Optimal and near-optimal encoders for short and moderate-length tailbiting trellises

Ståhl, Per; Anderson, John B; Johannesson, Rolf

Published in:
IEEE Transactions on Information Theory

DOI:
10.1109/18.796408

1999

Link to publication

Citation for published version (APA):

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 13. Jan. 2019
Optimal and Near-Optimal Encoders for Short and Moderate-Length Tail-Biting Trellises

Per Ståhl, Student Member, IEEE, John B. Anderson, Fellow, IEEE, and Rolf Johannesson, Fellow, IEEE

Abstract—The results of an extensive search for short and moderate-length polynomial convolutional encoders for time-invariant tail-biting representations of block codes at rates \( R = 1/4, 1/3, 1/2, \) and 2/3 are reported. The tail-biting representations found are typically as good as the best known block codes.

Index Terms—Convolutional codes, error-correction coding, tail-biting encoders, trellis codes.

I. INTRODUCTION

There exist two main principles to terminate convolutional codes into block codes. Assume for simplicity that the generator matrix for

...
is the $L \times L$ generator matrix for the tail-biting representation of the block code $B_{tb}$. The initial state is uniquely determined by the last $m$ input $t$-tuples.

Since we require that we have the same state at the beginning as at the end we can use a circular trellis for the tail-biting representation of a block code. In Fig. 1 we show as an example a circular trellis of $L = 6$ sections for a four-state tail-biting representation.

The block code is the union of $2^m$ subsets corresponding to the paths that go through each of the $2^m$ states at time 0. In fact, these are $2^m$ cosets of the $(N, K - bm)$ zero-tail terminated convolutional code corresponding to the paths that go through the all-zero state at time 0.

Tail-biting is particularly valuable in applications such as packet transmission, where codewords are apt to be short. Being trellis methods, tail-biting decoders easily accept soft-channel output. Finally, as we show here, time-invariant tail-biting encoders can generate many of the most powerful binary block codes.

Tail-biting (TB) representations of block codes were introduced by Solomon and van Tilborg [1]. An early paper that explains tail-biting in detail is [2]. In [3], a two-way decoding algorithm for exact a priori probability (APP) decoding of TB trellises is presented. The complexity of this algorithm is of the order $2^{2m}$, and performance is governed by the true minimum distance $d$ of the code.

More commonly, an approximate APP decoding algorithm such as that given in [4] is used (the one in [4] applies the BCJR recursions to the TB trellis). Such algorithms dramatically reduce the complexity to the order of $2^m$; moreover, they achieve near-
optimal performance provided that there are no "pseudocodewords" (multicycle paths through the trellis that are not codewords) whose "effective weight" is less than $\frac{1}{100}$.

We have not investigated whether low effective weight pseudocodewords exist for the TB trellises presented in this correspondence.

### II. Searching for Good Encoders

Since the encoder is used for tail-biting trellis representations of block codes, the codeword set is limited to those register output sequences which end in the same register states at which they began. The full set of codewords comprises a linear block code and $d$, the weight of the least weight codeword in the set, is the minimum distance of the code. We search for the polynomial encoders of memory $m$ that produce the largest effective weight for each pair $(r, K)$.

In keeping with the usual notation, the code is called an $(N, K, d)$ code.

For rates of the form $1/c$, $c$ an integer, $c$ sets of $(m + 1)$ taps, called the *generators* $g_1, g_2, \ldots, g_c$, perform the encoding, as shown at the top of Fig. 2 for the case $c = 2$ (rate $R = 1/2$); $m$ is the memory of the encoder and it is assumed that at least one generator has $g_{ij}^{(m)} = 1$. Generator $g_j$ describes the effect of the input on the $j$th bit of each group of $c$ outputs. Similarly, the rate $R = 2/3$ encoder at the bottom in the figure is defined by the six generators $g_1, g_2, g_3, g_4, g_5, g_6$ where $g_{ij}$ describes the effect of input stream $i$ on output stream $j$.

