Extracting Randomness Using Few Independent Sources

Boaz Barak*
Institute for Advanced Study, Princeton, NJ.
boaz@ias.edu

Russell Impagliazzo†
University of California, San Diego, La Jolla, CA.
russell@cs.ucsd.edu

Avi Wigderson‡
Institute for Advanced Study, Princeton, NJ.
avi@ias.edu

Abstract

In this work we give the first deterministic extractors from a constant number of weak sources whose entropy rate is less than 1/2. Specifically, for every $\delta > 0$ we give an explicit construction for extracting randomness from a constant (depending polynomially on $1/\delta$) number of distributions over $\{0,1\}^n$, each having min-entropy δn. These extractors output n bits, which are 2^{-n} close to uniform. This construction uses several results from additive number theory, and in particular a recent one by Bourgain, Katz and Tao [BKT03] and of Konyagin [Kon03].

We also consider the related problem of constructing randomness dispersers. For any constant output length m, our dispersers use a constant number of identical *distributions, each with min-entropy $\Omega(\log n)$ and outputs* every *possible m-bit string with positive probability. The main tool we use is a variant of the "stepping-up lemma" used in establishing lower bound on the Ramsey number for hypergraphs (Erdos and Hajnal, [GRS80]).*

1 Introduction

1.1 Background

Randomness is prevalent in computer science, and is widely used in algorithms, distributed computing, and cryptography. Perhaps the main motivation and justification for the use of randomness in computation is that randomness does exist in nature, and thus it is possible to sample natural phenomena (such as tossing coins) in order to make random choices in computation. However, there is a discrepancy between the type of random input that we expect when designing randomized algorithms and protocols, and the type of random data that can be found in nature. While randomized algorithms and protocols expect a stream of independent uniformly distributed random bits, this is too much to hope for from samples of natural phenomena.

Thus, a natural and widely studied problem has been the problem of constructing *randomness extractors*.[1] Loosely speaking, a *randomness extractor* is a (deterministic polynomial-time computable) function $\mathsf{Ext} : \{0,1\}^n \to \{0,1\}^m$, such that whenever $\mathcal{X}$ is a "good" random variable (where the definition for a "good" will be discussed shortly), then $\mathsf{Ext}(\mathcal{X})$ is statistically close to the uniform distribution. The random variable $\mathcal{X}$ is supposed to model the natural data, while the output of the extractor will be used to make random choices in a probabilistic algorithm or protocol.

Intuitively, the "good" distributions should be the distributions that contain more than m bits of entropy. Otherwise, information theoretic considerations show that such an extractor cannot exist, even if it is supposed to extract randomness from a fixed source that is known to the designer. Actually, past works have converged to the measure of *min-entropy* [CG85, Zuc91] which is a stronger notion than standard (Shannon) entropy.[2] Indeed, the same information theoretic considerations show that if $\mathsf{Ext}(\mathcal{X})$ is statistically close to being uniform, the distribution $\mathcal{X}$ must be statistically close to having min-entropy at least m. Thus possessing high min-entropy is a minimal requirement from the input distribution.

*Supported by NSF grants DMS-0111298 and CCR-0324906.

†Much of this research was performed at the Institute for Advanced Study in Princeton, NJ while supported by the State of New Jersey. Research partially supported by NSF Award CCR-0098197.

‡Partially supported by NSF grant CCR-0324906.

[1] We note that while this has been the original motivation for studying randomness extractors, such constructs have found numerous other applications in Computer Science, and are fundamental and interesting objects in their own right.

[2] The *min-entropy* of a random variable $\mathcal{X}$, denoted by $H^\infty(\mathcal{X})$, is equal to $\min_{x \in \mathsf{Supp}(\mathcal{X})}(-\log \Pr[\mathcal{X} = x])$. It can be easily seen that the min-entropy of a random variable $\mathcal{X}$ is always smaller than the (Shannon) entropy of X. It is known that any random variable $\mathcal{X}$ has min-entropy k a convex combination of random variables $\mathcal{X}_1, \ldots, \mathcal{X}_t$, where each $\mathcal{X}_i$ is the uniform distribution over some subset of size 2^k. Such random variables are called *flat*.

1.2 Seeded and Seedless Extractors

Unfortunately, it is not hard to see that there is no *single* function Ext which will produce a (close to) uniform output on *every* input distribution having high min-entropy. Previous works have dealt with this problem in two ways: the first way is to add a short truly random *seed* as a secondary input to the extractor, and the second way is to use no seed, but make further assumptions on the structure of the weak sources (in addition to the minimal assumption of it containing sufficient min-entropy).

Seeded extractors. One approach has been to allow the extractor to be *probabilistic*. That is, in addition to its input $\mathcal{X}$, we allow the function Ext to have an additional input $\mathcal{Y}$ which is uniformly distributed.To avoid trivialities, such as the extractor that simply outputs $\mathcal{Y}$, we will require the input $\mathcal{Y}$ to be (much) *shorter* than the output length m. We call the additional input $\mathcal{Y}$ the *seed* of the extractor, and thus we call such constructions *seeded* extractors.[3] Seeded extractors have been studied extensively in the past two decades (see the survey [Sha02]) with a series of exciting results, techniques, and applications. The current state of the art is a construction of extractors that are nearly optimal in the sense that they use a seed $\mathcal{Y}$ of length $O(\log n)$ bits, extracting essentially all the min-entropy of the source [LRVW03]. This in particular means that using such extractors, together with enumeration over all possible seed values, it is possible to simulate any probabilistic algorithm with polynomial overhead, using only a high min-entropy source.

Seedless extractors. There are many reasons to try and have seedless extractors. One is that the polynomial overhead implied by the seed enumeration above is too expensive. Another is that certain applications in computer science, such as cryptography and distributed computing, intrinsically prohibit such enumeration. For example, when using a weak source of randomness to choose an encryption key via a seeded extractor, it will certainly be *insecure* to encrypt a secret message with all secret keys produced by all seeds. More generally, it seems that we cannot always use directly seeded extractors in cryptography (see [MP90, DS02, BST03] for different cryptographic models under weak sources of randomness).

There have been many works constructing such seedless extractors (that work for specific families of high min-entropy sources). The first to consider this problem (and indeed, the first to consider the problem of randomness extraction) was von Neumann, who gave a seedless extractor from a stream of *biased* but independent bits [vN51] (see also [Per92]). Other works, such as [Blu84, SV84, CG85, CW89, CGH+85, MU02, KZ03, TV00] constructed seedless extractors for more general families of sources.[4]

[3] We remark that in most of the literature, the name randomness extractor (without any qualifiers) refers to what we call here a *seeded* randomness extractor.

1.3 Seedless extractors from few independent sources

When seedless extraction from one source is impossible, it is natural to consider doing so from several independent sources of the same quality. After all, assuming we have such a source in nature does not seem much weaker than assuming we have several.

The first to consider this problem were Santha and Vazirani [SV84], who showed how to use $O(\log n)$ independent "semi-random"[5] sources of length n and min-entropy δn for every constant $\delta > 0$.

Chor and Goldreich [CG85] were the first to consider general min-entropy sources, and proved that if $\delta > n/2$ than two sources suffice: indeed the Hadamard-Sylvester matrix $H : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ defined by $H(x,y) = \langle x, y \rangle$ (with the inner product in GF(2)) is such an extractor[6]. Vazirani [Vaz85] extended this to show that one can use a similar function to output a linear number of bits which are exponentially close to the uniform distribution.

Improving [CG85] seems hard. Even its disperser version, namely having a non-constant output for any two sources of entropy less rate below 1/2, is the notorious "Bipartite Ramsey" problem is still open. A slight improvement for this problem was made this year by Pudlak and Rödl [PR04], who lowered the min-entropy requirement for such a disperser to $n/2 - \sqrt{n}$, but getting a constant $\delta < 1/2$ remains a barrier for 2 sources. (Very recently, Barak, Kindler, Shaltiel, Sudakov and Wigderson [BKS+04] used the results of the current work to obtain such a disperser for entropy δn with $\delta > 0$ an arbitrarily small constant.)

An alternative 2-source function is the Paley matrix, $P : GF(p) \times GF(p) \to \{\pm 1\}$ defined by $P(x,y) = \chi_2(x+y)$ (where p is an n-bit prime and $\chi_2 : \mathrm{GF}(p) \to \{\pm 1\}$ is the quadratic character). P too is an extractor with exponentially small error for entropy $> n/2$, and is conjectured to have the same property for entropy δn for all $\delta > 0$. While generally believed, proving this conjecture seems beyond current techniques. Assuming even more, Zuckerman [Zuc90] showed that if this conjecture holds for *all* multiplicative characters (not just χ_2), than a constant (actually $poly(1/\delta)$) sources suffice for extraction of linearly many bits with exponential error. The extractor we use here is

[4] The work of Trevisan and Vadhan [TV00] differs from all the rest, as well as from ours, in that it works in the computational setting; the restriction on the family of sources is computational - they are efficiently sampleable, and the extractors work assuming an unproven computational assumption.

