

List Decoding, Linear Hashing, and Furstenberg over $\mathbb{F}_q$

Vinayak M. Kumar*

Geoffrey Mon†

Abstract

We give new bounds for list sizes of random linear codes at capacity, max loads of linear hash functions, and Furstenberg sets, over every finite field $\mathbb{F}_q$.

1. **Random linear codes** over $\mathbb{F}_q$ with rate $1 - H_q(p) - \varepsilon$ are $(p, O(qH_q(p)/\varepsilon))$ -list decodable with high probability for all values of p, q, ε , including the high error regime. This nearly matches the list size lower bound of $H_q(p)/\varepsilon$ due to Guruswami, Li, Mosheiff, Resch, Silas, and Wootters [IEEE Trans. Inf. Theory 2022]. Our bound is the first uniform improvement for $q > 2$ since Guruswami, Håstad, and Kopparty [STOC 2010].
2. **Linear hash functions** over $\mathbb{F}_q$ hashing n balls to n bins achieve maximum load $O(q \ln \ln q / \ln q) \cdot \ln n / \ln \ln n$, both in expectation and with probability $1 - o(1)$. This nearly matches the lower bound of $\ln n / \ln \ln n$. Previously, only a polylogarithmic upper bound was known for $q > 2$, due to Alon, Dietzfelbinger, Miltersen, Petrank, and Tardos [J. ACM 1999].

We reduce list decodability and linear hashing to strong Furstenberg set lower bounds, which we prove using a new polynomial method of multiplicity gaps. While previous polynomial methods analyze a set S by studying polynomials that vanish on it, we consider polynomials that vanish everywhere, but with higher multiplicity inside S than outside.

1 Introduction

Subspaces are a fundamental primitive in mathematics and theoretical computer science. In this paper, we make progress on three intertwined questions on subspaces: list decodability of random linear codes, expected max load of linear hashing, and lower bounds on Furstenberg sets.

*Simons Institute for the Theory of Computing, UC Berkeley. vmkumar@berkeley.edu. Supported by a Simons Institute Research Fellowship. Part of this work was conducted while the author was a graduate student at UT Austin, supported by a Jane Street Graduate Research Fellowship and a UT Austin Dean's Prestigious Fellowship Supplement.

†Department of Computer Science, University of Texas at Austin. gmon@cs.utexas.edu. Supported by NSF Award CCF-2505865, an NSF Graduate Research Fellowship (DGE-2137420), and a UT Austin Dean's Prestigious Fellowship Supplement. Part of this work was conducted while the author was visiting the Simons Institute for the Theory of Computing.

1.1 List Decodability of Random Linear Codes

An error-correcting code (or simply code) over a finite field $\mathbb{F}_q$ is a subset $C \subseteq \mathbb{F}_q^n$ of well-separated strings (known as codewords). The separatedness of C is measured by its *distance* d , which is the minimum Hamming distance between two distinct elements of C . In an adversarial (Hamming) error model, where a transmitted codeword could be corrupted on any $< d/2$ coordinates, unique decoding is always information-theoretically possible because the closest codeword is unique. However, it is easy to see that this promise breaks when we allow at least $d/2$ corruptions to take place.

A revolutionary idea in coding theory is the notion of *list decoding*: instead of hoping for a unique closest codeword to a noisy input, we are satisfied if the number of codewords close to a noisy string is small. This weaker guarantee has nonetheless found many applications in both theory and practice, and has spawned fruitful lines of research. One such program is to find families of list decodable codes that achieve the best possible parameters.

On the one hand, we want to maximize the *rate*, $R := \log_q |C|/n$, which measures how much information each codeword contains, because we want to minimize the code's redundancy (i.e., inefficiency). On the other hand, we want to minimize the *list size*, which is the largest number of codewords that are close (say, within distance pn) to a single string in $\mathbb{F}_q^n$. These two goals are in conflict with each other, because we can always maximize rate by blowing up list size and vice versa. When the list size is at most L , we say C is (p, L) -*list decodable*.

List decoding capacity refers to a particular rate threshold $1 - H_q(p)$ below which (for sufficiently large n) list decoding with constant list size is possible, and above which constant list size is impossible. A code *achieves capacity* if it has rate close to this threshold and constant list size, and we are interested in which families of codes do so. Through a successful series of works, we now know that one such family is random linear codes.

Random linear codes. Random linear codes (RLCs) are one of the most studied classes of practical codes. Linear codes have efficient encoders and representations, as well as useful properties such as symmetry. These properties make random linear codes helpful, for example as a building block for constructing good codes via concatenation. Random linear codes are also intimately related to notions in cryptography such as LWE and LPN. Hence, it is of great interest to understand the properties of random linear codes. Indeed, there has been a long line of work starting in the 1950s studying properties of RLCs, and how close they are in strength to fully random codes.

Let us consider the context of list decoding capacity in detail. In general, if a code has rate near capacity, i.e., $1 - H_q(p) - \varepsilon$, then its list size will grow as the *capacity gap* ε shrinks. Elias [Eli91] showed that a fully random code with rate gap ε from capacity will be $(p, H_q(p)/\varepsilon)$ -list decodable with high probability, for all choices of parameters q, p, ε . Guruswami, Li, Mosheiff, Resch, Silas, and Wootters [GLM⁺22] also show that an RLC cannot outperform a fully random code, because its list size must be at least $H_q(p)/\varepsilon$. Thus, researchers have been pushing towards showing RLCs have list decodability which is just as strong as fully random codes.

List size bounds. In the binary case ($q = 2$), Guruswami, Håstad, Sudan, and Zuckerman [GHSZ02] showed the existence (improved by [LW18] to high probability) of linear codes with capacity gap ε which are $(p, O(1/\varepsilon))$ -list decodable using potential functions. This method was developed further by Li and Wootters [LW18] to show random linear codes with rate gap ε are

Table 1: Uniform list decodability bounds for $q \geq 3$ and capacity gap ε ($\delta := 1 - 1/q - p$)

Source	List size bound
<i>Fully random codes</i>	
[Eli91]	$\sim H_q(p)/\varepsilon$
<i>Random linear codes</i>	
[GLM ⁺ 22]	$\geq H_q(p)/\varepsilon$
[ZP81]	$\leq q^{O(1/\varepsilon)}$
[GHK10]	$\leq \exp(O(\frac{\log_2^2 q}{\min\{p, \delta^2\}}))/\varepsilon$
this work	$\leq O(qH_q(p)/\varepsilon)$

$(p, H_2(p)/\varepsilon + 2)$ -list decodable with high probability, thereby matching fully random codes up to a small additive constant. This essentially completely resolves RLC list decoding for $q = 2$.

For $q > 2$, we know much less, as summarized in Table 1. A bound of Zyablov and Pinsker [ZP81] gives a bound of $q^{O(1/\varepsilon)}$ for all p . Guruswami, Håstad, and Kopparty [GHK10] showed that random $\mathbb{F}_q$ -linear codes with capacity gap ε are $(p, C_{p,q}/\varepsilon)$ -list decodable with high probability, where

$$C_{p,q} \approx \exp\left(O\left(\frac{\log_2^2 q}{\min((1 - 1/q - p)^2, p)}\right)\right),$$

as analyzed by [GM22, Theorem 7.1]. For fixed p and q , this bound has the optimal ε dependence in contrast with Zyablov and Pinsker [ZP81]. However, this bound's dependence on q is quasipolynomial (worse than [ZP81]), and the bound exponentially grows as p approaches $1 - 1/q$ (the *high error regime*). In particular, if $\delta := 1 - 1/q - p$, note that $C_{p,q}$ is exponential in $1/\delta^2$. The difference between known list sizes for $q = 2$ versus $q = 3$ is quite unsettling—intuitively, incrementing the alphabet size should not drastically affect the list size in such a conceptual way!

High error regime. This motivated a line of work proving list size bounds for $\mathbb{F}_q$ -RLCs with $q \geq 3$ in the high error regime [CGV13, Woo13, RW14, RW18]. Recall that a fully random code has a uniform list size bound of $H_q(p)/\varepsilon$, *even* if we set parameters $\delta = o(1)$, $\varepsilon = o(1)$. Can we prove the same bound for random $\mathbb{F}_q$ -linear codes in the same regime? Some improvements to [GHK10] were made with the condition that δ and ε decreased to 0 in a coupled manner [Woo13, RW18]. But the dream of a uniform list bound of C/ε for any p, q, ε , and absolute constant C remains open.

1.2 Linear Hashing

A more general problem is the question of hashing using a random linear map, which has its own history. Hashing is the idea of randomly mapping a massive universe U to a relatively tiny set of $n \ll |U|$ bins. This compression is necessarily very lossy, and many collisions will occur: many items in the universe will map to the same bin simply by the pigeonhole principle. But if there is a relatively small set S of items that we actually care about, then hopefully by exploiting the randomness, collisions between items in S will be rare. Then, we can use the hash to pretend that our universe is essentially of size n , and reap the benefits in our algorithm, data structure, etc.

The ubiquitousness of hashing in theoretical computer science cannot be understated. Hash functions appear in data structures or in subroutines of algorithms, as ingredients in pseudorandom constructions, and even as convenient mathematical tools for analyzing objects that have nothing to do with hashing. A common complaint in these applications is that a completely random function would have amazing properties in these contexts, but at an enormous cost. Hashing research seeks efficient hash functions which imitate the properties of these expensive random maps.

One such property of great interest is the maximum number of balls in any bin, known as the *max load*, which is a canonical measure of hashing quality that is critical to randomized algorithm design [MU05, Chapter 5]. Suppose that we have n balls (the relevant items from the universe) and a hash table with n bins. Max load measures the imbalance of the table, which is relevant for, e.g., the performance of hash tables that use chaining. In this situation, each bin has an associated linked list that stores all balls that hash to it. To check if a hash table contains a particular key, we need to compute its would-be bin and then sweep through that linked list. Consequently, the worst-case running time of this operation is proportional to the max load, so we want it to be small.

A fully random function hashing n balls into n bins will have expected max load $\sim \ln n / \ln \ln n$. However, the space and time required to deal with such functions is enormous— $O(U \log n)$ bits are required to even describe such a function, which is large enough to defeat the purpose of hashing. Can we get similar performance from a simpler and cheaper family of functions?

Linear hash functions. A natural candidate is the random linear map. For a prime power q , let $n = q^\ell$ for an integer $\ell \geq 1$, let $u \geq \ell$ be arbitrarily large, and let $h : \mathbb{F}_q^u \rightarrow \mathbb{F}_q^\ell$ be a random linear map. For *any* set of n balls $S \subset \mathbb{F}_q^u$, what is the average max load when h hashes S into the set of bins $\mathbb{F}_q^\ell$? This question had been studied since the seminal work of Carter and Wegman introduced universal hash functions [CW79]. Linear hash functions are inexpensive to evaluate and require only $O(\log U \log n)$ bits to describe, so they are an attractive alternative to full randomness.

On the one hand, we know that linear hash functions cannot have better expected max load than fully random hash functions. If we pick the set of balls completely randomly, this simulates the behavior of a completely random hash function and shows that the expected max load for linear hash functions over any q is at least $(1 - o(1)) \ln n / \ln \ln n$. In addition, Alon, Dietzfelbinger, Miltersen, Petrank, and Tardos [ADM⁺99] constructed a set of n balls for which the expected max load is $\Omega(q^{1/3})$, which is relevant when q grows with n .

On the other hand, how much worse are linear hash functions compared to fully random hash functions? Regarding expected max load upper bounds, Carter and Wegman [CW79] established an expected max-load of $O(\sqrt{n})$ on all pairwise independent hash functions. Subsequent improvements, summarized in Table 2, culminated in Alon, Dietzfelbinger, Miltersen, Petrank, and Tardos [ADM⁺99], whose bound was later optimized by Babka [Bab18] to $O(q^{O(\log_2 \log_2 q)} (\log_q n)^{\log_2 q})$.¹ This was the state of affairs until Jaber, Kumar, and Zuckerman [JKZ25] resolved the $q = 2$ case with a bound of $O(\ln n / \ln \ln n)$; the leading factor was later eliminated by Bshouty [Bsh26]. Peculiarly, however, their proof does not appear to extend to $q > 2$, and gives *no* bound for these cases—not even for $q = 3$! Hence, the state of the art for $q \geq 3$ remained the polylogarithmic bound of [ADM⁺99].

As with list decodability, our techniques thus far are too rigid to work over $\mathbb{F}_3$ or other fields. In

¹For $q > 3$, [ADM⁺99] only claimed a bound of $O_q((\log n \log \log n)^{\log_2 q})$ without proof, and [Bab18] removed the $\log \log n$ term for $q = 2$ but did not address the $q > 3$ case. We obtain a bound with explicit q dependence by reconstructing the argument we predict they had in mind in Appendix B, with optimizations to remove the $\log \log n$.

Table 2: Expected max load bounds for random linear hash functions ($\gtrsim$ and $\lesssim$ hide constants)

Source	Bound for $q > 2$	Bound for $q = 2$
<i>Lower bounds</i>		
folklore	$\geq (1 - o(1)) \frac{\ln n}{\ln \ln n}$	$\geq (1 - o(1)) \frac{\ln n}{\ln \ln n}$
[ADM ⁺ 99]	$\gtrsim q^{1/3}$	$\gtrsim 1$
<i>Upper bounds</i>		
[CW79]	$\lesssim \sqrt{n}$	$\lesssim \sqrt{n}$
[MCW78]	$\lesssim q^{1/2} n^{1/4}$	$\lesssim n^{1/4}$
[MV84]	$\lesssim q^{O(\sqrt{\log_q n})}$	$\leq 2^{O(\sqrt{\ln n})}$
[ADM ⁺ 99, Bab18] ¹	$\lesssim q^{O(\log_2 \log_2 q)} (\log_q n)^{\log_2 q}$	$\lesssim \ln n$
[JKZ25, Bsh26]	—	$\leq (1 + o(1)) \frac{\ln n}{\ln \ln n}$
this work	$\lesssim \frac{q \ln \ln q}{\ln q} \cdot \frac{\ln n}{\ln \ln n}$	$\lesssim \frac{\ln n}{\ln \ln n}$

fact, these techniques require additional properties, beyond linearity, which hold over $\mathbb{F}_2$ but vanish over larger fields. Indeed, Jaber, Kumar, and Zuckerman [JKZ25] comment that their argument works for any combinatorial hash function which iteratively groups universe elements in a q -wise independent manner. This property happens to hold for linear maps over $q = 2$, where pairwise independence coincides with the binary field, but this “happy coincidence” ends at $q \geq 3$. Thus, we need a new technique that uses *only* the linearity of a hash function to deduce pseudorandom properties, rather than extra combinatorial properties which happen to be true over $\mathbb{F}_2$.

