},p)$ is a sequence of functions $\{g_k : \tilde{B}^{N(k)} \to A\}$ where $\{N(k)\}$ is a sequence of positive integers and g_k determines the k^{th} decoded source output from the first $N(k)$ channel outputs. As with the sequence $\{K(n)\}$ for the encoder, it is assumed that $N(k)$ is, for each k, the smallest integer r such that the first k decoded outputs depend on at most the first r channel outputs. Remarks comparing a block decoder with the decoder defined here are similar to those made for the encoders. A diagram illustrating the general encoder and decoder is shown in Figure 2.

Throughout, an arbitrary but fixed source $\{X_i\}$ and DMC $(B,\tilde{B},p)$ will be understood if not explicitly stated. For the sequences $\{K(n)\}$ and $\{N(k)\}$ understood, an encoder and decoder for $\{X_i\}$ and $(B,\tilde{B},p)$ will be denoted by $\{f_n\}$ and $\{g_k\}$ respectively. The pair $(\{f_n\},\{g_k\})$ will be called a code.

For an encoder $\{f_n\}$ the following probability distribution between source outputs and channel outputs will always be assumed. For each positive integer t, for each $(x_1,\cdots x_{K(t)})$ and $(\tilde{y}_1,\cdots \tilde{y}_t)$

$$p(x_1,\cdots x_{K(t)};\tilde{y}_1,\cdots \tilde{y}_t) = \frac{1}{|A|^{K(t)}} \prod_{i=1}^{t} p(\tilde{y}_i / f_i(x_1 \cdots x_{K(i)})). \qquad (2.5$$

Because the probability distributions are consistent in t, that is,

$$\sum_{x_{K(t)+1},\cdots x_{K(t+1)},\tilde{y}_{t+1}} p(x_1,\cdots x_{K(t+1)};\tilde{y}_1,\cdots \tilde{y}_{t+1}) = p(x_1,\cdots x_{K(t)};\tilde{y}_1,\cdots \tilde{y}_t)$$

for $t=1,2,\cdots$, the marginal probability distributions $p(x_1,\cdots x_i;\tilde{y}_1,\cdots \tilde{y}_j)$ are determined for all i,j (assume $K(t) \xrightarrow{t} \infty$).

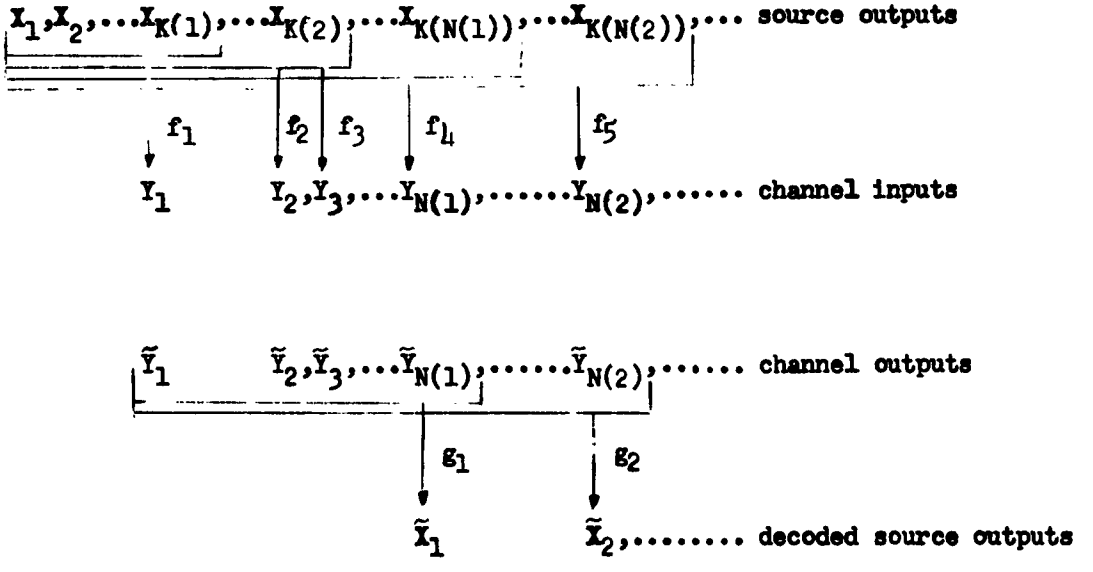

Figure 2. General Encoding and Decoding Relations for a Code $(\{f_n\}, \{g_k\})$ with $K(2) = K(3)$.

For a code $(\{f_n\},\{g_k\})$ the channel input random variables are defined by

$$Y_n = f_n(X_1,\dots X_{K(n)}) \quad n=1,2,\dots \tag{2.6}$$

and the decoded source random variables by

$$\tilde{X}_k = g_k(\tilde{Y}_1,\dots \tilde{Y}_{N(k)}). \tag{2.7}$$

For each k, $\frac{k}{N(k)}\log|A|$ is the average number of bits decoded per channel output from the first N(k) channel outputs. The <u>rate</u> R_d of the decoder $\left\{g_k : \tilde{B}^{N(k)} \to A\right\}$ is defined by

$$R_d = \lim_k \inf \frac{k}{N(k)}\log|A| \tag{2.8}$$

which is just the average number of bits decoded per channel output when the above limit exists. For a block decoder $g:\tilde{B}^{n'} \to A^{k'}$, $N(k) = N(i) + jk'$ for $k = i + jk'$, $0 < i \le k'$, $j = 0,1,2,\dots$ so that R_d $\lim_{k\to\infty} \frac{k}{N(k)}\log|A| = \frac{k'}{n'}\log|A|$, the usual rate for a block code (f,g)

With a block code (f,g) the error criterion which was used was the probability of error for a block: $P\left[\left(\tilde{X}_{jk'+1},\dots \tilde{X}_{(j+1)k'}\right) \neq \left(X_{jk'+1},\dots X_{(j+1)k'}\right)\right] = e(f,g)$. For a code $(\{f_n\},\{g_k\})$ the error criterion used will be the average coordinate probability of error $\bar{e}_t = \frac{1}{t}\sum_{i=1}^{t} P\left[\tilde{X}_i \neq X_i\right]$. In particular, the interest here will be in what happens to $\bar{e}_t$ as $t\to\infty$. Note that for a block code (f,g), $\bar{e}_{k'} \le e(f,g)$ and thus $\bar{e}_{k'}$ is a weaker error criterion for a block code than e(f,g); however, Theorem 1 is a converse and, consequently, a converse in terms of $\lim_t \sup \bar{e}_t$ is a stronger statement than a converse in terms of an average error for blocks of length r:

$$\bar{e}_{t,r} = \frac{1}{t}\sum_{i=1}^{t-1} P\left[\left(\tilde{X}_{ir+1},\dots \tilde{X}_{(i+1)r}\right) \neq \left(X_{ir+1},\dots X_{(i+1)r}\right)\right]$$

that is, $\lim_t \sup \bar{e}_t \leq \lim_t \sup \bar{e}_{t,r}$, $r=1,2,\ldots$. Channel capacity C' for the class of codes $(\{f_n\},\{g_k\})$ for a DMC $(B,\tilde{B},p)$ is thus defined by

$$C' = \sup\left\{R : \inf\left\{\lim_t \sup \bar{e}_t\right\} = 0 \text{ with } (\{f_n\},\{g_k\}) \text{ with } R_d \geq R\right\}. \tag{2.5}$$

It is easy to see that for a code $(\{f_n\},\{g_k\})$ which corresponds to a block code (f,g), $\lim_t \sup \bar{e}_t \leq e(f,g)$ and, since R_d is equal to the block code rate, it follows from (2.4) that $C' \geq C$.

The main result of this section follows from Theorem 1: $C' \leq C \Rightarrow C' = C$.

Theorem 1. For a DMC $(B,\tilde{B},p)$ with block coding capacity C, let $\left(\left\{f_n : A^{K(n)} \to B\right\},\left\{g_k : \tilde{B}^{N(k)} \to A\right\}\right)$ be any code with decoding rate R_d. If $R_d > C$ then there exists a positive number $\alpha(C/R_d)$ which depends only on C/R_d and not on $\{f_n\},\{g_k\}$ or $(B,\tilde{B},p)$ such that $\lim_t \inf \bar{e}_t \geq \alpha$.

Lemma 1. For all t, $1 - \frac{N(t)C}{t \log|A|} \leq h(\bar{e}_t) + \bar{e}_t$.

Proof of Lemma. From (1.2)

$$J\left((X_1,\ldots X_t),(\tilde{Y}_1,\ldots \tilde{Y}_{N(t)})\right) \leq N(t)C \Rightarrow$$

$$\frac{1}{t} I(X_1,\ldots X_t) - \frac{N(t)C}{t} \leq \frac{1}{t} \sum_{i=1}^{t} I\left(X_i/\tilde{Y}_1,\ldots \tilde{Y}_{N(i)}\right)$$

From $I(X_1,\ldots X_t) = t\log|A|$ and Fano's inequality (1.3)

$$1 - \frac{N(t)C}{t\log|A|} \leq \frac{1}{\log|A|}\left[\frac{1}{t}\sum_{i=1}^{t} h\left(P\left[\tilde{Y}_i \neq X_i\right]\right) + \bar{e}_t \log|A|\right] \leq$$

$h(\bar{e}_t) + \bar{e}_t$ since $h(x)$ is convex and $\log|A| \geq 1$. This completes the proof of the lemma.

