Linear Hashing

Noga Alon* Martin Dietzfelbinger† Peter Bro Miltersen‡ Erez Petrank§
Gábor Tardos¶

Abstract

Consider the set $\mathcal{H}$ of all linear (or affine) transformations between two vector spaces over a finite field F. We study how good $\mathcal{H}$ is as a class of hash functions, namely we consider hashing a set S of size n into a range having the same cardinality n by a randomly chosen function from $\mathcal{H}$ and look at the expected size of the largest hash bucket. $\mathcal{H}$ is a universal class of hash functions for any finite field, but with respect to our measure different fields behave differently.

If the finite field F has n elements then there is a bad set $S \subset F^2$ of size n with expected maximal bucket size $\Omega(n^{1/3})$. If n is a perfect square then there is even a bad set with largest bucket size *always* at least $\sqrt{n}$. (This is worst possible, since with respect to a universal class of hash functions every set of size n has expected largest bucket size below $\sqrt{n}+1/2$.)

If, however, we consider the field of two elements then we get much better bounds. The best previously known upper bound on the expected size of the largest bucket for this class was $O(2^{\sqrt{\log n}})$. We reduce this upper bound to $O(\log n \log\log n)$. Note that this is not far from the guarantee for a random function. There, the average largest bucket would be $\Theta(\log n/\log\log n)$.

In the course of our proof we develop a tool which may be of independent interest. Suppose we have a subset S of a vector space D over $\mathbf{Z}_2$, and consider a random linear mapping of D to a smaller vector space R. If the cardinality of S is larger than $c_\epsilon |R| \log |R|$ then with probability $1-\epsilon$, the image of S will cover all elements in the range.

1 Introduction

Consider distributing n balls in s buckets, randomly and independently. The resulting distribution of the balls in the buckets is the object of occupancy theory.

*Dep. of Math., Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv, Israel and Institute for Advanced Study, Princeton, NJ 08540. Research supported in part by a USA-Israeli BSF grant, by the Sloan Foundation grant No. 96-6-2, by an NEC Research Institute grant and by the Hermann Minkowski Minerva Center for Geometry at Tel Aviv University. E-mail: noga@math.tau.ac.il.

†Fachbereich Informatik, Lehrstuhl II, Universität Dortmund, D-44221 Dortmund, Germany. E-mail: dietzf@ls2.informatik.uni-dortmund.de. Partially supported by DFG grant Di 412/5-1.

‡BRICS, Centre of the Danish National Research Foundation, University of Aarhus, Ny Munkegade, Aarhus, Denmark. Supported by the ESPRIT Long Term Research Programme of the EU under project number 20244 (ALCOM-IT). E-mail: bromille@brics.dk. Part of this work was done while the author was at the University of Toronto.

§DIMACS, P.O.Box 1179, Piscataway, NJ 08855-1179, USA. E-mail: erez@dimacs.rutgers.edu. Part of this work was done while the author was visiting the University of Toronto.

¶Mathematical Institute of the Hungarian Academy of Sciences, Pf. 127, Budapest, H-1364 Hungary and Institute for Advanced Study, Princeton, NJ 08540. Supported by NSF grants CCR-95-03254 and DMS-9304580, a grant from Fuji Bank and the grant OTKA-F014919. E-mail: tardos@cs.elte.hu. Part of this work was done while the author was visiting the University of Toronto.

In the theory of algorithms and in complexity theory, it is often necessary and useful to consider putting balls in buckets without complete independence. More precisely, the following setting is studied: A class $\mathcal{H}$ of hash functions, each mapping a universe U to $\{1, 2, \ldots, s\}$, is fixed. A set $S \subseteq U$ to be hashed is given by an adversay, a member $h \in \mathcal{H}$ is chosen uniformly at random, S is hashed using h, and the distribution of the multi-set $\{h(x) | x \in S\}$ is studied. If the class $\mathcal{H}$ is the class of all functions between U and $\{1, 2, \ldots, s\}$, we get the classical occupancy problems. Carter and Wegman defined a class $\mathcal{H}$ to be universal if[1]

$$\forall x \neq y \in U : \text{Prob}(h(x) = h(y)) \leq 1/s.$$

For universal families, the following properties are well known and variations of them have been used extensively in various settings

1. If S of size n is hashed to n^2 buckets, with probability more than half, no collision occurs.
2. If S of size $2n^2$ is hashed to n buckets, with probability more than half, every bucket receives an element.
3. If S of size n is hashed to n buckets, the expected size of the largest bucket is less than $\sqrt{n} + \frac{1}{2}$.

The intuition behind universal hashing is that we often loose relatively little compared to using a completely random map. Note that for the property 1, this is true in a very strong sense; even with complete randomness, we do not expect $o(n^2)$ buckets to suffice (the birthday paradox), so nothing is lost by using a universal family instead. The bounds in the second and third properties, however, are rather coarse-grained compared to what one would get with complete randomness. For property 2, with complete randomness, $\Theta(n \log n)$ balls would suffice to cover the buckets with good probability (the coupon collector's theorem), i.e. a polynomial improvement over n^2, and for property 3, with complete randomness, we expect the largest bucket to have size $\Theta(\log n / \log \log n)$, i.e. an exponential improvement over $\sqrt{n}$. In these last cases we do seem to loose quite a lot compared to using a completely random map and better bounds would seem desirable. However, it is rather easy to construct (unnatural) families of universal families and sets to be hashed showing that size $\Theta(n^2)$ is necessary to cover n buckets with non-zero probability, and that buckets of size $\sqrt{n}$ are, in general unavoidable, when a set of size n is hashed to n buckets. This shows that the *abstract* property of universality does not allow for stronger statements. Now fix a *concrete* universal family of hash functions. We ask the following question: *To which extent are the fine-grained occupancy properties of completely random maps preserved?*.

We provide answers to these questions for the case of *linear maps* between two vector spaces over a finite field, a natural and well known class of universal hash functions. The general flavour of our results is that "large fields are bad", in the sense that the bounds becomes the worst possible for universal families, while "small fields are good" in the sense that the bounds become as good or almost as good as the ones for independently distributed balls.

More precisely, for the covering problem, we show the following (easy) theorem

Theorem 1 *Let F be a field of size n and let $\mathcal{H}$ be the class of linear maps between F^2 and F. There is a subset S of F^2 of size $\Theta(|F|^2)$, so that for no $h \in \mathcal{H}$, $h(S) = F$.*

On the other hand, we prove the following harder theorem

[1] We remark that a stricter definition is often used in the complexity theory literature, and a more liberal definition is often used in the data structure literature.

Theorem 2 *Let S be a subset of a vector space over $\mathbf{Z}_2$ and choose a random linear map to a smaller vector space R. If $|S| \geq c_\epsilon |R| \log |R|$ then with probability at least $1-\epsilon$ the image of S covers the entire range R.*

For the "largest bucket problem", let us first introduce some notation: Let U be the universe from which the keys are chosen. We fix a class $\mathcal{H}$ mapping U to $\{1, \ldots, s\}$. Then, a set $S \subseteq U$ of size n is chosen *by an adversary*, and we uniformly at random pick a hash function $h \in \mathcal{H}$, hash S using h and look at the size of the largest resulting hash bucket. We denote the expectation of this size by L_n^s. Formally,

$$L_n^s(\mathcal{H}) = \max_{S \subseteq U, |S|=n} E_{h \in \mathcal{H}}[\max_{y \in \{1,\ldots,s\}} |\{x \in S \mid h(x) = y\}|]$$

Usually we think of s being of size close to n. Note that if $s = \Omega(n^2)$, any universal class yields $L_n^s = O(1)$.

