Increasing the Output Length of Zero-Error Dispersers

Ariel Gabizon[1] and Ronen Shaltiel[2]

[1] Department of Computer Science, Weizmann institute of science, Rehovot, Israel ⋆
[2] Department of Computer Science, Haifa University, Haifa, Israel ⋆⋆

Abstract. Let $\mathcal{C}$ be a class of probability distributions over a finite set Ω. A function $D : \Omega \mapsto \{0,1\}^m$ is a *disperser* for $\mathcal{C}$ with *entropy threshold* k and *error* ϵ if for any distribution X in $\mathcal{C}$ such that X gives positive probability to at least 2^k elements we have that the distribution $D(X)$ gives positive probability to at least $(1-\epsilon)2^m$ elements. A long line of research is devoted to giving explicit (that is polynomial time computable) dispersers (and related objects called "extractors") for various classes of distributions while trying to maximize m as a function of k.

In this paper we are interested in explicitly constructing *zero-error dispersers* (that is dispersers with error $\epsilon = 0$). For several interesting classes of distributions there are explicit constructions in the literature of zero-error dispersers with "small" output length m and we give improved constructions that achieve "large" output length, namely $m = \Omega(k)$.

We achieve this by developing a general technique to improve the output length of zero-error dispersers (namely, to transform a disperser with short output length into one with large output length). This strategy works for several classes of sources and is inspired by a transformation that improves the output length of extractors (which was given in [29] building on earlier work by [15]). Nevertheless, we stress that our techniques are different than those of [29] and in particular give non-trivial results in the errorless case.

Using our approach we construct improved zero-error dispersers for the class of 2-*sources*. More precisely, we show that for any constant $\delta > 0$ there is a constant $\eta > 0$ such that for sufficiently large n there is a poly-time computable function $D : \{0,1\}^n \times \{0,1\}^n \mapsto \{0,1\}^{\eta n}$ such that for any two independent distributions X_1, X_2 over $\{0,1\}^n$ such that both of them support at least $2^{\delta n}$ elements we get that the output distribution $D(X_1, X_2)$ has full support. This improves the output length of previous constructions by [2] and has applications in Ramsey Theory and in constructing certain data structures [13].

We also use our techniques to give explicit constructions of zero-error dispersers for bit-fixing sources and affine sources over polynomially large fields. These constructions improve the best known explicit constructions due to [26, 14] and achieve $m = \Omega(k)$ for bit-fixing sources and $m = k - o(k)$ for affine sources.

⋆ Research supported by Binational Science Foundation (BSF) grant and by Minerva Foundation grant.

⋆⋆ Research supported by Binational US-Israel Science Foundation (BSF) grant 2004329 and Israel Science Foundation (ISF) grant 686/07.

1 Introduction

1.1 Background

Randomness extractors and dispersers are functions that refine the randomness in "weak sources of randomness" that "contain sufficient entropy". Various variants of extractors and dispersers are closely related to expander graphs, error correcting codes and objects from Ramsey theory. A long line of research is concerned with explicit constructions of these objects and these constructions have many applications in many areas of computer science and mathematics (e.g. network design, cryptography, pseudorandomness, coding theory, hardness of approximation, algorithm design and Ramsey theory).

Randomness extractors and dispersers We start with formal definitions of extractors and dispersers. (We remark that in this paper we consider the "seedless version" of extractors and dispersers).

Definition 1 (min-entropy and statistical distance) *Let Ω be a finite set. The* min-entropy *of a distribution X on Ω is defined by $H_\infty(X) = min_{x \in \Omega} \log_2 \frac{1}{\Pr[X=x]}$. For a class $\mathcal{C}$ of distributions on Ω we use $\mathcal{C}_k$ to denote the class of all distributions $X \in \mathcal{C}$ such that $H_\infty(X) \geq k$. We say that two distributions X, Y on Ω are* ϵ-close *if $\frac{1}{2}\sum_{w \in \Omega} |\Pr[X = w] - \Pr[Y = w]| \leq \epsilon$.*

When given a class $\mathcal{C}$ of distributions (which we call "sources") the goal is to design one function that refines the randomness of any distribution X in $\mathcal{C}$. An *extractor* produces a distribution that is (close to) uniform whereas a *disperser* produces a distribution with (almost) full support. A precise definition follows:

Definition 2 (Extractors and Dispersers) *Let $\mathcal{C}$ be a class of distributions on a finite set Ω.*

- *A function $E : \Omega \mapsto \{0,1\}^m$ is an* extractor *for $\mathcal{C}$ with* entropy threshold k *and* error $\epsilon > 0$ *if for every $X \in \mathcal{C}_k$, $E(X)$ is ϵ-close to the uniform distribution on $\{0,1\}^m$.*
- *A function $D : \Omega \mapsto \{0,1\}^m$ is a* disperser *for $\mathcal{C}$ with* entropy threshold k *and* error $\epsilon > 0$ *if for every $X \in \mathcal{C}_k$, $|\mathrm{Supp}(D(X))| \geq (1-\epsilon)2^m$ (where $\mathrm{Supp}(Z)$ denotes the support of the random variable Z).*

We remark that every extractor is in particular a disperser and that the notion of dispersers only depends on the support of the distributions in $\mathcal{C}$. A long line of research is concerned with designing extractors and dispersers for various classes of sources. For a given class $\mathcal{C}$ we are interested in designing extractors and dispersers with as small as possible entropy threshold k, as large as possible output length m and as small as possible error ϵ. (We remark that it easily follows that $m \leq k$ whenever $\epsilon < 1/2$).

