Goddard Grant
NN-32-CR
133255
29P.

A Concatenated Coded Modulation Scheme For Error Control

Technical Report

to

NASA

Goddard Space Flight Center

Greenbelt, Maryland 20771

Grant Number NAG 5–931

Shu Lin

Principal Investigator

Department of Electrical Engineering

University of Hawaii at Manoa

Honolulu, Hawaii 96822

April 22, 1988

A Concatenated Coded Modulation Scheme For Error Control

Tadao Kasami
Osaka University
Toyonaka, Osaka 560, Japan

Shu Lin
University of Hawaii
Honolulu, Hawaii 96822

ABSTRACT

This paper presents a concatenated coded modulation scheme for error control in data communications. The scheme is achieved by concatenating a Reed–Solomon outer code and a bandwidth efficient block inner code for M–ary PSK modulation. Error performance of the scheme is analyzed for an AWGN channel. We show that extremely high reliability can be attained by using a simple M–ary PSK modulation inner code and relatively powerful Reed–Solomon outer code. Furthermore, if an inner code of high effective rate is used, the bandwidth expansion required by the scheme due to coding will be greatly reduced. The proposed scheme is particularly effective for high–speed satellite communications for large file transfer where high reliability is required.

This paper also presents a simple method for constructing block codes for M–ary PSK modulation. Some short M–ary PSK codes with good minimum squared Euclidean distance are constructed. These codes have trellis structure and hence can be decoded with a soft–decision Viterbi decoding algorithm.

A CONCATENATED CODED MODULATION SCHEME FOR ERROR CONTROL

1. Introduction

Recently a great deal of research effort has been expended in bandwidth efficient coded modulation for achieving reliable communication on bandlimited channels [1–19]. This new technique of coded modulation is achieved by coding onto an expanded set of channel signals (relative to that needed for uncoded transmission). A properly designed coded modulation can provide significant coding gain over an uncoded modulation system with no or little bandwidth expansion. To achieve a 3 to 5 dB coding gain with a single modulation code, the decoding complexity is quite reasonable. However, to achieve coding gains exceeding 5 dB with a single modulation (trellis or block) code, the decoding complexity increases drastically, and the implementation of the decoder becomes very expensive and unpractical (if not impossible). Then the question is, "how can we achieve coded modulation with reduced complexity?". An answer to this question is to use coded modulation in conjunction with concatenated coding. In this combined coding/modulation scheme, a good short modulation code is used as the inner code and a relatively powerful Reed-Solomon (RS) code is used as the outer code. With properly chosen inner and outer codes, this scheme not only can achieve large coding gain (or high reliability) with good bandwidth efficiency but also can be practically implemented. That is to say, this concatenated coding/modulation scheme offers a way of achieving the best of two worlds.

In this paper, we present a coded modulation scheme with reduced complexity for error control in data communications. This scheme is achieved by concatenating a RS outer code and a bandwidth efficient block inner code for the M-ary PSK modulation. The outer code is interleaved to enhance the overall system performance. We show that this concatenated coded modulation scheme can achieve extremely high reliability (or large coding gain) with a very simple M-ary PSK modulation inner code. Furthermore, if an inner code of high effective rate is used, the overall bandwidth expansion of the scheme due to coding will be greatly reduced. Suppose an inner code with effective rate equal to one is used. Then the overall bandwidth expansion of the scheme is due to outer code coding. Generally, in a concatenated coding scheme, the RS outer code is a high rate code. Hence the bandwidth expansion required is small. Of course, if an inner code with effective rate greater than one is used, the overall scheme may not need any bandwidth expansion at all [16]. The proposed scheme is particularly effective for high–speed satellite communications for large file transfer where high reliability is required.

The presentation of this paper is organized as follows. In Section 2, we present a simple method for constructing block codes for M-ary PSK modulation with $M = 2^l$ which are suitable for concatenated coding. Some short M-ary PSK modulation codes with good minimum squared Euclidean distance are constructed. These codes have simple trellis structure and can be decoded with a soft-decision Viterbi decoding algorithm. In Section 3, the encoding and decoding of the proposed concatenated coded modulation scheme is described. Error performance of the proposed scheme is analyzed in Section 4. In Section 5, a specific concatenated coded modulation scheme with the NASA standard (255,223) RS code over the Galois field GF(2^8) as the outer code and an 8-PSK modulation block code as the inner code is presented. Conclusion is given in Section 6.

2. Bandwidth Efficient Block Codes for M–ary PSK Modulation

In this section, we present a simple method for constructing block codes for M–ary PSK modulation with $M = 2^l$ which are suitable for concatenated coding. Using this method, some good short codes are constructed. These codes have simple trellis structure and can be decoded with a soft–decision Viterbi decoding algorithm.

Let $A = \{0, 1, \ldots, 2^l - 1\}$ be the integer group under the modulo-2^l addition. Define a distance between two elements s and s' in A, denoted $d(s, s')$, as follows: $d(s, s') \triangleq 4\sin^2(2^{-l}\pi(s - s'))$. It is clear that $d(s, s') = d(s - s', 0)$. For $1 \leq i \leq l$, let $B_i = \{0, 2^{i-1}\}$. Then $A = \{b_1 + b_2 + \cdots + b_l : b_i \in B_i \text{ with } 1 \leq i \leq l\}$. Let d_i be the minimum distance between elements in the set, $B_i + B_{i+1} + \cdots + B_l$, for $1 \leq i \leq l$. It is easy to see that $d_i = 4\sin^2(2^{i-1-l}\pi)$. Let A^n denote the set of all n–tuples over A. Define a squared Euclidean distance between two n–tuples, $\mathbf{s} = (s_1, s_2, \ldots, s_n)$ and $\mathbf{s}' = (s_1', s_2', \ldots, s_n')$, over A as follows:

$$d^{(n)}(\mathbf{s}, \mathbf{s}') \triangleq \sum_{j=1}^{n} d(s_j, s_j') \qquad (1)$$

where $d(s_j, s_j') = 4\sin^2(2^{-l}\pi(s - s'))$.

For $1 \leq i \leq l$, let C_i be a block code of length n over B_i with minimum Hamming distance δ_i. From $C_1, C_2, \ldots, C_l$, we construct a block code C with symbols from A as follows:

$$C = \{\mathbf{v}_1 + \mathbf{v}_2 + \cdots + \mathbf{v}_l \quad : \quad \mathbf{v}_i \in C_i \quad \text{with} \quad 1 \le i \le l\}.$$

Then $|C| = \prod_{i=1}^{l} |C_i|$ where $|X|$ denotes the number of elements in set X. Let

$$D[C] \triangleq \min\{d^{(n)}(\mathbf{v}, \mathbf{v}') \quad : \quad \mathbf{v}, \mathbf{v}' \in C \quad \text{and} \quad \mathbf{v} \neq \mathbf{v}'\}. \tag{2}$$

Then $D[C]$ is the minimum squared Euclidean distance of C. It is possible to show that

$$D[C] \ge \min_{1 \le i \le l} \delta_i d_i = \min_{1 \le i \le l} 4\delta_i \sin^2(2^{i-1-l}\pi). \tag{3}$$

The code C_i with symbols from $B_i = \{0, 2^{i-1}\}$ can be constructed from a binary code C_{bi} of the same length and minimum Hamming distance δ_i by substituting 2^{i-1} for 1 in each nonzero component of a code vector in C_{bi}. Denote C_i with $2^{i-1}C_{bi}$. Then the following direct–sum,

$$C = C_{b1} + 2C_{b2} + \cdots + 2^{l-1}C_{bl} \tag{4}$$

is a linear code over the additive group A. The code C_{bi} is called a binary component code of C. Suppose C_{bi} is a binary (n, k_{bi}) linear code. Then

$$|C| = 2^{\sum_{i=1}^{l} k_{bi}}.$$

The parameter $k = \sum_{i=1}^{l} k_{bi}$ is called the dimension of C. If each component of a code vector $\mathbf{v}$ in C is mapped into a point in the 2–dimensional 2^l–PSK signal set, we obtain a block coded 2^l–PSK code. The effective rate of this code C is given by

$$R[C] = \frac{1}{2n} \sum_{i=1}^{l} k_{bi} \tag{5}$$

which is the number of information bits transmitted by C per dimension. The asymptotic coding gain, denoted $\gamma[C]$, of C over the uncoded QPSK is given by [2],

$$\gamma[C] = 10 \log_{10} \frac{R[C] \cdot D[C]}{2}. \tag{6}$$

As an example, let $l = 3$, $M = 2^3$ and $n = 8$. Then $A = \{0, 1, 2, 3, 4, 5, 6, 7\}$, $B_1 = \{0, 1\}, B_2 = \{0, 2\}$ and $B_3 = \{0, 4\}$. The minimum squared Euclidean distances of $A = B_1 + B_2 + B_3, B_2 + B_3$, and B_3 are $d_1 = 0.586, d_2 = 2$ and $d_3 = 4$ respectively. Choose the binary component codes C_{b1}, C_{b2} and C_{b3} as follows: (1) C_{b1} is the binary $(8, 1)$ code which consists of the all–zero and all–one vectors; (2) C_{b2} is the binary (8,7) code with all the even weight vectors; and (3) C_{b3} is the (8,8) code which consists of all binary 8–tuples. Clearly the minimum Hamming distances of C_{b1}, C_{b2} and C_{b3} are 8, 2, and 1 respectively. Consequently, the code $C = C_{b1} + 2C_{b2} + 4C_{b3}$ has minimum squared Euclidean distance $D[C] = 4$, dimension $k = 16$ and effective rate $R = 1$. Mapping the code symbols of C into points of the 8–PSK signal set as shown in Figure 1, we obtain a block code of length 8 for the 8–PSK modulation. This code provides a 3 dB coding gain over the uncoded QPSK modulation with no bandwidth expansion. Furthermore, the code has a trellis of 4 states and 8 sections as shown in Figure 2 (see Appendix for the trellis construction), and hence can be decoded easily by a soft–decision Viterbi decoding algorithm.

