Exponential decay of intersection volume with applications on list–decodability and Gilbert–Varshamov type bound

Jaehoon Kim* Hong Liu† Tuan Tran‡

Abstract

We give some natural sufficient conditions for balls in a metric space to have small intersection. Roughly speaking, this happens when the metric space is (i) expanding and (ii) well-spread, and (iii) a certain random variable on the boundary of a ball has a small tail. As applications, we show that the volume of intersection of balls in Hamming, Johnson spaces and symmetric groups decay exponentially as their centers drift apart. To verify condition (iii), we prove some large deviation inequalities 'on a slice' for functions with Lipschitz conditions.

We then use these estimates on intersection volumes to

- obtain a sharp lower bound on list-decodability of random q-ary codes, confirming a conjecture of Li and Wootters; and
- improve the classical bound of Levenshtein from 1971 on constant weight codes by a factor linear in dimension, resolving a problem raised by Jiang and Vardy.

Our probabilistic point of view also offers a unified framework to obtain improvements on other Gilbert–Varshamov type bounds, giving conceptually simple and calculation-free proofs for q-ary codes, permutation codes, and spherical codes. Another consequence is a counting result on the number of codes, showing ampleness of large codes.

1 Introduction

A well-known fact in convex geometry states that the volume of the intersection of two Euclidean balls of the same radius in $\mathbb{R}^n$ is exponentially (in n) smaller than the two given balls. It can be proved by observing that the intersection is contained in a ball of smaller radius centered at the mid-point of the centers of the two original balls. This simple proof, however, does not extend to some discrete settings, as the intersection might no longer be enclosed by a ball of smaller radius. One such example is that of the Hamming space over a finite alphabet, one of the most studied space in theoretical computer science and information theory. Indeed, take the discrete cube $\{0,1\}^n$ endowed with the Hamming metric and let $k, r \in \mathbb{N}$ with $2k \leqslant r$. Consider the two radius-r Hamming balls A and B centered at $a = 0^n$ and $b = 1^{2k}0^{n-2k}$ respectively. Take a mid-point c of a and b, say by symmetry $c = 1^k0^{n-k}$. Then the point $x = 0^k1^r0^{n-r-k}$ lies in the intersection $A \cap B$, but it is of Hamming distance $r + k$ from the chosen mid-point c.

The expression of the intersection volume in such discrete metric spaces can usually be written out explicitly. The problem is that such expression is often cumbersome and it is a grueling task to

*Department of Mathematical Sciences, KAIST, South Korea. E-mail: `jaehoon.kim@kaist.ac.kr`. Supported by the POSCO Science Fellowship of POSCO TJ Park Foundation, and by the KAIX Challenge program of KAIST Advanced Institute for Science-X.

†Mathematics Institute, University of Warwick, UK. E-mail: `h.liu.9@warwick.ac.uk`. Supported by the UK Research and Innovation Future Leaders Fellowship MR/S016325/1.

‡Discrete Mathematics Group, Institute for Basic Science (IBS), South Korea. E-mail: `tuantran@ibs.re.kr`. Supported by the Institute for Basic Science (IBS-R029-Y1).

estimate. To illustrate, let us consider the q-ary Hamming space $\{0,1,\ldots,q-1\}^n$. Denote by $\mathrm{vol}_q(n,r)$ the volume of a radius-r q-ary Hamming ball, and by $\mathrm{vol}_q(n,r;k)$ the volume of the intersections of two radius-r balls whose centers are distance k apart. It is not hard to show that the intersection volume is

$$\mathrm{vol}_q(n,r;k) = \sum_{i+j\leqslant k} \frac{k!}{i!j!(k-i-j)!}(q-2)^{k-i-j} \sum_{t\leqslant t_{\max}} \binom{n-k}{t}(q-1)^t, \tag{1}$$

where $t_{\max} := \min(n-k, r-k+i, r-k+j)$.

Estimating the asymptotics of the right hand side above is not at all a straightforward task. Indeed, when k and r are linear in the dimension n, Jiang and Vardy [25] studied the binary case $q=2$ with the help of computer. Later, Vu and Wu [43] estimated the general q-ary case for all $q \geqslant 2$ using a discrete analog of Lagrange's multiplier and some inequalities on entropy functions; their proof, though computer-free and much cleaner, is still rather involved.

Consider the following alternative probabilistic approach to estimate the intersection volume. Let A, B be two radius-r Euclidean balls centered at $a, b \in \mathbb{R}^n$ respectively. Let $\boldsymbol{x}$ be a uniform random point drawn from A, then the ratio of the volume of the intersection $A \cap B$ and the volume of the radius-r ball is precisely the probability that $\boldsymbol{x}$ lies in $A \cap B$, that is, $\frac{\mathrm{vol}(A\cap B)}{\mathrm{vol}(A)} = \mathbb{P}(\boldsymbol{x} \in A \cap B)$. We can then bound the probability $\mathbb{P}(\boldsymbol{x} \in A \cap B)$ using for instance Talagrand's celebrated deviation inequality [38] for functions with Lipschitz condition with respect to both ℓ_1^n and ℓ_2^n norms. We refer the readers to [4, 30] for related results on concentration of measure.

We use this probabilistic approach to give some natural sufficient conditions that guarantee small intersection of balls in a metric space. The advantage of this approach is that it can be implemented in the discrete settings, provided that appropriate concentration inequalities can be proved.

1.1 Sufficient conditions for small intersection

To state our result, we need some notations. Let (X, d) be a finite metric space with d taking values in $\mathbb{N} \cup \{0\}$. For $a \in X$ and $r \in \mathbb{N}$, we write $B(a,r)$ for the ball of radius r around a and write $S(a,r)$ for the shell of all points of distance exactly r from a. We say the metric space (X, d) has *exponential growth at radius r with rate c* if for every $a \in X$ and every $t < r$,

$$\frac{\mathrm{vol}(B(a, r-t))}{\mathrm{vol}(B(a,r))} \leqslant 2e^{-ct}.$$

For $a, b \in X$, let $\ell_{a,b} : X \to \mathbb{R}$ be given by

$$\ell_{a,b}(x) = \mathsf{d}(x,b) - \mathsf{d}(x,a). \tag{2}$$

Given $r, k \in \mathbb{N}$ and $\alpha > 0$, we say that the metric space (X, d) is (r,k)-*dispersed* with constant α if for any $a, b \in X$ with $\mathsf{d}(a,b) = k$ and any $0 \leqslant i \leqslant \alpha k$,

$$\mathbb{E}_{\boldsymbol{x}\sim S(a,r-i)}\Big[\ell_{a,b}(\boldsymbol{x})\Big] \geqslant 2\alpha k,$$

where $\boldsymbol{x}$ is a uniform random point of $S(a, r-i)$.

A real-valued random variable $\boldsymbol{X}$ is K-*subgaussian* if for any $t \geqslant 0$,

$$\mathbb{P}(|\boldsymbol{X}| \geqslant t) \leqslant 2\exp\left(-t^2/K\right).$$

Our result reads as follows.

Theorem 1.1. *Let (X, d) be a finite metric space with d taking values in $\mathbb{N} \cup \{0\}$ and let $k, r \in \mathbb{N}$. Suppose*

(A1) (X, d) *has exponential growth at radius* r *with rate* $c > 0$*;*

(A2) (X, d) *is* (r, k)*-dispersed with constant* $\alpha > 0$*;*

(A3) *For any* $a, b \in X$ *with* $\mathsf{d}(a, b) = k$ *and any* $0 \leqslant i \leqslant \alpha k$*,* $\ell_{a,b}(\boldsymbol{x}) - \mathbb{E}\ell_{a,b}(\boldsymbol{x})$ *is* K*-subgaussian, where* $\ell_{a,b}$ *is as in* (2) *and* $\boldsymbol{x}$ *is drawn uniformly from* $S(a, r - i)$*.*

Then, for any $a, b \in X$ *with* $\mathsf{d}(a, b) = k$*,*

$$\frac{\mathrm{vol}(B(a, r) \cap B(b, r))}{\mathrm{vol}(B(a, r))} = 2e^{-\Omega_{c,\alpha}(1) \cdot (k + k^2/K)}.$$

The merit of Theorem 1.1 is its formulation. The conditions (A1)–(A3) are quite natural; they are inspired by properties of the Euclidean space. By our result, showing that the intersection volume is small then amounts to verifying these conditions, which are more manageable. For instance, using Theorem 1.1, we can get a conceptually simple and *calculation-free* proof that the intersection volume in (1) is exponentially small, i.e. $\frac{\mathrm{vol}_q(n, pn; k)}{\mathrm{vol}_q(n, pn)} \leqslant e^{-\Omega(k)}$, for the optimal range $0 < p < 1 - 1/q$ and all $1 \leqslant k \leqslant n$ (Lemma 4.2). It is important that the exponential bound holds for not just when $k = \Omega(n)$, but for all k, which is needed in some applications, e.g. the tightness on list-decoding capacity theorem (Theorem 2.3).

To illustrate the power of Theorem 1.1, apart from the Hamming cube example above, we shall apply it to Johnson space (Lemma 4.3) and permutation group (Lemma 4.4). Such estimates on the intersection volume of balls are useful for various problems. We will use them in Section 2 to obtain results on list-decodability of q-ary random codes with rate just below the limiting rate, and improvements on Gilbert–Varshamov type bounds for constant weight codes, q-ary codes, permutation codes and spherical codes, and the corresponding counting results.

In order to apply Theorem 1.1, it is not hard to check that the discrete metric spaces we consider have the exponential growth and they are well-dispersed. To verify the third condition that the centered random variable $\ell_{a,b}(\boldsymbol{x}) - \mathbb{E}\ell_{a,b}(\boldsymbol{x})$ is subgaussian in our applications, we prove some concentration inequalities for Lipschitz functions defined on 'slices' of the space, see Lemmas 3.2 and 3.5.

Notations. Before discussing the applications in details, let us review the terminology that will be used throughout the paper. A *code* over a finite alphabet Σ is simply a subset of Σ^n; the number n is referred to as the *length* of the code. The elements of the code are called *codewords*. If $|\Sigma| = q$, the code is called q-ary code, with the term binary used for the case $q = 2$. We say that the code has *rate* R if the number of codewords is $|\Sigma|^{Rn}$. Given two words $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$ in Σ^n, the *Hamming distance* $\Delta(x, y)$ between x and y is the number of coordinates i in which x_i and y_i differ. For a word x we denote by x_i the value of its i-th coordinate. For $x \in \{0, 1, \ldots, q-1\}^n$, we denote its *weight*, which is the number of non-zero entries in x, by $\mathrm{wt}(x)$. The *Johnson distance* between two binary words $x, y \in \{0, 1\}^n$ of the same weight is half of their Hamming distance. The q-ary entropy function $h_q \colon [0, 1] \to \mathbb{R}$ is

$$h_q(x) = x \log_q(q-1) - x \log_q x - (1-x) \log_q(1-x).$$

We generally use boldface letters for random variables. Given a finite set A, we write $\boldsymbol{x} \sim A$ for a discrete random variable $\boldsymbol{x}$ chosen uniformly from A.

2 Applications

2.1 List decoding of random codes

One of the main goals of the theory of error-correcting codes is to understand the trade-off between the rate of a code and the fraction of errors the code can tolerate during transmission over a noisy channel. There are two natural error models: Hamming's adversarial noise model, and Shannon's stochastic noise model. Channels in Shannon's world can flip each transmitted bit with certain probability, independently of other bits, while channels in Hamming's world can corrupt the codeword arbitrarily, subject only to a bound on the total number of errors.

There is a gap between Hamming and Shannon's world: one can correct twice as many errors in Shannon's world. We refer the reader to [21] for a thorough comparison. List decoding, which was introduced by Elias [14] and Wonzencraft [40] in the late 1950's, can be used to bridge the gap. In list decoding we give up unique decoding, allowing decoder to output a list of all codewords that are within Hamming distance pn from the received word. Thus, if at most pn errors occur, the list will include the correct codeword. Formally, we say that a q-ary code $\mathcal{C} \subset \Sigma^n$ is (p, L)-*list decodable* if any Hamming ball of radius pn in Σ^n contains at most L codewords.

List decoding has three important parameters: the rate R of the code, the error fraction p, and the list size L. A fundamental question in list decoding is to determine the feasible region of (R, p, L). Despite significant efforts, a full description remains elusive. In 1981, Zyablov and Pinsker [48] proved the list-decoding capacity theorem, thus giving a partial solution to the above question.

Theorem 2.1 (Zyablov and Pinsker)**.** *Let $q \geqslant 2, 0 < p < \frac{q-1}{q}$, and $\varepsilon > 0$.*

1. *There exist q-ary codes of rate $1 - h_q(p) - \varepsilon$ that are $\left(p, \lceil \frac{1}{\varepsilon} \rceil\right)$-list decodable.*

2. *Any q-ary code of rate $1 - h_q(p) + \varepsilon$ that is (p, L)-list decodable must have $L \geqslant q^{\Omega(\varepsilon n)}$.*

Theorem 2.1 establishes the optimal trade-off between the rate and the error fraction for list decoding. In particular, it shows that the list decoding capacity is $1 - h_q(p)$, which matches the capacity of Shannon's model.

