On APP Decoding

Carlos R.P. Hartmann  
*Syracuse University*, chartman@syr.edu

Luther D. Rudolph  
*Syracuse University*

Kishan Mehrotra  
*Syracuse University*, mehrotra@syr.edu

Guy J. Snedeker  
*Syracuse University*

Follow this and additional works at: [https://surface.syr.edu/eecs_techreports](https://surface.syr.edu/eecs_techreports)

Part of the [Computer Sciences Commons](https://surface.syr.edu/eecs_techreports)

Recommended Citation  
[https://surface.syr.edu/eecs_techreports/38](https://surface.syr.edu/eecs_techreports/38)

This Report is brought to you for free and open access by the College of Engineering and Computer Science at SURFACE. It has been accepted for inclusion in Electrical Engineering and Computer Science Technical Reports by an authorized administrator of SURFACE. For more information, please contact surface@syr.edu.
ON APP DECODING

Carlos R.P. Hartmann
Luther D. Rudolph
Kishan G. Mehrotra
Guy J. Snedeker
June 81
On APP Decoding+

Carlos R.P. Hartmann*

Luther D. Rudolph*

Kishan G. Mehrotra**

Guy J. Snedeker**

+This work was partially supported by the National Science Foundation under Grant ECS 80-19484.

*Communication Studies Laboratory, School of Computer and Information Science, Syracuse University, Syracuse, New York, 13210, (315) 423-2368.

**School of Computer and Information Science, Syracuse University, Syracuse, New York, 13210, (315) 423-2368
Abstract

In this paper we show that APP decoding for a linear code $C$ is optimum not for $C$, but for a minimum-distance-2 code $\tilde{C}$ which contains $C$ as a subcode when the codewords of $\tilde{C}$ are transmitted with equal probability. However, APP decoding is shown to be asymptotically optimum for $C$ for high SNR when $C$ is a binary one-step orthogonalizable code with equiprobable codewords transmitted over the AWGN channel.
1. Introduction

In recent years there has been a growing interest in soft decision decoding schemes for error-correcting codes. The intent is to avoid, in part or in whole, the degradation of communication system performance which results when symbol-by-symbol "hard decision" quantization precedes decoding. In most cases the additional information provided by using soft decisions is worth about 2 dB of additional coding gain.

Among the soft decision decoding techniques which seek to minimize the probability of symbol error, Massey's APP (a posteriori probability) decoding [1] is one of the most important. In this paper we show that APP decoding for a linear code $C$ is optimum not for $C$, but for a minimum-distance-2 code $\tilde{C}$ which contains $C$ as a subcode when the codewords of $\tilde{C}$ are transmitted with equal probability. However, APP decoding is shown to be asymptotically optimum for $C$ for high SNR when $C$ is a binary one-step orthogonalizable code with equiprobable codewords transmitted over the AWGN channel.
2. **Main Result**

In this section we first show that the complexity of the optimum error symbol decoding rule is independent of the number of parity checks used and is prohibitive for almost any code. Based on this optimum decoding rule we then define a new error symbol decoding rule which is analogous to maximum-likelihood word decoding. We then show that for a particular case this new decoding rule becomes APP decoding.

Let H be the parity check matrix for an (n,k,d) linear code C over GF(p), p a prime. Also let B be a Jxn matrix whose rows correspond to J parity checks of C. A codeword \( \zeta = (c_0, c_1, \ldots, c_{n-1}) \) is transmitted over a time-discrete memoryless channel with output alphabet \( \Delta \). The received word is denoted by \( \xi = (r_0, r_1, \ldots, r_{n-1}) \), \( r_j \in \Delta \) and its quantized version by \( \tilde{\xi} = (\tilde{r}_0, \tilde{r}_1, \ldots, \tilde{r}_{n-1}) \), \( \tilde{r}_j \in GF(p) \). Define \( \xi = \tilde{r}B^T \), and let \( \oplus \) denote addition over GF(p) and \( \ominus \) x the additive inverse of x in GF(p). Now let \( \chi = (v_0, v_1, \ldots, v_{n-1}) \) and \( \psi = (w_0, w_1, \ldots, w_{n-1}) \) be arbitrary vector of \( V_n \), the n-dimentional vector space over GF(p). Finally, let \( \varepsilon = (e_0, e_1, \ldots, e_{n-1}) \), \( e_j \in GF(p) \), be the error vector, that is, \( \tilde{\xi} = \zeta \oplus \varepsilon \).

