Optimal Erasure Selection of M-ary PAM Signaling for Errors and Erasures Decoding Algorithms

Yeong-Hyeon Kwon, Member, IEEE, Mi-Kyung Oh, Member, IEEE, and Dong-Jo Park, Member, IEEE

Abstract—Erasure selection for errors-and-erasures decoding algorithms is investigated and analyzed, where declaring erasures provides additional gain for hard decision decoding. To derive an analytical criterion for optimal erasure selection, we first investigate erasure effects on decoding performance, by which a metric function for erasure selection is derived. From the derived metric, a suboptimal threshold is defined, which can be used for errors and erasures decoding of large code-distance channel codes. Moreover, application to M-ary (M>2) PAM constellation is also described. For verification of the proposed erasure threshold, simulation results with popular channel codes are included.

Index Terms—Errors and erasures decoding (EED), hard decision decoding (HDD), maximum likelihood decoding (MLD), pulse amplitude modulation (PAM), generalized minimum distance (GMD) decoding.

I. INTRODUCTION

HARD decision decoding (HDD) algorithms have been extensively adopted for decoding practical codes, including linear block codes and convolutional codes [1]-[12]. Though HDD algorithms shows significant performance degradation compared to the optimal soft decision decoding (SDD) algorithms [1], [2], [8], [10], [15], HDD is usually appropriate for very high speed applications, including high speed wireless communication, magnetic storage and optical systems [8]. To compensate the performance degradation, HDD algorithms are often implemented as erasure insertion schemes, i.e., “errors-and-erasures decoding (EED)” [5]-[12], [17], [18].

Different EED algorithms have been developed depending on the baseline channel codes, including Turbo and LDPC decoding. It was reported that EED gives significant performance gain with properly chosen erasures [2], [5], [8], [15], [17]. To classify received symbols into erasures and valid symbols, decision measures, such as ratio threshold test (RTT), Bayesian decision, and output threshold test (OTT), have been introduced for M-ary frequency shift keying (MFSK) [17]-[21]. In [17]-[21], although the existence of optimal erasure threshold is asserted, exact formulation was not given to determine the optimal erasure threshold except the full simulation search as shown in [17], [18].

In this paper, we focus on calculating the optimal erasure threshold for M-ary pulse amplitude modulation (PAM) systems. For erasure declaration, we also use the threshold method which was proposed in the RTT and OTT [17]-[21]. Since the objective of any channel decoding algorithm is to find out the transmitted codeword, a codeword with minimum Hamming distance to the received sequence, we model EED in a general decoding form regardless of specific channel codes, which makes mathematical analysis more attractive as shown in Section II. Based on the general EED form, codeword error rate (CER) is parameterized with erasure threshold and then the optimal erasure threshold is obtained, which becomes independent of any underlying channel code.

This paper is organized as follows. In Section II, we first model the errors-and-erasures decoder in a mathematically tractable form. Based on the model, we derive the codeword error rate (CER) for EED, which gives the optimal erasure threshold in Section III. Due to difficulty of deriving a closed form solution of the optimal erasure threshold, we also introduce a suboptimal method for erasure threshold calculation in Subsection III-C. Then, we compare the proposed scheme with that of the capacity maximizing threshold in Subsection III-D and extend our result to M-ary PAM (M>2) signalling in Section IV. Finally, numerical simulations and conclusion are followed in Section V and VI.

Notation: Lower bold face letters denote vectors and $v(k)$ indicates the $k$th entry of a vector $v$. Symbol $\oplus$ stands for the exclusive OR operation.

II. PROBLEM FORMULATION

In this section, a general EED model is introduced in a mathematically tractable form. As a reference decoding scheme, maximum likelihood decoding (MLD) is defined as full search decoding, which shows the best decoding performance.

At the transmitter, a data packet $d$ of length $k$ is encoded into a codeword sequence $c$ of length $n$ by a channel encoder with code rate $R := k/n$, where $d(i) \in \{0, 1\}$ for $i = 1, \ldots, k$, and $c(i) \in \{0, 1\}$ for $i = 1, \ldots, n$. The BPSK modulated signal $x$ of $c$ is then transmitted through AWGN channel. We note here that although we first consider the BPSK modulation for simplicity, an extension of M-ary PAM is straightforward and discussed in Section IV.

The received signal is then given by $v = x + n$, where $n$ is zero-mean, white Gaussian noise with variance $N_0/2$. For HDD, each element of the received vector $v$ is decided into a bit sequence $v_{HD}$ by hard decision, i.e.,

$$v_{HD}(i) = \begin{cases} 
1, & v(i) > 0 \\
0, & \text{otherwise} 
\end{cases} \quad \text{for } i = 1, \ldots, n. \quad (1)$$
Then, HDD selects the codeword \( \mathbf{c}_{HDD}^* \) with the minimum Hamming distance as follows:

\[
\mathbf{c}_{HDD}^* = \arg \min_{\mathbf{c}_j \in S_c} D(\mathbf{c}_j, \mathbf{v}_{HD}),
\]

(2)

where \( S_c \) is the codeword space of the considered channel code and \( D(\mathbf{c}_j, \mathbf{v}_{HD}) := \sum_{i=1}^{n} e_j(i) \oplus v_{HD}(i) \) calculates the Hamming distance between the candidate codeword \( \mathbf{c}_j \) and \( \mathbf{v}_{HD} \).

If the codeword distance \( d_e(\mathbf{c}_i, \mathbf{c}_j) \) for \( i \neq j \) equals to \( d \), where \( \mathbf{c}_i, \mathbf{c}_j \in S_c \), then we can correct at most \( \left\lfloor \frac{d}{2} \right\rfloor \) errors by using the HDD. However, it has been shown that proper erasure declaration can increase the number of correctable errors. Specifically, \( d - 1 \) errors can be corrected if \( d - 1 \) errors are erased correctly [12], [15], [16]. In other words, the true transmitted codeword can be found if the following equation holds

\[
d > 2n_u + n_e,
\]

(3)

where \( n_u \) is the number of unerased errors and \( n_e \) is the number of erased errors.

