6-1981

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

ON APP DECODING

Carlos R.P. Hartmann

Luther D. Rudolph

Kishan G. Mehrotra

Guy J. Snedeker

June 81

SCHOOL OF COMPUTER
AND INFORMATION SCIENCE
SYRACUSE UNIVERSITY

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 $\bar{C}$ which contains C as a subcode when the codewords of $\bar{C}$ 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 orthogonolizable 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 $\bar{C}$ which contains C as a subcode when the codewords of $\bar{C}$ 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 orthogonolizable 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 prohibitve 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 J×n matrix whose rows correspond to J parity checks of C. A codeword $\underset{\sim}{c} = (c_0, c_1, \ldots, c_{n-1})$ is transmitted over a time-discrete memoryless channel with output alphabet Δ. The received word is denoted by $\underset{\sim}{r} = (r_0, r_1, \ldots, r_{n-1})$, $r_j \in \Delta$ and its quantized version by $\underset{\sim}{\bar{r}} = (\bar{r}_0, \bar{r}_1, \ldots, \bar{r}_{n-1})$, $\bar{r}_j \in GF(p)$. Define $\underset{\sim}{S} = \underset{\sim}{\bar{r}}B^T$, and let $\oplus$ denote addition over GF(p) and $\ominus x$ the additive inverse of x in GF(p). Now let $\underset{\sim}{v} = (v_0, v_1, \ldots, v_{n-1})$ and $\underset{\sim}{w} = (w_0, w_1, \ldots, w_{n-1})$ be arbitrary vector of V_n, the n-dimentional vector space over GF(p). Finally, let $\underset{\sim}{e} = (e_0, e_1, \ldots, e_{n-1})$, $e_j \in GF(p)$, be the error vector, that is, $\underset{\sim}{\bar{r}} = \underset{\sim}{c} \oplus \underset{\sim}{e}$.

The decoding problem is: given that a codeword from C was transmitted, and given $\underset{\sim}{r}$ 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 $\hat{e}_m = \bar{r}_m \ominus a$, where $a \in GF(p)$ maximizes

$$\Pr(e_m = \bar{r}_m \ominus a | \underset{\sim}{r}) = \sum_{\underset{\sim}{v} \in E(a)} \Pr(\underset{\sim}{v} | \underset{\sim}{r}), \qquad (1)$$

where

$$E(a) = \{\underset{\sim}{v} | v_m = \bar{r}_m \ominus a,\ \underset{\sim}{v}B^T = \underset{\sim}{S}\} .$$

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

$$\Pr(c_m = a | \underset{\sim}{r}) = \sum_{\underset{\sim}{w} \in G(a)} \Pr(\underset{\sim}{w} | \underset{\sim}{r}), \tag{2}$$

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

$$G(a) = \{\underset{\sim}{w} | w_m = a,\ \underset{\sim}{w}H^T = \underset{\sim}{0}\}.$$

We demonstrate this equivalence by showing that (1) and (2) are equal. Define $\underset{\sim}{u} = \bar{\underset{\sim}{r}} - \underset{\sim}{w}$ for $\underset{\sim}{w} \in V_n$ and

$$U(a) = \{\underset{\sim}{u} | \underset{\sim}{u} = \bar{\underset{\sim}{r}} \ominus \underset{\sim}{w},\ \underset{\sim}{w} \in C(a)\},$$

where

$$C(a) = \{\underset{\sim}{w} | w_m = a,\ \underset{\sim}{w}B^T = \underset{\sim}{0}\}.$$

We now prove two technical lemmas.

Lemma 1: $\sum_{\underset{\sim}{w} \in C(a)} \Pr(\underset{\sim}{w} | \underset{\sim}{r}) = \sum_{\underset{\sim}{u} \in U(a)} \Pr(\underset{\sim}{u} | \underset{\sim}{r})$.

Proof: Since $\bar{\underset{\sim}{r}}$ is determined by knowing $\underset{\sim}{r}$, we may write $\Pr(\bar{\underset{\sim}{r}} \ominus \underset{\sim}{w} | \underset{\sim}{r}) = \Pr(\underset{\sim}{w} | \underset{\sim}{r})$. Since $\underset{\sim}{u} = \bar{\underset{\sim}{r}} \ominus \underset{\sim}{w}$, $\Pr(\underset{\sim}{u} | \underset{\sim}{r}) = \Pr(\underset{\sim}{w} | \underset{\sim}{r})$. Now, since for each element $\underset{\sim}{w} \in C(a)$ there is a corresponding element

$\underset{\sim}{u} = \bar{\underset{\sim}{r}} \ominus \underset{\sim}{w} \in U(a)$ and vice versa, we may conclude that

$$\sum_{\underset{\sim}{w}\in C(a)} Pr(\underset{\sim}{w}|\underset{\sim}{r}) = \sum_{\underset{\sim}{u}\in U(a)} Pr(\underset{\sim}{u}|\underset{\sim}{r}).$$

Q.E.D.