1.3 Furstenberg Sets

Finally, consider the fundamental question underlying the max load of linear hash functions *and* the list decodability of RLCs: determining the optimal size of a *Furstenberg set* over a finite field. Before giving the definition, let us begin with a much more classical notion: Kakeya sets over $\mathbb{R}^n$. These are subsets that contains a unit line segment in every direction. The famous Kakeya conjecture asks whether Kakeya sets over $\mathbb{R}^n$ must have Hausdorff dimension n . This was proven for $n = 2$ by Davies [Dav71], and for $n = 3$ only very recently by an amazing work of Wang and Zahl [WZ25]. The proof for general n is a longstanding open question.

A sister question to the Kakeya conjecture is the Furstenberg conjecture asked by Wolff [Wol99]. This asks what the minimum Hausdorff dimension of a subset $S \subset \mathbb{R}^2$ must be such that for every direction, there exists a line whose intersection with S is lower bounded in Hausdorff dimension. This question was recently resolved by Orponen and Shmerkin [OS26] and by Ren and Wang [RW25].

Time has made it apparent that this question becomes very hard for large n (even $n = 3, 4, 5, \dots$). In order to stimulate progress for the high-dimensional case, Wolff [Wol99] introduced the finite field analog of the Kakeya conjecture. If one takes the heuristic that the field $\mathbb{R}^n$ can be approximated by $\mathbb{F}_q^n$ with $q \rightarrow \infty$, and sets in $\mathbb{R}^n$ with Hausdorff dimension d correspond to sets in $\mathbb{F}_q^n$ of size $\Theta(q^d)$, the finite field version of the Kakeya conjecture asks whether sets $S \subset \mathbb{F}_q^n$ which contain a line of every slope has size $|S| \geq \Omega_n(q^n)$ (since $q \rightarrow \infty$, the hidden constant should not depend on q). This was resolved by Dvir [Dvi09], whose work kicked off the use of polynomial methods for

Takeya and Furstenberg set lower bounds over finite fields.

Upon the resolution of Wolff’s finite-field Takeya lower bound conjecture by Dvir [Dvi09], generalizations of Takeya and Furstenberg to higher dimensions were studied, where sets must contain shifts of every r -dimension linear subspace (“slope”) for a fixed parameter r . This notion has seen great attention over the reals, [Obe07, FM16, HKM19, Hé19, DOV22, BD25]. Formally:

Definition 1.1 (Takeya sets). A set $S \subset \mathbb{F}_q^n$ is (δ, n, r) -Takeya if for a δ -fraction of dimension- r subspaces V , there exists a shift $x \in \mathbb{F}_q^n$ such that $(x + V) \subseteq S$. If δ is suppressed, assume $\delta = 1$.

The traditional Takeya sets over $\mathbb{F}_q^n$ correspond to $\delta = 1$ and $r = 1$. The study of (n, r) -Takeya with $r > 1$ was initiated by Ellenberg, Oberlin, and Tao [EOT10]. Let us also formalize Furstenberg:

Definition 1.2 (Furstenberg sets). A set $S \subset \mathbb{F}_q^n$ is (δ, n, r, T) -Furstenberg if for a δ -fraction of dimension- r subspaces V , there exists a shift $x \in \mathbb{F}_q^n$ such that $|(x + V) \cap S| \geq T$. If δ is suppressed, assume $\delta = 1$.

Note that (n, r) -Takeya corresponds to setting $\delta = 1$ and $T = q^r$. (n, r, T) -Furstenberg sets were first studied by Ellenberg and Erman [EE16]. (n, r) -Takeya and (n, r, T) -Furstenberg sets have received continued attention throughout the past decade [EOT10, KLSS11, EE16, SS08, BC21, DDL21, DD22, DDL23], and have been used in extractor constructions [DKSS13, DW11] and lattice coverings [ORW22, ORW25]. Extensions of the polynomial method of Dvir [Dvi09] led to optimal lower bounds for (n, r) -Takeya sets in various regimes [DKSS13, KLSS11]. The polynomial method, along with point-line incidence theory, gives an array of (n, r, T) -Furstenberg set lower bounds [EOT10, DDL21, DD22]. One of our focuses is to prove stronger lower bounds on (n, r, T) -Furstenberg sets.

1.4 Our Results

We prove strong lower bounds for Furstenberg sets in a broad range of parameters. This is the main technical input for our other results on linear hashing and RLC list decodability.

Theorem 1.3 (simplified Theorem 2.20). *Let $\delta > 0$, $n \geq r \geq 1$, q be a prime power, and $1 \leq T \leq q^r$. Suppose S is $(\delta, n, n - k, T)$ -Furstenberg. Then,*

$$|S| \geq \frac{\delta T}{8} \cdot q^k \exp\left(-\frac{8q(k + 3 \log_q T)}{T}\right).$$

We did not attempt to optimize constants in the interest of clarity. However, the leading constant can be optimized at the cost of blowing up the constant inside the exponent, and vice versa. For an intuitive way of interpreting this bound, first set $\delta = 1$. Any set of size Tq^k must be $(1, n, n - k, T)$ -Furstenberg by the pigeonhole principle. Our lower bound implies that this is optimal up to a mild factor. Furthermore, this bound is independent of n , and shines when k and T are small. In our applications, we will have n arbitrarily large, and k, T small, so this is precisely the bound shape that we need.

Let us compare our bound with two relevant known bounds: one in the regime T is large, and the other when T is small and $\delta = 1$. The lower bound of Dhar and Dvir [DD22, Theorem 6.14] is quite weak when T is bounded away from 1, but has the strong shape of $|S|/q^n \geq \delta T q^k \exp(-qn/T)$

when $T = 1 - o(1)$. This nearly matches our bound, with the exception of the n in their exponent rather than our (potentially much) smaller $k + 3 \log_q T$. The other lower bound is of Dhar, Dvir, and Lund [DDL23, Theorem 2], who show a bound only when $\delta = 1$, in which the shape of $|S| \geq Tq^k(1 - \sqrt{q2^{k+7}/T})$ is obtained. This gives near-pigeonhole bounds for many values of T , and achieves a bound independent of n like Theorem 1.3. However, this bound is trivial unless $T \gg 2^k$, and using the heuristic

$$1 - \sqrt{q2^{k+7}/T} \approx \exp(-\sqrt{q2^{k+7}/T}),$$

this bound has a exponentially worse dependence on k inside the exponential term. Our bound unifies and optimizes these two prior bounds.

The Furstenberg parameter regime relevant to our applications is the non-classical regime of fixed q and growing n . In this regime, the “mild” exponential factor in our bound is actually necessary, up to constants. Intriguingly, for these parameters, *random sets* are optimal Furstenberg sets (see Appendix A). This is in stark contrast to Furstenberg and Kakeya set constructions in the classical regime of fixed n and growing q , where constructions are algebraic in nature.

Linear hashing. For our first application, we achieve optimal expected max load when hashing n balls into n bins using a random $\mathbb{F}_q$ -linear map. For a function $h : A \rightarrow B$ and subset $S \subset A$, define

$$M(S, h) := \max_{b \in B} |h^{-1}(b) \cap S|$$

to be the *max load* when h is viewed as a hash function mapping $S \subset A$ into B . Then,

Theorem 1.4 (simplified Theorem 3.5). *For sufficiently large ℓ , the following holds. For any prime power q , integers $u \geq \ell$ and $n := q^\ell$, subset $S \subset \mathbb{F}_q^u$ of size n , and uniformly random linear map $h : \mathbb{F}_q^u \rightarrow \mathbb{F}_q^\ell$, we have*

$$\mathbb{E}_h[M(S, h)] \leq \frac{11q \ln \ln q}{\ln q} \cdot \frac{\ln n}{\ln \ln n}.$$

Therefore, for constant q , we achieve optimal max load (up to constants), resolving a question of [ADM⁺99] in this regime. In fact, our techniques also show that optimal max load is achieved with probability $1 - o(1)$. For $q = 2$, our theorem gives an alternative proof of optimal expected max load to the potential-based argument of Jaber, Kumar, and Zuckerman [JKZ25].

We note that Alon, Dietzfelbinger, Miltersen, Petrank and Tardos [ADM⁺99] show the existence of a subset S of size q^ℓ which has max load $\geq q^{1/2}$ when q is a square, and $\geq \Omega(q^{1/3})$ otherwise. Hence, a polynomial dependence on q is unavoidable. However, it is possible the expected max load bound could be improved to $O(q^{1/2} + \ln n / \ln \ln n)$, which would be optimal up to constants for any q and ℓ . We leave this as an open question.

List decoding. Finally, we show that with high probability, a random $\mathbb{F}_q$ -linear code near capacity is list decodable with nearly optimal list size, even in the near-maximal error regime.

Theorem 1.5 (simplified Theorem 4.4). *For every prime power q , $p \in (0, 1 - 1/q)$, $\varepsilon \in (0, 1 - H_q(p))$, and sufficiently large $n \geq n_0(\varepsilon)$, a random linear code $C \subseteq \mathbb{F}_q^n$ of rate $1 - H_q(p) - \varepsilon$ is $(p, O(qH_q(p)/\varepsilon))$ -list decodable with probability $1 - q^{-\Omega(\varepsilon n)}$.*

For all p, q, ε , this is within a factor q of the lower bound of $H_q(p)/\varepsilon$ [GLM⁺22], and thus within a q -factor of a dream uniform bound of $O(1/\varepsilon)$ that holds for all parameters. We emphasize the implications of this bound in the high error rate regime, that is, when $\delta := 1 - 1/q - p \ll 1$. For fixed q , we obtain list size $O(1/\varepsilon)$ for all parameters, even when $\delta \rightarrow 0$. To the best of our knowledge, this is the first bound of this type. Previous bounds either blow up (usually exponentially) [GHK10], or coupled ε and δ together [Woo13, RW18]. We leave removing this factor of q as an open problem.

1.5 Concurrent Work

After our work was completed, we learned of two concurrent and independent works regarding the list decodability of random linear codes which were recently posted on the arXiv [YZ26, Sil26]. Yuan–Zhu and Silas both proved that for any p, q, ε , random linear codes of rate $1 - H_q(p) - \varepsilon$ are $(p, \frac{H_q(p)}{\varepsilon} + C_{p,q})$ -list decodable for some $C_{p,q}$. Silas further showed that for any fixed p, q , for sufficiently small $\varepsilon < \varepsilon_0(p, q)$, a random linear code of rate $1 - H_q(p) - \varepsilon$ will have list size either $\lfloor H_q(p)/\varepsilon \rfloor$ or $\lfloor H_q(p)/\varepsilon \rfloor + 1$.

In the work of Yuan and Zhu, Remark 3.9 of their paper declares that their additive term is

$$C_{p,q} = \exp \left(O \left(\frac{H_q(p) \ln^2 q}{(1 - \frac{1}{q} - p)^2} \right) \right).$$

Hence, this bound has a quasipolynomial dependence on q , and blows up exponentially in the high error regime $p \rightarrow 1 - 1/q$. In the work of Silas, Section 8 of the paper claims that for constant q , their additive term is

$$C_{p,q} = \exp \left(O_q \left(\frac{1}{(1 - \frac{1}{q} - p)^2} \right) \right),$$

and so the list size here does blow up exponentially as $p \rightarrow 1 - 1/q$ at a similar rate as Yuan and Zhu. From tracing the q dependence of Silas’s $C_{p,q}$, it appears to be the case $C_{p,q} = \exp(O_p(\log^3 q))$, which is a bit worse than that of Yuan and Zhu and of Guruswami, Håstad, and Kopparty [GHK10].

In our work, we show that for any p, q, ε , an RLC of rate $1 - H_q(p) - \varepsilon$ is $(p, 16q(H_q(p)/\varepsilon + 1))$ -list decodable. Our argument can be adjusted to replace the leading constant of 16 with 1.01. Hence, for fixed p, q and $\varepsilon \rightarrow 0$, our bound is within a factor of q of optimal, making it worse than the bounds of Yuan and Zhu and of Silas in this regime. However, our bound’s dependence on q is only a single factor of q , and perhaps more notably, our bound has optimal dependence on p . No matter how close p gets to $1 - 1/q$, our list size bound is always at most, say, $16q/\varepsilon$. To the best of our knowledge, our list bound is the only one in the literature which gives a uniform bound of $O_q(1/\varepsilon)$ in the high error rate regime, even with the concurrent work included.

1.6 Proof Overview

We will first dive into the question of linear hashing and reduce it to Furstenberg set lower bounds. Next, we outline how our strategy for Furstenberg sets, and finally we relate linear hashing to the list decodability of RLCs.

1.6.1 Linear Hashing

For simplicity, we will assume our universe has dimension 2ℓ , and consider a simpler question for intuition. For every set S of $n := q^\ell$ balls from the n^2 -size universe $\mathbb{F}_q^{2\ell}$, does there exist a linear map $h: \mathbb{F}_q^{2\ell} \rightarrow \mathbb{F}_q^\ell$ such that $M(S, h) \leq T$, where $T = \Theta_q(\ell/\ln \ell)$? Since each h can be identified with its kernel, and each preimage of h is simply a coset of its kernel, it is sufficient to find one subspace $V \leq_\ell \mathbb{F}_q^{2\ell}$ such that for all $x \in \mathbb{F}_q^{2\ell}$, $|(x + V) \cap S| \leq T$.

Suppose for contradiction that there is a subset $S \subset \mathbb{F}_q^{2\ell}$ for which $M(S, h) > T$ for *all* h . This implies that for all $V \leq_\ell \mathbb{F}_q^{2\ell}$, there exists a shift x such that $|(x + V) \cap S| > T$. Note that this is *exactly* the definition of a $(2\ell, \ell, T)$ -Furstenberg set. Hence, if we can show that all $(2\ell, \ell, L)$ -Furstenberg sets are of size strictly larger than $q^\ell = n$, we will reach a contradiction as desired. We note this translation between Furstenberg sets and max load of linear maps has been made previously in work of Dhar, Dvir, and Lund [DDL21, Lemma 15]

Unfortunately, Furstenberg set lower bounds in the literature are extremely poor in this regime of parameters. The main issue is the fact that T points is a vanishingly small fraction of the q^ℓ points in a coset. We call this the “sparse” regime. Most arguments in the literature go through the polynomial method, which give optimal lower bounds for Kakeya sets. Since Kakeya sets require a shift of every subspace to be fully contained in it, they can be thought of as the “densest” instances of Furstenberg sets. The method, as used traditionally used, severely degrades as the Furstenberg instances get sparser. The best known Furstenberg set lower bound in this regime shows $(2\ell, \ell, L)$ -Furstenberg sets have size at least $(L/q^\ell)^{2\ell} q^{2\ell} \exp(-2\ell/q^{\ell-1}) = q^{-\Omega(\ell^2)}$ [DD22, Theorem 6.14], which is astonishingly worse than the q^ℓ lower bound we desire.

