Finally, we note that by taking $M = L + T$, the ensemble of R,L,T,M trellis codes becomes exactly the ensemble of R,L,T tree codes. We have already noted that, for $M = T$, the ensemble of R,L,T,M trellis codes becomes the ensemble of trellis codes defined by Massey [1]. Hence our Theorem is a generalization from which upper bounds on $\overline{P[\epsilon]}$ for both these ensembles follow as special cases.

III. Results of Simulations

Although the above theory was developed for true maximum-likelihood (i.e., Viterbi) decoding where one almost never uses a tail, its practical application is to sequential decoding where a tail is often used. The undetected error phenomenon is more complex for sequential decoding and, hence, we have to be careful with our conclusions. Nevertheless, it is well-known [5], [6] that, with the appropriate bias term, the exponent of error probability for sequential decoding is the same as that for true maximum-likelihood or Viterbi decoding. Thus we have conducted sequential decoding simulations to test the dependence of $P[\epsilon]$ on T and M.

The particular sequential decoding algorithm employed was the stack algorithm [7], [8]. The simulations were all performed with rate $R = \frac{1}{2}$ optimum distance profile codes [9], [10]. The simulated binary symmetrical channel (BSC) had "crossover probability" $p = 0.045$, which corresponds to $R = R_0 = \frac{1}{2}$. For three different code memory lengths, a very large number (100 000) of received "frames;" i.e., complete received sequences of length $n(L + T)$, were decoded so that the decoding error probability could be accurately inferred.

In Fig. 1, we give the simulation results for the sequential decoding undetected error probability $P[\epsilon]$ as a function of the tail length T of the convolutional code. Because of the extreme variability of the computation in sequential decoding when M is large, there were occasions where the decoding had to be stopped, and hence, the frames had to be erased because the computation exceeded the alloted maximum. The number of erased frames is indicated in Fig. 1 and had negligible effect on the curves. These curves show that the actual $P[\epsilon]$ decreases exponentially with T having an exponent very close to that of the bound (5) for the range $T \leq M - [nE_{VU}(R)]^{-1} \log_2 L + 1$, while further increases in T beyond this point have virtually no effect on $P[\epsilon]$.

The range of T for which the bound becomes independent of L, viz., $T \leq M - [nE_{VU}(R)]^{-1} \log_2 L$, is close to the range where the true $P[\epsilon]$ becomes independent of L. Hence, relation (6) can be taken as a slightly conservative design rule for choosing M so that $P[\epsilon]$ is reduced to the minimum possible for the tail length T that can be allocated to an encoded frame.

IV. Remark

Finally, we should remark that, if we wanted solely to minimize the undetected error probability with sequential decoding for a given memory length and were not concerned with holding the tail size to a minimum to maximize the true rate of the trellis code, then the optimal value of the tail length is, of course, the memory length, i.e., $T = M$. Probably this fact has caused some investigators to ignore the distinction between the tail and the memory so that the memory length came to be honored for work actually done by the tail.

Acknowledgment

The author is greatly indebted to Prof. James L. Massey for his help in presenting these results in an easily understandable manner. Furthermore, the use in principle of convolutional codes with memory length greater than tail length to remove the dependence of $P[\epsilon]$ on tree length has been suggested independently by Prof. Massey [11].

References

[1] J. L. Massey, "Coded digital communication," Notes used in course 4007—Fall 1971 at the Royal Technical University of Denmark, Lyngby, Denmark.
[2] A. J. Viterbi, "Error bounds for convolutional codes and an asymptotically optimum decoding algorithm," *IEEE Trans. Inform. Theory* vol. IT-13, pp. 260–269, April 1967.
[3] G. D. Forney, Jr., "Convolutional codes II: Maximum-likelihood decoding and convolutional codes III: Sequential decoding," Appendix A, *Inform. Contr.*, vol. 25, pp. 222–266, July 1974.
[4] R. Johannesson, "On the error probability of general tree and trellis codes with applications to sequential decoding," Tech. Rpt. No. EE7316, Dept. of Elec. Engr., Univ. of Notre Dame, Notre Dame, IN Dec. 1973.
[5] H. L. Yudkin, "Channel state testing in information decoding," Sc.D. dissertation, Dept. of Elec. Engr., Mass. Inst. of Tech., Cambridge, MA, Sept. 1964.
[6] R. G. Gallager, *Information Theory and Reliable Communication.* New York: Wiley, 1968.
[7] F. Jelinek, "A fast sequential decoding algorithm using a stack," *IBM J. Res. Dev.*, vol. 13, pp. 675–685, Nov. 1969.
[8] K. Sh. Zigangirov, "Some sequential decoding procedures," *Probl. Peredach. Inform.*, vol. 2, pp. 13–25, no. 4, 1966.
[9] R. Johannesson, "Robustly optimal rate one-half binary convolutional codes," *IEEE Trans. Inform. Theory* vol. IT-21, pp. 464–468, July 1975.
[10] R. Johannesson and E. Paaske, "Further results on binary convolutional codes with an optimum distance profile," Presented at the IEEE Int. Symp. Inform. Theory, Ronneby, Sweden, June 20–24, 1976.
[11] J. L. Massey, "An error bound for random tree codes," presented at the IEEE Int. Symp. Inform. Theory, Ashkelon, Israel, June 25–29, 1973.