An effective strategy for finding good generators for TB trellis representations is to search exhaustively for the best noncatastrophic generators. (For TB lengths equal to a multiple of the length of the nontrivial all-zero cycle in the state transition graph of a catastrophic generator we have two all-zero trellis paths and, hence, the catastrophic generator cannot generate a TB trellis.) An early such study is [5]; we search to much higher $m$ than they do and correct some apparent errors. For TB trellises out to $K = 50$ information bits, we have found the optimal generators by exhaustive search out to memory $m$ at rate $R = 1/4$, $m = 6$ at rate $R = 1/3$, $m = 8$ at rate $R = 1/2$, and overall constraint length $\nu = 7$ at rate $R = 2/3$.

Although these codes suffice for many applications, good codes with longer memory are sometimes needed, and an exhaustive search...
is now too time consuming. There are several other ways to gain an idea of the capabilities of long-memory TB encoders. One is to search for good TB generators among those similar to the best known generators for convolutional codes. Another is to choose longer generators at random and keep the best found after, say, several days of computer search. We found this second method quite effective. It probably works because most long random codes are good.

For a given set of generators and TB length, it remains to find the minimum Hamming distance for the code. Since the code is linear, it is enough to find the minimum weight of any codeword other than the all-zero one. The nonzero block code trellis paths fall into two cases, which are illustrated in Fig. 3.

Case (i): (Distance within one code subset, $d_{\text{intra}}$.) The neighbor path touches the all-zero path at least once; all paths considered are within the subset of paths leaving state 0. Finding the minimum weight among such paths is the same as finding the minimum weight path in a convolutional code. By the symmetry of the circular trellis, the behavior of paths out of one trellis stage is the same as out of all the others, so we can place the touch point at the left of the trellis. We start the usual dynamic program from this point and solve for the minimum-weight descendant in the trellis that splits from this point and merges again later to the all-zero path. Note that this can happen before $L$ stages, but must happen at least by the $L$th stage, by the tail-biting condition. We call this the intra minimum distance of the TB trellis representation and denote it $d_{\text{intra}}$.

Case (ii): (Distance between code subsets, $d_{\text{inter}}$.) The neighbor path never touches the all-zero path. This case is unique to tail-biting. The worst case is found by executing a dynamic program starting from some state not equal to state 0. The usual dynamic program in case (i) may be used, but with some modified search conditions. The program must delete any paths merging to the all-zero path; second,
The minimum distance for the tail-biting representation is $d_{\text{TB}}$ if the tail-biting trellis is long enough, case (i) paths lead to the minimum distance but for short trellises case (ii) codewords may lead to the minimum weight path, i.e., $d_{\text{TB}}$ might be less than $d_{\text{terminated}}$. Hence, short TB trellis representations of block codes could have quite different optimal generators than terminated convolutional codes.

The minimum distance discussed above is the principal determinant of the merit of the resulting code when maximum-likelihood decoding or APP decoding is used. However, for certain suboptimal decoding algorithms it has recently been observed by Frey, Kötter, and Vardy [6] and Forney, Kschischang, and Marcus [7] that the decoding performance may be affected by so-called “pseudocodewords.” A pseudocodeword is a trellis path of one or more cycles that starts and ends in the same state at some integer multiple of the cycle length. The pseudocodewords are precisely the length-$L$ pseudocodewords. See also [8], [9]. We have not checked whether our tail-biting representations have pseudocodewords with low effective weight.

### III. Results and Observations

Tables I–IV list the best encoders found by exhaustive search for rates $R = 2/3$, $1/2$, $1/3$, $1/4$, $1/5$, $1/6$, and $1/7$. For each number of information bits $K$ up to 50, the best time-invariant generators are listed. For each encoder memory $m$ or overall constraint length $n$. Among the generators leading to the same distance $d$, the set shown leads to the fewest codeword neighbors at $d$, denoted $n_d$. The generator notation is octal, with each octal symbol representing groups of three generator bits beginning at the left; e.g., $g_{\text{TB}}(0, 4, 6)$.