The existential part of Theorem 2.1 was achieved by demonstrating that a random code of rate $1 - h_q(p) - \varepsilon$ is $(p, \lceil \frac{1}{\varepsilon} \rceil)$-list decodable with high probability. Rudra [34] proved that this result is best possible up to a constant factor, in the sense that a random code of rate $1 - h_q(p) - \varepsilon$ requires $L = \Omega_{p,q}(1/\varepsilon)$. In [20], Guruswami and Narayanan provided a more direct proof of Rudra's result. For binary codes, Li and Wootters [29] recently sharpened the argument of Guruswami and Narayanan to show that the list size of $1/\varepsilon$ in Theorem 2.1 is tight even in the leading constant factor:

Theorem 2.2 (Li and Wootters)**.** *For any $p \in (0, 1/2)$ and $\varepsilon > 0$, there exist $\gamma_{p,\varepsilon} = \exp\left(-\Omega_p\left(\frac{1}{\varepsilon}\right)\right)$ and $n_{p,\varepsilon} \in \mathbb{N}$ such that for all $n \geqslant n_{p,\varepsilon}$, a random code $\mathcal{C} \subseteq \{0,1\}^n$ of rate $R = 1 - h(p) - \varepsilon$ is with probability $1 - \exp(-\Omega_{p,\varepsilon}(n))$ not $(p, \frac{1-\gamma_{p,\varepsilon}}{\varepsilon} - 1)$-list decodable.*

Li and Wootters [29] conjectured that Theorem 2.2 generalizes to q-ary codes, for any $q \geqslant 3$. To quote their words, "our arguments only work for binary codes and do not extend to larger alphabets."

Our first application, making use of the intersection volume estimate, confirms their conjecture, showing that the list size $1/\varepsilon$ in list decoding capacity theorem is optimal for all $q \geqslant 2$.

Theorem 2.3. *Let $q \geqslant 2$. Then for any $p \in (0, \frac{q-1}{q})$ and $\varepsilon > 0$, there exist $\gamma_{p,q,\varepsilon} = \exp\left(-\Omega_{p,q}\left(\frac{1}{\varepsilon}\right)\right)$ and $n_{p,q,\varepsilon} \in \mathbb{N}$ such that for all $n \geqslant n_{p,q,\varepsilon}$, a random code $\mathcal{C} \subseteq [q]^n$ of rate $R = 1 - h_q(p) - \varepsilon$ is with probability $1 - \exp(-\Omega_{p,q,\varepsilon}(n))$ not $(p, \frac{1-\gamma_{p,q,\varepsilon}}{\varepsilon} - 1)$-list decodable.*

2.2 The sphere-covering bounds

Our second group of applications concern codes over metric spaces. Consider a metric space (X, d) and a real number $r > 0$. We say a subset C of X is an (X, d, r)-*code* if $\mathsf{d}(c, c') > r$ for any distinct codewords $c, c' \in C$. A simple covering argument shows the existence of such a code C with

$$|C| \geqslant \inf_{a \in X} \frac{m(X)}{m(B(a, r))} \tag{3}$$

for *any* finite measure m on the Borel σ-algebra of X. To see why (3) holds, one can assume C is a maximal (X, d, r)-code of finite size. From the maximality of C, we deduce that $X = \bigcup_{a \in C} B(a, r)$. By the subadditivity of measures, we then get $m(X) \leqslant \sum_{a \in C} m(B(a, r)) \leqslant |C| \cdot \sup_{a \in X} m(B(a, r))$, resulting in (3).

Improving upon the sphere-covering bound (3) is a notoriously difficult problem; more on this later. Our next result improves the bound, assuming some mild conditions on the metric space.

Theorem 2.4. *Let* (X, d) *be a finite metric space, and let* $r > 0$*. Suppose*

(P1) *(Homogeneous) For every* $s \in \mathbb{R}$*, all the balls of radius* s *have the same volume* $\mathrm{vol}(s)$*.*

Suppose further that there exist $t \in (0, r)$ *and* $K > 0$ *such that*

(P2) *(Exponential growth)* $\frac{\mathrm{vol}(r-t)}{\mathrm{vol}(r)} \leqslant e^{-K}$*; and*

(P3) *(Small intersection volume) for any* $a, b \in X$ *with* $r - t < \mathsf{d}(a, b) \leqslant r$*,* $\frac{\mathrm{vol}(B(a,r) \cap B(b,r))}{\mathrm{vol}(r)} \leqslant e^{-K}$*.*

Then there is an (X, d, r)*-code of size* $(1 - o_{K\to\infty}(1))K \cdot \frac{|X|}{\mathrm{vol}(r)}$*, and the number of* (X, d, r)*-codes is at least* $\exp\left((\frac{1}{8} + o_{K\to\infty}(1))K^2 \cdot \frac{|X|}{\mathrm{vol}(r)}\right)$*.*

Theorem 2.4 can be used in conjunction with Theorem 1.1 (for verifying condition (P3)) to give a unified proof of improvements on Gilbert-Varshamov type bounds on various models of error correction codes, which we now discuss in details. Theorem 2.4 builds on recent developments on some graph theoretic results; such approach was pioneered by the work of Jiang and Vardy [25] and by Krivelevich, Litsyn and Vardy [26].

q-ary codes

A q-ary code $\mathcal{C}$ is said to have minimum distance at least d if any two codewords in $\mathcal{C}$ have distance at least d. Given three parameters q, n and d, what is the largest possible size $A_q(n, d)$ of a q-ary length-n code with minimum distance at least d? This question has been studied extensively for almost seven decades, and remains one of the most important questions in coding theory.

For a word $x \in [q]^n$, the Hamming ball of radius d centered at x is the collection of words in $[q]^n$ with distance at most d from x. The volume of this ball does not depend on the location of x and can be expressed as

$$\mathrm{vol}_q(n, d) = \sum_{i=0}^{d} \binom{n}{i} (q-1)^i.$$

The sphere-covering bound (3), applied to the the Hamming space $([q]^n, \Delta)$, gives

$$A_q(n, d+1) \geqslant \frac{q^n}{\mathrm{vol}_q(n, d)}.$$

This is known in the literature as the famous Gilbert–Varshamov bound [19, 39] from the 1950's. For five decades this was the best asymptotic lower bound for $A_q(n, d+1)$ (see for example [22, page 95]).

The case when d is proportional to n, that is, d/n is a positive constant, is of special interest in coding theory. It is an easy exercise to see that for $d/n \geqslant (q-1)/q$, the fraction $q^n/\mathrm{vol}_q(n, d)$ is less than 2. In this case, the Gilbert–Varshamov bound gives no useful information. Thus, the value $(q-1)/q$ is a natural threshold for the ratio d/n.

In a breakthrough, Jiang and Vardy [25] improved the Gilbert–Varshamov bound, for the binary case, for $d \leqslant 0.4994n$. Extending the work of Jiang and Vardy, Vu and Wu [43] proved that if d/n is less than $(q-1)/q$, then one can improve the Gilbert–Varshamov bound by a factor linear in n. We give a short proof of the following strengthening of Vu-Wu's result, showing ampleness of large codes.

Theorem 2.5. *Let $q \geqslant 2$ and let $0 < p < \frac{q-1}{q}$ and $d = pn$. Then there exists a positive constant $c = c_{p,q}$ such that the number of q-ary length-n codes with minimum distance at least $d+1$ is at least*

$$\exp\left(cn^2 \cdot \frac{q^n}{\mathrm{vol}_q(n,d)}\right).$$

As the number of subsets of $[q]^n$ of size $o_{p,q}(1)n \cdot \frac{q^n}{\mathrm{vol}_q(n,d)}$ is $\exp\left(o_{p,q}(1)n^2 \cdot \frac{q^n}{\mathrm{vol}_q(n,d)}\right)$, Theorem 2.5 recovers the bound $A_q(n, d+1) \geqslant \Omega_{p,q}(1)n \cdot \frac{q^n}{\mathrm{vol}_q(n,d)}$ of Vu and Wu. The original proof of Vu-Wu's bound was quite complicated, and involved heavy calculations. Our proof of Theorem 2.5 is conceptual and reflects, in a clean way, the necessity of the assumption $d/n < (q-1)/q$.

Constant-weight codes

Given positive integers n, d and w, we denote by $A(n, d, w)$ the size of a largest *constant-weight code* of length n and minimum Johnson distance d all of whose codewords are in $\{0,1\}^n$ with weight w. Estimating $A(n, d, w)$ accurately is the central problem regarding constant-weight codes. With the exceptions of a few particular small cases [5] and the fixed w case [7], it remains open in general.

Thanks to symmetry, all Johnson ball of radius d in $\binom{[n]}{w}$ have the same volume

$$V_w(n,d) := \sum_{i=0}^{d} \binom{w}{i}\binom{n-w}{i}.$$

Thus, the sphere-covering bound, specialized to the Johnson space, gives

$$A(n, d+1, w) \geqslant \frac{\binom{n}{w}}{V_w(n,d)}.$$

This lower bound was obtained by Levenshtein back in 1971 [28].

Our next result provides an improvement on this 50-year-old bound of Levenshtein by a factor linear in the dimension. This resolves a problem posed by Jiang and Vardy [25].

Theorem 2.6. *Let α and λ be constants satisfying $0 < \alpha < \lambda(1-\lambda)$. There is a positive constant $c = c_{\alpha,\lambda}$ such that for $d = \alpha n$ and $w = \lambda n$*

$$A(n, d+1, w) \geqslant cn \cdot \frac{\binom{n}{w}}{V_w(n,d)}.$$

Permutation codes

Let S_n be the symmetric group of permutations on $[n]$. Consider a permutation $\sigma \in S_n$ as a codeword $(\sigma(1), \ldots, \sigma(n)) \in [n]^n$, then S_n is a subset of $[n]^n$. With this view, the Hamming distance between two permutations $\sigma, \tau \in S_n$ is naturally defined as

$$\Delta(\sigma,\tau) = \left|\{i \in [n] : \sigma(i) \neq \tau(i)\}\right|.$$

A code $\mathcal{C}$ is called a *permutation code* if $\mathcal{C} \subseteq S_n$. It is said to have minimum distance at least d if any two codewords in $\mathcal{C}$ have the Hamming distance at least d.

Permutation codes have been extensively studied, see for example [1, 2, 12, 13, 36]. It also has various applications including data transmission over power lines [8, 9, 16, 32, 42], and design of block ciphers [10]. From an extremal perspective, the most natural question for permutation codes is that for given n and d, what is the largest possible size $A^{\mathrm{per}}(n,d)$ of a length-n permutation code with minimum distance at least d? Let $\mathrm{vol}^{\mathrm{per}}(n,d)$ be the volume of a radius-d Hamming ball in S_n. Once again, the sphere-covering bound (3) yields

$$A^{\mathrm{per}}(n,d+1) \geqslant \frac{n!}{\mathrm{vol}^{\mathrm{per}}(n,d)}.$$

Tait-Vardy-Verstraëte [37], Yang-Chen-Yuan [47] and Wang-Zhang-Yang-Ge [44] further improved this to

$$A^{\mathrm{per}}(n,d+1) \geqslant \Omega(n) \cdot \frac{n!}{\mathrm{vol}^{\mathrm{per}}(n,d)} \quad \text{for } \Omega(n) \leqslant d < n/2.$$

We prove the following strengthening which recovers this bound for a larger range of distance d.

Theorem 2.7. *For given $\varepsilon \in (0,1/2)$, there exists a positive constant $c = c_\varepsilon$ such that the following holds. For $\varepsilon n < d < (1-\varepsilon)n$, $A^{\mathrm{per}}(n,d+1) \geqslant cn \cdot \frac{n!}{\mathrm{vol}^{\mathrm{per}}(n,d)}$. Furthermore, the number of length-n permutation codes with minimum distance at least d is at least*

$$\exp\left(cn^2 \cdot \frac{n!}{\mathrm{vol}^{\mathrm{per}}(n,d)}\right).$$

Spherical codes

A *spherical code* of angle θ in dimension n is a collection of vectors $x_1, \dots, x_k$ in the unit sphere $\mathbb{S}^{n-1}$ such that $\langle x_i, x_j \rangle \leqslant \cos\theta$ for every $i \neq j$, that is, any two distinct vectors form an angle at least θ. Let $A(n,\theta)$ be the size of the largest spherical code of angle θ in dimension n.

For $\theta \geqslant \pi/2$, Rankin [33] determined $A(n,\theta)$ exactly, so from now on we will assume that $\theta \in (0,\pi/2)$. For $x \in \mathbb{S}^{n-1}$, we write

$$C_\theta(x) = \{y \in \mathbb{S}^{n-1} : \langle x, y \rangle \geqslant \cos\theta\}$$

for the spherical cap of angular radius θ around x, and let $s_n(\theta)$ denote the normalized surface area of $C_\theta(x)$.

The sphere-covering bound (3) (observed by Chabauty [6], Shannon [35], and Wyner [46]) implies

$$A(n,\theta) \geqslant \frac{1}{s_n(\theta)} = (1+o(1))\sqrt{2\pi n} \cdot \frac{\cos\theta}{\sin^{n-1}\theta}.$$

For over six decades there have been no improvements to this easy lower bound. By estimating the expected size of a random spherical code drawn from a Gibbs point process, Jenssen, Joos and Perkins [24] recently improved the lower bound by a linear factor in dimension.

Theorem 2.8 (Jenssen, Joos and Perkins)**.** *For $\theta \in (0,\pi/2)$, let $c_\theta = \log \frac{\sin^2\theta}{\sqrt{(1-\cos\theta)^2(1+2\cos\theta)}}$. Then,*

$$A(n,\theta) \geqslant (1+o(1))c_\theta \cdot \frac{n}{s_n(\theta)}.$$

This bound was very recently further improved by Gil Fernández, Kim, Liu and Pikhurko [18].

