

Deterministic Extractors For Small-Space Sources

Jesse Kamp ^{*} Anup Rao [†] Salil Vadhan [‡] David Zuckerman [§]

March 29, 2007

Abstract

We give polynomial-time, deterministic randomness extractors for sources generated in small space, where we model space s sources on $\{0, 1\}^n$ as sources generated by width 2^s branching programs. For every constant $\delta > 0$, our algorithm extracts $.99\delta n$ bits that are exponentially close to uniform (in variation distance) from space s sources of min-entropy δn , where $s = \Omega(n)$. In addition, assuming an efficient deterministic algorithm for finding large primes, there is a constant $\eta > 0$ such that for any $\zeta > n^{-\eta}$, our algorithm extracts $m = (\delta - \zeta)n$ bits that are exponentially close to uniform from space s sources with min-entropy δn , where $s = \Omega(\zeta^3 n)$. Previously, nothing was known for $\delta \leq 1/2$, even for space 0.

Our results are obtained by a reduction to the class of *total-rate* independent sources. This model generalizes both the well-studied models of independent sources and symbol-fixing sources. These sources consist of a set of r independent subsources over $\{0, 1\}^\ell$, where the total min-entropy over all the subsources is k . We give deterministic extractors for such sources when k is as small as $\text{polylog}(r)$, for small enough ℓ .

^{*}Department of Computer Science, University of Texas, Austin, TX 78712, (kamp@cs.utexas.edu). Supported in part by NSF Grant CCR-0310960.

[†]Department of Computer Science, University of Texas, Austin, TX 78712, (arao@cs.utexas.edu). Supported in part by NSF Grant CCR-0310960 and an MCD Fellowship.

[‡]Division of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138, (salil@eecs.harvard.edu). Supported by NSF grant CCF-0133096, ONR grant N00014-04-1-0478, and US-Israel BSF grant 2002246.

[§]Department of Computer Science, University of Texas, Austin, TX 78712, (diz@cs.utexas.edu). Supported in part by a David and Lucile Packard Fellowship for Science and Engineering, NSF Grant CCR-0310960, a Radcliffe Institute Fellowship, and a Guggenheim Fellowship.

Contents

1	Introduction	2
1.1	Small-Space Sources	2
1.1.1	Our Results	4
1.2	Total-Rate Independent Sources	5
1.2.1	Independent Sources	5
1.2.2	Oblivious Bit-Fixing and Symbol-Fixing Sources	5
1.2.3	Our Results	6
1.3	Organization	8
2	Preliminaries	8
2.1	Convex Combinations	8
2.2	Classes of Sources	9
2.3	Seeded Extractors	10
3	Small-Space Sources As Convex Combinations Of Independent Sources	10
4	Extracting From Total-Rate Independent Sources By Reducing To Standard Independent Sources	11
5	Extracting From Polynomial Entropy Rate	12
5.1	Extracting From The Intermediate Model	13
5.2	Condensing To Aligned Sources With High Somewhere-Min-Entropy	15
5.3	Extracting From Independent Sources, A Few Of Which Are Aligned SR-Sources	16
6	Better Extractors For Total-Rate Independent Sources With Many Short Subsources	19
6.1	Random Walks	19
6.2	Reducing to Flat Total-Rate Independent Sources	20
6.3	Extracting From Flat Total-Rate Independent Sources	20
7	Extracting More Bits From Total-Rate Independent Sources	24
7.1	Seed Obtainers	24
7.2	Constructing Samplers	26
7.3	Extractors From Seed Obtainers	26
7.4	Extractors For Smaller Entropy	28
8	Nonconstructive Results	29
8.1	Small-Space Sources	30
8.2	Total-Rate Independent Sources	31
9	Doing Better For Width Two	33
9.1	Extracting From Previous-Bit Sources	33
9.2	Restricted Width Two Sources As Convex Combinations Of Previous-Bit Sources	34

1 Introduction

True randomness is needed for many applications, yet most physical sources of randomness are not truly random, and some are quite weak in that they can have substantial biases and correlations. Weak random sources can also arise in cryptography when an adversary learns some partial information about a random string. A natural approach to dealing with weak random sources is to apply an *extractor* — a function that transforms a weak random source into an almost-perfect random source. For example, Intel’s random number generator (cf., [JK99]) uses the extractor of von Neumann [vN51] as one of its components.

There was a significant body of work in the 80’s focused on this problem of randomness extraction, with researchers considering richer and richer models of weak sources, e.g. [Blu86, SV86, CG88, Vaz87, CFG⁺85, BBR88, BOL90, LLS89]. However, attempts to handle sources that do not have a significant amount of independence ran into strong impossibility results showing that it is impossible to devise a single function that extracts even one bit of randomness from sufficiently general classes of sources [SV86].

These impossibility results led researchers to focus on the weaker task of simulating probabilistic algorithms with weak random sources [VV85, CG88, Vaz86, CW89, Zuc96]. This line of work culminated in the introduction, by Nisan and Zuckerman [NZ96], of the notion of a *seeded* extractor, which uses a small number of additional *truly random* bits, known as the *seed*, as a catalyst for the randomness extraction. When simulating probabilistic algorithms with weak random sources, the need for truly random bits can be eliminated by enumerating over all choices of the seed. Seeded extractors have turned out to have a wide variety of other applications and were found to be closely related to many other important pseudorandom objects. Thus, they were the main focus of attention in the area of randomness extraction in the 90’s, with a variety of very efficient constructions. (See [NTS99, Sha02] for surveys.)

In the last few years, however, there has been a resurgence of interest in the original concept of a “seedless” (or deterministic) extractor, cf. [TV00, Dod00b]. This is motivated in part by the realization that seeded extractors do not seem suitable for many settings where we need randomness, such as cryptography. In addition, seedless extractors for specific classes of sources were found to be useful in mitigating partial key exposure in cryptography [CDH⁺00, Dod00b]. Recent attention on seedless extractors has focused on several classes of sources, the main ones being *independent sources*, which consist of several independent parts, each of which has some randomness [CG88, BIW04, BKS⁺05, Raz05, Rao06]; *bit-fixing sources*, where some of the bits are perfectly random and the rest are fixed [CFG⁺85, CW89, KZ03, GRS04]; and *samplable sources*, where the source is generated by an efficient algorithm [TV00]. Our work relates to all of these models; indeed, we establish connections between them. However, our main motivation is a form of samplable sources — namely ones generated by algorithms that have small space.

Before proceeding, we recall a few standard definitions: the *min-entropy* k of a source X is defined as $H_\infty(X) = \min_s(\log(1/\Pr[X = s]))$. (Here and throughout, all logarithms are base 2 unless otherwise specified.) The *min-entropy rate* δ for a source on $\{0, 1\}^n$ is defined as $\delta = H_\infty(X)/n$. The *variation distance* between random variables X_1 and X_2 on Ω is defined as $|X_1 - X_2| = \max_{S \subseteq \Omega} |\Pr[X_1 \in S] - \Pr[X_2 \in S]| = \frac{1}{2} \sum_{s \in \Omega} |\Pr[X_1 = s] - \Pr[X_2 = s]|$. A function $\text{Ext} : \{0, 1\}^n \rightarrow \{0, 1\}^m$ is an ϵ -*extractor* for a class of random sources \mathcal{X} , if for every $X \in \mathcal{X}$, $\text{Ext}(X)$ is ϵ -close to uniform in variation distance.