Using the above method, we have constructed some short codes for QPSK, 8–PSK and 16–PSK modulations (listed in Table 1). These codes provide 3 to 6 dB coding gain over the uncoded QPSK modulation with little or no bandwidth expansion. They all have trellis structure.

Encoding of a 2^l–PSK modulation code C of length n constructed based on the above method can be done as follows. A message $\mathbf{u}$ of

$$k = \sum_{i=1}^{l} k_{bi}$$

bits (called a segment) is divided into l subsegments, the i–th subsegment consists of k_{bi} bits. For $1 \leq i \leq l$, the i–th subsegment is encoded into a code vector $\mathbf{v}_i$ in the binary component code C_{bi} of C. Then the sum

$$\begin{aligned} \mathbf{v} &= \mathbf{v}_1 + 2\mathbf{v}_2 + \cdots + 2^{l-1}\mathbf{v}_l \\ &= (s_1, s_2, \ldots, s_n) \end{aligned}$$

is the codeword in C for the message segment $\mathbf{u}$. This codeword $\mathbf{v}$ is called a frame. The components, $s_1, s_2, \ldots, s_n$, of $\mathbf{v}$ are then mapped into points in the 2–dimensional 2^l–PSK signal set and transmitted. Hence, each message segment of k bits is encoded into a sequence of n 2^l–PSK signals.

A soft–decision decoding algorithm for the above M–ary PSK codes can be devised as follows. For any element s in the group $A = \{0, 1, \ldots, 2^l - 1\}$, let $X(s)$ and $Y(s)$ be defined as

$$X(s) = \cos(2^{1-l}\pi s), \qquad Y(s) = \sin(2^{1-l}\pi s). \tag{7}$$

For any two elements, s and s', in A, we find that

$$d(s,s') = \left(X(s) - X(s')\right)^2 + \left(Y(s) - Y(s')\right)^2. \tag{8}$$

For $1 \leq j \leq n$, let (x_j, y_j) be the normalized output of a coherent demodulator [20] for the j−th symbol of a received frame. The received frame is then represented by the following $2n$−tuple: $\mathbf{z} = (x_1, y_1, x_2, y_2, \ldots, x_n, y_n)$. For the received frame $\mathbf{z}$ and a codeword $\mathbf{v} = (s_1, s_2, \ldots, s_n)$ in C, let $|\mathbf{z},\mathbf{v}|^2$ be defined as follows:

$$|\mathbf{z},\mathbf{v}|^2 = \sum_{j=1}^{n} \left(x_j - X(s_j)\right)^2 + \left(y_j - Y(s_j)\right)^2. \tag{9}$$

Assume that the channel is an additive white Gaussian noise (AWGN) channel. When symbol $s \in \{0, 1, \ldots, 2^l - 1\}$ is transmitted, the normolized output (x, y) of a coherent demodulator for 2^l−ary PSK is distributed with the following joint probability density function,

$$p(x,y) = \frac{1}{2\pi\sigma^2} \exp\left\{-\left[\left(x - X(s)\right)^2 + \left(y - Y(s)\right)^2\right] \Big/ 2\sigma^2\right\}, \tag{10}$$

where $\sigma^2 = 1/2\rho$, and ρ is the SNR per symbol [20]. Suppose every codeword of C is transmitted with the same probability. Then we have the following decoding rule: For a received frame $\mathbf{z}$, choose a codeword $\mathbf{v}$ in C with minimum $|\mathbf{z},\mathbf{v}|^2$. The segment $\mathbf{u}$ corresponding to $\mathbf{v}$ is then the decoded segment. This decoding rule achieves maximum likelihood decoding for C over an AWGN channel. If C has a simple trellis structure with moderate number of states, the decoding of C can be implemented easily with a Viterbi decoding algorithm.

To analyze the error performance of a 2^l−PSK code C, we need to know its complete weight distribution. Let $\mathbf{v} = (s_1, s_2, \ldots, s_n)$ be an n−tuple over the additive group A. The composition of $\mathbf{v}$, denoted $\mathbf{comp}(\mathbf{v})$, is a 2^l−tuple,

$$\mathbf{t} = (t_0, t_1, \ldots, t_{2^l-1}),$$

where t_i is the number of components s_j in $\mathbf{v}$ equal to the symbol i in A. Let $W(\mathbf{t})$ be the number of codewords $\mathbf{v}$ in C with $\mathbf{comp}(\mathbf{v}) = \mathbf{t}$. Let T be the set,

$$T \triangleq \left\{(t_0, t_1, \ldots, t_{2^l-1}) \quad : \quad 0 \leq t_i \leq n \quad \text{with} \quad 0 \leq i < 2^l\right\}.$$

Then, $\left\{W(\mathbf{t}) : \mathbf{t} \in T\right\}$ is the detail weight distribution of C. $W(\mathbf{t})$ can be enumerated from the joint weight distribution [21] of the binary component codes, $C_{b1}, C_{b2}, \ldots, C_{bl}$ of C.

Once the detail weight distribution of C is known, its error performance can be analyzed and computed for an AWGN channel. The detail weight distributions of the codes listed in Table 1 have been determined [19].

3. The Encoding and Decoding of the Proposed Concatenated Coded Modulation Scheme

For the proposed concatenated coded modulation scheme, the inner code, denoted C_1, is a 2^l−PSK code of length n_1 with binary component codes, $C_{b1}, C_{b2}, \ldots, C_{bl}$, where C_{bi} is an (n_1, k_{bi}) binary linear code for $1 \leq i \leq l$. The dimension of C_1 is

$$k_1 = \sum_{i=1}^{l} k_{bi}.$$

The outer code of the scheme, denoted C_2, is an (n_2, k_2) RS code with symbols from the Galois field GF(2^b) and minimum (Hamming) distance $d_2 = n_2 - k_2 + 1$. Each code symbol of the outer code is represented by a binary b−tuple (called a b−bit byte) based on a certain basis of GF(2^b). We require that $k_1 = mb$ where m is a positive integer.

The encoding of the proposed scheme is performed in two stages. First a message of $k_2 b$ bits is divided into k_2 b−bit bytes. Each b−bit byte is regarded as a symbol in GF(2^b). These k_2 bytes are encoded according to the outer code C_2 to form an n_2−byte codeword in C_2. This codeword is then temporarily stored in a buffer as a row in an array. After m outer codewords have been formed, the buffer stores an $m \times n_2$ array, called a segment-array as shown in Figure 3. Each row of a segment-array is called a section. Each column of a segment-array consists of m b−bit bytes (or mb bits), and is called a segment. There are k_2 data segments and $n_2 - k_2$ parity segments. At the second stage of encoding, each segment of a segment-array is encoded according to the inner 2^l−PSK code C_1 to form a sequence of n_1 2^l−PSK signals as described in the previous section. This sequence of n_1 2^l−PSK signals is called a frame. The n_2 frames corresponding to the segments of a segment-array form a code block. A code block is transmitted column by column (or frame by frame). In fact each frame is transmitted as soon as it has been formed. Note that the outer code is interleaved with a depth (or degree) of m.

The decoding for the proposed scheme also consists of two stages, the inner and outer decodings. When a frame in a code block is received, it is decoded into a segment of m bytes based on the soft-decision decoding algorithm as described in the previous section.

The decoded segment is then stored as a column of an array in a receiver buffer for the second stage of decoding. After n_2 frames of a received code block have been decoded, the receiver buffer contains a $m \times n_2$ decoded segment-array. Each column of this decoded segment-array may contain symbol (or byte) errors which are distributed among the m sections (rows), at most one symbol error in each section. Now the second stage of decoding begins. Each section of the decoded segment array is decoded based on the RS outer code C_2. Suppose the RS outer code is designed to correct t_2 or fewer symbol errors with $0 \leq t_2 \leq \lfloor (n_2 - k_2)/2 \rfloor$. If the syndrome of a section corresponds to an error pattern of t_2 or fewer symbol errors, error correction is performed. The values and locations of symbol errors are determined based on a certain algorithm. If more than t_2 symbol errors are detected, the receiver stops the decoding process and declares an erasure (or raises a flag) for the entire segment array. If all the m sections of a segment-array are successfully decoded and the number of segments which contain corrected symbols is t_2 or less, then the k_2 decoded data segments are accepted by the receiver and delivered to the user in proper order. Otherwise, the receiver declares an erasure for the entire decoded segment-array.

When the receiver fails to decode a received block, the block is erased from the receiver buffer and a retransmission for that block is requested. However, if retransmission is either not possible or not practical and no block is allowed to be discarded, then the erroneous block with all the parity symbols removed is accepted by the user with alarm.

4. Error Performance Analysis

In this section, we analyze the performance of the proposed concatenated coded modulation scheme for an AWGN channel. We assume that all the codewords of the inner modulation code are equally likely to be transmitted.

Let $P_c^{(1)}$ be the probability that a decoded segment is error–free and $P_{ic}^{(1)}$ be the probability that a decoded segment is erroneous. Since the inner code C_1 is linear over $\{0, 1, \ldots, 2^l - 1\}$ under binary component–wise modulo–2 addition, we assume that the all–zero codeword $\mathbf{0}$ is transmitted without loss of generality. For a received frame $\mathbf{z}$, the decoded segment is error–free if and only if

$$|\mathbf{z}, \mathbf{v}|^2 > |\mathbf{z}, \mathbf{0}|^2 \tag{11}$$

for every nonzero codeword $\mathbf{v}$ in C_1 (the probability that $|\mathbf{z}, \mathbf{v}|^2 = |\mathbf{z}, \mathbf{0}|^2$ is assumed to be zero). It follows from (9) that the inequality of (11) can be put into the following form:

$$2\sum_{j=1}^{n_1}(X(s_j)-1)(x_j-1)+Y(s_j)y_j<\sum_{j=1}^{n_1}(X(s_j)-1)^2+Y(s_j)^2=d^{(n)}(\mathbf{v},0), \quad (12)$$

where $\mathbf{z}=(x_1,y_1,x_2,y_2,\ldots,x_{n_1},y_{n_1})$ and $\mathbf{v}=(s_1,s_2,\ldots,s_{n_1})$. For any codeword $\mathbf{v}$ in C_1, let $Q(\mathbf{v})$ denote the set of $(x_1,y_1,x_2,y_2,\ldots,x_{n_1},y_{n_1})$ which satisfies the inequality of (12). Define the following set,

$$Q_c \triangleq \bigcap_{\mathbf{v}\in C_1-\{0\}} Q(\mathbf{v}). \quad (13)$$

Then it follows from (10) that

$$P_c^{(1)}=\frac{1}{(2\pi\sigma^2)^{n_1}}\int\underset{Q_c}{\cdots}\int \exp\left\{-\left(\sum_{j=1}^{n_1}(x_j-1)^2+y_j^2\right)/2\sigma^2\right\}dx_1dy_1\cdots dx_{n_1}dy_{n_1} \quad (14)$$

where the integration is taken over the set Q_c.

