Some Inequalities Concerning Binomial Coefficients and the Weight Distribution of Proper Linear Codes

WACKER H. D., BOERCSOEK J.
Development
HIMA Paul Hildebrandt GmbH + Co KG
Albert-Bassermann-Strasse 28, D-68782 Bruehl
GERMANY

Abstract: - Proper linear codes play an important role in error detection. They are characterized by an increasing probability of undetected error $p_{ue}(\varepsilon,C)$ and are considered "good for error detection". A lot of CRCs commonly used to protect data transmission via a variety of field busses are known for being proper. In this paper the weight distribution of proper linear codes on a binary symmetric channel without memory is investigated. A proof is given that its components are upper bounded by the binomial coefficients in a certain sense. Secondly an upper bound of the tail of the binomial is given, and the results are then used to derive estimates of $p_{ue}(\varepsilon,C)$. Finally, applications on safety integrity levels are studied.

Key-Words: - Binary Symmetric Channel, Proper Linear Code, CRC, Probability of Undetected Error, Weight Distribution, Binomial, Safety Integrity Level, Block Length, Bit Error Pobability, Minimum Distance

1 Introduction

Let C be a [n, k] linear code on a binary symmetric channel without memory, where n is the block length and k is the number of data bits. The probability of undetected error of such a code is then given by (see [14] for example):

$$p_{ue}(\varepsilon, \mathrm{C}) = \sum_{l=1}^{n} A_l \varepsilon^l (1-\varepsilon)^{n-l}$$

where

A_l = component of the weight distribution of C
= number of code words of weight l,
ε = bit error probability,
n = block length.
d = minimum distance of C.

Clearly the A_l are upper bounded by the binomial coefficients

$$A_l \le \binom{n}{l}, \tag{1}$$

an inequality representing the "worst case".

In several publications ([1], [2], [10], [11], [12]) the range of binomiality of a linear code has been investigated, i.e. the range of all indices l with A_l satisfying

$$A_l \le \gamma \cdot \frac{\sqrt{n}}{2^r} \cdot \binom{n}{l}, \tag{2}$$

where $r = n - k$ is the length of the check sum and $\gamma > 0$ is a positive constant. If C is a cyclic redundancy check (CRC) $r = n - k$ is equal to the degree of the polynomial generating the CRC. A common result of all papers is that there is binomial behavior of A_l when l is taken from some neighborhood of $n/2$. Moreover, in each subinterval large enough there is an index i such that the binomial bound is asymptotically met (see for example [1] or [10]).

2 Proper Linear Codes

A linear code C is said to be proper if and only if the probability of undetected error $p_{ue}(\varepsilon,C)$ is an increasing function of ε in the interval [0, 1/2].

Because of

$$\begin{aligned} p_{ue}(\varepsilon, C) &\le p_{ue}(1/2, C) \\ &\le (2^{\mathrm{k}} - 1)/2^n \\ &< 1/2^r \end{aligned} \tag{3}$$

for all $\varepsilon \in [0, 1/2]$ proper linear codes obey the 2^{-r} bound. Those codes are considered "good for error detection" (see for example [13]), and they are widely used in this field.

A lot of important CRCs used to protect data transmission are known for being proper (at least for most block lengths, see [3], [4], [5], [6], [7], [8]). On the other hand nothing seems to be known about specific properties of the weight distribution of a linear code resulting from properness.

In subsection (3.1) of this paper we shall prove that the weight distribution of each proper linear code is showing binomial behavior in the sense of (2) for all components A_l with $l \le n/2$. In subsection 3.2 we shall give estimates of the tail of the binomial, and use it in 3.3 to derive upper bounds on the probability of undetected error for proper linear codes. Finally the consequences of these

estimates for the problem of achieving a specific Safety Integrity Level (SIL) are investigated.
.

3 Binomial Behavior and Properness

3.1 The Weight Distribution

In order to demonstrate our main result we took advantage of Stirling's approximation

$$1 \le \frac{n!}{\sqrt{2\pi n}\,(n/\mathrm{e})^n} \le 1+1/11n \quad \text{for all } n = 1,2,3,\ldots,$$

from which we were able to deduce

Theorem 1: Let C be an arbitrary linear code, then for each component A_i of the weight distribution of C the inequality

$$A_l \le \frac{72}{121}\sqrt{2\pi}\sqrt{n}\cdot p_{ue}(\frac{l}{n}, C)\cdot\binom{n}{l}.$$

holds.

