ON THE EFFECTIVE WEIGHTS OF PSEUDOCODEWORDS FOR CODES DEFINED ON GRAPHS WITH CYCLES

G. DAVID FORNEY, JR.*, RALF KOETTER†, FRANK R. KSCHISCHANG‡, AND ALEX REZNIK§

Abstract. The behavior of an iterative decoding algorithm for a code defined on a graph with cycles and a given decoding schedule is characterized by a cycle-free computation tree. The pseudocodewords of such a tree are the words that satisfy all tree constraints; pseudocodewords govern decoding performance. Wiberg [12] determined the effective weight of pseudocodewords for binary codewords on an AWGN channel. This paper extends Wiberg's formula for AWGN channels to nonbinary codes, develops similar results for BSC and BEC channels, and gives upper and lower bounds on the effective weight. The 16-state tail-biting trellis of the Golay code [2] is used for examples. Although in this case no pseudocodeword is found with effective weight less than the minimum Hamming weight of the Golay code on an AWGN channel, it is shown by example that the minimum effective pseudocodeword weight can be less than the minimum codeword weight.

Key words. Codes on graphs, iterative decoding, pseudocodewords, effective weights, tail-biting.

AMS(MOS) subject classifications. 94B99

1. Introduction.

The subject of codes defined on graphs was founded by Tanner [10], inspired by Gallager's low-density parity-check (LDPC) codes [4]. The thesis of Wiberg [12, 13], along with the practical successes of turbo codes and LDPC codes, has stimulated great current interest in this subject. For recent developments, see [1, 6, 7, 8].

By now it is well known that if C is a block code defined on a cycle-free graph G (*i.e.,* a tree), then the min-sum decoding algorithm is guaranteed to converge to the maximum-likelihood (ML) code sequence [12, 13].

The min-sum algorithm may also be applied to a graph with cycles, but its behavior then depends on the decoding schedule, and convergence is not guaranteed. Given a decoding schedule, there exists a cycle-free computation tree G' such that the behavior of the min-sum algorithm on G' with the given schedule is identical to that of the iterative algorithm on G [3, 11, 12, 13]. In general, a node in G has more than one representation in G'.

*G. David Forney, Jr. is with the Laboratory for Information and Decision Systems, M.I.T., Cambridge, MA 02139 USA.

†Ralf Koetter is with the Coordinated Science Laboratory, U. Illinois, Urbana, IL 61801 USA.

‡Frank R. Kschischang is with the Dept. of Electrical and Computer Engineering, U. Toronto, Toronto, Ont. M5S 3G4 Canada.

§Alex Reznik was with the Laboratory for Information and Decision Systems, M.I.T., Cambridge, MA 02139 USA; he is now with InterDigital Communications Corporation, Melville, NY 11747 USA.

A codeword in C is a sequence $\mathbf{c}$ of node values in G that satisfies all the constraints of G. A pseudocodeword [3] is a sequence of node values in G' that satisfies all the constraints of G'. There exists a pseudocodeword corresponding to every codeword $\mathbf{c} \in C$, obtained by assigning the values of the nodes of G to the corresponding nodes of G'. In general there will also exist pseudocodewords that do not correspond to valid codewords, because different values are assigned to nodes of G' that correspond to the same node of G.

In this note we will focus on pseudocodewords of tail-biting trellises, whose significance is particularly clear. A tail-biting trellis (TBT) corresponds to a graph G that consists of a single cycle. A computation tree G' is obtained by "unwrapping" G into a conventional trellis defined on an ordered time axis. A codeword $\mathbf{c}$ of G corresponds to a pseudocodeword on G' that repeats periodically with a period equal to one cycle of G. But there also exist periodic pseudocodewords on G' whose period is a multiple of the cycle length of G and which do not correspond to any valid codeword.

In Chapter 6 of [12], Wiberg developed a formula for the effective Hamming weight ("generalized weight") w_{eff} of a pseudocodeword ("tree configuration") for the case of binary codes and binary antipodal signaling on an additive white Gaussian noise (AWGN) channel, namely

$$w_{\text{eff}} = \frac{\left(\sum_j n_j\right)^2}{\sum_j n_j^2}, \tag{1.1}$$

where n_j is the number of nodes in G' corresponding to the jth node in G that have value equal to 1. If the pseudocodeword corresponds to a valid codeword $\mathbf{c}$, then $w_{\text{eff}} = w_{\text{H}}(\mathbf{c})$, the Hamming weight of $\mathbf{c}$. For a binary linear code, the probability of a decoding error on G' is governed by the minimum effective weight w_{eff}; therefore it is important that all pseudocodewords of G' have effective weight w_{eff} at least as great as the minimum Hamming weight d_{H} of C if performance is not to be degraded.

In this note we develop some extensions of Wiberg's result, as follows:

1. We extend Wiberg's formula to the nonbinary case;
2. We give lower and upper bounds on w_{eff};
3. We develop similar results for the binary symmetric channel (BSC) and binary erasure channel (BEC).

