REPORTS IN INFORMATICS

ISSN 0333-3590

Large Binary Codes for Error Detection

Fang-Wei Fu and Torleiv Kløve

REPORT NO 315 **February 2006**

Department of Informatics
UNIVERSITY OF BERGEN
Bergen, Norway

This report has URL http://www.ii.uib.no/publikasjoner/texrap/ps/2006-315.ps

Reports in Informatics from Department of Informatics, University of Bergen, Norway, is available at http://www.ii.uib.no/publikasjoner/texrap/.

Requests for paper copies of this report can be sent to:

Department of Informatics, University of Bergen, Høyteknologisenteret,
P.O. Box 7800, N-5020 Bergen, Norway

Large Binary Codes for Error Detection

Fang-Wei Fu*, Torleiv Kløve†

Abstract

In this paper, the undetected error probability for large binary codes is studied. It is shown that if the size of the code is sufficiently large (for given length), then the code is good for error detection (in the technical sense).

Index Terms - Undetected error probability, error detection, binary codes, good codes.

1 Introduction

In automatic-repeat-request (ARQ) error-control systems, the undetected error probability of an error-detecting code is one of the most important performance characteristics. There are a number of papers dedicated to examine the error detection capability for some well known classes of linear codes, and a number of lower and upper bounds for the undetected error probability are known. For a general introduction to the theory of error detecting codes, we refer the readers to [3] and its references.

Let $V_n = \{0,1\}^n$ be the n-dimensional vector space over the binary field $\{0,1\}$. The (Hamming) distance $d_H(\mathbf{a},\mathbf{b})$ between two vectors $\mathbf{a}$ and $\mathbf{b}$ is the number of components where they differ, and the (Hamming) weight $w_H(\mathbf{x})$ of a vector $\mathbf{x}$ is the number of nonzero components in $\mathbf{x}$.

A binary (n, M) code is a subset of V_n with cardinality M. A linear binary $[n, k]$ code is a subspace of V_n of dimension k.

For any codeword $\mathbf{a}$ of a code C, the distance distribution from $\mathbf{a}$ is defined by

$$A_i(\mathbf{a}) = \Big|\{\mathbf{b} \in C \mid d_H(\mathbf{a},\mathbf{b}) = i\}\Big|$$

for $i = 0, 1, \ldots, n$.

The (average) distance distribution of an (n, M) code C is defined as

$$A_i = A_i(C) = \frac{1}{M}\sum_{\mathbf{a}\in C} A_i(\mathbf{a}),$$

for $i = 0, 1, \ldots, n$. We note that $A_i(\mathbf{a}) \le \binom{n}{i}$ and so

$$A_i \le \binom{n}{i}. \tag{1}$$

*Fang-Wei Fu is with the Temasek Laboratories, National University Of Singapore, 10 Kent Ridge Crescent, Singapore 119260 (on leave from the Department of Mathematics, Nankai University, Tianjin 300071, China). E-mail: tslfufw@nus.edu.sg

†Torleiv Kløve is with the Department of Informatics, University of Bergen, Norway). E-mail: Torleiv.Klove@ii.uib.no

If the code C is used for error detection on a binary symmetric channel with symbol error probability p, the undetected error probability is given by (see e.g. [3, p.38])

$$\begin{aligned} P_{ue}(C,p) &= \frac{1}{M}\sum_{\substack{\mathbf{a},\mathbf{b}\in C,\\ \mathbf{a}\neq\mathbf{b}}} p^{d_H(\mathbf{a},\mathbf{b})}(1-p)^{n-d_H(\mathbf{a},\mathbf{b})} \\ &= \sum_{i=1}^{n} A_i p^i (1-p)^{n-i}. \end{aligned} \tag{2}$$