Proof: For all l = 1,…n the subsequent inequality is obvious

$$A_l\left(\frac{l}{n}\right)^l\left(1-\frac{l}{n}\right)^{n-l} \le \sum_{i=1}^{n} A_i\left(\frac{l}{n}\right)^i\left(1-\frac{l}{n}\right)^{n-i} \tag{4}$$

$$= p_{ue}(\frac{l}{n}, C).$$

Consequently

$$A_l \le \frac{1}{\left(\frac{l}{n}\right)^l\left(1-\frac{l}{n}\right)^{n-l}}\cdot p_{ue}(\frac{l}{n}, C)$$

$$= \frac{n^n}{l^l(n-l)^{n-l}}\cdot p_{ue}(\frac{l}{n}, C),$$

from which, by Stirling's approximation, we get

$$A_l \le \frac{n!\mathrm{e}^n}{\sqrt{2\pi n}}\frac{\frac{12}{11}\sqrt{2\pi l}}{l!\mathrm{e}^l}\frac{\frac{12}{11}\sqrt{2\pi(n-l)}}{(n-l)!\mathrm{e}^{n-l}}\cdot p_{ue}(\frac{l}{n}, C)$$

$$= \frac{144}{121}\sqrt{2\pi}\sqrt{\frac{l(n-l)}{n}}\frac{n!}{l!(n-l)!}\cdot p_{ue}(\frac{l}{n}, C)$$

$$= \frac{144}{121}\sqrt{2\pi}\sqrt{\frac{l(n-l)}{n}}\binom{n}{l}\cdot p_{ue}(\frac{l}{n}, C).$$

And then, by the inequality of the arithmetic and geometric means

$$A_l \le \frac{144}{121}\sqrt{2\pi}\frac{1}{\sqrt{n}}\frac{l+(n-l)}{2}\cdot\binom{n}{l}\cdot p_{ue}(\frac{l}{n}, C)$$

$$\le \frac{72}{121}\sqrt{2\pi}\sqrt{n}\,p_{ue}(\frac{l}{n}, C)\cdot\binom{n}{l}.$$

∎

Remark 1: Perry in [13] used (4) to find codes not satisfying the 2^{-r} bound. We pursued a different plan and therefore continued in a different way.

Now we are able to state our main result:

Theorem 2: Let C be a proper linear code then each component A_l of the weight distribution of C with $l \le n/2$ is showing binomial behavior, i.e.:

$$A_l \le \frac{72}{121}\sqrt{2\pi}\cdot\frac{\sqrt{n}}{2^r}\cdot\binom{n}{l}.$$

Proof: For all $1 \le n/2$ we have $l/n \le ½$ and therefore by (3) $p_{ue}(l/n, C) \le 2^{-r}$, from which the statement follows by Theorem 1. ∎

Remark 2: As the proof shows, Theorem 2 remains valid, if we replace properness by the more general condition of C satisfying the 2^{-r} bound. Because of the importance of the class of proper linear codes and due to the fact that normally properness is used to validate the 2^{-r} bound we didn't state Theorem 2 under the most general conditions.

As an easy conclusion of Theorem 1 we now get immediately a first simple estimate of the probability of undetected error:

Theorem 3: Let C be an arbitrary linear code, then for each ε with $0 \le \varepsilon \le 1$ the probability of undetected error is upper bounded by

$$p_{ue}(\varepsilon, C) \le \frac{72}{121}\sqrt{2\pi}\sqrt{n}\cdot\sum_{l=d}^{n} p_{ue}(\frac{l}{n}, C)\cdot\binom{n}{l}\varepsilon^l(1-\varepsilon)^{n-l}.$$

3.2 The Tail of the Binomial

We now want to apply the statement of Theorem 2 to get an upper bound on the probability of undetected error of proper linear codes. To this end we need an estimate of the tail of the binomial delivered by Theorem 4.

