CAMBRIDGE
UNIVERSITY PRESS

ARTICLE

On the number of error correcting codes

Dingding Dong[1], Nitya Mani[2] and Yufei Zhao[2]

[1]Department of Mathematics, Harvard University, Cambridge, MA 02138, USA and [2]Department of Mathematics,
Massachusetts Institute of Technology, Cambridge, MA 02139, USA
Corresponding author: Nitya Mani; Email: ddong@math.harvard.edu

Abstract

We show that for a fixed q, the number of q-ary t-error correcting codes of length n is at most $2^{(1+o(1))H_q(n,t)}$ for all $t \leq (1 - q^{-1})n - 2\sqrt{n \log n}$, where $H_q(n, t) = q^n/V_q(n, t)$ is the Hamming bound and $V_q(n, t)$ is the cardinality of the radius t Hamming ball. This proves a conjecture of Balogh, Treglown, and Wagner, who showed the result for $t = o(n^{1/3}(\log n)^{-2/3})$.

1. Introduction

A *q-ary t-error correcting code of length n* is a subset of $[q]^n$ whose elements pairwise differ in at least $2t + 1$ coordinates. Error correcting codes play a central role in coding theory, as they allow us to correct errors in at most t coordinates when sending codewords over a noisy channel. In the course of developing the container method, Sapozhenko in [10] posed the following natural question: how many error correcting codes (with given parameters) are there?

Given a t-error correcting code $\mathcal{C} \subseteq [q]^n$ of maximum size, any subset of $\mathcal{C}$ is also a t-error correcting code. Consequently, the number of q-ary t-error correcting codes of length n is at least $2^{|\mathcal{C}|}$. It is natural to suspect that the total number of t-error correcting codes is not much larger, provided that this maximum size code $\mathcal{C}$ is not too small.

A fundamental problem in coding theory is determining the maximum size of an error correcting code with given parameters (e.g., see [[12], Chapter 5] and [[7], Chapter 17] for classic treatments). There is an easy geometric upper bound on the size of a q-ary t-error correcting code. We use $d(x, y)$ to denote the *Hamming distance* between x and $y \in [q]^n$, defined as the number of coordinates on which x and y differ. Let $B_q(x, r) := \{y \in [q]^n : d(x, y) \leq r\}$ be the *Hamming ball* of radius r centred at $x \in [q]^n$ with associated cardinality

$$V_q(n, r) := \sum_{i=0}^{r} \binom{n}{i}(q - 1)^i.$$

Any t-error correcting code in $[q]^n$ has size at most

$$H_q(n, t) := \frac{q^n}{V_q(n, t)}.$$

This simple upper bound is known as the *Hamming bound* (also known as the *sphere packing bound*). Codes that attain equality for this upper bound are called *perfect codes*; the only non-trivial perfect codes are Hamming codes and Golay codes (see [11]). Except for some special parameters such as these, there is a large gap between the best known upper and lower bounds.

The Hamming bound can be significantly improved when $t \geq c\sqrt{n}$ (see [3, 8, 12]). Most studies on bounds for codes focus on the linear distance case, that is, $d = \lfloor \delta n \rfloor$ for some constant $0 < \delta < 1$. Curiously, even after consulting with several experts on the subject, we are not aware of any upper bound better than the Hamming bound when $t < c\sqrt{n}$. An important family of constructions is BCH codes, which can achieve sizes about $H_q(n, t)/t!$ for $q = 2, t = O(\sqrt{n})$ (see [[7], Corollary 9.8]); this is within a subexponential but superpolynomial factor of the Hamming bound. As we are primarily interested in the regime $t \leq c\sqrt{n}$ in this paper, we view the Hamming bound $H_q(n, t)$ as a proxy for the size of the code.

Balogh, Treglown and Wagner [2] showed that as long as $t = o(n^{1/3}(\log n)^{-2/3})$, the number of binary t-error correcting codes of length n is at most $2^{(1+o(1))H_2(n,t)}$ (their paper only considered the $q = 2$ case, though the situation for larger fixed q is similar). They conjectured that the same upper bound should also hold for a larger range of t. We prove a stronger version of this conjecture.

Theorem 1.1. *Fix integer $q \geq 2$. Below t is allowed to depend on n.*

(a) *For $0 < t \leq 10\sqrt{n}$, the number of q-ary t-error correcting codes of length n is at most $2^{(1+o(1))H_q(n,t)}$.*

(b) *For all $10\sqrt{n} < t \leq (1 - 1/q)n - 2\sqrt{n \log n}$, the number of q-ary t-error correcting codes of length n is $2^{o(H_q(n,t))}$.*

(c) *There exists some constant $c_q > 0$ such that for all $t > (1 - 1/q)n - c_q\sqrt{n \log n}$, one has $H_q(n, t) = O(n^{1/10})$. (Since there are q^n one-element codes, the number of codes is not $2^{O(H_q(n,t))}$ whenever $H_q(n, t) = O(n^{1/10})$.)*

The most interesting part is (a), where $t \leq 10\sqrt{n}$; here, our proof builds on the container argument of [2]. The key idea in [2] was analysing the container algorithm in two stages depending on the number of remaining vertices. Our work introduces two new ideas. We first give a pair of more refined supersaturation estimates; these estimates arise from understanding Hamming ball intersections and weak transitivity-type properties of a Hamming distance graph we study. Secondly, we employ a variant of the classical graph container algorithm with early stopping. See Section 2 for the container argument as well as the statements of the required supersaturation results, and Section 4 for proofs of the supersaturation estimates.

For the remaining cases when $t > 10\sqrt{n}$, we observe in Section 5 that the Elias bound (a classic upper bound on code sizes) implies the desired result when $t \leq (1 - q^{-1})n - 2\sqrt{n \log n}$ and that $H_q(n, t)$ is very small when t is any larger. Also, note that by the Plotkin bound [[12], (5.2.4)], if $2t + 1 - (1 - 1/q)n = \alpha > 0$, then the size of a t-error correcting code in $[q]^n$ is at most $(2t + 1)/\alpha$, which is very small, and so (b) is uninteresting when t is large.