[5] We will not define them formally here, only note that they are weaker than high min-entropy sources, but nevertheless one cannot extract seedlessly from only one such source.

[6] For this construction the 1/2 entropy bound is tight - there are two sources of entropy exactly $n/2$ on which H is constant

exactly the same as Zuckerman's (but our analysis uses no unproven assumptions).

1.4 Our Results

1.4.1 Multiple-Sample Extractors.

In this work we will be interested in extracting randomness from a *constant* (larger than 2) number of samples from a high min-entropy source.[7] Loosely speaking, our main result is an efficient construction to extract (almost) uniformly distributed bits from a constant number of samples from independent distributions over $\{0,1\}^n$ each having min-entropy at least δn, for an arbitrarily small constant $\delta > 0$. The statistical distance of our output from the uniform distribution is exponentially low (i.e., $2^{-\Omega(n)}$).[8] More formally, our main theorem is the following:

Theorem 1.1 (Multiple-sample extractor). *For every constant $\delta > 0$ there exists a constant $\ell = (1/\delta)^{O(1)}$ and a polynomial-time computable function $\mathsf{Ext} : \{0,1\}^{n\cdot\ell} \to \{0,1\}^n$ such that for every independent random variables $\mathcal{X}_1, \ldots, \mathcal{X}_\ell$ over $\{0,1\}^n$ satisfying $H^\infty(\mathcal{X}_i) \geq \delta n$ for $i = 1, \ldots, \ell$, it holds that $\mathsf{dist}\Big(\mathsf{Ext}(\mathcal{X}_1, \ldots, \mathcal{X}_\ell), U_n\Big) < 2^{-\Omega(n)}$*

U_n denotes the uniform distribution over the set $\{0,1\}^n$ and $\mathsf{dist}(\mathcal{X}, \mathcal{Y})$ denotes the *statistical distance* of two distributions $\mathcal{X}$ and $\mathcal{Y}$. That is, $\mathsf{dist}(\mathcal{X}, \mathcal{Y}) \stackrel{def}{=} \frac{1}{2}\sum_{i\in\mathsf{Supp}(\mathcal{X})\cup\mathsf{Supp}(\mathcal{Y})} |\Pr[\mathcal{X} = i] - \Pr[\mathcal{Y} = i]|$. As mentioned in Footnote 2, we denote by $H^\infty(\mathcal{X})$ the min-entropy of the random variable $\mathcal{X}$.

Remark 1.2. It is possible to reduce the statistical distance of our extractor by XOR'ing together independent samples. In particular, for every c, we can ensure that the output of our extractor will be 2^{-cn}-close to the uniform distribution by multiplying the number of samples we use by a factor of $O(c)$. We leave the details to the full version of this work.

1.4.2 Dispersers.

We also present a construction of *dispersers* that works for much lower entropy samples of the *same* source. A *disperser* is a relaxed variant of an extractor, which does not need to output a distribution statistically close to uniform, but rather only a distribution with large support. Thus, when talking about dispersers, it is natural to consider only the support of their input, and so talk about the inputs as sets rather than random variables (where a set of size 2^k corresponds to a random variable of min-entropy k). The formal definition of dispersers is in the next section. Our main result regarding dispersers is the following theorem:

Theorem 1.3 (Multiple-sample same-source disperser). *There exists constants ℓ and d such that for every m, there is polynomial-time computable function $\mathsf{Disp} : \{0,1\}^{n\cdot\ell} \to \{0,1\}^m$ satisfying that for every subset $\mathcal{X} \subseteq \{0,1\}^n$ with $|\mathcal{X}| \geq n^{d2^m}$ $\mathsf{Disp}(\mathcal{X}, \ldots, \mathcal{X}) = \{0,1\}^m$*

That is, if $|\mathcal{X}| \geq n^{d2^m}$ then for every $y \in \{0,1\}^m$, there exist $x_1, \ldots, x_\ell \in \mathcal{X}$ such that $\mathsf{Disp}(x_1, \ldots, x_\ell) = y$.

This is a better disperser than the multiple-source extractor mentioned above in the sense that it works for very low min-entropy (indeed note that if we let m be a constant, then sets of size $n^{O(1)}$ correspond to distributions of min-entropy of $O(\log n)$). However, it still has two drawbacks: one drawback is that its output is much smaller than the input entropy (although it can still be much larger than the number of samples); another drawback is that it requires all its input samples to come from the *same* distribution $\mathcal{X}$ (rather than from different distributions, as in our extractor). We consider the second drawback to the more serious one. The construction of this disperser closely follows the "stepping-up" technique of Erdos and Hajnal [GRS80, Sec. 4.7] for giving an (explicit) lower-bound on the Ramsey number of hypergraphs. We remark that we work for worse entropy than this lower-bound[9] because we want to output a super-constant number of bits (that does not depend on the number of samples).

1.5 Our Techniques

An Erdos-Szemeredi theorem for finite fields. Our main tools for Theorem 1.1 are several results from additive number theory and in particular a relatively recent result by Bourgain, Katz and Tao [BKT03]. They proved an analog of the Erdos-Szemeredi theorem for finite prime fields. Let $\mathcal{A}$ be a subset of some field $\mathbb{F}$. We define the set $\mathcal{A} + \mathcal{A}$ to equal $\{a + b \mid a, b \in \mathcal{A}\}$ and the set $\mathcal{A} \cdot \mathcal{A}$ to equal $\{a \cdot b \mid a, b \in A\}$. Note that $|\mathcal{A}| \leq |\mathcal{A} + \mathcal{A}| \leq |\mathcal{A}|^2$ (and similarly $|\mathcal{A}| \leq |\mathcal{A} \cdot \mathcal{A}| \leq |\mathcal{A}|^2$). An example for a set $\mathcal{A}$ where $\mathcal{A} + \mathcal{A}$ is small (of size about $2|\mathcal{A}|$) is an *arithmetic progression*. An example for a set $\mathcal{A}$ where $\mathcal{A} \cdot \mathcal{A}$ is small is a *geometric progression*. The Erdos-Szemeredi theorem is that for every set $\mathcal{A} \subseteq \mathbb{N}$, either $\mathcal{A} + \mathcal{A}$ or $\mathcal{A} \cdot \mathcal{A}$ is of size at least $|\mathcal{A}|^{1+\epsilon_0}$, for some universal constant ϵ_0. In some sense, one can view this theorem as saying that a set of integers can't be simultaneously close to both an arithmetic progression and a geometric progression.

A natural question, with many diverse applications in geometry and analysis, is whether this theorem also holds in *finite* fields. A first observation is that this theorem is *false*

[7] We note that our results can be easily generalized to the case of a larger (i.e., super-constant) number of samples.

[8] Note that we get statistical distance that is much better than what is possible to obtain with seeded extractors, which cannot do better than a polynomially small distance when using a logarithmic-length seed. This low statistical distance is important for cryptographic applications.

[9] Using c samples these lower-bound yield a 1-bit-output disperser for $\log^{(\Theta(c))} n$ entropy, where we use $\log^{(i)} n$ to denote iterating log i times on n.

in a field $\mathbb{F}$ that contains a non-trivial subfield $\mathbb{F}'$. This because if we let $\mathcal{A} = \mathbb{F}'$ then $\mathcal{A}+\mathcal{A} = \mathcal{A}\cdot\mathcal{A} = \mathcal{A}$. However, [BKT03] showed that a variant of this theorem does hold in a finite field with no non-trivial subfields.[10] In particular it holds in the fields $\mathrm{GF}(p)$ and $\mathrm{GF}(2^p)$ for every prime p. That is, they proved the following:

Theorem 1.4 ([BKT03]). *Let $\delta > 0$ be some constant and let $\mathbb{F}$ be an N-element field with no subfield $\mathbb{F}'$ of size in $[N^\delta, N)$. Let $\mathcal{A} \subseteq \mathbb{F}$ be a set such that $|\mathcal{A}| \in (N^\delta, N^{1-\delta})$. Then, there exist some constant ϵ (depending on δ) such that* $\max\{|\mathcal{A}+\mathcal{A}|, |\mathcal{A}\cdot\mathcal{A}|\} > |\mathcal{A}|^{1+\epsilon}$

In [BKT03], the constant ϵ was $\Theta(\frac{1}{\delta^2})$. Konyagin [Kon03] gave a stronger result for prime fields, and showed that, as long as $|\mathcal{A}| < |\mathbb{F}|^{0.99}$, ϵ can be made independent of the size of $\mathcal{A}$ (even if $|\mathcal{A}|$ is very small). That is, he proved the following theorem:

