UNDETECTED ERROR PROBABILITY AND THROUGHPUT ANALYSIS OF A CONCATENATED CODING SCHEME

Interim Report

to

NASA

Goddard Space Flight Center

Greenbelt, Maryland

Error Control Techniques for Satellite Communications

Grant Number NAG 5-234

Principal Investigator

Daniel J. Costello, Jr.

Department of Electrical & Computer Engineering

Illinois Institute of Technology

Chicago, IL 60616

October 15, 1984

UNDETECTED ERROR PROBABILITY AND THROUGHPUT ANALYSIS OF A CONCATENATED CODING SCHEME

HUIJIE DENG and DANIEL J. COSTELLO, JR.

Department of Electrical & Computer Engineering
Illinois Institute of Technology
Chicago, IL 60616

ABSTRACT

The performance of a proposed concatenated coding scheme for error control on a NASA telecommand system is analyzed. In this scheme, the inner code is a distance-4 Hamming code used for both error correction and error detection. The outer code is a shortened distance-4 Hamming code used only for error detection. Interleaving is assumed between the inner and outer codes. A retransmission is requested if either the inner or outer code detects the presence of errors. Both the undetected error probability and the throughput of the system are analyzed. Results indicate that high throughputs and extremely low undetected error probabilities are achievable using this scheme.

I. INTRODUCTION

Consider a concatenated coding scheme for error control on a binary symmetric channel, called the inner channel. The bit error rate (BER) of the channel is called the inner BER, and is denoted by ε_i. Two linear block codes, C_f and C_b, are used. The inner code C_f, called the frame code, is an (n,k) systematic binary block code with minimum distance d_f. The frame code is designed to correct t or fewer errors and simultaneously detect $\lambda (\lambda > t)$ or fewer errors, where $t+\lambda+1 \leq d_f$ [1]. The outer code is an (n_b, k_b) binary block code with

$$n_b = mk, \tag{1}$$

where m, a positive integer, is the number of frames. The outer code is designed for error detection only.

The encoding of the concatenated code is achieved in two stages (see Figure 1). A message of k_b bits is first encoded into a codeword of n_b bits in the outer code C_b. Then this codeword is interleaved to depth m. After interleaving, the n_b-bit block is divided into m k-bit segments. Each k-bit segment is encoded into an n-bit word in the frame code C_f. This n-bit word is called a frame. The two dimensional block format is depicted in Figure 2.

The decoding consists of error correction and error detection on each frame and error detection on the m decoded k-bit segments. When a frame in a block is received, it is first decoded based on the frame code C_f. The n-k parity bits are then removed from the decoded frame. If there are t or fewer transmission errors in a received frame, the errors will be corrected, and the decoded segment is error free. If there are more than t errors in the received frame, the errors will be either detected or undetected. If the errors are detected, the decoder stops decoding immediately and requests a retransmission of the entire block. On the other hand, if the errors in a frame are undetected, the decoded segment will be stored in a buffer and the decoder continues to decode the next frame. After m frames of a block have been decoded, the m k-bit decoded segments are then deinterleaved. Error detection is performed on these deinterleaved m segments based on the outer code C_b. If no errors are detected, the m decoded segments are assumed to be error free, and are accepted by the receiver. If the presence of errors

is detected, the m decoded segments are discarded and the receiver requests a retransmission of the entire block.

The error control scheme described above is actually a combination of forward-error-correction (FEC) and automatic-repeat-request (ARQ), called a hybrid ARQ scheme [1]. In this paper, we analyze the performance of the proposed error control scheme. Specifically, the system reliability and the system throughput are calculated. The system reliability is measured in terms of the probability of undetected error after decoding. The system throughput is determined by the retransmission strategy, which may be one of three basic modes, namely, stop-and-wait, go-back-N, or selective-repeat. Only the selective-repeat retransmission strategy is considered in this paper. A particular scheme, proposed for a NASA telecommand system, is analyzed as an example.