Proof of Theorem 1. First, by Lemma 1

$$\lim_t \inf\left(1 - \frac{N(t)C}{t\log|A|}\right) = 1 - \lim_t \sup \frac{N(t)C}{t\log|A|} = 1 - C/R_d \leq$$

$\lim_{t'} h(\bar{e}_{t'}) + \bar{e}_{t'}$, where t' is any subsequence of $\{t\}$ such that $h(\bar{e}_{t'}) + \bar{e}_{t'}$ converges. In particular, for $R_d > C$

$$0 < 1 - C/R_d \leq h(\lim_t \inf \bar{e}_t) + \lim_t \inf \bar{e}_t.$$

Now, let α be the unique real number, $0 < \alpha < 1/2$, such that

$$1 - C/R_d = h(\alpha) + \alpha. \tag{2.10}$$

Clearly, $\alpha = \alpha(C/R_d) > 0$ and $\lim_t \inf \bar{e}_t \geq \alpha$ so that Theorem 1 is proved.

If R_d had been defined as $\lim_k \sup \frac{k \log|A|}{N(k)}$ then Theorem 1 would be true with $\lim_t \sup \bar{e}_t \geq \alpha$ replacing $\lim_t \inf \bar{e}_t \geq \alpha$. The proof is straightforward following the proof above.

There is an interesting generalization of the decoders $\{g_k\}$ used here suggested by the work of Blackwell[5]. Suppose that the decoder can change its past decisions, that is, suppose $X_1, \ldots X_k$ can be changed after $N(t)$ channel outputs where $t > k$. To make sense from a communication standpoint the first k decoded source outputs should be changed only a finite number of times with probability 1 for each k. This point, however, will not be needed for the converse (Theorem 2) below.

Let $\left\{g_k^*: \tilde{B}^{N(k)} \to A^k\right\}$ be any sequence of functions where the same assumptions are made for $\{N(k)\}$ as before. (The superscript * will be used to denote a decoder $\{g_k^*\}$ which can change its previous decisions.) Let

$$(\tilde{X}_{1k},\ldots\tilde{X}_{kk}) = g_k^*(\tilde{Y}_1,\ldots\tilde{Y}_{N(k)}),\ k=1,2,\ldots \qquad (2.$$

where $\tilde{X}_{ik}$ $(i\leq k)$ represents the decision for the i^{th} source output after N(k) channel outputs. For the decoders previously used, $\tilde{X}_{ik} = \tilde{X}_{ik'} = \tilde{X}_i$ for all $k,k' \geq i$, that is, no changes in the decoded outputs were made. A code $(\{f_n\},\{g_k^*\})$ is illustrated in Figure 3.

As before, the rate of a decoder $\{g_k^*\}$ will be defined by

$$R_d = \lim_k \inf \frac{k}{N(k)} \log|A|. \qquad (2.$$

Theorem 2. For a DMC $(B,\tilde{B},p)$ with block coding capacity C and any code $(\{f_n\},\{g_k^*\})$ with $R_d > C$, $\lim_t \inf \bar{e}_t^* \geq \alpha > 0$ where $\bar{e}_t^* = \frac{1}{t}\sum_{i=1}^{t} P[\tilde{X}_{it} \neq X_i]$ and α is defined by (2.10)

Lemma 2. For all t, $1 - \frac{N(t)C}{t\log|A|} \leq h(\bar{e}_t^*) + \bar{e}_t^*$.

Proof of Lemma. As in the proof of Lemma 1,

$$\frac{1}{t}\left[I(X_1,\ldots X_t) - I(X_1,\ldots X_t/\tilde{Y}_1,\ldots\tilde{Y}_{N(t)})\right] \leq \frac{N(t)C}{t} \implies$$

$$\frac{1}{t} I(X_1,\ldots X_t) - \frac{N(t)C}{t} \leq \frac{1}{t}\sum_{i=1}^{t} I(X_i/\tilde{Y}_1,\ldots\tilde{Y}_{N(t)}).$$

(In the proof of Lemma 1 the sum on the right hand side of the last inequality was $\frac{1}{t}\sum_{i=1}^{t} I(X_i/\tilde{Y}_1,\ldots\tilde{Y}_{N(i)})$.) From $I(X_1,\ldots X_t) = t\log|A|$ and Fano's inequality

$$1 - \frac{N(t)C}{t\log|A|} \leq \frac{1}{\log|A|}\left[\frac{1}{t}\sum_{i=1}^{t} h\left(P[\tilde{X}_{it} \neq X_i]\right) + \bar{e}_t^* \log|A|\right] \leq$$

$h(\bar{e}_t^*) + \bar{e}_t^*$ since h(x) is convex and $\log|A| \geq 1$. This proves the lemma.

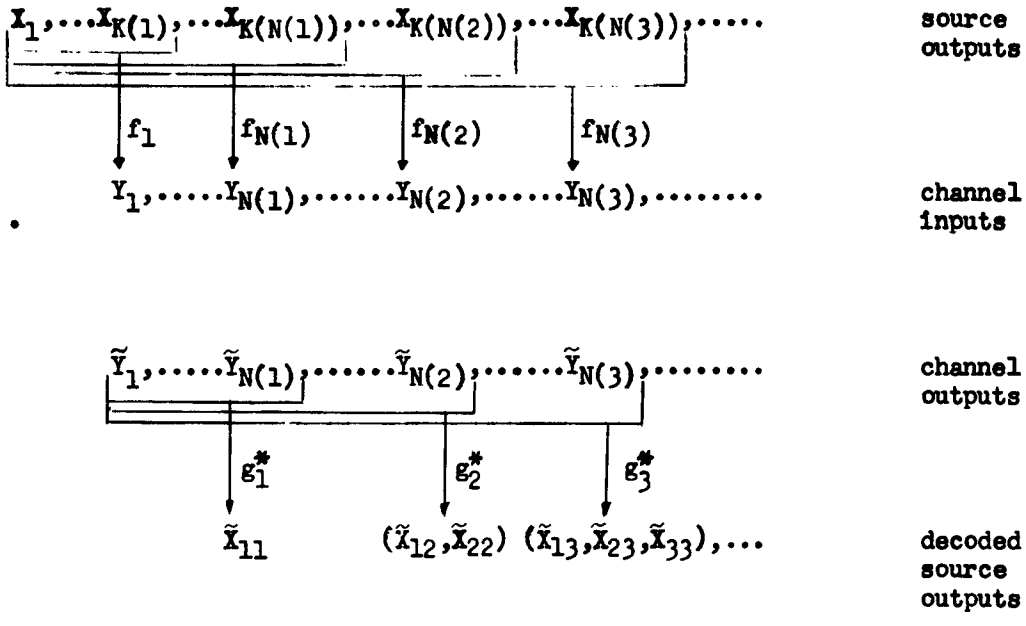

Figure 3. Encoding and Decoding Relations for a Code $(\{f_n\}, \{g_k^*\})$.

The proof of Theorem 2 proceeds from this point exactly as in Theorem 1.

It can be concluded from Theorem 2 that the channel capacity C* for the class of codes $(\{f_n\},\{g_k^*\})$, where

$$C^* = \sup\left\{R: \begin{array}{l} \inf\{\lim_t \sup \varepsilon_t^*\}=0 \\ (\{f_n\},\{g_k^*\}) \text{ with } R_d \geq R \end{array}\right\},$$

is equal to C.

In both Theorems 1 and 2 no use was made of the encoder $\{f_n: A^{K(n)} \to B\}$. In fact, Theorems 1 and 2 hold for any function $f: A^I \to B^I$ which is $(\sigma(A^I),\sigma(B^I))$- measurable. What is needed is the definition of the probability distributions $p(x_1,\ldots x_i;\tilde{y}_1,\ldots\tilde{y}_j)$, $i,j = 1,2,\ldots$ for the general f, since then $J((X_1,\ldots X_i),(\tilde{Y}_1,\ldots\tilde{Y}_j))$ is defined for all i,j and is $\leq jC$. From this point the proofs go through as before. The definition of $p(x_1,\ldots x_i;\tilde{y}_1,\ldots\tilde{y}_j)$ for $f: A^I \to B^I$, $(\sigma(A^I),\sigma(B^I))$- measurable is given in the appendix.

