

Capacity, Cutoff Rate, and Coding for a Direct-Detection Optical Channel

J. L. Massey*

Communications Systems Research Section

It is shown that Pierce's pulse-position modulation scheme with 2^L pulse positions used on a self-noise-limited direct-detection optical communication channel results in a 2^L -ary erasure channel that is equivalent to the parallel combination of L completely-correlated binary erasure channels. The capacity of the full channel is the sum of the capacities of the component channels, but the cutoff rate of the full channel is shown to be much smaller than the sum of the cutoff rates. An interpretation of the cutoff rate is given that suggests a complexity advantage in coding separately on the component channels. It is shown that if short-constraint-length convolutional codes with Viterbi decoders are used on the component channels, then the performance and complexity compare favorably with the Reed-Solomon coding system proposed by McEliece for the full channel. The reasons for this unexpectedly fine performance by the convolutional code system are explored in detail, as are various facets of the channel structure.

I. Introduction

A recent paper by Pierce (Ref. 1) has heightened interest in direct-detection optical communications, particularly for space applications. Pierce considered the situation where the only "noise" limiting communications is that due to the inherent randomness of the optical field at the receiver. He proposed using M -ary pulse position modulation (PPM) together with direct-detection by photon-counting at the receiver. The T second modulation symbol interval is divided into M "slots," in only one of which an optical frequency pulse is transmitted. By virtue of the noiseless assumption, no photons will be detected by the receiver in the $M - 1$ slots where no signal is

present. In the single slot where the transmitter was active, the number of photons detected will be a Poisson random variable whose mean we denote by λ . Thus λ is the average number of received photons per modulation symbol interval. With probability

$$\epsilon = e^{-\lambda}, \quad (1)$$

no photons will be detected in the slot where the pulse was transmitted. Thus, Pierce's PPM scheme creates a constant discrete memoryless channel (CDMC) that is just the M -ary erasure channel where ϵ is the erasure probability. For purposes of this paper, we restrict consideration to the case where

$$M = 2^L \quad (2)$$

*Consultant from University of California, Los Angeles. Now at Swiss Federal Technical University, Sternwartstrasse 7, CH-8006 Zurich, Switzerland.

for some positive integer L so that the modulation symbol can be specified by L binary digits.

A simple calculation gives the capacity of Pierce's PPM channel as

$$\begin{aligned} C &= L(1 - \epsilon) \ln(2) \\ &= L(1 - e^{-\lambda}) \ln(2). \quad (\text{nats}) \end{aligned} \quad (3)$$

On a per-photon basis, this capacity is just

$$C = L(1 - e^{-\lambda}) \ln(2)/\lambda, \quad (\text{nats/photon}) \quad (4)$$

which, as Pierce noted, can be made arbitrarily large by increasing the modulation alphabet size M or, equivalently, by increasing L . Pierce concluded that the problem of communicating efficiently over this self-noise-limited optical channel was thus the coding problem of finding easily implementable schemes to exploit this unlimited capacity.

Although the capacity of the CDMC created by a modulation system is an undeniably interesting characterization of the system's capabilities, it unfortunately gives no information about the complexity of the coding system needed to achieve a desired decoding error probability. This limitation led Wozencraft and Kennedy (Ref. 2) to propose using the cutoff rate, R_0 , of the resulting CDMC to characterize the modulation system. They were motivated by the fact that R_0 is the upper limit of code rates for which the average decoding computation per information bit is finite when sequential decoding is employed. Massey (Ref. 3) suggested further reasons for preferring R_0 over C as a single parameter characterization of a modulation system. He noted that, whether block codes or convolutional codes are employed, R_0 specifies both a range of code rates for which reliable decoding is possible and also a measure of the complexity of the coding system that will be required to achieve a desired error probability. Massey suggested that, as a rule of thumb, R_0 is the practical upper limit on code rates for reliable communications, whereas capacity is the theoretical upper limit.