1.6.2 Strong Furstenberg Set Lower Bounds

We now explain our approach to proving strong Furstenberg lower bounds in this sparse regime. To do so, we first give a brief overview of the polynomial method of multiplicities technique used by Kopparty, Lev, Saraf, and Sudan [KLSS11] to show optimal Kakeya set lower bounds. Let $S \subset \mathbb{F}_q^n$ be $(2\ell, \ell)$ -Kakeya. Their argument studies a flavor of algebraic complexity of a Kakeya set by encoding it as a polynomial $P \in \mathbb{F}_q[x_1, \dots, x_n]$ with certain constraints and studying its properties.

In particular, let the *multiplicity complexity* $d_M(S)$ be the smallest possible degree that P can have such that it vanishes with multiplicity M on S . That is, if we take any $M - 1$ derivatives on P , the resulting polynomial will still vanish at every point in S . On the one hand, we can upper bound $d_M(\cdot)$ for *any* set as a strictly increasing function of its size: small sets have smaller d_M , simply by bounding the number of constraints imposed on P . On the other hand, we can show that $d_M(S)$ is large when S is Kakeya by utilizing its structure.

Consider a polynomial $P \in \mathbb{F}_q[x_1, \dots, x_n]$ that vanishes on every $x \in S$ with multiplicity M . The goal will be to show that P will vanish with high multiplicity over a very large product set over the rational function field $\mathbb{F}_q(t_1, \dots, t_\ell)$. Multiplicity Schwartz–Zippel then converts the large product set size and large total multiplicity into a large degree lower bound on P .

For any subspace $V \leq_\ell \mathbb{F}_q^n$, there exists some shift $b_V \in \mathbb{F}_q^{2\ell}$ such that $b_V + V \subset S$. V can be parametrized by a linear form in ℓ formal variables $t_1, \dots, t_\ell$, and thus the restriction of P to $b_V + V$ can be written as a polynomial in $\mathbb{F}_q(t_1, \dots, t_\ell)$ of total multiplicity Mq^ℓ . One can gather analogous information over all subspaces V and derivatives of p , glue them together using multiplicity Schwartz–Zippel, and derive a lower bound on d_M , which will scale with the total

multiplicity garnered at each restriction.

What happens when S is $(2\ell, \ell, L)$ -Furstenberg rather than Kakeya? It is natural to attempt the same argument, where we look at a minimal degree polynomial $P \in \mathbb{F}_q[x_1, \dots, x_n]$ which vanishes on S with multiplicity M . The constraint-counting argument remains the same, but the Schwartz–Zippel argument now suffers. For every $V \leq_\ell \mathbb{F}_q^n$, we know there is a shift b_V such that $b_V + V$ intersects S in $\geq L$ vectors. Hence, P restricted to this affine space will be an ℓ -variate polynomial with total multiplicity $ML \ll Mq^\ell$. This multiplicity degradation causes the gluing procedure to give a much smaller degree lower bound. We remark that this approach recovers the Furstenberg lower bounds of Dhar and Dvir mentioned earlier using a different argument [DD22, Theorem 6.14].

Polynomial method of multiplicity gaps. Our idea to avoid this severe multiplicity degradation is surprisingly simple: enforce multiplicity constraints on $x \notin S$ as well. This initially seems like a horrible idea, as this implies our polynomial will now be zero *everywhere*, rather than on S alone. How can such a polynomial encode any information about S ? We crucially set the multiplicity constraints on $x \notin S$ to be more relaxed than on $x \in S$. Hence, information about our set is encoded solely in the derivatives—in the *gap* between the two classes of multiplicity constraints—rather than in the zero set.

More concretely, for parameters $M' < M$, we define the *multiplicity gap complexity* $d_{M,M'}(S)$ to be the smallest nonzero degree of a polynomial P which vanishes with multiplicity $\geq M$ on each $x \in S$, and vanishes with multiplicity $\geq M'$ on $x \notin S$. Then for every $V \leq_r \mathbb{F}_q^n$, the restriction of P to $b_V + V$ will now have total multiplicity $ML + M'(q^\ell - L)$, which gets very close to the Kakeya-case quantity Mq^ℓ as M' approaches M .

However, there is a cost to increasing M' . The larger M' is, the more linear constraints each $x \notin S$ will induce in the coefficients, which then degrades the dependence of the degree upper bound on $|S|$. By performing the computation, one observes that the number of linear constraints is dominated by enforcing multiplicity M on S , unless M' is extremely close to M . We thus set $M' = (1 - \varepsilon)M$, compute the lower bound implied by the polynomial method, and optimize the value of ε to yield our optimal bounds.

To the best of our knowledge, applications of polynomial method to lower bound sets with structure always consider polynomials that vanish on that set (perhaps with high multiplicity), but never enforce any conditions outside of S . In the setting of Furstenberg sets, enforcing multiplicity outside of the set allows us to get optimal lower bounds. We hope that this technique of enforcing multiplicity outside of S will have further applications.

Furstenberg acrobatics. To generalize our results to arbitrarily large universe and to develop tail bounds strong enough to deduce expectation, we require ways to strengthen parameters of Furstenberg sets.

Given a (δ, n, r, T) -Furstenberg set S , where $\delta \ll 1$, we use a classic “random rotations” trick to construct a $(1/2, n, r, T)$ -Furstenberg set whose size is not much bigger than S . This is done by taking unions of sufficiently many random rotations of S to cover more directions, and was most relevantly used by Ordentlich, Regev, and Weiss [ORW22, Theorem 2.8] for Kakeya sets. This allows us to amplify tail bounds.

We also provide a new way to compress the ambient dimension of a Furstenberg set. Given a

(δ, n, r, T) -Furstenberg set S and $n' \ll n$, we can construct a $(\delta, n', n' - n + r, T)$ -Furstenberg set of size at most $|S|$. This is done by randomly mapping $S \subset \mathbb{F}_q^n$ into $\mathbb{F}_q^{n'}$ using a random linear map, and showing that for some choice of linear map, the image of S is the desired Furstenberg set. To the best of our knowledge, this maneuver appears to be new, but is heavily inspired by a compression trick done in the work of Alon, Dietzfelbinger, Miltersen, Petrank, and Tardos [ADM⁺99] on linear hashing. This trick allows us to generalize from input dimension 2ℓ to arbitrarily large dimensions.

1.6.3 List Decodability of RLCs

After getting max load bounds for linear hashing, list decoding bounds for RLCs immediately follow. This is because a linear code $C \leq \mathbb{F}_q^n$ of rate R being (p, L) -list decodable is exactly equivalent to $M(h, B) \leq L$, where $h : \mathbb{F}_q^n \rightarrow \mathbb{F}_q^{(1-R)n}$ is any linear map of kernel C , and $B \subset \mathbb{F}_q^n$ is Hamming ball of radius pn . Therefore, analyzing the list decodability of an RLC is equivalent to analyzing how well a random linear map hashes Hamming balls (rather than arbitrary sets). RLC list decodability is consequently a simple application of our hashing result.

2 Tight Bounds for Furstenberg Sets

2.1 Preliminaries

$U \leq_r V$ denotes that U is a dimension- r subspace of V .

Hasse derivatives. We make use of Hasse derivatives, which are an analog of traditional derivatives which are useful for multivariate polynomials over finite fields. However, all of the properties of Hasse derivatives that we use have analogs that hold for traditional derivatives over $\mathbb{R}$. These properties and their proofs can be found in the work of Dvir, Kopparty, Saraf, and Sudan [DKSS13].

Definition 2.1 (Hasse derivative). Let $X = (X_1, \dots, X_n)$ and $P(X) \in \mathbb{F}[X]$ be an n -variate polynomial. Let $\vec{i}$ be a non-negative integer vector of dimension n . The $\vec{i}$ -th Hasse derivative of P , denoted $P^{(\vec{i})}$, is the unique coefficient of $Z^{\vec{i}}$ in the expansion of $P(X + Z) \in \mathbb{F}[X, Z]$:

$$P(X + Z) = \sum_{\vec{i}} P^{(\vec{i})}(X) Z^{\vec{i}}.$$

From the definition, it is clear that the Hasse derivative is linear.

Fact 2.2. *Let P be a polynomial. Then, every coefficient of every Hasse derivative of P is a linear combination of the coefficients of P .*

For a vector $\vec{i} \in \mathbb{Z}_{\geq 0}^n$, let $\|\vec{i}\| = \sum_{j=1}^n i_j$. Just as in the reals, taking a derivative reduces the degree of a polynomial by 1.

Fact 2.3 ([Sar11, Lemma 2.1.4]).

$$\deg(P^{(\vec{i})}) = \deg(P) - \|\vec{i}\|.$$

For a polynomial $P \in \mathbb{F}_q[x_1, \dots, x_n]$, denote P_H to be the homogeneous part of P , i.e., the unique homogeneous polynomial such that $\deg(P - P_H) < \deg P$. The derivative of the homogeneous part is the homogeneous part of the derivative.

Fact 2.4 ([DKSS13, Proposition 2.3]).

$$(P^{(\vec{i})})_H \equiv (P_H)^{(\vec{i})}$$

We now define the multiplicity of a polynomial at a value.

Definition 2.5 (Multiplicity). Let $P \in \mathbb{F}_q[x_1, \dots, x_n]$ and $a \in \mathbb{F}_q^n$. We define $\text{mult}(P, a)$ to be the largest nonnegative integer m such that for all $\vec{i} \in \mathbb{Z}_{\geq 0}^n$, with $\|\vec{i}\| < m$, $P^{(\vec{i})}(a) = 0$.

If a polynomial has a zero with multiplicity at least m at some point, the derivative of the polynomial will have multiplicity at least $m - 1$ at that point.

Fact 2.6 (Multiplicity loss [DKSS13, Proposition 2.3]).

$$\text{mult}(P^{(\vec{i})}, \vec{a}) \geq \text{mult}(P, \vec{a}) - \|\vec{i}\|.$$

We finally present a very important lemma, relating the number of zeros (with multiplicity) of a polynomial to its degree.

Lemma 2.7 (Multiplicity Schwartz–Zippel, [DKSS13, Lemma 2.7]). *If P is a n -variate nonzero² polynomial over $\mathbb{F}$ of (total) degree at most d , then for all $S \subseteq \mathbb{F}$,*

$$\sum_{\vec{x} \in S^n} \text{mult}(P, \vec{x}) \leq d \cdot |S|^{n-1}.$$

2.2 Method of Multiplicity Gaps

Let us begin with the a lower bound which holds for every value of a parameter ε , to pick later.

Theorem 2.8. *Let $S \subset \mathbb{F}_q^n$ be $(\delta, r, \alpha q^r)$ -Furstenberg. For all $0 < \varepsilon \leq 1$ and $\beta_\varepsilon := \alpha + (1 - \varepsilon)(1 - \alpha)$,*

$$|S| \geq q^n \cdot \frac{\left(\frac{\delta \beta_\varepsilon q^r}{\delta q^r + q - 1}\right)^n - (1 - \varepsilon)^n}{1 - (1 - \varepsilon)^n},$$

or equivalently

$$\begin{aligned} |\bar{S}| &\leq q^n \cdot \frac{1 - \left(\frac{\delta \beta_\varepsilon q^r}{\delta q^r + q - 1}\right)^n}{1 - (1 - \varepsilon)^n} \\ &= q^n \cdot \frac{1 - \left(1 - \frac{\delta \varepsilon (1 - \alpha) q^r + q - 1}{\delta q^r + q - 1}\right)^n}{1 - (1 - \varepsilon)^n}. \end{aligned}$$

²A polynomial that is not the literal zero— P might still evaluate to zero everywhere!

We prove this using the polynomial method of *multiplicity gaps*, building on the extended multiplicity method of [DKSS13, KLSS11] with a new twist. At a high level, we will analyze the *multiplicity gap complexity* of a set S : what is the smallest d such that there exists a nonzero degree- d polynomial P that encodes S in its multiplicity gaps? That is, for some parameter M , we want a P with multiplicity M on every point in S and multiplicity $(1 - \varepsilon)M$ on every point in $\bar{S}$.

On the one hand, for *any* set S , there is an easy upper bound on the multiplicity gap complexity that depends on the size of S , by counting the linear constraints imposed on P by the multiplicity requirements. The smaller the set, the fewer the imposed constraints and so the smaller the multiplicity gap complexity.

On the other hand, we can lower bound the complexity when S is Furstenberg. Because S hits many affine subspaces, restrictions of P to those subspaces must vanish with high multiplicity. All of these subspaces can be treated as distinct roots when we work over a function field, and by Schwartz–Zippel (in particular, the strengthened variant using multiplicities), only a high degree polynomial can have this many roots. Therefore, a Furstenberg set has high multiplicity gap complexity, so it cannot be small.

Lemma 2.9 (Constraint counting [DDL21, Lemma 12]). *For any integer parameter M , subset $S \subseteq \mathbb{F}^n$, and real $0 < \varepsilon \leq 1$, choose the smallest d such that*

$$\binom{M+n-1}{n}|S| + \binom{(1-\varepsilon)M+n-1}{n}|\bar{S}| \leq \binom{d+n}{n}. \quad (1)$$

Then, there exists a nonzero n -variate polynomial P of degree d over $\mathbb{F}$ such that

$$\begin{aligned} \forall \vec{x} \in S, \text{mult}(P, \vec{x}) &\geq M, \\ \forall \vec{x} \in \bar{S}, \text{mult}(P, \vec{x}) &\geq (1 - \varepsilon)M. \end{aligned} \quad (\text{M})$$

Proof. If $\text{mult}(P, \vec{x}) \geq M$ for a point $\vec{x} \in S$, then for every $\vec{i}$ with $\|\vec{i}\| < M$, $P^{(\vec{i})}(\vec{x}) = 0$. The coefficients of $P^{(\vec{i})}$ are linear combinations of coefficients of P , so this enforces a linear constraint on the coefficients of P . Using stars and bars, there are $\binom{M+n-1}{n}|S|$ many pairs $(\vec{i}, \vec{x})$ with $\vec{x} \in S$. Thus, the total rank of the linear constraints imposed on P by S and $\bar{S}$ is at most the LHS of (1). Meanwhile, the RHS is the number of coefficient variables in a degree- d n -variate polynomial, also by stars and bars. Since the constraints do not outnumber the variables of this system, there must exist a nontrivial solution, i.e., a nonzero P satisfying the multiplicity constraints. $\square$

Fact 2.10 (Multiplicity gap complexity upper bound). *The degree d chosen in Lemma 2.9 satisfies*

$$\binom{M+n-1}{n}|S| + \binom{(1-\varepsilon)M+n-1}{n}|\bar{S}| > \binom{d+n-1}{n}, \quad (2)$$

because (1) is strictly increasing in d .

