On the Capacity of Channels with Block Memory

WAYNE E. STARK, MEMBER, IEEE, AND
ROBERT J. MCELIECE, FELLOW, IEEE

Abstract — **The capacity of channels with block memory is investigated. It is shown that, when modeled as a game theoretic problem, the optimum coding and noise distributions when block memory is permitted are independent from symbol to symbol within a block.**

I. Introduction

We consider communication channels in which the memory of the channel lasts for a fixed finite time duration; that is, the channel is a block memoryless channel in the sense that, if we consider blocks of fixed length as single symbols in some much larger alphabet, then the channel is memoryless. One motivation for considering such channels is a spread-spectrum frequency-hopped (SSFH) communication system in which the transmitted signal occupies a certain frequency band for a fixed time duration during which a certain number of symbols are transmitted, whereupon the carrier frequency changes. Interference during one carrier dwell usually is assumed to be independent of the interference during any other. The types of interference for which this is true are partial-band jamming (tone or noise), fading (self-jamming), and certain kinds of multiple-access interference. Another channel which exhibits similar behavior is that of a sequence of computer memory chips each of which stores a finite number of bits. Bits within a single chip may have highly correlated errors (i.e., row errors, column errors, or even whole chips being in error); however, bit errors from one chip to the next are independent.

There are several problems concerning channels with block interference. One of these is to determine optimal coding strategies and the largest rate that information can be reliably transmitted. In a spread-spectrum system subject to jamming, the worst-case jamming strategy subject to constraints on the first-order distribution of the jammer only is of considerable importance. We formulate these as game theory problems in which both the coder and jammer have constraints placed on their first-order distributions. The payoff function is the mutual information between the input and output of the channel. The main result is that, provided the interference (jammer) is independent of the transmitted signal (repeat-back jamming is not allowed), the optimal distribution of symbols within a block for the coder and jammer are independent from symbol-to-symbol. In Section II we formulate and solve the game theory problem. In Section III we give several examples. We conclude in Section IV with some discussion about the results and other applications.

II. Game Theory Formulation

The communication channel we consider has input alphabet A and output alphabet B. Player I, called the *coder*, wishes to communicate information through the channel reliably with the largest possible rate. Player II, called the *jammer*, wants to minimize the rate at which information can be transmitted through the channel. The channel is described by specifying two random variables, X and Y. The random variable X is the input to the channel from the coder and the random variable Y is the output of the channel. The coder's strategies are distributions F_X on the random variable X, while the jammer's strategies are the distributions $G_Y(y|a)$ on the output of the channel when $X = a$ is the input. The jammer thus chooses the conditional probabilities of the output given the input while the coder chooses the distribution of the input. We restrict the set of distributions the players can have as follows. The allowable distributions (strategies) for the coder are given by a set S ($F_X \in S$). The collection $\{G_Y(y|a)\colon a \in A\}$, which we denote by G_Y, is required to be in a set T of allowable channels. The payoff function $\phi(F_X, G_Y)$ for this game is taken to be the mutual information $I(X;Y)$ between the input to the channel X and the output of the channel Y. The objective of player I is to choose $F_X \in S$ to make $\phi(F_X, G_Y)$ as large as possible. Player II chooses $G_Y \in T$ to minimize $\phi(F_X, G_Y)$. Thus associated with the game are two programs.

Program I (Coder's Program):

$$C' = \sup_{F_X \in S} \inf_{G_Y \in T} \phi(F_X, G_Y).$$

Program II (Jammer's Program):

$$C'' = \inf_{G_Y \in T} \sup_{F_X \in S} \phi(F_X, G_Y).$$

A strategy $F_X^* \in S$ such that $\inf\{\phi(F_X^*, G_Y)\colon G_Y \in T\} = C'$ is called an optimum strategy for the coder. Similarly, if $\sup\{\phi(F_X, G_Y^*)\colon F_X \in S\} = C''$ then G_Y^* is called an optimum strategy for the jammer.

It is clear from the above programs that $C' \le C''$, and it is easy to give examples where $C' < C''$. However, since ϕ is convex $\cap$ (concave) in F_X and convex $\cup$ (convex) in G_Y [1, Theorems 1.6, 1.7] if S and T are compact convex sets, then $C' = C''$ [2]. This equality is equivalent to the existence of a *saddle point*, i.e., a pair of strategies $F_X^* \in S$, $G_Y^* \in T$ such that

$$\phi(F_X, G_Y^*) \le \phi(F_X, G_Y) \le \phi(F_X^*, G_Y), \qquad F_X \in S,\ G_Y \in T. \tag{1}$$

If (1) holds, then F_X^* and G_Y^* are optimal strategies for the coder and jammer, respectively. This game theory formulation was considered by Dobrushin [3] and Blachman [4]. While we do not have any physical justification for assuming the set of possible strategies is compact, time-sharing between two strategies is a physical justification for the set of possible strategies to be convex.

