Error correcting codes arising from cubes

Sheng Bau
School of Mathematics, University of the Witwatersrand,
Private Bag 3, Wits 2050, Johannesburg, South Africa

e-mail: sheng.bau@wits.ac.za

and School of Mathematics,
Inner Mongolia University of Nationalities,
Tongliao, China

Abstract

In this note, experimental results on binary nonlinear error correcting codes arising from special families of graphs will be reported. The graph Q_n of the n-dimensional cube provides binary nonlinear error correcting codes with high capacity of error correction. The codes are given by maximum induced forests of a specific type in cubes, obtained by deletion of a (minimum) decycling set. For C a cycle of length $2k+\delta$ where $\delta \in \{0,1\}$, we also determine the decycling number $\phi(K_2 \Box C) = k+1$. Many interesting problems remain open in this topic.

Keywords: codes, decycling, Hamming distance, induced forest

1 Introduction and Backgrounds

The capacity of error correction of a code is measured by a given metric over the set of codewords (see [6]). Nonlinear codes usually provide a large number of codewords with a prescribed minimum distance (i.e., error correction capacity). This is an advantage of nonlinear codes over linear codes. An (n, M, d)-*code* is a set $C \subseteq V$ with $|C| = M$ and each $v \in C$ is an n-dimensional vector in a vector space V over a division ring $\mathbb{F}$, such that the (Hamming) distance between any two elements of C is $\geq d$, and d is the smallest integer with this property. The division ring $\mathbb{F}$ is usually taken to be a finite field $\mathbb{F}_q$. If $\mathbb{F} = \mathbb{F}_2$ then the code is called *binary*.

Each n-dimensional binary vector $(a_1, \cdots, a_n)$ is a vertex of Q_n, the graph of the n-dimensional cube. Each (n, M, d)-code is a subset of $V(Q_n)$. The problem of finding a large code with high capacity of error correction is a sphere packing problem in cubes. If the codes have minimum Hamming distance d, the Euclidean distance between codewords is at least $\sqrt{d}$. Thus, finding an (n, M, d)-code is equivalent to finding M non-overlapping spheres of radius $\sqrt{d}$ with centers at vertices of Q_n. The graphic distance between vertices of Q_n is exactly the Hamming distance between the corresponding codewords. Let $S \subset V(Q_n)$. Spheres in Q_n are precisely stars in Q_n. Hence if $Q_n - S$ is a set of vertex disjoint stars then the centers of these stars give rise to an (n, M, d)-code C with a reasonable d and large M. The code obtained in this way is a large nonlinear binary code with a prescribed error correction capacity d. This is the motivation of this note.

Note that the graph Q_n is five things at the same time. Firstly, Q_n is the geometrical object that is the 1-dimensional skeleton of the n-dimensional unit cube. Secondly, it is the graph defined recursively: $Q_1 = K_2$, $Q_n = K_2 \Box Q_{n-1}$. Thirdly, $V(Q_n)$ is the finite field $\mathbb{F}_{2^n}$ with

$$E(Q_n) = \{xy : x, y \in \mathbb{F}_{2^n}, x + y \in S\}$$

where

$$S = \{e_1, \cdots, e_n\}$$

is the set of standard basis vectors; note that $-S = S$ and Q_n is a Cayley graph; of course, the multiplicative group of $\mathbb{F}_{2^n}$ is cyclic. Fourthly, $V(Q_n)$ is the vector space of dimension n over $GF(2) = \mathbb{F}_2$. And fifthly, Q_n is the rich home of binary codes. The natural properties of each of these five aspects may be employed for our purposes.

