WORD ERROR PERFORMANCE OF VITERBI AND TAILBITING DECODERS

M. A. King
The Aerospace Corporation
El Segundo, CA 90045

INTRODUCTION

One of the main advantages of convolutional codes is the existence of soft decision maximum likelihood decoders of manageable complexity. A disadvantage of convolutional codes is that the standard implementation requires that the encoder be flushed after each data sequence. This requirement adds a significant amount of overhead when short sequences are to be encoded.

Consider an m-symbol information sequence that is to be encoded using a rate 1/v constrain length k convolutional encoder. The number of channel symbols that would be required by the standard encoding would be v(m + k - 1). Thus, the actual effective code rate is

$$\frac{1}{v}_{eff} = \frac{m}{v(m + k - 1)} \tag{1}$$

Therefore, a rate 1/2 constraint length 3 encoder operating on a 10 symbol sequence has an effective rate of only 0.417. This is only 93% of the listed code rate of 0.5. Thus, either the channel symbols must be shortened to 41.7% of the information symbol rate, or the information rate must be slowed by (100 - 83) = 17%.

Tailbiting is a technique that eliminates the need to flush the convolutional encoder. Thus, the benefits of soft decision decoding can be gained without the additional overhead of the flush sequence. The penalties are an increase in the number of required decoder operations, and a slight decrease in the error performance of the decoder. This technique simply involves the preloading of the tail of the information sequence into the encoder. Thus, the register of a constraint length k encoder is preloaded with the last k symbols on the information sequence. This establishes the initial state of the encoder. As the information sequence is fed into the encoder, the encoder traces a path through the code trellis and ends in the same state in which it began. Thus, the complete code trellis is altered from a “trestle bridge” -like structure, expanding from the all-zeros state, and then contracting back into it, to a “Chinese handcuff”-like structure of interwoven loops.

This paper compares the performance of the tailbiting technique to standard encoding for 10 symbol information words. The 10-symbol words are assumed to be chosen randomly with a uniform distribution from an 8-ary alphabet. Two codes will be considered: the rate 1/2 dual-3, and rate 1/2 triple-3 convolutional codes. Schematic diagrams of encoder implementations for these codes are given as Figures 1 and 2, respectively. In addition, it is assumed that each code symbol is separated into four equal time segments and an orthogonal synchronization code is impressed onto the channel symbol sequence. All computations will assume the transmission medium to be an additive white Gaussian noise (AWGN) channel.

DERIVATIONS

It is well known that it is infeasible to compute the exact performance of most convolutional codes because of the complicated relationships between paths. The standard approach, and the approach that will be used here, is to bound the code performance using union bound arguments.

Assume without loss of generality that the transmitted code path is the all-zeros code path. Consider a second sequence that diverges from the all-zeros path and then remerges at some later node. Let n denote the Hamming weight of this second path with respect to (wrt) the all-zeros path, and p_e (n) denote the probability that a path of this weight is mistaken or the correct path by the decoder. Clearly the probability of a correct decision between these two paths is given by

$$p_c(n) = 1 - p_e(n). \qquad (2)$$

Consider the family of all code paths remerging with the all-zeros path at the node i. This family can be characterized as having a spectrum of Hamming weights. In particular, let N(n, i) denote the number of paths in this family of weight n. A lower bound on the probability that none of the paths in this family causes an error is given by

$$P_c(i) \geq \prod_{n=1}^{\infty} (1 - p_e(n)^{N(n,i)} \qquad (3)$$

This is a bound instead of an equality because the remerging paths will be correlated, and the knowledge that any one of them has failed to cause an error can only decrease the probability that the others will succeed in causing an error.

A lower bound on the overall probability of correctly decoding a data word for a Viterbi decoder is the product of the probabilities of making correct decisions at each node.

$$
\begin{aligned}
P_c &\geq \prod_{i=1}^{M} P_c(i) \\
&\geq \prod_{i=1}^{M} \prod_{n=1}^{\infty} \left(1 - p_e(n)\right)^{N(n,i)} \qquad (4) \\
&= \prod_{n=1}^{\infty} (1 - p_e(n))^{\sum_{i=1}^{M} N(n,i)}
\end{aligned}
$$

Equation (4) provides an upper bound on the overall probability of word error for a Viterbi decoder,

$$
\begin{aligned}
P_E &= 1 - P_c \\
&\leq 1 - \prod_{n=1}^{\infty} \left(1 - p_e^{(n)}\right)^{\sum_{i=1}^{M} N(n,i)} \qquad (5)
\end{aligned}
$$