II. PROBABILITY OF UNDETECTED ERROR FOR THE INNER CODE

Let $P_c^{(f)}(\varepsilon_i)$ denote the probability of correct decoding for the frame code. Suppose that a bounded-distance decoding algorithm is employed. Bounded-distance decoding corrects all received n-bit sequences with t or fewer errors. When an n-bit sequence with more than t errors is detected, no attempt is made to correct the errors. Since there are $\binom{n}{i}$ distinct ways in which i errors may occur among n bits,

$$P_c^{(f)}(\varepsilon_i) = \sum_{i=0}^{t} \binom{n}{i} \varepsilon_i^{\,i} (1-\varepsilon_i)^{n-i} \tag{2}$$

for bounded-distance decoding.

For a code word $\overline{v}$ in the frame code C_f, let $w(\overline{v})$ denote the Hamming weight of $\overline{v}$. If a decoded frame contains an undetectable error pattern, this error pattern must be a nonzero codeword in C_f. Let $\overline{e}_0$ be a nonzero error pattern after decoding. The probability $P_f(w, \varepsilon_i)$ that a decoded frame contains a nonzero error pattern $\overline{e}_0$ after decoding is given by [2-4]

$$P_f(w, \varepsilon_i) = \sum_{i=0}^{t} \sum_{j=0}^{\min(t-i,\,n-w)} \binom{w}{i} \binom{n-w}{j} \varepsilon_i^{\,w-i+j} (1-\varepsilon_i)^{n-w+i-j}, \tag{3}$$

where $w = w(\overline{e}_0)$, and ε_i is the BER of the inner channel. If $\varepsilon_i \ll \frac{1}{n}$, then

$$P_f(w, \varepsilon_i) \approx \binom{w}{t} \varepsilon_i^{\,w-t} (1-\varepsilon_i)^{n-w+t}. \tag{4}$$

Let $P_{ud}^{(f)}(\varepsilon_i)$ denote the probability of undetected error for the frame code. Let $\{A_w^{(f)},\ d_f \le w \le n\}$ be the weight distribution of C_f. It follows from (3) and (4) that

$$P_{ud}^{(f)}(\varepsilon_i) = \sum_{w=d_f}^{n} A_w^{(f)} P_f(w, \varepsilon_i) \tag{5}$$

$$P_{ud}^{(f)}(\varepsilon_i) \approx A_{d_f}^{(f)} P_f(d_f, \varepsilon_i)$$

$$\approx A_{d_f}^{(f)} \binom{d_f}{t} \varepsilon_i^{\,d_f-t} (1-\varepsilon_i)^{n-d_f+t}, \quad \text{for} \quad \varepsilon_i \ll \frac{1}{n}. \tag{6}$$

Now consider any one of the m frames, say the j^{th} frame. If the decoded frame contains undetected errors, the BER ε_a after decoding is given by

$$\varepsilon_a = \frac{1}{n} \sum_{w=d_f}^{n} A_w^{(f)} w \, P_f(w, \varepsilon_i). \tag{7}$$

For $\varepsilon_i << \frac{1}{n}$, then

$$\varepsilon_a \approx \frac{1}{n} d_f A_{d_f}^{(f)} P_f(d_f, \varepsilon_i) \approx \frac{d_f}{n} A_{d_f}^{(f)} \binom{d_f}{t} \varepsilon_i^{d_f - t} (1-\varepsilon_i)^{n-d_f+t}, \tag{8}$$

will be a good approximation to ε_a. Let E be defined as the event that a frame contains undetected errors. Now let $\varepsilon_{a/E}$ denote the BER embedded in a decoded frame conditioned on the occurrence of event E. It follows from (7) that

$$\varepsilon_{a/E} = \varepsilon_a / \Pr\{E\} \tag{9}$$

$$= \varepsilon_a / P_{ud}^{(f)}(\varepsilon_i). \tag{10}$$

For $\varepsilon_i << \frac{1}{n}$, substituting (6) and (8) into (10) yields

$$\varepsilon_{a/E} \approx \frac{\frac{1}{n} d_f A_{d_f}^{(f)} P_f(d_f, \varepsilon_i)}{A_{d_f}^{(f)} P_f(d_f, \varepsilon_i)} = \frac{d_f}{n}. \tag{11}$$