Theorem 4: For all natural numbers q and n with $q \le n$ and all ε with $0 \le \varepsilon \le 1$ the tail of the binomial obeys the following inequality:

$$\sum_{l=q}^{n}\binom{n}{l}\varepsilon^e(1-\varepsilon)^{n-e} \le \binom{n}{q}\varepsilon^q \le \frac{1}{q!}n^q\varepsilon^q \tag{5}$$

Proof: a) To begin with, let k, n and q be natural numbers satisfying $k + q \le n$, with the help of which we get an estimate of the binomial coefficients:

$$\binom{n}{k+q} = \frac{n(n-1)\cdots(n-q+1)}{(k+q)\cdots(k+1)} \cdot \frac{(n-q)\cdots(n-k-q+1)}{k!}$$
$$= \frac{n(n-1)\cdots(n-q+1)}{(k+q)\cdots(k+1)} \binom{n-q}{k}$$
$$\leq \frac{n(n-1)\cdots(n-q+1)}{q!} \binom{n-q}{k}$$
$$= \binom{n}{q} \cdot \binom{n-q}{k}$$

b) When then focusing onto (5) by means of part a), we achieve

$$\sum_{l=q}^{n} \binom{n}{l} \varepsilon^l (1-\varepsilon)^{n-l} = \varepsilon^q \sum_{l=q}^{n} \binom{n}{l} \varepsilon^{l-q} (1-\varepsilon)^{n-l}$$
$$= \varepsilon^q \sum_{k=0}^{n-q} \binom{n}{k+q} \varepsilon^k (1-\varepsilon)^{n-k-q}$$
$$\leq \varepsilon^q \cdot \sum_{k=0}^{n-q} \binom{n}{q} \cdot \binom{n-q}{k} \varepsilon^k (1-\varepsilon)^{n-k-q}$$
$$= \varepsilon^q \binom{n}{q} \cdot \{\varepsilon + (1-\varepsilon)\}^{n-q}$$
$$= \frac{n \cdot (n-1) \cdots (n-q+1)}{q!} \varepsilon^q$$
$$\leq \frac{n^q}{q!} \varepsilon^q.$$

∎

Now, by inequality (1) and Theorem 4 a simple inequality turns out

Theorem 5: Let *C* be an arbitrary linear code, then for each ε with $0 \leq \varepsilon \leq 1$ the probability of undetected error is upper bounded by

$$(6) \qquad p_{ue}(\varepsilon, C) \leq \sum_{l=d}^{n} \binom{n}{l} \varepsilon^e (1-\varepsilon)^{n-e} \leq \frac{1}{d!} n^d \varepsilon^d$$

where *d* is the minimum distance of *C*.

Remark 3: If nothing is known about the code but its minimum distance *d*, Theorem 5 is useful for calculating maximal block lengths in order to achieve a specific upper bound σ on $p_{ue}(\varepsilon, C)$. You only have to choose

$$(7) \qquad n_{max} < \varepsilon^{-1} (\sigma d!)^{1/d}.$$

In fact (7) is used when dealing with safety related systems.

3.3 The Probability of Undetected Error

Now we are in a position to estimate the probability of undetected error in the case of proper linear codes. As common use $\lfloor x \rfloor$ has the meaning of the floor function.

Theorem 6: Let C be a proper linear code, then for all ε ϑ [0, 1/2] the probability of undetected error obeys:

$$(8) \qquad p_{ue}(\varepsilon, C) \leq \frac{72}{121} \sqrt{2\pi} \sum_{l=d}^{\lfloor n/2 \rfloor} \frac{\sqrt{n}}{2^r} \binom{n}{l} \varepsilon^l (1-\varepsilon)^{n-l} + R_n(\varepsilon),$$

where *d* is the minimum distance of *C*, and the remainder term $R_n(\varepsilon)$ obeys

$$R_n(\varepsilon) \leq \binom{n}{\lfloor n/2 \rfloor} \varepsilon^{\lfloor n/2 \rfloor} \leq \begin{cases} \left(2\sqrt{\varepsilon}\right)^n & \text{, if } n \geq 3 \text{ and even,} \\ 2\left(2\sqrt{\varepsilon}\right)^{n-1} & \text{, if } n \geq 4 \text{ and odd} \end{cases}.$$