It is established terminology (see e.g. [3]) to say that an (n, M) code is *good* (for error detection) if

$$P_{ue}(C,p) \leq P_{ue}(C,1/2) = \frac{M-1}{2^n} \tag{3}$$

for all $p \in [0, 1/2]$. It is appropriate in our context to emphasize that this is a relative, not absolute, description of the code's performance for error detection. This is well illustrated by the $(n, 2^n)$ code C consisting of all binary vectors of length n. For this code

$$P_{ue}(C,p) = 1-(1-p)^n;$$

hence (3) is satisfied and so the code is good (in the technical sense). On the other hand, this code can not detect any errors! In general, one measure for how suitable a code is for error detection is

$$P_{max}(C) = \max_{0\leq p\leq 1/2} P_{ue}(C,p).$$

Clearly, $P_{max}(C) \geq P_{ue}(C,1/2)$ and the code is good if and only if $P_{max}(C) = P_{ue}(C,1/2)$. Hence a good code is better (by this measure) than a not good code *of the same size.*

For a binary code C of length n, let $\overline{C} = V_n \setminus C$, be the complementary code. In [2] we proved the following result.

Lemma 1 *If C is a binary (n, M) code, then*

$$P_{ue}(\overline{C},p) = \frac{MP_{ue}(C,p) + (2^n - 2M)(1-(1-p)^n)}{2^n - M}. \tag{4}$$

Clearly, using Lemma 1, any upper bound on $P_{ue}(C,p)$ gives an upper bound on $P_{ue}(\overline{C},p)$. In this paper we discuss such upper bounds. In particular, we show that if the size of C is sufficiently large (for given n), then C is good for error detection. More precisely, define $\Omega(n)$ to be the largest integer such that all code of length n and size $\geq 2^n - \Omega(n)$ are good. We give upper and lower bounds on $\Omega(n)$, both of the order $2^{n/2}$.

2 Linear codes

Define $K(n)$ to be the largest integer such that $\overline{C}$ is good for all linear $[n, k]$ codes C with $k \leq K(n)$. In [1] we determined $K(n)$:

Theorem 1 *For $n \geq 6$ we have*

$$K(n) = \lfloor (n+1)/2 \rfloor.$$

From the definition of $\Omega(n)$ it follows that $\Omega(n) < 2^{K(n)+1}$. Hence Theorem 1 implies the following upper bound.

Corollary 1 *For $n \geq 6$ we have*

$$\frac{\Omega(n)}{2^{n/2}} < 2^{\lfloor (n+1)/2 \rfloor + 1 - n/2} = \begin{cases} 2 & \text{if } n \text{ is even} \\ 2\sqrt{2} & \text{if } n \text{ is odd.} \end{cases}$$

3 General codes

For $1 \leq r \leq n$, define

$$M_r = \sum_{i=0}^{r} \binom{n}{i}.$$

Theorem 2 *If C is an (n, M) code, where $n > 2r + 2$ and*

$$M_r \leq M \leq \mu_r = \frac{b_r + \sqrt{4n(n-2r-2)2^n + b_r^2}}{2(n-2r-2)},$$

where

$$b_r = M_r(n-2r-2) - (n-r)\binom{n}{r} - n, \tag{5}$$

then $\overline{C}$ is good for error detection.

We break the proof of Theorem 2 down into a series of simple lemmas.

Lemma 2 *If $0 \leq d \leq d' \leq n$ and $p \in (0, 1/2)$, then*

$$p^d(1-p)^{n-d} > p^{d'}(1-p)^{n-d'}.$$

Proof: For any d, $1 \leq d < n$, and any $p \in (0, 1/2)$ we have

$$p^d(1-p)^{n-d} - p^{d+1}(1-p)^{n-d-1} = p^d(1-p)^{n-d-1}(1-2p) > 0.$$

The lemma follows by induction. QED

Lemma 3 *If $M \geq M_r$, then*

$$P_{ue}(C, p) \leq \sum_{i=1}^{r} \binom{n}{i} p^i(1-p)^{n-i} + (M - M_r)p^{r+1}(1-p)^{n-r-1}.$$