<table>
<thead>
<tr>
<th>$K / 2$</th>
<th>$G$</th>
<th>$d$</th>
<th>$n_d$</th>
<th>$G$</th>
<th>$d$</th>
<th>$n_d$</th>
<th>$G$</th>
<th>$d$</th>
<th>$n_d$</th>
<th>$G$</th>
<th>$d$</th>
<th>$n_d$</th>
<th>$G$</th>
<th>$d$</th>
<th>$n_d$</th>
<th>$G$</th>
<th>$d$</th>
<th>$n_d$</th>
<th>$G$</th>
<th>$d$</th>
<th>$n_d$</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>(0.4, 4)</td>
<td>2</td>
<td>(0.4, 6)</td>
<td>2</td>
<td>(2.4, 6)</td>
<td>2</td>
<td>(0.6, 7)</td>
<td>2</td>
<td>(10.4, 70)</td>
<td>2</td>
<td>(0.5, 4, 7)</td>
<td>2</td>
<td>(30.4, 70)</td>
<td>2</td>
<td>(0.5, 4, 7)</td>
<td>2</td>
<td>(30.4, 70)</td>
<td>2</td>
<td>(0.5, 4, 7)</td>
<td>2</td>
<td>(30.4, 70)</td>
</tr>
<tr>
<td>6</td>
<td>(0.4, 4)</td>
<td>2</td>
<td>(0.4, 6)</td>
<td>2</td>
<td>(2.4, 6)</td>
<td>2</td>
<td>(1.5, 3)</td>
<td>2</td>
<td>(10.5, 30)</td>
<td>2</td>
<td>(0.5, 4, 7)</td>
<td>2</td>
<td>(30.5, 30)</td>
<td>2</td>
<td>(0.5, 4, 7)</td>
<td>2</td>
<td>(30.5, 30)</td>
<td>2</td>
<td>(0.5, 4, 7)</td>
<td>2</td>
<td>(30.5, 30)</td>
</tr>
<tr>
<td>8</td>
<td>(0, 4, 6)</td>
<td>3</td>
<td>(2.4, 6)</td>
<td>3</td>
<td>(1.5, 3)</td>
<td>3</td>
<td>(6.3, 7)</td>
<td>3</td>
<td>(30.5, 30)</td>
<td>3</td>
<td>75/59/52/75</td>
<td>3</td>
<td>48/51/60/74</td>
<td>3</td>
<td>(30.5, 30)</td>
<td>3</td>
<td>75/59/52/75</td>
<td>3</td>
<td>48/51/60/74</td>
<td>3</td>
<td>(30.5, 30)</td>
</tr>
<tr>
<td>10</td>
<td>(0.4, 6)</td>
<td>3</td>
<td>(2.4, 6)</td>
<td>4</td>
<td>(1.5, 3)</td>
<td>4</td>
<td>(30.5, 30)</td>
<td>4</td>
<td>(0.5, 4, 7)</td>
<td>4</td>
<td>(30.5, 30)</td>
<td>4</td>
<td>(30.5, 30)</td>
<td>4</td>
<td>(30.5, 30)</td>
<td>4</td>
<td>(30.5, 30)</td>
<td>4</td>
<td>(30.5, 30)</td>
<td>4</td>
<td>(30.5, 30)</td>
</tr>
<tr>
<td>12</td>
<td>(0.4, 6)</td>
<td>3</td>
<td>(2.4, 6)</td>
<td>4</td>
<td>(1.5, 3)</td>
<td>4</td>
<td>(30.5, 30)</td>
<td>4</td>
<td>(0.5, 4, 7)</td>
<td>4</td>
<td>(30.5, 30)</td>
<td>4</td>
<td>(30.5, 30)</td>
<td>4</td>
<td>(30.5, 30)</td>
<td>4</td>
<td>(30.5, 30)</td>
<td>4</td>
<td>(30.5, 30)</td>
<td>4</td>
<td>(30.5, 30)</td>
</tr>
<tr>
<td>14</td>
<td>(0.4, 6)</td>
<td>3</td>
<td>(2.4, 6)</td>
<td>4</td>
<td>(1.5, 3)</td>
<td>4</td>
<td>(30.5, 30)</td>
<td>4</td>
<td>(0.5, 4, 7)</td>
<td>4</td>
<td>(30.5, 30)</td>
<td>4</td>
<td>(30.5, 30)</td>
<td>4</td>
<td>(30.5, 30)</td>
<td>4</td>
<td>(30.5, 30)</td>
<td>4</td>
<td>(30.5, 30)</td>
<td>4</td>
<td>(30.5, 30)</td>
</tr>
<tr>
<td>16</td>
<td>(0.4, 6)</td>
<td>3</td>
<td>(2.4, 6)</td>
<td>4</td>
<td>(1.5, 3)</td>
<td>4</td>
<td>(30.5, 30)</td>
<td>4</td>
<td>(0.5, 4, 7)</td>
<td>4</td>
<td>(30.5, 30)</td>
<td>4</td>
<td>(30.5, 30)</td>
<td>4</td>
<td>(30.5, 30)</td>
<td>4</td>
<td>(30.5, 30)</td>
<td>4</td>
<td>(30.5, 30)</td>
<td>4</td>
<td>(30.5, 30)</td>
</tr>
</tbody>
</table>