A q-ary code of *length n* and *distance d* is a subset of $[q]^n$ whose elements are pairwise separated by Hamming distance at least d. The maximum size of such a code is denoted $A_q(n, d)$. It is natural

to conjecture the following much stronger bound on the number of codes of length n and distance d, which is likely quite difficult.

Conjecture 1.2. *For all fixed $c > 0$, the number of q-ary codes of length n and distance d is $2^{O(A_q(n,d))}$ whenever $d < (1 - q^{-1} - c)n$.*

The main difficulty seems to be a lack of strong supersaturation bounds, for example, showing that sets with size much larger than $A_q(n, d)$ must have many pairs of elements with distance at most d.

The above conjecture is analogous to the following conjecture of Kleitman and Winston [6]. We use ex (n, H) to denote the maximum number of edges in an n-vertex H-free graph.

Conjecture 1.3. *For every bipartite graph H that contains a cycle, the number of H-free graphs on n labelled vertices is $2^{O_H(\text{ex}(n,H))}$.*

Kleitman and Winston [6] proved the conjecture for $H = C_4$ and in doing so developed what is now called the graph container method, which is central to the rest of this paper. See Ferber, McKinley, and Samotij [4] for recent results and discussion on the problem of enumerating H-free graphs.

Let $r_k(n)$ denote the maximum size of a k-AP-free subset of $[n]$. Balogh, Liu, and Sharifzadeh [1] recently proved the following theorem.

Theorem 1.4 (Balogh, Liu, and Sharifzadeh [1]). *Fix $k \geq 3$. The number of k-AP-free subsets of $[n]$ is $2^{O(r_k(n))}$ for infinitely many n.*

A notable feature of the above theorem is that the asymptotic order of $r_k(n)$ is not known for any $k \geq 3$, similar to the situation for $A(n, d)$. Ferber, McKinley, and Samotij [4] proved analogous results for the number of H-free graphs. It remains an open problem to extend the above result to all n. Also, Cameron and Erdős asked whether the number of k-AP-free subsets of $[n]$ is $2^{(1+o(1))r_k(n)}$. This is unknown, although for several similar questions the answer is no (see [1] for discussion). Likewise, one can ask whether the number of q-ary error correcting codes of length n and distance d is $2^{(1+o(1))A_q(n,d)}$.

Remarks about notation. For non-negative quantities depending on n, we write $f \lesssim g$ to mean that there is some constant C such that $f \leq Cg$ for all sufficiently large n. Throughout the paper, we view q as a constant, and so hidden constants are allowed to depend on q.

Given a finite set X, we write 2^X for the collection of all subsets of X, $\binom{X}{t}$ for the collection of t-element subsets of X, and $\binom{X}{\leq t}$ for the collection of subsets of X with at most t elements. We also write $\binom{n}{\leq t} = \sum_{0 \leq i \leq t} \binom{n}{i}$.

2. Graph container argument

In this section, we analyse the case $t \leq 10\sqrt{n}$. As in [2], we use the method of graph containers. This technique was originally introduced by Kleitman and Winston [5, 6]. See Samotij [9] for a modern introduction to the graph container method. We will not need the more recent hypergraph container method for this work. Instead, as in [2], we refine the graph container algorithm and analysis.

Let $G_{q,n,t}$ be the graph with vertex set $[q]^n$ where two vertices are adjacent if their Hamming distance is at most $2t$. For a subset $S \subseteq [q]^n$, we denote by $G_{q,n,t}[S]$ the induced subgraph of $G_{q,n,t}$ on S. For simplicity of notation, we write $G[S] = G_{q,n,t}[S]$, as the tuple (q, n, t) does not change in the proof. Given a graph G, we write $\Delta(G)$ for the maximum degree of G and $i(G)$ for the number of independent sets in G.

Applications of the container method usually require supersaturation estimates. Our main advance, building on [2], is a more refined supersaturation bound for codes. The methods used to prove the first lemma below already allow us to prove the main theorem for $t = o(\sqrt{n}/\log n)$; in concert with the second lemma, we are able to establish the full theorem.

Lemma 2.1 (Supersaturation I). *Let $t \leq 10\sqrt{n}$. If $S \subseteq [q]^n$ has size $|S| \geq n^4 H_q(n, t)$, then*

$$\Delta(G[S]) \gtrsim \frac{n^{3/2}}{H_q(n, t)}|S|.$$

Lemma 2.2 (Supersaturation II). *Let $\varepsilon > 0$. For all sufficiently large n, if $60 \leq t \leq 10\sqrt{n}$ and $S \subseteq [q]^n$ satisfies $\Delta(G[S]) \leq n^5$, then $i(G[S]) \leq 2^{(1+\varepsilon)H_q(n,t)}$.*

We defer the proof of the above two lemmas to subsection 4.1. Assuming these two lemmas, we now run the container argument to deduce the main theorem. The container algorithm we use is a modification of the classical graph container algorithm of Kleitman and Winston [5, 6]. As in [2], we divide the container algorithm into several stages and analyse them separately. Moreover, we allow 'early stopping' of the algorithm if we can certify that the residual graph has few independent sets.

Lemma 2.3 (Container). *For every $\varepsilon > 0$, the following holds for n sufficiently large. For all $60 \leq t \leq 10\sqrt{n}$, there exists a collection $\mathcal{F}$ of subsets of $[q]^n$ with the following properties:*

- *$|\mathcal{F}| \leq 2^{\varepsilon H_q(n,t)}$;*
- *Every t-error correcting code in $[q]^n$ is contained in S for some $S \in \mathcal{F}$;*
- *$i(G[S]) \leq 2^{(1+\varepsilon)H_q(n,t)}$ for every $S \in \mathcal{F}$.*

This container lemma implies Theorem 1.1(a).