The minimum number of edges whose removal eliminates all cycles in a given graph has been known as the *cycle rank* (that is, the dimension of the cycle space) of the graph. There is an expression of this number in terms of the three basic graph parameters: order, size and the number of components. This is the well known *Betti number*:

$$b(G) = \|G\| - |G| + \omega.$$

The corresponding problem for deletion of vertices is usually difficult. Let $G = (V, E)$ be a graph and $S \subseteq V(G)$. If for each cycle $C \subseteq G$, $V(C) \cap S \neq \emptyset$, then S is called a *decycling set* of G. The *decycling number* of G is the smallest cardinality of a decycling set of G. The *decycling* or *feedback* number of G is denoted by $\phi(G)$. A decycling set with cardinality $\phi(G)$ is called a *minimum decycling set* of G. Clearly, $S \subset V(G)$ is a decycling set if and only if $F = G - S$ is an induced forest of G; and $S \subset V(G)$ is a minimum decycling set if and only if $F = G - S$ is a maximum induced forest of G. Denote by $\varphi(G)$ the order of a maximum induced forest of G. Then we have

$$\varphi(G) + \phi(G) = |G|.$$

Determination of the decycling number of an arbitrary graph is NP-complete [5].

Clearly, $\phi(G) = 0$ if and only if G is a forest, and $\phi(G) = 1$ if and only if G has at least one cycle and a vertex is on all of its cycles. It is also easy to see that $\phi(K_n) = n - 2$ and $\phi(K_{r,s}) = r - 1$ if $r \leq s$.

Let $\alpha(G)$ and $\beta(G)$ be the independence and covering numbers of G respectively. Then

$$\alpha(G) + \beta(G) = |G|.$$

For each nonempty graph G, $\phi(G) \leq \beta(G) - 1$.

2 Direct Products

For graphs H and J, the *direct product* $H\Box J$ is defined by assignment

$$V(H\Box J) = V(H) \times V(J)$$

$$E(H\Box J) = \{\{(u,v),(x,y)\} : [u = x \wedge vy \in E(J)] \vee [v = y \wedge ux \in E(H)]\}.$$

Suppose that H and J are both bipartite graphs. Let the bipartition of H be (A, B) and that of J be (A', B'). Let

$$X = \{(u,v) : u \in A, v \in B'\},\ Y = \{(x,y) : x \in B, y \in A'\}.$$

Since J is bipartite, if $u = x$ then $vy \notin E(J)$, hence $\{(u,v),(x,y)\} \notin E(H\Box J)$. Since H is bipartite, if $v = y$ then $ux \notin E(H)$, hence $\{(u,v),(x,y)\} \notin E(H\Box J)$. Hence $\{(u,v),(x,y)\} \notin E(H\Box J)$ for all $(u,v) \in X$ and all $(x,y) \in Y$. Thus the proof of the following result is complete.

LEMMA 2.1. *If H and J are both bipartite graphs then the direct product $G = H\Box J$ is also bipartite.*

□

Now $Q_1 = K_2$ is bipartite. Suppose that Q_{n-1} is bipartite. Then by Lemma 2.1, $Q_n = K_2\Box Q_{n-1}$ is also bipartite. Hence Q_n is bipartite for all $n \geq 1$. We have

COROLLARY 2.1. *For each $n \geq 1$, Q_n is bipartite.*

□

This corollary may of course be proved in another manner. Let X be all binary strings with an even number of 1s and let Y be that with an odd number of 1s. Then for each $x \in X$ and each $y \in Y$, $x + y \notin \{e_1, \cdots, e_n\}$. Hence $E(Q_n) \subseteq [X, Y]$ and therefore Q_n is bipartite.

LEMMA 2.2. *For any graph H,*

$$2\phi(H) \leq \phi(K_2\Box H) \leq \phi(H) + \beta(H).$$

□

The equalities in Lemma 2.2 are satisfied by a graph of each order. Let $H = \bar{K}_n$. Then $\phi(K_2\Box H) = \phi(H) = 0$ and both equalities hold. For the equality to the lower bound, if $n \geq 2$, then $\phi(K_2\Box K_n) = 2n - 4 = 2\phi(K_n)$. The path of order n gives equality to the upper bound.

In addition to these initial examples, we also have

THEOREM 2.1. *Let $C_{2k+\delta}$ be a cycle with $|C| = 2k + \delta \geq 4$ where $\delta \in \{0, 1\}$. Then*

$$\phi(K_2 \Box C_{2k+\delta}) = k + 1.$$

PROOF: A more specific statement will be proved using induction on k. In fact, we prove

> (A) $K_2 \Box C_{2k+\delta}$ has a minimum decycling set consisting of just one vertex from $0 \times C_{2k+\delta}$ and a minimum cover from $1 \times C_{2k+\delta}$, and for $\delta = 0$ such a decycling set is a subset of one part of the bipartition.