The decoding problem is: given that a codeword from C was transmitted, and given \( \xi \) and \( B \), compute an estimate \( \hat{e}_m \) of the \( m^{th} \) error symbol \( e_m \) in such a way that the probability that \( \hat{e}_m \) equals \( e_m \) is maximized. Now we state the optimum error symbol decoding rule.

**Decoding Rule 1:** Set \( \hat{e}_m = \tilde{r}_m \ominus a \), where \( a \in GF(p) \) maximizes

\[
\Pr(e_m = \tilde{r}_m \ominus a | \xi) = \sum_{\chi \in E(a)} \Pr(\chi | \xi),
\]

(1)
where

\[ E(a) = \{ y | v_m = \bar{r}_m \ominus a, yB^T = \xi \} . \]

We now show that this decoding rule is equivalent to the optimum symbol-by-symbol decoding rule, which is: Set \( \hat{c}_m = a \), where \( a \in GF(p) \) maximizes

\[ \Pr(c_m = a | \xi) = \sum_{\omega \in G(a)} \Pr(\omega | \xi), \quad (2) \]

where \( \hat{c}_m \) is the estimate of \( c_m \) and

\[ G(a) = \{ \omega | w_m = a, \omega H^T = 0 \}. \]

We demonstrate this equivalence by showing that (1) and (2) are equal. Define \( \eta = \bar{\xi} - \omega \) for \( \omega \in V_n \) and

\[ U(a) = \{ \eta | \bar{\eta} = \bar{\xi} \ominus \omega, \omega \in C(a) \}, \]

where

\[ C(a) = \{ \omega | w_m = a, \omega B^T = \xi \}. \]

We now prove two technical lemmas.

**Lemma 1:** \( \sum_{\omega \in C(a)} \Pr(\omega | \xi) = \sum_{\eta \in U(a)} \Pr(\eta | \xi). \)

**Proof:** Since \( \bar{\xi} \) is determined by knowing \( \xi \), we may write

\[ \Pr(\bar{\xi} \ominus \omega | \xi) = \Pr(\omega | \xi). \]

Since \( \bar{\eta} = \bar{\xi} \ominus \omega \), \( \Pr(\bar{\eta} | \xi) = \Pr(\omega | \xi) \). Now, since for each element \( \omega \in C(a) \) there is a corresponding element...
we may conclude that

\[ \sum_{\bar{\omega} \in C(a)} \Pr(\bar{\omega} | \bar{x}) = \sum_{\bar{\nu} \in U(a)} \Pr(\bar{\nu} | \bar{x}) . \]

Q.E.D.

**Lemma 2:** \( \sum_{\bar{\nu} \in U(a)} \Pr(\bar{\nu} | \bar{x}) = \sum_{\bar{\omega} \in E(a)} \Pr(\bar{\omega} | \bar{x}) . \)

**Proof:** We prove Lemma 2 by proving that \( U(a) = E(a) \). We first assume
that \( \bar{\nu} \in U(a) \), so that \( \bar{\nu} = \bar{\omega} \ominus \bar{\nu} \), where \( \bar{\nu} \in C(a) \). Thus \( \bar{\omega} = a \) and \( \bar{\omega}B^T = \bar{\omega}B^T = \emptyset \). Then \( \bar{\nu} \in E(a) \). Now we assume that \( \bar{\nu} \in E(a) \). Then \( z_m = \bar{r}_m \ominus a \) and \( zB^T = \emptyset \). Define \( \bar{w} = \bar{r} \ominus \bar{z} \). Then \( w_m = a \) and \( wB^T = \emptyset \). So \( \bar{w} \in C(a) \) and \( \bar{z} \in U(a) \).

Q.E.D.

As a consequence of Lemmas 1 and 2 we have the following theorem:

**Theorem 1:** \( \sum_{\bar{\nu} \in E(a)} \Pr(\bar{\nu} | \bar{x}) = \sum_{\bar{\omega} \in C(a)} \Pr(\bar{\omega} | \bar{x}) . \) (3)

Now (3) may be written as

\[ \sum_{\bar{\nu} \in E(a)} \Pr(\bar{\nu} | \bar{x}) = \sum_{\bar{\omega} \in C(a)} \Pr(\bar{\omega} | \bar{x}) \Pr(\bar{\omega} | \bar{\nu}) / \Pr(\bar{\nu}) . \] (4)

Since only codewords of \( C \) are transmitted, we may write (4) as

\[ \sum_{\bar{\nu} \in E(a)} \Pr(\bar{\nu} | \bar{x}) = \sum_{\bar{\omega} \in G(a)} \Pr(\bar{\omega} | \bar{x}) . \]

