

A BOUND ON VITERBI DECODER ERROR BURST LENGTH

S. J. CURRY and W. D. HARMON

Aerospace Corp.
El Segundo, California

Summary. A maximum likelihood (Viterbi) decoder used with a convolutional code on a Gaussian channel produces decoding errors which tend to occur in clusters or bursts. A method is described for deriving an upper bound on the probability of occurrence of error bursts of a given length. The method applies to the optimum convolutional codes found by Odenwalder,¹ for which the codeword weight distribution is partially known. Laboratory measurements of error burst length at signal-to-noise ratios greater than 4 dB indicate that the upper bound is useful for approximating the length of high-probability bursts, but is not precise enough to predict the probability of very long, low-probability bursts.

Introduction. In some applications, the length of error bursts produced by a Viterbi decoder can be an important consideration. One example is a concatenated code in which the inner code consists of a convolutional code with a Viterbi decoder. The outer encoder and decoder should be designed to correct most of the Viterbi decoder error bursts. An approximation to the probability distribution of the length of these bursts is needed before the outer decoder, which may include interleaving, can be designed.

Most previous derivations of burst length probabilities are based upon randomly chosen time-varying codes. The derivation in this paper differs in that it applies to the optimum short constraint length codes, which are non-time-varying.

Convolutional Code Structure. As a preliminary, some well-known properties of convolutional codes will be reviewed. Refer to Figure 1, which shows an encoder having constraint length 7 and rate 1/2. In general, constraint length is designated K , and rate $R = 1/V$.

Assume that the input sequence X has the following properties: (a) The first non-zero information bit is X_1 , (b) the last non-zero bit is X_L , and (c) between X_1 and X_L there is no run of $K-1$ consecutive zeros. The resulting encoder output sequence, consisting of $V(L + K - 1)$ code symbols is the codeword Y .

¹ J. P. Odenwalder, Optimal Decoding of Convolutional Codes, Ph. D. Thesis, UCLA, 1970.

Definition 1: The Hamming weight of codeword Y is the number of ones appearing in Y.

Definition 2: The free distance of a code is the minimum weight of a non-zero codeword, where the minimization is done over all possible non-zero input sequences.

For the optimum² rate 1/2 codes of constraint length 3 through 9, and rate 1/3 codes of constraint length 3 through 8, Odenwalder computed the Hamming weight distribution for the low-weight codewords. Table 1 summarizes this information for the $K = 7$, $R = 1/2$ code. The number of non-zero codewords having Hamming weight 10 through 16 is listed. (Since the free distance of this particular code is 10, there are no non-zero codewords of weight less than 10.)

Codeword Weight	Number of Codewords	
10	11	
12	38	
14	193	Code
16	1331	{

1	0	0	1	1	1	1
1	1	0	1	1	0	1

**Table 1. Codeword Weight Spectrum for
 $K=7$, $R=1/2$ Optimum Code**

Error Events. It is assumed throughout that the channel has additive white Gaussian noise with single-sided spectral density N_0 . The signal consists of coherent bi-phase shift keying with signal-to-noise ratio specified by E_b/N_0 , where E_b is the signal energy per information bit. The convolutional code is decoded by the ideal (maximum likelihood) Viterbi decoding algorithm. The example of the $K = 7$, $R = 1/2$ optimum code will be used, but the generalization to other codes is obvious. It is necessary to treat each code separately since the codeword weight spectrum is unique to that code.

It is assumed, without loss of generality, that the all-zero codeword is transmitted. (Since the code is linear, error bursts do not depend on the particular codeword transmitted, and the all-zero codeword can be assumed.)

An error burst in the decoded data consists of a cluster of errors preceded and followed by a guard space (error-less span) of at least $K - 1$ bits, where K is the code constraint length. There is a unique error burst corresponding to each non-zero codeword. Thus, the error bursts can be ranked according to the weight of the corresponding codeword.

² Optimum in the sense of providing minimum bit error rate.

In our example of the $K = 7, R = 1/2$ code, there are codewords of weight 10, 12, 14, 16, A 1st-order error burst corresponds to a codeword of weight 10, a 2nd-order burst to a codeword of weight 12, and so on for higher orders.

A Bound on Burst Length. For each order burst a bound on the burst length can now be derived. Two more definitions are needed.

Definition 3: The minimum distance, d_{\min} , of a convolutional code of constraint length K is the minimum Hamming weight of the first K codeword branches,³ where the minimum is taken over all non-zero codewords.