Now define S to be a random variable such that when h of the m frames contain undetected errors, and the remaining m-h frames are decoded correctly, S = h, h = 0, 1, 2, ..., m. It follows from (2) and (5) that

$$\Pr\{S=h\} = \binom{m}{h} \cdot [P_{ud}^{(f)}(\varepsilon_i)]^h \cdot [P_c^{(f)}(\varepsilon_i)]^{m-h}. \tag{12}$$

Note that (12) is not a binomial distribution because $P_{ud}^{(f)}(\varepsilon_i) + P_c^{(f)}(\varepsilon_i) < 1$, i.e., some received sequences with more than t errors are detected by the frame code.

After deinterleaving of the m decoded segments (with the n-k parity bits removed from each frame), the BER embedded in the n_b-bit block, conditioned on S=h, is given by

$$\varepsilon_0(h) = \varepsilon_{a/E} \cdot \frac{h}{m}, \quad h = 0, 1, 2, \ldots, m. \tag{13}$$

We call the channel specified by (12) and (13) the outer channel, and it is depicted in Figure 3. Note that $\varepsilon_0(0) = 0$. This channel can be viewed as a block interference (BI) channel, as described in [5]. Δ_h, h = 0, 1, ..., m, is called the h-th component channel of the BI channel. Each block of n_b bits (n_b is the length of the outer code) is transmitted over one of the m component channels. The random variable S determines which component channel is used to transmit a given n_b-bit block.

III. PERFORMANCE OF THE CONCATENATED CODE

Let $\{A_i^{(b)}, d_b \le i \le n_b\}$ be the weight distribution of the outer code, where d_b is the minimum distance of C_b. Let $P_{ud}^{(b)}(\varepsilon)$ be the probability of undetected error for the outer code C_b. If the n_b-bit block is transmitted over the h-th component channel Δ_h of the outer channel, it follows from (13) that

$$P_{ud}^{(b)}(\varepsilon_0(h)) = \sum_{i=d_b}^{n_b} A_i^{(b)} (\varepsilon_0(h))^i (1-\varepsilon_0(h))^{n_b - i}. \tag{14}$$

Let $P_{ud}(\varepsilon_i)$ be the average probability of undetected error of the con-

catenated code. From (12) and (14) we obtain

$$P_{ud}(\varepsilon_i) = \sum_{h=0}^{m} \Pr\{S=h\} P_{ud}^{(b)}(\varepsilon_0(h))$$

$$= \sum_{h=1}^{m} \{\binom{m}{h}[P_{ud}^{(f)}(\varepsilon_i)]^h \, [P_c^{(f)}(\varepsilon_i)]^{m-h} \cdot \sum_{i=d_b}^{n_b} A_i^{(b)} (\varepsilon_0(h))^i (1-\varepsilon_0(h))^{n_b-i}\}, \quad (15)$$

where $P_c^{(f)}(\varepsilon_i)$ and $P_{ud}^{(f)}(\varepsilon_i)$ are given by (2) and (5), respectively.