We now consider how the MLD can be modelled with the declared erasures. First, the signal \( \mathbf{v} \) is processed by an erasure selector that generates an erasure indication sequence \( \mathbf{e} \) as following:

\[
e(i) = \begin{cases} 
0, & |v(i)| \leq \gamma_e \\
1, & \text{otherwise},
\end{cases}
\]

(4)

where \( i = 1, \ldots, n \) and \( \gamma_e \) is an erasure threshold to be optimized which will be discussed in Section III. It is observed from (4) that if the amplitude of the received signal is lower than \( \gamma_e \), the symbol is declared as an erasure. Using the erasure indicator \( \mathbf{e} \), the Hamming distance with erasures, \( D_e(\mathbf{c}_j, \mathbf{v}_{HD}, \mathbf{e}) \) can be defined as

\[
D_e(\mathbf{c}_j, \mathbf{v}_{HD}, \mathbf{e}) = \sum_{i=0}^{n-1} e(i) \cdot (c_j(i) \oplus v_{HD}(i)).
\]

(5)

Based on (5), the erasure decoding of the MLD gives the following codeword:

\[
\mathbf{c}_{EHD}^* = \arg \min_{\mathbf{c}_j \in S_c} D_e(\mathbf{c}_j, \mathbf{v}_{HD}, \mathbf{e}).
\]

(6)

A. Error Probability of Codeword Decision

Before further description, let us compare the scheme of (4) with the existing methods for erasure declaration. Note that the same form was already used by Forney when introducing the generalized minimum distance (GMD) decoding in which erasures are selected one-by-one by increasing the erasure threshold during iterative decoding trials [5], [6], [7].

On the other hand, the erasure selection problem was also studied for M-ary FSK modulation [17]-[21]. With Viterbi’s ratio threshold test (RTT), erasure is declared by inspecting the “maximum” and the “second maximum” values among \( M \) matched outputs, while the output threshold test (OTT) uses only the maximum matched output for erasure declaration [17], [18]. As a combination of RTT and OTT, the maximum output RTT (MO-RTT) was also proposed in [18]. In addition to these suboptimum decision methods, there exists the optimal erasure detection method, i.e., the Bayesian approach by Baum [20], [21], where all matched output values are considered. Note that the erasure decision with (4) is equivalent to the OTT and Baum’s Bayesian approaches, which can be simply proved by rearranging the declaration rules. Since the erasure threshold was not fully discussed for these test rules, we will describe how to select the best erasure threshold with (4). In the following, the erasure threshold will be chosen so that average EED decoding performance could be optimized with one decoding trial.

III. OPTIMAL SELECTION OF ERASURE THRESHOLD \( \gamma_e \)

In this section, decision error probability of the MLD model (6) is derived with a given erasure indicator \( \mathbf{e} \). Minimizing the derived error probability gives an erasure threshold \( \gamma_e^* \) to achieve the best average performance.

A. Error Probability of Codeword Decision

Without loss of generality, let us assume that the transmitted codeword is the all-zero codeword, i.e., \( \mathbf{e} = \mathbf{0} \). For the given threshold \( \gamma_e \), the erasure indicator \( \mathbf{e} \) is determined as in (4). Then, consider two Hamming distances, which are \( D_e(\mathbf{0}, \mathbf{v}_{HD}, \mathbf{e}) \) and \( D_e(\mathbf{c}_c, \mathbf{v}_{HD}, \mathbf{e}) \):

\[
D_e(\mathbf{0}, \mathbf{v}_{HD}, \mathbf{e}) = \sum_{i=1}^{n} e(i) \cdot (0 \oplus v_{HD}(i)),
\]

(7)

\[
D_e(\mathbf{c}_c, \mathbf{v}_{HD}, \mathbf{e}) = \sum_{i=1}^{n} e(i) \cdot (c_c(i) \oplus v_{HD}(i)),
\]

where \( D_e(\mathbf{0}, \mathbf{v}_{HD}, \mathbf{e}) \) refers to the distance between the desired codeword \( \mathbf{0} \) and the received codeword \( \mathbf{v}_{HD} \) when considering \( \mathbf{e} \), and \( D_e(\mathbf{c}_c, \mathbf{v}_{HD}, \mathbf{e}) \) indicates the distance between a competing codeword \( \mathbf{c}_c \) and the received codeword. For correct codeword decision, \( \mathbf{0} \) should be closest to \( \mathbf{v}_{HD} \). Therefore, the following inequality should hold:

\[
d(\mathbf{c}_c, \mathbf{0}, \mathbf{e}) = D_e(\mathbf{c}_c, \mathbf{v}_{HD}, \mathbf{e}) - D_e(\mathbf{0}, \mathbf{v}_{HD}, \mathbf{e}) > 0.
\]

(8)

We note that the above decision measure \( d(\mathbf{c}_c, \mathbf{0}, \mathbf{e}) \) is a random variable and takes values between \(-d \) and \( d \) according to errors and erasures. If there is no error, then \( d(\mathbf{c}_c, \mathbf{0}, \mathbf{e}) = d \). Otherwise, it becomes smaller than \( d \). Therefore, we can expect that decoding performance depends on \( \mathbf{e} \) since \( d \) can reduce the distance between \( \mathbf{0} \) and \( \mathbf{c}_c \) (see (3)).

To obtain the average decoding performance of (6), we model \( e(i) \) as a random variable because it is affected by the instantaneous additive noise. The fact that \( e(i) \in \{0, 1\} \) and probability distribution of \( e(i) \) is influenced by channel noise distribution, \( e(i) \) has a binomial distribution:

\[
P(e(i) = 0) = P_e(i | \gamma_e), \quad P(e(i) = 1) = 1 - P_e(i | \gamma_e),
\]

(9)

where \( P_e(i | \gamma_e) \) is erasure declaration probability for the \( i \)th symbol. Assuming white Gaussian channel noise, Fig. 1 shows the probability density of \( v(i) \) and the regions for correct/erasure symbol decision and erasure declaration. Denoting \( s_0 := 0 \) and \( s_1 := 1 \) which are the transmitted
symbols, the probabilities of correct/error/erasure decision can be defined as

\[ P_e(i \mid \gamma_e) = \sum_{k=0}^{1} \int_{y \in R_k} p(y \mid s_k) p(s_k)dy, \]

\[ P_e(i \mid \gamma_e) = \sum_{k=0}^{1} \int_{y \in R_{k \text{mod}(k+1,2)}} p(y \mid s_k) p(s_k)dy, \]

\[ P_r(i \mid \gamma_e) = 1 - P_e(i \mid \gamma_e) - P_r(i \mid \gamma_e), \]

where \( p(y \mid s_k) \) is the probability density function of the received signal when \( s_k \) is sent, and \( R_0 \) and \( R_1 \) indicate the decision regions for 0 and 1, respectively, as shown in Fig. 1.