An Improved Upper Bound on the Block Coding Error Exponent for Binary-Input Discrete Memoryless Channels

ROBERT J. McELIECE, MEMBER, IEEE, AND JIM K. OMURA, MEMBER, IEEE

Abstract—**The recent upper bounds on the minimum distance of binary codes given by McEliece, Rodemich, Rumsey, and Welch are shown to result in improved upper bounds on the block coding error exponent for binary-input memoryless channels.**

Consider a binary-input memoryless channel with input alphabet $A = \{0,1\}$, output alphabet B, and transition probabilities $\{p(y|x): x \in A, y \in B\}$. Let $\mathcal{C} = \{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_M\}$ be a binary code of length n and rate $R = n^{-1} \log_2 M$ for this channel, and assume that each of the M codewords is sent with probability $1/M$. Let $d_{\min}(\mathcal{C})$ denote the minimum Hamming distance between distinct codewords, and let $P_e(\mathcal{C})$ denote the probability of maximum-likelihood decoder error when the code $\mathcal{C}$ is used on the given channel.

Now define

$$\delta(n,R) = \frac{1}{n} \max d_{\min}(\mathcal{C}) \tag{1}$$

$$P_e(n,R) = \min P_e(\mathcal{C}), \tag{2}$$

where the maximum and minimum in (1) and (2) are taken over the set of all codes of length n and rate R or greater. And finally

Manuscript received June 30, 1976; revised January 6, 1977. This work was supported in part by the National Aeronautics and Space Administration under Contract NAS 7-100 and in part by the National Science Foundation under Grant ENG 75-03224.

R. J. McEliece is with the Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91103.
J. K. Omura is with the System Science Department, University of California, Los Angeles, CA 90024.

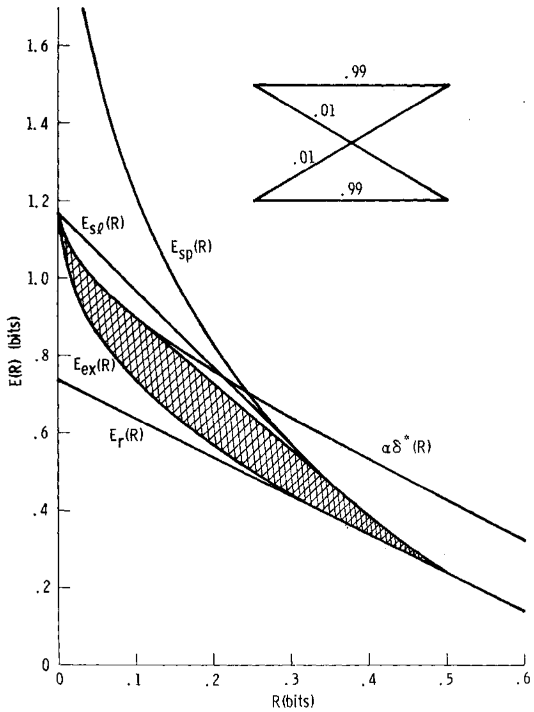

Fig. 1. Error exponents for binary symmetric channel.