Proof: a) Firstly, let n be even, then by Theorem 2

$$p_{ue}(\varepsilon, C) = \sum_{l=d}^{n/2} A_l \varepsilon^l (1-\varepsilon)^{n-l} + \sum_{l=n/2+1}^{n} A_l \varepsilon^l (1-\varepsilon)^{n-l}$$
$$\leq \frac{72}{121} \sqrt{2\pi} \sum_{l=d}^{\lfloor n/2 \rfloor} \frac{\sqrt{n}}{2^r} \cdot \binom{n}{l} \varepsilon^l (1-\varepsilon)^{n-l} + R_n(\varepsilon),$$

where (by Theorem 4 with $q = n/2$ and Stirling's approximation)

$$R_n(\varepsilon) = \sum_{l=q+1}^{n} A_l \varepsilon^l (1-\varepsilon)^{n-l}$$
$$\leq \sum_{l=q}^{n} A_l \varepsilon^l (1-\varepsilon)^{n-l}$$
$$\leq \binom{n}{q} \varepsilon^q$$
$$= \frac{n!}{q!(n-q)!} \varepsilon^q$$
$$\leq \frac{2\sqrt{2\pi n} n^n \mathrm{e}^{-n}}{\sqrt{2\pi q} q^q \mathrm{e}^{-q} \sqrt{2\pi(n-q)} (n-q)^{n-q} \mathrm{e}^{-(n-q)}} \varepsilon^q$$
$$= \sqrt{\frac{2n}{\pi q(n-q)}} \frac{n^n}{q^q (n-q)^{n-q}} \varepsilon^q$$
$$= \sqrt{\frac{2n}{\pi(n/2)(n/2)}} \frac{n^n}{(n/2)^{n/2}(n/2)^{n/2}} \varepsilon^{n/2}$$
$$\leq \left(2\sqrt{\varepsilon}\right)^n,$$

if $n \geq 3$. Let now n be odd, then again by Theorem 2

$$p_{ue}(\varepsilon, C) = \sum_{l=d}^{(n-1)/2} A_l \varepsilon^l (1-\varepsilon)^{n-l} + \sum_{l=(n+1)/2}^{n} A_l \varepsilon^l (1-\varepsilon)^{n-l}$$

$$\leq \frac{72}{121}\sqrt{2\pi}\sum_{l=d}^{\lfloor n/2 \rfloor}\sqrt{n}\cdot\frac{1}{2^r}\cdot\binom{n}{l}\varepsilon^l(1-\varepsilon)^{n-l} + R_n(\varepsilon),$$

where (by Theorem 4 with $q = (n - 1)/2$)

$$\begin{aligned} R_n(\varepsilon) &= \sum_{l=q+1}^{n} A_l \varepsilon^l (1-\varepsilon)^{n-l} \\ &\leq \sum_{l=q}^{n} A_l \varepsilon^l (1-\varepsilon)^{n-l} \\ &\leq \binom{n}{q}\varepsilon^q \\ &= \frac{n!}{q!(n-q)!}\varepsilon^q \\ &= \frac{2q+1}{q+1}\frac{(2q)!}{q!q!}\varepsilon^q \\ &\leq 2\frac{(2q)!}{q!q!}\varepsilon^q. \end{aligned}$$

And therefore as in the proof of the "even case" by Stirling's approximation

$$R_n(\varepsilon) \leq 2\left(2\sqrt{\varepsilon}\right)^{n-1}$$

if $n \geq 4$. ∎

Remark 4: Even for relatively large ε (= 10^{-2}) and relatively small n (= 40, imagine a payload of 1 byte and a CRC-32) the remainder term $R_n(\varepsilon)$ is so small (< 10^{-26}) that it doesn't carry any weight compared with the first term on the right hand side of (8).
Now, as a consequence of Theorems 4 and 6, an analogon of Theorem 5 for proper linear codes emerges.