Theorem 2.9 (Gil Fernández, Kim, Liu and Pikhurko)**.** *Let $\theta \in (0,\pi/2)$ be fixed. Then,*

$$A(n,\theta) \geqslant (1+o(1)) \log \frac{\sin\theta}{\sqrt{2}\sin\frac{\theta}{2}} \cdot \frac{n}{s_n(\theta)}, \qquad \text{as } n \to \infty.$$

Although Theorem 2.4 is not directly applicable to the continuous setting of spherical codes, we use discretization and the graph theoretic idea in Theorem 2.4 to give a short proof of the improvement of Jenssen, Joos and Perkins [24] in Theorem 2.8. This answers another question of Jiang and Vardy [25], who asked whether discretization approach would work for spherical codes. A closely related topic in continuous setting is the sphere packing problem, where a similar approach using integer lattice instead was utilized by Krivelevich, Litsyn and Vardy [26].

We remark that the best lower bound by Gil Fernández, Kim, Liu and Pikhurko [18] in Theorem 2.9, however, seems not attainable via discretization and requires to work directly with intrinsic properties of spherical geometry.

Organization. The rest of the paper is organized as follows. In Section 3, we prove Theorem 1.1 and concentration inequalities for Lipschitz functions over slices of Hamming spaces and symmetric group, see Lemmas 3.2 and 3.5. We then use these concentration inequalities in Section 4 to deduce bounds on the volume of intersections of Hamming/Johnson/permutation balls, see Lemmas 4.2 to 4.4. Section 5 comtains some graph theoretic tools, which will be used in Section 6 to prove Theorems 2.5 to 2.8 on improvements on sphere-covering bounds. The proof of Theorem 2.3 is given in Section 7.

3 Proof of Theorem 1.1 and concentration on the slice

In this section we will prove Theorem 1.1, and establish some new concentration inequalities that will be used to verify (A3) when applying Theorem 1.1. Concentration inequalities are fundamental tools in probabilistic combinatorics and theoretical computer science for proving that nice random variables are near their means. The main principle is that a random function that smoothly depends on many independent random variables should be sharply concentrated. The new concentration inequalities we need are for functions of *dependent* random variables. Our proofs use coupling techniques.

Proof of Theorem 1.1. Let $T = B(a,r) \cap B(b,r)$, and let $\boldsymbol{\eta} \sim B(a,r)$. Then

$$\frac{\mathrm{vol}(B(a,r) \cap B(b,r))}{\mathrm{vol}(B(a,r))} = \mathbb{P}(\boldsymbol{\eta} \in T).$$

By definition, $\boldsymbol{\eta}$ lies in T if and only if it is of distance at most r from b, i.e.

$$\mathbb{P}(\boldsymbol{\eta} \in T) = \mathbb{P}(\mathsf{d}(\boldsymbol{\eta}, b) \leqslant r).$$

As the metric space has exponential growth at radius r, $\mathbb{P}(\mathsf{d}(\boldsymbol{\eta}, a) \leqslant r - \alpha k) \leqslant 2e^{-\Omega(k)}$. Thus,

$$\begin{aligned}\mathbb{P}(\boldsymbol{\eta} \in T) &\leqslant \mathbb{P}(\boldsymbol{\eta} \in T \,|\, \mathsf{d}(\boldsymbol{\eta}, a) > r - \alpha k) \cdot \mathbb{P}(\mathsf{d}(\boldsymbol{\eta}, a) > r - \alpha k) + \mathbb{P}(\mathsf{d}(\boldsymbol{\eta}, a) \leqslant r - \alpha k) \\ &\leqslant \sum_{i=0}^{\alpha k} \mathbb{P}\big(\mathsf{d}(\boldsymbol{\eta}, b) \leqslant r \,\big|\, \mathsf{d}(\boldsymbol{\eta}, a) = r - i\big) \cdot \mathbb{P}(\mathsf{d}(\boldsymbol{\eta}, a) = r - i) + 2e^{-\Omega(k)} \\ &\leqslant \max_{0 \leqslant i \leqslant \alpha k} \mathbb{P}\big(\mathsf{d}(\boldsymbol{\eta}, b) \leqslant r \,\big|\, \mathsf{d}(\boldsymbol{\eta}, a) = r - i\big) + 2e^{-\Omega(k)}.\end{aligned}$$

Fix an arbitrary $0 \leqslant i \leqslant \alpha$, and let $\boldsymbol{x} \sim S(a, r-i)$. Note that, conditioning on $\mathsf{d}(\boldsymbol{\eta}, a) = r - i$, $\boldsymbol{\eta}$ and $\boldsymbol{x}$ are identically distributed. We thus have

$$\begin{aligned}\mathbb{P}\big(\mathsf{d}(\boldsymbol{\eta}, b) \leqslant r \,\big|\, \mathsf{d}(\boldsymbol{\eta}, a) = r - i\big) &= \mathbb{P}\big(\mathsf{d}(\boldsymbol{\eta}, b) - \mathsf{d}(\boldsymbol{\eta}, a) \leqslant i \,\big|\, \mathsf{d}(\boldsymbol{\eta}, a) = r - i\big) \\ &= \mathbb{P}(\mathsf{d}(\boldsymbol{x}, b) - \mathsf{d}(\boldsymbol{x}, a) \leqslant i) \\ &= \mathbb{P}(\ell_{a,b}(\boldsymbol{x}) \leqslant i).\end{aligned}$$

Using that (X, d) is (r,k)-dispersed with constant α, we see that $\mathbb{E}\ell_{a,b}(\boldsymbol{x}) \geqslant 2\alpha k$. Consequently, $i - \mathbb{E}\ell_{a,b}(\boldsymbol{x}) \leqslant i - 2\alpha k \leqslant -\alpha k$. Thus, since $\ell_{a,b}(\boldsymbol{x}) - \mathbb{E}\ell_{a,b}(\boldsymbol{x})$ is K-subgaussian, we get

$$\begin{aligned}\mathbb{P}(\ell_{a,b}(\boldsymbol{x}) \leqslant i) &= \mathbb{P}(\ell_{a,b}(\boldsymbol{x}) - \mathbb{E}\ell_{a,b}(\boldsymbol{x}) \leqslant i - \mathbb{E}\ell_{a,b}(\boldsymbol{x})) \\ &\leqslant \mathbb{P}(\ell_{a,b}(\boldsymbol{x}) - \mathbb{E}\ell_{a,b}(\boldsymbol{x}) \leqslant -\alpha k) \\ &\leqslant 2e^{-\Omega(k^2/K)},\end{aligned}$$

as desired. □

3.1 Slices of the q-ary cube

One of the most natural and easy-to-verify smoothness assumptions that one may consider is the so-called bounded differences condition.

Definition 3.1 (Bounded differences condition)**.** A function $f\colon \Omega^n \to \mathbb{R}$ is said to satisfy the *bounded differences condition* with parameters $(c_1, \ldots, c_n) \in \mathbb{R}^n$ if for every $x, x' \in \Omega^n$

$$|f(x) - f(x')| \leqslant \sum_{i=1}^{n} c_i \mathbb{1}_{\{x_i \neq x_i'\}}.$$

In the proof of Theorems 2.3, 2.5 and 2.6 we will use the following "non-uniform" concentration inequality.

Lemma 3.2. *Suppose $f\colon \{0,1,\ldots,q-1\}^n \to \mathbb{R}$ satisfies the bounded differences condition with parameters $(c_1,\ldots,c_n)$ and that $\boldsymbol{\eta}$ is drawn uniformly at random from $\{0,1,\ldots,q-1\}^n$ subject to* $\mathrm{wt}(\boldsymbol{\eta}) = k$*. Then*

$$\mathbb{P}(|f(\boldsymbol{\eta}) - \mathbb{E}f(\boldsymbol{\eta})| \geqslant t) \leqslant 2\exp\left(-\frac{t^2}{68\sum_{i=1}^n c_i^2}\right) \quad \textit{for all } t \geqslant 0.$$

The binary case above is Lemma 2.1 from [27]. For completeness, we include its short proof.

Lemma 3.3 ([27])**.** *Suppose $g\colon \{0,1\}^n \to \mathbb{R}$ satisfies the bounded differences condition with parameters $(c_1,\ldots,c_n)$ and that $\boldsymbol{\xi} \in \{0,1\}^n$ is a random vector uniformly distributed in $\binom{[n]}{k}$. Then*

$$\mathbb{P}(|g(\boldsymbol{\xi}) - \mathbb{E}g(\boldsymbol{\xi})| \geqslant t) \leqslant 2\exp\left(-\frac{t^2}{8\sum_{i=1}^n c_i^2}\right) \quad \textit{for all } t \geqslant 0.$$

Proof. We may assume without loss of generality that $c_1 \geqslant \cdots \geqslant c_n$. Consider the Doob martingale $\boldsymbol{Z}_i = \mathbb{E}\left[g\left(\boldsymbol{\xi}\right)|\xi_1,\ldots\xi_i\right]$, so $\boldsymbol{Z}_0 = \mathbb{E}g\left(\boldsymbol{\xi}\right)$ and $\boldsymbol{Z}_n = \boldsymbol{Z}_{n-1} = g\left(\boldsymbol{\xi}\right)$. Let $\mathcal{L}\left(x_1,\ldots,x_i\right)$ be the conditional distribution of $\boldsymbol{\xi}$ given $\xi_1 = x_1, \ldots, \xi_i = x_i$.

We want to show that

$$|\mathbb{E}\left[g\left(\mathcal{L}\left(x_1,\ldots,x_{i-1},0\right)\right)\right] - \mathbb{E}\left[g\left(\mathcal{L}\left(x_1,\ldots,x_{i-1},1\right)\right)\right]| \leqslant 2c_i$$

for all feasible $x_1,\ldots,x_{i-1} \in \{0,1\}$; this will imply that $|\boldsymbol{Z}_i - \boldsymbol{Z}_{i-1}|$ is uniformly bounded by $2c_i$, so the desired result will follow from the Azuma–Hoeffding bound (see for example [17, Theorem 22.16]).

If $\boldsymbol{\xi}$ is distributed as $\mathcal{L}\left(x_1,\ldots,x_{i-1},0\right)$, we can change ξ_i to 1 and then randomly choose one of the ones among $\xi_{i+1},\ldots,\xi_n$ and change it to 0; we thereby obtain the distribution $\mathcal{L}\left(x_1,\ldots,x_{i-1},1\right)$. This provides a coupling between $\mathcal{L}\left(x_1,\ldots,x_{i-1},0\right)$ and $\mathcal{L}\left(x_1,\ldots,x_{i-1},1\right)$ that differs in only two coordinates i and $j > i$, and since $c_j \leqslant c_i$ this implies the required bound. □

We also require some standard facts about subgaussian random variables (see for instance [41, Proposition 2.5.2]).

Lemma 3.4 (Subgaussian properties)**.** *Let $\boldsymbol{X}$ be a random variable with mean zero. Then the following properties are equivalent.*

(i) There exists $K_1 > 0$ such that the tails of $\boldsymbol{X}$ satisfy

$$\mathbb{P}(|\boldsymbol{X}| \geqslant t) \leqslant 2\exp\left(-t^2/K_1\right) \quad \text{for all } t \geqslant 0.$$

(ii) There exists $K_2 > 0$ such that the moment generating function of $\boldsymbol{X}$ satisfies

$$\mathbb{E}\exp(\lambda \boldsymbol{X}) \leqslant \exp\left(K_2\lambda^2\right) \quad \text{for all } \lambda \geqslant 0.$$

In particular, for (i) $\implies$ (ii), we can take $K_2 = 2K_1$ and for (ii) $\implies$ (i), we can take $K_1 = 4K_2$.

We now have all the tools to prove Lemma 3.2.