Proposition 2.11 (Multiplicity gap complexity lower bound). *For all $0 < \varepsilon \leq 1$ and $\beta_\varepsilon := \alpha + (1 - \varepsilon)(1 - \alpha)$, let $S \subseteq \mathbb{F}_q^n$ be $(\delta, r, \alpha q^r)$ -Furstenberg, and let P be nonzero and degree- d satisfying (M). Then,*

$$d > \delta q^r \left\lceil \frac{q\beta_\varepsilon M - d}{q - 1} \right\rceil. \quad (3)$$

Proof. We will lower bound the multiplicity of the homogeneous part of P on different subspaces and use multiplicity Schwartz–Zippel to get a degree lower bound. The first step is to show that on many affine subspaces, many derivatives of P vanish.

Claim 2.12. *For a δ fraction of $n \times r$ matrices V , there exists a shift $\vec{b}$ such that for every $\vec{i}$ with $w := \|\vec{i}\|$, the r -variate polynomial $Q_{\vec{i},V}(\vec{t}) := P^{(\vec{i})}(\vec{b} + V\vec{t})$ has total multiplicity at least*

$$\sum_{\vec{t} \in \mathbb{F}_q^r} \text{mult}(Q_{\vec{i},V}, \vec{t}) \geq \alpha q^r (M - w) + (1 - \alpha) q^r ((1 - \varepsilon)M - w) = q^r (\beta_\varepsilon M - w). \quad (4)$$

Claim proof. First, consider a random V with full column rank. Its columns span a uniformly random r -dimension subspace, so with probability δ , there must be a shift $\vec{b} + \text{colspan}(V)$ that intersects S in at least αq^r points. Every $\vec{t}$ is mapped to a unique point in this shift, and using [Fact 2.6](#) with the multiplicities enforced by [\(M\)](#) gives [\(4\)](#).

Now, let us extend this to general $n \times r$ matrices. For each $s < r$, take a uniformly random rank- s matrix V and let $\Lambda := \text{colspan}(V)$, and complete Λ to a random rank- r subspace Λ' . With probability δ over V and Λ' , there is a good shift $\vec{y} + \Lambda'$ that intersects S at αq^r points. This means that for at least a δ fraction of V , there exists a suitable completion Λ' with such a good shift.

Consider such V and completely tile the good completion $\vec{y} + \Lambda'$ with q^{r-s} shifts of Λ . By averaging, at least one of these shifts $\vec{b} + \Lambda$ hits S at αq^s points. Finally, notice that for every $\vec{x} \in \Lambda$, there are q^{r-s} solutions $\vec{t}$ to $V\vec{t} = \vec{x}$. Thus, for at least δ fraction of V , there exists $\vec{b}$ so that

$$\sum_{\vec{t} \in \mathbb{F}_q^r} \text{mult}(Q_{\vec{i},V}, \vec{t}) = \sum_{\vec{x} \in \Lambda} q^{r-s} \text{mult}(P^{(\vec{i})}, \vec{b} + \vec{x}) \geq q^{r-s} \cdot q^s (AM - w),$$

by [Fact 2.6](#) and the multiplicity constraints [\(M\)](#). □

Using the contrapositive of multiplicity Schwartz–Zippel, if $Q_{\vec{i},V}$ satisfies [\(4\)](#) and

$$q(\beta_\varepsilon M - w) > d - w \geq \deg(Q_{\vec{i},V}),$$

then $Q_{\vec{i},V}$ is the zero polynomial. This is true whenever

$$w < m := \left\lceil \frac{q\beta_\varepsilon M - d}{q - 1} \right\rceil.$$

If $Q_{\vec{i},V} \equiv 0$, we can show that part of $P^{(\vec{i})}$ must also be zero.

Claim 2.13 (Unshifting). *If $Q_{\vec{i},V}$ is zero, then so is $(P^{(\vec{i})})_H(V\vec{t})$.*

Claim proof. Recall that $Q_{\vec{i},V}(\vec{t}) := P^{(\vec{i})}(\vec{b} + V\vec{t})$. Plug $\vec{b} + V\vec{t}$ into each monomial of $(P^{(\vec{i})})_H$ and expand; each such monomial (say of degree ℓ) will expand to

$$(\vec{b} + V\vec{t})_{j_1} (\vec{b} + V\vec{t})_{j_2} \dots (\vec{b} + V\vec{t})_{j_\ell} = (V\vec{t})_{j_1} (V\vec{t})_{j_2} \dots (V\vec{t})_{j_\ell} + (\text{products of } < \ell \text{ terms}).$$

The first term on the RHS is the same monomial but with $V\vec{t}$ substituted instead of $\vec{b} + V\vec{t}$, so the sum over these terms across monomials will appear in both $Q_{\vec{i},V}$ and $(P^{(\vec{i})})_H(V\vec{t})$. This is the only possible source of degree- ℓ monomials in $Q_{\vec{i},V}$, so if $(P^{(\vec{i})})_H(V\vec{t})$ is nonzero, then so is $Q_{\vec{i},V}$. □

We have shown so far that for a δ fraction of $n \times r$ matrices V , every $\|\vec{i}\| < m$ will give $(P^{(\vec{i})})_H(V\vec{t}) \equiv 0$ (equivalently, $(P_H)^{(\vec{i})}(V\vec{t}) \equiv 0$ via [Fact 2.4](#)). Let us now treat $V\vec{t}$ itself as the parameter by viewing $(P_H)^{(\vec{i})}$ as a polynomial over the r -variate function field $\mathbb{F}_q(\vec{t})$ —now, $(P_H)^{(\vec{i})}(V\vec{t}) \equiv 0$ means that $V\vec{t}$ itself is a root. Define the set of n -output linear functions on r variables, which is a product set:

$$T^n := \{V\vec{t} : V \in \mathbb{F}_q^{n \times r}\}, \quad T := \{\langle \vec{u}, \vec{t} \rangle : \vec{u} \in \mathbb{F}_q^r\} \subseteq \mathbb{F}_q(\vec{t}).$$

On a δ fraction of T^n , for all $\vec{i}$ with $\|\vec{i}\| < m$, $(P_H)^{(\vec{i})}$ is the zero polynomial, which is the zero element of the function field. This is saying that P_H (as a polynomial over $\mathbb{F}_q(\vec{t})$) has multiplicity m on a δ fraction of T^n . P_H overall is nonzero so by multiplicity Schwartz–Zippel over $\mathbb{F}_q(\vec{t})$,

$$d \geq \delta m |T| > \delta q^r \left\lceil \frac{q\beta_\varepsilon M - d}{q - 1} \right\rceil. \quad \square$$

At last, let us prove [Theorem 2.8](#) by combining our bounds for d .

Proof of Theorem 2.8. As a reminder, M is a parameter that we can set to be whatever we want. With that in mind, let us compare our upper and lower bounds on d which hold for all M . (2) gives

$$\begin{aligned} \binom{d+n-1}{n} &< \binom{M+n-1}{n} |S| + \binom{(1-\varepsilon)M+n-1}{n} |\bar{S}| \\ &= \left(\binom{M+n-1}{n} - \binom{(1-\varepsilon)M+n-1}{n} \right) \cdot |S| + q^n \binom{(1-\varepsilon)M+n-1}{n}. \end{aligned}$$

Rearranging this gives a lower bound for $|S|$ for every choice of M , including in the limit $M \rightarrow \infty$:

$$\begin{aligned} |S| &\geq \frac{\eta - q^n \tau}{1 - \tau}, \\ \eta &:= \frac{\binom{d+n-1}{n}}{\binom{M+n-1}{n}}, \quad \tau := \frac{\binom{(1-\varepsilon)M+n-1}{n}}{\binom{M+n-1}{n}}. \end{aligned} \quad (5)$$

We can analyze the behavior of η and τ as follows.

$$\begin{aligned} \tau &= \frac{(1-\varepsilon)M+n-1}{M+n-1} \cdot \frac{(1-\varepsilon)M+n-2}{M+n-2} \cdots \frac{(1-\varepsilon)M}{M} \implies \lim_{M \rightarrow \infty} \tau = (1-\varepsilon)^n. \\ \eta &= \frac{d+n-1}{M+n-1} \cdot \frac{d+n-2}{M+n-2} \cdots \frac{d}{M} \geq \left(\frac{d}{M} \right)^n. \end{aligned}$$

(3) gives a lower bound for d/M :

$$\frac{d}{M} > \frac{\delta\beta_\varepsilon q^{r+1}}{\delta q^r + q - 1}.$$

Plugging into (5) and taking the limit $M \rightarrow \infty$, we get our bound.

$$\begin{aligned} |S| &\geq \lim_{M \rightarrow \infty} \frac{\eta - q^n \tau}{1 - \tau} \\ &\geq q^n \cdot \frac{\left(\frac{\delta\beta_\varepsilon q^r}{\delta q^r + q - 1} \right)^n - (1-\varepsilon)^n}{1 - (1-\varepsilon)^n}. \end{aligned} \quad \square$$

Next, we choose ε to get a good lower bound. This bound is nearly tight; see [Appendix A](#).

Corollary 2.14. *Let $S \subset \mathbb{F}_q^n$ be $(\delta, n, r, \alpha q^r)$ -Furstenberg. For all $1 < C < \alpha \delta q^{r-1}/2$,*

$$\frac{|S|}{q^n} \geq \exp\left(-\frac{2Cn}{\delta \alpha q^{r-1}}\right) \cdot \left(1 - \frac{1}{C}\right) \cdot \alpha.$$

Proof. Set $\gamma := \frac{q-1}{\delta q^r}$. From [Theorem 2.8](#), we have for every $\varepsilon \in (0, 1)$ that

$$\frac{|S|}{q^n} \geq \frac{\left(1 - \frac{\delta \varepsilon (1-\alpha) q^r + q - 1}{\delta q^r + q - 1}\right)^n - (1 - \varepsilon)^n}{1 - (1 - \varepsilon)^n} = \frac{\left(1 - \frac{\varepsilon(1-\alpha) + \gamma}{1 + \gamma}\right)^n - (1 - \varepsilon)^n}{1 - (1 - \varepsilon)^n}.$$

By the mean value theorem, we have that for $x > y > 0$, $x^n - y^n \geq n y^{n-1} (x - y)$, and by Bernoulli's inequality, we have $1 - (1 - \varepsilon)^n \leq n\varepsilon$. Hence,

$$\frac{|S|}{q^n} \geq \frac{n(1 - \varepsilon)^{n-1} \left(\varepsilon - \frac{\varepsilon(1-\alpha) + \gamma}{1 + \gamma}\right)}{n\varepsilon} = \frac{\alpha + \gamma - \gamma/\varepsilon}{1 + \gamma} \cdot (1 - \varepsilon)^{n-1}.$$

As ε gets smaller, the second factor becomes larger, but the first factor decreases, and even becomes negative once $\varepsilon \leq \gamma/(\alpha + \gamma)$. However, we note that for any $\varepsilon \geq C\gamma/(\alpha + \gamma)$ that is growing, the first term remains the same asymptotically, while the second term decays. This hints that we should set $\varepsilon = C\gamma/(\alpha + \gamma)$ for some constant-bounded $C > 1$. Plugging this in, we have

$$\begin{aligned} \left(\frac{\alpha + \gamma - \gamma/\varepsilon}{1 + \gamma}\right) (1 - \varepsilon)^{n-1} &= \left(1 - \frac{1}{C}\right) \frac{\alpha + \gamma}{1 + \gamma} \left(1 - \frac{C\gamma}{\alpha + \gamma}\right)^{n-1} \\ &\geq \left(1 - \frac{1}{C}\right) \alpha \left(1 - \frac{C\gamma}{\alpha}\right)^{n-1}. \end{aligned}$$

Assuming $1 - C\gamma/\alpha \geq 1/2$, we use $1 - x \geq \exp(-x/(1 - x))$ to obtain

$$\frac{|S|}{q^n} \geq \left(1 - \frac{1}{C}\right) \alpha \exp\left(-\frac{2Cn\gamma}{\alpha}\right) \geq \left(1 - \frac{1}{C}\right) \alpha \exp\left(-\frac{2Cn}{\alpha \delta q^{r-1}}\right). \quad \square$$

Remark 2.15. The constant 2 inside the exponential term in [Corollary 2.14](#), which will affect the load bound for hashing and the list size for random linear codes, can be set arbitrarily close to 1 as long as the corresponding assumption on C is updated. For clarity, we will not optimize this and other constants throughout.

2.3 Furstenberg Acrobatics

2.3.1 Random Rotations

We now improve the Furstenberg set lower bound for small δ by using a “random rotations” argument, just like how Ordentlich, Regev, and Weiss did for Kakeya sets [[ORW22](#), Lemma 2.6]. This is crucial to getting tail bounds sharp enough to obtain optimal expectation and also to get optimal probability of achieving list decoding capacity.

Lemma 2.16. *Let $0 < \varepsilon < \delta < 1$. Assume $S \subset \mathbb{F}_q^n$ is (ε, n, r, T) -Furstenberg. There exists S' which is (δ, n, r, T) -Furstenberg with*

$$|S'| \leq \left\lceil \frac{\log(1 - \delta)}{\log(1 - \varepsilon)} \right\rceil \cdot |S|.$$

Proof. The idea will be to take the union of random rotations of S . These random rotations will introduce new “rich” directions, thereby boosting ε to δ . We will then upper bound the number of random rotations needed, giving the upper bound on the size of S' .

