RFI Channels, II

R. J. McEliece
Communications Systems Research Section

Continuing a previous study, we calculate the cutoff parameters for a class of channel models exhibiting burst noise behavior and evaluate the performance of interleaved coding strategies. We conclude that, provided the channel memory is large enough and is properly exploited, interleaved coding is nearly optimal.

I. Introduction

In Ref. 1 we introduced a general class of discrete channels, which we called RFI channels. These channels, which were motivated by an earlier study (Ref. 2) of pulse-position modulation in optical channels, exhibit a simplified kind of burst-noise behavior, and in Ref. 3 we calculated their channel capacities. In this paper, after reviewing our previous results in Section II, we continue our analysis of RFI channels, as follows. In Section III we give formulas for R_0, the cutoff parameter for these channels. This parameter is considered by many engineers to be a more meaningful measure of the channel's quality than capacity, and its behavior on RFI channels is quite interesting. In Section IV we begin to deal with the practical problems of coding for RFI channels by considering the merits of interleaved codes. In Section V we give a numerical calculation to illustrate our results and state our conclusions.

II. Review of Previous Results

We start with a set $\{\zeta_1, \zeta_2, \ldots, \zeta_K\}$ of K discrete memoryless channels, each with the same input alphabet A, and output alphabet B, and a probability vector $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_K)$, with K nonzero components. For each positive integer b, we define two *mixture channels* $\zeta(b)$ and $\overline{\zeta}(b)$, as follows. When a sequence of letters from A is to be sent over $\zeta(b)$, each block of b consecutive letters (such blocks we call *packets*) is transmitted over one of the auxiliary channels ζ_k. Which channel is selected to transmit a given packet is determined by an external random variable Z, which is described statistically by $Pr\{Z = k\} = \alpha_k$. Informally we think of Z as determining the noise *level* and b (the *burst length*) as the length of time that a given noise level persists. Formally, the channel $\zeta(b)$ can be viewed as an orthodox discrete memoryless channel (DMC) with input alphabet A^b, output alphabet B^b, and transition probabilities

$$p(\underline{y}|\underline{x}) = \sum_{k=1}^{K} \alpha_k \prod_{i=1}^{b} p_k(y_i|x_i)$$

where $\underline{y} = (y_1, y_2, \ldots, y_b)$, $\underline{x} = (x_1, x_2, \ldots, x_n)$, and $p_k(y|x)$ denotes the transition probability function of the channel ζ_k.

The second RFI channel $\overline{\zeta}(b)$, which might be called *the channel with side information*, is identical to $\zeta(b)$ except that the channel provides to the user along with every packet the index of the channel used to transmit that packet. Informally, this side information can be thought as being provided by a noise level detector, perhaps an automatic gain control device.

Formally $\bar{\zeta}(b)$ is a DMC with input alphabet A^b, output alphabet $B^b \times \{1, 2, \ldots, K\}$, and transition probabilities

$$\bar{p}(\underline{y}, k|\underline{x}) = \alpha_k \prod_{i=1}^{b} p_k(y_i|x_i) .$$

Let us denote the capacities of $\zeta(b)$ and $\bar{\zeta}(b)$ by $C(b)$ and $\bar{C}(b)$, respectively. The main results of Ref. 1 can be summarized as follows. First, $\bar{C}(b)$ is independent of b, and under a mild additional hypothesis* we have the formula

$$\bar{C}(b) = \bar{C} = \sum_{k=1}^{k} \alpha_k C_k ,$$

where C_k denotes the capacity of ζ_k. Second, $C(b)$ is less than $\bar{C}(b)$, in general strictly less, and in fact

$$\bar{C}(b) - \frac{\log K}{n} \leqslant C(b) \leqslant \bar{C}(b)$$

It follows in particular that $\lim_{b\to\infty} C(b) = \bar{C}$.

III. The Calculation of R_0

For a general DMC, the calculation of R_0 depends on the function $j(x_1, x_2)$, defined for pairs of input letters:

$$j(x_1, x_2) = \sum_{y \in B} p(y|x_1)^{1/2} \, p(y|x_2)^{1/2} .$$

If then X is any random variable taking values in the input alphabet A we define $J(X) = E[j(X_1, X_2)]$, where X_1 and X_2 are independent random variables with the same distribution as X. Finally,

$$R_0 = \max_X \{-\log J(X)\} .$$

We now consider the channel $\bar{\zeta}(1)$, which is the easiest case. Here the output alphabet is $B \times \{1, 2, \ldots, K\}$. Let us denote the function j for this channel by $\bar{j}$, and the corresponding functions for the auxiliary channels $\{\zeta_k\}$ by $\{j_k\}$.