As examples, we compute the effective weights of certain pseudocodewords in the 16-state TBT of the binary (24, 12, 8) Golay code of [2]. For the AWGN channel, we have not found any examples of pseudocodewords with effective weight less than 8; however, neither have we been able to prove that 8 is the minimum effective weight for this case. For the binary symmetric channel, on the other hand, we exhibit a pseuodcodeword with effective weight 6.

2. Effective weight on AWGN channels. Let C be a block code of length n defined on a graph G, and let G' be a computation tree corresponding to some schedule for min-sum decoding of C. Let $N_j, 1 \le j \le n$, be the number of occurrences of the jth node of G in G'.

Let $\mathbf{c} = \{c_j, 1 \le j \le n\}$ be a codeword of G, and let $\{p_{ji}, 1 \le j \le n, 1 \le i \le N_j\}$ be a pseudocodeword of G', where p_{ji} represents the value of the ith occurrence of the jth node, $1 \le i \le N_j$.

For each symbol x_m in the symbol alphabet A, let n_{jm} be the number of times that $p_{ji} = x_m$, and let $f_{jm} = n_{jm}/N_j$; *i.e.,* f_{jm} is the frequency with which x_m appears in the N_j occurrences of the jth node. In the following, we think of the fractions f_{jm} as defining a random pseudocodeword $\mathbf{p} = \{p_j\}$ in which p_j takes on value $x_m \in A$ with probability f_{jm}, and we will take expectations over this distribution.

Note that $\mathbf{p}$ is non-random ($\mathbf{p} = \mathsf{E}[\mathbf{p}]$) if and only if $\mathbf{p}$ corresponds to a valid codeword $\mathbf{c}'$, for then and only then $f_{jm} = 1$ when $x_m = c'_j$ and $f_{jm} = 0$ otherwise.

Define the variance

$$\sigma_p^2 = \sum_j \left(\mathsf{E}\left[p_j^2\right] - \mathsf{E}\left[p_j\right]^2 \right) = \mathsf{E}\left[\|\mathbf{p}\|^2\right] - \|\mathsf{E}\left[\mathbf{p}\right]\|^2. \tag{2.1}$$

We have the following obvious lemma:

LEMMA 2.1. *The variance σ_p^2 is greater than or equal to 0, with equality if and only if $\mathbf{p}$ is non-random;* i.e., *iff $\mathbf{p}$ corresponds to a valid codeword $\mathbf{c}'$.* □

Let $\mathbf{c}$ be the input codeword to an AWGN channel whose output sequence is $\mathbf{r} = \mathbf{c} + \mathbf{n}$, where $\mathbf{n}$ is an i.i.d. Gaussian sequence with mean 0 and variance σ^2 per symbol. A maximum-likelihood (ML) decoder on G' chooses the pseudocodeword $\mathbf{p}$ that minimizes the squared distance

$$||\mathbf{r} - \mathbf{p}||^2 = \sum_j N_j \sum_m f_{jm}(r_j - x_m)^2.$$

For simplicity we will assume that G' is balanced [3, 11]— *i.e.,* that N_j is constant for all j. Then an ML decoder chooses the pseudocodeword $\mathbf{p}$ that minimizes the expected squared distance

$$\mathsf{E}\left[\|\mathbf{r} - \mathbf{p}\|^2\right] = \sum_m \Pr[p_j = x_m](r_j - x_m)^2 = \sum_j \sum_m f_{jm}(r_j - x_m)^2.$$

The probability $\Pr(\mathbf{c} \to \mathbf{p})$ that an ML decoder will choose the pseudocodeword $\mathbf{p}$ over $\mathbf{c}$ is thus

$$\Pr(\mathbf{c} \to \mathbf{p}) = \Pr\left\{ \mathsf{E}\left[\|\mathbf{r} - \mathbf{p}\|^2\right] \le \|\mathbf{r} - \mathbf{c}\|^2 \right\}.$$

(If G' is not balanced, then a similar result holds if we replace $\mathsf{E}[\|\mathbf{r} - \mathbf{p}\|^2]$ by the expectation of the weighted squared distance $N_j(r_j - x_m)^2$.)

Defining $\langle \mathbf{r}, \mathbf{c} \rangle = \sum_j r_j c_j$, we can write

$$\begin{aligned} \|\mathbf{r} - \mathbf{c}\|^2 &= \|\mathbf{r}\|^2 - 2\langle \mathbf{r}, \mathbf{c} \rangle + \|\mathbf{c}\|^2; \\ \mathsf{E}\left[\|\mathbf{r} - \mathbf{p}\|^2\right] &= \|\mathbf{r}\|^2 - 2\langle \mathbf{r}, \mathsf{E}[\mathbf{p}] \rangle + \mathsf{E}\left[\|\mathbf{p}\|^2\right], \end{aligned}$$

where $\langle \mathbf{r}, \mathsf{E}[\mathbf{p}] \rangle = \sum_j r_j \mathsf{E}[p_j] = \sum_j r_j \sum_m f_{jm} x_m$. Thus

$$\Pr(\mathbf{c} \to \mathbf{p}) = \Pr\left\{2\langle \mathbf{r}, \mathbf{c} - \mathsf{E}[\mathbf{p}] \rangle \le \|\mathbf{c}\|^2 - \mathsf{E}\left[\|\mathbf{p}\|^2\right]\right\}.$$