It is often the case that the probabilistic method gives that a randomly chosen function E is an excellent extractor. (This is in particular true whenever the class $\mathcal{C}$ contains "not too many" sources). However, most applications of extractors and dispersers require *explicit constructions*, namely functions that can be computed in time polynomial in their input length. Much of the work done in this area can be described as an attempt of matching the parameters obtained by existential results using explicit constructions.

Some related work

Classes of sources. Various classes $\mathcal{C}$ of distributions were studied in the literature: The first construction of deterministic extractors can be traced back to von Neumann [33] who showed how to use many independent tosses of a biassed coin (with unknown bias) to obtain an unbiased coin. Blum [5] considered sources that are generated by a finite Markov-chain. Santha and Vazirani [28], Vazirani [28, 32], Chor and Goldreich [8], Dodis et al. [11], Barak, Impagliazzo and Wigderson [1], Barak et al. [2], Raz [27], Rao [25], Bourgain [6], Barak et al. [3], and Shaltiel [29] studied sources that are composed of several independent samples from "high entropy" distributions. Chor et al. [9], Ben-Or and Linial [4], Cohen and Wigderson [10], Mossel and Umans [22], Kamp and Zuckerman [20], Gabizon, Raz and Shaltiel [15], and Rao [26] studied bit-fixing sources which are sources in which a subset of the bits are uniformly distributed. Trevisan and Vadhan [31] and Kamp et al. [19] studied sources which are "samplable" by "efficient" procedures. Barak et al. [2], Bourgain [7], Gabizon and Raz [14], and Rao [26] studied sources which are uniform over an affine subspace. Dvir, Gabizon and Wigderson [12] studied a generalization of affine sources to sources which are sampled by low degree multivariate polynomials.

Seeded extractors and dispersers. A different variant of extractors and dispersers are *seeded* extractors and dispersers (defined by Nisan and Zuckerman [23]). Here the class $\mathcal{C}$ is the class of all distributions on $\Omega = \{0,1\}^n$. It is easy to verify that there do not exist extractors or dispersers for $\mathcal{C}$ (even when $k = n-1$, $m = 1$ and $\epsilon < 1/2$). However, if one allows the extractor (or disperser) to receive an additional independent uniformly distributed input (which is called "a seed") then extraction is possible as long as the seed is of length $\Theta(\log(n/\epsilon))$. More precisely, a seeded extractor (or disperser) with entropy threshold k and error ϵ is a function $F : \{0,1\}^n \times \{0,1\}^t \mapsto \{0,1\}^m$ such that for any distribution X on $\{0,1\}^n$ with $\mathrm{H}_\infty(X) \geq k$ the distribution $F(X,Y)$ (where Y is an independent uniformly distributed variable) satisfies the guarantees of Definition 2. A long line of research is concerned with explicit constructions of seeded extractors and dispersers (the reader is referred to [30] for a survey article and to [21, 18] for the current milestones in explicit constructions of extractors).

Zero-error dispersers In this paper we are interested in *zero-error dispersers.* These are dispersers where the output distribution has full support. That is for every source X in the class $\mathcal{C}$:

$$\{D(x) : x \in \mathrm{Supp}(X)\} = \{0,1\}^m$$

We also consider a stronger variant which we call *strongly-hitting disperser* in which every output element $z \in \{0,1\}^m$ is obtained with "not too small" probability. A precise definition follows:

Definition 3 (Zero-error dispersers and strongly hitting dispersers) *Let $\mathcal{C}$ be a class of distributions on a finite set Ω.*

- *A function D is a* zero-error disperser *for $\mathcal{C}$ with* entropy threshold k *if it is a disperser for $\mathcal{C}$ with entropy threshold k and error $\epsilon = 0$.*
- *A function $D : \Omega \mapsto \{0,1\}^m$ is a μ-*strongly hitting disperser *for $\mathcal{C}$ with* entropy threshold k *if for every $X \in \mathcal{C}_k$ and for every $z \in \{0,1\}^m$, $\Pr[D(X) = z] \geq \mu$.*

Note that a μ-strongly hitting disperser with $\mu > 0$ is in particular a zero-error disperser and that any μ-strongly hitting disperser has $\mu \leq 2^{-m}$. The following facts immediately follow:

Fact 1 *Let $f : \Omega \mapsto \{0,1\}^m$ be a function and let $\epsilon \leq 2^{-(m+1)}$.*

- *If f is a disperser with error ϵ then f is a zero-error disperser (for the same class $\mathcal{C}$ and entropy threshold k).*
- *If f is an extractor with error ϵ then f is a $2^{-(m+1)}$-strongly hitting disperser (for the same class $\mathcal{C}$ and entropy threshold k).*

It follows that extractors and dispersers with small ϵ immediately translate into zero-error dispersers (as one can truncate the output length to $m' = \log(1/\epsilon) - 1$ bits and such a truncation preserves the output guarantees of extractors and dispersers).

