On the Probability of Symbol Error in Viterbi Decoders

Amos Lapidoth

***Abstract*—An upper bound is derived on the probability that at least one of a sequence of B consecutive bits at the output of a Viterbi decoder is in error. Such a bound is useful for the analysis of concatenated coding schemes employing an outer block code over GF(2^B) (typically a Reed–Solomon (RS) code), an inner convolutional code, and a symbol (GF(2^B)) interleaver separating the two codes. The bound demonstrates that in such coding schemes a *symbol* interleaver is preferable to a *bit* interleaver. It also suggests a new criterion for good inner convolutional codes.**

***Index Terms*—Concatenated codes, convolutional codes, interleavers, Reed–Solomon codes, transfer function, trellis codes, Viterbi decoding.**

I. Introduction

THIS LETTER deals with the errors that result when convolutionally encoded data are transmitted over a noisy channel and are decoded using the Viterbi algorithm (VA). More specifically, we derive an upper bound on the probability P_B that at least one of a sequence of B consecutive bits at the output of the Viterbi decoder is in error. We assume that the channel is memoryless with binary $\{0, 1\}$ inputs and real valued outputs, and assume that the probability law that governs the channel is symmetric in the sense that for any real number y,

$$p(y|0) = p(-y|1).$$

Such a bound on P_B is useful for the analysis of concatenated coding schemes consisting of an outer blockcode over $\mathrm{GF}(2^B)$, a symbol interleaver (an interleaver over $\mathrm{GF}(2^B)$), and an inner binary convolutional code. In particular, if the decoder first uses the Viterbi algorithm to decode the inner code, then de-interleaves the symbols and finally decodes the outer code, then the performance of the overall system can be bounded based on P_B and the distance profile of the outer code. For example, if the outer code is a $(2^B - 1, 2^B - 2t - 1)$ Reed–Solomon (RS) code over $\mathrm{GF}(2^B)$, then it can correct any pattern of t or fewer symbol errors, and the probability of correct decoding P_{CD} is thus lower bounded by

$$P_{CD} \geq \sum_{\nu=0}^{t} \binom{2^B - 1}{\nu} P_B^{\nu} (1 - P_B)^{2^B - 1 - \nu}. \tag{1}$$

Paper approved by S. G. Wilson, the Editor for Coding Theory and Applications of the IEEE Communications Society. Manuscript received December 20, 1994; revised August 13, 1995. This work was carried out in part while the author was with AT&T Bell Laboratories, Murray Hill, NJ.

The author was with the Information Systems Laboratory, Stanford University, Stanford, CA 94305 USA. He is now with Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA 02139-4307 USA.

Publisher Item Identifier S 0090-6778(97)01402-5.

It is with this motivation in mind that we shall sometimes refer to a sequence of B consecutive bits as a symbol, and refer to P_B as the probability of a symbol error.

The "standard bound" on P_B [see (8)] is based on two bounds: the union of events bound according to which

$$P_B \leq BP_b \tag{2}$$

where P_b is the bit-error rate (BER) at the output of the Viterbi decoder, and the transfer function bound on P_b [9]. Note, however, that for some channels, such as the binary symmetric channel (BSC) and the binary erasure channel, only the union bound is needed as P_b can be computed exactly (see [2] and [6]).

If we were to use *bit* interleaving rather than *symbol* interleaving between the outer and inner encoders, then the probability of a symbol error would be $1 - (1 - P_b)^B = BP_b + O(P_b^2)$, and (2) would hold with equality up to first order terms in P_b. For the concatenated coding scheme under consideration, bit interleaving is thus the worst kind of interleaving. The "standard bound," however, is not sufficiently tight to demonstrate the advantages of the symbol interleaver over the bit interleaver, as it is equally valid for both types of interleavers: It does not capture the bursty nature of the errors made by the Viterbi decoder.