McEliece and Welch (Ref. 4) and McEliece (Ref. 5) have investigated the cutoff rate of the self-noise-limited direct-detection optical channel and reached conclusions startlingly different from those that arise from capacity considerations. In Ref. 5, McEliece showed that, even allowing multi-amplitude pulsing and soft-decision demodulation, the modulation system is limited to

$$\mathcal{R}_0 \leq 1, \quad (\text{nats/photon}) \quad (5)$$

where the upper limit is attained by Pierce's PPM scheme in the limit of large M and small λ , a result anticipated in Ref. 4. In (5), we have continued the practice begun in Ref. 4 of employing script letters to denote channel measures on a per-photon basis. Note that $\mathcal{R}_0 = R_0/\lambda$ for Pierce's PPM channel.

The enormous discrepancy between the values of R_0 and C for Pierce's PPM channel renders it an ideal channel for resolving the question of which parameter gives a more meaningful measure of the quality of the modulation system. The evidence thus far has seemed to favor R_0 . Note that for a fixed symbol time T , the bandwidth of Pierce's PPM scheme grows linearly with M and hence exponentially with L because of Eq. (2). McEliece, Rodemich and Rubin (Ref. 6) and McEliece (Ref. 7) have shown that this "explosive" increase in bandwidth is unavoidable in the self-noise-limited direct-detection optical channel; they showed that for code rates \mathcal{R} above 1 nat/photon, the required bandwidth and the required peak-to-average signal power must both grow exponentially with \mathcal{R} . They conclude that no practical system could ever be built to operate at a rate \mathcal{R} above, say, 10 nats/photon. The same conclusion was reached by Butman, Katz and Lesh (Ref. 8) starting from a much different point, namely with practical constraints on achievable time resolution and specification that the information rate be interestingly large, say 10^4 nats/sec or greater.

Besides the R_0 versus C debate, Pierce's PPM channel impinges on another ongoing controversy, namely assessing the relative merits of block codes and convolutional codes. For Pierce's PPM channel, the evidence thus far has seemed to favor block codes. McEliece (Refs. 7 and 9) has proposed using Reed-Solomon (RS) codes on the optical PPM channel and has shown that code rates up to 2 or 3 nats/photon are feasible. Moreover, the large alphabet over which RS codes are defined makes these codes appear as virtually ideal for this application, as will be seen in Section IIIA.

In this article, we offer additional evidence in favor of R_0 over C as a meaningful characterization of Pierce's PPM channel. But we also offer some rather surprising evidence to support the claim that convolutional codes are superior to block codes even in this application that is almost tailor made to fit the virtues of RS codes.

In Section II, we show that the optical PPM channel can be viewed as the parallel combination of L "completely correlated" binary erasure channels (BEC's), and we investigate both R_0 and C from this perspective. In Section III, we show that the use of short-constraint-length binary convolutional codes with Viterbi decoding on each component BEC yields coding performance and complexity that compare favorably to those

for RS codes, and we isolate the somewhat strange cause of this excellent performance by convolutional codes. Finally, in Section IV, we offer some additional interpretations of our results and raise some further questions of interest.

II. The Optical PPM Channel as Parallel Completely Correlated BEC's

Suppose we number the slots in Pierce's PPM scheme from 0 to $2^L - 1$. For the modulation symbol, \underline{x} , we can choose the index of the slot containing the optical frequency pulse. Writing \underline{x} as the L -place radix-two number

$$\begin{aligned}\underline{x} &= [x_1, x_2, \dots, x_L] \\ &= x_1 2^{L-1} + x_2 2^{L-2} + \dots + x_L,\end{aligned}\quad (6)$$

we can view the transmission of a single modulation symbol \underline{x} as the transmission of the L binary digits x_1, x_2, \dots, x_L . For example, with $L = 3$, the slots would be numbered from 0 to 7 and $\underline{x} = [1, 1, 0]$ would instruct the transmitter to send the optical frequency pulse in slot 6. Notice that so long as even one photon is detected in the slot where the pulse was sent, the demodulator will correctly identify all L binary digits since the pulse position will be known. But when no photons are detected in this slot, the *entire* modulation symbol is "erased" and hence all L binary digits are simultaneously erased. Thus, we can represent the demodulation symbol, \underline{y} , as

$$\underline{y} = [y_1, y_2, \dots, y_L] \quad (7)$$

where $y_i = x_i$ when one or more photons are detected at the receiver, but $y_i = [E, E, \dots, E]$ (where E is the "erasure indicator") when no photons are detected as happens with probability $\epsilon = e^{-\lambda}$.