Lemma 2: $\sum_{\underset{\sim}{u}\in U(a)} Pr(\underset{\sim}{u}|\underset{\sim}{r}) = \sum_{\underset{\sim}{v}\in E(a)} Pr(\underset{\sim}{v}|\underset{\sim}{r})$.

Proof: We prove Lemma 2 by proving that U(a)=E(a). We first assume that $\underset{\sim}{x}\in U(a)$, so that $\underset{\sim}{x} = \bar{\underset{\sim}{r}} \ominus \underset{\sim}{y}$, where $\underset{\sim}{y}\in C(a)$. Thus $y_m=a$ and $\underset{\sim}{x}B^T = \bar{\underset{\sim}{r}}B^T = \underset{\sim}{S}$. Then $\underset{\sim}{x}\in E(a)$. Now we assume that $\underset{\sim}{z}\in E(a)$. Then $z_m=\bar{r}_m \ominus a$ and $\underset{\sim}{z}B^T = \underset{\sim}{S}$. Define $\underset{\sim}{w} = \bar{\underset{\sim}{r}} \ominus \underset{\sim}{z}$. Then $w_m=a$ and $\underset{\sim}{w}B^T=\underset{\sim}{0}$. So $\underset{\sim}{w}\in C(a)$ and $\underset{\sim}{z}\in U(a)$.

Q.E.D.

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

Theorem 1: $$\sum_{\underset{\sim}{v}\in E(a)} Pr(\underset{\sim}{v}|\underset{\sim}{r}) = \sum_{\underset{\sim}{w}\in C(a)} Pr(\underset{\sim}{w}|\underset{\sim}{r}). \qquad (3)$$

Now (3) may be written as

$$\sum_{\underset{\sim}{v}\in E(a)} Pr(\underset{\sim}{v}|\underset{\sim}{r}) = \sum_{\underset{\sim}{w}\in C(a)} Pr(\underset{\sim}{r}|\underset{\sim}{w})Pr(\underset{\sim}{w})/Pr(\underset{\sim}{r}). \qquad (4)$$

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

$$\sum_{\underset{\sim}{v}\in E(a)} Pr(\underset{\sim}{v}|\underset{\sim}{r}) = \sum_{\underset{\sim}{w}\in G(a)} Pr(\underset{\sim}{w}|\underset{\sim}{r}).$$

So we may conclude that for any given matrix B we have

$$Pr(e_m = \bar{r}_m \ominus a|\underset{\sim}{r}) = \sum_{\underset{\sim}{w}\in G(a)} Pr(\underset{\sim}{w}|\underset{\sim}{r}).$$

An alternative way to see this result is to observe that

$\Pr(\underset{\sim}{v}|\underset{\sim}{r}) = \Pr(\underset{\sim}{w} = \bar{\underset{\sim}{r}} \ominus \underset{\sim}{v}|\underset{\sim}{r}) = 0$ whenever $\underset{\sim}{w} = \bar{\underset{\sim}{r}} \ominus \underset{\sim}{v} \notin C$. This also shows that the complexity of calculating $\Pr(e_m = \bar{r}_m \ominus a|\underset{\sim}{r})$ 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 $\Pr(\underset{\sim}{v}|\underset{\sim}{r}) = P(\underset{\sim}{w} = \bar{\underset{\sim}{r}} \ominus \underset{\sim}{v}|\underset{\sim}{r})$ where $\underset{\sim}{v} \in E(a)$ and $\underset{\sim}{w} = \bar{\underset{\sim}{r}} \ominus \underset{\sim}{v} \in C(a)$. Since the channel is memoryless we have

$$\Pr(\underset{\sim}{v}|\underset{\sim}{r}) = \prod_{i=0}^{n-1} \Pr(r_i|w_i)(\Pr(\underset{\sim}{w})/\Pr(\underset{\sim}{r}))$$

$$= \prod_{i=0}^{n-1} (\Pr(w_i|r_i)\Pr(r_i)/\Pr(w_i))(\Pr(\underset{\sim}{w})/\Pr(\underset{\sim}{r}))$$

where $\underset{\sim}{w} = \bar{\underset{\sim}{r}} \ominus \underset{\sim}{v}$. We note that $\Pr(w_i|r_i) = \Pr(\bar{r}_i \ominus w_i|r_i) = \Pr(v_i|r_i)$. Thus (1) may be written as:

$$\Pr(e_m = \bar{r}_m \ominus a|\underset{\sim}{r}) = \sum_{\underset{\sim}{v} \in E(a)} \left(\prod_{i=0}^{n-1} \Pr(v_i|r_i)\Pr(r_i)/\Pr(\bar{r}_i \ominus v_i) \right) (\Pr(\bar{\underset{\sim}{r}} \ominus \underset{\sim}{v})/\Pr(\underset{\sim}{r}))$$

or

$$\Pr(e_m = \bar{r}_m \ominus a|\underset{\sim}{r}) = R(\underset{\sim}{r}) \sum_{\underset{\sim}{v} \in E(a)} \left(\prod_{i=0}^{n-1} \Pr(v_i|r_i)/\Pr(\bar{r}_i \ominus v_i) \right) \Pr(\bar{\underset{\sim}{r}} \ominus \underset{\sim}{v}) \qquad (5)$$

where

$$R(\underset{\sim}{r}) = \prod_{j=0}^{n-1} \Pr(r_j)/\Pr(\underset{\sim}{r}),$$

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

$$\Pr(e_m=\bar{r}_m \ominus a \mid \underset{\sim}{r}) = R(\underset{\sim}{r})p^{n-k} \sum_{\underset{\sim}{v}\in E(a)} \prod_{i=0}^{n-1} \Pr(v_i \mid r_i). \tag{6}$$

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

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

$$\sum_{\underset{\sim}{v}\in E(a)} \prod_{i=0}^{n-1} \Pr(v_i \mid r_i). \tag{7}$$

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:

$$\sum_{\underset{\sim}{v}\in E(a)} \prod_{i=0}^{n-1} \Pr(v_i \mid r_i) = \sum_{\underset{\sim}{w}\in C(a)} \prod_{i=0}^{n-1} \Pr(w_i \mid r_i). \tag{8}$$

Now let $\bar{C}$ be the linear $(n,\bar{k},\bar{d})$ code over GF(p) defined by the matrix B, that is, $\underset{\sim}{w} \in \bar{C}$ if and only if $\underset{\sim}{w}B^T=\underset{\sim}{0}$. 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 $\bar{C}$ whenever the codewords of $\bar{C}$ are transmitted with equal probability. We note that C is a subcode of $\bar{C}$ and thus $\bar{d} \leq d$ and $k \leq \bar{k}$. Since the complexity of calculating (7) is $O(\min(p^{\bar{k}},p^{n-\bar{k}})$ and $\bar{d} \leq d$, 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 $Q_j=\{ j_1, j_2, \ldots, j_{F_j} \}$, $1\le j\le J$, be the set of positionschecked 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 $Q_j\cap Q_i = \phi$ for $i\neq j$, $1\le i,j\le J$. Now let

$$S_j= \sum_{i=1}^{F_j} b_{j_i} \bar{r}_{j_i} \oplus \bar{r}_m, \quad b_{j_i} \in GF(p),$$

and

$$M_j= \{\underset{\sim}{x}=(x_{j_1}, x_{j_2}, \ldots, x_{j_{F_j}}) \mid x_{j_i} \in GF(p), \sum_{i=1}^{F_j} b_{j_i} x_{j_i} = S_j \ominus \bar{r}_m \oplus a\}.$$

We now define APPDR in terms of our formulation.

APPDR: Set $\hat{e}_m= \bar{r}_m \ominus a$, where $a \in GF(p)$ maximizes

$$Pr(x_m=\bar{r}_m \ominus a \mid r_m) \prod_{j=1}^{J} \Big(\sum_{\underset{\sim}{x}\in M_j} \prod_{i=1}^{F_j} Pr(x_{j_i} \mid r_{j_i}) \Big). \tag{9}$$

Since $Q_j\cap Q_i=\phi$ 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 $\bar{d}=2$. Thus the performance of an APP decoder is the same as the performance of the optimum symbol-by-symbol decoder for $\bar{C}$, with $\bar{d}=2$, whenever the codewords of $\bar{C}$ 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

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

where θ_m is the minimum Hamming weight of vectors in C(1), $N(\theta_m)$ the number of codewords of weight θ_m in C(1), $R=k/n$, $\gamma_b=E_b/N_o$ 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

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

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

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

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 make maximum-radius decoding very attractive for practical applications.

References

[1] J.L. Massey, Threshold Decoding. Cambridge, MA.: MIT Press,1963.

[2] C.R.P. Hartmann and L.D. Rudolph, "An optimum symbol-by-symbol decoding rule for linear codes," IEEE Trans. Inform. Theory, vol. IT-22, pp. 514-517, September, 1976.

[3] C.R.P. Hartmann, L.D. Rudolph and K.G. Mehrotra, "Asymptotic performance of optimum bit-by-bit decoding for the white Gaussian channel," IEEE Trans. Inform. Theory, vol. IT-23, pp. 520-522, July, 1977.

