New Bounds on a Hypercube Coloring Problem

Hung Quang Ngo[*] Ding-Zhu Du[†] Ronald L. Graham[‡]

Abstract

In studying the scalability of optical networks, one problem which arises involves coloring the vertices of the n-cube with as few colors as possible such that any two vertices whose Hamming distance is at most k are colored differently. Determining the exact value of $\chi_{\bar{k}}(n)$, the minimum number of colors needed, appears to be a difficult problem. In this note, we improve the known lower and upper bounds of $\chi_{\bar{k}}(n)$ and indicate the connection of this coloring problem to linear codes.

Keywords: Combinatorial problems, n-cube, coloring.

[*]Department of Computer Science and Engineering, State University of New York, Buffalo, NY14260, USA. e-mail: `hungngo@cs.buffalo,edu`.

[†]Department of Computer Science and Engineering, University of Minnesota, 200 Union street, EE/CS Building, room 4-192, Minneapolis, MN 55455, USA. e-mail: `dzd@cs.umn.edu`. Support in part by by the National Science Foundation under grant CCR-9530306.

[‡]Department of Computer Science, University of California at San Diego, La Jolla, CA92093. e-mail: `rgraham@cs.ucsd.edu`.

1 Introduction

An n-cube (or n-dimensional hypercube) is a graph whose vertices are the vectors of the n-dimensional vector space over the field $GF(2)$. There is an edge between two vertices of the n-cube whenever their Hamming distance is exactly 1, where the Hamming distance between two vectors is the number of coordinates in which they differ. Given n and k, the problem we consider is that of finding $\chi_{\bar{k}}(n)$, the minimum number of colors needed to color the vertices of the n-cube so that any two vertices at (Hamming) distance of at most k have different colors. This problem originally arised in the study of the scalability of optical networks [2].

It was shown by Wan [4] that

$$n + 1 \leq \chi_{\bar{2}}(n) \leq 2^{\lceil \log_2 (n+1) \rceil} \tag{1}$$

and it was conjectured that the upper bound is the truth, i.e.,
$\chi_{\bar{2}}(n) = 2^{\lceil \log_2 (n+1) \rceil}$

Kim et al. [1] showed that

$$2n \leq \chi_{\bar{3}}(n) \leq 2^{\lceil \log_2 n \rceil + 1} \tag{2}$$

$$\left(\binom{n}{k/2}\right) \leq \chi_{\bar{k}}(n) \leq (k+1)\left(\frac{k+2}{2}\right)^{\frac{k(k+2)}{8}\lceil \log_2 n \rceil} \tag{3}$$

and

$$2\left(\binom{n-1}{\frac{k-1}{2}}\right) \leq \chi_{\bar{k}}(n) \leq (k+1)\left(\frac{k+2}{2}\right)^{\frac{k(k+2)}{8}\lceil \log_2 n \rceil} \tag{4}$$

where $\left(\binom{n}{m}\right) = \sum_{i=0}^{m} \binom{n}{i}$

The upper bounds in (1) and (2) are fairly tight. In (1), the upper and lower bounds coincide when $n+1$ is an exact power of 2, and the same is true for (2) when n is a power of 2. However, the upper bounds in (3) and (4) are not very tight. In fact, when $k = 2$ and 3, they are already different from those of (1) and (2). A natural approach for getting an upper bound for $\chi_{\bar{k}}(n)$ is to find a valid coloring of the n-cube with as few colors as possible. We shall use this idea and various properties of linear codes (to be introduced in the next section) to give tighter bounds for general k which imply (1) when $k = 2$ and (2) when $k = 3$.

2 Main Results

We show the following.

Theorem 1 *Let $t = \lfloor \frac{k}{2} \rfloor$ and let $\left(\binom{n}{m}\right)$ denote $\sum_{i=0}^{m} \binom{n}{i}$. Then, when k is even, we have*

$$\sum_{i=0}^{t} \binom{n}{i} + \frac{1}{\left\lfloor \frac{n}{t+1} \right\rfloor} \binom{n}{t} \left(\frac{n-t}{t+1} - \left\lfloor \frac{n-t}{t+1} \right\rfloor \right) \leq \chi_{\bar{k}}(n) \leq 2^{\left\lfloor \log_2 \left(\binom{n-1}{k-1}\right) \right\rfloor + 1}$$

and when k is odd, we have

$$2 \left(\sum_{i=0}^{t} \binom{n-1}{i} + \frac{1}{\left\lfloor \frac{n-1}{t+1} \right\rfloor} \binom{n-1}{t} \left(\frac{n-1-t}{t+1} - \left\lfloor \frac{n-1-t}{t+1} \right\rfloor \right) \right) \leq \chi_{\bar{k}}(n) \leq 2^{\left\lfloor \log_2 \left(\binom{n-2}{k-2}\right) \right\rfloor + 2}$$