Define $\mathbf{d} = \mathbf{c} - \mathsf{E}[\mathbf{p}]$ and $D = \|\mathbf{c}\|^2 - \mathsf{E}\left[\|\mathbf{p}\|^2\right]$; then

$$\Pr(\mathbf{c} \to \mathbf{p}) = \Pr\left\{2\langle \mathbf{r}, \mathbf{d} \rangle \le D\right\}.$$

Given c_j, the received symbol r_j is a Gaussian random variable (r.v.) with mean c_j and variance σ^2. Therefore $r_j d_j$ is a Gaussian r.v. with mean $c_j d_j$ and variance $\sigma^2 d_j^2$. The inner product $\langle \mathbf{r}, \mathbf{d} \rangle = \sum_j r_j d_j$ therefore has mean $\langle \mathbf{c}, \mathbf{d} \rangle = \sum_j c_j d_j$ and variance $\sigma^2 \sum_j d_j^2 = \sigma^2 \|\mathbf{d}\|^2$. The probability that $\langle \mathbf{r}, \mathbf{d} \rangle \le D/2$ is thus the probability that a Gaussian r.v. with mean $\langle \mathbf{c}, \mathbf{d} \rangle - D/2$ and variance $\sigma^2 \|\mathbf{d}\|^2$ is less than zero, which is given by

$$Q\left(\frac{\langle \mathbf{c}, \mathbf{d} \rangle - D/2}{\sigma \|\mathbf{d}\|}\right),$$

where $Q(x) = \frac{1}{2\pi} \int_x^\infty \exp(-x^2/2)\, dx$ is the usual Q function.

Therefore if we define the effective squared Euclidean distance as

$$d^2_{\text{eff}}(\mathbf{c}, \mathbf{p}) = \frac{\left(2\langle \mathbf{c}, \mathbf{d} \rangle - D\right)^2}{\|\mathbf{d}\|^2}, \tag{2.2}$$

then we obtain the familiar expression

$$\Pr(\mathbf{c} \to \mathbf{p}) = Q\left(\frac{d_{\text{eff}}(\mathbf{c}, \mathbf{p})}{2\sigma}\right). \tag{2.3}$$

If $\mathbf{p}$ corresponds to a codeword $\mathbf{c}'$, then $\mathsf{E}[\mathbf{p}] = \mathbf{c}'$ and $\mathsf{E}[\|\mathbf{p}\|^2] = \|\mathbf{c}'\|^2$, so

$$2\langle \mathbf{c}, \mathbf{d} \rangle - D = 2\langle \mathbf{c}, \mathbf{c} - \mathbf{c}' \rangle - \|\mathbf{c}\|^2 + \|\mathbf{c}'\|^2 = \|\mathbf{c} - \mathbf{c}'\|^2.$$

Also $\|\mathbf{d}\|^2 = \|\mathbf{c} - \mathbf{c}'\|^2$, so we have as usual

$$d^2_{\text{eff}}(\mathbf{c}, \mathbf{c}') = \|\mathbf{c} - \mathbf{c}'\|^2.$$

More generally, if $\sigma_p^2 = \mathsf{E}\left[\|\mathbf{p}\|^2\right] - \|\mathsf{E}[\mathbf{p}]\|^2$ is the variance of $\mathbf{p}$ defined in (2.1), then we have

$$\begin{aligned} 2\langle \mathbf{c}, \mathbf{d} \rangle - D &= 2\langle \mathbf{c}, \mathbf{c} - \mathsf{E}[\mathbf{p}] \rangle - \|\mathbf{c}\|^2 + \mathsf{E}\left[\|\mathbf{p}\|^2\right] \\ &= \|\mathbf{c} - \mathsf{E}[\mathbf{p}]\|^2 + \mathsf{E}\left[\|\mathbf{p}\|^2\right] - \|\mathsf{E}[\mathbf{p}]\|^2 \\ &= \|\mathbf{d}\|^2 + \sigma_p^2. \end{aligned}$$

This leads to our main theorem:

THEOREM 2.1. *Let* $\mathbf{c}$ *be a codeword and* $\mathbf{p}$ *a pseudocodeword in a balanced computation tree. Then the effective squared Euclidean distance between* $\mathbf{c}$ *and* $\mathbf{p}$ *is*

$$d^2_{\text{eff}}(\mathbf{c},\mathbf{p}) = \frac{\left(\|\mathbf{d}\|^2 + \sigma_p^2\right)^2}{\|\mathbf{d}\|^2}, \tag{2.4}$$

where $\mathbf{d} = \mathbf{c} - \mathsf{E}[\mathbf{p}]$ *and* $\sigma_p^2 = \mathsf{E}[\|\mathbf{p}\|^2] - \|\mathsf{E}[\mathbf{p}]\|^2$. *If* $\mathbf{c}$ *is transmitted, then the probability that the received word* $\mathbf{r}$ *is closer to* $\mathbf{p}$ *than* $\mathbf{c}$ *is*

$$Pr(\mathbf{c} \to \mathbf{p}) = Q\left(\frac{d_{\text{eff}}(\mathbf{c},\mathbf{p})}{2\sigma}\right). \tag{2.5}$$