The <u>rate</u> R_e of the encoder $\{f_n\}$ is defined by

$$R_e = \lim_n \inf \frac{K(n)}{n} \log|A| .$$

If the source subscripts correspond to time in seconds then K(n) is the time in seconds the n^{th} channel input is transmitted (assuming zero delays in the encoding equipment) and $K(N(k))$ is the time in seconds the k^{th} source output is decoded (again, assuming zero delays in channel transmission and in the decoding equipment). One natural requirement for any code is that of bounding the time lag (positive or negative) between the time an output occurs and time it is decoded, that is, require $\sup_k |K(N(k)) - k| < \infty$.

Theorem 3. For a code $(\{f_n\}, \{g_k\})$ with $\sup_k |K(N(k)) - k| < \infty$, $R_e \leq R_d$.

Proof. $\sup_k |K(N(k)) - k| < \infty \Rightarrow \lim_{k\to\infty} \frac{K(N(k))}{k} = 1 \Rightarrow$

$$\lim_{k\to\infty} \frac{K(N(k))}{N(k)} \cdot \frac{N(k)}{k} = 1.$$

Pick a subsequence k' such that

$$\frac{k'}{N(k')} \xrightarrow{k'} \frac{R_d}{\log|A|}$$

so that $\lim_{k'\to\infty} \frac{K(N(k'))}{N(k')}$ exists and equals $\frac{R_d}{\log|A|} \Rightarrow$

$$\frac{R_d}{\log|A|} \geq \frac{R_e}{\log|A|} .$$

This completes the proof.

From the proof it is immediate that Theorem 3 remains true if $\sup_k |K(N(k)) - k| < \infty$ is replaced by the weaker condition $\lim_{k\to\infty} \frac{K(N(k))}{k} = 1$.

Since both Theorems 1 and 2 are weak converses the next question to ask would be whether $\bar{e}_t \xrightarrow{t} 1$ or $\bar{e}_t^* \xrightarrow{t} 1$ (the strong converse statements) for $R_d > C$. Since one could guess each source output correctly with a probability of at least $1/|A|$, regardless of the decoding rate R_d, it follows that $\bar{e}_t \leq 1 - 1/|A|$ for all t for at least one decoder $\{g_k\}$. Thus there is no strong converse statement in terms of $\bar{e}_t$†. Similar remarks are true for $\bar{e}_t^*$.

Strong converses can be obtained for a different probability of error as follows. For a code $(\{f_n\}, \{g_k\})$ one may consider the encoding of

† This fact was pointed out to the writer by Professor D. Blackwell.

$(X_1, \ldots X_k)$ to $(Y_1, \ldots Y_{N(k)})$ to be random, that is, if $K(N(k)) > k$ then what $(x_1, \ldots x_k)$ is mapped into (i.e., what sequence $(y_1, \ldots y_{N(k)})$) depends on $X_{k+1}, \ldots X_{K(N(k))}$. (If $K(N(k)) \leq k$ there is no question of random encoding since $X_1, \ldots X_k$ determines $Y_1, \ldots Y_{N(k)}$.) Since the strong converse for block codes continues to hold when the encoder is random the following result is obtained.

Theorem 4. For a DMC $(B, \tilde{B}, p)$ with block coding capacity C let $(\{f_n\}, \{g_k\})$ be a code with $R_d > C$. Then

$$e_t = P\left[\bigcup_{i=1}^{t}\left[g_i(\tilde{Y}_1, \ldots \tilde{Y}_{N(i)}) \neq X_i\right]\right] \xrightarrow{t} 1$$

The same argument applies for the redecoding case.

Theorem 5. For a DMC $(B, \tilde{B}, p)$ let $(\{f_n\}, \{g_k^*\})$ be any code with $R_d > C$. Then

$$e_t^* = P\left[g_t^*\left(\tilde{Y}_1, \ldots \tilde{Y}_{N(t)}\right) \neq (X_1, \ldots X_t)\right] \xrightarrow{t} 1.$$

The codes $(\{f_n\}, \{g_k\})$ of Theorem 1 can be changed for use with a doubly infinite source $\{X_i, -\infty < i < \infty\}$. An encoder becomes a sequence of functions $\left\{f_n : \prod_{i=-\infty}^{K(n)} A_i \to B\right\}$ where it is required that each $f_n (-\infty < n < \infty)$ depend on at most a finite number of coordinates of $\prod_{i=-\infty}^{K(n)} A_i$. The sequence $\{K(n)\}$ will be a sequence of integers such that $K(n)$ is, for each n, the smallest integer r such that all channel inputs up to and including the n^{th} do not depend on the source outputs $(X_{r+1}, X_{r+2}, \ldots)$. A decoder is a sequence of functions $\left\{g_k : \prod_{i=D(k)}^{N(k)} \tilde{B}_i \to A\right\}$ where it is assumed that

(i) $N(k)$ is, for each k, the smallest integer r such that all decoded outputs $\tilde{X}_i = g_i(\tilde{Y}_{D(i)}, \ldots \tilde{Y}_{N(i)})$ up to and including the k^{th}

do not depend on channel outputs $(\tilde{Y}_{r+1}, \tilde{Y}_{r+2}, \cdots)$, and

(ii) $D(k) \le N(k)$, $D(k) \le D(k+1)$ for all k.

$R_d = \lim_k \inf \frac{k}{N(k)} \log|A|$ will be called the rate of the decoder $\left\{g_k : \prod_{i=D(k)}^{N(k)} \tilde{B}_i \to A\right\}$. The average number of bits decoded per channel output between the t^{th} and k^{th} decoded outputs $(t<k)$ is

$$\frac{(k-t)\log|A|}{N(k) - N(t)} \qquad \left(\text{assume } N(k) > N(t)\right)$$

and

$$\lim_k \inf \frac{(k-t)\log|A|}{N(k) - N(t)} = R_d \text{ for all } t: -\infty < t < \infty,$$

that is, the same rate is obtained independently of which decoded source output is used as a starting point.

The analogous statement of Theorem 1 for the doubly-infinite source is

Theorem 6. For a DMC $(B, \tilde{B}, p)$ with block coding capacity C let $\left(\{f_n\}, \{g_k\}\right)$ be any code for a doubly-infinite source with $R_d > C$. Then

$$\lim_t \inf \frac{1}{t} \sum_{i=r}^{r+t-1} P\left[\tilde{X}_i \neq X_i\right] \ge \alpha(C/R_d) > 0$$

for all $r: -\infty < r < \infty$ where α is defined by (2.10).

Lemma 3. For each $r: -\infty < r < \infty$ and for all $t: 1 < t < \infty$

$$1 - \frac{N(t+r-1) - D(r)}{t \log|A|} \le h\left(\frac{1}{t} \sum_{i=r}^{r+t-1} P\left[\tilde{X}_i \neq X_i\right]\right) + \frac{1}{t} \sum_{i=r}^{r+t-1} P\left[\tilde{X}_i \neq X_i\right]$$

The proofs of Lemma 3 and Theorem 6 offer no difficulty following the proofs of Lemma 1 and Theorem 1, and thus are omitted.

III. UNIFORM FINITE-MEMORY CODES

For this section $\{X_i\}$ will denote a doubly-infinite source $\{X_i, -\infty<i<\infty\}$ and all codes discussed will be for such sources. To begin, suppose the $(\{f_n\},\{g_k\})$ of section II takes almost the simplest form possible:

$$f_n = f^{(m)}: A^m \to B \text{ for all } n: -\infty<n<\infty ,$$
$$g_k = g^{(m)}: \tilde{B}^m \to A \text{ for all } k: -\infty<k<\infty , \tag{3.}$$
$$\text{with } \{K(n)\} = \{n\} \text{ and } \{N(k)\} = \{k+m-1\}.$$

The code is illustrated in Figure 4. $\tilde{Y}_1,\ldots\tilde{Y}_m$ are used to decode X_1 since in the encoder $Y_1,\ldots Y_m$ are the only channel inputs which depend on X_1 and, intuitively then, $\tilde{Y}_1,\ldots\tilde{Y}_m$ should be the most important channel outputs for decoding X_1. Of course, using other channel outputs will improve the probability of deciding X_1 correctly but, for the present, only $\tilde{Y}_1,\ldots\tilde{Y}_m$ will used. Because of the stationarity involved with the above code it is clear that $P[\tilde{X}_i \neq X_i]$ is the same for all i so that it suffices to restrict attention to X_1 when discussing the probability of error for X_i. The following questions arise. When can $P[\tilde{X}_1 \neq X_1] \xrightarrow{m} 0$, that is, when does there exist a sequence $\{(f^{(m)},g^{(m)})\}$ of codes of the above form such that $P[\tilde{X}_1 \neq X_1] \xrightarrow{m} 0$? What if the n-extension $(B^n,\tilde{B}^n,q)$ is used instead of $(B,\tilde{B},p)$, that is, when does there exist, for some fixed n, a sequence of codes $\{(f^{(m)},g^{(m)})\}$ (where for $m=1,2,\ldots$

