terleaved over $L = 3$ channel blocks for four different Rician factors $\gamma = 0, 5, 10,$ and infinity (= additive white Gaussian noise (AWGN) channel). The code polynomials in octal are $(133, 145, 175)$. Here the union bounds are truncated at $d_{\max} = 42$.[4] The traditional bounds (7) get tighter for increasing values of γ and finally coincide with the modified bound for $\gamma = \infty$ (AWGN channel) with which no averaging is needed. The modified bounds (11) for the block error probability are surprisingly tight for all values of γ and they also get tighter for higher values of γ. Furthermore, they are clearly tighter than the bit error probability upper bounds. The convolutional codes used in these examples are optimized for the AWGN channel ($d_{\rm free}$ maximized), and they are not necessarily optimum for the block fading channel, see, e.g., [15] and [9].

V. Conclusions

In this correspondence, the union bounding techniques for error control codes with limited interleaving over block fading Rician channels were considered. The traditional union bounding technique, which sums averaged pairwise error probabilities, was shown to yield very loose results especially for low Rician factors (a block fading Rayleigh channel). A modified union bounding technique was presented which limits the conditional union bound before averaging over the fading process and thus avoids the "explosion" of the union bound for low SNR. This modified bounding technique provides much tighter, and hence useful, numerical results, but requires L-fold numerical integration where L is the number of diversity subchannels. Examples were shown for terminated convolutional codes but the necessity for optimization before averaging in the block fading channels can be extended to other block codes as well. This was also clearly shown for random coding techniques in [5].

Acknowledgment

The authors wish to thank the anonymous reviewers for their comments that helped to improve this correspondence.

References

[1] J. G. Proakis, *Digital Communications*, 2nd ed. New York: McGraw-Hill, 1989.
[2] A. J. Viterbi, "Convolutional codes and their performance in communication systems," *IEEE Trans. Commun. Technol.*, vol. COM-19, pp. 751–772, Oct. 1971.
[3] G. Kaplan and S. Shamai (Shitz), "Error probabilities for the block-fading Gaussian channel," *Arch. Elek. Übertragung*, vol. 49, pp. 192–205, July 1995.
[4] L. Ozarow, S. Shamai, and A. D. Wyner, "Information theoretic considerations for cellular mobile radio," *IEEE Trans. Veh. Technol.*, vol. 43, pp. 359–378, May 1994.
[5] E. Malkamäki and H. Leib, "Coded diversity on block fading channels," *IEEE Trans. Inform. Theory*, vol. 45, pp. 771–781, Mar. 1999.
[6] E. Malkamäki, "Performance of error control over block fading channels with ARQ applications," Ph.D. dissertation (Tech. Rep. T43), Helsinki Univ. Technol., Commun. Lab., Espoo, Finland, 1998.
[7] G. Kaplan, S. Shamai (Shitz), and Y. Kofman, "On the design and selection of convolutional codes for an uninterleaved, bursty Rician channel," *IEEE Trans. Commun.*, vol. 43, pp. 2914–2921, Dec. 1995.
[8] Y. S. Leung, S. G. Wilson, and J. W. Ketchum, "Multi-frequency trellis coding with low delay for fading channels," *IEEE Trans. Commun.*, vol. 41, pp. 1450–1459, Oct. 1993.

[4] Enough terms have to be taken into account in order to have an upper bound. Notice that without truncation the traditional union bound does not necessarily converge.

[9] R. Knopp and P. A. Humblet, "Maximizing diversity on block fading channels," in *Proc. IEEE ICC'97* (Montreal, Que., Canada, 1997), pp. 647–651.
[10] J. Hagenauer, "Rate-compatible punctured convolutional codes (RCPC codes) and their applications," *IEEE Trans. Commun.*, vol. 36, pp. 389–400, Apr. 1988.
[11] A. J. Viterbi and J. K. Omura, *Principles of Digital Communication and Coding*. Tokyo, Japan: McGraw-Hill, 1979.
[12] D. N. Rowitch and L. B. Milstein, "Convolutional coding for direct sequence multicarrier CDMA," in *Proc. IEEE MILCOM'95* (San Diego, CA, 1995), pp. 55–59.
[13] M. K. Simon and D. Divsalar, "Some new twists to problems involving the Gaussian probability integral," *IEEE Trans. Commun.*, vol. 46, pp. 200–210, Feb. 1998.
[14] E. Biglieri, D. Divsalar, P. J. McLane, and M. K. Simon, *Introduction to Trellis-Coded Modulation with Applications*. New York: Macmillan, 1991.
[15] E. Malkamäki and H. Leib, "Rate $1/n$ convolutional codes with interleaving depth of n over a block fading Rician channel," in *Proc. IEEE VTC'97* (Phoenix, AZ, 1997), pp. 2002–2006.
[16] J. W. Modestino and S. Y. Mui, "Convolutional code performance in the Rician fading channel," *IEEE Trans. Commun.*, vol. COM-24, pp. 592–606, June 1976.
[17] S. Kallel and C. Leung, "Efficient ARQ schemes with multiple copy decoding," *IEEE Trans. Commun.*, vol. 40, pp. 642–650, Mar. 1992.
[18] NAG Ltd, *NAG Fortran Library Manual, Mark 17*, The Numerical Algorithms Group Limited, UK, 1995.
[19] M. Cedervall and R. Johannesson, "A fast algorithm for computing distance spectrum of convolutional codes," *IEEE Trans. Inform. Theory*, vol. 35, pp. 1146–1159, Nov. 1989.