In this letter, we derive a new analytic bound on P_B that is tighter than the "standard bound." Our bound captures the bursty nature of the errors made by the Viterbi decoder, and demonstrates the advantages of the symbol interleaver over the bit interleaver. Our bound indicates that P_B depends on the length of the error events more than on the number of decoding errors that they cause. This demonstrates that in the design of concatenated codes, one should not necessarily choose as the inner code the code which gives rise to the minimum BER; one should take into account the expected length and the probability of an error event.

To the best of our knowledge, all previous estimates of P_B were based on the "standard bound," on simulations, or on *ad hoc* statistical modeling of the errors at the Viterbi decoder's output; see [1] and [3], and the references therein.

II. An Upper Bound on P_B

We describe the upper bound on P_B for a rate $1/n$ feed-forward convolutional encoder with memory m (constraint length $K = m + 1$, or 2^m states) and treat the rate k/n case in the Appendix. Using the assumption that the channel is a binary-input output-symmetric channel, it follows from the linearity of the convolutional code that for the purposes of upper bounding P_B we may assume that the transmitted coded sequence was the all-zero sequence [9]. Consider now an *error string* of length L, i.e., a sequence $\boldsymbol{e} = (e_1, \cdots, e_L)$

such that $e_1 = 1, e_{L-m+1}^m = (0, \cdots, 0)$ and such that for no$0 < L - m + 1$ is e_j^m all zero. Here e_i^j denotes the substring $(e_i, \cdots, e_{i+j-1})$, i.e., the substring of length j starting at index i. Such a string $\boldsymbol{e}$ is thus an L-length sequence of inputs to the convolutional encoder which leads it away from the all-zero state and back to the all zero-state without any intermediate visits to the all-zero state. We shall denote by $\mathcal{E}$ the set of all error strings.

We shall say that an *error event* with error string $\boldsymbol{e}$ occurred at time i if the maximum likelihood state sequence was at the all-zero state at time $i-1$ and then returned to the all-zero state at time $i+L-1$ following the state sequence determined by the error string $\boldsymbol{e}$ in between. Such an error event will be denoted $(\boldsymbol{e}, i)$. Note that we shall say that the event $(\boldsymbol{e}, i)$ occurred irrespective of how the maximum likelihood state sequence arrived at the all-zero state at time $i-1$ or how it continued onward after time $i + L - 1$.

For the purposes of evaluating P_B, we are interested in the Viterbi decoder's steady-state behavior. Since the channel has been assumed stationary, it follows that the probability of the error event $(\boldsymbol{e}, i)$ does not depend on the epoch i. An error event $(\boldsymbol{e}, i)$ will cause an error in the decoding of the block of B consecutive bits that extends from time t to time $t + B - 1$ if and only if (iff) $i \in \mathcal{I}(\boldsymbol{e}, t, B)$, where

$$\mathcal{I}(\boldsymbol{e}, t, B) = \left\{ i : \sum_{j=t-i+1}^{t-i+B} e_j > 0 \right\} \tag{3}$$

and for the purposes of (3) we take $e_j = 0$ for $j \leq 0$ or $j > L$. For example, for $B = 8$, and error string $\boldsymbol{e} = 1\ 0\ 0\ 1\ 1\ 0\ 0\ 0\ 0$, we have $\mathcal{I}(\boldsymbol{e}, t, B) = \{t-4, t-3, t-2, t-1, t, t+1, t+2, t+3, t+4, t+5, t+6, t+7\}$, since an error event with error string $\boldsymbol{e}$ will cause an error in the block starting at t and ending at $t+7$ iff it begins at any point in time in the above set.

Let $N(\boldsymbol{e}, B)$ denote the cardinality of the set $\mathcal{I}(\boldsymbol{e}, t, B)$:

$$N(\boldsymbol{e}, B) = |\mathcal{I}(\boldsymbol{e}, t, B)|.$$

Clearly, the cardinality of $\mathcal{I}(\boldsymbol{e}, t, B)$ does not depend on t and depends only on the error string $\boldsymbol{e}$ and on the block size B. Furthermore, denote the length of the error string $\boldsymbol{e}$ by $L(\boldsymbol{e})$, the Hamming weight of the error string $\boldsymbol{e}$ by $I(\boldsymbol{e})$, and the Hamming weight of the convolutional encoder's output associated with $\boldsymbol{e}$ by $W(\boldsymbol{e})$. Thus, $W(\boldsymbol{e})$ is the Hamming weight of the output sequence which results when one follows the encoder's state diagram, starting at the all-zero state and following the path back to the all-zero state determined by the input string $\boldsymbol{e}$.

