International Journal of Pure and Applied Mathematics

Volume 46 No. 2 2008, 219-224

Invited Lecture Delivered at
Forth International Conference of Applied Mathematics
and Computing (Plovdiv, Bulgaria, August 12–18, 2007)

CONFIGURATIONS IN BINARY LINEAR CODES

Martin Dowd

1613 Wintergreen Pl.
Costa Mesa, CA 92626, USA
e-mail: MartDowd@aol.com

Abstract: Some theorems are proved about configurations in binary linear codes. A conjecture is made about configurations in linear double error correcting codes, which would lead to an upper bound, better by a constant factor than the sphere packing bound.

AMS Subject Classification: 94B05
Key Words: linear double error correcting codes, sphere packing bound

*

It has long been a great mystery of coding theory, why there is a gap of a factor of $\sqrt{2}$ between the maximum length of double error correcting codes given by the sphere packing bound, and the length of the best known linear double error correcting codes. In Dowd [3] it was observed that a better theory of these packings would seem to be a prerequisite to a theory of the analog of the Erdös-Turan conjecture for linear double error correcting codes. Some theorems were proved about configurations in such codes. For some later references concerning the Erdös-Turan conjecture, see Borwein et al [1], Nathanson [6], and Grekos et al [4].

In this extended abstract some further facts about such configurations are proved, and some conjectures made relevant to the main questions. The complete manuscript is available at the author's web site www.hyperonsoft.com.

Received: August 17, 2007

This describes some computational experiments concerning the conjectures.

We adopt the following notational conventions. $\mathcal{F}_q$ denotes the finite field of order q, for a prime power q. A binary code of length n is a subset of the vector space $\mathcal{F}_2^n$ over $\mathcal{F}_2$. Such a code is linear if it is a subspace; if k is the dimension the redundancy r is defined to be $n-k$, and the code is said to be an $[n,k]$ code. Codewords are generally denoted v, w, etc. Positions are generally denoted i, j, etc., with $1 \leq i \leq n$. As usual, v_i denotes the element of $\mathcal{F}_2$ in position i of v. A vector in $\mathcal{F}_2^n$ will be called a bit vector, of length n.

Suppose χ is a code containing 0, of minimum distance at least d where $d = 2\delta + 1$ is odd. Define the A configuration determined by χ to be the set of weight d vectors. Define $A(n)$ to be the largest size of an A configuration in a code of length n. For $v \in \chi$ let $\alpha_v = \{u - v : d(u,v) = d, u \in \chi\}$. Let $\bar{A}(\chi)$ be the

The following "codewise" version of the Johnson bound is stated in Brouwer and Tolhuizen [2], and may be proved by modifications to the proof in MacWilliams and Sloane [5].

Theorem 1. *With notation as above,*

$$|\chi|\left(1 + n + \binom{n}{2} + \cdots \binom{n}{\delta} + \frac{\binom{n}{\delta+1} - \binom{d}{\delta}\bar{A}(\chi)}{\lfloor n/(\delta+1)\rfloor}\right) \leq 2^n.$$

Suppose χ is a code with A configuration α, $v \in \alpha$, $1 \leq i \leq n$, and $t \in \mathcal{F}_2^n$ with $|t| = \delta$. Let

— $\alpha_i = \{v \in \alpha : i \in v\}$;

— $\alpha_t = \{v \in \alpha : t \subseteq v\}$;

— $\nu_v = \{w \in \alpha : |w - v| = d + 1\}$;

— $\nu_{vi} = \nu_v \cap \alpha_i$, where $i \in v$;

— $\nu_{vij} = \nu_v \cap \alpha_{\{i,j\}}$, where $i, j \in v$.

If $v, w \in \alpha$ then $|v - w| \geq d + 1$, so ν_v is the neighborhood of v within α. Also, $|v - w| = d + 1$ iff $|v \cap w| = \delta$. A ν_v is partitioned into the 10 classes ν_{vij}, for $i, j \in v$, $i < j$; ν_{vi} is the union of 4 of these.

A number of further definitions used below will now be given.

— An A_1 configuration is any α_i occurring in some code. $A_1(\chi)$ is the maximum of $|\alpha_i|$ for an α_i in χ. $A_1(n)$ is the maximum of $|\alpha_i|$ for an α_i in a code of length n.

— An N configuration is any ν_v occurring in some code. $N(\alpha)$ is the maximum of $|\nu_v|$ for a $v \in \alpha$. $N(n)$ is the maximum of $|\nu_v|$ for a ν_v in a code

of length n.

— An N_1 configuration is any ν_{vi} occurring in some code. $N_1(\alpha)$ is the maximum of $|\nu_{vi}|$ for $v \in \alpha$ and i a posdition of v. $N_1(n)$ is the maximum of $|\nu_{vi}|$ for a ν_{vi} in a code of length n.

