LINEAR BLOCK CODES FOR ERROR DETECTION

R. Dodunekova
Department of Mathematics
Chalmers University of Technology
and the University of Gothenburg
412 96 Gothenburg, Sweden

S. M. Dodunekov
Institute of Mathematics
Bulgarian Academy of Sciences
8 G. Bontchev Str.
1113 Sofia, Bulgaria

Abstract

The performance of linear block codes over a finite field is investigated when they are used for pure error detection. The maximal probability of undetected error is investigated. Sufficient condition for a code to be good or proper for error detection are derived.

1 INTRODUCTION

Two basic strategies are used to control transmission errors in data communication systems. In Forward-Error-Correction (FEC) an error-correcting code is used for correcting errors. With Automatic-Repeat-Request (ARQ) protocols using error detection together with request for retransmission almost error-free data transmission can be achieved. The second approach, because of its simple interpretation and high reliability, is widely used in packet-switching data networks, computer communication net-works, satelite communications.

In this paper we investigate the performance of linear block codes when they are used for pure error detection. A linear $[n, k, d; q]$ code with symbols from a

finite field of q elements $GF(q)$, is a k-dimensional subspace of the n-dimensional vector space over $GF(q)$, with minimum Hamming distance d.

Let C be an $[n, k, d; q]$ code which is used for error detection. Let $x \in C$ be the transmitted codeword and $y \in GF(q)^n$ be the received vector. Then the vector

$$e = y - x$$

is the error vector caused by the channel noise. If $e \in C$, then $y = x + e \in C$ and the decoder assumes that y is error free, i.e., the decoder fails to detect the error and y will be accepted as a code vector. Such an error is called undetectable. If $e \notin C$, then $y \notin C$ and the decoder will discover the presence of an error. Such an error is called detectable.

We shall consider the following probabilistic model. The $[n, k, d; q]$ code C is used for error detection on a discrete memoryless channel with q inputs and q outputs. Any transmitted codeword has a probability $1 - \epsilon$ of being received correctly and a probability $\frac{\epsilon}{q-1}$ of being transformed into each of the $q - 1$ other symbols. We assume that $0 \leq \epsilon \leq \frac{q-1}{q}$.

Define $P_{ud}(C, \epsilon)$ to be the probability that the decoder fails to detect the existence of an transmission error. This probability is called the probability of undetected error for C. The probability $P_{ud}(C, \epsilon)$ can be expressed in terms of the weight distribution of either C or its dual code $C^{\perp}$. Denote by $\{A_i : 0 \leq i \leq n\}$, resp. $\{B_i : 0 \leq i \leq n\}$ the weight distribution of C, resp. $C^{\perp}$. Then

$$P_{ud}(C, \epsilon) = \sum_{i=1}^{n} A_i \big(\frac{\epsilon}{q-1}\big)^i (1-\epsilon)^{n-i}, \tag{1.1}$$

or equivalently,

$$P_{ud}(C, \epsilon) = q^{-(n-k)} \sum_{i=0}^{n} B_i \big(1 - \frac{q\epsilon}{q-1}\big)^i - (1-\epsilon)^n \tag{1.2}$$

(see, for example, Lin-Castello, p. ...).

To compute the exact value of $P_{ud}(C, \epsilon)$ by use of (1.1) or (1.2) is equivalent to find the weight distribution of C, resp. $C^{\perp}$. This is done only for few classes of codes and for large code parameters it is known to be a hard computational problem [MacWilliams & Sloane, p. ...]. An easier problem is to find good bound on $P_{ud}(C, \epsilon)$. Note that even when $P_{ud}(C, \epsilon)$ is known a criterion is needed to decide if the code is suitable for error detection. One such reasonable criterion is to compare $P_{ud}(C, \epsilon)$ with the average probability $P_{ud}(\epsilon)$ of undetected error for

the ensemble of all linear q-nary $[n,k]$ codes [Lin-Castello, p. 78]. It is known that

$$P_{ud}(\epsilon) = q^{-(n-k)}[1-(1-\epsilon)^k]$$

[Wolf, Michelson, Levesgue (1982); for $q=1$: Korznik (1965), Massey (1978)].