Finally, let $P(W)$ denote the probability that a code sequence of Hamming weight W accumulates a higher likelihood than the all-zero code sequence. For example, for the Gaussian channel with antipodal signaling and noise variance $N_0/2$ we have [9]

$$\begin{aligned} P(W) &= Q(\sqrt{2WE_s/N_0}) \\ &\leq Q\left(\sqrt{\frac{2d_f E_s}{N_0}}\right) e^{-(W-d_f)E_s/N_0} \end{aligned} \tag{4}$$

where E_s is the energy which is used to transmit each bit at the convolutional encoder's output, d_f is the code's free Hamming distance, and $Q(x)$ is the probability that a zero-mean Normal random variable of variance 1 exceeds x. For a general channel, a commonly used bound on $P(W)$ is the Bhattacharyya bound [4], [9]:

$$P(W) \leq \tfrac{1}{2} Z^W \tag{5}$$

where

$$Z = \int \sqrt{p(y|0)p(y|1)}\, dy. \tag{6}$$

To continue with the argument, if the block of output bits beginning at time t and ending at time $t + B - 1$ was decoded incorrectly then the maximum-likelihood sequence of states was not the all-zero state for this time interval. Let $i' - 1$ be the most recent time prior to $t + B$ at which the maximum likelihood state sequence was in the all-zero state, and let $\boldsymbol{e}' \in \mathcal{E}$ be the input (error) string that drove it from that time to its next return to the all-zero state. We can now conclude that if the block beginning at t of length B was incorrectly decoded then the event

$$\bigcup_{\boldsymbol{e}' \in \mathcal{E}} \ \bigcup_{i' \in \mathcal{I}(\boldsymbol{e}', t, B)} (\boldsymbol{e}', i')$$

must have happened. Using the union of events bound, the standard bound on the probability of an error event [9], and the fact that by stationarity the probability of the event $(\boldsymbol{e}', i')$ does not depend on i', we conclude that

$$P_B \leq \sum_{\boldsymbol{e} \in \mathcal{E}} P(W(\boldsymbol{e})) N(\boldsymbol{e}, B). \tag{7}$$

A few remarks about this bound are called for.

1) For $B = 1$, we have $P_B = P_b$ and the bound reduces to Viterbi's standard bound on the BER [7], [9], since if $B = 1$, then

$$N(\boldsymbol{e}, B) = I(\boldsymbol{e})$$

and we thus get from (7) that

$$P_b \leq \sum_{\boldsymbol{e} \in \mathcal{E}} P(W(\boldsymbol{e})) I(\boldsymbol{e}). \tag{8}$$

2) Since every "1" (data-bit error) in $\boldsymbol{e}$ contributes exactly B elements to $\mathcal{I}(\boldsymbol{e}, t, B)$, it follows that

$$N(\boldsymbol{e}, B) \leq BI(\boldsymbol{e}). \tag{9}$$

(There is usually a strict inequality in (9) since the sets which the different one's contribute to $\mathcal{I}(\boldsymbol{e}, t, B)$ are not disjoint.) Note that if we were to use (9) to upper bound $N(\boldsymbol{e}, B)$ in (7), for every $\boldsymbol{e} \in \mathcal{E}$, then the resulting bound on P_B would coincide with the "standard bound" of (8) and (2). Improving on (9), even for only one error string, already improves on the "standard bound."

3) A different bound on $N(\boldsymbol{e}, B)$ can be obtained based on $L(\boldsymbol{e})$, the length of the error string: Since every $\boldsymbol{e} \in \mathcal{E}$ terminates with m zeros, it follows that

$$N(\boldsymbol{e}, B) \leq B + L(\boldsymbol{e}) - m - 1 \tag{10}$$

by assuming a worst case where all of the first $L-m$ bits of $\boldsymbol{e}$ are ones.