The bounds of Equations (3) - (5) can be easily identified as marginally tighter versions of the standard union bounds used by Viterbi [1]. In particular, from Equation (3)

$$
\begin{aligned}
P_E(i) &= 1 - P_c(i) \\
&\leq 1 - \prod_{n=1}^{\infty} \left(1 - p_e^{(n)}\right)^{N(n,i)} \\
&\leq \sum_{n=1}^{\infty} N(n,i) p_e(n) \qquad (6) \\
&= T\left(p_e(n)\right) ,
\end{aligned}
$$

where T(·) is the moment generating function with the substitution

$$D^n = P_e (n) \qquad (7)$$

The expression in Equation (6) is the usual bound on first event error probability - in this case, referenced to a particular mode i. Equation (5) merely extends the union bound arguments to cover all paths at each node of the decoding procedure. It is clear that the results will be computationally equivalent to Viterbi's when P_e (n) 1.

With standard Viterbi decoder implementations, all paths must eventually remerge. This is because the final code state is known. Thus, all error events occur at path mergers. When tailbiting is used, however, the beginning and ending states are still equal, but unknown, a priori. Therefore, with tailbiting, significant numbers of legitimate but erroneous paths exist in the modified trellis that never merge with the correct path. These paths cannot be referenced to any particular decoding node, but will have their effect at the end of the decoding process. The lowest weight error paths are still among those that merge with the correct path. However, because of the circular nature of the modified tailbiting trellis, and the fact that the tailbiting decoder will make more than a single pass through each node, the spectrum of paths remerging at all decoding nodes becomes uniform. In particular, the values of N(n, i) for a given n, become equal for all i at the largest value achieved by the standard implementation. Let this value be denoted N(n), where

$$\hat{N}(n) = \max_{i} N(n,i). \tag{8}$$

The bound of Equation (5) can be simplified for the case of tailbiting to read,

$$P_E^{(TB)} \leq 1 - \prod_{n=1}^{\infty} \left(1 - p_e(n)\right)^{M\hat{N}(n) + I(n)} \tag{9}$$

where I(n) is the number of legitimate code paths of weight n that never merge with the correct path.

COMPUTATIONS

The path error probability, p_e (n), can be computed for any particular signal to noise ratio as [2].

$$p_e(n) = \int_0^{\infty} p_{\gamma}^{(n)}(\xi) \int_0^{\xi} p_x^{(n)}(\psi)\, d\psi\, d\xi \tag{10}$$

where

$$p_x^{(n)}(\psi) = 1/2 \exp \left| - (\psi + p)/2 \left(\frac{\psi}{p}\right)^{\frac{4n-1}{2}} I_{4n-1} \left(\sqrt{\psi p}\right), \right.$$

$$p = \text{synchronization symbol signal to noise ratio} \tag{11}$$

$I_K(\cdot) = K^{th}$ order modified Bessel's function, where

$\Gamma(L)$ = Gamma function with argument L. A computer program has been written that is able to compute p_e (n) for the required values of n and signal to noise ratio.

The values of N(n, i) can be obtained from the code's moment generating function, if it is known, or from a computer search. Odenwalder [3] has derived the form of the generating function for a family of rate (1/v) dual - k convolutional codes. For the rate 1/2 dual - 3 the function is

$$T(D,L) = \frac{7D^4L^2}{1 - L(2D + 5D^2)} \qquad (12)$$

The valu4s of N(n, i) are the values of the coefficients of the D^nL^i terms of the power series expansion of T(D,L). These values are given in Table 1 for n 10. The triple - 3 code is much more complicated in path structure than the dual - 3, and no generating function is known [4]. However, a computer search of the code trellis has identified all remerging paths of weight n 10. The values of N(n, i) for the triple - 3 code for n 10 are given in Table 2. The exponents in Equations (5) and (9) are summarized in Table 3 for n 10 and M = 10, for both the dual - 3 and triple - 3 codes.

The expressions of Equations (5) and (9) have been evaluated for several values of bit signal energy to noise ratio (E_b /N_o) for both the dual - 3 and the triple -3, and for both constant information rate and constant channel symbol rate. The results are portrayed in Figures 3 and 4.