Proof: The lemma follows directly from (1) and Lemma 2. QED

Lemma 4 *If $M \geq M_r$, then*

$$(2^n - M)P_{ue}(\overline{C}, p) \leq f_M(p)$$

where

$$\begin{aligned} f_M(p) &= M\sum_{i=1}^{r} \binom{n}{i} p^i(1-p)^{n-i} \\ &\quad + M(M - M_r)p^{r+1}(1-p)^{n-r-1} \\ &\quad + (2^n - 2M)(1 - (1-p)^n). \end{aligned}$$

Proof: The lemma follows directly from Lemmas 1 and 3. QED

We will show that under the conditions of Theorem 2, $f_M(p)$ is non-decreasing on $[0, 1/2]$.

Lemma 5 *If $f'_M(1/2) \geq 0$, then $f'_M(p) \geq 0$ for all $p \in [0, 1/2]$.*

Proof: We have

$$\begin{aligned}
& f'_M(p) \\
&= M\sum_{i=1}^{r}\binom{n}{i} i p^{i-1}(1-p)^{n-i} \\
&\quad -M\sum_{i=1}^{r}\binom{n}{i}(n-i)p^{i}(1-p)^{n-i-1} \\
&\quad +M(M-M_r)(r+1)p^r(1-p)^{n-r-1} \\
&\quad -M(M-M_r)(n-r-1)p^{r+1}(1-p)^{n-r-2} \\
&\quad +(2^n-2M)n(1-p)^{n-1} \\
&= Mn(1-p)^{n-1} - M\binom{n}{r}(n-r)p^r(1-p)^{n-r-1} \\
&\quad +M(M-M_r)(r+1)p^r(1-p)^{n-r-1} \\
&\quad -M(M-M_r)(n-r-1)p^{r+1}(1-p)^{n-r-2} \\
&\quad +(2^n-2M)n(1-p)^{n-1} \\
&= (2^n-M)n(1-p)^{n-1} \\
&\quad +M\Big\{(M-M_r)(r+1) - \binom{n}{r}(n-r)\Big\}p^r(1-p)^{n-r-1} \\
&\quad -M(M-M_r)(n-r-1)p^{r+1}(1-p)^{n-r-2} \\
&= (1-p)^{n-1}g_M(x)
\end{aligned}$$

where $x = p/(1-p)$ and

$$\begin{aligned}
& g_M(x) \\
&= (2^n-M)n + M\Big\{(M-M_r)(r+1) - \binom{n}{r}(n-r)\Big\}x^r \\
&\quad -M(M-M_r)(n-r-1)x^{r+1}.
\end{aligned}$$

Hence we have to show that $g_M(x) \geq 0$ for $x \in [0, 1]$ if $g_M(1) \geq 0$. We show that for $x \in [0, 1]$ we in fact have

$$g_M(x) \geq \min\{g_M(0), g_M(1)\}. \tag{6}$$

If $M = M_r$, then $g_M(x) = (2^n - M)n$ for all x and (6) is immediate. Now consider $M > M_r$.

We have

$$\begin{aligned} g'_M(x) &= rM\Big\{(M-M_r)(r+1)-\binom{n}{r}(n-r)\Big\}x^{r-1} \\ &\quad -(r+1)M(M-M_r)(n-r-1)x^r \\ &= Mx^{r-1}\Big\{r(r+1)(M-M_r)-r\binom{n}{r}(n-r) \\ &\quad -(r+1)(M-M_r)(n-r-1)x\Big\}. \end{aligned}$$

Since $(r+1)(M-M_r)(n-r-1)>0$, there are three possibilities, $g'_M(x)>0$ for all $x\in(0,1)$ (in which case $g_M(x)$ is increasing on $[0,1]$), $g'_M(x)<0$ for all $x\in(0,1)$ (in which case $g_M(x)$ is decreasing on $[0,1]$), or $g'_M(x^*)=0$ for a unique $x^*\in(0,1)$ (in which case $g_M(x)$ is increasing for $x\in(0,x^*)$ and then decreasing). In all cases, (6) follows. Since $g_M(0)=(2^n-M)n>0$, $g_M(1)\geq 0$, implies that $g_M(x)\geq 0$ for all $x\in[0,1]$. QED