The following natural criteria were introduced by Leung, Barnes and Friedman (1979), Kasami-Lin (1984), and Kløve (1995).

If

$$P_{ud}(C,\epsilon) \leq P_{ud}\big(\frac{q-1}{q}\big)$$

for all $\epsilon \in [0, \frac{q-1}{q}]$, then C is *good* for error detection. If $P_{ud}(C,\epsilon)$ is an increasing function on ϵ in the interval $[0, \frac{q-1}{q}]$, the code is *proper* for error detection. Thus the proper codes are good in a somewhat regular way: the bigger ϵ is, the worse they perform in detecting errors.

The paper is organized as follows. In Section 2 we derive a unified representation of the function $P_{ud}(C,\epsilon)$, $0 \leq \epsilon \leq \frac{q-1}{q}$ in (1.1) as a function of $z, 0 \leq z \leq 1$, and discuss some properties of the functions involved in this representation. In Section 3 we give a sufficient condition for a linear $[n,k,d;q]$ code to be good for error detection. As corrollaries we derive some known results (see [K-L], [D-D], [Kløve]). In Section 4 we obtain a sufficient condition for a linear code to be proper. As an application we show that all q-nary $[n,k]$ codes with minimum distance $d \geq \frac{(q-1)n}{q}$ are proper. In particular, MacDonalds codes [see...] are proper.

For all notions and results from Coding theory which are not defined here we refer to [, ,]. A nice reference to the theory of error detecting codes is the recent monograph [Kløve-Korzhik].

2 UNIFIED REPRESENTATION OF $P_{ud}(C,\epsilon)$

For $z \in [0,1]$ introduce the function

$$R_e(z) = \binom{n}{i} z^i (1-z)^{n-i},\ \ell = 1,2,\ldots,n \tag{2.1}$$

and

$$L_\ell(z) = \sum_{j=\ell}^{n} R_j(z),\ \ell = 1,2,\ldots,n \tag{2.2}$$

We will express now the probability of undetected error in (1.1) in terms of the functions (2.1) and (2.2).

For brevity, denote

$$A_\ell^* = \sum_{i=d}^{\ell} \frac{\ell_{(i)}}{n_{(i)}} A_i,\ \ell = d, \dots, n, \tag{2.3}$$

where $\ell_{(i)} = \ell(\ell-1)\dots(e-i+1)$ and $n_{(i)} = n(n-1)\dots(n-i+1)$.

Lemma 1. *The following representation of $P_{ud}(C,\epsilon)$ takes place:*

$$P_{ud}(C,\epsilon) P(C,z),\ z = \frac{\epsilon q}{q-1} \tag{2.4}$$

with

$$\begin{aligned} P(c,z) &= \sum_{\ell=d}^{n} q^{-\ell} A_e^* R_e(z) && (2.5)\\ &= q^{-d} A_d^* L_d(z) \\ &\quad + \sum_{\ell=d+1}^{n} q(A_e^* - qA_{\ell-1}^*) L_\ell(z) && (2.6) \end{aligned}$$

Proof. The representation (2.4) with $P(C,z)$ as in (2.5) or (2.6) is obtained from (1.1) as follows:

$$\begin{aligned} P_{ud}(C,\epsilon) &= \sum_{i=d}^{n} A_i \Big(\frac{\epsilon}{q-1}\Big)^i (1-\epsilon)^{n-i} \\ &= \sum_{i=d}^{n} A_i \Big(\frac{\epsilon}{q-1}\Big)^i \sum_{j=0}^{n-i} \binom{n-i}{j} \Big(\frac{\epsilon}{q-1}\Big)^j \Big(1 - \frac{\epsilon q}{q-1}\Big)^{n-i-j} \\ &= \sum_{i=d}^{n} q^{-i} A_i z^i \sum_{j=0}^{n-i} q^{-j} \binom{n-i}{j} z^j (1-z)^{n-i-j} \\ &= \sum_{i=d}^{n} A_i \sum_{j=0}^{n-i} q^{-(i+j)} \binom{n-i}{j} z^{i+j} (1-z)^{n-i-j} \end{aligned}$$