Proof of Theorem 1.1(a). If $t < 60$, then the result is already known ([2] proved it for $t = o(n^{1/3}(\log n)^{-2/3})$). So assume $60 \leq t \leq 10\sqrt{n}$. Let $\varepsilon > 0$. Thus, letting $\mathcal{F}$ be as in Lemma 2.3, the number of t-error correcting codes in $[q]^n$ is at most

$$\sum_{S \in \mathcal{F}} i(G[S]) \leq |\mathcal{F}| \cdot 2^{(1+\varepsilon)H_q(n,t)} = 2^{(1+2\varepsilon)H_q(n,t)},$$

for sufficiently large n. Since $\varepsilon > 0$ can be arbitrarily small, we have the desired result. □

Proof of Lemma 2.3. Fix $\varepsilon > 0$. Let $C > 0$ be a sufficiently large constant and assume that n is sufficiently large throughout. We show that there exists some function

$$f : \binom{V(G)}{\leq \frac{CH_q(n,t)\log n}{n}} \to 2^{V(G)},$$

such that for every independent set $I \in \mathcal{I}(G)$, there is some 'fingerprint' $P \subseteq I$ that satisfies the following three conditions:

1. $|P| \leq \frac{CH_q(n,t)\log n}{n}$,
2. $i(G[P \cup f(P)]) \leq 2^{(1+\varepsilon)H_q(n,t)}$,
3. $I \subseteq P \cup f(P)$.

We consider the following algorithm that constructs P and $f(P)$ given some $I \in \mathcal{I}(G)$.

Algorithm 1 Fix an arbitrary order $v_1, \ldots, v_{q^n}$ of the elements of $V(G)$. Let $I \in \mathcal{I}(G)$ be an independent set, as the input.

(1) Initialize $G_0 := G$ and $P := \emptyset$.

(2) For $i = 1, 2, \ldots$

   - If $i(G_{i-1}) \leq 2^{(1+\varepsilon)H_q(n,t)}$, then set $f(P) = V(G_{i-1})$ and terminate.
   - Else, let u be a maximum degree vertex in G_{i-1} (break ties by taking the earliest vertex in the given ordering).
   - If $u \notin I$, define G_i by taking G_{i-1} and removing u, and continue onto the next step.
   - If $u \in I$, then add u to P, define G_i by taking G_{i-1} and removing u and its neighbours, and continue.

(3) Output P and $f(P)$.

Given any independent set I, the algorithm outputs some P and $f(P)$. Note that $f(P)$ depends only on P (i.e., two different I's that output the same P will always output the same $f(P)$). Also, for any independent set I, the output satisfies $P \subseteq I \subseteq P \cup f(P)$.

To prove $|P| \leq \frac{CH_q(n,t) \log n}{n}$, we distinguish stages in the algorithm based on the size of $V(G_i)$.

- Let P_1 denote the vertices added to P when $|V(G_{i-1})| \geq n^4 H_q(n, t)$.

- Let P_2 denote the vertices added to P when $|V(G_{i-1})| < n^4 H_q(n, t)$.

While we are adding vertices to P_1, according to Lemma 2.1, we are removing at least a $\beta \gtrsim n^{3/2}/H_q(n, t)$ fraction of vertices with every successful addition. We have

$$|P_1| \leq \frac{\log\left(\frac{q^n}{n^4 H_q(n,t)}\right)}{\log\left(\frac{1}{1-\beta}\right)} \lesssim \frac{\log V_q(n,t) - 4\log n}{\beta} \leq \frac{\log\left((nq)^t\right)}{\beta} \lesssim \frac{t \log n}{n^{3/2}} H_q(n,t) \lesssim \frac{\log n}{n} H_q(n,t).$$

While we are adding vertices to P_2, by Lemma 2.2, we are removing at least n^5 vertices with every successful addition, as $i(G_{i-1}) \geq 2^{(1+\varepsilon)H_q(n,t)}$ during the associated iteration. Consequently,

$$|P_2| \lesssim \frac{n^4 H_q(n,t)}{n^5} = \frac{H_q(n,t)}{n}.$$

These two inequalities imply that

$$|P| = |P_1| + |P_2| \leq \frac{C \log n}{n} H_q(n,t).$$

Thus,

$$i(G[P \cup f(P)]) \leq 2^{|P|} \cdot i(f(P)) \leq 2^{\varepsilon H_q(n,t)} \cdot 2^{(1+\varepsilon)H_q(n,t)} \leq 2^{(1+2\varepsilon)H_q(n,t)}.$$

Take $\mathcal{F}$ to be the collection of all $P \cup f(P)$ obtained from this procedure as I ranges over all independent sets of G. We have just proved that every t-error correcting code is contained in some $S \in \mathcal{F}$. We also have $i(G[S]) \leq 2^{(1+2\varepsilon)H_q(n,t)}$ for every $S \in \mathcal{F}$ (equivalent to the stated result after replacing ε by $\varepsilon/2$). Finally, using that $\log \binom{n}{m} \leq m \log (O(n)/m)$, we have

$$\log |\mathcal{F}| \leq \log \binom{q^n}{\leq \frac{C \log n}{n} H_q(n,t)}$$

$$\lesssim \frac{C \log n}{n} H_q(n,t) \log \left(O\left(\frac{n V_q(n,t)}{C \log n}\right)\right)$$

$$\leq \frac{C \log n}{n} \log \left(O \left(\frac{n(qn)^t}{C \log n} \right) \right) H_q(n, t)$$

$$\lesssim \frac{Ct(\log n)^2}{n} H_q(n, t).$$

Since $t \leq 10\sqrt{n}$, $|\mathcal{F}| \leq 2^{\varepsilon H_q(n,t)}$ for sufficiently large n. □

It remains to prove the supersaturation estimates Lemmas 2.1 and 2.2, which we will do in Section 4 after first observing some technical preliminaries in Section 3. The cases when $t > 10\sqrt{n}$ are in Section 5.

3. Hamming ball volume estimates

3.1 Hamming ball volume ratios

We first record some estimates about sizes of Hamming balls of different radii, beginning with the following decay estimate.