1.1 Small-Space Sources

Trevisan and Vadhan [TV00] proposed the study of extraction from weak random sources that are generated by a process that has a bounded amount of computational resources. This seems to be a plausible model for physical random sources and generalizes a number of the previously studied models. They focused on the case that the source is sampled by either a small circuit or an algorithm with a limited running time.

Their main result is a construction of polynomial-time extractors for such sources based on some strong but plausible complexity assumptions. It would be nice to have unconditional constructions (as well as ones that are more efficient and have better error). However, they showed that complexity assumptions are needed for the original model of sources generated by time-bounded algorithms. Thus, they suggested, as a research direction, that we might be able to construct unconditional extractors for sources generated by *space-bounded* algorithms. This model is our focus.

Small space sources are very general in that most other classes of sources that have been considered previously can be computed with a small amount of space. This includes von Neumann’s model of a coin with unknown bias [vN51], Blum’s finite Markov chain model [Blu86], symbol-fixing sources [KZ03], and sources that consist of many independent sources. Strong results in this last model will not follow directly from strong results in the small-space model, but our results do generalize, for example, the results of [BIW04]. In fact, the only model for which deterministic extractors have been given that appears unrelated to our model is “affine sources”. Yet despite the small-space model being so natural, very little in the way of explicit constructions for such sources was known.

The first example of an explicit construction was due to Blum [Blu86], who showed how to extract from sources generated by a finite Markov chain with a constant number of states. His results generalized the earlier results of von Neumann [vN51] for extracting from an independent coin with unknown bias. However, the finite Markov chain model is very restricted; it has a constant-size description and the transitions must be the same at each time step.

The model for small-space sources we consider is similar to the one previously considered by Koenig and Maurer [KM04, KM05]. It is a generalization of the Markov chain model to time-dependent Markov chains, which yields a much richer class of sources. Our model of a space s source is basically a source generated by a width 2^s branching program. The exact model we consider is that at each step the process generating the source is in one of 2^s states. This can be modelled by a layered graph with each layer corresponding to a single time-step and consisting of vertices corresponding to each of the states. From each node v in layer i , the edges leaving v (going to layer $i + 1$) are assigned a probability distribution as well as an output bit for each edge. Unlike in Blum’s model [Blu86], the transitions can be different at each time-step.

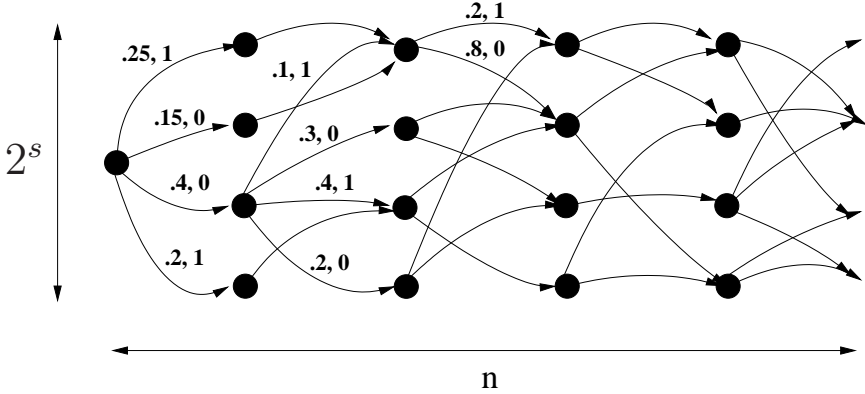


Figure 1: Part of a space $s = 2$ source

It can be shown using the probabilistic method that there exist extractors even when the space s is a constant fraction of the min-entropy k , even when the min-entropy is logarithmically small. Our goal is to provide *efficient* and *deterministic* constructions with parameters that come as close to these bounds as

Table 1: Small space extractors for sources on $\{0, 1\}^n$ that extract 99% of the min-entropy. In this table c and C represent sufficiently small and large constants, respectively.

Reference	Min-entropy Rate	Space	Error
Thm 1.1	$\delta \geq n^{-c}$	$c\delta^3 n$	$\exp(-n^c)$
Thm 1.3	Any constant δ	cn	$\exp(-\tilde{\Omega}(n))$
Thm 1.4	$\delta \geq C/\log n$	$c\delta \log n$	$\exp(-n^{.99})$

possible.

Koenig and Maurer [KM04, KM05] gave the first explicit constructions of extractors for space-bounded sources. Their extractors require the min-entropy rate to be least $1/2$. We do not know of any other constructions for space-bounded sources, even space 0 sources, which are simply sources of independent bits each of which has a different, unknown, bias.

1.1.1 Our Results

For space s sources with min-entropy $k = \delta n$, we have several constructions, all of which are able to extract almost all of the entropy in the source. These extractors are summarized in Table 1. The first extracts whenever $\delta > n^{-\eta}$ for some fixed constant η and extracts almost all of the entropy.

Theorem 1.1. *Assume we can find primes with length in $[\tau, 2\tau]$ deterministically in time $\text{poly}(\tau)$. Then there is a constant $\eta > 0$ such that for every $n \in \mathbb{N}$, and $\delta > \zeta > n^{-\eta}$, there is a polynomial-time computable ϵ -extractor $\text{Ext} : \{0, 1\}^n \rightarrow \{0, 1\}^m$ for space s sources with min-entropy rate δ , where $s = \Omega(\zeta^3 n)$, $m = (\delta - \zeta)n$, and $\epsilon = 2^{-n^{\Omega(1)}}$.*

Remark 1.2. The assumption about finding primes follows from Cramer’s conjecture on the density of primes [Cra37], together with the deterministic primality test of [AKS04].

We also have constructions that do not depend on the ability to find large primes. Though the parameters of these constructions are mostly subsumed by the previous construction, they are considerably simpler and achieve somewhat better error. For constant min-entropy rate sources, we have a construction that extracts any constant fraction of the entropy.

Theorem 1.3. *For any constants $\delta > \zeta > 0$ and every $n \in \mathbb{N}$, there is a polynomial-time computable ϵ -extractor $\text{Ext} : \{0, 1\}^n \rightarrow \{0, 1\}^m$ for space s sources with min-entropy rate δ , where $s = \Omega(n)$, $m = (\delta - \zeta)n$, and $\epsilon = 2^{-\Omega(n/\log^3 n)}$.*

The last extractor works with min-entropy rate as low as $\delta = \Omega(1/\log n)$ and space $O(\delta \log n)$.

Theorem 1.4. *For every $n \in \mathbb{N}$ and $\delta > \zeta > 28/\log n$ and $s \leq (\zeta \log n)/28$, there is a polynomial-time computable ϵ -extractor $\text{Ext} : \{0, 1\}^n \rightarrow \{0, 1\}^m$ for space s sources with min-entropy rate δ , where $m = (\delta - \zeta)n$ and $\epsilon = \exp(-n/(2^{O(s/\zeta)} \cdot \log^5 n))$.*

In comparison to the previous results (e.g. [KM04, KM05]) we have reduced the min-entropy required from $n/2$ to $n^{1-\Omega(1)}$ (in Theorem 1.1). However, we are still far from what can be achieved nonconstructively, where we can extract when the min-entropy is logarithmically small. We also have a gap in terms

of the space tolerated. Nonconstructively we can get s to be almost $\delta n/2$ while our results require s to be smaller than $\delta^3 n$.

