Extracting Randomness Using Few Independent Sources

Boaz Barak* Russell Impagliazzo† Avi Wigderson‡

January 26, 2006

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 (Erdős and Hajnal, [GRS80]).

Keywords: Randomness extractors, Ramsey graphs, Sum-Product Theorem

*Department of Computer Science, Princeton University boaz@cs.princeton.edu. Work done as a postdoctoral member at the Institute for Advanced in Princeton, NJ and supported by NSF grants DMS-0111298 and CCR-0324906.

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

‡Institute for Advanced Study, Princeton, NJ. Email: avi@ias.edu . Partially supported by NSF grant CCR-0324906.

Contents

1 Introduction **2**
- 1.1 Background . . . 2
- 1.2 Seeded and Seedless Extractors . . . 2
- 1.3 Seedless extractors from few independent sources . . . 3
- 1.4 Our Results . . . 4
  - 1.4.1 Multiple-Sample Extractors. . . . 4
  - 1.4.2 Dispersers. . . . 5
- 1.5 Our Techniques . . . 6
- 1.6 How is this related to extractors? . . . 6

2 Preliminaries **8**

3 Constructing a Multiple-Sample Extractor **9**
- 3.1 Basic Facts and Notations . . . 10
- 3.2 Additive Number-Theoretic Results . . . 12
- 3.3 Proof of Lemma 3.1 . . . 12
  - 3.3.1 Finishing up . . . 14
- 3.4 Constructing the field $\mathbb{F}$ . . . 15
- 3.5 Decreasing the statistical distance. . . . 15
- 3.6 Proof of Lemma 3.2 . . . 16

4 A Constant-Samples Same-Source Disperser for Low Min-Entropy **18**

5 Subsequent and Future Work **19**

A A Proof of Theorem 1.4 **23**
- A.1 Proof of Claim A.2 . . . 24
- A.2 Proof of Claim A.3 . . . 24
  - A.2.1 More number theoretic lemmas. . . . 25
  - A.2.2 The actual proof. . . . 26

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. (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 and 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 indeed a minimal requirement from the input distribution.

We note that it is known that if a 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*. Because of this fact, when constructing randomness extractors typically we can assume without loss of generality that the input distribution is flat. This is convenient for several reasons, one of which is that the min-entropy of flat distributions is equal to their Shannon entropy.

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

[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.

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.[2] Seeded extractors have been studied extensively in the past two decades (see the survey [Sha02] and there references therein) 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 enumerate all secret keys produced using all seeds and then send the encryptions of a secret message using all these keys. 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.[3]

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 one such 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"[4] sources of length n and min-entropy δn for every constant $\delta > 0$.

[2] 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.

[3] 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.

[4] 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.

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. (We note that for this construction the $n/2$ entropy bound is tight - there are two sources of entropy exactly $n/2$ on which H is constant.) Vazirani [Vaz85] extended this to show that one can use a similar function to output from two indpendent sources of entropy $(\frac{1}{2}+\epsilon)n$, a linear number of bits which are exponentially close to the uniform distribution.[5]

Improving [CG85] seems hard. Even its disperser version, namely having a non-constant output for any two sources of entropy rate below 1/2, is the notorious "Bipartite Ramsey" problem, which 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+05] used the results of the current work to overcome this barrier and obtain such a disperser for entropy δn with $\delta > 0$ an arbitrarily small constant, see Section 5.)

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.[6] Assuming even more, Zuckerman [Zuc90] showed that if this conjecture holds for *all* multiplicative characters (not just χ_2), than a constant (actually $poly(1/\delta)$) number of sources suffice for extraction of linearly many bits with exponential error. The extractor we use here is *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*

$$\mathsf{dist}\Big(\mathsf{Ext}(\mathcal{X}_1, \ldots, \mathcal{X}_\ell), U_n\Big) < 2^{-\Omega(n)}$$

[5] Vazirani [Vaz85] states his result for the "semi-random" sources of [SV84] but it extends for general min-entropy sources.

[6] We note that it is known (using Weil's estimate of character sums) that P is an extractor for two sources where one of them has entropy $> n/2$ and the other one only entropy $> \log n$.

[7] We note that our results can be generalized to the case of extracting from a larger (i.e., super-constant) number of samples having sublinear entropy. However, we do not consider such extensions in this paper.

[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.

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,

$$\mathsf{dist}(\mathcal{X},\mathcal{Y}) \stackrel{def}{=} \frac{1}{2}\sum_{i\in\mathsf{Supp}(\mathcal{X})\cup\mathsf{Supp}(\mathcal{Y})}\left|\Pr[\mathcal{X}=i]-\Pr[\mathcal{Y}=i]\right| \quad .$$

As mentioned in the introduction, we denote by $H^\infty(\mathcal{X})$ the min-entropy of the random variable $\mathcal{X}$ (i.e., $H^\infty(\mathcal{X}) = \min_{x\in\mathsf{Supp}(\mathcal{X})}(-\log\Pr[\mathcal{X}=x])$).

Remark 1.2. As is noted in Section 3.5, we can actually reduce the statistical distance of our extractor by using more 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 $O(c)$. Note that if a distribution $\mathcal{Y}$ over $\{0,1\}^n$ is 2^{-cn}-close to the uniform distribution then in particular for every $y\in\{0,1\}^n$, $\Pr[\mathcal{Y}=y]\in(2^{-n}-2^{-cn},2^{-n}+2^{-cn})$ (i.e., $\mathcal{Y}$ is even close to the uniform distribution in L_∞ norm).

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 Erdős 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).

[9] Using c samples this lower-bound yields 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.

1.5 Our Techniques

An Erdős-Szemerédi 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 Erdős-Szemerédi [ES83] 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 Erdős-Szemerédi 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* 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 a field with no subfield of size between $|\mathbb{F}|^{\delta/2}$ and $|\mathbb{F}|-1$. Let $\mathcal{A} \subseteq \mathbb{F}$ be a set such that $|\mathbb{F}|^{\delta} < |\mathcal{A}| < |\mathbb{F}|^{1-\delta}$. 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 dependence of the constant ϵ on δ was not specified (and examining the proof sees that it is probably of the form $\epsilon = 2^{-\Omega(1/\delta)}$). In Appendix A we give a proof of Theorem 1.4 (using some of their lemmas) with $\epsilon = \Theta(\delta)$.[11] 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:

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

[11] The proof of Appendix A also has the advantage of proving the theorem for all fields $\mathbb{F}$ simultaneously. [BKT03] prove Theorem 1.4 for prime fields, and then only mention how it can be generalized to other fields.

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 $|\mathcal{A}| < |\mathbb{F}|^{0.99}$,*

$$|\mathcal{A} \cdot \mathcal{A} + \mathcal{A}| > |\mathcal{A}|^{1+\epsilon_0}$$

Indeed, by Theorem 1.5, for every set $\mathcal{A}$ either $|\mathcal{A} \cdot \mathcal{A}|$ or $|\mathcal{A} + \mathcal{A}|$ is at least $|\mathcal{A}|^{1+\epsilon'}$ for some absolute constant ϵ'. If the former holds then since $|\mathcal{A} \cdot \mathcal{A} + \mathcal{A}| \geq |\mathcal{A} \cdot \mathcal{A}|$ we're done. In the latter case we can use previously known number theoretic results (i.e. Lemma 3.7) that imply that if $|\mathcal{A} + \mathcal{A}| \geq |\mathcal{A}|^{1+\epsilon'}$ then $|\mathcal{A} + \mathcal{B}| \geq |\mathcal{A}|^{1+\epsilon'/2}$ for every set $\mathcal{B}$ with $|\mathcal{B}| = |\mathcal{A}|$. This means that if we let s be any member of $\mathcal{A}$ and consider $\mathcal{B} = s^{-1}\mathcal{A}$ then we get that $|\mathcal{A} + \mathcal{B}| \geq |\mathcal{A}|^{1+\epsilon'/2}$ (since multiplying by a fixed element is a permutation that does not change the size of a set). For the same reason, $|\mathcal{A} + \mathcal{B}| = |\mathcal{A} + s^{-1}\mathcal{A}| = |s\mathcal{A} + \mathcal{A}|$ but the set $s\mathcal{A} + \mathcal{A}$ is a subset of $\mathcal{A} \cdot \mathcal{A} + \mathcal{A}$.

In other words, Corollary 1.6 states that there is an efficiently computable $f(\cdot)$ such that $|f(\mathcal{A}, \mathcal{A}, \mathcal{A})| \geq |\mathcal{A}|^{1+\epsilon_0}$ (i.e., $f(a, b, c) = a \cdot b + c$). 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 $g : \mathbb{F}^{\ell} \to \mathbb{F}$ (for $\ell = 2^{O(\log(1/\delta))} = (1/\delta)^{O(1)}$) such that $g(\mathcal{A}, \ldots, \mathcal{A}) = \mathbb{F}$ (where $g(\mathcal{A}, \ldots, \mathcal{A})$ denotes the image of g on $\mathcal{A}^{\ell}$) .[12] If we identify the field $\mathbb{F}$ with the strings $\{0,1\}^n$, then we see that this function g (which is obviously polynomial-time computable) 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 $O(2^k)$.[13]

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 Ruzsa [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 $\delta \log |\mathbb{F}|$, the random variable $g(\mathcal{X}, \ldots, \mathcal{X})$ (where g is applied to ℓ independent copies of $\mathcal{X}$) 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 statement. Namely, that $g(\mathcal{X}_1, \ldots, \mathcal{X}_{\ell})$ is close to uniform for every independent random variables $\mathcal{X}_1, \ldots, \mathcal{X}_{\ell}$ with $H^{\infty}(\mathcal{X}_i) \geq \delta \log |\mathbb{F}|$ for $i = 1, \ldots, \ell$.)

[12] 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.

[13] 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.

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 an 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.
- 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.[14]

[14] 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 $> \epsilon|\mathbb{F}|$ with $k > 2/\epsilon$ then $\mathcal{A}_1 + \ldots + \mathcal{A}_k = \mathbb{F}$.