Lemma 3.1. *For every* $1 \leq i \leq t$, *we have*

$$V_q(n - i, t - i) \leq \left(\frac{t}{(q-1)n} \right)^i V_q(n, t).$$

Proof. Since $\binom{n-1}{j} = \frac{j+1}{n} \binom{n}{j+1}$, we have

$$\frac{V_q(n-1, t-1)}{V_q(n, t)} = \frac{\sum_{j=0}^{t-1} \binom{n-1}{j}(q-1)^j}{1 + \sum_{j=0}^{t-1} \binom{n}{j+1}(q-1)^{j+1}} \leq \frac{\sum_{j=0}^{t-1} \binom{n-1}{j}(q-1)^j}{\sum_{j=0}^{t-1} \binom{n}{j+1}(q-1)^{j+1}} \leq \frac{t}{(q-1)n}.$$

Consequently,

$$\frac{V_q(n-i, t-i)}{V_q(n, t)} = \frac{V_q(n-1, t-1)}{V_q(n, t)} \cdots \cdots \frac{V_q(n-i, t-i)}{V_q(n-i+1, t-i+1)}$$

$$\leq \frac{1}{(q-1)^i} \cdot \frac{t}{n} \cdot \frac{t-1}{n-1} \cdots \cdots \frac{t-i+1}{n-i+1} \leq \left(\frac{t}{(q-1)n} \right)^i.$$ □

Lemma 3.2. *For* $1 \leq \alpha \leq t \leq n - \alpha$, *we have*

$$V_q(n, t + \alpha) \geq \left(\frac{(q-1)n}{t+\alpha} \right)^\alpha \left(\frac{n-\alpha+1-t}{n-\alpha+1} \right)^\alpha V_q(n, t).$$

Proof. Observe that

$$\frac{V_q(n-\alpha, t)}{V_q(n, t)} = \frac{\sum_{i=0}^t (q-1)^i \binom{n-\alpha}{i}}{\sum_{i=0}^t (q-1)^i \binom{n}{i}} \geq \left(\frac{n-\alpha+1-t}{n-\alpha+1} \right)^\alpha.$$

Also, by Lemma 3.1, we have

$$V_q(n-\alpha, t) \leq \left(\frac{t+\alpha}{(q-1)n} \right)^\alpha V_q(n, t+\alpha).$$

Combining this pair of inequalities gives

$$V_q(n, t+\alpha) \geq \left(\frac{(q-1)n}{t+\alpha} \right)^\alpha V_q(n-\alpha, t) \geq \left(\frac{(q-1)n}{t+\alpha} \right)^\alpha \left(\frac{n-\alpha+1-t}{n-\alpha+1} \right)^\alpha V_q(n, t).$$ □

3.2 Intersection of Hamming balls

We record some estimates about the sizes of intersections of Hamming balls as a function of the distance between their centres.

Let $W_q(n, t, k)$ be the size of the intersection of two Hamming balls in $[q]^n$ of radius t, the centres being distance k apart. It is easy to check that

$$W_q(n, t, k) = \sum_{r=0}^{k} \sum_{s=0}^{k-r} \binom{k}{r} \binom{k-r}{s} (q-2)^{k-r-s} V_q(n-k, t - \max\{k-r, k-s\}).$$

Indeed, suppose the two balls are centred at (1^n) and $(2^k 1^{n-k})$. The intersection then consists of all points with r 1's and s 2's among the first k coordinates (as r and s range over all possible values), and $\leq t - \max\{k-r, k-s\}$ non-1 coordinates among the remaining $n-k$ coordinates.

In particular, it is easy to see that

$$W_q(n, t, 1) = q V_q(n-1, t-1).$$

(As above, it does not matter what happens in the first coordinate and the remaining $n-1$ coordinates can have at most $t-1$ non-1 coordinates).

Lemma 3.3. $W_q(n, t, k+1) \leq W_q(n, t, k)$ *for every integer* $0 \leq k < n$.

Proof. It suffices to consider radius t Hamming balls centred at $u = 1^n$, $v = 2^k 1^{n-k}$, and $w = 2^{k+1} 1^{n-k-1}$. We show that $|B_q(u, t) \cap B_q(w, t)| \leq |B_q(u, t) \cap B_q(v, t)|$. Let us consider the following subsets of $[q]^n$:

$$X = B_q(u, t) \cap B_q(w, t) \setminus B_q(v, t) \quad \text{and} \quad Y = B_q(u, t) \cap B_q(v, t) \setminus B_q(w, t).$$

Observe that every $x = (x_1, \ldots, x_n) \in X$ must have $x_{k+1} = 2$ and $d(x, w) = t$, $d(x, v) = t+1$. Define $\phi : X \to Y$ where $\phi(x)$ is obtained from x by changing the $(k+1)$st coordinate from 2 to 1. This map is well-defined since for every $x \in X$, $d(\phi(x), u) < d(x, u) \leq t$, $d(\phi(x), w) = t+1$ and $d(\phi(x), v) = t$, and so $\phi(x) \in Y$. Since ϕ is injective, we have $|X| \leq |Y|$ and therefore $|B_q(u, t) \cap B_q(w, t)| \leq |B_q(u, t) \cap B_q(v, t)|$. $\square$

Lemma 3.4. *For every integer* $k \geq 0$, *we have*

$$W_q(n, t, 2k+2) \leq W_q(n, t, 2k+1) \leq 2 \left(\frac{q^2 t}{(q-1)n} \right)^k W_q(n, t, 1) \leq 2 \left(\frac{q^2 t}{(q-1)n} \right)^{k+1} V_q(n, t).$$

Proof. The first inequality is simply Lemma 3.3. We further have that

$W_q(n, t, 2k+1)$

$$= \sum_{r=0}^{2k+1} \sum_{s=0}^{2k+1-r} \binom{2k+1}{r} \binom{2k+1-r}{s} (q-2)^{2k+1-r-s} V_q(n-2k-1, t-2k-1+\min\{r, s\})$$

$$\leq 2 \sum_{r=0}^{k} \sum_{s=0}^{2k+1-r} \binom{2k+1}{r} \binom{2k+1-r}{s} (q-2)^{2k+1-r-s} V_q(n-2k-1, t+r-2k-1)$$

$$\leq 2 \sum_{r=0}^{k} \sum_{s=0}^{2k+1-r} \binom{2k+1}{r} \binom{2k+1-r}{s} (q-2)^{2k+1-r-s} V_q(n+r-2k-1, t+r-2k-1).$$