Proof of Lemma 3.2. Let $\boldsymbol{\xi} \in \{0,1\}^n$ be a random vector uniformly distributed in $\binom{[n]}{k}$. Let $\boldsymbol{u}$ be drawn uniformly from $[q-1]^n$, independently from $\boldsymbol{\xi}$. Then the distribution of $\boldsymbol{\eta}$ coincides with the distribution of

$$\boldsymbol{u} \star \boldsymbol{\xi} := (u_1\xi_1, \ldots, u_n\xi_n).$$

Writing $\|c\|^2 = \sum_{i=1}^n c_i^2$, by Lemma 3.4, it suffices to show that

$$\mathbb{E}_{\boldsymbol{u}}\mathbb{E}_{\boldsymbol{\xi}} e^{\lambda(f(\boldsymbol{u}\star\boldsymbol{\xi}) - \mathbb{E}_{\boldsymbol{u},\boldsymbol{\xi}} f(\boldsymbol{u}\star\boldsymbol{\xi}))} \leqslant e^{17\|c\|^2\lambda^2}. \tag{4}$$

Fix an instance of $\boldsymbol{u}$. Note that, as $f(\cdot)$, $f(\boldsymbol{u}\star\cdot)$ also satisfies the bounded differences condition with parameters $c = (c_1, \ldots, c_n)$. Then, by Lemma 3.3 with $f(\boldsymbol{u}\star\cdot)$ playing the role of $g(\cdot)$ and Lemma 3.4, we get that

$$\mathbb{E}_{\boldsymbol{\xi}} e^{\lambda(f(\boldsymbol{u}\star\boldsymbol{\xi}) - \mathbb{E}_{\boldsymbol{\xi}} f(\boldsymbol{u}\star\boldsymbol{\xi}))} \leqslant e^{16\|c\|^2\lambda^2}.$$

Thus,

$$\begin{aligned}\mathbb{E}_{\boldsymbol{u}}\mathbb{E}_{\boldsymbol{\xi}} e^{\lambda(f(\boldsymbol{u}\star\boldsymbol{\xi}) - \mathbb{E}_{\boldsymbol{u},\boldsymbol{\xi}} f(\boldsymbol{u}\star\boldsymbol{\xi}))} &= e^{-\lambda\mathbb{E}_{\boldsymbol{u},\boldsymbol{\xi}} f(\boldsymbol{u}\star\boldsymbol{\xi})} \cdot \mathbb{E}_{\boldsymbol{u}} e^{\lambda\mathbb{E}_{\boldsymbol{\xi}} f(\boldsymbol{u}\star\boldsymbol{\xi})} \mathbb{E}_{\boldsymbol{\xi}} e^{\lambda(f(\boldsymbol{u}\star\boldsymbol{\xi}) - \mathbb{E}_{\boldsymbol{\xi}} f(\boldsymbol{u}\star\boldsymbol{\xi}))} \\ &\leqslant e^{16\|c\|^2\lambda^2} \cdot \mathbb{E}_{\boldsymbol{u}} e^{\lambda(\mathbb{E}_{\boldsymbol{\xi}} f(\boldsymbol{u}\star\boldsymbol{\xi}) - \mathbb{E}_{\boldsymbol{u}}\mathbb{E}_{\boldsymbol{\xi}} f(\boldsymbol{u}\star\boldsymbol{\xi}))}.\end{aligned} \tag{5}$$

It is easy to check that $g(\cdot) := \mathbb{E}_{\boldsymbol{\xi}} f(\cdot \star \boldsymbol{\xi})$ also has the bounded differences condition with parameters c. Thus by McDiarmid's inequality (see for example [17, Theorem 22.17]),

$$\mathbb{P}(|g(\boldsymbol{u}) - \mathbb{E}_{\boldsymbol{u}} g(\boldsymbol{u})| \geqslant t) \leqslant 2e^{-\frac{2t^2}{\|c\|^2}}$$

and so by Lemma 3.4,

$$\mathbb{E}_{\boldsymbol{u}} e^{\lambda(g(\boldsymbol{u}) - \mathbb{E}_{\boldsymbol{u}} g(\boldsymbol{u}))} \leqslant e^{\|c\|^2\lambda^2}.$$

This, together with (5), implies (4) and completes the proof. □

3.2 Slices of the symmetric group

The proof of Theorem 2.7 relies on the following concentration inequality for functions over slices of the symmetric group. We define the *weight* of a permutation σ in S_n to be the Hamming distance between σ and the identity.

Lemma 3.5. *Let $S_{n,k}$ be the set of all permutations in S_n with weight k. Suppose $f\colon S_{n,k} \to \mathbb{R}$ satisfies*

$$|f(\sigma) - f(\tau)| \leqslant \Delta(\sigma, \tau) \quad \text{for all } \sigma, \tau \in S_{n,k}. \tag{6}$$

Let $\boldsymbol{\sigma}$ be drawn uniformly at random from $S_{n,k}$. Then

$$\mathbb{P}(|f(\boldsymbol{\sigma}) - \mathbb{E} f(\boldsymbol{\sigma})| \geqslant t) \leqslant 2\exp(-t^2/72k) \quad \text{for all } t \geqslant 0.$$

To prove Lemma 3.5 we will use a coupling argument together with two well-known concentration inequalities. The first is a simple consequence of the Azuma–Hoeffding bound, obtained by Wormald [45, Theorem 2.19].

Theorem 3.6 (Wormald [45])**.** *Let $D_n \subset S_n$ be the set of derangements, that is, $\sigma \in D_n$ if and only if $\sigma(i) \neq i$ for all $i \in [n]$. Suppose $f: D_n \to \mathbb{R}$ satisfies*

$$|f(\sigma) - f(\tau)| \leqslant \Delta(\sigma, \tau) \quad \text{for all } \sigma, \tau \in D_n.$$

Let $\boldsymbol{\sigma}$ be a uniformly random element of D_n. Then

$$\mathbb{P}(|f(\boldsymbol{\sigma}) - \mathbb{E}f(\boldsymbol{\sigma})| \geqslant t) \leqslant 2\exp(-t^2/4n) \quad \text{for all } t \geqslant 0.$$

We also need a Gaussian bound for Lipshitz functions on slices of the discrete cube, due to Bobkov [3, Theorem 2.1].

Theorem 3.7 (Bobkov [3])**.** *Let $g: \binom{[n]}{k} \to \mathbb{R}$ be a function such that*

$$|g(J) - g(J')| \leqslant 1.$$

for any $J, J' \in \binom{[n]}{k}$ with $|J \cap J'| = k-1$. Let $\boldsymbol{I}$ be a uniformly random element of $\binom{[n]}{k}$. Then

$$\mathbb{P}\left(|g(\boldsymbol{I}) - \mathbb{E}g(\boldsymbol{I})| \geqslant t\right) \leqslant 2\exp\left(-\frac{t^2}{\min\{k, n-k\}}\right) \quad \text{for all } t \geqslant 0.$$

Proof of Lemma 3.5. For $J \subseteq [n]$, let D_J be the set of all permutations which has exactly $[n]\backslash J$ as the set of fixed points, in other words,

$$D_J = \{\sigma \in S_n : \sigma(i) \neq i \text{ if and only if } i \in J\}. \tag{7}$$

It is not difficult to see that the following two-step random process yields the uniform distribution on the set of all permutations in S_n with weight k:

1. Choose a set $\boldsymbol{I}$ uniformly at random from $\binom{[n]}{k}$, and
2. Choose a permutation $\boldsymbol{\sigma}$ uniformly at random from $D_{\boldsymbol{I}}$.

For a set $J \in \binom{[n]}{k}$, let $g(J)$ be the average of f over D_J, that is,

$$g(J) = \mathbb{E}_{\boldsymbol{I},\boldsymbol{\sigma}}[f(\boldsymbol{\sigma}) \mid \boldsymbol{I} = J] = \mathbb{E}_{\boldsymbol{\sigma} \sim D_J} f(\boldsymbol{\sigma}).$$

Claim 3.8. *For any $J, J' \in \binom{[n]}{k}$ with $|J \cap J'| = k-1$, we have*

$$|g(J) - g(J')| \leqslant 3.$$

Proof of claim. Let j be the element in $J\backslash J'$ and j' be the element in $J'\backslash J$. For each permutation $\sigma \in D_J$, we define a permutation $\widetilde{\sigma} \in D_{J'}$ as follows:

- $\widetilde{\sigma}(i) = i$ for all $i \notin \{j, j', \sigma^{-1}(j)\}$,
- $\widetilde{\sigma}(j) = j$,
- $\widetilde{\sigma}(j') = \sigma(j)$, and
- $\widetilde{\sigma}(\sigma^{-1}(j)) = j'$.

Since $\sigma(j) \neq j$ and $\sigma(j') = j'$, we see that $\sigma^{-1}(j) \notin \{j, j'\}$, $\sigma(j) \neq j'$, and $j' \neq \sigma^{-1}(j)$. Thus $\widetilde{\sigma}$ is a permutation in $D_{J'}$. Also it is easy to see that the map $\sigma \mapsto \widetilde{\sigma}$ is a bijection from D_J to $D_{J'}$. As σ and $\widetilde{\sigma}$ differ only at three places, by the hypothesis we have $|f(\sigma) - f(\widetilde{\sigma})| \leqslant \Delta(\sigma, \widetilde{\sigma}) = 3$. Therefore,

$$|g(J) - g(J')| = \left|\mathbb{E}_{\boldsymbol{\sigma} \sim D_J}[f(\boldsymbol{\sigma}) - f(\widetilde{\boldsymbol{\sigma}})]\right| \leqslant 3,$$

as desired. ■

Let μ be the mean of f. Then note that

$$\mu = \mathbb{E}g(\boldsymbol{I}).$$

By the triangle inequality,

$$\mathbb{P}(|f(\boldsymbol{\sigma}) - \mu| \geqslant t) \leqslant \mathbb{P}(|g(\boldsymbol{I}) - \mu| \geqslant t/2) + \mathbb{P}(|f(\boldsymbol{\sigma}) - g(\boldsymbol{I})| \geqslant t/2).$$

For the first term, recalling Claim 3.8 and applying Theorem 3.7 to $\frac{1}{3}g$, we get

$$\mathbb{P}\{|g(\boldsymbol{I}) - \mu| \geqslant t/2\} \leqslant 2\exp(-t^2/36k).$$

For the second term, note that $g(\boldsymbol{I}) = \mathbb{E}_{\boldsymbol{\sigma} \sim D_{\boldsymbol{I}}} f(\boldsymbol{\sigma})$ for each instance of $\boldsymbol{I}$. Once $\boldsymbol{I}$ is fixed, for $\boldsymbol{\sigma} \sim D_{\boldsymbol{I}}$, we can view $f(\boldsymbol{\sigma})$ as a function from $D_{\boldsymbol{I}}$ to $\mathbb{R}$. Then, by Eq. (6), we can apply Theorem 3.6 to f and get

$$\mathbb{P}(|f(\boldsymbol{\sigma}) - g(\boldsymbol{I})| \geqslant t/2) \leqslant 2\exp(-t^2/16k).$$

Therefore,

$$\mathbb{P}(|f(\boldsymbol{\sigma}) - \mu| \geqslant t) \leqslant 4\exp(-t^2/36k).$$

As the left side is at most one, we get $\mathbb{P}(|f(\boldsymbol{\sigma}) - \mu| \geqslant t) \leqslant 2\exp(-t^2/72k)$. □

4 Small intersection

In this section, we will verify the conditions of Theorem 1.1 for Hamming/Johnson/permutation spaces, using the concentration inequalities proved in previous section, to show that the intersection of balls in these spaces has small volume.

As these metric spaces (X, d) have the property that the balls of the same radius have the same volume independent of the center point, we will use $\mathrm{vol}(r)$ throughout this section to denote the volume of a radius-r ball in X.

We start with the Hamming space. We will need the following standard estimate on the volume of a Hamming ball.

Lemma 4.1. *Suppose that $0 < p < 1 - 1/q$ and that $1 \leqslant \alpha n \leqslant pn$. Then*

$$\mathrm{vol}_q(n, \alpha n) = \Theta_{p,q}(1) \cdot \frac{q^{h_q(\alpha)n}}{\sqrt{\alpha n}}.$$

The Hamming space satisfies the conditions of Theorem 1.1 as follows.

Lemma 4.2. *Let $0 < p < \frac{q-1}{q}$, and let k be any positive integer. Consider $X = \{0, 1, \ldots, q-1\}^n$ endowed with the Hamming distance Δ. Then (X, Δ) satisfies the conditions* (A1)–(A3) *of Theorem 1.1 as follows.*

(A1) *(X, Δ) has exponential growth at radius pn with rate $c = \Omega_{p,q}(1)$.*

(A2) *(X, Δ) is (pn, k)-dispersed with constant $\alpha = \frac{1}{2}(1 - \frac{pq}{q-1}) > 0$.*

(A3) *For any* $a, b \in X$ *with* $\Delta(a,b) = k$ *and any* $0 \leqslant i \leqslant \alpha k$, $\ell_{a,b}(\boldsymbol{x}) - \mathbb{E}\ell_{a,b}(\boldsymbol{x})$ *is* $400k$*-subgaussian, where* $\ell_{a,b}$ *is as in* (2) *and* $\boldsymbol{x}$ *is drawn uniformly from* $S(a, pn - i)$*.*

Consequently, for every $a, b \in X$*,*

$$\frac{\operatorname{vol}(B(a,r) \cap B(b,r))}{\operatorname{vol}(B(a,r))} = 2e^{-\Omega_{p,q}(1)\cdot\Delta(a,b)}. \tag{8}$$

Proof. (A1) Consider $t < pn$. By the mean value theorem, $h_q(p) - h_q(p - t/n) = h_q'(x)t/n$ for some $x \in (p - t/n, p)$. Together with Lemma 4.1, this yields

$$\frac{\operatorname{vol}(pn)}{\operatorname{vol}(pn-t)} = \Omega_{p,q}(1) \cdot \frac{\sqrt{pn-t}}{\sqrt{pn}} \cdot q^{(h_q(p) - h_q(p-t/n))n} = \Omega_{p,q}(1) \cdot \frac{\sqrt{pn-t}}{\sqrt{pn}} \cdot q^{h_q'(x)t}.$$

As $x \leqslant p < 1 - 1/q$, we have $h_q'(x) = \log_q(q-1) - \log_q \frac{x}{1-x} \geqslant \log_q(q-1) - \log_q \frac{p}{1-p} > 0$. Letting $\varepsilon = \log_q(q-1) - \log_q \frac{p}{1-p}$, we thus get

$$\frac{\operatorname{vol}(pn)}{\operatorname{vol}(pn-t)} \geqslant \Omega_{p,q}(1) \cdot \frac{\sqrt{pn-t}}{\sqrt{pn}} \cdot q^{\varepsilon t}.$$

If $t \leqslant pn/2$, then $\frac{\sqrt{pn-t}}{\sqrt{pn}} \geqslant 1/2$; while $\frac{\sqrt{pn-t}}{\sqrt{pn}} \cdot q^{\varepsilon t} \geqslant q^{\varepsilon t/2}$ if $pn/2 \leqslant t \leqslant pn - 1$ and pn is sufficiently large. Hence $\frac{\operatorname{vol}(pn-t)}{\operatorname{vol}(pn)} \leqslant O_{p,q}(1) \cdot q^{-\varepsilon t/2}$ in either case. As the left side is at most one, we conclude that there exists $c = \Omega_{p,q}(1)$ such that $\frac{\operatorname{vol}(pn-t)}{\operatorname{vol}(pn)} \leqslant 2e^{-ct}$ for all $t < pn$.