at the $L$th stage only paths entering state $a'$ count. The procedure must be repeated for each nonzero starting state, and the minimum taken over all the outcomes. We call this the inter minimum distance of the TB trellis representation and denote it $d_{\text{inter}}$. The minimum distance for the tail-biting representation is $d = \min\{d_{\text{int}}, d_{\text{inter}}\}$. If the tail-biting trellis is long enough, case (i) paths lead to the minimum distance but for short trellises case (ii) codewords may lead to the minimum weight path, i.e., $d_{\text{TB}}$ might be less than $d_{\text{inter}}$. Hence, short TB trellis representations of block codes could have quite different optimal generators than terminated convolutional codes.

The minimum distance discussed above is the principal determinant of the merit of the resulting code when maximum-likelihood decoding or APP decoding is used. However, for certain suboptimal decoding algorithms it has recently been observed by Frey, Kötter, and Vardy [6] and Forney, Kschischang, and Marcus [7] that the decoding performance may be affected by so-called “pseudocodewords.” A pseudocodeword is a trellis path of one or more cycles that starts and ends in the same state at some integer multiple of the cycle length. The pseudocodewords are precisely the length-$L$ pseudocodewords. See also [8], [9]. We have not checked whether our tail-biting representations have pseudocodewords with low effective weight.

### III. Results and Observations

Tables I–IV list the best encoders found by exhaustive search for rates $R = 1/4$, $1/3$, $1/2$, and $2/3$, respectively. For each number of information bits $K$ up to 50, the best time-invariant generators are listed at each encoder memory $m$ or overall constraint length $n$. Among the generators leading to the same distance $d$, the set shown leads to the fewest codeword neighbors at $d$, denoted $n_d$. The generator notation is octal, with each octal symbol representing groups of three generator bits beginning at the left; e.g., $g_{\text{TB}}(0, 4, 6)$.
improves further with tail-biting length; in fact, the best generators are now those of the memory $m$ convolutional code with best free distance and best $n_{s,free}$. When the columns of a table reach this point, no further entries are shown. The neighbor number continues to grow with $K$, according to the linear rate indicated. An explanation of the $K$ at which distance saturation occurs is given in the next section.
Fig. 6. Best minimum distance found for $R = 1/4$.