Lemma 6 *We have $f'_M(1/2)\geq 0$ if and only if*

$$M\leq \mu_r. \tag{7}$$

Proof: In the notation of the proof of the previous lemma, $f'_M(1/2)\geq 0$ if and only if $g_M(1)\geq 0$. Since

$$\begin{aligned} g_M(1) &= (2^n-M)n+M(M-M_r)(r+1) \\ &\quad -M\binom{n}{r}(n-r)-M(M-M_r)(n-r-1), \end{aligned}$$

$f'_M(1/2)\geq 0$ if and only if

$$(n-2r-2)M^2-\Big\{M_r(n-2r-2)-(n-r)\binom{n}{r}-n\Big\}M-2^n n\leq 0,$$

that is

$$(n-2r-2)M^2-b_rM-2^nn\leq 0.$$

Solving for M we get $M\leq\mu_r$. QED

We can now prove Theorem 2. By Lemmas 5 and 6, if $M\leq\mu_r$, then $f_M(p)$ is increasing on $[0,1/2]$. Hence, using Lemma 4, we get

$$\begin{aligned} (2^n-M)P_{ue}(\overline{C},p) &\leq f_M(p) \\ &\leq f_M(1/2)=(2^n-M)P_{ue}(\overline{C},1/2) \end{aligned}$$

for $p\in[0,1/2]$. QED

Using Theorem 2, we can get lower bounds on $\Omega(n)$. The simplest case is $r=0$. We have

$$\begin{aligned} \Omega(n)\geq\mu_0 &= \frac{-2-n+\sqrt{4n(n-2)2^n+(n+2)^2}}{2(n-2)} \\ &> 2^{n/2}\sqrt{\frac{n}{n-2}}-2. \end{aligned}$$

In particular, we have the corollary

Corollary 2 *For $n\geq 3$ we have*

$$\frac{\Omega(n)}{2^{n/2}}>1,$$

This lower bound can be improved using the full strength of Theorem 2. First we give a couple of lemmas. Since we must have $M_r \leq \mu_r$ to get any information from Theorem 2, we consider this condition first.

Lemma 7 *We have $M_r \leq \mu_r$ if and only if*

$$M_r\Big\{(n-r)\binom{n}{r}+n\Big\} \leq n2^n. \tag{8}$$

Proof: $M_r \leq \mu_r$ can be rewritten as

$$2(n-2r-2)M_r \leq b_r + \sqrt{4n(n-2r-2)2^n + b_r^2}$$

which is equivalent to

$$\begin{aligned} & 4(n-2r-2)^2M_r^2 - 4(n-2r-2)M_r b_r + b_r^2 \\ = \;& (2(n-2r-2)M_r - b_r)^2 \\ \leq \;& 4n(n-2r-2)2^n + b_r^2, \end{aligned}$$

which further is equivalent to

$$n2^n \geq M_r((n-2r-2)M_r - b_r) = M_r\Big\{(n-r)\binom{n}{r}+n\Big\}.$$

QED

Lemma 8 *We have $M_r \leq \mu_{r-1}$ if and only if (8) is satisfied.*

Proof: $M_r \leq \mu_{r-1}$ can be rewritten as

$$2(n-2r)M_r \leq b_{r-1} + \sqrt{4n(n-2r)2^n + b_{r-1}^2}$$

which is equivalent to

$$\begin{aligned} & 4(n-2r)^2M_r^2 - 4(n-2r)M_r b_{r-1} + b_{r-1}^2 \\ = \;& (2(n-2r)M_r - b_{r-1})^2 \\ \leq \;& 4n(n-2r)2^n + b_{r-1}^2, \end{aligned}$$

which further is equivalent to

$$\begin{aligned} n2^n \;\geq\; & M_r((n-2r)M_r - b_{r-1}) \\ = \;& M_r\Big\{(n-2r)M_{r-1} + (n-2r)\binom{n}{r} \\ & \quad -(n-2r)M_{r-1} + (n-r+1)\binom{n}{r-1} + n\Big\} \\ = \;& M_r\Big\{(n-r)\binom{n}{r}+n\Big\}. \end{aligned}$$