(A2) Consider any two points $a, b \in X$ with $\Delta(a,b) = k$. Let $0 \leqslant i \leqslant \alpha k$, and let $\boldsymbol{x} \sim S(a, pn - i)$. We can assume $a = 0^n$ and $b = 1^k 0^{n-k}$. Write $\gamma = \mathbb{P}(x_1 \neq 1) = \cdots = \mathbb{P}(x_1 \neq q-1)$ and $\delta = \mathbb{P}(x_1 \neq 0)$. Then $\delta = \frac{pn-i}{n} \leqslant p$. Moreover, note that $(q-1)\gamma + \delta = q - 1$, and so $\gamma = 1 - \frac{\delta}{q-1} \geqslant 1 - \frac{p}{q-1}$. By the linearity of expectation we have

$$\begin{aligned}
\mathbb{E}\ell_{a,b}(\boldsymbol{x}) &= \sum_{i=1}^{k} \left(\mathbb{P}(x_i \neq 1) - \mathbb{P}(x_i \neq 0)\right) \\
&= k(\gamma - \delta) \\
&\geqslant k\big(1 - \frac{p}{q-1} - p\big) = 2\alpha k,
\end{aligned}$$

where the second equality follows from the symmetry.

(A3) Assume $a = 0^n$ and $b = 1^k 0^{n-k}$. It is easy to see that the function $\ell_{a,b}$ satisfies the bounded difference condition with parameters $(2, \ldots, 2, 0, \ldots, 0)$ where only the first k coordinates are non-zero. Let $0 \leqslant i \leqslant \alpha k$, and let $\boldsymbol{x} \sim S(a, pn - i)$. By Lemma 3.2, $\ell_{a,b}(\boldsymbol{x}) - \mathbb{E}\ell_{a,b}(\boldsymbol{x})$ is $400k$-subgaussian. □

Our next result justifies the conditions of Theorem 1.1 for the Johnson space.

Lemma 4.3. *Let* β, λ *and* ε *be real numbers with* $0 < \varepsilon < 1/10$ *and* $0 < \beta < (1-\varepsilon)\lambda(1-\lambda)$*. Let* k *be any positive integer. Consider the slice* $X = \binom{[n]}{\lambda n}$ *endowed with the Johnson distance* d*. Then* (X, d) *satisfies the conditions* (A1)–(A3) *of Theorem 1.1 as follows.*

(A1) (X, d) *has exponential growth at radius* βn *with rate* ε^2*;*

(A2) (X, d) *is* $(\beta n, k)$*-dispersed with constant* $\varepsilon/2$*;*

(A3) *For any* $a, b \in X$ *with* $\Delta(a,b) = k$ *and any* $0 \leqslant i \leqslant \varepsilon k$, $\ell_{a,b}(\boldsymbol{x}) - \mathbb{E}\ell_{a,b}(\boldsymbol{x})$ *is* $8\beta n$*-subgaussian, where* $\boldsymbol{x} \sim S(a, \beta n - i)$*.*

Consequently, for every $a, b \in X$,

$$\frac{\operatorname{vol}(B(a,\beta n)\cap B(b,\beta n))}{\operatorname{vol}(B(a,\beta n))} = 2e^{-\Omega_\varepsilon(1)\cdot\left(\mathsf{d}(a,b)+\mathsf{d}(a,b)^2/(\beta n)\right)}. \tag{9}$$

Proof. (A1) We wish to show that $\operatorname{vol}(\beta n - t)/\operatorname{vol}(\beta n) \leqslant 2e^{-\varepsilon^2 t}$ for all $t \leqslant \beta n$. Since the left side is at most one, we can assume $t \geqslant 1/(2\varepsilon^2)$. Recall that $\operatorname{vol}(d) = \sum_{i=0}^{d}\binom{\lambda n}{i}\binom{(1-\lambda)n}{i}$ for all non-negative integer d. For $1 \leqslant i \leqslant \beta n$, we have

$$\begin{aligned}\frac{\binom{\lambda n}{i}\binom{(1-\lambda)n}{i}}{\binom{\lambda n}{i-1}\binom{(1-\lambda)n}{i-1}} &= \frac{(\lambda n - i + 1)((1-\lambda)n - i + 1)}{i^2}\\ &\geqslant \frac{(\lambda-\beta)((1-\lambda)-\beta)}{\beta^2}\\ &= 1 + \frac{\lambda(1-\lambda)-\beta}{\beta^2} \geqslant 1 + 4\varepsilon. \end{aligned}\tag{10}$$

It follows that

$$\operatorname{vol}(\beta n - t) \leqslant \binom{\lambda n}{\beta n - t}\binom{(1-\lambda)n}{\beta n - t}\cdot\sum_{i=0}^{\beta n - t}(1+4\varepsilon)^{-i} \leqslant \binom{\lambda n}{\beta n - t}\binom{(1-\lambda)n}{\beta n - t}\cdot\frac{1+4\varepsilon}{4\varepsilon}.$$

Furthermore, (10) implies $\operatorname{vol}(\beta n) \geqslant \binom{\lambda n}{\beta n}\binom{(1-\lambda)n}{\beta n} \geqslant \binom{\lambda n}{\beta n - t}\binom{(1-\lambda)n}{\beta n - t}\cdot(1+4\varepsilon)^t$. Therefore, we have $\frac{\operatorname{vol}(\beta n - t)}{\operatorname{vol}(\beta n)} \leqslant \frac{1+4\varepsilon}{4\varepsilon}\cdot(1+4\varepsilon)^{-t} \leqslant 2e^{-\varepsilon^2 t}$ assuming $0 < \varepsilon \leqslant 1/10$ and $t \geqslant 1/(2\varepsilon^2)$.

(A2) Consider any two points $a, b \in X$ with $\mathsf{d}(a,b) = k$. Let $0 \leqslant i \leqslant \varepsilon k/2$, and let $\boldsymbol{x} \sim S(a, \beta n - i)$. We can assume $a = 1^{\lambda n}0^{(1-\lambda)n}$ and $b = 0^k 1^{\lambda n} 0^{(1-\lambda)n-k}$. Since $\boldsymbol{x} \in \binom{[n]}{\lambda n}$ and $\mathsf{d}(\boldsymbol{x}, a) = \beta n - i$, we find $\sum_{j=1}^{\lambda n} x_j = (\lambda-\beta)n + i$ and $\sum_{j=\lambda n+1}^{n} x_j = \beta n - i$. We thus get $\mathbb{E}x_1 = \cdots = \mathbb{E}x_{\lambda n} = \frac{(\lambda-\beta)n+i}{\lambda n}$ and $\mathbb{E}x_{\lambda n+1} = \cdots = \mathbb{E}x_n = \frac{\beta n - i}{(1-\lambda)n}$, by the symmetry. Furthermore, notice that

$$\ell_{a,b}(\boldsymbol{x}) = \mathsf{d}(\boldsymbol{x}, b) - \mathsf{d}(\boldsymbol{x}, a) = \frac{1}{2}\sum_{j=1}^{k}(2x_j - 1) + \frac{1}{2}\sum_{j=\lambda n+1}^{\lambda n+k}(1 - 2x_j) = \sum_{j=1}^{k} x_j - \sum_{j=\lambda n+1}^{\lambda n+k} x_j.$$

Therefore, by linearity of expectation, we obtain

$$\mathbb{E}\ell_{a,b}(\boldsymbol{x}) = k\cdot\left(\frac{(\lambda-\beta)n+i}{\lambda n} - \frac{\beta n - i}{(1-\lambda)n}\right) \geqslant k\cdot\frac{\lambda(1-\lambda)-\beta}{\lambda(1-\lambda)} \geqslant \varepsilon k,$$

as desired.

(A3) Without loss of generality we can assume $a = 1^{\lambda n}0^{(1-\lambda)n}$ and $b = 0^k 1^{\lambda n}0^{(1-\lambda)n-k}$. We wish to show $\mathbb{P}(|\ell_{a,b}(\boldsymbol{x}) - \mathbb{E}\ell_{a,b}(\boldsymbol{x})| \geqslant t) \leqslant 2e^{-t^2/(8\beta n)}$ for all $t \geqslant 0$. As the left side is at most one, we may assume $2e^{-t^2/(8\beta n)} \leqslant 1$. Observe that $\boldsymbol{x} \sim S(a, \beta n - i)$ is a concatenation of two independent random vectors $(x_1, \ldots, x_{\lambda n}) \sim \binom{[\lambda n]}{(\lambda-\beta)n+i}$ and $(x_{\lambda n+1}, \ldots, x_n) \sim \binom{[n]\setminus[\lambda n]}{\beta n - i}$. Moreover, we can decompose $\ell_{a,b}(\boldsymbol{x}) = f(x_1, \ldots, x_{\lambda n}) + g(x_{\lambda n+1}, \ldots, x_n)$, where $f(x_1, \ldots, x_{\lambda n}) = \sum_{j=1}^{k} x_j$ and $g(x_{\lambda n+1}, \ldots, x_n) = -\sum_{j=\lambda n+1}^{\lambda n+k} x_j$. Applying Theorem 3.7 to f and g, we therefore get

$$\begin{aligned}\mathbb{P}(|\ell_{a,b}(\boldsymbol{x}) - \mathbb{E}\ell_{a,b}(\boldsymbol{x})| \geqslant t) &\leqslant \mathbb{P}(|f - \mathbb{E}f| \geqslant t/2) + \mathbb{P}(|g - \mathbb{E}g| \geqslant t/2)\\ &\leqslant 4\exp\left(-\frac{t^2}{4(\beta n - i)}\right) \leqslant 4e^{-t^2/(4\beta n)} \leqslant 2e^{-t^2/(8\beta n)},\end{aligned}$$

where the last inequality holds as $2e^{-t^2/(8\beta n)} \leqslant 1$. This completes our proof. □

The last result of this section confirms the conditions of Theorem 1.1 for the permutation space.

Lemma 4.4. *Let $0 < \varepsilon < 0.01$, $1 \leqslant r \leqslant (1-\varepsilon)n$ and $k \geqslant 6/\varepsilon$. Consider the symmetric group S_n endowed with the Hamming distance Δ. Then (S_n, Δ) satisfies the conditions* (A1)–(A3) *of Theorem 1.1 as follows.*

(A1) *(S_n, Δ) has exponential growth at radius r with rate ε;*

(A2) *(S_n, Δ) is (r,k)-dispersed with constant $\varepsilon/4$;*

(A3) *For any $a, b \in M$ with $\Delta(a,b) = k$ and any $0 \leqslant i \leqslant \varepsilon k/4$, $\ell_{a,b}(\boldsymbol{x}) - \mathbb{E}\ell_{a,b}(\boldsymbol{x})$ is $72r$-subgaussian, where $\boldsymbol{x} \sim S(a, r-i)$.*

Consequently, for every $a, b \in S_n$ with $\Delta(a,b) \geqslant 6/\varepsilon$,

$$\frac{\operatorname{vol}(B(a,r) \cap B(b,r))}{\operatorname{vol}(B(a,r))} \leqslant 2e^{-\Omega_\varepsilon(1)\cdot\left(\Delta(a,b)+\Delta(a,b)^2/r\right)}. \tag{11}$$

Proof. (A1) We wish to show $\operatorname{vol}(r-t)/\operatorname{vol}(r) \leqslant 2e^{-\varepsilon t}$ for all $t < r$. As the left side is at most one, we may assume $2e^{-\varepsilon t} \leqslant 1$. It is well known that $\frac{1}{3}|I|! \leqslant |D_{|I|}| \leqslant \frac{1}{2}|I|!$ for $|I| \geqslant 2$ (where D_I is as defined in (7)). Hence,

$$\begin{aligned}\frac{\operatorname{vol}(r-t)}{\operatorname{vol}(r)} = \frac{\sum_{I \in \binom{[n]}{\leqslant r-t}} |D_I|}{\sum_{I \in \binom{[n]}{\leqslant r}} |D_I|} &\leqslant \frac{1 + \frac{1}{2}\sum_{i=2}^{r-t}\binom{n}{i} i!}{1 + \frac{1}{3}\sum_{i=2}^{r}\binom{n}{i} i!} \\ &\leqslant \frac{3}{2}\cdot\frac{3n(n-1)\cdots(n-r+t+1)}{n(n-1)\cdots(n-r+1)} \leqslant \frac{5}{t!} \leqslant 2e^{-\varepsilon t},\end{aligned}$$

where the last inequality holds as $\varepsilon \leqslant 0.01$ and $2e^{-\varepsilon t} \leqslant 1$.

(A2) Consider any $a, b \in S_n$ with $\Delta(a,b) = k$. Let $0 \leqslant i \leqslant \varepsilon k/4$, and let $\boldsymbol{x} \sim S(a, r-i)$. We can assume $a \in S_n$ is the identity permutation and b is a permutation in $D_{[k]}$. To compute the mean of $\ell_{a,b}(\boldsymbol{x})$, we generate $\boldsymbol{x}$ by first drawing $\boldsymbol{I} \sim \binom{[n]}{r-i}$ and then choosing $\boldsymbol{x} \sim D_{\boldsymbol{I}}$.