Fig. 7. Best minimum distance found for $R = 2/3$. 
TABLE V

<table>
<thead>
<tr>
<th>K</th>
<th>systematic</th>
<th>m</th>
<th>d</th>
<th>n_d</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>(4,4,6)</td>
<td>3</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>(4,6,7)</td>
<td>4</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>(4,5,6,7)</td>
<td>5</td>
<td>7</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>(5,4,6,2,7)</td>
<td>6</td>
<td>8</td>
<td>5</td>
</tr>
<tr>
<td>6</td>
<td>(4,5,6,2,7)</td>
<td>7</td>
<td>9</td>
<td>6</td>
</tr>
<tr>
<td>7</td>
<td>(4,6,5,6,7)</td>
<td>8</td>
<td>10</td>
<td>7</td>
</tr>
<tr>
<td>8</td>
<td>(4,5,6,1,7)</td>
<td>9</td>
<td>11</td>
<td>8</td>
</tr>
<tr>
<td>9</td>
<td>(4,6,1,5,7)</td>
<td>10</td>
<td>12</td>
<td>9</td>
</tr>
</tbody>
</table>

TABLE VI

<table>
<thead>
<tr>
<th>K</th>
<th>systematic</th>
<th>m</th>
<th>d</th>
<th>n_d</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>(4,4,6)</td>
<td>3</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>(4,6,7)</td>
<td>4</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>(4,5,6,7)</td>
<td>5</td>
<td>7</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>(5,4,6,2,7)</td>
<td>6</td>
<td>8</td>
<td>5</td>
</tr>
<tr>
<td>6</td>
<td>(4,5,6,2,7)</td>
<td>7</td>
<td>9</td>
<td>6</td>
</tr>
<tr>
<td>7</td>
<td>(4,6,5,6,7)</td>
<td>8</td>
<td>10</td>
<td>7</td>
</tr>
<tr>
<td>8</td>
<td>(4,5,6,1,7)</td>
<td>9</td>
<td>11</td>
<td>8</td>
</tr>
<tr>
<td>9</td>
<td>(4,6,1,5,7)</td>
<td>10</td>
<td>12</td>
<td>9</td>
</tr>
</tbody>
</table>

For a fixed \( K \), there is also an encoder memory \( m \) or overall constraint length \( \nu \), going across a table row, at which no further distance growth occurs, although sometimes a larger \( m \) or \( \nu \) allow an encoder with fewer nearest neighbors. There is thus a saturation as \( m \) or \( \nu \) grow for a given \( K \), as well as \( \nu \) versa. As a rule, a code with a given \( d \) should be as short as possible, in order to obtain a good decoded data error rate.

Tables V–VIII show the best time-invariant tail-biting generators found at each \( K \), regardless of encoder memory or overall constraint length. These tables also list the best feedforward systematic encoders found by exhaustive searching.

Figs. 4–7 compare \( d \) and \( K \) for the best TB encoders to the best bounds given in Brouwer and Verhoef [10], for all four rates. Note that the TB encoders in the right half of the figure were found by random search and are not optimal; wherever the TB encoders are known to be optimal, they essentially match the B-V table. It is likely that better TB encoders exist where \( L \) is large.

Several especially interesting codes were found during the searches. A number of rate \( R = 1/3 \) (15, 5, 7) TB trellis representations with memory 3 were found which are equivalent to the (15, 5, 7) Bose–Chaudhuri–Hocquenghem (BCH) code (i.e., by permuting codeword bit positions, the TB codeword set can be made identical to the BCH code). This eight-state TB trellis is the trellis description of this BCH code with the fewest known states (cf. [11]). The square-root lower bound combined with Schurman’s 16-state conventional trellis for the (15, 5, 7) BCH code [12] suggests that there might be a four-state time-varying TB trellis for this code.