QED

Lemmas 7 and 8 show in particular that if $\mu_r \geq M_r$, then also $\mu_{r-1} \geq M_r$ (even if $\mu_{r-1} < \mu_r$). Let

$$r(n) = \max\Big\{r\Big|M_r\Big\{(n-r)\binom{n}{r}+n\Big\} \leq n2^n\Big\}.$$

Lemma 9 *If C is an (n, M) code and $M \leq \mu_{r(n)}$, then $\overline{C}$ is good for error detection.*

Proof: by Theorem 2, $\overline{C}$ is good if $M_{r(n)} \le M \le \mu_{r(n)}$. If $M < M_{r(n)}$, then $M_{r-1} \le M < M_r$ for some $r \le r(n)$. By Lemma 8, $\mu_{r-1} \ge M_r > M$ and so $\overline{C}$ is good for error detection by Theorem 2. QED

This lemma immediately gives the following theorem.

Theorem 3 *We have* $\Omega(n) \ge \mu_{r(n)}$.

It is not too hard to show that $\mu_0 < \mu_1 < \cdots < \mu_{r(n)}$; we omit the proof since we do not need it for any of our results. However, it shows that Theorem 3 gives the best lower bound we can obtain from Theorem 2.

We illustrate these results with the following table of values of $\lfloor \mu_r \rfloor$ and M_r for $n = 100$.

r	M_r	$\lfloor \mu_r \rfloor$
1	101	1149116780504808
2	5051	1161277085078223
3	166751	1173831796809154
4	4087976	1186802707617363
5	79375496	1200213332467409
6	1271427896	1214089077286780
7	17278988696	1228457289606556
8	203366882996	1243345741876057
9	2105598691396	1258765269121305
10	19415908147836	1274551874821334
11	161045712791436	1289092885139942
12	1211466763898136	1290126367194585
13	8322009263697336	1198782290724140

We see that $r(100) = 12$. We have $\mu_{r-1} < \mu_r$ for $r \le 12$ (however, $\mu_{12} > \mu_{13}$). Further, we see that $\mu_0 > M_{11}$, $\mu_5 < M_{12}$ and $\mu_6 > M_{12}$. We get

$$\Omega(100) \ge \mu_{12} = 1290126367194585 \approx 1.145862 \cdot 2^{50}.$$

Some further values of μ_0 and $\mu_{r(n)}$ are given in the following table:

n	$r(n)$	$\mu_0 2^{-n/2}$	$\mu_{r(n)} 2^{-n/2}$
200	23	1.005038	1.141152
500	56	1.002006	1.136641
1000	111	1.001001	1.134679
2000	221	1.000500	1.133560

The table indicates that $\mu_{r(n)} 2^{-n/2}$ converges to a limit larger than 1. We will prove that this is indeed the case.

4 Determination of $r(n)$ and $\mu_{r(n)}$

In this section we will find an almost explicit expression for $r(n)$ and explicit approximations for $\mu_{r(n)}$.