$$f^{(m)}: A^m \to B^n,$$
$$g^{(m)}: \tilde{B}^{nm} \to A)$$

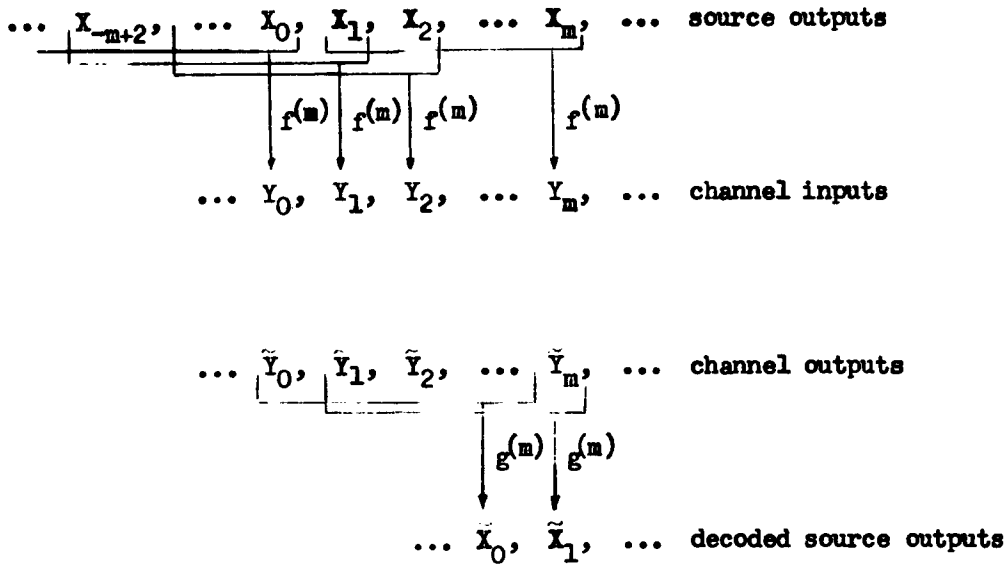

Figure 4. Coding Relations for the Code $(f^{(m)}, g^{(m)})$ of (3.1).

such that $P[X_1 \neq g^{(m)}(\tilde{Y}_1, \ldots \tilde{Y}_{rm})] = P[\tilde{X}_1 \neq X_1] \xrightarrow{m} 0$? Before giving partial answers to these questions certain properties of the above codes will be singled out.

An encoder $\left\{f_n : \prod_{i=-\infty}^{K(n)} A_1 \to B, -\infty < n < \infty\right\}$ will be called a <u>finite-memory encoder</u> if there exists a positive integer m' such that the f_n, for each n, depend on at most the last m' coordinates of $(\ldots X_1, \ldots X_{K(n)})$. The smallest such m' will be denoted by m and called the <u>duration of memory</u> of $\{f_n\}$. The corresponding definition for decoders is obvious.

An encoder $\{f_n\}$ will be said to satisfy a <u>uniform timing constraint</u> (k_e, n_e, s_e) if $K(n) = \left\lceil \frac{n^+}{n_e} \right\rceil k_e + s_e$ for all n. Here, k_e, n_e and s_e are integers with k_e and n_e positive, and $\lceil a^+ \rceil$ denotes the smallest integer $\geq a$. The corresponding uniform timing constraint for decoders will be denoted by (k_d, n_d, s_d) where $N(k) = \left\lceil \frac{k^+}{k_d} \right\rceil n_d + s_d$ for all k.

The natural requirement for an encoder $\{f_n\}$ with memory duration m and uniform timing constraint (k_e, n_e, s_e) is that the functions f_n themselves satisfy a uniformity requirement. An encoder $\{f_n\}$ with memory duration m and uniform timing constraint (k_e, n_e, s_e) will be called a <u>uniform finite-memory encoder</u> (UFME) if there exists a function $f : A^m \to B^{n_e}$ such that the f_n are periodic with period n_e and

$$f_{j+s_e} = f(\cdot)(j) \text{ for } j = 1, \ldots n_e, \tag{3.2}$$

where $f(\cdot)(j)$ denotes the j^{th} component of f. As before, a similar definition is obvious for uniform finite-memory decoders (UFMD).

For everything that remains it will be assumed that $k_e = k_d = k$, $n_e = n_d = n$, and $s_e = C$. (The positive integers k and n used here have no relation to the dummy variables in $N(k)$ and $K(n)$.) In addition,

the memory duration of a UFME and a UFMD will always be m'k and m'n respectively for some positive integer m'. At the risk of confusion, the superscript ' will be dropped. For a given m, s_d will be taken as (m-1)n. Because k and n will always be understood a UFME will be denoted by $f^{(m)}$ and a UFMD by $g^{(m)}$ where

$$f^{(m)}:A^{km}\rightarrow B^{n} \text{ and } g^{(m)}:B^{nm}\rightarrow A^{k}. \tag{3.3}$$

The code (3.3) corresponds to the code (3.1) for k=1, n=1.

Notation for a uniform finite-memory code (UFMC) $(f^{(m)},g^{(m)})$ can be greatly simplified to correspond to that of code (3.1). Let, for all i,

$$U_i=(X_{(i-1)k+1},\dots X_{ik}),\ \tilde{U}_i=(\tilde{X}_{(i-1)k+1},\dots \tilde{X}_{ik})$$
$$V_i=(Y_{(i-1)n+1},\dots Y_{in}),\ \tilde{V}_i=(\tilde{Y}_{(i-1)n+1},\dots \tilde{Y}_{in}).$$

Then

$$V_i=f^{(m)}(U_{i-m+1},\dots U_i),\ \tilde{U}_i=g^{(m)}(\tilde{V}_i,\dots \tilde{V}_{i+m-1}) \tag{3.4}$$

for $-\infty<i<\infty$. The diagram of Figure 5 illustrates the relations for a UFMC $(f^{(m)},g^{(m)})$. Note that the encoding and decoding rates for the above UFMC are both equal to $\frac{k}{n}\log|A|$, and, as before with code (3.1), $P[\tilde{U}_i\neq U_i]$ is the same for all i.

Particular cases of uniform finite-memory encoders are quite prominent in coding literature. For the case of a binary source and channel Elias[6] calls $f^{(m)}$ a convolutional encoder if $f^{(m)}$ is a linear function (in the binary field sense) from B^{km} to B^{n}, $B=\{0,1\}$. For other sources and channels $f^{(m)}$ is called a sequential encoder by Reiffen[7] when it is linear in the sense of some finite field appropriate for both source and channel. Decoding for these encoders, however, is quite different from that done here. For example, the decision for U_1 ($\tilde{U}_1$) is always made from $\tilde{V}_1,\dots\tilde{V}_m$

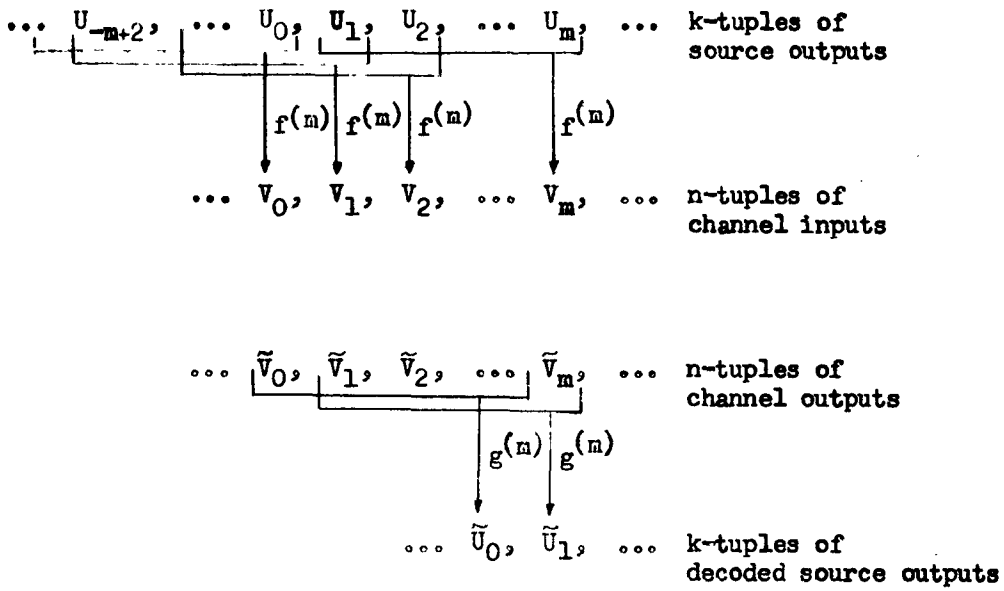

Figure 5. Coding Relations for the UFMC $\left(f^{(m)}, g^{(m)}\right)$.

assuming outputs $U_{-m+2},\dots U_0$ are known or that previously decoded outputs $\tilde{U}_{-m+2},\dots\tilde{U}_0$ are correct. $\tilde{U}_2$ is then determined from $\tilde{V}_2,\dots\tilde{V}_{m+1}$ and $\tilde{U}_{-m+3},\dots\tilde{U}_1$ assuming $\tilde{U}_1$ was correct and so on for $\tilde{U}_3,\dots$. The whole coding procedure is made into a block code by periodically putting in m-1 known source symbols at appropriate intervals.