1.2 Increasing the output length of zero-error dispersers

For several interesting classes of sources there are explicit constructions of dispersers with "large" error (which by Fact 1 give zero-error dispersers with "short" output length). In this paper we develop techniques to construct zero-error dispersers with large output length.

The composition approach The following methodology for increasing the output length of extractors was suggested in [15, 29]: When given an extractor E' with "small" output length t (for some class $\mathcal{C}$) consider the function $E(x) = F(x, E'(x))$ where F is a seeded extractor. Shaltiel [29] (building on earlier work by Gabizon et al. [15]) shows that if E' and F fulfill certain requirements then this construction yields an extractor for $\mathcal{C}$ with large output length. The high level idea is that if certain conditions are fulfilled then the distribution $F(X, E(X))$ (in which the two inputs of F are *dependent*) is close to the distribution $F(X, Y)$ (where Y is an independent uniformly distributed variable) and note that the latter distribution is close to uniform by the definition of seeded extractors. This technique proved useful for several interesting classes of sources.

We would like to apply an analogous idea to obtain zero-error dispersers. However, by the lower bounds of [23, 24] if F is a seeded extractor (or seeded disperser) then its seed length is at least $\log(1/\epsilon)$. This means that if we want $F(X,Y)$ to output m bits with error $\epsilon < 1/2^m$ we need seed length larger than m. This in turn means that we want E' to have output length $t > m$ which makes the transformation useless.

There are also additional problems. The argument in [29] requires the "original function" E' to be an extractor (and it does not go through if E' is a disperser) and furthermore the error of the "target function" E is at least as large as that of the "original function" E' (and once again we don't gain when shooting for zero-error dispersers).

Summing up we note that if we want to improve the output length of a zero-error disperser D' by a composition of the form $D(x) = F(x, D'(x))$ we need to use a function F with different properties (a seeded extractor or disperser will not do) and we need to use a different kind of analysis.

Composing zero-error dispersers In this paper we imitate the method of [29] and give a general method to increase the output length of zero-error dispersers. That is when given:

- A zero-error disperser $D' : \Omega \mapsto \{0,1\}^t$ for a class $\mathcal{C}$ and "small" output length t.
- A function $F : \Omega \times \{0,1\}^t \mapsto \{0,1\}^m$ for "large" output length m.

We identify properties of F that are sufficient so that the construction

$$D(x) = F(x, D'(x))$$

gives a zero-error disperser. (The argument is more general and transforms $2^{-(t+O(1))}$-strongly hitting dispersers into $2^{-(m+O(1))}$-strongly hitting dispersers). We then use this technique to give new constructions of zero-error dispersers and strongly-hitting dispersers.

Subsource hitters As explained earlier we cannot choose F to be a seeded extractor. Instead, we introduce a new object which we call a *subsource hitter*. The definition of subsource hitters is somewhat technical and is tailored so that the construction $D(x) = F(x, D'(x))$ indeed produces a disperser.

Definition 4 (subsource hitter) *A distribution X' on Ω is a* subsource *of a distribution X on Ω if there exist $\alpha > 0$ and a distribution X'' on Ω such that X can be expressed as a convex combination $X = \alpha X' + (1-\alpha)X''$.*

Let $\mathcal{C}$ be a class of distributions on Ω. A function $F : \Omega \times \{0,1\}^t \mapsto \{0,1\}^m$ is a subsource-hitter *for $\mathcal{C}$ with* entropy threshold k *and* subsource entropy $k-v$ *if for any $X \in \mathcal{C}_k$ and $z \in \{0,1\}^m$ there exists a $y \in \{0,1\}^t$ and a distribution $X' \in \mathcal{C}_{k-v}$ that is a subsource of X such that for every $x \in \mathrm{Supp}(X')$ we have that $F(x,y) = z$.*

A subsource hitter has the property that for any $z \in \{0,1\}^m$ there exist $y \in \{0,1\}^t$ and $x \in \mathrm{Supp}(X)$ such that $F(x,y) = z$ and in particular

$$\{F(x,y) : x \in \mathrm{Supp}(X), y \in \{0,1\}^t\} = \{0,1\}^m$$

In addition a subsource hitter has the stronger property that there exists a subsource X' of X (which is itself a source in $\mathcal{C}$) such that for any $z \in \{0,1\}^m$ there exists $y \in \{0,1\}^t$ such that for *any* $x \in \mathrm{Supp}(X') \subseteq \mathrm{Supp}(X)$, $F(x,y) = z$.

This property allows us to show that $D(x) = F(x, D'(x))$ is a zero-error disperser with entropy threshold k whenever D' is a zero-error disperser with entropy threshold $k - v$. This is because when given a source $X \in \mathcal{C}_k$ and $z \in \{0,1\}^m$ we can consider the seed $y \in \{0,1\}^t$ and subsource X' guaranteed in the definition. We have that D' is a zero-error disperser and that X' meets the entropy threshold of D'. It follows that there exist $x \in \mathrm{Supp}(X') \subseteq \mathrm{Supp}(X)$ such that $D'(x) = y$. It follows that:

$$D(x) = F(x, D'(x)) = F(x,y) = z$$

and this means that D indeed outputs z. (We remark that a more complicated version of this argument shows that the composition applies to strongly-hitting dispersers). The exact details are given in the full version. It is interesting to note that this argument is significantly simpler than that of [29]. Indeed, the definition of subsource hitters is specifically tailored to make the composition argument go through and the more complicated task is to design subsource hitters. This is in contrast to [29] in which the function F is in most cases an “off the shelf” seeded extractor and the difficulty is to show that the composition succeeds.