First we need some auxiliary results. The binary entropy function is defined on $(0, 1)$ by

$$h(z) = -z \log_2(z) - (1 - z) \log_2(1 - z).$$

Define λ by

$$h(\lambda) = 1/2 \text{ and } 0 < \lambda < 1/2.$$

Then $\lambda \approx 0.11003$. Let

$$\gamma_1 = \log_2\Big(\frac{1 - \lambda}{\lambda}\Big) \approx 3.01589,$$

$$\gamma_2 = \frac{1}{2\ln(2)\lambda(1-\lambda)} \approx 7.36657.$$

Since

$$h'(z) = \log_2\Big(\frac{1-z}{z}\Big) \text{ and } h''(z) = \frac{-1}{\ln(2)z(1-z)}$$

we see that if $0 < u < 1/2 - \lambda$, then

$$\frac{1}{2} + \gamma_1 u - \gamma_2 u^2 < h(\lambda + u) < \frac{1}{2} + \gamma_1 u. \tag{9}$$

The following well known bounds can be found in [4, p.466]:

$$\frac{2^{h(\tau)n}}{\sqrt{2\pi n\tau(1-\tau)}} e^{\frac{-1}{12n\tau(1-\tau)}} < \binom{n}{\tau n} < \frac{2^{h(\tau)n}}{\sqrt{2\pi n\tau(1-\tau)}}. \tag{10}$$

Further, an easy induction shows that for $1 < r < n/5$ we have

$$\frac{n-r+3}{n-2r+3}\binom{n}{r} < M_r < \frac{n-r+2}{n-2r+2}\binom{n}{r}. \tag{11}$$

Let

$$\Delta = \frac{(1-\lambda)}{2\pi\lambda(1-2\lambda)} \approx 1.65056,$$

$$\alpha = \frac{1}{2\gamma_1} \approx 0.16579,$$

$$\beta = \frac{\log_2(\Delta)}{2\gamma_1} \approx 0.11986.$$

$$\theta = \frac{\alpha^2\gamma_2}{\gamma_1} \approx 0.06714.$$

Theorem 4 *For all $n \geq 1$ we have*

$$r(n) \geq \lfloor \lambda n + \alpha\log_2(n) - \beta \rfloor. \tag{12}$$

For any $\gamma > \theta$ there exists an $n(\gamma)$ such that for $n \geq n(\gamma)$ we have

$$r(n) \leq \Big\lfloor \lambda n + \alpha\log_2(n) - \beta + \gamma\frac{(\log_2(n))^2}{n} \Big\rfloor. \tag{13}$$

Proof: Let $r = \lambda n + \alpha\log_2(n) - \beta - \epsilon$ where $0 \leq \epsilon < 1$. For $n \geq 108$, $r \geq \lambda n$ and $r \geq 2$. Let $\lambda' = r/n$. From (9)–(11) we get

$$\begin{aligned}
&\Big\{n + (n-r)\binom{n}{r}\Big\}M_r \\
&< \Big\{n + (n-r)\binom{n}{r}\Big\}\frac{n-r+2}{n-2r+2}\binom{n}{r} \\
&< (n-r)\binom{n}{r}\frac{n-r}{n-2r}\binom{n}{r}^2 \\
&= n\frac{(1-\lambda')^2}{(1-2\lambda')}\binom{n}{r}^2 \\
&< n\frac{(1-\lambda')^2}{(1-2\lambda')}\frac{2^{2h(\lambda')n}}{2\pi n\lambda'(1-\lambda')} \\
&= \frac{(1-\lambda')}{2\pi\lambda'(1-2\lambda')}2^{2h(\lambda')n}
\end{aligned}$$

$$
\begin{aligned}
&< \frac{(1-\lambda)}{2\pi\lambda(1-2\lambda)} 2^{2h(\lambda')n} \\
&= \Delta\, 2^{2h(\lambda')n} \\
&< \Delta\, 2^{n+2\gamma_1\alpha\log_2(n)-2\gamma_1\beta-2\gamma_1\epsilon} \\
&\leq \Delta\, 2^{n+2\gamma_1\alpha\log_2(n)-2\gamma_1\beta} \\
&= \Delta\, 2^{n+\log_2(n)-\log_2(\Delta)} = 2^n n.
\end{aligned}
$$

Hence $r(n) \geq \lfloor \lambda n + \alpha\log_2(n) - \beta \rfloor$ for $n \geq 108$. Direct computation shows that the inequality is satisfied also for $n \leq 107$. This proves the lower bound.