Theorem 1.5 ([Kon03]). *There exist some constant ϵ_0 such that for every $\mathbb{F}$ which is a field of prime order, and every $\mathcal{A} \subseteq \mathbb{F}$ with $|\mathcal{A}| < |\mathbb{F}|^{0.99}$,* $\max\{|\mathcal{A}+\mathcal{A}|, |\mathcal{A}\cdot\mathcal{A}|\} > |\mathcal{A}|^{1+\epsilon_0}$

1.6 How is this related to extractors?

Using Theorem 1.5 to obtain dispersers. Theorem 1.5 actually already implies some kind of multiple-sample disperser that was not previously known. This is because it implies the following corollary:

Corollary 1.6. *There exist some constant ϵ_0 such that for every $\mathbb{F}$ which is a field of prime order, and every $\mathcal{A} \subseteq \mathbb{F}$ with $|\mathcal{A}| < |\mathbb{F}|^{0.99}$,* $|\mathcal{A}\cdot\mathcal{A}+\mathcal{A}| > |\mathcal{A}|^{1+\epsilon_0}$

In other words, it implies that $|f(\mathcal{A},\mathcal{A},\mathcal{A})| \geq |\mathcal{A}|^{1+\epsilon_0}$ where $f(a,b,c) = a\cdot b + c$. Yet this implies that for every set $\mathcal{A}$ with $|\mathcal{A}| \geq |\mathbb{F}|^\delta$, if we compose f with itself $O(\log(1/\delta))$ times we get a function $g : \mathbb{F}^\ell \to \mathbb{F}$ such that $g(\mathcal{A},\ldots,\mathcal{A}) = \mathbb{F}$, where $\ell = 2^{O(\log(1/\delta))} = (1/\delta)^{O(1)}$.[11] If we identify the field $\mathbb{F}$ with the strings $\{0,1\}^n$, then we see that this function g is already some kind of a disperser.

Obtaining extractors. To obtain an *extractor*, rather than a disperser we would like to obtain a *statistical* analog of Theorem 1.5. Consider the following notation: if $\mathcal{X}$ is a random variable then we define $\mathcal{X}+\mathcal{X}$ to be the distribution obtained by choosing a, b independently at random from $\mathcal{X}$ and outputting $a+b$, and define $\mathcal{X}\cdot\mathcal{X}$ in a similar way. Then, a statistical analog of Theorem 1.5 would be that there exists some $\epsilon > 0$ such that for every random variable $\mathcal{X}$ with min-entropy at most $0.9\log|\mathbb{F}|$, either the distribution $\mathcal{X}+\mathcal{X}$ or the distribution $\mathcal{X}\cdot\mathcal{X}$ has min-entropy at least $(1+\epsilon)H^\infty(\mathcal{X})$. Unfortunately, this is false: for every prime field $\mathbb{F}$, there is a random variable $\mathcal{X}$ which is uniform over some set of size 2^k (for $k \ll 0.9\log|\mathbb{F}|$) and such that for both $\mathcal{X}+\mathcal{X}$ and $\mathcal{X}\cdot\mathcal{X}$, a constant fraction of the probability mass is concentrated in a set of size $O(2^k)$.[12]

Even though the statistical analog for Theorem 1.5 does not hold, we show that a statistical analog for Corollary 1.6 *does* hold. That is, we prove (in Lemma 3.1) that there exists some constant $\epsilon_0 > 0$ such that for every random variable $\mathcal{X}$ over a prime field $\mathbb{F}$, the distribution $\mathcal{X}\cdot\mathcal{X}+\mathcal{X}$ has (up to some negligible statistical distance) min-entropy at least $(1+\epsilon_0)H^\infty(\mathcal{X})$. (In fact, we need to (and do) prove a more general statement regarding distributions of the form $\mathcal{X}\cdot\mathcal{Y}+\mathcal{Z}$.) The proof of this lemma is the main technical step in our extractor. The proof uses Theorem 1.5, along with some other additive number-theoretic results of Rusza [Ruz96] and Gowers [Gow98].

We use this lemma to show that the function g sketched above is actually not just a disperser but an *extractor*. That is, we show that for every random variable $\mathcal{X}$ of min-entropy at least $\delta\log|\mathbb{F}|$, $g(\mathcal{X},\ldots,\mathcal{X})$ when considered as a random variable, not only has large support but is in fact very close to the uniform distribution. (In fact, we need to (and do) prove a stronger version of this version. Namely, that $g(\mathcal{X}_1,\ldots,\mathcal{X}_\ell)$ is close to uniform for every random variables $\mathcal{X}_1,\ldots,\mathcal{X}_\ell$ with $H^\infty(\mathcal{X}_i) \geq \delta\log|\mathbb{F}|$ for $i = 1,\ldots,\ell$.)

2 Preliminaries

In this section we establish our definitions for multiple-sample extractors and dispersers. Unfortunately, such definitions tend to have a large number of parameters.

Definition 2.1 (Multiple-sample extractor). A function $\mathsf{Ext} : \{0,1\}^{n\cdot\ell} \to \{0,1\}^m$ is called a *ℓ-sample (k,m,ϵ)-extractor* if for every independent random variables $\mathcal{X}_1,\ldots,\mathcal{X}_\ell$ satisfying $H^\infty(\mathcal{X}_i) \geq k$ for $i = 1,\ldots,\ell$ it holds that $\mathsf{dist}\Big(\mathsf{Ext}(\mathcal{X}_1,\ldots,\mathcal{X}_\ell), U_m\Big) < \epsilon$

Parameters and qualifiers:

- The parameter ℓ is called the *number of samples* of the extractor. In all of our constructions we will use ℓ which is a constant independent of the input length.
- The parameter m is called the *output length* of the extractor. Usually, we'll use $m = n$.
- The parameter k is called the min-entropy requirement by the extractor.

[10] By a *trivial* subfield in this context we mean a subfield that is either the entire field or very small (i.e., of constant size).

[11] This is because each time we apply f we grow in the set size from m to $m^{1+\epsilon_0}$. We note that one needs to analyze separately the case that $|\mathcal{A}| > |\mathbb{F}|^{0.99}$ but this can be done. This function g is described in more detail in the beginning of Section 3.

[12] An example for such a random variable $\mathcal{X}$ is a random variable which is with probability half an arithmetic progression and with probability half a geometric progression.

- The parameter ϵ is called the statistical distance of the extractor.
- One can also make a weaker definition in which the extractor is required to output a close to uniform value only if the variables $\mathcal{X}_1, \ldots, \mathcal{X}_\ell$ are identically distributed. We call such an extractor a *same-source* extractor. Thus, we will sometimes say that a standard (as per Definition 2.1) multiple-sample extractor is a *different-source* extractor.

We now define the weaker notion of a disperser:

Definition 2.2 (Multiple-sample disperser). A function $\mathsf{Disp} : \{0,1\}^{n \cdot \ell} \to \{0,1\}^m$ is called a *ℓ-sample (k,m)-disperser* if for all sets $\mathcal{X}_1, \ldots, \mathcal{X}_\ell$ satisfying $|\mathcal{X}_i| \geq 2^k$ for $i = 1, \ldots, \ell$ it holds that $\mathsf{Disp}(\mathcal{X}_1, \ldots, \mathcal{X}_\ell) = \{0,1\}^m$

Notes:

- We will use the same names for the parameters and qualifiers of dispersers as we used for extractors (e.g., number of samples, output length, same-source/different-source).
- In previous literature dispersers are usually defined with a parameter ϵ analogous to the statistical distance of extractors, requiring that $|\mathsf{Disp}(\mathcal{X}_1, \ldots, \mathcal{X}_\ell)| \geq (1-\epsilon)2^m$ instead of the requirement we make. Thus, one can think of our definition as setting $\epsilon = 0$. However, in our particular setting of independent samples, this is essentially without loss of generality as we can convert a disperser satisfying the weaker definition with any constant $\epsilon < 1/2$ into a disperser satisfying the stronger definition with only a constant factor increase in the number of samples.[13]

Basic facts and observations. The following facts regarding multiple-sample extractors and dispersers are either known or easy to verify:

- There does not exist a 1-sample disperser (and hence extractor) even if with only one bit of output and even if the source is assumed to have $n-1$ bits of min-entropy.
- There is a simple construction for a 2-sample *same-source* 1-bit output extractor whenever the source min-entropy is larger than $O(\log \frac{1}{\epsilon})$ (where ϵ is the statistical distance). This is the extractor that on input x, y outputs 1 if $x > y$.