Suppose that the selective-repeat ARQ scheme is used as the retransmission strategy. The specific manner in which the receiver signals to the transmitter for a retransmission will not be considered. It will be assumed, however, that this backward signal is error-free, and that repeated retransmissions of a block are possible. Thus for the concatenated code, let $P_{ud}(\varepsilon_i)$, $P_r(\varepsilon_i)$, and $P_c(\varepsilon_i)$ denote the probabilities of an undetected error, of a block retransmission, and of correct decoding, respectively. Obviously,

$$P_{ud}(\varepsilon_i) + P_r(\varepsilon_i) + P_c(\varepsilon_i) = 1. \quad (16)$$

Then the throughput of the system is [1]

$$\eta = \frac{k}{n} \cdot \frac{k_b}{n_b} (1-P_r(\varepsilon_i)) = \frac{k}{n} \cdot \frac{k_b}{n_b} \cdot (P_{ud}(\varepsilon_i) + P_c(\varepsilon_i)). \quad (17)$$

Note that a transmitted block will be received correctly if and only if all m frames are decoded correctly. Therefore, the probability of accepting a correct block is given by

$$P_c(\varepsilon_i) = [P_c^{(f)}(\varepsilon_i)]^m = [\sum_{i=0}^{t} \binom{n}{i} \varepsilon_i^{\,i} (1-\varepsilon_i)^{n-i}]^m. \quad (18)$$

Because $P_{ud}(\varepsilon_i) \ll P_c(\varepsilon_i)$, it follows from (17) and (18) that

$$\eta \approx \frac{k}{n} \cdot \frac{k_b}{n_b} \cdot [\sum_{i=0}^{t} \binom{n}{i} \varepsilon_i^{\,i} (1-\varepsilon_i)^{n-i}]^m. \quad (19)$$

It can easily be seen that η increases monotonically as t increases, but that for small ε_i, η is only a weakly increasing function of t.

In order to see the relationship between t and $P_{ud}(\varepsilon_i)$, from (15) we have

$$P_{ud}(\varepsilon_i) \approx m \cdot P_{ud}^{(f)}(\varepsilon_i) \cdot [P_c^{(f)}(\varepsilon_i)]^{m-1} \cdot \{\sum_{i=d_b}^{n_b} A_i^{(b)} (\varepsilon_0(1))^i (1-\varepsilon_0(1))^{n_b-i}\},$$

$$\text{for} \quad \varepsilon_i \ll \frac{1}{n}. \quad (20)$$

Using (6), (11), and (13), $P_{ud}(\varepsilon_i)$ can be further approximated as

$$P_{ud}(\varepsilon_i) \approx K \cdot \binom{d_f}{t} \varepsilon_i^{\,d_f-t} (1-\varepsilon_i)^{n-d_f+t} [P_c^{(f)}(\varepsilon_i)]^{m-1}, \quad (21)$$

where

$$K = m \cdot A_{d_f}^{(f)} \{ \sum_{i=d_b}^{n_b} A_i^{(b)} \left(\frac{d_f}{m \cdot n} \right)^i \left(1 - \frac{d_f}{m \cdot n} \right)^{n_b - i} \}$$

is a constant which is independent of t. Let Q(t) denote the right hand side of (21),

$$\frac{Q(t+1)}{Q(t)} \approx \frac{(d_f - t)}{(t+1)} \cdot \frac{1}{\varepsilon_i} >> n, \quad \text{for} \quad \varepsilon_i << \frac{1}{n}. \tag{22}$$

That is, for $\varepsilon_i << 1/n$, when t increases by 1, $P_{ud}(\varepsilon_i)$, the probability of undetected error, will increase by approximately ε_i^{-1}. Thus $P_{ud}(\varepsilon_i)$ is a strongly increasing function of t. For this reason, a large value of t is not desirable in such a system.