The assertion is trivial for $k = 2$. Assume that $k \geq 3$. Suppose that $\delta = 0$. Suppose that (A) holds for $K_2 \Box C_{2k-2}$. Then $K_2 \Box C_{2k-2}$ has a minimum decycling set T consisting of one white vertex from $0 \times C_{2k-2}$ and all white vertices of $1 \times C_{2k-2}$. Thus, $|T| = k$. Since T is a minimum decycling set in $K_2 \Box C_{2k-2}$, any decycling set of $K_2 \Box C_{2k}$ must have cardinality at least $k + 1$ since the new 4-cycle must be broken. But by Lemma 2.2, $\phi(K_2 \Box C_{2k}) \leq k + 1$. Hence $\phi(K_2 \Box C_{2k}) = k + 1$. Let $a \in 0 \times C_{2k-2}$ which is white. Then $S := T \cup \{a\}$ is a minimum decycling set of $K_2 \Box C_{2k}$ satisfying condition (A).

The case for $\delta = 1$ is similarly proved, only that the condition about the minimum decycling set being a subset of a part of the bipartition is dropped. □

Note that the reduction used in this proof does not work for Q_n. It is natural to ask the following questions.

PROBLEM 2.1. *Suppose that H is a bipartite graph. Does the graph $G = K_2 \Box H$ have a minimum decycling set which is a subset of a part of the bipartition of G?*

As a special case, one might consider regular bipartite graphs for H.

PROBLEM 2.2. *For which bipartite graphs H, the graph $G = K_2 \Box H$ has a minimum decycling set which consists of one vertex from $0 \times H$ and a minimum cover of $1 \times H$?*

3 Cubes

The field $\mathbb{F}_{2^n}$ may be generated computationally, element by element, using the simple algorithm of addition in $\mathbb{F}_2$ of 1 at a time, starting from the zero vector. Each element is generated exactly once in this procedure, and hence it takes 2^n bit operations and a memory of $2^n \cdot n$ bits. This is linear in the order of $\mathbb{F}_{2^n}$. The verification of the adjacency and hence the generation of edges of the graph Q_n is also simple: since Q_n is a Cayley graph on the additive group of $\mathbb{F}_{2^n}$ with the

standard basis as generating set, it needs only to check $x + y = e_i$ for some i with $1 \leq i \leq n$ in order to decide whether $xy \in E(Q_n)$. Let $v \in \mathbb{F}_{2^n}$. The vector v is called *odd* or *even* if it has an odd or even number of 1s as components. Note that this is the 1-norm of v whose value is always a nonnegative integer. The bipartition of Q_n is easily seen by letting

$$X = \{v : v \equiv 0 (\text{mod } 2)\},\ Y = \{v : v \equiv 1 (\text{mod } 2)\}.$$

Since $xy \in E(Q_n)$ if and only if $x+y = e_i$ for some i with $1 \leq i \leq n$, there is no edge in $Q_n|_X$ or $Q_n|_Y$. Hence Q_n is bipartite. There is also a simple relation between the coordinatization and the recursive definition. Since the recursion is $Q_n = K_2 \Box Q_{n-1}$, if the vertices of K_2 are 0 and 1, then the vertices of Q_n are just the vertices of Q_{n-1} with a 0 or a 1 adjoined onto it (say from the left). The recursive definition also implies that for $n \geq 2$ the graph Q_n is hamiltonian: let C be a hamiltonian cycle of Q_{n-1} (which we assume is hamiltonian) and assume that $ab \in E(C)$, then

$$H = 0 \times (C - ab) \cup \{(0, a), (1, a), (0, b), (1, b)\} \cup 1 \times (C - ab)$$

is a hamiltonian cycle of Q_n. By König's Theorem on matchings and coverings, we also know that the covering number (the least cardinality of a cover) $\beta(Q_n) = 2^{n-1}$ since Q_n has a 1-factor with this size. This simple observation provides an upper bound on a minimum decycling set of Q_n, and hence a lower bound on the cardinality of a nonlinear error correcting binary code, if the complement of a (minimum) decycling set is a vertex disjoint union of stars.