[13] This can be done for example by identifying the set $\{0,1\}^n$ with elements of a prime field $\mathbb{F}$ (as we do in the sequel) and using the Cauchy-Davenport theorem. This theorem says that if $\mathcal{A}$ and $\mathcal{B}$ are subsets of a prime field $\mathbb{F}$, then the set $\mathcal{A}+\mathcal{B} = \{a+b | a \in \mathcal{A}, b \in \mathcal{B}\}$ is of size at least $\min\{|\mathbb{F}|, |\mathcal{A}|+|\mathcal{B}|-1\}$. Hence, if $\mathcal{A}_1, \ldots, \mathcal{A}_k$ are sets of size $\epsilon|\mathbb{F}|$, $k > 1/\epsilon$ and $|\mathcal{A}_i| > \epsilon|\mathbb{F}|$ for all $i \in [k]$ then $\mathcal{A}_1 + \ldots + \mathcal{A}_k = \mathbb{F}$.

- In contrast, by Ramsey-type arguments, a 2-sample same-source disperser (and hence an extractor) with more than 1 bit of output requires the input min-entropy to be at least $\log n$. The same holds for a different-source 1-bit output disperser.

  The best known previous explicit constructions for both cases require the min-entropy to be more than $n/2$ [CG85, Vaz87]. Indeed, the function $H : \{0,1\}^{2n} \to \{0,1\}$ defined as follows $H(x,y) = \sum x_i y_i \pmod 2$ (i.e., the adjacency function of the Hadamard graph) is known to be a 1-bit output extractor for sources with more than $n/2$ entropy [CG85]. This is essentially the best known previous construction in terms of the minimum entropy requirement.[14] There has been some improvement in obtaining variants of this extractor that have a larger output size [Vaz87, Elb03, DO03] (although these still require at least $n/2$ entropy).
- Using enumeration over all possible seeds, one can use a *seeded* extractor to obtain a polynomial-samples same-source extractor with the same requirement over the min-entropy. It is also possible to generalize this to work for *different sources*[SRW03].

Explicit constructions. In the definition of extractors and dispersers we did not require these functions to be efficiently computable. However, we will naturally be interested in obtaining extractors and dispersers that are computable in polynomial-time. We call such constructions *explicit*.

3 Constructing a Multiple-Sample Extractor

In this section we prove Theorem 1.1. That is, we construct an extractor Ext (where $\mathsf{Ext} : \{0,1\}^{\ell \cdot n} \to \{0,1\}^n$) such that $\mathsf{Ext}(\mathcal{X}_1, \ldots, \mathcal{X}_\ell)$ is statistically close to the uniform distribution for every independent random variables $\mathcal{X}_1, \ldots, \mathcal{X}_\ell$ over $\{0,1\}^n$ with high enough min-entropy (i.e., at least δn where $\ell = \mathrm{poly}(1/\delta)$). Our extractor will be very simple, involving only a recursive composition of the operation $(a,b,c) \mapsto a \cdot b + c$. As noted in the introduction, this is the same exact construction used by Zuckerman [Zuc90].

Formally, we fix a suitable field $\mathbb{F}$ (see below) and define the functions Ext^i for all $i \in \mathbb{N}$ recursively as follows:

- For every i, the function Ext^i will be a mapping from $\mathbb{F}^{3^i}$ to $\mathbb{F}$.
- $\mathsf{Ext}^0 : \mathbb{F}^{3^0} = \mathbb{F} \to \mathbb{F}$ is the identity function $\mathsf{Ext}^0(x) = x$.

[14] This is assuming we make the same entropy requirement from both sources. It is known that the adjacency function of the Paley graph is a 1-bit output extractor which works as long as one of its inputs has more than $n/2$ entropy while the other input only needs to have more than $\log n$ entropy.

- Assume Ext^i is already defined. We define
$\mathsf{Ext}^{i+1} : \mathbb{F}^{3^{i+1}} = \left(\mathbb{F}^{3^i}\right)^3 \to \mathbb{F}$ as follows:
for every $x_1, x_2, x_3 \in \mathbb{F}^{3^i}$,
$\mathsf{Ext}^{i+1}(x_1, x_2, x_3) \stackrel{def}{=} \mathsf{Ext}^i(x_1) \cdot \mathsf{Ext}^i(x_2) + \mathsf{Ext}^i(x_3)$

Our extractor will be equal to Ext^i for a suitably chosen constant i. When extracting from a string in $\{0,1\}^n$ we will choose as the field $\mathbb{F}$ to be of the form $\mathrm{GF}(N)$ for some prime $N \in [2^n, 2^{n+1})$ (and hence we can identify any string in $\{0,1\}^n$ with an element in the field $\mathbb{F}$). We postpone the issue of finding such a prime to Section 3.4.

Theorem 1.1 will follow from the following two lemmas:

Lemma 3.1. *There exists some constant $\epsilon > 0$ such that for every distributions $\mathcal{A}, \mathcal{B}, \mathcal{C}$ over a prime field $\mathbb{F}$ each with min-entropy at least m, the distribution $\mathcal{A} \cdot \mathcal{B} + \mathcal{C}$ is $2^{-\epsilon m}$-close to having min-entropy at least $\min\{(1+\epsilon)m, 0.9 \log |\mathbb{F}|\}$.*

Lemma 3.2. *Let $\mathcal{A}_1, \ldots, \mathcal{A}_9$ be 9 independent distributions over a prime field $\mathbb{F}$ each with min-entropy at least $0.9 \log |\mathbb{F}|$. Then, $\mathsf{Ext}^2(\mathcal{A}_1, \ldots, \mathcal{A}_9)$ is of distance at most $|F|^{-0.01}$ from the uniform distribution over $\mathbb{F}$.*

Lemmas 3.1 and 3.2 imply Theorem 1.1. Indeed, Lemma 3.1 implies that for every constant $\delta > 0$, if we let $i = \log_{(1+\epsilon)}(1/\delta)$ then the output of Ext^i is close to having min-entropy $0.9 \log |\mathbb{F}|$. Note that the number of samples Ext^i requires is 3^i which is polynomial in $1/\delta$. We use Lemma 3.2 to get from min-entropy $0.9 \log |\mathbb{F}|$ to a close to uniform output. Using the union bound, the statistical distance of the extractors output from the uniform distribution on $\mathbb{F}$ will be at most $N^{-\Omega(1)}$. We defer the proof of Lemma 3.2 to the full version of this work (see Appendix A for the ideas behind the proof) and now turn to proving Lemma 3.1.[15]

3.1 Basic Facts and Notations

Before proving Lemma 3.1, we introduce some notations and some tools that will be used in the course of the proof. We identify a set $\mathcal{A}$ with the uniform distribution on elements of $\mathcal{A}$. If $\mathcal{D}$ is a distribution then we denote by $\mathsf{cp}(\mathcal{D})$ the collision probability of $\mathcal{D}$; that is, $\mathsf{cp}(\mathcal{D}) = \Pr_{x,y \leftarrow_{\mathrm{R}} \mathcal{D}}[x = y]$. Note that for a set $\mathcal{A}$, $\mathsf{cp}(\mathcal{A}) = \frac{1}{|\mathcal{A}|}$. If $\mathcal{A}$ and $\mathcal{B}$ are distributions over subsets of the field $\mathbb{F}$, then we denote by $\mathcal{A} + \mathcal{B}$ the *distribution* obtained by picking a at random from $\mathcal{A}$, b at random from $\mathcal{B}$ and outputting $a + b$. Note that this distribution may be far from the uniform distribution with the same support. We define the distribution $\mathcal{A} \cdot \mathcal{B}$ in a similar way. We say that a distribution $\mathcal{X}$ is a convex combination of distributions $\mathcal{X}_1, \ldots, \mathcal{X}_m$ if there exist numbers $p_1, \ldots, p_m \in [0,1]$ such that $\sum_i p_i = 1$ and the random variable $\mathcal{X}$ (when looked at as a vector of probabilities) is equal to $\sum_i p_i \mathcal{X}_i$.

The following lemmas would be useful: (We omit the proofs for lack of space)

Lemma 3.3. *Suppose that $\mathcal{X}$ is a convex combination of distributions $\mathcal{X}_1, \ldots, \mathcal{X}_m$. Then $\mathsf{cp}(\mathcal{X}) \leq \max\{\mathsf{cp}(\mathcal{X}_1), \ldots, \mathsf{cp}(\mathcal{X}_m)\}$.*

Lemma 3.4. *Let $\mathcal{X}$ be a distribution such that $\mathsf{cp}(\mathcal{X}) \leq \frac{1}{KL}$. Then $\mathcal{X}$ is of statistical distance $\frac{1}{\sqrt{L}}$ from having min-entropy at least $\log K$.*

Note that if $H^\infty(\mathcal{X}) \geq k$ then clearly $\mathsf{cp}(\mathcal{X}) \leq 2^{-k}$. Together with Lemma 3.4 this implies that, up to an arbitrarily close to 1 multiplicative factor and an exponentially small (in the min-entropy) statistical distance, $H^\infty(X)$ is approximately equal to $\log(1/\mathsf{cp}(\mathcal{X}))$.

3.2 Additive Number-Theoretic Results

The following two lemmas hold for any Abelian group, and so the "+" operator may be replaced in them by "·". Note that we didn't state these lemmas with the most optimal choice of constants.