Among the optimal rate \( R = 1/2 \) codes, a (24, 12, 8) code with memory 6 was found. Since the extended Golay code has the same \( (N, K, d) \) and it can be shown that there is only one such code, we have found a 64-state TB trellis representation for the extended Golay code, i.e., a TB trellis representation with the same number of states as the trellis given in [13]; such a trellis was previously reported in [1] and [5]. (These three are nonequivalent as convolutional encoders.) Recently, Calderbank, Forney, and Vardy [14] reported a 16-state TB trellis for this code, but the trellis is time-varying with period 4, whereas our 64-state TB trellis is time-invariant. The existence of the 16-state, rate \( R = 1/2 \) TB trellis for the (24, 12, 8) Golay code shows that nonnegligible savings in complexity might be obtained by considering time-varying generators.

Some of our TB trellis representations beat those given in [5]. For example, our 64-state, rate \( R = 1/4 \) (48, 12, 17) trellis beats the (48, 12, 16) code given there; this code ties the shortening of the (49, 13, 17) code in [10]. Since a 16-state TB trellis for a (48, 12, 16) code can be obtained just by doubling the bits in the Golay code, it seems probable that there is a simpler time-varying TB trellis for the (48, 12, 17) code as well.

Three of our interesting encoders are given in Table IX.

IV. Distance Behavior as the Tail-Biting Trellis Grows

As noted in Section II, optimal generators for short TB trellises are very different from those of convolutional codes. Consider a fixed memory \( m \). Long TB trellis representations have both the same generators and the same distance as the best memory \( m \) convolutional codes. For TB lengths \( L \) in between (about 4\( m \) in the rate \( R = 1/2 \) case), the generators become the same as or at least roughly similar to...
the best free-distance generators of memory $m$ convolutional codes, but the achieved distance falls short.

Thinking conversely, we can fix the TB sections at $L$ and ask what is the best distance obtainable by any code found at any memory. These distances can be seen in Fig. 4. They are obtained from generators whose \( m \) is considerably shorter than the tail-biting trellis. The relationship of \( m \) to \( \nu \) and \( \mu \) seems to be similar to that of the decision depth parameter for convolutional codes to \( \mu \) and \( \nu \) respectively. We now discuss this relationship.

Roughly speaking, the decision depth parameter is the depth needed in convolutional decoding so that distance properties are not dominated by unmerged trellis paths. A precise definition of this depth can be given with reference to the case (i) paths in Fig. 3. Consider all trellis paths stemming from a root node as in the figure; without loss of generality, the all-zero path is one of these and was transmitted. Then the decision depth function \( d \) at distance \( \mu \) is the first trellis depth at which every nonzero path, merged with the all-zero path or not, is \( \mu \) away from the all-zero path. In ordinary trellis decoding, the decision on the first branch cannot be made until at least \( \nu \) stages have been explored, if the code is to act as if it had minimum distance \( \nu \). The decoder observation window is thus \( \nu \) trellis stages. In a typical code, the paths merging to the all-zero without loss of generality, the all-zero path is one of these and was transmitted. Then the decision depth function \( L_D(d) \) at distance \( d \) is the first trellis depth at which every nonzero path, merged with the all-zero path or not, is \( d \) away from the all-zero path. In ordinary trellis decoding, the decision on the first branch cannot be made until at least \( L_D(d) \) stages have been explored, if the code is to act as if it had minimum distance \( d \). The decoder observation window is thus \( L_D \) trellis stages. In a typical code, the paths merging to the all-zero without loss of generality, the all-zero path is one of these and was transmitted. Then the decision depth function \( L_D(d) \) at distance \( d \) is the first trellis depth at which every nonzero path, merged with the all-zero path or not, is \( d \) away from the all-zero path. In ordinary trellis decoding, the decision on the first branch cannot be made until at least \( L_D(d) \) stages have been explored, if the code is to act as if it had minimum distance \( d \). The decoder observation window is thus \( L_D \) trellis stages. In a typical code, the paths merging to the all-zero without loss of generality, the all-zero path is one of these and was transmitted. Then the decision depth function \( L_D(d) \) at distance \( d \) is the first trellis depth at which every nonzero path, merged with the all-zero path or not, is \( d \) away from the all-zero path. In ordinary trellis decoding, the decision on the first branch cannot be made until at least \( L_D(d) \) stages have been explored, if the code is to act as if it had minimum distance \( d \). The decoder observation window is thus \( L_D \) trellis stages. In a typical code, the paths merging to the all-zero
path that leads to the free distance occurs well before \( L_D(d_{\text{free}}) \), since the many close unmerged paths are what dominate \( L_D \). The \( L_D \) function is tabulated in [15] for most convolutional codes. These tabulations show that \( L_D(d) \) rapidly assumes the asymptotic rule