Theorem 7: Let C be a proper linear code, then for all ε ∈ [0, 1/2] the probability of undetected error is upper bounded by:

$$(9) \qquad p_{ue}(\varepsilon, C) \leq \frac{72}{121}\sqrt{2\pi}\frac{\sqrt{n}}{2^r}\frac{1}{d!}n^d\varepsilon^d + R_n(\varepsilon),,$$

where d is the minimum distance of C, and the remainder term $R_n(\varepsilon)$ obeys

$$R_n(\varepsilon) \leq \binom{n}{\lfloor n/2 \rfloor}\varepsilon^{\lfloor n/2 \rfloor} \leq \begin{cases} \left(2\sqrt{\varepsilon}\right)^n & \text{, if } n \geq 3 \text{ and even,} \\ 2\left(2\sqrt{\varepsilon}\right)^{n-1} & \text{, if } n \geq 4 \text{ and odd} \end{cases}.$$

Finally let us state the subsequent remarks, pointing out the influence of properness on the size of the probability of undetected error on one hand and on maximal block lengths on the other hand

Remark 5: In the case of a CRC of length r and for all n not too large, inequality (9) improves inequality (6) by a factor of

$$\frac{72}{121}\frac{\sqrt{2\pi n}}{2^r}.$$

Remark 6: Similar to (6) inequality (9) too is useful for calculating maximal block lengths in order to achieve a specific upper bound σ on $p_{ue}(\varepsilon, C)$:

$$(10) \qquad p_{ue}(\varepsilon, C) \leq \frac{72}{121}\sqrt{2\pi}\frac{\sqrt{n}}{2^r}\frac{1}{d!}n^d\varepsilon^d + R_n(\varepsilon) < \sigma.$$

Apart from $R_n(\varepsilon)$, being small compared with the other term on the right hand side of (9), you only have to choose

$$(11) \qquad n_{max} < \left(\varepsilon^{-d}\cdot\sigma\cdot d!\cdot\frac{121}{72}\frac{1}{\sqrt{2\pi}}2^r\right)^{1/(d+0.5)}.,$$

and ensure $R_n(\varepsilon)$ to be small enough such that (10) is fulfilled.
In the case of a CRC of length r, inequality (11) improves inequality (7) by an order of magnitude of

$$\left(\frac{121}{72}\frac{1}{\sqrt{2\pi}}2^r\right)^{1/(d+0.5)}$$

3.4 Application to Safety Integrity Levels

As an application of our results, let us now have a closer look at data integrity according to IEC 68508.
According to remark 6 in subsection 3.2 we wanted to analyze the effect of properness on maximal block lengths achievable for a specific Safety Integrity Level (SIL). Safety Integrity Levels are defined by means of the number Λ of undetected errors per hour:

$$\Lambda = 3600 \cdot p_{ue}(\varepsilon, C) \cdot \nu \cdot (m\text{-}1) \cdot 100$$

where
ν = number of safety related messages per second
m = number of communicating devices
100 = 1%-rule
(Details are outlined in IEC 61508 2000, [9].) For our example, we decided to choose ν = 100, a value suggested by experience, and m = 2. In this way we get

(12) $\Lambda = 3{,}6 \cdot 10^{7} \cdot p_{ue}(\varepsilon, C)$

If no details are known about the quality of the transmission especially about the electromagnetic compatibility (EMC), and nothing can be said about the bit-error probability ε, the Technical Control Board of Germany requires to do all calculations concerning Λ with $\varepsilon = 10^{-2}$. Therefore for our analysis we took account of this bad value of ε.

If on the other hand no details are known about the weight distribution of the code *C*, the only chance of estimating the probability of undetected error is to use (6) or (9).

We based our calculations on the results of Castagnoli et al. in [4] about the CRC-32/6 polynomial. According to [4], CRC-32/6 is proper for all $n \leq$ 32 767. It is exemplary for a lot of other CRCs for which similar results are known (see for example [3], [4], [5], [6], [7], [8]).

By means of (7) and (11) we then derived the content of table 1 from the results in [4] about the minimum distance *d* as a function of *n*.

Table 1: Maximal block lengths for CRC-32/6

SIL	Λ high demand	$n_{\max}$ (7)	$n_{\max}$ by (11)
4	10^{-8}	37	56
3	10^{-7}	37	66
2	10^{-6}	39	87
1	10^{-5}	43	114

Using (7) SIL 3 and 4 are achievable with a payload of only 5 bits. In contrast to this fact, with the help of (11) they are achievable with a payload of 34 respectively 24 bits. This result shows the improvement of (11) compared with (7) with regard of practical application.