1.3 Applications

We use the new composition technique to construct zero-error dispersers with large output length for various classes of sources. We discuss these constructions and some applications below.

Zero-error 2-source dispersers The class of 2-*sources* is the class of distributions $X = (X_1, X_2)$ on $\Omega = \{0,1\}^n \times \{0,1\}^n$ such that X_1, X_2 are independent. It is common to consider the case where each of the two distributions X_1, X_2 has min-entropy at least some threshold k.

Definition 5 (2-source extractors and dispersers) *A function $f : \{0,1\}^n \times \{0,1\}^n \mapsto \{0,1\}^m$ is a 2-source extractor (resp. disperser) with entropy threshold $2 \cdot k$ and error $\epsilon \geq 0$ if for every two independent distributions X_1, X_2 on $\{0,1\}^n$ both having min-entropy at least k, $f(X_1, X_2)$ is ϵ-close to the uniform distribution on $\{0,1\}^m$ (resp. $|\mathrm{Supp}(f(X_1,X_2))| \geq (1-\epsilon)2^m$). We say that f is a zero-error disperser if it is a disperser with error $\epsilon = 0$. We say that f is a μ-strongly hitting disperser if for every X_1, X_2 as above and every $z \in \{0,1\}^m$, $\Pr[f(X_1,X_2) = z] \geq \mu$.*

Background. The probabilistic method gives 2-source extractors with $m = 2 \cdot k - O(\log(1/\epsilon))$ for any $k \geq \Omega(\log n)$. However, until 2005 the best explicit constructions [8, 32] only achieved $k > n/2$. The current best extractor construction [6] achieves entropy threshold $k = (1/2 - \alpha)n$ for some constant $\alpha > 0$. Improved constructions of dispersers for entropy threshold $k = \delta n$ (for an arbitrary constant $\delta > 0$) were given in [2]. These dispersers can output any constant number of bits with zero-error (and are μ-strongly hitting for some constant $\mu > 0$).[3] Subsequent work by [3] achieved entropy threshold to $k = n^{o(1)}$ and gives zero-error dispersers that output one bit.

Our results. We use our composition techniques to improve the output length in the construction of [2]. We show that:

Theorem 2 (2-source zero-error disperser) *For every $\delta > 0$ there exists a $\nu > 0$ and $\eta > 0$ such that for sufficiently large n there is a poly(n)-time computable $(\nu 2^{-m})$-strongly hitting 2-source disperser $D : (\{0,1\}^n)^2 \mapsto \{0,1\}^m$ with entropy threshold $2 \cdot \delta n$ and $m = \eta n$.*

Note that our construction achieves an output length that is optimal up to constant factors for this entropy threshold. For lower entropy threshold our techniques gives that any explicit construction of a zero-error 2-source disperser D' with entropy threshold k and output length $t = \Omega(\log n)$ can be transformed into an explicit construction of a zero-error 2-source disperser D with entropy threshold $2 \cdot k$ and output length $m = \Omega(k)$. (See the full version for a precise formulation that also considers strongly hitting dispersers). This cannot be applied on the construction of [3] that achieves entropy threshold $k = n^{o(1)}$ as this construction only outputs one bit. Nevertheless, this means that it suffices to extend the construction of [3] so that it outputs $\Theta(\log n)$ bits in order to obtain an output length of $m = \Omega(k)$ for low entropy threshold k.

We prove Theorem 2 by designing a subsource hitter for 2-sources and using our composition technique. The details are given in the full version and a high level outline appears next.

Outline of the argument. We want to design a function $F : \{0,1\}^n \times \{0,1\}^n \times \{0,1\}^t \mapsto \{0,1\}^m$ such that for any 2-source $X = (X_1, X_2)$ with sufficient min-entropy and for any $z \in \{0,1\}^m$ there exists a "seed" $y \in \{0,1\}^t$ and a subsource X' of X such that $X' = (X_1', X_2')$ is a 2-source with roughly the same min-entropy as X and $\Pr[F(X_1', X_2', y) = z] = 1$. We will be shooting for $m = \Omega(n)$ and $t = O(\log n)$.

We construct the seed obtainer F using ideas from [2, 3]. Let E be a seeded extractor with seed length $t = O(\log n)$, output length $v = \Omega(k)$ and error

[3] In [25] it is pointed out that by enhancing the technique of [2] using ideas from [3] and replacing some of the components used in the construction with improved components that are constructed in [25] it is possible to increase the output length and achieve a zero-error disperser with output length $m = k^{\Omega(1)}$ for the same entropy threshold k.