The class $\mathcal{H}$ we will consider is the set of linear maps between $F^m \to F^k$ for $m > k$. Here F is a finite field and $s = |F|^k$. This class is universal for all values of the parameters.

When $k = 1$ and thus $|F| = s$ the expected largest bucket can be large.

Theorem 3 *Let F be a finite field with $|F| = s$. For the class $\mathcal{H}$ of all linear transformations $F^2 \to F$ we have*

$$L_s^s(\mathcal{H}) = \Omega(s^{1/3}).$$

Furthermore if $|F|$ is a perfect square we have

$$L_s^s(\mathcal{H}) > \sqrt{s}.$$

Note how close our lower bound for quadratic fields is to the upper bound of $\sqrt{s} + 1/2$ that holds for every universal class. We also mention that for the bad set we construct in Theorem 8 for a quadratic field there is no good hash function, since there *always* exists a bucket of size at least $\sqrt{s}$.

When the field is the field of two elements, the situation is completely different. Markowsky, Carter and Wegman [MCW78] showed that for this case $L_s^s(\mathcal{H}) = O(s^{1/4})$. Mehlhorn and Vishkin [MV84] improved on this result (although this is implicit in their paper) and showed that $L_s^s(\mathcal{H}) = O(2^{\sqrt{\log s}})$. We further improve the bound and show that:

Theorem 4 *For the class $\mathcal{H}$ of all linear transformations between two vector spaces over $\mathbf{Z}_2$,*

$$L_s^s(\mathcal{H}) = O(\log s \log\log s).$$

Furthermore, we also show that even if the range is smaller than $|S|$ by a logarithmic factor, the same still holds:

Theorem 5 *For the class $\mathcal{H}$ of all linear transformations between two vector spaces over $\mathbf{Z}_2$,*

$$L_{s\log s}^s(\mathcal{H}) = O(\log s \log\log s).$$

Note that even if one uses the class $\mathcal{R}$ of *all* functions one obtains only a slightly better result: the expected size of the largest bucket in this case is $L_s^s(\mathcal{R}) = \Theta(\log s / \log\log s)$ and $L_{s\log s}^s(\mathcal{R}) = \Theta(\log s)$, which is the best possible bound for any class. Interestingly, our upper bound is based on our upper bound for the covering property.

We do not know what the right bound is for the class of linear maps over $\mathbf{Z}_2$, i.e., is it as good as $O(\log s / \log\log s)$? We leave this as an open question.

1.1 Motivation

There is no doubt that the method of implementing a dictionary by hashing with chaining, recommended in textbooks [CLR90, GBY90] especially for situations with many update operations, is a practically important scheme.

In situations in which a good bound on the cost of single operations is important, e. g., in real-time applications, the expected maximal bucket size as formed by all keys ever present in the dictionary during a time interval plays a crucial role. Our results show that, at least as long as the size of the hash table can be determined right at the start, using a hash family of linear functions over $\mathbf{Z}_2$ will perform very well in this respect. For other simple hash classes such bounds on the worst case bucket size are not available or are even wrong (see Theorem 8); other, more sophisticated hash families [S89, DM90, DGMP92] that do guarantee small maximal bucket sizes consist of functions with higher evaluation time. Of course, if worst case constant time for certain operations is absolutely necessary, the known two-level hashing schemes can be used, e. g., the FKS scheme [FKS84] for static dictionaries; dynamic perfect hashing [DKMHRT94] for the dynamic case with constant time lookups and expected time $O(n)$ for n update operations; and the "real-time dictionaries" from [DM90] that perform each operation in constant time, with high probability. It should be noted, however, that a price is to be paid for the guaranteed constant lookup time in the dynamic schemes: the (average) cost of insertions is significantly higher than in simple schemes like chained hashing; the overall storage requirements are higher as well.

1.2 Related work

Another direction in trying to show that a specific class has a good bound on the expected size of the largest bucket is to build a class specifically designed to have such good property.

One immediate such result is obtained by looking at the class of d-degree polynomials over finite fields, where $d = c \log n / \log\log n$ (see, e.g., [ABI86].) It is easy to see that this class maps each d elements of the domain independently to the range, and thus, the bound that applies to the class of all functions also applies to this class. We can combine this with the following well known construction, (which is usually called "collapsing the universe") : At the expense of $\Theta(\log n + \log\log |U|)$ random bits one can construct a map $g : U \to \{1, \ldots, n^{k+2}\}$ that on any set S will be one-to-one with probability $1 - O(1/n^k)$. This gives us a class with $L_n^n = \Theta(\log n / \log\log n)$ of size $2^{O(\log\log |U| + \log^2 n / \log\log n)}$ and with evaluation time $O(\log n / \log\log n)$ in a reasonable model of computation, say, a RAM with unit cost operations on members of the universe to be hashed.

More efficient (but much larger) families where given by Siegel [S89] and by Dietzfelbinger and Meyer auf der Heide [DM90]. Both provide families of size $|U|^{n^\epsilon}$ such that the functions can be evaluated in O(1) time on a RAM and with $L_n^n = \Theta(\log n / \log\log n)$. The families from [S89] and [DM90] are somewhat complex to implement while the class of linear maps requires only very basic bit operations (as discussed already in [CW79]). It is therefore desirable to study this class, and this is the main purpose of the present paper.

1.3 Notation

If S is a subset of the domain D of a function h we use $h(S)$ to denote $\{h(s) \mid s \in S\}$. If x is an element of the range we use $h^{-1}(x)$ to denote $\{s \in D \mid h(s) = x\}$. If A and B are subsets of a vector space V and $x \in V$ we use the notations $A + B = \{a + b \mid a \in A \wedge b \in B\}$ and $x + A = \{x + a \mid a \in A\}$. We use $\mathbf{Z}_2$ to denote the two element field. All logarithms in this paper are base two.

2 The covering property

2.1 Lower bounds for covering with a large field

We prove Theorem 1. Take a set $A \subset F$ of size $|A| = \lfloor |F|/2 \rfloor$ and consider $S = \{(x,y) \mid y \neq 0 \wedge x/y \in A \wedge (x-1)/y \notin A\}$. S has density around one quarter and no linear map $g : F^2 \to F$ satisfies $g(S) = F$. To see this take a nonzero linear map $g : (x,y) \mapsto ax + by$ and note that if $0 \in g(S)$ then $a \neq 0$ and $-b/a \in A$ but in this case $a \notin g(S)$.

2.2 Upper bounds for covering with a small field - the existential case

We start by showing that if we have a large enough subset A of a vector space over $\mathbf{Z}_2$ then *there exists* a linear transformation T to a large vector space such that $T(A)$ is the entire range. The constant e below is the base of the natural logarithm.