4) If $B>m-1$, as is the case in most practical applications, then (10) holds with *equality*! The properties of $\boldsymbol{e}$ needed to see this are that $e_1 = e_{L-m} = 1$ and that except for the last m zeros, $\boldsymbol{e}$ does not contain a substring of m consecutive zeros. It follows from $e_1 = 1$ that $\{t, \cdots, t+B-1\} \subseteq \mathcal{I}(\boldsymbol{e}, t, B)$, and from $e_{L-m} = 1$ that $\{t-(L-m)+1, \cdots, t-(L-m)+B\} \subseteq \mathcal{I}(\boldsymbol{e}, t, B)$. Assume now by contradiction that there exists $t-(L-m)+B<i<t$ such that $i \notin \mathcal{I}(\boldsymbol{e}, t, B)$. This would imply that the substring e_{t-i+1}^{B} is all zero, so that $\boldsymbol{e}$ contains an all-zero substring of length B. This leads to a contradiction since, by assumption, $B>m-1$. We conclude that $\{t-(L-m)+B+1, \cdots, t-1\} \subseteq \mathcal{I}(\boldsymbol{e}, t, B)$. The number of elements in the three disjoint subsets of $\mathcal{I}(\boldsymbol{e}, t, B)$ that we have demonstrated is $B+L-m+1$ and thus $N(\boldsymbol{e}, B) \geq B+L(\boldsymbol{e})-m-1$, demonstrating that for $B>m-1$, (10) holds with equality. For $B>m-1$, our best bound is

$$P_B \leq \sum_{\boldsymbol{e} \in \mathcal{E}} P(W(\boldsymbol{e}))(L(\boldsymbol{e})+B-m-1). \tag{11}$$

Note that this bound requires no information about the number of information bit errors that the different error strings cause.

5) Combining bound (11) with the Bhattacharyya bound (5) results in a particularly simple bound. If $T(W,L)$ is the convolutional code's transfer function [9], then

$$P_B \leq \frac{1}{2}\left(\frac{\partial}{\partial L}T(W,L) + (B-m-1)T(W,L)\right)\Bigg|_{W=Z, L=1}. \tag{12}$$

A similar bound for the Gaussian channel is given in (15).

6) Since each "one" in an error event is followed by at most $m-1$ zeros, except for the last "one" which is followed by m zeros, it follows that $L(\boldsymbol{e}) \leq I(\boldsymbol{e})(m-1)+1$ and hence by (10) that

$$N(\boldsymbol{e}, B) \leq I(\boldsymbol{e})(m-1)+B-m. \tag{13}$$

This bound, which is looser than (10), can be used if no information about the error strings' lengths is available. It can compete favorably with (9), when $B>m$.

7) We can bound the different terms in (7) separately. For the more likely error events (those for which $P(W(\boldsymbol{e}))$ is high) we could compute $N(\boldsymbol{e}, B)$ exactly, and for the rest use (9), (13) or (10).

III. A Numerical Example

Consider the rate $1/2$ convolutional code of memory $m=3$, i.e., constraint length four, with generators that are given in octal representation as (13, 17). (This code is isomorphic to the standard (15, 17) code found by Odenwalder; see [5].) The code has free distance $d_f = 6$, and its transfer function is given by

$$T(W,I,L) = \frac{W^6 I^2 L^5 + W^7 I L^4 - W^8 I^2 L^5}{\Delta} \tag{14}$$

where

$$\Delta = 1 - W^2 I^2 L^4 - W^3 I L^3 - W I L^2 - W^4 I^2 L^3 - W I L + W^4 I^2 L^4 + W^2 I^2 L^3.$$

Assume now that the outer RS code is defined over $\mathrm{GF}(2^8)$ so that $B=8$, and that the inner code is used over a Gaussian channel with noise variance $N_0/2$ using an antipodal signaling scheme transmitting energy E_s for every bit at the convolutional encoder's output. Since $m=3$ and $B=8$, (10) holds with equality and we conclude from (11) and (4) that

$$P_B \leq Q\left(\sqrt{\frac{2d_f E_s}{N_0}}\right) e^{d_f E_s/N_0} \left\{\frac{\partial}{\partial L}T(W,L) + (B-m-1)T(W,L)\right\}\Bigg|_{W=e^{-E_s/N_0}, L=1}. \tag{15}$$