In a partial attempt to close the entropy gap for the case of space 1 sources, we also have an extractor that extracts about $\Omega(k^2/n)$ bits from a more restricted model when $k > n^{0.81}$. The extra restriction is that the output bit is required to be the same as the state.

1.2 Total-Rate Independent Sources

Our extractors for small-space sources are all obtained via a reduction from a new model of sources we introduce called *total-rate independent sources*. The reduction we use is based on that of Koenig and Maurer [KM04, KM05], who used it to generalize extractors for two independent sources. Total-rate independent sources consist of a string of r independent sources of length ℓ such that the total min-entropy of all r sources is at least k . Our reduction shows that optimal extractors for total-rate independent sources are also essentially optimal extractors for small-space sources. In addition to being a natural model, these sources are a common generalization of two of the main models studied for seedless extraction, namely symbol-fixing sources [CFG⁺85, KZ03] and independent sources [CG88, BIW04], which we proceed to discuss below.

1.2.1 Independent Sources

One of the most well-studied models of sources is that of extracting from a small number of *independent sources*, each of which has a certain amount of min-entropy, a model essentially proposed by Chor and Goldreich [CG88]. They constructed extractors for two independent sources with entropy rate greater than $1/2$. Recently, similar extractors have been obtained for multiple independent sources with any constant and even subconstant entropy rate, but each of these require at least 3 independent sources [BIW04, BKS⁺05, Raz05, Rao06]. This model is appealing because the individual sources can have arbitrary correlations and biases, and it seems plausible that we can ensure independence between a few such sources. However, such extractors require knowing that all of the sources have large entropy. This motivates our generalization of independent sources to total-rate independent sources, where we only require that the *total* min-entropy over all of the sources is high. Another difference between what we consider is that the usual independent source model consists of few sources that are long, whereas total-rate independent sources are interesting even if we have many short sources.

1.2.2 Oblivious Bit-Fixing and Symbol-Fixing Sources

Another particular class that has been studied a great deal is that of *bit-fixing sources*, where some subset of the bit-positions in the source are fixed and the rest are chosen uniformly at random. The first extractors for bit-fixing sources extracted perfectly random bits [CFG⁺85, CW89] and required the source to have a large number of random positions. Kamp and Zuckerman [KZ03] constructed extractors that worked for sources with a much smaller number of random bits. They also generalized the notion of bit-fixing sources to symbol-fixing sources, where instead of bits the values are taken from a d symbol alphabet. Gabizon, Raz and Shaltiel [GRS04] gave a construction that converts any extractor for bit-fixing sources into one that extracts almost all of the randomness, which they apply to the extractor from [KZ03].

Total-rate independent sources can be seen as a generalization of symbol-fixing sources, where each symbol is viewed as a separate source.¹ The difference is that instead of each symbol being only fixed or

¹Though for ease of presentation we define total-rate independent sources only over sources with alphabet size 2^ℓ , more generally the sources could be over alphabets of any size d , as with symbol-fixing sources. All of our results naturally generalize to this

uniformly random, the symbols (sources) in total-rate independent sources are allowed to have any distribution as long as the symbols are independent. Naturally, we place a lower bound on the total min-entropy rather than just the number of random positions. Usually, symbol-fixing sources are thought of as having many symbols that come from a small alphabet (e.g. $\{0, 1\}$). This restriction is not necessary to the definition, however, and here we consider the full range of parameters, including even the case that we have a constant number of symbols from an exponentially large “alphabet” (e.g. $\{0, 1\}^\ell$).

1.2.3 Our Results

Our extractors for total-rate independent sources are all based on generalizing various techniques from extractors for independent and symbol-fixing sources.

Koenig and Maurer [KM04, KM05] showed how any extractor for two independent sources with certain algebraic properties can be translated into an extractor for many sources where only two of the sources have sufficient entropy. Their result generalizes to extractors for more than two sources. We show that this also yields extractors for independent-symbol sources. In particular, we apply this to extractors for independent sources that follow from the exponential sum estimates of Bourgain, Glibichuk, and Konyagin [BGK06] (see Bourgain [Bou05]), and thereby obtain extractors for total-rate independent sources of any constant min-entropy rate. These extractors are quite simple. Each source is viewed as being an element of a finite field, and the output of the extractor is simply the product of these finite field elements.

We also show how to use ideas from the work of Rao [Rao06] for extracting from several independent sources, together with recent constructions of randomness efficient condensers [BKS⁺05, Raz05], to get extractors for total-rate independent sources that extract from sources of min-entropy $(r\ell)^{1-\Omega(1)}$.

When the subsources each have short length ℓ , we use ideas from the work of Kamp and Zuckerman [KZ03] about bit-fixing sources to construct extractors for total-rate independent sources with min-entropy k . We can extract many bits when $k > 2^\ell \sqrt{r\ell}$, and for $k = \Omega(2^{2\ell})$ we can still extract $\Omega(\log k)$ bits. The base extractor simply takes the sum of the sources modulo p for some $p > 2^\ell$, similar to the cycle walk extractor in [KZ03]. Using this extractor we can extract $\Omega(\log k)$ bits. To extract more bits when k is sufficiently large, we divide the source into blocks, apply the base extractor to each block, and then use the result to take a random walk on an expander as in [KZ03].

Unlike the first two extractors, the extractors obtained using this technique use the full generality of the total-rate independent sources. In the first two constructions, using a Markov argument we can essentially first reduce the total-rate independent sources into sources where some of the input sources have sufficiently high min-entropy while the rest may or may not have any min-entropy. These reductions also cause some entropy to be lost. In this last construction, however, we benefit even from those sources that have very little min-entropy. Thus we are able to take advantage of all of the entropy, which helps us extract from smaller values of k .

We also show how to generalize the construction of Gabizon et al. [GRS04] to total-rate independent sources to enable us to extract more of the entropy. Note that we use it to improve not only the extractors based on [KZ03] (analogous to what was done in [GRS04] for bit-fixing sources), but also our extractors based on techniques developed for independent sources. Independently of our work, Shaltiel [Sha05] has recently generalized the ideas in [GRS04] to give a framework for constructing deterministic extractors which extract almost all of the entropy from extractors which extract fewer bits. Our extractor can be seen to fit inside this framework, although we cannot directly use his results as a black box to obtain our results.

Applying the technique based on [GRS04] to our extractors that use the independent sources techniques

more general case.

Table 2: Total-rate independent source extractors for sources on $(\{0, 1\}^\ell)^r$ that extract 99% of the min-entropy. In this table c and C represent sufficiently small and large constants, respectively.

Reference	Min-entropy Rate	Error
Thm 1.5	$\delta = (r\ell)^{-c}$	$\exp(-(r\ell)^c)$
Thm 1.6	Any constant δ	$\exp(-\tilde{\Omega}(r\ell))$
Thm 1.7	$\delta = C \frac{d \log^{3/2} r}{(r\ell)^{\frac{1}{2}-\gamma}}$	$\exp(-(r\ell)^{2\gamma})$
Thm 1.8	$\delta = (2^\ell \log r)^C / r$	$(\delta r\ell)^{-c}$

of Rao [Rao06], the results of [BGK06], and the bit-fixing source extractor from [KZ03], respectively, we get the following three theorems. These theorems are directly used to obtain the small-space extractors from Theorem 1.1, Theorem 1.3, and Theorem 1.4. Table 2 presents a summary of these extractors.