Note that for all $i \in [k]\backslash\boldsymbol{I}$ and $j \in \{k+1, \ldots, n\}\backslash\boldsymbol{I}$, we have $\boldsymbol{x}(i) = i \neq b(i)$ and $\boldsymbol{x}(j) = j = b(j)$. Hence, by the linearity of expectation, we have

$$\begin{aligned}\mathbb{E}[\Delta(\boldsymbol{x}, b)\colon \boldsymbol{I} = I] &= \sum_{i \in I}\mathbb{P}[\boldsymbol{x}(i) \neq b(i)] + |[k]\backslash I| \\ &\geqslant |I|\left(1 - \frac{(|I|-1)!}{|D_I|}\right) + |[k]\backslash I| \\ &\geqslant (r-i-3) + |[k]\backslash I|.\end{aligned}$$

Here the penultimate inequality holds as there are at most $(|I|-1)!$ permutations fixing one value, and the final inequality follows from the facts that $|D_I| \geqslant \frac{1}{3}|I|!$ and that $|I| = r-i$. From this we get

$$\begin{aligned}\mathbb{E}[\Delta(\boldsymbol{x}, b)] &= \sum_{I \in \binom{[n]}{r-i}} \mathbb{E}[\Delta(\boldsymbol{x}, b)\colon \boldsymbol{I} = I]\cdot\mathbb{P}(\boldsymbol{I} = I) \\ &\geqslant (r-i-3) + \mathbb{E}|[k]\backslash\boldsymbol{I}| \\ &= (r-i-3) + \frac{k(n-r+i)}{n}.\end{aligned}$$

As $\ell_{a,b}(\boldsymbol{x}) = \Delta(\boldsymbol{x}, b) - \Delta(\boldsymbol{x}, a) = \Delta(\boldsymbol{x}, b) - (r-i)$, we obtain

$$\begin{aligned}\mathbb{E}[\ell_{a,b}(\boldsymbol{x})] &= \mathbb{E}[\Delta(\boldsymbol{x}, b)] - (r-i) \\ &\geqslant \frac{k(n-r+i)}{n} - 3 \\ &\geqslant \varepsilon k - 3 \geqslant \varepsilon k/2,\end{aligned}$$

assuming $r \leqslant (1-\varepsilon)n$ and $k \geqslant 6/\varepsilon$.

(A3) For all $x, x' \in B(a, r-i)$, we have $|\ell_{a,b}(x) - \ell_{a,b}(x')| = |\Delta(x,b) - \Delta(x',b)| \leqslant \Delta(x,x')$. Hence Lemma 3.5 implies $\ell_{a,b}(\boldsymbol{x}) - \mathbb{E}\ell_{a,b}(\boldsymbol{x})$ is $72r$-subgaussian. □

5 Graph theoretic tools

We will reduce the lower bound on various codes to lower bound on independence number of some auxiliary graphs. We then show that all the auxiliary graphs are locally sparse. We can then use known bound on independence number of locally sparse graphs. We will use the following variant which is tailored to our needs.

Theorem 5.1. *Let G be an N-vertex with maximum degree D and minimum degree at least $D/2$. Let $K \in [1, D]$, and let $\Gamma \subseteq G$ be a subgraph induced by the neighborhood of an arbitrary vertex. Suppose there is a partition $V(\Gamma) = B \cup I$ such that*

- *every vertex $u \in B$ has degree $\deg_\Gamma(u) \leqslant D/K$; and*
- *$|I| \leqslant D/K$.*

Then the independence number of G is at least $\big(1 - o_{K\to\infty}(1)\big)\frac{N}{D}\log K$, and the number of independent sets in G is at least $\exp\big((\frac{1}{8} + o_{K\to\infty}(1))\frac{N}{D}\log^2 K\big)$.

Remark. In some of our applications we have $K = D^{\Theta(1)}$, in which case the second conclusion implies that the average size of an independent set in G is at least $\Omega(1)\cdot\frac{N}{D}\log D$.

Proof of Theorem 5.1. As $|\Gamma| \leqslant D$, we get

$$2e(\Gamma) = \sum_{v\in B}\deg_\Gamma(v) + \sum_{v\in I}\deg_\Gamma(v) \leqslant |B|\cdot(D/K) + |I|\cdot|\Gamma| \leqslant 2D^2/K.$$

Hence Γ has average degree at most $4D/K$. By a result of Hurley and Pirot [23, Theorem 2], G has chromatic number at most $(1 + o_{K\to\infty}(1))\frac{D}{\log K}$. It follows that the independence number of G is at least $\big(1 - o_{K\to\infty}(1)\big)\frac{N}{D}\log K$, as desired.

For the second statement, we need to introduce some notation. Let $\mathcal{I}(G)$ be the collection of independent sets of G. The *hard-core model on G at fugacity* $\lambda > 0$ is a probability distribution on $\mathcal{I}(G)$, where each $I \in \mathcal{I}(G)$ occurs with probability proportional to $\lambda^{|I|}$. In other words,

$$\mathbb{P}[I] = \frac{\lambda^{|I|}}{\sum_{J\in\mathcal{I}(G)}\lambda^{|J|}}.$$

The denominator, $P_G(\lambda) = \sum_{J\in\mathcal{I}(G)}\lambda^{|J|}$, is the *partition function* of the hard-core model on G. Note that $P_G(\lambda)$ is an increasing function with $P_G(0) = 1$ and $P_G(1) = |\mathcal{I}(G)|$.

The expected size of an independent set drawn from the hard-core model on G at fugacity λ is the scaled logarithmic derivative of the partition function:

$$\bar{\alpha}_G(\lambda) = \sum_{I\in\mathcal{I}(G)}|I|\cdot\mathbb{P}[I] = \frac{\sum_{I\in\mathcal{I}(G)}|I|\lambda^{|I|}}{P_G(\lambda)} = \frac{\lambda P_G'(\lambda)}{P_G(\lambda)} = \lambda\cdot(\log P_G(\lambda))'. \tag{12}$$

We need a lower bound on $\bar{\alpha}_G(\lambda)$ for certain range of λ, due to Davies et al. [11]. The lower bound is written in terms of the Lambert W function: for $z > 0$, $W(z)$ is the unique positive real satisfying $W(z)e^{W(z)} = z$. Note that $W(z) = (1 + o(1))\log z$ as $z \to \infty$.

Consider a graph G that satisfies the assumptions of Theorem 5.1. Let $\lambda_0 = \frac{\log K}{D}$ and $\lambda_1 = \frac{\sqrt{K}}{D}$. As $e(\Gamma) \leqslant D^2/K$, Theorem 5 in [11] shows that for all $\lambda \in [\lambda_0, \lambda_1]$ we have

$$\frac{1}{N}\bar{\alpha}_G(\lambda) \geqslant (1+o(1))\frac{\lambda}{1+\lambda}\frac{W(D\log(1+\lambda))}{D\log(1+\lambda)}.$$

Combining this with (12) and letting $u_i = W(D\log(1+\lambda_i))$, we find

$$\begin{aligned}\log P_G(\lambda_1) - \log P_G(\lambda_0) &\geqslant \frac{N}{D}\int_{\lambda_0}^{\lambda_1} \frac{W(D\log(1+t))}{(1+t)\log(1+t)}\,dt \\ &= \frac{N}{D}\int_{W(D\log(1+\lambda_0))}^{W(D\log(1+\lambda_1))} (1+u)\,du \\ &= \frac{N}{2D}\,[u_1^2 + 2u_1 - u_0^2 - 2u_0],\end{aligned}$$

where the first equality follows from change of variable $u = W(D\log(1+t))$. Using the approximations $D\log(1+\lambda_0) = (1+o(1))\log K$, $D\log(1+\lambda_1) = (1+o(1))\sqrt{K}$, and $W(z) = (1+o(1))\log z$, we have $u_0 = (1+o(1))\log\log K$ and $u_1 = (\frac{1}{2}+o(1))\log K$. Therefore, we get

$$\log P_G(\lambda_1) - \log P_G(\lambda_0) \geqslant \Big(\frac{1}{8}+o(1)\Big)\frac{N}{D}\log^2 K.$$

Since $1 \leqslant P_G(\lambda_0) \leqslant P_G(\lambda_1) \leqslant |\mathcal{I}(G)|$, this gives $\log|\mathcal{I}(G)| \geqslant (\frac{1}{8}+o(1))\frac{N}{D}\log^2 K$, as desired. □

6 Improvement on Gilbert–Varshamov bounds

We present in this section a unified short proofs of improvements on sphere-covering bounds on various codes by reducing it to lower bound on independence number of an auxiliary graph. In order to use Theorem 5.1, we need to show that the graph is locally sparse. Our strategy is to split the edge count in the subgraph induced by the neighbourhood of a vertex into two parts, one from vertices from the boundary of the Hamming/Johnson/Euclidean ball, and the other from interior vertices of the ball. The contribution from boundary vertices is exponentially small because the volume of the intersection of balls that are far apart is small as we have shown using Theorem 1.1 and concentration of measure. On the other hand, the contribution from the interior vertices is also small as there are negligible amount of interior vertices using the growth of the balls in such spaces.

Proof of Theorem 2.4. Define a graph G whose vertices are points in the metric space (X, d) and two points are adjacent if their distance is at most r. It is easy to see that G has $|X|$ vertices, the degree of every vertex is $\mathrm{vol}(r) - 1$, and the maximum size of an (X, d, r)-code is the independence number $\alpha(G)$ of G. Let Γ be a subgraph induced by the neighborhood of an arbitrary vertex $x \in X$. We partition $V(\Gamma) = B \cup I$, where I is the punctured ball of radius $r - t$ centered at x. By the assumption, $\frac{|I|}{\mathrm{vol}(r)} = \frac{\mathrm{vol}(r-t)-1}{\mathrm{vol}(r)} \leqslant e^{-K}$. Consider any vertex $u \in B$. As $r - t < \mathsf{d}(x,u) \leqslant r$, we obtain

$$\frac{\deg_\Gamma(u)}{\mathrm{vol}(r)} = \frac{\mathrm{vol}(B(x,r) \cap B(u,r))}{\mathrm{vol}(r)} \leqslant e^{-K}.$$

Therefore, Theorem 2.4 is a realization of Theorem 5.1. □

Proof of Theorems 2.5 to 2.7. Each of Lemmas 4.2 to 4.4 verifies the conditions for each of q-ary codes, constant-weight codes and permutation codes for applying Theorem 2.4, respectively. Hence, Theorems 2.5 to 2.7 all follow from Theorem 2.4. □

6.1 Spherical codes

We need two lemmas for the short proof of Theorem 2.8. The first one is a folklore result that partitions the sphere into small pieces of equal measure (see e.g. [15, Lemma 21]).

Lemma 6.1. *For each $\delta \in (0,1)$ the sphere $\mathbb{S}^{n-1}$ can be partitioned into $N = (O(1)/\delta)^n$ pieces of equal measure, each of diameter at most δ.*

The second one is an Euclidean version of results from Section 4. For a measurable set $A \subset \mathbb{S}^{n-1}$, let $s(A)$ denote the normalized surface area of A. Recall that $s_n(\theta)$ is the normalized surface area of a spherical cap of angular radius θ. It is well known that for fixed angle $\theta \in (0, \pi/2)$

$$s_n(\theta) = \frac{1+o(1)}{\sqrt{2\pi n}} \cdot \frac{\sin^{n-1}\theta}{\cos\theta}. \tag{13}$$

We need a parameter q_θ, which is the angular radius of the smallest cap containing the intersection of two spherical caps of angular radius θ whose centers are at angle θ. It is straightforward to compute that

$$q_\theta = \arcsin\Big(\frac{\sqrt{(\cos\theta-1)^2(1+2\cos\theta)}}{\sin\theta}\Big). \tag{14}$$

Lemma 6.2 ([24, Lemma 6])**.** *Let $x \in \mathbb{S}^{n-1}$ and $A \subset C_\theta(x)$ be measurable with $s(A) > 0$. Then*

$$\mathop{\mathbb{E}}_{\boldsymbol{u}\sim A}[s(C_\theta(\boldsymbol{u}) \cap A)] \leqslant 2 \cdot s_n(q_\theta),$$

where q_θ is as in (14).

Proof of Theorem 2.8. Choose $\delta \ll_{\theta,n} 1$, that is, δ is less than a suitable function of θ and n. Apply Lemma 6.1 to partition the unit sphere into $N = (O(1)/\delta)^n$ pieces $P_1, \ldots, P_N$ of equal measure, each with diameter at most δ. For each $i \in [N]$, pick an arbitrary point v_i from P_i. Let G be a graph with vertex set being these N chosen points, and two vertices form an edge if the angle between them is less than θ. Then by definition, $A(n,\theta) \geqslant \alpha(G)$. We first use a packing/covering argument to show that every vertex in G has degree $(1+o(1))s_n(\theta)N$. Write $N[x] := N(x) \cup \{x\}$ for the closed neighborhood of x.