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 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 $x > y$.
- 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. 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. (Also, as mentioned above, 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 [GS71, Alo95] . However, in this paper we restrict ourselves to the *symmetric* case where all sources have the same entropy requirement.)
- 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*[RSW03].

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 $\mathsf{Ext}^0(x) = x$.

- 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 Section 3.6 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, $\mathsf{cp}(\mathcal{D}) = \Pr_{x,y \leftarrow_R \mathcal{D}}[x = y]$. 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. For $k \in \mathbb{Z}$ and $\mathcal{A}$ a set or a distribution we define $k\mathcal{A}$ and $\mathcal{A}^k$ in the natural way.[16]

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:

[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.

[16] E.g., for $k > 0$, $k\mathcal{A} = \underbrace{\mathcal{A} + \cdots + \mathcal{A}}_{k \text{ times}}$ and for $k < 0$, $k\mathcal{A} = \underbrace{-\mathcal{A} - \cdots - \mathcal{A}}_{|k| \text{ times}}$.

Lemma 3.3. *Let $\mathcal{X}$ and $\mathcal{Y}$ be distributions, then* $\Pr[\mathcal{X} = \mathcal{Y}] \leq \sqrt{\mathsf{cp}(\mathcal{X})\mathsf{cp}(\mathcal{Y})} \leq \max\{\mathsf{cp}(\mathcal{X}), \mathsf{cp}(\mathcal{Y})\}$.

Proof. For every i in the union of the supports of $\mathcal{X}$ and $\mathcal{Y}$, let x_i denote the probability that $\mathcal{X} = i$ and let y_i denote the probability that $\mathcal{Y} = i$. Then

$$\Pr[\mathcal{X} = \mathcal{Y}] = \sum_i x_i y_i \leq \sqrt{\sum_i x_i^2 \sum_j y_j^2}$$

by Cauchy-Schwartz, however this is the geometric mean of $\mathsf{cp}(\mathcal{X})$ and $\mathsf{cp}(\mathcal{Y})$ (which is less than the maximum of these quantities). □

Corollary 3.4. *Let $\mathcal{A}, \mathcal{C}$ be distributions over $\mathbb{F}$. Then* $\mathsf{cp}(\mathcal{A} + \mathcal{C}) \leq \sqrt{\mathsf{cp}(\mathcal{A} - \mathcal{A})\mathsf{cp}(\mathcal{C} - \mathcal{C})}$.

Proof. Let $\mathcal{A}'$ and $\mathcal{C}'$ be two new independent random variables distributed identically to $\mathcal{A}$ and $\mathcal{C}$ respectively. Then,

$$\mathsf{cp}(\mathcal{A} + \mathcal{C}) = Pr[\mathcal{A} + \mathcal{C} = \mathcal{A}' + \mathcal{C}'] = \Pr[\mathcal{A} - \mathcal{A}' = \mathcal{C} - \mathcal{C}']$$

which is smaller than $\sqrt{\mathsf{cp}(\mathcal{A} - \mathcal{A})\mathsf{cp}(\mathcal{C} - \mathcal{C})}$ by Lemma 3.3. □

Lemma 3.5. *Suppose that $\mathcal{X}$ is a convex combination of distributions $\mathcal{X}_1, \ldots, \mathcal{X}_m$. Then* $\mathsf{cp}(\mathcal{X}) \leq \max\{\mathsf{cp}(\mathcal{X}_1), \ldots, \mathsf{cp}(\mathcal{X}_m)\}$.

Proof. Using induction on m, this reduces to the case that $m = 2$ (since we can treat the combination of $\mathcal{X}_1, \ldots, \mathcal{X}_{m-1}$ as one distribution whose collision probability is bounded using induction). However in this case $\mathcal{X} = \alpha\mathcal{X}_1 + (1 - \alpha)\mathcal{X}_2$ and then

$$\mathsf{cp}(\mathcal{X}) = \alpha^2 \mathsf{cp}(\mathcal{X}_1) + 2\alpha(1 - \alpha)\Pr[\mathcal{X}_1 = \mathcal{X}_2] + (1 - \alpha)^2 \mathsf{cp}(\mathcal{X}_2)$$

However, $\Pr[\mathcal{X}_1 = \mathcal{X}_2] \leq \max\{\mathsf{cp}(\mathcal{X}_1), \mathsf{cp}(\mathcal{X}_2)\}$ by Lemma 3.3 and so we're done. □

Lemma 3.6. *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* $\log K$.

Proof. We can split $\mathcal{X}$ into a convex combination $\mathcal{X} = \alpha\mathcal{X}' + (1 - \alpha)\mathcal{X}''$, where $\mathcal{X}'$ contains the "big" elements of $\mathcal{X}$ that are obtained with probability at least $\frac{1}{K}$ and $\mathcal{X}''$ contains the rest of $\mathcal{X}$. We see that $\mathsf{cp}(\mathcal{X}) \geq \alpha^2 \mathsf{cp}(\mathcal{X}')$, and so $\frac{1}{KL} \geq \frac{\alpha^2}{K}$ and thus $\alpha \leq \frac{1}{\sqrt{L}}$. However, $H^\infty(\mathcal{X}'') \geq \log K$ and so we're done. □

Note that if $H^\infty(\mathcal{X}) \geq k$ then clearly $\mathsf{cp}(\mathcal{X}) \leq 2^{-k}$. Together with Lemma 3.6 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 $\log(1/\mathsf{cp}(\mathcal{X}))$. (The quantity $\log(1/\mathsf{cp}(\mathcal{X}))$ is sometimes called the 2-*entropy* or the *Renyi entropy* of $\mathcal{X}$.)

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 3.7 ([Ruz96] (see also [Nat96, Ruz89]))**.** *Let $\mathcal{A},\mathcal{B},\mathcal{C}$ be subsets of some Abelian group G. Then, $|\mathcal{A}+\mathcal{C}||\mathcal{B}| \leq |\mathcal{A}+\mathcal{B}||\mathcal{B}+\mathcal{C}|$.*

In particular, $|\mathcal{A}| = |\mathcal{B}| = M$ and $\rho > 0$ is such that $|\mathcal{A}+\mathcal{B}| \leq M^{1+\rho}$ then $|\mathcal{A}+\mathcal{A}| \leq \frac{|\mathcal{A}+\mathcal{B}|^2}{M} \leq M^{1+2\rho}$.

In other words, if $\mathcal{A}+\mathcal{B}$ is "small" for some $\mathcal{B}$ then $\mathcal{A}+\mathcal{A}$ is small. Note that Lemma 3.7 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 $\mathcal{A}\cdot\mathcal{B}$. Note that Lemma 3.7 also implies that if $|\mathcal{A}| = M$, $|\mathcal{B}| \geq M^{1-\epsilon}$ and $|\mathcal{A}+\mathcal{B}| \leq M^{1+\rho}$ then $|\mathcal{A}+\mathcal{A}| \leq M^{1+2\rho+\epsilon}$.

We'll use the following lemma of Gowers (which is a quantitative improvement of a result by Balog and Szemerédi). 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 $\mathcal{A} = \mathcal{B}$). See also Lemma A.5 for a generalization of this lemma obtained by [SSV04].[17]

Lemma 3.8 (Proposition 12, [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. We also note that we can apply the lemma even if $\mathcal{A}$ and $\mathcal{B}$ are of different sizes, as long as they are close enough. Indeed, if $|\mathcal{A}| = M$ and $|\mathcal{B}| = M^{1-\rho/10}$, then we can partition $\mathcal{A}$ to subsets $\mathcal{A}_1,\ldots,\mathcal{A}_k$ of size $|\mathcal{B}|$ and since $\mathcal{A}+\mathcal{B}$ is a convex combination of $\mathcal{A}_i+\mathcal{B}$, one of these subsets has collision probability as least as large as $\mathsf{cp}(\mathcal{A}+\mathcal{B})$, and we can apply the lemma to it.

3.3 Proof of Lemma 3.1

Fixing ϵ. We fix ϵ small enough such that for every $M < |\mathbb{F}|^{0.99}$, 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 $\epsilon = \epsilon_0/10^9$). **Note:** This is the only place in the proof we use the fact that $\mathbb{F}$ is a prime field. In particular, using Theorem 1.4 instead of Theorem 1.5, our proof yields also a variant of Lemma 3.1 for non-prime fields (see Lemma 3.14).

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 3.7, 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.

[17] We note that in some sources the lemma is stated with the condition being that the distribution $\mathcal{A}+\mathcal{B}$ is of statistical distance at least $M^{-C\rho}$ from having min-entropy $(1+\rho)\log M$ (a condition that up to constant factor in the distance is equivalent to the condition that there is a subset $\mathcal{C}$ of $\mathcal{A}\times\mathcal{B}$ such that $|\mathcal{C}| \leq M^{1+\rho/(2C)}$, but $\Pr[\mathcal{A}+\mathcal{B}\in\mathcal{C}] \geq M^{-C\rho}$). However this is equivalent to our statement below by the observations above on the relation between min-entropy and collision probability. In particular, if there is such a set $\mathcal{C}$ the the collision probability of $\mathcal{A}+\mathcal{B}$ is at least $\Pr[\mathcal{A}+\mathcal{B}\in\mathcal{C}]|\mathcal{C}|^{-1}$.