We now define the effective distance between codewords with erasures, \( d(e) \) as

\[ d(e) = \sum_{i \in S_d} e(i), \]

where \( d(e) \in \{0, 1, \ldots, d\} \) and \( S_d \) is the set of positions where \( c_0 \) and \( \hat{0} \) are different. Then the probability of \( d(e) \) is given by

\[ P(d(e) = x) = \binom{d}{x} P_r^x(\gamma_e)(1 - P_r(\gamma_e))^{d-x}, \]

where \( x \) is the number of erased code bits in \( S_d \), \( \binom{n}{k} = \frac{n!}{(a-b)!b!} \) and \( P_r(\gamma_e) = P_r(i \mid \gamma_e) \) for \( 1 \leq i \leq n \). For the given \( e \), the probability of the codeword decision by (8) can be evaluated as

\[ P(d(c_e, \hat{0}) \leq 0) = P(d(c_e, \hat{0})|d(e) = x) \leq 0) \]

\[ = \sum_{m=0}^{d-x} \left[ \sum_{k=m+1}^{d-x} P_e^m(\gamma_e) P_r^{d-x-m}(\gamma_e) \right] \]

\[ = \frac{1}{(1 - P_r(\gamma_e))^{d-x}} \sum_{m=0}^{d-x} \left[ \sum_{k=m+1}^{d-x} P_r^m(\gamma_e) P_r^{d-x-m}(\gamma_e) \right], \]

where \( P_e^m(\gamma_e) := P_e(\gamma_e)/(1 - P_r(\gamma_e)) \) and \( P_r^m(\gamma_e) := P_r(\gamma_e)/(1 - P_r(\gamma_e)) \) are normalized error and correct probabilities with given \( e \), respectively. We note that the error probability of HDD can be obtained from (13) by setting \( \gamma_e = 0 \) and \( x = 0 \), resulting in

\[ P_{HDD}(\text{DE}|d) = \sum_{m=0}^{d} P_e^m(0) P_r^{d-m}(0), \]

since \( P_r(\gamma_e = 0) = 0 \), where \( \text{DE} \) denotes decoding error.

Averaging (13) over all possible combinations of \( e \), the average probability of codeword decision error is given by

\[ P_{EHDD}(\text{DE}|\gamma_e, d) \triangleq E[P(d(c_e, \hat{0}|e) \leq 0)|e] \]

\[ = \sum_{x=0}^{d} P(d(e) = x) P(d(c_e, \hat{0}|d(e) = x) \leq 0). \]

By inserting (12) and (13) into (15), we finally obtain

\[ P_{EHDD}(\text{DE}|\gamma_e, d) \]

\[ = \sum_{x=0}^{d} \binom{d}{x} P_r^x(\gamma_e) \sum_{k=0}^{d-x} P_e^k(\gamma_e) P_r^{d-x-k}(\gamma_e), \]

Using (14) and (16), the average decoding error probability is bounded as follows:

\[ P_{HDD}(\text{DE}) \leq \sum_{d=d_{\text{min}}}^{d_{\text{max}}} f_d(d) P_{HDD}(\text{DE}|d), \]

and

\[ P_{EHDD}(\text{DE}) \leq \sum_{d=d_{\text{min}}}^{d_{\text{max}}} f_d(d) P_{EHDD}(\text{DE}|\gamma_e, d), \]

where \( f_d(d) \) is the probability density of the distance between pairwise codewords equal to \( d \) among all pairwise combination. If a linear block code has single distance \( d \) among all the codeword-pairs, then (17) and (18) are equal to (14) and (16). However, since the distance spectrum is not a delta function for almost any practical code, we assume that the decoding error probability is dominated by minimum distance \( d_{\text{min}} \). Hence, we obtain an erasure threshold \( \gamma_e^* \) in the sense of minimizing the average probability of codeword decision error as in (16).

**B. Optimal Erasure Threshold: \( \gamma_e^* \)**

Since error declaration is purposely to reduce error probability of codeword decision, minimizing (16) is an attractable approach. Although the optimal erasure threshold can be simply obtained from the numerical evaluation of (16), we also approximate (16) to obtain a closed form solution for \( \gamma_e^* \) and investigate the asymptotic property of the derived optimal erasure threshold.

Assuming that the distance between two codewords \( d \) is sufficiently large, Chernoff approximation can be applied to (16). Chernoff approximation of a half sum of binomial series is given by [15]

\[ \sum_{k=0}^{m} X^k Y^{m-k} \leq 2(\sqrt{XY})^{m/2}. \]

Applying (19) to (16), we obtain the approximate error probability

\[ P_{EHDD}(\text{DE}|\gamma_e, d) \leq \left[ P_r(\gamma_e) + 2 \sqrt{P_r(\gamma_e) P_e(\gamma_e)} \right]^d. \]

Note that the erasure threshold \( \gamma_e^* \) obtained by minimizing the upper bound (20) is independent of \( d \). Moreover, it can be proved that Chernoff approximation becomes accurate as the SNR becomes high (see Appendix A). Therefore, the erasure
threshold is not necessarily calculated regardless of \( d \). Based on (16) and (20), the evaluated erasure thresholds \( \gamma_e^* \) and \( \bar{\gamma}_e \) are plotted in Fig. 2, where \( \gamma_e^* \) converges to \( \bar{\gamma}_e \) as \( d \) becomes large and SNR becomes high.

