Electronic Colloquium on Computational Complexity, Report No. 86 (2026)

A Note on Second-Order Expected Maximum-Load Bounds for Binary Linear Hashing

Nader H. Bshouty
Technion

Abstract

Let $S \subseteq \mathbb{F}_2^u$ have size $n = 2^\ell$, and let $h : \mathbb{F}_2^u \to \mathbb{F}_2^\ell$ be a uniformly random linear map. For $y \in \mathbb{F}_2^\ell$, write $\mathrm{load}_h(y) := |h^{-1}(y) \cap S|$, and let $\mathsf{M}(S,h) := \max_{y\in\mathbb{F}_2^\ell} \mathrm{load}_h(y)$ be the maximum load. Jaber, Kumar and Zuckerman (STOC 2025) proved that the expected maximum load of h on S is at most $16 \log n/\log\log n$, matching the fully independent keys-into-bins scale up to constants. Their proof also gives the tail estimate

$$\Pr\left[\mathsf{M}(S,h) \ge R\frac{\log n}{\log\log n}\right] \le O\left(\frac{1}{R^2}\right).$$

We record a base optimization in their exponential-potential method showing that binary linear hashing nearly matches fully independent hashing also at the level of the second-order maximum-load scale. For every $R > 1$ satisfying $R\ell^{1-1/R} \ge D \ln \ell$, where D is an absolute constant, we prove

$$\Pr\left[\mathsf{M}(S,h) \ge R\frac{\log n}{\log\log n}\right] \le O\left(\frac{(\log\log n)^2}{R^2(\log n)^{2-2/R}}\right).$$

Integrating this tail yields

$$\mathbb{E}[\mathsf{M}(S,h)] \le \left(1 + (1+o(1))\frac{\log\log\log n}{\log\log n}\right)\frac{\log n}{\log\log n}.$$

Thus binary linear hashing matches fully independent hashing in the leading term and matches the dominant second-order correction up to a $1 + o(1)$ factor.

We also prove, by an independent self-contained argument, a sharp tail bound for one prescribed bucket: for fixed $y \in \mathbb{F}_2^\ell$,

$$\Pr[\mathrm{load}_h(y) > 2^a - 2] \le \gamma^{-1}2^{-a^2},$$

where $\gamma = \prod_{j\ge1}(1-2^{-j})$. A subspace construction shows that this is asymptotically tight even in the leading constant as $a \to \infty$. However, this controls only a fixed bucket; a direct union bound over all buckets loses a factor 2^ℓ.

ISSN 1433-8092

Contents

1 Introduction

Hashing is often used to distribute a set of keys among bins. If the hash function maps many keys to the same bin, then operations such as lookup with chaining may take a long time. Thus a central quantity is the maximum load, the largest number of keys mapped to any single bin.

In this note we study this question for binary linear hashing. Let $u \geq \ell$, let $h : \mathbb{F}_2^u \to \mathbb{F}_2^\ell$ be a uniformly random linear map, and let $S \subseteq \mathbb{F}_2^u$ have size $n = 2^\ell$. For $y \in \mathbb{F}_2^\ell$, define

$$\mathrm{load}_h(y) := |h^{-1}(y) \cap S|,$$

and define the maximum load

$$\mathsf{M}(S, h) := \max_{y \in \mathbb{F}_2^\ell} \mathrm{load}_h(y).$$

Throughout this paper, all logarithms are base 2. We call the elements of S *keys* and the elements of $\mathbb{F}_2^\ell$ *bins*. For $y \in \mathbb{F}_2^\ell$, the *bucket* at y is the set $h^{-1}(y) \cap S$.

For fully independent hashing of n keys into n bins, the expected maximum load is

$$(1+o(1))\frac{\log n}{\log\log n}.$$

More precisely, the classical second-order asymptotic is

$$\mathbb{E}[M_n] = \frac{\log n}{\log\log n} + (1+o(1))\frac{\log n \cdot \log\log\log n}{(\log\log n)^2}.$$

A uniformly random linear map is much simpler than a fully random function, but it is only pairwise independent. Therefore the fully independent keys-into-bins analysis does not directly apply.

1.1 The expected max-load of a linear map

The maximum-load behavior of linear hashing has been studied for several decades. Alon, Dietzfelbinger, Miltersen, Petrank and Tardos [1] proved that, for binary linear hashing,

$$\mathbb{E}_h[\mathsf{M}(S,h)] = O(\log n \log\log n).$$

They asked whether the fully independent scale $O(\log n/\log\log n)$ is also valid for random binary linear maps.

Jaber, Kumar and Zuckerman [3] then proved the optimal expected maximum-load bound for the binary case. They showed that

$$\mathbb{E}_h[\mathsf{M}(S,h)] \le 16\frac{\log n}{\log\log n}.$$

They also proved the tail estimate

$$\Pr\left[\mathsf{M}(S,h) \ge R\frac{\log n}{\log\log n}\right] \le O\left(\frac{1}{R^2}\right).$$

Their proof uses an exponential-potential method. One advantage of this method is that it directly tracks the growth of bucket loads as the kernel of the linear map is revealed one dimension at a time.

1.2 Our result and comparison with fully independent hashing

We record a small optimization of the Jaber–Kumar–Zuckerman potential argument. Their method uses an exponential potential with a certain base. To detect a bucket of load

$$R\frac{\log n}{\log\log n},$$

it is enough to use base roughly $\ell^{1/R}$, rather than base roughly ℓ. This lowers the initial potential and gives a sharper tail bound.

Our main tail estimate is the following. There are absolute constants $C, D > 0$ such that, for every $R > 1$ satisfying

$$R\ell^{1-1/R} \ge D\ln\ell,$$

we have

$$\Pr\left[\mathsf{M}(S,h) \ge R\frac{\log n}{\log\log n}\right] \le C\frac{(\log\log n)^2}{R^2(\log n)^{2-2/R}}.$$

This improves the $O(R^{-2})$ tail bound in the range where R is fixed, and it remains useful even when R is very close to 1.

Integrating the optimized tail gives

$$\mathbb{E}_h[\mathsf{M}(S,h)] \le \frac{\log n}{\log\log n} + (1+o(1))\frac{\log n \cdot \log\log\log n}{(\log\log n)^2}.$$

Thus binary linear hashing matches fully independent hashing in the leading term and also matches the dominant part of the second-order correction.

1.3 A sharp tail bound for one prescribed bucket

We also prove a separate fixed-bucket result. This part is independent of the potential method.

Fix one bucket $y \in \mathbb{F}_2^\ell$. We prove that

$$\Pr[\mathrm{load}_h(y) > 2^a] \le \gamma^{-1}2^{-a^2}, \qquad \gamma := \prod_{j=1}^{\infty}(1-2^{-j}).$$

The proof is elementary. If a prescribed bucket is large, then it contains many ordered linearly independent tuples. On the other hand, a random linear map sends any fixed independent tuple to that prescribed bucket with very small probability. Counting such tuples and applying Markov's inequality gives the tail bound.

This fixed-bucket bound is essentially sharp. A subspace example shows that the upper bound is asymptotically tight even in the leading constant as $a \to \infty$. Thus the 2^{-a^2} fixed-bucket tail is the correct scale.

However, this fixed-bucket theorem does not by itself imply the correct expectation bound for the maximum load. A direct union bound over all 2^ℓ buckets loses a factor 2^ℓ. Therefore the fixed-bucket estimate is best viewed as a sharp far-tail result for one prescribed bucket, while the optimized potential method is needed to control the maximum load near the expectation scale.

1.4 The case of m keys and n bins

The same base-optimized argument also applies when the number of keys is not equal to the number of bins. Let the number of bins be $n = 2^\ell$, let $|S| = m$, and write

$$\lambda := \frac{m}{n}$$

for the average load. We consider the regime where the maximum-load scale is much larger than the average load. For fully independent hashing, this scale is the value $t = t(m,n)$ defined by

$$t \ln\left(\frac{t}{e\lambda}\right) = \ln n.$$

Thus the regime we consider is $t/\lambda \to \infty$, meaning that the largest bucket is expected to be much larger than the average bucket.

Our method gives the following analogue of the $m = n$ result. There are absolute constants $C, D > 0$ such that, for every threshold T satisfying $T \ge D\lambda n^{1/T}$, one has

$$\Pr_h[\mathsf{M}(S,h) \ge T] \le C\left(\frac{\lambda n^{1/T}}{T}\right)^2.$$

Taking $T = Rt$, this becomes

$$\Pr_h[\mathsf{M}(S,h) \geq Rt] \leq \frac{C}{R^2}\left(\frac{\lambda}{t}\right)^{2-2/R},$$

for every $R > 1$ satisfying

$$R\left(\frac{t}{\lambda}\right)^{1-1/R} \geq D.$$

Thus, when $t/\lambda \to \infty$, the tail bound is stronger than the constant-factor $O(R^{-2})$ tail: for fixed $R > 1$, the extra factor

$$\left(\frac{\lambda}{t}\right)^{2-2/R}$$

tends to zero.

Integrating the tail gives

$$\mathbb{E}_h[\mathsf{M}(S,h)] \leq (1+o(1))t$$

whenever $t/\lambda \to \infty$. Thus, in the sparse large-load regime, binary linear hashing matches the fully independent expected maximum-load scale up to a $1 + o(1)$ factor.

1.5 Fixed buckets when the number of keys is arbitrary

The fixed-bucket estimate also extends to the case of m keys and n bins. Let $n = 2^\ell$, let $S \subseteq \mathbb{F}_2^d \setminus \{0\}$ be a set of m distinct nonzero vectors, and let $h : \mathbb{F}_2^d \to \mathbb{F}_2^\ell$ be a uniformly random linear map. For a fixed bucket $y \in \mathbb{F}_2^\ell$, write $Z_y = |\{x \in S : h(x) = y\}|$, and $\lambda := m/n$. Then, for every integer $r \geq 0$, with $a = \lceil \log(r+2) \rceil$, we prove

$$\Pr[Z_y > r] \leq \lambda^a \left(\prod_{j=0}^{a-1}(r + 2 - 2^j)\right)^{-1}.$$

In particular, for thresholds of the form $2^a - 2$,

$$\Pr[Z_y > 2^a - 2] \leq \gamma^{-1}\lambda^a 2^{-a^2}, \qquad \gamma := \prod_{j=1}^{\infty}(1 - 2^{-j}).$$

This recovers the balanced fixed-bucket estimate when $m = n$, since then $\lambda = 1$. For general m, the extra factor λ^a reflects the average load of the prescribed bucket. We also show that this dependence is sharp up to absolute constants: when $m = 2^d$, a subspace construction gives a matching lower bound of order $\lambda^a 2^{-a^2}$ for the zero bucket, in the natural range $a \geq d - \ell$. Thus the fixed-bucket tail has the correct dependence on both the threshold and the average load.