*This hypothesis is that the *same* input distribution can be used to achieve channel capacity on each of the auxiliary channels. Throughout the paper we will describe this state of affairs by saying the channels $\{\zeta_i\}$ are *compatible*.

Lemma 1:

$$\bar{j}(x_1, x_2) = \sum_{k=1}^{K} \alpha_k j_k(x_1, x_2) .$$

Proof:

$$\begin{aligned}
\bar{j}(x_1, x_2) &= \sum_{y,k} \bar{p}(y, k|x_1)^{1/2} \, \bar{p}(y, k|x_2)^{1/2} \\
&= \sum_{y} \sum_{k} [\alpha_k p_k(y|x_1)]^{1/2} \, [\alpha_k p_k(y|x_2)]^{1/2} \\
&= \sum_{k} \alpha_k \sum_{y} p_k(y|x_1)^{1/2} \, p_k(y|x_2)^{1/2} \\
&= \sum_{k} \alpha_k j_k(x_1, x_2) .
\end{aligned}$$

QED.

Corollary 1:

$$\bar{R}_0(1) \leqslant -\log_2 \sum_k \alpha_k 2^{-R_0^{(k)}} ,$$

and equality holds if the auxiliary channels are compatible.

Proof: Let X achieve $\bar{R}_0(1)$. Then

$$\begin{aligned}
2^{-\bar{R}_0(1)} &= \bar{J}(X) \\
&= \sum \alpha_k J_k(X) \qquad \text{by Lemma 1} \\
&\geqslant \sum \alpha_k 2^{-R_0^{(k)}}
\end{aligned}$$

On the other hand, if the channels are compatible, and if X simultaneously achieves R_0 for each of the auxiliary channels, we have

$$\bar{J}(X) = \sum_k \alpha_k 2^{-R_0^{(k)}}$$

and so

$$\bar{R}_0(1) \geqslant -\log_2 \sum_k \alpha_k 2^{-R_0^{(k)}}, \text{ as well.}$$

QED.

Corollary 2:

$$\bar{R}_0(b) \leqslant -\frac{1}{b} \log_2 \sum_{k=1}^{K} \alpha_k 2^{-bR_0^{(k)}},$$

with equality if the channels are compatible.

Proof: An examination of the proof of Lemma 1 shows that if we denote the j-function for $\bar{\zeta}(b)$ by $\bar{j}^{(b)}$, we have

$$\bar{j}^{(b)}(\underline{x}_1, \underline{x}_2) = \sum_k \alpha_k j^{(b)}(\underline{x}_1, \underline{x}_2),$$

here $j^{(b)}$ denotes the j-function for b parallel copies of ζ_k. Arguing as in the proof of Corollary 1, we get the desired inequality. It is possible to show (see Ref. 3, p. 150, Eq. 5.6.59) that R_0 for b parallel copies of ζ_k is exactly b times the R_0 for ζ_k, and is achieved by an input $(X_1, X_2, \ldots, X_b)$ of independent input random variables, each distributed according to the input that achieves $R_0^{(k)}$. Thus if the channels are compatible and we choose $\underline{X} = (X_1, X_2, \ldots, X_b)$, we get

$$\bar{j}^{(b)}(\underline{X}) = \sum \alpha_k 2^{-bR_0^{(k)}}$$

and so (remembering to divide by b),

$$\bar{R}_0(b) \geqslant -\frac{1}{b} \log_2 \sum_k \alpha_k 2^{-bR_0^{(k)}},$$

which combined with the opposite inequality (which is true in general) yields the desired result.

QED.

Corollary 3:

Let

$$R_0^{(min)} = \min_k \{R_0^{(k)} : k = 1, 2, \ldots, K\}.$$

Then

$$\lim_{b\to\infty} \bar{R}_0(b) \leqslant R_0^{(min)},$$

with equality if the channels are compatible.

Proof: This follows immediately from Corollary 2, since the probabilities α_k are all nonzero by assumption.

We close this section with several elementary remarks about $R_0(b)$. We have no simple formula analogous to that of Corollary 2 for $R_0(b)$, but in any given case it is not difficult to compute since, as remarked in Section II, $\bar{\zeta}(b)$ can be viewed as a DMC with alphabets A^b, B^b. If R_0 indeed measures the channel's quality, the side information present in $\bar{\zeta}(b)$ should not *decrease* R_0, and indeed we can prove $R_0(b) \leqslant \bar{R}_0(b)$. This result follows from Lemma 2, which relates the j-functions for $\zeta(b)$ and $\bar{\zeta}(b)$.