where $z = \frac{\epsilon q}{q-1}$. Further, we set $\ell = i + j$ to get

$$\begin{aligned}
P_{ud}(C,\epsilon) &= \sum_{k=d}^{n} A_i \sum_{\ell=i}^{n} q^{-\ell} \binom{n-i}{\ell-i} z^{\ell} (1-z)^{n-\ell} \\
&= \sum_{i=d}^{n} A_i \sum_{\ell=i}^{n} q^{-\ell} \frac{\binom{n-i}{\ell-i}}{\binom{n}{\ell}} R_e(z) \\
&= \sum_{i=d}^{n} A_i \sum_{\ell=i}^{n} q^{-\ell} \frac{(n-i)!e!(n-\ell)!}{(\ell-i)!(n-e)!n!} R_e(z) \\
&= \sum_{i=d}^{n} A_i \sum_{\ell=i}^{n} q^{-\ell} \frac{\ell_{(i)}}{n_{(i)}} R_e(z) \\
&= \sum_{\ell=d}^{n} q^{-\ell} \Big(\sum_{i=d}^{\ell} \frac{\ell_{(i)}}{n_{(i)}} A_i \Big) R_e(z) \\
&= \sum_{\ell=d}^{n} q^{-\ell} A_e^* R_e(z),
\end{aligned}$$

which is the form (2.4)-(2.5). Using this and also that

$$R_e(z) = L_e(z) - L_{e+1}(z),\ \ell = d, \dots, n-1$$

which is easily seen from (2.1) and (2.2). We get

$$\begin{aligned}
P_{ud}(C,\epsilon) &= \sum_{\ell=d}^{n-1} q^{-\ell} A_e^* [L_e(z) - L_{e+1}(z)] \\
&\quad + q^{-n} A_n^* R_n(z) \\
&= \sum_{\ell=d}^{n-1} q^{-\ell} A_e^* L_e(z) \\
&\quad - \sum_{\ell=d+1}^{n} q^{-(e-1)} A_{e-1}^* L_e(z) \\
&\quad + q^{-d} A_d^* L_d \\
&\quad + \sum_{\ell=d+1}^{n} q^{-\ell} (A_e^* - q A_{e-1}^*) L_e(z),
\end{aligned}$$

which proves (2.4) with $P(C,z)$ as in (2.6). □

Next we will show that the functions $L_e(z)$ are strictly increasing in $z \in [0,1]$.

Lemma 2. *For $\ell = 1, 2, \dots, n$,*

$$L_e'(z) = n \binom{n-1}{\ell-1} z^{\ell-1} (1-z)^{n-\ell} \tag{2.7}$$

Proof. We will use induction on σ. If $j = 1$, then

$$L_1'(z) = 1 - [(1-z)^n]_z' = n(1-z)^{n-1}$$

and the statement holds in this case. Assume (2.7) for some ℓ, $1 \leq \ell < n$. We have

$$\begin{aligned}
L_{\ell+1}' &= L_\ell'(z) - R_e'(z) \\
&= n\binom{n-1}{\ell-1} z^{\ell-1}(1-z)^{n-\ell} \\
&- n\binom{n}{e} z^{\ell-1}(1-z)^{n-\ell-1}\big(\frac{\ell}{n} - z\big) \\
&= n\binom{n-1}{\ell} z^{\ell}(1-z)^{n-\ell-1}\big[\frac{\ell}{n-\ell}\frac{1-z}{z} - \frac{1}{n-\ell}\cdot\frac{\ell - nz}{z}\big] \\
&= n\binom{n-1}{\ell} z^{\ell}(1-z)^{n-\ell-1},
\end{aligned}$$

which proves (2.7) for $\ell + 1$. Then it is true for all ℓ, $1 \leq \ell \leq n$. □