[4] L.D. Rudolph, C.R.P. Hartmann, T.-Y. Hwang and N.Q. Duc, "Algebraic analog decoding of linear binary codes," IEEE Trans. Inform. Theory, vol. IT-25, pp. 430-440, July, 1979.

[5] G.D. Forney, Jr., Concatenated Codes, Cambridge, MA.: MIT Press, 1966, p.52.

[6] CNR, Incorporated, "Demod/decoder intergration," Rome Air Development Center, Griffiss AFB, NY, final technical report RADC-TR-78-70 on Contract F30602-76-C-0361.

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

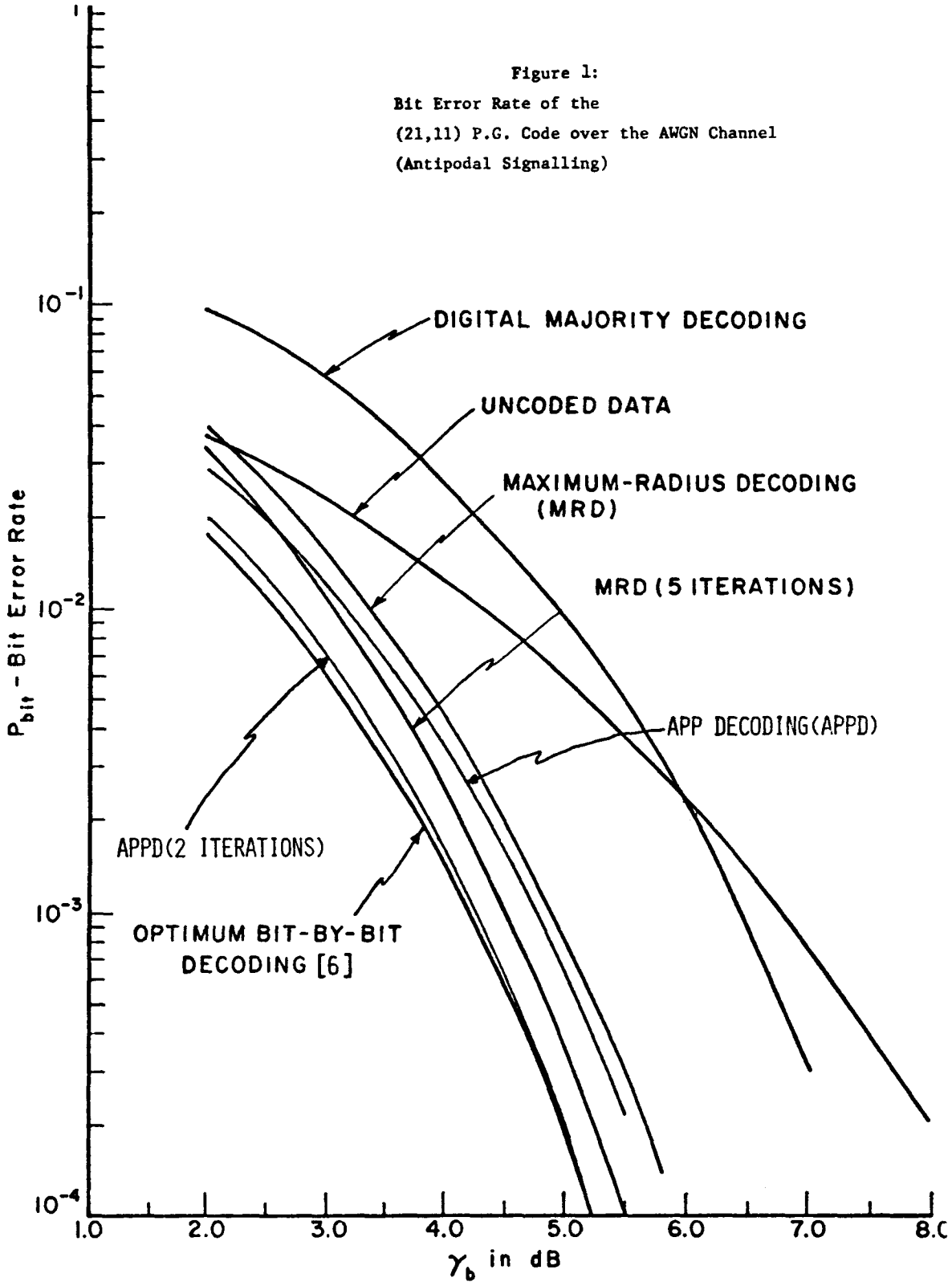

Figure 2:
Word Error Rate of the
(21,11) P.G. Code over the AWGN Channel
(Antipodal Signalling)