Using Lemma 3.1 and substituting the above inequality, we deduce that

$$\frac{W_q(n, t, 2k+1)}{W_q(n, t, 1)} = \frac{W_q(n, t, 2k+1)}{qV_q(n-1, t-1)}$$

$$\overset{(*)}{\leq} \frac{2}{q} \sum_{r=0}^{k} \sum_{s=0}^{2k+1-r} \binom{2k+1}{r} \binom{2k+1-r}{s} (q-2)^{2k+1-r-s} \left(\frac{t-1}{(q-1)(n-1)}\right)^{2k-r}$$

$$\leq \left(\frac{t-1}{(q-1)(n-1)}\right)^{k} \cdot \frac{2}{q} \sum_{r=0}^{k} \sum_{s=0}^{2k+1-r} \binom{2k+1}{r} \binom{2k+1-r}{s} (q-2)^{2k+1-r-s}$$

$$= \left(\frac{t-1}{(q-1)(n-1)}\right)^{k} \cdot \frac{2}{q} \sum_{r=0}^{k} \binom{2k+1}{r} ((q-2)+1)^{2k+1-r}$$

$$\leq \left(\frac{t-1}{(q-1)(n-1)}\right)^{k} \cdot \frac{2}{q} \sum_{r=0}^{2k+1} \binom{2k+1}{r} (q-1)^{2k+1-r}$$

$$= \left(\frac{t-1}{(q-1)(n-1)}\right)^{k} \cdot \frac{2}{q} \cdot q^{2k+1}$$

$$\leq \left(\frac{t-1}{(q-1)(n-1)}\right)^{k} \cdot 2q^{2k} \leq 2 \left(\frac{q^2 t}{(q-1)n}\right)^{k},$$

where $(*)$ follows from the previous expansion of $W_q(n, t, 2k+1)$ combined with Lemma 3.1. This proves the second inequality in the lemma. To prove the last inequality in the lemma, apply Lemma 3.1 again to yield

$$\frac{W_q(n, t, 1)}{V_q(n, t)} = \frac{qV_q(n-1, t-1)}{V_q(n, t)} \leq \frac{qt}{(q-1)n}. \qquad \square$$

4. Supersaturation

4.1 Supersaturation I

From now on, let $S \subseteq [q]^n$ and let $G[S] = (S, E)$ be the associated graph with the edge set

$$E = \left\{\{x, y\} \in \binom{S}{2} : d(x, y) \leq 2t\right\}.$$

For each $1 \leq k \leq 2t$, write

$$E_k = \left\{\{x, y\} \in \binom{S}{2} : d(x, y) = k\right\}.$$

In other words, E is the set of pairs in S with Hamming distance at most $2t$, and each E_k is the set of pairs in S with Hamming distance exactly k. Also, for each $v \in S$, denote number of edges in E_k incident to v by

$$\deg_k(v) = \left|\left\{u \in S : d(u, v)\right\} = k\right|.$$

We have the trivial bound $\deg_k(v) \leq \binom{n}{k}(q-1)^k$ for every $v \in S$, and hence, $|E_k| \leq \frac{1}{2}\binom{n}{k}(q-1)^k |S|$.

We first recall the following supersaturation estimate from [2] (stated there for binary codes). We include the proof below for completeness. Later on, we will derive new and stronger supersaturation estimates.

Lemma 4.1 ([2], Lemma 5.3). *If* $|S| \geq 2H_q(n, t)$, *then*

$$\sum_{k=1}^{2t} W_q(n, t, k)|E_k| \geq \frac{|S|^2 V_q(n, t)^2}{10 \cdot q^n} \quad and \quad |E| \geq \frac{n|S|^2}{20tH_q(n, t)}.$$

Proof. Define $K_x = \{a \in S \mid d(x, a) \leq t\}$ for $x \in [q]^n$. Observe that

$$\sum_{k=1}^{2t} W_q(n, t, k)|E_k| = \sum_{x \in [q]^n} \binom{|K_x|}{2},$$

since both terms count pairs $(x, \{a, b\})$ for $x \in [q]^n$ and distinct $a, b \in S$ such that $d(x, a)$, $d(x, b) \leq t$. From Lemmas 3.1 to 3.3, we know that

$$W_q(n, t, k) \leq W_q(n, t, 1) = qV_q(n-1, t-1) \leq \frac{qt}{(q-1)n}V_q(n, t)$$

for $1 \leq k \leq 2t$. By convexity, since the average value of K_x over $x \in [q]^n$ is $|S|V_q(n, t)/q^n \geq 2$, we have

$$\frac{qt}{(q-1)n} \cdot V_q(n, t)|E| \geq W_q(n, t, 1)|E| \geq \sum_{k=1}^{2t} W_q(n, t, k)|E_k| = \sum_{x \in [q]^n} \binom{|K_x|}{2} \geq \frac{|S|^2 V_q(n, t)^2}{10q^n}.$$

This proves the first inequality stated in the lemma. Rearranging, we derive the second inequality:

$$|E| \geq \frac{(q-1)n|S|^2 V_q(n, t)}{10qt \cdot q^n} \geq \frac{n|S|^2}{20tH_q(n, t)}. \qquad \square$$

Now let us prove the first supersaturation estimate, Lemma 2.1, which says that if $t \leq 10\sqrt{n}$ and $|S| \geq n^4 H_q(n, t)$, then $\Delta(G[S]) \gtrsim \frac{n^{3/2}}{H_q(n,t)}|S|$.

Proof of Lemma 2.1. By Lemma 4.1, since $|S| \geq n^4 H_q(n, t)$,

$$|E| \geq \frac{|S|n}{20tH_q(n, t)} \cdot |S| \geq \frac{n^5}{20t} \cdot |S|.$$

Together with $|E_1| + |E_2| \leq n^2 q^2 |S|$ and $|E_3| + |E_4| \leq n^4 q^4 |S|$, we have (recall that hidden constants are allowed to depend on q)

$$|E_1| + |E_2| \lesssim \frac{t}{n^3}|E| \quad and \quad |E_3| + |E_4| \lesssim \frac{t}{n}|E|.$$

Furthermore, by Lemma 3.4,

$$W_q(n, t, 2) \leq W_q(n, t, 1),$$

$$W_q(n, t, 4) \leq W_q(n, t, 3) \lesssim \frac{t}{n}W_q(n, t, 1), \quad and$$

$$W_q(n, t, k) \leq W_q(n, t, 5) \lesssim \frac{t^2}{n^2}W_q(n, t, 1) \quad for\ all\ k \geq 5.$$