The proof of the upper bound is similar, but a little more complicated.

Let $r = \lambda n + \alpha\log_2(n) - \beta + \gamma(\log_2(n))^2/n$ for some $\gamma > \theta$. Let $\lambda' = r/n$. From (9)–(11) we get

$$
\begin{aligned}
&\Big\{n + (n-r)\binom{n}{r}\Big\} M_r \\
&> (n-r)\frac{n-r+3}{n-2r+3}\binom{n}{r}^2 \\
&= n\frac{(1-\lambda')(1-\lambda'+3/n)}{(1-2\lambda'+3/n)}\binom{n}{r}^2 \\
&> \frac{(1-\lambda'+3/n)}{2\pi\lambda'(1-2\lambda'+3/n)} 2^{2h(\lambda')n} e^{-\frac{-1}{12n\lambda'(1-\lambda')}} \\
&> \frac{(1-\lambda'+3/n)}{2\pi\lambda'(1-2\lambda'+3/n)} e^{-\frac{-1}{12n\lambda'(1-\lambda')}} \\
&\quad \cdot 2^{n+2\gamma_1\alpha\log_2(n)-2\gamma_1\beta+2\gamma_1\gamma\frac{(\log_2(n))^2}{n}} \\
&\quad \cdot 2^{-\gamma_2 n(\alpha\frac{\log_2(n)}{n}-\frac{\beta}{n}+\gamma\frac{(\log_2(n))^2}{n^2})^2} \\
&= 2^n\, n\, g(n)
\end{aligned}
$$

where, putting $\frac{1}{n} = x$ and $\log_2(n) = y$,

$$
\begin{aligned}
g(n) &= \frac{\lambda(1-2\lambda)(1-\lambda'+3/n)}{\lambda'(1-2\lambda'+3/n)(1-\lambda)} e^{-\frac{-1}{12\lambda'(1-\lambda')}x} \\
&\quad \cdot 2^{2\gamma_1\gamma y^2 x - \gamma_2 x(\alpha y - \beta x + \gamma y^2 x)^2}.
\end{aligned}
$$

The Taylor expansion of $g(n)$ in x (looking at y as an independent parameter) is

$$
g(n) = 1 + \sum_{i=1}^{\infty}\Big(\sum_{j=0}^{2i} t(i,j) y^j\Big) x^i.
$$

The coefficients $t(i,j)$ are expressions in the constants introduced above. Since these expressions are quite complicated, even for small i and j, we just give the numerical approximations for the first few terms:

$$
\begin{aligned}
g(n) &= 1 + 2\ln(2)\gamma_1(\gamma-\theta)y^2 x - 0.862yx - 0.557x \\
&\quad + (8.740\gamma^2 - 1.174\gamma + 0.039)y^4 x^2 \\
&\quad + (0.711 - 6.990\gamma)y^3 x^2 + (2.085 - 7.537\gamma)y^2 x^2 \\
&\quad - 1.151 yx^2 + 2.418 x^2 + \cdots.
\end{aligned}
$$

The dominating terms are $1 + 2\ln(2)\gamma_1(\gamma-\theta)y^2 x > 1$. Hence $g(n) \geq 1$ for n sufficiently large. QED

How large is "sufficiently large"? For $\gamma = 0.09$ for example, numerical results indicate that $g(n) \geq 1$ for all $n \geq 939$ (this has been verified for $939 \leq n \leq 5000$ and $\frac{n}{(\log_2(n))^2}(g(n)-1)$ is increasing in this range). Further, note that the condition $g(n) \geq 1$ is only a sufficient condition which may not be necessary. Numerical computations show that (13) for $\gamma = 0.09$ is in fact satisfied *with equality* for all $2 \leq n \leq 10000$, except $n = 5$, $n = 38$, and $n = 415$; in those three cases, $r(n)$ is one above the bound in (13) (hence the bound (13) seems to be valid with $\gamma = 0.09$ for all $n \geq 416$).