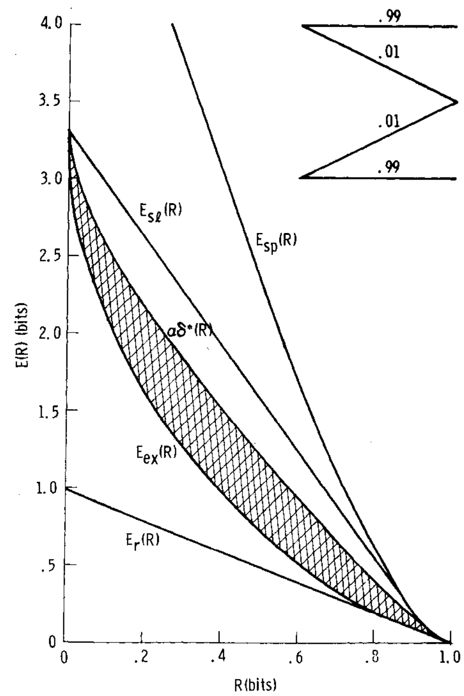

Fig. 2. Error exponents for binary erasure channel.

define

$$\bar{\delta}(R) = \lim_{n\to\infty} \sup \delta(n,R) \tag{3}$$

$$\underline{\delta}(R) = \lim_{n\to\infty} \inf \delta(n,R)$$

$$\bar{E}(R) = \lim_{n\to\infty} \sup \frac{1}{n}(-\log_2 P_e(n,R))$$

$$\underline{E}(R) = \lim_{n\to\infty} \inf \frac{1}{n}(-\log_2 P_e(n,R)). \tag{4}$$

It is widely believed that the limits in (3) and (4) exist. However, for $\bar{\delta}(R)$, this is known only for $R = 0$ and 1: $\bar{\delta}(0) = \underline{\delta}(0) = \frac{1}{2}$, $\bar{\delta}(1) = \underline{\delta}(1) = 0$. For $\underline{E}(R)$, the limit is known to exist, and its value is known at $R = 0$ and for $R \geq R_{\text{crit}}$, R_{crit} being a number to be defined below. We shall now briefly survey the known upper and lower bounds on $E(R)$, and indicate how the new upper bound $\delta^*(R)$ on $\bar{\delta}(R)$ obtained in [1] can be used to improve the known upper bounds on $\bar{E}(R)$ for small values of R.

First, we have the *sphere-packing* bound $E_{sp}(R)$ and the *random coding* bound $E_r(R)$, both valid for all rates R less than channel capacity [2], [3]:

$$E_r(R) \leq \underline{E}(R) \leq \bar{E}(R) \leq E_{sp}(R). \tag{5}$$

The two bounds in (5) are equal for sufficiently large R, and in fact the number R_{crit} cited above is the point where these two bounds meet. (Formulas for E_r and E_{sp} for binary symmetric and binary erasure channels are given in the Appendix.)

Next, we have bounds which depend on the *Bhattacharya* parameter [4], [5] for the channel, which is defined by

$$\alpha = -\log_2 \sum_{y\in B} (p(y|0)p(y|1))^{1/2}.$$

These bounds are

$$\alpha D \leq \underline{E}(R) \leq \bar{E}(R) \leq \alpha\bar{\delta}(R), \tag{6}$$

where $0 \leq D \leq \frac{1}{2}$ is defined implicitly by $R = 1 - H_2(D)$, where $H_2(x)$ is the binary entropy function. (The lower bound in (6) is called the *expurgated* bound $E_{ex}(R)$; it is valid only for $0 \leq R \leq R'$, where R' is the rate at which the expurgated bound meets the random coding bound.) As mentioned, the function $\bar{\delta}(R)$ is unknown, so the upper bound in (6) is ineffective. However, by using the bound $\bar{\delta}(R) \leq \delta^*(R)$ obtained in [1] (for numerical values of $\delta^*(R)$, see Table 1 in [1]), we obtain an upper bound

$$\bar{E}(R) \leq \alpha\delta^*(R) \tag{7}$$

which can be evaluated, and which is already better than any previously known upper bound for small values of R.

Finally, Shannon *et al.* [3] have shown that if $E_0(R)$ is any upper bound to $\bar{E}(R)$, then so is the convex hull of the curves $E_0(R)$ and $E_{sp}(R)$. In particular, by taking $E_0(R) = \frac{1}{2}\alpha$ (from (6) and the fact that $\delta(0) = \frac{1}{2}$), we see that $\bar{E}(R)$ is bounded from above by the line passing through the point $(0,\alpha/2)$ which is tangent to $E_{sp}(R)$. This bound is called the *straight-line* bound $E_{sl}(R)$. However, by taking $E_0(R) = \alpha\delta^*(R)$ (cf. (7)), we can obtain an upper bound which is significantly better than min $(E_{sl}(R), E_{sp}(R))$ for a considerable range of R. We illustrate this in Fig. 1 with a binary symmetric channel with crossover probability $\epsilon = 0.01$, $\alpha = -\log_2 \sqrt{4\epsilon(1-\epsilon)} = 2.329$, and in Fig. 2 with a binary erasure channel with erasure probability $\epsilon = 0.01$, $\alpha = -\log_2 \epsilon = 6.644$. In both figures, the unknown region for $0 \leq R \leq R_{\text{crit}}$ bounding $(\underline{E}(R),\bar{E}(R))$ is shaded. A final point worth mentioning is that the new upper bound (7) on $\bar{E}(R)$ always matches the expurgated bound $E_{ex}(R)$ in slope at $R = 0$. (Both slopes are $-\infty$; this slope is well-known for the expurgated bound, and follows for the bound (7) from the results of [1].) This fact supports the conjecture that $\underline{E}(R) = \bar{E}(R) = E_{ex}(R)$ for $R \leq R_{\text{crit}}$ for binary-input channels.