IV. EXAMPLE

We now consider a particular scheme proposed for a NASA telecommand system. The frame code C_f is a distance-4 Hamming code with generator polynomial

$$g(x) = (x+1)(x^6+x+1) = x^7 + x^6 + x^2 + 1,$$

where $x^6 + x + 1$ is a primitive polynomial of degree 6. The natural length of this code is 63. This code is used for single error correction, and is also used to detect all error patterns of double weight and some higher odd weight error patterns. The outer code is a distance-4 shortened Hamming code with generator polynomial

$$\begin{aligned} g(x) &= (x+1)(x^{15} + x^{14} + x^{13} + x^{12} + x^4 + x^3 + x^2 + x+1) \\ &= x^{16} + x^{12} + x^5 + 1, \end{aligned}$$

where $x^{15} + x^{14} + x^{13} + x^{12} + x^4 + x^3 + x^2 + x+1$ is a primitive polynomial of degree 15. This code is the X.25 standard for packet-switched data networks [6]. The natural length of this code is $2^{15} - 1 = 32{,}767$. In this example, a shortened code of maximum length 3,584 bits is considered. This code is used for error detection only.

We assume that the number of information bytes in a frame is between 3 and 7, that is, the inner code can also be shortened. The number of frames in a block is between 4 and 64.

To obtain a precise result for $P_{ud}(\varepsilon_i)$, a computer program was written to help determine the reliability of the proposed concatenated coding scheme. We found that if only one frame contains a weight 4 undetected error pattern, then this error pattern can always be detected by the outer code. Thus (15) can be modified as follows:

$$\begin{aligned} P_{ud}(\varepsilon_i) = {} & \binom{m}{1} \overline{P}_{ud}^{(f)}(\varepsilon_i) P_c^{(f)}(\varepsilon_i)^{m-1} \\ & \cdot \sum_{i=d_b}^{n_b} A_i^{(b)} (\overline{\varepsilon}_0(1))^i (1-\overline{\varepsilon}_0(1))^{n_b - i} \\ & + \sum_{h=2}^{m} \{ \binom{m}{h} [P_{ud}^{(f)}(\varepsilon_i)]^h [P_c^{(f)}(\varepsilon_i)]^{m-h} \\ & \cdot \sum_{i=d_b}^{n_b} A_i^{(b)} (\varepsilon_0(h))^i (1-\varepsilon_0(h))^{n_b - i} \}, \end{aligned} \tag{23}$$

where

$$\overline{P}_{ud}^{(f)}(\varepsilon_i) = \sum_{w=d_f+1}^{n} A_w^{(f)} P_f(w, \varepsilon_i), \tag{24}$$

and

$$\overline{\varepsilon}_0(1) = \frac{\frac{1}{n} \sum_{w=d_f+1}^{n} A_w^{(f)} w \, P_f(w, \varepsilon_i)}{\overline{P}_{ud}^{(f)}(\varepsilon_i)} \cdot \frac{1}{m} . \tag{25}$$

The weight distributions of the frame code are listed in Table 1. Results for $P_{ud}(\varepsilon_i)$, based on (23) for various values of ε_i, are given in Table 2, where we have used the method in [7] to obtain

$$P_{ud}^{(b)}(\varepsilon_0(h)) = \sum_{i=d_b}^{n_b} A_i^{(b)} (\varepsilon_0(h))^i (1-\varepsilon_0(h))^{n_b-i}.$$

$1-P_c(\varepsilon_i)$, the probability of not decoding correctly, and η, the system throughput, are given in Tables 3 and 4, respectively, for $\varepsilon_i = 10^{-4}$. Clearly, the proposed system is capable of achieving high throughput and low undetected error probability over a wide range of system parameters.

REFERENCES

[1] S. Lin and D.J. Costello, Jr., Error Control Coding: Fundamentals and Applications, Prentice-Hall, New Jersey, 1983.

[2] E.R. Berlekamp, Algebraic Coding Theory, McGraw-Hill, New York, 1968.

[3] J. MacWilliams, "A Theorem on the Distribution of Weights in a Systematic Code", Bell Sys. Tech. J., Vol. 42, pp. 79-94, 1963.

[4] Z. McHuntoon and A.M. Michelson, "On the Computation of the Probability of Post-Decoding Error Events for Block Codes", IEEE Trans. on Inform. Theory, IT-23, No. 3, May 1977, pp. 399-403.