On the Weight Distribution of Terminated Convolutional Codes

Marc P. C. Fossorier, *Member, IEEE*, Shu Lin, *Fellow, IEEE*, and Daniel J. Costello, Jr., *Fellow, IEEE*

***Abstract*—In this correspondence, the low-weight terms of the weight distribution of the block code obtained by terminating a convolutional code after x information blocks are expressed as a function of x. It is shown that this function is linear in x for codes with noncatastrophic encoders, but quadratic in x for codes with catastrophic encoders. These results are useful to explain the poor performance of convolutional codes with a catastrophic encoder at low-to-medium signal-to-noise ratios.**

***Index Terms*—Block codes, convolutional codes, soft-decision decoding, Viterbi decoding, weight distribution.**

I. Introduction

Any binary rate k/n convolutional code with constraint length L (or memory order $L-1$) whose associated trellis diagram terminates with a zero tail after encoding x blocks of k information bits into x blocks of n transmitted symbols generates an $(n(x+L-1), kx)$ binary block code. In this correspondence, for each such $(n(x+L-1), kx)$ block code, we express the low-weight terms of the corresponding weight distribution as a function of x. It is shown that this

Manuscript received November 1, 1996; revised February 15, 1999. This work was supported by the National Science Foundation under Grants NCR-94-15374 and NCR95-22939.
M. P. C. Fossorier and S. Lin are with the Department of Electrical Engineering, University of Hawaii, Honolulu, HI 96822 USA.
D. J. Costello, Jr. is with the Department of Electrical Engineering, University of Notre Dame, Notre Dame, IN 46556 USA.
Communicated by N. Seshadri, Associate Editor for Coding.
Publisher Item Identifier S 0018-9448(99)04382-5.

function is linear in x for codes with noncatastrophic encoders, but quadratic in x for codes with catastrophic encoders. This fact helps to explain why convolutional codes generated by a catastrophic encoder perform poorly. These results can be viewed as a simpler method than the one derived in [1] to obtain the low-weight terms of the weight distribution of a zero-tail terminated convolutional code. These results are also useful in deriving a good approximation of the union upper bound on the bit error probability for maximum-likelihood decoding (MLD) of convolutional codes with a noncatastrophic encoder by applying standard bounding techniques for block codes. Finally, since the weight distribution of a block code is independent of the mapping between information bits and codewords, these results are valid for terminated convolutional codes with encoder realizations in feedforward as well as in feedback forms.

II. Weight Distribution of the $(n(x+L-1),\, kx)$ Block Code

A. Noncatastrophic Encoders

Let $W_i(x)$ denote the number of codewords of weight i in the $(n(x+L-1),\, kx)$ block code obtained by terminating with a zero tail a rate k/n convolutional code of constraint length L after x information blocks. We define: 1) N_i as the number of paths of weight i in the trellis diagram that diverge from the all-0 path at the origin $x = 0$ and remerge only once (note that $N_i = 0$ for all $i < d_H$, the minimum free Hamming distance of the convolutional code) and 2) $l_{i,\max}$ as the maximum length (in information blocks) of the N_i paths of weight i previously defined. (We exclude catastrophic encoders here since $l_{i,\max}$ is unbounded for some i in this case.) Based on these definitions, we derive the following theorem.

Theorem 2.1: For $i \in \{d_H, \cdots, 2d_H - 1\}$ and $x \ge l_{i,\max}$, $W_i(x)$ is of the form

$$W_i(x) = N_i x - a_i \tag{1}$$

where $a_i \ge 0$.