1.6 Dense two-sided bounds

There is another line of work on binary linear hashing in the dense regime, where the number of keys is larger than the number of bins. This asks for a two-sided guarantee: not only should no bucket be too large, but no bucket should be too small.

Dhar and Dvir [5] proved strong ℓ_∞-type guarantees for random linear maps using connections to finite-field Kakeya and the polynomial method. In a related dense setting, their result shows that random linear maps can distribute a large set nearly as well as fully independent hashing, up to constant factors in the relevant parameters.

Jaber, Kumar and Zuckerman [3] also prove a dense two-sided theorem. In their Section 6, they show that if the number of keys m is at least on the order of $n \log n$, then with high probability every bucket has load within constant factors of the average load m/n. More precisely, for every $0 < \varepsilon < 1/2$, there are constants $C_1 < C_2$, depending on ε, such that if $m \geq C_1^{-1} n \log n$, then

$$\Pr_h \left[\forall y \in \mathbb{F}_2^\ell, \quad C_1 \frac{m}{n} \leq |h^{-1}(y) \cap S| \leq C_2 \frac{m}{n} \right] \geq 1 - \varepsilon.$$

Thus in the dense regime all buckets are balanced up to constant factors.

The results in this note are complementary to these dense two-sided bounds. Our base optimization improves the maximum-load tail in the sparse large-load regime, where the relevant threshold is much larger than the average load. It does not improve the dense two-sided theorem above, whose goal is to control all buckets at the scale m/n, including the lower tail. Similarly, the fixed-bucket estimates proved here give sharp far upper-tail bounds for one prescribed bucket, but they do not address the main lower-tail difficulty in the dense two-sided problem.

1.7 Organization

The rest of the paper is organized as follows. Section 2 proves the fixed-bucket tail bound. We first give the upper bound for one prescribed bucket, and then show that the 2^{-a^2} behavior is sharp by using a subspace construction.

Section 3 proves the expected maximum-load bound. We recall the Jaber–Kumar–Zuckerman potential framework, optimize the base in the potential argument, remove the surjectivity assumption, and integrate the resulting tail bound. We also compare the fixed-bucket and potential methods.

The appendices contain the auxiliary and extended results. Appendix A gives self-contained proofs of the potential lemmas used in Section 3. Appendix B extends the fixed-bucket tail estimate to m keys and n bins. Appendix C extends the maximum-load bound to m keys and n bins in the regime where the fully independent maximum-load scale is much larger than the average load.

2 A fixed-bucket tail bound

We begin with a fixed-bucket estimate, which is independent of the potential method used later for the maximum load. The goal is to understand the load of one prescribed bucket y, rather than the maximum over all buckets. First we prove a uniform upper bound showing that, for every fixed y, the probability that $\text{load}_h(y)$ exceeds a dyadic threshold 2^a is at most on the order of 2^{-a^2}. The proof uses a simple counting idea: a heavy bucket must contain many linearly independent tuples, while a random linear map sends any fixed independent tuple to the prescribed bucket with probability 2^{-ma}.

We then show that this 2^{-a^2} scale is sharp. The lower-bound example takes the input set to be essentially an m-dimensional subspace. In that case the bucket load is governed by the nullity of a random $m \times m$ binary matrix, whose distribution has exactly the same 2^{-a^2} behavior. In fact, the construction nearly matches the leading constant in the dyadic upper bound as $a \to \infty$.

This section is independent of the potential argument used later.

2.1 Upper bound for a prescribed bucket

We first prove the upper bound for one fixed bucket. The idea is simple: if the bucket contains many keys, then it must contain many ordered linearly independent tuples. But any fixed independent tuple is sent to the prescribed bucket with very small probability. We count such tuples and then apply Markov's inequality.

Lemma 1 (Independent tuples in a set of distinct nonzero vectors)**.** *Let $A \subseteq \mathbb{F}_2^n \setminus \{0\}$ have size q. Let $a = \lceil \log(q+1) \rceil$. Then A contains at least*

$$\prod_{j=0}^{a-1} (q + 1 - 2^j)$$

ordered linearly independent a-tuples.

Proof. Choose the tuple sequentially. Suppose that $v_1, \ldots, v_j \in A$ have already been chosen and are linearly independent. Their span contains 2^j vectors, of which at most $2^j - 1$ are nonzero vectors of A. Hence the number of choices for $v_{j+1} \in A$ outside $\operatorname{span}(v_1, \ldots, v_j)$ is at least $q - (2^j - 1) = q + 1 - 2^j$. Since $j < a$, this quantity is positive. Multiplying over $j = 0, 1, \ldots, a-1$ proves the lemma. □

Theorem 2 (Fixed-bucket tail)**.** *Let $U = \{u_1, \ldots, u_{2^m}\} \subseteq \mathbb{F}_2^n \setminus \{0\}$ be a set of distinct nonzero vectors, and let $B : \mathbb{F}_2^n \to \mathbb{F}_2^m$ be a uniformly random linear map. Fix $y \in \mathbb{F}_2^m$, and define*

$$Z_y := |\{i : Bu_i = y\}|.$$

Then, for every integer $r \geq 0$, if $a = \lceil \log(r+2) \rceil$, we have

$$\Pr[Z_y > r] \leq \left(\prod_{j=0}^{a-1} (r + 2 - 2^j) \right)^{-1}.$$

Proof. Let $q = r + 1$, and $a = \lceil \log(q+1) \rceil$. Let $\mathcal{I}_a$ be the set of all ordered a-tuples $(i_1, \ldots, i_a)$ such that $u_{i_1}, \ldots, u_{i_a}$ are linearly independent. For a linear map B, define

$$T_a(B) := \sum_{(i_1, \ldots, i_a) \in \mathcal{I}_a} \mathbf{1}[Bu_{i_1} = \cdots = Bu_{i_a} = y].$$

Thus $T_a(B)$ is the number of ordered linearly independent a-tuples $(u_{i_1}, \ldots, u_{i_a})$ such that $Bu_{i_1} = \cdots = Bu_{i_a} = y$.

For a fixed ordered linearly independent a-tuple, the random vectors $Bu_{i_1}, \ldots, Bu_{i_a}$ are independent and uniformly distributed in $\mathbb{F}_2^m$.[1] Therefore

$$\Pr_B[Bu_{i_1} = \cdots = Bu_{i_a} = y] = 2^{-ma}.$$

Taking expectation over the random choice of B, linearity of expectation gives

$$\begin{aligned} \mathbb{E}_B[T_a(B)] &= \sum_{(i_1, \ldots, i_a) \in \mathcal{I}_a} \Pr_B[Bu_{i_1} = \cdots = Bu_{i_a} = y] \\ &= |\mathcal{I}_a| \, 2^{-ma}. \end{aligned}$$

[1] Indeed, after fixing bases of $\mathbb{F}_2^n$ and $\mathbb{F}_2^m$, the map B is represented by a uniformly random $m \times n$ binary matrix. Since $u_{i_1}, \ldots, u_{i_a}$ are linearly independent, the random vectors $Bu_{i_1}, \ldots, Bu_{i_a}$ are independent uniform elements of $\mathbb{F}_2^m$.

Since $|U| = 2^m$, the total number of ordered a-tuples from U is $(2^m)^a = 2^{ma}$. Hence $|\mathcal{I}_a| \le 2^{ma}$, and therefore

$$\mathbb{E}_B[T_a(B)] \le 2^{ma} 2^{-ma} = 1.$$

If $Z_y \ge q$, then the set $A := \{u_i : Bu_i = y\}$ has size at least q. By Lemma 1, it contains at least

$$M(q) := \prod_{j=0}^{a-1} (q + 1 - 2^j)$$

ordered linearly independent a-tuples. Thus

$$Z_y \ge q \implies T_a(B) \ge M(q).$$

By Markov's inequality,

$$\Pr[Z_y \ge q] \le \Pr[T_a(B) \ge M(q)] \le \frac{\mathbb{E}_B[T_a(B)]}{M(q)} \le \frac{1}{M(q)}.$$

Since $q = r + 1$, this is the desired bound. □

Let

$$\gamma := \prod_{j=1}^{\infty} (1 - 2^{-j}).$$

The preceding theorem has a clean form at dyadic thresholds.

Corollary 3 (Dyadic fixed-bucket thresholds)**.** *Under the assumptions of Theorem 2, for every integer $a \ge 1$,*

$$\Pr[Z_y > 2^a - 2] \le \gamma^{-1} 2^{-a^2}.$$

Proof. Apply Theorem 2 with $r = 2^a - 2$. Then

$$\prod_{j=0}^{a-1} (r + 2 - 2^j) = \prod_{j=0}^{a-1} (2^a - 2^j) = 2^{a^2} \prod_{j=0}^{a-1} (1 - 2^{j-a}) = 2^{a^2} \prod_{s=1}^{a} (1 - 2^{-s}).$$

Since $\prod_{s=1}^{a} (1 - 2^{-s}) \ge \gamma$, the result follows. □

2.2 A matching lower bound for the fixed-bucket tail

We next show that the fixed-bucket upper bound is essentially best possible. The example is based on a subspace. In this case, the load of the zero bucket is controlled by the nullity of a random binary matrix, and the nullity distribution has the same 2^{-a^2} behavior as the upper bound.

Proposition 4 (Near-sharpness of the fixed-bucket bound)**.** *Let $\gamma := \prod_{j=1}^{\infty} (1 - 2^{-j})$. For every integer $a \ge 1$, there is a sequence of examples with $|U| = 2^m$, $m \to \infty$, such that, for a uniformly random linear map $B : \mathbb{F}_2^n \to \mathbb{F}_2^m$, one has*

$$\liminf_{m \to \infty} \Pr\left[|\{u \in U : Bu = 0\}| > 2^a - 2\right] \ge \gamma^{-1} 2^{-a^2} (1 - 2^{-a})^2.$$

In particular, the upper bound

$$\Pr[Z_0 > 2^a - 2] \le \gamma^{-1} 2^{-a^2}$$

from Corollary 3 is asymptotically tight in the leading constant as $a \to \infty$.

Proof. Let $W < \mathbb{F}_2^n$ be an m-dimensional subspace and choose $v \notin W$. Let $U := (W \setminus \{0\}) \cup \{v\}$. Then $|U| = (2^m - 1) + 1 = 2^m$, and all elements of U are distinct and nonzero.