[5] R.J. McEliece and W.E. Stark, "Channels with Block Interference", IEEE Trans. on Inform. Theory, IT-30, pp. 44-53, Jan. 1984.

[6] CCITT: Recommendation X.25, "Interface Between Data Terminal Equipment for Terminals Operating in Packet Mode on Public Data Networks", with Plenary Assembly, Doc. No. 7, Geneva, 1980.

[7] T. Fujiwara, T. Kasami, A. Kitai, and S. Lin, "On the Undetected Error Probability for Shortened Hamming Codes", Technical Report, NASA Grant NAG 5-234, August 1983.

ACKNOWLEDGEMENT

This work was supported by NASA under Grant NAG 5-234.

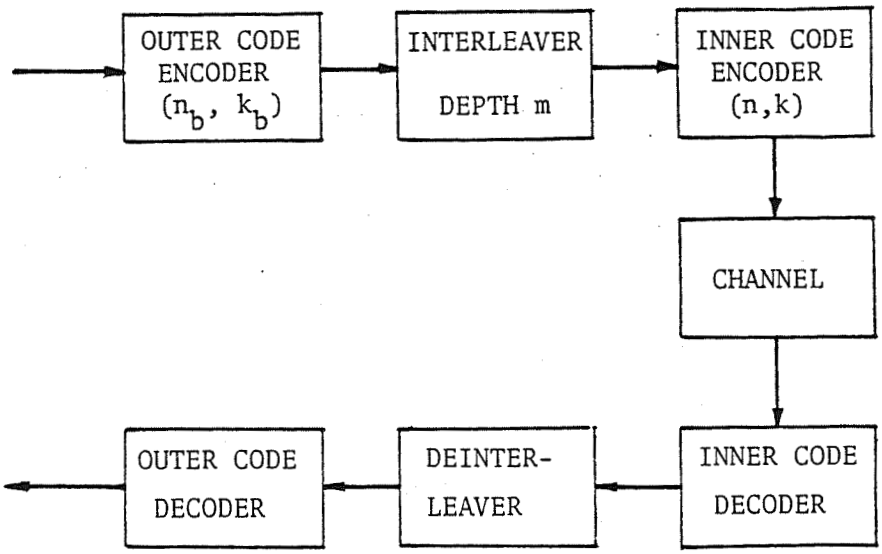

Figure 1. A concatenated coding system.