Notice that with respect to the transmission of a given component x_i of \underline{x} and the reception of the corresponding component y_i of \underline{y} , the PPM channel becomes simply a binary erasure channel (BEC) with the same erasure probability ϵ as for the entire modulation symbol. Thus, *each use of the 2^L -ary optical PPM channel is entirely equivalent to one use in parallel of L BEC's that are completely correlated* in the sense that an erasure either occurs on all L channels or on none.

The capacity of the BEC with erasure probability ϵ is just $(1 - \epsilon) \ln(2)$ nats. The total capacity, $(C)_{TOT}$, of the L parallel BEC's is thus

$$(C)_{TOT} = L(1 - \epsilon) \ln(2). \quad (\text{nats}) \quad (8)$$

Comparing Eq. (8) with Eq. (3), we see that *there is no penalty in capacity if each of the L parallel BEC's is coded independently*, as opposed to coding jointly over the component channels, but neither is there any gain.

The situation for the cutoff rate, R_0 , is much more interesting. In general, R_0 for a CDMC is given by the expression

$$R_0 = -\min_Q \ln \left\{ \sum_y \left[\sum_x Q(x) \sqrt{P(y|x)} \right]^2 \right\} \quad (\text{nats}) \quad (9)$$

where $P(y|x)$ is the probability that the channel output symbol is y given that the input symbol was x , and where Q is a probability distribution over the channel input alphabet (Ref. 3). For the 2^L -ary erasure channel, $Q(x) = 2^{-L}$ for all x is the minimizing distribution in Eq. (9) and gives

$$R_0 = -\ln [\epsilon + 2^{-L}(1 - \epsilon)] \quad (\text{nats}) \quad (10)$$

or, on a per-photon basis,

$$\mathcal{R}_0 = -\ln [e^{-\lambda} + 2^{-L}(1 - e^{-\lambda})] / \lambda. \quad (\text{nats/photon}) \quad (11)$$

From Eq. (11), we see that for any fixed $\lambda > 0$, \mathcal{R}_0 increases with L but

$$\lim_{L \rightarrow \infty} \mathcal{R}_0 = 1 \quad (\text{nat/photon}) \quad (12)$$

in agreement with Eq. (5).

The cutoff rate of the BEC with erasure probability ϵ is $\ln[2/(1 + \epsilon)]$ nats, as can be found by taking $L = 1$ in Eq. (10). The total cutoff rate, $(R_0)_{TOT}$, of the L parallel BEC's is thus

$$(R_0)_{TOT} = L \ln [2/(1 + \epsilon)], \quad (\text{nats}) \quad (13)$$

which is much larger than the cutoff rate for the full channel as given by Eq. (10). In fact, from Eq. (10) and Eq. (13) we see that

$$\lim_{L \rightarrow \infty} \frac{R_0}{(R_0)_{TOT}} = 0. \quad (14)$$

We will take up the interpretation of this result in Section IV-A, where we will argue that a small value of $R_0/(R_0)_{TOT}$

suggests a complexity advantage in coding over the component channels rather than jointly coding the component channels.