To return to the questions at the beginning of this section one may ask what is

$$C_u = \sup_{k,n,|A|}\left\{\frac{k}{n}\log|A|: \text{ there exists a sequence } \{(f^{(m)},g^{(m)})\} \text{ such that } P[\tilde{U}_1\neq U_1]\xrightarrow{m}0.\right\} \tag{3.5}$$

C_u is called the uniform finite-memory coding capacity of $(B,\tilde{B},p)$. Because it is required that the error go to 0 with $m\to\infty$, k and n fixed, C_u is <u>not</u> just the expression for C' restricted to the class of uniform finite-memory codes.

Let $P[\tilde{U}_1\neq U_1/f^{(m)},g^{(m)}]$ denote $P[\tilde{U}_1\neq U_1]$ when a UFME $f^{(m)}$ and a UFMD $g^{(m)}$ are used. Because

$$\frac{1}{k}\sum_{i=1}^{k}P[X_1\neq X_i]\leq P[\tilde{U}_1\neq U_1]=P[\tilde{U}_i\neq U_i]$$

for all i with a UFMC $(f^{(m)},g^{(m)})$ it follows from Theorem 6 that for any m and $(f^{(m)},g^{(m)})$ with $R=\frac{k}{n}\log|A|\geq C$

$$P[\tilde{U}_1\neq U_1/f^{(m)},g^{(m)}]\geq\alpha>0 \tag{3.6}$$

where $\alpha=\alpha(C/R)$ is defined by (2.10). $\therefore C_u\leq C$.

In the appendix a stronger form of the above statement is proven which allows the decoder, for a UFME $f^{(m)}$, to use all channel outputs $(\dots\tilde{V}_{-1},\tilde{V}_0,\tilde{V}_1,\dots)$.

Unfortunately, what C_u is for a DMC $(B,\tilde{E},p)$ is an open question. The next few paragraphs are devoted to exhibiting a class of channels for which $C_u \geq 0$.

Denote the minimum probability of error over all ways of deciding U_1 from $(U_a,\ldots U_b;\tilde{V}_c,\ldots\tilde{V}_d)$ $(a\leq b, c\leq d)$ when a UFMC $f^{(m)}$ is used by $e(U_1/f^{(m)};U_a,\ldots U_b;\tilde{V}_c,\ldots\tilde{V}_d)$. Also, let $e(U_1/f^{(m)};\tilde{V}_c,\ldots\tilde{V}_d)$ be the minimum probability of error over all ways of deciding U_1 from $(\tilde{V}_c,\ldots\tilde{V}_d)$. For example, for $c=1$ and $d=m$, $e(u_1/f^{(m)};\tilde{V}_1,\ldots\tilde{V}_m)=\min_{g^{(m)}} P\left[U_1\neq U_1/f^{(m)},g^{(m)}\right]$.

Reiffen[7,8] has shown that for $R=\frac{k}{n}\log|A|<C$, there exists a sequence of UFMC's $\{f_o^{(m)}:m=1,2,\ldots\}$ such that for all m

$$e(U_1/f_o^{(m)};U_{-m+2},\ldots U_0;\tilde{V}_1,\ldots\tilde{V}_m)\leq \mathcal{P}(m)2^{-mnE(R)} \qquad (3.7)$$

where $\mathcal{P}(m)$ is a polynomial in m (whose coefficients depend on n) and $E(R)>0$ for all $R<C$. Before making use of (3.7) three lemmas are given below, the first two of which are routine.

Lemma 4. For a UFMC $f^{(m)}$ and positive integer r $(m\geq 2)$

$$e(U_1/f^{(m)};U_{-m-r+3},\ldots U_{-r+1};\tilde{V}_{-r+2},\ldots\tilde{V}_m)\leq$$
$$r\,e(U_1/f^{(m)};U_{-m+2},\ldots U_0;\tilde{V}_1,\ldots\tilde{V}_m).$$

Lemma 5. For any finite-valued random variables X and Y let

$$e(X/Y)=\min_{\{g:B\to A\}} P\left[g(Y)\neq X\right].$$

Then for any sequence $\{Y_i\}$ of finite-valued random variables with the same finite range B

$$e(X/(Y_1,\ldots Y_n))\downarrow e(X/(Y_1,Y_2,\ldots)),$$

where $e(X/Y_1,Y_2,\dots)) = \inf_{\{g:g^{-1}(x)\in\sigma(B^I) \text{ for all } x\}} P[g(Y_1,Y_2,\dots)\neq X]$

Let $e(U_1/f^{(m)};\dots\tilde{V}_1,\dots\tilde{V}_m) =$

$\inf\{P[g(\dots\tilde{V}_1,\dots\tilde{V}_m)\neq U_1] \text{ when } f^{(m)} \text{ is the encoder}\}$.

$\{g:g^{-1}(u)\in\sigma\left(\prod_{i=-\infty}^{mn}\tilde{B}_i\right) \text{ for all } u\}$

Lemma 6. Suppose there exists a sequence of UFME's $\{f_0^{(m)}\}$ such that $e(U_1/f_0^{(m)};\dots\tilde{V}_1,\dots\tilde{V}_m)\xrightarrow{m}0$. Then there exists a sequence $\{f^{(m)}\}$ of UFME's such that $e(U_1/f^{(m)};\tilde{V}_1,\dots\tilde{V}_m)\xrightarrow{m}0$.

Proof. Let $a_{m,r} = \min_{\{\text{all UFME's } f'^{(m)} \text{ of memory } m\}} e(U_1/f'^{(m)};\tilde{V}_{m-r+1},\dots\tilde{V}_m)$, $m,r = 1,2,\dots$

Then (i) $a_{m,r}\geq a_{m,r+1}$ and $a_{m,r}\overset{r}{\downarrow} a_m \leq e(U_1/f_0^{(m)};\dots\tilde{V}_1,\dots\tilde{V}_m)$

$\Rightarrow a_m\xrightarrow{m}0$. These statements follow from Lemma 2 and the definition of $a_{m,r}$.

(ii) $a_{m,r}\geq a_{m+1,r}$.

To prove (ii) notice that for $f_r^{(m)}$, the UFME which yields $a_{m,r} = e(U_1/f_r^{(m)};\tilde{V}_{m-r+1},\dots\tilde{V}_m)$, the UFME $h^{(m+1)}$ defined by

$$V_j = h^{(m+1)}(U_{j-m},\dots U_j) = f_r^{(m)}(U_{j-m},\dots U_{j-1})$$

merely advances the V's one coordinate so that the probability distribution of $(U_1;\tilde{V}_{m-r+1},\dots\tilde{V}_m)$ with $f_r^{(m)}$ is the same as $(U_1;\tilde{V}_{m-r+2},\dots\tilde{V}_{m+1})$ with $h^{(m+1)} \Rightarrow a_{m+1,r}\leq e(U_1/h^{(m+1)};\tilde{V}_{m-r+2},\dots\tilde{V}_{m+1}) = a_{m,r}$.

From (i) and (ii) it follows that $a_{m,m}\xrightarrow{m}0 \Rightarrow$ the sequence $\{f_m^{(m)}\}$ of UFME's has $e(U_1/f_m^{(m)};\tilde{V}_1,\dots\tilde{V}_m)\xrightarrow{m}0$. This completes the proof.

From Lemma 6 whenever there exists a sequence $\{f_0^{(m)}\}$ such that $e(U_1/f_0^{(m)};\dots\tilde{V}_1,\dots\tilde{V}_m)\xrightarrow{m}0$ then there exists a sequence $\{(f^{(m)},g^{(m)})\}$ of UFME's such that $P[\tilde{U}_1\neq U_1/f^{(m)},g^{(m)}]\xrightarrow{m}0$. A class of channels will now be

given for which $e(U_1/f_o^{(m)},\ldots\tilde{V}_1,\ldots\tilde{V}_m)\xrightarrow{m}0$ for some sequence $\left\{f_o^{(m)}\right\}$.

The following definitions will be needed. The product of the DMC's $(B,\tilde{B},p)$ and $(B',\tilde{B}',p')$ is the DMC $(B\times B',\tilde{B}\times\tilde{B}',q)$ where $q(\tilde{y},\tilde{y}'/y,y')=p(\tilde{y}/y)p'(\tilde{y}'/y')$ for all y, $\tilde{y}$, y', $\tilde{y}'$. A binary symmetric channel (BSC) has $B=\tilde{B}=\{0,1\}$ with

$$p(0/0)=p(1/1)=q,$$

$$p(0/1)=p(1/0)=p,$$

and $p+q=1$.

BSC (p_o,q_o) denotes a BSC with $q=q_o, p=p_o$. The k-tuple erasure channel accepts k-tuples of 0's and 1's and erases the whole k-tuple with probability p', or lets it through correctly with probability q'.