The "standard bound" that is based on (2) and (8) takes the form:

$$P_B \leq BQ\left(\sqrt{\frac{2d_f E_s}{N_0}}\right) e^{d_f E_s/N_0} \cdot \left\{\frac{\partial}{\partial I}T(W,I)\right\}\Bigg|_{W=e^{-E_s/N_0}, I=1}. \tag{16}$$

The two bounds (15) and (16) are compared in Fig. 1, where the probability of a symbol error $(B=8)$ is plotted versus E_b/N_0. Since the code is of rate 1/2, the energy per transmitted information bit E_b is $E_b = 2E_s$. To estimate the tightness of the bounds, simulation results are also plotted.

At high signal-to-noise ratios (SNR's), the dominant error string, $\boldsymbol{e_0}$, is the one that attains the free distance of the code. For the code under consideration this error string is of length five and causes two decoding errors. The coefficient by which $P(W(\boldsymbol{e_0}))$ is multiplied in our bound (15) is $B+L(\boldsymbol{e_0})-m-1 = 8+5-3-1 = 9$, whereas in (16) the multiplying coefficient is $BI = 8 \times 2 = 16$. Asymptotically then, our bound improves on the standard bound by a factor of 9/16. This translates to an improvement by a factor of roughly $(9/16)^{t+1}$ in estimating $1-P_{CD}$ from (1). The improvement is even larger at low SNR's.

Appendix

We now treat the rate k/n case. We can think of the code's state machine as being driven by binary k-tuples in $\mathrm{GF}(2^k)$, and define error strings as finite sequences over $\mathrm{GF}(2^k)$ that begin by departing from the all-zero state and return to it exactly once—at the end of the sequence. For any epoch t and error event $\boldsymbol{e} = (e_1, \cdots, e_L)$,where e_j is a k-tuple, i.e.,

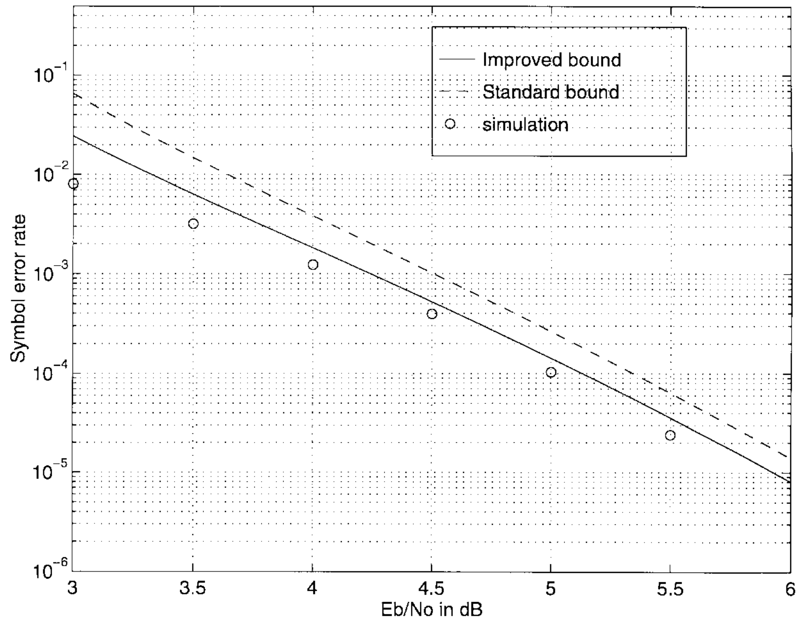

Fig. 1. Probability P_B of an error in at least one of $B = 8$ consecutive bits at the output of a Viterbi decoder for the rate 1/2 convolutional code with generators (13,17) transmitted over the Gaussian channel.