Theorem 1.5. *Assuming we can find primes with length in $[\tau, 2\tau]$ deterministically in time $\text{poly}(\tau)$, there is a constant η such that for every $r, \ell \in \mathbb{N}$ and $\delta > \zeta > (r\ell)^{-\eta}$, there is a polynomial-time computable ϵ -extractor $\text{Ext} : (\{0, 1\}^\ell)^r \rightarrow \{0, 1\}^m$ for sets of r independent sources over $\{0, 1\}^\ell$ with total min-entropy rate $\delta > \zeta$ where $m = (\delta - \zeta)r\ell$ and $\epsilon = \exp(-(r\ell)^{\Omega(1)})$.*

We note that in the independent sources model this extractor gives comparable results to the extractor from [BIW04] as a corollary.

The following extractor extracts a constant fraction of the entropy from any constant rate source.

Theorem 1.6. *For any constants $\delta > \zeta > 0$ and every $r \in \mathbb{N}$, there is a polynomial-time computable ϵ -extractor $\text{Ext} : (\{0, 1\}^\ell)^r \rightarrow \{0, 1\}^m$ for sets of r total min-entropy rate δ independent subsources over $\{0, 1\}^\ell$, where $m = (\delta - \zeta)r\ell$ and $\epsilon = 2^{-\Omega((r\ell)/\log^3(r\ell))}$.*

For the following extractor we can take $\zeta = \tilde{O}(1/\sqrt{r})$ and can then extract randomness from sources with min-entropy rate as small as $\delta = \tilde{O}(1/\sqrt{r})$.

Theorem 1.7. *For every $r \in \mathbb{N}$, $1 \leq \ell \leq \frac{1}{2} \log r$ and $\zeta > \sqrt{2^{2\ell} \log^3 r / r\ell}$ there is a polynomial-time computable ϵ -extractor $\text{Ext} : (\{0, 1\}^\ell)^r \rightarrow \{0, 1\}^m$ for r total min-entropy rate $\delta > \zeta$ independent subsources over $\{0, 1\}^\ell$ where $m = (\delta - \zeta)r\ell$ and $\epsilon = \exp(-\Omega((\zeta^2 r\ell)/(2^{2\ell} \log^3 r)))$.*

Gabizon et al. also give a technique which improves the output length of extractors that extract only $\Omega(\log k)$ bits. We show that this technique also generalizes to total-rate independent sources, so we use it together with our extractor based on ideas from [KZ03] that extracts $\Omega(\log k)$ bits to get the following theorem. This theorem shows that even for polylogarithmic k , for small enough ℓ we can extract almost all of the min-entropy.

Theorem 1.8. *There exists a constant $C > 0$ such that for every $r \in \mathbb{N}$, $\ell \geq 1$, $k \geq (2^\ell \log r)^C$, there exists a polynomial-time computable ϵ -extractor $\text{Ext} : (\{0, 1\}^\ell)^r \rightarrow \{0, 1\}^m$ for r independent subsources over $\{0, 1\}^\ell$ with total min-entropy k , where $m = k - k^{1-\Omega(1)}$ and $\epsilon = k^{-\Omega(1)}$.*

Using the probabilistic method, we show that there exist (nonconstructive) extractors that extract even when the min-entropy k is as small as $O(\ell + \log r)$. Note that we always need $k > \ell$, since otherwise all

of the entropy could be in a single source, and thus extraction would be impossible. The extractor from [Theorem 1.8](#) comes closest to meeting this bound on k , but only works for small ℓ and has suboptimal error, so there is still much room for improvement.

1.3 Organization

In [Section 3](#) we describe our reduction from small-space sources to total-rate independent sources. We then restrict our focus to extracting from total-rate independent sources, starting with the basic extractors. In [Section 4](#) we describe the extractor that provides the basis for the extractor from [Theorem 1.6](#). In [Section 5](#) we describe the extractor that provides the basis for the extractor from [Theorem 1.5](#). In [Section 6](#) we describe the extractors that provide the basis for the extractors from [Theorem 1.7](#) and [Theorem 1.8](#). Then in [Section 7](#), we describe how to generalize the techniques of Gabizon et al. [[GRS04](#)] so that we can extract almost all of the entropy, and so achieve the theorems described in the introduction. Next, in [Section 8](#), we give nonconstructive results on extractors for both small-space and total-rate independent sources. Finally, in [Section 9](#), we give the improved extractor for our more restrictive model of space 1 sources.

2 Preliminaries

Notation: Given a string $x \in (\{0, 1\}^\ell)^r$ and a set $S \subseteq [r]$ we use x_S to denote the string obtained by restricting x to the indices in S . We use \circ to denote concatenation.

2.1 Convex Combinations

Definition 2.1. Let \mathcal{P} be a property of sources. Let X be some random variable over some universe. We will say that X is a convex combination of sources with property \mathcal{P} if there exists some random variable I over an arbitrary universe such that for all $i \in \text{supp}(I)$, $X|I = i$ has property \mathcal{P} .

A key observation that is essential to our results is that random variables that are convex combinations of sources with certain good properties are good themselves. This is captured in the following easy propositions:

Proposition 2.2. *Let X, Y be random variables such that X is a convex combination of sources that are ϵ -close to Y . Then X is ϵ -close to Y .*

Proposition 2.3. *Let X, I be random variables such that X is a convex combination of random variables $\{X_i\}_{i \in I}$. Let f be some function such that for all $i \in I$, $f(X_i)$ is a convex combination of sources that have some property \mathcal{P} . Then $f(X)$ is a convex combination of sources that have property \mathcal{P} .*

We'll also need the following simple lemma.

Lemma 2.4. *Let X, Y , and V be distributions over Ω such that X is ϵ -close to uniform and $Y = \gamma \cdot V + (1 - \gamma) \cdot X$. Then Y is $(\gamma + \epsilon)$ -close to uniform.*

Note that X and V could also be combinations of distributions, so this lemma also says that if Y is a convex combination of distributions that with high probability are close to uniform, then Y itself is also close to uniform.

Proof. Let U denote the uniform distribution on Ω and $S \subseteq \Omega$. Then

$$\begin{aligned} |\Pr[Y \in S] - \Pr[U \in S]| &= |\gamma \cdot \Pr[V \in S] + (1 - \gamma) \cdot \Pr[X \in S] - \Pr[U \in S]| \\ &\leq \gamma |\Pr[V \in S] - \Pr[X \in S]| + |\Pr[X \in S] - \Pr[U \in S]| \\ &\leq \gamma + \epsilon. \end{aligned}$$

□

2.2 Classes of Sources

We formally define the various classes of sources we will study.

Definition 2.5. A *space s source* X on $\{0, 1\}^n$ is a source generated by a width 2^s branching program. That is, the branching program is viewed as a layered graph with $n + 1$ layers with a single start vertex in the first layer and 2^s vertices in each subsequent layer. Each edge is labeled with a probability and a bit value. From a single vertex we can have multiple edges corresponding to the same output bit. The source is generated by taking a random walk starting from the start vertex and outputting the bit values on every edge.