We generalize this game theory formulation by allowing the players to adopt m-dimensional strategies (i.e., nonmemoryless strategies). We extend the definition of admissible strategies to higher dimensions by using the notion of the mixture of a set of distribution functions. Let the m-dimensional distribution $F_X^{(m)}(\boldsymbol{x})$, $\boldsymbol{X} = (X_1, \cdots, X_m)$, $\boldsymbol{x} = (x_1, x_2, \cdots, x_m)$ have marginal distribution $F_{X_i}(x) = F_X^{(m)}(\infty, \infty, \cdots, \infty, x, \infty, \cdots, \infty)$ with the ith component being x. We say $F_X^{(m)} \in S^{(m)}$ if the uniform mixture of the marginals is in S:

$$F_X^{(m)} \in S^{(m)} \text{ if } \frac{1}{m}\sum_{i=1}^{m} F_{X_i}(x) \in S. \tag{2}$$

The admissible strategies for the jammer are defined similarly. We say $G_Y^{(m)} = \{G_Y^{(m)}(\boldsymbol{y}|\boldsymbol{a})\colon \boldsymbol{a} \in A^m\} \in T^{(m)}$ where $G_Y^{(m)}(\boldsymbol{y}|\boldsymbol{a})$ is the m-dimensional conditional distribution of the output of the channel given the input $\boldsymbol{X} = \boldsymbol{a}$, if the uniform mixture of the conditional marginals $G_{Y_i}(y|a)$ is in T:

$$G_Y^{(m)} \in T^{(m)} \text{ if } \bar{G}_Y \in T \tag{3}$$

where

$$\bar{G}_Y = \left\{ G_Y(y|a)\colon G_Y(y|a) = \frac{1}{m}\sum_{i=1}^{m} G_{Y_i}(y|a),\ a \in T \right\}$$

Manuscript received June 12, 1984; revised February 10, 1987. This work was supported in part by the National Science Foundation under Grant ECS-8307150 and in part by the Air Force Office of Scientific Research under Grant AFOSR-83-0296.

W. E. Stark is with the Department of Electrical Engineering and Computer Science at the University of Michigan, Ann Arbor, MI 48109.

R. J. McEliece is with the Department of Electrical Engineering at the California Institute of Technology, Pasadena, CA 91125.

IEEE Log Number 8819894.

is the collection of conditional distributions with uniform mixture of the marginals $G_{Y_i}(y|a)$ of $G_Y^{(m)}(y|\boldsymbol{a})$. We note here that we have restricted the strategies to those with no "intersymbol interference;" i.e., previous inputs are not allowed to affect current outputs. For these generalized strategies we have the following programs.

Program I_m (Coder's Program):

$$C_m' = \sup_{F_X^{(m)} \in S^{(m)}} \inf_{G_Y^{(m)} \in T^{(m)}} \phi\left(F_X^{(m)}, G_Y^{(m)}\right).$$

Program II_m (Jammer's Program):

$$C_m'' = \inf_{G_Y^{(m)} \in T^{(m)}} \sup_{F_X^{(m)} \in S^{(m)}} \phi\left(F_X^{(m)}, G_Y^{(m)}\right)$$

where the payoff function is now $\phi(F_X^{(m)}, G_Y^{(m)}) = (1/m) I(X;Y)$. We have the following result concerning C_m' and C_m''.

Theorem 1: $C_m' = C'$ *and* $C_m'' = C''$ for all $m \geq 1$.