4 Conclusions

Via the binomiality of the weight distribution an upper bound on the probability of undetected error of a class of codes has been proven, which is important for practical applications in safety related systems. The bound can be calculated without knowledge of the complete weight distribution of the code. Only the knowledge of the minimum distance is required. It improves a bound used so far in this field.

References:

[1] Ashikhmin, A., Barg, A., and Litsyn, S., "Estimates of the Distance Distribution of Codes and Designs," *IEEE Trans. Inform. Theory*, vol. 47, no. 3, March 2001. pp. 1050–1061.

[2] Ashikhmin, A., Cohen, G.D., Krivelevich, M. and Litsyn, S., "Bounds on Distance Distributions on Codes of Known Size," *IEEE Trans. Inform. Theory*, vol. 51, no. 1, Jan. 2005. pp. 250–258.

[3] Baicheva, T., Dodunekov, S., and Kazakov, P., "Undetected error probability performance of cyclic redundancy-check codes of 16-bit redundancy," *IEE Proc.- Commun.*, Vol. 147, No. 5, October 2000.

[4] Castagnoli, G., Braeuer, S., and Herrman, M., "Optimization of Cyclic Redundancy-Check Codes with 24 and 32 Parity Bits," *IEEE Trans. on Communications*, Vol. 41, No. 6, June 1993. pp. 883-992.

[5] Castagnoli, G., Ganz. J., and Graber, P., "Optimum Cyclic Redundancy-Check Codes with 16-Bit Redundancy," *IEEE Trans. on Communications*, Vol. 38, No. 1, 1990, pp. 111-114.

[6] Fujiwara, T., Kasami, T., and Lin, S., "Error Detecting Capabilities of the Shortened Hamming Codes Adopted for Error Detection in IEEE Standard 802.3," *IEEE Trans. on Communications*, Vol. 37, No. 9, Sept. 1989. p. 986-989.

[7] Fujiwara, T., Kasami, T., and Lin, S., "On the Undetected Error Probability for Shortened Hamming Codes," *IEEE Trans. on Communications*, Vol. COM-33, No. 6, Sept. 1985. pp. 570-574.

[8] Funk, G., "Determination of Best Shortened Linear Codes," *IEEE Trans. on Communications*, Vol. 4, No. 1, Jan. 1996. pp. 1-6.

[9] IEC 61508, International Standard 61508: Functional safety of electrical/electronic/ Programmable electronic safety-related systems, Geneva, International Electrotechnical Commission, 2000

[10] Krasikov, I., and Litsyn, S., "Bounds on Spectra of Codes with Known Dual Distance," *Des.Codes Cryptogr.*, vol. 13, no. 3, pp. 285–297, 1998.

[11] Krasikov, I. and Litsyn, S., "Estimates for the Range of Binomiality in Codes Spectra," *IEEE Trans. on Information Theory*, vol. 43, no. 3, May 1997. pp. 987–990.

[12] Krasikov, I. and Litsyn, S., "Linear Programming Bounds for Doubly-Even Self-Dual Codes," *IEEE Trans. on Information Theory*, vol. 43, no. 4July 1997. pp. 1238–1244.

[13] Perry, P., "Necessary Conditions for Good Error Detection," *IEEE Trans. Inform. Theory*, vol. 37, no. 2, March. 1991. pp. 375–378.

[14] Peterson, W. W. and Weldon, E. J., *Error Correcting Codes*. The MIT Press Cambridge, Massachusetts , and London, England, Second Edition 1972.

[15] Witzke, K. A., and Leung, C., "A Comparison of Some Error Detecting CRC Code Standards," *IEEE Trans. on Communications*, Vol. COM-33, No. 9, Sept. 1985. pp. 996-998.

[16]Wolf, J. K., and Blakeney, R. D., "An exact Evaluation of the Probability of Undetected Error for certain Shortened Binary CRC Codes," Qual.Comm, Inc., San Diego, CA 92121. *Proc. Milcom IEEE* 1988. pp. 287-292.