Let M be the restriction of B to W:

$$M := B|_W : W \to \mathbb{F}_2^m.$$

Since $\dim W = m$, after choosing a basis of W, the map M is represented by an $m \times m$ binary matrix. As B is uniformly random, M is a uniformly random linear map from W to $\mathbb{F}_2^m$, equivalently a uniformly random $m \times m$ binary matrix.

If[2] $\mathrm{nul}(M) \geq a$, then

$$|(W \setminus \{0\}) \cap \ker B| = 2^{\mathrm{nul}(M)} - 1 \geq 2^a - 1.$$

Therefore

$$|\{u \in U : Bu = 0\}| > 2^a - 2.$$

Consequently,

$$\Pr\left[|\{u \in U : Bu = 0\}| > 2^a - 2\right] \geq \Pr[\mathrm{nul}(M) \geq a] \geq \Pr[\mathrm{nul}(M) = a].$$

We use the standard rank distribution formula for random matrices over finite fields; see, for example, Fulman and Goldstein [7]. For a uniformly random $m \times m$ matrix over $\mathbb{F}_2$, the probability of nullity exactly a, equivalently rank $m - a$, is

$$p_{m,a} = 2^{-a^2} \frac{\prod_{j=a+1}^{m}(1 - 2^{-j})^2}{\prod_{j=1}^{m-a}(1 - 2^{-j})}.$$

For fixed a, as $m \to \infty$,

$$p_{m,a} \longrightarrow 2^{-a^2} \frac{\prod_{j=a+1}^{\infty}(1 - 2^{-j})^2}{\gamma}.$$

Using the elementary inequality $\prod_j (1 - x_j) \geq 1 - \sum_j x_j$ for $0 \leq x_j \leq 1$, we get

$$\prod_{j=a+1}^{\infty} (1 - 2^{-j}) \geq 1 - \sum_{j=a+1}^{\infty} 2^{-j} = 1 - 2^{-a}.$$

Hence

$$\liminf_{m \to \infty} \Pr\left[|\{u \in U : Bu = 0\}| > 2^a - 2\right] \geq \gamma^{-1} 2^{-a^2} (1 - 2^{-a})^2.$$

Since $(1 - 2^{-a})^2 \to 1$ as $a \to \infty$, this matches the leading constant γ^{-1} in the dyadic fixed-bucket upper bound. □

Remark 5 (Why the fixed-bucket bound is not enough for the expectation)**.** *The fixed-bucket bound controls the load of one prescribed bucket. To control the maximum load, one could try to apply it to all 2^ℓ buckets and then take a union bound. This gives*

$$\Pr[\mathsf{M}(S, h) > 2^a - 2] \leq \gamma^{-1} 2^{\ell - a^2}.$$

This bound becomes useful only when a^2 is at least comparable to ℓ. Equivalently, it only controls very large loads, roughly of size $2^{\sqrt{\ell}}$ or larger.

However, the expected maximum load is much smaller: it is of order $\ell / \log \ell$. The logarithm of this load is only $O(\log \ell)$, far below $\sqrt{\ell}$. Therefore the fixed-bucket estimate is useful for the far tail, but it does not by itself prove the correct expectation bound. For the expectation-scale bound, we need the potential method used later.

[2] Here $\mathrm{nul}(M) := \dim \ker M = m - \mathrm{rank}(M)$, since $\dim W = m$.

3 The Expected Maximum-Load Bound

In this section we prove the main maximum-load estimate. We first recall the potential framework of Jaber–Kumar–Zuckerman for uniformly random surjective linear maps. We then optimize the choice of the potential base to get a sharper tail bound. Finally, we remove the surjectivity assumption and integrate the tail to obtain the second-order expectation bound.

3.1 The Jaber–Kumar–Zuckerman potential framework

We now recall the potential framework used by Jaber, Kumar and Zuckerman. We first work with uniformly random surjective maps.

Let $H : \mathbb{F}_2^U \to \mathbb{F}_2^\ell$ be a uniformly random surjective linear map, and let $k := U - \ell$. A uniformly random surjective map can be sampled by first choosing a uniformly random k-dimensional kernel $V \leq \mathbb{F}_2^U$ and then choosing an isomorphism $\mathbb{F}_2^U/V \cong \mathbb{F}_2^\ell$. We build the kernel gradually through a chain

$$V_0 \leq V_1 \leq \cdots \leq V_k = V,$$

where $V_0 = \{0\}$ and $\dim V_i = i$. Thus, at step i, the subspace V_i is the i-dimensional part of the kernel that has been revealed.

Let $S \subseteq \mathbb{F}_2^U$ have size 2^ℓ. For $x \in \mathbb{F}_2^U$, define

$$S_i(x) := |(x + V_i) \cap S|.$$

The quantity $S_i(x)$ is the load of the partial bucket $x + V_i$. At $i = 0$, the buckets are singletons, so $S_0(x) = 1_S(x)$. At the final step, $V_k = \ker H$, and the cosets $x + V_k = x + \ker H$ are exactly the final buckets of the hash map H: all points in the same coset have the same image under H. Thus $S_k(x)$ is the actual load of the final bucket containing x.

Equivalently, let

$$G_i := \mathbb{F}_2^U/V_i$$

be the quotient group of cosets of V_i, and define $f_i : G_i \to \mathbb{Z}_{\geq 0}$ by

$$f_i(x + V_i) := |(x + V_i) \cap S|.$$

Then

$$S_i(x) = f_i(x + V_i).$$

For a base $b > 1$, define the exponential potential

$$\Phi_i := \mathbb{E}_{x \in \mathbb{F}_2^U}\left[b^{S_i(x)}\right].$$

Since all cosets of V_i have the same size and S_i is constant on each coset, this is equivalently

$$\Phi_i = \mathbb{E}_{C \in G_i}\left[b^{f_i(C)}\right].$$

We use the following two lemmas from the potential analysis of Jaber, Kumar and Zuckerman. Lemma 6 packages JKZ Lemmas 3 and 4 in the normalization used here. For completeness, see the proof in Appendix A.

Lemma 6 (JKZ potential evolution)**.** *For every $b > 1$, the potentials satisfy*

$$\mathbb{E}[\Phi_{i+1} \mid \Phi_0, \ldots, \Phi_i] \le \Phi_i^2$$

and

$$\Phi_{i+1} - 1 \ge 2(\Phi_i - 1)$$

for every $0 \le i < k$.

The next lemma is a slightly stronger version of the quadratic tail lemma needed for the argument. It says that if a nonnegative process grows at least multiplicatively in the sense $X_{i+1} - 1 \ge 2(X_i - 1)$, but its conditional expectation is at most quadratic, then the final value has a polynomial tail. We include a self-contained proof in Appendix A.

Lemma 7 (Strengthened JKZ quadratic potential tail lemma)**.** *Let $X_0 > 1$ be deterministic, and let $X_1, \ldots, X_k$ be nonnegative random variables satisfying*

$$X_{i+1} - 1 \ge 2(X_i - 1)$$

and

$$\mathbb{E}[X_{i+1} \mid X_0, \ldots, X_i] \le X_i^2$$

for every $0 \le i < k$. If $\tau \ge 1 + 4(X_0 - 1)$, then

$$\Pr\left[X_k \ge \tau^{2^k}\right] \le 48 \left(\frac{X_0 - 1}{\tau - 1}\right)^2.$$

For $y \in \mathbb{F}_2^\ell$, the *fiber* of H over y is the set

$$H^{-1}(y) = \{x \in \mathbb{F}_2^U : H(x) = y\}.$$

Since H is surjective and $\ker H = V_k$, every fiber is a coset of V_k.

We also need the following elementary observation.

Lemma 8 (Heavy bin implies large potential)**.** *If some fiber of the surjective map H contains at least T elements of S, then*

$$\Phi_k \ge \frac{b^T}{2^\ell}.$$

Proof. A fiber of H is a coset $x + V_k$. If $|(x + V_k) \cap S| \ge T$, then for every $z \in x + V_k$,

$$S_k(z) = |(z + V_k) \cap S| = |(x + V_k) \cap S| \ge T.$$

The coset has size $|V_k| = 2^k$. Therefore, its contribution to the average Φ_k is at least

$$\frac{2^k b^T}{2^U} = \frac{b^T}{2^{U-k}} = \frac{b^T}{2^\ell}.$$

□

Finally, since $V_0 = \{0\}$, we have

$$S_0(x) = 1_S(x).$$

Therefore

$$\Phi_0 = \left(1 - \frac{|S|}{2^U}\right) + \frac{|S|}{2^U} b.$$

Using $|S| = 2^\ell$ and $k = U - \ell$, this gives

$$\Phi_0 - 1 = \frac{b-1}{2^k} \le \frac{b}{2^k}. \tag{1}$$

3.2 The optimized tail bound for surjective maps

We first prove the optimized tail bound in the simpler setting where the linear map is conditioned to be surjective. The proof follows the Jaber–Kumar–Zuckerman potential argument, but chooses the base according to the target load level. This smaller base lowers the initial potential and gives the extra factor in the tail bound.

We now optimize the base in the potential argument.