Proof: For $x \ge l_{i,\max}$, N_i can equivalently be defined as the number of paths of weight i in the trellis diagram that remerge with the all-0 path at section $x+L-1$. Hence after adding one dimension (i.e., one information block) to the $(n(x-1+L-1),\, k(x-1))$ code we obtain the $(n(x+L-1),\, kx)$ code with, for $i \in \{d_H, \cdots, 2d_H - 1\}$

$$W_i(x) = W_i(x-1) + N_i. \tag{2}$$

Applying a chain argument to (2), we obtain,

$$W_i(x) = W_i(l_{i,\max} - 1) + (x - l_{i,\max} + 1)N_i. \tag{3}$$

However, for the first $l_{i,\max} - 1$ steps in the recursion, i.e., for $x < l_{i,\max}$, fewer than N_i paths are added since some of the paths corresponding to N_i have not terminated yet. This completes the proof after defining a_i as the total number of codewords discarded in the first $l_{i,\max} - 1$ steps. $\square$

Theorem 2.1 corresponds to the steady-state part of the weight distributions associated with the family of $(n(x+L-1),\, kx)$ block codes. By evaluating (1) and (2) for $x = l_{i,\max}$, we obtain

$$W_i(l_{i,\max}) = W_i(l_{i,\max} - 1) + N_i = N_i l_{i,\max} - a_i \tag{4}$$

so that

$$a_i = (l_{i,\max} - 1)N_i - W_i(l_{i,\max} - 1). \tag{5}$$

Based on (5), we observe that (1) is also satisfied for $x = l_{i,\max} - 1$. For $x < l_{i,\max} - 1$, N_i in (1) must be replaced by a smaller value which in general differs for each value of x.

For $i \ge 2d_H$, $W_i(x)$ is no longer a linear function of x, since not only paths of weight i contribute to $W_i(x)$ in this case. Theorem 2.2 gives an example of this fact.

Theorem 2.2: For $x \ge \max\{2l_{d_H,\max} + L - 1,\, l_{2d_H,\max}\}$

$$W_{2d_H}(x) = \frac{(N_{d_H})^2}{2} x^2 + b_{2d_H} x - a_{2d_H} \tag{6}$$

where $a_{2d_H} \ge 0$ and $b_{2d_H} \ge 0$.

Proof: For $x \ge \max\{2l_{d_H,\max} + L - 1,\, l_{2d_H,\max}\}$, (2) becomes

$$W_{2d_H}(x) = W_{2d_H}(x-1) + N_{2d_H} + P_{d_H}(x) \tag{7}$$

where $P_{d_H}(x)$ represents the number of pairs of disjoint paths, of weight d_H each, with the first path diverging from the all-0 path at $x = 0$. Then, for $x \ge \max\{2l_{d_H,\max} + L - 1,\, l_{2d_H,\max}\}$

$$P_{d_H}(x) = P_{d_H}(x-1) + (N_{d_H})^2 \tag{8}$$

since there are $(N_{d_H})^2$ such possible pairs. By following the same approach as in the proof of Theorem 2.1, we obtain, for $a'_{d_H} \ge 0$

$$P_{d_H}(x) = (N_{d_H})^2 x - a'_{d_H}. \tag{9}$$

Regrouping (7) and (9), it follows that

$$W_{2d_H}(x) = W_{2d_H}(x-1) + N_{2d_H} + (N_{d_H})^2 x - a'_{d_H}. \tag{10}$$

A chain argument completes the proof after defining b_{2d_H} as

$$b_{2d_H} = (N_{d_H})^2/2 + N_{2d_H} - a'_{d_H} \tag{11}$$

and a_{2d_H} as the total number of paths discarded in the initial steps. $\square$

Also, by algebraic manipulations of (7)–(10), we can show that (6) remains valid for $x = \max\{2l_{d_H,\max} + L - 1,\, l_{2d_H,\max}\} - 1$. For $i > 2d_H$, $W_i(x)$ can be derived by following a similar approach.

For example, the rate $1/3$ convolutional code with $L = 3$ and generators in octal form $(5, 7, 7)$ (see [2, Table 11.1]) generates the family of $(3(x+2),\, x)$ block codes with $d_H = 8$ and weight distribution starting with

$$\begin{aligned} W_8(x) &= 2x - 1 \\ W_{10}(x) &= 5x - 14 \\ W_{12}(x) &= 13x - 65 \\ W_{14}(x) &= 34x - 244 \\ W_{16}(x) &= 2x^2 + 75x - 807 \end{aligned}$$

for $x \ge 13$, since $l_{16,\max} = 14$.