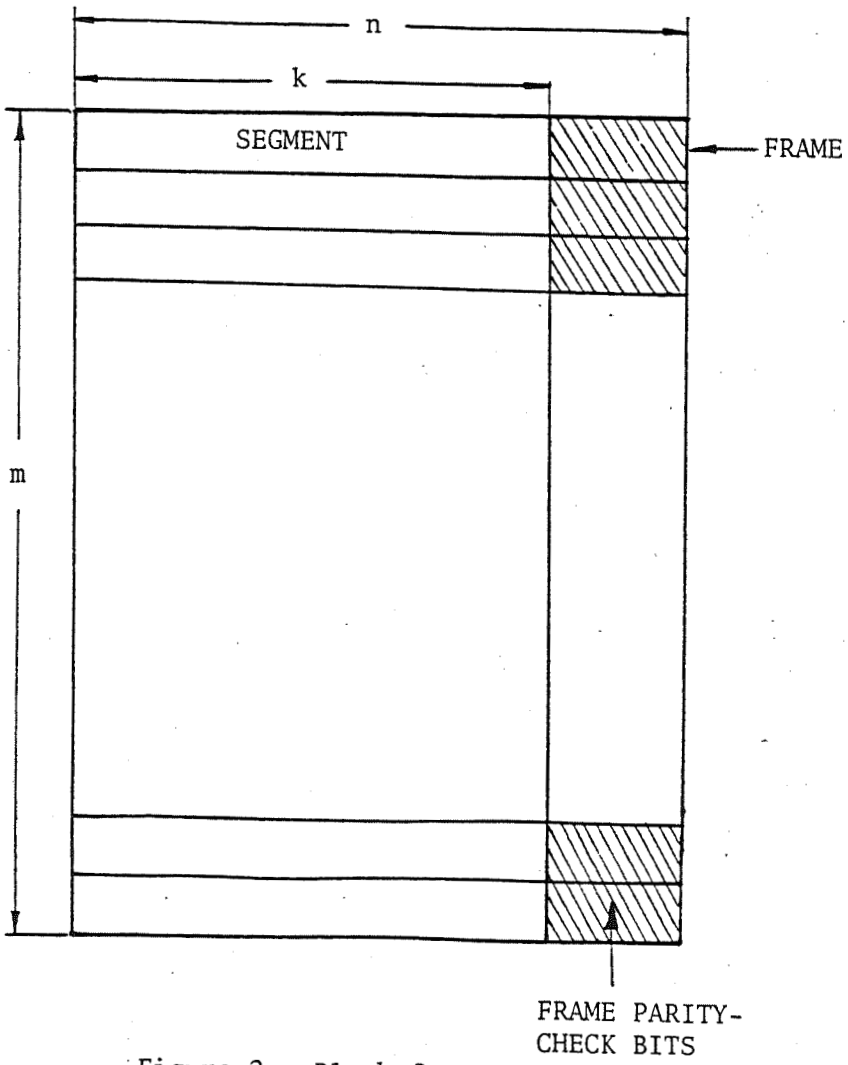

Figure 2. Block format.

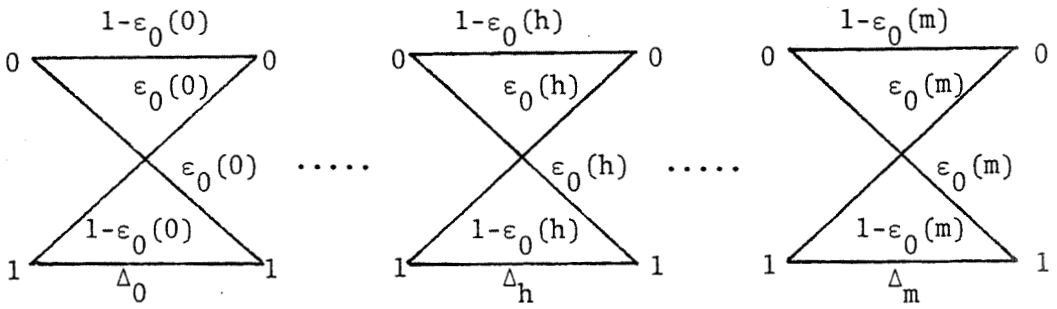

Figure 3. The outer channel.

Table 1. The inner code weight distribution.

n	IB	$A_4^{(f)}$	$A_6^{(f)}$	$A_8^{(f)}$	$A_{10}^{(f)}$
63	7	9765	1.057×10^6	6.048×10^7	1.997×10^9
55	6	5592	4.530×10^5	1.902×10^7	4.570×10^8
47	5	2927	1.678×10^5	4.913×10^6	8.091×10^7
39	4	1343	5.098×10^4	9.613×10^5	9.934×10^6
31	3	505	1.151×10^4	1.233×10^5	6.930×10^5

n : The inner code length

IB : The number of information bytes in the inner code

Table 2(A). $P_{ud}(\varepsilon_i)$, the probability of undetected error, for $\varepsilon_i=10^{-4}$.