It can be inferred from the path weight summary of Table 3 that the smaller numbers of higher weight paths will cause the standard implementation to have superior performance than the tailbiting implementation for the same channel symbol rate. This is seen to be true in Figures 3 and 4. However, the difference is slight and disappears as E_b /N_o increases. Holding the channel symbol rate constant means that the information rate for the standard or Viterbi implementation must be slower than the tailbiting implementation by 20/24 for the triple - 3 and 20/22 for the dual - 3 codes.

When the information rate is held constant, the channel rate for the standard implementation must be increased by the above ratios. This means that individual channel symbols will be shorter and the channel symbol to noise ratio will be reduced. This reduction will cause a performance degradation of about 0.8 dB for the triple - 3 and 0.4 dB for the dual - 3 coded.

CONCLUSIONS

Figures 3 and 4 show that when the channel symbol rate is held constant, there is no significant degradation due to the use of the tailbiting technique for E_b/N_o 8 dB. Thus, for a reasonably high signal to noise ratio, the additional over head of the flush sequence is avoided with little adverse effect on the word error probability. This is highlighted by comparison of the constant data rate performance curves. The advantage of the tailbiting implementation over the standard Viterbi implementation for constant data rate is essentially the same as their difference in channel symbol period. Thus, the additional paths are not significantly effecting decoder performance.

Weight n	NODE 2	3	4	5	6	7	≥ 8
0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0
4	7	7	7	7	7	7	7
5	0	14	14	14	14	14	14
6	0	35	63	63	63	63	63
7	0	0	140	196	196	196	196
8	0	0	175	595	707	707	707
9	0	0	0	1050	2170	2394	2394
10	0	0	0	875	5075	7875	8323

Dual - 3 Generating Function Coefficients

TABLE 1

Weight n	NODE 3	4	5	6	7	≥ 8
0	0	0	0	0	0	0
1	0	0	0	0	0	0
2	0	0	0	0	0	0
3	0	0	0	0	0	0
4	0	0	0	0	0	0
5	0	0	0	0	0	0
6	7	7	7	7	7	7
7	0	28	28	28	28	28
8	0	21	112	119	119	119
9	0	0	112	364	392	392
10	0	0	189	665	1358	1442

Triple - 3 Generating Function Coefficients

TABLE 2

Weight n	Triple - 3		Dual - 3	
	Tail - biting	Viterbi	Tail - biting	Viterbi
4	0	0	70	70
5	0	0	140	126
6	70	70	630	539
7	280	252	1960	1512
8	1190	917	7070	5012
9	5691	2828	23940	15190
10	42280	9422	84077	47117