In a recent paper [1], it was shown that the entire weight distribution of the $(2(x+L-1),\, x)$ block code obtained by truncating a rate $1/2$ convolutional code of constraint length L after x information blocks can be obtained by multiplying by itself $(x+L-1)$ times a $2^{L-1} \times 2^{L-1}$ square matrix. The elements of this matrix correspond to the Hamming weights of the transitions in the state diagram of the rate $1/2$ convolutional code. However, in many cases, only the first terms of the weight distribution of a code are required. Consequently, the results of this correspondence can be used in conjunction with the results of [1] to efficiently determine the terms of interest in the weight distribution of a terminated convolutional code. For example, given a noncatastrophic encoder, we know from Theorem 2.1 that for $i \in \{d_H, \cdots, 2d_H - 1\}$ and $x \ge l_{i,\max}$, $W_i(x)$ is of the form

$$W_i(x) = Ax + B. \tag{12}$$

Then the results of [1] can be used to evaluate the coefficients A and B for two small values of $x \ge l_{i,\max}$. More precisely, let us consider the example in [1] with $L = 5$, $d_H = 7$, and $l_{7,\max} = 4$. Based on [1, Table 1], we have for $x + 4 = 12$ and $x + 4 = 15$

$$\begin{aligned} W_7(8) &= 8A + B = 13 \\ W_7(11) &= 11A + B = 19 \end{aligned} \tag{13}$$

so that $A = 2$ and $B = -3$. (Note that since $l_{7,\max} = 4$, even smaller values of x can be chosen to solve for A and B.) Then we can verify in [1, Table 1] that for $x + 4 = 18$, $W_7(14) = 2 \cdot 14 - 3 = 25$. However, for $x = 996$, we can directly obtain $W_7(996) = 1989$ instead of raising a 16×16 matrix to the 1000th power. The same method can be used to determine any $W_i(x)$ with $i \in \{d_H, \cdots, 2d_H - 1\}$. Also, based on Theorem 2.2, three equations with

$$x \geq \max\{2l_{d_H,\max} + L - 1, l_{2d_H,\max}\}$$

are sufficient to determine $W_{2d_H}(x)$ in the form $Ax^2 + Bx + C$. Indeed, the results of Theorems 2.1 and 2.2 can be generalized to express any $W_i(x)$ in polynomial form. To this end, the results of [1] become useful to evaluate the coefficients of the corresponding polynomials and to verify that large enough values of x are chosen.

B. Catastrophic Encoders

In this section, we show that $W_{d_H}(x)$ is proportional to x^2 if the encoder is catastrophic. We consider a rate $1/2$ convolutional code with a catastrophic encoder and assume that, in addition to the weight-zero self-loop around the all-zero state, the state diagram contains a weight-zero self-loop around the all-one state. Although derived for this particular case, the result that $W_{d_H}(x)$ is proportional to x^2 holds for any convolutional code with a catastrophic encoder since for such codes, in addition to the all-zero state, at least one other state touches a weight-zero loop as in the case considered in the following proof. (If the length of the weight-zero loop is $l_0 > 1$, a change of variable $y = x/l_0$ is needed.) Also, extension to multiple weight-zero loops follows the same lines. However, the corresponding general expressions become significantly more complex.

We define $T_i(x)$ as the number of distinct paths of length x (in information blocks) such that the xth branch is the weight-zero self-loop around the all-one state and the codeword has weight i in the associated $(n(x + L - 1), kx)$ block code. (We assume here that the free distance d_H is the weight of the lowest weight path that diverges from and remerges with the all-zero path.) Based on this definition, we obtain, for $x \geq l_{d_H,\max}$

$$W_{d_H}(x) = W_{d_H}(x - 1) + T_{d_H}(x) + N_{d_H} \tag{14}$$

where N_{d_H} is the number of paths of weight d_H diverging from the all-zero path at $x = 0$ that do not contain the weight-zero self-loop around the all-one state, and $l_{d_H,\max}$ is defined with respect to these paths only. Also

$$T_{d_H}(x) = T_{d_H}(x - 1) + N'_{d_H} \tag{15}$$

where $N'_{d_H} \leq N_{d_H}$ is the number of paths of weight d_H that diverge from the all-zero path at $x = 0$, that do not contain the weight-zero self-loop around the all-one state, and that pass through the all-one state, and where we assume $N'_{d_H} > 0$. As in the proof of Theorem 2.1, we obtain