Let S be a subset of $C_1-\{0\}$ such that

$$\bigcap_{\mathbf{v}\in C_1-\{0\}} Q(\mathbf{v})=\bigcap_{\mathbf{v}\in S} Q(\mathbf{v}). \quad (15)$$

Then S is called a representative set for $C_1-\{0\}$. To evaluate (14), it is desired to find a small set S to represent $C_1-\{0\}$.

For a nonzero codeword $\mathbf{v}$ in C_1, let $P_e^{(1)}(\mathbf{v})$ denote the probability that a received frame $\mathbf{z}$ satisfies the following condition:

$$|\mathbf{z},\mathbf{v}|^2<|\mathbf{z},0|^2. \quad (16)$$

The inequality of (16) can be put into the following form:

$$2\sum_{j=1}^{n_1}(X(s_j)-1)(x_j-1)+Y(s_j)y_j\geq|\mathbf{v}|^2 \quad (17)$$

where $|\mathbf{v}|=\sqrt{d^n(\mathbf{v},0)}$. Since the random variable,

$$2\sum_{j=1}^{n_1}(X(s_j)-1)(x_j-1)+Y(s_j)y_j$$

is distributed with a Gaussian distribution of zero mean and variance $4\sigma^2|\mathbf{v}|^2$, we have

$$P_e^{(1)}(\mathbf{v}) = \int_{|\mathbf{v}|^2}^{\infty} \frac{1}{2\sqrt{2\pi}\sigma|\mathbf{v}|} \exp\{-\frac{x^2}{8\sigma^2|\mathbf{v}|^2}\}dx$$

$$= \frac{1}{2}\text{erfc}(\frac{|\mathbf{v}|}{2\sqrt{2}\sigma}) = \frac{1}{2}\text{erfc}(\frac{\sqrt{\rho}|\mathbf{v}|}{2}) \tag{18}$$

where

$$\text{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_x^{\infty} \exp\{-t^2\}dt$$

and ρ is the SNR per symbol [20].

Let $\bar{Q}_c$ and $\bar{Q}(\mathbf{v})$ denote the complementary sets of Q_c and $Q(\mathbf{v})$ respectively. Then it follows from (13) and (15) that

$$\bar{Q}_c = \bigcup_{\mathbf{v} \in S} \bar{Q}(\mathbf{v}), \tag{19}$$

where S is a representative set for $C_1 - \{0\}$. Consequently, we obtain the following upper bound on $P_{ic}^{(1)}$,

$$P_{ic}^{(1)} = 1 - P_c^{(1)} \leq \sum_{\mathbf{v} \in S} P_e^{(1)}(\mathbf{v}). \tag{20}$$

Let $\triangle$ be the set of real numbers such that, for any $d \in \triangle$, there is a nonzero codeword $\mathbf{v}$ in C_1 with squared Euclidean distance d from the all–zero codeword 0. For a $d \in \triangle$ and a representative set S for $C_1 - \{0\}$, let $A_d[S]$ be the number of codewords of C_1 in S with squared Euclidean distance d from the all–zero codeword 0. Then it follows from (18) and (20) that

$$P_{ic}^{(1)} \leq \frac{1}{2} \sum_{d \in \triangle} A_d[S]\text{erfc}(\sqrt{d\rho}/2). \tag{21}$$

$A_d[C_1 - \{0\}]$ can be computed from the complete weight distribution of C_1. If we can choose a small representative set S for $C_1 - \{0\}$, $A_d[S]$ may be much smaller than $A_d[C_1 - \{0\}]$ except for "dominate" d's close to the minimum squared Euclidean distance $D[C_1]$ of C_1.

Next we analyze the error performance of the overall concatenated coded modulation scheme. Let P_c, P_{es} and P_{er} be the probabilities of a correct decoding, an erasure and an incorrect decoding for an entire received code block respectively. Then

$$1 - P_c = P_{es} + P_{er} = \sum_{i=t_2+1}^{n_2} \binom{n_2}{i} [P_{ic}^{(1)}]^i [1 - P_{ic}^{(1)}]^{n_2 - i}. \tag{22}$$

Let $\bar{P}_{ic}^{(1)}$ denote an upper bound on $P_{ic}^{(1)}$, say the right–hand side of (20). Then it follows from (22) that

$$P_c \geq 1 - \sum_{i=t_2+1}^{n_2} \binom{n_2}{i} g_{n_2,i}(\bar{P}_{ic}^{(1)}) \tag{23}$$

where $g_{n_2,i}(x) = x^i(1-x)^{n_2-i}$ for $0 \leq x \leq i/n_2$, and $g_{n_2,i}(x) = (i/n_2)^i(1-i/n_2)^{n_2-i}$ otherwise.

For a nonnegative integer i, let $d[i]$ denote the smallest d' in $\triangle$ such that

$$\sum_{\substack{d \in \triangle, \\ d < d'}} A_d[C_1 - \{0\}] \geq i. \tag{24}$$

An upper bound on the probability of an incorrect decoding, P_{er}, can be derived in a way similar to the one given in Appendix F of [22],

$$P_{er} \leq \sum_{w=d_2}^{n_2} \binom{n_2}{w} \sum_{h=0}^{\min\{t_2, t_2-w\}} \binom{n_2-w}{h} \sum_{j=w+h-t_2}^{w} \binom{w}{j} \bar{P}(w,h,j), \tag{25}$$

where $d_2 = n_2 - k_2 + 1$, and

$$\begin{aligned} \bar{P}(w,h,j) = {} & g_{n_2-d_2, w+h-d_2}(\bar{P}_{ic}^{(1)}) \\ & \cdot \Bigg\{ \sum_{\substack{d \in \triangle, \\ d < d[j+d_2-w]}} A_d[C_1 - \{0\}] \Big[\frac{1}{2} \operatorname{erfc}\Big(\frac{\sqrt{d\rho}}{2}\Big)\Big]^{j+d_2-w} \\ & + \Big(j + d_2 - w - \sum_{\substack{d \in \triangle, \\ d < d[j+d_2-w]}} A_d[C_1 - \{0\}]\Big) \\ & \cdot \Big[\frac{1}{2} \operatorname{erfc}\Big(\frac{\sqrt{d[j+d_2-w]\rho}}{2}\Big)\Big]^{j+d_2-w} \Bigg\} \end{aligned} \tag{26}$$

Let $\check{P}_c$ and $\hat{P}_{er}$ denote the right–hand sides of (23) and (25) respectively. Then $\check{P}_c$ is a lower bound on P_c, and $\hat{P}_{er}$ is an upper bound on P_{er}. The probability, $1 - \check{P}_c$, is an upper bound on the total probability of a decoding failure and a decoding error. Clearly, $1 - \check{P}_c$ serves as an upper bound on the probability that a received block will be rejected (we will call this as the rejection rate). The performance of the proposed concatenated coded modulation scheme is then measured by the pair of probabilities, $\hat{P}_{er}$ and $1 - \check{P}_c$.

5. A Specific Concatenated Coded Modulation Scheme

In this section, we consider a specific concatenated coded modulation scheme, in which the outer code C_2 is the NASA standard (255,223) RS code over GF(2^8) and the inner code C_1 is the 8–PSK code of length 8 (described in Section 2) with a 4–state trellis structure as shown in Figure 2. The inner code has dimension $k_1 = 16$, effective rate $R[C_1] = 1$ and minimum squared Euclidean distance $D[C_1] = 4$. Since $k_1 = 16$ and $b = 8$, the outer code is interleaved to a depth of $m = 2$. The overall code rate of the scheme is $R_{eff} = (k_2/n_2) \cdot R[C_1] = (223/255) \cdot 1 = 0.875$. The inner code provides a 3 dB (asymptoic) coding gain over the uncoded QPSK modulation without bandwidth expansion. The bandwidth expansion required by the overall scheme is due to the coding of the outer RS code.

The complete weight distribution of the inner code can be enumerated from the joint weight distribution of its three binary component codes, C_{b1}, C_{b2} and C_{b3}. For integers i, j and h such that i is even, $0 \leq j \leq i \leq 8$ and $0 \leq h \leq 8 - i$,

$$\begin{aligned} &W\big((h, 0, j, 0, 8 - i - h, 0, i - j, 0)\big) \\ &\quad = W\big((0, h, 0, j, 0, 8 - i - h, 0, i - j)\big) \\ &\quad = \binom{8}{i}\binom{i}{j}\binom{8-i}{h} \end{aligned} \tag{27}$$

For other composition $\mathbf{t} = (t_0, t_1, t_2, t_3, t_4, t_5, t_6, t_8)$, $W(\mathbf{t}) = 0$.