Exponent Summary
TABLE 3

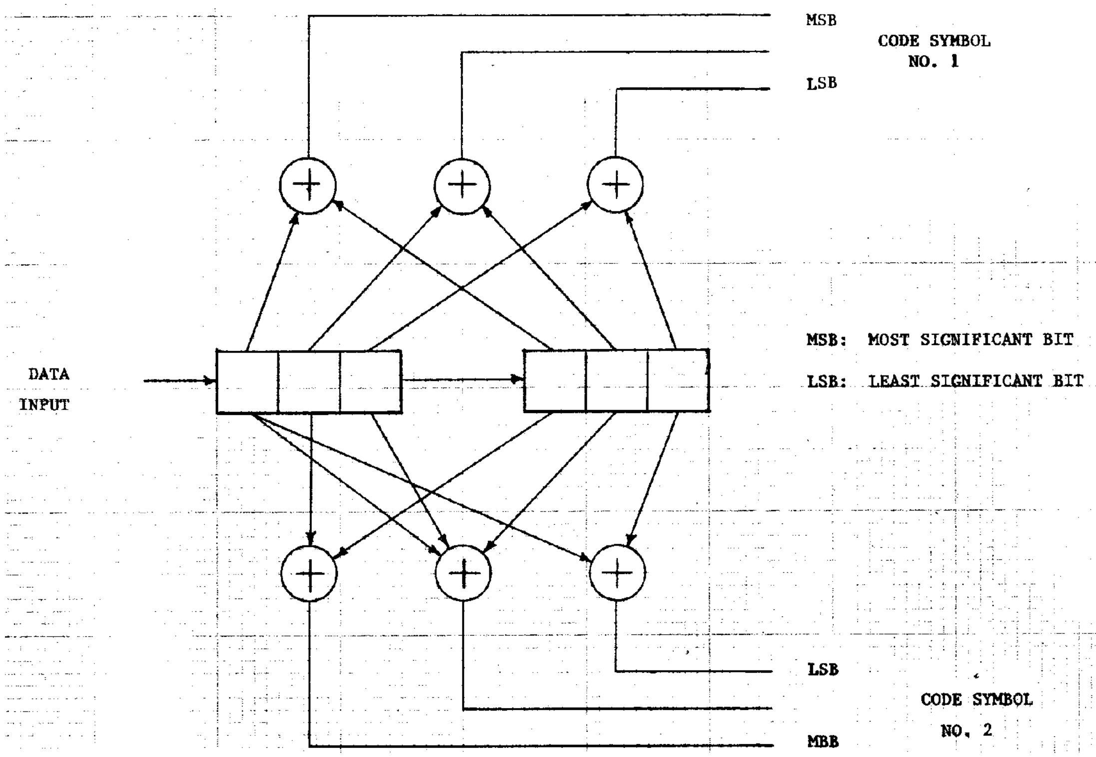


FIGURE 1. RATE ½ DUAL-3 ENCODER