□

By Lemma 1, $\sigma_p^2 \geq 0$, with equality if and only if $\mathbf{p}$ corresponds to a codeword $\mathbf{c}'$. Thus we obtain the following two lower bounds on $d^2_{\text{eff}}(\mathbf{c},\mathbf{p})$:

COROLLARY 2.1. *The effective squared Euclidean distance* $d^2_{\text{eff}}(\mathbf{c},\mathbf{p})$ *satisfies*

$$d^2_{\text{eff}}(\mathbf{c},\mathbf{p}) \geq \|\mathbf{d}\|^2 + \sigma_p^2 \geq \|\mathbf{d}\|^2 = \|\mathbf{c} - \mathsf{E}[\mathbf{p}]\|^2, \tag{2.6}$$

with equality in both cases if and only if $\mathbf{p}$ *corresponds to a codeword* $\mathbf{c}'$. □

This result shows that the variation σ_p^2 in non-codeword pseudocodewords $\mathbf{p}$ causes $d^2_{\text{eff}}(\mathbf{c},\mathbf{p})$ to be greater than the squared distance $\|\mathbf{d}\|^2 = \|\mathbf{c} - \mathsf{E}[\mathbf{p}]\|^2$ between $\mathbf{c}$ and the average $\mathsf{E}[\mathbf{p}]$. Thus the more variation σ_p^2 in a pseudocodeword $\mathbf{p}$, the less troublesome it is likely to be, for a given average $\mathsf{E}[\mathbf{p}]$.

3. Binary signaling on the AWGN channel. With binary antipodal signaling using the symbol alphabet $A = \{\pm 1\}$, we have $\|\mathbf{c}\|^2 = \mathsf{E}[\|\mathbf{p}\|^2]$, so $D = 0$. Define

$$d_j = 1 - \mathsf{E}[p_j] = 1 - \sum_m f_{jm} x_m = 2f_j,$$

where f_j is the fraction of p_{ji} equal to -1; *i.e.,* $f_j = (1 - \mathsf{E}[p_j])/2$. If we take $\mathbf{c} = \mathbf{1}$ (the all-zero codeword), then

$$d^2_{\text{eff}}(\mathbf{1},\mathbf{p}) = 4w_{\text{eff}}(\mathbf{p}) = 4\frac{\left(\sum_j d_j\right)^2}{\sum_j d_j^2}.$$

If we further define $|\mathbf{f}| = \sum_j f_j$ and $\|\mathbf{f}\|^2 = \sum_j f_j^2$, then we have

$$\begin{aligned} \|\mathbf{d}\|^2 &= \|1 - \mathsf{E}[\mathbf{p}]\|^2 = 4\|\mathbf{f}\|^2; \\ \sigma_p^2 &= \mathsf{E}\left[\|\mathbf{p}\|^2\right] - \|\mathsf{E}[\mathbf{p}]\|^2 = 4(|\mathbf{f}| - \|\mathbf{f}\|^2); \\ d^2_{\text{eff}}(\mathbf{1},\mathbf{p}) &= \frac{\left(\|\mathbf{d}\|^2 + \sigma_p^2\right)^2}{\|\mathbf{d}\|^2} = 4\frac{|\mathbf{f}|^2}{\|\mathbf{f}\|^2}. \end{aligned}$$

This yields Wiberg's formula (1.1) for $w_{\text{eff}}(\mathbf{p})$ in the balanced case:

COROLLARY 3.1. *On an AWGN channel, the effective Hamming weight of a binary pseudocodeword* $\mathbf{p}$ *with frequency* f_j *of ones in the* j*th position is*

$$w_{\text{eff}}(\mathbf{p}) = \frac{|\mathbf{f}|^2}{\|\mathbf{f}\|^2}, \tag{3.1}$$

where $|\mathbf{f}| = \sum_j f_j$ *and* $\|\mathbf{f}\|^2 = \sum_j f_j^2$. □

The quantity $|\mathbf{f}| = \sum_j f_j$ may be interpreted as the average Hamming weight of $\mathbf{p}$. Corollary 2.1 then has the following corollary:

COROLLARY 3.2. *If* $\mathbf{p}$ *is binary, then its effective weight* $w_{\text{eff}}(\mathbf{p})$ *satisfies*

$$w_{\text{eff}}(\mathbf{p}) \geq |\mathbf{f}| \geq \|\mathbf{f}\|^2, \tag{3.2}$$

with equality if and only if $\mathbf{p}$ *corresponds to a codeword* $\mathbf{c}'$. □

In other words, $w_{\text{eff}}(\mathbf{p})$ is lowerbounded by the average Hamming weight $|\mathbf{f}|$, with strict inequality if $\mathbf{p}$ does not correspond to a codeword.