Lemma 2:

$$\bar{j}^{(b)}(x_1, x_2) \leqslant j^{(b)}(x_1, x_2).$$

Proof:

$$\begin{aligned}
\bar{j}^{(b)}(x_1, x_2) &= \sum_{\underline{y},k} \bar{p}(\underline{y}, k|\underline{x}_1)^{1/2}\, \bar{p}(\underline{y}, k|\underline{x}_2)^{1/2} \\
&= \sum_{\underline{y}} \sum_k [\alpha_k p_k(\underline{y}|\underline{x}_1)]^{1/2} \cdot [\alpha_k p_k(\underline{y}|\underline{x}_2)]^{1/2} \\
&\leqslant \sum_{\underline{y}} \left[\sum_k \alpha_k p_k(\underline{y}|\underline{x}_1) \cdot \sum_j \alpha_j p_j(\underline{y}|\underline{x}_2)\right]^{1/2} \\
&\qquad \text{(by Schwarz inequality)} \\
&= \sum_{\underline{y}} p(\underline{y}|\underline{x}_1)^{1/2}\, p(\underline{y}|\underline{x}_2)^{1/2} = j^{(b)}(\underline{x}_1, \underline{x}_2)
\end{aligned}$$

QED.

Corrollary 4: $\bar{R}_0(b) \geqslant R_0(b)$.

Proof: Let X achieve $R_0(b)$. Then

$$2^{-R_0(b)} = J(X) \geqslant \bar{J}(X) \geqslant 2^{-\bar{R}_0(b)}.$$

QED.

Corllary 5: If $R_0^{(min)} = 0$, then

$$\lim_{b\to\infty} R_0(b) = \lim_{b\to\infty} \bar{R}_0(b) = 0 .$$

Proof: This follows from Corollary 3 and 4.

When $R_0^{(min)} > 0$, the limits of Corollary 5 are at present unknown to us. Even if the auxiliary channels are compatible, the limit of $R_0(b)$ is as yet unknown, although we conjecture that the two limits are always the same.

IV. A Study of Interleaving

The values of C and R_0 for our channel models only indicate possible ranges of rates for reliable communication. To design a practical system for these channels requires a study of coding. If we choose to view $\zeta(b)$ or $\bar{\zeta}(b)$ as DMCs, the coding alphabet A^b is exponentially large and the prospects of devising practical codes using such a large alphabet are rather poor. On the other hand, motivated by practical experience with real bursty channels, we might try to communicate over $\zeta(b)$ or $\bar{\zeta}(b)$ by *interleaving* codes over the basic alphabet A. Of course if this is done, the channels we are really coding for are $\zeta(1)$ and $\bar{\zeta}(1)$, respectively. Now one normally expects such interleaving to *decrease* channel capacity, and in Ref. 1 we showed that indeed $\zeta(b)$ is an increasing function of b. However, we also showed there that $\bar{\zeta}(b) = \bar{\zeta}(1)$ for all b, so that in the presence of side information apparently no penalty is paid if interleaving is employed. Given the results of Section III, we can now easily describe what happens to R_0 when interleaving is used.

Let us assume for purposes of discussion that the K auxiliary channels are compatible. In that case, according to Corollary 2 in Section III, the value of R_0 for the channel $\bar{\zeta}(b)$ is given by

$$\bar{R}_0(b) = \frac{1}{b}\log_2 \sum_{k=1}^{K} \alpha_k \, 2^{-bR_0^{(k)}} .$$

From this it follows that $\bar{R}_0(b)$ is a decreasing function of b. If we further assume that $R_0^{(min)} = 0$, we have the expression

$$\bar{R}_0(b) \sim \frac{K}{b} ,$$

where $K = -\log_2(\alpha_k)$, k being the index of the auxiliary channel with $R_0 = 0$. Since we already know that under these assumptions the capacity of $\bar{\zeta}(b)$ is a constant independent of b, we have the peculiar situation that

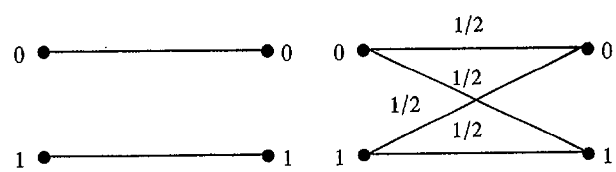
$$\lim_{b\to\infty} \frac{\bar{C}(b)}{R_0(b)} = \infty .$$

We say peculiar because, as we mentioned in the introduction, both C and R_0 are believed to be measures of the channel's quality, and yet as the burst length b of our RFI channels increases, these measures diverge. For the channels $\zeta(b)$ without side information, the situation is if anything even more puzzling. We omit the details, but what happens in general is that $C(b)$ is a strictly increasing function of b, while $R_0(b)$ is a strictly decreasing function of b. We will comment on this apparent paradox in Section V.