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 $\Omega(1/M)$.[18]

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 3.9. *Let M and ϵ be as above and let $\mathcal{A} \subseteq \mathbb{F}$ be a set such that $|\mathcal{A}| \geq M^{1-20\epsilon}$ but $|\mathcal{A} \cdot \mathcal{B}| \leq M^{1+20\epsilon}$ for some $\mathcal{B}$ with $|\mathcal{B}| \geq M^{1-20\epsilon}$. Then $\mathsf{cp}(\mathcal{A} + \mathcal{C})$ is smaller than $\frac{1}{M^{1+20\epsilon}}$ for* all *sets $\mathcal{C}$ of size M.*

Proof. If $\mathsf{cp}(\mathcal{A} + \mathcal{C}) \geq \frac{1}{M^{1+20\epsilon}}$ then by applying Lemma 3.8 (with $\rho = 60\epsilon$, and assuming $\epsilon < 1/10$) it holds that there exists subsets $\mathcal{A}', \mathcal{C}'$ of $\mathcal{A}$ and $\mathcal{C}$ respectively such that $|\mathcal{A}'|, |\mathcal{C}'| \geq M^{1-600\epsilon}$ but $|\mathcal{A}' + \mathcal{C}'| \leq M^{1+600\epsilon}$. This means by Lemma 3.7 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 3.7. 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 3.9 $(M, 20\epsilon)$-*plus-friendly*. That is, $\mathcal{A}$ is said to be (M, ϵ')-*plus-friendly* if $\mathsf{cp}(()\mathcal{A} + \mathcal{C}) \leq M^{-1-\epsilon'}$ for every $\mathcal{C}$ with $|\mathcal{C}| = M$. Since M and ϵ are fixed for this proof, we'll sometimes drop the prefix $(M, 20\epsilon)$, and simply call $(M, 20\epsilon)$-plus-friendly sets *plus-friendly*. Reversing the roles of addition and multiplication we obtain

Lemma 3.10. *Let M and ϵ be as above and let $\mathcal{A} \subseteq \mathbb{F}$ be a set such that $|\mathcal{A}| \geq M^{1-20\epsilon}$ but $|\mathcal{A} + \mathcal{B}| \leq M^{1+20\epsilon}$ for some $\mathcal{B}$ with $|\mathcal{B}| \geq M^{1-20\epsilon}$. Then $\mathsf{cp}(\mathcal{A} \cdot \mathcal{C})$ is smaller than $\frac{1}{M^{1+20\epsilon}}$ for* all *sets $\mathcal{C}$ of size M.*

Again, we call a set $\mathcal{A}$ satisfying the conclusion of Lemma 3.10 $(M, 20\epsilon)$-*times-friendly* and again in the following we'll sometimes simply call such sets *times-friendly*. The main step in our proof will be the following lemma:

Lemma 3.11. *Let $\mathcal{A} \subseteq \mathbb{F}$ with $|\mathcal{A}| = M$. Then, there exist two disjoint subsets $\mathcal{A}_+$ and $\mathcal{A}_\times$ such that*

- *$\mathcal{A}_+$ is either empty or (M, ϵ)-plus-friendly.*
- *$\mathcal{A}_\times$ is either empty or (M, ϵ)-times-friendly.*
- *$|\mathcal{A} \setminus (\mathcal{A}_+ \cup \mathcal{A}_\times)| < M^{1-\epsilon}$. (I.e., $\mathcal{A}_+ \cup \mathcal{A}_\times$ capture almost all of of $\mathcal{A}$.)*

Note that this lemma implies that the counterexample described above (of a distribution that is an arithmetic progression with probability half and a geometric progression with probability half) captures in some sense the worst case for the theorem.

[18] 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 $\mathsf{cp}(\mathcal{A} \cdot \mathcal{A}) \geq \Omega(1/M)$.

Proof of Lemma 3.11. We prove the lemma by repeatedly applying Lemma 3.8 to construct the sets $\mathcal{A}_+, \mathcal{A}_\times$. We start with $\mathcal{A}_+ = \emptyset$ and $\mathcal{A}_\times = \mathcal{A}$. At each point we remove some elements from $\mathcal{A}_\times$ and add them to $\mathcal{A}_+$. We always maintain the invariant that $\mathcal{A}_+$ is either empty or plus-friendly.

If $|\mathcal{A}_\times| < M^{1-\epsilon}$ then we're done (since we can then let $\mathcal{A}_\times = \emptyset$ and have $|\mathcal{A}_\times \cup \mathcal{A}_+| \geq M - M^{1-\epsilon}$). If $\mathcal{A}_\times$ is (M,ϵ)-times-friendly then we're done. Otherwise, there exists $\mathcal{B}''$ of size M such that $\mathsf{cp}(\mathcal{A}_\times \cdot \mathcal{B}'') \geq \frac{1}{M^{1+\epsilon}}$ and so we can apply Lemma 3.8 to obtain subsets $\mathcal{A}'$ of $\mathcal{A}_\times$ and $\mathcal{B}'$ of $\mathcal{B}''$ such that $|\mathcal{A}'|, |\mathcal{B}'| \geq M^{1-5\epsilon}$ but $|\mathcal{A}' \cdot \mathcal{B}'| \leq M^{1+5\epsilon}$. By Claim 3.9, $\mathcal{A}'$ will be $(M, 10\epsilon)$ (and so in particular (M,ϵ)) plus-friendly, and so we can remove it from $\mathcal{A}_\times$ and add it to $\mathcal{A}_1$ (i.e., let $\mathcal{A}_+ = \mathcal{A}_+ \cup \mathcal{A}'$, $\mathcal{A}_\times = \mathcal{A}_\times \setminus \mathcal{A}'$). Note that the union of disjoint plus friendly sets is plus-friendly with the same parameters (since a convex combination of low collision-probability distributions has low collision probability by Lemma 3.5). We continue in this way until either $\mathcal{A}_\times$ is (M,ϵ)-times-friendly or $|\mathcal{A}_\times| \leq M^{1-\epsilon}$.

□

Using Lemma 3.11, we can obtain a collision-probability analog of Corollary 1.6:

Lemma 3.12. *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. We split $\mathcal{A}$ into $\mathcal{A}_+$ and $\mathcal{A}_\times$ as per Lemma 3.11. The distribution $\mathcal{A} \cdot \mathcal{B} + \mathcal{C}$ is within $M^{-\epsilon}$ statistical distance to a convex combination of the distribution $\mathcal{X}_+ = \mathcal{A}_+ \cdot \mathcal{B} + \mathcal{C}$ and the distribution of $\mathcal{X}_\times = \mathcal{A}_\times \cdot \mathcal{B} + \mathcal{C}$. (Unless $\mathcal{A}_+$ or $\mathcal{A}_\times$ is empty, in which case $\mathcal{A} \cdot \mathcal{B} + \mathcal{C}$ is within $M^{-\epsilon}$ statistical distance to one of these distributions.) We will finish the proof by showing that for both distributions $\mathcal{X}_+$ and $\mathcal{X}_\times$, if the corresponding set is not empty, then the collision probability is at most $\frac{1}{M^{1+\epsilon}}$.

Showing that $\mathsf{cp}(\mathcal{A}_+ \cdot \mathcal{B} + \mathcal{C}) \leq \frac{1}{M^{1+\epsilon}}$. Using Lemma 3.5, $\mathsf{cp}(\mathcal{A}_+ \cdot \mathcal{B} + \mathcal{C}) \leq \max_{b \in \mathbb{F}} \mathsf{cp}(\mathcal{A}_+ b + \mathcal{C})$. However $\mathsf{cp}(\mathcal{A}_+ b + \mathcal{C}) = \mathsf{cp}(\mathcal{A}_+ + \mathcal{C}b^{-1})$ since the latter distribution is a permutation of the former distribution (obtained by multiplying each element by b^{-1}). Yet the fact that $\mathcal{A}_+$ is (M,ϵ)-plus-friendly implies that $\mathsf{cp}(\mathcal{A}_+ + \mathcal{C}b^{-1}) \leq \frac{1}{M^{1+\epsilon}}$.

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

□

3.3.1 Finishing up

Lemma 3.12 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 3.12 for such distributions. By Lemma 3.12, 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 3.6 (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 3.13 ([Zuc91])**.** *For every $\delta > 0$, there exists a constant $d > 1$ and a polynomial-time computable function $\mathsf{D} : \{0,1\}^{(10/\delta)n} \times \{0,1\}^{d \log n} \to \{0,1\}^n$ such that for every set $\mathcal{A} \subseteq \{0,1\}^{(10/\delta)n}$ with $|\mathcal{A}| \geq 2^{2n}$, it holds that*

$$\left|\mathsf{D}(\mathcal{A}, \{0,1\}^{d \log n}) \cap \{0,1\}^n\right| \geq (1 - \tfrac{1}{n^2}) \cdot 2^n$$

Let D be the function obtained from Theorem 3.13 and identify its output with a number in $[2^n]$. We say that $x \in \{0,1\}^{(10/\delta)n}$ is "bad" if $2^n + \mathsf{D}(x,y)$ is *not* a prime number for *every* $y \in \{0,1\}^{d \log n}$. Because the set of primes has more than $1/n^2$ density in the interval $[2^n, 2 \cdot 2^n]$,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 $2^n + \mathsf{D}(x,y)$ is prime. Note that we can check primality in deterministic polynomial-time [AKS02].[19] 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 $P \in [2^n, 2^{n+1}]$. We then embed the set $\{0,1\}^n$ in the field $\mathbb{F} = \mathrm{GF}(P)$. Note that since $n \geq \log P - 1$ there is no significant loss in relative entropy by this embedding.

Using non-prime fields. A different approach to solve the problem of obtaining the field $\mathbb{F}$ is to use a *non-prime* field such as $\mathrm{GF}(2^p)$ (where p now is a *small* prime and hence can be easily found) and then use Theorem 1.4 (the version proved in Appendix A) instead of Theorem 1.5. This would yield the following variant of Lemma 3.1:

Lemma 3.14. *There exists an absolute constant $c > 0$ such that for every prime p every $\delta > 0$ and every distributions $\mathcal{A}, \mathcal{B}, \mathcal{C}$ over $\mathrm{GF}(2^p)$ with $H^\infty(\mathcal{A}), H^\infty(\mathcal{B}), H^\infty(\mathcal{C}) > \delta p$, the distribution $\mathcal{A} \cdot \mathcal{B} + \mathcal{C}$ is $2^{-\epsilon m}$-close to having min-entropy at least $\min\{(1+\epsilon)m, 0.9 \log |\mathbb{F}|\}$ where $\epsilon = c\delta$.*

3.5 Decreasing the statistical distance.

We now show how we can decrease the statistical distance by repetition. We use the following variant of a XOR-lemma:

Lemma 3.15. *Let $\mathcal{Y}_1, \ldots, \mathcal{Y}_t$ be independent distributions over $\mathbb{F}$ such that $\mathsf{dist}(\mathcal{Y}_i, U_\mathbb{F}) < \epsilon$ for every $i = 1, \ldots, t$. Then*

$$\mathsf{dist}\left(\mathcal{Y}_1 + \cdots + \mathcal{Y}_t, U_\mathbb{F}\right) \leq \epsilon^t$$

[19] 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.