Lemma 3.5 ([Ruz96] (see also [Nat96])). *Let $\mathcal{A}, \mathcal{B}$ be subsets of some Abelian group G with $|\mathcal{A}| = |\mathcal{B}| = M$ and let $\rho > 0$ be some number. If $|\mathcal{A} + \mathcal{B}| \leq M^{1+\rho}$ then $|\mathcal{A} + \mathcal{A}| \leq M^{1+2\rho}$.*

In other words, if $\mathcal{A}+\mathcal{B}$ is "small" for some $\mathcal{B}$ then $\mathcal{A}+\mathcal{A}$ is small. Note that Lemma 3.5 implies that we can replace $\mathcal{A} + \mathcal{A}$ and $\mathcal{A} \cdot \mathcal{A}$ in Theorem 1.4 with $\mathcal{A} + \mathcal{B}$ and $\mathcal{A} \cdot \mathcal{B}$.

We'll use the following lemma of Gowers (which is a quantitative improvement of a result by Balog and Szemeredi). We state it here using different sets $\mathcal{A}$,$\mathcal{B}$ (similarly to the way it is quoted in [BKT03], although in [Gow98] it is stated with $\mathcal{A} = \mathcal{B}$).

Lemma 3.6 ([Gow98]). *Let $\mathcal{A}, \mathcal{B}$ be subsets of some group G with $|\mathcal{A}| = |\mathcal{B}| = M$ and let $\rho > 0$ be such that $\mathsf{cp}(\mathcal{A} + \mathcal{B}) \geq \frac{1}{M^{1+\rho}}$. Then, there exists $\mathcal{A}' \subseteq \mathcal{A}$ and $\mathcal{B}' \subseteq \mathcal{B}$ such that $|\mathcal{A}'|, |\mathcal{B}'| \geq M^{1-10\rho}$ and $|\mathcal{A}' + \mathcal{B}'| \leq M^{1+10\rho}$.*

We remark that it can be shown that if $\mathcal{A} = \mathcal{B}$ in the conditions of this lemma, then $\mathcal{A}' = \mathcal{B}'$ in its conclusion.

3.3 Proof of Lemma 3.1

Fixing ϵ. We fix ϵ small enough such that Theorem 1.5 ensures us that if $\mathcal{X}$ is a set of size at least $M^{1-10^4\epsilon}$ then $\max\{|\mathcal{X} \cdot \mathcal{X}|, |\mathcal{X} + \mathcal{X}|\}$ is at least $M^{1+10^4\epsilon}$ (e.g., we can take $\epsilon = \epsilon_0/10^9$).

[15] Note that we could prove Theorem 1.1 without using Lemma 3.2 by applying the previously known two-sample extractors that work for entropy larger than $\frac{\log |\mathbb{F}|}{2}$. In addition, Salil Vadhan observed that using the fact that the function family $\{h_{b,c}\}$ (where $h_{b,c}(a) = a \cdot b + c$) is pairwise independent and the Leftover Hash Lemma of [HILL89], one can prove that under the conditions of Lemma 3.2 the first $0.8 \log |F|$ bits of $\mathcal{A}_1 \cdot \mathcal{A}_2 + \mathcal{A}_3$ are within statistical distance $|F|^{-0.01}$ to the uniform distribution. This is also sufficient to prove Theorem 1.1.

Statistical analog of Theorem 1.5. Roughly speaking, Theorem 1.5 says that if $|\mathcal{A}\cdot\mathcal{A}|$ is "small" then $|\mathcal{A}+\mathcal{A}|$ is "large". By combining Theorem 1.4 with Lemma 3.5, it is possible get a "different sets" analog of this statement and show that for every set $\mathcal{A}$ of size M, if $|\mathcal{A}\cdot\mathcal{B}|$ is "small" (i.e. $\leq M^{1+\epsilon}$) for some set $\mathcal{B}$ of size M, then $|\mathcal{A}+\mathcal{C}|$ is "large" ($\geq M^{1+\epsilon}$) for *every* set $\mathcal{C}$ of size M.

At this point we would have liked to obtain a collision probability analog of this statement. That is, we would like to prove that if $\mathcal{A},\mathcal{B},\mathcal{C}$ are uniform distributions over sets of size M, and $\mathsf{cp}(\mathcal{A}\cdot\mathcal{B})\geq\frac{1}{M^{1+\epsilon}}$ then $\mathsf{cp}(\mathcal{A}+\mathcal{C})\leq\frac{1}{M^{1+\epsilon}}$. Unfortunately, this is false, as can be witnessed by considering the following counterexample: Let $M=N^{0.1}$ and let $\mathcal{A}=\mathcal{B}=\mathcal{C}$ be a distribution that is an arithmetic progression of size M with probability $\frac{1}{2}$ and a geometric progression of size M with probability $\frac{1}{2}$. For such a distribution, both $\mathsf{cp}(\mathcal{A}+\mathcal{A})$ and $\mathsf{cp}(\mathcal{A}\cdot\mathcal{A})$ are at least $\Omega(1/M)$.[16]

However, we are able to prove a slightly weaker statement. That is, we show that if $\mathcal{A}\cdot\mathcal{B}$ is small in *set-size*, then $\mathcal{A}+\mathcal{C}$ shrinks significantly in *collision probability*. This is stated in the following lemma:

Lemma 3.7. *Let $\mathcal{A}\subseteq\mathbb{F}$ with $|\mathcal{A}|\geq M^{1-10\epsilon}$ such that $|\mathcal{A}\cdot\mathcal{B}|\leq M^{1+\epsilon}$, for some $\mathcal{B}$ with $|\mathcal{B}|\geq M^{1-10\epsilon}$. Then $\mathsf{cp}(\mathcal{A}+\mathcal{C})$ is smaller than $\frac{1}{M^{1+10\epsilon}}$ for* all *sets $\mathcal{C}$ of size M.*

Proof. If $\mathsf{cp}(\mathcal{A}+\mathcal{C})\geq\frac{1}{M^{1+10\epsilon}}$ then by applying Lemma 3.6 (with $\rho=20\epsilon$) it holds that there exists subsets $\mathcal{A}',\mathcal{C}'$ of $\mathcal{A}$ and $\mathcal{C}$ respectively such that $|\mathcal{A}'|,|\mathcal{C}'|\geq M^{1-200\epsilon}$ but $|\mathcal{A}'+\mathcal{C}'|\leq M^{1+200\epsilon}$. This means by Lemma 3.5 that $|\mathcal{A}'+\mathcal{A}'|\leq M^{1+2000\epsilon}$. However, this means that $|\mathcal{A}'\cdot\mathcal{A}'|\geq M^{1+10^4\epsilon}$ by Theorem 1.5 which implies that $|\mathcal{A}'\cdot\mathcal{B}|\geq M^{1+10^3\epsilon}$ by Lemma 3.5. However, since $\mathcal{A}'\subseteq\mathcal{A}$ this implies that $|\mathcal{A}\cdot\mathcal{B}|\geq M^{1+10^3\epsilon}$, contradicting our initial assumption. □

We call a set $\mathcal{A}$ satisfying the conclusion of Lemma 3.7 *plus-friendly* (since its collision probability grows when it is added with any other set of the same size). Reversing the roles of addition and multiplication we obtain the following lemma:

Lemma 3.8. *Let $\mathcal{A}\subseteq\mathbb{F}$ with $|\mathcal{A}|\geq M^{1-10\epsilon}$ such that $|\mathcal{A}+\mathcal{B}|\leq M^{1+\epsilon}$, for some $\mathcal{B}$ with $|\mathcal{B}|\geq M^{1-10\epsilon}$. Then $\mathsf{cp}(\mathcal{A}\cdot\mathcal{C})$ is smaller than $\frac{1}{M^{1+10\epsilon}}$ for* all *sets $\mathcal{C}$ of size M.*

We call a set $\mathcal{A}$ satisfying the conclusion of Lemma 3.8 *times-friendly*.

Using Lemmas 3.7 and 3.8, we now turn to obtaining a collision-probability analog of Corollary 1.6:

[16] Let $\mathcal{A}_{\mathsf{arith}}$ denote the arithmetic progression. A random element from $\mathcal{A}+\mathcal{A}$ is in the set $\mathcal{A}_{\mathsf{arith}}+\mathcal{A}_{\mathsf{arith}}$ with probability at least $\frac{1}{4}$. However, because the set $\mathcal{A}_{\mathsf{arith}}+\mathcal{A}_{\mathsf{arith}}$ is of size at most $O(M)$, we get that $\mathsf{cp}(\mathcal{A}+\mathcal{A})\geq\frac{1}{16}\Omega(1/M)=\Omega(1/M)$. The symmetrical reasoning shows that $\mathsf{cp}(\mathcal{A}\cdot\mathcal{A})\geq\Omega(1/M)$.