$$T_{d_H}(x) = N'_{d_H} x - \alpha \tag{16}$$

where $\alpha \geq 0$. It follows that

$$W_{d_H}(x) = W_{d_H}(x - 1) + N'_{d_H} x - \alpha + N_{d_H}. \tag{17}$$

By applying the chain rule, we finally obtain

$$W_{d_H}(x) = N'_{d_H} x(x + 1)/2 + (N_{d_H} - \alpha)x - \beta \tag{18}$$

where $\beta \geq 0$. For example, the trivial rate $1/2$ convolutional encoder with one memory element whose input and output are summed to provide the two same output bits is catastrophic [3]. This code defines a family of $(2(x + 1), x)$ block codes with $d_H = 4$ and $W_4(x) = x(x + 1)/2$. As in Section II-A, the results of [1] can be used to evaluate the coefficients values of $W_{d_H}(x)$ for catastrophic encoders.

III. Approximation of the Union Bound on the Bit Error Probability of Convolutional Codes

For binary phase-shift keying (BPSK) transmission over an additive white Gaussian noise (AWGN) channel, the block error probability associated with MLD of an $(n(x + L - 1), kx)$ block code obtained by terminating a rate-k/n convolutional code with row distance d_H can be bounded using the union bound by

$$P_s \leq \sum_{i=d_H}^{N} W_i(x)\tilde{Q}\left(\sqrt{i}\right) \tag{19}$$

with

$$\tilde{Q}(x) = (\pi N_0)^{-1/2} \int_x^\infty e^{-n^2/N_0}\, dn.$$

Also, if encoding is done in reduced echelon form, a good approximation of the union bound on the bit error probability P_b is obtained by scaling each term in the sum of (19) by $i/(n(x + L - 1))$ [4]. Consequently, a good approximation of the union bound on P_b for medium to high signal-to-noise ratio (SNR) values can be derived based on Theorem 2.1 and (19) for zero tail terminated convolutional codes with a noncatastrophic encoder by considering only the values $i \in \{d_H, \cdots, 2d_H - 1\}$ [5]. For x large enough, the approximate bound on P_b becomes independent of x.

On the other hand, the approximate bound on P_b increases linearly with x for terminated convolutional codes with a catastrophic encoder. This suggests that for these codes, the best error performance is achieved by terminating the trellis after encoding x_{N_0} information blocks, where x_{N_0} depends on the operating SNR. Although catastrophic encoders should be avoided when convolutional codes are decoded with the Viterbi algorithm, they can be used if the trellis terminates, since information sequences of finite length only are considered. In particular, for a given rate and constraint length, a catastrophic encoder may generate a convolutional code with a larger row distance d_{row} than the free distance d_{free} of the best code generated by a noncatastrophic encoder[1] [6]. However, based on the previous results, this gain in d_{row} for the corresponding terminated codes with minimum Hamming distance d_{row} is achieved at the expense of significantly larger low-weight coefficients in the weight distribution. As a result, despite a smaller row distance, convolutional codes with noncatastrophic encoders should give lower values of P_b than the equivalent codes with catastrophic encoders at low-to-medium SNR values.

References

[1] J. K. Wolf and A. J. Viterbi, "On the weight distribution of linear block codes formed from convolutional codes," *IEEE Trans. Commun.*, vol. 44, pp. 1049–1051, Sept. 1996.

[2] S. Lin and D. J. Costello Jr., *Error Control Coding: Fundamentals and Applications.* Englewood Cliffs, NJ: Prentice-Hall, 1983.

[3] J. L. Massey and M. K. Sain, "Inverses of linear sequential circuits," *IEEE Trans. Comput.*, vol. C-17, pp. 330–337, Apr. 1968.

[4] M. P. C. Fossorier, S. Lin, and D. Rhee, "Bit error probability for maximum likelihood decoding of linear block codes," *IEEE Trans. Inform. Theory*, to be published.

[5] M. P. C. Fossorier, S. Lin, and D. J. Costello, "Weight distribution and error performance of terminated convolutional codes," in *Proc. Conf. Information Sciences and Systems* (Baltimore, MA, Mar. 1997), pp. 767–768.

[6] D. J. Costello Jr., "Free distance bounds for convolutional codes," *IEEE Trans. Inform. Theory*, vol. IT-20, pp. 356–365, May 1974.

[1] In general $d_{\text{free}} \leq d_{\text{row}}$, and $d_{\text{free}} = d_{\text{row}}$ if the encoder is noncatastrophic.