Proof. We can represent each distribution $\mathcal{Y}_i$ as a convex combination of the following form: with probability $(1-\epsilon)$ we get U_i (where each U_i is an independent copy of the uniform distribution) and with probability ϵ an element of some distribution $\tilde{\mathcal{Y}}_i$. Thus, one can think of the distribution $\mathcal{Y}_1 + \ldots + \mathcal{Y}_t$ as a convex combination where with probability ϵ^t we get an element of the form $\tilde{\mathcal{Y}}_1 + \cdots + \tilde{\mathcal{Y}}_t$ and with probability $1-\epsilon^t$ we get a distribution of the form $\tilde{Y} + U_{\mathbb{F}}$ for some distribution $\tilde{Y}$ which is independent of $U_{\mathbb{F}}$. In other words w.p. $-\epsilon^t$ we get the uniform distribution. □

3.6 Proof of Lemma 3.2

Before proving Lemma 3.2, we'll prove the following related lemma:

Lemma 3.16. *Let $\mathbb{F}$ be any field of size N, and let $\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D}$ be four independent random variables over $\mathbb{F}$, where each variable has collision probability at most $1/M$ for some $M > N^{3/4}$ Then,*

$$\mathsf{dist}\left((\mathcal{A}-\mathcal{C})\cdot(\mathcal{B}-\mathcal{D}), U_{\mathbb{F}}\right) < O(N^{3/2}/M^2)$$

where $U_{\mathbb{F}}$ denotes the uniform distribution over $\mathbb{F}$.

We will start with a variant, where we divide rather than multiply, and where $A = C$ and $B = D$. (Throughout this section, we will always use N to denote the size of the field $\mathbb{F}$.)

Lemma 3.17. *Let $\mathcal{A}_1, \mathcal{A}_2$ be identical independent random variables over $\mathbb{F}$ with collision probability at most $1/M$ for some $M > N^{1/2}$ and let $\mathcal{B}_1, \mathcal{B}_2$ be likewise. Then,*

$$\mathsf{dist}\left((\mathcal{A}_1-\mathcal{A}_2)\cdot(\mathcal{B}_1-\mathcal{B}_2)^{-1}, U_{\mathbb{F}}\right) \leq O(N/M^2)$$

(We can ignore the event that $\mathcal{B}_1 = \mathcal{B}_2$ and we need to divide by zero, since it happens with probability at most $1/M$. Thus, no matter how the value of $x/0$ is defined, it won't contribute more than $1/M$ to the statistical distance.)

Proof. Let s be an arbitrary non-zero element of $\mathbb{F}$. The distribution $\mathcal{A}_1 + s\mathcal{B}_1$ is a random variable over $\mathbb{F}$ and hence has collision probability at least $\frac{1}{N}$. Hence we see that

$$\Pr[(\mathcal{A}_1 + s\mathcal{B}_1) = \mathcal{A}_2 + s\mathcal{B}_2)] \geq \tfrac{1}{N}$$

There are two ways that this equality can occur. If $\mathcal{B}_1 = \mathcal{B}_2$ then this equation can only hold if $\mathcal{A}_1 = \mathcal{A}_2$. Otherwise, it holds only if $s = (\mathcal{A}_1 - \mathcal{A}_2)\cdot(\mathcal{B}_1 - \mathcal{B}_2)^{-1}$ and hence we get that

$$\Pr[s = (\mathcal{A}_1 - \mathcal{A}_2)\cdot(\mathcal{B}_1 - \mathcal{B}_2)^{-1}] + \Pr[\mathcal{A}_{(1)} = \mathcal{A}_{(2)} \wedge \mathcal{B}_{(1)} = \mathcal{B}_{(2)}] \geq \tfrac{1}{N}$$

So $\Pr[s = (\mathcal{A}_1 - \mathcal{A}_2)\cdot(\mathcal{B}_1 - \mathcal{B}_2)^{-1}] \geq \frac{1}{N} - 1/M^2$ for all $s \neq 0$, 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 $\mathbb{F}$, and D_2 as the above distribution, each s except 0 contributes at most $1/M^2$, and 0 contributes at most $1/N$ to this sum. Hence,the statistical distance is at most $2N/M^2 + 2/N = O(N/M^2)$. □

Lemma 3.18. *Let X and Y be independent random variables on a set S of size N. Then $\mathsf{cp}(X) + \mathsf{cp}(Y) + 2\Pr[X = Y] \geq 4/N$.*

Proof. Let X be distributed according to D_0 and Y according to D_1. Then let $D_{1/2} = 1/2D_0 + 1/2D_1$. Then $D_{1/2}$ is a probability distribution on S, so $\mathsf{cp}(D_{1/2}) \geq 1/N$. On the other hand, $\mathsf{cp}(D_{1/2}) = \sum_{s\in S}(1/2D_0(s)+1/2D_1(s))^2 = 1/4(\sum_{s\in S} D_0(s)^2+2\sum_{s\in S} D_0(s)D_1(s)+\sum_{s\in S} D_1(s)^2) = 1/4(\mathsf{cp}(X) + 2\Pr[X = Y] + \mathsf{cp}(Y)$. □

Lemma 3.19. *Let $\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D}$ be four independent variables over $\mathbb{F}$ each with collision probability at most $1/M$ where $M > N^{1/2}$. Let $\mathcal{A}_1, \mathcal{A}_2$ be two independent variables distributed the same as $\mathcal{A}$, and likewise for $\mathcal{B}_1, \mathcal{B}_2, \mathcal{C}_1, \mathcal{C}_2, \mathcal{D}_1, \mathcal{D}_2$. Then for any $s \in \mathbb{F}$, $s \neq 0$, $\Pr[s = (\mathcal{A}_1 - \mathcal{A}_2)(\mathcal{B}_2 - \mathcal{B}_1)^{-1}] + \Pr[s = (\mathcal{C}_1 - \mathcal{C}_2)(\mathcal{D}_2 - \mathcal{D}_1)^{-1}] + 2\Pr[s = (\mathcal{A} - \mathcal{C})(\mathcal{D} - \mathcal{B})^{-1}] \geq 4/N - O(1/M^2)$.*

Proof. Let X be $\mathcal{A} + s\mathcal{B}$ and Y be $\mathcal{C} + s\mathcal{D}$. Note that, by a similar case analysis to the proof of Lemma 3.17, $\mathsf{cp}(X) \leq \Pr[s = (\mathcal{A}_1 - \mathcal{A}_2)(\mathcal{B}_2 - \mathcal{B}_1)^{-1}] + 1/M^2$, $\mathsf{cp}(Y) \leq \Pr[s = (\mathcal{C}_1 - \mathcal{C}_2)(\mathcal{D}_2 - \mathcal{D}_1)^{-1}] + 1/M^2$ and $\Pr[X = Y] \leq \Pr[s = (\mathcal{A} - \mathcal{C})(\mathcal{D} - \mathcal{B})^{-1}] + 1/M^2$ (using Lemma 3.3 to bound the probabilities that $\mathcal{A} = \mathcal{C}$ and $\mathcal{B} = \mathcal{D}$). The claim then follows from Lemma 3.18. □

Lemma 3.20. *Let $\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D}$ be four independent random variables over $\mathbb{F}$ each with collision probability at most $1/M$ where $M > N^{1/2}$. Then,*

$$\mathsf{dist}\left((\mathcal{A} - \mathcal{C}) \cdot (\mathcal{B} - \mathcal{D})^{-1}, U_{\mathbb{F}}\right) < O(N/M^2)$$

where $U_{\mathbb{F}}$ denotes the uniform distribution over $\mathbb{F}$.

Proof. The statistical distance between $(\mathcal{A} - \mathcal{C})(\mathcal{D} - \mathcal{B})^{-1}$ and $U_{\mathbb{F}}$ can be computed as the sum of $\Pr[U_{\mathbb{F}} = s] - \Pr[(\mathcal{A} - \mathcal{C})(\mathcal{D} - \mathcal{B})^{-1} = s]$ for all $s \in \mathbb{F}$ where this difference is positive. Since $\Pr[U_{\mathbb{F}} = s] = 1/N$, using Lemma 3.19, we see that for all such $s \neq 0$ (the case $s = 0$ can add at most $1/M$ to the distance) it is the case that

$$\begin{aligned}
&\frac{1}{N} - \Pr[s = (\mathcal{A} - \mathcal{C})(\mathcal{D} - \mathcal{B})^{-1}] \leq \\
&\quad \frac{1}{2}\Pr[s = (\mathcal{A}_1 - \mathcal{A}_2)(\mathcal{B}_2 - \mathcal{B}_1)^{-1}] - \frac{1}{N} - \frac{1}{2}\Pr[s = (\mathcal{C}_1 - \mathcal{C}_2)(\mathcal{D}_2 - \mathcal{D}_1)^{-1}] + O(1/M^2) \leq \\
&\quad\quad \frac{1}{2}\Big|\Pr[s = (\mathcal{A}_1 - \mathcal{A}_2)(\mathcal{B}_2 - \mathcal{B}_1)^{-1}] - \frac{1}{N}\Big| + \frac{1}{2}\Big|\frac{1}{2}\Pr[s = (\mathcal{C}_1 - \mathcal{C}_2)(\mathcal{D}_2 - \mathcal{D}_1)^{-1}]\Big| + O(1/M^2)
\end{aligned}$$

Summing this over all such s, we get that

$$\begin{aligned}
&\mathsf{dist}\left((\mathcal{A} - \mathcal{C}) \cdot (\mathcal{B} - \mathcal{D})^{-1}, U_{\mathbb{F}}\right) \leq \\
&\qquad \tfrac{1}{2}\mathsf{dist}\left((\mathcal{A}_1 - \mathcal{A}_2) \cdot (\mathcal{B}_2 - \mathcal{B}_1)^{-1}, U_{\mathbb{F}}\right) + \tfrac{1}{2}\mathsf{dist}\left((\mathcal{C}_1 - \mathcal{C}_2) \cdot (\mathcal{D}_2 - \mathcal{D}_1)^{-1}, U_{\mathbb{F}}\right) + O(N/M^2)
\end{aligned}$$

which is at most $O(N/M^2)$ by Lemma 3.17. □

To finish the proof we also use the following simple observation about the relation between L_2 and L_1 norm, or in our notations, between the statistical distance of a random variable from uniform and the difference between the collision probability of this random variable and $1/N$.