Hence we have

$$\sum_{k=1}^{2t} W_q(n,t,k)|E_k| \le W_q(n,t,1)(|E_1| + |E_2|) + W_q(n,t,3)(|E_3| + |E_4|) + W_q(n,t,5)|E|$$

$$\lesssim \left(\frac{t}{n^3} + \frac{t}{n}\cdot\frac{t}{n} + \frac{t^2}{n^2}\right) W_q(n,t,1)|E| \lesssim \frac{t^2}{n^2}\cdot W_q(n,t,1)|E|.$$

By Lemma 4.1, we have

$$\frac{|S|^2 V_q(n,t)^2}{q^n} \lesssim \sum_{k=1}^{2t} W_q(n,t,k)|E_k| \lesssim \frac{t^2}{n^2}\cdot W_q(n,t,1)|E|.$$

Rearranging and then applying $W_q(n,t,1) \lesssim (t/n)V_q(n,t)$ from Lemma 3.4, and $t \lesssim \sqrt{n}$, we obtain

$$|E| \gtrsim \frac{n^2}{t^2}\cdot\frac{V_q(n,t)}{W_q(n,t,1)}\cdot\frac{V_q(n,t)}{q^n}\cdot|S|^2 \gtrsim \frac{n^3}{t^3}\cdot\frac{1}{H_q(n,t)}\cdot|S|^2 \gtrsim \frac{n^{3/2}}{H_q(n,t)}\cdot|S|^2.$$

Thus, the average degree (and hence the maximum degree) in $G[S]$ is $\gtrsim \frac{n^{3/2}}{H_q(n,t)}|S|$. $\square$

4.2 Supersaturation II

We maintain the notation from the previous subsection.

Lemma 4.2. *Suppose $t \le 10\sqrt{n}$ and $\Delta(G[S]) \le n^5$. Fix $\varepsilon > 0$. Let*

$$S_1 = \left\{v \in S \colon \deg_k(v) \le \varepsilon n^{\lceil k/2 \rceil/2} \text{ for each } k = 1, \dots, 20\right\}.$$

Then $|S_1| \le (1 + O(\varepsilon))H_q(n,t)$.

Proof. The idea is that the Hamming balls $B_q(v,t)$, $v \in S_1$, are 'mostly disjoint', in the sense that the overlap is negligible.

For any $v \in S_1$, the overlap of $B_q(v,t)$ with other balls $B_q(u,t)$, $u \in S_1$, has size

$$\left| B_q(v,t) \cap \bigcup_{u \in S_1 \setminus \{v\}} B_q(u,t) \right| \le \sum_{k=1}^{2t} \deg_k(v) W_q(n,t,k)$$

$$\le \left(\sum_{k=1}^{20} \deg_k(v) W_q(n,t,k)\right) + n^5 W_q(n,t,21)$$

by Lemma 3.3. Recall from Lemma 3.4 that $W_q(n,t,k) \le 2(q^2 t/((q-1)n))^{\lceil k/2 \rceil} V_q(n,t)$, which is $\lesssim n^{-\lceil k/2 \rceil/2} V_q(n,t)$ for each $1 \le k \le 20$. Combining with $\deg_k(v) \le \varepsilon n^{\lceil k/2 \rceil/2}$ for each $v \in S$, the above sum is $\lesssim \varepsilon V_q(n,t)$. In other words, for every $v \in S_1$, the ball $B_q(v,t)$ contains $(1 - O(\varepsilon))V_q(n,t)$ unique points not shared by other such balls, and thus, the union of these balls has size $\ge (1 - O(\varepsilon))V_q(n,t)|S_1|$. Since the union is contained in $[q]^n$, we deduce $|S_1| \le (1 + O(\varepsilon))q^n/V_q(n,t) = (1 + O(\varepsilon))H_q(n,t)$. $\square$

Lemma 4.3. *Suppose $60 \le t \le 10\sqrt{n}$ and $\Delta(G[S]) \le n^5$. Fix $\varepsilon > 0$. Let*

$$S_2 = \left\{v \in S \colon \sum_{k=1}^{20} \deg_k(v) \ge \frac{\log n}{\varepsilon}\right\}.$$

Then, the number of independent subsets of S_2 satisfies

$$i(G[S_2]) \leq 2^{O(\varepsilon H_q(n,t))}.$$

Proof. Pick a maximum size subset $X \subseteq S_2$ where every pair of distinct elements of X have Hamming distance greater than t. Using the definition of S_2, for each $x \in X$, we can find a $\lceil (\log n)/\varepsilon \rceil$-element subset $A_x \subseteq S \cap B_q(x, 20)$. For distinct $x, y \in X$, one has $d(x, y) > t \geq 60$, and thus $A_x \cap A_y = \emptyset$.

Consider the balls $B_q(u, t)$, $u \in \bigcup_{x \in X} A_x$. As in the previous proof, we will show that these balls are mostly disjoint. The intersection of one of these balls with the union of all other such balls has size at most

$$\frac{\log n}{\varepsilon} W_q(n, t, 1) + n^5 W_q(n, t, 21) \lesssim \left(\frac{\log n}{\varepsilon} \cdot \frac{t}{n} + n^5 \cdot \frac{t^{11}}{n^{11}} \right) V_q(n, t) = o(V_q(n, t)).$$

Indeed, for each of the $\lceil (\log n)/\varepsilon \rceil - 1$ points u' that lie in the same A_x as u, the overlap is $\leq W_q(n, t, 1)$; each ball $B_q(u', t)$ with u' in some A_y other than A_x contributes to $\leq W_q(n, t, 21)$ overlap since $d(u, u') \geq d(x, y) - 40 \geq t - 39 \geq 21$. There are at most $\Delta(G[S]) \leq n^5$ other balls that intersect a given ball. Here we invoke the monotonicity of $W_q(n, t, \cdot)$ (Lemma 3.3) and also Lemma 3.4.