$$\phi(Q_n) \leq \phi(Q_{n-1}) + 2^{n-1}. \tag{1}$$

If C is a binary error correcting code arising from Q_n as described above, then

$$|C| \geq 2^{n-1} - \phi(Q_{n-1}). \tag{2}$$

Thus we have proved

PROPOSITION 3.1. *If S is a minimum decycling set of Q_n such that $Q_n - S$ is the vertex disjoint union of stars. Then a binary error correcting code C given by these stars satisfies*

$$|C| \geq 2^{n-1} - \phi(Q_{n-1}).$$

□

As in [6], denote by $A(n, d)$ the maximum cardinality of a binary code of length n with minimum Hamming distance d.

The decycling numbers of cubes and their relations to codes have been considered in [1, 2, 7]. The following results are from [7]. The main work of [7] is a study of conditions under which $A(n, d)$ and $\phi(Q_n)$ are related, where the following have been proved. Let X and Y be the parts of the bipartition of Q_n as given above.

LEMMA 3.1. *Let $C \subseteq X$ be a binary $(n, 4)$-code and let $S = X - C$. Then S is a decycling set of Q_n.*

Let $C \subseteq X$ be a binary $(n, 4)$-code with $|C| = A(n, 4)$. Let $S = X - C$. Then

$$|S| = 2^{n-1} - |C| = 2^{n-1} - A(n, 4)$$

and by Lemma 3.1, S is a decycling set of Q_n. We have

THEOREM 3.1. $\phi(Q_n) \leq 2^{n-1} - A(n, 4)$.

□

For n a power of 2, it is known that $A(n, d) = 2^{n-1} - 2^{n-r-1}$ where for x with $0 \leq x < 2^{r-1}$, $n = 2^r - x$. Suppose inductively that $\phi(Q_{n-1}) \leq 2^{n-2} - 2^{n-r-2}$. Then by

$$\phi(Q_n) \leq \phi(Q_{n-1}) + \beta(Q_{n-1}) = \phi(Q_{n-1}) + 2^{n-2}$$

we have

$$\phi(Q_n) \leq 2^{n-1} - 2^{n-r-2} + 2^{n-2} < 2^{n-1} - 2^{n-r-1}.$$

Hence the proof of the following is complete.

COROLLARY 3.1. *Let $n = 2^r - x$ with $0 \leq x < 2^{r-1}$. Then $\phi(Q_n) \leq 2^{n-1} - 2^{n-r-1}$.*

□

The last step of this proof is a little loose and it seems there is room for improvement. Note that for $n \leq 8$ the upper bound in Theorem 3.1 is equal to $\phi(Q_n)$:

n	Q_2	Q_3	Q_4	Q_5	Q_6	Q_7	Q_8
ϕ	1	3	6	14	28	56	112

These were determined in [3]. The decycling set described in Theorem 3.1 is an independent set. Thus the question of whether this is always true.

PROBLEM 3.1. (Pike, 2003) *Does there exist a minimum decycling set of Q_n that is independent?*

CONJECTURE 3.1. (Pike, 2003) $\phi(Q_n) = 2^{n-1} - A(n, 4)$.

Another main result of [7] is an assertion that the answer to Problem 3.1 is "yes" if and only if Conjecture 3.1 is true.

THEOREM 3.2. *The cube Q_n has a minimum independet decycling set if and only if $\phi(Q_n) = 2^{n-1} - A(n, 4)$.*

A slightly improved lower bound for the decycling number of Q_n was also given in [7].

THEOREM 3.3. *For each $n \geq 7$, $\phi(Q_n) \geq 2^{n-1} - \dfrac{2^{n-1} - n - 1}{n-1}$.*

With this lower bound and direct calculations in [2], we have an update on bounds for $\phi(Q_n)$ for $9 \leq n \leq 13$.

$$\begin{array}{rcl} 226 & \leq \phi(Q_9) \leq & 236 \\ 456 & \leq \phi(Q_{10}) \leq & 472 \\ 922 & \leq \phi(Q_{11}) \leq & 952 \\ 1862 & \leq \phi(Q_{12}) \leq & 1904 \\ 3755 & \leq \phi(Q_{13}) \leq & 3840 \end{array}$$