III. Coding for the Optical PPM Channel

A. Joint Coding of the Component Channels

McEliece (Refs. 7 and 9) has proposed using Reed-Solomon (RS) codes on Pierce's 2^L -ary PPM channel in the following manner. Each modulation symbol $\mathbf{x} = [x_1, x_2, \dots, x_L]$ is treated as a digit in the finite field $GF(2^L)$. An (n, k) RS code over this field has block length $n = 2^L - 1$, k information digits for any k such that $1 \leq k < n$, and minimum Hamming distance $d = n - k + 1$, which is the maximum possible for a linear code with $n - k$ parity digits. A linear code with $d = n - k + 1$ is called maximum-distance-separable (MDS) to emphasize this optimality (Ref. 10, pp. 70-72). See Ref. 10, pp. 277-308 for further properties of RS codes and for decoding procedures.

The maximum number of erasures guaranteed correctable by a linear code with minimum Hamming distance d is $d - 1$. Thus an (n, k) RS code can correct all patterns of $n - k$ or fewer erasures, but cannot correct all patterns of $n - k + 1$ erasures. All the well-known algebraic decoding procedures for RS codes correctly decode all patterns of $n - k$ or fewer erasures but virtually no patterns of more than $n - k$ erasures. Thus, it is customary to assume that a decoding error occurs whenever $n - k + 1$ or more erasures occur so that the block error probability, P_e , after decoding is

$$P_e = \sum_{s=n-k+1}^n \binom{n}{s} \epsilon^s (1 - \epsilon)^{n-s} \quad (15)$$

where ϵ is the symbol erasure probability. Our interest, however, is in the bit error probability, P_b , defined as the average probability of error among the kL binary digits that form the k $GF(2^L)$ information digits in the RS code. When a decoding error is made, it is made with high probability to a nearest-neighbor codeword so that $d = n - k + 1$ symbol errors are made. Because a RS code is cyclic, the error probability in each symbol is the same so that the probability that a particular information symbol is in error, given a decoding error, is very nearly d/n . But on the average very close to half of the binary digits forming an information symbol will be incorrect when that symbol is decoded incorrectly. Hence, to a very good approximation,

$$P_b \approx \frac{1}{2} \frac{n - k + 1}{n} P_e \quad (16)$$

for the RS codes.

McEliece (Refs. 7 and 9) has observed that the best performance (i.e., smallest P_e for a given bandwidth after coding) on Pierce's PPM channel is obtained from the RS codes with dimensionless rate $k/n \approx 1/2$. In particular, he proposed using the (31, 16), (63, 32) and (127, 64) RS codes over $GF(2^5)$, $GF(2^6)$ and $GF(2^7)$, respectively. In Fig. 1, we give plots of P_b versus the code rate

$$\mathcal{R} = \frac{k}{n} \frac{\ln(L)}{\lambda} \quad (\text{nats/photon})$$

for these three codes. These plots were taken from Ref. 8, where they were given as P_e as calculated by Eq. (15), after conversion to P_b via Eq. (16). Note that the above expression reflects the fact that on the average, λ photons are used to transmit each $GF(2^L)$ encoded symbol. Note also that λ determines the erasure probability ϵ according to Eq. (1).

Figure 1 shows that reliable communications using RS codes is feasible for rates up to about 2 nats/photon. Notice that the coding and modulation together expand the transmitted bandwidth relative to on/off binary signalling by a factor

$$F = \frac{n}{k} \frac{2^L}{L} \approx \frac{2^{L+1}}{L} \quad (17)$$

where the factor $2^L/L$ is due to the PPM modulation which uses 2^L slots to transmit L binary digits, and where the factor $n/k \approx 1/2$ is due to the RS code which uses (very close to) 2 encoded symbols for each information symbol. The bandwidth expansion factor F is indicated on each curve in Fig. 1. The 37-fold expansion for the (127, 64) RS code is perhaps near the practical limit for time resolution at reasonably high data rates; the required 63 erasure-correcting RS decoder is certainly near the practical limit of complexity.

The RS codes, because they are MDS codes, have maximum erasure-correcting power for their length and number of information symbols. Moreover, their symbol alphabet $GF(2^L)$ is ideally matched to the 2^L -ary PPM channel since each erasure by the receiver erases only one code symbol although it erases all L binary components of that symbol. It is doubtful that any block coding scheme can significantly outperform McEliece's RS coding scheme on the PPM channel for a given bandwidth expansion and a given decoder complexity.