Thus, the union of the balls $B_q(u, t)$, $u \in \bigcup_{x \in X} A_x$, has size $|X| \lceil (\log n)/\varepsilon \rceil \cdot (1 - o(1)) V_q(n, t)$. Since this quantity is at most q^n,

$$|X| \leq \frac{q^n}{\lceil (\log n)/\varepsilon \rceil \cdot (1 - o(1)) V_q(n, t)} \lesssim \frac{\varepsilon H_q(n, t)}{\log n}.$$

Note that $S_2 \subseteq \bigcup_{x \in X} B_q(x, t)$ (since if there were some uncovered $x' \in [q]^n$, then one can add x' to X). Every pair of elements of $B_q(x, t)$ has Hamming distance at most $2t$. It follows that every independent set I in $G[S_2]$ (so the elements of I are separated by Hamming distance $> 2t$) can be formed by choosing at most one element from $B_q(x, t) \cap S_2$ for each $x \in X$. Since $|B_q(x, t) \cap S_2| \leq n^5$, we have

$$i(G[S_2]) \leq (n^5 + 1)^{|X|} \leq 2^{O(\varepsilon H_q(n,t))}. \qquad \square$$

Now we prove the second supersaturation estimate, Lemma 2.2, which says that if $60 \leq t \leq 10\sqrt{n}$ and $\Delta(G[S]) \leq n^5$, then $i(G[S]) \leq 2^{(1+\varepsilon)H_q(n,t)}$ for all sufficiently large n.

Proof of Lemma 2.2. Define S_1 and S_2 as in the previous two lemmas. One has $S = S_1 \cup S_2$ provided that n is sufficiently large. Thus

$$i(G[S]) \leq i(G[S_1]) i(G[S_2]) \leq 2^{|S_1|} i(G[S_2]) \leq 2^{(1+O(\varepsilon))H_q(n,t)}.$$

This is equivalent to the desired result after changing ε by a constant factor. $\qquad \square$

5. Bounds on codes with larger distances

When $t > 10\sqrt{n}$, we observe that a maximum sized t error correcting code must be much smaller than the Hamming bound; this implies an associated bound on the number of t-error correcting codes. Recall that $A_q(n, 2t + 1)$ is the maximum size of a t-error correcting code over $[q]^n$. The number of q-ary t-error correcting codes of length n is at most

$$\binom{q^n}{\leq A_q(n, 2t + 1)} \leq q^{nA_q(n, 2t+1)}. \tag{5.1}$$

If $A_q(n, 2t + 1) = o(H_q(n, t)/n)$, then the number of t-error correcting codes of length n is $2^{o(H_q(n,t))}$.

In fact, in the regime of very large t, we will be able to obtain Theorem 1.1(c), that is, that there exists $c_q > 0$ such that for all $t > (1 - 1/q)n - c_q\sqrt{n \log n}$, we have $H_q(n, t) = O(n^{1/10})$.

Proof of Theorem 1.1(c). Let $t = (1 - 1/q)n - \alpha$ with $\alpha < c_q\sqrt{n \log n}$ for some sufficiently small $c_q > 0$. Let $\theta = 1 - q^{-1}$. Let $Y \sim \text{Bin}(n, \theta)$ be a binomial random variable and $Z \sim \mathcal{N}(0, 1)$ a standard normal random variable. Observe that for all $0 \leq i \leq n$, we have $V_q(n, t) = q^n \mathbb{P}(Y \leq t)$. By the Berry–Esseen theorem (a quantitative central limit theorem), for all $x \in \mathbb{R}$,

$$\left| \mathbb{P}\left[\frac{Y - \theta n}{\sqrt{\theta(1 - \theta)n}} \leq x \right] - \mathbb{P}[Z \leq x] \right| \lesssim \frac{1}{\sqrt{n}}.$$

We have the following standard estimate on the Gaussian tail:

$$\mathbf{P}[Z \leq -x] > \frac{1}{\sqrt{2\pi}} \frac{x}{x^2 + 1} e^{-x^2/2} \qquad \text{for all } x > 0.$$

Thus

$$\frac{1}{H_q(n, t)} = \frac{V_q(n, \theta n - \alpha)}{q^n} \geq \mathbf{P}\left[Z \leq \frac{-c_q\sqrt{n \log n}}{\sqrt{\theta(1 - \theta)n}} \right] - O\left(\frac{1}{\sqrt{n}} \right)$$

$$\gtrsim e^{-\Theta(c_q^2 \log n)} - O\left(\frac{1}{\sqrt{n}} \right)$$

$$\gtrsim n^{-1/10}$$

by choosing c_q to be sufficiently small. Therefore $H_q(n, t) = O(n^{1/10})$. $\qquad \square$

Remark 5.1. *It is possible to give more precise estimates of binomial tails with larger deviations. The above application of the Barry–Esseen theorem is a concise way to obtain what we need.*

While we will not get a bound as strong as above for smaller values of t, already when $t > 10\sqrt{n}$, we have that $A_q(n, 2t + 1) = o(H_q(n, t)/n)$, due to the following classic upper bound on $A_q(n, d)$.

Theorem 5.2 (Elias bound [[12], Theorem 5.2.11]). *Let $\theta = 1 - q^{-1}$. For every $r \leq \theta n$ satisfying*

$$r^2 - 2\theta nr + \theta nd > 0,$$

we have the upper bound

$$A_q(n, d) \leq \frac{\theta nd}{r^2 - 2\theta nr + \theta nd} \cdot \frac{q^n}{V_q(n, r)}.$$

We have the following consequence of the above Elias bound.