Consider a channel which is the product of the n-extension of the BSC (p_o,q_o) and the k-tuple erasure channel with $q'=q^k$. Suppose the encoding of a $\{0,1\}$ source to the BSC (p_o,q_o) is done by a UFME $f_o^{(m)}$ (an element of a sequence satisfying (3.7)) and the encoding of the source to the erasure channel is done by using the k-tuple of source directly as an erasure channel input. A diagram is given below (Figure 6).

If the product channel inputs are denoted by $V_i=(U_i,V_i)$ it is clear that the encoder given by

$$V_i'=\left(U_i,f_o^{(m)}(U_{i-m+1},\ldots U_i)\right)$$

is also a UFME $f_o'^{(m)}$ of memory duration m for the product channel.

Decoding U_1 from the sequence $(\ldots\tilde{V}_1',\ldots\tilde{V}_m')$ is accomplished as follows. Starting at the zero coordinate, m-1 consecutive unerased U's are sought. Suppose that $U_{-r+1},\ldots U_{-r+m-1}$ is the first run of m-1 U's not erased ($r\geq m-1$). By Lemma 3, $\tilde{U}_1$ can be decided using only

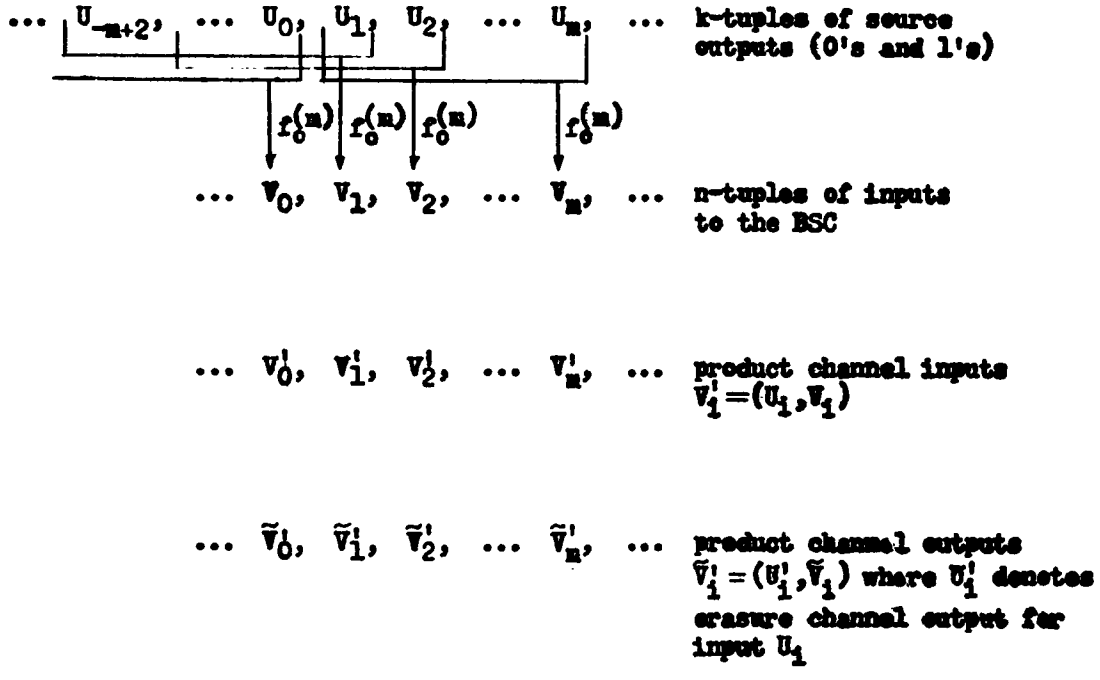

Figure 6. Encoding Relations for the Product of the n-extension of the BSC and the k-tuple Erasure Channel.

$(U_{-r+1},\dots U_{-r+m-1};\tilde{V}_{-r+m},\dots\tilde{V}_m)$ with a probability of error $\leq$ $(r-m+2)\mathcal{O}(m)2^{-mnE(R)}$. Let $p_{m,r}$ be the probability of finding m-1 consecutive U's not erased for the first time at coordinate -r+1, r=1,2,... . Then $e(U_1/f_o^{(m)};\dots\tilde{V}_1',\dots V_m') \leq \sum_{r=1}^{\infty}(r-m+2)p_{m,r}\mathcal{O}(m)2^{-mnE}$

$\leq \sum_{r=1}^{\infty} rp_{m,r}\mathcal{O}(m)2^{-mnE}$ (assume $m\geq 2$)

= (Feller[9], page 300 with Feller's p,q and r replaced by q^k, $1-q^k$, and m-1 respectively)

$$\frac{1-(q^k)^{m-1}}{(1-q^k)(q^k)^{m-1}}\mathcal{O}(m)2^{-mnE} \leq \mathcal{O}(m)\frac{q^k}{1-q^k}2^{-m\left[nE-k\log 1/q\right]}$$

$\xrightarrow{m} 0$ if $\left[nE-k\log 1/q\right]>0$.

Since only nonpositive coordinates of $\tilde{V}_i'$ were used to find m-1 consecutive erasures one can write

$$e(U_1/f_o^{(m)};\dots\tilde{V}_o';\tilde{V}_1,\dots\tilde{V}_m) \leq \mathcal{O}(m)\frac{q^k}{1-q^k}2^{-m\left[nE-k\log 1/q\right]}$$

Examining the coefficient $\left[nE(R)-k\log 1/q\right]$ of -m it is clear that if it is greater than zero for some k_o, n_o with $k_o/n_o=R$ then it is greater than zero for all k, n with $k/n = R$. For the case here, R is the rate for the BSC so that $R<1-h(p_o)$. (The rate for the product channel is just k.) By selecting a q sufficiently close to 1 depending on $R= k_o/n_o<1-h(p_o)$ there exists a sequence $\left\{f_o^{(m)}\right\}$ such that $e(U_1/f_o^{(m)};\dots\tilde{V}_1',\dots\tilde{V}_m')\xrightarrow{m}0$. Once q is fixed k can be selected such that $R=k/n$ is fixed and the block coding channel capacity kq^k of the erasure channel is arbitrarily small. Thus one has the situation of being able to make a product channel from an n-extension of the BSC (block coding channel capacity $n\left[1-h(p_o)\right]$) with an erasure channel of arbitrarily small block coding capacity and

still have a sequence $\{f^{(m)}\}$ with $R = k/n$ such that $e(U_1/f^{(m)}; \ldots \tilde{V}'_0, \tilde{V}'_1, \ldots \tilde{V}'_m) \xrightarrow{m} 0$.

The same demonstration can be carried out for a general DMC $(B, \tilde{B}, p)$ and source $\{X_i\}$ if the k-tuple erasure channel takes k-tuples $(x_1, \ldots x_k)$, $x_i \in A$ as inputs. The block coding channel capacity of the erasure channel is then $kq^k \log A$ where, as before q^k is the probability that any k-tuple goes through unerased.

Aside from calculating C_u for a given DMC it would be nice to know if $C_u > 0$ for a channel which does not have the erasure properties used above (for example, is $C_u > 0$ for the BSC). It would also be useful to have an example, if possible, for which $C_u < C$.

Another question arises from the fact that Lemma 6 does not provide any information about how fast $P\left[\tilde{U}_1 \neq U_1/f^{(m)}, g^{(m)}\right]$ tends to 0 with m. For example, when

$$e(U_1/f_0^{(m)}; \ldots \tilde{V}'_1, \ldots \tilde{V}'_m) \leq \mathcal{K}(m) \frac{q^k}{1-q^k} 2^{-m\left[nE - k \log 1/q\right]},$$

and hence tends to 0 exponentially with m, does $e(U_1/f_0^{(m)}; \tilde{V}'_1, \ldots \tilde{V}'_m)$ tend to 0 exponentially with m? More generally, one would like to know how much smaller $e(U_1/f^{(m)}; \tilde{V}_{-r'+2}, \ldots \tilde{V}_{m+r''-1})$ is than $e(U_1/f^{(m)}; \tilde{V}_1, \ldots \tilde{V}_m)$ for $r', r'' = 1, 2, \ldots$.

APPENDIX I

In this section the probability distributions $p(x_1,\dots x_i;\tilde{y}_1,\dots \tilde{y}_j)$, $i,j = 1,2,\dots$ are defined for a function $f:A^I \to B^I$ which is $(\sigma(A^I),\sigma(B^I))$-measurable. Terms used can be found in Loève.[10]

For a DMC $(B,\tilde{B},p)$ let $\bar{p}(\cdot/\bar{y})$ be the product probability on $\sigma(\tilde{B}^I)$ determined by the probabilities $p(\cdot/y_i)$ on $\tilde{B}_i$, $i = 1,2,\dots$, $\bar{y} = (y_1,y_2,\dots)$. Then $\bar{p}(\cdot/\cdot)$ is a $\sigma(B^I)$ - measurable probability, that is,

(i) $\bar{p}(\cdot/\bar{y})$ is a probability on $\sigma(\tilde{B}^I)$ for each $\bar{y} \in B^I$ and

(ii) for each set $\tilde{S} \in \sigma(\tilde{B}^I)$, $p(\tilde{S}/\cdot)$ is a $\sigma(B^I)$- measurable function.