$e_j = (e_{j,1}, \cdots, e_{j,k})$ we define

$$\mathcal{I}(\boldsymbol{e},t,B)=\Bigg\{i\colon \sum_{\kappa=t \bmod k}^{k} e_{\lceil t/k\rceil-i+1,\kappa} + \sum_{\lceil t/k\rceil-i+1<j<\lceil t+B/k\rceil}\ \sum_{\kappa=1}^{k} e_{j,\kappa} + \sum_{\kappa=1}^{t+B-1 \bmod k} e_{\lceil t+B/k\rceil,\kappa}>0\Bigg\}$$

where $x \bmod k$ is taken to return a value between one and k. We also define

$$N(\boldsymbol{e},t,B)=|\mathcal{I}(\boldsymbol{e},t,B)|.$$

Note that when $k \neq 1$, the cardinality of $\mathcal{I}(\boldsymbol{e},t,B)$ may, in fact, depend on t (actually, on $t \bmod k$).

Defining $W(\boldsymbol{e})$ as before to be the Hamming weight of the encoder's output when driven by $\boldsymbol{e}$ from the all-zero state we obtain that $P_{B,t}$, the probability of incorrect decoding of the block of B bits that begins at time t, is upper bounded by

$$P_{B,t} \le \sum_{\boldsymbol{e}\in\mathcal{E}} P(W(\boldsymbol{e}))N(\boldsymbol{e},t,B)$$

where as before $\mathcal{E}$ is the set of all error strings.

Note that the analysis of trellis codes is straightforward too. The only additional difficulty one encounters is that one cannot in general assume that the correct sequence is the all-zero, and one has to average $P(W(\boldsymbol{e}))$ over all possible correct sequences [10], [8].

Acknowledgment

The author acknowledges the help of M. Ho and E. Telatar in debugging the simulation program. He also thanks J. Mazo, P. O'Neill, S. Wilson, and the anonymous referees for their help in clarifying the text.

References

[1] T. Berman, J. Freedman, and T. Kaplan, "An analytic analysis of a concatenated Reed–Solomon (RS)/Viterbi coding system both with and without RS interleaving," in *Proc. IPCCC'92*, Scottsdale, AZ, 1992. pp. 3.4.3.1–3.4.3.7.

[2] M. V. Burnashev, and D. L. Cohn, "Symbol error probability for convolutional codes," *Probl. Peredači Inform.*, vol. 26, no. 4, pp. 3–15, 1990.

[3] K. Fazel, "Application of error modeling at the output of maximum likelihood decoder to concatenated coded 16 PSK," in *Proc. GLOBECOM'89*, Dallas, TX, pp. 1528–1533.

[4] M. E. Hellman, and J. Raviv, "Probability of error, equivocation, and the Chernoff bound," *IEEE Trans. Inform. Theory*, vol. IT-16, pp. 368–372, July 1970.

[5] R. Johannesson, "Robust optimal rate one-half binary convolutional codes," *IEEE Trans. Inform. Theory*, vol. IT-21, pp. 464–468, July 1975.

[6] A. Lapidoth, "The performance of convolutional codes on the block erasure channel using various finite interleaving techniques," *IEEE Trans. Inform. Theory*, vol. 40, pp. 1459–1473, Nov. 1994.

[7] J. Massey, *Applied Digital Information Theory*, course notes for the Winter semester 1994/95, ETH, Zurich, Switzerland.

[8] M. Rouanne and D. J. Costello, "An algorithm for computing the distance spectrum of trellis codes," *IEEE J. Select. Areas Commun.*, vol. 7, pp. 929–940, Aug. 1989.

[9] A. J. Viterbi, and J. K. Omura, *Principles of Digital Communication and Coding*. New York: McGraw-Hill, 1979.

[10] E. Zehavi and J. K. Wolf, "On the performance evaluation of trellis codes," *IEEE Trans. Inform. Theory*, vol. IT-33, pp. 196–202, Mar. 1987.