Lemma 3.21. *Let $\mathcal{Z}$ be a random variable over $\mathbb{F}$. Let $\delta = \mathsf{dist}(\mathcal{Z}, U_{\mathbb{F}})$. Then $1/N + \delta^2/N \leq \mathsf{cp}(\mathcal{Z}) \leq 1/N + \delta^2$.*

Proof. Let $\delta_s = |\Pr[\mathcal{Z} = s] - 1/N|$. Then $\sum_{s\in\mathbb{F}} \delta_s = \delta$, and $\sum_{s\in\mathbb{F}} \delta_s^2 = \sum_s (\Pr[\mathcal{Z} = s])^2 - 2/N \sum_s [\Pr[\mathcal{Z} = s] + N/N^2 = \mathsf{cp}(\mathcal{Z}) - 1/N$. By convexity, $\sum \delta_s^2$ is minimized when all $\delta_s = \delta/N$ and maximized when one $\delta_s = \delta$. Thus, $\delta^2/N \leq \mathsf{cp}(\mathcal{Z}) - 1/N \leq \delta^2$. □

To obtain Lemma 3.16, we note that for every two distributions $\mathcal{X}, \mathcal{Y}$ over $\mathbb{F} \setminus \{0\}$, $\mathsf{cp}(\mathcal{X} \cdot \mathcal{Y}) = \mathsf{cp}(\frac{\mathcal{X}}{\mathcal{Y}})$. Indeed, $x \cdot y = x' \cdot y'$ if and only if $\frac{x}{y'} = \frac{x'}{y}$. By Lemma 3.20, $\mathsf{dist}((\mathcal{A}-\mathcal{C})(\mathcal{D}-\mathcal{B})^{-1}, U_{\mathbb{F}}) = O(N/M^2)$. Thus, from the previous lemma, $\mathsf{cp}((\mathcal{A}-\mathcal{C})(\mathcal{D}-\mathcal{B})) = \mathsf{cp}((\mathcal{A}-\mathcal{C})(\mathcal{D}-\mathcal{B})^{-1}) \leq 1/N + O(N^2/M^4)$. Then it follows from the lemma above that $\mathsf{dist}((\mathcal{A}-\mathcal{C})(\mathcal{D}-\mathcal{B}),\mathcal{U})^2/N \leq O(N^2/M^4)$, so $\mathsf{dist}((\mathcal{A}-\mathcal{C})(\mathcal{D}-\mathcal{B}),\mathcal{U}) \leq O(N^{3/2}/M^2)$. □

Proving Lemma 3.2. We can now prove Lemma 3.2. Indeed, for every 9 distributions $\mathcal{X}_1, \ldots, sX_9$, $\mathsf{cp}(\mathsf{Ext}^2(\mathcal{X}_1, \ldots, \mathcal{X}_9)) \leq \mathsf{cp}(\mathsf{Ext}^1(\mathcal{X}_1, \mathcal{X}_2, \mathcal{X}_3) \cdot \mathsf{Ext}^1(\mathcal{X}_4, \mathcal{X}_5, \mathcal{X}_6))$, since adding the additional independent distribution $\mathsf{Ext}^1(\mathcal{X}_7, \mathcal{X}_8, \mathcal{X}_9)$ cannot increase the collision probability. Hence, it is enough to prove that for $\mathcal{X}_1, \ldots, \mathcal{X}_6$ of min-entropy $0.9 \log N$, the distribution $(\mathcal{X}_1 \cdot \mathcal{X}_2 + \mathcal{X}_3) \cdot (\mathcal{X}_4 \cdot \mathcal{X}_5 + \mathcal{X}_6)$ is $N^{-0.1}$ close to the uniform distribution over $\mathbb{F}$. Yet this distribution is a convex combination of distributions of the form $(\mathcal{X}_1 \cdot x_2 + \mathcal{X}_3) \cdot (\mathcal{X}_4 \cdot x_5 + \mathcal{X}_6)$ for fixed (and with very high probability non-zero) $x_2, x_5 \in \mathbb{F}$. Any such distribution is of the form $(\mathcal{A}-\mathcal{B})(\mathcal{C}-\mathcal{D})$ (for $\mathcal{A} = \mathcal{X}_1 \cdot x_2$, $\mathcal{B} = -\mathcal{X}_3$, $\mathcal{C} = \mathcal{X}_4 \cdot x_5$, $\mathcal{D} = -\mathcal{X}_6$) and hence the result is implied by Lemma 3.16. □

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 Erdős 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 pigeon-hole 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 $\mathsf{ct}(x,y) \notin S$. □

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 constant m every 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 $|A| > \log n + 10m$, it holds that $|E_{\mathsf{opt}}(A,A)| \geq \frac{m}{2}$*

Proof. 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 $\min\{2^n, n^{10m+\log n}\}$ many.

Thus if we reduce the number of functions to test to something smaller than this number then we can reduce the overall time to $2^{O(n)}$. This can be done by considering functions from a sample space over $[m]^{n^2}$ which is 2^{-n}-close to being $(\log n + 10m)^2$-wise independent. There are explicit construction for such sample spaces with $2^{O(n)}$ many points [NN93].

□

Proof of Theorem 1.3. To prove Theorem 1.3, we consider the following disperser Disp:

$$\mathsf{Disp}(x_1,\ldots,x_8) = E_{\mathsf{opt}}\left(\mathsf{ct}\big(\mathsf{ct}(x_1,x_2),\mathsf{ct}(x_3,x_4)\big),\mathsf{ct}\big(\mathsf{ct}(x_5,x_6),\mathsf{ct}(x_7,x_8)\big)\right)$$

It clearly runs in polynomial time. We will prove that for every set A of size at least n^{d2^m}, $|\mathsf{Disp}(A,\ldots,A)| \geq \frac{m}{2}$. 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 $\mathsf{Disp}'(x_1,\ldots,x_{32}) = \mathsf{Disp}(x_1,\ldots,x_8)+\mathsf{Disp}(x_9,\ldots,x_{16})+\mathsf{Disp}(x_1,\ldots,x_{24})+\mathsf{Disp}(x_{25},\ldots,x_{32})$. By the Cauchy-Davenport inequalities it will hold that for every set A as above, $\mathsf{Disp}'(A,\ldots,A) = [p]$.

To prove the bound on A, we note that by Lemma 4.1, if $|A| \geq n^{d2^m}$ then $\mathsf{ct}(\mathsf{ct}(A,A),\mathsf{ct}(A,A)) \geq \log\log n + m + \log d$. Since we apply E_{opt} on a domain of size $\log n$, by Lemma 4.2, this means that for d large enough, $|E_{\mathsf{opt}}(A,A)| \geq \frac{m}{2}$, thus finishing our proof. □

5 Subsequent and Future Work

In this work we have given the first extractors for a constant number of independent sources, each of linear entropy. Unfortunately the number of sources needed is a function of the (constant) entropy rate, and the most obvious next challenge was to remove this dependency.

In very recent work, [BKS+05] have achieved this. They give explicit deterministic extractors from 3 independent sources of any linear entropy. This was further improved by Ran Raz [Raz05], who only needs one of the sources to have linear entropy, while the others can have only logarithmic amount of entropy.

Further results in [BKS+05] give a deterministic disperser from only two independent sources of constant entropy rate (greatly improving bounds on the Bipartite Ramsey problem). They also give a disperser for one *affine* source (a uniform distribution of an affine subspace) of linear entropy (=dimension). The error in all these constructions is a constant.

All these new results heavily rely on the techniques and results in this paper. They demonstrate the power of the tools from combinatorial number theory even more, and argue the relevance of these tools to theoretical computer science.

In light of this progress, the two natural directions which present themselves. The first is reducing the entropy requirement in extraction from a constant number of sources. Can we go down to polynomial, or even polylogarithmic entropy? The second is reducing the error in all these constructions. Can we go down to polynomial, or even exponential error? The first problem seems to require a way around the [BKT03] argument (which necessarily needs linear entropy). The second problem seem to require bypassing the use of a non-constructive extractor as a building block. Both need new ideas!

Another major problem is of course extraction from two sources. The dispersers in [BKS+05] are strong in the sense that each of their outputs has some constant probability. But the technique achieving that, called the challenge-response mechanism in that paper, seem to inherently prevent

extraction. Can one bypass this method, or enhance it? The same question applies equally well to the problem of extraction from affine sources. An exciting new development in this direction is a very recent result by Jean Bourgain [Bou05], which gives (again using results from [BKT03]) a simple algebraic construction of an extractor for 2 independent sources requiring $(\frac{1}{2} - \epsilon)n$ entropy for some small absolute constant $\epsilon > 0$.

To summarize, after nearly 20 years of no progress, there seem to be new excitement, ideas, tools, and results on deterministic extraction from independent sources, and we expect more to follow soon.