B. Separate Coding of the Component Channels

We now consider employing a separate binary coding scheme on each of the L BEC's that constitute the 2^L -ary

optical PPM channel. Note that this is equivalent to interleaving L separate binary encoded streams to form a single binary stream whose digits, taken in blocks of length L , constitute the modulation symbols.

The use of binary block codes with algebraic decoding gives disappointingly poor performance in this separate channel mode of coding for the PPM channel. For instance, the (31, 16) binary BCH code has minimum Hamming distance 7 and is thus 6 erasure correcting. However, the (31, 16) RS code considered above can also be considered to be encoding 16 information bits on each component channel into 31 binary digits, yet is 15 erasure correcting on each component channel. The cross channel constraints imposed by the RS code effectively more than doubles the number of correctable erasures compared to the single channel BCH code. The BCH code performs thus much worse than the RS code when L is chosen for the BCH code to give the same bandwidth expansion as does the RS code, and is not significantly easier to decode. The (24, 12) 7 erasure correcting Golay binary code fares little better than the BCH code, as can also be seen from Fig. 1.

In light of the above, it seems quite surprising that good performance relative to the RS codes can be obtained by separately coding the component BEC's, using short-constraint-length convolutional codes with Viterbi (i.e., maximum-likelihood (ML)) decoding (Ref. 11, pp. 227-252). In Fig. 1, we show the performance of dimensionless rate 1/2 binary convolutional codes with constraint length K (measured in information bits) for $K = 4, 6$ and 8. In each case, the number L of component channels was chosen so that the bandwidth expansion factor

$$F = \frac{2^{L+1}}{L} \quad (18)$$

matched that of one of the RS codes considered above.

We see from Fig. 1 that the binary $K = 4$ convolutional code gives virtually the same performance as the (31, 16) RS code with the same bandwidth expansion factor. The required $2^{K-1} = 8$ state Viterbi decoder appears much easier to implement than the corresponding 15 erasure correcting RS decoder. Similarly, we see from Fig. 1 that the $K = 6$ binary convolutional code is an attractive competitor to the (63, 32) RS code, and that the $K = 8$ binary convolutional code fares well against the (127, 64) RS code.

Inasmuch as they sacrifice the substantial advantage that can be gained by coding across the component channels (which the RS codes exploit with maximum effectiveness), it

appears puzzling at first that the short-constraint-length binary convolutional codes perform so well in the separate channel coding mode for the PPM channel. The explanation is that Viterbi decoders, unlike algebraic decoders, degrade gracefully. The free distance d_f of the convolutional code determines that no patterns of $d_f - 1$ or fewer erasures can cause a decoding error but that some patterns of d_f erasures will. However, the Viterbi decoder, because it is a ML decoder, corrects the overwhelming majority of patterns of $d_f, d_f + 1$, and more erasures. This ability to go beyond the minimum distance bound on erasure correction fully compensates for the sacrifice made in coding separately on the component channels.

The convolutional code performance curves in Fig. 1 are actually the Bhattacharyya upper bounds on P_b (Ref. 11, p. 246). According to this bound,

$$P_b \leq f(z) \quad (19)$$

where f is a rational function determined by the state-transition structure of the convolutional encoder, and where z is the channel parameter

$$z = \sum_y \sqrt{P(y|0)P(y|1)}. \quad (20)$$

For the BEC,

$$\begin{aligned} z &= e \\ &= e^{-\lambda} \end{aligned} \quad (21)$$

where we have made use of Eq. (1).

To obviate explicitly finding f , we employed the following "trick" due to Omura (Ref. 12). For the additive white Gaussian noise (AWGN) channel with binary antipodal signals of energy E and one-sided noise power spectral density N_o , one finds

$$z = e^{-E/N_o}. \quad (22)$$

Thus, for the same code, the bound Eq. (19) on P_b will be the same for the BEC as for the AWGN if one chooses

$$\lambda = E/N_o. \quad (23)$$

By the artifice of Eq. (23), we converted the bound Eq. (19) on P_b versus E_b/N_o (where $E_b = 2E$ is the energy per

information bit) given in Ref. 12 to the convolutional code performance curves given in Fig. 1.