Lemma 3.9. *Let $\mathcal{A},\mathcal{B},\mathcal{C}\subseteq\mathbb{F}$ with $|\mathcal{A}|=|\mathcal{B}|=|\mathcal{C}|=M$. Then $\mathcal{A}\cdot\mathcal{B}+\mathcal{C}$ is $M^{-\epsilon}$-close to having collision probability at most $\frac{1}{M^{1+\epsilon}}$*

Proof. The proof follows by showing that the counterexample described above (of a distribution that is an arithmetic progression with probability half and a geometric progression with probability half) is in some sense the worst case for the theorem. Namely, we prove the following claim

Claim 3.9.1. *If $\mathcal{A},\mathcal{B},\mathcal{C}$ are as stated in the theorem then there exist two disjoint subsets $\mathcal{A}_1,\mathcal{A}_2$ of $\mathcal{A}$ such that*

1. *$\mathcal{A}_1\cup\mathcal{A}_2$ is almost all of $\mathcal{A}$. That is, $|\mathcal{A}_1\cup\mathcal{A}_2|\geq M-M^{1-\epsilon}$ (in other words, there is a convex combination of $\mathcal{A}_1$ and $\mathcal{A}_2$ that is $M^{-\epsilon}$-close to $\mathcal{A}$)*
2. *Either $\mathcal{A}_1=\emptyset$ or $\mathcal{A}_1$ is plus-friendly.*
3. *Either $\mathcal{A}_2=\emptyset$ or $\mathcal{A}_2$ is times-friendly.*

Proof of Claim. We prove the claim by repeatedly applying Lemma 3.6 to construct the sets A_1,A_2. We start with $\mathcal{A}_1=\emptyset$ and $\mathcal{A}_2=A$. We maintain the invariant that $\mathcal{A}_1$ is either empty or plus friendly.

If $|\mathcal{A}_2|\leq M^{1-\epsilon}$ then we're done (since we can then let $\mathcal{A}_2=\emptyset$ and then $|\mathcal{A}_1\cup\mathcal{A}_2|\geq M-M^{1-\epsilon}$). If $\mathcal{A}_2$ is times-friendly then we're done. Otherwise, there exists $\mathcal{B}''$ such that $\mathsf{cp}(\mathcal{A}_2\cdot\mathcal{B}'')\geq\frac{1}{M^{1+\epsilon}}$ and so we can apply Lemma 3.6 to obtain subsets $\mathcal{A}'$ of $\mathcal{A}_2$ and $\mathcal{B}'$ of $\mathcal{B}''$ such that $|\mathcal{A}'|,|\mathcal{B}'|\geq M^{1-10\epsilon}$ but $|\mathcal{A}'\cdot\mathcal{B}'|\leq M^{1+10\epsilon}$. By Claim 3.7, $\mathcal{A}'$ will be plus-friendly, so we can remove it from $\mathcal{A}_2$ and add it to $\mathcal{A}_1$ (i.e., let $\mathcal{A}_1=\mathcal{A}_1\cup\mathcal{A}'$, $\mathcal{A}_2=\mathcal{A}_2\setminus\mathcal{A}'$). Note that the union of disjoint plus friendly sets is plus-friendly (since a convex combination of low collision-probability distributions has low collision probability by Lemma 3.3). We continue in this way until either $\mathcal{A}_2$ is times-friendly or $|\mathcal{A}_2|\leq M^{1-\epsilon}$ (in which case we let $\mathcal{A}_2=\emptyset$). □

The distribution $\mathcal{A}\cdot\mathcal{B}+\mathcal{C}$ is within $M^{-\epsilon}$ statistical distance to the two distribution $\mathcal{X}_1$ and $\mathcal{X}_2$ where $\mathcal{X}_i=\mathcal{A}_i\cdot\mathcal{B}+\mathcal{C}$, and $\mathcal{A}_1$, $\mathcal{A}_2$ are the two sets obtained by Claim 3.9.1. We will finish the proof by showing that for both $i=1$ and $i=2$, if $\mathcal{A}_i\neq\emptyset$ then $\mathsf{cp}(\mathcal{A}_i\cdot\mathcal{B}+\mathcal{C})\leq\frac{1}{M^{1+\epsilon}}$.

Showing that $\mathsf{cp}(\mathcal{A}_1\cdot\mathcal{B}+\mathcal{C})\leq\frac{1}{M^{1+\epsilon}}$. Using Lemma 3.3, $\mathsf{cp}(\mathcal{A}_1\cdot\mathcal{B}+\mathcal{C})\leq\max_{b\in\mathbb{F}}\mathsf{cp}(\mathcal{A}_1 b+\mathcal{C})$. However $\mathsf{cp}(\mathcal{A}_1 b+\mathcal{C})=\mathsf{cp}(\mathcal{A}_1+\mathcal{C}b^{-1})$ since the latter distribution is a permutation of the former distribution (obtained by multiplying each element by b^{-1}). Yet the fact that $\mathcal{A}_1$ is plus-friendly implies that $\mathsf{cp}(\mathcal{A}_1+\mathcal{C}b^{-1})\leq\frac{1}{M^{1+\epsilon}}$.

Showing that $\mathsf{cp}(\mathcal{A}_2\cdot\mathcal{B}+\mathcal{C})\leq\frac{1}{M^{1+\epsilon}}$. This follows immediately from the fact that $\mathsf{cp}(\mathcal{A}_2\cdot\mathcal{B})\leq\frac{1}{M^{1+\epsilon}}$ and since $\mathsf{cp}(\mathcal{A}_2\cdot\mathcal{B}+\mathcal{C})\leq\mathsf{cp}(\mathcal{A}_2\cdot\mathcal{B})$ (as $\mathcal{A}_2\cdot\mathcal{B}+\mathcal{C}$ is a convex combination of distributions of the form $\mathcal{A}_2\cdot\mathcal{B}+c$ for some fixed $c\in\mathcal{C}$).

□

3.3.1 Finishing up

Lemma 3.9 almost directly implies Lemma 3.1: First we use the known fact that if the distributions $\mathcal{A}, \mathcal{B}, \mathcal{C}$ have min-entropy at least m, then the joint distribution $\mathcal{A}, \mathcal{B}, \mathcal{C}$ is a convex combination on independent distributions of the form $\mathcal{A}', \mathcal{B}', \mathcal{C}'$ where each of these is a uniform distribution on a set of size at least $M = 2^m$. Thus, it is sufficient to prove Lemma 3.9 for such distributions. By Lemma 3.9, for such distributions the distribution $\mathcal{A}' \cdot \mathcal{B}' + \mathcal{C}'$ is within $M^{-\epsilon}$ distance of having collision probability at most $\frac{1}{M^{1+\epsilon}}$ for some constant $\epsilon > 0$. Now, by applying the Lemma 3.4 (with $K = M^{1+\epsilon/2}$, $L = M^{\epsilon/2}$) we obtain that $\mathcal{A}' \cdot \mathcal{B}' + \mathcal{C}'$ is within statistical distance $M^{-\epsilon} + M^{-\epsilon/4}$ from having min-entropy at least $\log(M^{(1+\epsilon/2)}) = (1 + \epsilon/2)m$.

3.4 Constructing the field $\mathbb{F}$

Recall that our extractor obtains inputs in $\{0,1\}^n$, and needs to treat these inputs as elements of a field $\mathbb{F}$ of prime order. Unfortunately, there is no known deterministic polynomial-time algorithm that on input 1^n outputs an n-bit prime (without assuming a strong number-theoretic assumption about the distance between consecutive primes). Fortunately, in our setting we can still find such a prime. The reason is that we can use some of our samples from the high-entropy distribution to obtain such a prime. To do so, we will use the following result on *seeded* dispersers (which we state here with the parameters suitable for our purposes):

Theorem 3.10 ([Zuc91]). *For every $\delta > 0$, there exists a constant $d > 1$ and a polynomial-time computable function $D : \{0,1\}^{(10/\delta)n} \times \{0,1\}^{d \log n} \to \{0,1\}^n$ such that for every set $\mathcal{A} \subseteq \{0,1\}^{(10/\delta)n}$ with $|\mathcal{A}| \geq 2^{2n}$, it holds that $\left|D(\mathcal{A}, \{0,1\}^{d \log n}) \cap \{0,1\}^n\right| \geq (1 - \frac{1}{n^2}) \cdot 2^n$*

Let D be the function obtained from Theorem 3.10 and identify its output with a number in $[2^n]$. We say that $x \in \{0,1\}^{10/\delta n}$ is "bad" if $2^n + D(x,y)$ is *not* a prime number for *every* $y \in \{0,1\}^{d \log n}$. Because the set of primes has at least $1/n^2$ density in the interval $[2^n, 2^{2n}]$,the set of all bad $x \in \{0,1\}^{(10/\delta)n}$ is of size at most 2^{2n}. This means that if we take $10/\delta$ samples $x_1, \ldots, x_{10/\delta}$ from $10/\delta$ independent distributions over $\{0,1\}^n$ each of min-entropy at least δn, then the probability that the concatenation x of $x_1, \ldots, x_{10/\delta}$ is bad is exponentially low (at most $2^{2n}/2^{(10/\delta)\cdot\delta n} = 2^{-8n}$). This means that with $1 - 2^{\Omega(n)}$ probability if we enumerate over all seeds $y \in \{0,1\}^{d \log n}$ (which can be done in polynomial time), then we will find some y such that $2^n + D(x,y)$ is prime. Note that we can check primality in deterministic polynomial-time [AKS02].[17] Thus we can construct the field $\mathbb{F}$ with very high probability by using the first $10/\delta n$ samples and the function D to obtain a prime $N \in [2^n, 2^{n+1}]$. We then embed the set $\{0,1\}^n$ in the field $\mathbb{F} = \mathrm{GF}(N)$. Note that if since $n \geq \log N - 1$ there is no significant loss in relative entropy by this embedding.