Let $M_1, \dots, M_k$ be random and independent $n \times n$ invertible matrices over $\mathbb{F}_q^n$. Consider $S' = \bigcup_{i \in [k]} M_i S$, where $MS = \{Ms : s \in S\}$. Fix an r -dimensional subspace $V \leq \mathbb{F}_q^n$. For a fixed i , the probability that V is not a rich direction in $M_i S$ (does not intersect $M_i S$ in at least αq^r points) is at most $1 - \varepsilon$. This is because at least ε fraction of subspaces are rich for S , and an M_i transitively permutes these rich subspaces. Hence the probability V is not rich when we take the union of k independent rotations is $(1 - \varepsilon)^k$. Thus, there exists a choice of $M_1, \dots, M_k$ such that the number of rich directions of S' comprise of at least $1 - (1 - \varepsilon)^k$. For this quantity to be at least δ , we need $k = \lceil \log(1 - \delta) / \log(1 - \varepsilon) \rceil$. The result now follows by noting $|S'| \leq k|S|$. $\square$

Combine this random rotations lemma with our prior lower bound [Corollary 2.14](#). This shifts the dependence on δ from inside the exponential to outside.

Corollary 2.17. *Let $S \subset \mathbb{F}_q^n$ be (δ, n, r, T) -Furstenberg. As long as $T \geq 8q$,*

$$|S| \geq \frac{\delta T}{4} \cdot q^{n-r} \exp\left(-\frac{8qn}{T}\right)$$

Proof. We will use the following claim:

Claim 2.18. *For $x \in (0, 1)$,*

$$\left\lceil \frac{\log(1/2)}{\log(1-x)} \right\rceil \leq \frac{2}{x}.$$

Claim proof. First,

$$\ln(1-x) \leq -x \implies \frac{\ln(1/2)}{\ln(1-x)} \leq \frac{\ln 2}{x}.$$

Then,

$$\begin{aligned} \left\lceil \frac{\log(1/2)}{\log(1-x)} \right\rceil &\leq \frac{\log(1/2)}{\log(1-x)} + 1 \\ &\leq \frac{x + \ln 2}{x} \leq \frac{2}{x}, \end{aligned}$$

since $x < 1 < 2 - \ln 2$. $\square$

Now, assume for the sake of contradiction that there exists set S contradicting the corollary. By [Lemma 2.16](#), there exists a $(1/2, n, r, T)$ -Furstenberg set S' of size

$$\begin{aligned} |S'| &< \left\lceil \frac{\log(1/2)}{\log(1-\delta)} \right\rceil \cdot \frac{\delta T}{4} \cdot q^{n-r} \exp\left(-\frac{8qn}{T}\right) \\ &\leq \frac{T}{2} \cdot q^{n-r} \exp\left(-\frac{8qn}{T}\right). \end{aligned}$$

However, this directly contradicts [Corollary 2.14](#) with $C = 2$. $\square$

2.3.2 Random Projections

Jumping ahead a bit, [Corollary 2.17](#) is strong enough to show optimal expected max load for hash functions on ambient dimension $n = O(\ell)$, which is the expository case described in [Section 1.6](#), as well as optimal list size for random linear codes with rate bounded away from zero. Unfortunately, this lower bound is meaningless when $n \gg T/q$, which is the setting of hashing from large universes or random linear codes with vanishing rate.

We will use a dimension reduction idea to improve the Furstenberg set lower bound in the case where $r = n - k$ for small k , at the cost of a leading constant factor. We can do so by taking a random projection of a $(\delta, n, n - k, T)$ -Furstenberg set which reduces the ambient dimension drastically to something which only depends on k and T . This is inspired by a compression step done in [\[ADM⁺99\]](#). The first step is to show that a random linear map that compresses *any* arbitrarily large vector space to $2t$ dimensions will be surjective on an arbitrary fixed set of size q^t with constant probability. This is optimal as it is the converse of the birthday paradox.

Lemma 2.19 (Dimension reduction). *Let V be a vector space, and let $f: V \rightarrow \mathbb{F}_q^{2t}$ be a uniformly random linear map. For every set $B \subseteq V$ of size at most q^t ,*

$$\Pr_f[|f(B)| = |B|] \geq \frac{1}{2}.$$

Proof. By 2-universality, any fixed pair of vectors collide with probability $1/q^{2t}$. There are $\binom{|B|}{2}$ pairs of vectors from B , so the expected number of collisions within B is at most

$$q^{-2t} \binom{|B|}{2} \leq \frac{|B|^2}{2q^{2t}} \leq 1/2.$$

Therefore, with probability $\geq 1/2$, there are no collisions, which implies $|f(B)| = |B|$. $\square$

If we take a $(\delta, n, n - k, T)$ -Furstenberg set S over $\mathbb{F}_q^n$ and apply a random projection, the resulting set will still be meaningfully Furstenberg, because this lemma shows that many affine subspaces that intersect S at T points will continue to do so after projection. But our new Furstenberg set is over a significantly smaller ambient dimension, so we can apply our Furstenberg lower bound which implies a much stronger lower bound for S .

Theorem 2.20 (formal version of [Theorem 1.3](#)). *Let $\delta > 0$, $n \geq k \geq 1$, q be a prime power, $8q \leq T \leq q^{n-k}$, and $t = \lceil \log_q T \rceil$. Suppose S is $(\delta, n, n - k, T)$ -Furstenberg. Then, there exists a $(\delta/2, k + 2t, 2t, T)$ -Furstenberg set of size at most $|S|$, which implies*

$$|S| \geq \frac{\delta T}{8} \cdot q^k \exp\left(-\frac{8q(k + 2t)}{T}\right).$$

Proof. Let $h: \mathbb{F}_q^n \rightarrow \mathbb{F}_q^{k+2t}$ be a uniformly random *surjective* linear map. Consider a uniformly random subspace $W \leq_{2t} \mathbb{F}_q^{k+2t}$. We will show that

$$\Pr_{h,W}[\exists b \in \mathbb{F}_q^{k+2t}, |(b + W) \cap h(S)| \geq T] \geq \frac{\delta}{2}. \quad (6)$$

Then it will follow that the proportion of directions $W \leq_{2t} \mathbb{F}_q^{k+2t}$ which have a shift that intersects $h(S)$ on at least T points is at least $\delta/2$ in expectation. We can then fix h that achieves at least this average, and obtain that $h(S)$ is $(\delta/2, k + 2t, 2t, T)$ -Furstenberg of size $|h(S)| \leq |S|$. Apply [Corollary 2.17](#) to get a lower bound.

Define the random subspace $V := h^{-1}(W)$. We note that V will be a random subspace of dimension $n - k$ in $\mathbb{F}_q^n$. Define the event

$$\mathcal{E} := (\exists b_V \in \mathbb{F}_q^n, |(b_V + V) \cap S| \geq T).$$

By definition of S , we know that $\Pr_{h,W}[\mathcal{E}] \geq \delta$. Condition on $\mathcal{E}$ which fixes some subspace V which is heavy for S , and let S_V be an arbitrary subset of size T from $(b_V + V) \cap S$. Despite this conditioning, $g := h|_V: V \rightarrow W$ is still a uniformly random surjective map. And because $S_V - b_V \subseteq V$ is a set of size $T \leq q^t$, [Lemma 2.19](#) implies that g will preserve its size often:

$$\Pr[|g(S_V - b_V)| = T \mid \mathcal{E}] \geq 1/2.$$

But this implies that

$$T = |g(S_V - b_V)| \leq |h((S - b_V) \cap V)| = |h(S - b_V) \cap W| = |h(S) \cap (h(b_V) + W)|,$$

and thus there is a shift of W that intersects $h(S)$ in at least T places. Hence,

$$\Pr_h[\exists b \in \mathbb{F}_q^{k+2t}, |(b + W) \cap h(S)| \geq T] \geq \Pr[\mathcal{E}] \cdot \Pr[|g(S_V - b_V)| = T \mid \mathcal{E}] \geq \frac{\delta}{2},$$

establishing [\(6\)](#). □

3 Linear Hashing over $\mathbb{F}_q$

We will explicitly denote dependence on q anywhere it appears.

Definition 3.1 (Max load). For a function $h : A \rightarrow B$ and subset $S \subset A$, we define

$$M(S, h) := \max_{b \in B} |h^{-1}(b) \cap S|$$

to be the *max load* when h is viewed as a hash function mapping $S \subset A$ into B .

3.1 Max Load Tail Bounds

A tail bound for the max load of a random (surjective) linear hash function is a special case of a Furstenberg lower bound. The contents of each bin are determined by shifts of the kernel of the hash function, so a bad set of balls S is precisely one where many subspaces have shifts that heavily intersect S . If we directly apply our Furstenberg lower bound to the hashing setting, we get the following tail bound:

Theorem 3.2 (Load tail bound from Furstenberg). *Let q be a prime power, $u \geq \ell \geq 1$ be integers, and $h : \mathbb{F}_q^u \rightarrow \mathbb{F}_q^\ell$ be a uniformly random surjective $\mathbb{F}_q$ -linear map. For any $T \geq 8q$ and any $S \subseteq \mathbb{F}_q^u$,*

$$\Pr[M(S, h) \geq T] \leq \frac{8|S|}{q^\ell T} \exp\left(\frac{8q(\ell + 2\lceil \log_q T \rceil)}{T}\right).$$

Proof. Denote $\delta := \Pr[M(S, h) \geq T]$ and $t := \lceil \log_q T \rceil$. The kernel of h is a uniformly random subspace of $\mathbb{F}_q^u$ of dimension $u - \ell$. If $M(S, h) \geq T$, then some shift of this kernel intersects S in at least T points. Therefore, $\Pr_h[M(S, h) \geq T] = \delta$ implies that S is $(\delta, u, u - \ell, T)$ -Furstenberg. By [Theorem 2.20](#), this implies that

$$\begin{aligned} |S| &\geq \frac{\delta T}{8} \cdot q^\ell \exp\left(-\frac{q(\ell + 2t)}{T}\right) \\ \implies \delta &\leq \frac{8|S|}{q^\ell T} \exp\left(\frac{q(\ell + 2t)}{T}\right). \quad \square \end{aligned}$$

For linear hashing, we are interested in a uniform random linear map, rather than a uniform surjective linear map. We can generalize the tail bound by showing that linear maps are surjective with high probability. If u is sufficiently large, the probability that h is not surjective is small enough that we can union bound it with the tail.

Theorem 3.3 (General tail bound for large u). *Let q be a prime power, $u \geq \ell \geq 1$ be integers, and $h : \mathbb{F}_q^u \rightarrow \mathbb{F}_q^\ell$ be a uniformly random $\mathbb{F}_q$ -linear map. In addition, suppose that u is sufficiently large, i.e., $u \geq u_0(\ell, |S|, T, q)$. Then for any $T \geq 8q$ and any $S \subseteq \mathbb{F}_q^u$, we have*

$$\Pr[M(S, h) \geq T] \leq \frac{9|S|}{q^\ell T} \exp\left(\frac{8q(\ell + 2\lceil \log_q T \rceil)}{T}\right).$$

Proof. When u is sufficiently large relative to the other parameters, then the probability that h is not surjective is tiny. Thus, the max load tail bound conditioned on surjectiveness extends to this general case with a tiny loss. Let $\mathcal{E}$ be the probability that h is not surjective. Viewing h as an $\mathbb{F}_q^{\ell \times u}$ matrix, $\mathcal{E}$ is equivalent to the probability this matrix is not full rank, which is

$$1 - (1 - q^{-u})(1 - q^{-u+1}) \cdots (1 - q^{-u+\ell-1}) \leq \sum_{i=0}^{\ell-1} q^{-u+i} \leq q^{-u+\ell}.$$

Conditioned on $\neg \mathcal{E}$, h is a random linear surjective map for which [Theorem 3.2](#) applies. Hence,

$$\begin{aligned} \Pr[M(S, h) \geq T] &\leq \Pr[\mathcal{E}] + \Pr[M(S, h) \geq T \mid \neg \mathcal{E}] \\ &\leq q^{-u+\ell} + \frac{8|S|}{q^\ell T} \exp\left(\frac{8q(\ell + 2\lceil \log_q T \rceil)}{T}\right) \\ &\leq \frac{9|S|}{q^\ell T} \exp\left(\frac{8q(\ell + 2\lceil \log_q T \rceil)}{T}\right) \end{aligned}$$

as long as $u_0(\ell, |S|, T, q)$ is set large enough. $\square$

Finally, we can extend this tail bound to every u with an appropriate embedding. If the tail bound holds over large $u' > u$, we can lift our set S and our random hash function h from $\mathbb{F}_q^u$ to $\mathbb{F}_q^{u'}$ while preserving which balls go into which bins (and hence preserving the max load). [Theorem 3.3](#) then applies, and this bound has no dependence on u' so it also applies to our original setting.

Corollary 3.4 (General tail bound for all u). *Let q be a prime power, $u \geq \ell \geq 1$ be integers, and $h : \mathbb{F}_q^u \rightarrow \mathbb{F}_q^\ell$ be a uniformly random $\mathbb{F}_q$ -linear map. Then for any $T \geq 8q$ and any $S \subseteq \mathbb{F}_q^u$, we have*

$$\Pr[M(S, h) \geq T] \leq \frac{9|S|}{q^\ell T} \exp\left(\frac{8q(\ell + 2\lceil \log_q T \rceil)}{T}\right).$$

Proof. If $u \geq u_0(\ell, |S|, T, q)$ from [Theorem 3.3](#), then we are done. Otherwise, pick $u' = u_0(\ell, |S|, T, q)$, and lift S and h to $\mathbb{F}_q^{u'}$: add zeroes to the end of each tuple in S , and add uniformly random columns to the matrix of h , turning it into a uniformly random linear map on a larger universe. [Theorem 3.3](#) gives a tail bound for the larger universe, but the bin assignments (and hence max load) are the same as if we had evaluated h on S in the original, smaller universe. $\square$

3.2 Optimal Expected Max Load

Combining our tail bounds with tail bounds of prior art [[ADM⁺99](#)] allows us to show that linear hashing over $\mathbb{F}_q$ achieve optimal expected max load for any constant q .