This definition is very similar to the general Markov sources studied by Koenig and Maurer [KM04, KM05]. This is not quite the most general model of such sources imaginable, because we could consider sources that output a variable number of bits depending on which edge is chosen at each step, including possibly not outputting any bits. However, this restriction makes sense in light of the fact that we are primarily interested in sources of fixed length. In this case, the sources in the more general model can be transformed into our model by modifying the states appropriately.

The other important class of sources we study are independent sources.

Definition 2.6. A source consisting of r subsources on $\{0, 1\}^\ell$ is an *independent source* on $(\{0, 1\}^\ell)^r$ if each of the r subsources are independent. An independent source on $(\{0, 1\}^\ell)^r$ has total-rate δ if the total min-entropy over all of the sources is $\delta r \ell$.

Definition 2.7. A source on $\{0, 1\}^\ell$ is *flat* if it is uniformly distributed over a non-empty subset of $\{0, 1\}^\ell$.

Note that when $\ell = 1$, a flat independent source is the same as an oblivious bit-fixing source.

Definition 2.8. Let X be a random variable taking values in $\{0, 1\}^{t \times a}$, viewed as $t \times a$ matrices with entries in $\{0, 1\}$. We say that X on $(\{0, 1\}^a)^t$ is $(t \times a)$ *somewhere-random*² (*SR-source* for short) if it is a random variable on t rows of r bits each such that one of the rows of X is uniformly random. Every other row may depend on the random row in arbitrary ways. We will say that a collection X_1, \dots, X_m of $(t \times a)$ SR-sources is *aligned* if there is some i for which the i 'th row of each X_j is uniformly distributed.

We will also need a relaxed notion of the previous definition to where the “random” row is not completely random, but only has some min-entropy.

Definition 2.9. We say that a $(t \times a)$ source X on $(\{0, 1\}^a)^t$ has *somewhere-min-entropy* k , if X has min-entropy k in one of its t rows. We will say that a collection X_1, \dots, X_m of $(t \times a)$ somewhere-min-entropy k sources is *aligned* if there is some i for which the i 'th row of each X_j has min-entropy k .

²This definition is slightly different from the original one used by Ta-Shma [TS96]. The original definition considered the closure under convex combinations of the class defined here (i.e. convex combinations of sources that have one random row). We use this definition because we can do so without loss of generality and it considerably simplifies the presentation.

2.3 Seeded Extractors

We will also need to define what it means to have a seeded extractor for a given class of sources.

Definition 2.10. A polynomial-time computable function $\text{Ext} : \{0, 1\}^n \times \{0, 1\}^s \rightarrow \{0, 1\}^m$ is a *seeded ϵ -extractor* for a set of random sources \mathcal{X} , if for every $X \in \mathcal{X}$, $\text{Ext}(X, U_s)$ is ϵ -close to uniform. The extractor is called *strong* if for Y chosen according to U_s , $Y \circ \text{Ext}(X, Y)$ is also ϵ -close to uniform.

We use the following seeded extractor in our constructions, which allows us to get almost all the randomness out.

Theorem 2.11. [Tre01, RRV02] For every $n, k \in \mathbb{N}$, $\epsilon > 0$, there is a polynomial-time computable strong seeded ϵ -extractor $\text{Ext} : \{0, 1\}^n \times \{0, 1\}^t \rightarrow \{0, 1\}^{k - O(\log^3(n/\epsilon))}$ for sources with min-entropy k , with $t = O(\log^3(n/\epsilon))$.

3 Small-Space Sources As Convex Combinations Of Independent Sources

Here we show how small-space sources can be converted into convex combinations of independent sources. Thus we will be able to use our extractor constructions from subsequent sections to extract from small-space sources. The idea is simple: to extract from a space s source X , we divide the n bits in X into n/t blocks of size t . We view each block as a source on t bits. If we condition on the states of the source at the start of each block, all of the blocks become independent, so we end up with a set of n/t independent subsources on $\{0, 1\}^t$. We show, using techniques similar to Koenig and Maurer [KM04, KM05], that with high probability these sources will have sufficient min-entropy.

Lemma 3.1. Let X be a space s source on $\{0, 1\}^n$ with min-entropy rate δ . Then for any $0 < \alpha < 1$, X is $2^{-\alpha\delta n/2}$ -close to a convex combination of independent sources on $(\{0, 1\}^\ell)^r$ with total-rate δ' , where $\ell = 2s/(\alpha\delta)$, $r = \alpha\delta n/2s$ and $\delta' = (1 - \alpha)\delta$.

All of our extractors for small-space sources are obtained by combining Lemma 3.1 with the corresponding extractor for total-rate independent sources. We note that the reduction in this lemma is only interesting when the min-entropy rate $\delta > 1/\sqrt{n}$, since otherwise the total entropy of the independent sources would be less than the length of an individual source. In this case all of the entropy could be in a single source and thus extraction would be impossible.

To prove Lemma 3.1 we use the following standard lemma (for a direct proof see Lemma 5 in Maurer and Wolf [MW97], although it has been used implicitly earlier in, e.g., [WZ99]).

Lemma 3.2. Let X and Y be random variables and let \mathcal{Y} denote the range of Y . Then for all $\epsilon > 0$

$$\Pr_Y \left[H_\infty(X|Y = y) \geq H_\infty(X) - \log |\mathcal{Y}| - \log \left(\frac{1}{\epsilon} \right) \right] \geq 1 - \epsilon$$

Proof. (Of Lemma 3.1.) Divide X into $\alpha\delta n/2s$ blocks of size $2s/\alpha\delta$. Let Y represent the values of the initial states for each block. Then each $(X|Y = y)$ is a set of independent subsources with each block viewed as a subsorce of length $2s/(\alpha\delta)$. By Lemma 3.2, since $|\mathcal{Y}| = (2^s)^{(\alpha\delta n)/(2s)} = 2^{\alpha\delta n/2}$, with probability $1 - 2^{-\alpha\delta n/2}$ the sources $(X|Y = y)$ have min-entropy $(1 - \alpha)\delta n$ and thus min-entropy rate $(1 - \alpha)\delta$. \square

4 Extracting From Total-Rate Independent Sources By Reducing To Standard Independent Sources

In this section, we show how to construct extractors for total-rate independent sources using techniques from standard independent sources.

The following Markov-like lemma will be used to show that if we divide a source into blocks, many of the blocks will have a large entropy rate.

Lemma 4.1. *For any partition of a total-rate δ independent source on $(\{0, 1\}^\ell)^r$ into t blocks of r/t sub-sources each, the number b of blocks with min-entropy rate greater than $\delta/2$ satisfies $b > \delta t/2$.*

Therefore we can view this source as a set of t independent subsources on $\{0, 1\}^{\ell r/t}$ where at least $\delta t/2$ of the subsources have min-entropy rate greater than $\delta/2$.

Proof. We know that b blocks have min-entropy rate greater than $\delta/2$ and at most 1. In each of the remaining blocks the min-entropy rate is at most $\delta/2$. Since the total entropy rate is δ and min-entropies add for independent sources, $\delta \leq (b + (t - b)(\delta/2))/t$, which after a simple calculation gives the desired result. \square

Once we are in this model, we can generalize the result from Koenig and Maurer [KM04, KM05] that states that any two source extractor of the form $f(x_1 \cdot x_2)$, where the x_i are elements of some group, can be extended to any number of sources where only two of the sources have sufficient min-entropy.