By choosing a representative set for $C_1 - \{0\}$, we find the following upper bound on the probability of an incorrect decoding for the inner code,

$$P_{ic}^{(1)} \leq 60 \text{ erfc}(\sqrt{\rho}) + 64 \text{ erfc}\Big(\sqrt{2(2-\sqrt{2})\rho}\Big) + 484 \text{ erfc}\Big(\frac{\sqrt{2(8-3\sqrt{2})\rho}}{2}\Big). \tag{28}$$

Let $\bar{P}_{ic}^{(1)}$ denote the upper bound on $P_{ic}^{(1)}$ given by the right hand side expression of (28). Then $\bar{P}_{ic}^{(1)}$ is used as a measure of the error performance of the inner code C_1. The error performance of the inner code and that of the uncoded QPSK versus SNR per symbol is shown in Figure 4. Simulation result on $P_{ic}^{(1)}$ is also included. Note that the difference between $\bar{P}_{ic}^{(1)}$ and the simulation result on $P_{ic}^{(1)}$ is very small for SNR > 8 dB/symbol (or 5 dB/information bit). We also see that the simple 8–PSK inner code provide a 2.31 dB (real) coding gain over the uncoded QPSK modulation at 10^{-5} decoded segment–error rate, and a 2.42 dB coding gain at 10^{-6} decoded segment–error rate.

The RS outer code is a very powerful code which is capable of correcting up to 16 symbol (or byte) errors. From (23) and (28), we can compute the probability $1 - \check{P}_c$ which is an upper bound on the probability of a decoding failure. From (25) and (28), we can compute the probability $\hat{P}_{er}$ which is an upper bound on the probability P_{er} of an incorrect decoding of a received block. The error performance of the scheme is given by the pair of probabilities, $\hat{P}_{er}$ and $1 - \check{P}_c$.

We can also use the simulation result on $P_{ic}^{(1)}$, denoted $P_{ic,s}^{(1)}$, to compute the error performance of the overall scheme. Let $1 - P_{c,s}$ denote the value computed from the expression of (22) with $P_{ic}^{(1)}$ replaced by the simulation result $P_{ic,s}^{(1)}$. Then $1 - P_{c,s}$ gives the total probability of a decoding failure and a decoding error for the overall scheme based on the simulation results of the inner code.

Using the outer code to correct up to 16 symbol errors ($t_2 = 16$), the error performance of the overall scheme is shown in Figures 5 to 8, where Figures 5 and 6 give the block–error performance and Figures 7 and 8 give the bit–error performance. Figure 5 gives the upper bound $1 - \check{P}_c$ and simulation result $1 - P_{c,s}$ on the total probability of a decoding failure and a decoding error versus SNR/symbol, and Figure 6 gives the upper bound $\hat{P}_{er}$ on the probability P_{er} of a decoding error for the overall scheme versus SNR/symbol. From Figures 5 to 6, we see that the specific concatenated modulation scheme proposed in this section provides extremely high reliability. For example, with SNR = 9 dB/symbol (or 6.57 dB/information bit), the probability of a decoding error is upper bounded by 6.28×10^{-25} and the rejection rate is upper bounded by 4.95×10^{-16} (using simulation results of the inner code). We see that the rejection rate is extremely small and it does not affect the system throughput. With SNR = 10 dB/symbol (or 7.57 dB/information bit), the probability of a decoding error is less than 6.80×10^{-41}! Practically, the scheme achieves error–free communication. Figure 5 also shows the coding gain of the scheme over the uncoded QPSK in terms of decoded block–error rate. For example, at decoded block–error rates, 10^{-7} and 10^{-10}, the scheme achieves coding gains, 8 dB/symbol and 9 dB/symbol, over the uncoded QPSK (a block of 2×223 bytes) respectively. These are large coding gains.

For data file transfer, the block–error–rates should be used as the measure of error performance of the scheme. However, for the purpose of providing a basis for comparing with other coding schemes, we follow the conventional practice to compute the decoded bit–error–rate (BER) of the scheme. The bit–error performance of the scheme is shown in Figures 7 and 8. Two types of decoded BERs are computed. The first type, denoted P_{b1}, is computed based on the probability P_{er} of an incorrect decoding of a code block using the approximation, $P_{b1} = (d_2/2n_2) \cdot P_{er}$. This type of BER is a measure of bit–error performance of the scheme when retransmission is allowed. The type–1 decoded BER of the scheme versus SNR/symbol is shown in Figure 7. We see that the scheme achieves large coding gain over the uncoded QPSK for $P_{b1} \geq 10^{-6}$. At $P_{b1} = 10^{-6}$ and $P_{b1} = 10^{-10}$, the coding gains are 6.16 dB/symbol (or 5.58 dB/information bit) and 8.30 dB/symbol (or 7.72 dB/information bit) respectively. The required SNR to achieve $P_{b1} = 10^{-6}$ is 7.39 dB/symbol (or 4.96 dB/information bit), and the required SNR to achieve $P_{b1} = 10^{-10}$ is 7.80 dB/symbol (or 5.37 dB/information bit).

The second type of decoded BER, denoted P_{b2}, is computed based on the total probability $1 - P_c$ of a decoding failure and a decoding error of a code block using the approximation, $P_{b2} = (d_2/2n_2) \cdot (1 - P_c)$. This type of BER is used as the measure of bit–error performance of the scheme when retransmission is not allowed. Figure 8 gives the type–2 decoded BER of the scheme versus SNR/symbol. We see that, at $P_{b2} = 10^{-6}$ and $P_{b2} = 10^{-10}$, the coding gains of the scheme over the uncoded QPSK are 5.52 dB/symbol (or 4.94 dB/information bit) and 7.60 dB/symbol (or 7.02 dB/information bit) respectively. The required SNR to achieve $P_{b2} = 10^{-6}$ is 8.04 dB/symbol (or 5.61 dB/information bit), and the required SNR to achieve $P_{b2} = 10^{-10}$ is 8.50 dB/symbol (or 6.07 dB/information bit).

From Figure 2, we see that the 4–state trellis diagram for the 8–PSK inner code consists of two identical parallel 2–state trellis sub–diagrams without cross connection between them. This structure suggests that the decoding of the inner code can be done with two 2–state Viterbi decoders to process the two trellis sub–diagrams in parallel. This implementation not only simplify the decoding complexity but also speeds up the decoding process. Since the inner code is very short, a very high speed decoder can be implemented without much cost.

The specific concatenated coded modulation scheme is suitable for high–speed satellite communications for large file transfer where high reliability is required. Furthermore, the scheme is very robust in correcting burst–errors. It is capable of correcting any single burst of errors of length up to 242 bits.

To reduce the bandwidth expansion due to coding of the (255,223) RS outer code, we may use a higher effective rate M–ary PSK code as the inner code. Suppose we use the

7–th code in Table 1 as the inner code. This code is an 8–PSK code with effective rate $R[C_7] = 9/8$. As a result, the overall code rate of the concatenated coded modulation scheme with the (255,223) RS code as the outer code is $R_{eff} = 0.9844$. In this case, the inner code has a 16–state trellis diagram and is much harder to decode than the 8–PSK code with a 4–state trellis diagram considered above.

6. Conclusion Remarks

In this paper, a concatenated coded modulation scheme for error control in data communications has been presented. This scheme is achieved by concatenating a RS (or maximum–distance–separable) outer code and a bandwidth efficient block inner code for the M–ary PSK modulation. Error performance of this scheme has been analyzed. A simple method for constructing bandwidth efficient block codes for the M–ary PSK modulation has been devised. By a specific example, we have shown that extremely high reliability can be achieved by concatenating a good short 8–PSK modulation inner code and relatively powerful RS outer code, such as the NASA standard (255,223) RS code over GF(2^8). Since the inner code is short, a high speed decoder with a soft–decision decoding algorithm can be implemented without much cost (or complexity). If a proper high effective rate inner code is used, the bandwidth expansion required by the overall scheme due to coding will be greatly reduced. The proposed scheme is actually devised to achieve coded modulation with reduced complexity. It offers a way of obtaining the best of three worlds, reliability, complexity and bandwidth efficiency. The proposed scheme is particularly suitable for high–speed satellite communications for large file transfer where high reliability is required.

The inner code decoder can be implemented to perform both decoding and erasure operations. In this case, a decoded segment may contains symbol errors and an erased segment creates m symbol erasures, one in each section. The RS outer code is then designed to correct both symbol errors and erasures.

Of course, other types of modulation codes can be used as inner codes in the proposed concatenated coded modulation scheme.

Appendix

Trellis Diagram for the 8–PSK Code Described in Section 2

The 8–PSK code C described in Section 2 consists of three binary component codes C_{b1}, C_{b2} and C_{b3} where (1) C_{b1} is the binary (8,1) code which consists of the all–zero and all–one vectors; (2) C_{b2} is the binary (8,7) code with all the even weight vectors; and (3) C_{b3} is the (8,8) code which consists of all binary 8–tuples. Let $\mathbf{u}$ be a 16–bit message segment to be encoded. Divide $\mathbf{u}$ into three sub–segments $\mathbf{u}_1$, $\mathbf{u}_2$ and $\mathbf{u}_3$ where $\mathbf{u}_1$ consists of only one bit, $\mathbf{u}_2$ consists of seven bits and $\mathbf{u}_3$ consists of eight bits. Then $\mathbf{u}_1$, $\mathbf{u}_2$ and $\mathbf{u}_3$ are encoded based on C_{b1}, C_{b2} and C_{b3} respectively. Let

$$\mathbf{a} = (a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8),$$

$$\mathbf{b} = (b_1, b_2, b_3, b_4, b_5, b_6, b_7, b_8),$$

$$\mathbf{c} = (c_1, c_2, c_3, c_4, c_5, c_6, c_7, c_8),$$

be their corresponding binary codewords. Note that $\mathbf{a}$ is either the all–zero vector or the all–one vector. The codeword $\mathbf{b}$ has even weight.