Theorem 3.5 (formal [Theorem 1.4](#)). *There is an absolute constant ℓ_0 such that for all integers $\ell \geq \ell_0$, the following holds. For any prime power q , integer $u \geq \ell$, subset $S \subset \mathbb{F}_q^u$ of size $n := q^\ell$, and uniformly random linear map $h : \mathbb{F}_q^u \rightarrow \mathbb{F}_q^\ell$, we have*

$$\mathbb{E}_h[M(S, h)] \leq \frac{10q\ell}{\ln \ell} = O\left(\frac{q \ln \ln q}{\ln q} \cdot \frac{\ln n}{\ln \ln n}\right).$$

Proof. We will write the expectation as the integral of the upper-tail bound.

$$\mathbb{E}_h[M(S, h)] = \int_0^\infty \Pr[M(S, h) \geq t] dt.$$

Evaluate this integral by splitting it into three regions:

- (1) In the first region when t is small, we expect $\Pr[M(S, h) \geq t] = \Omega(1)$. We will thus use a trivial bound of 1 until our tail bound starts to predict tail decay. Let us integrate up to $t \leq 9q\ell/\ln \ell$. This term will dominate the integral.

$$\int_0^{9q\ell/\ln \ell} \Pr[M(S, h) \geq t] dt \leq \int_0^{9q\ell/\ln \ell} 1 dt = \frac{9q\ell}{\ln \ell}.$$

- (2) In the next region, t is now large enough that our tail bound becomes meaningful. However, this integral does not converge, so we will need an appropriate upper limit. This second region bridges the gap until very strong tail bounds hold. [Corollary 3.4](#) applies in this regime if $9\ell/\ln \ell \geq 8$, which is true for $\ell \geq \Omega(1)$.

$$\int_{9q\ell/\ln \ell}^\gamma \Pr[M(S, h) \geq t] dt \leq \int_{9q\ell/\ln \ell}^\gamma \frac{8}{t} \exp\left(\frac{8q(\ell + 2\lceil \log_q t \rceil)}{t}\right) dt.$$

The exponential term is nonincreasing for $t \geq e$, and our integral's limits clearly fall in this regime. So, let us simplify the integral by moving the exponential term out with the lower limit substituted. This term on its own is upper bounded by

$$\exp\left(\frac{8(\ell + 2\log_q(8\ell/\ln \ell) + 4)}{9\ell/\ln \ell}\right) \leq \exp\left(\frac{8\ln \ell}{9} + \frac{16\log_2(8\ell/\ln \ell)}{9\ell/\ln \ell} + \frac{32}{9\ell/\ln \ell}\right) \leq \ell^{9/10},$$

for ℓ larger than an absolute constant. Therefore,

$$\begin{aligned} \int_{9q\ell/\ln \ell}^{\gamma} \Pr[M(S, h) \geq t] dt &\leq 8\ell^{9/10} \int_{9q\ell/\ln \ell}^{\gamma} \frac{1}{t} dt \\ &\leq 8\ell^{9/10} \ln \gamma. \end{aligned}$$

Choose $\gamma := \exp(q\ell^{1/20})$ so that this is upper bounded by $8q\ell^{19/20}$.

- (3) Lastly, once t is large enough, we can utilize the extremely strong tail bound [Corollary B.4](#) from [\[ADM⁺99\]](#). In this region, the integral will easily converge to something vanishing.

$$\begin{aligned} &\int_{\gamma}^{\infty} \Pr[M(S, h) \geq t] dt \\ &\leq \int_W^{\infty} \Pr[M(S, h) \geq wq^{15+\log_2 \log_2 q} \ell^{\log_2 q}] (q^{15} \ell^{\log_2 q}) dw \\ &\leq 2q^{15+\log_2 \log_2 q} \ell^{\log_2 q} \int_W^{\infty} \frac{1}{w^{\log_q w}} dw \\ &\leq 2q^{15+\log_2 \log_2 q} \ell^{\log_2 q} \int_W^{\infty} \frac{1}{w^{\log_q W}} dw \\ &\leq \frac{\exp(O(\ln q(\ln \ell + \log_2 \log_2 q)))}{(\log_q W - 1)W^{\log_q W - 1}}. \end{aligned}$$

Here, $W := \gamma/(q^{15+\log_2 \log_2 q} \ell^{\log_2 q}) = \exp(q\ell^{1/20} - O(\ln q(\ln \ell + \log_2 \log_2 q))) \geq q^{\ell^{1/20} - O(\ln \ell)}$, which gives

$$\leq \frac{q^{O(\ln \ell)}}{q^{(\ell^{1/20} - O(\ln \ell))^2}},$$

which is at most 1 for ℓ larger than an absolute constant.

In total, whenever ℓ is larger than some absolute constant ℓ_0 , the expectation is upper bounded by

$$\mathbf{E}_h[M(S, h)] \leq \frac{9q\ell}{\ln \ell} + 8q\ell^{19/20} + 1 \leq \frac{10q\ell}{\ln \ell}.$$

We can also write this bound in terms of n instead of ℓ as follows:

$$\begin{aligned} \frac{10q\ell}{\ln \ell} &= \frac{10q}{\ln q} \cdot \frac{\ln n}{\ln \ln n - \ln \ln q} \\ &= \frac{10q}{\ln q} \cdot \frac{\ln n}{\ln \ln n} \cdot \frac{\ln(\ell \ln q)}{\ln(\ell \ln q) - \ln \ln q} \end{aligned}$$

$$\begin{aligned}
&= \frac{10q}{\ln q} \cdot \frac{\ln \ell + \ln \ln q}{\ln \ell} \cdot \frac{\ln n}{\ln \ln n} \\
&= O\left(\frac{q \ln \ln q}{\ln q} \cdot \frac{\ln n}{\ln \ln n}\right).
\end{aligned}$$

(When $q = 2$, $\ln \ln q$ is actually negative, but we still recover a $O(\ln n / \ln \ln n)$ upper bound.) $\square$

4 List Decodability of Random $\mathbb{F}_q$ -Linear Codes

An error-correcting code $C \subseteq \mathbb{F}_q^n$ (with relative distance δ , say) has the property that every two codewords differ in at least δn places. Equivalently, every Hamming ball of radius δn contains at most one codeword of C . List decoding generalizes this property to larger Hamming balls.

Definition 4.1. For $0 \leq p < 1 - 1/q$, a (p, L) -list decodable code $C \subseteq \mathbb{F}_q^n$ satisfies the property that every Hamming ball of radius pn contains at most L codewords of C .

Let $H_q(p)$ denote the q -ary entropy function. The *list decoding capacity* $1 - H_q(p)$ is a special rate threshold for when list decodability with constant list size is achievable. We recommend Chapter 7 of Guruswami, Rudra, and Sudan [GRS26] for more details.

Theorem 4.2 (List decoding capacity, see [GRS26, Section 7.4]). *Let $0 \leq p < 1 - 1/q$ and let $\varepsilon > 0$ be sufficiently small. Then for every sufficiently large n ,*

- (1) *there exists a $(p, O(1/\varepsilon))$ -list decodable code with rate $1 - H_q(p) - \varepsilon$, and*
- (2) *every (p, L) -list decodable code with rate $1 - H_q(p) + \varepsilon$ must have $L \geq q^{\Omega(\varepsilon n)}$.*

We now show that random $\mathbb{F}_q$ -linear codes achieve list decoding capacity—in particular, have optimal list size near this rate threshold—for $q = O(1)$, by directly showing list size is a special case of hashing max load. We will define a random linear code (RLC) $C \subseteq \mathbb{F}_q^n$ of rate R as the kernel of a uniformly random parity check matrix of size $(1 - R)n \times n$. Up to an exponentially small probability, this is equivalent to choosing the code as a uniformly random generator matrix $n \times Rn$ or as a uniformly random subspace of dimension Rn . First, show that bounding the list size of a linear code is equivalent to bounding the max load when linearly hashing a Hamming ball. This connection was briefly mentioned in [DLM⁺26], and we formally prove it here.

Lemma 4.3. *Let $h : \mathbb{F}_q^n \rightarrow \mathbb{F}_q^{(1-R)n}$ be a linear map, $C := h^{-1}(0)$, and $B \subset \mathbb{F}_q^n$ be the Hamming ball of radius pn centered at 0. C is (p, L) -list decodable if and only if $M(B, h) \leq L$.*

Proof. The definition of (p, L) -list decodability of C can be written as

$$\max_{x \in \mathbb{F}_q^n} |(x + B) \cap C| \leq L.$$

However, by relativity of motion, this is equivalent to

$$\max_{x \in \mathbb{F}_q^n} |B \cap (x + C)| \leq L.$$

Finally, the left hand side is equal to the max load $M(B, h)$. We can tile $\mathbb{F}_q^n$ with shifts of C , and each shift corresponds to a different nonempty preimage of h . We can extend the quantifier to consider all preimages without changing the maximum value.

$$\max_{x \in \mathbb{F}_q^n} |B \cap (x + C)| = \max_{y \in \mathbb{F}_q^n} |B \cap h^{-1}(y)| \leq L. \quad \square$$

Now that we know list size is equivalent to max load, we can plug the set B into our tail bound.

Theorem 4.4 (formal [Theorem 1.5](#)). *Consider all $0 \leq p < 1 - 1/q$ and all $0 < \varepsilon < 1 - H_q(p)$. Let $C \leq \mathbb{F}_q^n$ be a random linear code of rate $1 - H_q(p) - \varepsilon$, and let*

$$L := 16q \cdot \left(\frac{H_q(p)}{\varepsilon} + 1 \right).$$

If $n \geq 40/\varepsilon \cdot \log_q(1/\varepsilon)$, then C is (p, L) -list decodable with probability $1 - q^{-\Omega(\varepsilon n)}$.

Proof. Let $B \subset \mathbb{F}_q^n$ be the Hamming ball of radius pn centered about zero, so that $|B| \leq q^{H_q(p)n}$, and let $R := 1 - H_q(p) - \varepsilon$. A random linear code C of rate R is defined as the kernel of a uniformly random linear function $h: \mathbb{F}_q^n \rightarrow \mathbb{F}_q^{(1-R)n}$.

We can bound the probability that C is *not* (p, L) -list decodable. By [Lemma 4.3](#), this is equivalent to the max load $M(B, h)$ being larger than L , and we can show this is rare using [Corollary 3.4](#).

$$\begin{aligned} \Pr[C \text{ is not } (p, L)\text{-list decodable}] &= \Pr[M(B, h) > L] \\ &\leq \frac{9|B|}{q^{(1-R)n}L} \cdot \exp\left(\frac{8q((1-R)n + 2\lceil \log_q L \rceil)}{L}\right) \\ &\leq \frac{9q^{H_q(p)n}}{q^{(H_q(p)+\varepsilon)n}L} \cdot \exp\left(\frac{8q((H_q(p) + \varepsilon)n + 2\lceil \log_q L \rceil)}{L}\right) \\ &= \frac{9}{q^{\varepsilon n}L} \cdot \exp\left(\frac{8q((H_q(p) + \varepsilon)n + 2\lceil \log_q L \rceil)}{L}\right). \end{aligned}$$

To control the exponential term, let us set list size to be

$$L = \frac{16q(H_q(p) + \varepsilon)}{\varepsilon} = 16q \cdot \left(\frac{H_q(p)}{\varepsilon} + 1 \right),$$

which is at least $8q$ as needed to apply [Corollary 3.4](#). Then,

$$\begin{aligned} \Pr[C \text{ is not } (p, L)\text{-list decodable}] &\leq \frac{9\varepsilon}{16qH_q(p) + 16q\varepsilon} \exp\left(\frac{16\varepsilon\lceil \log_q L \rceil}{16H_q(p) + 16\varepsilon}\right) \cdot e^{\varepsilon n/2} \cdot q^{-\varepsilon n} \\ &\leq \frac{9\varepsilon}{16q\varepsilon} \exp(\lceil \log_q L \rceil) \cdot q^{-\varepsilon n/4} \\ &\leq \frac{144e^2}{q\varepsilon^2} \cdot q^{-\varepsilon n/4}. \end{aligned}$$

If $n \geq 40/\varepsilon \cdot \log_q(1/\varepsilon)$, then $\varepsilon^2 q^{\varepsilon n/4} \geq q^{\varepsilon n/5}$, so this is upper bounded by

$$\leq \frac{144e^2}{q^{\varepsilon n/5}} = q^{-\Omega(\varepsilon n)}. \quad \square$$

Acknowledgements

VMK thanks Ray Li for various early conversations regarding list decodability and Kakeya sets. We thank Sidhanth Mohanty for many conversations during Summer 2026 and for hosting us at Northwestern. We thank Venkatesan Guruswami for discussions and pointers to literature.

AI involvement. All ideas, analysis, and writing here are due to the authors. No AI was used.

References

- [ADM⁺99] Noga Alon, Martin Dietzfelbinger, Peter Bro Miltersen, Erez Petrank, and Gábor Tardos. Linear hash functions. *J. ACM*, 46(5):667–683, September 1999. [doi:10.1145/324133.324179](#). 4, 5, 7, 11, 18, 21, 22, 33
- [Ash90] Robert B. Ash. *Information Theory*. Courier Corporation, January 1990. Google-Books-ID: yZ1JZA6Wo6YC. 30
- [Bab18] Martin Babka. A note on the size of largest bins using placement with linear transformations. 2018. [arXiv:1810.04161](#). 4, 5
- [BC21] Boris Bukh and Ting-Wei Chao. Sharp density bounds on the finite field Kakeya problem. *Discrete Analysis*, 2021. [doi:10.19086/da.30707](#). 6
- [BD25] Paige Bright and Manik Dhar. Spread Furstenberg sets, 2025. URL: <https://arxiv.org/abs/2412.18193>, [arXiv:2412.18193](#). 6
- [Bsh26] Nader H. Bshouty. A note on second-order expected maximum-load bounds for binary linear hashing, 2026. URL: <https://arxiv.org/abs/2605.18335>, [arXiv:2605.18335](#). 4, 5
- [CGV13] Mahdi Cheraghchi, Venkatesan Guruswami, and Ameya Velingker. Restricted isometry of fourier matrices and list decodability of random linear codes. *SIAM Journal on Computing*, 42(5):1888–1914, 2013. [arXiv:https://doi.org/10.1137/120896773](#), [doi:10.1137/120896773](#). 3
- [CW79] J. Lawrence Carter and Mark N. Wegman. Universal classes of hash functions. *Journal of Computer and System Sciences*, 18(2):143–154, 1979. [doi:10.1016/0022-0000\(79\)90044-8](#). 4, 5
- [Dav71] Roy O. Davies. Some remarks on the Kakeya problem. *Mathematical Proceedings of the Cambridge Philosophical Society*, 69(3):417–421, 1971. [doi:10.1017/S0305004100046867](#). 5
- [DD22] Manik Dhar and Zeev Dvir. Linear hashing with ℓ_∞ guarantees and two-sided Kakeya bounds. In *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 419–428, 2022. [doi:10.1109/FOCS54457.2022.00047](#). 6, 9, 10