Lemma 4.2. *Let $(\mathcal{G}, *)$ be a group and let $Ext(x_1, x_2, \dots, x_b) := f(x_1 * x_2 \cdots * x_b)$ be an extractor for b independent sources over \mathcal{G} , each of which has min-entropy rate at least δ . Then $F(x_1, \dots, x_r) := f(x_1 * \cdots * x_r)$ is an extractor for r independent sources over \mathcal{G} , b of which have min-entropy rate at least δ .*

The proof is essentially the same as in [KM04, KM05]. The key idea is that the r sources can be divided into b blocks, each of which contains exactly one of the high entropy sources.

Bourgain, Glibichuk, and Konyagin [BGK06] gave bounds on the exponential sums of the function $f(x_1, \dots, x_b) = \prod_{i=1}^b x_i$ over large subsets of fields without large subfields, in particular $GF(p)$ and $GF(2^p)$. As observed by Bourgain in [Bou05], this estimate gives an extractor for b independent sources where each source has high entropy. Bourgain only explicitly gives an extractor that outputs a single bit, but his result can be easily generalized using his techniques together with Vazirani's XOR lemma [Vaz86] to get the following.

Theorem 4.3. [BGK06] *Let the finite field K be either $GF(p)$ or $GF(2^p)$ for some prime p . Let $f(x_1, \dots, x_b) = \prod_{i=1}^b x_i$ and view the output of the function as an integer from 0 to $|K| - 1$. Then there exist functions $B(\delta)$ and $c(\delta)$ such that the function $BGK(x_1, \dots, x_b) = \lfloor (2^m f(x_1, \dots, x_b)) / |K| \rfloor$ (i.e. taking the m most significant bits of $f(x_1, \dots, x_b) / |K|$) is an ϵ -extractor for b independent min-entropy rate δ sources over K for $b \geq B(\delta)$, $m = \Theta(c(\delta) \log |K|)$, and $\epsilon = 2^{-\Omega(m)}$.*

Note that for constant δ , we can extract $\Theta(\log |K|)$ bits with only a constant number of sources. For $GF(p)$, [BGK06] make explicit the relationship between δ and the number of sources and entropy. It turns out in this case that we can handle slightly subconstant δ , down to $\delta = \Omega(1/(\log \log |K|)^{(1/C)})$ for some constant C . For $GF(2^p)$, it's not clear whether or not a similar result can be achieved.

Combining this theorem with Lemma 4.2, restricting the sources to be over the multiplicative group K^* , we get the following corollary.

Corollary 4.4. *Let the finite field K be either $GF(p)$ or $GF(2^p)$ for some prime p . Let $f(x_1, \dots, x_r) = \prod_{i=1}^r x_i$ and view the output of the function as a number from 0 to $|K| - 1$. Then there exist functions $B(\delta)$ and $c(\delta)$ such that the function $\text{BGK}(x_1, \dots, x_r) = \lfloor (2^m f(x_1, \dots, x_r)) / |K| \rfloor$ is an ϵ -extractor for r independent sources over K^* , at least $B(\delta)$ of which have min-entropy rate at least δ , and with $m = \Theta(c(\delta) \log |K|)$ and $\epsilon = 2^{-\Omega(m)}$.*

It will also be useful to formulate the following corollary.

Corollary 4.5. *For every constant $\delta > 0$, there exists a constant $v(\delta)$ and a polynomial time computable function $\text{BGK} : (\{0, 1\}^\ell)^r \rightarrow \{0, 1\}^m$ that is an ϵ -extractor for r independent sources on $(\{0, 1\}^\ell)^r$, such that at least $v(\delta)$ of the subsources have min-entropy rate δ where $m = \Omega(\ell)$ and $\epsilon = 2^{-\Omega(\ell)}$.*

Proof. Find the next smallest prime $p > \ell$ (we know $p \leq 2\ell$), and apply the extractor from [Corollary 4.4](#) over $GF(2^p)$, viewing each source as being embedded in $GF(2^p)^*$. \square

Now we can combine this extractor with [Lemma 4.1](#) to get an extractor for independent sources with constant total min-entropy rate.

Theorem 4.6. *For any constant δ , we can construct a polynomial-time computable ϵ -extractor $\text{Ext} : (\{0, 1\}^\ell)^r \rightarrow \{0, 1\}^m$ for total-rate δ independent sources on $(\{0, 1\}^\ell)^r$, with $m = \Theta(r\ell)$ and $\epsilon = 2^{-\Omega(m)}$. This extractor can be computed in time $\text{poly}(r, \ell)$.*

Proof. Given an independent source $X = X_1, \dots, X_n$ on $(\{0, 1\}^\ell)^r$, divide it into $t = 2B(\delta/2)/\delta$ blocks of r/t subsources each, where $B(\delta)$ is the constant from [Corollary 4.4](#). Then by [Lemma 4.1](#), we can view X as an independent sources on $(\{0, 1\}^{\ell r/t})^t$, where at least $\delta t/2 = B(\delta/2)$ of the subsources have min-entropy rate at least $\delta/2$. Find the smallest prime $p > (r\ell)/t$. By Bertrand's postulate, $p \leq 2(r\ell)/t$, we can find such a prime in time $\text{poly}(r, \ell)$ by brute force search. Then we can embed each of our subsources into $GF(2^p)^*$ and apply the extractor from [Corollary 4.4](#) to get the stated result. \square

5 Extracting From Polynomial Entropy Rate

In this section we will show how to extract from total-rate independent sources when the min-entropy of the sources is polynomially small. As in the previous section, we will reduce the problem to another model: we will try to extract from a few independent sources when just some of them have a polynomial amount of entropy, but we don't know exactly which ones. The probabilistic method shows that extractors exist for this model even when just two sources contain logarithmic min-entropy and the total number of sources is polynomially large. Our main theorem is as follows.

Theorem 5.1. *Assuming we can find primes with length in $[\tau, 2\tau]$ in time $\text{poly}(\tau)$, there is a constant β such that there exists a polynomial-time computable ϵ -extractor $\text{Ext} : (\{0, 1\}^\ell)^r \rightarrow \{0, 1\}^m$ for total-rate $\delta \geq \ell^{-\beta}$ independent sources on $(\{0, 1\}^\ell)^r$, where $n = \Theta(1/\delta^2)$, $m = \ell^{\Omega(1)}$ and $\epsilon = 2^{-\ell^{\Omega(1)}}$.*

We can also get the following corollary for when we have a larger number of smaller sources.