3 GOOD ERROR DETECTING COES

Let C be an $[n, k, d; q]$ code with weight distribution $\{A_i : 0 \leq i \leq n\}$. A sufficient condition for C to be good for error detection is given by the following theorem

THEOREM 1. *If for $\ell = d, d+1, \ldots, n$,*

$$q^{-(n-k)} - q^{-n} \geq q^{-e} \sum_{i=d}^{\ell} \frac{\ell_{(i)}}{n_{(i)}} A_i \tag{3.1}$$

then C is good for error detection.

Proof. We sue (3.1) in (2.4) with $P(C, z)$ in it as given in (2.5) and also (2.7) to get

$$\begin{aligned}
P_{ud}(C, \epsilon) &= P(C, z) \\
&\leq \big(q^{-(n-k)} - q^{-n}\big) \sum_{\ell=d}^{n} R_e(z) \\
&\leq \big(q^{-(n-k)} - q^{-n}\big) L_d(z) \\
&\leq \big(q^{-(n-k)} - q^{-n}\big) L_d(1) \\
&= q^{-(n-k)} - q^{-n},
\end{aligned}$$

which shows that C is good. □

The theorem implies some known results

Corollary 1 ([Kasami-Lin]). All MDS codes are good for error detection.

Corollary 2 ([Dod-Dod, Th 2], [Kløve]). If C is in an NMDS q-nary code for which

$$A_{n-k} \leq \left(1 - q^{-k}\right) \binom{n}{k}$$

then C is good for error detection.

Proof of Corollary 1. Let C be an MDS q-nary (n, k) code. In this case $d = n - k + 1$ and [see [Dod-Dod]).

$$P_{ud}(C, \epsilon) = P(C, z)$$

with

$$P(C, z) = \sum_{\ell=n-k+1}^{n} \left(q^{-(n-k)} - q^{-\ell}\right) R_e(z), \tag{3.2}$$

that is,

$$\begin{aligned} A_e^* = q^{-(n-k)} - q^{-e} \leq q^{-(n-1)} - q^{-n}, \\ \ell n - k + 1, \ldots n. \end{aligned} \tag{3.3}$$

Theorem 1 holds and C is proper. □

Proof of Corollary 2. Let C be an NMDS q-nary (n, k) code. In this case $d = n - k$ and (see [Dod-Dod])

$$P_{ud}(C, \epsilon) = P^*(C, z)$$

where

$$P^*(C, z) = P(C, z) + q^{-(n-k)} \frac{A_{n-k}}{\binom{n}{k}} \cdot R_{n-k}(z)$$

and $P(C, z)$ is the function in (3.2).

The condition of corollary gives

$$q^{-(n-k)} \frac{A_{n-k}}{\binom{n}{k}} \leq q^{-(n-k)}(1 - q^{-k}) = q^{-(n-k)} - q^{-n}$$

and this together with (3.3) gives (3.1). Theorem 1 holds and the code is good for error detection. □

4 PROPER ERROR DETECTING COES

Again, let C be an $[n, k, d; q]$ code with weight distribution $\{A_i, 0 \le i \le n\}$.

THEOREM 2. *If for* $\ell = d,\ d+1, \ldots, n$

$$\sum_{i=d}^{\ell} \frac{\ell_{(i)}}{n_{(i)}} A_i \ge \sum_{i=d}^{\ell-1} \frac{(\ell-1)_{(i)}}{n_{(i)}} A_i \tag{4.1}$$

then C *is proper for error detection.*

Proof of Theorem 2. We use (2.4) with $P(C, z)$ as given in (2.6). The condition (4.1) of the theorem is, in terms of notation (2.3)