Note that since

$$2^{\left\lfloor \log_2 \left(\binom{n-1}{2-1}\right) \right\rfloor + 1} = 2^{\lfloor \log_2 n \rfloor + 1} = 2^{\lceil \log_2 (n+1) \rceil}$$

and

$$2^{\left\lfloor \log_2 \left(\binom{n-2}{3-2}\right) \right\rfloor + 2} = 2^{\lfloor \log_2 (n-1) \rfloor + 2} = 2^{\lceil \log_2 n \rceil + 1},$$

the inequalities (1) and (2) are direct consequences of this theorem. In general, the bounds in this theorem improve previous ones (3) and (4), especially, the upper bounds are much better. In fact,

$$\begin{aligned} (k+1) \left(\frac{k+2}{2} \right)^{\frac{k(k+2)}{8} \lceil \log_2 n \rceil} &= n^{\Omega(k^2 \log k)} \\ 2^{\left\lfloor \log_2 \left(\binom{n-1}{k-1}\right) \right\rfloor + 1} &= n^{O(k)} \\ 2^{\left\lfloor \log_2 \left(\binom{n-2}{k-2}\right) \right\rfloor + 2} &= n^{O(k)}. \end{aligned}$$

Since

$$\begin{aligned} \sum_{i=0}^{t} \binom{n}{i} + \frac{1}{\left\lfloor \frac{n}{t+1} \right\rfloor} \binom{n}{t} \left(\frac{n-t}{t+1} - \left\lfloor \frac{n-t}{t+1} \right\rfloor \right) &= n^{\Omega(k)} \\ 2 \left(\sum_{i=0}^{t} \binom{n-1}{i} + \frac{1}{\left\lfloor \frac{n-1}{t+1} \right\rfloor} \binom{n-1}{t} \left(\frac{n-1-t}{t+1} - \left\lfloor \frac{n-1-t}{t+1} \right\rfloor \right) \right) &= n^{\Omega(k)} \end{aligned}$$

our lower and upper bounds are close in some sense.

We first prove the lower bounds.

Given a valid coloring of the n-cube with parameters n and k using m colors, let $S_i, 1 \le i \le m$, be the set of vertices which are colored i. Note that each vertex of the n-cube is a binary string of length n and for any two vertices $u = u_1u_2\ldots u_n$ and $v = v_1v_2\ldots v_n$ in S_i, their *Hamming distance* $d(u,v) = |\{i : u_i \ne v_i\}|$ is at least $k+1$. Therefore, for each i, S_i forms a binary $(n, |S_i|, d_i)$-code for some $d_i \ge k+1$. (A binary (n,m,d)-code is a set of m binary strings of length n such that the least Hamming distance between any two strings in the set is d [3].) Let $A(n,d)$ denote the largest size M such that a binary (n,M,d)-code exists. Since $A(n,d)$ is decresing in d, we have

$$2^n = \sum_{i=1}^{m} |S_i| \le \sum_{i=1}^{m} A(n,d_i) \le mA(n,k+1)$$

Thus, we have

$$\chi_{\bar{k}}(n) \ge m \ge \frac{2^n}{A(n,k+1)}.$$

Moreover, it follows from Johnson bound[3] that

$$A(n,2t+1) \le \frac{2^n}{\sum_{i=0}^{t}\binom{n}{i} + \frac{1}{\lfloor \frac{n}{t+1} \rfloor}\binom{n}{t}\left(\frac{n-t}{t+1} - \left\lfloor \frac{n-t}{t+1} \right\rfloor\right)} \tag{5}$$

When $k = 2t$, by (5) we obtain

$$\chi_{\bar{k}}(n) \ge \sum_{i=0}^{t}\binom{n}{i} + \frac{1}{\lfloor \frac{n}{t+1} \rfloor}\binom{n}{t}\left(\frac{n-t}{t+1} - \left\lfloor \frac{n-t}{t+1} \right\rfloor\right)$$