Before leaving the subject of interleaving, however, we would like to describe a modification which can be used on the channel $\zeta(b)$, and which for large values of b makes this channel almost as good as $\bar{\zeta}(b)$. The idea, which we first suggested in Ref. 1, 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 K auxiliary channels was used to transmit the packet. If, say, the test pattern is of length $\log b$, then for large values of b 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 b the channel $\zeta(b)$ behaves as if the side information were available, and so interleaving should not cause a severe degradation in performance. In the next section we will illustrate this point with a specific numerical example.

V. A Numerical Example and Some Conclusions

We shall now illustrate our results with a specific example, the same example introduced in Ref. 1, which is indeed the "RFI channel" associated with pulse-position modulation in optical channels (Ref. 2). Here $K = 2$; the auxiliary channel ζ_1 is a noiseless binary symmetric channel, and ζ_2 is a "useless" BSC with transition probability 1/2.



CHANNEL ζ_1 CHANNEL ζ_2

Let us assume that $\alpha_1 = 1 - \epsilon, \alpha_2 = \epsilon$ are the probabilities that the "channel selector" chooses ζ_1 and ζ_2, respectively. We present below a table of the various values of R_0 and C for $\epsilon = 0.1$ and $b = 2^m, m = 0, 1, 2, \ldots, 10$. We now present some notes on the calculations:

1. $\boldsymbol{R_0(b), C(b)}$: For purposes of computation the channel $\zeta(b)$ can be viewed as a DMC with input and output alphabets both equal to the set of binary b-tuples. The transition probabilities are

$$p(y|x) = (1-\epsilon) + 2^{-b}\epsilon \quad y = x$$

$$= 2^{-b}\epsilon \qquad y \neq x .$$

From this it follows that the j-function for $\zeta(b)$ is

$$j(x_1, x_2) = \frac{(2^b - 2)}{2^b}\,\epsilon + \frac{2}{2^b}\{\epsilon\,[\epsilon + 2^b(1-\epsilon)]\}^{1/2} \quad x_1 \neq x_2$$

$$= 1 \qquad x_1 = x_2 .$$

Since $\zeta(b)$ is symmetric, R_0 is achieved for equiprobable inputs, and indeed

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

where

$$J_0 = \frac{2^b - 1}{2^b}\,\overline{J} + \frac{2}{2^b}, \text{ where}$$

$\overline{J}$ is the value of $\overline{j}(x_1, x_2)$ for unequal x's given above. We computed $C(b)$ in Ref. 3; we repeat the formula here:

$$C(b) = (1-\epsilon_b) - \frac{1}{b}[H_2(\epsilon_b) + \epsilon_b \log_2 (1 - 2^{-b})],$$

where $\epsilon_b = (1 - 2^{-b})\epsilon$, and H_2 is the binary entropy function.

2. $\boldsymbol{\overline{R}_0(b), \overline{C}(b)}$: Since both channels ζ_1 and ζ_2 are symmetric, both R_0 and C are achieved by equiprobable inputs, and the channels are compatible in the sense of this paper. Since $R_0^{(1)} = 1$, $R_0^{(2)} = 0$, we have by Corollary 2 of Section III,

$$\overline{R}_0(b) = -\frac{1}{b}\log_2 [(1-\epsilon)\,2^{-b} + \epsilon] .$$

Of course from our previous paper

$$\overline{C}(b) = 1 - \epsilon \text{ for all } b .$$

3. $\boldsymbol{\widetilde{R}_0(b), \widetilde{C}(b)}$: This is a new notation and it refers to the channel $\zeta(b)$ when a specific kind of "smart" interleaving of the general kind described in Section IV is implemented. Here we use an all-zeros test pattern of length t in 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. What this means is that after interleaving the channel $\zeta(b)$ becomes a binary symmetric erasures-and-errors channel with erasure probability $p = \epsilon(2^t - 1)/2^t$ and error probability $q = \epsilon\, 2^{-(t+1)}$. The R_0 for this channel is given by

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

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

$$\widetilde{R}_0(b) = \max_{0 \leq t \leq b} \left(1 - \frac{t}{b}\right) R_0(p, q) .$$

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

Similarly the *capacity* of the above erasures-and-errors channel is given by

$$C(p, q) = (1-p)\log_2 \frac{2}{1-p} - (1-p-q)\log_2 \frac{1}{1-p-q} - q\log_2 \frac{1}{q},$$

and so

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

We now present our table, with $\epsilon = 0.1$.