$$A_e^* - qA_{e-1}^* \ge 0,\ \ell = d+1, \ldots, n$$

which implies that the function

$$\sum_{\ell=d+1}^{n} q^{-\ell}(A_e^* - qA_{e-1}^*)L_e(z)$$

in the respresentation (2.6) is either identically zero (if all inequalities (4.1) are actually equalities) or strictly increasing on $[0, 1]$, by Lemma 2. As for the first term in (2.6), it obviously increases strictly on $[0, 1]$, by Lemma 2, again. Then $P(C.z)$ increases strictly on $z \in [0, 1]$ and hence so does $P_{ud}(C, \epsilon)$, where $\epsilon = \frac{z(q-1)}{q} \in [0, \frac{q-1}{q}]$. Thus the code C is proper. □

The theorem implies known results.

Corollary 3. All MDS codes are proper.

Corollary 4 (Corollary 3 in [Dod-Dod]). If C is an NMDS q-nary $[n, k]$, code for which

$$A_{n-k} \le (1 - q^{-1})\binom{n}{k}$$

then C is proper.

Corollary 5. If C is an $[n, k, d; q]$, code with

$$d \ge \frac{q-1}{q} n \tag{4.2}$$

then C is proper.

Proof of Corollary 3. For a MDS code C, we have in the respresentation (2.4) of $P_{ud}(C,\epsilon)$ with $P(C,z)$ as in (2.6) that (see [D-D])

$$P(C,v) = (q-1)\sum_{\ell=d}^{n} q^{-\ell} L_e(z).$$

This (and the notations (2.3)) show that (4.1) holds true and then C is proper. □

Proof of Corollary 4. For an NMDS q-nary code C, the representation (2.4) with $P(C,z)$ in it as in (2.6) is (see [Dod-Dod])

$$\begin{aligned} P_{ud}(C,\epsilon) &= q^{-(n-k)}\frac{A_{n-k}}{\binom{n}{k}} L_{n-k}(z) \\ &\quad + q^{-(n-k)}\big[1 - q^{-1} - \frac{A_{n-k}}{\binom{n}{k}}\big] L_{n-k+1} \\ &\quad + \sum_{\ell=n-k+2}^{n} q^{-ell} L_e(z). \end{aligned}$$

This (and notations (2.3) show that (4.1) holds true and then C is proper. □

Proof of Corollary 5. It is easily seen that (4.2) implies (4.1). Really,

$$\begin{aligned} &\sum_{i=d}^{\ell} \frac{\ell_{(i)}}{n_{(i)}} A_i - q\sum_{i=d}^{\ell-1} \frac{(\ell-1)_{(i)}}{n_{(i)}} A_i \\ &= \sum_{i=d}^{\ell-1} \frac{9\ell-1)\cdots(e-i+1)}{n_{(i)}} A_i(e-(e-i)q) + \frac{A_e}{\binom{n}{e}} \end{aligned}$$

and

$$e-(e-i)q = iq - e(q-1) \geq dq - n(q-1) \geq 0$$

by (4.2). □

5 Acknowledgements

The work has been done during a visit of S. M. Dodunekov to the Department of Information Theory, Chalmers University of Techology. He would like to thank Arne Svensson for his hospitality and Mrs. Eva Axelsson for her assistance. This work was partially supported by Bulgarian NSF under contract NMM-502/95.

References

Khaled A.S. Abdel-Ghaffar. A lower bound on the undetected error probability of block codes. In: Proc. of the ISIT'95, 1995 IEEE International Symposium on Information Theory, Whistler, Canada, 1995, p. 341.

J. K. Wolf, A. M. Michelson and A. H. Levesgue. On the probability of undetected error for linear block codes. IEEE Trans. Commun. vol. COM-30, No. 2, pp. 317-324, Feb. 1982.

T. Kløve. The weight distribution of cosets, IEEE Trans. Inf. Theory, vol. IT-40, No. 3, pp. 911-913, May 1994.

T. Fujiwara, T. Kasami, Shou-ping Feng. On the monotonic property of the probability of undetected error for a shortened code. IEEE Trans. Inf. Theory, vol. IT-37, No. 5, pp. 1403-1411, Sept. 1991.