For example, three pseudocodewords of low effective weight in the Golay TBT have the following parameters:

Example 1 Suppose f_j equals $\frac{1}{2}$ in 8 places and 0 elsewhere. Then

$$\begin{aligned} \|\mathbf{f}\|^2 &= 8 \times (1/4) = 2; \\ |\mathbf{f}| &= 8 \times (1/2) = 4; \\ w_{\text{eff}}(\mathbf{p}) &= \frac{|\mathbf{f}|^2}{\|\mathbf{f}\|^2} = \frac{4^2}{2} = 8. \end{aligned}$$

Example 2 Suppose f_j equals $\frac{1}{2}$ in 8 places, 1 in 2 places, and 0 elsewhere. Then

$$\begin{aligned} \|\mathbf{f}\|^2 &= 8 \times (1/4) + 2 \times 1 = 4; \\ |\mathbf{f}| &= 8 \times (1/2) + 2 \times 1 = 6; \\ w_{\text{eff}}(\mathbf{p}) &= \frac{|\mathbf{f}|^2}{\|\mathbf{f}\|^2} = \frac{6^2}{4} = 9. \end{aligned}$$

Example 3 Suppose f_j equals $\frac{1}{3}$ in 6 places, $\frac{2}{3}$ in 2 places, 1 in 2 places, and 0 elsewhere. Then

$$\begin{aligned} \|\mathbf{f}\|^2 &= 6 \times (1/9) + 2 \times (4/9) + 2 = 32/9; \\ |\mathbf{f}| &= 6 \times (1/3) + 2 \times (2/3) + 2 = 16/3; \\ w_{\text{eff}}(\mathbf{p}) &= \frac{|\mathbf{f}|^2}{\|\mathbf{f}\|^2} = 2^3 = 8. \end{aligned}$$

Example 1 illustrates the general proposition that if all nonzero f_j are equal, then $w_{\text{eff}}(\mathbf{p})$ is equal to their number, $w_{\text{eff}}(\mathbf{p}) = |\operatorname{supp}(\mathbf{f})|$. (We will see that the support size $|\operatorname{supp}(\mathbf{f})|$ is the effective weight on a binary erasure channel.) More generally, we can show that $|\operatorname{supp}(\mathbf{f})|$ is an upper bound on $w_{\text{eff}}(\mathbf{p})$:

THEOREM 3.1. *If* $\mathbf{p}$ *is binary, then its effective weight* $w_{\text{eff}}(\mathbf{p})$ *satisfies*

$$w_{\text{eff}}(\mathbf{p}) \le |\operatorname{supp}(\mathbf{f})|, \tag{3.3}$$

with equality if and only if all nonzero f_j *are equal.*

Proof. Define $\mathbf{1}$ as the vector whose components are equal to one on the support of $\mathbf{f}$ and equal to zero elsewhere. Then $\|\mathbf{1}\|^2 = |\operatorname{supp}(\mathbf{f})|$, and $\langle \mathbf{1}, \mathbf{f} \rangle = |\mathbf{f}|$. Now by Schwarz's inequality,

$$\langle \mathbf{1}, \mathbf{f} \rangle^2 \le \|\mathbf{1}\|^2 \|\mathbf{f}\|^2 = |\operatorname{supp}(\mathbf{f})| \cdot \|\mathbf{f}\|^2,$$

with equality if and only if $\mathbf{f} = \alpha \mathbf{1}$ for some α. The conclusion follows from

$$w_{\text{eff}}(\mathbf{p}) = \frac{|\mathbf{f}|^2}{\|\mathbf{f}\|^2} = \frac{\langle \mathbf{1}, \mathbf{f} \rangle^2}{\|\mathbf{f}\|^2} \le |\operatorname{supp}(\mathbf{f})|,$$

with equality if and only if $\mathbf{f}$ is proportional to $\mathbf{1}$. □

4. Binary symmetric channels. Now let us consider binary signaling on a binary symmetric channel (BSC) with crossover probability $\varepsilon < 1/2$. As in the previous section, a pseudocodeword $\mathbf{p}$ will be represented by a vector $\mathbf{f}$, where f_j is the fraction of pseudocodeword components equal to 1 in the jth position.