For a source $\{X_i\}$ and a function $f:A^I \to B^I$ which is $(\sigma(A^I),\sigma(B^I))$-measurable, the function $\bar{p}(\cdot/f(\cdot))$ is a $\sigma(A^I)$- measurable probability. From $\bar{p}(\cdot/f(\cdot))$ a probability Q is defined on $\sigma(A^I)\times\sigma(B^I)$ by

$$Q(W) = \int_{A^I} \bar{p}\left(W(\bar{x})/f(\bar{x})\right) d\mu, \qquad W \in \sigma(A^I)\times\sigma(B^I)$$

where μ is the probability on $\sigma(A^I)$ corresponding to the source variables being independent, identically and uniformly distributed and $W(\bar{x})$ is the section of W at $\bar{x}$. All probabilities $p(x_1,\dots x_i;\tilde{y}_1,\dots \tilde{y}_j)$ are then just the marginal probabilities of Q.

It can be shown that the probabilities $p(x_1,\dots x_i;\tilde{y}_1,\dots \tilde{y}_j)$ thus defined are the same as those used in section II when $f:A^I \to B^I$ is given by the encoder $\left\{f_n : A^{K(n)} \to B\right\}$.

APPENDIX II

Let $f^{(n)}$ be any UFME for $\{X_i\}$ and $(B,\tilde{B},p)$ such that $R=(k/n)\log|A|<C$. Then for any function $g:\prod_{i=-\infty}^{\infty}\tilde{B}_i \to A^k$ such that $g^{-1}(u)\in \sigma\left(\prod_{i=-\infty}^{\infty}\tilde{B}_i\right)$* for all u

$P\left[g(\ldots \tilde{V}_{-1}, \tilde{V}_0, \tilde{V}_1, \ldots) \neq U_1\right] \geq \alpha(C/R)$.

Proof. For any positive integers r and t:

$$I(U_1, \ldots U_t) - I(U_1, \ldots U_t/\tilde{V}_{-r+1}, \ldots \tilde{V}_{r+t}) \leq (2r+t)nC$$

Since $I(U_1, \ldots U_t) = tk\log|A|$

it follows $I(U_1, \ldots U_t/\tilde{V}_{-r+1}, \ldots \tilde{V}_{r+t}) \geq tk\log|A| - (2r+t)nC$

Now, $I(U_1, \ldots U_t/\tilde{V}_{-r+1}, \ldots \tilde{V}_{r+t}) \leq \sum_{i=1}^{t} I(U_i/\tilde{V}_{-r+1}, \ldots \tilde{V}_{r+t})$

$\leq \sum_{i=1}^{t} I(U_i/\tilde{V}_{-r+i}, \ldots \tilde{V}_{r+i})$

$= t\, I(U_1/\tilde{V}_{-r+1}, \ldots \tilde{V}_{r+1})$

since the probabilities $p(u_i;\tilde{v}_{-r+i}, \ldots \tilde{v}_{r+i})$ are the same for all i.

$\therefore\quad I(U_1/\tilde{V}_{-r+1}, \ldots \tilde{V}_{r+1}) \geq k\log|A|\left(1 - \frac{2r+t}{t}\frac{C}{R}\right)$

for all positive integers t and r $\Longrightarrow$

$$I(U_1/\tilde{V}_{-r+1}, \ldots \tilde{V}_{r+1}) \geq k\log|A|\left(1 - \frac{C}{R}\right)$$

for all positive integers r.

Because $I(U_1/\tilde{V}_{-r+1}, \ldots \tilde{V}_{r+1}) \overset{r}{\downarrow} I(U_1/\ldots\tilde{V}_{-1}, \tilde{V}_0, \tilde{V}_1, \ldots)$ [11] it follows

$$I(U_1/\ldots\tilde{V}_{-1}, \tilde{V}_0, \tilde{V}_1, \ldots) \geq k\log|A|(1 - C/R)$$

Since $k\log|A| \geq 1$ it follows immediately from Fano's inequality (1.3) that

$$P\left[g(\ldots\tilde{V}_0, \tilde{V}_1, \ldots) \neq U_1\right] + h\left(P\left[g(\ldots \tilde{V}_0, \tilde{V}_1, \ldots) \neq U_1\right]\right) \geq 1 - C/R$$

so that $P\left[g(\ldots\tilde{V}_0, \tilde{V}_1, \ldots) \neq U_1\right] \geq \alpha(C/R)$ where α is defined by (2.10).

Fano's inequality (see (1.3)) was assumed (and can be proven)[11] for the case

---

* $\sigma\left(\prod_{i=-\infty}^{\infty}\tilde{B}_i\right)$ is the σ-field of subsets of $\prod_{i=-\infty}^{\infty}\tilde{B}_i$ determined by cylinder sets.

X is finite-valued, Y is an arbitrary random variable and $g:B \to A$ is measurable on the range of Y (that is, $g^{-1}(x) \in$ σ-field appropriate for the range of Y, $x \in A$). I(X/Y) is defined for this case as the expectation of the random variable almost surely defined by

$$i(X/Y) = -\log\Big(P\big([X=x]/Y\big)\Big) \text{ on the set } [X=x].$$

REFERENCES

1. Shannon, C. E., "A Note on a Partial Ordering for Communication Channels," Information and Control, vol. 1, number 4, pp. 390-397, 1958.

2. Fano, R. M., Transmission of Information, M.I.T. Press and John Wiley and Sons, Inc., New York, 1961, pp. 186-187.

3. Wolfowitz, J. Coding Theorems of Information Theory, Prentice-Hall, Inc., Englewood Cliffs with Springer-Verlag, Berlin, 1961, pp. 17-20.

4. Shannon, C. E., "A Mathematical Theory of Communication," Bell System Technical Journal, vol. 27, pp. 379-423, 623-656, 1948.

5. Blackwell, D., "Infinite Codes for Memoryless Channels," Annals of Math. Stat., vol. 30, number 4. pp. 1242-1244, 1959.

6. Elias, P., "Coding for Noisy Channels," I.R.E. Convention Record, Pt. IV, pp. 37-44, 1955.

7. Reiffen, B., "Sequential Encoding and Decoding for the Discrete Memoryless Channel," Technical Report 231, Lincoln Laboratory, M.I.T., 1960.

8. Wozencraft, J. M. and Reiffen, B., Sequential Decoding, M.I.T. Press and John Wiley and Sons, Inc., New York, 1961, p. 65.

9. Feller, W., An Introduction to Probability Theory and Its Applications, John Wiley and Sons, Inc., New York, 1957, p. 300.

10. Loève, M., Probability Theory, D. Van Nostrand Company, Inc., New York, 1960, pp. 90-91, 359.

11. Blackwell, D., "Information Theory," notes on a course given at the University of California at Berkeley, 1961.

DISTRIBUTION LIST
CONTRACT NO. Nonr-222(53)

ORGANIZATION	NO. COPIES
Assistant Secretary of Defense for Research and Engineering Information Office Library Branch Pentagon Building Washington 25, D. C.	2
Armed Services Technical Information Agency Arlington Hall Station Arlington 12, Virginia	10
Chief of Naval Research Department of the Navy Washington 25, D. C. ATTN Code 437, Information Systems Branch	2
Chief of Naval Operations OP-07T-12 Navy Department Washington 25, D. C.	1
Director, Naval Research Laboratory Technical Information Officer/Code 2000 Washington 25, D. C.	6
Commanding Officer, Office of Naval Research Navy #100, Fleet Post Office New York, New York	10
Commanding Officer, ONR Branch Office 346 Broadway New York 13, New York	1
Commanding Officer, ONR Branch Office 495 Summer Street Boston 10, Massachusetts	1
Bureau of Ships Department of the Navy Washington 25, D. C. ATTN Code 607A NTDS	1
Bureau of Naval Weapons Department of the Navy Washington 25, D. C. ATTN: RAAV Avionics Division	1
Bureau of Ships Department of the Navy Washington 25, D. C. ATTN: Communications Branch Code 686	1
Naval Ordnance Laboratory White Oaks Silver Spring 19, Maryland ATTN: Technical Library	1
David Taylor Model Basin Washington 7, D. C. ATTN: Technical Library	1
Naval Electronics Laboratory San Diego 52, California ATTN: Technical Library	1
University of Illinois Control Systems Laboratory Urbana, Illinois ATTN: D. Alpert	1
Air Force Cambridge Research Laboratories Laurence G. Hanscom Field Bedford, Massachusetts ATTN: Research Library, CRX2-R	1
Technical Information Officer U.S. Army Signal Research & Development Lab Fort Monmouth, New Jersey ATTN: Data Equipment Branch	1
National Security Agency Fort George G. Meade, Maryland ATTN: R-4, Howard Campaigne	1
U.S. Naval Weapons Laboratory Dahlgren, Virginia ATTN: Head, Compution Division, G.H. Gleissner	1
National Bureau of Standards Data Processing Systems Division Room 239, Building 10 ATTN: A.K. Smilow Washington 25, D.C.	1
Aberdeen Proving Ground, BRL Aberdeen Proving Ground, Maryland ATTN: J.H. Giese, Chief Compution Lab	1
Office of Naval Research Resident Representative University of California Room 201, Building T-9 Berkeley 4, California	1
Commanding Officer ONR Branch Office John Crerar Library Building 86 East Randolph Street Chicago 1, Illinois	1