m \ IB	3	4	5	6	7
4	4.25×10^{-20}	1.87×10^{-19}	6.14×10^{-19}	1.70×10^{-18}	4.01×10^{-18}
14	1.53×10^{-19}	6.93×10^{-19}	2.34×10^{-18}	6.50×10^{-18}	1.56×10^{-17}
24	2.70×10^{-19}	1.24×10^{-18}	4.24×10^{-18}	1.21×10^{-17}	2.97×10^{-17}
34	3.93×10^{-19}	1.83×10^{-18}	6.40×10^{-18}	1.84×10^{-17}	4.60×10^{-17}
44	5.22×10^{-19}	2.47×10^{-18}	8.74×10^{-18}	2.55×10^{-17}	6.46×10^{-17}
54	6.58×10^{-19}	3.15×10^{-18}	1.13×10^{-17}	3.33×10^{-17}	8.56×10^{-17}
64	8.00×10^{-19}	3.87×10^{-18}	1.41×10^{-17}	4.19×10^{-17}	1.09×10^{-16}

Table 2(B). $P_{ud}(\varepsilon_i)$, the probability of undetected error, for $\varepsilon_i=10^{-5}$.

m \ IB	3	4	5	6	7
4	4.23×10^{-25}	1.87×10^{-24}	6.16×10^{-24}	1.66×10^{-23}	3.89×10^{-23}
14	1.48×10^{-24}	6.57×10^{-24}	2.21×10^{-23}	5.88×10^{-23}	1.38×10^{-22}
24	2.55×10^{-24}	1.13×10^{-23}	3.74×10^{-23}	1.02×10^{-22}	2.39×10^{-22}
34	3.62×10^{-24}	1.62×10^{-23}	5.34×10^{-23}	1.45×10^{-22}	3.42×10^{-22}
44	4.69×10^{-24}	2.10×10^{-23}	6.96×10^{-23}	1.90×10^{-22}	4.48×10^{-22}
54	5.78×10^{-24}	2.58×10^{-23}	8.59×10^{-23}	2.39×10^{-22}	5.56×10^{-22}
64	6.87×10^{-24}	3.08×10^{-23}	1.03×10^{-22}	2.81×10^{-22}	6.22×10^{-22}

m : The number of frames in a block

IB : The number of information bytes in a frame.

Table 3. $1-P_c(\varepsilon_i)$, The probability of not decoding correctly, for $\varepsilon_i=10^{-4}$.

$1-P_c(\varepsilon_i)$ IB / m	3	4	5	6	7
4	1.86×10^{-5}	2.96×10^{-5}	4.31×10^{-5}	5.92×10^{-5}	7.78×10^{-5}
14	6.50×10^{-5}	1.04×10^{-4}	1.51×10^{-4}	2.07×10^{-4}	2.72×10^{-4}
24	1.11×10^{-4}	1.77×10^{-4}	2.59×10^{-4}	3.55×10^{-4}	4.67×10^{-4}
34	1.58×10^{-4}	2.51×10^{-4}	3.66×10^{-4}	5.03×10^{-4}	6.61×10^{-4}
44	2.04×10^{-4}	3.25×10^{-4}	4.74×10^{-4}	6.51×10^{-4}	8.56×10^{-4}
54	2.51×10^{-4}	3.99×10^{-4}	5.82×10^{-4}	7.98×10^{-4}	1.05×10^{-3}
64	2.97×10^{-4}	4.73×10^{-4}	6.89×10^{-4}	9.46×10^{-4}	1.24×10^{-3}

Table 4. η, The system throughput, for $\varepsilon_i=10^{-4}$.

IB / m	3	4	5	6	7
4	0.64515	0.71793	0.76591	0.79995	0.82533
14	0.73728	0.79113	0.82662	0.85177	0.87051
24	0.75260	0.80328	0.83666	0.86020	0.87790
34	0.75889	0.80824	0.84074	0.86360	0.88084
44	0.76231	0.81093	0.84293	0.86555	0.88236
54	0.76444	0.81259	0.84427	0.86665	0.88326
64	0.76590	0.81372	0.84494	0.86736	0.88382

m : The number of frames in a block

IB : The number of information bytes in a frame.