Suppose that the all-zero word is sent and that the received word is $\mathbf{e}$, where $e_j = 1$ if there is an error in the jth position and $e_j = 0$ otherwise. The Hamming distance between $\mathbf{e}$ and the all-zero word is $|\mathbf{e}| = \sum_j e_j$. Given a pseudocodeword $\mathbf{p}$ represented by $\mathbf{f}$, the average Hamming distance in the jth component is equal to f_j if $e_j = 0$ and $1 - f_j$ if $e_j = 1$, so the average Hamming distance $d_{\text{H}}(\mathbf{e}, \mathbf{p})$ is

$$d_{\text{H}}(\mathbf{e}, \mathbf{p}) = \sum_j f_j(1 - e_j) + e_j(1 - f_j). \tag{4.1}$$

Thus the error event $\{d_{\text{H}}(\mathbf{e}, \mathbf{p}) \le |\mathbf{e}|\}$ is the event

$$\{\langle \mathbf{f}, \mathbf{1} - 2\mathbf{e} \rangle = \sum_j f_j(1 - 2e_j) = \sum_j f_j(-1)^{e_j} \le 0\}. \tag{4.2}$$

The probability of error is thus the probability that $\sum_j f_j(-1)^{e_j} \le 0$. This is a sum of independent random variables v_j, where $v_j = f_j$ with probability ε and $v_j = -f_j$ with probability $1 - \varepsilon$.

We would like again to define the effective Hamming weight $w_{\text{eff}}(\mathbf{p})$ of $\mathbf{p}$ as a single parameter that has the same significance as the usual

Hamming weight $w_{\mathrm{H}}(\mathbf{c})$ of a codeword $\mathbf{c}$ for error probability, and that reduces to Hamming weight when $\mathbf{p}$ is actually a codeword. This cannot be done quite as neatly in the BSC case as in the Gaussian case, but a reasonable approach is as follows.

We ask for the minimum number of errors $|\mathbf{e}|$ that can cause a decoding error to $\mathbf{p}$. Clearly, given the total weight $|\mathbf{f}| = \sum_j f_j$ of $\mathbf{f}$, the worst case occurs when e errors occur in the e positions for which f_j is greatest. A decoding error may occur if the sum of the weights f_j in these e positions is equal to $|\mathbf{f}|/2$, and must occur if the sum exceeds $|\mathbf{f}|/2$. The effective weight in the former case will be taken as $w_{\mathrm{BSC}}(\mathbf{p}) = 2e$, and in the latter case as $w_{\mathrm{BSC}}(\mathbf{p}) = 2e - 1$.

To a first approximation, the decoding error probability to $\mathbf{p}$ will therefore be of the order of $K\varepsilon^{e(\mathbf{p})}$, where $e(\mathbf{p}) = \lceil w_{\mathrm{BSC}}(\mathbf{p})/2 \rceil$ is the minimum number of channel errors required to make a decoding error to $\mathbf{p}$.

The correspondence between effective weight and Hamming weight is not precise, because whereas with an ordinary codeword the multiplicity K is the number of ways that $e(\mathbf{c})$ errors can occur in $w_{\mathrm{H}}(\mathbf{c})$ positions, with pseudocodewords the number of possible combinations of $e(\mathbf{p})$ errors will in general be less.

The three examples given earlier illustrate these points and show that the effective weight for Gaussian channels and for BSCs are in general different.

Example 1 (cont.) If f_j equals $\frac{1}{2}$ in 8 places and 0 elsewhere, then $|\mathbf{f}| = 4$. There exist error patterns of weight $e = 4$ such that the sum of the e largest components of $\mathbf{f}$ is equal to $|\mathbf{f}|/2 = 2$, namely any error pattern with 4 errors in places where $f_j = \frac{1}{2}$. Thus the effective weight of such a pseudocodeword is $w_{\mathrm{BSC}}(\mathbf{p}) = 8$. In this case the number of error patterns of weight 4 that could cause a decoding error is $\frac{8!}{4!4!} = 70$, as in the usual case.

Example 2 (cont.) If f_j equals $\frac{1}{2}$ in 8 places, 1 in 2 places, and 0 elsewhere, then $|\mathbf{f}| = 6$. There exist error patterns of weight $e = 4$ such that the sum of the e largest components of $\mathbf{f}$ is equal to $|\mathbf{f}|/2 = 3$, namely error patterns with errors in the two places where $f_j = 1$ and in two other places where $f_j = \frac{1}{2}$. Thus the effective weight of such a pseudocodeword is $w_{\mathrm{BSC}}(\mathbf{p}) = 8$. The number of error patterns of weight 4 that could cause a decoding error is $\frac{8!}{6!2!} = 28$, compared to $\frac{8!}{4!4!} = 70$ in the usual case.

Example 3 (cont.) If f_j equals $\frac{1}{3}$ in 6 places, $\frac{2}{3}$ in 2 places, 1 in 2 places, and 0 elsewhere, then $|\mathbf{f}| = \frac{16}{3}$. There exist error patterns of weight $e = 3$ such that the sum of the e largest components of $\mathbf{f}$ is equal to $|\mathbf{f}|/2 = \frac{8}{3}$, namely error patterns with errors in the two places where $f_j = 1$ and in one other place where $f_j = \frac{2}{3}$. The effective weight of such a pseudocodeword is $w_{\mathrm{BSC}}(\mathbf{p}) = 6$. The number of error patterns of weight 3 that could cause a decoding error is 2, compared to $\frac{6!}{3!3!} = 20$ in the usual case.