Definition 4: The reverse minimum distance, d'_{\min} , of a convolutional code is the minimum distance of the reverse code. (In our example the code has subgenerators $\begin{Bmatrix} 1001111 \\ 1101101 \end{Bmatrix}$ and the reverse code has subgenerators $\begin{Bmatrix} 1111001 \\ 1011011 \end{Bmatrix}$)

For a short constraint length code, d_{\min} and d'_{\min} can be computed either by inspection or by a simple computer search algorithm. For the $K = 7, R = 1/2$ code, $d_{\min} = 5$ and $d'_{\min} = 4$.

The following theorem is proved in the Appendix.

Theorem 1:⁴ Suppose a non-catastrophic convolutional code has constraint length K , minimum distance d_{\min} , and reverse minimum distance d'_{\min} . An error burst corresponding to a codeword of weight W has a burst length, L , which can be upper bounded by

$$L < K + [W + 1 - d_{\min} - d'_{\min}] (K - 1) \quad (1)$$

It is assumed that $W > d_{\min} + d'_{\min}$ and a maximum likelihood decoder is used.

To illustrate the use of this theorem, consider the $K = 7$ code, which has $d_{\min} = 5$ and $d'_{\min} = 4$. The codewords corresponding to 1st-order error bursts have Hamming weight $W = 10$. Using Theorem 1, the bound on the length of these first order bursts is found to be $L_1 \leq 19$ bits.⁵ The bound can also be evaluated for higher order error bursts as

³ A codeword branch consists of V code symbols, as shown in Figure 1.

⁴ S. J. Curry, Selection of Convolutional Codes Having Large Free Distance, Ph.D. Thesis, UCLA, 1971.

⁵ It can be shown by an exhaustive search for 1st-order error bursts that the actual maximum burst length is 10 bits.

$$L_n < 7 + 12n \quad (2)$$

where n = order of the error burst.

Error Burst Probabilities. The Viterbi decoder error bursts have been ranked in different orders, where the ranking reflects the relative probability of occurrence. A union bound on this probability will now be calculated.

Again suppose the all-zero codeword is transmitted, and suppose the decoder has correctly decoded all previous data. Let p_W be the probability that an error burst corresponding to one particular codeword of weight W begins on a given bit. Viterbi⁶ has shown that for a maximum likelihood decoder

$$p_W = \text{erfc} \left\{ \sqrt{\frac{2W}{V} \frac{E_b}{N_o}} \right\}$$

Let P_n be the probability that an n -th order error event begins on a given decoded bit. It is assumed that the number of n -th order code-words, N_n , is known. Using the union bound on the probability of N_n events which are not disjoint, it follows that

$$P_n \leq N_n p_{W_n} = N_n \text{erfc} \left\{ \sqrt{\frac{2W}{V} \frac{E_b}{N_o}} \right\} \quad (3)$$

where W_n is the weight of an n -th order codeword.

Equation (3) gives an upper bound on the probability that an n -th order burst will occur, and Equation (2) gives a bound on the maximum length for that burst. Burst probabilities and bounds on burst lengths for orders 1 through 4 are shown in Table 2.

Burst Order n	Bound on Burst Length L_n (Bits)	Burst Probability Bound, P_n		
		$\frac{E_b}{N_o} = 4 \text{ dB}$	$\frac{E_b}{N_o} = 5 \text{ dB}$	$\frac{E_b}{N_o} = 6 \text{ dB}$
1	19	2.6×10^{-6}	1.0×10^{-7}	1.5×10^{-9}
2	31	7.9×10^{-7}	1.4×10^{-8}	9.3×10^{-11}
3	43	3.0×10^{-7}	2.8×10^{-9}	8.1×10^{-12}
4	55	1.6×10^{-7}	7.7×10^{-10}	9.8×10^{-13}

Table 2. Bounds on Burst Length and Probability of Occurrence

⁶ A. J. Viterbi, "Convolutional Codes and Their Performance in Communication Systems," IEEE Trans. on Comm. Tech., COM-19, October 1971.

It is assumed that bursts of order 5 or higher occur so rarely that they comprise an insignificant percentage of the total bursts. This assumption is supported by the computed bounds on burst probability in Table 2. Using the computed values in Table 2, an approximate distribution of burst lengths can be constructed, as shown in Table 3.

Burst Length L (Bits)	Percentage of Bursts with Length < L		
	$\frac{E_b}{N_o} = 4 \text{ dB}$	$\frac{E_b}{N_o} = 5 \text{ dB}$	$\frac{E_b}{N_o} = 6 \text{ dB}$
19	67%	85%	94%
31	88%	97%	99%
43	96%	99%	99.9%
55	≈ 100%	≈ 100%	≈ 100%

Table 3. Approximate Burst Length Distribution

Comparison with Experimental Data. Burst length measurements were made at various signal-to-noise ratios for the Linkabit LV 7026 Viterbi decoder. The decoder was operated at a data rate of 1 Mbps on a Gaussian channel. The encoder/decoder has rate 1/2, constraint length 7, and uses the optimal code. The mean and standard deviation of the burst length were measured at various signal-to-noise ratios using large data samples with the results shown in Table 4. In addition, the exact length of 100 bursts was measured. A histogram of measured burst length can be constructed, as shown in Figure 2. This data indicates that the approximate burst length distribution given in Table 3 is pessimistic in that it predicts a higher percentage of long burst (length greater than 19 bits) than was actually observed.