Proposition 9 (Surjective base-optimized tail)**.** *There exist absolute constants $C_0 > 0$, $D_0 > 0$, and ℓ_0 such that the following holds. Let $U \geq \ell$, let $S \subseteq \mathbb{F}_2^U$ have size 2^ℓ, and let $H : \mathbb{F}_2^U \to \mathbb{F}_2^\ell$ be a uniformly random surjective linear map. Then, for every real $R > 1$ satisfying*

$$R\ell^{1-1/R} \geq D_0 \ln \ell$$

and every $\ell \geq \ell_0$,

$$\Pr_H \left[\mathsf{M}(S, H) \geq R \frac{\ell}{\log \ell} \right] \leq C_0 \frac{(\ln \ell)^2}{R^2 \ell^{2-2/R}}.$$

Proof. Let $k := U - \ell$, $L := \ell / \log \ell$, and $T := RL = R\ell / \log \ell$. Choose

$$\alpha := \frac{1}{R} + \frac{1}{\ln \ell}$$

and set $b := \ell^\alpha$. Then $b = \ell^{1/R} \ell^{1/\ln \ell} = e\ell^{1/R}$. If $\mathsf{M}(S, H) \geq T$, Lemma 8 gives $\Phi_k \geq b^T / 2^\ell$. Since $\log b = \alpha \log \ell$ and $T = R\ell / \log \ell$, we have $b^T = 2^{\alpha R \ell}$. Thus

$$\frac{b^T}{2^\ell} = 2^{(\alpha R - 1)\ell}.$$

Let

$$A := \alpha R - 1 = \frac{R}{\ln \ell}.$$

Hence

$$\mathsf{M}(S, H) \geq T \implies \Phi_k \geq 2^{A\ell}. \tag{2}$$

Define

$$\tau := 1 + \frac{(\ln 2) A \ell}{2^k}.$$

Then, using $1 + x \leq e^x$, with $x = \tau - 1$,

$$\tau^{2^k} \leq \exp\left(2^k (\tau - 1)\right) = \exp\left((\ln 2) A \ell\right) = 2^{A\ell}.$$

Therefore, by (2),

$$\mathsf{M}(S, H) \geq T \implies \Phi_k \geq \tau^{2^k}.$$

Choose $D_0 \geq 4e / \ln 2$. Using (1), the definition of τ, and the hypothesis $R\ell^{1-1/R} \geq D_0 \ln \ell$, we get

$$\begin{aligned} 4(\Phi_0 - 1) &\leq \frac{4b}{2^k} = \frac{4e\ell^{1/R}}{2^k} \\ &\leq \frac{(\ln 2) R \ell}{2^k \ln \ell} = \frac{(\ln 2) A \ell}{2^k} = \tau - 1. \end{aligned}$$

Thus $1 + 4(\Phi_0 - 1) \le \tau$.

Applying Lemma 7 with $X_i = \Phi_i$, we get

$$\Pr[\mathsf{M}(S, H) \ge T] \le \Pr[\Phi_k \ge \tau^{2^k}] \le 48 \left(\frac{\Phi_0 - 1}{\tau - 1}\right)^2.$$

Using (1) and the definition of τ, and substituting $b = e\ell^{1/R}$, and $A = R/\ln \ell$,

$$\frac{\Phi_0 - 1}{\tau - 1} \le \frac{b/2^k}{(\ln 2)A\ell/2^k} = \frac{b}{(\ln 2)A\ell} = \frac{e \ln \ell}{(\ln 2)R\ell^{1-1/R}}.$$

Therefore

$$\Pr[\mathsf{M}(S, H) \ge T] \le C_0 \frac{(\ln \ell)^2}{R^2 \ell^{2-2/R}}.$$

This proves the claimed bound with $C_0 = 48\,(e/\ln 2)^2$. □

3.3 Removing the surjectivity assumption

We now pass from uniformly random surjective linear maps to uniformly random linear maps. The idea is to embed the original space into a larger space. A random map from the larger space is surjective with very high probability, and its restriction to the original space is still a uniformly random linear map.

Lemma 10 (Rank deficiency)**.** *Let $H : \mathbb{F}_2^U \to \mathbb{F}_2^\ell$ be a uniformly random linear map. Then*

$$\Pr[H \text{ is not surjective}] \le 2^{\ell - U}.$$

Proof. After fixing bases of $\mathbb{F}_2^U$ and $\mathbb{F}_2^\ell$, the map H is represented by a uniformly random $\ell \times U$ binary matrix. The map H is not surjective exactly when this matrix has rank less than ℓ.

The probability that a random $\ell \times U$ binary matrix has full row rank is

$$\prod_{j=0}^{\ell-1} (1 - 2^{j-U}).$$

Therefore

$$\Pr[H \text{ is not surjective}] = 1 - \prod_{j=0}^{\ell-1}(1 - 2^{j-U}) \le \sum_{j=0}^{\ell-1} 2^{j-U} = 2^{-U}(2^\ell - 1) \le 2^{\ell - U}.$$

□

Theorem 11 (Base-optimized maximum-load tail)**.** *There exist absolute constants $C > 0$, $D > 0$, and ℓ_0 such that the following holds. Let $u \ge \ell$, $n := 2^\ell$, and let $S \subseteq \mathbb{F}_2^u$ have size n. Let $h : \mathbb{F}_2^u \to \mathbb{F}_2^\ell$ be a uniformly random linear map. Then, for every real $R > 1$ satisfying*

$$R\ell^{1-1/R} \ge D \ln \ell$$

and every $\ell \ge \ell_0$,

$$\Pr_h \left[\mathsf{M}(S, h) \ge R \frac{\ell}{\log \ell}\right] \le C \frac{(\ln \ell)^2}{R^2 \ell^{2-2/R}}.$$

Equivalently,

$$\Pr_h \left[\mathsf{M}(S,h) \geq R \frac{\log n}{\log \log n} \right] \leq C \frac{(\log \log n)^2}{R^2 (\log n)^{2-2/R}},$$

for every $R > 1$ satisfying

$$R(\log n)^{1-1/R} \geq D \log \log n,$$

where changing the base of the logarithms only changes the absolute constants.

Proof. Let

$$p_{\ell,R} := C_0 \frac{(\ln \ell)^2}{R^2 \ell^{2-2/R}},$$

where C_0 is the constant from Proposition 9. Choose $D \geq \max(D_0, \sqrt{C_0})$. Then, whenever $R\ell^{1-1/R} \geq D \ln \ell$, we have

$$p_{\ell,R} = C_0 \frac{(\ln \ell)^2}{R^2 \ell^{2-2/R}} = C_0 \left(\frac{\ln \ell}{R\ell^{1-1/R}} \right)^2 \leq \frac{C_0}{D^2} \leq 1.$$

Fix $u \geq \ell$ and $S \subseteq \mathbb{F}_2^u$ of size 2^ℓ. Choose an integer $U \geq u$ so large that $2^{\ell-U} \leq p_{\ell,R}$. Embed $\mathbb{F}_2^u$ into $\mathbb{F}_2^U$ by appending $U - u$ zero coordinates:

$$x = (x_1, \dots, x_u) \mapsto (x_1, \dots, x_u, 0, \dots, 0).$$

Let $V \leq \mathbb{F}_2^U$ be the image of this embedding. We then regard $S \subseteq \mathbb{F}_2^u$ as a subset of $V \subseteq \mathbb{F}_2^U$.

Let $H : \mathbb{F}_2^U \to \mathbb{F}_2^\ell$ be a uniformly random linear map. By Lemma 10,

$$\Pr[H \text{ is not surjective}] \leq 2^{\ell-U} \leq p_{\ell,R}.$$

Conditioned on H being surjective, the map H is uniformly distributed among all surjective maps $\mathbb{F}_2^U \to \mathbb{F}_2^\ell$. Therefore Proposition 9 gives

$$\Pr \left[\mathsf{M}(S,H) \geq R \frac{\ell}{\log \ell} \;\middle|\; H \text{ is surjective} \right] \leq p_{\ell,R}.$$

Hence

$$\Pr_H \left[\mathsf{M}(S,H) \geq R \frac{\ell}{\log \ell} \right] \leq 2p_{\ell,R}.$$

Finally, the restriction $H|_V : V \to \mathbb{F}_2^\ell$ is a uniformly random linear map from $V \cong \mathbb{F}_2^u$ to $\mathbb{F}_2^\ell$. Also, for the set $S \subseteq V$, $\mathsf{M}(S,H) = \mathsf{M}(S,H|_V)$, because loads are computed only using points of S. Thus the same bound holds for a uniformly random linear map $h : \mathbb{F}_2^u \to \mathbb{F}_2^\ell$. Absorbing the factor 2 into the absolute constant proves the theorem. □

3.4 The expectation bound

The optimized tail can be integrated starting slightly above $R = 1 + \ln \ln \ell / \ln \ell$. We choose the starting point with a fixed additive constant in the numerator.

Corollary 12 (Expectation bound to second order)**.** *For every $u \geq \ell$ and every $S \subseteq \mathbb{F}_2^u$ of size 2^ℓ,*

$$\mathbb{E}_h[\mathsf{M}(S,h)] \leq \left(1 + (1 + o(1)) \frac{\ln \ln \ell}{\ln \ell} \right) \frac{\ell}{\log \ell}$$

as $\ell \to \infty$, where $h : \mathbb{F}_2^u \to \mathbb{F}_2^\ell$ is a uniformly random linear map. Equivalently, since $n = 2^\ell$,

$$\mathbb{E}_h[\mathsf{M}(S,h)] \le \left(1 + (1+o(1))\frac{\log\log\log n}{\log\log n}\right)\frac{\log n}{\log\log n},$$

where the logarithms in the final display are base 2.

Proof. Let

$$L := \frac{\ell}{\log \ell}.$$

Let D be the constant from Theorem 11. Choose a fixed constant $F > 0$ large enough so that $e^F > 2D$. Let

$$R_0 := 1 + \frac{\ln\ln\ell + F}{\ln \ell}.$$

We first check that Theorem 11 applies for every $R \ge R_0$, once ℓ is sufficiently large. We have

$$\left(1 - \frac{1}{R_0}\right)\ln \ell = \frac{\ln\ln\ell + F}{R_0} = \ln\ln\ell + F + o(1).$$

Therefore

$$\ell^{1-1/R_0} = \exp\left(\left(1 - \frac{1}{R_0}\right)\ln\ell\right) = e^F(1+o(1))\ln\ell. \tag{3}$$

Hence, since $R_0 = 1 + o(1)$ and $e^F > 2D$,

$$R_0\ell^{1-1/R_0} = e^F(1+o(1))\ln\ell \ge D\ln\ell$$

for all sufficiently large ℓ. Since the function $R \mapsto R\ell^{1-1/R}$ is increasing for $R > 0$, we also have $R\ell^{1-1/R} \ge D\ln\ell$ for every $R \ge R_0$. Thus Theorem 11 applies throughout the range $R \ge R_0$.