Corollary 5.2. *Assuming we can find primes with length in $[\tau, 2\tau]$ in time $\text{poly}(\tau)$, there exists a constant η such that for any $\delta \geq (r\ell)^{-\eta}$, there exists a polynomial-time computable ϵ -extractor $\text{Ext} : (\{0, 1\}^\ell)^r \rightarrow \{0, 1\}^m$ for total-rate δ independent sources on $(\{0, 1\}^\ell)^r$, where $m = (\delta^2 r \ell)^{\Omega(1)}$ and $\epsilon = 2^{-(\delta^2 r \ell)^{\Omega(1)}}$.*

Proof. Divide the source into $\Theta(1/\delta^2)$ blocks of $\Theta(\delta^2 n)$ subsources each and apply [Theorem 5.1](#). \square

In this section we will describe a generic technique to turn any extractor for the model where a few subsources have min-entropy rate less than half into an extractor that can extract when the min-entropy is as small as $\ell^{1-\alpha_0}$ for some universal constant α_0 . There are two major ingredients that will go into our construction:

- The first ingredient is based on recent constructions of randomness efficient condensers [BKS⁺05, Raz05]. We use these condensers to transform a set of sources with polynomial min-entropy rate into a set of aligned sources with somewhere-min-entropy rate 0.9. An important property that we will need is that the length of each of the rows is much higher than the number of rows. We prove the following theorem in Section 5.2.

Theorem 5.3. *Assume we can find primes with length in $[\tau, 2\tau]$ in time $\text{poly}(\tau)$. Let X_1, \dots, X_B all be sources on $\{0, 1\}^\ell$, for B a constant. Then for any small enough constant $\alpha > 0$ there exist constants $\gamma = \gamma(\alpha)$ and $\mu(\alpha) > 2\gamma$ and a polynomial time computable function $\text{ACond} : \{0, 1\}^\ell \rightarrow (\{0, 1\}^{\ell^\mu})^{\ell^\gamma}$ such that if each X_i has min-entropy rate $\delta = \ell^{-\alpha}$, then*

$$\text{ACond}(X_1), \text{ACond}(X_2), \dots, \text{ACond}(X_B)$$

is $2^{-\Omega(\ell^{1-2\alpha})}$ close to a convex combination of sets of aligned somewhere-min-entropy rate 0.9 sources.

- The second ingredient is the technique of condensing independent SR-sources from the work of Rao [Rao06]. We will generalize a theorem from that work. We show how to extract from independent sources with only a few of them being aligned SR-sources that have rows that are much longer than the number of rows. Formally, we get the following, proved in Section 5.3:

Theorem 5.4. *For every constant $\gamma < 1$ there exists a polynomial time $2^{-\ell^{\Omega(1)}}$ -extractor $\text{SRExt} : (\{0, 1\}^{\ell^{\gamma+1}})^u \rightarrow \{0, 1\}^m$ for u independent sources, of which v are independent aligned $(\ell^\gamma \times \ell)$ SR-sources, where $m = \ell - \ell^{\Omega(1)}$.*

We will first describe how to use these two ingredients to extract from an intermediate model. Then we will see that total-rate independent sources can be easily reduced to this intermediate model to prove Theorem 5.1.

5.1 Extracting From The Intermediate Model

The intermediate model we work with is defined as follows.

Definition 5.5. A (u, v, α) intermediate source X consists of u^2 subsources X^1, \dots, X^{u^2} , each on $\{0, 1\}^\ell$. These subsources are partitioned into u sets S_1, \dots, S_u such that v of the sets have the property that v of their subsources have min-entropy at least $\ell^{1-\alpha}$.

Now we show that for certain constant v and $\alpha > 0$ we can extract from this model.

Theorem 5.6. *Assuming we can find primes with length in $[\tau, 2\tau]$ in time $\text{poly}(\tau)$, for some constants v and $\alpha > 0$ there exists a polynomial time computable $2^{-\ell^{\Omega(1)}}$ -extractor IExt for (u, v, α) intermediate sources, where $m = \ell^{\Omega(1)}$.*

Using this theorem together with Lemma 4.1, we can prove Theorem 5.1.

Proof. (Of [Theorem 5.1](#).) Let $X = X_1, \dots, X_r$ be an independent source on $(\{0, 1\}^\ell)^r$ with total min-entropy rate $\delta \geq 4\ell^{-\alpha}$, where α is the constant from [Theorem 5.6](#) and $n = u^2$ where u will be chosen later. Divide the source into u blocks with u subsources each. By [Lemma 4.1](#), $\delta u/2$ of the blocks have min-entropy rate at least $\delta/2$. Now further divide each of the blocks into u sub-blocks of one subsource each. By [Lemma 4.1](#), for the blocks with min-entropy rate at least $\delta/2$, at least $\delta u/4$ of the sub-blocks have min-entropy rate $\delta/4 \geq \ell^{-\alpha}$. Let $u = 4v/\delta$, where v is the constant from [Theorem 5.6](#). Then X is a (u, v, α) intermediate source satisfying the conditions of [Theorem 5.6](#), which immediately gives us the theorem. \square

Here is the algorithm promised by [Theorem 5.6](#):

Construction: $\text{IExt}(x^1, \dots, x^{u^2})$

Input: x^1, \dots, x^{u^2} partitioned into sets S_1, \dots, S_u

Output: z .

Let v be a constant that we will pick later.

Let BGK be as in [Corollary 4.5](#) - an extractor for independent sources when $v-1$ of the subsources have min-entropy.

Let ACond be as in [Theorem 5.3](#), letting $B = v^2$ - a condenser that converts a set of sources with sublinear min-entropy into a convex combination of sets of aligned sources with somewhere-min-entropy rate 0.9.

Let SRExt be as in [Theorem 5.4](#) - an extractor for independent sources that works when just v of the inputs come from aligned SR-sources.

Set $\epsilon = 1/v^3$. Let α be a small enough constant to apply [Theorem 5.3](#) with α in the hypothesis. Let γ be as in the conclusion of the theorem.

1. Compute $y^i = \text{ACond}(x^i)$ for every source i in the input. Let y_j^i denote the j th row of y^i .
2. For every $l \in [u]$, and every $j \in [2^{\ell^\gamma}]$, let b_j^l be the string obtained by applying BGK using the y_j^i from all $i \in S_l$ as input.

We think of b^l as a sample from an SR-source with ℓ^γ rows, one for each seed s_i .

3. Output $\text{SRExt}(b^1, \dots, b^u)$.

Proof. (Of [Theorem 5.6](#))

If we restrict our attention to the v^2 high min-entropy subsources, from [Theorem 5.3](#) we know that from the first step from these subsources is $2^{-\Omega(\ell^{1-2\alpha})}$ close to a convex combination of sets of aligned somewhere-min-entropy rate 0.9 sources.

Then in the second step, for each distribution in the convex combination BGK succeeds in extracting from the aligned min-entropy rate 0.9 row in each set.

Remark 5.7. Actually, we don't really need the Bourgain-Glibichuk-Konyagin extractor for this step. If the min-entropy is so high, it is easy to see that the generalized inner product function is an extractor. Still we use BGK since we will need it later on in the construction.

Thus the result of the first step in the algorithm is a distribution that is $2^{-\ell^{\Omega(1)}}$ -close to a convex combination of collections of u independent subsources, v of which are independent aligned SR-sources.

Our extractor SRExt then extracts from each distribution in the convex combination, and thus extracts from the entire convex combination. So our algorithm succeeds in extracting from the input. \square

5.2 Condensing To Aligned Sources With High Somewhere-Min-Entropy

In this section we give the condenser from [Theorem 5.3](#). The first ingredient we'll need is the following condenser from [\[BKS⁺05\]](#).

