Int. Symp. Inform. Theory, Tsahkadsor, Armenia, USSR, Sept. 2–8, 1971, pp. 23–52, Hungarian Academy of Sciences, 1973.

[3] T. M. Cover, "Broadcast channels," *IEEE Trans. Inform. Theory*, vol. IT-18, pp. 2–14, 1972.

[4] R. Ahlswede, "The capacity region of a channel with two senders and two receivers," *Ann. Prob.*, vol. 2, no. 5, pp. 805–814, 1974.

[5] A. B. Carleial, "Interference channels," *IEEE Trans. Inform. Theory*, vol. IT-24, pp. 60–70, 1978.

[6] L. M. Lipkind, "Two-way discrete memoryless communication channels," *Probl. Peredach. Inform.*, vol. 3, pp. 37–46, 1967.

[7] F. Jelinek, "Loss in information transmission through two-way channels," *Inform. Contr.*, vol. 6, pp. 337–371, 1963.

[8] ——, "Coding for and decomposition of two-way channels," *IEEE Trans. Inform. Theory*, vol. IT-10, pp. 5–17, 1964.

[9] G. Dueck, "The capacity region of the two-way channel can exceed the inner bound," *Inform. Contr.*, vol. 40, no. 3, pp. 258–266, 1979.

[10] F. R. Gantmacher, *Matrix Theory*, Vol. II. New York: Chelsea Publishing, 1959.

[11] T. Berger, "Multiterminal source coding," lecture notes presented at CISM summer school, Udine, Italy, 1977.

[12] T. S. Han and K. Kobayashi, "A unified achievable rate region for a general class of multiterminal source coding systems," *IEEE Trans. Inform. Theory*, vol. IT-26, no. 3, pp. 277–288, 1980.

[13] T. M. Cover and A. El Gamal, "Capacity theorems for the relay channel," *IEEE Trans. Inform. Theory*, vol. IT-25, no. 5, pp. 572–584, 1979.

[14] J. P. M. Schalkwijk, "The binary multiplying channel—A coding scheme that operates beyond Shannon's inner bound region," to appear in *IEEE Trans. Inform. Theory*.

[15] H. G. Eggleston, *Convexity*. Cambridge: Cambridge University, 1958.

[16] R. Ahlswede and J. Körner, "Source coding with side information and a converse for degraded broadcast channels," *IEEE Trans. Inform. Theory*, vol. IT-21, pp. 629–637, 1975.

[17] T. S. Han, "The capacity region of general multiple-access channel with a certain correlated sources," *Inform. Contr.*, vol. 40, no. 1, pp. 37–60, 1979.

[18] T. S. Han, "A coding theorem for the two-way channel," Research Memorandum RMI 81-04, Dept. Mathematical Eng. and Instrumentation Physics, Univ. of Tokyo, June 1981; presented at 5th Res. Meeting Inform. Theory and Its Appl., Hachimantai, Japan, Oct. 25–27, 1982.

Channels with Block Interference

ROBERT J. McELIECE, SENIOR MEMBER, IEEE, AND WAYNE E. STARK, MEMBER, IEEE

Abstract—A new class of channel models with memory is presented in order to study various kinds of interference phenomena. It is shown, among other things, that when all other parameters are held fixed, channel capacity C is an *increasing* function of the memory length, while the cutoff rate R_0 generally is a *decreasing* function. Calculations with various explicit coding schemes indicate that C is better than R_0 as a performance measure for these channel models. As a partial resolution of this C versus R_0 paradox, the conjecture is offered that R_0 is more properly a measure of coding delay rather than of coding complexity.

I. INTRODUCTION AND SUMMARY

REAL communication channels are often plagued by noise whose severity is time-dependent, and yet most research in information theory deals with channels without memory. As a result, an engineer who must design a practical coded communication system may find little guidance in the information theory literature. In this paper we attempt to partially remedy this situation by studying a

new class of channel models with memory. These channels, which we call *block interference* (*BI*) *channels*, are simple enough to be mathematically tractible and yet complex enough to reflect many of the difficulties caused by channel memory in practice.

Example: Consider the two memoryless binary symmetric channels (BSC) in Fig. 1.

Channel Δ_0 (in Greek, channel = $\delta i \alpha \upsilon \lambda o \zeta$) is a noiseless BSC, and $\Delta_{1/2}$ is a useless BSC with crossover probability $\frac{1}{2}$. Let S be a random variable with $\Pr\{S = 0\} = 1 - \epsilon, \Pr\{S = 1/2\} = \epsilon$. Then for any integer $m \geq 1$, we define the BI channel Δ^m as follows. When a sequence of binary digits is to be sent over Δ^m, each block of m consecutive digits is in fact sent over one of the two component channels. The random variable S determines which component channel is used to transmit a given m-block. It follows that this interference channel is characterized by occasional severe error bursts of length m. This very simple example is in many ways typical of the whole family of BI channels. In particular, it illustrates the important assumptions that makes our channel models mathematically tractible: the noise severity must remain constant in blocks of m consecutive transmitted symbols but is independent from block to block. This assumption is very restrictive, and certainly many real channels do not have constant memory. Still, our models do match certain
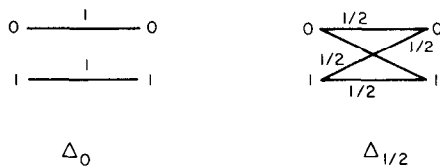
Fig. 1.   Component channels for a simple interference channel.

real channels very well. For instance:

Slow frequency-hopped spread-spectrum systems subject to jamming, fading, or multiple-access interference suffer from phased bursts of interference of fixed duration [4], [13], [17]. (We named our channels with this application in mind.)

In optical communication systems using direct detection and pulse-position modulation, the appropriate channel model is (with a slight modification to include side information) the channel of Example 1 [9].

In any concatenated coding system with a memoryless inner channel and an inner block code, the outer channel is a BI channel in the sense of this paper.

Furthermore, we believe that many of our results will remain true, at least qualitatively, for more sophisticated channel models.

The paper is in seven sections. In Section II we give the general definition of an interference channel and several additional concrete examples. Roughly speaking, a BI channel is defined to be a stochastic mixture of the mth extensions of a family of memoryless channels, called the component channels. In Section II and throughout the paper, we emphasize the difference between BI channels *with* and *without side information*. By side information, we mean information available to the receiver about the current state of the channel. In practice such side information is often provided by an automatic gain control device; sometimes it is called soft decision information. In our models, when side information is present, it is provided directly by the channel and tells the receiver which of the component channels is currently active. (The BI channel of Example 1, with side information, is equivalent to an erasure-burst channel, and is the model used for the study of optical communication in [7] and [9].)

In Section III we study the capacity of BI channels. We show that the capacity of a BI channel with side information is simply the average $\overline{C}$ of the capacities of the component channels, *independent of the memory length* m. The capacity without side information is more difficult to compute, but we show that it is in general less than $\overline{C}$, apparently increases with m, and approaches $\overline{C}$ as m approaches infinity.