For $1 \leq l \leq 8$, the input to the signal selector of the overall encoder–modulator at the l–th time unit is the triplet (a_l, b_l, c_l). If $a_l = 0$, then (b_l, c_l) selects a point from the QPSK signal set shown in Figure 1b. If $a_l = 1$, the (b_l, c_l) selects a point from the QPSK signal set shown in Figure 1c. Hence the system switches between two QPSK signal sets. To construct the trellis diagram for C, we need to define the states of the overall encoder–modulator. Let $(b_1, b_2, \ldots, b_l)$ denote the l–bit prefix of codeword $\mathbf{b}$. Let $W(b_1, b_2, \ldots, b_l)$ denote the Hamming weight of $(b_1, b_2, \ldots, b_l)$. At the l–th time unit, the state of the encoder–modulator depends on the bit a_l and the number $W(b_1, b_2, \ldots, b_l)$. Define the following states:

(1) A_e represents the state that $a_l = 0$ and $W(b_1, b_2, \ldots, b_l)$ is even;
(2) A_o represents the state that $a_l = 0$ and $W(b_1, b_2, \ldots, b_l)$ is odd;
(3) B_e represents the state that $a_l = 1$ and $W(b_1, b_2, \ldots, b_l)$ is even; and
(4) B_o represents the state that $a_l = 1$ and $W(b_1, b_2, \ldots, b_l)$ is odd.

Assume that the encoder–modulator starts from the state A_e at the time $l = 0$. Then the trellis diagram for C can be constructed easily as shown in Figure 2. There are two parallel branches (or transitions) between the transition of two states; they correspond to $c_l = 0$ and $c_l = 1$ respectively.

The encoding of message **u** is equivalent to tracing a path in the trellis diagram. The codeword corresponding to **u** is a sequence of QPSK signal points either from the set shown in Figure 1b or from the set shown in Figure 1c.

REFERENCES

1. H. Imai and S. Hirakawa, "A New Multilevel Coding Method Using Error Correcting Codes," *IEEE Trans. on Information Theory*, Vol. IT-23, No. 2, March 1977.
2. G. Ungerboeck, "Channel Coding with Multilevel/Phase Signals," *IEEE Trans. on Information Theory*, Vol. IT–28, pp. 55–67, January 1982.
3. V. V. Ginzburg, "Multidimensional Signal for a Continuous Channel," *Problemy Peredachi Informatsii*, Vol. 20, No. 1, pp. 28–46, 1984.
4. L. F. Wei, "Rotationally Invariant Convolutional Channel Coding with Expanded Signal Space – II: Nonlinear Codes," *IEEE Journal of Select. Areas Communications*, Vol. SAC–2, pp. 672–686, 1984.
5. G. D. Forney, Jr., R. G. Gallager, G. R. Lang, F. M. Longstaff, and S. U. Qureshi, "Efficient Modualtion for Band- limited Channels", *IEEE J. Select. Areas Communications*, Vol. SAC–2, pp. 632–647, September 1984.
6. R. M. Tanner, "Algebraic Construction of Large Euclidean Distance Combined Coding/Modulation Systems," presented at *IEEE International Symposium on Information Theory*, October 6–9, 1986, Ann Arbor, Michigan.
7. A. R. Calderbank and N. J. A. Sloane, "An Eight- Dimensional Trellis Code," *Proc. IEEE*, Vol. 74, pp. 757- 759, 1986.
8. S. I. Sayegh, "A Class of Optimum Block Codes in Signal Space," *IEEE Trans. on Communications*, Vol. CT–30, No. 10, October 1986.
9. S. G. Wilson, "Rate 5/6 Trellis–Coded 8–PSK," *IEEE Transactions on Communications*, Vol. COM–34, pp. 1045- 1049, October 1986.
10. A. R. Calderbank and N. J. A. Sloane, "New Trellis Codes Based on Lattices and Cosets," *IEEE Trans. on Information Theory*, Vol. IT–33, No. 2, March 1987.
11. G. J. Pottie and D. P. Taylor, "An Approach to Ungerboeck Coding for Rectangular Signal Sets," *IEEE Trans. on Information Theory*, Vol. IT–33, No. 2, March 1987.
12. L. F. Wei, "Trellis-Coded Modulation with Multidimensional Constellations," *IEEE Trans. on Information Theory*, Vol. IT–33, No. 4, July 1987.
13. G. D. Forney, Jr., "Coset Codes I: Geometry and Classification," to appear in *IEEE Trans. on Information Theory.*
14. G. D. Forney, Jr., "Coset Codes II: Binary Lattices and Related Codes," to appear in *IEEE Trans. on Information Theory.*

15. A. Lafanechere, R. H. Deng, and D. J. Costello, Jr., "Multidimensional Trellis Coded Phase Modulation using Unit-Memory and Partial Unit-Memory Convolutional Codes," submitted to *IEEE Trans. on Information Theory*, 1987.
16. R. H. Deng and D. J. Costello, Jr., "High Rate Concatenated Coding Systems with Bandwidth Efficient Inner Codes," submitted to *IEEE Trans. on Information Theory*, 1987.
17. G. Ungerboeck, "Trellis-Coded Modulation with Redundant Signal Sets, Pat I: Introduction," *IEEE Communications Magzine*, Vol. 25, No. 2, February 1987.
18. G. Ungerboeck, "Trellis-Coded Modulation with Redundant Signal Sets, Pat II: State of the Art," *IEEE Communications Magzine*, Vol. 25, No. 2, February 1987.
19. T. Kasami and S. Lin, "Bandwidth Effcient Block Codes for M–ary PSK Modulation," NASA Technical Report, NAG5–931, December 1987.
20. J. C. Proakis, *Digital Communications*, Mc Graw– Hill, 1983.
21. F. J. MacWilliams and N. J. A. Sloane, *The Theory of Error–Correcting Codes*, North–Holland, 1977.
22. T. Kasami, T. Fujiwara, T. Takata and S. Lin, "A Cascaded Coding Scheme for Error Control and Its Performance Analysis," *IEEE Trans. on Information Theory*, Vol. IT–34, No. 2, May 1988.

Table 1 *Some short QPSK, 8-PSK and 16-PSK codes (the reference system is the uncoded QSPK modulation)*

Modulation	Code C	Length n	$D[C]$	$R[C]$	$\gamma[C]$ dB	No. of states of trellis	Binary component codes C_{b1}	C_{b2}	C_{b3}	C_{b4}
QPSK	C_1	5	4	4/5	2.0	2	$P_2 \times P_3$	V_5		
	C_2	15	8	4/5	5.0	2^4	$s-RM_{4,2}$	P_{15}		
	C_3	16	16	1/2	6.0	2^7	$RM_{4,2}$	$RM_{4,1}$		
	C_4	32	16	21/32	7.2	2^{10}	$RM_{5,2}$	$RM_{5,1}$		
8-PSK	C_5	8	4	1	3.0	2^2	$P_8^{\perp}$	P_8	V_8	
	C_6	15	4	16/15	3.3	2^4	$s-RM_{4,1}$	$P_7 \times P_8$	V_{15}	
	C_7	16	4	9/8	3.52	2^4	$RM_{4,1}$	P_{16}	V_{16}	
	C_8	23	4	28/23	3.8	2^7	G_e	P_{23}	V_{23}	
	C_9	33	8	32/33	5.9	2^9	$RM_{5,1}^*$	$R_{5,3}^*$	P_{33}	
	C_{10}	$4h+3$ $4\leq h\leq 7$	8	$\frac{4h}{4h+3}$	$10\log_{10}\frac{16h}{4h+3}$	2^6	$P_{4h+3}^{\perp}$	$s-RM_{5,3}$	P_{4h+3}	
16-PSK	C_{11}	32	4	5/4	3.9	2^8	$P_{32}^{\perp}$	$RM_{5,2}$	P_{32}	V_{32}

Notations:

(1) $V_n = \{0,1\}^n$;

(2) P_n denotes the $(n, n-1)$ linear binary code which consists of all the even–weight binary n–tuple;

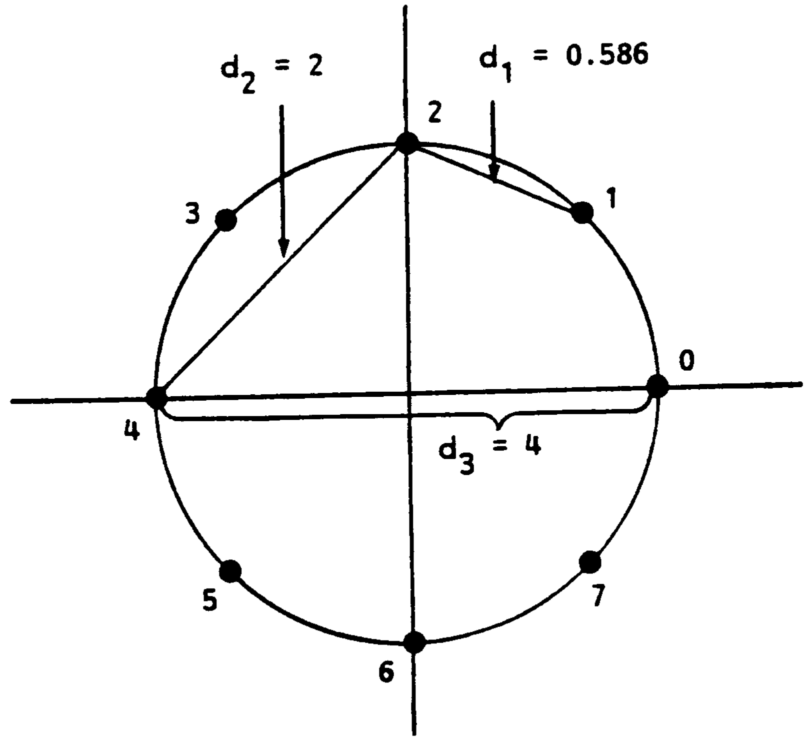
(3) $P_n^{\perp}$ denotes the dual code of P_n, which consists of the all–zero and all–one vectors;

(4) $RM_{i,j}$ denotes the j–th order binary Reed–Muller code of length 2^i;