— If χ is required to be linear, L is appended to the subscript, for A, A_1, N, or N_1, configurations or functions of n.

— Let ν'_v denote ν_v, restricted to v^c; and similarly for ν'_{vi} and ν'_{vij}.

The following theorem gives various facts about quantities just defined; some parts are given in Dowd [3], and some are well-known.

Theorem 2. *a. ν_v is the disjoint union of the sets $\alpha_t - \{v\}$; hence*

$$\sum_{t \subseteq v} |\alpha_t| = |\nu_v| + \binom{d}{\delta}.$$

b. $\binom{d}{\delta}|\alpha| = \sum_t |\alpha_t|$.

c. $\binom{d}{\delta}^2 |\alpha| \le \binom{n}{\delta}\left(N(\alpha) + \binom{d}{\delta}\right)$.

d. Two members of α_t intersect in t. Hence $|\alpha_t| \le (n-\delta)/(\delta+1)$.

e. $N(n) \le \binom{d}{\delta}\left\lfloor \dfrac{n-d}{\delta+1} \right\rfloor$.

f. $|\alpha| \le (n/d)A_1(\chi)$.

g. $|\alpha_i| \le \dfrac{n-1}{d-1}\left[\dfrac{\delta-1}{d-1}N_1(\alpha) + 1\right]$.

Proof. For part a, if $w \in \alpha_t - \{v\}$ then $w \in \nu_v$. If $w \in \nu_v$ then $w \cap v = t$ for a unique t, and $w \in \alpha_t - \{v\}$. For part b, counting pairs $\langle v, t\rangle$, the left side counts v first, and the right side counts t first. For part c, let i_t denote $|\alpha_t|$. Using parts a and b,

$$\sum_t i_t^2 = \sum_t i_t \,|\{v \in \alpha : t \subseteq v\}| = \sum_{v\in\alpha}\sum_{t\subseteq v} i_t = \sum_{v\in\alpha}|\nu_v| + \sum_t i_t.$$

For fixed $c = \sum_{v\in\alpha}|\nu_v|$, $\sum_t i_t$ is maximized, subject to $\sum_t (i_t^2 - i_t)$ equaling c, when the i_t are equal, say to the common value i, and it follows that

$$\binom{n}{\delta}(i^2 - i) = c.$$

By part b,

$$\binom{n}{\delta} i = \binom{d}{\delta}|\alpha|.$$

Finally, $c \le |\alpha| N(\alpha)$, and part c follows. For part d, by definition two members intersect in at least t, and they cannot intersect in a larger set. Part e follows by parts a and d. For part f, counting 1's in the incidence matrix of α, $\sum_i |\alpha_i| = d|\alpha|$. Since $|\alpha_i| \le A_1(\chi)$, $d|\alpha| \le nA_1(\chi)$. For part g, given an α_i, let α_i' be the incidence matrix with the common position i deleted. Let p_l be the number of rows which are 1 in column l. Let $q_j = |\nu_{vi}|$, where v is row j. Then

$$\sum_l p_l = (d-1)|\alpha_i|, \qquad (\delta - 1)\sum_j q_j = 2\sum_i \binom{p_i}{2}, \qquad \text{and} \quad q_j \le N_1(\alpha).$$

Given $c = (\delta-1)\sum_j q_j$, $\sum_l p_l$ is maximized subject to $2\sum_l \binom{p_l}{2} = c$ when the p_l are all equal, say to p. At equality $2(n-1)\binom{p}{2} = c \le (\delta-1)|\alpha_i|N_1(\alpha)$ and $(n-1)p = (d-1)|\alpha_i|$; part g follows.

From hereon only $\delta = 2$, $d = 5$ will be considered. In this case, Theorem 2. □

f becomes $|\alpha| \le (n/5)A_1(\chi)$; equality holds for any binary code with a transitive automorphism group, in particular for cyclic codes. Also, $|\nu_v| \le 10\lfloor (n-5)/3 \rfloor$. The bound is achieved in a 3-(n,5,1) design; these exist for $n = 4^m+1$ where $m \ge 1$ (see [1], Theorem 6.9). In such a design $|\alpha_t| = (n-2)/3$ for any pair t, and the claim follows. In such a design

$$|\alpha| = \frac{1}{10}\binom{n}{3} = \frac{1}{60}n(n-1)(n-2).$$

It is a question of interest how large a code exists, which has this as an A configuration.

Theorem 2. g becomes $|\alpha_i| \le (n-1)/4[N_1(\alpha)/4+1]$. This yields the usual bound $(n-1)(n-2)/12$ on $A_1(\chi)$ for an arbitrary code χ, when $N_1 = 4(n-5)/3$.