APPENDIX
$E_r(R)$ AND $E_{sp}(R)$ FOR BINARY SYMMETRIC AND BINARY ERASURE CHANNELS

For a binary symmetric channel with crossover probability ϵ, the random coding exponent is given by

$$E_r(R) = \begin{cases} 1 - R - \log_2(1 + \sqrt{4\epsilon(1-\epsilon)}), \\ \qquad 0 \leq R \leq 1 - H_2(\sqrt{\epsilon}/(\sqrt{\epsilon} + \sqrt{1-\epsilon})) \\ T_\epsilon(D) - H_2(D), \\ \qquad 1 - H_2(\sqrt{\epsilon}/(\sqrt{\epsilon} + \sqrt{(1-\epsilon)})) \leq R \leq 1 - H_2(\epsilon), \end{cases}$$

where $T_\epsilon(D) = -D \log_2 \epsilon - (1 - D) \log_2 (1 - \epsilon)$, and where D satisfies (7). The sphere packing exponent is

$$E_{sp}(R) = T_\epsilon(D) - H_2(D), \qquad 0 \le R \le 1 - H_2(\epsilon).$$

(Hence, $R_{\text{crit}} = 1 - H_2(\sqrt{\epsilon}/(\sqrt{\epsilon} + \sqrt{1-\epsilon})$ and $E(R) = E_r(R) = E_{sp}(R)$ for $R \ge R_{\text{crit}}$.) For the binary erasure channel with erasure probability ϵ,

$$E_r(R) = \begin{cases} 1 - R - \log_2 (1 + \epsilon), & 0 \le R \le 1 - 2\epsilon/(1+\epsilon) \\ E_{sp}(R), & 1 - 2\epsilon/(1+\epsilon) \le R \le 1 - \epsilon, \end{cases}$$

where

$$E_{sp}(R) = \frac{\rho\epsilon 2^\rho}{(1-\epsilon) + \epsilon 2^\rho} - \log_2 ((1-\epsilon) + \epsilon 2^\rho),$$

where ρ is determined by $R = 1 - \epsilon 2^\rho/(1 - \epsilon + \epsilon 2^\rho)$.

REFERENCES

[1] R. J. McEliece, E. R. Rodemich, H. Rumsey, Jr., and L. R. Welch, "New upper bounds on the rate of a code via the Delsarte-MacWilliams inequalities," *IEEE Trans. Inform. Theory*, vol IT-23, pp. 157–166, Mar. 1977.

[2] R. G. Gallager, "A simple derivation of the coding theorem and some applications," *IEEE Trans. Inform. Theory*, vol. IT-11, pp. 3–18, Jan. 1965.

[3] C. E. Shannon, R. G. Gallager, and E. R. Berlekamp, "Lower bounds to error probability for coding on discrete memoryless channels," *Inform. Contr.*, vol. 10, pp. 65–103 (Part I), pp. 522–552 (Part II), Jan. 1967.

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

[5] ——, "On general Gilbert bounds," *IEEE Trans. Inform. Theory*, vol. IT-19, pp. 661–665, Sept. 1973.