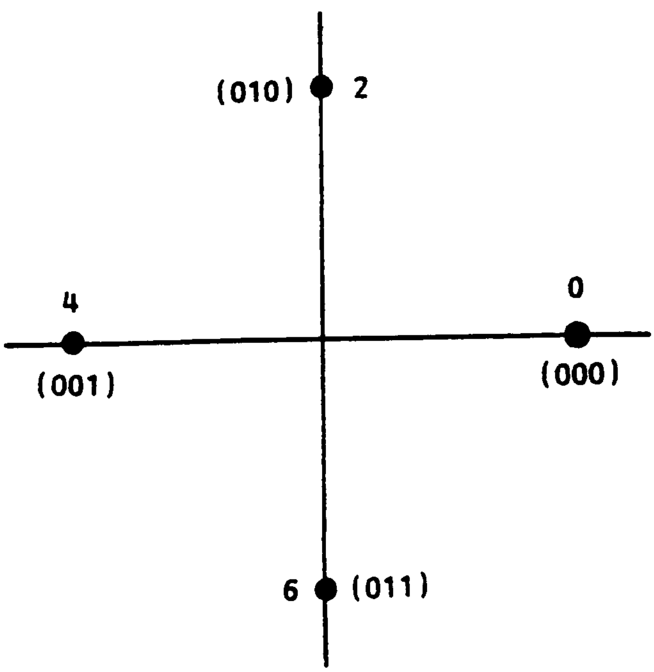
(5) $s - RM_{i,j}$ denotes a shortened version of $RM_{i,j}$;

(6) G_e denotes the even weight subcode of Golay code;

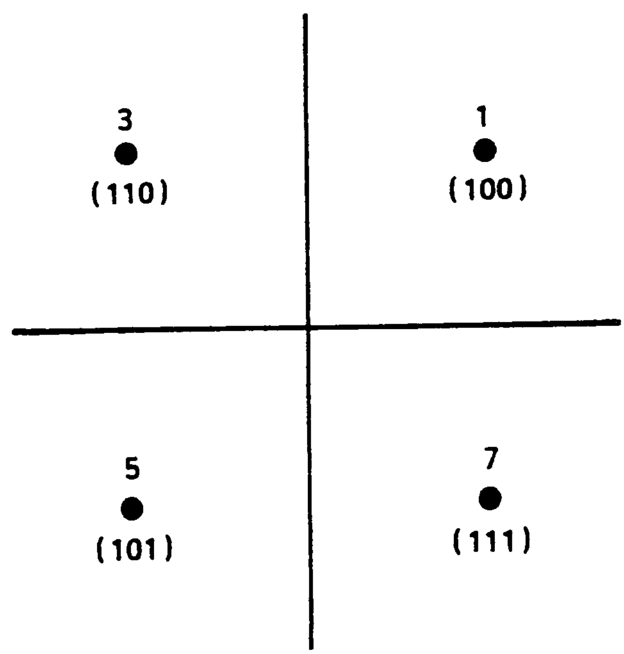
(7) C^* denotes the code obtained from C by appending an extra zero bit to each code vector of C.