Theorem 6 *Let A be a finite set of vectors in a vector space V of an arbitrary dimension over $\mathbf{Z}_2$ and let $t > 0$ be an integer. If $|A| > t2^t/\log e$ then there exists a linear map $T : V \to \mathbf{Z}_2^t$, so that T maps A onto $\mathbf{Z}_2^t$.*

For the poof of this theorem we need the following simple lemma. Note that although we state the lemma for vector spaces, it holds for any finite group.

Lemma 2.1 *Let V be a finite vector space, $A \subseteq V$, $\alpha = 1 - |A|/|V|$. Then for a random $v \in V$ it holds that*

$$E_v(1 - |A \cup (v + A)|/|V|) = \alpha^2.$$

Proof. If v and u are both chosen uniformly independently at random from V then both events $u \notin A$ and $u \notin v + A$ have probability α and they are independent. □

Proof of Theorem 6. Let m be the dimension of V, $N = |A|$ and $\alpha = 1 - |A|/|V| = 1 - N/2^m$. Starting with $A_0 = A$, we choose a vector $v_1 \in V$ so that for $A_1 = A_0 \cup (v_1 + A_0)$

$$1 - \frac{|A_1|}{|V|} \leq \alpha^2.$$

Such a choice for v_1 exists by Lemma 2.1. Then, by the same procedure, we choose a v_2 so that for

$$A_2 = A_1 \cup (v_2 + A_1) = A + \mathrm{Span}\{v_1, v_2\},$$

$$1 - \frac{|A_2|}{|V|} \leq \alpha^4,$$

and so on up to $A_s = A + \mathrm{Span}\{v_1, \ldots, v_s\}$ with $s = m - t$ for which

$$1 - \frac{|A_s|}{|V|} \leq \alpha^{2^s}.$$

Note that one can assume that the vectors $v_1, \ldots, v_s$ are linearly independent since choosing a vector v_i which depends on the previous choices makes $A_i = A_{i-1}$.

We now claim that $A_s = V$. Suppose in way of contradiction that $A_s \neq V$. Take a vector $v \in V$ outside A_s, then $v + \mathrm{Span}\{v_1, \ldots, v_s\}$ is disjoint from A_s thus

$$1 - \frac{|A_s|}{|V|} \geq 2^s/2^m.$$

Combining the last two inequalities and recalling that $t = m - s$ we get

$$2^{-t} \leq \alpha^{2^s} = (1 - N/2^m)^{2^s} < e^{-N2^{-t}}.$$

Taking the logarithm of both sides yields $-t < -N \log_2 e/2^t$ contradicting our assumption on N. Thus, $A_s = V$ must hold.

Now we choose a linear map $T : V \to \mathbf{Z}_2^t$ such that its kernel $T^{-1}(0) = \text{Span}\{v_1, \ldots, v_s\}$. The equality $V = A_s = A + \text{Span}\{v_1, \ldots, v_s\}$ implies that T maps A onto $\mathbf{Z}_2^t$. □ The bound in Theorem 6 is assimptotically tight as shown by the following proposition.

Proposition 2.2 *For every large enough integer t there is a set A of at least $(t - 3\log t)2^t/\log e$ vectors in a vector space V over $\mathbf{Z}_2$ so that for any linear map $T : V \to \mathbf{Z}_2^t$, T does not map A onto $\mathbf{Z}_2^t$.*

Proof: Let $V = Z_2^{t+s}$ with $s = \lfloor t/10 \rfloor$ and let A be chosen at random by picking each element of V independently and with probability $p = 1 - 2^{-x}$ into the set with $x = (t - 2\log t)2^{-s}$. ¿From the Chebyshev inequality we know that with probability at least $3/4$, A has cardinality at least $pN - 2\sqrt{pN}$ for $N = 2^{s+t}$. Using $p > x/\log e - x^2/(2\log^2 e)$ one can show that this is as many as claimed in the proposition. Let us compute the probability that there exists a linear map $T : V \to \mathbf{Z}_2^t$ such that T maps A onto $\mathbf{Z}_2^t$. There are $2^{t(t+s)}$ possible maps and each of them covers $\mathbf{Z}_2^t$ with probability at most $\left(1 - (1-p)^{2^s}\right)^{2^t} = \left(1 - 2^{-2^s x}\right)^{2^t} = (1 - t^2/2^t)^{2^t} < e^{-t^2}$. So with probability almost $3/4$, A is not small and still cannot cover. □

2.3 Choosing the linear map at random

In this subsection we strengthen Theorem 6 and prove that if A is bigger by a constant factor, then almost all choices of the linear transformation T work. This may seem immediate at first glance since Lemma 2.1 tells us that a random choice for the next vector is good on average. In particular, it might seem that for a random choice of v_1 and v_2 in the proof of Theorem 6, $E_{v_1,v_2}(1 - |A + \text{Span}\{v_1, v_2\}|/|V|) \leq \alpha^4$. Unfortunately this is not the case: For example, think of A being a linear subspace containing half of V. In this case, the ratio α of points that are not covered is $1/2$. As random vectors v_i are chosen to be added to A, vectors in A are chosen with probability $1/2$. Thus, after i steps, α remains $1/2$ with probability $1/2^i$ and becomes 0 otherwise. Thus, the expected value of α_i is 2^{-i-1} which is much bigger than 2^{-2^i}.

Theorem 7 *For every $\epsilon > 0$ there is a constant $c_\epsilon > 0$ such that the following holds. Let A be a finite set of vectors in a vector space V of an arbitrary dimension over $\mathbf{Z}_2$, let $t > 0$ be an integer and $0 < \epsilon < 1$. If $|A| \geq c_\epsilon t 2^t$ then for a uniform random linear transformation $T : V \to \mathbf{Z}_2^t$*

$$\text{Prob}(T(A) = \mathbf{Z}_2^t) \geq 1 - \epsilon.$$

Proof: Let $N = |A|$, $m = \lceil \log_2(3N/\epsilon) \rceil$, and $s = m - t$. Consider the vector space $W = \mathbf{Z}_2^m$. Instead of choosing the transformation T uniformly at random, we choose T in two steps, which imply the same distribution on T: First, we pick uniformly at random a linear transformation $T_0 : V \to W$. Then, we pick a random *onto* (full rank) linear map $T_1 : W \to \mathbf{Z}_2^t$, and set $T = T_0 \circ T_1$. This results in a uniformly chosen linear map $T : V \to \mathbf{Z}_2^t$. In order to pick the onto map T_1 we use the following process (similar to the one in the proof of Theorem 6). Pick s vectors $v_1, \ldots, v_s$ uniformly at random from the vectors in W and choose $T_1 : W \to \mathbf{Z}_2^t$ to be a random onto linear transformation $T_1 : W \to \mathbf{Z}_2^t$ with the constraints $T_1(v_i) = 0$ $(i = 1, \ldots, s)$, i.e. the vectors

$v_1, \ldots, v_s$ are in the kernel of T_1. Note that the v_i's are not necessarily linearly independent and that they are not necessarily all the vectors in the kernel. After picking them, an *onto* transformation satisfying that the v_i's are in the kernel is chosen at random. The transformation T_1 is indeed distributed uniformly at random amongst all onto linear maps of W onto $\mathbf{Z}_2^t$ since each possible kernel is chosen with equal probability, and given the kernel, the choice of the transformation is uniform.