IV. Interpretations, Remarks, and Questions

A. On the Significance of R_0

As was first shown by Viterbi, the average bit error probability for ML decoding of the ensemble of time-varying convolutional codes of rate R and constraint length N (measured in encoded digits) on a CDMC with cutoff rate R_0 satisfies

$$P_b \leq c_R e^{-NR_0} \quad \text{if } R < R_0 \quad (24)$$

where c_R is an unimportant factor that depends on R but not on N (Ref. 11, p. 312). Moreover, the error exponent R_0 in Eq. (24) is also the exponent of P_b versus N for the sequence of best codes at each length N when $R \approx R_0$ (Ref. 11, p. 320). This strongly suggests that R_0 should be considered as at least a rough measure of the necessary code constraint length in channel symbols required to achieve a given P_b , in the sense that doubling R_0 will roughly halve the required N .

To validate this interpretation in a fairly trivial instance, consider a CDMC that is the parallel combination of L identical and independent CDMC's. Letting R_0 be the cutoff rate of the full channel and $(R_0)_{TOT}$ be the sum of the cutoff rates of the component channels, one easily verifies from Eq. (9) that

$$R_0 = (R_0)_{TOT} \quad (25)$$

so that R_0 is exactly L times that of each component channel. Thus, by the above interpretation, for a given P_b separate coding on each channel should require a constraint length in channel symbols L times that required for joint coding of the channels. But a channel symbol for the full channel is equivalent to L channel symbols for a component channel. Thus, the required constraint length, measured in symbols for the component channel, is the same whether separate channel or joint channel coding is used. This is hardly surprising since one use of the full channel is, because of the independence of the component channels, entirely equivalent to L uses of one component channel. But this does illustrate that the above interpretation of R_0 is precisely correct in this case.

Next, consider the optical PPM channel viewed as L parallel but completely correlated BEC's. Recall also from Eq. (14) that R_0 for the full channel is generally much smaller than L times that for each channel, i.e., $R_0 \ll (R_0)_{TOT}$. The above interpretation of R_0 then suggests that a much smaller binary

digit constraint length on the component channels will suffice to give the same P_b compared to the constraint length in *binary digits* required for cross-channel coding. This suggests a complexity advantage in coding separately for each of the component channels. To illustrate the quantitative validity of R_0 in this context, note that for dimensionless rate $1/2$ coding on the 2^L -ary PPM channel, a rate of \mathcal{R} nats/photon corresponds to an average of

$$\lambda = \frac{1}{2} L \ln(2)/\mathcal{R} \quad (26)$$

photons in the transmitted pulse. For example, with $L = 5$ and $\mathcal{R} = 1.0$ nats/photon, Eq. (26) gives $\lambda = 1.73$ photons. From Eq. (1), we find the corresponding erasure probability to be $\epsilon = 0.177$. Then from Eqs. (10) and (13) we find $R_0 = 1.597$ and $(R_0)_{TOT} = 2.652$, respectively. This suggests that the required constraint length in binary digits required for joint coding of the $L = 5$ BEC's will be about $2.652/1.597 = 1.66$ times that required for separate coding of each BEC to obtain the same P_b . To test this conclusion, consider again Fig. 1. Note that for $\mathcal{R} = 1.0$, the $K = 4$ ($L = 5$) convolutional code gives virtually the same P_b as does the $(31, 16)$ RS code. But the RS code has a constraint length of $5(31) = 155$ binary digits. Using the rule of thumb that the effective decoding constraint length of a convolutional code is about twice that of a block code with the same encoding constraint length, we can approximate the equivalent block code constraint length of the convolutional code as about $2K/(1/2) = 16$ binary digits. The ratio $155/16 = 9.7$ of the required constraint lengths is rather larger than the ratio $(R_0)_{TOT}/R_0 = 1.66$, but the discrepancy is probably due more to the difficulty of comparing a convolutional code to a block code than to the coarseness of our interpretation of R_0 .