(a) 8-PSK signal set and squared Euclidean distances between signal points



(b) QPSK signal set for $a_{\ell}=0$ and signal mapping



(c) QPSK signal set for $a_{\ell}=1$ and signal mapping

Figure 1

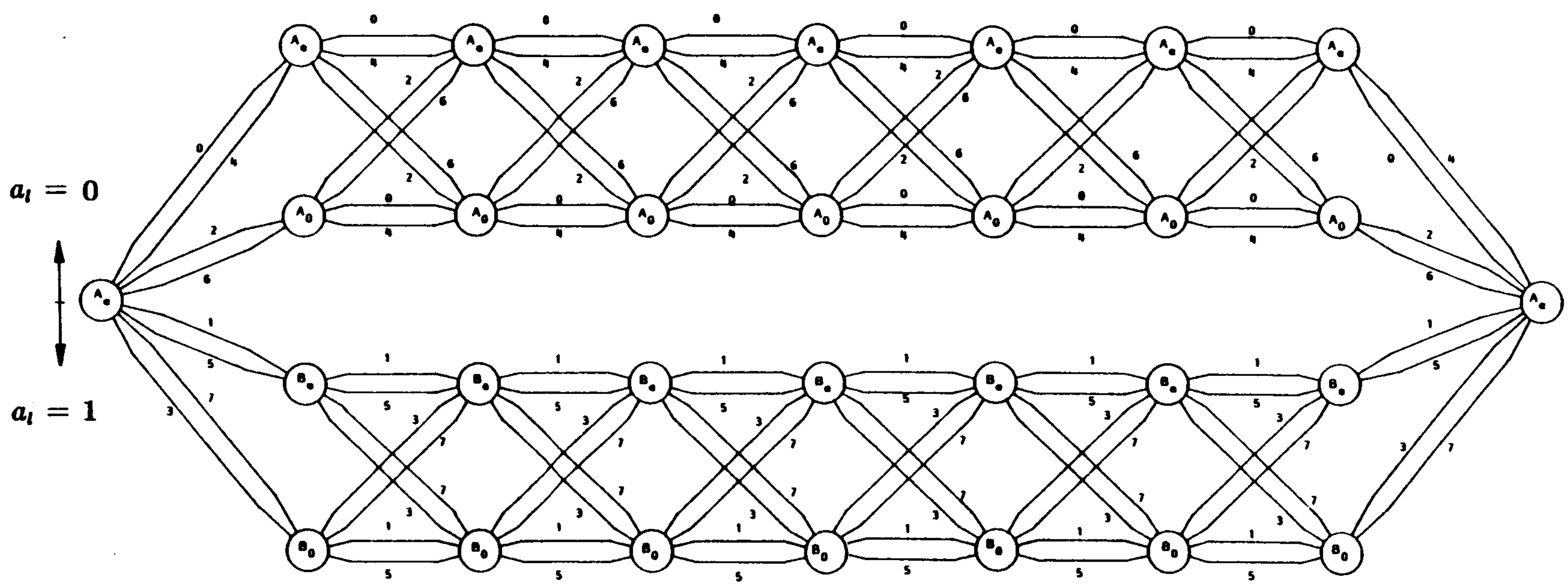


Figure 2 A 4–state trellis diagram for the 8–PSK code described in Section 2

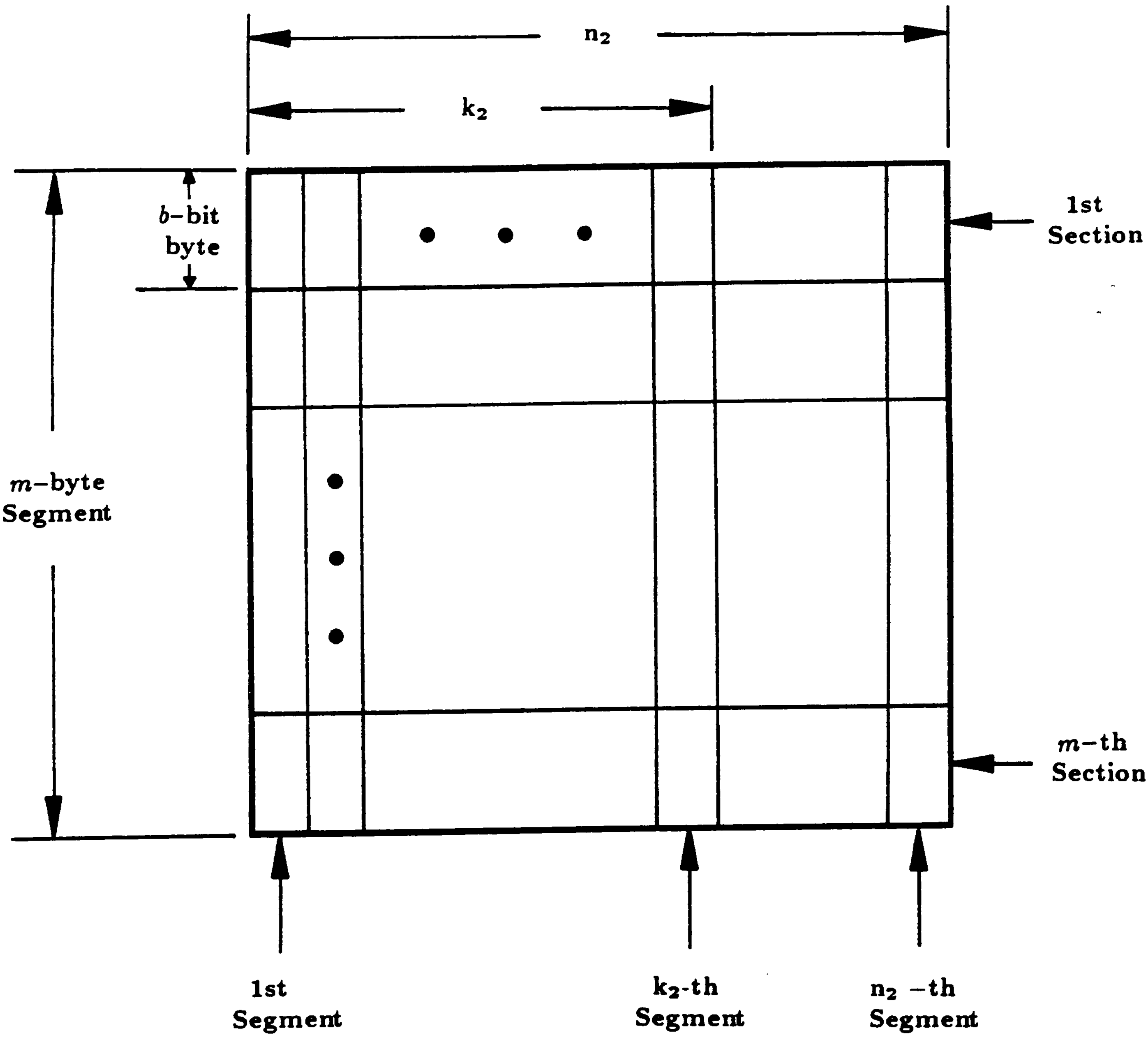


Figure 3 A Segment–Array

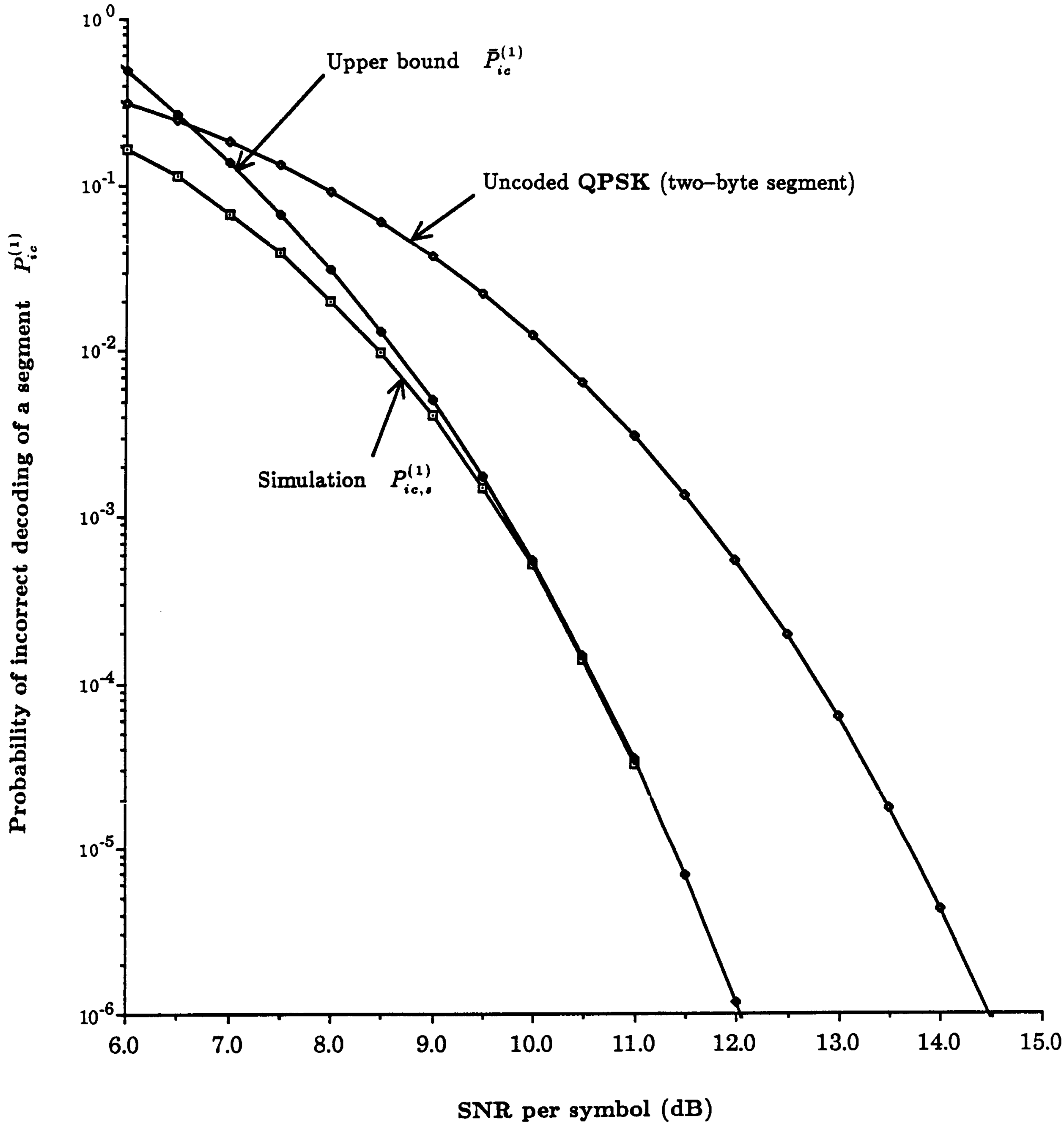


Figure 4 Error performance of the 4-state 8-PSK block inner code

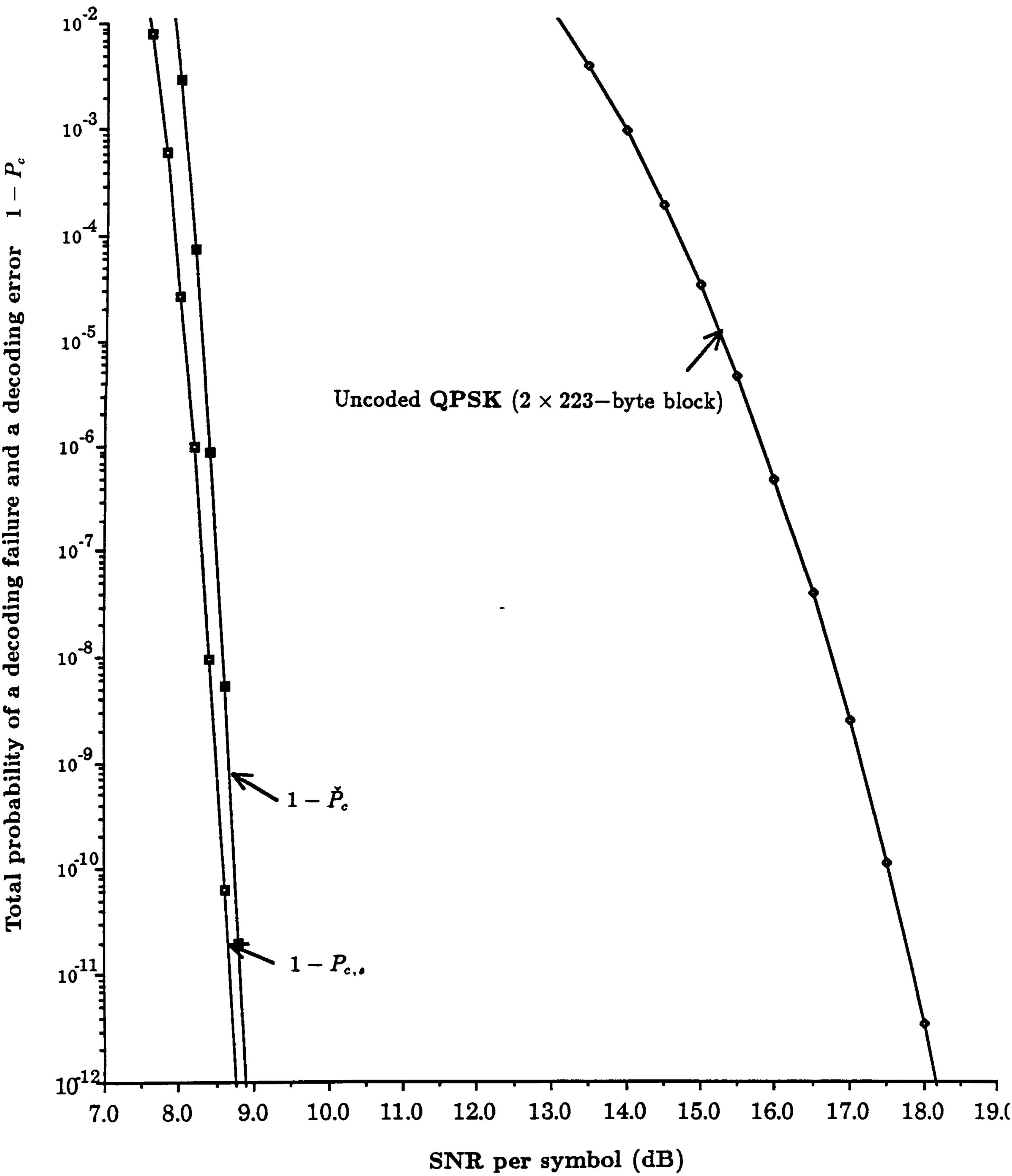


Figure 5 The total probability of a decoding failure and a decoding error for the concatenated modulation scheme with the (255,233) RS outer code and the 4–state 8–PSK block inner code