b	R_0	$\overline{R}_0$	$\widetilde{R}_0[t_{opt}]$	$\widetilde{C}[t_{opt}]$	C	$\overline{C}$
1	0.47805	0.86250	0.4871 [0]	0.7136 [0]	0.71360	0.9000
2	0.47783	0.81074	0.4781 [0]	0.7136 [0]	0.74841	0.9000
4	0.45193	0.66952	0.4781 [0]	0.7136 [0]	0.79622	0.9000
8	0.35444	0.40901	0.4923 [1]	0.7136 [0]	0.84199	0.9000
16	0.20552	0.20761	0.5622 [3]	0.7343 [1]	0.87069	0.9000
32	0.10381	0.10381	0.6506 [5]	0.7781 [3]	0.88534	0.9000
64	0.05191	0.05191	0.7260 [7]	0.8214 [4]	0.89267	0.9000
128	0.02595	0.02595	0.7794 [9]	0.8519 [5]	0.89634	0.9000
256	0.01298	0.01298	0.8138 [12]	0.8716 [6]	0.89817	0.9000
512	0.00649	0.00649	0.8347 [14]	0.8837 [8]	0.89908	0.9000
1024	0.00324	0.00324	0.8469 [16]	0.8909 [9]	0.89954	0.9000

The numbers R_0, $\bar{R}_0$, C, $\bar{C}$ behave as expected, but the behavior of $\tilde{R}_0$, $\tilde{C}$ is rather interesting. For small values of b (up to about $b = 8$), the optimal test pattern length is seen to be $t = 0$; i.e., no test pattern should be used. However, for larger b's the test pattern does help, and indeed as $b \to \infty, \tilde{R}_0$ appears to be, and indeed is, approaching the capacity 0.900 of $\bar{\zeta}(b)$. If R_0 is in some sense a practical measure of the channel's quality, this indicates that for large b, the "smart" interleaving idea makes $\zeta(b)$ a very tractable channel for coding.

To further illustrate our ideas, we next present a table for $b = 128$, $\epsilon = 0.1$ giving the values of R_0 and C for 5 different combinations of side information and interleaving.

Option	R_0	C
No side information, no interleaving	0.02595	0.89634
No side information, "dumb" interleaving	0.47805	0.71360
No side information, "smart" interleaving	0.7794	0.8519
Side information, no interleaving	0.02595	0.9000
Side information, interleaving	0.86250	0.9000

At first we found the fact that interleaving could increase R_0, and increase it dramatically, very puzzling. But if we take the view that R_0 is an inverse measure of the *delay*, rather than the *complexity*, required to achieve a given performance, the data become comprehensible. Suppose, for example, one can achieve a given bit error probability and rate with delay D on the channel $\bar{\zeta}(1)$. Then exactly the *same* performance can be achieved on $\bar{\zeta}(b)$, with delay $D \cdot b$, by interleaving b copies of the code used on $\zeta(1)$. Thus we would predict $\bar{R}_0(b) \geqslant 1/b\ \bar{R}_0(1)$, and indeed the data in the above table satisfy this inequality. Indeed, since as we showed above

$$\bar{R}_0(b) \sim \frac{-\log(\epsilon)}{b} ,$$

we have

$$\frac{\bar{R}_0(1)}{R_0(b)} \sim f(\epsilon) \cdot b$$

where

$$f(\epsilon) = \frac{\log\left(\frac{2}{1+\epsilon}\right)}{\log(\epsilon^{-1})} ,$$

and $0 \leqslant f(\epsilon) \leqslant 1/2$. Similar but computationally messier results for $\zeta(b)$ confirm these observations.

On the basis of this numerical example and several others, we tender the following conclusion. On a general RFI channel, i.e., one that exhibits long periodic bursts of poor data quality, the best coding strategy is probably a "smart" interleaving strategy. By this we mean a strategy that uses the received data to estimate the noise severity, and passes along these estimates to the parallel decoders. In a later paper we hope to verify this conjecture by considering the performance of specific coding schemes on specific RFI channels.

References

1. McEliece, R. J., "RFI Channels," *TDA Progress Report 42-60*, pp. 103-106, Jet Propulsion Laboratory, Pasadena, Calif., Dec. 15, 1980.

2. McEliece, R. J., and Welch, L. R., "Coding for Optical Channels with Photon Counting," *DSN Progress Report 42-52*, pp. 61-66, Jet Propulsion Laboratory, Pasadena, Calif., Aug. 15, 1979.

3. Gallager, R. G., *Information Theory and Reliable Communication*, McGraw-Hill, New York, 1968.