B. On ML Decoding of the RS Codes

We observed in Section III-B that the ML nature of Viterbi decoding, which allows most patterns of more than $d_f - 1$ erasures to be corrected, was the primary reason for the strong performance of binary convolutional codes as compared to the RS block codes on the 2^L -ary optical PPM channel. The question then arises as to whether the performance of the RS codes could not also be greatly enhanced if they were decoded by a ML decoder rather than a distance-limited algebraic decoder. The answer, surprisingly, is no.

Suppose that s erasures occur in the RS code symbols where $s > n - k$. This leaves only $n - s < k$ unerased digits in the block. However, the MDS property of RS codes implies that every set of k code positions is an information set, i.e., that it can be used as the positions containing the k information digits. Thus, there will be at least one erased

position to which we can assign an arbitrary digit in $GF(2^L)$ and still be able to find a codeword that matches it and all the unerased digits. Thus there will be at least 2^L codewords matching all the unerased digits, and even a ML decoder can do no more than guess which of these was the transmitted codeword. It follows that, given that more than $n - k$ erasures have occurred, any decoder for the RS code will err with probability at least $(2^L - 1)/2^L \approx 1$, even though the code can correct all patterns of $n - k$ and fewer erasures. The conclusion is that no RS decoder can degrade gracefully on the optical PPM channel and that a ML decoder is negligibly better than an algebraic decoder on this channel. We caution the reader, however, to note that this conclusion would not hold on many other types of channels where ML decoding would be significantly better than algebraic decoding of RS codes.

C. Correlated Decoding of the Component Channels

When coding separately for each of the L component BEC's of the optical PPM channel, one can either use L separate binary coding systems or time-share one such system that operates at L times the speed required for the separate systems. In either case, the decoding complexity would be reckoned at about L times that of each separate system. We point out now that there is a possibility to reduce substantially the decoding complexity when separate channel coding is used.

Because the L components BEC's of the 2^L -ary optical PPM channel are completely correlated, the decoder for one channel can pass useful information to the other $L - 1$ decoders to simplify their decoding tasks; i.e., the decoders can profitably operate in a "correlated" fashion. To see this more clearly, note that the decoder for a linear (whether block or convolutional) binary code used on the BEC effectively solves the linear equations, determined by the code, that relate the erased digits to the unerased digits. The decoder effectively evaluates each erased digit as a modulo-two sum of certain unerased digits. Thus, after the first decoder has determined which set of unerased digits should be added to find a given erased digit, it can pass this information to the other $L - 1$

decoders. Then, because the erasure patterns on all L BEC's are identical, these other decoders need merely to add (modulo-two) the unerased digits that have been received over their own channels in those positions specified by the first decoder. Such correlated or "cooperative" decoding is clearly possible in principle and would have obvious complexity advantages. However, we have not yet succeeded in finding a general way to implement such correlated decoding when a Viterbi decoder is used, although we have been able to find simple implementations for certain very-short-constraint-length convolutional codes.

D. Correlated Channels

The somewhat curious properties of the optical PPM channel viewed as a parallel combination of completely correlated BEC's suggest that it might be interesting to consider more generally a CDMC that is the parallel combination of identical CDMC's that have some specified dependency. The relationship of R_0 to $(R_0)_{TOT}$ should be especially interesting. It should also be interesting to consider whether correlated decoding to reduce decoding complexity can be performed when each component channel is separately encoded.

E. Background Noise on the Optical PPM Channel

It is clear that the self-noise-limited optical PPM channel model used throughout this paper becomes physically inappropriate when the signalling bandwidth becomes sufficiently large. Account then must be taken of background radiation that can lead to "errors" as well as erasures by the (preferably soft-decision) demodulator. We will not pursue these matters further here except to note that the short-constraint-length convolutional codes with Viterbi decoding can easily be adapted to make use of the soft-decision demodulation information, but the RS block codes cannot. Thus, the convolutional codes should become even more attractive vis-a-vis the RS codes when background noise is sufficiently strong so that it must be taken into account. Convolutional codes with Viterbi decoding seem to make a more robust coding system than do RS codes with algebraic decoding.