Many researchers believe that in some respects the computational cutoff rate R_0 is more useful than capacity as an indicator of a channel's ability to transmit information. In Section IV we study R_0 for BI channels. Given the results

of Section III one might guess that with side information present R_0 would be the average of the R_0's of the component channels. However, we show that it is 2^{-R_0} rather than R_0 that behaves this way. One consequence of this is that R_0 is a *decreasing* function of channel memory. This is also true for BI channels without side information; for these channels we have the paradoxical situation that as m increases, C increases but R_0 decreases!

In Section V we present numerical values for C and R_0 for the interference channel of Example 1, with $\epsilon = 0.1$ and various values of m. We also present performance curves for several different combinations of coding and interleaving for this channel. These performance curves appear to indicate that C is better than R_0 as a performance measure for block-interference channels. In Section VI we give a brief description of how a fixed "test pattern" can be used to estimate the channel state when n is large and side information is absent. This technique is one practical way of exploiting the channel's memory.

In Section VII we attempt to draw general conclusions from our work. First, we point out that since, with side information present, C is independent of m, the interleaved coding strategies often used in practice apparently suffer from no theoretical disadvantage when compared to codes which deal more directly with burst errors. On the other hand, when side information is not available, C increases with m and interleaving is a much riskier bet. However, if m is large enough, the receiver can make a reliable statistical estimate of the noise severity even if the channel does not provide side information. If this is done, interleaving may still be the best available coding strategy. As for the C versus R_0 paradox, we suggest that, contrary to current folklore, R_0 may be more properly an inverse measure of coding *delay* rather than coding *complexity*. We cannot rigorously defend this view, but it does render some otherwise confusing data (e.g., Table I) comprehensible.

II.   CHANNEL MODELS

In this section we shall define the class of channel models we call block interference channels. In any given situation we shall in fact describe two slightly different channels. Both channels are characterized by a random variable S and a positive integer m. The random variable S describes the noise severity at a given time, and the integer m measures the duration of a given level of noise severity. In neither case does the *transmitter* know what the current noise severity is, but we shall distinguish the two cases where the *receiver* does or does not know the noise severity.

We start with a collection $\{\Delta_s\}$ of memoryless channels, each with the same discrete input alphabet A and discrete output alphabet B. We assume that the index parameter s lies in a set Ω on which a probability distribution P has been defined. Let $S_1, S_2, S_3, \cdots$ be a sequence of independent identically distributed Ω-valued random variables, the common distribution being that given by P. For each integer $m \geqslant 1$ we now define the BI channels Δ^m and $\overline{\Delta}^m$,

both channels having input alphabet A and output alphabet B.

When a sequence of letters $x_0, x_1, x_2, \cdots$ from A is transmitted over Δ^m the kth block of m consecutive letters, viz., $[x_{(k-1)m}, x_{(k-1)m+1}, \cdots, x_{km-1}]$ is in fact transmitted over one of the component channels Δ_s; the random variable S_k determines which component channel is used. We assume throughout that both sender and receiver know when these blocks begin and end, i.e., block synchronization is assumed. The receiver gets no direct information about which component channel is active at any given time. For this reason, we call Δ^m the BI channel *without side information*.

The channel $\overline{\Delta}^m$ is almost the same, the difference being that, together with each noisy block, the receiver is also given the index s_k of the component channel used to transmit that block. We call $\overline{\Delta}^m$ the BI channel *with side information*. (For $m = 1$, these channel models are the same as those considered by Wolfowitz [18, sec. 4.6]. The channel Δ^1 is a Wolfowitz compound channel of "Type I," and $\overline{\Delta}^1$ is of "Type III.")

Formally, the channel Δ^m (without side information) can be viewed as a traditional discrete memoryless channel (DMC) with input alphabet A^m, output alphabet B^m, the channel statistics being the P-mixture of the statistics of the mth memoryless extensions $\{\Delta_s^m\}$ of the component channels. Similarly the channel $\overline{\Delta}^m$ (with side information) can be viewed as a memoryless channel with input alphabet A^m and output alphabet $B^m \times \Omega$. Coding theorems for our channel models therefore follow directly from the standard coding theorems for memoryless channels, and so there is no need for proofs of coding theorems in this paper.

We conclude this section with some concrete examples of interference channels.

Example 1: For each $s \in [0, 1]$, let Δ_s be a binary symmetric channel with crossover probability s, and let P be an arbitrary but fixed probability distribution on $[0, 1]$. The interference channels Δ^m and $\overline{\Delta}^m$ are then characterized by phased bursts of errors of varying severity. When Δ^m is viewed as a DMC as suggested above, it becomes a 2^m-ary channel with input and output alphabets $\{0, 1\}^m$. The probability that $y = (y_1, \cdots, y_m)$ is received, given that $x = (x_1, \cdots, x_m)$ is the input, is given by

$$p(y|x) = \int_0^1 s^k (1 - s)^{m-k} \, dP, \qquad (2.1)$$

where k is the Hamming distance between the binary m-tuples x and y. The channel $\overline{\Delta}^m$ has the same input alphabet, but its output alphabet is the infinite alphabet $\{0, 1\}^m \times [0, 1]$. We now give two special cases of this rather general example.

Example 1a: Here the distribution P is concentrated at the two points $s = 0$ and $s = \frac{1}{2}$. The channel Δ_0 is a noiseless BSC, and $\Delta_{1/2}$ is a useless BSC as shown in Fig. 1.

If $P\{S = 0\} = 1 - \epsilon$, $P\{S = \frac{1}{2}\} = \epsilon$, this example is the same as that given in Section I, and the statistic (2.1)

specializes to

$$p(y|x) = \begin{cases} 1 - \epsilon(1 - 2^{-m}), & y = x; \\ \epsilon \cdot 2^{-m}, & y \neq x. \end{cases} \qquad (2.2)$$

$$\bar{p}(y, 0|x) = \begin{cases} 1 - \epsilon, & y = x; \\ 0, & y \neq x. \end{cases} \qquad (2.3)$$

$$\bar{p}(y, \tfrac{1}{2}|x) = \epsilon \cdot 2^{-m} \qquad \text{all } x, y. \qquad (2.4)$$

In this special case Δ^m is a 2^m-ary symmetric channel with channel error probability $\epsilon(1 - 2^{-m})$. In particular Δ^1 is a BSC with crossover probability $\epsilon/2$. Notice also that here the channel $\overline{\Delta}^m$ is equivalent to a 2^m-ary erasure channel, since, if the side information associated with a given b-bit block is $s = \frac{1}{2}$, then that block is wholly unreliable and may as well be erased, whereas if $s = 0$, then the block is guaranteed to be without error. In particular Δ^1 is equivalent to a binary erasure channel with erasure probability ϵ. The channel $\overline{\Delta}^m$ occurs in practice in optical communication systems using direct detection and pulse position modulation [7], [9].