From these examples one might conjecture that the effective Hamming weight of a binary pseudocodeword on a BSC is always less than or equal to its effective weight on an AWGN channel. We can construct a counterexample to such a conjecture as follows. Let d be an even integer greater than 4, let δ be a very small number such as $\delta = 0.001$, and let

$$N = d - 2 + \frac{1}{\delta},$$

where we assume that $1/\delta$ is an integer. Consider a set of nonzero weights f_j with one weight equal to 1 and N weights equal to δ. Then

$$\begin{aligned} |\mathbf{f}| &= 2 + \delta(d-2) \approx 2; \\ ||\mathbf{f}||^2 &= 1 + N\delta^2 \approx 1; \\ w_{\text{eff}}(\mathbf{p}) &= \frac{|\mathbf{f}|^2}{||\mathbf{f}||^2} \approx 4; \\ e &= 1 + \frac{d-2}{2} = \frac{d}{2}; \\ w_{\text{BSC}}(\mathbf{p}) &= 2e = d > w_{\text{eff}}(\mathbf{p}), \end{aligned}$$

where we note that on a BSC it takes one error in the position where $f_j = 1$ and $(d-2)/2$ errors in positions where $f_j = \delta$ to accumulate a weight of $|\mathbf{f}|/2 = 1 + \delta(d-2)/2$.

5. Binary erasure channels. Now let us consider binary signaling on a binary erasure channel (BEC) with erasure probability ε. Again, a pseudocodeword $\mathbf{p}$ will be represented by a vector $\mathbf{f}$, where f_j is the fraction of pseudocodeword symbols equal to 1 in the jth position.

Suppose that the all-zero word is sent and that $|S|$ erasures occur in a certain set S of coordinates. The remaining unerased symbols will all agree with the all-zero word. They will evidently also all agree with a pseudocodeword $\mathbf{p}$ represented by $\mathbf{f}$ if and only if $f_j = 0$ for all $j \notin S$.

Therefore we define the effective Hamming weight $w_{\text{BEC}}(\mathbf{p})$ of a pseudocodeword on a BEC as $|\operatorname{supp}(\mathbf{f})|$, the number of nonzero components of $\mathbf{f}$. Then:

(a) A decoding error to $\mathbf{p}$ may occur if and only if $|S| \geq w_{\text{BEC}}(\mathbf{p})$;

(b) If $\mathbf{p}$ is actually a codeword $\mathbf{c}$, then $w_{\text{BEC}}(\mathbf{p}) = w_{\text{H}}(\mathbf{c})$.

By Theorem 3.1, the effective weight $w_{\text{BEC}}(\mathbf{p})$ of a pseudocodeword $\mathbf{p}$ on a BEC is greater than or equal to its effective weight on an AWGN channel, with equality if and only if $\mathbf{p}$ is actually a codeword. Similarly, by the discussion in Section 4, $w_{\text{BEC}}(\mathbf{p}) \geq w_{\text{BSC}}(\mathbf{p})$, with equality if and only if $\mathbf{p}$ is actually a codeword.

For example, for Examples 1, 2 and 3, the effective weights on a BEC are 8, 10 and 10, respectively, compared to 8, 9 and 8 on an AWGN channel and 8, 8 and 6 on a BSC.

6. Examples with low pseudocodeword weights. In this section, we present a family of examples of binary tail-biting trellises for which the minimum effective pseudocodeword weight on an AWGN channel is strictly less than the minimum codeword weight.

For a first example, let C be the binary linear $(16, 6, 4)$ code generated by the following 6 generators:

$$\begin{array}{llll} 100 & 110000 & 110000 & 1 \\ 010 & 001100 & 001100 & 1 \\ 001 & 000011 & 000011 & 1 \\ & & & \\ 000 & 110000 & 001100 & 0 \\ 000 & 001100 & 000011 & 0 \\ 000 & 000011 & 110000 & 0 \end{array}$$

With a TBT constructed from these generators, the sum of the shifted generators shown below gives a low-weight three-cycle pseudocodeword:

000	000011	110000	0								
		110000	1	100	110000						
				000	110000	001100	0				
						001100	1	010	001100		
								000	001100	000011	0
001	000011									000011	1
001	000000	000000	1	100	000000	000000	1	010	000000	000000	1

The resulting pseudocodeword has $f_j = \frac{1}{3}$ in 3 places, $f_j = 1$ in 1 place, and 0 elsewhere. Thus $|\mathbf{f}| = 2$, $||\mathbf{f}||^2 = \frac{4}{3}$, and the effective Hamming weight on an AWGN channel is $|\mathbf{f}|^2/||\mathbf{f}||^2 = 3$. Notice that since one position has weight $|\mathbf{f}|/2 = 1$, the effective weight on a BSC is only 2.