Using notations similar to the ones used in the proof of Theorem 6 (but note the difference in the definition of A_0), let $A_0 = T_0(A)$, $A_i = A_0 + \mathrm{Span}\{v_1, \ldots, v_i\}$ and $\alpha_i = 1 - |A_i|/|W|$ for $i = 0, \ldots, s$. Observe that since $|W| = 2^m$ and $\alpha_s = 1 - |A_s|/|W|$, showing $\alpha_s < 1/2^m$ implies $A_s = W$. In this case (as in the proof of Theorem 6) $T_1(A_0) = \mathbf{Z}_2^t$ and therefore $T(A) = \mathbf{Z}_2^t$. So we only have to show that

$$\mathrm{Prob}(\alpha_s < 2^{-m}) \geq 1 - \epsilon$$

where the probability is taken over the choices of T_0 and T_1.

We first claim that $\mathrm{Prob}(|A_0| \leq N/2) < \epsilon/3$. Since the transformation T_0 maps any two distinct vectors from V to the same point with probability $2^{-m} < \epsilon/(3N)$, the expected number of such colliding pairs from A is $\binom{N}{2}/2^m < \epsilon N/6$. In the event $|A_0| \leq N/2$, we have at least $N/2$ such colliding pairs, thus this event happens with probability at most $\epsilon/3$. We regard this small chance event an "error" and concentrate on the $|A_0| > N/2$ case. In this case, we have $\alpha_0 < 1 - N/2^{m+1} < 1 - \epsilon/12$.

Clearly α_i is nonnegative and monotone decreasing in i. We want to measure how fast α_i is decreasing. Imagine the vectors $v_1, \ldots, v_s$ being chosen one by one and consider $v_1, \ldots, v_i$ fixed and thus α_i determined. By Lemma 2.1

$$E(\alpha_{i+1}) = \alpha_i^2. \tag{1}$$

Following the (simple) idea in [ABM87], we want to say that the choice of v_{i+1} is successful if α_{i+1} is small enough with respect to α_i. The decrease of α_i is different depending on α_i being greater or less than $1/2$. Therefore, let us call the choice of v_{i+1} *successful* if either $\alpha_i \geq 1/2$ and $1 - \alpha_{i+1} \geq 5(1 - \alpha_i)/4$ or if $\alpha_i < 1/2$ and $\alpha_{i+1} \leq 3\alpha_i/4$.

We first claim that given any choice of $v_1, \ldots, v_i$ the choice of v_{i+1} is successful with probability more than $1/3$. For $\alpha_i < 1/2$ this is immediate from Equation (1). In case $\alpha_i \geq 1/2$ the expected decrease is still the same, and in order to apply Markov's Inequality, one has to notice that since $A_{i+1} = A_i \cup (A_i + v_{i+1})$ we have $|A_{i+1}| \leq 2|A_i|$ and thus $1 - \alpha_{i+1} \leq 2(1 - \alpha_i)$.

Now we have a series of s decreasing numbers $\{\alpha_i\}_{i=1}^s$ which decrease "a lot" in each step with high probability. We will show that with high probability α_s is quite small, i.e., less than 2^{-m}. Let $C \stackrel{\mathrm{def}}{=} 12/\epsilon$. We partition the analysis into two parts, first checking the probability that α_i gets from as high as $1 - 1/C$ to below $1/(2C^2)$ in $s_0 \stackrel{\mathrm{def}}{=} \lfloor 38 \log C \rfloor$ steps. Then, we check the probability that in the remaining $s - s_0$ steps we get that $\alpha_s < 2^{-m}$.

Applying Chernoff's bound (similarly to Corollary A.7 in Appendix A of [AS92]) implies that the probability that we do not have at least $9 \log C$ successful choices out of the choices of $v_1, \ldots, v_{s_0}$ is at most $\epsilon/3$.

We regard this small probability event as another error and we concentrate on the case none of the two errors happens. In that case we have $\alpha_0 < 1 - 1/C$ and the $9 \log C$ successful choices reduce this number below $1/(2C^2)$ thus we have $\alpha_{s_0} < 1/(2C^2)$.

Finally for $i = 1, \ldots, s - s_0$ we call the choice of v_{s_0+i} a *failure* if $\alpha_{s_0+i} > C^i \alpha_{s_0+i-1}^2$. By Markov's Inequality and Equation (1) we get that for any choice of $v_1, \ldots, v_{s_0+i-1}$ the failure probability of v_{s_0+i} is less than $1/C^i$. Thus, the probability of ever encountering this kind of failure is below $\epsilon/3$.

So if failure never occurs one has

$$C^{i+2}\alpha_{s_0+i} \leq (C^{i+1}\alpha_{s_0+i-1})^2$$

for every $i = 1, \ldots, s - s_0$. Thus, we must have

$$\alpha_s < C^{s-s_0+2}\alpha_s < (C^2\alpha_{s_0})^{2^{s-s_0}}.$$

If in addition none of the first two type of errors occurs we have $C^2\alpha_{s_0} < 1/2$ thus $\alpha_s < 2^{-2^{s-s_0}}$.

So to prove $\text{Prob}(\alpha_s < 2^{-m}) \geq 1 - \epsilon$ as needed we only have to show that $2^{s-s_0} \geq m$. Using the formulae defining s, s_0 and m this is implied by the bound $N \geq C^{38}t2^t$. Thus, the choice $c_\epsilon = C^{38}$ satisfies the statement of the theorem. □

3 The largest bucket

3.1 Lower bound for the largest bucket with a large field

We start by showing why linear hashing over a large finite field is bad with respect to our expected largest bucket size measure. This natural example shows that universality of the class is not enough to assure small buckets. For a finite field F we prove the existence of a bad set $S \subset F^2$ of size $|S| = |F|$ such that the expected largest bucket in S with respect to random linear map $F^2 \to F$ is big. We prove the results in Theorem 3 separately for quadratic and non-quadratic fields.

Before the proofs we start with an intuitive description of the constructions. Linear hashing of the plane collapses all straight lines of a random direction. Thus, a bad set in the plane must contain many points on at least one line in many different directions. It is not hard to come up with bad sets that contain many points of many different lines, however the obvious constructions (subplane or grid) yield sets where many of the "popular lines" tend to be parallel and thus they only cover a few directions. This problem can be solved by a projective transformation, the transformed set has many popular lines, but they are no longer parallel.

Theorem 8 *Let F be a finite field with $|F|$ being a perfect square. There exists a set $S \subset F^2$ of size $|S| = |F|$ such that for every linear map $h : F^2 \to F$, S has a large bucket, i.e. there exists a value $y \in F$ with $|h^{-1}(y)| \geq \sqrt{|F|}$.*

Proof. We have a finite field F_0 of which F is a quadratic extension. Let $|F_0| = m$ and $|F| = m^2 = n$. Let a be an arbitrary element in $F \setminus F_0$ and define $S = \{(\frac{1}{x+a}, \frac{y}{x+a}) \mid x, y \in F_0\}$. Note that $|S| = m^2 = |F|$. Notice also, that S is the image of the subplane F_0^2 under the projective transformation $(x, y) \mapsto (\frac{1}{x+a}, \frac{y}{x+a})$.