Concatenated Codes with Large Minimum Distance

LIH-JYH WENG, MEMBER, IEEE

Abstract—**Some concatenated codes of length 128 and less are constructed. Nineteen of these codes are superior to the best previously known linear codes, as shown by the fact that the well-known lower bound on the minimum distance of the concatenated code as the product of the minimum distances of its component codes exceeds the minimum distance of the best previously known code.**

I. INTRODUCTION

Concatenated codes [1] are usually considered to be effective codes for channels with burst errors as well as random errors. Most studies concerning concatenated codes treat the inner and outer codes separately. The purpose of the inner code is to correct random errors and detect burst errors, while the principal purpose of the outer code is to correct both the "erasures" detected by the inner codes and some erroneously decoded inner blocks. In this correspondence, we study concatenated codes as block codes. An $(N,K;D)$ concatenated code of block length N, dimension K and minimum distance D consists of an outer $(n_0,k_0;d_0)$ code over $GF(2^m)$ and an inner $(n_i,k_i;d_i)$ binary code. For convenience, we also refer to this as an $(n_0,k_0;d_0) \otimes (n_i,k_i;d_i)$ code.

Manuscript received May 20, 1976; revised November 24, 1976.

The author is with CNR, Inc., Needham Heights, MA 02194.

The relation of the parameters N, K, and D of a concatenated code to those of its outer and inner codes is established in Section II. A class of concatenated codes with easily computed weight distributions is then introduced. In Section III, 38 concatenated codes of length less than 128 with known weight distributions are listed. All the codes listed in Section III are as good as any previously known linear codes in the sense that the minimum distance of each code is the same as, or greater than, that of the best previously known linear code with the same length and dimension. In Section IV, 19 concatenated codes of length 128 or less which have larger minimum distances than the best previously known codes are tabulated.

II. MINIMUM DISTANCE OF A CONCATENATED CODE

Let us arrange a codeword of the $(N,K;D) = (n_0,k_0;d_0) \otimes (n_i,k_i;d_i)$ concatenated code into an $n_i \times n_0$ matrix as

$$A = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n_0} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n_0} \\ \vdots & & & \\ a_{m,1} & a_{m,2} & \cdots & a_{m,n_0} \\ a_{m+1,1} & a_{m+1,2} & \cdots & a_{m+1,n_0} \\ \vdots & & & \\ a_{2m,1} & a_{2m,2} & \cdots & a_{2m,n_0} \\ \vdots & \cdot & & \\ a_{k_i,1} & a_{k_i,2} & \cdots & a_{k_i,n_0} \\ \vdots & & & \\ a_{n_i,1} & a_{n_i,2} & \cdots & a_{n_i,n_0} \end{bmatrix}. \tag{1}$$

Each column of the matrix A is a codeword of the inner binary $(n_i,k_i;d_i)$ code. The first k_i rows of the matrix A can be grouped into ℓ sets, each of which contains m consecutive binary rows. If each set of these m binary rows is considered as one row of symbols of $GF(2^m)$, then each set of these rows forms a codeword of the outer $(n_0,k_0;d_0)$ code over $GF(2^m)$. Note that in this correspondence, we always assume $k_i = \ell m$, where ℓ and m are positive integers. For each nonzero codeword of a concatenated code arranged in the form of (1), there exists at least one codeword over $GF(2^m)$, which contains at least d_0 nonzero elements. Thus the matrix A has at least d_0 nonzero columns; each of the nonzero columns, which are codewords of the inner code, contains at least d_i ones. Therefore, the weight of a nonzero concatenated codeword is at least $D_L = d_0 d_i$. Also, each element of the $k_i \times k_0$ submatrix of A in (1) in the upper left corner can be assigned arbitrarily as "0" or "1." Hence, the dimension of the concatenated code is $k_i \times k_0$. Thus we have the following well-known theorem.

Theorem: Let the inner and outer codes of an $(N,K;D)$ concatenated code be a binary $(n_i,k_i;d_i)$ code and an $(n_0,k_0;d_0)$ code over $GF(2^m)$, respectively, with $k_i = \ell m$. Then

$$N = n_i n_0 \qquad K = k_i k_0 \qquad D \ge D_L = d_i d_0.$$

We also state the following immediate corollary.

Corollary: If $\ell = 1$, i.e., if $k_i = m$, and if the inner code is a code whose nonzero codewords are of constant weight, then $D = D_L = d_i d_0$.

The truth of this corollary can be seen from the fact that the weight of A in (1) will be $d_i d_0$ when there are exactly d_0 nonzero columns since each of these columns will have weight exactly d_i.

It should be noted that the lower bound on the minimum distance D_L of a concatenated code is a more important parameter to consider than the actual minimum distance D when we employ deletions-and-errors decoding for the outer code [1]. With this decoding algorithm, all possible error patterns of weight $T_L = \lfloor (D_L - 1)/2 \rfloor$ or less can be correctly decoded. In addition, many error patterns of weight exceeding T_L can be corrected; for ex-