On the error probability of general trellis codes with applications to sequential decoding

Johannesson, Rolf

Published in:
IEEE Transactions on Information Theory

1977

[4] F. R. Magee, "A comparison of compromise Viterbi algorithm and standard equalization over band-limited channels," *IEEE Trans. Commun.*, vol. COM-23, pp. 361–367, March 1975.
[5] D. D. Falconer and F. R. Magee, Jr., "Adaptive memory truncation for maximum-likelihood sequence estimation," *Bell Syst. Tech. J.*, vol. 59, pp. 1541–1562, November 1973.
[6] D. G. Messerschmitt, "Design of a finite impulse response for the Viterbi algorithm and decision-feedback equalizer," in *Proceedings of the International Conference on Communications, Minneapolis, MN*, pp. 370-1 to 370-5, June 1974.
[7] S. A. Fredriccson, "Optimum transmitting filter in digital PAM systems with a Viterbi detector," *IEEE Trans. Inform. Theory*, vol. IT-20, pp. 479–489, July 1974.
[8] R. R. Anderson and G. J. Foschini, "The minimum distance for MLSE digital data systems of limited complexity," *IEEE Trans. Inform. Theory*, vol. IT-21, No. 5, pp. 544–551, September 1975.
[9] D. G. Messerschmitt, "A geometric theory of intersymbol interference—Part II: Performance of the maximum-likelihood detector," *Bell Syst. Tech. J.*, vol. 52, pp. 1521–1539, November 1973.
[10] F. L. Vermuellen and M. E. Hellman, "Reduced state Viterbi decoders for channels with intersymbol interference," in *Proceedings of the International Conference on Communications, Minneapolis, MN*, pp. 37B-1 to 37B-9, June 1974.
[11] G. J. Foschini, "Maximally null response for a MLSE digital data system involving high loss coaxial cable," *IEEE Trans. Inform. Theory*, vol. IT-21, pp. 601–604, November 1975.
[12] S. A. Fredriccson, "A reduced state Viterbi detector for multilevel partial response channels," Royal Institute of Technology, Stockholm, Sweden, Technical Report No. 86, September 1974.
[13] D. Kazakos, "Computational savings and implementation of maximum-likelihood detection in intersymbol interference," *IEEE Trans. Inform. Theory*, to appear.
[14] W. Feller, *An Introduction to Probability Theory and Its Application*, Vol. II. New York: Wiley, 1966, chapters II and VII.

Correspondence

On the Error Probability of General Trellis Codes with Applications to Sequential Decoding

ROLF JOHANNESSON, MEMBER, IEEE

Abstract—**An upper bound on the average error probability for maximum-likelihood decoding of the ensemble of random L-branch binary trellis codes of rate $R = 1/n$ is given which separates the effects of the tail length T and the memory length M of the code. It is shown that the bound is independent of the length L of the information sequence when $M \geq T + [nE_{VU}(R)]^{-1} \log_2 L$. The implication that the actual error probability behaves similarly is investigated by computer simulations of sequential decoding utilizing the stack algorithm. These simulations confirm the implication which can thus be taken as a design rule for choosing M so that the error probability is reduced to its minimum value for a given T.**

I. INTRODUCTION

Massey [1] has defined the class of random trellis codes as a generalization of the type of convolutional code used in Viterbi decoding. He has also proved random upper bounds on the average probability of error for maximum-likelihood decoding of these codes for codes rates less than R_0, where R_0 is less than capacity C of the channel.

In Section II of this correspondence, we extend Massey's bound for trellis codes to all rates less than capacity, but we do this in the context of a more general class of trellis codes for which a distinction can be made between "memory length" and "tail length." The bounds obtained suggest that it is advantageous to use a memory length which is a specified amount greater than the tail length, this amount depending on the length of the trellis. In Section III, we report sequential decoding simulations which confirm this suggestion and which should be useful guides in the design of future sequential decoding systems.

Manuscript received April 9, 1976; revised December 15, 1976. This work was supported in part by the National Aeronautics and Space Administration under NASA Grant NGL 15-004-026 at the University of Notre Dame in liaison with the Communications and Navigation Division of the Goddard Space Flight Center, and in part by the Swedish Board for Technical Development under Grant 75-4165.

The author is with the Department of Automata and General Systems Sciences, University of Lund, Lund, Sweden.

II. UPPER BOUNDS FOR RANDOM TRELLIS CODES

We define an $R = 1/n,L,T,M$ *random trellis code* to be a tree code of length $L + M$ in which each channel input symbol is chosen independently according to a specified probability distribution and with the property that

i) if the preceding M information digits on the paths leading into two nodes at the same depth in the tree coincide, then the same further encoded sequence results whenever the same further information sequence is applied starting from either node, and
ii) all digits on all of the last M–T branches have the same common value.

In other words, the memory or dependence on the past information bits is limited to the M previous information bits, but the useful tail of the tree is only T rather than M branches in length. By allowing $T < M$, we are able, as shown in the sequel, to demarcate rather precisely the different effects of the tail length T and the memory length M on decoding error probability.

Our artifice of requiring all of the digits on the last M–T branches of each path in the trellis to coincide renders these digits "useless" and hence it is unnecessary to transmit them over the discrete memoryless channel being considered. This artifice not only results in a true tail length of only T branches, but also allows us to use with only slight change the bounding techniques normally used for the "usual" trellis codes with $M = T$ [2], [3]. Hence, we give the following Theorem without further proof [4].