Claim 6.3. *For every $x \in V(G)$,*

$$C_{\theta-2\delta}(x) \subset \bigcup_{\boldsymbol{v}_i \in N[x]} P_i \subset C_{\theta+2\delta}(x).$$

Proof of claim. We only prove the first inclusion. Let y be any point in $C_{\theta-2\delta}(x)$, that is, the angle between y and x is at most $\theta - 2\delta$. As the P_i's cover the sphere, there exists an index i such that $y \in P_i$. By the assumption on P_i, we have $||y - v_i|| \leqslant \delta$. Thus, the angle between v_i and y is $2\arcsin(||y-v_i||/2) \leqslant 2\arcsin(\delta/2) < 2\delta$. It follows from the triangle inequality that the angle between v_i and x is less than $2\delta + (\theta - 2\delta) = \theta$, implying $v_i \in N[x]$. Therefore, for every $y \in C_{\theta-2\delta}(x)$ we must have $y \in \bigcup_{v_i \in N[x]} P_i$, as desired. ∎

Let x be an arbitrary vertex of G. Since the P_i's are disjoint subsets of $\mathbb{S}^{n-1}$ of normalized surface area $1/N$, Claim 6.3 gives $s_n(\theta-2\delta)N \leqslant |N[x]| \leqslant s_n(\theta+2\delta)N$. Moreover, it follows from (13) that $s_n(\theta \pm 2\delta) = (1+O(\delta))^n s_n(\theta) = (1+o(1))s_n(\theta)$. Therefore, every vertex in G has degree $D := (1+o(1))s_n(\theta)N$.

Let $K = \frac{s_n(\theta)}{4s_n(q_\theta)}$. By (13), we obtain $\log K = (1+o(1))\log\frac{\sin\theta}{\sin q_\theta} \cdot n = (1+o(1))c_\theta \cdot n$. It suffices to show that we can apply Theorem 5.1 with this choice of K. This amounts to proving that for any

$x \in V(G)$, the average degree of $G[N(x)]$ is at most D/K. For this, we view the average degree of $G[N(x)]$ probabilistically as the expected degree of a uniform random vertex in $N(x)$.

We partition $N[x] = B \cup I$, where $I = \{v_i \colon P_i \subset C_\theta(x)\}$. From Claim 6.3, we know that $\bigcup_{v_i \in B} P_i$ is contained in $C_{\theta+2\delta}(x) \backslash C_{\theta-2\delta}(x)$. Thus, $\delta \ll_{\theta,n} 1$, the number of boundary point is

$$|B| \leqslant \big(s_n(\theta + 2\delta) - s_n(\theta - 2\delta)\big)N = O(\delta n)s_n(\theta)N = o(D/K),$$

which is negligible. So it suffices to estimate the average degree of $G[I]$.

Let $A = \bigcup_{v_i \in I} P_i$, and let $\boldsymbol{u}$ be a uniform random point in A. Now, as each vertex in G corresponds to a piece of the sphere with the same measure, we can generate $\boldsymbol{v}_i \sim I$ by rounding $\boldsymbol{u}$ to the vertex $\boldsymbol{v_i}$ such that $\boldsymbol{u} \in P_i$. Thus, we have by Lemma 6.2 that

$$\mathop{\mathbb{E}}_{\boldsymbol{v}_i \sim I}[\deg_{G[I]}(\boldsymbol{v}_i)] = \mathop{\mathbb{E}}_{\boldsymbol{u} \sim A}[s(C_\theta(\boldsymbol{u}) \cap A)] \cdot N \leqslant 2s_n(q_\theta)N \leqslant D/K,$$

as desired. □

7 List-decodability of random codes

In this section, we prove Theorem 2.3, which states that a uniformly chosen random code of rate $1 - h_q(p) - \varepsilon$ is with high probability *not* $(p, (1 - o(1))/\varepsilon)$-list decodable. In large part we follow the proof of Guruswami and Narayanan [20, Theorem 20]. As in [20] we define a random variable $\boldsymbol{W}$ that counts the number of witnesses that certify the violation of the (p, L)-list decodability property. Thus the code is (p, L)-list decodable if and only if $\boldsymbol{W} = 0$. So our job becomes to bound the probability of the event that $\boldsymbol{W} = 0$. For this we employ the Chebyshev's inequality

$$\mathbb{P}(\boldsymbol{W} = 0) \leqslant \frac{\operatorname{Var}[\boldsymbol{W}]}{\mathbb{E}[\boldsymbol{W}]^2}.$$

We then show that $\operatorname{Var}[\boldsymbol{W}]/\mathbb{E}[\boldsymbol{W}]^2$ is exponentially small, which would finish the proof. To bound the variance, we introduce a new ingredient (Lemma 7.1), whose proof relies crucially on our bound on intersection volume from Lemma 4.2.

Notation. For the rest of this section, we shall employ the following notation. Given $a \in [q]^n$ and $r \in \mathbb{N}$, we write $B_q(a, r)$ for the Hamming ball of radius r centered at a. Recall that $\operatorname{vol}_q(n, r)$ is the volume of a radius-r Hamming ball in $[q]^n$, and $\operatorname{vol}_q(n, r; k)$ stands for the volume of the intersections of two radius-r balls whose centers are distance k apart.

Lemma 7.1. *Let $0 < p < 1 - 1/q$, $1 \leqslant \ell \leqslant L$ and $\mu := q^{-n}\operatorname{vol}_q(n, pn)$. There exists a constant $c = c_{p,q} > 0$ such that the following holds. Let*

$$\boldsymbol{a}, \boldsymbol{b},\ \ \boldsymbol{x}_1, \ldots, \boldsymbol{x}_\ell,\ \ \boldsymbol{y}_{\ell+1}, \ldots, \boldsymbol{y}_L,\ \ \boldsymbol{z}_{\ell+1}, \ldots, \boldsymbol{z}_L$$

be chosen independently and uniformly at random from $[q]^n$. Denote by $\mathcal{E}_\ell$ the event

$$\Big\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_\ell \in B_q(\boldsymbol{a}, pn) \cap B_q(\boldsymbol{b}, pn),\ \ \boldsymbol{y}_{\ell+1}, \ldots, \boldsymbol{y}_L \in B_q(\boldsymbol{a}, pn),\ \ \boldsymbol{z}_{\ell+1}, \ldots, \boldsymbol{z}_L \in B_q(\boldsymbol{b}, pn)\Big\}.$$

Then

$$\mathbb{P}(\mathcal{E}_\ell) \leqslant \min\Big\{\mu^{2L-\ell+1}, q^{-n}\mu^{2L-\ell}\left(1 + 2(q-1)q^{-c\ell}\right)^n\Big\}.$$

Remark. A version of Lemma 7.1, for the case $q = 2$, appeared as [29, Lemma A.5]. The proof of [29, Lemma A.5], however, does not extend to larger q.

Proof of Lemma 7.1. We first show that the probability of $\mathcal{E}_\ell$ is at most $\mu^{2L-\ell+1}$. For the event $\mathcal{E}_\ell$ to occur, one must have (i) $\boldsymbol{a}, \boldsymbol{b} \in B_q(\boldsymbol{x}_1, pn)$, (ii) $\boldsymbol{x}_2, \ldots, \boldsymbol{x}_\ell, \boldsymbol{y}_{\ell+1}, \ldots, \boldsymbol{y}_L \in B_q(\boldsymbol{a}, pn)$, and (iii) $\boldsymbol{z}_{\ell+1}, \ldots, \boldsymbol{z}_L \in B_q(\boldsymbol{b}, pn)$. Note that the events (i), (ii), (iii) are independent. Conditioned on the position of $\boldsymbol{x}_1$, (i) occurs with probability μ^2. Given $\boldsymbol{a}$ and $\boldsymbol{b}$, (ii) and (iii) happen with probability μ^{L-1} and $\mu^{L-\ell}$, respectively. It follows that $\mathbb{P}(\mathcal{E}_\ell) \leqslant \mu^{2L-\ell+1}$.

For the other bound, we first apply the law of total probability to get

$$\mathbb{P}(\mathcal{E}_\ell) = \sum_{k=0}^{n} \mathbb{P}(\Delta(\boldsymbol{a}, \boldsymbol{b}) = k) \cdot \mathbb{P}(\mathcal{E}_\ell \big| \Delta(\boldsymbol{a}, \boldsymbol{b}) = k).$$

Since there are $\binom{n}{k}(q-1)^k$ codewords $b \in [q]^n$ which are at distance k from $a \in [q]^n$, the probability that $\Delta(\boldsymbol{a}, \boldsymbol{b}) = k$ is exactly $q^{-n}\binom{n}{k}(q-1)^k$. Conditioned on the positions of $\boldsymbol{a}$ and $\boldsymbol{b}$ being distance k apart, the probability that $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_\ell \in B_q(\boldsymbol{a}, pn) \cap B_q(\boldsymbol{b}, pn)$ is $\left(\frac{\mathrm{vol}_q(n,pn;k)}{q^n}\right)^\ell = \left(\frac{\mathrm{vol}_q(n,pn;k)}{\mathrm{vol}_q(n,pn)}\right)^\ell \mu^\ell$. The probability that $\boldsymbol{y}_{\ell+1}, \ldots, \boldsymbol{y}_L \in B_q(\boldsymbol{a}, pn)$ is $\mu^{L-\ell}$, and the probability that $\boldsymbol{z}_{\ell+1}, \ldots, \boldsymbol{z}_L \in B_q(\boldsymbol{a}, pn)$ is $\mu^{L-\ell}$. Thus, we have

$$\begin{aligned} \mathbb{P}(\mathcal{E}_\ell \big| \Delta(\boldsymbol{a}, \boldsymbol{b}) = k) &= \left(\frac{\mathrm{vol}_q(n, pn; k)}{\mathrm{vol}_q(n, pn)}\right)^\ell \mu^\ell \cdot \mu^{L-\ell} \cdot \mu^{L-\ell} \\ &= \left(\frac{\mathrm{vol}_q(n, pn; k)}{\mathrm{vol}_q(n, pn)}\right)^\ell \mu^{2L-\ell}. \end{aligned}$$

Therefore, we get the following for some $c = c_{p,q}$ as in Lemma 4.2.

$$\begin{aligned} \mathbb{P}(\mathcal{E}_\ell) &= \sum_{k=0}^{n} q^{-n}\binom{n}{k}(q-1)^k \cdot \left(\frac{\mathrm{vol}_q(n, pn; k)}{\mathrm{vol}_q(n, pn)}\right)^\ell \mu^{2L-\ell} \\ \text{(by Lemma 4.2)} &\leqslant q^{-n}\mu^{2L-\ell} \sum_{k=0}^{n} \binom{n}{k}(q-1)^k \cdot (2q^{-ck})^\ell \\ &= q^{-n}\mu^{2L-\ell}\left(1 + 2(q-1)q^{-c\ell}\right)^n, \end{aligned}$$

as desired. ☐

We are now ready to prove Theorem 2.3.

Proof of Theorem 2.3. Let c be the positive constant given by (8). Let

$$\mu := q^{-n}\,\mathrm{vol}_q(n, pn), \quad \ell_0 := \frac{1 - h_q(p)}{2\varepsilon}, \quad \gamma := \frac{4(q-1)}{\ln q} \cdot q^{-c\ell_0}, \quad \text{and} \quad L = \frac{1-\gamma}{\varepsilon}.$$

From Lemma 4.1, and recalling that $R = 1 - h_q(p) - \varepsilon$, we get

$$\mu = \frac{\Theta(1)}{\sqrt{n}} \cdot q^{-(1-h_q(p))n} \quad \text{and} \quad q^{Rn}\mu = \frac{\Theta(1)}{\sqrt{n}} \cdot q^{-\varepsilon n}. \tag{15}$$

Notice that a random q-ary code of rate R is simply a random map $\boldsymbol{\mathcal{C}} \colon [q]^{Rn} \to [q]^n$ where, for each $x \in [q]^{Rn}$, its image $\boldsymbol{\mathcal{C}}(x)$ is chosen independently and uniformly at random from $[q]^n$. For any center $a \in [q]^n$ and any ordered list of L distinct messages $X = (x_1, \ldots, x_L) \in ([q]^{Rn})^L$, we define $\mathbb{I}(a, X)$ to be the indicator random variable for the event that $\boldsymbol{\mathcal{C}}(x_1), \ldots, \boldsymbol{\mathcal{C}}(x_L)$ all fall in $B_q(a, pn)$, and let $\boldsymbol{W} = \sum_{a,X} \mathbb{I}(a, X)$. Then $\boldsymbol{\mathcal{C}}$ is $(p, L-1)$-list decodable if and only if $\boldsymbol{W} = 0$.

We have $\mathbb{E}[\mathbb{I}(a,X)] = \mathbb{P}\{\boldsymbol{\mathcal{C}}(x_1),\ldots,\boldsymbol{\mathcal{C}}(x_L) \in B_q(a,pn)\} = \mu^L$ and the number of pairs (a,X) is $q^n \cdot \prod_{i=0}^{L-1}(q^{Rn}-i) \geqslant q^n \cdot \frac{1}{2}q^{RnL}$. Thus, by linearity of expectation,

$$\mathbb{E}[\boldsymbol{W}] \geqslant \frac{1}{2}\mu^L q^{RnL+n}. \tag{16}$$

Observe that if X and Y are two disjoint lists (viewed as sets), then the events $\mathbb{I}(a,X)$ and $\mathbb{I}(b,Y)$ are independent for any pair of centers a,b. It follows that

$$\begin{aligned}
\operatorname{Var}[\boldsymbol{W}] &= \sum_{X\cap Y\neq\varnothing}\sum_{a,b}\Big(\mathbb{E}[\mathbb{I}(a,X)\mathbb{I}(b,Y)] - \mathbb{E}[\mathbb{I}(a,X)]\cdot\mathbb{E}[\mathbb{I}(b,Y)]\Big)\\
&\leqslant \sum_{X\cap Y\neq\varnothing}\sum_{a,b}\mathbb{E}[\mathbb{I}(a,X)\mathbb{I}(b,Y)]\\
&= \sum_{\ell=1}^{L}\sum_{|X\cap Y|=\ell}\sum_{a,b}\mathbb{P}\{\mathbb{I}(a,X)=1 \text{ and } \mathbb{I}(b,Y)=1\}\\
&= \sum_{\ell=1}^{L}\sum_{|X\cap Y|=\ell} q^{2n}\cdot\mathbb{P}_{\boldsymbol{a},\boldsymbol{b},\boldsymbol{\mathcal{C}}}\{\mathbb{I}(\boldsymbol{a},X)=1 \text{ and } \mathbb{I}(\boldsymbol{b},Y)=1\},
\end{aligned}$$

where in the last equality we converted the inner summation into an expectation by randomizing over the centers a and b.