Example 1b (Rayleigh Fading): Here we take P to be a continuous probability distribution on $[0, \frac{1}{2}]$ with density given by

$$p(s) = \alpha 2^\alpha \cdot s^{\alpha - 1} \qquad 0 \leqslant s < \tfrac{1}{2}; \qquad (2.5)$$

where α is any positive real number. In practice this arises when binary FSK modulation is used on a slow Rayleigh fading channel with signal-to-noise ratio $E/N_0 = 2/\alpha$ [19, ch. 7].

Example 2 (Pulsed Jammer): Here the input alphabet A consists of only two letters, $A = \{-1, +1\}$, and the output alphabet B is the real line. There are only two component channels, Δ_{on} and Δ_{off}. Δ_{on} ("jammer on") is an additive Gaussian noise channel with noise variance equal to σ^2. Δ_{off} is a noiseless channel ("jammer off"), for which the output equals the input exactly. If $P\{S = \text{on}\} = \rho$, $\Pr\{S = \text{off}\} = 1 - \rho$ this channel is equivalent to the pulsed jammer channel considered by Viterbi [17], where the average signal-to-noise ratio is $E/N_0 = 1/(2\rho\sigma^2)$. Here side information is intrinsically available, since the reception of ± 1 indicates (with probability 1) that the jammer is off, and so $\Delta^m = \overline{\Delta}^m$.

III. CHANNEL CAPACITY

In this section we will discuss the capacities of the channels Δ^m and $\overline{\Delta}^m$. However, in all of our proofs we will make the following *compatibility assumption* about the component channels $\{\Delta_s\}$:

Input distribution which achieves capacity is the same for each of the component channels Δ_s. (3.1)

For example, if the Δ_s are all binary symmetric channels, (3.1) is satisfied, since a uniform input distribution achieves capacity for any BSC [12, th. 2.3].

Let us denote the capacities of Δ^m and $\overline{\Delta}^m$, measured in bits per channel use, by $C(m)$ and $\overline{C}(m)$, respectively. If we take the memoryless view described in Section II, the

calculation of these capacities is simply a matter of maximizing the appropriate mutual informations. For $\overline{\Delta}^m$, the capacity is

$$C(m) = \frac{1}{m} \max_X I(X; Y), \qquad (3.2)$$

where X and Y denote the m-dimensional random inputs and outputs of Δ^m. For $\overline{\Delta}^m$, the capacity is

$$\overline{C}(m) = \frac{1}{m} \max_X I(X; Y, S), \qquad (3.3)$$

where S is the random variable describing the noise severity. (We assume X and S are independent.) Since $I(X; Y) \leqslant I(X; Y, S)$ [12, th. 1.4] it follows immediately that for all m,

$$C(m) \leqslant \overline{C}(m). \qquad (3.4)$$

This result is intuitively obvious, since one would not expect side information to degrade performance.

We shall consider $\overline{C}(m)$ in detail first, since its calculation is the easier of the two. We have

$$I(X; Y, S) = I(X; S) + I(X; Y|S) \qquad [6, \text{eq. } (2.5.4)]$$
$$= I(X; Y|S) \qquad (\text{since } X \text{ and } S \text{ are independent})$$
$$= \int_\Omega I(X; Y^{(s)}) \, dP, \qquad (3.5)$$

where $Y^{(s)}$ denotes the output of the mth memoryless extension of the channel Δ_s, if X is the input. Since each Δ_s is memoryless, we have [6, th. 4.2.1]

$$I(X; Y^{(s)}) \leqslant mC_s, \qquad (3.6)$$

where C_s denotes the capacity of Δ_s, with equality if and only if the components of X are independent and have a distribution that achieves capacity C_s. It follows immediately from the compatability assumption (3.1) that

$$\overline{C}(m) = E\{C_s\} \triangleq \overline{C}, \qquad (3.7)$$

i.e., the capacity with side information is the average of the individual capacities of the component channels and is independent of m. This result is not entirely new since the memoryless version of $\overline{\Delta}^m$ is a channel of a type considered by Wolfowitz [18, sec. 4.6, channel type III] who gives a general formula [18, eq. (4.6.11)] for the capacity. However, our compatability assumption (3.1) greatly simplifies the formula for the capacity, and our block memory viewpoint allows us to draw a new conclusion, viz., that *capacity is independent of the burst length.*

Next, we return to the capacity without side information, $C(m)$. It follows from (3.4) and (3.7) that

$$C(m) \leqslant \overline{C}, \qquad m = 1, 2, 3, \cdots. \qquad (3.8)$$

We now claim that, at least when S has finite entropy or when the alphabets A and B are both finite,

$$\lim_{m \to \infty} C(m) = \overline{C}. \qquad (3.9)$$

To prove (3.9), we note for any random variables X, Y, and Z (see, e.g., [20 lemmas 3.1, and 3.2]).

$$I(X; Y, Z) \leqslant I(X; Y) + I(X, Y; Z), \qquad (3.10)$$

with equality if and only if Y and Z are independent. Using (3.10), we have

$$I(X; Y, S) - I(X, Y; S) \leqslant I(X; Y). \qquad (3.11)$$

Our allegation (3.9) will now follow from the definitions (3.2) and (3.3), provided we can show that

$$\lim_{m \to \infty} \frac{1}{m} I(X, Y; S) = 0. \qquad (3.12)$$

We conjecture that (3.12) holds in general, but we do not at present have a general proof. Of course, if S has finite entropy then $I(X, Y; S) = H(S) - H(S|X, Y) \leqslant H(S)$, from which (3.12) follows immediately.

If the alphabets A and B are both finite, we can prove (3.12) as follows: let $|A| = a$, $|B| = b$. Then without loss of generality we suppose the set Ω to consist of all possible memoryless channels connecting A and B; i.e., all $a \times b$ stochastic matrices $\{ p(y|x): y \in B, x \in A \}$. Let $X = (X_1, \cdots, X_m)$ and $Y = (Y_1, \cdots, Y_m)$ be observations of the input and output of an unknown one of these channels, and consider the statistical problem of estimating which channel gave rise to the observation. Clearly a sufficient statistic T for this estimation is a histogram describing the number of times each of the pairs $(x, y) \in A \times B$ occurs among the m observed pairs (X_k, Y_k). (See Ferguson [5, sec. 3.3].) This means that S and the pair (X, Y) are independent when conditioned on T, i.e., $I(S; X, Y|T) = 0$. Hence, since $(X, Y) \to T \to S$ is a Markov chain, ([6, eq. (2.3.18)]),

$$I(S; X, Y) = I(S; X, Y, T) = I(S; T) + I(S; X, Y|T)$$
$$= I(S; T). \qquad (3.13)$$

Since there are ab components in the histogram T, and each component lies in the set $\{0, 1, \cdots, m\}$, there are at most $(m + 1)^{ab}$ possible histograms, and so

$$I(S; T) \leqslant H(T) \leqslant ab \log(m + 1). \qquad (3.14)$$

(The exact value is $\binom{m + ab - 1}{m}$, but we do not need this fact.) Thus, combining (3.13) and (3.14), we get

$$\frac{1}{m} I(X, Y; S) \leqslant ab \frac{\log(m + 1)}{m}.$$

Since a and b are fixed, this proves (3.12) when A and B are finite.