Numerical computations also show that the bounds in (12) and (13) coincide in most cases, as we would expect. The bounds differ by at most one. They are different in 8 cases for $n \leq 100$, 9 cases for $100 < n \leq 1000$, 15 cases for $1000 < n \leq 5000$, and 6 cases for $5000 < n \leq 10000$.

Once we have determined $r(n)$, we get $\mu_{r(n)}$ from (5). In our final theorem, we give the asymptotic behavior of $\mu_{r(n)}$.

Theorem 5 *When $n \to \infty$, then*

$$\mu_{r(n)} 2^{-n/2} \to \rho = (1-2\lambda)^{-1/2} \approx 1.1323175.$$

Proof: By (5) and (11), if $r = r(n) \sim \lambda n$, then

$$\begin{aligned} b_r &= M_r(n-2r-2) - (n-r)\binom{n}{r} - n \\ &> \left\{\frac{n-r+3}{n-2r+3}(n-2r-2) - (n-r)\right\}\binom{n}{r} - n \\ &= -\frac{2n+r+6}{n-2r+3}\binom{n}{r} - n \\ &\sim -\frac{(2+\lambda)}{(1-2\lambda)}\frac{2^{n/2}}{\sqrt{2\pi n\lambda(1-\lambda)}}. \end{aligned}$$

Similarly, the upper bound on M_r in (11) gives

$$b_r < -\frac{2n+4}{n-2r+2}\binom{n}{r} - n \sim -\frac{2}{(1-2\lambda)}\frac{2^{n/2}}{\sqrt{2\pi n\lambda(1-\lambda)}}.$$

Hence $b_r/2^{n/2} \to 0$ and so

$$\mu_r \sim \frac{\sqrt{4n(n-2r-2)2^n}}{2(n-2r-2)} \sim 2^{n/2}\rho.$$

QED

5 Summary

We defined $\Omega(n)$ to be the largest integer such that all codes of length n and size at least $2^n - \Omega(n)$ are good for error detection. We have shown that $\Omega(n)$ is of the order $2^{n/2}$, more precisely, that

$$\rho \approx 1.1323175 < \frac{\Omega(n)}{2^{n/2}} < \begin{cases} 2 & \text{if } n \text{ is even} \\ 2\sqrt{2} & \text{if } n \text{ is odd.} \end{cases}$$

It is not clear if $\omega = \lim_{n\to\infty} \Omega(n)2^{-n/2}$ exists, but probably it does. It is an open problem to show this and to determine the exact value of ω. Our bounds give $\rho \leq \omega \leq 2$. Possibly none of these bounds are sharp. The lower bound was derived using an upper bound on $P_{ue}(C,p)$ which is not sharp in most cases. The upper bound 2 follows from a construction using complements of linear codes and possibly some non-linear codes will give lower values.

Acknowledgements

This research work is supported in part by the National Natural Science Foundation of China under the Grant 60172060, the Trans-Century Training Program Foundation for the Talents by the Education Ministry of China, and the Foundation for University Key Teacher by the Education Ministry of China, the DSTA project (POD 0103223), and The Norwegian Research Council.

References

[1] F.-W. Fu and T. Kløve, "The complement of binary linear codes for error detection", in: *Proc. 2002 IEEE Information Theory Workshop*, Bangalore, India, October 20-25, p. 187, 2002.

[2] F.-W. Fu, T. Kløve and V. K. Wei, "On the Undetected Error Probability for Binary Codes", *IEEE Trans. Inform. Theory* vol. 49, pp. 382-390, 2003.

[3] T. Kløve and V. Korzhik, *Error Detecting Codes: General Theory and Their Application in Feedback Communication Systems*. Boston: Kluwer Acad. Press, 1995.

[4] W. W. Peterson and E. J. Weldon: *Error-Correcting Codes*, Cambridge, MA: MIT Press, 1972.