Fix $A, B \in F$ and consider the function $h : F^2 \mapsto F$ defined by $h(x, y) = Ax + By$. We must show that there is some $C \in F$ such that $|h^{-1}(C) \cap S| \geq m$. If $B = 0$ then h maps all the m elements of $S' = \{(1/a, y/a \mid y \in F_0\}$ to $C = A/a$, as needed. Otherwise, we claim that there is a $C \in F$ such that both $\frac{C}{B}$ and $\frac{aC-A}{B}$ are in F_0. To see this observe that if g_1 and g_2 are two distinct members of F_0, then ag_1 and ag_2 lie in distinct additive cosets of F_0 in F, since otherwise their difference, $a(g_1 - g_2)$ would have been in F_0, contradicting the fact that $a \notin F_0$. Thus, as g ranges over all members of F_0, ag intersects distinct additive cosets of F_0 in F, and hence aF_0 intersects all those cosets. In particular, there is some $g \in F_0$ so that $ag \in F_0 + \frac{A}{B}$, implying that $C = gB$ satisfies the assertion of the claim. For the above C, for every choice of $x \in F_0$, $y(x) = \frac{C}{B}x + \frac{aC-A}{B} \in F_0$. We have now $A\frac{1}{a+x} + B\frac{y(x)}{a+x} = C$, showing that h maps all the m elements of $S' = \{(\frac{1}{a+x}, \frac{y(x)}{a+x}) \mid x \in F_0\} \subset S$ to C. □

Theorem 9 *Let F be a finite field. There exists a set $S \subset F^2$ of size $|S| = |F|$ such that for more than half of the linear maps $h : F^2 \to F$, S has a large bucket, i.e. there exists a value $y \in F$ with $|h^{-1}(y)| \geq |F|^{1/3}/3 - 1$.*

Proof. First we construct a set $S' \subset F^2$ such that $|S'| \leq |F| = n$ and there are n distinct lines in the plane F^2 each containing at least $m \geq n^{1/3}/3$ points of S'.

Let us first consider the case when n is a prime, so F consists of the integers modulo n. We let $A = \{i \mid 1 \leq i < \sqrt{n}\}$ and consider the square grid $S' = A \times A$. Clearly $|S'| < n$. It is well known that each of the n most popular lines contains at least $m \geq n^{1/3}/3$ points of S'. This is usually proved for the same grid in the Euclidean plane (see e.g. [PA95], pp. 178–179) but that result implies the same for our grid in F^2.

Now let $n = p^k$ and let F_0 be the subfield in F of p elements. Let $x \in F$ be a primitive element, then every element of F can be uniquely expressed as a polynomial of x of degree below k with coefficients from F_0. Let $k_1 = \left\lfloor \frac{k+1}{3} \right\rfloor$, $k_2 = k - k_1 = \left\lfloor \frac{2k+1}{3} \right\rfloor$ and let $A_1 = \{f(x) \mid \deg(f) < k_1\}$, $A_2 = \{f(x) \mid \deg(f) < k_2\}$ where the polynomials f have coefficients from F_0. Finally we take $S' = A_1 \times A_2$. Clearly $|S'| = n$. For $a \in A_1$ and $b \in A_2$ we consider the line $L_{a,b} = \{(y, ay + b) \mid y \in F\}$ in F^2. Notice that there are n such lines and we have $ay + b \in A_2$ whenever $y \in A_1$. Thus, we have n distinct lines each containing $m = |A_1| = p^{k_1}$ points of S'. We have $m \geq n^{1/3}$ as claimed unless $k \equiv 1 \pmod 3$. Notice that for $k \equiv 2 \pmod 3$ our m is much higher than $n^{1/3}$. For the bad case $k \equiv 1 \pmod 3$ we apply the construction below instead.

Finally suppose $n = p^k$, p is a prime and $k \equiv 1 \pmod 3$. To get our set S' in this case we have to merge the two constructions above. Let F_0 be the p element subfield of F, then F_0 consists of the integers modulo p. We set $A = \{i \mid 1 \leq i \leq< p\}$. Let $k_1 = (k + 2)/3$ and $k_2 = (2k + 1)/3$ and let $x \in F$ be a primitive element, so we can express any element of F uniquely as a polynomial of x of degree less then k with coefficients from F_0. We set $A_1 = \{f(x) \mid \deg(f) < k_1 \wedge f(0) \in A\}$, $A_2 = \{f(x) \mid \deg(f) < k_2 \wedge f(0) \in A\}$ where the polynomials f have coefficients from F_0. Finally we set $S' = A_1 \times A_2$. Clearly $|S'| < n$. For $j, j' \in F_0$ let $L_{j,j'} = \{(i, ji + j') \mid i \in F_0\}$. Let a and b be polynomials with coefficients from F_0 with $\deg(a) < k_1$ and $\deg(b) < k_2$. Consider the line $L_{a,b} = \{(y, a(x)y + b(x)) \mid y \in F\}$. Notice that $|L_{a,b} \cap S'| = p^{k_1 - 1}|L_{a(0),b(0)} \cap (A \times A)|$. Thus, from knowing that the p most popular lines in F_0^2 contain at least $m_0 \geq p^{1/3}/3$ points from $A \times A$ we conclude that there exist n distinct lines each containing at least $m = m_0 p^{k_1 - 1} \geq n^{1/3}/3$ points of S'.

In all cases now we have constructed our set $S' \subset F^2$ of size $|S'| \leq n$ with n distinct popular lines each containing at least $m > n^{1/3}/3$ points of S'. Let P be the projective plane containing F^2. Out of the $n^2 + n + 1$ points in P every popular line covers $n + 1$. The ith popular line ($1 \leq i \leq n$) can only have $i - 1$ intersections with earlier lines, thus it covers at least $n + 2 - i$ points previously uncovered. Therefore a total of at least $\binom{n+2}{2} - 1$ points are covered by popular lines. Simple counting gives the existence of a line L in P not among the popular lines, such that more than half of the points on L are covered by popular lines. Let f be a projective transformation taking the ideal line $L' = P \setminus F^2$ to L. We define $S = \{x \in F^2 \mid f(x) \in S'\}$. Clearly $|S| \leq |S'| \leq n$.

One linear hash function $h : F^2 \to F$ is constant zero (and thus all of S is a single bucket), for the rest there is a point $x_h \in L'$ such that h collapses the points of F^2 of each single line going through x_h. As we get all points of L' equal number of times ($n - 1$ times to be precise) it is enough to prove that whenever $f(x_h)$ is covered by a popular line S has a big bucket with respect to h as claimed. So suppose $f(x_h)$ is in the popular line L''. As L'' contains at least m points of S' thus $f^{-1}(L'')$ must be a line through x_h containing at least $m - 1$ points of S (the -1 comes from the possibility of $f(x_h) \in S'$). This ensures a bucket of S with respect to h of size at least $m - 1$ as

claimed. □

3.2 Upper bound for the largest bucket with a small field

Let us now recall and prove our main result.

For convenience here we speak about hashing $n \log n$ keys to n values. Also, we assume that n is a power of 2.

Theorem 5: Let $\mathcal{H}$ be the class of linear transformations between two vector spaces over $\mathbf{Z}_2$, then

$$L^n_{n \log n}(\mathcal{H}) = O(\log n \log \log n).$$

This theorem implies Theorem 4.