R. Padovani and J. K. Wolf. Poor error correction codes are poor error detection codes. IEEE Trans. Inform. Theory, IT-30, No. 1, pp. 1110-111.

R. P. Varshanov. Estimate of the number of signals in error-correcting codes. Dokl. Akad. Nauk SSSR, vol. 117, pp. 739-741, 1957.

E. N. Gilbert. A comparison of signaling alphabets. Bell Syst. Tech. J. v. 31, pp. 504-532, May 1952.

P. Perry. Necesary conditions for good error detection. IEEE Trans. Inform. Theory, vol. IT-37, No. 2, pp. 375-378, March 1991.

T. Hashimoto. Good error-detection codes satisfying the expurgated bound. IEEE Trans. Inform. Theory, vol. IT-41, pp. 1347-1353, Sept. 1995.

T. Kasami, S. Lin. On the probability of undetected error for the Maximum Distance Separable codes. IEEE Trans. Commun. vol. COM-32, No. 9, 1984, pp. 998-1006.

E. R. Berlekamp. Algebraic Coding Theory. NY: McGraw-Hill, 1968.

S. Lin, D. J. Costello, Jr. Error Control Coding: Fundamentals and Applications. Englewood Cliffs, NY: Prentice-Hall, 1983.

F. J. MacWilliams and N. J. A. Sloane. The Theory of Error-Correcting Codes. Amsterdam: North-Holland, 1977.

W. W. Peterson and E. J. Weldon., Jr. Error-Correcting Codes. 2nd ed. Cambridge, MA: M.I.I. Press, 1972.

S. C. Chang and J. K. Wolf. A simple derivation of the MacWilliams identity for

linear codes. IEEE Trans. Inform. Theory, vol. IT-26, pp. 476-477, July 1980.

J. Massey. Coding techniques for digital data networks. In: Proc. Int. Conf. Inform. Theory and Syste., NTG-Fachberichte, vol. 65, Berlin, Germany, Sept. 18-20, 1978.

V. I. Korzlik. Bounds on undetected error probability and optimum group codes in a channel with feedback. Radiotechnika, vol. 20, No. 1, pp. 27-33, 1965. English translation in Telecommun. Radio Eng., vol. 20, pp. 87-92, Jan., 1965.

S. K. Leung-Yan-Cheong and M. E. Hellman. Concerning a bound on undetecated error probability. IEEE Trans. Inform. Theory., vol. IT-22, pp. 235-237, March 1976.

S. K. Leung-Yan-Cheong, E. R. Barnes and D. U. Friedman. Some properties of undetected error probability of linear codes. IEEE Trans. Inform. Theory, vol. IT-25, pp. 110-112, Jan. 1979.

T. Kasami, T. Kløve and S. Lin. Linear block codes for error detection. IEEE Trans. Inform. Theory, vol. IT-29, No. 1, pp. 131-136, Jan. 1983.

V. I. Levenshtein. Bounds on the probability of undetected error. Probl. Peredachi Inform., vol. 13, No. 1, pp. 3-18, 1977 (in Russian; English translation: Probl. Inform. Transmission, vol. 13, No. 1, pp. 1-12, 1978).

T. Kløve. Near-MDS codes for error detection. In: Proc. International workshop Optimal codes and related topics, May 26– June 1, 1995, Sozopol, Bulgaria, pp. 103-107.

R. Dodunekova and S. M. Dodunekov. On the probability of undetected error for near-MDS codes. Preprint No. 1995-25/ISSN 0347-2809, Chalmers University of Technology and Göteborg University, 1995.

T. Kløve. (The book).

G. L. Katsman. Upper bounds on the probability of undetected error. In Proc. fourth international workshop ACCT'94. Novgonod, Sept. 11-17, 1994, pp. 106-107.