Proof: First we prove $C_m' = C'$. Let F_X be an admissible strategy; i.e., $F_X \in S$, and let $G_Y^{(m)} \in T^{(m)}$ be an admissible strategy for the jammer. Then if $\boldsymbol{X} = (X_1, \cdots, X_m)$ is a random vector consisting of m independent identically distributed (i.i.d.) copies of X, we have

$$\begin{aligned}\phi\left(F_X^{(m)}, G_Y^{(m)}\right) &\geq \frac{1}{m}\sum_{i=1}^{m} \phi\left(F_X, G_{Y_i}\right) \qquad \text{[1, Theorem 1.8]} \\ &\geq \phi\left(F_X, G_Y\right) \qquad \text{[1, Theorem 1.7]}\end{aligned}$$

where G_Y is the uniform mixture of the conditional distributions G_{Y_i}. Since $G_Y \in T$ we have

$$\inf_{G_Y^{(m)} \in T^{(m)}} \phi\left(F_X^{(m)}, G_Y^{(m)}\right) \geq \inf_{G_Y \in T} \phi\left(F_X, G_Y\right).$$

Now since the above holds when $(X_1, \cdots, X_m)$ are i.i.d. we have that

$$\sup_{F_X^{(m)} \in S^{(m)}} \inf_{G_Y^{(m)} \in T^{(m)}} \phi\left(F_X^{(m)}, G_Y^{(m)}\right) \geq \sup_{F_X \in S} \inf_{G_Y \in T} \phi\left(F_X, G_Y\right),$$

so that $C_m' \geq C'$. Now let $F_X^{(m)} \in S^{(m)}$ and $G_Y(y|a)$ be arbitrary. Then if $G_Y^{(m)}$ is the m-dimensional distribution of $\boldsymbol{Y} = (Y_1, \cdots, Y_m)$ given $\boldsymbol{X} = \boldsymbol{a}$ with Y_i conditionally independent, then

$$\begin{aligned}\phi\left(F_X^{(m)}, G_Y^{(m)}\right) &\leq \frac{1}{m}\sum_{i=1}^{m} \phi\left(F_{X_i}, G_Y\right) \qquad \text{[1, Theorem 1.9]} \\ &\leq \phi\left(F_X, G_Y\right) \qquad \text{[1, Theorem 1.6]}\end{aligned}$$

where F_X is the uniform mixture of $F_X^{(m)}$; i.e.,

$$F_X(x) = \frac{1}{m}\sum_{i=1}^{m} F_{X_i}(x).$$

Since the above is true for the case when $Y_1, Y_2, \cdots, Y_m$ are conditionally independent we have that

$$\inf_{G_Y^{(m)} \in T^{(m)}} \phi\left(F_X^{(m)}, G_Y^{(m)}\right) \leq \inf_{G_Y \in T} \phi\left(F_X, G_Y\right).$$

Now since $F_X \in S$ we have

$$\sup_{F_X^{(m)} \in S^{(m)}} \inf_{G_Y^{(m)} \in T^{(m)}} \phi\left(F_X^{(m)}, G_Y^{(m)}\right) \leq \sup_{F_X \in S} \inf_{G_Y \in T} \phi\left(F_X, G_Y\right),$$

so that $C_m' \leq C'$. Thus $C_m' = C'$ as asserted. Similarly, $C_m'' = C''$.

What we have shown is that of all block m memoryless channels, the channel with minimum mutual information is the block 1 memoryless channel. The conclusions one can draw from this are that memoryless jamming is optimal $(G_Y^{(m)*}(y|\boldsymbol{x}) = \prod_{i=1}^m G^*(y_i|x_i))$ and memoryless coding is optimal $(F_X^{(m)*}(\boldsymbol{x}) = \prod_{i=1}^m F_{X_i}^*(x_i))$. Thus the interference (which is independent of the transmitted signal) that minimizes the mutual information is independent from symbol to symbol.

III. Examples

Consider the channel with $A = B = \{0,1,\cdots,M-1\}$ and let T be the set of channels with error probability per symbol less than ϵ, $0 \leq \epsilon \leq 1$, and S be all distributions on A. Then a result of Dobrushin [3] is that

$$C' = C'' = \begin{cases} \log M + (1-\epsilon)\log(1-\epsilon) + \epsilon \log \epsilon - \epsilon \log(M-1), \\ \quad \epsilon \leq 1 - (1/M) \\ 0, \qquad \epsilon \geq 1 - (1/M) \end{cases} \tag{4}$$

Here C' and C'' are measured in bits per channel use and all logarithms have base 2. The optimal distribution $F_X^*(x)$ is the uniform distribution on $\{0,1,\cdots,M-1\}$ and the optimal channel $G^*(y|x)$ satisfies

$$G^*(y|x) = \begin{cases} \epsilon/(M-1), & y \neq x,\ \epsilon \leq 1 - \dfrac{1}{M} \\ 1-\epsilon, & y = x,\ \epsilon \leq 1 - \dfrac{1}{M} \\ \dfrac{1}{M}, & \epsilon \geq 1 - \dfrac{1}{M} \end{cases} \tag{5}$$