$\epsilon_E = 1/100$ (such extractors were constructed in [21, 18]). When given inputs x_1, x_2, y we consider $r_1 = E(x_1, y)$ and $r_2 = E(x_2, y)$. By using a stronger variant of seeded extractors called "strong extractors" it follows that there exists a "good seed" $y \in \{0,1\}^t$ such that $R_1 = E(X_1, y)$ and $R_2 = E(X_2, y)$ are ϵ_E-close to uniform. We then use a 2-source extractor $H : \{0,1\}^v \times \{0,1\}^v \mapsto \{0,1\}^m$ for *very high* entropy threshold (say entropy threshold $2 \cdot 0.9v$) and very low error (say error $2^{-(m+1)}$ for output length $m = \Omega(v) = \Omega(k)$). Such extractors were constructed in [32]. Our final output is given by:

$$F(x_1, x_2, y) = H(E(x_1, y), E(x_2, y))$$

This seems strange at first sight as it is not clear why running H on inputs R_1, R_2 that are already close to uniform helps. Furthermore, the straightforward analysis only gives that $H(R_1, R_2)$ is ϵ-close to uniform for *large* error $\epsilon \geq \epsilon_E = 1/100$ and this means that the output of F may miss a large fraction of strings in $\{0,1\}^m$.

The point to notice is that both R_1, R_2 are close to uniform and therefore have large support $(1-\epsilon_E)2^v \geq 2^{0.9v}$. Using Fact 1 we can think of H as a zero-error disperser. Recall that for dispersers are oblivious to the precise probability distribution of R_1, R_2 and it is sufficient that R_1, R_2 have large support. It follows that indeed every string $z \in \{0,1\}^m$ is hit by $H(R_1, R_2)$.

This does not suffice for our purposes as we need that any string z is hit with probability one on a subsource $X' = (X'_1, X'_2)$ of X in which the two distributions X'_1 and X'_2 are independent. For any output string $z \in \{0,1\}^m$ we consider a pair of values (r_1, r_2) for R_1, R_2 on which $H(r_1, r_2) = z$ (we have just seen that such a pair exists) and set $X'_1 = (X_1 | E(X_1, y) = r_1)$ and $X'_2 = (X_2 | E(X_2, y) = r_2)$. Note that these two distributions are indeed independent (as each depends only on one of the original distributions X_1, X_2) and that on every $x'_1 \in \mathrm{Supp}(X'_1)$ and $x'_2 \in \mathrm{Supp}(X'_2)$ we have that:

$$F(x'_1, x'_2, y) = H(E(x'_1, y), E(x'_2, y)) = H(r_1, r_2) = z$$

Furthermore, for a typical choice of (r_1, r_2) we can show that both X'_1, X'_2 have min-entropy roughly $k - v$. Thus, setting v appropriately, X' is a subsource of X with the required properties. (A more careful version of this argument can be used to preserve the "strongly hitting" property).

Interpretation in Ramsey Theory A famous theorem in Ramsey Theory (see [17]) states that for sufficiently large N and any 2-coloring of the edges of the complete graph on N vertices there is an induced subgraph on $K = \Theta(\log N)$ vertices which is "monochromatic" (that is all edges are of the same color).

Zero-error 2-source dispersers (with output length $m = 1$) can be seen as providing counterexamples to this statement for larger values of K in the following way: When given a zero-error 2-source disperser $D : \{0,1\}^n \times \{0,1\}^n \mapsto \{0,1\}^m$ with entropy threshold $2 \cdot k$ we can consider coloring the edges of the full graph on $N = 2^n$ vertices with 2^m colors by coloring an edge (v_1, v_2) by $D(v_1, v_2)$. (A technicality is that $D(v_1, v_2)$ may be different than $D(v_2, v_1)$ and to avoid this problem the coloring is defined by ordering the vertices according to some order and coloring the edge (v_1, v_2) where $v_1 \leq v_2$ by $D(v_1, v_2)$). The disperser guarantee can be used to show that any induced subgraph with $K = 2^{k+1}$ vertices contains edges of *all* 2^m colors.[4]

Note that dispersers with $m > 1$ translate into colorings with more colors and that in this context of Ramsey Theory the notion of a zero-error disperser seems more natural than one that allows error. Our constructions achieve $m = \Omega(k)$ and thus the number of colors in the coloring approaches the size of the induced subgraph.

Generalizing this relation between dispersers and Ramsey theory we can view any zero-error disperser for a class $\mathcal{C}$ as a coloring of all $x \in \Omega$ such that any set S that is obtained as the support of a distribution in $\mathcal{C}$ is colored by all possible 2^m colors.

Rainbows and implicit $O(1)$-probe search As we now explain, explicit constructions of zero-error 2-source dispersers can be used to construct certain data structures (this connection is due to [13]).

Consider the following problem: We are given a set $S \subseteq \{0,1\}^n$ of size 2^k. We want to store the elements of S in a table T of the same size where every entry in the table contains a single element of S (and so the only freedom is in ordering the elements of S in the table T). We say that T supports q-queries if given $x \in \{0,1\}^n$ we can determine whether $x \in S$ using q queries to T (note for example that ordered tables and binary search support $q = k$ queries). Yao [34] and Fiat and Naor [13] showed that it is impossible to achieve $q = O(1)$ when n is large enough relative to k. (This result can be seen as a kind of Ramsey Theorem).

Fiat and Naor [13] gave explicit constructions of tables that support $q = O(1)$ queries when $k = \delta \cdot n$ for any constant $\delta > 0$. This was achieved by reducing the implicit probe search problem to the task of explicitly constructing a certain combinatorial object that they call a “rainbow”.