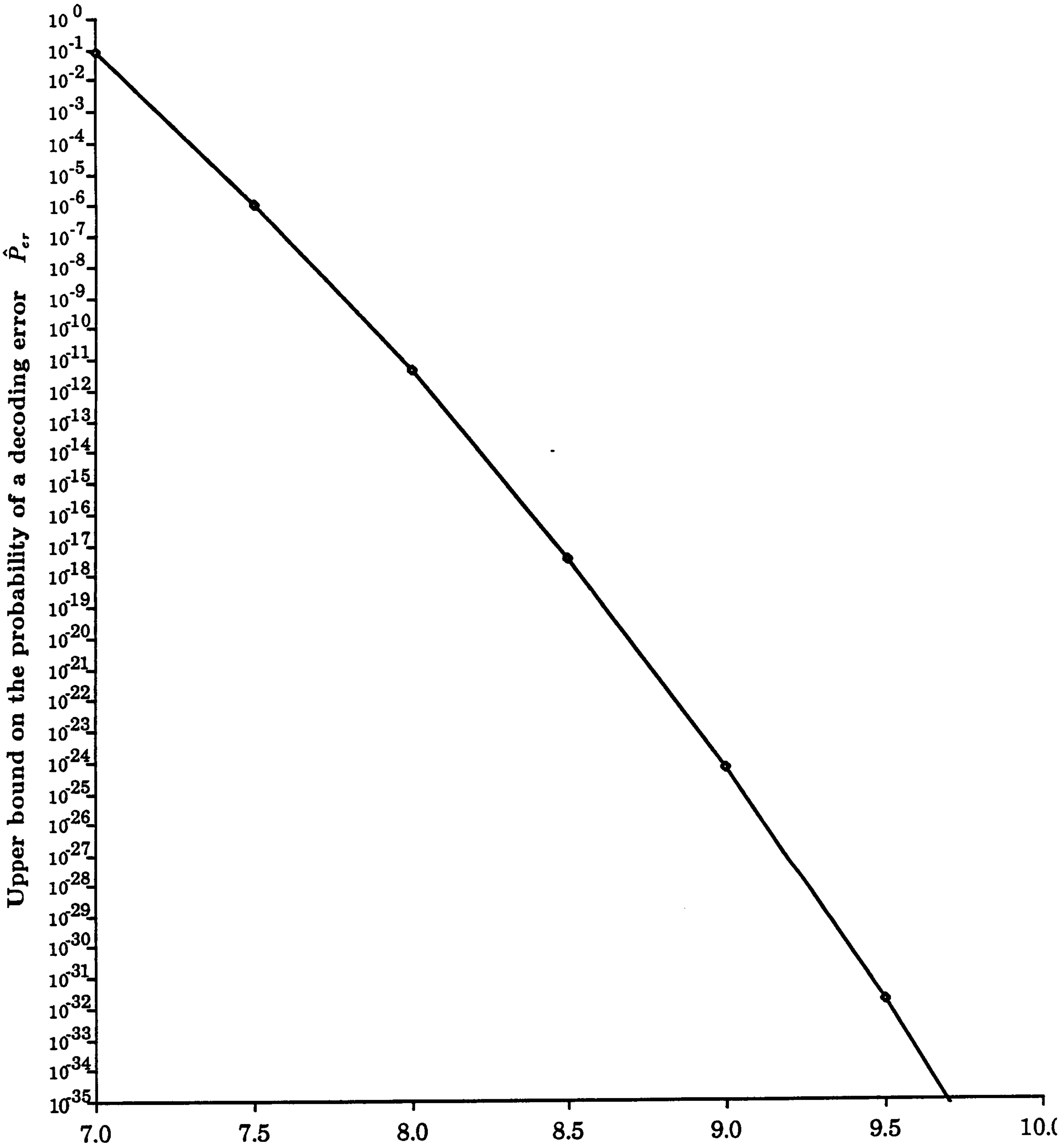


Figure 6 The probability of a decoding error for the concatenated coded modulation scheme with the (255,223) RS outer code and the 4–state 8–PSK block inner code

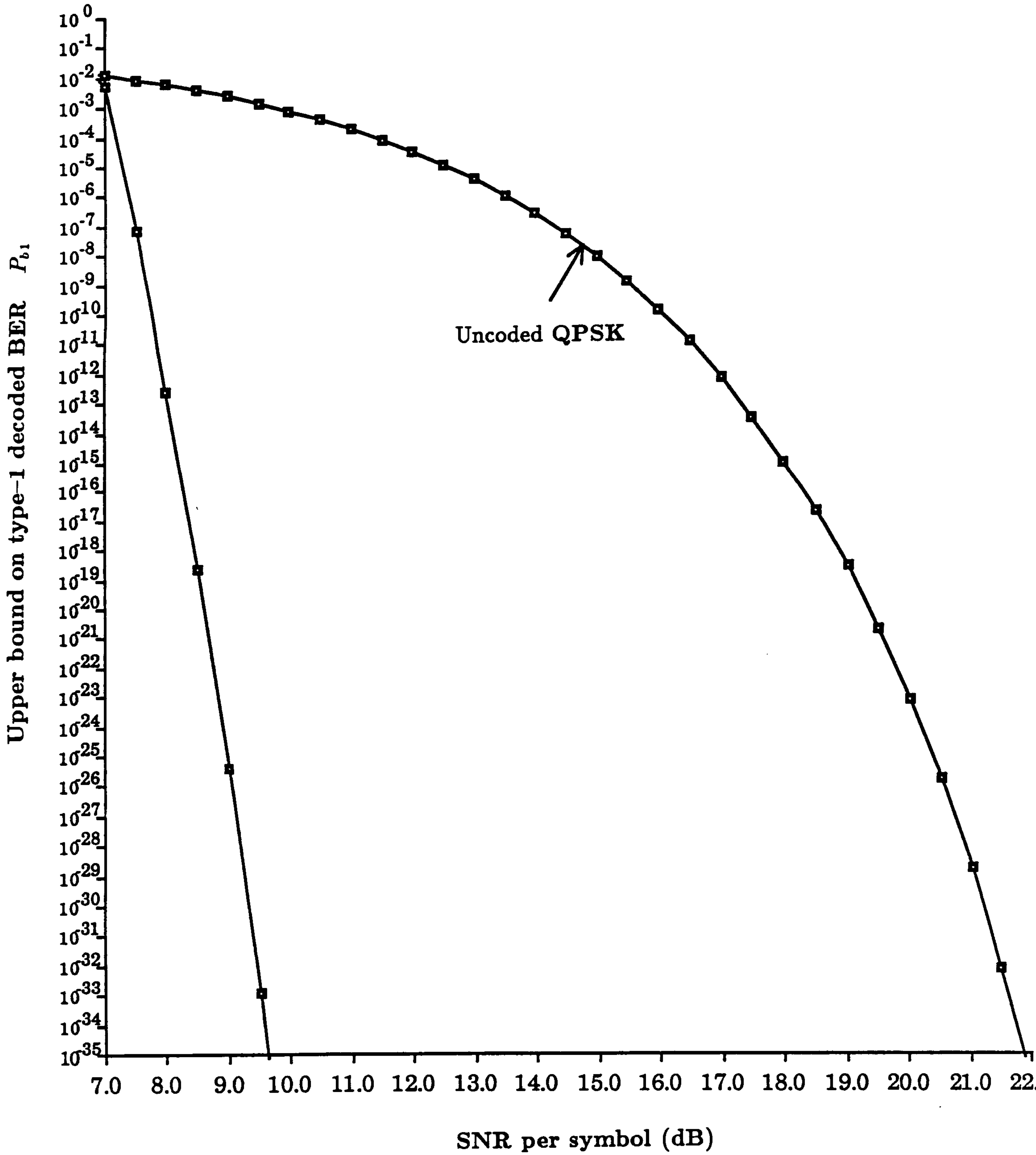


Figure 7 Type–1 bit–error performance of the concatenated coded modulation scheme with the (255,223) RS outer code and the 4–state 8–PSK block inner code

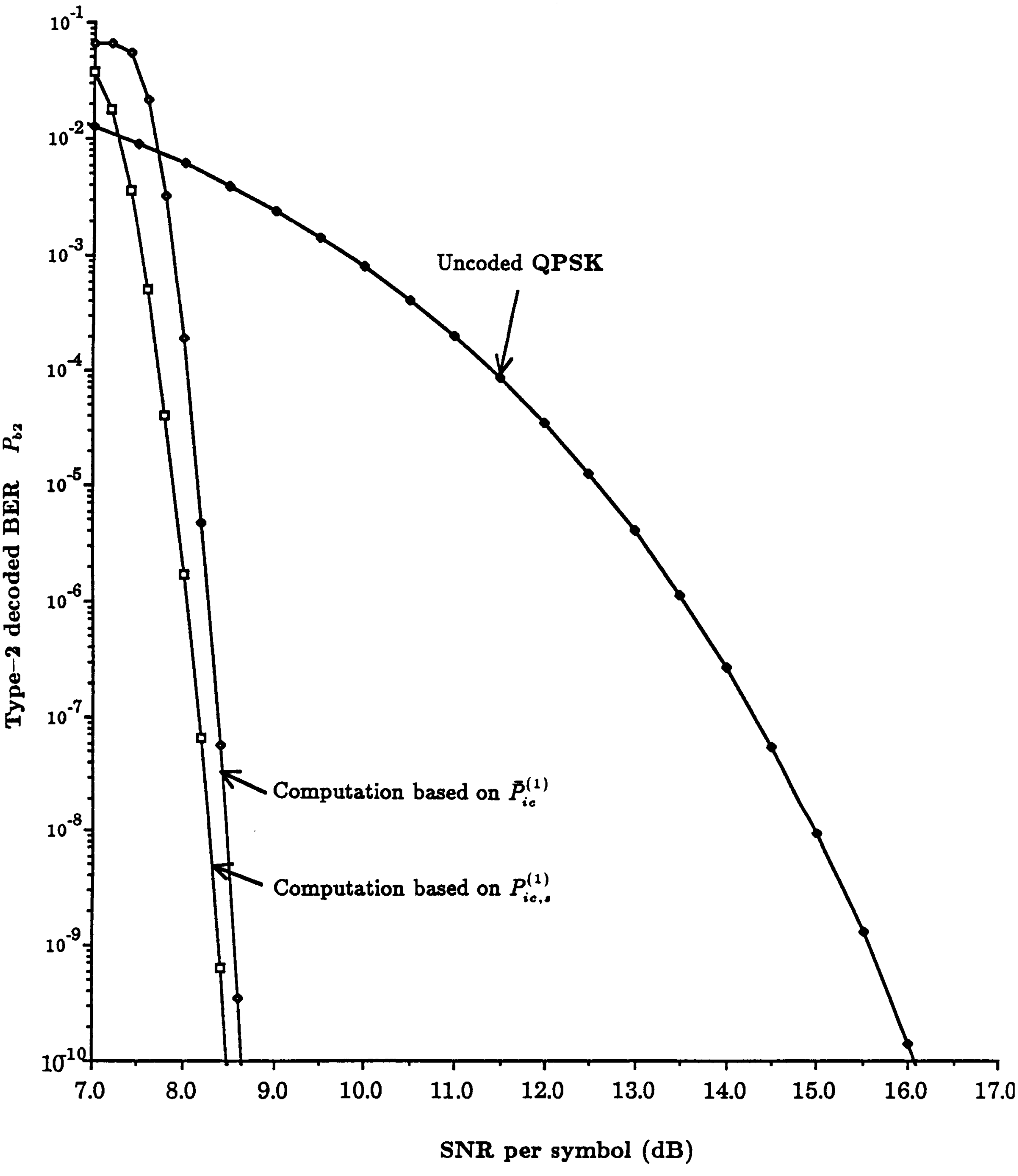


Figure 8 Type–2 bit error performance of the concatenated coded modulation scheme with the (255,223) RS outer code and the 4–state 8–PSK block inner code