E_b/N_o (dB)	Mean Burst Length (Bits)	Std. Dev. (Bits)
4.0	7.3	4.8
4.5	6.3	4.4
5.0	5.9	3.6
5.5	5.6	3.1

Table 4. Measured Burst Statistics for K=7, R=1/2 Viterbi Decoder

To compare predicted maximum burst length with the measurements, consider the case of $E_b/N_o = 4 \text{ dB}$. The maximum measured burst length was 34 bits (over a data sample of 200 bursts). From Table 3, the predicted maximum length is 55 bits, excluding bursts of order 5 or higher, which occur rarely. However, if a much larger data sample is used, one would expect to observe a few bursts of very long length, possibly exceeding 55 bits. The method of deriving the bound cannot predict these rare events because the codeword weight structure is not known with enough precision.

Conclusions. A method of deriving an upper bound on the probability of Viterbi decoder error bursts of a given length has been described. The bound is derived for a specific convolutional code, using knowledge of the codeword weight distribution. It was evaluated for a particular case, the $K = 7$, $R = 1/2$ optimal convolutional code. The method is limited to short constraint length codes, since these are the only ones for which the codeword weight distribution is known.

Appendix: Proof of Theorem 1.⁷ Referring to Figure 1, the codeword Y consists of $L + K - 1$ branches, where each branch contains 2 symbols. Let $W(Y)$ be the Hamming weight of the codeword. Then

$$W(Y) = \sum_{i=1}^{L+K-1} W(Y_i)$$

when Y_i is the i -th codeword branch.

Assume that the input sequence length satisfies the inequality

$$L \geq K + 1 + [W(Y) + 1 - (d_{\min} + d'_{\min})] (K - 1)$$

But,

$$\begin{aligned} W(Y) &= \sum_{i=1}^{L+K-1} W(Y_i) \\ &= \sum_{i=1}^K W(Y_i) + \sum_{i=K+1}^{L-1} W(Y_i) + \sum_{i=L}^{L+K-1} W(Y_i) \end{aligned}$$

By definition,

$$\sum_{i=1}^K W(Y_i) \geq d_{\min}$$

and

$$\sum_{i=L}^{L+K-1} W(Y_i) \geq d'_{\min}$$

Thus,

$$W(Y) \geq d_{\min} + d'_{\min} + \sum_{i=K+1}^{L-1} W(Y_i)$$

⁷ Curry, *op. cit.*

The above assumption on L implies that

$$L - (K+1) \geq [W(Y) + 1 - (d_{\min} + d'_{\min})] (K-1)$$

Non-catastrophic convolutional codes have the property that any (K-1) consecutive branches of a non-zero codeword must have Hamming weight at least one.⁸ But,

$$\sum_{K+1}^{L-1} W(Y_i)$$

is the sum of Hamming weights of L-(K+1) codeword branches, where L-K+1 satisfies the above inequality. Thus,

$$\sum_{K+1}^{L-1} W(Y_i) \geq \left\lceil \frac{L-(K+1)}{K-1} \right\rceil \geq W(Y) + 1 - (d_{\min} + d'_{\min})$$

where $\lceil \cdot \rceil$ is the next smallest integer operator. It follows that $W(Y) \geq W(Y) + 1$, which is a contradiction. Thus, the assumption on L is invalid, and it follows that

$$L < K + [W(Y) + 1 - (d_{\min} + d'_{\min})] (K-1).$$

This proves the theorem.