We have to bound the probability of the event that many elements in the set S are mapped to a single element in the range. Denote this bad event by E_1. The overall idea is to present another (less natural) event E_2 and show that the probability of E_2 is small, yet the probability of E_2 given E_1 is big. Thus, the probability of E_1 must be small. We remark here that a somewhat similar line of reasoning was used in the seminal paper of Vapnik and Chervonenkis [VC71].

For the proof we fix the domain to be $D = \mathbf{Z}_2^m$, the range (the buckets) to be $B = \mathbf{Z}_2^{\log n}$, and $S \subset D$ of size $|S| = n \log n$.

Let us choose arbitrary $\ell > \log n$ and consider the space $A = \mathbf{Z}_2^\ell$. We construct the linear transformation $h : D \to B$ through the intermediate range A in the following way. We choose uniformly at random a linear transformation $h_1 : D \to A$ and uniformly at random an onto linear transformation $h_2 : A \to B$. Now we define $h \stackrel{\text{def}}{=} h_1 \circ h_2$. Note that as mentioned in the proof of Theorem 7 this yields an h which is uniformly chosen between all linear transformations from D to B.

Let us fix a threshold t. We define two events. E_1 is the existence of a bucket of size at least t:

Event E_1: There exists an element $\alpha \in B$ such that

$$\left| h^{-1}(\alpha) \cap S \right| > t.$$

We are going to limit the probability of E_1 through the seemingly unrelated event E_2:

Event E_2: There exists an element $\alpha \in B$ such that

$$h_2^{-1}(\alpha) \subseteq h_1(S).$$

Consider the distribution space in which h_1 and h_2 are uniformly chosen as above. We shall show that

Proposition 3.1 *If for some $\epsilon > 0$ it holds that $2^\ell \geq (c_\epsilon + 1) n \log n$ (for the c_ϵ guaranteed by Theorem 7) then*

$$\text{Prob}[E_2] \leq \epsilon.$$

Proposition 3.2 *If $t > c_{1/2}(2^\ell/n) \log(2^\ell/n)$ (with $c_{1/2}$ from Theorem 7) then*

$$\text{Prob}[E_2 | E_1] \geq \frac{1}{2}.$$

From Propositions 3.1 and 3.2 we deduce that the probability of E_1 must be small:

Corollary 3.3 *For every $\epsilon > 0$ there is a constant d_ϵ such that for every n (which is a power of 2) a random linear transformation hashing a subset S of a $\mathbf{Z}_2$ vector space of size $|S| = n \log n$ to $\mathbf{Z}_2^{\log n}$ we have*

$$\text{Prob}[\textit{maximum bucket size} \geq d_\epsilon \log n \log\log n] < \epsilon.$$

Proof: We set $\ell = \lceil \log n + \log\log n + \log(c_\epsilon + 1) \rceil$ to satisfy the condition of Proposition 3.1. Then we set $t(n) = \lceil c_{1/2}(2^\ell/n) \log(2^\ell/n) \rceil = O(\log n \log\log n)$ to satisfy the condition of Proposition 3.2. Thus, we have $\text{Prob}[E_1] \leq 2\epsilon$. Thus any choice for $d_{2\epsilon}$ satisfying $t(n) \leq d_{2\epsilon} \log n \log\log n$ is good. Since $\epsilon > 0$ was arbitrary this proves our corollary. □

Let us now prove the propositions above.

Proof of Proposition 3.1: Note first that an alternative way to describe E_2 is

$$h_2(A \setminus h_1(S)) \neq B.$$

We will prove that Proposition 3.1 holds for any specific h_1, and thus it also holds for a randomly chosen h_1. So fix h_1 and consider the distribution in which h_2 is chosen uniformly amongst all full rank linear transformation from A to B.

Clearly, $|h_1(S)| \leq |S| = n \log n$ and thus $|A \setminus h_1(S)| \geq 2^\ell - n \log n \geq c_\epsilon n \log n$. By Theorem 7, a uniformly chosen h_2 does not satisfy $h_2(A \setminus h_1(S)) = B$ with probability at most ϵ, and we are done with the proof of Proposition 3.1. Note that in Theorem 7 one chooses a random element from all linear maps and here we choose a full rank map but this only decreases the probability of not covering since none of the excluded maps can cover. □

Proof of Proposition 3.2: Fix h for which E_1 holds, and fix any full rank h_2. We will show that the probability of event E_2 is at least $1/2$ even when these two are fixed and thus the conditional probability is also at least $1/2$.

Now since E_1 holds there is a subset $S' \subseteq S$ of cardinality at least t mapped by h to a single element $\alpha \in \mathbf{Z}_2^{\log n}$. Fix this α and define $D' \stackrel{\text{def}}{=} h^{-1}(\alpha)$ and $A' \stackrel{\text{def}}{=} h_2^{-1}(\alpha)$. Consider the distribution of h_1 satisfying $h = h_1 \circ h_2$. When we restrict h_1 to D', we get that the distribution implied by such h_1 is a uniform choice of an affine or linear map from D' into A' (we show this in Proposition 3.6 below). For event E_2 to hold it is enough to have $A' \subseteq h_1(S)$. We will show that $h_1(S')$ covers all the points in A' with probability at least $1/2$ and thus we get that event E_2 happens with probability $1/2$. Since h_2 is onto we have $|A'| = 2^\ell/n$. On the other hand, $D' \cap S$ has cardinality at least $t = \lceil c_{1/2}(2^\ell/n) \log(2^\ell/n) \rceil$. By Theorem 7, the probability that a set of cardinality t mapped by a random linear transformation will cover a range of cardinality $2^\ell/n$ is at least $1/2$. (Note that Theorem 7 clearly applies to a random affine transformation too.) □

At this point, we have proven Corollary 3.3. This limits the probability of large buckets with linear hashing. Unfortunately the bound there is not strong enough to imply Theorem 5. We strengthen this bound below (see Corollary 3.5) to get the implication. This improvement has little practical value as people want the largest bucket to be small with high probability and this is already guaranteed by Corollary 3.3.

We need to improve the dependence of d_ϵ in Corollary 3.3 on ϵ. One way of doing this would be to study and improve the dependence of c_ϵ on ϵ in Theorem 7. Although this dependence can be vastly improved, one cannot hope to get a $c_\epsilon = o(1/\epsilon)$ and thus Theorem 7 cannot be deduced this way. To solve this problem we notice that in Proposition 3.1 we use Theorem 7 in a very special case. Namely, by the setting of all parameters, we actually use the theorem when the set hashed is quite big, and therefore we can get a much better bound on the dependence of c_ϵ on ϵ as follows.

Proposition 3.4 *Let t and m be positive integers with $c = m - t - \log t \geq 4$. Consider the vector spaces $U = \mathbf{Z}_2^m$ and $V = \mathbf{Z}_2^t$ and let $A \subset U$ satisfy $|U \setminus A| \leq t2^t$. Then for a uniform random onto linear transformation $T : U \to V$ we have*

$$\text{Prob}[T(A) \neq V] < 2^{-c^2/2+3c}.$$

Proof: The proof is similar but simpler than that of Theorem 7.