So we may conclude that for any given matrix \( B \) we have

\[ \Pr(e_m = \bar{r}_m \ominus a | \bar{x}) = \sum_{\bar{\omega} \in G(a)} \Pr(\bar{\omega} | \bar{x}) . \]

An alternative way to see this result is to observe that
Pr(y|x) = Pr(w = x ∩ y|x) = 0 whenever w = x ∩ y / C. This also shows that the complexity of calculating Pr(e_m = r_m ∩ a|x) is O(min(p^k, p^{n-k})) [2] and it is independent of the given matrix B. This complexity of decoding is prohibitive for almost any code. We now modify Decoding Rule 1 in order to decrease its complexity. From the proof of Lemmas 1 and 2 we know that we may write Pr(y|x) = P(w = x ∩ y|x) where y ∈ E(a) and w = x ∩ y ∈ C(a). Since the channel is memoryless we have

\[
Pr(y|x) = \prod_{i=0}^{n-1} Pr(r_i|w_i)(Pr(w)/Pr(x))
\]

where w = x ∩ y. We note that Pr(w_i|r_i) = Pr(r_i ∩ w_i|r_i) = Pr(v_i|r_i).

Thus (1) may be written as:

\[
Pr(e_m = r_m ∩ a|x) = \sum_{y ∈ E(a)} \left( \prod_{i=0}^{n-1} Pr(v_i|r_i)/Pr(r_i)Pr(r_i)/Pr(x) \right)(Pr(x)/Pr(\bar{x} ∩ y)/Pr(y))
\]

or

\[
Pr(e_m = r_m ∩ a|x) = R(x) \sum_{y ∈ E(a)} \left( \prod_{i=0}^{n-1} Pr(v_i|r_i)/Pr(r_i)Pr(r_i)/Pr(x) \right)Pr(x ∩ y)
\]  \hspace{1cm} (5)

where

\[
R(x) = \prod_{j=0}^{n-1} Pr(r_j)/Pr(x),
\]

and if the codewords of C are equiprobable and B = H, (5) may be written as

-6-
\[
\Pr(e_m = \bar{r}_m \land a | x) = R(x) p^{n-k} \sum_{\gamma \in E(a)} \prod_{i=0}^{n-1} \Pr(v_i | r_i).
\]  \hspace{1cm} (6)

Based on (6) we define the following decoding rule:

**Decoding Rule 2:** Set \( \hat{e}_m = \bar{r}_m \land a \), where \( a \in GF(p) \) maximizes

\[
\sum_{\gamma \in E(a)} \prod_{i=0}^{n-1} \Pr(v_i | r_i).
\]  \hspace{1cm} (7)

The reader should note the analogy between Decoding Rule 2 and maximum-likelihood word decoding. Using arguments similar to those used in the proof of Theorem 1 we can prove the following:

**Theorem 2:**

\[
\sum_{\gamma \in E(a)} \prod_{i=0}^{n-1} \Pr(v_i | r_i) = \sum_{\omega \in C(a)} \prod_{i=0}^{n-1} \Pr(w_i | r_i).
\]  \hspace{1cm} (8)

Now let \( \tilde{C} \) be the linear \((n,k,d)\) code over \( GF(p) \) defined by the matrix \( B \), that is, \( \omega \in \tilde{C} \) if and only if \( \omega B^T = \bar{q} \). Theorem 2 shows that the performance of a decoder which implements Decoding Rule 2 is equal to the performance of a decoder which implements the optimum symbol-by-symbol decoding rule for \( \tilde{C} \) whenever the codewords of \( \tilde{C} \) are transmitted with equal probability. We note that \( C \) is a subcode of \( \tilde{C} \) and thus \( \tilde{d} \leq d \) and \( k \leq \tilde{k} \). Since the complexity of calculating (7) is \( O(\min(p^k, p^{n-k}) \) and \( \tilde{d} \leq \bar{d} \), it appears that Decoding Rule 2 may be more complex than, and yields an inferior performance to, Decoding Rule 1. However, we will show that when the parity check corresponding to the rows of \( B \) are orthogonal on the \( m \)th position, Decoding Rule 2 is the same as
Massey's original APP decoding rule (APPDR) which has complexity $O(Jn)$ [1]. To show this we let $Q_j = \{ j_1, j_2, \ldots, j_F \}, 1 \leq j \leq J$, be the set of positions checked by the $j$th row of $B$, excluding the $m$th position. Since the parity checks corresponding to the rows of $B$ are orthogonal in the $m$th position, we know that $Q_j \cap Q_i = \emptyset$ for $i \neq j, 1 \leq i, j \leq J$. Now let

$$S_j = \sum_{i=1}^{F_j} b_{j_i} r_{j_i} \oplus r_m, \quad b_{j_i} \in GF(p),$$

and

$$M_j = \{ x = (x_{j_1}, x_{j_2}, \ldots, x_{j_F}) | x_{j_i} \in GF(p), \sum_{i=1}^{F_j} b_{j_i} x_{j_i} = S_j \oplus r_m \oplus a \}.$$

We now define APPDR in terms of our formulation.