Proposition 5.3. *Let $\theta = 1 - q^{-1}$. If $10\sqrt{n} < t \leq \theta n - 2\sqrt{n \log n}$, then we have*

$$A_q(n, 2t + 1) = o\left(\frac{H_q(n, t)}{n} \right).$$

Proof. We apply the Elias bound with $r = t + \alpha$, where $\alpha = 7$ if $10\sqrt{n} < t < n^{4/5}$ and $\alpha = \sqrt{n \log n}$ if $n^{4/5} \leq t \leq \theta n - 2\sqrt{n \log n}$. When $t > 10\sqrt{n}$, we satisfy the condition

$$r^2 - 2\theta nr + \theta n(2t + 1) = r^2 - \theta(2\alpha - 1)n \gtrsim r^2. \qquad (5.2)$$

Indeed, when $10\sqrt{n} < t < n^{4/5}$, we have $\alpha = 7$, and thus $\theta(2\alpha - 1)n \leq 2\alpha n = 14n$ whereas $r^2 > (10\sqrt{n})^2 = 100n$. When $t \geq n^{4/5}$, we have $r^2 \geq n^{8/5}$ whereas $\theta(2\alpha - 1)n = O(n^{3/2}\sqrt{\log n})$. By Lemma 3.2, we have

$$V_q(n, r) = V_q(n, t + \alpha) \geq \left(\frac{(q - 1)n}{t + \alpha} \right)^\alpha \left(\frac{n - \alpha + 1 - t}{n - \alpha + 1} \right)^\alpha V_q(n, t).$$

Applying Theorem 5.2 gives that

$$A_q(n, 2t+1) \le \frac{\theta n(2t+1)}{r^2 - 2\theta nr + \theta n(2t+1)} \cdot \frac{q^n}{V_q(n,r)}$$

$$\overset{(5.2)}{\lesssim} \frac{nt}{(t+\alpha)^2} \cdot \frac{q^n}{V_q(n, t+\alpha)}.$$

$$\le \frac{n}{t}\left(\frac{(t+\alpha)(n-\alpha+1)}{(q-1)n(n-\alpha+1-t)}\right)^\alpha H_q(n,t).$$

When $10\sqrt{n} < t < n^{4/5}$ and $\alpha = 7$, the above upper bound simplifies to

$$A_q(n, 2t+1) \lesssim \frac{n}{t}\left(\frac{(t+\alpha)(n-\alpha+1)}{(q-1)n(n-\alpha+1-t)}\right)^\alpha H_q(n,t) \lesssim \frac{t^6}{n^6}H_q(n,t) \le \frac{1}{n^{6/5}}H_q(n,t)$$

$$= o\left(\frac{H_q(n,t)}{n}\right).$$

When $n^{4/5} \le t \le \theta n - 2\sqrt{n\log n}$ and $\alpha = \sqrt{n\log n}$, from the fact that $t + \alpha \le \theta n - \alpha$, the above upper bound simplifies to

$$A_q(n, 2t+1) \lesssim \frac{n}{t}\left(\frac{(t+\alpha)(n-\alpha+1)}{(q-1)n(n-\alpha+1-t)}\right)^\alpha H_q(n,t)$$

$$\le n^{1/5}\left(\frac{(\theta n - \alpha)(n-\alpha+1)}{(q-1)n((1-\theta)n+\alpha+1)}\right)^\alpha H_q(n,t).$$

Since $(n-\alpha+1)/((1-\theta)n+\alpha+1) \le (n-\alpha)/((1-\theta)n+\alpha)$ and $\theta = (q-1)(1-\theta)$, we get that

$$\left(\frac{(\theta n - \alpha)(n-\alpha+1)}{(q-1)n((1-\theta)n+\alpha+1)}\right)^\alpha \le \left(\frac{(n-\alpha/\theta)(n-\alpha)}{n(n+\alpha/(1-\theta))}\right)^\alpha \le \left(\frac{(n-\alpha/\theta)(n-\alpha)}{n^2}\right)^\alpha$$

$$= (1 - (\alpha(1+\theta^{-1})/n - \alpha^2\theta^{-1}/n^2))^\alpha$$

$$\le e^{-\alpha^2(1+\theta^{-1})n^{-1}+\alpha^3\theta^{-1}n^{-2}}$$

$$\lesssim e^{-\log n(1+\theta^{-1})}$$

$$\le n^{-2}.$$

This shows that

$$A_q(n, 2t+1) \lesssim n^{1/5} \cdot n^{-2} \cdot H_q(n,t) = o\left(\frac{H_q(n,t)}{n}\right)$$

as desired. □

Proof of Theorem 1.1(b). When $10\sqrt{n} < t \le (1-q^{-1})n - 2\sqrt{n\log n}$, by Proposition 5.3 the number of t-error correcting codes is at most

$$\binom{q^n}{\le A_q(n, 2t+1)} \le q^{nA_q(n,2t+1)} \le 2^{(\log_2 q)\cdot n \cdot o(H_q(n,t)/n)} = 2^{o(H_q(n,t))}.$$
□

Acknowledgement

References

[1] Balogh, J., Liu, H. and Sharifzadeh, M. (2017) The number of subsets of integers with no k-term arithmetic progression. *Int. Math. Res. Not. IMRN* **20** 6168–6186.

[2] Balogh, J., Treglown, A. and Wagner, A. Z. (2016) Applications of graph containers in the Boolean lattice. *Random Struct. Algorithms* **49** 845–872.

[3] Delsarte, P. (1973) An algebraic approach to the association schemes of coding theory. *Philips Res. Rep. Suppl.* **10** vi+97.

[4] Ferber, A., McKinley, G. and Samotij, W. (2020) Supersaturated sparse graphs and hypergraphs. *Int. Math. Res. Not. IMRN* **2** 378–402.

[5] Kleitman, D. J. and Winston, K. J. (1980) The asymptotic number of lattices. *Ann. Discrete Math.* **6** 243–249.

[6] Kleitman, D. J. and Winston, K. J. (1982) On the number of graphs without 4-cycles. *Discrete Math.* **41** 167–172.

[7] MacWilliams, F. J. and Sloane, N. J. A. (1977) The Theory of Error-Correcting Codes. North-Holland.

[8] Robert, J. M., Eugene, R. R., Howard, R. Jr. and Lloyd, R. W. (1977) New upper bounds on the rate of a code via the Delsarte-MacWilliams inequalities. *IEEE Trans. Inform. Theory* **23** 157–166.

[9] Samotij, W. (2015) Counting independent sets in graphs. *Eur. J. Combin.* **48** 5–18.

[10] Sapozhenko, A. (2005) Systems of containers and enumeration problems. In *International Symposium on Stochastic Algorithms*. Springer. 1–13.

[11] van Lint, J. H. (1975) A survey of perfect codes. *Rocky Mountain J. Math.* **5** 199–224.

[12] van Lint, J. H. (1999) Introduction to Coding Theory. Springer-Verlag, (3rd).