- [DDL21] Manik Dhar, Zeev Dvir, and Benjamin Lund. Simple proofs for Furstenberg sets over finite fields. *Discrete Analysis*, oct 14 2021. doi:[10.19086/da.29067](https://doi.org/10.19086/da.29067). 6, 9, 13
- [DDL23] Manik Dhar, Zeev Dvir, and Ben Lund. Furstenberg sets in finite fields: Explaining and improving the Ellenberg–Erman proof. *Discrete Comput. Geom.*, 71(2):327–357, November 2023. doi:[10.1007/s00454-023-00585-y](https://doi.org/10.1007/s00454-023-00585-y). 6, 7
- [DKSS13] Zeev Dvir, Swastik Kopparty, Shubhangi Saraf, and Madhu Sudan. Extensions to the method of multiplicities, with applications to Kakeya sets and mergers. *SIAM Journal on Computing*, 42(6):2305–2328, 2013. arXiv:<https://doi.org/10.1137/100783704>, doi:[10.1137/100783704](https://doi.org/10.1137/100783704). 6, 11, 12, 13
- [DLM⁺26] Dean Doron, Tal Leonov, Jonathan Mosheiff, Henrique Navas, Nicolas Resch, and João Ribeiro. Discrepancy for random linear codes, 2026. URL: <https://arxiv.org/abs/2606.24471>, arXiv:2606.24471. 23
- [DOV22] Damian Dąbrowski, Tuomas Orponen, and Michele Villa. Integrability of orthogonal projections, and applications to Furstenberg sets. *Advances in Mathematics*, 407:108567, 2022. URL: <https://www.sciencedirect.com/science/article/pii/S000187082200384X>, doi:[10.1016/j.aim.2022.108567](https://doi.org/10.1016/j.aim.2022.108567). 6
- [Dvi09] Zeev Dvir. On the size of Kakeya sets in finite fields. *Journal of the American Mathematical Society*, 22(4):1093–1097, 2009. doi:[10.1090/S0894-0347-08-00607-3](https://doi.org/10.1090/S0894-0347-08-00607-3). 5, 6
- [DW11] Zeev Dvir and Avi Wigderson. Kakeya sets, new mergers, and old extractors. *SIAM Journal on Computing*, 40(3):778–792, 2011. arXiv:<https://doi.org/10.1137/090748731>, doi:[10.1137/090748731](https://doi.org/10.1137/090748731). 6
- [EE16] Jordan S. Ellenberg and Daniel Erman. Furstenberg sets and Furstenberg schemes over finite fields. *Algebra and Number Theory*, 10(7):1415–1436, 2016. doi:[10.2140/ant.2016.10.1415](https://doi.org/10.2140/ant.2016.10.1415). 6
- [Eli91] P. Elias. Error-correcting codes for list decoding. *IEEE Transactions on Information Theory*, 37(1):5–12, 1991. doi:[10.1109/18.61123](https://doi.org/10.1109/18.61123). 2, 3
- [EOT10] Jordan S. Ellenberg, Richard Oberlin, and Terence Tao. The Kakeya set and maximal conjectures for algebraic varieties over finite fields. *Mathematika*, 56(1):1–25, 2010. URL: <https://londmathsoc.onlinelibrary.wiley.com/doi/abs/10.1112/S0025579309000400>, arXiv:<https://londmathsoc.onlinelibrary.wiley.com/doi/pdf/10.1112/S0025579309000400>, doi:[10.1112/S0025579309000400](https://doi.org/10.1112/S0025579309000400). 6
- [FM16] Kenneth J. Falconer and Pertti Mattila. Strong Marstrand theorems and dimensions of sets formed by subsets of hyperplanes. *Journal of Fractal Geometry*, 3(4):319–329, 2016. doi:[10.4171/JFG/38](https://doi.org/10.4171/JFG/38). 6

- [GHK10] Venkatesan Guruswami, Johan Hastad, and Swastik Kopparty. On the list-decodability of random linear codes. In *Proceedings of the Forty-Second ACM Symposium on Theory of Computing*, STOC '10, page 409–416, New York, NY, USA, 2010. Association for Computing Machinery. doi:[10.1145/1806689.1806747](https://doi.org/10.1145/1806689.1806747). 3, 8
- [GHSZ02] V. Guruswami, J. Hastad, M. Sudan, and D. Zuckerman. Combinatorial bounds for list decoding. *IEEE Transactions on Information Theory*, 48(5):1021–1034, 2002. doi:[10.1109/18.995539](https://doi.org/10.1109/18.995539). 2
- [GLM⁺22] Venkatesan Guruswami, Ray Li, Jonathan Mosheiff, Nicolas Resch, Shashwat Silas, and Mary Wootters. Bounds for list-decoding and list-recovery of random linear codes. *IEEE Transactions on Information Theory*, 68(2):923–939, 2022. doi:[10.1109/TIT.2021.3127126](https://doi.org/10.1109/TIT.2021.3127126). 2, 3, 8
- [GM22] Venkatesan Guruswami and Jonathan Mosheiff. Punctured low-bias codes behave like random linear codes. In *63rd IEEE Annual Symposium on Foundations of Computer Science, FOCS 2022, Denver, CO, USA, October 31 - November 3, 2022*, pages 36–45. IEEE, 2022. doi:[10.1109/FOCS54457.2022.000011](https://doi.org/10.1109/FOCS54457.2022.000011). 3
- [GRS26] Venkatesan Guruswami, Atri Rudra, and Madhu Sudan. Essential coding theory, 2026. URL: <https://cse.buffalo.edu/faculty/atri/courses/coding-theory/book/>. 23
- [HKM19] Kornélia Héra, Tamás Keleti, and András Máthé. Hausdorff dimension of unions of affine subspaces and of Furstenberg-type sets. *Journal of Fractal Geometry*, 6(3):263–284, 2019. doi:[10.4171/JFG/77](https://doi.org/10.4171/JFG/77). 6
- [Hé19] Kornélia Héra. Hausdorff dimension of Furstenberg-type sets associated to families of affine subspaces. *Annales Fennici Mathematici*, 44(2):903–923, Aug. 2019. URL: <https://afm.journal.fi/article/view/135827>, doi:[10.5186/aasfm.2019.4469](https://doi.org/10.5186/aasfm.2019.4469). 6
- [JKZ25] Michael Jaber, Vinayak M. Kumar, and David Zuckerman. Linear hashing is optimal. In *Proceedings of the 57th Annual ACM Symposium on Theory of Computing*, STOC '25, page 245–255, New York, NY, USA, 2025. Association for Computing Machinery. doi:[10.1145/3717823.3718208](https://doi.org/10.1145/3717823.3718208). 4, 5, 7
- [KLSS11] Swastik Kopparty, Vsevolod F. Lev, Shubhangi Saraf, and Madhu Sudan. Kakeya-type sets in finite vector spaces. *Journal of Algebraic Combinatorics*, 34(3):337–355, 2011. doi:[10.1007/s10801-011-0274-8](https://doi.org/10.1007/s10801-011-0274-8). 6, 9, 13
- [LW18] Ray Li and Mary Wootters. Improved List-Decodability of Random Linear Binary Codes. In Eric Blais, Klaus Jansen, José D. P. Rolim, and David Steurer, editors, *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2018)*, volume 116 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 50:1–50:19, Dagstuhl, Germany, 2018. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. URL: <https://drops.dagstuhl.de/entities/>

[document/10.4230/LIPIcs.APPROX-RANDOM.2018.50](#), [doi:10.4230/LIPIcs.APPROX-RANDOM.2018.50](#). 2

- [MCW78] George Markowsky, J. Lawrence Carter, and Mark N. Wegman. Analysis of a universal class of hash functions. In *Mathematical Foundations of Computer Science 1978*, volume 64 of *Lecture Notes in Computer Science*, pages 345–354. Springer, 1978. [doi:10.1007/3-540-08921-7_82](#). 5
- [MU05] Michael Mitzenmacher and Eli Upfal. *Probability and Computing: Randomized Algorithms and Probabilistic Analysis*. Cambridge University Press, Cambridge, 2005. [doi:10.1017/CBO9780511813603](#). 4
- [MV84] Kurt Mehlhorn and Uzi Vishkin. Randomized and deterministic simulations of prams by parallel machines with restricted granularity of parallel memories. *Acta Informatica*, 21(4):339–374, 1984. [doi:10.1007/BF00264615](#). 5
- [Obe07] Daniel M. Oberlin. Unions of hyperplanes, unions of spheres, and some related estimates. *Illinois Journal of Mathematics*, 51(4):1265 – 1274, 2007. [doi:10.1215/ijm/1258138543](#). 6
- [ORW22] Or Ordentlich, Oded Regev, and Barak Weiss. New bounds on the density of lattice coverings. *J. Amer. Math. Soc.*, 35(1):295–308, 2022. [doi:10.1090/jams/984](#). 6, 10, 16
- [ORW25] Or Ordentlich, Oded Regev, and Barak Weiss. Bounds on the density of smooth lattice coverings. *Journal d’Analyse Mathématique*, 156(1):301–326, 2025. [doi:10.1007/s11854-025-0367-2](#). 6
- [OS26] Tuomas Orponen and Pablo Shmerkin. Projections, Furstenberg sets, and the *abc* sum-product problem, 2026. URL: <https://arxiv.org/abs/2301.10199>, [arXiv:2301.10199](#). 5
- [RW14] Atri Rudra and Mary Wootters. Every list-decodable code for high noise has abundant near-optimal rate puncturings. In *Proceedings of the Forty-Sixth Annual ACM Symposium on Theory of Computing*, STOC ’14, page 764–773, New York, NY, USA, 2014. Association for Computing Machinery. [doi:10.1145/2591796.2591797](#). 3
- [RW18] Atri Rudra and Mary Wootters. Average-radius list-recoverability of random linear codes. In *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA ’18, page 644–662, USA, 2018. Society for Industrial and Applied Mathematics. 3, 8
- [RW25] Kevin Ren and Hong Wang. Furstenberg sets estimate in the plane, 2025. URL: <https://arxiv.org/abs/2308.08819>, [arXiv:2308.08819](#). 5
- [Sar11] Shubhangi Saraf. The method of multiplicities, 2011. URL: <http://hdl.handle.net/1721.1/68494>. 11

- [Sil26] Shashwat Silas. The list size of random linear codes at capacity, 2026. URL: <https://arxiv.org/abs/2609.06570>, arXiv:2609.06570. 8
- [SS08] Shubhangi Saraf and Madhu Sudan. An improved lower bound on the size of Kakeya sets over finite fields. *Analysis & PDE*, 1(3):375–379, 2008. URL: <https://msp.org/apde/2008/1-3/p03.xhtml>, doi:10.2140/apde.2008.1.375. 6
- [Wol99] Thomas Wolff. Recent work connected with the Kakeya problem. In Hugo Rossi, editor, *Prospects in Mathematics*, pages 129–162. American Mathematical Society, Providence, RI, 1999. 5
- [Woo13] Mary Wootters. On the list decodability of random linear codes with large error rates. In *Proceedings of the Forty-Fifth Annual ACM Symposium on Theory of Computing*, STOC '13, page 853–860, New York, NY, USA, 2013. Association for Computing Machinery. doi:10.1145/2488608.2488716. 3, 8
- [WZ25] Hong Wang and Joshua Zahl. Volume estimates for unions of convex sets, and the Kakeya set conjecture in three dimensions, 2025. URL: <https://arxiv.org/abs/2502.17655>, arXiv:2502.17655. 5
- [YZ26] Chen Yuan and Ruiqi Zhu. Asymptotically optimal list size of random linear codes, 2026. URL: <https://arxiv.org/abs/2609.01070>, arXiv:2609.01070. 8
- [ZP81] V. V. Zyablov and M. S. Pinsker. List cascade decoding. *Problems of Information Transmission*, 17(4):236–240, 1981. URL: <https://www.mathnet.ru/eng/ppi1415>. 3

A Upper Bound via Probabilistic Method

When $\delta = 1$, we can show that [Corollary 2.14](#) is almost tight (up to a q factor in the exponential term) using random sets. We begin with a generic upper bound, and then set parameters appropriately in both lower and upper bounds. Any density- α set is a $(1, n, r, \alpha q^r)$ -Furstenberg set by pigeonhole over the shifts of each subspace, so we will hope to beat this using the probabilistic method. We do not attempt whatsoever to optimize the constants here.

Lemma A.1. *For sufficiently large n , let $r < n/2$ and let $S \subseteq \mathbb{F}_q^n$ be drawn i.i.d. randomly with probability $p < \alpha$. For every $\eta < 1$, suppose that*

$$q^r D(\alpha \parallel p) < (n - 3r/2) \ln q - \ln \sqrt{2} - \ln \ln(q^{r(n-r)}/\eta). \quad (7)$$

Then S is a $(1, n, r, \alpha q^r)$ -Furstenberg set with probability at least $1 - \eta$.

Proof. The probability that S is *not* Furstenberg can be estimated by union bounding over each dimension- r subspace V and considering the probability of every shift of V being small in S .

$$\Pr[S \text{ is not Furstenberg}] \leq q^{r(n-r)} (1 - \Pr[|V \cap S| \geq \alpha q^r])^{q^{n-r}}. \quad (8)$$

$|V \cap S|$ is binomially distributed with expectation $pq^r < \alpha q^r$, and its tail satisfies [Ash90, (4.7.2)]

$$\begin{aligned} \Pr[|V \cap S| \geq \alpha q^r] &\geq \frac{\exp(-q^r D(\alpha \parallel p))}{\sqrt{8q^r \alpha (1 - \alpha)}} \\ &\geq \frac{1}{\sqrt{2q^r}} \cdot \exp(-q^r D(\alpha \parallel p)). \end{aligned}$$

Substitute into (8) and upper bound using the exponential inequality.

$$\Pr[S \text{ is not Furstenberg}] \leq q^{r(n-r)} \exp\left(-\frac{q^{n-r}}{\sqrt{2q^r}} \exp(-q^r D(\alpha \parallel p))\right).$$

If p satisfies (7), then this entire expression is strictly less than η . □

Proposition A.2 (Tightness). *Let $\alpha \in (0, 1)$, $q, r \in \mathbb{Z}$, $r < n/2$, and n sufficiently large.*

(1) *If $n = \Omega(\alpha q^{r-1})$, the smallest $(1, n, r, \alpha q^r)$ -Furstenberg set S satisfies*

$$\frac{\alpha}{2} \exp\left(-\frac{4n}{\alpha q^{r-1}}\right) \leq \frac{|S|}{q^n} \leq 3\alpha \exp\left(-\frac{n - O(r)}{\alpha q^r} \cdot \ln q\right).$$

(2) *If $n = o(\alpha q^{r-1})$, the smallest $(1, n, r, \alpha q^r)$ -Furstenberg set S satisfies*

$$\alpha - 3\sqrt{\frac{\alpha n}{q^{r-1}}} \leq \frac{|S|}{q^n} \leq \alpha - \sqrt{\frac{2\alpha(n - O(r)) \ln q}{q^r}} + o(1).$$

Proof. Let us begin by calculating a Furstenberg upper bound using the probabilistic method, and then set parameters appropriately for lower and upper bounds in both regimes.