Loosely speaking a rainbow is a zero-error disperser for the class of distributions X that are composed of q independent copies of a high min-entropy distribution. We stress that for this application one needs (strongly-hitting) dispersers with large output length. More precisely, in order to support $q = O(1)$ queries one requires such dispersers that have output length m that is a *constant fraction* of the entropy threshold.

Our techniques can be used to explicitly construct rainbows which in turn allow implicit probe schemes that support $q = O(1)$ queries for smaller values of k than previously known. More precisely for any constant $\delta > 0$ and $k = n^{\delta}$ there is a constant q and a scheme that supports q queries. The precise details are given in the full version. (We remark that one can also achieve the same

[4] In fact, Dispersers translate into a significantly stronger guarantee that discusses colorings of the edges of the complete N by N bipartite graph such that any induced K by K subgraph has all colors.

results by using the technique of [13] and plugging in recent constructions of seeded dispersers).

Zero-error dispersers for bit-fixing sources The class of *bit-fixing sources* is the class of distributions X on $\Omega = \{0,1\}^n$ such that there exists a set $S \subseteq [n]$ such that X_S (that is X restricted to the indices in S) is uniformly distributed and $X_{[n] \setminus S}$ is constant. Note that for such a source X, $\mathrm{H}_\infty(X) = |S|$. (We remark that these sources are sometimes called "oblivious bit-fixing sources" to differentiate them from "non-oblivious bit-fixing sources" in which $X_{[n] \setminus S}$ is allowed to be a function of X_S).

Background. The function $Parity(x)$ (that is the exclusive-or of the bits of x) is obviously an extractor for bit-fixing sources with entropy threshold $k = 1$, error $\epsilon = 0$ and output length $m = 1$. It turns out that there are no errorless extractors for $m = 2$. More precisely, [9] showed that for $k < n/3$ there are no extractors for bit-fixing sources with $\epsilon = 0$ and $m = 2$. For larger values of k, [9] give constructions with $m > 1$ and $\epsilon = 0$. For general entropy threshold k the current best explicit construction of extractors for bit-fixing sources is due to [26] (in fact, this extractor works for a more general class of "low weight affine sources"). These extractors work for any entropy threshold $k \geq (\log n)^c$ for some constant c, and achieve output length $m = (1 - o(1))k$ for error $\epsilon = 2^{-k^{\Omega(1)}}$. Using Fact 1 this gives a zero-error disperser with output length $m = k^{\Omega(1)}$.

Our results. We use our composition techniques to construct zero-error dispersers for bit-fixing sources with output length $m = \Omega(k)$. We show that:

Theorem 3 (Zero-error disperser for bit-fixing sources) *There exist $c > 1$ and $\eta > 0$ such that for sufficiently large n and $k \geq (\log n)^c$ there is a poly(n)-time computable zero-error disperser $D : \{0,1\}^n \mapsto \{0,1\}^m$ for bit-fixing sources with entropy threshold k and output length $m = \eta k$.*

Note that our construction achieves an output length that is optimal up to constant factors. We prove Theorem 3 by designing a subsource hitter for bit-fixing sources and using our composition technique. The details are given in the full version and a high level outline appears next.

Outline of the argument. Our goal is to design a subsource hitter $G : \{0,1\}^n \times \{0,1\}^t \mapsto \{0,1\}^m$ for bit-fixing sources with entropy threshold k, output length $m = \Omega(k)$ and "seed length" $t = O(\log n)$. We make use of the subsource hitter for 2-sources $F : \{0,1\}^n \times \{0,1\}^n \times \{0,1\}^{O(\log n)} \mapsto \{0,1\}^m$ that we designed earlier. We apply it for entropy threshold $k' = k/8$ and recall that it has output length $m = \Omega(k') = \Omega(k)$.

When given a seed $y \in \{0,1\}^t$ for G we think about it as a pair of strings (y', y'') where y' is a seed for F and y'' is a seed for an explicit construction of pairwise independent variables $Z_1, \ldots, Z_n$ where for each i, Z_i takes values in $\{1,2,3\}$ (indeed there are such constructions with seed length $O(\log n)$). When given such a seed y'' we can use the values $Z_1, \ldots, Z_n$ to partition the set $[n]$ into three disjoint sets T_1, T_2, T_3 by having each index $i \in [n]$ belong to T_{Z_i}. We construct G as follows:

$$G(x, (y', y'')) = F(x_{T_1}, x_{T_2}, y')$$

In words, we use y'' to partition the given n bit string into three strings and we run F on the first two strings (padding each of them to length n) using the seed y'.

We need to show that for any bit-fixing source X of min-entropy k and for any $z \in \{0,1\}^m$ there exist a seed $y = (y', y'')$ and a subsource X' of X such that X' is a bit-fixing source with roughly the same min-entropy as X and $\Pr[G(X', (y', y'')) = z] = 1$.