Using the tail-integral formula for a nonnegative random variable,

$$\begin{aligned}
\mathbb{E}[\mathsf{M}(S,h)] &= \int_0^\infty \Pr[\mathsf{M}(S,h) \ge t]\,dt \\
&= \int_0^{R_0L} \Pr[\mathsf{M}(S,h) \ge t]\,dt + \int_{R_0L}^\infty \Pr[\mathsf{M}(S,h) \ge t]\,dt \\
&\le R_0L + \int_{R_0L}^\infty \Pr[\mathsf{M}(S,h) \ge t]\,dt \\
&= R_0L + L\int_{R_0}^\infty \Pr[\mathsf{M}(S,h) \ge RL]\,dR.
\end{aligned}$$

By Theorem 11,

$$\mathbb{E}[\mathsf{M}(S,h)] \le R_0L + CL\int_{R_0}^\infty \frac{(\ln\ell)^2}{R^2\ell^{2-2/R}}\,dR.$$

Set $x = 1 - 1/R$. Then $dx = dR/R^2$, and therefore

$$\int_{R_0}^\infty \frac{(\ln\ell)^2}{R^2\ell^{2-2/R}}\,dR = (\ln\ell)^2\int_{x_0}^1 \ell^{-2x}\,dx,$$

where $x_0 = 1 - 1/R_0$. Thus, by (3),

$$\begin{aligned}(\ln \ell)^2 \int_{x_0}^{1} \ell^{-2x}\, dx &\leq \frac{\ln \ell}{2} \ell^{-2x_0} = \frac{\ln \ell}{2} \ell^{-2(1-1/R_0)} \\ &= \frac{1}{2} e^{-2F+o(1)} (\ln \ell)^{-1} \\ &= o\left(\frac{\ln \ln \ell}{\ln \ell}\right).\end{aligned}$$

It follows that

$$\mathbb{E}[\mathsf{M}(S,h)] \leq \left(R_0 + o\left(\frac{\ln \ln \ell}{\ln \ell}\right)\right) L.$$

Substituting the definition of R_0, we obtain

$$\mathbb{E}[\mathsf{M}(S,h)] \leq \left(1 + \frac{\ln \ln \ell + F}{\ln \ell} + o\left(\frac{\ln \ln \ell}{\ln \ell}\right)\right) \frac{\ell}{\log \ell} = \left(1 + (1+o(1))\frac{\ln \ln \ell}{\ln \ell}\right) \frac{\ell}{\log \ell}.$$

Finally, since $n = 2^{\ell}$, we have $\ell = \log n$ and $\log \ell = \log \log n$. Also,

$$\frac{\ln \ln \ell}{\ln \ell} = (1+o(1)) \frac{\log \log \log n}{\log \log n}.$$

This gives the equivalent n-form. □

3.5 Comparison of the two tail mechanisms

The fixed-bucket and potential arguments apply in different regimes. Corollary 3 gives, for a prescribed bucket,

$$\Pr[\mathrm{load}_h(y) > 2^a - 2] \leq \gamma^{-1} 2^{-a^2}.$$

This estimate is sharp for a single bucket. However, controlling the maximum load by a direct union bound gives

$$\Pr[\mathsf{M}(S,h) > 2^a - 2] \leq \gamma^{-1} 2^{\ell - a^2}.$$

This becomes nontrivial only when $a \gtrsim \sqrt{\ell}$, that is, only for loads at least about $2^{\sqrt{\ell}}$. Thus the fixed-bucket estimate is a sharp far-tail result, but it does not by itself reach the expectation scale.

By contrast, the optimized potential tail in Theorem 11 controls loads at the expectation scale $\ell/\log \ell$. The improvement over the $O(R^{-2})$ tail bound of Jaber–Kumar–Zuckerman comes from reducing the potential base. To detect a bin of load

$$R \frac{\ell}{\log \ell},$$

we take

$$b = \ell^{1/R + 1/\ln \ell} = e\ell^{1/R}.$$

The initial potential satisfies

$$\Phi_0 - 1 = \frac{b-1}{2^k}.$$

With the original choice $b \approx \ell$, this is of order $\ell/2^k$. With the optimized choice $b = e\ell^{1/R}$, it is instead of order $\ell^{1/R}/2^k$. Thus the initial potential is smaller by a factor

$$\ell^{-1+1/R}.$$

The quadratic tail lemma bounds the failure probability by a constant times the square of the normalized initial potential. Hence this saving is squared, giving the factor

$$\ell^{-2+2/R}.$$

The condition

$$R\ell^{1-1/R} \geq D \ln \ell$$

is the condition that the chosen base is large enough to detect the threshold while keeping the initial potential small enough for the quadratic lemma. The threshold for this condition occurs when R is just above

$$1 + \frac{\ln \ln \ell}{\ln \ell}.$$

More precisely, if F is a sufficiently large absolute constant and

$$R_0 = 1 + \frac{\ln \ln \ell + F}{\ln \ell},$$

then

$$R_0 \ell^{1-1/R_0} = e^F (1 + o(1)) \ln \ell \geq D \ln \ell$$

for all sufficiently large ℓ. Hence Theorem 11 applies for all $R \geq R_0$. Integrating the tail from this point gives Corollary 12, namely

$$\mathbb{E}_h[\mathsf{M}(S,h)] \leq \left(1 + (1+o(1))\frac{\ln \ln \ell}{\ln \ell}\right) \frac{\ell}{\log \ell}.$$

Thus the two estimates are complementary. The fixed-bucket argument gives a nearly sharp tail for one prescribed bucket, including the correct 2^{-a^2} behavior. The potential argument is weaker for a single bucket but is global: it controls the maximum over all buckets at the scale relevant for the expected maximum load.

4 Conclusion

This note revisits the potential method of Jaber–Kumar–Zuckerman and shows that the choice of base can be tuned to the target load level. This simple optimization strengthens the maximum-load tail from the $O(R^{-2})$ bound to

$$\Pr\left[\mathsf{M}(S,h) \geq R\frac{\ell}{\log \ell}\right] \leq C\frac{(\ln \ell)^2}{R^2 \ell^{2-2/R}}$$

for every admissible $R > 1$. Integrating the estimate from

$$R_0 = 1 + \frac{\ln \ln \ell + O(1)}{\ln \ell}$$

gives

$$\mathbb{E}[\mathsf{M}(S,h)] \leq \left(1 + (1+o(1))\frac{\ln \ln \ell}{\ln \ell}\right) \frac{\ell}{\log \ell}.$$

Equivalently, since $n = 2^\ell$,

$$\mathbb{E}[\mathsf{M}(S,h)] \leq \frac{\log n}{\log \log n} + (1+o(1))\frac{\log n \cdot \log \log \log n}{(\log \log n)^2}.$$

Thus binary linear hashing matches the fully independent maximum-load scale not only in the leading term, but also in the dominant second-order term.

We also proved a separate fixed-bucket estimate. For one prescribed bucket, the load has tail

$$\Pr[\mathrm{load}_h(y) > 2^a - 2] \le \gamma^{-1} 2^{-a^2},$$

and a subspace construction shows that this is asymptotically tight even in the leading constant as $a \to \infty$. This fixed-bucket result is sharp, but it is a far-tail statement: a direct union bound over all buckets loses a factor 2^ℓ, so the potential method remains essential for controlling the maximum load at the expectation scale.

The same optimization also extends to m keys and n bins. In the regime where the fully independent maximum-load scale t is much larger than the average load m/n, the argument gives

$$\mathbb{E}[\mathsf{M}(S,h)] \le (1+o(1))t.$$

The dense two-sided balancing regime is different: there one must control both large and small buckets at the scale of the average load. The methods here improve the sparse maximum-load tail, but they do not improve the dense two-sided bounds.

References

[1] Noga Alon, Martin Dietzfelbinger, Peter Bro Miltersen, Erez Petrank, and Gábor Tardos. Is linear hashing good? In *Proceedings of the Twenty-Ninth Annual ACM Symposium on Theory of Computing*, pages 465–474, 1997.

[2] Noga Alon, Martin Dietzfelbinger, Peter Bro Miltersen, Erez Petrank, and Gábor Tardos. Linear hash functions. *Journal of the ACM*, 46(5):667–683, 1999.

[3] Michael Jaber, Vinayak M. Kumar, and David Zuckerman. Linear hashing is optimal. In *Proceedings of the 57th Annual ACM Symposium on Theory of Computing*, pages 245–255, 2025. Also available as arXiv:2505.14061.

[4] Martin Raab and Angelika Steger. "Balls into bins"—A simple and tight analysis. In *Randomization and Approximation Techniques in Computer Science*, RANDOM 1998, Lecture Notes in Computer Science, vol. 1518, pages 159–170. Springer, 1998.

[5] Manik Dhar and Zeev Dvir. Linear hashing with ℓ_∞ guarantees and two-sided Kakeya bounds. *TheoretiCS*, 3:Article 8, 2024. Preliminary version in *Proceedings of the 63rd IEEE Symposium on Foundations of Computer Science*, pages 419–428, 2022.

[6] Gaston H. Gonnet. Expected length of the longest probe sequence in hash code searching. *Journal of the ACM*, 28(2):289–304, 1981.

[7] Jason Fulman and Larry Goldstein. Stein's method and the rank distribution of random matrices over finite fields. *Annals of Probability*, 43(3):1274–1314, 2015.

A Potential lemmas

In this appendix we include self-contained proofs of the potential lemmas used in the main text. The first lemma packages the potential-evolution estimates of Jaber–Kumar–Zuckerman in the normalization used here. The second lemma is a normalized quadratic tail lemma.

A.1 Potential evolution

Proof of Lemma 6. Fix $0 \le i < k$, and condition on the subspace V_i. Let

$$G_i := \mathbb{F}_2^U / V_i, \qquad N_i := |G_i| = 2^{U-i}.$$

Since V_{i+1} is obtained from V_i by adjoining one new vector outside V_i, let $w \in G_i \setminus \{0\}$ denote the image of that new vector in the quotient $G_i = \mathbb{F}_2^U / V_i$. Conditional on V_i, this w is uniformly distributed over $G_i \setminus \{0\}$.

Define $f_i : G_i \to \mathbb{Z}_{\ge 0}$ by

$$f_i(C) := |C \cap S|.$$

Thus

$$\Phi_i = \mathbb{E}_{C \in G_i} \left[b^{f_i(C)} \right].$$

Let

$$\pi_i : G_i \to G_{i+1}$$

be the quotient map induced by $V_i \le V_{i+1}$. Passing from V_i to V_{i+1} merges the two V_i-cosets C and $C + w$ into the same V_{i+1}-coset $\pi_i(C)$. Define $f_{i+1} : G_{i+1} \to \mathbb{Z}_{\ge 0}$ by

$$f_{i+1}(D) := |D \cap S|.$$

Then, for every $C \in G_i$,[3]

$$f_{i+1}(\pi_i(C)) = |\pi_i(C) \cap S| = f_i(C) + f_i(C + w).$$

Therefore, for the choice of new vector w,

$$\Phi_{i+1}(w) = \mathbb{E}_{D \in G_{i+1}} \left[b^{f_{i+1}(D)} \right].$$

Since every $D \in G_{i+1}$ has exactly two preimages in G_i, namely C and $C + w$, we may equivalently average over $C \in G_i$. Hence

$$\Phi_{i+1}(w) = \mathbb{E}_{C \in G_i} \left[b^{f_i(C) + f_i(C+w)} \right].$$

Let $a_C := b^{f_i(C)}$. Then $\Phi_i = \mathbb{E}_{C \in G_i}[a_C]$. Averaging over the uniformly random choice of $w \in G_i \setminus \{0\}$, we get

$$\mathbb{E}_w[\Phi_{i+1}(w)] = \mathbb{E}_{w \ne 0} \mathbb{E}_{C \in G_i}[a_C a_{C+w}].$$

Expanding this average gives

$$\mathbb{E}_{w \ne 0} \mathbb{E}_{C \in G_i}[a_C a_{C+w}] = \frac{1}{N_i(N_i - 1)} \sum_{C \in G_i} \sum_{\substack{w \in G_i \\ w \ne 0}} a_C a_{C+w}.$$

[3] Here we identify a coset with the corresponding subset of $\mathbb{F}_2^U$. Thus $\pi_i(C)$, which is formally a V_{i+1}-coset, is the union of the two V_i-cosets C and $C + w$, namely $\pi_i(C) = C \cup (C + w)$.