C. Evaluation of \( \bar{\gamma}_e^* \)

Since the evaluation of \( \gamma_e^* = \arg \min_{\gamma_e} \mathbb{E}_{\text{HD}}(\gamma_e) \) does not give a simplified form, \( \bar{\gamma}_e^* \) will be obtained from (20). By differentiating the bound on \( \mathbb{E}_{\text{HD}}(\gamma_e) \) in (20) with respect to \( \gamma_e \), and setting it equal to zero, the optimal threshold \( \bar{\gamma}_e^* \) is found to satisfy the following equality

\[
P_e(\bar{\gamma}_e^*) = \left( \frac{P_e(\gamma_e^*)}{P_e(\bar{\gamma}_e^*)} \right)^2,
\]

where \( P_e(\gamma_e) \) and \( P_e(\bar{\gamma}_e^*) \) are the first derivatives of \( P_e(\cdot) \) and \( P_e(\gamma_e) \), respectively (for details, see derivation in Appendix B). Then, the threshold satisfying (21) optimizes the decoder performance bound. By adopting this threshold in EED algorithms, codeword error rate can be improved.

To indicate the dependency between the erasure threshold and received SNR, \( \gamma_e^* \rightarrow \bar{\gamma}_e^*(d, \eta) \) and \( \bar{\gamma}_e \rightarrow \bar{\gamma}_e(\eta) \) will be used in the following explanation, where \( \eta \) is SNR. As it can be observed from (21), the calculation of \( \bar{\gamma}_e(\eta) \) requires only the cumulative distribution function (cdf) and the probability density function (pdf) of the received signal. Although a unique solution of (21) is not guaranteed for general distributions, we can find the unique solution of (21) for the case of Gaussian distribution (See Appendix C) in the following manner.

Let us assume that the received signal follows the Gaussian distribution, i.e., \( v(i) \sim \mathcal{N}(-\sqrt{E_s/\eta}, \sigma_i^2) \) for \( i = 1, \ldots, n \), where \( E_s \) is the symbol energy and all-zero codeword (AZC) is assumed. Then we can use an approximation of the Q-function for calculating probabilities in (10)

\[
Q(u) \approx \frac{\exp(-u^2/2)}{2\sqrt{\pi u}}.
\]

By inserting probabilities (10) into (21) and exploiting (22), we obtain the optimal erasure threshold to satisfy (21), as follows:

\[
\bar{\gamma}_e^*(\eta) = \frac{\sqrt{E_s}}{\eta} \ln \left( \frac{\sqrt{2}}{\sqrt{3} b_1 + \sqrt{\frac{1}{18} b_2}} \right), \quad (23)
\]

where

\[
\eta = E_s/\sigma_i^2, \quad b_1 = \frac{1}{\sqrt{2}} 9 b_2 + \sqrt{12} + 81 b_2^2 \quad \text{and} \quad b_2 = 2\sqrt{\pi \eta} \exp \left( \frac{2}{3} \right).
\]

Assuming that \( 81 b_2^2 \gg 12 \), (23) can be approximated as

\[
\bar{\gamma}_e^*(\eta) = \frac{\ln 4 \pi \eta}{6 \eta} + 1/6. \quad (24)
\]

We note that the solution of (23) or (24) may not be suitable for low SNRs because the approximation of the Q-function holds for only high SNRs.

In real time applications, fitted lines can be used for the calculation of \( \gamma_e^*(d, \eta) \) when \( d \) and \( \eta \) are given. For this purpose, we derive a function form and fit it to the exact \( \gamma_e^*(d, \eta) \) calculated from (16). Replacing \( \eta \) with \( \eta_{dB} = 10 \log_{10} \eta \) in (24), we obtain the following

\[
\bar{\gamma}_e^*(\eta_{dB}) = C_1(\cdot)\eta_{dB} \exp(-C_2(\cdot)\eta_{dB}) + C_3(\cdot), \quad (25)
\]

where \( C_1(\cdot), C_2(\cdot) \) and \( C_3(\cdot) \) are the fitting parameters. Since (24) is not accurate for low SNRs, we remove the \( \eta_{dB} \) multiplicative term in (25), resulting in:

\[
\bar{\gamma}_e^*(\eta_{dB}) = C_1(\cdot)\eta_{dB} \exp(-C_2(\cdot)\eta_{dB}) + C_3(\cdot). \quad (26)
\]

Performing least square fitting to \( \gamma_e^*(d, \eta_{dB}) \), which is numerically obtained from (16), gives the fitting parameters of (26), as shown in Table I. Figure 2 shows that the fitting result is fairly good.

D. Comparison with Channel Capacity Maximization

As noted in the channel coding theorem [15], we can achieve arbitrarily low BER if code rate is smaller than channel capacity. In the following, an erasure selection criterion will be obtained by maximizing channel capacity.

Assuming that (4) is used for the erasure declaration rule, channel capacity is given by

\[
I(x; v|\gamma_e) = \sum_{i,j} p(v(i)|x(j))p(x(j)) \log_2 \frac{p(v(i)|x(j))}{p(v(i))},
\]

where \( x(j) \) is the transmission symbol, \( v(i) \) is the received signal, and \( p(v(i)) \) is the probability distribution of \( v(i) \). By inserting probabilities in (10) into (27), we can obtain the channel capacity which is parameterized with erasure threshold as following:

\[
I(x; v|\gamma_e) = \frac{1}{1 - P_e(\gamma_e)} \left( P_e(\gamma_e) \log_2 \frac{P_e(\gamma_e)}{P_e(\gamma_e) + P_e(\gamma_e)} + P_e(\gamma_e) \log_2 \frac{P_e(\gamma_e)}{P_e(\gamma_e) + P_e(\gamma_e)} \right) + P_e(\gamma_e) \log_2 \frac{P_e(\gamma_e)}{P_e(\gamma_e) + P_e(\gamma_e)},
\]

\[
(28)
\]