Taken together, (3.8) and (3.9) say that, while the lack of side information can definitely hurt performance, if the channel memory is large then the penalty paid for the lack of side information is small. Again, on reflection, this result is obvious. For if m is very large, the sender could reserve a certain small fraction of symbols in each transmitted block as a known "test pattern," and the receiver could estimate the noise severity S on the basis of the noisy test pattern. We will have more to say about test patterns in Section VI.

In Section V, below, we present numerical values for $C(m)$ and $\overline{C}$ for the BI channel of Example 1a.

IV. CUTOFF RATES

In this section we will discuss the computational cutoff parameters R_0 of the burst interference channels Δ^m and $\bar{\Delta}^m$. However, as in Section III, we make a simplifying second *compatibility assumption*:

Input distribution which achieves R_0 is the same for
each of the component channels Δ_s.   (4.1)

Again we note that this condition is satisfied for any set of binary symmetric channels, and indeed for any set of channels which are symmetric enough so that a uniform input distribution achieves R_0. (In fact, (4.1) is satisfied for any set of binary input DMC's, since the uniform input distribution achieves R_0 for all such channels [6, problem 5.13].)

We denote the cutoff parameters of Δ^m and $\bar{\Delta}^m$, measured in bits per channel use, by $R_0(m)$ and $\bar{R}_0(m)$, respectively. As in our calculations with capacity, we shall view our block interference channels as memoryless channels over product alphabets. We now recall the definition of R_0 for a discrete memoryless channel. In Gallager's notation [6, ch. 4], $R_0 = \sup_Q E_0(1, Q)$. Our definition, which follows, is equivalent to Gallager's, but is motivated by Omura's work [14] on Bhattachayya distance. See also [12, pp. 68–69].

If x_1 and x_2 are two letters from the input alphabet of a given memoryless channel with input alphabet A and output alphabet B, we define

$$j(x_1, x_2) = \sum_y p(y|x_1)^{1/2} p(y|x_2)^{1/2}.   (4.2)$$

If X is a random variable taking values in the input alphabet A, and if X_1, X_2 are independent copies of X, we define

$$J(X) = E(j(X_1, X_2)).   (4.3)$$

The cutoff rate R_0 is then defined as

$$R_0 = \sup_X \{-\log_2 J(X)\}.   (4.4)$$

Stated another way, if X is an A-valued random variable, we have

$$J(X) \geqslant 2^{-R_0},   (4.5)$$

with equality if and only if X achieves the supremum in (4.4). For the BI channel Δ^m, since we are measuring on a per-channel-use basis, the corresponding definition is

$$R_0(m) = \frac{1}{m} \sup_X \{-\log_2 J(X)\},$$

where the supremum is over all random vectors over the alphabet A^m. $\bar{R}_0(m)$ is defined similarly. As a measure of a channel's information-carrying capabilities, R_0 is second only to capacity in importance. It is the largest number for which there is a linear error exponent, viz., a bound of the form $P_e \lesssim 2^{-n(R_0 - R)}$, relating the error probability of the best code of length n and rate R, which is true for all $R < R_0$. It is the rate beyond which sequential decoding of convolutional codes becomes intractible [6], [19]. And fi-

nally, it is widely believed to be the largest rate at which "practical" coding systems can be implemented (Massey [8], Viterbi [17]).

In our calculations of R_0 for interference channels, we begin with $\bar{R}_0(1)$, the easiest case. We denote the j-function for $\bar{\Delta}^1$ by $\bar{j}$, and the corresponding functions for the component channels $\{\Delta_s\}$ by $\{j_s\}$. The key result is the following

$$\bar{j}(x_1, x_2) = E\{j_S(x_1, x_2)\},   (4.6)$$

where the expectation is with respect to the probability distribution P on the parameter space Ω. This can be proved as follows, provided the output alphabet B and the parameter space Ω are finite (or countably infinite):

$$\bar{j}(x_1, x_2) = \sum_{(y,s)} \sqrt{p(y, s|x_1) p(y, s|x_2)}$$
$$= \sum_s \sqrt{p(s|x_1) p(s|x_2)} \sum_y \sqrt{p(y|x_1, s) p(y|x_2, s)}.$$

From the independence assumption, $p(s|x) = p(s)$ so the above is

$$\sum_s p(s) \sum_y \sqrt{p(y|x_1, s) p(y|x_2, s)} = E\{j_S(x_1 x_2)\}.$$

If now X has an input distribution on A that achieves $\bar{R}_0(1)$, then

$$2^{-\bar{R}_0(1)} = \bar{J}(X)$$
$$= E\{J(X)\}   \text{(from (4.6))}$$
$$\geqslant E\{2^{-R_{0,s}}\},   \text{(from (4.5))},   (4.7)$$

where $R_{0,s}$ is the cutoff parameter for the component channel Δ_s. On the other hand, if X has an input distribution that simultaneously achieves R_0 for each of the component channels (cf. assumption (4.1)), we have from (4.3) $\bar{J}(X) = E(J_S(X)) = E(2^{-R_{0,s}})$, and so by (4.4), $2^{-R_0(1)} \leqslant E\{2^{-R_{0,s}}\}$. Comparing this to (4.7), we get our main result, viz.,

$$2^{-\bar{R}_0(1)} = E\{2^{-R_{0,s}}\},   (4.8)$$

or

$$\bar{R}_0(1) = -\log_2 E\{2^{-R_{0,s}}\}.   (4.9)$$

This result should be compared to (3.7). We see then that the quantity 2^{-R_0}, rather than R_0 itself, behaves in a manner similar to channel capacity for the channel $\bar{\Delta}^1$.

We now turn to the general case $\bar{\Delta}^m$ of an interference channel with side information. If we denote the j-function for $\bar{\Delta}^m$ by $\bar{j}^{(m)}$, and the j-functions for the mth memoryless extension of Δ_s by $j_s^{(m)}$, (4.6) implies that for $x_1, x_2 \in A^m$

$$\bar{j}^{(m)}(x_1, x_2) = E\{j_S^{(m)}(x_1, x_2)\}.   (4.10)$$

Thus the same argument that led to (4.7) gives

$$2^{-m\bar{R}_0(m)} \geqslant E\{2^{-R_{0,s}^{(m)}}\},   (4.11)$$

where $R_{0,s}^{(m)}$ denotes the cutoff parameter for Δ_s^m. From known results for parallel channels, (see [6, eq. (5.6.59) and exercise 5.30]), R_0 for Δ_s^m is equal to $mR_{0,s}$, and is achieved

by an input vector $X = (X_1, X_2, \cdots, X_m)$ of independent copies of the input X achieving $R_{0,s}$. Thus (4.11) becomes