When $k = 2t+1$, we note the relation $A(n,2t+1) = A(n+1,2t+2)$ (see [3]). Then by (5) again,

$$\begin{aligned}
\chi_{\bar{k}}(n) &\ge \frac{2^n}{A(n,k+1)} \\
&= \frac{2^n}{A(n,2t+2)} \\
&= \frac{2^n}{A(n-1,2t+1)} \\
&\ge 2\left(\sum_{i=0}^{t}\binom{n-1}{i} + \frac{1}{\lfloor \frac{n-1}{t+1} \rfloor}\binom{n-1}{t}\left(\frac{n-1-t}{t+1} - \left\lfloor \frac{n-1-t}{t+1} \right\rfloor\right)\right)
\end{aligned}$$

To show the upper bounds, let us first recall some concepts about linear code.

The set of all n-dimensional vectors over $GF(2)$ forms an n-dimensional vector space, which we denote by $V_n(2)$. A code $C \subset V_n(2)$ is called a *linear code* if it is a linear subspace of $V_n(2)$. Moreover, C is called a *$[n, m]$-code* if it has dimension m. If C also has minimum distance d then it is called an *$[n, m, d]$-code*. The square brackets will automatically refer to linear codes. An $m \times n$ matrix G is called a *generator matrix* of an $[n, m]$-code C if its rows form a basis for C. Given an $[n, m]$-code C, an $(n - m) \times n$ matrix H is called a *parity check matrix* for C if $c \in C$ implies $cH^T = 0$. From coding theory, we know that specifying a linear code by using its generator matrix and using a parity check matrix are equivalent. For a vector $x \in V_2(n)$, the *syndrome* of x associated with a parity check matrix H is defined to be $synd(x) = xH^T$.

We next explain that if a linear $[n, m, k + 1]$-code exists, then $\chi_{\bar{k}}(n) \leq 2^{n-m}$.

Given an $[n, m, k + 1]$-code C, the *standard array* of C is a $2^{n-m} \times 2^m$ table where each row is a (left) coset of C. This table is well defined since elements of C form an Abelian subgroup in $V_2(n)$ under addition (and the cosets of a group partition the group uniformly). The first row of the standard array contains C itself. The first column of the standard array contains the minimum weight elements from each coset. These are called *coset leaders*. Each entry in the table is the sum of the codeword on the top of its column and its coset leader. Since each pair of distinct codewords has Hamming distance at least $k + 1$, each pair of elements in the same row also has Hamming distance at least $k + 1$. Thus, coloring each row of C's standard array by a different color would give us a valid coloring. The number of colors used is 2^{n-m}, which is the number of rows of C's standard array. Therefore, $\chi_{\bar{k}}(n) \leq 2^{n-m}$.

It is a basic fact from coding theory that all elements in the same row of the standard array have the same syndrome and different rows have different syndromes. Therefore, we can use its parity check matrix H to color each vector $x \in V_2(n)$ with $synd(x) = xH^T$.

Now, we want to construct a linear $[n, n - p, k + 1]$-code where

$$p = \left\lfloor \log_2 \left(1 + \binom{n-1}{1} + \binom{n-1}{2} + \dots \binom{n-1}{k-1} \right) \right\rfloor + 1 = \left\lfloor \log_2 \left(\left(\binom{n-1}{k-1} \right) \right) \right\rfloor + 1.$$

To do so, we employ the following result from [3].

Lemma 1 *If H is an $(n - m) \times n$ matrix where any $d - 1$ columns of H are linearly independent and there exist d linearly dependent columns in H, then H is the parity check matrix of an $[n, m, d]$-code.*

By Lemma 1, it suffices to build an $p \times n$ matrix H with the property that $k + 1$ is the largest number d such that any d columns of H are linearly independent and there exist d

dependent columns. This H is actually the parity check matrix of an $[n, n-p, k+1]$-code