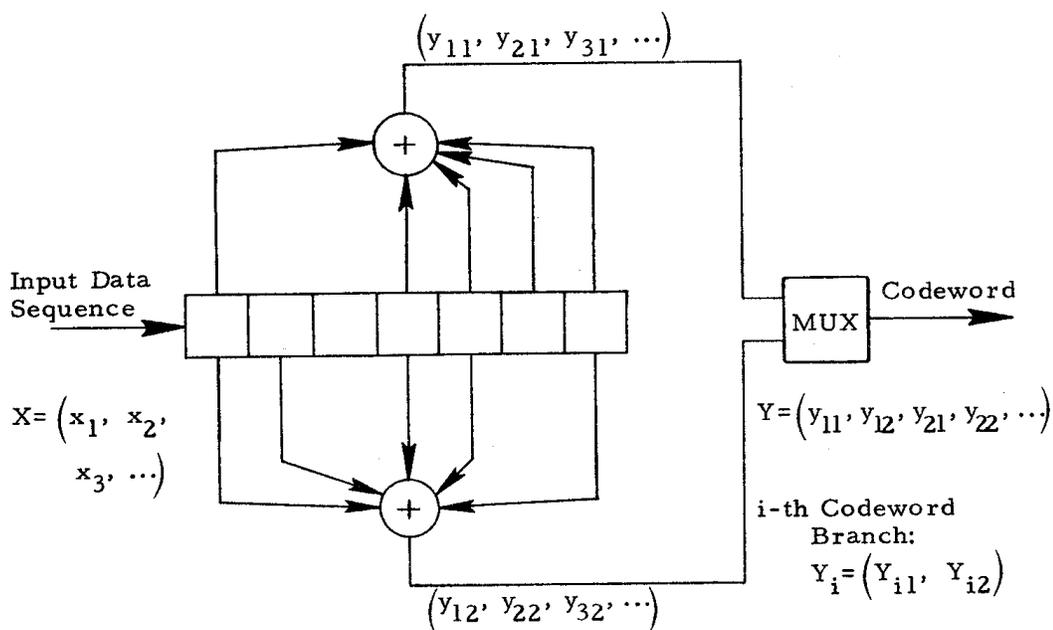


Figure 1. Convolutional Encoder with Subgenerators $\begin{Bmatrix} 1001111 \\ 1101101 \end{Bmatrix}$

⁸ Odenwalder, *op. cit.*

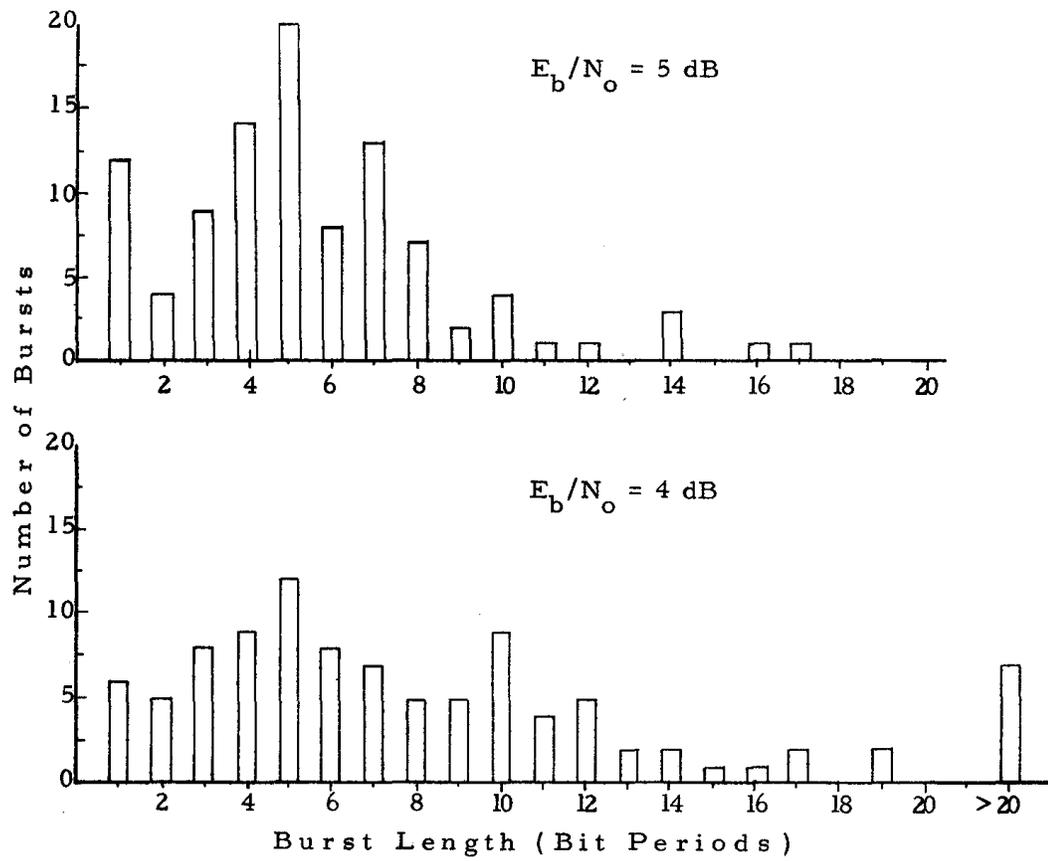


Figure 2. Burst Length Distribution from Measurements