$$2^{-m\bar{R}_0(m)} \geqslant E\{2^{-mR_{0,s}}\}. \qquad (4.12)$$

Furthermore if $X = (X_1, \cdots, X_m)$ consists of m independent copies of the optimizing input X, we have from (4.6) that $\bar{J}(X) = E\{2^{-mR_{0,s}}\}$, and this implies the inequality opposite to (4.11). Hence

$$2^{-m\bar{R}_0(m)} = E\{2^{-mR_{0,s}}\}, \qquad (4.13)$$

or

$$\bar{R}_0(m) = -\frac{1}{m}\log_2 E\{2^{-mR_{0,s}}\}. \qquad (4.14)$$

Here again our result is not exactly new, since both Bernstein [2] and Viterbi [16] have obtained similar results for special kinds of BI channels. We emphasize, however, that the explicit formula (4.14) holds only when side information is present since only then can the BI channel be decomposed into m parallel channels. In Section III (3.7), we saw that the capacity $\bar{C}(m)$ of a BI channel with side information is equal to the average $\bar{C}$ of the capacities of the component channels. For the cutoff rate, (4.14) shows that the situation is more complicated. Since 2^{-x} is convex $\cup$, Jensen's inequality, applied to (4.14) gives the upper bound

$$\bar{R}_0(m) \leqslant \bar{R}_0 \triangleq E(R_{0,s}). \qquad (4.15)$$

However, a study of (4.14) for various special cases indicates that $\bar{R}_0(m)$ is a decreasing function of m. We have no proof of this in general, but for a numerical example, see Table I in Section V, below. It is in any event easy to identify the limiting value of $\bar{R}_0(m)$:

$$\lim_{m\to\infty} \bar{R}_0(m) = R_{0,\inf}, \text{ where}$$

$$R_{0,\inf} = \text{ess. inf}\{R_{0,s}\}. \qquad (4.16)$$

(See Appendix A.) In particular if there are only finitely many component channels, each of which occurs with positive probability, we have

$$\lim_{m\to\infty} \bar{R}_0(m) = \min_s R_{0,s}. \qquad (4.17)$$

We find these results rather surprising, in view of the fact that capacity is a nondecreasing function of m.

We turn next to the cutoff rate for the BI channel without side information. We have no explicit formula analogous to (4.14), only the expected bound

$$R_0(m) \leqslant \bar{R}_0(m). \qquad (4.18)$$

This result follows immediately from the following inequality relating the j-function of the channels Δ^m and $\bar{\Delta}^m$, denoted by $j^{(m)}$ and $\bar{j}^{(m)}$:

$$\bar{j}^{(m)}(x_1, x_2) \leqslant j^{(m)}(x_1, x_2), \qquad (4.19)$$

which is a consequence of the Cauchy–Schwarz inequality. The following argument proves (4.19) for $m = 1$ and

countable Ω and B:

$$\bar{j}(x_1, x_2)$$
$$= \sum_s p(s)\sum_y p(y|x_1, s)^{1/2} p(y|x_2, s)^{1/2}$$

$$\text{(by (4.6).}$$

$$= \sum_y \sum_s (p(s)p(y|x_1, s))^{1/2}(p(s)p(y|x_2, s))^{1/2}$$

$$\leqslant \sum_y \left(\sum_s p(s)p(y|x_1, s)\right)^{1/2}$$

$$\cdot \left(\sum_s p(s)p(y|x_2, s)\right)^{1/2} \quad \text{(Cauchy–Schwarz)}$$

$$= \sum_y p(y|x_1)^{1/2} p(y|x_2)^{1/2}$$

$$\text{(because } p(s) = p(s|x_i)\text{)}$$

$$= j(x_1, x_2).$$

The proof of the more general assertion (4.19) follows the same lines.

It is likely, in view of (4.18) and (4.14), that $R_0(m)$ is also a decreasing function of m, and calculations with special cases support this conjecture. However, we have no proof of this fact, and except in the case where $R_{0,\inf} = 0$, where we obviously have

$$\lim_{m\to\infty} R_0(m) = 0, \qquad (4.20)$$

we do not even know the limiting value of $R_0(m)$. We would be surprised, however, if the limit turned out to be different from that of $\bar{R}_0(m)$.

In the next section, we present numerical results for the interference channel of Example 1a.

V. NUMERICAL RESULTS: A DISCUSSION OF INTERLEAVING

In this section we shall illustrate our results by making a detailed study of the BI channel of Example 1a, which was introduced in Section II. We present first in Table I various values of C and R_0 for $\epsilon = 0.1$ and $m = 2^k$, $k = 0, 1, 2, \cdots, 10$. We note that both Δ_0 and $\Delta_{1/2}$ are symmetric and so the compatibility assumptions (3.1) and (4.1) are satisfied. Here are some further notes on the calculations.

$\bar{R}_0(m), \bar{C}(m)$: Since $C_0 = 1$, $C_{1/2} = 0$, we have, according to (3.11)

$$\bar{C}(m) = \bar{C} = 1 - \epsilon, \quad \text{for all } m. \qquad (5.1)$$

Similarly $R_{0,0} = 1$, $R_{0,1/2} = 0$, and so by (4.14)

$$\bar{R}_0(m) = -\frac{1}{m}\log_2\{(1-\epsilon)2^{-m} + \epsilon\}$$

$$= -\frac{\log_2\epsilon}{m} + O(2^{-m}), \quad m \to \infty. \qquad (5.2)$$

$\bar{R}_0(m)$ is a decreasing function of m and, in agreement with (4.17), we have $\lim_{m\to 0}\bar{R}_0(m) = 0$.

TABLE I
NUMERICAL VALUES OF $R_0(m)$, $\bar{R}_0(m)$, $C(m)$, $\bar{C}(m)$ FOR A
SPECIFIC BI CHANNEL

m	R_0	$\bar{R}_0$	C	$\bar{C}$
1	0.47805	0.86250	0.71360	0.9000
2	0.47783	0.81074	0.74841	0.9000
4	0.45193	0.66952	0.79622	0.9000
8	0.35444	0.40901	0.84199	0.9000
16	0.20552	0.20761	0.87069	0.9000
32	0.10381	0.10381	0.88534	0.9000
64	0.05191	0.05191	0.89267	0.9000
128	0.02595	0.02595	0.89634	0.9000
256	0.01298	0.01298	0.89817	0.9000
512	0.00649	0.00649	0.89908	0.9000
1024	0.00324	0.00324	0.89954	0.9000

$R_0(m)$, $C(m)$: For purposes of computation, the channel Δ^m can be viewed as a DMC with input and output alphabets both equal to the set $\{0,1\}^m$ of binary m-tuples. The transition probabilities are given by (2.2). Since the channel is symmetric, capacity is achieved by a uniform input and a straightforward calculation yields

$$C(m) = (1 - \epsilon + \epsilon \cdot 2^{-m})$$

$$- \frac{1}{m} \{ H(\epsilon - \epsilon \cdot 2^{-m})$$

$$+ (\epsilon - \epsilon 2^{-m}) \log(1 - 2^{-m}) \}$$

$$= (1 - \epsilon) - \frac{H(\epsilon)}{m} + O(2^{-m}), \quad \epsilon \neq 0, 1.$$

$$(5.3)$$

Thus as predicted by (3.9), the capacity without side information does approach the capacity with side information, although the rate of convergence is rather slow. This $O(1/m)$ convergence of $C(m)$ to $\bar{C}$ is typical when the random variable S has finite (absolute) entropy as follows from the discussion following (3.12). When $H(S) = \infty$, the rate of convergence is typically $O(\log m/m)$.