We start by observing that if $m > 2t + \log t$ then for *every* onto linear transformation $T : U \to V$ we have $T(A) = V$ so we may assume $m \leq 2t + \log t$.

Let $s = m - t$ and choose s independent vectors $x_i \in U$, $i = 1, 2, \ldots, s$ uniformly and independently at random. Then choose T uniformly at random among the onto linear transformations $T : U \to V$ with $T(x_i) = 0$ for $i = 1, \ldots, s$. This process yields the uniform distribution on all onto linear transformations $T : U \to V$.

Consider the sets $A_i = A + \text{Span}\{x_1, \ldots, x_i\}$ and the numbers $\alpha_i = 1 - |A_i|/|U|$ for $i = 0, \ldots, s$. We shall prove that α_s is small with high probability.

As in the proof of Theorem 7 we imagine choosing the vectors x_i one by one. After choosing x_i the number α_i is determined and by Lemma 2.1 for the random choice of x_{i+1} we have

$$E(\alpha_{i+1}) = \alpha_i^2.$$

Let us take $C = 2^{c/2-1}$ and define the choice of x_{i+1} a *failure* if $\alpha_{i+1} \geq C^{j_i+1}\alpha_i^2$ where j_i is the number of indices $1 \leq j \leq i$ for which the choice of x_j was *not* a failure.

Clearly for any choices of $x_1, \ldots, x_i$ the subsequent choice of x_{i+1} is a failure with probability at most C^{-j_i-1}. Thus, the probability of the first failure to ever occur is at most $\sum_{k=1}^{s} C^{-k} < 2/C$. Similarly after any failure the probability of the next failure to occur is again at most $2/C$. Thus, the probability of having at least d failures is at most $(2/C)^d$ for any integer $d \geq 1$. Let us choose $d = \lfloor c \rfloor - 1$ then $(2/C)^d < 2^{-c^2/2+3c}$.

Now consider the quantity $v_i = C^{j_i+2}\alpha_i$. If the choice of x_{i+1} was a failure then $v_{i+1} \leq v_i$ because $\alpha_{i+1} \leq \alpha_i$. If however the choice of x_{i+1} was not a failure then $j_{i+1} = j_i + 1$ and thus $v_{i+1} \leq v_i^2$. Thus, $v_i \leq v_0^{2^{j_i}}$ for each $i = 1, \ldots, s$. Notice that by assumption we have $\alpha_0 \leq 2^{-c}$ and therefore $v_0 \leq 1/4$ so $v_i \leq 2^{-2^{j_i+1}}$. So except for the $(2/C)^d$ error probability mentioned in the preceding paragraph we have $\alpha_s < v_s \leq 2^{-2^{s-d+1}}$. As $s - c = \log t$ we have $s - d \geq \log t + 1$ and thus $2^{s-d+1} \geq 4t > m$. Thus, in this case $\alpha_s < 1/2^m$ and therefore $A_s = U$ so $T(A) = V$ as required. □

Corollary 3.5 *For a large enough d and any power n of 2 and for every subset S of any $\mathbf{Z}_2$ vector space of size $|S| = n \log n$ if we take a uniform random linear transformation to hash S to $\mathbf{Z}_2^{\log n}$ we have*

$$\text{Prob}[\textit{maximum bucket size} \geq 2^d \log n \log\log n] < 2^{-d^2/3}.$$

Proof: We go back to the scenario of choosing the random linear transformation h from the domain D to the buckets B through an intermediate domain A. We set the dimension of A to be $\ell = \lfloor d - \log d - \log c_{1/2} \rfloor$. So $h = h_1 \circ h_2$ and $h_1 : D \to A$ is a uniform random linear map, while $h_2 : A \to B$ is a uniform random onto linear map. We choose the threshold $t = 2^d \log d \log\log d$. Now the events E_1 and E_2 of Propositions 3.2 are defined and the $t > c_{1/2}(2^\ell/n)\log(2^\ell/n)$ condition is satisfied, thus we have

$$\text{Prob}[E_2|E_1] \geq \frac{1}{2}.$$

To bound Prob$[E_2]$ we use Proposition 3.4 instead of Proposition 3.1 as follows. We prove that the probability of E_2 is limited for every choice of h_1. Let us fix a linear map $h_1 : D \to A$ and take $t = \log n$, $m = \ell$ (thus $c = d - \log d - O(1)$) and apply the proposition for the set $A - h_1(S)$ (clearly $|h_1(S)| \leq |S| = n \log n$). The proposition tells us that

$$\Pr[E_2] \leq 2^{-c^2/2+3c}.$$

Combining the last two inequalities for large enough d one gets

$$\text{Prob}[E_1] \leq 2\text{Prob}[E_2] \leq 2^{-c^2/2+3c+1} \leq 2^{-d^2/3}.$$

This proves the corollary. □

Remark: A more careful analysis can limit the error probability in Proposition 3.4 to $2^{-c^2-c\log c}$. This yields an error bound of $2^{-d^2+2d\log d}$ for large enough d in Corollary 3.5.

Proof of Theorem 5: $L^n_{n\log n}$ is the expectation of the distribution of the largest bucket size. Corollary 3.5 limits the probability of the tail of this distribution, thus yielding the desired bound on the expectation. □

In order for the paper to be self-contained we include a proof of the simple statement about random linear transformations used above.

Proposition 3.6 *Let D, A and B be vector spaces over $\mathbf{Z}_2$. Let $h : D \to B$ be an arbitrary linear map, and let $h_2 : A \to B$ be an arbitrary onto linear map. Let α be any point in B and denote $D' \stackrel{\text{def}}{=} h^{-1}(\alpha)$ and $A' \stackrel{\text{def}}{=} h_2^{-1}(\alpha)$. Then, choosing a uniform linear map $h_1 : D \to A$ such that $h = h_1 \circ h_2$ and restricting the domain to D' we get a uniformly chosen linear map from D' to A' if $\alpha = 0$ or uniformly chosen affine map from D' to A' otherwise.*

Proof: Consider $D_0 \stackrel{\text{def}}{=} h^{-1}(0)$ and $A_0 \stackrel{\text{def}}{=} h_2^{-1}(0)$. Let us choose a complement space D_1 to D_0 in D, i.e. $D_0 \cap D_1 = \{0\}$ and $D_0 + D_1 = D$. Let us call x the unique vector in $D' \cap D_1$. We have $D' = D_0 + x$. A linear transformation $h_1 : D \to A$ is determined by its two restrictions $h' : D_0 \to A$ and $h'' : D_1 \to A$. Clearly the uniform random choice of h_1 corresponds to uniform and independent choices for h' and h''. The restriction $h = h_1 \circ h_2$ means that $h'(D_0) \subseteq A_0$ and $h'' \circ h_2$ is the restriction of h to D_1. Thus, after the restriction the random choices of h' and h'' are still independent. Note now that if $\alpha = 0$ then the restriction of h_1 in question is exactly $h' : D' \to A'$. If $\alpha \neq 0$ then use $h_1(u + x) = h'(u) + h''(x)$ for $u \in D_0$ to note that the restriction in question is again h', this time translated by the random value $h''(x) \in A'$. □

4 Remarks and open questions

We have discussed the case of a very small field (size 2) and a very large field (size n). What happens with intermediate sized fields? Some immediate rough generalizations of our bounds are the following: If we hash an adversarily chosen subset of F^m of size $n = |F|^k$ to F^k by a randomly chosen linear map, the expected size of the largest bucket is at most $O((\log n \log\log n)^{\log |F|})$ and at least $\Omega(|F|^{1/3})$. Tighter bounds should be possible.