Now, we describe a procedure for constructing a $p \times n$ parity check matrix H by choosing its column vectors sequentially. The first column vector can be any non-zero vector. Suppose we already have a set V of i vectors so that any k of them are linearly independent. The $(i+1)^{th}$ vector can be chosen as long as it is not in the span of any $k-1$ vectors in V. In other words, since we are working over the field $GF(2)$, the new vector cannot be the sum of any $k-1$ or fewer vectors in V. The total number of *unallowable* vectors is at most $\binom{i}{1} + \binom{i}{2} + \dots \binom{i}{k-1}$ (this is an increasing function of i). Consequently, as long as $\binom{i}{1} + \binom{i}{2} + \dots \binom{i}{k-1} < 2^p - 1$ then we can still add a new column to H. Note that

$$\binom{n-1}{1} + \binom{n-1}{2} + \dots \binom{n-1}{d-2} < 2^p - 1.$$

Therefore, a $p \times n$ parity check matrix H can be constructed. It follows that

$$\chi_{\bar{k}}(n) \leq 2^p = 2^{\lfloor \log_2(\binom{n-1}{k-1}) \rfloor + 1}$$

This inequality holds regardless of k being odd or even and thus proves the upper bound for the even k case. However, when k is odd we are able to do better.

Notice that if we add an even parity bit to each vector of $V_2(n-1)$ then we get half of $V_2(n)$. Adding an odd parity bit would give us the other half. When k is odd, we just proved that we can color the $(n-1)$-cube using $a = 2^{\lfloor \log_2(\binom{n-2}{k-2}) \rfloor + 1}$ colors so that if two vertices have the same color then their distance is at least k. From this, we can obtain a coloring of the n-cube as follows. We first add an even parity bit to each vertex of the $(n-1)$-cube, color them using a colors, and then add an odd parity bit and color them using a completely different set of a colors. This is clearly a coloring of the n-cube using $2a = 2^{\lfloor \log_2(\binom{n-2}{k-2}) \rfloor + 2}$ colors. What remains to be shown is that this coloring is valid with parameters n and k.

For any vertex x of the n-cube, let x' be the vector obtained from x by deleting the parity bit just added. By the way we constructed the coloring, if two vertices x and y of the n-cube have the same color then $d(x', y') \geq k$, and the same type of parity bit (even or odd) was added to them to get x and y. It is clear that if $d(x', y') \geq k+1$, then $d(x, y) \geq k+1$. If $d(x', y') = k$, then since k is odd, x' and y' must have had different bits added . Consequently, $d(x, y) = k+1$. In other words, if two vertices x and y of the n-cube have the same color then $d(x, y) \geq k+1$, and so we had a valid coloring with parameters n and k. This completes the proof of the theorem.

3 Concluding Remarks

The key for us to get a good coloring is to find a good parity check matrix H when k is even. As we saw, the proof of Theorem 1 implicitly gave us an algorithm to construct H, but it is still not very constructive. However, in the case $k = 2$ (and thus in case $k = 3$) we can explicitly construct H. To do this, consider the Hamming code $H_2(r)$, which is a $[2^r - 1, 2^r - 1 - r, 3]$ code. Its parity check matrix $H(r, 2)$ has dimensions $r \times (2^r - 1)$. Let $r = \lceil \log_2(n+1) \rceil$, so that $2^r - 1 \geq n$. If we remove the last $2^r - 1 - n$ columns of $H(r, 2)$, then we get a parity check matrix of an $[n, n - \lceil \log_2(n+1) \rceil, 3]$ code. This code gives us a coloring of the n-cube with parameters n and 2 using $2^{\lceil \log_2(n+1) \rceil}$ colors. This proves the upper bound of (1).

Instead of the *Johnson bound* we used, we could have used other known upper bounds for $A(n, d)$ to give us lower bounds for $\chi_{\bar{k}}(n)$ (e.g., the *Plotkin bound*, the *Elias bound* or the *Linear Programming bound*). However, applying these other bounds doesn't seem to give us significantly better results.

References

[1] D.S. Kim, D.-Z. Du, and P.M. Pardalos, A coloring problem on the n-cube, *Discrete Applied Mathematics*, 103 (2000) 307-311.

[2] A. Pavan, P.-J. Wan, S.-R. Tong, and D.H. Du, A new multihop lightwave network based on the generalized de-Bruijn graph, *Proc. on IEEE INFOCOM*, 1996, pp.498-507.

[3] S. Roman, *Coding and Information Theory*, (1992, Springer-Verlag, New York) pp.253-263.

[4] P.-J. Wan, Near-optimal conflict-free channel set assignments for an optical cluster-based hypercube network, *Journal of Combinatorial Optimization* 1 (1997) 179-186.