Here we generalize this game to the m-dimensional case and apply the theorem. For the generalized strategies we use the channel m times to transmit m symbols. Let ϵ_i be the error probability of the ith channel. Then $T^{(m)}$ is the set of channels with

$$\frac{1}{m}\sum_{i=1}^{m} \epsilon_i \leq \epsilon.$$

Also $S^{(m)}$ is the set of distributions on A^m. By Theorem 1 $C_m' = C'$ and $C_m'' = C''$ and is given in (4). The optimal strategies are memoryless with marginals $F_X^*(x)$ and $G_Y^*(y|x)$ given above. A conclusion one can draw is that if the *average* error probability is less than ϵ, then with memoryless encoding $I(X;Y) \geq C' = C''$ with equality for the optimal strategy given above. Notice that the worst possible channel (for fixed average error probability) is thus a symmetric channel. This example can be used to model a faded channel. The type of fading determines the first-order statistics, i.e., ϵ. For a slowly faded channel the statistics are such that

$$\epsilon_i = \epsilon, \qquad i = 1,2,\cdots,m.$$

However, the lowest mutual information is achieved by fast fading, i.e., when the channel fading variable is independent from symbol to symbol within a block.

As a second example, consider $A = B = R$, the real line. Let

$$S = \left\{ F_X(x) : \int_R x^2\, dF_X(x) \leq E \right\} \tag{6}$$

and

$$T = \left\{ G_Y(y|x) : \int_{y \in R} (y-x)^2\, dG_y(y|x) \leq N,\ x \in A \right\}. \tag{7}$$

The set of channels is restricted to channels whose added noise has mean square less than or equal to N. For S and T given in (6) and (7) a result of Dobrushin [3] and Blachman [4] shows that

$$C' = C'' = \frac{1}{2}\log\left(1 + \frac{E}{N}\right) \tag{8}$$

with $F^*(x) = \Phi(x/\sqrt{E})$ and $G_Y^*(y|x) = \Phi((y-x)/\sqrt{N})$ where $\Phi(u)$ is given by

$$\Phi(u) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{u} e^{-x^2/2}\, dx. \tag{9}$$

Again we generalize this game to allow for m-dimensional distributions. The sets $S^{(m)}$ and $T^{(m)}$ are given by

$$S^{(m)} = \left\{ F_X^{(m)}(x) : \frac{1}{m} \int_{R^m} x \cdot x'\, dF_X^{(m)}(x) \le E \right\}$$

where α' denotes the transpose of the vector α, and

$$T^{(m)} = \left\{ G_Y^{(m)}(y|x) : \frac{1}{m} \int_{R^m} (y-x) \cdot (y-x)'\, dG_Y^{(m)}(y|x) \le N,\ x \in A^m \right\}.$$

By Theorem 1, $C_m' = C_m'' = C' = C''$ and the optimal distributions are memoryless with marginal distributions being Gaussian. We note here that Theorem 1 is the discrete time analog of the result for continuous time channels that white Gaussian noise is the optimal jamming and coding strategy.

IV. Conclusion

We have formulated block memoryless channels as a game theory problem and have shown that optimum coding strategies are independent from symbol to symbol within a block and optimal jamming strategies are also independent from symbol to symbol within a block. Many important questions still need to be answered concerning the optimum jamming distribution for each symbol within a block and robust decoding strategies.

References

[1] R. J. McEliece, *The Theory of Information and Coding*. Reading, MA: Addison-Wesley, 1977.

[2] J. Stoer and C. Witzgall, *Convexity and Optimization in Finite Dimensions I*. New York: Springer-Verlag, 1970.

[3] R. L. Dobrushin, "Optimum information transmission through a channel with unknown parameters," *Radio Eng. Electron.*, vol. 4, no. 12, 1959.

[4] N. M. Blachman, "Communication as a game," in *Proc. Wescon 1957 Conf. Rec.*, Aug. 1957, pp. 61-66.