<table>
<thead>
<tr>
<th>( d )</th>
<th>4</th>
<th>8</th>
<th>12</th>
<th>16</th>
<th>32</th>
<th>64</th>
<th>CHF</th>
</tr>
</thead>
<tbody>
<tr>
<td>( C_1 )</td>
<td>0.356</td>
<td>0.413</td>
<td>0.451</td>
<td>0.468</td>
<td>0.479</td>
<td>0.485</td>
<td>0.491</td>
</tr>
<tr>
<td>( C_2 )</td>
<td>0.117</td>
<td>0.125</td>
<td>0.134</td>
<td>0.141</td>
<td>0.143</td>
<td>0.144</td>
<td>0.144</td>
</tr>
<tr>
<td>( C_3 )</td>
<td>0.132</td>
<td>0.136</td>
<td>0.136</td>
<td>0.141</td>
<td>0.142</td>
<td>0.142</td>
<td>0.142</td>
</tr>
</tbody>
</table>

Authorized licensed use limited to: Korea Advanced Institute of Science and Technology. Downloaded on March 4, 2009 at 06:33 from IEEE Xplore. Restrictions apply.
Fig. 3. Signal constellations of 4-ary and 8-ary PAMs: $R^0_i$ denotes a region, where each bit has the same value. $\gamma^e$ is the erasure threshold for each bit position.

To maximize (28), (28) is differentiated with respect to $\gamma_e$ and set to zero. The resultant criterion is given by

$$P_e(\tilde{\gamma}_e(\eta)) = \frac{\log_2 \left( 1 + \frac{P_e(\bar{\gamma}_e(\eta))}{P_e(\gamma_e(\eta))} \right) - 1}{\log_2 \left( 1 + \frac{P_e(\bar{\gamma}_e(\eta))}{P_e(\gamma_e(\eta))} \right) - 1},$$

(29)

where $\tilde{\gamma}_e(\eta)$ is the erasure threshold to maximize the channel capacity.

$\tilde{\gamma}_e(\eta)$ is also shown in Fig. 2. Compared with the proposed results, $\gamma^e_s(d, \eta)$, $\bar{\gamma}^e_s(\eta)$ and $\bar{\gamma}^e_s(d, \eta)$, we can see that $\gamma^e_s(\eta)$ deviates too much to be used for the modulated EED except in the low SNR region. It is due to the fact that channel capacity maximization is not directly related with channel decoding algorithms. Instead, this threshold may be used for channel capacity achieving codes with corresponding erasure decoders, which is not further covered in this paper.

IV. EXTENSION TO M-ARY PAM

Until now, the erasure threshold is analyzed for BPSK modulation. However, N-ary quadrature amplitude modulations (N-ary QAM) are often adopted to achieve higher spectral efficiency for the given transmission bandwidth. For application to QAM signaling, extension to M-ary (M>2) PAM signaling is constructed by following the same procedure with BPSK modulation. Note that N-ary QAM can be represented with two PAM signals, in-phase and quadrature-phase components, as following:

$$s_{N-QAM} = s_{M-PAM}^I + s_{M-PAM}^Q,$$

(30)

where $N = M^2$, $s_{N-QAM}$ is the modulated signal of N-ary QAM, $s_{M-PAM}^I$ and $s_{M-PAM}^Q$ are the in-phase and quadrature-phase signals of M-ary PAM, respectively. Since the in-phase and quadrature-phase signals are independent each other, the erasure selection rule for N-ary QAM signaling can be simplified by considering each phase component separately.

Figure 3 shows signal constellations of the 4-ary PAM and 8-ary PAM. After the Gray coding of the signal constellation, the decision boundary of each bit position shows regularity.

In Fig. 3, the decision boundaries and the corresponding erasure regions for each bit position are also depicted. We note that the $q$th bit position has $2^{q-1}$ decision boundaries and $2^{q-1} + 1$ constellation groups, in which the received signals are quantized into the same value.

Hence, the optimal erasure threshold of each bit position can be calculated by the same procedure conducted for BPSK modulated signals. Let us define a partition set $I_p^q$ of the M-ary PAM constellation as

$$I_p^q = \left\{ \frac{2p-3}{2^q} M + 1, \frac{2p-3}{2^q} M + 2, \ldots, \frac{2p-1}{2^q} M \right\},$$

(31)

where $1 < p < 2^{q-1} + 1$ and $I_p^q$ and $I_{2^{q-1}+1}^q$ are defined as

$$I_p^q = \left\{ 1, 2, \ldots, \frac{M}{2^q} \right\},$$

(32)

$$I_{2^{q-1}+1}^q = \left\{ \frac{2q-1}{2^q} M + 1, 4q - 5 \frac{M}{2^q} + 2, \ldots, M \right\},$$

where $q > 1$. Note that $I_p^q$ is a set of constellations, in which each constellation gives the same bit decision for the $q$th bit of an $m$-tuple bit vector, where $m = \log_2 M$. Using $I_p^q$, we now define the probability of correct decision, erroneous decision and erasure decision as following:

$$P_c^q(\gamma_e^q) = \sum_{j=1}^{2^{q-1}+1} \int_{I_p^q} \sum_{k \in I_p^q} p(y|s_k)P(s_k)dy,$$

(33)

$$P_e^q(\gamma_e^q) = \sum_{j=1}^{2^{q-1}+1} \int_{I_p^q} \sum_{k \notin I_p^q} p(y|s_k)P(s_k)dy,$$

$$P_r^q(\gamma_e^q) = 1 - P_c^q(\gamma_e^q) - P_e^q(\gamma_e^q),$$

where $1 \leq q \leq m$, $\gamma_e^q$ denotes the erasure threshold for the $q$th bit and $R_p^q$ is the region within which the $q$th bit position will be decided as the same value (see Fig. 3). Utilizing the generalized probability definition (33), the erasure threshold of each bit can be calculated by (16) or (20) so that the average probability of codeword decision error is minimized. Instead of using the separate erasure threshold for each bit position, we can define a common erasure threshold $\gamma_e$ to replace $\gamma_e^q$’s for simplicity. In this case, $P_c(\gamma_e)$, $P_e(\gamma_e)$, and $P_r(\gamma_e)$ should be the averaged probabilities of $P_c^q(\gamma_e^q)$, $P_e^q(\gamma_e^q)$, and $P_r^q(\gamma_e^q)$ over $m$ bit positions. Note that common erasure threshold shall be applied to bit decision boundaries independently for each bit position as shown Fig. 3. Figure 4 shows the common erasure thresholds for 4-ary PAM and 8-ary PAM.