[17] It is possible to use the same idea to run also a *probabilistic* primality testing algorithm using some additional samples from the high entropy sources.

4 A Constant-Samples Same-Source Disperser for Low Min-Entropy

In this section we prove Theorem 1.3. That is, we construct a constant-samples same-source disperser for subsets of $\{0,1\}^n$ of size $n^{O(1)}$. A central tool we will use is the *deterministic coin tossing* technique. This technique was used in several contexts in computer science (e.g., the parallel algorithm of [CV86]), and it was also used in a very similar context to ours by Erdos and Hajnal (where it was called the "stepping up lemma", see Section 4.7 in [GRS80]) and Fiat and Naor [FN93]. By "deterministic coin tossing" we mean the function $\mathsf{ct} : \{0,1\}^{2n} \to [n]$ defined as follows: for every $x, y \in \{0,1\}^n$, $\mathsf{ct}(x,y)$ is equal to the first position i such that $x_i \neq y_i$ (if $x = y$ then we let $\mathsf{ct}(x,y) = n$). The following property of this function will be useful for us:

Lemma 4.1. *For every $A \subseteq \{0,1\}^n$, $|\mathsf{ct}(A,A)| \geq \log |A|$*

Proof. Let $S \subseteq [n]$ denote $\mathsf{ct}(A,A)$ and suppose for the sake of contradiction that $|A| > 2^{|S|}$. Then, by the pigeonhole principle, there exist two distinct strings $x, y \in A$ such that x agrees with y on all the coordinates of S. However, for such x and y clearly $\mathsf{ct}(x,y) \notin S$. $\square$

The main idea behind our construction of a disperser is to apply the function ct in order to map our initial domain $\{0,1\}^n$ to the much smaller domain $[n]$, and then use brute force to find an optimal disperser (or even extractor) on this smaller domain. In fact, we will need to apply the function *twice* to reduce our domain to the domain $[\log n]$ so we can apply brute force and obtain an optimal disperser for the smaller domain. That is, we use the following simple lemma:

Lemma 4.2. *For every n, m (with $m < n$) it is possible to construct in $2^{O(n)}$ time the table for a function $E_{\mathsf{opt}} : [n] \times [n] \to [m]$ such that for every $A \subseteq [n]$ with $|A| > \log n + m + O(1)$, it holds that $|E_{\mathsf{opt}}(A,A)| \geq \frac{m}{2}$*

Proof Sketch: One way to do so will be to go over all possible functions until we find a function satisfying this condition (there will be such a function because a random function satisfies it with high probability). However, enumerating all possible functions will take take m^{n^2}-time. Note that testing this condition only requires going over all such subsets A which are at most $\min\{2^n, n^{2m \log n}\}$ many. Thus if we reduce the number of functions to test then we can reduce the overall time to a polynomial in this number. This can be done by considering functions from a sample space which is $\frac{1}{2m}$-close to being $2 \log n$-wise independent. Such a sample space can be constructed in time $2^{O(n)}$. □

Proof of Theorem 1.3. To prove Theorem 1.3, we consider the following disperser Disp: $\mathsf{Disp}(x_1, \dots, x_8) = E_{\mathsf{opt}}\left(\mathsf{ct}\big(\mathsf{ct}(x_1, x_2), \mathsf{ct}(x_3, x_4)\big), \mathsf{ct}\big(\mathsf{ct}(x_5, x_6), \mathsf{ct}(x_7, x_8)\big)\right)$

It clearly runs in polynomial time. We will prove that for every set A of size at least n^{5m}, $|\mathsf{Disp}(A, \dots, A)| \geq \frac{m}{2}$. Thus Disp is a disperser with "statistical distance" equal to $\frac{1}{2}$. Such a disperser can be modified to obtain a disperser according to our standard definition. For example, this can be done by embedding $[m]$ in some prime field $[p]$ (with $m \leq p \leq 2m$) and letting $\mathsf{Disp}'(x_1, \dots, x_{32}) = \mathsf{Disp}(x_1, \dots, x_8) + \mathsf{Disp}(x_9, \dots, x_{16}) + \mathsf{Disp}(x_1, \dots, x_{24}) + \mathsf{Disp}(x_{25}, \dots, x_{32})$. By the Cauchy-Davenport inequalities it will hold that for every set A as above, $\mathsf{Disp}'(A, \dots, A) = [p]$.

To prove the bound on A, we note that by Lemma 4.1, if $|A| \geq n^{5m}$ then $\mathsf{ct}(\mathsf{ct}(A, A), \mathsf{ct}(A, A)) \geq \log \log n + m + 5$. This means that $|E_{\mathsf{opt}}(A, A)| \geq \frac{m}{2}$, thus finishing our proof. □

Open Questions and Future Directions

Because dispersers and extractors have so many parameters, and our constructions is not optimal in many of them, there is much room for improvement in many directions. The main possible improvement to our results seems to be on the entropy front. That is, construct extractors and disperser that work for sources with lower entropy. Using this work, some improvements were made in this direction by [BKS+04], who constructed a ℓ-sample extractor for sources of entropy δn where δ can be any positive constant and ℓ is a universal constant independent of δ (in fact, for dispersers ℓ can be equal to 2). However, there is still no improvement for the case of significantly smaller entropy (i.e., for sources of entropy n^δ for some $\delta > 0$ or $\log^c n$ for some $c > 1$).

In this work we used some additive number-theoretic results that have not (to the best of our knowledge) been used before in Computer Science. It is quite likely that these results and techniques will find other Computer Science applications.

Acknowledgements. We thank Amir Shpilka for many valuable discusions during the early stages of this research.

References

[AKS02] M. Agrawal, N. Kayal, and N. Saxena. PRIMES is in P. Report, Department of Computer Science and Engineering, Indian Institute of Technology Kanpur, Kanpur-208016, India, Aug. 2002.

[BKS+04] B. Barak, G. Kindler, R. Shaltiel, B. Sudakov, and A. Wigderson. Simulating Independence: New Constructions of Condensers, Ramsey Graphs, Disersers, and Extractors, 2004. In Preparations.

[BST03] B. Barak, R. Shaltiel, and E. Tromer. True Random Number Generators Secure in a Changing Environment. In *Workshop on Cryptographic Hardware and Embedded Systems (CHES)*, pages 166–180, 2003. LNCS no. 2779.

[Blu84] M. Blum. Independent Unbiased Coin Flips From a Correlated Biased Source: A Finite State Markov Chain. In *Proc.* 25*th FOCS*, pages 425–433. IEEE, 1984.

[BKT03] J. Bourgain, N. Katz, and T. Tao. A sum-product estimate in finite fields, and applications. Arxiv technical report, `http://arxiv.org/abs/math.CO/0301343`, 2003. To appear in GAFA.

[CG85] B. Chor and O. Goldreich. Unbiased Bits from Sources of Weak Randomness and Probabilistic Communication Complexity. In *Proc.* 26*th FOCS*, pages 429–442. IEEE, 1985.

[CGH+85] B. Chor, O. Goldreich, J. Hastad, J. Friedman, S. Rudich, and R. Smolensky. The Bit Extraction Problem of t-Resilient Functions (Preliminary Version). In *Proc.* 26*th FOCS*, pages 396–407. IEEE, 1985.

[CW89] A. Cohen and A. Wigderson. Dispersers, Deterministic Amplification, and Weak Random Sources. In *Proc.* 30*th FOCS*, pages 14–19. IEEE, 1989.

[CV86] R. Cole and U. Vishkin. Deterministic coin tossing and accelerating cascades: micro and macro techniques for designing parallel algorithms. In *Proc.* 18*th STOC*, pages 206–219. ACM, 1986.

[DO03] Y. Dodis and R. Oliveira. On Extracting Private Randomness over a Public Channel. In *Proc. of RANDOM*. LNCS, 2003.