Theorem: The average probability of error for maximum-likelihood decoding of the ensemble of binary $R = 1/n,L,T,M$ trellis codes on a discrete memoryless channel satisfies

$$\overline{P[\epsilon]} < c2^{-nTE_{VU}(R)} \left[\frac{1 - 2^{-n(M-T)E_{VU}(R)}}{1 - 2^{-nE_{VU}(R)}} + (L + T - M)2^{-n(M-T)E_{VU}(R)}\right], \quad (1)$$

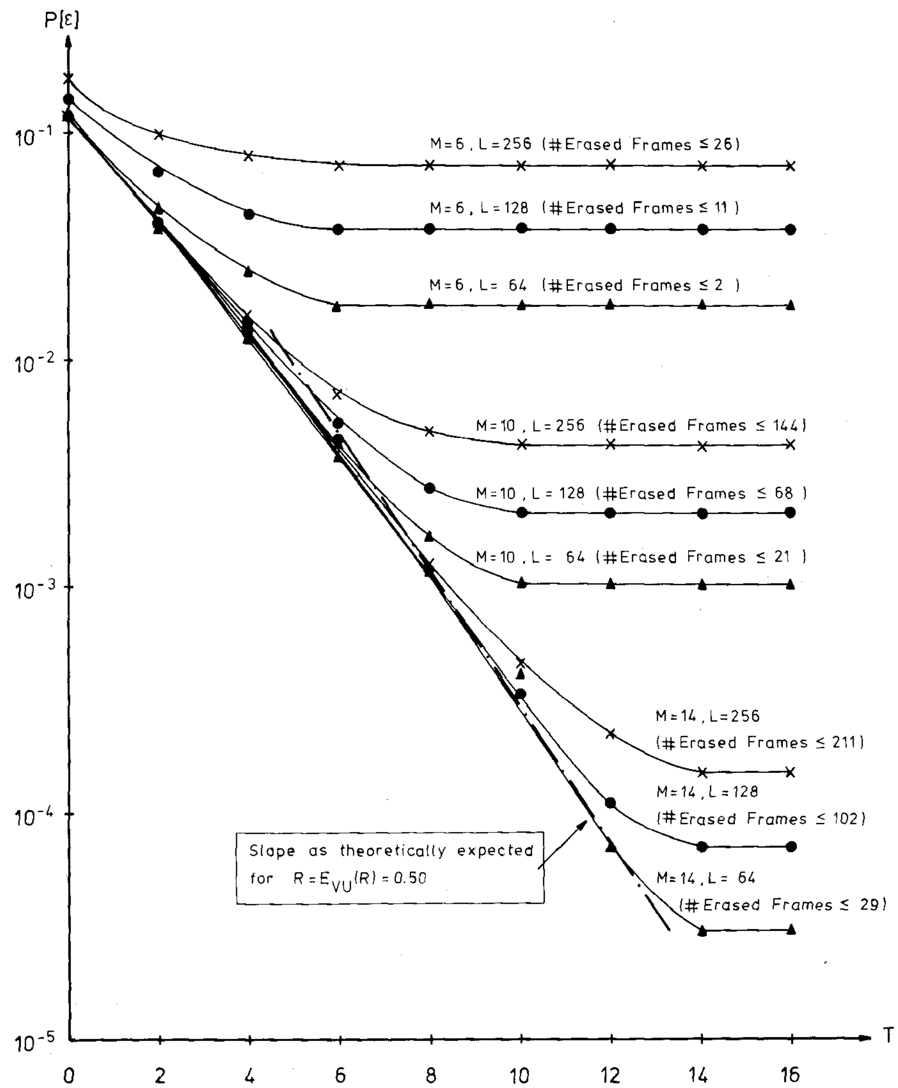

Fig. 1. Decoding error probability obtained from sequential decoding simulations versus tail length of convolutional code (100 000 decoded frames).

where $T \leq M$ and where $E_{VU}(R)$ is Viterbi's upper bound exponent [2], namely

$$E_{VU}(R) = \begin{cases} R_0, & 0 \leq R < R_0 \\ \sup_{\rho} E_0(\rho), & R_0 \leq R < C, \end{cases} \tag{2}$$

where the supremum is taken over ρ such that $0 \leq \rho \leq 1$ and $E_0(\rho) > \rho R$, and where

$$c = \frac{2^{-nE_{VU}(R)}}{1 - 2^{-n(E_0(\rho)-\rho R)}}.$$

The "constant" c depends on R but is independent of L and T. Upon observing that, since $T \leq M$,

$$\frac{1 - 2^{-n(M-T)E_{VU}(R)}}{1 - 2^{-nE_{VU}(R)}} \leq M - T \tag{3}$$

we can state the following.

Corollary 1: The average probability of error for maximum-likelihood decoding of the ensemble of binary $R = 1/n,L,T,M$ trellis codes satisfies

$$\overline{P[\epsilon]} < Lc2^{-nTE_{VU}(R)}, \tag{4}$$

where $E_{VU}(R)$ is given in (2).

In the special case when $T = M$, the bound of Corollary 1 is identical to Viterbi's well-known upper bound for the ensemble of time-varying convolutional codes.

Next, we note that the first term within the brackets in (1) is independent of L, whereas the second term can be made arbitrarily small, for a given L, by increasing M. Thus by choosing that value of M which, for a given L, makes the second term less than one, we have the following.

Corollary 2: The average probability of error for maximum-likelihood decoding of the ensemble of binary $R = 1/n,L,T,M$ trellis codes satisfies

$$\overline{P[\epsilon]} < c'2^{-nTE_{VU}(R)}, \tag{5}$$

provided

$$M \geq T + [nE_{VU}(R)]^{-1} \log_2 L, \tag{6}$$

where

$$c' = \frac{2c}{1 - 2^{-nE_{VU}(R)}} \tag{7}$$

and $E_{VU}(R)$ is given in (2).

The remarkable feature of the bound (5) of Corollary 2 is that it is independent of the length L of the trellis.

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.