V. SIMULATION RESULTS

In order to investigate the effectiveness of the proposed erasure threshold, simulation results are included in terms of 1) the accuracy of the proposed erasure threshold and 2) the codeword error rate (CER) of EED equipped with the proposed erasure threshold. In the following simulations, three kinds of channel codes are used: 1) linear block codes, 2) convolutional codes and 3) low-density parity-check (LDPC) codes. For each channel code, we will first describe the corresponding EED algorithm, then give the simulation results.
Fig. 4. Optimal common erasure thresholds vs. Es/No for 4-ary PAM and 8-ary PAM; Exact threshold is calculated with (16) and Chernoff threshold is from (21) numerically.

Fig. 5. Codeword error rate (CER) variations vs. γc: BCH(15,7,5), BCH(15,5,7), BCH(31,11,11), and BCH(31,6,15) are used for simulations. The CER varies significantly according to γc. Note that the analytical γc∗ are exactly matched to the CER minimum points.

A. Linear Block Codes

Since most derivations in this paper focus on the erasure selection threshold for linear block codes with a single dominant distance d, we show how accurate the proposed erasure threshold is for linear block codes (BCH codes). To avoid decoder dependent performance, the simple search as in (5) and (6) was used to implement EED for linear block codes. Figure 5 shows CER performance according to erasure threshold variation, where the proposed erasure thresholds are marked also. Note that γc∗ does match almost perfectly to the optimal CER points although it is calculated from a bound (20).

B. Convolutional Codes

Convolutional codes are widely used because of its length flexibility and availability of easy and optimal decoding algorithms. The commonly used decoding algorithms are Viterbi algorithm for the maximum likelihood sequence estimator and BCJR algorithm for the maximum a posterior probability decoder [13], [14]. EED implementation for convolutional codes can be realized by simply ignoring the erased bits. Consider Viterbi’s algorithm [16]. Let the branch output be \( b_{q,x} = \{b_{q,x}(0), \ldots, b_{q,x}(n_{f} - 1)\} \) and the observation 

\[ \mathbf{v}_{HD} = \{v_{HD}(0), \ldots, v_{HD}(n_{f} - 1)\} \]

at the \( i \)th transition, where \( n_{f} \) is the number of generator functions. Then branch weight \( w_{x}^{i+1} \) of state \( x \) at the \((i+1)\)th state is given by:

\[
w_{x}^{i+1} = \min \left\{ w_{p}^{i} + w_{b}^{i}(p, x), w_{q}^{i} + w_{b}^{i}(q, x) \right\},
\]

where \( e^{i} = \{e^{i}(0), \ldots, e^{i}(n_{f} - 1)\} \) is the erasure indicator for each branch at the \( i \)th stage, and \( w_{p}^{i} \) (\( w_{q}^{i} \)) is the branch metric of state \( p \) (\( q \)) at the \( i \)th stage. Using (34), the EED can be implemented for the convolutional codes.

For the erasure threshold calculation of rate-1/2 convolution codes, the generator polynomials \((q_{1} = 15, q_{2} = 17), (g_{1} = 53, g_{2} = 75)\), and \((g_{1} = 133, g_{2} = 171)\) are used for \( d_{\text{min}} = 6, d_{\text{min}} = 8 \), and \( d_{\text{min}} = 10 \) respectively, where \( d_{\text{min}} \) is the shortest distance between codewords [16]. The data sequence is of length 512 and padded with zeros to terminate the encoder state at zero.

Figure 6 shows the optimal erasure thresholds found from simulations, which are compared with analytical results \( \gamma^{*}_{c}(d, \eta) \) and \( \gamma^{*}_{c}(\eta) \). Although the distance spectrum of convolutional codes is not a delta function, the decoding performance is dominated also by \( d_{\text{min}} \) [16]. Therefore, the simulated threshold shows close proximity to the \( d = d_{\text{min}} \) lines of the analyzed threshold \( \gamma^{*}_{c}(d, \eta) \).

The CER variation according to the erasure thresholds is shown in Fig. 7, where the convolutional code with \((g_{1} = 133, g_{2} = 171)\) is used. As is seen in Fig. 7, the analytical threshold \( \gamma^{*}_{c} \) from (16) is well-matched to the minimum points of the CER plots. In addition, the Chernoff-approximated threshold \( \gamma^{*}_{c}(\eta) \) also shows little deviation from the minimum points, which makes us relieve to use it as the erasure threshold. We also note that if the minimum distance \( d_{\text{min}} \) is larger than 10, then the performance difference becomes negligible between \( \gamma^{*}_{c}(d, \eta) \) and \( \gamma^{*}_{c}(\eta) \).
C. Low-Density Parity-Check (LDPC) Codes

LDPC codes are one type of linear block codes [1], [3] which provide a simple iterative decoding structure. The typical decoder of an LDPC code is a message mapping decoder (a.k.a. sum-product decoder or iterative belief propagation decoder). The EED of the message mapping decoding for an $N_c \times N_v$ parity check matrix can be implemented by defining the $i$th edge messages as following [2]:

$$\Psi^t_{b_i}(b_i, s_j) = \text{sgn} \left( f_0(b_i) + \sum_{z=1, z \neq s_j}^{N_c} \Psi^{t-1}_{b_i}(z, b_i) \right),$$

$$\Psi^t_{s_j}(s_j, b_i) = \prod_{z=1, z \neq b_i}^{N_c} \Psi^{t-1}_{b_i}(z, s_j),$$

(35)

where $b_i$ is the $i$th variable node, $s_j$ is the $j$th check node, $f_0(j) = 2 v_{HD}(j) - 1$, $\Psi^0_{b_i}(b_i, s_j)$ is the message along edge $\vec{e} = (b_i, s_j)$ at the $t$th iteration, and $\Psi^t_{s_j}(s_j, b_i)$ is the message along edge $\vec{e} = (s_j, b_i)$ at the $t$th iteration.