[DS02] Y. Dodis and J. Spencer. On the (non)Universality of the One-Time Pad. In *Proc.* 43*rd FOCS*, pages 376–388. IEEE, 2002.

[Elb03] A. Elbaz. Improved Constructions for Extracting Quasi-Random Bits from Sources of Weak Randomness. Master's thesis, Weizmann Institute of Science, 2003.

[FN93] A. Fiat and M. Naor. Implicit $O(1)$ probe search. *SIAM J. Comput.*, 22(1):1–10, Feb. 1993.

[Gow98] W. T. Gowers. A new proof of Szemerédi's theorem for arithmetic progressions of length four. *Geom. Funct. Anal.*, 8(3):529–551, 1998.

[GRS80] R. L. Graham, B. L. Rothschild, and J. L. Spencer. *Ramsey Theory*. John Wiley & Sons, 1980.

[HILL89] J. Håstad, R. Impagliazzo, L. A. Levin, and M. Luby. A pseudorandom generator from any one-way function. *SIAM J. Comput.*, 28(4):1364–1396, 1999. Preliminary versions appeared in STOC' 89 and STOC' 90.

[KZ03] Kamp and Zuckerman. Deterministic Extractors for Bit-Fixing Sources and Exposure-Resilient Cryptography. In *Proc.* 44*rd FOCS*. IEEE, 2003.

[Kon03] S. Konyagin. A sum-product estimate in fields of prime order. Arxiv technical report, `http://arxiv.org/abs/math.NT/0304217`, 2003.

[LRVW03] C.-J. Lu, O. Reingold, S. Vadhan, and A. Wigderson. Extractors: optimal up to constant factors. In *Proc.* 35*th STOC*, pages 602–611. ACM, 2003.

[MP90] J. L. McInnes and B. Pinkas. On the Impossibility of Private Key Cryptography with Weakly Random Keys. In *Crypto '90*, pages 421–436, 1990. LNCS No. 537.

[MU02] Mossel and Umans. On the Complexity of Approximating the VC Dimension. *J. Comput. Syst. Sci.*, 65, 2002.

[Nat96] M. B. Nathanson. *Additive number theory. Inverse Problems and the Geometry of Sumsets*, volume 165 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1996.

[Per92] Y. Peres. Iterating von Neumann's procedure for extracting random bits. *Ann. Statist.*, 20(1):590–597, 1992.

[PR04] P. Pudlak and V. Rodl. Pseudorandom sets and explicit constructions of Ramsey graphs, 2004. Submitted for publication.

[Ruz96] I. Z. Ruzsa. Sums of finite sets. In *Number theory (New York, 1991–1995)*, pages 281–293. Springer, New York, 1996.

[SRW03] M. Saks, O. Reingold, and A. Wigderson. Personal Communication, 2003.

[SV84] M. Santha and U. V. Vazirani. Generating Quasi-Random Sequences from Slightly-Random Sources. In *Proc.* 25*th FOCS*, pages 434–440. IEEE, 1984.

[Sha02] R. Shaltiel. Recent developments in extractors. *Bulletin of the European Association for Theoretical Computer Science*, 2002.

[TV00] L. Trevisan and S. Vadhan. Extracting randomness from samplable distributions. In *Proc.* 41*st FOCS*, pages 32–42. IEEE, 2000.

[Vaz85] U. Vazirani. Strong Communication Complexity or Generating Quasi-Random Sequences from Two Communicating Semi-Random Sources. *Combinatorica*, 7, 1987. Preliminary version in STOC' 85.

[Vaz87] U. V. Vazirani. Efficiency Considerations in Using Semi-random Sources. In *Proc.* 19*th STOC*, pages 160–168. ACM, 25–27 May 1987.

[vN51] J. von Neumann. Various Techniques Used in Connection with Random Digits. *Applied Math Series*, 12:36–38, 1951.

[Zuc90] D. Zuckerman. General Weak Random Sources. In *Proc.* 31*st FOCS*, pages 534–543. IEEE, 1990.

[Zuc91] D. Zuckerman. Simulating BPP Using a General Weak Random Source. *Algorithmica*, 16(4/5):367–391, Oct./Nov. 1996. Preliminary version in FOCS' 91.

A Proof Idea of Lemma 3.2

To prove Lemma 3.2 one needs to show that if $H^\infty(\mathcal{A}_i) \geq 0.9\log|\mathbb{F}|$ for $i = 1,\ldots,9$ then $\mathsf{dist}\left(\mathsf{Ext}^2(\mathcal{A}_1,\ldots,\mathcal{A}_9), U_\mathbb{F}\right) < |\mathbb{F}|^{-0.01}$. We prove it by showing the stronger statement $H^\infty(\mathcal{A}_1),\ldots,H^\infty(\mathcal{A}_4) > 0.9\log|\mathbb{F}|$ then $\mathsf{dist}\left((\mathcal{A}_1+\mathcal{A}_2)(\mathcal{A}_3+\mathcal{A}_4), U_\mathbb{F}\right) > |\mathbb{F}|^{0.01}$. This is indeed a stronger statement as Ext^2 essentially contains this expression combined with adding and multiplying of additional independent samples (which cannot increase the statistical distance from the uniform distribution). To give an idea of how this is proved, we are going to prove the following weaker lemma, where we assume that $\mathcal{A}_1,\ldots,\mathcal{A}_4$ are distributed identically and where we replace addition with substraction and multiplication with division:

Lemma A.1. *Let $\mathcal{A}$ be a random variable over $\mathbb{F}$ with $\mathsf{cp}(\mathcal{A}) < |\mathbb{F}|^{-0.9}$. Then $\mathsf{dist}\left((\mathcal{A}^{(1)}-\mathcal{A}^{(2)})/(\mathcal{A}^{(3)}-\mathcal{A}^{(4)}), U_\mathbb{F}\right) < 4|F|^{-0.8}$. (Where $\mathcal{A}^{(1)},\ldots,\mathcal{A}^{(4)}$ are independent random variables distributed according to $\mathcal{A}$.)*

Proof. Let s be an arbitrary non-zero element of $\mathbb{F}$. The distribution $\mathcal{A}^{(1)} + s\mathcal{A}^{(3)}$ is a random variable over $\mathbb{F}$ and hence has collision probability at least $\frac{1}{|\mathbb{F}|}$. Hence we see that $\Pr[(\mathcal{A}^{(1)} + s\mathcal{A}^{(3)}) = \mathcal{A}^{(2)} + s\mathcal{A}^{(4)})] \geq \frac{1}{|\mathbb{F}|}$. There are two ways that this collision can occur. If $\mathcal{A}^{(3)} = \mathcal{A}^{(4)}$ then this equation can only hold if $\mathcal{A}^{(1)} = \mathcal{A}^{(2)}$. Otherwise, it holds only if $s = (\mathcal{A}^{(1)} - \mathcal{A}^{(2)})/(\mathcal{A}^{(3)} - \mathcal{A}^{(4)})$ and hence we get that $\Pr[s = (\mathcal{A}^{(1)} - \mathcal{A}^{(2)})/(\mathcal{A}^3 - \mathcal{A}^4)] + \Pr[\mathcal{A}^{(1)} = \mathcal{A}^{(2)} \wedge \mathcal{A}^{(3)} = \mathcal{A}^{(4)}] \geq \frac{1}{|\mathbb{F}|}$. Thus, $\Pr[s = (\mathcal{A}^1 - \mathcal{A}^2)\cdot(\mathcal{A}^3 - \mathcal{A}^4)^{-1}] \geq \frac{1}{|\mathbb{F}|} - 1/\mathsf{cp}(\mathcal{A})^2$ for all $s \neq 0$, and so almost all field elements are at least close to the uniform probability. The distance between two distributions D_1 and D_2 can be expressed as $2\sum_{s|D_1(s)\geq D_2(s)}(D_1(s) - D_2(s))$. Taking D_1 as the uniform distribution on F, and D_2 as the above distribution, each s except 0 contributes at most $1/\mathsf{cp}(\mathcal{A})^2 \leq |\mathbb{F}|^{-1.8}$, and 0 contributes at most $1/|\mathbb{F}|$ to this sum. Hence,the statistical distance is at most $2|\mathbb{F}|/|\mathbb{F}|^{1.8} + 2/|\mathbb{F}| < 4|\mathbb{F}|^{-0.8}$. □

To move from division to multiplications and from subtractions to additions, we use the observations that for every two distributions $\mathcal{X},\mathcal{Y}$ over an Abelian group $\mathsf{cp}(\mathcal{X}+\mathcal{Y}) = \mathsf{cp}(\mathcal{X}-\mathcal{Y})$. To move to random variables $\mathcal{A}_1,\ldots,\mathcal{A}_4$ that are *not* identically distributed one needs to work a little harder. We defer the details to the full version of this paper.