Configurations almost this large occur in the Preparata codes. These are defined when $n = 4^m - 1$ where $m \ge 2$. The weight 5 vectors form a 2-(n,5,$(n-3)/3$) design, and $|\alpha| = (1/60)n(n-1)(n-3)$ (MacWilliams and Sloane [5], Theorem 15.33). It follows that $|\nu_v| = (10/3)(n-6)$.

The Johnson bound when $d = 5$ is

$$|\chi|\left(1 + n + \binom{n}{2} + \frac{\binom{n}{3} - 10\,\bar{A}(\chi)}{\lfloor n/3 \rfloor}\right) \le 2^n.$$

Since 3-(n,5,1) designs exist, this does not yield an improvement almost everywhere to the sphere packing bound. The nonexistence of perfect codes (MacWilliams and Sloane [5]) does show that the sphere packing bound for double error correcting codes is met only finitely often. The Preparata codes (which are examples of "nearly perfect" codes) show, however, that for nonlin-

ear codes of minimum weight 5 the sphere packing bound is nearly tight.

The Johnson bound does yield an improvement sometimes. The bound $A(n) \leq \lfloor \frac{n}{5} \lfloor \frac{n-1}{4} \lfloor \frac{n-2}{3} \rfloor \rfloor \rfloor$ is well-known (MacWilliams and Sloane [5], Corollary 17.5). The weaker bound $\frac{1}{10}\binom{n}{2} \lfloor \frac{n-2}{3} \rfloor$ follows by Theorem 2.c and 2.e. It also follows by Theorem 2.g, and the fact that for $d = 5$, $N_1(n) \leq 4 \lfloor \frac{n-5}{3} \rfloor$. The importance of Theorem 2 lies in the hope that for linear codes, better bounds on $N(n)$ or $N_1(n)$ can be obtained. Indeed, a slight improvement is readily obtained.

Theorem 3. *Suppose $n - 5 = 3l + t$. Then $N_{1L}(n) \leq 4l - 3$ if $n \equiv 2$ or $n \equiv 3$ mod 6; and $N_{1L}(n) \leq 4l - 1$ if $n \equiv 0$ mod 6.*

Proof. For the first claim, suppose two ν_{vij} have size l, for some v, i. The sum of all the vectors in the two classes has weight at most $2t$ in the positions of v^c, and weight 2 in the positions of v. For the second claim, suppose all four ν_{vij} have size l. Let a denote the number of weight 3 columns. Counting flags, $12\,l = 4(3\,l + 1 - a) + 3\,a$, whence $a = 4$. The sum of all the rows has weight 0 in v, and weight 4 in v^c.

We conjecture that in fact, Theorem 3 is a weak bound, and that $N_{1L}(n)$ is $\leq c_1(n - 5)$ almost everywhere, for a constant c_1 smaller than 4/3. By Theorem 2, this conjecture would yield an upper bound on $A_L(n)$, better by a constant factor than the bound for arbitrary A configurations. This in turn would yield an upper bound on the length of a linear double error correcting code of redundancy r, better by a constant factor than the sphere packing bound.

We define a partial linear space to be an incidence matrix (matrix over $\mathcal{F}_2$), where two columns are incident to at most one row (no "rectangle" of 1's occurs). In the full manuscript it is shown that an N_1' configuration is a partial linear space, of constant row weight 3, together with a partition of the rows into 4 or fewer parts, such that in each part the rows are disjoint. It is also shown that an N_1' configuration is an N_{1L}' configuration iff the following holds. If S is a subset of the rows, let s_j be the number of rows of S in part j, for $1 \leq j \leq 4$. Then the weight of $\sum S$ must be at least 3,3,4,4, according to whether the number of odd s_j is 1,2,3,4 respectively; and if all s_j are even the weight must be either 0 or at least 5.

References

[1] P. Borwein, S. Choi, F. Chu, An old conjecture of Erdös-Turan on additive bases, *Math. Comp.*, **75** (2006), 475-484.

[2] A. Brouwer, L. Tolhuizen, A sharpening of the Johnson bound for binary linear codes and the nonexistence of linear codes with preparata parameters, *Designs, Codes and Cryptography*, **3** (1993), 95-98.

[3] M. Dowd, Questions related to the Erdös-Turan conjecture, *SIAM Journal in Discrete Mathematics*, **1** (1988), 142-150.

[4] G. Grekos, L. Haddad, C. Helou, J. Pihko, On the Erdös-Turan conjecture, *J. Number Theory*, **102** (2003), 339-352.

[5] F. MacWilliams, N. Sloane, *The Theory of Error-Correcting Codes*, North Holland (1977).

[6] M. Nathanson, Generalized additive bases, Konigs lemma, the Erdös-Turan conjecture, *Preprint.*