With EED for the LDPC codes as described in (35), the progressive-edge-growth (PEG) LDPC codes with column weight 3 are considered [4] and AWGN channel was assumed. Because the distance spectrum of LDPC codes is not exactly calculable and the decoding performance generally increases according to the parity check matrix (PCM) size, we performed simulations under the various PEG PCM sizes by using the finite alphabet (FA) message passing decoder (MPD) for the EED [2]. Figure 8 illustrates the optimal erasure thresholds $\bar{\gamma}_e (PCM, \eta)$, which are numerically searched with FA-MPD. The optimal erasure thresholds of various PCMs show close proximity to the line of $\gamma^*_{\bar{\eta}} (d = 4, \eta)$. Note that $\gamma^*_{\bar{\eta}} (PCM, \eta)$ with larger PCM is closer to $\bar{\gamma}_e (\eta)$ of the Chernoff-approximated bound, which means that the performance of the LDPC codes approaches to that of the ideal channel coding.

Figure 9 illustrates the effect of the erasure threshold $\bar{\gamma}_e$ on the CER, where the above FA-MPD is also used as EED.

We considered two setups: 1) $64 \times 512$ PCM with code rate 0.875 is selected to compare the CER under various SNRs and 2) the different size PCMs ($100 \times 200, 200 \times 400, 300 \times 600$ and $400 \times 800$) with the same code rate 0.5 are used under the same SNR (= 4 dB). The vertical dotted lines in Fig. 9 are corresponding to $\bar{\gamma}^*_e (d = 8, \eta)$, $\bar{\gamma}^*_e (d = 16, \eta)$ and $\bar{\gamma}^*_e (\eta)$ (denoted as CHF). Note that the performance difference is quite small. This figure shows that the proposed erasure selection scheme is also applicable to the LDPC codes.

D. Extension to M-ary PAM Case

For the erasure threshold simulation with M-ary PAM, we used a convolutional code with $(g_1 = 133, g_2 = 171)$ of $d_{\text{min}} = 10$. After a data packet is encoded with the convolutional code, code bits are grouped into $m$ bits to form M-ary PAM signal. At the reception of the M-ary PAM signal,
the received signal is hard-decided into binary sequences, which are re-arranged to obtain the received sequence. During hard-decision, a common erasure threshold is used to declare erased bits as described in Section IV. $P_c(\gamma_e)$, $P_r(\gamma_e)$ and $P_r(\gamma_e)$ are averaged over all $1 \leq q \leq m$, where $m = 2$ for 4-ary PAM and $m = 3$ for 8-ary PAM. We used the Viterbi decoding algorithm (as described in Subsection V-B) to decode the received sequence with erasure declaration.

Figure 10 shows the CER variation according to $\gamma_e$, where 4-ary PAM signaling is used. It is shown that the analytical threshold is well matched to the CER minimum points. In addition, the CER variation of 8-ary PAM signaling is plotted in Fig. 10, where the analyzed erasure threshold is also at the minimum CER points.

VI. CONCLUSION

In this paper, we showed that there exists an optimal erasure threshold which minimizes the decoding error probability. We also proposed how to calculate the exact optimal erasure threshold, which are compared with that obtained from the channel capacity maximization. Erasure threshold calculation for M-ary PAM was obtained with a simple extension to erasure selection for the BPSK modulation. To demonstrate the effectiveness of the proposed erasure threshold, simulation results for linear block codes, convolutional codes and LDPC codes are included.

APPENDIX

A. Approximation of Tail Probability

Here, we prove that the tail probability of the binomial distribution can be approximated with the Chernoff approximation up to a constant factor for high SNR and large $d$. Let the tail probability $S_n(k)$ be as following:

$$S_n(k) = \sum_{j=k}^{n} \binom{n}{j} P_c(\gamma_e)^j P_r(\gamma_e)^{n-j}. \quad (A1)$$

Then we can approximate $S_n(k)$ as following [22]:

$$b_n(k) < S_n(k) < \frac{b_n(k)}{1 - \rho}, \quad (A2)$$

where $b_n(k) = \binom{n}{k} P_c(\gamma_e)^k P_r(\gamma_e)^{n-k}$ and $\rho = \frac{b_n(k+1)}{b_n(k)} = \frac{(n-k)P_r(\gamma_e)}{(k+1)P_c(\gamma_e)}$. We note that $\rho$ approaches to zero for high SNR and large $d$. Therefore we can assume $S_n(k) \approx b_n(k)$. Here we can rewrite $S_n(k)$ in simpler form using Stirling’s approximation to the factorial:

$$(2\pi)^{1/2} n^{n+1/2} \exp(-n + \frac{1}{12n + 1}) < n! \leq (2\pi)^{1/2} n^{n+1/2} \exp(-n + \frac{1}{12n}). \quad (A3)$$

Then $b_n(k)$ is given as

$$C(n, \lambda, \rho, \gamma_e) \exp\left(\frac{1}{12n} - \frac{1}{12\lambda n + 1} - \frac{1}{12\rho n + 1}\right) \leq b_n(k),$$

where $C(n, \lambda, \rho, \gamma_e) = \frac{1}{(2\pi n/\lambda)^{1/2}} 2^{n(H(\lambda, \rho) + E(\lambda, P_r(\gamma_e)))}, \lambda = k/n, \rho = 1 - \lambda, H(\lambda, \rho) = -\lambda \log_2 \lambda - \rho \log_2 \rho, \rho = 1 - \lambda, H(\lambda, \rho) = -\lambda \log_2 \lambda - \rho \log_2 \rho$. As we are concerned only about the error probability, let $k = n/2$. Then (A4) can be simplified as

$$\frac{1}{(\pi n/2)^{1/2}} \sqrt{4P_e(\gamma_e)P_c(\gamma_e)^n} \exp\left(\frac{1}{12n} - \frac{2}{6n + 1}\right) \leq b_n\left(\frac{n}{2}\right) \leq \frac{1}{(\pi n/2)^{1/2}} \sqrt{4P_e(\gamma_e)P_c(\gamma_e)^n} \exp\left(\frac{1}{12n} - \frac{1}{3n}\right). \quad (A5)$$