References

1. Pierce, J. R., "Optical Channels: Practical Limits with Photon Counting," *IEEE Trans. Comm.*, vol. COM-26, pp. 1819-1821, Dec. 1978.
2. Wozencraft, J. M., and Kennedy, R. S., "Modulation and Demodulation for Probabilistic Coding," *IEEE Trans. Info. Th.*, vol. IT-12, pp. 291-297, July 1966.
3. Massey, J. L., "Coding and Modulation in Digital Communications," in *Proc. Int. Zurich Sem. on Digital Comm.*, Zurich, Switzerland, Mar. 12-15, 1974.
4. McEliece, R. J., and Welch, L. R., "Coding for Optical Channels with Photon Counting," in *The Deep Space Network Progress Report 42-52*, pp. 61-66, Jet Propulsion Laboratory, Pasadena, Calif., Aug. 15, 1979.
5. McEliece, R. J., "The R_0 -Parameter for Optical Communication Using Photon Counting," in *The Deep Space Network Progress Report 42-53*, pp. 62-65, Jet Propulsion Laboratory, Pasadena, Calif., Oct. 15, 1979.
6. McEliece, R. J., Rodemich, E. R., and Rubin, A. L., "The Practical Limits of Photon Communication," in *The Deep Space Network Progress Report 42-55*, pp. 63-67, Jet Propulsion Laboratory, Pasadena, Calif., Feb. 15, 1980.
7. McEliece, R. J., "Practical Codes for Photon Communication," submitted to *IEEE-Trans. Info. Th.*, 1980.
8. Butman, S. A., Katz, J., and Lesh, J. R., "Practical Limitations on Noiseless Optical Channel Capacity," in *The Deep Space Network Progress Report 42-55*, pp. 12-14, Jet Propulsion Laboratory, Pasadena, Calif..
9. McEliece, R. J., "Coding for the Photon Channel," in *Proc. Nat. Tel. Conf.*, pp. 23.3.1-23.3.3, 1979.
10. Peterson, W. W., and Weldon, E. J., Jr., *Error-Correcting Codes*, 2nd ed. M.I.T. Press, Cambridge, Mass., 1972.
11. Viterbi, A. J., and Omura, J. K., *Principles of Digital Communication and Coding*, McGraw-Hill, New York, 1979.
12. Heller, J. A., and Jacobs, I. M., "Viterbi Decoding for Satellite and Space Communication," *IEEE Trans. Comm.*, vol. COM-19, pp. 835-848, 1971.

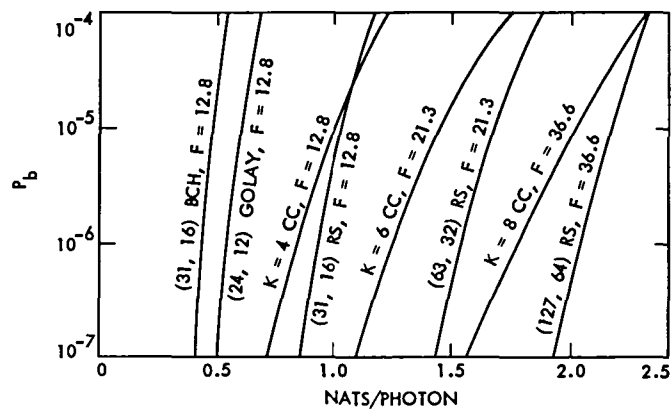


Fig. 1. Bit decoding error probability versus code rate in nats/ photon for selected Reed-Solomon (RS) codes, convolutional codes (CC), the (31, 16) BCH code, and the (24, 12) Golay code on the optical PPM channel. F is the bandwidth expansion factor due both to the modulation system and coding scheme