**APPDR:** Set $\hat{\gamma}_m = r_m \ominus a$, where $a \in GF(p)$ maximizes

$$Pr(x_m = r_m \ominus a | r_m) = \prod_{j=1}^{F_j} \sum_{x_j \in M_j} \frac{\prod_{i=1}^{F_j} Pr(x_{j_i} | r_{j_i})}{\prod_{x \in M_j} \prod_{i=1}^{F_j} Pr(x_{j_i} | r_{j_i})}. \quad (9)$$

Since $Q_j \cap Q_i = \emptyset$ for $i \neq j$, it is easily seen that (7) and (9) are equal.

When the rows of $B$ correspond to parity checks orthogonal on the $m$th position, then $d = 2$. Thus the performance of an APP decoder is the same as the performance of the optimum symbol-by-symbol decoder for $\tilde{C}$, with $\tilde{d} = 2$, whenever the codewords of $\tilde{C}$ are equiprobable. However, when a binary code $C$ is transmitted over the AWGN channel and $J = d - 1$, the APP decoder is asymptotically optimum for high SNR. This is obtained as follows: from (8) and [3] we may conclude that $P_{APP}(SNR \to \infty),$
the asymptotic probability of error of the $m^{th}$ bit for APPDR, is given by

$$P_{APP}(\text{SNR} \to \infty) \approx N(\Theta_m) Q(\sqrt{2R}\Theta_m \gamma_b)$$

where $\Theta_m$ is the minimum Hamming weight of vectors in $C(1)$, $N(\Theta_m)$ the number of codewords of weight $\Theta_m$ in $C(1)$, $R=k/n$, $\gamma_b = E_b/N_0$ the SNR per transmitted digit of information and

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty \exp(-x^2/2) \, dx.$$ 

It is easily seen that $\Theta_m = J+1$. Thus for $J=d-1$, $\Theta_m = d$ and

$$P_{APP}(\text{SNR} \to \infty) \approx N(d) Q(\sqrt{2Rd}\gamma_b)$$

which is proportional to the asymptotic behavior of the probability of error of the $m^{th}$ bit for the optimum symbol-by-symbol decoding rule for $C$ whenever its codewords are transmitted over the AWGN channel with equal probability. [3]
3. Simulation Results

In this section we present performance curves for the (21,11) binary projective geometry code transmitted over the AWGN channel for APP decoding and maximum-radius decoding [4], both using five orthogonal parity checks, their iterative extensions [4], majority logic decoding and optimum symbol-by-symbol decoding. The APP decoder and the maximum radius decoder with the demodulation function

\[ p(x) = \begin{cases} 
1, & x \leq 0 \\
\cos(\pi x), & 0 < x < 1 \\
-1, & 1 \leq x
\end{cases} \quad [4] \]

are asymptotically optimum for high SNR. Figures 1 and 2 show the bit-error-rate and the word-error-rate performance respectively. The dotted curve in Figure 2 shows the asymptotic behavior of optimum decoding [5].

The APP decoding rule is a function of the channel SNR. However, as Figures 1 and 2 show for the (21,11) PG code transmitted over the AWGN channel, maximum radius decoding, which is fixed and thus independent of the channel SNR, gives almost the same performance as APP decoding when both decoders use the same five orthogonal parity checks. This makes maximum-radius decoding very attractive for practical applications.
References


Figure 1:
Bit Error Rate of the
(21,11) P.G. Code over the AWGN Channel
(Antipodal Signalling)

- DIGITAL MAJORITY DECODING
- UNCODED DATA
- MAXIMUM-RADIUS DECODING (MRD)
  (5 ITERATIONS)
- APP DECODING (APPD)
  (2 ITERATIONS)
- OPTIMUM BIT-BY-BIT DECODING [6]

$P_{bit}$ - Bit Error Rate

$\gamma_b$ in dB
Figure 2:
Word Error Rate of the
(21,11) P.G. Code over the AWGN Channel
(Antipodal Signalling)