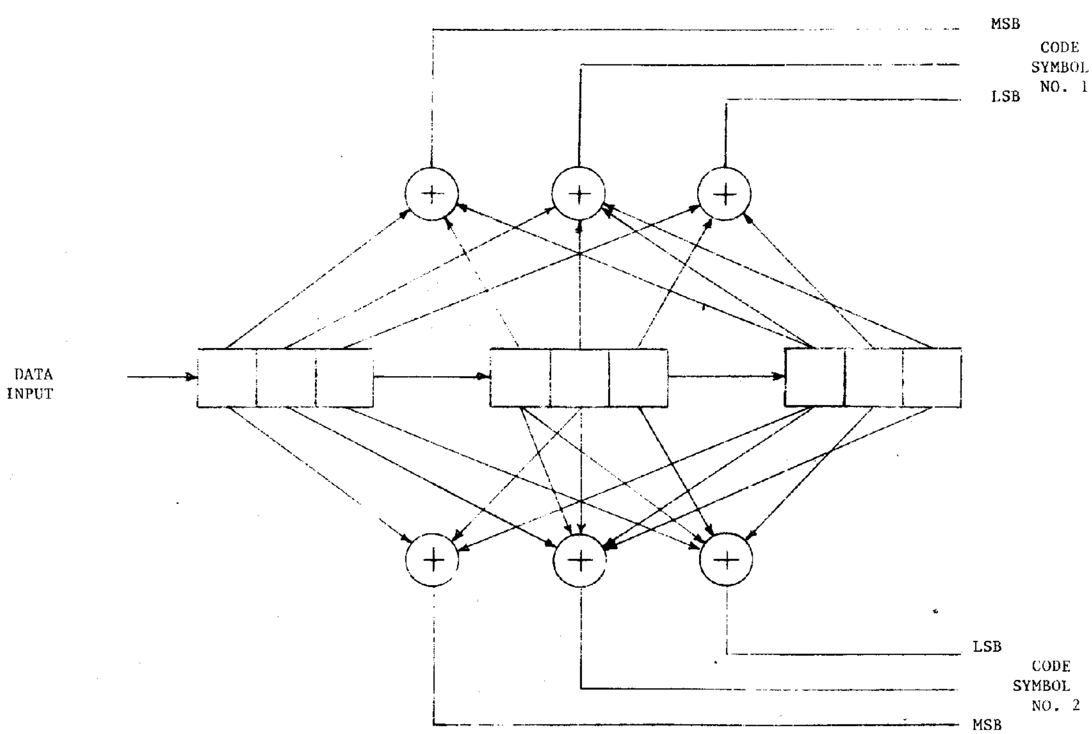
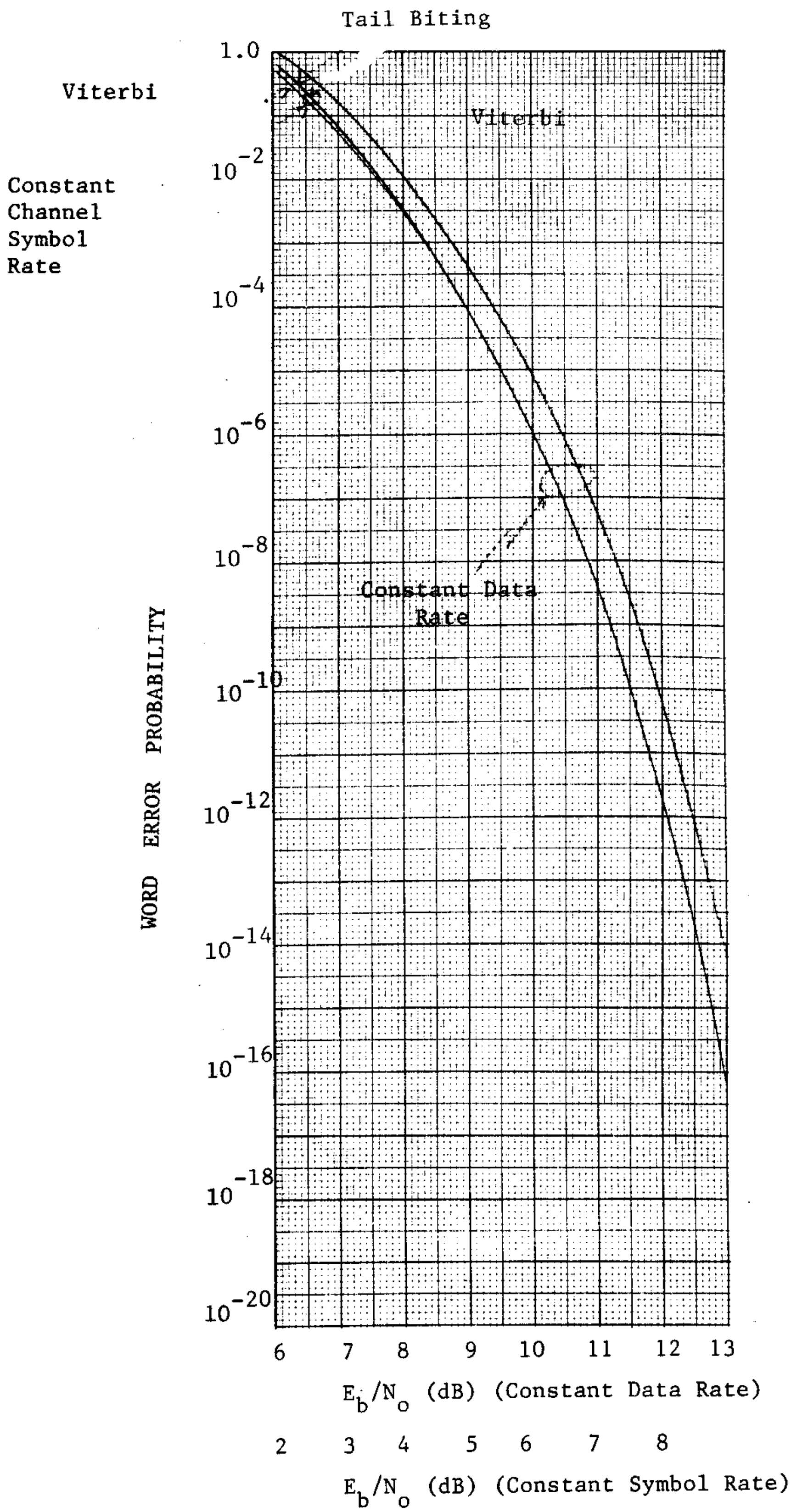