\[
L_D(d) \approx d / ch^{-1}(1 - R)
\]

(6)
as \( d \) grows, where \( h^{-1}(\cdot) \) is the inverse binary entropy function. That the right-hand side is an overbound can be shown for several theoretical classes of codes, including time-invariant convolutional codes for \( L_D(d) \leq m + 1 \) and random convolutional codes [16], [3].

Turning now to tail-biting trellis representations of block codes, we modify the discussion, but not by much. Case (ii) path pairs, for which merger to all-zero is forbidden, must be considered, and paths must start and end in the same state. In addition, TB trellis representations are block codes, so the Gilbert–Varshamov bound applies; it states that a linear \((N, K, d)\) block code exists whenever

\[
\sum_{i=0}^{\infty} \binom{N-1}{i} 2^{N-K} < 2^{N-K}.
\]

(7)

As \( N, K, d \) grow, (7) takes the asymptotic form \( d / N \approx h^{-1}(1 - R) \), which may be rewritten for our code parameters as

\[
L \approx K / b \approx d / ch^{-1}(1 - R)
\]

(8)
or, equivalently,

\[
K \approx d R / h^{-1}(1 - R).
\]

(9)

If we assume that the Gilbert–Varshamov bound is tight, then (8) is the same as (6). The precise trends here for block codes are shown in Figs. 4–7. None of this guarantees that an \((N, K, d)\) TB representation exists with size \( L \) as small as (8). But if the TB trellis representations have performance equal to that of the best block codes, then their length \( L \) and distance \( d \) will be related as in (8). Figs. 4–7 show that this is apparently the case.

To summarize these remarks, consider a TB trellis representation with growing length, derived from generators of fixed memory \( m \). At small \( L \), case (ii) unmerged path pairs determine which generators have the best distance. As \( L \) grows, the best selection changes unpredictably, but our results show that the codes are excellent and that (9) approximately holds. At large \( L \), the case (i) \( d_{\text{inter}} \) must eventually dominate \( d \), since the minimum distance among case (ii) path pairs clearly grows linearly with \( L \) and exceeds any limit. This statement applies even though the TB condition applies: If a given distance \( d' \) is reached by \( L' \), descendants can reach any specified state by \( L' + m \), and their distance cannot be smaller. This being the case, the entire distance structure must come from case (i). The best TB generators of memory \( m \) become identical to the best convolutional code generators. So also must all the distance properties, and thus (6) and (7) apply to tail-biting trellis representations of block codes as they do to convolutional codes.

V. Conclusion

We have constructed a new list of optimal short and moderate-length encoders for tail-biting trellis representations of block codes. These are as good as the best known block codes at their rate and block length, and at the same time they have important advantages over block and convolutional codes. Soft channel outputs yield a 2–3-dB signal-to-noise ratio (SNR) improvement, and are easily obtained by trellis decoders. The trellis structure itself seems to lead to simpler decoding, compared to block decoders with the same \((N, K, d)\). Finally, the rate loss from the terminating bits in a convolutional encoder, which can be significant at moderate block lengths, is eliminated by tail-biting representations. The advantages summarized here are particularly important in packet and other short-block transmission systems.

Acknowledgment

The suggestions made by the reviewers helped improve this correspondence.

References