For fixed C, as w ranges over $G_i \setminus \{0\}$, the coset $C + w$ ranges over $G_i \setminus \{C\}$. Therefore

$$\sum_{C \in G_i} \sum_{\substack{w \in G_i \\ w \neq 0}} a_C a_{C+w} = \sum_{C \in G_i} a_C \sum_{\substack{D \in G_i \\ D \neq C}} a_D = \left(\sum_{C \in G_i} a_C \right)^2 - \sum_{C \in G_i} a_C^2.$$

Hence

$$\begin{aligned} \mathbb{E}_w[\Phi_{i+1}(w)] &= \frac{1}{N_i(N_i - 1)} \left[\left(\sum_{C \in G_i} a_C \right)^2 - \sum_{C \in G_i} a_C^2 \right] \\ &= \frac{1}{N_i(N_i - 1)} \left[N_i^2 \Phi_i^2 - N_i \mathbb{E}_{C \in G_i}[a_C^2] \right] \\ &= \frac{N_i \Phi_i^2 - \mathbb{E}_{C \in G_i}[a_C^2]}{N_i - 1}. \end{aligned}$$

By Jensen's inequality, $\mathbb{E}_{C \in G_i}[a_C^2] \geq (\mathbb{E}_{C \in G_i}[a_C])^2 = \Phi_i^2$. Therefore

$$\mathbb{E}_w[\Phi_{i+1}(w)] \leq \frac{N_i \Phi_i^2 - \Phi_i^2}{N_i - 1} = \Phi_i^2.$$

Equivalently, $\mathbb{E}[\Phi_{i+1} \mid V_i] \leq \Phi_i^2$. Since $\Phi_0, \ldots, \Phi_i$ are determined once V_i is known, the tower property gives

$$\mathbb{E}[\Phi_{i+1} \mid \Phi_0, \ldots, \Phi_i] = \mathbb{E}\left[\mathbb{E}[\Phi_{i+1} \mid V_i] \,|\, \Phi_0, \ldots, \Phi_i\right] \leq \Phi_i^2.$$

It remains to prove the deterministic growth estimate. For every $C \in G_i$, since $b > 1$, both

$$b^{f_i(C)} - 1 \qquad \text{and} \qquad b^{f_i(C+w)} - 1$$

are nonnegative. Therefore

$$b^{f_i(C) + f_i(C+w)} - 1 = b^{f_i(C)} b^{f_i(C+w)} - 1 \geq \left(b^{f_i(C)} - 1\right) + \left(b^{f_i(C+w)} - 1\right).$$

Averaging over $C \in G_i$, we obtain

$$\Phi_{i+1} - 1 \geq \mathbb{E}_{C \in G_i}\left[b^{f_i(C)} - 1\right] + \mathbb{E}_{C \in G_i}\left[b^{f_i(C+w)} - 1\right].$$

Since $C + w$ is uniformly distributed over G_i when C is uniformly distributed over G_i, the two averages are equal. Hence

$$\Phi_{i+1} - 1 \geq 2\mathbb{E}_{C \in G_i}\left[b^{f_i(C)} - 1\right] = 2(\Phi_i - 1).$$

This completes the proof. □

A.2 Strengthened quadratic potential tail lemma

We next prove Lemma 7. This is the strengthened quadratic tail lemma used in the main text; in particular, it does not require any upper bound on τ.

We need one elementary one-step estimate.

Lemma 13 (One-step estimate)**.** *Let $s > 0$, $0 \leq r \leq 1/4$, and let $d := s(2 + s)$.*

Let Z be a nonnegative random variable satisfying $Z \geq 2rs$ and $\mathbb{E}[Z] \leq 2rs + r^2 s^2$. Define $F(u) := \min\{1, 48u^2\}$. Then

$$\mathbb{E}\left[F\left(\frac{Z}{d}\right)\right] \leq 48r^2.$$

Proof. The function $F(u) = \min\{1, 48u^2\}$ is increasing and globally Lipschitz on $[0, \infty)$ with Lipschitz constant $8\sqrt{3}$. Indeed, on the quadratic part its derivative is $96u$, and the quadratic part ends at $u = 1/\sqrt{48}$, where the derivative equals $8\sqrt{3}$.

Since $Z \geq 2rs$, we have

$$F\left(\frac{Z}{d}\right) \leq F\left(\frac{2rs}{d}\right) + \frac{8\sqrt{3}}{d}(Z - 2rs).$$

Taking expectations and using $\mathbb{E}(Z - 2rs) \leq r^2 s^2$, we get

$$\mathbb{E}\left[F\left(\frac{Z}{d}\right)\right] \leq F\left(\frac{2rs}{s(2+s)}\right) + \frac{8\sqrt{3}}{s(2+s)} r^2 s^2 = F\left(\frac{2r}{2+s}\right) + \frac{8\sqrt{3}\, r^2 s}{2+s}.$$

We now show that the right-hand side is at most $48r^2$.

First suppose

$$\frac{2r}{2+s} \leq \frac{1}{\sqrt{48}}.$$

Then

$$F\left(\frac{2r}{2+s}\right) = 48\left(\frac{2r}{2+s}\right)^2.$$

It is enough to prove

$$48\frac{4r^2}{(2+s)^2} + \frac{8\sqrt{3}\, r^2 s}{2+s} \leq 48r^2.$$

After dividing by $48r^2$, this becomes

$$\frac{4}{(2+s)^2} + \frac{\sqrt{3}}{6}\frac{s}{2+s} \leq 1.$$

Since

$$1 - \frac{4}{(2+s)^2} = \frac{(2+s)^2 - 4}{(2+s)^2} = \frac{s(4+s)}{(2+s)^2},$$

it is enough to show

$$\frac{s(4+s)}{(2+s)^2} \geq \frac{\sqrt{3}}{6}\frac{s}{2+s}.$$

Since $s > 0$, this is equivalent to

$$\frac{4+s}{2+s} \geq \frac{\sqrt{3}}{6}.$$

Since $(4+s)/(2+s) > 1 > \sqrt{3}/6$, the desired inequality follows.

Now suppose

$$\frac{2r}{2+s} > \frac{1}{\sqrt{48}}.$$

Then

$$F\left(\frac{2r}{2+s}\right) = 1.$$

The assumption implies

$$48r^2 - 1 > \frac{(2+s)^2}{4} - 1 = s + \frac{s^2}{4}.$$

Since $r \le 1/4$, we also have

$$s + \frac{s^2}{4} \ge \frac{\sqrt{3}}{4} s \ge \frac{\sqrt{3}}{2} \frac{s}{2+s} \ge \frac{8\sqrt{3}\, r^2 s}{2+s}.$$

Therefore

$$1 + \frac{8\sqrt{3}\, r^2 s}{2+s} \le 48 r^2.$$

This proves the estimate in both cases. □

We now prove the strengthened quadratic tail lemma.

Proof of Lemma 7. We prove the lemma by induction on k. Write $x := X_0$. The case $k = 0$ is immediate. Indeed, $\tau \ge 1 + 4(x-1) = x + 3(x-1) > x$ and so $\Pr[X_0 \ge \tau] = 0$.

Assume the result is known for sequences of length $k-1$, and prove it for length k. Let $y = x - 1$, and $s = \tau - 1$. Then $0 < s$, and $0 \le y = x - 1 \le (\tau - 1)/4 = s/4$. Define $r := y/s$. Thus $0 \le r = (x-1)/(\tau-1) \le 1/4$. Also let $d := \tau^2 - 1 = s(2+s)$.

Let

$$Z := X_1 - 1.$$

Since $X_1 - 1 \ge 2(X_0 - 1)$,

$$Z \ge 2y = 2rs.$$

The conditional expectation hypothesis gives

$$\mathbb{E}[Z] = \mathbb{E}[X_1 - 1] \le x^2 - 1 = (1+y)^2 - 1 = 2y + y^2 = 2rs + r^2 s^2.$$

Now condition on X_1. If

$$1 + 4(X_1 - 1) \le \tau^2,$$

then the induction hypothesis, applied to the remaining sequence $X_1, X_2, \ldots, X_k$ with threshold τ^2, gives

$$\Pr\left[X_k \ge (\tau^2)^{2^{k-1}} \,\middle|\, X_1\right] \le 48 \left(\frac{X_1 - 1}{\tau^2 - 1}\right)^2.$$

If instead

$$1 + 4(X_1 - 1) > \tau^2,$$

then, $48((X_1 - 1)/(\tau^2 - 1))^2 > 48/16 > 1$, and we use the trivial bound by 1. We have, in all cases,

$$\Pr\left[X_k \ge \tau^{2^k} \,\middle|\, X_1\right] \le F\left(\frac{X_1 - 1}{\tau^2 - 1}\right),$$

where

$$F(u) := \min\{1, 48u^2\}.$$

Taking expectations,

$$\Pr[X_k \ge \tau^{2^k}] \le \mathbb{E}\left[F\left(\frac{Z}{d}\right)\right].$$

By Lemma 13,

$$\mathbb{E}\left[F\left(\frac{Z}{d}\right)\right] \le 48r^2 = 48\left(\frac{y}{s}\right)^2 = 48\left(\frac{x-1}{\tau-1}\right)^2 = 48\left(\frac{X_0 - 1}{\tau - 1}\right)^2.$$

This completes the induction. □

B A fixed-bucket tail bound for m keys and n bins

In this appendix we record the fixed-bucket version of the tail estimate for m keys and n bins. Since binary linear hashing naturally gives a power-of-two number of bins, write $n := 2^\ell$. Let

$$S = \{u_1, \ldots, u_m\} \subseteq \mathbb{F}_2^d \setminus \{0\}$$

be a set of distinct nonzero vectors, and let $h : \mathbb{F}_2^d \to \mathbb{F}_2^\ell$ be a uniformly random linear map. For a fixed bucket $y \in \mathbb{F}_2^\ell$, define

$$Z_y := |\{i : h(u_i) = y\}|.$$

Let

$$\lambda := \frac{m}{n}.$$

We will use again Lemma 1: if $A \subseteq \mathbb{F}_2^d \setminus \{0\}$ has size q, and $a = \lceil \log(q+1) \rceil$, then A contains at least

$$\prod_{j=0}^{a-1} (q + 1 - 2^j)$$

ordered linearly independent a-tuples.