The calculation of $R_0(m)$ is algebraically messy, but straightforward. Omitting the details, we have

$$R_0(m) = -\frac{1}{m} \log_2(J_0),$$

where

$$\bar{J}_0 = \frac{2^m - 1}{2^m} \cdot \bar{J} + \frac{1}{2^m},$$

and

$$J = \frac{2^m - 2}{2^m} \cdot \epsilon + \frac{2}{2^m} \left(\epsilon(\epsilon + 2^m(1 - \epsilon))^{1/2} \right). \quad (5.4)$$

Asymptotically,

$$R_0(m) = -\frac{\log_2 \epsilon}{m} + O(2^{-m/2}),$$

and of course (see (4.20)),

$$\lim_{m \to \infty} R_0(m) = 0.$$

The values of C and R_0 for our channels only indicate possible rates for reliable communication. To design practical systems for these channels, we must study the perfor-

mance of explicit coding schemes. If we chose to view Δ^m or $\bar{\Delta}^m$ as DMC's, the coding alphabet A^m is very large and the prospects of devising practical codes using such a large alphabet are rather poor for large m. On the other hand, motivated by practical experience with real bursty channels, we might try to communicate over Δ^m or $\bar{\Delta}^m$ by *interleaving* codes over the basic alphabet A.

An interleaved coding strategy is one in which a finite-state device, called an *interleaver*, is inserted between the encoder and the channel. In the present content, its function is to introduce varying delays between successive encoded symbols, so that no two letters in any codeword will reside in the same m-letter transmitted block. At the receiving end, there is a *deinterleaver*, between the channel and the decoder, that inverts the action of the interleaver and presents to the decoder (noisy versions of) the codewords in the order they were generated by the encoder. The interleavers/deinterleavers are inserted by the system designers so that they may ignore the channel memory,[1] and code for the memoryless channels Δ^1 or $\bar{\Delta}^1$. Now according to (3.8) and (3.9) the capacity $C(m)$ tends to increase with m, (see also Table I), and so one would predict interleaved coding strategies to be inferior, perhaps markedly so, *if side information is absent*, to strategies which deal more directly with the burst noise. On the other hand, according to (3.7), $\bar{C}(m) = \bar{C}(1)$ for all m, so that, *in the presence of side information*, one would not predict a performance penalty if interleaving is employed.

To test these predictions, we have evaluated the performance of several practical coding schemes on the channels Δ^m and $\bar{\Delta}^m$ of Example 1a, (see Fig. 1), for various values of ϵ, and displayed our results in Figs. 2 and 3. In these figures we have plotted the decoded bit error probability vertically, and "E/N_J", which is defined as $10 \log_{10} 1/\epsilon$, horizontally. (This definition is motivated by the fact that in the presence of partial-band or pulsed jamming interference, the worst-case jamming "duty factor" ϵ is typically inversely proportional to E/N_J, the signal-to-jamming noise ratio [13], [17]. It is also motivated by our desire to have performance curves of the same general shape as usually seen in the communications literature.)

We have considered two general types of codes: *Reed–Solomon (RS) codes*, which correct burst errors directly, and *binary convolutional codes*, which must be interleaved to correct bursts.

First consider the Reed–Solomon codes. We have studied the performance of two such codes, both with rate approximately $\frac{1}{2}$, a (31,15) RS code over the alphabet GF (2^5), and a (255,127) RS code over GF (2^8). The first of these codes can be used directly on either of the channels Δ^5 or $\bar{\Delta}^5$, since each code symbol is from a 32-letter alphabet, and the second can be used on Δ^8 or $\bar{\Delta}^8$. If $m \gg 10$, the use of

---

[1] Of course the "interleaved" channel (from the input to the interleaver to the output of the deinterleaver) is not memoryless at all and in fact has exactly the same capacity as the original channel! However, since the users of the interleaved channel deliberately ignore the channel memory, in what follows we will treat the interleaved channel as if it were, in fact, the memoryless channel Δ^1 or $\bar{\Delta}^1$.
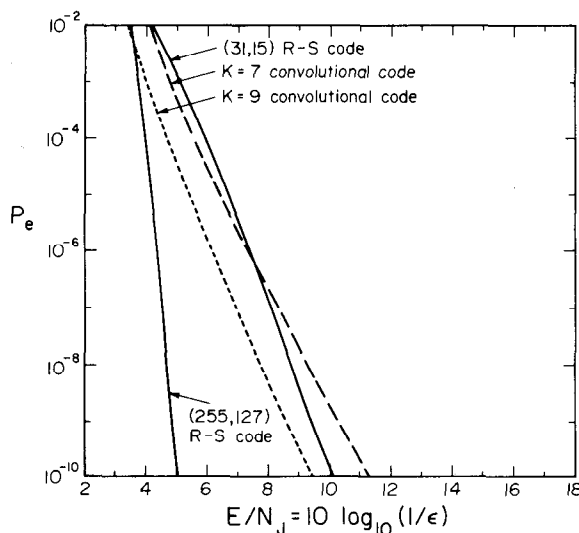
Fig. 2. Error probability of Reed–Solomon codes (solid lines) and convolutional codes (dashed lines) on the interference channel of Example 1a (side information available).
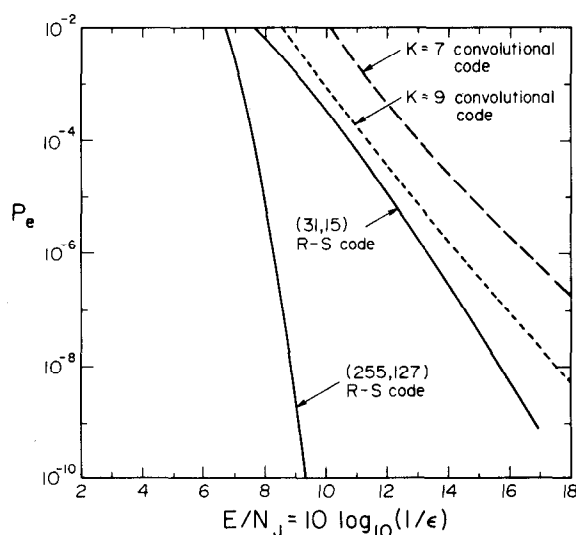

Fig. 3. Error probability of Reed–Solomon codes (solid lines) and convolutional codes (dashed lines) on the interference channel of Example 1a (side information not available).

RS codes directly on Δ^m or $\overline{\Delta}^m$ becomes impractical. In Fig. 2, we have plotted (using solid lines) the performance of these two codes on the BI channel with side information, and in Fig. 3, the performance on the channel without side information. (We used the techniques described in [1] to obtain these curves.)