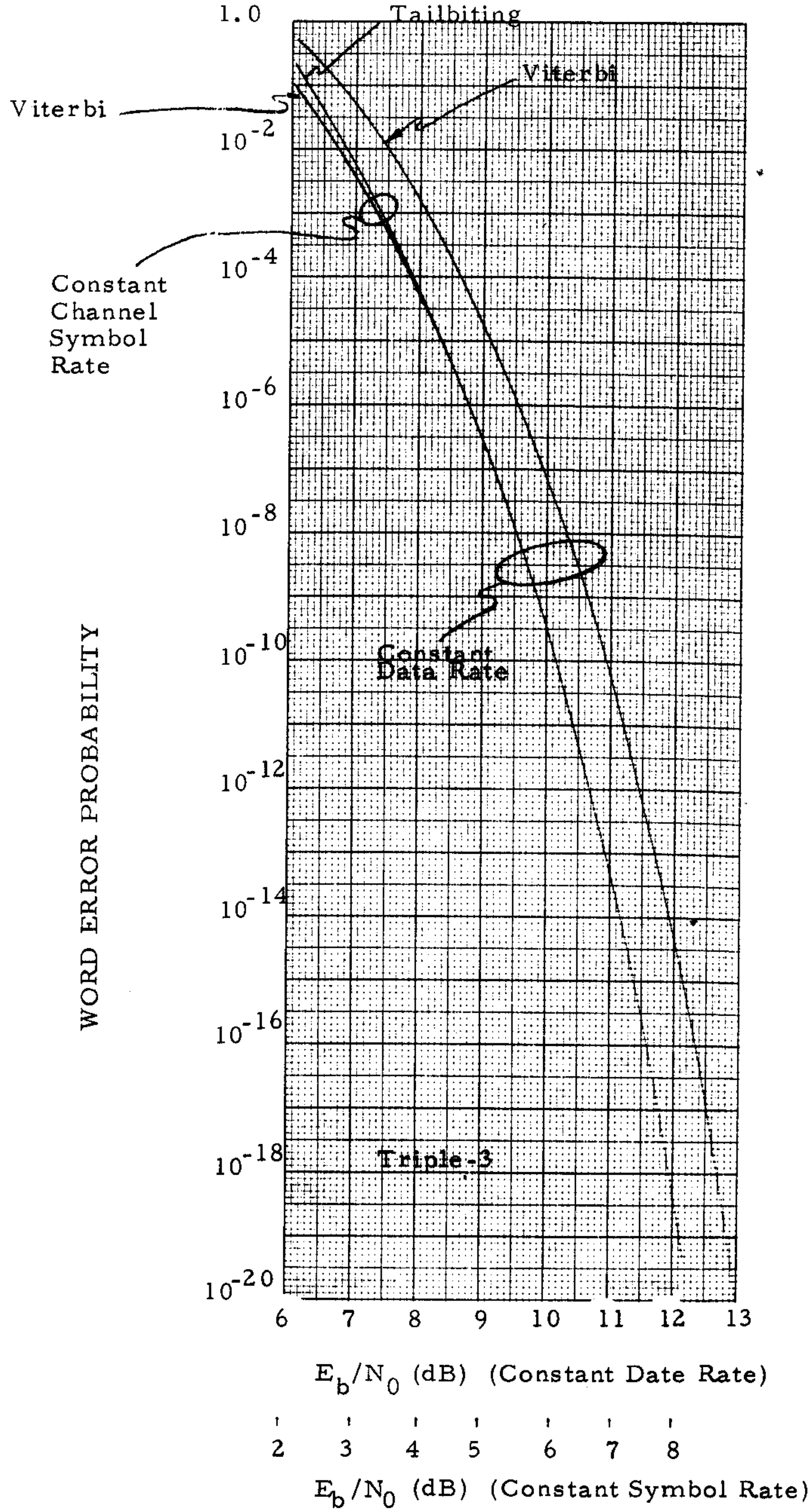
FIGURE 2. RATE ½ TRIPLE-3 ENCODER

FIGURE 3



Tail Biting
Viterbi
Constant
Channel
Symbol
Rate
1.0
Viterbi
10^-2
10^-4
10^-6
10^-8
Constant Data
Rate
10^-10
10^-12
10^-14
10^-16
10^-18
10^-20
WORD ERROR PROBABILITY
6 7 8 9 10 11 12 13
Eb/No (dB) (Constant Data Rate)
2 3 4 5 6 7 8
Eb/No (dB) (Constant Symbol Rate)

FIGURE 4



Tailbiting
Viterbi
Viterbi
Constant Channel Symbol Rate
Constant Data Rate
Triple-3
WORD ERROR PROBABILITY
1.0
10^{-2}
10^{-4}
10^{-6}
10^{-8}
10^{-10}
10^{-12}
10^{-14}
10^{-16}
10^{-18}
10^{-20}
6 7 8 9 10 11 12 13
E_b/N_0 (dB) (Constant Date Rate)
2 3 4 5 6 7 8
E_b/N_0 (dB) (Constant Symbol Rate)

References

1. Viterbi, A. J., "Convolutional Codes and Their Performance in Communication Systems," IEEE Trans. Com. Tech., Vol. COM-19, October, 1971.

2. Schwartz, M., Bennett, W. R., and Stein S., Communication Systems and Techniques, McGraw-Hill, Inc., 1966.

3. Odenwalker, J. P., "Dual-k Convolutional Codes for Noncoherently Demodulated Channels," Proceedings of the 1976 International Telemetering Conference, October, 1976.

4. Bucker, E., Technical Discussions, August 1980.