Theorem 14 (Fixed-bucket tail for m keys and n bins)**.** *Let $n = 2^\ell$, and let*

$$S = \{u_1, \ldots, u_m\} \subseteq \mathbb{F}_2^d \setminus \{0\}$$

be a set of distinct nonzero vectors. Let $h : \mathbb{F}_2^d \to \mathbb{F}_2^\ell$ be a uniformly random linear map. Fix $y \in \mathbb{F}_2^\ell$, and define $Z_y := |\{i : h(u_i) = y\}|$. Let $\lambda = m/n$. Then, for every integer $r \geq 0$, if

$$a := \lceil \log(r+2) \rceil,$$

we have

$$\Pr[Z_y > r] \leq \lambda^a \left(\prod_{j=0}^{a-1} (r + 2 - 2^j) \right)^{-1}.$$

Proof. Let $q = r + 1$, and $a = \lceil \log(q+1) \rceil$. Let $\mathcal{I}_a$ be the set of ordered a-tuples $(i_1, \ldots, i_a)$ such that $u_{i_1}, \ldots, u_{i_a}$ are linearly independent. For a linear map h, define

$$T_a(h) := \sum_{(i_1,\ldots,i_a) \in \mathcal{I}_a} \mathbf{1}[h(u_{i_1}) = \cdots = h(u_{i_a}) = y].$$

Thus $T_a(h)$ counts ordered linearly independent a-tuples from S that all land in the prescribed bucket y.

For a fixed ordered linearly independent a-tuple $(u_{i_1}, \ldots, u_{i_a})$, the random vectors $h(u_{i_1}), \ldots, h(u_{i_a})$ are independent and uniformly distributed in $\mathbb{F}_2^\ell$. Hence

$$\Pr_h[h(u_{i_1}) = \cdots = h(u_{i_a}) = y] = 2^{-\ell a} = n^{-a}.$$

Therefore, by linearity of expectation,

$$\mathbb{E}_h[T_a(h)] = |\mathcal{I}_a| n^{-a}.$$

Since $|\mathcal{I}_a| \le m^a$, we get

$$\mathbb{E}_h[T_a(h)] \le \left(\frac{m}{n}\right)^a = \lambda^a.$$

Now suppose $Z_y \ge q$. Then the set $A = \{u_i : h(u_i) = y\}$ has size at least q. By Lemma 1, A contains at least

$$M(q) := \prod_{j=0}^{a-1}(q + 1 - 2^j)$$

ordered linearly independent a-tuples. Thus

$$Z_y \ge q \quad \Longrightarrow \quad T_a(h) \ge M(q).$$

By Markov's inequality,

$$\Pr[Z_y \ge q] \le \Pr[T_a(h) \ge M(q)] \le \frac{\mathbb{E}_h[T_a(h)]}{M(q)} \le \frac{\lambda^a}{M(q)}.$$

Since $q = r + 1$, this gives

$$\Pr[Z_y > r] \le \lambda^a \left(\prod_{j=0}^{a-1}(r + 2 - 2^j)\right)^{-1}.$$

□

Let

$$\gamma := \prod_{j=1}^{\infty}(1 - 2^{-j}).$$

The preceding theorem has the following clean form at dyadic thresholds.

Corollary 15 (Dyadic fixed-bucket tail for m keys and n bins)**.** *Under the assumptions of Theorem 14, for every integer $a \ge 1$,*

$$\Pr[Z_y > 2^a - 2] \le \gamma^{-1}\lambda^a 2^{-a^2}.$$

Proof. Apply Theorem 14 with $r = 2^a - 2$. Then $\lceil \log(r + 2) \rceil = a$, and

$$\prod_{j=0}^{a-1}(r + 2 - 2^j) = \prod_{j=0}^{a-1}(2^a - 2^j) \ge \gamma 2^{a^2}.$$

Therefore

$$\Pr[Z_y > 2^a - 2] \le \lambda^a \gamma^{-1} 2^{-a^2}.$$

□

Remark 16 (Including the zero vector)**.** *The theorem above assumes that all keys are nonzero. If $0 \in S$, then for $y \ne 0$, the zero vector never contributes to Z_y, so one may apply the theorem to $S \setminus \{0\}$. For $y = 0$, the zero vector contributes deterministically one point to the load. Thus the same bound applies to $Z_0 - 1$ after removing the zero vector from S.*

Remark 17 (Comparison with the balanced case)**.** *When $m = n$, we have $\lambda = 1$, and Corollary 15 becomes*

$$\Pr[Z_y > 2^a - 2] \le \gamma^{-1} 2^{-a^2},$$

which is the fixed-bucket estimate used in the balanced case. For general m, the additional factor

$$\lambda^a = \left(\frac{m}{n}\right)^a$$

reflects the average load of the prescribed bucket.

Proposition 18 (Matching lower bound for the fixed-bucket tail)**.** *Let $n = 2^\ell$, let $m = 2^d$, and let $\lambda = m/n = 2^{d-\ell}$. Let $d \ge a \ge \max\{1, d - \ell\}$. Then there exists a set $S \subseteq \mathbb{F}_2^N \setminus \{0\}$ of m distinct nonzero vectors such that, for a uniformly random linear map $h : \mathbb{F}_2^N \to \mathbb{F}_2^\ell$, one has*

$$\Pr\left[|\{x \in S : h(x) = 0\}| > 2^a - 2\right] \ge \gamma^2 \lambda^a 2^{-a^2},$$

where $\gamma := \prod_{j=1}^{\infty}(1 - 2^{-j})$.

Thus, the upper bound

$$\Pr[Z_0 > 2^a - 2] \le \gamma^{-1} \lambda^a 2^{-a^2}$$

is sharp up to an absolute multiplicative constant.

Proof. Let $W \le \mathbb{F}_2^N$ be a d-dimensional subspace, and choose $v \notin W$. Define $S := (W \setminus \{0\}) \cup \{v\}$. Then $|S| = (2^d - 1) + 1 = 2^d = m$, and all elements of S are distinct and nonzero.

Let

$$M := h|_W : W \to \mathbb{F}_2^\ell.$$

After choosing a basis of W, the map M is represented by a uniformly random $\ell \times d$ binary matrix. If $\mathrm{nul}(M) \ge a$, then

$$|\ker(M) \setminus \{0\}| = 2^{\mathrm{nul}(M)} - 1 \ge 2^a - 1.$$

Since

$$\ker(M) \setminus \{0\} \subseteq \{x \in S : h(x) = 0\},$$

we get

$$|\{x \in S : h(x) = 0\}| > 2^a - 2.$$

Therefore

$$\Pr\left[|\{x \in S : h(x) = 0\}| > 2^a - 2\right] \ge \Pr[\mathrm{nul}(M) \ge a] \ge \Pr[\mathrm{nul}(M) = a].$$

It remains to lower-bound the probability that a uniformly random $\ell \times d$ binary matrix has nullity exactly a. This is the same as having rank $d - a$. The standard rank formula in [7] gives

$$\Pr[\mathrm{rank}(M) = d - a] = 2^{-a(\ell - d + a)} \frac{\prod_{i=0}^{d-a-1}(1 - 2^{i-\ell}) \prod_{i=0}^{d-a-1}(1 - 2^{i-d})}{\prod_{i=0}^{d-a-1}(1 - 2^{i-(d-a)})}.$$

The denominator is at most 1. Also,

$$\prod_{i=0}^{d-a-1}(1 - 2^{i-d}) = \prod_{j=a+1}^{d}(1 - 2^{-j}) \ge \gamma,$$

and, since $a \geq d - \ell$, we have $d - a \leq \ell$, so

$$\prod_{i=0}^{d-a-1} (1 - 2^{i-\ell}) \geq \gamma.$$

Hence

$$\Pr[\mathrm{nul}(M) = a] = \Pr[\mathrm{rank}(M) = d - a] \geq \gamma^2 2^{-a(\ell-d+a)}.$$

Finally,

$$2^{-a(\ell-d+a)} = 2^{a(d-\ell)} 2^{-a^2} = \lambda^a 2^{-a^2}.$$

Therefore

$$\Pr\left[|\{x \in S : h(x) = 0\}| > 2^a - 2\right] \geq \gamma^2 \lambda^a 2^{-a^2}.$$

□

C Maximum-load bounds for m keys and n bins

In this section we record the extension of the base-optimized argument to the case where the number of keys is not necessarily equal to the number of bins. Let the number of bins be $n = 2^\ell$, let $S \subseteq \mathbb{F}_2^u$ have size $|S| = m$, and write

$$\lambda := \frac{m}{n}.$$

Thus λ is the average load. For a linear map $h : \mathbb{F}_2^u \to \mathbb{F}_2^\ell$, define

$$\mathsf{M}(S, h) := \max_{y \in \mathbb{F}_2^\ell} |h^{-1}(y) \cap S|.$$

The fully independent comparison scale in the sparse large-load regime is the number $t = t(m, n)$ defined by

$$t \ln\left(\frac{t}{e\lambda}\right) = \ln n.$$

This scale is meaningful in the range

$$\frac{t}{\lambda} \to \infty,$$

that is, when the maximum-load scale is much larger than the average load.

We first prove the tail bound for uniformly random surjective maps.