We next consider the convolutional codes. We have studied in detail two such, both of rate $\frac{1}{2}$, one of constraint length 7 ($\nu = 6$), and one of constraint length 9 ($\nu = 8$), as given in [4, table B-2]. Since these codes are not designed to deal directly with bursts, if they are used on Δ^m or $\overline{\Delta}^m$, we imagine them to be interleaved to depth m, which means that they are actually being used on the channels Δ^1 or $\overline{\Delta}^1$. In Fig. 2, we have plotted (using dashed lines), the performance of these two codes on $\overline{\Delta}^1$, and in Fig. 3, the performance on Δ^1. (We used the transfer function upper bounds described in [4] to obtain these curves. Experience shows that these bounds are quite tight in the range $P_e < 10^{-3}$, but cannot be relied upon for larger values of P_e.)

What conclusions do we draw from these figures? Consider first Fig. 2, where side information is present. We know from Section III that when side information is present, channel capacity is independent of m, and we have predicted that the interleaved convolutional codes would be at no disadvantage versus the RS codes. We claim that Fig. 2 supports this prediction: the $K = 7$ convolutional code is about as good as, and the $K = 9$ code is better than, the (31,15) $m = 5$ RS code. All three codes have rate $\frac{1}{2}$, and in current hardware the complexity of all three is roughly comparable. The larger RS code is superior to all three, but we attribute this more to the increased code complexity than to the somewhat increased memory ($m = 8$ versus $m = 5$).

When we move to Fig. 3, however, where side information is *absent*, but everything else is the same, the situation has changed significantly. All four codes now have worse

performance, as expected, but the interleaved convolutional codes have suffered more than the RS codes. Now the (31,15) RS code is superior to *both* convolutional codes; again, this supports our prediction, based on the capacity results of Section III, that interleaved strategies will be inferior if side information is absent.

Although the above example is of a channel model that is highly simplified, we believe that the quantitative conclusions drawn from this channel are applicable to more sophisticated and realistic channel models.

VI. Test Patterns

Before leaving the subject of interleaving we would like to describe a modification which can be used on the channel Δ^m, and which for large values of m, makes this channel almost as good as $\overline{\Delta}^m$. The idea is to attach to each transmitted packet a fixed "test pattern," and on the basis of the received version of the test pattern to make a statistical decision about which of the component channels was used to transmit the packet. If the test pattern is of length say $\log m$, then for large values of m one would expect this "noise estimate" to be increasingly reliable, and yet the fraction of the transmitted letters devoted to the test pattern is quite small. What this means is that, for large m, the channel Δ^m behaves as if the side information were available, and so interleaving should not cause a severe degradation in performance.

To illustrate this idea, we again use Example 1a, with now an all-zeros test pattern of length t within each transmitted packet. If the received test pattern is not all zeros, the entire packet is *erased*; if it is all zeros, the packet is accepted as correct. What this means is that, after interleaving, the channel Δ^m becomes a binary symmetric erasures-and-errors channel with erasure probability $p = \epsilon(2^t - 1)/2^t$, and error probability $q = \epsilon 2^{-t}$. The cutoff

TABLE II
R_0 AND C FOR AN INTERFERENCE CHANNEL WITHOUT SIDE INFORMATION
USING TEST PATTERN AND INTERLEAVING

m	$\tilde{R}_0[t_{opt}]$	$\tilde{C}[t_{opt}]$
1	0.4781 [0]	0.7136 [0]
2	0.4781 [0]	0.7136 [0]
4	0.4781 [0]	0.7136 [0]
8	0.4923 [1]	0.7136 [0]
16	0.5622 [3]	0.7343 [1]
32	0.6506 [5]	0.7781 [3]
64	0.7260 [7]	0.8214 [4]
128	0.7794 [9]	0.8519 [5]
256	0.8138 [12]	0.8716 [6]
512	0.8347 [14]	0.8837 [8]
1024	0.8469 [16]	0.8909 [9]

TABLE III
R_0 AND C FOR FIVE DIFFERENT COMBINATIONS OF SIDE INFORMATION
AND INTERLEAVING, FOR Δ^{128} (EXAMPLE 1a, = 0.1)

Option	R_0	C
No side information, no interleaving	0.0259	0.8963
No side information, "normal" interleaving	0.4780	0.7136
No side information, "smart" interleaving	0.7794	0.8519
Side information, no interleaving	0.0259	0.9000
Side information, interleaving	0.8625	0.9000

rate for this channel is

$$R_0(p,q) = 1 - \log_2\left(1 + p + 2\sqrt{q(1 - p - q)}\right),$$

and so R_0 for the channel Δ^m when depth $(m - t)$ interleaving is employed together with this "noise detection" procedure is given by

$$\tilde{R}_0(m) = \max_{0 \leqslant t \leqslant m} \left(1 - \frac{t}{m}\right) R_0(p,q). \qquad (6.1)$$

The maximization in (6.1) is over all possible test pattern lengths, and the factor $(1 - t/m)$ reflects the rate loss due to the presence of the test pattern.

Similarly the capacity of the above erasures-and-errors channel is

$$C(p,q) = (1 - p)(1 - H_2(q/(1 - p))),$$

and so the capacity of Δ^m, when this procedure is adopted, is

$$\tilde{C}(m) = \max_t \left(1 - \frac{t}{m}\right) C(p,q).$$

Table II gives numerical values for these parameters for the same channels of Table I. In Table II it is seen that, for small values of m (up to about $m = 8$), the optimal test pattern length is $t = 0$, i.e., no test pattern should be used. However, for larger m's, the test pattern does help, and indeed as $m \to \infty$, $\tilde{R}_0$ appears to be, and in fact is, approaching the capacity 0.9000 of $\overline{\Delta}^m$. If R_0 is in some sense a practical measure of the channel's quality, this indicates that for large m, the "smart" interleaving idea of the previous section makes Δ^m a very tractable channel for coding.

To further illustrate our ideas, we conclude this section with a table (Table III), again using Example 1a, $\epsilon = 0.1$, and $m = 128$, giving the values of R_0 and C for five

different combinations of side information and interleaving.

VII. CONCLUDING DISCUSSION

For us, our most interesting result is the divergent behavior of C and R_0. In the discussion that follows, we will use Example 1a to illustrate the point, but qualitatively similar results should hold for all BI channels.

According to (5.2), the cutoff parameter for this particular interference channel with side information decreases to 0 as $m \to \infty$. Since $\overline{C}(m)$ is by (3.1) a constant independent of m we have the peculiar situation that

$$\lim_{m \to \infty} \frac{\overline{C}(m)}{\overline{R}_0(m)} = \infty. \qquad (7.1)$$

We say peculiar because, both C and R_0 are believed to be measures of the channel's quality, and yet as the burst length m of our interference channel increases, these measures diverge. For the BI channel without side information, the situation is if anything even more puzzling (see Table I). Here $C(m)$ is a strictly *increasing* function of m, while $R_0(m)$ is a strictly *decreasing* function of m.