Fix a pair (X,Y) with $|X\cap Y| = \ell$, and suppose that the elements of $\boldsymbol{\mathcal{C}}(X)$ are $\boldsymbol{x}_1,\ldots,\boldsymbol{x}_\ell,\boldsymbol{y}_{\ell+1},\ldots,\boldsymbol{y}_L$ while the elements of $\boldsymbol{\mathcal{C}}(Y)$ are $\boldsymbol{x}_1,\ldots,\boldsymbol{x}_\ell,\boldsymbol{z}_{\ell+1},\ldots,\boldsymbol{z}_L$. Then the event $\{\mathbb{I}(\boldsymbol{a},X)=1 \text{ and } \mathbb{I}(\boldsymbol{b},Y)=1\}$ is exactly the event $\mathcal{E}_\ell$ in Lemma 7.1. Thus, we can bound the variance of $\boldsymbol{W}$ as

$$\begin{aligned}
\operatorname{Var}[\boldsymbol{W}] &\leqslant \sum_{\ell=1}^{L}\sum_{|X\cap Y|=\ell} q^{2n}\cdot\mathbb{P}(\mathcal{E}_\ell)\\
&\leqslant \sum_{\ell=1}^{L} L^{2L}q^{Rn(2L-\ell)+2n}\cdot\mathbb{P}(\mathcal{E}_\ell),
\end{aligned}$$

where the second inequality stems from the fact that the number of pairs (X,Y) with $|X\cap Y| = \ell$ is at most $L^{2L}q^{Rn(2L-\ell)}$. We split the summation into $\ell \leqslant \ell_0$ and $\ell > \ell_0$, and get $\operatorname{Var}[\boldsymbol{W}] \leqslant V_{\leqslant\ell_0} + V_{>\ell_0}$. From Lemma 7.1 and (16), we find

$$\begin{aligned}
\frac{V_{\leqslant\ell_0}}{\mathbb{E}[\boldsymbol{W}]^2} &\leqslant \frac{4}{\mu^{2L}q^{2RnL+2n}}\sum_{\ell=1}^{\ell_0} L^{2L}q^{Rn(2L-\ell)+2n}\cdot\mu^{2L-\ell+1}\\
&= 4L^{2L}\sum_{\ell=1}^{\ell_0}(q^{Rn}\mu)^{-\ell}\cdot\mu\\
(\text{by } (15)) &= \Theta(1)\cdot\left(\sqrt{n}q^{\varepsilon n}\right)^{\ell_0}\cdot\frac{\Theta(1)}{\sqrt{n}}q^{-(1-h_q(p))n}\\
(\text{as } \ell_0 = \frac{1-h_q(p)}{2\varepsilon}) &= q^{-\Omega(n)}.
\end{aligned}$$

Again by appealing to Lemma 7.1 and (16), we see that

$$\begin{aligned}
\frac{V_{>\ell_0}}{\mathbb{E}[\boldsymbol{W}]^2} &\leqslant \frac{4}{\mu^{2L} q^{2RnL+2n}} \sum_{\ell_0<\ell\leqslant L} L^{2L} q^{Rn(2L-\ell)+2n} \cdot q^{-n} \mu^{2L-\ell} \left(1+2(q-1)q^{-c\ell}\right)^n \\
&= 4L^{2L} \sum_{\ell_0<\ell\leqslant L} (q^{Rn}\mu)^{-\ell} \cdot \left(\frac{1+2(q-1)q^{-c\ell}}{q}\right)^n \\
(\text{by the choice of } \gamma) &\leqslant 4L^{2L} \sum_{\ell_0<\ell\leqslant L} (q^{Rn}\mu)^{-\ell} \cdot q^{-(1-\gamma/2)n} \\
(\text{by (15)}) &= \Theta(1) \cdot \left(\sqrt{n} q^{\varepsilon n}\right)^L \cdot q^{-(1-\gamma/2)n} \\
(\text{since } L = \frac{1-\gamma}{\varepsilon}) &= q^{-\Omega(n)}.
\end{aligned}$$

Putting everything together, we get from Chebyshev's inequality that

$$\mathbb{P}(\boldsymbol{W}=0) \leqslant \frac{\operatorname{Var}[\boldsymbol{W}]}{\mathbb{E}[\boldsymbol{W}]^2} \leqslant \frac{V_{\leqslant\ell_0}+V_{>\ell_0}}{\mathbb{E}[\boldsymbol{W}]^2} \leqslant q^{-\Omega(n)}.$$

Since $\mathcal{C}$ is $(p, L-1)$-list decodable if and only if $\boldsymbol{W}=0$, we conclude that $\boldsymbol{\mathcal{C}}$ is with probability $1-q^{-\Omega(n)}$ not $(p, L-1)$-list decodable. □

Acknowledgement

We would like to thank Benny Sudakov for bringing [26] to our attention.

References

[1] I.F. Blake, Permutation codes for discrete channels. *IEEE Trans. Inform. Theory* **20** (1974), 138–140.

[2] I.F. Blake, G. Cohen and M. Deza, Coding with permutations. *Inf. Control* **43** (1979), 1–19.

[3] S.G. Bobkov, Concentration of normalized sums and a central limit theorem for noncorrelated random variables. *Ann. Probab.* **32** (2004), 2884–2907.

[4] S. Bobkov and M. Ledoux, Poincaré's inequalities and Talagrand's concentration phenomenon for the exponential distribution. *Probab. Theory Related Fields* **107** (1997), 384–400.

[5] A.E. Brouwer, Bounds for binary constant weight codes. `https://www.win.tue.nl/~aeb/codes/Andw.html`.

[6] C. Chabauty, Résultats sur l'empilement de calottes égales sur une périsphère de $\mathbb{R}^n$ et correction à un travail antérieur. *Comptes Rendus* **236** (1953), 1462–1464.

[7] Y.M. Chee and S. Ling, Constructions for q-ary constant-weight codes. *IEEE Trans. Inform. Theory* **53** (2007), 135–146.

[8] W. Chu, C.J. Colbourn and P. Dukes, Constructions for permutation codes in powerline communications. *Des. Codes Cryptogr.* **32** (2004), 51–64.

[9] C.J. Colbourn, T. Kløve and A.C.H. Ling, Permutation arrays for powerline communications and mutually orthogonal Latin squares. *IEEE Trans. Inform. Theory* **50** (2004), 1289–1291.

[10] C.J. Colbourn, A.C.H. Ling and D.R. de la Torre, An application of permutation arrays to block ciphers. *Proc. Southeastern International Conference on Combinatorics, Graph theory and Computing* **145** (2000), 5–7.

[11] E. Davies, R. Joannis de Verclos, R.J. Kang and F. Pirot, Occupancy fraction, fractional colouring, and triangle fraction. *J. Graph Theory* **97** (2021), 557– 568.

[12] M. Deza and P. Frankl, On the maximum number of permutations with given maximal or minimal distance. *J. Combin. Theory Ser. A* **22** (1977), 352–360.

[13] M. Deza and S.A. Vanstone, Bounds for permutation arrays *J. Statist. Plann. Inference* **2** (1978), 197–209.

[14] P. Elias, List decoding for noisy channels. *Technical Report 335, Research Laboratory of Electronics, MIT*, 1957.

[15] U. Feige and G. Schechtman, On the optimality of the random hyperplane rounding technique for MAX CUT. *Ramdom Structures Algorithms* **20** (2002), 403–440.

[16] H.C. Ferreira and A.J.H. Vinck, Inference cancellation with permutation trellis arrays. *Proc. IEEE Vehicular Technology Conf.* (2000), 2401-2407.

[17] A. Frieze and M. Karoński, *Introduction to Random Graphs*, Cambridge University Press, 2015.

[18] I. Gil Fernández, J. Kim, H. Liu and O. Pikhurko, New lower bounds on kissing numbers and spherical codes in high dimensions. arXiv preprint arXiv:2111.01255.

[19] E.N. Gilbert, A comparison of signalling alphabets. *Bell System Tech. J.* **31** (1952), 504–522.

[20] V. Guruswami and S. Narayanan, Combinatorial limitations of average-radius list-decoding. *IEEE Trans. Inform. Theory* **60** (2014), 5827–5842.

[21] V. Guruswami, A. Rudra and M. Sudan, *Essential coding theory.* Draft available at `https://cse.buffalo.edu/faculty/atri/courses/coding-theory/book/web-coding-book.pdf`.

[22] W. Cary Huffman and V. Pless, *Fundamentals of Error-Correcting Codes*, Cambridge University Press, 2010.

[23] E. Hurley and F. Pirot, *A first moment proof of the Johansson-Molloy theorem.* arXiv preprint arXiv:2109.15215.

[24] M. Jenssen, F. Joos and W. Perkins, On kissing numbers and spherical codes in high dimensions. *Adv. Math.* **335** (2018), 307–321.

[25] T. Jiang and A. Vardy, Asymptotic improvement of the Gilbert–Varshamov bound on the size of binary codes. *IEEE Trans. Inform. Theory* **50** (2004), 1655–1664.

[26] M. Krivelevich, S. Litsyn and A. Vardy, A lower bound on the density of sphere packings via graph theory. *Int. Math. Res. Not.* **43** (2004), 2271–2279.

[27] M. Kwan, B. Sudakov and T. Tran, Anticoncentration for subgraph statistics. *J. Lond. Math. Soc.* **99** (2019), 757–777.

[28] V.I. Levenshtein, Upper-bound estimates for fixed-weight codes. *Problemy Peredachi Informatsii* **7** (1971), 3–12.

[29] R. Li and M. Wootters, Improved list-decodability of random linear binary codes. *IEEE Trans. Inform. Theory* **67** (2021), 1522–1536.

[30] B. Maurey, Some deviation inequalities. *Geom. Funct. Anal.* **1** (1991), 188–197.

[31] X. Niu, C. Xing and C. Yuan, Asymptotic Gilbert–Varshamov Bound on Frequency Hopping Sequences. *IEEE Trans. Inform. Theory* **66** (2020), 1213–1218.

[32] N. Pavlidou, A.J.H. Vinck, J. Yazdani and B. Honary, Power line communications: State of the art and future trends. *IEEE Commun. Mag.* **41** (2003), 34–40.

[33] R.A. Rankin, The closest packing of spherical caps in n dimensions. *Glasg. Math. J.* **2** (1955), 139–144.

[34] A. Rudra, Limits to list decoding of random codes. *IEEE Trans. Inform. Theory* **57** (2011), 1398–1408.

[35] C.E. Shannon, Probability of error for optimal codes in a Gaussian channel. *Bell System Tech. J.* **38** (1959), 611–656.

[36] D. Slepian, Permutation modulation. *Proc. IEEE* **53** (1965), 228–236.

[37] M. Tait, A. Vardy and J. Verstraëte, Asymptotic Improvement of the Gilbert-Varshamov Bound on the Size of Permutation Codes. arXiv preprint arXiv:1311.4925.

[38] M. Talagrand, A new isoperimetric inequality and the concentration of measure phenomenon. *Geometric Aspects of Functional Analysis* (1989–90) Lecture Notes in Math., Springer, **1469** (1991), 94–124.

[39] R.R. Varshamov, Estimate of the number of signals in error correcting codes. *Doklady Akademii Nauk* **117** (1957), 739–741.

[40] J.M. Wozencraft, List Decoding. *Quaterly Progress Report, Research Laboratory of Electronics, MIT*, **48** (1958), 90–95.

[41] R. Vershynin, High-dimensional probability: An introduction with applications in data science. Vol. 47. Cambridge university press, 2018.

[42] A.J.H. Vinck, Coded modulation for powerline communications. *A.E. Ü. Int. J. Electron. Commun.* **54** (2005), 3200–3208.

[43] V. Vu, L. Wu, Improving the Gilbert–Varshamov bound for q-ary codes. *IEEE Trans. Inform. Theory* **51** (2005), 3200–3208.

[44] X. Wang, Y. Zhang, Y. Yang and G. Ge, New bounds of permutation codes under Hamming metric and Kendall's τ-metric. *Des. Codes Cryptogr.* **85** (2017) 533–545.

[45] N.C. Wormald, Models of random regular graphs. In Surveys in combinatorics, 1999 (Canterbury), volume 267 of London Math. Soc. Lecture Note Ser., pages 239–298. Cambridge Univ. Press, Cambridge, 1999.

[46] A.D. Wyner, Capabilities of bounded discrepancy decoding. *Bell System Tech. J.* **44** (1965), 1061–1122.

[47] L. Yang, K. Chen and L. Yuan, New lower bounds on sizes of permutation arrays. arXiv preprint arXiv:0801.3986.

[48] V.V. Zyablov and M.S. Pinsker, List concatenated decoding. *Problemy Peredachi Informatsii* **17** (1981), 29–33.