Proposition 19 (Surjective tail bound for m keys and n bins)**.** *There exist absolute constants $C_0, D_0 > 0$ such that the following holds. Let $U \geq \ell$, let $S \subseteq \mathbb{F}_2^U$ have size m, and let $H : \mathbb{F}_2^U \to \mathbb{F}_2^\ell$ be a uniformly random surjective linear map. Put $n = 2^\ell$ and $\lambda = m/n$. Then, for every $T > 0$ satisfying*

$$T \geq D_0 \lambda n^{1/T},$$

one has

$$\Pr_H[\mathsf{M}(S, H) \geq T] \leq C_0 \left(\frac{\lambda n^{1/T}}{T}\right)^2.$$

Proof. Let $k = U - \ell$. As before, expose the kernel of H through a chain

$$V_0 \le V_1 \le \cdots \le V_k = \ker H,$$

with $V_0 = \{0\}$ and $\dim V_i = i$. For a base $b > 1$, define

$$\Phi_i := \mathbb{E}_{x \in \mathbb{F}_2^U}\left[b^{S_i(x)}\right], \qquad S_i(x) := |(x + V_i) \cap S|.$$

We use the same potential evolution and quadratic tail lemmas as before.

Choose $b = en^{1/T}$. Since $V_0 = \{0\}$, we have $S_0(x) = 1_S(x)$, and therefore

$$\Phi_0 - 1 = \frac{|S|}{2^U}(b-1) = \frac{m}{n2^k}(b-1) \le \frac{\lambda b}{2^k}.$$

If some final bucket has load at least T, then, by Lemma 8,

$$\Phi_k \ge \frac{b^T}{n} = \frac{(en^{1/T})^T}{n} = e^T.$$

Now define

$$\tau := 1 + \frac{T}{2^k}.$$

Then

$$\tau^{2^k} = \left(1 + \frac{T}{2^k}\right)^{2^k} \le e^T.$$

Hence

$$\mathsf{M}(S, H) \ge T \implies \Phi_k \ge \tau^{2^k}.$$

We verify the hypothesis of the quadratic potential tail lemma. Choose $D_0 \ge 4e$. Using $\Phi_0 - 1 \le \lambda b / 2^k$, the definition of τ, and the assumption $T \ge D_0 \lambda n^{1/T}$, we get

$$4(\Phi_0 - 1) \le \frac{4\lambda b}{2^k} = \frac{4e\lambda n^{1/T}}{2^k} \le \frac{T}{2^k} = \tau - 1.$$

Thus

$$\tau \ge 1 + 4(\Phi_0 - 1).$$

Applying the quadratic potential tail lemma with $X_i = \Phi_i$, we get

$$\Pr_H[\mathsf{M}(S, H) \ge T] \le \Pr[\Phi_k \ge \tau^{2^k}] \le 48\left(\frac{\Phi_0 - 1}{\tau - 1}\right)^2.$$

Finally,

$$\frac{\Phi_0 - 1}{\tau - 1} \le \frac{\lambda b / 2^k}{T / 2^k} = \frac{\lambda b}{T} = \frac{e\lambda n^{1/T}}{T}.$$

Therefore

$$\Pr_H[\mathsf{M}(S, H) \ge T] \le 48e^2\left(\frac{\lambda n^{1/T}}{T}\right)^2.$$

This proves the result with $C_0 = 48e^2$. □

We next remove the surjectivity assumption, exactly as in the equal-size case.

Theorem 20 (Tail bound for m keys and n bins)**.** *There exist absolute constants $C, D > 0$ such that the following holds. Let $n = 2^\ell$, let $S \subseteq \mathbb{F}_2^u$ have size m, and let $\lambda := m/n$. Let $h : \mathbb{F}_2^u \to \mathbb{F}_2^\ell$ be a uniformly random linear map. Then, for every $T > 0$ satisfying $T \geq D\lambda n^{1/T}$, one has*

$$\Pr_h[\mathsf{M}(S,h) \geq T] \leq C \left(\frac{\lambda n^{1/T}}{T} \right)^2 .$$

Proof. Let

$$p_T := C_0 \left(\frac{\lambda n^{1/T}}{T} \right)^2 ,$$

where C_0 is the constant from Proposition 19. Choose the constant D large enough so that $p_T \leq 1$ whenever $T \geq D\lambda n^{1/T}$.

Fix $u \geq \ell$ and $S \subseteq \mathbb{F}_2^u$ of size m. Choose $U \geq u$ so large that $2^{\ell-U} \leq p_T$. Embed $\mathbb{F}_2^u$ into $\mathbb{F}_2^U$ by appending $U - u$ zero coordinates, and regard S as a subset of this copy of $\mathbb{F}_2^u$ inside $\mathbb{F}_2^U$.

Let $H : \mathbb{F}_2^U \to \mathbb{F}_2^\ell$ be a uniformly random linear map. By the Lemma 10,

$$\Pr[H \text{ is not surjective}] \leq 2^{\ell-U} \leq p_T.$$

Conditioned on H being surjective, the map H is uniformly distributed among all surjective maps $\mathbb{F}_2^U \to \mathbb{F}_2^\ell$. Hence Proposition 19 gives

$$\Pr\left[\mathsf{M}(S,H) \geq T \mid H \text{ is surjective}\right] \leq p_T.$$

Therefore

$$\Pr_H[\mathsf{M}(S,H) \geq T] \leq 2p_T.$$

Finally, the restriction of H to the embedded copy of $\mathbb{F}_2^u$ is a uniformly random linear map $\mathbb{F}_2^u \to \mathbb{F}_2^\ell$, and the loads are computed only using points of S. Thus the same estimate holds for a uniformly random linear map $h : \mathbb{F}_2^u \to \mathbb{F}_2^\ell$. Absorbing the factor 2 into the constant proves the theorem. □

The previous theorem can be written in terms of the fully independent large-load scale t.

Corollary 21 (Comparison with the fully independent scale)**.** *Let $n = 2^\ell$, let $S \subseteq \mathbb{F}_2^u$ have size m, and let $\lambda := m/n$. Let $t = t(m,n)$ be defined by*

$$t \ln\left(\frac{t}{e\lambda}\right) = \ln n.$$

Then, for every $R > 1$ satisfying

$$R \left(\frac{t}{\lambda}\right)^{1-1/R} \geq D,$$

one has

$$\Pr_h[\mathsf{M}(S,h) \geq Rt] \leq \frac{C}{R^2} \left(\frac{\lambda}{t}\right)^{2-2/R} .$$

Proof. Since

$$t \ln\left(\frac{t}{e\lambda}\right) = \ln n,$$

we have

$$n^{1/t} = \frac{t}{e\lambda}.$$

Apply Theorem 20 with $T := Rt$. Then

$$n^{1/T} = n^{1/(Rt)} = \left(n^{1/t}\right)^{1/R} = \left(\frac{t}{e\lambda}\right)^{1/R}.$$

The condition $T \geq D\lambda n^{1/T}$ becomes

$$Rt \geq D\lambda \left(\frac{t}{e\lambda}\right)^{1/R},$$

which is implied, after changing the absolute constant D, by

$$R\left(\frac{t}{\lambda}\right)^{1-1/R} \geq D.$$

The tail bound gives

$$\begin{aligned}
\Pr_h[\mathsf{M}(S,h) \geq Rt] &\leq C\left(\frac{\lambda n^{1/(Rt)}}{Rt}\right)^2 \\
&= C\left(\frac{\lambda}{Rt}\left(\frac{t}{e\lambda}\right)^{1/R}\right)^2 \\
&\leq \frac{C}{R^2}\left(\frac{\lambda}{t}\right)^{2-2/R}.
\end{aligned}$$

This proves the corollary. □

Finally, integrating the preceding tail gives a leading-constant comparison with the fully independent scale in the sparse large-load regime.

Corollary 22 (Expectation in the sparse large-load regime)**.** *Assume that*

$$\frac{t}{\lambda} \to \infty,$$

where $t = t(m,n)$ *is defined by*

$$t \ln\left(\frac{t}{e\lambda}\right) = \ln n.$$

Then

$$\mathbb{E}_h[\mathsf{M}(S,h)] \leq (1+o(1))t.$$

Proof. Let

$$\rho := \frac{t}{\lambda}.$$

By assumption, $\rho \to \infty$. Choose a fixed constant $F > 0$ large enough, and set

$$R_0 := 1 + \frac{F}{\ln \rho}.$$

Then

$$1 - \frac{1}{R_0} = \frac{R_0 - 1}{R_0} = \frac{F + o(1)}{\ln \rho}.$$

Hence

$$R_0 \rho^{1-1/R_0} = e^F(1 + o(1)).$$

Choosing F large enough ensures that $R_0\rho^{1-1/R_0} \geq D$ for all sufficiently large n. Since the function $R \mapsto R\rho^{1-1/R}$ is increasing for $R > 0$, the condition $R\rho^{1-1/R} \geq D$ holds for every $R \geq R_0$. Therefore Corollary 21 applies throughout the range $R \geq R_0$.

Using the tail-integral formula,

$$\begin{aligned} \mathbb{E}_h[\mathsf{M}(S,h)] &= \int_0^\infty \Pr[\mathsf{M}(S,h) \geq s]\, ds \\ &\leq R_0 t + t \int_{R_0}^\infty \Pr[\mathsf{M}(S,h) \geq Rt]\, dR. \end{aligned}$$

By Corollary 21,

$$\mathbb{E}_h[\mathsf{M}(S,h)] \leq R_0 t + Ct \int_{R_0}^\infty \frac{1}{R^2} \rho^{-2+2/R}\, dR.$$

Set

$$x := 1 - \frac{1}{R}.$$

Then $dx = dR/R^2$, and the integral becomes

$$\int_{R_0}^\infty \frac{1}{R^2} \rho^{-2+2/R}\, dR = \int_{x_0}^1 \rho^{-2x}\, dx, \qquad x_0 := 1 - \frac{1}{R_0}.$$

Thus

$$\int_{x_0}^1 \rho^{-2x}\, dx \leq \frac{\rho^{-2x_0}}{2 \ln \rho}.$$

Since

$$x_0 = \frac{F + o(1)}{\ln \rho},$$

we have $\rho^{-2x_0} = e^{-2F + o(1)}$. Therefore

$$\int_{x_0}^1 \rho^{-2x}\, dx = O\left(\frac{1}{\ln \rho}\right) = o(1).$$

Also

$$R_0 = 1 + O\left(\frac{1}{\ln \rho}\right) = 1 + o(1).$$

It follows that

$$\mathbb{E}_h[\mathsf{M}(S,h)] \leq (1 + o(1))t.$$

□

Thus, in the regime $t/\lambda \to \infty$, binary linear hashing matches the fully independent expected maximum-load scale up to a $1 + o(1)$ factor. This improves the constant-factor tail and expectation bounds of the general Jaber–Kumar–Zuckerman theorem in the sparse large-load range. It does not address the dense balancing regime, where the maximum load is only a constant factor above the average load λ.

ECCC ISSN 1433-8092
https://eccc.weizmann.ac.il