Moreover, since Stirling’s approximation has a tight bound, we take the average of the lower and the upper bound in (A5) to approximate $b_n\left(\frac{n}{2}\right)$:

$$b_n\left(\frac{n}{2}\right) = C_1(n) C_2(n) / 4P_e^2(\gamma_e)P_c(\gamma_e)^n, \quad (A6)$$

where $C_1(n) = \frac{\gamma_e}{(\pi n/2)^{1/2}}$ and $C_2(n) = \exp\left(\frac{\gamma_e}{\gamma_e} - \frac{\gamma_e}{\gamma_e + \gamma_e}\right) + \exp\left(\frac{\gamma_e}{\gamma_e + \gamma_e} - \frac{\gamma_e}{\gamma_e}\right)$. In (A6), we note that $C_1(n)$ and $C_2(n)$ depend only on the distance $n = d$. Therefore, the Chernoff approximation becomes accurate up to a constant factor to the tail probability of the binomial distribution.

B. Derivation of (21)

Let $f(\gamma_e) = P_c(\gamma_e) + 2\sqrt{P_c(\gamma_e)P_r(\gamma_e)}$. Using $P_c(\gamma_e) + P_r(\gamma_e) + P_r(\gamma_e) = 1$, the first derivative of $f(\gamma_e)$ is given by

$$f'(\gamma_e) = -P_c'(\gamma_e) - P_r'(\gamma_e) + \frac{P_c(\gamma_e)P_r(\gamma_e) + P_r(\gamma_e)P_c(\gamma_e)}{\sqrt{P_c(\gamma_e)P_r(\gamma_e)}}$$

$$= P_c'(\gamma_e) \left( \frac{P_c(\gamma_e)}{P_r(\gamma_e)} - 1 \right) + P_r'(\gamma_e) \left( \frac{P_r(\gamma_e)}{P_c(\gamma_e)} - 1 \right)$$

$$= P_c'(\gamma_e) (P_c(\gamma_e))^2 - (P_c(\gamma_e) + P_r(\gamma_e))(P_r(\gamma_e)) + P_c'(\gamma_e), \quad (A7)$$

$$(P_c(\gamma_e))^2 - (P_c(\gamma_e) + P_r(\gamma_e))(P_r(\gamma_e)) + P_c'(\gamma_e).$$
where $t = \sqrt{\frac{P_c(\gamma_e)}{P_e(\gamma_e)}}$. Setting $f'(\gamma_e) = 0$ and solving the second order equation, we have two solutions:

$$t = 1 \quad \text{or} \quad t = \frac{P'_e(\gamma_e)}{P'_c(\gamma_e)}.$$  \hspace{1cm} (A8)

Using the definition of $t$, the two solutions become

$$P_c(\gamma_e) = P_e(\gamma_e) \quad \text{or} \quad \sqrt{\frac{P_c(\gamma_e)}{P_e(\gamma_e)}} = \frac{P'_c(\gamma_e)}{P'_e(\gamma_e)}.$$  \hspace{1cm} (A9)

Note that the first solution is impossible if the signal power exists. The second solution is (21).

C. Unique $\gamma_e$ of (21)

For Gaussian distribution, the unique solution of (21) can be guaranteed. By inserting the cumulative distribution and density function to (21), the following equation can be easily obtained:

$$Q\left(\gamma_e + \frac{\sqrt{E_s}}{N_0}\right) = \exp\left(-\frac{4\gamma_e\sqrt{E_s}}{N_0}\right) Q\left(\frac{\gamma_e - \sqrt{E_s}}{N_0}\right).$$  \hspace{1cm} (A10)

Evaluating the above equation at $\gamma_e = 0$, $Q\left(\frac{\sqrt{E_s}}{N_0}\right) < Q\left(-\frac{\sqrt{E_s}}{N_0}\right)$ holds if and only if $\sqrt{E_s} > 0$, i.e., signal power exists. On the other hand, we note that the decreasing rate of $Q\left(\frac{\gamma_e + \sqrt{E_s}}{N_0}\right)$ is slower than that of $\exp\left(-\frac{4\gamma_e\sqrt{E_s}}{N_0}\right) Q\left(\frac{\gamma_e - \sqrt{E_s}}{N_0}\right)$. With these two facts, therefore, we can conclude that there will be a unique cross-point that satisfies (A10).

REFERENCES


Yeong-Hyeon Kwon (M’06) received B.S., M.S. and Ph. D. degrees in Electrical and Electronic Engineering from KAIST, Korea in 1997, 1999 and 2005, respectively. He is a senior engineer in LG Electronics since 2005. His research interests include MIMO/OFDM, synchronization, channel coding, UWB, macro diversity transmission and relayed transmission.

Mi-Kyung Oh (M’07) received the B.S. degree from Chung-Ang University, Seoul, Korea, in 2000 and the M.S. and Ph.D. degrees in electrical engineering from the Korea Advanced Institute of Science and Technology, Daejeon, Korea, in 2002 and 2006, respectively. From September 2002 to February 2004, she was a Visiting Researcher with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis. She is currently a Senior Engineer with the Electronics and Telecommunications Research Institute, Daejeon. Her research interests include channel estimation, channel synchronization, UWB communication systems, and ranging technology.

Dong-Jo Park (M’84) received his B.S. degree from the Department of Electrical Engineering, Seoul National University, Seoul, Korea in 1976 and M.S. and Ph.D. degrees in Electrical Engineering from the University of California, Los Angeles, USA, in 1981 and 1984, respectively. He was a senior researcher of Electronics and Telecommunications Research Institute (ETRI), Daejeon, Korea, from 1984 to 1985. Since 1985, He has been with the School of Electrical Engineering and Computer Science in Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Korea. His research interests include MIMO, OFDM, UWB and cooperative communication systems and wireless multimedia signal processing.