Acknowledgements

We thank Amir Shpilka for many valuable discussions during the early stages of this research. We also thank the anonymous SICOMP referee for many useful comments and corrections on an earlier version of this manuscript.

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.

[Alo95] N. Alon. Tools from Higher Algebra. In R. L. Graham, M. Grotschel and L. Lovasz (eds.), *Handbook of Combinatorics*, Elsevier and The MIT Press, Vol. 2, 1995.

[Bar06] B. Barak A Simple Explicit Construction of an $n^{\tilde{O}(\log n)}$-Ramsey Graph In `arXiv.org` report `math.CO/0601651`, 2006.

[BKS+05] B. Barak, G. Kindler, R. Shaltiel, B. Sudakov, and A. Wigderson. Simulating Independence: New Constructions of Condensers, Ramsey Graphs, Dispersers, and Extractors, 2005.

[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.

[Bou05] J. Bourgain. More on the Sum-Product Phenomenon in Prime Fields and its Applications, 2005. Unpublished manuscript.

[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.

[ES83] P. Erdős and E. Szemerédi. On sums and products of integers. In *Studies in pure mathematics*, pages 213–218. Birkhäuser, Basel, 1983.

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

[FW81] P. Frankl and R. M. Wilson. Intersection theorems with geometric consequences. *Combinatorica.*, 4(1):357–368, 1981.

[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. H. Spencer. *Ramsey Theory*. John Wiley & Sons, 1980.

[GS71] R. L. Graham, and J.H. Spencer. A constructive solution to a tournament problem. *Canad. Math. Bull.*, 14:45–48, 1971.

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

[NN93] J. Naor and M. Naor Small-Bias Probability Spaces: Efficient Constructions and Applications *SIAM J. Comput.*, 22(4): 838-856 ,1993.

[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.

[Plü70] H. Plünnecke. Eine zahlentheoretische Anwendung der Graphentheorie. *J. Reine Angew. Math.*, 243:171–183, 1970.

[Pud05] P. Pudlak. On Explicit Ramsey Graphs and Estimates on the Numbers of Sums And Products, 2005. Unpublished manuscript.

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

[Raz05] R. Raz. Extractors with Weak Random Seeds, 2005.

[Ruz89] I. Ruzsa. An Application of Graph Theory to Additive Number Theory. *Scientia Ser. A: Mathematical Sciences*, 3:97–109, 1989.

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

[RSW03] O. Reingold, M. Saks, and A. Wigderson. Personal Communication, 2003.

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

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

[SSV04] B. Sudakov, E. Szemerédi, and V. Vu. On a question of Erdős and Moser, 2004. Submitted for publication, available on `http://www.math.princeton.edu/~bsudakov/` .

[TV00] L. Trevisan and S. Vadhan. Extracting randomness from samplable distributions. In *Proc. 41st 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. 19th 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. 31st 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 A Proof of Theorem 1.4

In this section, we prove Theorem 1.4 with a better explicit dependence between the constants ϵ and δ (namely $\epsilon = \Theta(\delta)$). That is, we prove the following theorem:

Theorem A.1. *There exists an absolute constant $c_0 > 0$ such that for every field $\mathbb{F}$ and $\delta > 0$ such that $|\mathbb{F}|^{1/k}$ is* not *an integer for every integer $2 \leq k \leq (2/\delta)$, and every set $\mathcal{A} \subseteq \mathbb{F}$ with $|\mathbb{F}|^{\delta} < |\mathcal{A}| < |\mathbb{F}|^{1-\delta}$, it holds that*

$$\max\{|\mathcal{A}+\mathcal{A}|, |\mathcal{A}\cdot\mathcal{A}|\} \geq |\mathcal{A}|^{1+c_0\delta}$$

Note that Theorem A.1 indeed implies Theorem 1.4 since a finite field $\mathbb{F}$ has a subfield of size M if and only if $M = |\mathbb{F}|^{1/k}$ for some integer k.

We prove Theorem A.1 by proving the following two claims A.2 and A.3:

Claim A.2. *There exists a fixed size uniform rational expression $r(\cdot)$ such that for every δ, field $\mathbb{F}$ satisfying the conditions of Theorem A.1 and set $\mathcal{B} \subseteq \mathbb{F}$ with $|\mathcal{B}| \geq |\mathbb{F}|^{\delta}$,*

$$|r(\mathcal{B},\ldots,\mathcal{B})| \geq \min\{|\mathcal{B}|^{1+\delta}, |\mathbb{F}|\}$$

By a *rational expression* we mean an expression involving only the operations $+, -, \cdot$ and $/$ and variable names, but no coefficients (for example, $(x_1 - x_2)/(x_2 \cdot x_3 + x_4)$).[20] By *fixed size* we mean that the number of operations and variables in the expression does not depend on δ, $\mathbb{F}$ or $\mathcal{A}$. A rational function is obviously a division of two polynomials. We say that a polynomial is *uniform* if all its monomials have the same degree. We say that a rational function is *uniform* if it's a division of two uniform polynomials. In fact, the expression r obtained from the proof of Claim A.2 will be very simple: it will be a uniform 16-variable rational expression with both the nominator and the denominator of degree 2.

Claim A.3. *For every uniform rational expression $r(\cdot)$ there is a constant c (depending on $r(\cdot)$) such that if $\mathcal{A} \subseteq \mathbb{F}$ satisfies $|\mathcal{A}+\mathcal{A}|, |\mathcal{A}\cdot\mathcal{A}| \leq |\mathcal{A}|^{1+\rho}$ (for sufficiently small $\rho > 0$) then there exists a set $\mathcal{B} \subseteq \mathcal{A}$ with $|\mathcal{B}| \geq |\mathcal{A}|^{1-c\rho}$ but $|r(\mathcal{B},\ldots,\mathcal{B})| \leq |\mathcal{A}|^{1+c\rho}$.*

We note that this actually holds also for non-uniform rational expressions as well, but the proof is slightly easier for uniform expressions.

[20] Throughout this section we define $x/0$ as 0 (it will not matter much since we'll always have that the event that the denominator is zero in the expression is zero has negligible probability).

Proving Theorem A.1 using Claims A.2 and A.3. Claims A.2 and A.3 together imply Theorem A.1. Indeed let $r(\cdot)$ be the rational expression obtained from Claim A.2 and let c be the constant (depending on r) obtained from Claim A.3. If for $\delta, \mathbb{F}$ and $\mathcal{A}$ as in the conditions of the theorem both $|\mathcal{A}+\mathcal{A}|$ and $|\mathcal{A}\cdot\mathcal{A}|$ are less than $|\mathcal{A}|^{1+\delta/(10c)}$ then there is a subset $\mathcal{B}$ of size at least $|\mathcal{A}|^{1-c\delta/(10c)} \geq |\mathbb{F}|^{\delta/2}$ such that $|r(\mathcal{B},\ldots,\mathcal{B})| \leq |\mathcal{A}|^{1+\delta/10} < |\mathcal{B}|^{1+\delta/2}$, thus obtaining a contradiction. □

We note that Claim A.3 follows almost immediately from Lemmas 2.4 and 3.1 in [BKT03]. Nevertheless, for the sake of completeness, we do provide a proof of Claim A.3 (following similar lines to the proofs of [BKT03]) in Appendix A.2. We now move to the proof of Claim A.2.

A.1 Proof of Claim A.2

To prove Claim A.2 we'll prove the following even simpler claim:

Claim A.4. *Let $\mathbb{F}$ be any field and let $\mathcal{B} \subseteq \mathbb{F}$ and $k \in \mathbb{N}$ (with $k \geq 2$) be such that $|\mathbb{F}|^{1/k} < |\mathcal{B}| \leq |\mathbb{F}|^{1/(k-1)}$. Then $|\frac{\mathcal{B}-\mathcal{B}}{\mathcal{B}-\mathcal{B}}| \geq |\mathbb{F}|^{1/(k-1)}$.*

Proof. Suppose otherwise that $|\frac{\mathcal{B}-\mathcal{B}}{\mathcal{B}-\mathcal{B}}| < |\mathbb{F}|^{1/(k-1)}$. Thus we can find $s_1 \notin \frac{\mathcal{B}-\mathcal{B}}{\mathcal{B}-\mathcal{B}}$. Similarly, if $k > 2$ then we can find $s_2 \notin \frac{\mathcal{B}-\mathcal{B}}{\mathcal{B}-\mathcal{B}} + s_1\frac{\mathcal{B}-\mathcal{B}}{\mathcal{B}-\mathcal{B}}$ (since this set is of size at most $|\frac{\mathcal{B}-\mathcal{B}}{\mathcal{B}-\mathcal{B}}|^2 < |\mathbb{F}|$. In this way we define inductively $s_1,\ldots,s_{k-1}$ such that for $1 < i \leq k-1$,

$$s_i \notin \tfrac{\mathcal{B}-\mathcal{B}}{\mathcal{B}-\mathcal{B}} + s_1\tfrac{\mathcal{B}-\mathcal{B}}{\mathcal{B}-\mathcal{B}} + \ldots + s_{i-1}\tfrac{\mathcal{B}-\mathcal{B}}{\mathcal{B}-\mathcal{B}}$$

Consider the function $f(x_0,\ldots,x_{k-1}) = x_0 + s_1x_1 + \ldots + s_{k-1}x_{k-1}$. This is a function from $\mathcal{B}^k$ to $\mathbb{F}$ where $|\mathcal{B}|^k > |\mathbb{F}|$ and hence it has a *collision*. That is, there are two vectors $\vec{x} = (x_0,\ldots,x_{k-1})$ and $\vec{x}\,' = (x'_0,\ldots,x'_{k-1})$ such that $\vec{x} \neq \vec{x}\,'$ but $f(\vec{x}) = f(\vec{x}\,')$. If we let i be the maximum index such that $x_i \neq x'_i$ we see that

$$(x_0 - x'_0) + s_1(x_1 - x'_1) + \ldots + s_{i-1}(x_{i-1} - x'_{i-1}) = s_i(x'_i - x_i) \quad .$$