Of course the fact that channel memory tends to increase capacity is a well-known, if somewhat vague, folk theorem (see, e.g., Gallager [6, sec. 6.10]). The fact that channel memory tends to *decrease* R_0 is less well-known, but has been noted before [3], [16], [17] with side information present.

This issue is of more than academic interest, because depending on which figure of merit (C or R_0) one prefers, one reaches totally different conclusions about the merits of interleaving. For example, Viterbi [17] cites the decrease of $\overline{R}_0(m)$ with m as evidence in favor of interleaving. Since his models have side information, we cannot quite dispute this conclusion. On the other hand, in Section V, we saw clear evidence that, when side information is not available, interleaving definitely causes performance degradation.

How can one reconcile these apparently contradictory facts? We feel that if one takes the view that R_0 is an inverse measure of the coding *delay*, rather than the *complexity*, required to achieve a given performance, the data become comprehensible. Suppose, for example, that one can achieve a given bit error probability and rate with delay D on the channel $\overline{\Delta}^1$. Then exactly the same performance can be achieved on $\overline{\Delta}^m$, with delay $D \cdot m$, by interleaving m copies of the code used on $\overline{\Delta}^1$. Thus interpreting R_0 as inverse measure of the needed delay we could predict $\overline{R}_0(m) \geqslant (1/m)\overline{R}_0(1)$, and indeed the data in Table I satisfy this inequality. Indeed, according to (5.2) for Example 1a,

$$\overline{R}_0(m) \sim \frac{-\log \epsilon}{m},$$

so we have

$$\frac{\overline{R}_0(1)}{\overline{R}_0(m)} \sim f(\epsilon) \cdot m,$$

where $0 < f(\epsilon) < \frac{1}{2}$ is a function depending on ϵ but not

m. Similar but computationally messier results for Δ^m confirm these observations. We thus conclude that C, not R_0, is the better measure of a BI channel's quality, unless one is under severe constraints of coding delay or interleaver size. Perhaps future researchers will understand these issues better and express themselves more articulately.

ACKNOWLEDGMENT

This paper has benefited enormously from the extraordinarily thorough and patient work of the referees. We wish to thank them for saving us from making numerous embarassing blunders.

APPENDIX

Proof of (4.16)

We have, according to (4.13),

$$2^{-m\bar{R}_0(m)} = \int_\Omega 2^{-mR_{0,s}}\, dP.$$

Let x_0 denote the essential infimum of $\{R_{0,s}\}$, and suppose $R_{0,s_0} = x_0$ (if there is no such value, an obvious modification is required). Then for any $\epsilon > 0$ there is a neighborhood Ω_ϵ of s_0 with positive probability such that $x_0 \leqslant R_{0,s} \leqslant x_0 + \epsilon$ for $s \in \Omega_\epsilon$ and $R_{0,s} > x_0 + \epsilon$ for $s \notin \Omega_\epsilon$. Then

$$\int_\Omega 2^{-mR_{0,s}}\, dP = \int_{\Omega_\epsilon} + \int_{\Omega_\epsilon'}.$$

The first integral is bounded above by $P\{\Omega_\epsilon\}2^{-mx_0}$ and below by $P\{\Omega_\epsilon\}2^{-m(x_0+\epsilon)}$. The second integral is bounded above by $2^{-m(x_0+\epsilon)}$. Hence

$$2^{-m(x_0+\epsilon)} \leqslant 2^{-m\bar{R}_0(m)} \leqslant P\{\Omega_\epsilon\}2^{-mx_0} + 2^{-m(x_0+\epsilon)}.$$

Taking logarithms and limits, we get

$$x_0 \leqslant \varliminf_{m\to\infty} \bar{R}_0(m) \leqslant \varlimsup_{m\to\infty} \bar{R}_0(m) \leqslant x_0 + \epsilon.$$

Since ϵ was arbitrary, this completes the proof.

REFERENCES

[1] E. R. Berlekamp, "The technology of error-correcting codes," *Proc. IEEE*, vol. 68, May 1980.

[2] S. L. Bernstein, *et al.*, "Variable rate Viterbi decoding in the presence of RFI," *Nat. Telecommun. Conf. Rec.*, pp. 36.6.1–5, Dec. 1977.

[3] J. B. Cain and J. M. Geist, "Interleaving considerations for coding on Gaussian noise channels with burst erasures," *National Telecommun. Conf. Rec.*, pp. 44.5.1–5, Nov. 1979.

[4] G. C. Clark and J. B. Cain, *Error-Correcting Coding for Digital Communications.* New York: Plenum, 1981.

[5] T. S. Ferguson, *Mathematical Statistics.* New York: Academic, 1967.

[6] R. G. Gallager, *Information Theory and Reliable Communication.* New York: Wiley, 1968.

[7] J. L. Massey, "Capacity, cutoff rate, and coding for a direct-detection optical channel," *Jet Propulsion Laboratory Deep Space Network Progress Report*, 42-60, pp. 68–74, 1980.

[8] ——, "Coding and modulation in digital communications," in *Proc. Int. Zurich Sem. Digital Commun.*, Zurich, Switzerland, Mar. 1974.

[9] R. J. McEliece, "Practical codes for photon communication," *IEEE Trans. Inform. Theory*, vol. IT-27, pp. 393–398, July 1981.

[10] ——, "RFI channels," *Jet Propulsion Laboratory Deep Space Network Progress Report*, 42-60, pp. 103–106, 1980.

[11] ——, "RFI channels, II," *Jet Propulsion Laboratory Deep Space Network Progress Report*, 42-65, pp. 75–80, 1981.

[12] ——, *The Theory of Information and Coding.* Reading, MA: Addison-Wesley, 1977.

[13] R. J. McEliece and W. E. Stark, "An information theoretic study of communication in the presence of jamming," *IEEE Int. Conf. Commun., Conf. Rec.*, pp. 45-3.1–5, June 1981.

[14] J. K. Omura, "Expurgated bounds, Bhattacharyya distance, and rate distortion functions," *Inform. Contr.*, vol. 24, pp. 358–383, 1974.

[15] M. S. Pinsker, *Information and Information Stability of Random Variables and Processes.* San Francisco: Holden-Day, 1964.

[16] A. J. Viterbi, "R_0 Parameter for a pulse-burst channel," presented at the 1979 Int. Symp. Inform. Theory, June 1979.

[17] ——, "Spread-spectrum communications—Myths and realities," *IEEE Commun. Mag.*, vol. 4, pp. 11–18, May 1979.

[18] J. Wolfowitz, *Coding Theorems of Information Theory*, 3rd ed. Berlin: Springer-Verlag, 1978.

[19] J. M. Wozencraft and I. M. Jacobs, *Principles of Communication Engineering.* New York: Wiley, 1965.

[20] A. Wyner, "A definition of conditional mutual information for arbitrary ensembles," *Inform. Contr.*, vol. 38, pp. 51–59, 1978.