Almost Asymptotically Optimal Flag Encoding of the Integers

MUZHONG WANG

***Abstract*—A simple prefix-free encoding scheme for the positive integers is proposed in which a flag of f zeroes is used to mark the end of a codeword, and bit stuffing is used to prevent premature appearance of the flag in the conventional binary coding of the integers. It is shown that this coding scheme is universal in the sense defined by Elias, and that for large f its asymptotic efficiency is virtually 1.**

I. Introduction

Elias considered the problem of finding prefix-free encodings for the positive integers and introduced the notions of universality and asymptotic optimality to characterize codings [1]. A D-ary encoding $\mathscr{E}$ of the positive integers, $N^+ = \{1,2,\cdots\}$, is a function that assigns a different nonempty string $\mathscr{E}(n)$ of letters from a given alphabet of D letters to each n in N^+. The encoding $\mathscr{E}$ is *prefix-free* if no codeword $\mathscr{E}(n)$ is the beginning of another codeword. Hereafter, we consider only binary encoding and take $\{0,1\}$ as the coding alphabet. Note that the conventional binary coding $\mathscr{B}$ of N^+ is not prefix-free because, for example, $\mathscr{B}(2) = 10$ is a prefix of $\mathscr{B}(4) = 100$.

Let $\mathscr{E}$ be a given binary prefix-free encoding of N^+, and let $L(n)$ be the length of the codeword $\mathscr{E}(n)$. Hereafter, let P be any probability distribution of N^+ (i.e., $P(n) \ge 0$, all $n \in N^+$, and $P(1) + P(2) + P(3) + \cdots = 1$) that satisfies

$$P(n) \ge P(n+1), \qquad \text{all } n \in N^+. \tag{1}$$

Let $H(P)$ be the entropy of P in bits, and let $E_P(L)$ denote the expected codeword length for $\mathscr{E}$ when P is the probability distribution. Elias [1] has defined $\mathscr{E}$ to be a *universal* encoding of N^+ if a finite number ρ exists such that

$$\frac{E_P(L)}{\max\{1, H(P)\}} \le \rho \tag{2}$$

for all P, and to be *asymptotically optimal* if it is universal and

$$\frac{E_P(L)}{\max\{1, H(P)\}} \le R(H(P)) \tag{3}$$

where R is some real-valued function such that

$$\lim_{x \to \infty} R(x) = 1. \tag{4}$$

Actually, Elias [1] defined these notions only for encodings $\mathscr{E}$ with the "minimal" property that

$$L(n) \le L(n+1), \qquad n \in N^+, \tag{5}$$

which property minimizes $E_P(L)$ for the codeword set because of (1), but the definitions are useful for any $\mathscr{E}$.

We now introduce another notion of efficiency for prefix-free encodings of N^+. We first note that the conventional binary coding $\mathscr{B}$ of the integers gives a codeword of length $\lfloor \log_2 n \rfloor + 1 = \lceil \log_2 (n+1) \rceil$ to $n \in N^+$, where $\lfloor \cdot \rfloor (\lceil \cdot \rceil)$ denotes the largest integer not greater than (the smallest integer not less than) the enclosed number. It seems natural then to define the *asymptotic efficiency* of a prefix-free encoding of N^+ by

$$\Gamma = \lim_{n \to \infty} \sup \frac{\log_2 n + 1}{L(n)}. \tag{6}$$

Note that $0 \le \Gamma \le 1$ and that P is not involved in the definition of Γ.

In the next section, we introduce a flag encoding scheme for N^+ and develop simple but tight upper and lower bounds for $L(n)$. The upper bound is used in the following section to show that $\Gamma \approx 1$ when the flag length is large and that the encoding then is "almost asymptotically optimal" in a sense that we will define there.

II. The Flag Encoding Scheme for the Integers

The idea of our flag encoding scheme for N^+ is as follows. We choose the flag length f to be some positive integer at least 2. Note that the codeword $\mathscr{B}(n)$ of the conventional binary coding of a positive integer always begins with a 1. Reversing $\mathscr{B}(n)$, we get a string that ends with a 1. All occurrences of the flag 0^f, a string of f zeroes, are removed from this string by stuffing a 1 after each occurrence of 0^{f-1}. Finally, we add the flag 0^f to the end of the string to form the codeword $\mathscr{E}_f(n)$.

Assume that integer n to be encoded is represented in its binary conventional form $\mathscr{B}(n)$. We have the following encoding and decoding algorithms:

Encoding algorithm $\mathscr{E}_f(n)$:

Step 1: Set $\mathscr{E}_f(n)$ equal to $(b_0 b_1 \cdots b_i)$, the reverse of $\mathscr{B}(n) = (b_i \cdots b_1 b_0)$. (Note that $b_i = 1$.)

Manuscript received December 9, 1986; revised May 18, 1987.

The author is with the Institute of Signal and Information Processing, Swiss Federal Institute of Technology, 8092 Zürich, Switzerland.

IEEE Log Number 8820327.