We have that X is a bit-fixing source and let $S \subseteq [n]$ be the set of its "good indices". Note that $|S| \geq k$. By the "sampling properties" of pairwise independent distributions (see e.g. [16] for a survey on "averaging samplers") it follows that there exists a y'' such that for every $i \in [3]$, $|S \cap T_i| \geq k/8$. It follows that $X_{T_1}, X_{T_2}, X_{T_3}$ are bit-fixing sources with min-entropy at least $k/8$ (and note that these three distributions are independent). Thus, by the properties of the subsource hitter F there exist x_1, x_2, y' such that $F(x_1, x_2, y') = z$ (note that here we're only using the property that F "hits z" and do not use the stronger property that F "hits z on a subsource"). Consider the distribution

$$X' = (X | X_{T_1} = x_1 \wedge X_{T_2} = x_2)$$

This is a subsource of X which is a bit-fixing source with min-entropy at least $k/8$ (as we have not fixed the $k/8$ good bits in T_3). It follows that for every $x \in \mathrm{Supp}(X')$

$$G(x, (y', y'')) = F(x_1, x_2, y') = z$$

and G is indeed a subsource hitter for bit-fixing sources.

Affine sources The class of *affine sources* is the class of distributions X on $\Omega = \mathbb{F}_q^n$ (where $\mathbb{F}_q$ is the finite field of q elements) such that X is uniformly distributed over an affine subspace V in $\mathbb{F}_q^n$. Note that such a source X has min-entropy $\log q \cdot dim(V)$. Furthermore, any bit-fixing source is an affine source over $\mathbb{F}_2$.

Background. For $\mathbb{F}_2$ the best explicit construction of extractors for affine sources was given in [7]. This construction works for entropy threshold $k = \delta n$ (for any fixed $\delta > 0$) and achieves output length $m = \Omega(k)$ with error $\epsilon < 2^{-m}$.

Extractors for lower entropy thresholds were given by [14] in the case that $q = n^{\Theta(1)}$. For any entropy threshold $k > \log q$ these extractors can output $m = (1 - o(1))k$ bits with error $\epsilon = n^{-\Theta(1)}$. Using Fact 1 this gives zero-error dispersers with output length $m = \Theta(\log n)$.

Our results. Our composition techniques can be applied on affine sources. We focus on the case of large fields (as in that case we can improve the results of [14]). We remark that our techniques also apply when q is small (however, at the moment we do not gain by applying them on the existing explicit constructions). We prove the following theorem:

Theorem 4 *Fix any prime power q and integers n, k such that $q \geq n^{18}$ and $2 \leq k < n$. There is a poly$(n, \log q)$-time computable zero-error disperser $D : \mathbb{F}_q^n \mapsto \{0,1\}^m$ for affine sources with entropy threshold $k \cdot \log q$ and $m = (k-1) \cdot \log q$.*

Outline of the argument. We use our composition techniques to give a different analysis of the construction of [14] which shows that this construction also gives a zero-error disperser. The construction of [14] works by first constructing an affine source extractor D' with small output length $m = \Theta(\log n)$ and then composing it with some function F to obtain an extractor $D(x) = F(x, D'(x))$ that extracts many bits (with rather large error). We observe that the function F designed in [14] is in fact a subsource hitter for affine sources and therefore our composition technique gives that the final construction is a zero-error disperser.

2 Open problems

2-sources. One of the most important open problems in this area is to give constructions of extractors for entropy threshold $k = o(n)$. Such constructions are not known even for $m = 1$ and large error ϵ.

There are explicit constructions of zero-error dispersers with $k = n^{o(1)}$ [3]. However, these dispersers only output one bit. Improving the output length in these constructions to $\Theta(\log n)$ bits will allow our composition techniques to achieve output length $m = \Omega(k)$.

Another intriguing problem is that for the case of zero-error (or strongly hitting) dispersers we do not know whether the existential results proven via the probabilistic method achieve the best possible parameters. More precisely, a straightforward application of the probabilistic method gives zero-error 2-source dispersers which on entropy threshold $2 \cdot k$ output $m = k - \log(n-k) - O(1)$ bits. On the other hand the lower bounds of [23, 24] can be used to show that any zero-error 2-source disperser with entropy threshold $2 \cdot k$ has $m \leq k + O(1)$.[5]

O(1)-sources, rainbows and implicit probe search. When allowing ℓ-sources for $\ell = O(1)$ we give constructions of zero-error dispersers which on entropy threshold $k = n^{\Omega(1)}$ achieve output length $m = \Omega(k)$. An interesting open problem is to try and improve the entropy threshold. As explained in the full version this immediately implies improved implicit probe search schemes.

[5] Radhakrishnan and Ta-Shma [24] show that any seeded disperser $D : \{0,1\}^n \times \{0,1\}^t \to \{0,1\}^m$ that is nontrivial in the sense that $m \geq t+1$ has $t \geq \log(1/\epsilon) - O(1)$. A zero-error 2-source disperser D' with entropy threshold k can be easily transformed into a seeded disperser with seed length $t = k$ by setting $D(x, y) = D'(x, y')$ where y' is obtained by padding the k bit long "seed" y with $n - k$ zeroes. The bound follows as D' has error smaller than 2^{-m}.

Bit-fixing sources. We give constructions of zero-error dispersers which on entropy threshold k achieve $m = \Omega(k)$. A straightforward application of the probabilistic method gives zero-error dispersers with $m = k - \log n - o(\log n)$. We do not know how to match these parameters with explicit constructions.

Affine sources. We constructed a subsource hitter for affine sources over relatively large fields (that is $q = n^{\Theta(1)}$). It is interesting to try and construct subsource hitters for smaller fields.

Finally, it is also natural to ask whether our composition approach applies to other classes of sources.

3 Acknowledgements

We are grateful to Ran Raz for his support.