Dividing by $(x'_i - x_i)$ we get that $s_i = y_0 + s_1y_1 + \ldots + s_{i-1}y_{i-1}$ where all the y_i's are members of $\frac{\mathcal{B}-\mathcal{B}}{\mathcal{B}-\mathcal{B}}$, contradicting our choice of s_i. □

To prove Claim A.2 we let $\mathbb{F}$,δ, $\mathcal{B}$ be as stated in the theorem and choose k such that $|\mathbb{F}|^{1/k} < |\mathcal{B}| \leq |\mathbb{F}|^{1/(k-1)}$. By one invocation of $\frac{\mathcal{B}-\mathcal{B}}{\mathcal{B}-\mathcal{B}}$ we get to a set of size at least $|\mathbb{F}|^{1/(k-1)}$ but since that is not an integer, this set is of size *larger* than $|\mathbb{F}|^{1/(k-1)}$ (this is for $k > 2$, for $k = 2$ we get to the entire field $\mathbb{F}$). Thus if we compose this expression two times (i.e., let $r(x_1,\ldots,x_{16}) = \frac{r'(x_1,\ldots,x_4)-r'(x_5,\ldots,x_8)}{r'(x_9,\ldots,x_{12})-r'(x_{13},\ldots,x_{16})}$ where $r'(a,b,c,d) = \frac{a-b}{c-d}$) then we get that for $k > 2$,

$$|r(\mathcal{B},\ldots,\mathcal{B})| \geq |\mathbb{F}|^{\frac{1}{k-2}} = |\mathbb{F}|^{\frac{1}{k-1}(1+\frac{1}{k-2})} \geq |\mathcal{B}|^{1+\delta}$$

where for $k = 2$, $r(\mathcal{B},\ldots,\mathcal{B}) = \mathbb{F}$. □

A.2 Proof of Claim A.3

We now prove Claim A.3. Before turning to the actual proof, we state two number theoretic lemmas which we'll use. These lemmas are variants of the lemmas presented in Section 3.2.

A.2.1 More number theoretic lemmas.

Stronger form of Gowers's Lemma. We will use the following generalized and stronger form of Lemma 3.8 (see [Gow98, BKT03], [SSV04, Claim 4.4]). Because it is such a useful lemma, we state it below in the most general (and unfortunately also cumbersome) form:

Lemma A.5. *Let $\mathcal{A}_1, \ldots, \mathcal{A}_k$ be subsets of some group G with $|\mathcal{A}_i| = M$ for all $i \in [k]$. Then, there exists $C = C(k)$ such that for every $\rho > 0$, if $\mathsf{cp}(\sum_{i=1}^k \mathcal{A}_i) \geq \frac{1}{M^{1+\rho}}$ then there are k subsets $\mathcal{A}'_1, \ldots, \mathcal{A}'_k$ with $\mathcal{A}'_i \subseteq \mathcal{A}_i$ and $|\mathcal{A}'_i| \geq M^{1-C\rho}$ for every $i \in [k]$ satisfying*

$$|\sum_{i=1}^k \mathcal{A}'_i| \leq M^{1+C\rho} \tag{1}$$

Moreover, Equation 1 is demonstrated by the fact that every element $z \in \sum_{i=1}^k \mathcal{A}'_i$ can be represented in $M^{\ell-1-C'\rho}$ different ways as a sum $z = y_1 + \cdots + y_\ell$ (for $C' = C'(k)$ and $\ell = 2k^2 - k$) where each of the y_j's is a member of A_i or $-A_i$ for some $i = i(j)$ (with the choice of a sign also being a function of j).

Furthermore, if for all $i \in [k]$, $\mathcal{A}_i = \mathcal{A}$ or $\mathcal{A}_i = -\mathcal{A}$ for some set $\mathcal{A}$ then all the subsets $\mathcal{A}'_i$ are of the form $\mathcal{A}'_i = \mathcal{A}'$ or $\mathcal{A}'_i = -\mathcal{A}'$ for some subset $\mathcal{A}' \subseteq \mathcal{A}$.

It is easy to see that by applying Lemma A.5 for $k = 2$ we get Lemma 3.8 (perhaps with 10 replaced by a different constant). Lemma A.5 can be proven by a generalization of the proof of Lemma 3.8, see [SSV04].[21] When using Lemma A.5, we will always be in the case that that we have an upper bound on the *set size* of $\sum_{i=1}^k \mathcal{A}_i$ (i.e., $|\sum_{i=1}^k \mathcal{A}_i| \leq M^{1+\rho}$) and not just a lower bound on its collision probability.

We note that in the proofs below we will use several times this technique of showing that some set $\mathcal{B}$ is not much larger than M by showing that every $b \in \mathcal{B}$ can be represented in roughly $M^{\ell-1}$ ways as a sum of ℓ elements, each from some set $\mathcal{D}$ of size M.

Sumset estimates. We'll also use the following Lemma, which is a variant of Lemma 3.7:

Lemma A.6 ([Plü70, Ruz96])**.** *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*

$$|\mathcal{A} \pm \underbrace{\mathcal{B} \cdots \pm \mathcal{B}}_{h \text{ times}}| \leq M^{1+2h\rho}$$

We note that this lemma immediately implies a similar result for sets that are of slightly different sizes. That is, if $|\mathcal{A}| = M$ and $|\mathcal{B}| = M^{1-\epsilon}$ then we can break up $\mathcal{A}$ to $\mathcal{A}_1, \ldots, \mathcal{A}_{M^\epsilon}$ that are of the same size as $\mathcal{B}$. We then have that $|\mathcal{A}_i + \mathcal{B}| \leq |\mathcal{A} + \mathcal{B}| \leq M^{1+\rho}$ and hence the lemma implies that for every i, $\mathcal{A}_i \pm \underbrace{\mathcal{B} \cdots \pm \mathcal{B}}_{h \text{ times}}| \leq M^{1+2h\rho}$. However $\mathcal{A} \pm \mathcal{B} \cdots \pm \mathcal{B}$ is just the union of $\mathcal{A}_i \pm \mathcal{B} \cdots \pm \mathcal{B}$ for all i and hence we get that

$$|\mathcal{A} \pm \underbrace{\mathcal{B} \cdots \pm \mathcal{B}}_{h \text{ times}}| \leq M^{1+2h\rho+\epsilon}$$

[21] We note that this lemma is not stated exactly in this form in [SSV04]. Rather, Claim 4.4 there states that under these conditions every such z can be represented in roughly M^{2k-2} ways as a sum of $2k-1$ elements w_l, where each of these elements can be represented in roughly M^{k-1} ways as a sum of k elements in the A_i's. It is also stated in [SSV04] in terms of distance from having high min-entropy rather than high collision probability, see Footnote 17.

Similarly if $|\mathcal{B}| = M$ and $|\mathcal{A}| = M^{1-\epsilon}$ then,

$$|\mathcal{A} \pm \underbrace{\mathcal{B} \cdots \pm \mathcal{B}}_{h \text{ times}}| \leq M^{1+2h\rho+h\epsilon}$$

since if we split sB to $\mathcal{B}_1, \ldots, \mathcal{B}_{M^\epsilon}$ then we have that $\mathcal{A} \pm \mathcal{B} \cdots \pm \mathcal{B}$ is the union of sets of the form $\mathcal{A} \pm \mathcal{B}_{i_1} \pm \cdots \pm \mathcal{B}_{i_h}$ for all $i_1, \ldots, i_h \in [M^\epsilon]$.

A.2.2 The actual proof.

In fact, to prove Claim A.3 it is enough to prove the following claim:

Claim A.7. *For every integer $k > 0$ there exists a constant $C = C(k) > 0$ such that for every $\rho > 0$, if $\mathcal{A} \subseteq \mathbb{F}$ satisfies $|\mathcal{A}+\mathcal{A}|, |\mathcal{A}\cdot\mathcal{A}| \leq |\mathcal{A}|^{1+\rho}$ then there is a set $\mathcal{B} \subseteq \mathcal{A}$ such that $|\mathcal{B}| \geq |\mathcal{A}|^{1-C\rho}$ but $|\mathcal{B}^k - \mathcal{B}^k| \leq |\mathcal{A}|^{1+C\rho}$.*

Obtaining Claim A.3 from Claim A.7. Claim A.7 implies Claim A.3 by applying Lemma A.6. Indeed, suppose that r is a rational expression where both the numerator and denominator have at most k' monomials each of degree k' (if one of them has monomials of smaller degree than k', we can multiply with a uniform polynomial to make the degree k'). Let $k = 4k'^2$ and let $\mathcal{B}$ be the subset of $\mathcal{A}$ obtained from Claim A.7 such that $|\mathcal{B}^k - \mathcal{B}^k|$ is at most $|\mathcal{A}|^{1+C\rho}$ for some constant $C = C(k)$. Since $|\mathcal{B}| \geq |\mathcal{A}|^{1-C\rho}$, this means that $|\mathcal{B}^k - \mathcal{B}^k| \leq |\mathcal{B}|^{1+C'\rho}$ for some different constant C'. This implies by Lemma A.6 that $|k\mathcal{B}^k| \leq |\mathcal{B}|^{1+2C'k\rho}$. Now, $k\mathcal{B}^k \supseteq \mathcal{C} \stackrel{def}{=} (k'\mathcal{B}^{k'})(k'\mathcal{B}^{k'})$ and hence the size of $\mathcal{C}$ is also at most $|\mathcal{B}|^{1+2C'k\rho}$. Applying Lemma A.6 again we get that $\left|\frac{k'\mathcal{B}^{k'}}{k'\mathcal{B}^{k'}}\right| \leq |\mathcal{A}|^{1+4C'k\rho}$. However, $r(\mathcal{B}, \ldots, \mathcal{B}) \subseteq \frac{k'\mathcal{B}^{k'}}{k'\mathcal{B}^{k'}}$ and hence we're done.