A generalization of this construction yields for every integer $a \geq 3$ a binary linear (n, k, d) code C with $d = 2\lceil a/2 \rceil$ and a TBT with an a-cycle pseudocodeword with effective weight

$$w_{\text{eff}} = \frac{4a}{a+1} < d.$$

The generator matrix has the form

$$\begin{bmatrix} I & B & B & 1 \\ O & B & C & 0 \end{bmatrix},$$

where I is an $a \times a$ identity matrix, B is the matrix obtained from I by repeating every column b times where $b = \lceil a/2 \rceil$, 1 is a column of a ones, O is an $a \times a$ zero matrix, C is the cyclic shift of B to the right b times, and 0 is a column of a zeroes. It is straightforward to verify that the minimum nonzero codeword weight is $d = 2b$.

By a similar concatenation to that above, we obtain an a-cycle pseudocodeword with $f_j = \frac{1}{a}$ in a places, $f_j = 1$ in 1 place, and 0 elsewhere.

Thus $|\mathbf{f}| = 2$, $||\mathbf{f}||^2 = (a+1)/a$, and the effective Hamming weight on an AWGN channel is $|\mathbf{f}|^2/||\mathbf{f}||^2 = 4a/(a+1) < 4$. Notice that in general one position has weight $|\mathbf{f}|/2 = 1$, so the effective weight on a BSC is only 2. On the other hand, the effective weight on a BEC is $a + 1 \geq d$.

For $a = 3, 4, 5, \ldots$, the minimum nonzero codeword weight is $4, 4, 6, \ldots$, while the pseudocodeword weight on an AWGN channel is $3, 3.2, 3.33, \ldots$, approaching a limit of 4.

We note that the TBT that produces this low-weight pseudocodeword is not in general minimal in the sense of [5]; however, it is linear, biproper and one-to-one.

7. Conclusions. We have determined the effective weight and distance of pseudocodewords on the AWGN channel, the BSC, and the BEC. In general pseudocodewords are least troublesome on a BEC.

For the Golay TBT, we have found pseudocodewords whose effective weight on the BSC is less than the minimum distance of the code, which indicates that ML decoding using this TBT will be distinctly suboptimal. For the AWGN channel, we have found no such pseudocodewords; moreover, simulations have shown that ML decoding using the Golay TBT is near-optimal [9]. However, as far as we know, there is no proof yet that the minimum nonzero pseudocodeword weight is 8.

For more general graphs, the concept of pseudocodeword may need some refinement. Just as in Viterbi decoding the influence of symbols far in the past eventually dies out, at least probabilistically, we expect that the influence of nodes far away from the root node in the computation tree will eventually die out. The concepts of pseudocodeword weight used in this paper do not have this property, which suggests that they need refinement.

Acknowledgments. We wish to acknowledge helpful discussions with S. M. Aji, B. J. Frey, G. B. Horn, H.-A. Loeliger, R. J. McEliece, A. Vardy, N. Wiberg and M. Xu.

REFERENCES

[1] S. M. Aji and R. J. McEliece, "The generalized distributive law," *IEEE Trans. Inform. Theory*, vol. 46, pp. 325-343, March 2000.

[2] A. R. Calderbank, G. D. Forney, Jr. and A. Vardy, "Minimal tail-biting trellises: The Golay code and more," *IEEE Trans. Inform. Theory*, vol. 45, pp. 1435–1455, July 1999.

[3] B. J. Frey, R. Koetter and A. Vardy, "Skewness and pseudocodewords in iterative decoding," *Proc. 1998 IEEE Intl. Symp. Inform. Theory* (Cambridge, MA), p. 148, Aug. 1998.

[4] R. G. Gallager, *Low-Density Parity-Check Codes.* Cambridge, MA: MIT Press, 1963.

[5] R. Koetter and A. Vardy, "Construction of minimal tail-biting trellises," in *Proc. 1998 Inform. Theory Workshop* (Killarney), pp. 72–74, June 1998.

[6] F. R. Kschischang and B. J. Frey, "Iterative decoding of compound codes by probability propagation in graphical models," *IEEE J. Selected Areas Commun.*, vol. 16, pp. 219–230, Feb. 1998.

[7] F. R. Kschischang, B. J. Frey and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," submitted to *IEEE Trans. Inform. Theory*, July 1998.

[8] R. J. McEliece, D. J. C. MacKay and J.-F. Cheng, "Turbo decoding as an instance of Pearl's 'belief propagation' algorithm," *IEEE J. Selected Areas Commun.*, vol. 16, pp. 140–152, Feb. 1998.

[9] A. Reznik, "Iterative decoding of codes defined on graphs," M.Sc. thesis, M.I.T., Cambridge, MA, May 1998.

[10] R. M. Tanner, "A recursive approach to low complexity codes," *IEEE Trans. Inform. Theory*, vol. IT-27, pp. 533–547, Sept. 1981.

[11] Y. Weiss, "Correctness of local probability propagation in graphical models with loops," *Neural Comp.*, vol. 12, pp. 1-41, 2000.

[12] N. Wiberg, "Codes and decoding on general graphs," Ph.D. Thesis, U. Linköping, Sweden, 1996.

[13] N. Wiberg, H.-A. Loeliger and R. Koetter, "Codes and iterative decoding on general graphs," *Euro. Trans. Telecomm.*, vol. 6, pp. 513–525, Sept./Oct. 1995.