References

1. B. Barak, R. Impagliazzo, and A. Wigderson. Extracting randomness using few independent sources. *SIAM J. Comput*, 36(4):1095–1118, 2006.
2. B. Barak, G. Kindler, R. Shaltiel, B. Sudakov, and A. Wigderson. Simulating independence: New consturctions of condenesers, ramsey graphs, dispersers, and extractors. In *STOC 2005*, pages 1–10.
3. B. Barak, A. Rao, R. Shaltiel, and A. Wigderson. 2-source dispersers for sub-polynomial entropy and Ramsey graphs beating the Frankl–Wilson construction. In *STOC 2006*, pages 671–680.
4. M. Ben-Or and N. Linial. Collective coin flipping. *ADVCR: Advances in Computing Research*, 5:91–115, 1989.
5. M. Blum. Independent unbiased coin flips from a correlated biased source-a finite stae markov chain. *Combinatorica*, 6(2):97–108, 1986.
6. J. Bourgain. More on the sum-product phenomenon in prime fields and its applications. *International Journal of Number Theory*, 1:1–32.
7. J. Bourgain. On the construction of affine extractors. *Geometric And Functional Analysis*, 17(1):33–57, 2007.
8. B. Chor and O. Goldreich. Unbiased bits from sources of weak randomness and probabilistic communication complexity. *SIAM Journal on Computing*, 17(2):230–261, April 1988. Special issue on cryptography.
9. B. Chor, O. Goldreich, J. Hastad, J. Friedman, S. Rudich, and R. Smolensky. The bit extraction problem or t-resilient functions. In *FOCS 1985*, pages 396–407.
10. A. Cohen and A. Wigderson. Dispersers, deterministic amplification and weak random sources. In *FOCS 1989*, pages 14–25.
11. Y. Dodis, A. Elbaz, R. Oliveira, and R. Raz. Improved randomness extraction from two independent sources. In *RANDOM 2004*, pages 334–344.
12. Z. Dvir, A. Gabizon, and A. Wigderson. Extractors and rank extractors for polynomial sources. In *FOCS 2007*, pages 52–62.
13. A. Fiat and M.Naor. Implicit O(1) probe search. *SICOMP: SIAM Journal on Computing*, 22, 1993.
14. A. Gabizon and R. Raz. Deterministic extractors for affine sources over large fields. In *FOCS 2005*, pages 407–418.

15. A. Gabizon, R. Raz, and R. Shaltiel. Deterministic extractors for bit-fixing sources by obtaining an independent seed. *SICOMP: SIAM Journal on Computing*, 36(4):1072–1094, 2006.
16. O. Goldreich. A sample of samplers – A computational perspective on sampling (survey). In *ECCCTR: Electronic Colloquium on Computational Complexity, technical reports*, 1997.
17. R. L. Graham, B. L. Rothschild, and J. H. Spencer. *Ramsey Theory*. Wiley, 1980.
18. V. Guruswami, C. Umans, and S. P. Vadhan. Unbalanced expanders and randomness extractors from parvaresh-vardy codes. In *CCC 2007*, pages 96–108.
19. J. Kamp, A. Rao, S. Vadhan, and D. Zuckerman. Deterministic extractors for small-space sources. In *STOC 2006*, pages 691–700.
20. J. Kamp and D. Zuckerman. Deterministic extractors for bit-fixing sources and exposure-resilient cryptography. *SIAM J. Comput*, 36(5):1231–1247, 2007.
21. C. Lu, O. Reingold, S. Vadhan, and A. Wigderson. Extractors: Optimal up to constant factors. In *STOC 2003*, pages 602–611.
22. E. Mossel and C. Umans. On the complexity of approximating the vc dimension. In *CCC 2001*, pages 220–225.
23. N. Nisan and D. Zuckerman. Randomness is linear in space. *Journal of Computer and System Sciences*, 52(1):43–52, 1996.
24. J. Radhakrishnan and A. Ta-Shma. Bounds for dispersers, extractors, and depth-two superconcentrators. *SIAM Journal on Discrete Mathematics*, 13(1):2–24, 2000.
25. A. Rao. Extractors for a constant number of polynomially small min-entropy independent sources. In *STOC 2006*, pages 497–506.
26. A. Rao. Extractors for low weight affine sources. *Unpublished Manuscript*, 2008.
27. R. Raz. Extractors with weak random seeds. In *STOC 2005*, pages 11–20.
28. M. Santha and U. V. Vazirani. Generating quasi-random sequences from semi-random sources. *Journal of Computer and System Sciences*, 33:75–87, 1986.
29. R. Shaltiel. How to get more mileage from randomness extractors. In *CCC 2006*, pages 46–60.
30. R. Shaltiel. Recent developments in explicit constructions of extractors. *Bulletin of the EATCS*, 77:67–95, 2002.
31. L. Trevisan and S. Vadhan. Extracting randomness from samplable distributions. In *FOCS 2000*, pages 32–42.
32. U. Vazirani. Strong communication complexity or generating quasi-random sequences from two communicating semi-random sources. *Combinatorica*, 7:375–392, 1987.
33. J. von Neumann. Various techniques used in connection with random digits. *Applied Math Series*, 12:36–38, 1951.
34. A. C.-C. Yao. Should tables be sorted? *J. ACM*, 28(3):615–628, 1981.