Large n . When $n = \Omega(\alpha q^{r-1})$, the exponential term of Corollary 2.14 dominates. Set $C = 2$ to get

$$\frac{|S|}{q^n} \geq \frac{\alpha}{2} \cdot \exp\left(-\frac{4n}{\alpha q^{r-1}}\right).$$

For the upper bound, choose p to be

$$\begin{aligned} p &= \alpha(1 - \alpha)^{1/\alpha} \exp\left(-\frac{(n - 3r/2) \ln q - \ln \sqrt{2} - \ln \ln(3q^{r(n-r)})}{\alpha q^r}\right) \\ &\leq \frac{\alpha}{2} \cdot \exp\left(-\frac{n - O(r)}{\alpha q^r} \cdot \ln q\right). \end{aligned}$$

This choice of p satisfies Lemma A.1 for $\eta = 1/3$, using the fact that

$$D(\alpha \parallel p) = \alpha \ln \frac{\alpha}{p} + (1 - \alpha) \ln \frac{1 - \alpha}{1 - p} \geq \alpha \ln \frac{\alpha}{p} - \ln \frac{1}{1 - \alpha}.$$

So, S chosen i.i.d. with probability p will be Furstenberg with probability at least $2/3$. Next, by Markov's inequality, $|S| \leq 3p$ with probability at least $2/3$. Hence, $|S| \leq 3p$ and S is Furstenberg with probability at least $1/3$, so there exists a $(1, n, r, \alpha q^r)$ -Furstenberg set with density at most $3p$.

Small n . When $n = o(\alpha q^{r-1})$, apply the exponential inequality to [Corollary 2.14](#) and balance the first two terms by setting $C = \sqrt{\alpha q^{r-1}/n}$ to get

$$\frac{|S|}{q^n} \geq \alpha - 3\sqrt{\frac{\alpha n}{\delta q^{r-1}}}.$$

For the upper bound, use the approximation

$$D(\alpha \parallel p) \geq \frac{(\alpha - p)^2}{2\alpha}.$$

Then, we can satisfy [Lemma A.1](#) as long as p satisfies

$$\alpha - p \leq \sqrt{\frac{(2\alpha)((n - 3r/2) \ln q - \ln \sqrt{2} - \ln \ln(3q^{r(n-r)}))}{q^r}}.$$

Hence, it suffices to choose

$$p = \alpha - \sqrt{\frac{2\alpha(n - O(r)) \ln q}{q^r}} + o(1),$$

and S chosen i.i.d. with probability p will be Furstenberg with probability at least $2/3$. In addition, via the Chernoff bound, the density of S will be at most $p + O(\sqrt{p/q^n})$ with probability at least $2/3$. So, S will be Furstenberg and also have size

$$|S| \leq \alpha - \sqrt{\frac{2\alpha(n - O(r)) \ln q}{q^r}} + O\left(\sqrt{\frac{\alpha}{q^n}}\right) + o(1). \quad \square$$

B Tail Bound of Alon–Dietzfelbinger–Miltersen–Petrak–Tardos

Lemma B.1. Let α_i for $1 \leq i \leq k$ be random variables and let $0 < \alpha_0 < 1$ be a constant.. Suppose that for $0 \leq i < k$ we have $0 \leq \alpha_{i+1} \leq \alpha_i$ and conditioned on any set of values for $\alpha_1, \dots, \alpha_i$, we have $\mathbf{E}[\alpha_{i+1} \mid \alpha_1, \dots, \alpha_i] \leq \alpha_i^2$. Then for any threshold $0 < t < 1$ we have

$$\Pr[\alpha_k \geq t] \leq \alpha_0^{k - \log_2 \log_2(1/t) + \log_2 \log_2(1/\alpha_0)}.$$

From this, a weak version of coupon collection can be shown.

Lemma B.2. (a) Let $S \subset \mathbb{F}_q^u$ of density $\frac{|S|}{q^u} := 1 - \alpha < 1$ and $0 \leq \ell \leq u$ is an integer, then for uniform random surjective linear map $h : \mathbb{F}_q^u \rightarrow \mathbb{F}_q^\ell$,

$$\Pr[h(S) \neq \mathbb{F}_q^\ell] \leq \alpha^{u - \ell - \log_2 \ell - \log_2 \log_2 q + \log_2 \log_2(1/\alpha)}.$$

(b) There exists constant $c > 0$ such that the following holds. Let S be a subset of a vector space V of arbitrary dimension over $\mathbb{F}_q$. Let $\ell > 0$ be an integer. If $|S| \geq q^{15} \ell^{\log_2 q} q^\ell$, then for a uniform random linear $h : V \rightarrow \mathbb{F}_q^\ell$,

$$\Pr[h(S) = \mathbb{F}_q^\ell] \geq \frac{1}{2}.$$

We note that when $q = 2$, item (b) recovers the coupon collector prediction, but for larger q , the set S will be asymptotically larger (even for $q = 3$). This is the cause behind the max load degradation for larger q .

Proof. (a) Denote $k := u - \ell$. We pick h by picking k vectors $v_1, \dots, v_k$ uniformly at random from $\mathbb{F}_q^u$, and choosing h to be a random surjective linear map conditioned on the kernel of h containing $v_1, \dots, v_k$. Denote $A_0 = A$, $A_{i+1} = A_i + \text{span}\{v_i\}$, $\overline{A}_i = \mathbb{F}_q^u \setminus A_i$, and $\alpha_i := |\overline{A}_i|/q^u$. Observe that

$$\overline{A_{i+1}} = \bigcap_{\gamma \in \text{span}\{v_i\}} (\overline{A}_i + \gamma) \subset \overline{A}_i \cap (\overline{A}_i + v_{i+1}).$$

Hence, conditioned on a choice of $v_1, \dots, v_i$, we can bound

$$\mathbf{E}[\alpha_{i+1}] = \mathbf{Pr}_{x \sim \mathbb{F}_q^u}[x \in \overline{A_{i+1}}] \leq \mathbf{Pr}[x \in \overline{A}_i \cap (\overline{A}_i + v_{i+1})] = \frac{1}{q^u} \sum_{x \in \overline{A}_i} \mathbf{Pr}[x - v \in \overline{A}_i] = \alpha_i^2.$$

Therefore, the α_i satisfy the assumption of [Lemma B.1](#), and thus

$$\mathbf{Pr}[\alpha_k \geq q^{-\ell}] \leq \alpha^{k - \log_2 \log_2(q^\ell) + \log_2 \log_2(1/\alpha)} = \alpha^{u - \ell - \log_2 \ell - \log_2 \log_2 q + \log_2 \log_2(1/\alpha)}.$$

It suffices to show that $\alpha_k < q^{-\ell}$ implies $h(S) = \mathbb{F}_q^\ell$. Now define $B = h(A)$. Since h is surjective, we have $|h^{-1}(B)| = q^{u-\ell}|B|$. Now note that $h(A_k) = h(A + \text{Span}(v_1, \dots, v_k)) = h(A + \ker h) = B$. Therefore $A_k \subset h^{-1}(B)$, and so $|A_k| \leq q^{u-\ell}|B|$. However, the bound on α_k implies $|A_k| > q^u - q^{u-\ell}$. Hence, we must have $|B| > q^\ell - 1$, and thus $|B| = q^\ell$. This implies $h(A) = \mathbb{F}_q^\ell$ as desired.

(b) Let $W = \mathbb{F}_q^t$, where $t = \lfloor \log_q(4m) \rfloor$. We factor h into $h_1 \circ h_0$, where $h_0 : V \rightarrow W$ is a random linear map, and $h_1 : W \rightarrow \mathbb{F}_q^\ell$ is a random surjective linear map. For any fixed distinct $s, t \in S$, h_0 will collide the two with probability $1/|W|$. Therefore, the expected number of collisions caused by h_0 is at most

$$\binom{m}{2} \cdot \frac{1}{|W|} \leq \frac{m^2}{2} \cdot \frac{1}{4m} = \frac{m}{8}.$$

Hence, by Markov's inequality, the number of collisions exceeds $m/2$ with probability at most $1/4$. With that many collisions, it will be that $|h_0(S)| \leq m/2$. Condition on the event that this doesn't happen, and $S_0 := h_0(S)$ is of size $\geq m/2$. Denote $\alpha = 1 - |S_0|/|W|$. We observe

$$\alpha \leq 1 - \frac{m/2}{4m} = \frac{7}{8}.$$

Hence, by part (a), it follows

$$\begin{aligned} \mathbf{Pr}[h_1(S_0) \neq \mathbb{F}_q^\ell] &\leq \left(\frac{7}{8}\right)^{t - \ell - \log_2 \ell - \log_2 \log_2 q + \log_2 \log_2(8/7)} \\ &\leq \exp\left(-\frac{1}{8}(\log(4m) - \ell - \log_2 \ell - \log_2 \log_2 q + \log_2 \log_2(8/7))\right). \end{aligned}$$

Let $C_1 := \log_2 \log_2(8/7)$ and $C_2 := 8 \ln 4$. Hence, if we pick m such that

$$\log_q(4m) - \ell - \log_2 \ell - \log_2 \log_2 q + C_1 = C_2,$$

it follows that $\Pr[h_1(S_0) \neq \mathbb{F}_q^\ell] \leq \exp(-\frac{1}{8} \cdot 8 \ln 4) < 1/2$. This means

$$\begin{aligned} \log_q(4m) &= C_2 - C_1 + \log_2 \ell + \ell + \log_2 \log_2 q \\ \implies m &= \frac{1}{4} q^{C_2 - C_1 + \log_2 \ell + \ell + \log_2 \log_2 q} \leq q^{15 + \log_2 \log_2 q} q^{\log_2 \ell + \ell}. \end{aligned}$$

For such an m , it follows

$$\Pr[h(S) \neq \mathbb{F}_q^\ell] = \Pr[|S_0| < m/2] + \Pr\left[h_1(S_0) \neq \mathbb{F}_q^\ell \mid |S_0| \geq m/2\right] \leq \frac{1}{4} + \frac{1}{4} = \frac{1}{2}. \quad \square$$

With these covering properties, a max load tail bound can be derived.

Theorem B.3. *Let $t \geq 1$, $u \geq \ell \geq 1$ be integers, and let $S \subset \mathbb{F}_q^u$ be an arbitrary subset of size q^ℓ . For uniformly random linear $h : \mathbb{F}_q^u \rightarrow \mathbb{F}_q^\ell$,*

$$\Pr[M(S, h) \geq q^{15 + \log_2 \log_2 q + t + \log_2 t}] < 2q^{-t(t - \log_2 \ell + \log_2 t)}.$$

Proof. Denote $T := q^{15 + \log_2 \log_2 q + \log_2 t} q^t$. We will factor through $\mathbb{F}_q^{\ell+t}$. Let $h_1 : \mathbb{F}_q^u \rightarrow \mathbb{F}_q^{\ell+t}$ be a random linear map, and let $h_2 : \mathbb{F}_q^{\ell+t} \rightarrow \mathbb{F}_q^\ell$ be a random surjective linear map. Denote the event

$$\mathcal{E} = \{\exists y \in \mathbb{F}_q^\ell : h_2^{-1}(y) \subset h_1(S)\}.$$

We first show that $\Pr[\mathcal{E} \mid M(S, h) \geq T] \geq 1/2$. Condition on h and h_2 such that $|h^{-1}(y) \cap S| \geq T$ for some y . Consider the distribution over all h_1 after this conditioning. We will show that the probability $\mathcal{E}$ is satisfied is at least $1/4$, which implies the same lower bound in the weaker conditioning of $M(S, h) \geq T$. Denote $V := h^{-1}(y)$ and $W := h_2^{-1}(y) \cong \mathbb{F}_q^t$. Let $h'_1 = h_1|_V$, which we note is a map taking V to W . By Proposition 3.4 of [ADM⁺99], h'_1 is either a random linear or random affine linear map. Let $S' = S \cap V$. Applying Lemma B.2 to h'_1 and $S' \subset V$, it follows in this conditional distribution,

$$\Pr[\mathcal{E}] \geq \Pr[h_2^{-1}(y) \subset h_1(S)] \geq \Pr[h'_1(S') = \mathbb{F}_q^t] \geq \frac{1}{2}.$$

Therefore, $\Pr[\mathcal{E} \mid M(S, h) \geq T] \geq 1/2$, and so we have

$$\Pr[M(S, h) \geq T] = \frac{\Pr[\mathcal{E}]}{\Pr[\mathcal{E} \mid M(S, h) \geq T]} \leq 2\Pr[\mathcal{E}]. \quad (9)$$

To estimate $\Pr[\mathcal{E}]$, we phrase it as a covering problem. Let $S_1 := \mathbb{F}_q^{\ell+t} \setminus h_1(S)$. We see

$$\mathcal{E} \iff h_2(S_1) \neq \mathbb{F}_q^\ell.$$

Therefore, by Lemma B.2 and the fact that $1 - |S_1|/q^{\ell+t} \leq q^\ell/q^{\ell+t} = q^{-t}$, we have

$$\Pr[\mathcal{E}] \leq q^{-t(t - \log_2 \ell - \log_2 \log_2 q + \log_2 \log_2(q^t))} = q^{-t(t - \log_2 \ell + \log_2 t)}.$$

Combining this with (9) yields the desired result. $\square$

Corollary B.4. *Let $u \geq \ell \geq 1$ be integers, and let $h : \mathbb{F}_q^u \rightarrow \mathbb{F}_q^\ell$ be a random linear map. For arbitrary $S \subset \mathbb{F}_q^u$ of size q^ℓ , we have*

$$\Pr[M(S, h) > wq^{15+\log_2 \log_2 q \ell^{\log_2 q}}] \leq 2w^{-\log_q w}.$$

Proof. Set t to be the smallest integer such that $t + \log_2 t \geq \log_2 \ell + \log_q w + \log_2 \log_q w$. Thus, $t \geq \log_q w$, and plugging this t into [Theorem B.3](#),

$$\Pr[M(S, h) \geq wq^{15+\log_2 \log_2 q \ell^{\log_2 q}}] \leq 2q^{-t \log_q w} = 2w^{-t} \leq 2w^{-\log_q w}. \quad \square$$

It is not hard to see that integrating this tail bound yields the following expectation bound.

Corollary B.5.

$$\mathbf{E}_h[M(S, h)] = O(q^{15+\log_2 \log_2 q \ell^{\log_2 q}}).$$