ORGANIZATION	NO. COPIES
Cornell University Cognitive Systems Research Program Hollister Hall Ithaca, New York ATTN: Dr. Frank Rosenblatt	1
Communications Sciences Lab University of Michigan 180 Frieze Building Ann Arbor, Michigan ATTN: Gordon E. Peterson	1
Census Bureau Washington 25, D.C. ATTN: Office of Asst. Director for Statistical Services, Mr. J.L. McPherson	1
Stanford University Stanford, California ATTN Electronic Laboratory, Prof. Gene Franklin	1
University of California Institute of Engineering Research Berkeley 4, California ATTN Prof. A.J. Thomasian	1
National Science Foundation Program Director for Documentation Research Washington 25, D.C. ATTN: Helen L. Brownson	1
Wayne State University Detroit, Michigan ATTN. Dept. of Slavic Languages, Prof. Harry H. Josselson	1
University of California at Los Angeles Los Angeles 24, California ATTN Dept. of Engineering, Prof. Gerald Estrin	1
Columbia University New York 27, New York ATTN Dept. of Physics, Prof. L. Brillouin	1
Hebrew University Jerusalem, Israel ATTN: Prof. Y. Bar-Hillel	1
Naval Research Laboratory Washington 25, D.C. ATTN Security Systems Code 5266, Mr. G. Abraham	1
National Physical Laboratory Teddington, Middlesex England ATTN: Dr. A.M. Uttley, Superintendent, Autonomics Division	1
Dr. Jacob Beck Harvard University Memorial Hall Cambridge 38, Massachusetts	1
George Washington University Human Resources Research Office P. O. Box 3596 Washington 7, D.C. ATTN Dr. John Finan	1
Diamond Ordnance Fuze Laboratory Connecticut Ave. & Van Ness Street Washington 25, D.C. ORDTL-012, E.W. Channel	1
Harvard University Cambridge, Massachusetts ATTN: School of Applied Science, Dean Harvey Brook	1
Commanding Officer and Director U.S. Naval Training Device Center Port Washington Long Island, New York ATTN: Technical Library	1
Office of Naval Research Washington 25, D.C. ATTN. Code 450, Dr. R. Trumbull	1
The University of Chicago Institute for Computer Research Chicago 37, Illinois ATTN Mr. Nicholas C. Metropolis, Director	1
U.S. Army Biological Warfare Laboratories Building 29, Room 516 Bethesda 14, Maryland ATTN. Clifford J. Maloney, Division of Biologics Standards	1
Stanford Research Institute Computer Laboratory Menlo Park, California ATTN: H. D. Crane	1

ORGANIZATION	NO. COPIES
Air Force Office of Scientific Research Directorate of Information Sciences Washington 25, D.C. ATTN: Dr. Harold Wooster	1
National Bureau of Standards Washington 25, D.C. ATTN: Miss Ida Rhodes, 220 Stucco Bldg.	1
New York University New York, New York ATTN: Dr. J.H. Mulligan, Jr. Chairman, Electrical Engineering Dept.	1
Texas Technological College Lubbock, Texas ATTN: Paul G. Griffith, Dept. of Electrical Engineering	1
Prof. Frank J. Mullins c/o Bellcomm, Inc. 1737 L Street, N.W. Washington 6, D.C.	1
L.G. Hanscom Field/AF-CRL-CRRB/ Bedford, Massachusetts ATTN: Dr. H.H. Zschirnt	1
Rome Air Development Center Griffiss Air Force Base Rome, New York ATTN: Mr. Alan Barnun	1
Department of the Army Office of the Chief of Research & Development Pentagon, Room 3D442 Washington 25, D.C. ATTN: Mr. L.H. Geiger	1
Dr. George Malcolm Dyson Chemical Abstracts Ohio State University Columbus 10, Ohio	1
Royal Aircraft Establishment, Mathematics Dept. Farnborough, Hampshire, England ATTN: Mr. R.A. Fairthorne, Minister of Aviation	1
University of Pennsylvania Moore School of Electrical Engineering 200 South 33rd Street Philadelphia 4, Pennsylvania ATTN: Miss Anna Louise Campion	1
Department of the Army Office of the Asst. COFD for Intelligence Room 2B529, Pentagon Washington, D.C. ATTN: John F. Kullgren	1
Division of Automatial Data Processing/AOP/ Department of State Washington 25, D.C. ATTN: F.P. Diblasi, 19A16	1
University of Pennsylvania Mechanical Languages Projects Moore School of Electrical Engineering Philadelphia 4, Pennsylvania ATTN: Dr. Saul Gorn, Director	1
Mr. Bernard M. Fry, Deputy Head Office of Science Information Service National Science Foundation 1951 Constitution Avenue, N.W. Washington 25, D.C.	1
Harry Kesten Cornell University Dept. of Mathematics Ithaca, New York	1
Applied Physics Laboratory Johns Hopkins University 8621 Georgia Avenue Silver Spring, Maryland ATTN: Document Library	1
Bureau of Supplies and Accounts, Chief Navy Department Washington, D.C. ATTN: Code W3	1
National Aeronautics & Space Administration Goddard Space Flight Center Greenbelt, Maryland ATTN: Chief, Data Systems Division, C.V.L. Smith	1
Federal Aviation Agency Bureau of Research and Development Washington 25, D.C. ATTN: RD-375, Mr. Harry Hayman	1
American Systems Incorporated 3412 Century Boulevard Inglewood, California ATTN: M.D. Adcock	1

DISTRIBUTION LIST
CONTRACT NO. Nonr-222(53)

ORGANIZATION	NO. COPIES
Commanding Officer ONR Branch Office 1030 E. Green Street Pasadena, California	1
Commanding Officer ONR Branch Office 1000 Geary Street San Francisco 9, California	1
National Bureau of Standards Washington 25, D. C. ATTN: Mr. R. D. Elbourn	1
Syracuse University Electrical Engineering Department Syracuse 10, New York ATTN: Dr. Stanford Goldman	1
Dr. W. Papian Lincoln Laboratories, MIT Lexington, Massachusetts	1
Institute for Defense Analysis Communications Research Division Von Neumann Hall Princeton, New Jersey	1
Air Force Office of Scientific Research Information Research Division Washington 25, D. C. ATTN: R. W. Swanson	1
W. A. Kozumplik, Manager Lockheed Aircraft Corporation Missiles & Space Division 3251 Hanover Street Palo Alto, California	1
Joel Levy National Bureau of Standards Far West Building, 1B Washington, D. C	1
The Rand Corporation 1700 Main Street Santa Monica, California ATTN: Numerical Analysis Dept., Willis H. Ware	1
Massachusetts Institute of Technology Cambridge 39, Massachusetts ATTN: Prof. John McCarthy, 26-007B	1
Sylvania Electric Systems 1100 Wehrle Drive Buffalo 21, New York ATTN: R. L. San Soucie	1
Carnegie Institute of Technology Pittsburgh, Pennsylvania ATTN: Director, Computation Center, Alan J. Perlis	1
Chief, Bureau of Naval Weapons Navy Department Washington 25, D. C ATTN: RREN	1
Electronic Systems Development Corp. 1484 E. Main Street Ventura, California ATTN: Barbara J. Lange	1
Electronics Research Laboratory University of California Berkeley 4, California ATTN: Director	1
R. Turyn Applied Research Laboratory Sylvania Electric Products, Inc. 40 Sylvan Road Waltham 54, Massachusetts	1
George Washington University Washington, D. C. ATTN: Prof. N. Grisamore	1
Cornell Aeronautical Laboratory, Inc. P. O. Box 235 Buffalo 21, New York ATTN: Systems Requirements Dept., A. E. Murray	1
Chief, Bureau of Ships Code 671A2 Washington, D. C. ATTN: LCDR. E. B. Mahineke, USN	1
Lincoln Laboratory Massachusetts Institute of Technology Lexington 73, Massachusetts ATTN: Library	1
Maj. Gen. Casemiro Montenegro Filho, Director Centro Technico da Aeronautica/CTA Sao Jose Dos Campos Sao Paulo, Brazil	1
Professor C. L. Pekeris, Head Department of Applied Mathematics Weismann Institute of Science Rehovoth, Israel	1
Mr. Julian H. Bigelow Institute for Advanced Study Princeton, New Jersey	1
The Mitre Corporation P. O. Box 208 Bedford, Massachusetts ATTN: Library	1
E. Tomash Ampex Computer Products P. O. Box 329 Culver City, California	1