4 Codes from Cubes

Consider Q_9. There exists a decycling set of cardinality 237 in one part of the bipartition (and hence is an independent set). The deletion of the set from Q_9 results in 19 copies of $K_{1,9}$ and a set of isolated vertices. The cube Q_9 may be visualized as 64 copies of Q_3 arranged in an 8×8 grid. In the coordinatization of Q_9 in this arrangement, the last three entries are given by the usual coordinatization of Q_3. The fourth coordinate counted from the last is 0 if the vertex is in the columns $1, 3, 5, 7$ and 1 if it is in columns $2, 4, 6, 8$; the fifth entry of the coordinate of a vertex is 0 if it is in columns $1, 2, 5, 6$ and 1 if it is in $3, 4, 7, 8$; the sixth entry of the coordinate of a vertex is 0 if it is in columns $1, 2, 3, 4$ and 1 if it is in columns $5, 6, 7, 8$; the seventh entry of the coorinate of a vertex is 0 if it is in rows $1, 3, 5, 7$ and 1 if it is in rows $2, 4, 6, 8$; the eighth from the last of the coordinate of a vertex is 0 if it is in rows $1, 2, 5, 6$ and 1 if it is in rows $3, 4, 7, 8$; and the ninth entry of the coordinate of a vertex is 0 if it is in rows $1, 2, 3, 4$ and 1 if it is in rows $5, 6, 7, 8$. The centers of the 19 copies of $K_{1,9}$ are at grids:

	grid	last three digits
1	$(1,1)$	000
2	$(1,4)$	101
3	$(1,7)$	010
4	$(2,2)$	010
5	$(2,5)$	110
6	$(2,8)$	000
7	$(3,6)$	001
8	$(3,7)$	100
9	$(4,1)$	101
10	$(4,4)$	110
11	$(5,1)$	111
12	$(5,6)$	010
13	$(6,3)$	100
14	$(6,6)$	101
15	$(7,2)$	100
16	$(7,3)$	010
17	$(7,8)$	111
18	$(8,4)$	001
19	$(8,5)$	010

From this table we have a binary $(19, 9, 4)$-code:

$$\begin{pmatrix}
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 1 \\
0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 \\
0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 \\
0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 \\
0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\
0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 1 \\
0 & 1 & 0 & 1 & 1 & 0 & 1 & 0 & 0 \\
0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 1 \\
0 & 1 & 1 & 0 & 1 & 1 & 1 & 1 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\
1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\
1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 \\
1 & 0 & 1 & 1 & 0 & 1 & 1 & 0 & 1 \\
1 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\
1 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\
1 & 1 & 0 & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 1 \\
1 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 1
\end{pmatrix}$$

Presently, we have not worked out codes from further cubes such as Q_{10}. Note also that $\phi(Q_9) \leq 236$. Hence the decycling set of 237 we used is larger than the actual minimum value. Thus even for Q_9 the question of a binary code larger than the one above remains open.

The work of this paper was completed while the author was at the ***Center for Discrete Mathematics, Fuzhou University, Fuzhou China*** *under a support of NSF (Natural Science Foundation of China) Grant No.: 10971027.*

References

[1] S. Bau and L.W. Beineke: The decycling number of graphs, *Australasian Journal of Combinatorics* 25(2002), 285-298.

[2] S. Bau, L.W. Beineke, G-M. Du, Z-S. Liu and R.C. Vandell: Decycling cubes and grids, *Utilitas Mathematica* 59(2001), 129137.

[3] L.W. Beineke and R.C. Vandell: Decycling graphs, *Journal of Graph Theory* 25(1996), 59-77.

[4] P. Erdős, M. Saks and V.T. Sós: Maximum induced trees in graphs, *Journal of Combinatorial Theory* B41(1986), 61-79.

[5] R.M. Karp: Reducibility among combinatorial problems, in Complexity of Compputer Computations (edt. R.E. Miller and J.W. Thatcher), Plenum Press, New York-London 1972, 85-103.

[6] J.H. van Lint: Introduction to Coding Theory, Springer Verlag, New York 1982.

[7] D.A. Pike: Decycling hypercubes, *Graphs and Combinatorics* 19(2003), 547-550.