Another question is which fine-grained property other well known hash families have. Examples of the families we have in mind include: Arithmetic over $\mathbf{Z}_p$ [CW79, FKS84] (with $h_{a,b}(x) = (ax + b \bmod p) \bmod n$), integer multiplication [DHKP93, AHNR95] (with $h_a(x) = (ax \bmod 2^k) \operatorname{div} 2^{k-l}$), Boolean convolution [MNT93] (with $h_a(x) = a \circ x$ projected to some subspace).

An example of a natural non-linear scheme for which the assertion of Theorem 6 fails is the scheme that maps integers between 1 and p, for some large prime p, to integers between 0 and $n-1$ for $n = \lceil p/m \rceil$, by mapping $x \in \mathbf{Z}_p$ to the integer part of the fraction $(ax+b)(\bmod p)/m$, where a, b are two randomly chosen elements of $\mathbf{Z}_p$. For this scheme, there are primes p and choices of an n and a subset S of cardinality $\Omega(n \log n \log\log\log n)$ of $\mathbf{Z}_p$, which is not mapped by the above mapping onto $[0, n-1]$ under any choice of a and b.

To see this, let p be a prime satisfying $p \equiv 3 \pmod 4$ and consider the set

$$S = \{j^2 \bmod p \mid j \in \mathbf{Z}_p \setminus \{0\}\},$$

of all quadratic residues modulo p. Note that for every nonzero element $a \in \mathbf{Z}_p$, the set $aS \pmod p$ is either the set of all quadratic residues or the set of all quadratic non-residues modulo p. The main result of Graham and Ringrose [GR90] asserts that for infinitely many primes p, the smallest quadratic nonresidue modulo p is at least $\Omega(\log p \log\log\log p)$ (this result holds for primes $p \equiv 3 \pmod 4$ as well, as follows from the remark at the end of [GR90]). Since for such primes p, -1 is a quadratic nonresidue, it follows that for the above S and for any choice of $a, b \in \mathbf{Z}_p$, the set $aS + b$ (computed in $\mathbf{Z}_p$) avoids intervals of length at least $\Omega(\log p \log\log\log p)$. Choosing $m = c \log p \log\log\log p$ for an appropriate (small) constant c, and defining $n = \lceil p/m \rceil$, it follows that $|S| = (p-1)/2 = \Omega(n \log n \log\log\log n)$ is not mapped onto $[0, n-1]$ under any choice of a and b.

A final question is whether there exists a class $\mathcal{H}$ of size only $2^{O(\log\log|U| + \log n)}$ and with $L_n^n(\mathcal{H}) = O(\log n / \log\log n)$. Note that linear maps over $\mathbf{Z}_2$, even combined with collapsing the universe, use $O(\log\log|U| + (\log n)^2)$ random bits while the simple scheme using higher degree polynomials uses $O(\log\log|U| + (\log n)^2 / \log\log n)$.

References

[ABI86] N. Alon, L. Babai and A. Itai, A fast and simple randomized parallel algorithm for the maximal independent set problem. *J. Algorithms* **7** (1986) 567–583.

[ABM87] N. Alon, A. Barak and U. Manber, On disseminating information reliably without broadcasting, in: *Proc. 7th International Conference on Distributed Computing Systems (ICDS)*, Berlin, 1987, pp. 74–81.

[AS92] N. Alon and J. H. Spencer, *The Probabilistic Method*, Wiley, New York, 1992.

[AHNR95] A. Andersson, T. Hagerup, S. Nilsson, and R. Raman, Sorting in linear time?, in: *Proc. 27th ACM Symposium on Theory of Computing*, 1995, pp. 427–436.

[CW79] J. L. Carter and M. N. Wegman, Universal classes of hash functions, *J. Comput. Syst. Sci.* **18** (1979) 143–154.

[CLR90] T. H. Cormen, C. E. Leiserson, and R. L. Rivest, Introduction to Algorithms, MIT Press, 1990.

[DHKP93] M. Dietzfelbinger, T. Hagerup, J. Katajainen, and M. Penttonen, A reliable randomized algorithm for the closest-pair problem, Technical Report 513, Fachbereich Informatik, Universität Dortmund, 1993.

[DM90] M. Dietzfelbinger and F. Meyer auf der Heide, Dynamic hashing in real time, in: J. Buchmann, H. Ganzinger, W. J. Paul (Eds.): *Informatik · Festschrift zum 60. Geburtstag von Günter Hotz*, Teubner-Texte zur Informatik, Band 1, B. G. Teubner, 1992, pp. 95–119. (A preliminary version appeared under the title "A New Universal Class of Hash Functions and Dynamic Hashing in Real Time" in *ICALP'90*.)

[DKMHRT94] M. Dietzfelbinger, A. Karlin, K. Mehlhorn, F. Meyer Auf Der Heide, H. Rohnert, R.E. Tarjan, Dynamic perfect hashing: upper and lower bounds, *SIAM J. Comput.* **23** (1994) 738–761.

[DGMP92] M. Dietzfelbinger, J. Gil, Y. Matias, and N. Pippenger, Polynomial hash functions are reliable, ICALP'92, Springer LNCS 623, pp. 235–246.

[FKS84] M. L. Fredman, J. Komlós, and E. Szemerédi, Storing a sparse table with $O(1)$ worst case access time, *J. Ass. Comput. Mach.* **31** (1984) 538–544.

[GBY90] G. Gonnet and R. Baeza-Yates, *Handbook of Algorithms and Data Structures*, Addison-Wesley, 1991.

[GR90] S. W. Graham and C. J. Ringrose, Lower bounds for least quadratic nonresidues, in: *Analytic Number Theory: Proceedings of a Conference in Honor of P.T. Bateman*, B. C. Berndt et al. (Eds.), Birkhäuser, Boston, 1990.

[MCW78] G. Markowsky, J. L. Carter, and M. N. Wegman, Analysis of a universal class of hash functions, in: *Proc. 7th Conference on Math. Found. of Computer Science (MFCS)*, 1978, Springer LNCS 64, pp. 345–354.

[MV84] K. Mehlhorn and U. Vishkin, Randomized and deterministic simulations of PRAMs by parallel machines with restricted granularity of parallel memories., *Acta Informatica* **21** (1984) 339–374.

[MNT93] Y. Mansour, N. Nisan, and P. Tiwari, The computational complexity of universal hashing. *Theoretical Computer Science* **107** (1993) 121–133.

[PA95] J. Pach and P. K. Agarwal, *Combinatorial Geometry*, Wiley 1995.

[S89] A. Siegel, On universal classes of fast high performance hash functions, their time-space tradeoff, and their application, in: *Proc. 30th IEEE Symposium on Foundations of Computer Science*, 1989, pp. 20–25.

[VC71] V. A. Vapnik and A. Y. Chervonenkis, On the uniform convergence of relative frequencies of events to their probabilities, *Theory of Prob. Applications* **16** (1971) 264–280.