Proof of Claim A.7. We now turn to proving Claim A.7. Let $\mathcal{A}$ be a set satisfying the conditions of the claim, and denote $M = |\mathcal{A}|$. We will prove that for some constant $C = C(k) > 0$, there is a subset $\mathcal{B} \subseteq \mathcal{A}$ of size at least $M^{1-C\rho}$ such that any member of the set $\mathcal{B}^k - \mathcal{B}^k$ can be represented in at least $M^{\ell-1-C\rho}$ different ways as a sum $d_1 + \ldots + d_\ell$ where all the elements d_i come from a set $\mathcal{D}$ of size at most $M^{1+C\rho}$ for some $\ell = \ell(k)$. This clearly implies that $|\mathcal{B}^k - \mathcal{B}^k| \leq M^{1+C'\rho}$ for some constant C' thus proving the claim.

Proof idea: the case $k = 2$. To illustrate the proof idea, we now sketch the proof for the case $k = 2$. The proof for general k follows in exactly the same way, although with more cumbersome notations. Under the conditions of the claim, $|\mathcal{A}+\mathcal{A}| \leq |\mathcal{A}|^{1+\rho}$, and hence $\mathsf{cp}(\mathcal{A}-\mathcal{A}) = \mathsf{cp}(\mathcal{A}+\mathcal{A}) \geq |\mathcal{A}|^{1-\rho}$. Hence, by Lemma A.5, there exists a set $\mathcal{A}' \subseteq \mathcal{A}$ with $|\mathcal{A}'| \geq M^{1-C\rho}$ (where C is some absolute constant independent of ρ) such that not only $|\mathcal{A}' - \mathcal{A}'| \leq M^{1+C\rho}$ but actually this can be demonstrated by the fact that every element of $\mathcal{A}' - \mathcal{A}'$ can be represented in at least $M^{5-C\rho}$ different ways in the form $a_1 - a_2 + a_3 - a_4 + a_5 - a_6$ where for $i = 1, \ldots, 6$, $a_i \in \mathcal{A}$.[22]

We know that every member of $\mathcal{A}' - \mathcal{A}'$ can be represented in roughly M^5 different ways as $a_1 - a_2 + a_3 - a_4 + a_5 - a_6$ with the a_i's in $\mathcal{A}$. Now, if we multiply this by an arbitrary element of $\mathcal{A}$ we get that every member of $(\mathcal{A}' - \mathcal{A}')\mathcal{A}$ can be represented in roughly M^5 different ways as $b_1 - b_2 + b_3 - b_4 + b_5 - b_6$ with the b_i's in $\mathcal{A} \cdot \mathcal{A}$. Since by the conditions of the claim, $\mathcal{A} \cdot \mathcal{A}$ is

[22] For simplicity of exposition we assumed that the pattern of + and − signs that is obtained from Lemma A.5 is as written above. The proof clearly follows through regardless of the fixed pattern we use.

also of roughly size M, we get that the set $(\mathcal{A}' - \mathcal{A}')\mathcal{A}$ is also "not large" (i.e., of size $M^{1+C'\rho}$ for some absolute constant C'). Now consider an element $y - z$ of the set $\mathcal{A}'^2 - \mathcal{A}'^2$. For the sake of simplicity, we assume for a moment (*with* loss of generality) that any y in $\mathcal{A}' \cdot \mathcal{A}'$ can be represented in roughly M different ways as $y = y_1 y_2$ with $y_1, y_2 \in \mathcal{A}'$.[23] Since $z \in \mathcal{A}' \cdot \mathcal{A}'$, it is equal to $z_1 z_2$ for some $z_1, z_2 \in \mathcal{A}'$. Every representation $y_1 y_2$ of y induces a representation of $y - z = y_1 y_2 - z_1 z_2$ as $(y_1 - z_1)y_2 + z_1(y_2 - z_2)$, and so we get that every element of $\mathcal{A}'^2 - \mathcal{A}'^2$ can be represented as $d_1 - d_2$, with $d_1, d_2 \in \mathcal{D} \stackrel{def}{=} (\mathcal{A}' - \mathcal{A}')\mathcal{A}$ in roughly M different ways. However, since we already showed that the size of $\mathcal{D}$ is roughly M, this proves that $\mathcal{A}'^2 - \mathcal{A}'^2$ is also of size roughly M and hence we're done (for the case $k = 2$).

Proving for general k. We now turn to proving the claim rigorously and for any k. Recall that our goal is to find a not-too-small subset $\mathcal{B}$ of $\mathcal{A}$ such that $\mathcal{B}^k - \mathcal{B}^k$ can be represented in roughly $M^{\ell-1}$ ways as a sum of ℓ elements from a set $\mathcal{D}$ that is not too large. We'll start by defining the set $\mathcal{D}$.

Let ℓ be some number, and let $a, b \in \mathcal{A}'$ (where $\mathcal{A}'$ is obtained from Lemma A.5 as above) and $c \in \mathcal{A}'^{\ell}$. By the same reasoning as above, we get that the element $(a - b)c$ can be represented in at least $M^{5-C\rho}$ different ways as $a_1 c - a_2 c + a_3 c - a_4 c + a_5 c - a_6 c$ or in other words, it can be represented in at least $M^{5-C'\rho}$ different ways as $b_1 - b_2 + b_3 - b_4 + b_5 - b_6$ for $b_i \in \mathcal{A}^{\ell+1}$ for $i = 1, \ldots, 6$. By the multiplicative version of Lemma A.6, $|\mathcal{A}^{\ell+1}| \leq M^{1+3\ell\rho}$ and hence the set $(\mathcal{A}' - \mathcal{A}')\mathcal{A}'^{\ell}$ is "small" (i.e., of size at most $M^{6(1+3\ell\rho)} M^{-(5-C'\rho)} \leq M^{1+20C'\ell\rho}$). Using again the multiplicative version of Lemma A.6 (setting $\mathcal{A} = (\mathcal{A}' - \mathcal{A}')\mathcal{A}'^{\ell-1}$, $\mathcal{B} = \mathcal{A}'$), we get that the set $(\mathcal{A}' - \mathcal{A}')\mathcal{A}'^{\ell}\mathcal{A}'^{-\ell'}$ is also "small" (i.e., of size at most $M^{1+C''\rho}$ for some constant $C'' = C''(\ell, \ell')$).[24] Using the fact that $(\mathcal{A}'^{-1} - \mathcal{A}'^{-1}) \subseteq (\mathcal{A}' - \mathcal{A}')\mathcal{A}'^{-2}$, we get that for any ℓ_1, ℓ_2, the set $\mathcal{D}_{\ell_1,\ell_2}$ defined as follows:

$$\mathcal{D}_{\ell_1,\ell_2} = (\mathcal{A}'^{-1} - \mathcal{A}'^{-1})\mathcal{A}'^{\ell_1}\mathcal{A}'^{-\ell_2} \cup (\mathcal{A}' - \mathcal{A}')\mathcal{A}'^{\ell_1}\mathcal{A}'^{-\ell_2}$$

is of size at most $M^{1+C'''\rho}$ for some constant C''' depending on ℓ_1, ℓ_2. We define $\mathcal{D} \stackrel{def}{=} \cup_{\ell_1+\ell_2=\ell} s D_{\ell_1,\ell_2}$ for ℓ as obtained by Lemma A.5 (i.e., $\ell = 2k^2 - k$). As desired, we have that $|\mathcal{D}| \leq M^{1+C\rho}$ where C is a constant depending on ℓ.

Defining the set $\mathcal{B}$. We now turn to defining the required set $\mathcal{B}$. Utilizing Lemma A.5 again, we obtain that for some absolute constant $D = D(k)$, there is a set $\mathcal{B} \subseteq \mathcal{A}'$ with $|\mathcal{B}| \geq M^{1-D\rho}$ and every element in $\mathcal{B}^k$ can be represented in $M^{\ell-1-D\rho}$ ways as $a_1 \cdots a_\ell$, where a_i is in $\mathcal{A}'$ or in $\mathcal{A}'^{-1}$ for $i = 1, \ldots, \ell$ (for $\ell = 2k^2 - k$). Let $y - z$ be a member of $\mathcal{B}^k - \mathcal{B}^k$. Fix one representation $z = z_1 \ldots z_\ell$ of z as a multiplication of elements of $\mathcal{A}'$ or $\mathcal{A}'^{-1}$. The element y can be represented $M^{\ell-1-D\rho}$ times as $y = y_1 \cdots y_\ell$ with $y_i \in \mathcal{A}'$ for $i = 1, \ldots, \ell$. For each such representation we define $d_i = y_1 \cdot \ldots \cdot y_{i-1}(y_i - z_i)z_{i+1} \cdots z_\ell$. Note that $\sum_{i=1}^{\ell} d_i = \prod_{i=1}^{\ell} y_i - \prod_{i=1}^{\ell} z_j = y - z$. Also note that for every i, $d_i \in \mathcal{D}$, where $\mathcal{D}$ is the set defined above. The map $(y_1, \ldots, y_\ell) \mapsto (d_1, \ldots, d_\ell)$ is one-to-one (indeed, given $z_1, \ldots, z_\ell$ we can recover $y_1, \ldots, y_\ell$ from $d_1, \ldots, d_\ell$). Hence we get that each member of $\mathcal{B}^k - \mathcal{B}^k$ can be represented in $M^{\ell-1-D\rho}$ different ways as $\sum_{i=1}^{\ell} d_i$ with $d_i \in \mathcal{D}$, implying that $|\mathcal{B}^k - \mathcal{B}^k| \leq M^{1+D'\rho}$ for $D' = D'(k)$. □

[23] We will not be able to get to that situation in the actual proof below, but we will approximate it using the multiplicative version of Lemma A.5.

[24] Again, the case of division by zero does not matter, but for simplicity, we can just remove 0 from the set $\mathcal{A}'$ if it's there.