Lemma 5.8. *[BKS⁺05] Assuming we can find primes with length in $[\tau, 2\tau]$ in time $\text{poly}(\tau)$, there exists a constant $\alpha > 0$ such that for any $t, \ell > 0$ there exists a polynomial-time computable condenser $\text{Zuck} : \{0, 1\}^\ell \rightarrow (\{0, 1\}^{(2/3)^t \ell})^{2^t}$ such that if X has rate δ , $\text{Zuck}(X)$ is $t2^{-\Omega(\alpha\delta\ell)}$ close to somewhere-min-entropy rate $(1 + \alpha)^t \delta$.*

We'll also need to use the condenser from Raz's work [\[Raz05\]](#) with the improved analysis of Dvir and Raz (Lemma 3.2 in [\[DR05\]](#)), which shows that most of the output rows are statistically close to having high min-entropy.

Lemma 5.9. *[DR05] For any constant $c > 0$, there is a polynomial-time computable function $\text{Raz} : (\{0, 1\}^\ell)^r \rightarrow (\{0, 1\}^{\Omega(\ell)})^{2^{\Omega(n)}}$ such that the following holds. If the input source X has somewhere-min-entropy rate δ , the output $\text{Raz}(X)$ is $2^{-\Omega(\delta\ell)}$ close to a convex combination of distributions, each of which has the property that at least a $(1 - c)$ fraction of its rows have min-entropy rate at least 0.9δ .*

Now we can apply the functions from the previous two lemmas in succession to transform any source with min-entropy rate δ into a convex combination of sources with high somewhere-min-entropy sources where almost all of the rows in the sources have high min-entropy.

Lemma 5.10. *Assuming we can find primes with length in $[\tau, 2\tau]$ in time $\text{poly}(\tau)$, there exists a constant $\alpha > 0$ such that for any constants $t > 0$ and $c > 0$ there exists a polynomial-time computable function $\text{Cond} : \{0, 1\}^\ell \rightarrow (\{0, 1\}^{\Omega(\ell/3^t)})^{2^{\Omega(4^t)}}$ such that the following holds. If the input source X has min-entropy rate δ , the output $\text{Cond}(X)$ is $2^{-\Omega(\delta^2\ell)}$ close to a convex combination of distributions, each of which has the property that at least a $(1 - c)$ fraction of its rows have min-entropy rate at least $0.9\delta(1 + \alpha)^t$.*

Proof. Let $\text{Cond}(x) = \text{Raz}(\text{Zuck}(x))$. □

Corollary 5.11. *Assuming we can find primes with length in $[\tau, 2\tau]$ in time $\text{poly}(\tau)$, there is a constant C such that for any constant $c > 0$ there exists a polynomial-time computable function $\text{Cond} : \{0, 1\}^\ell \rightarrow (\{0, 1\}^{\Omega(\ell)})^C$ such that the following holds. If the input source X has min-entropy rate δ , then the output $\text{Cond}(X)$ is $2^{-\Omega(\delta^2\ell)}$ close to a convex combination of distributions where each source in the convex combination has the property that at least a $(1 - c)$ fraction of its rows have min-entropy rate at least 2δ .*

Proof. Pick t large enough (but still constant) in [Lemma 5.10](#) so that $0.9(1 + \alpha)^t \geq 2$. Then $C = 2^{\Omega(4^t)}$. □

Now we can use this basic condenser to help prove [Theorem 5.3](#). To do this, we apply this condenser to our input subsources and then recursively apply it to the outputs. We might think we could just apply the union bound to show that most of the output rows are aligned, but that is not true. However, we only need that one single row in the output is aligned, which we can accomplish by ensuring that at each step we have an aligned row, and then concentrating recursively on that aligned row.

Proof. (Of [Theorem 5.3](#).) First, apply the function Cond from [Corollary 5.11](#) to each X_i , choosing $c < \frac{1}{B}$. Then the output is $2^{-\Omega(\delta^2\ell)}$ close to a convex combination of distributions where each source in the convex combination has the property that at least a $(1 - c)$ fraction of its rows have min-entropy rate at least 2δ . Now we restrict our attention to a single source in the convex combination. In this source at most $cB < 1$

fraction of the rows have a subsource $\text{Cond}(X_i)$ with min-entropy rate less than 2δ in that row. Thus there is at least one row where the min-entropy rate for all the subsources is at least 2δ , i.e., the output is aligned with somewhere-min-entropy rate 2δ . Now we recursively apply Cond to each row in each output source. When we apply it to the aligned row, we'll get another aligned row with min-entropy rate 4δ . If we recursively do this t times, we end up close to a convex combination of a set of aligned sources with somewhere-min-entropy rate $2^t\delta$. If we let $t = \log(.9/\delta) = \log(.9\ell^\alpha)$, then these sources have somewhere-min-entropy rate 0.9. If we choose α small enough (depending on the constants in [Corollary 5.11](#)), then we can achieve $\mu > 2\gamma$, as desired. \square

5.3 Extracting From Independent Sources, A Few Of Which Are Aligned SR-Sources

Here we will prove [Theorem 5.4](#). Our extractor will be obtained by condensing the aligned SR-sources, closely following a similar construction of Rao [[Rao06](#)]. The additional complication is that whereas in [[Rao06](#)] every source was assumed to have a random row, in our model only some of the sources contain a random row and the rest may be arbitrary. We will build a condenser that when given u independent sources, v of which are aligned SR-sources, outputs a distribution that is statistically close to a convex combination of sources of the same type, with far fewer rows in each SR-source. Our condenser can handle an arbitrarily large u and some small universal constant v .

Iterating our condenser, we will eventually obtain just one row in our SR-sources, at which point we can use BGK from [Corollary 4.5](#) to extract from the sources, or even simply XOR all the sources together.

To condense a single source from the input, we will take a small slice of bits from all other sources in the input. We will use these slices to generate a short list of candidate seeds that are independent of the source we are trying to condense. Then we will use these seeds with a strong seeded extractor to extract from the source we are trying to condense. In this way we reduce the number of rows of one source.

To condense all of the sources, we repeat the same construction with all sources: each source is condensed using seeds generated from slices of the other sources. The output of all this condensing is u sources that are no longer independent. Still, we will argue that if we fix all the slices of bits we used to generate the seeds, the output is the distribution of independent sources of the type that we want.

Remark 5.12. Although we do not include the details here, it is not hard to modify the construction in this subsection to extract even when $v = 2$ and u is arbitrarily large, by replacing the function BGK from [Corollary 4.5](#) in the composition below with a generalization of Bourgain's extractor [[Bou05](#)]. We can also show that our construction is *strong*, i.e. the output of our extractor is statistically close to being independent of any one source from the input.

Now we describe our condenser in detail.

Construction: $\text{Cond}(x^1, \dots, x^u)$

Input: x^1, \dots, x^u , strings each divided into t rows of length r .

Output: z^1, \dots, z^u .

Let w, l be parameters that we will set later.

Let BGK be as in [Corollary 4.5](#) - an extractor for independent sources when $v - 1$ of them have min-entropy rate 0.2. Let a be the output length of BGK. Let ϵ_1 be the error of BGK.

Let Ext be the strong seeded extractor promised by [Theorem 2.11](#). We will set up Ext to extract from sources on $\{0, 1\}^{ta}$ with min-entropy at least $a - l$ and to have output length m , using seed length a . Let ϵ_2 be the error of Ext.

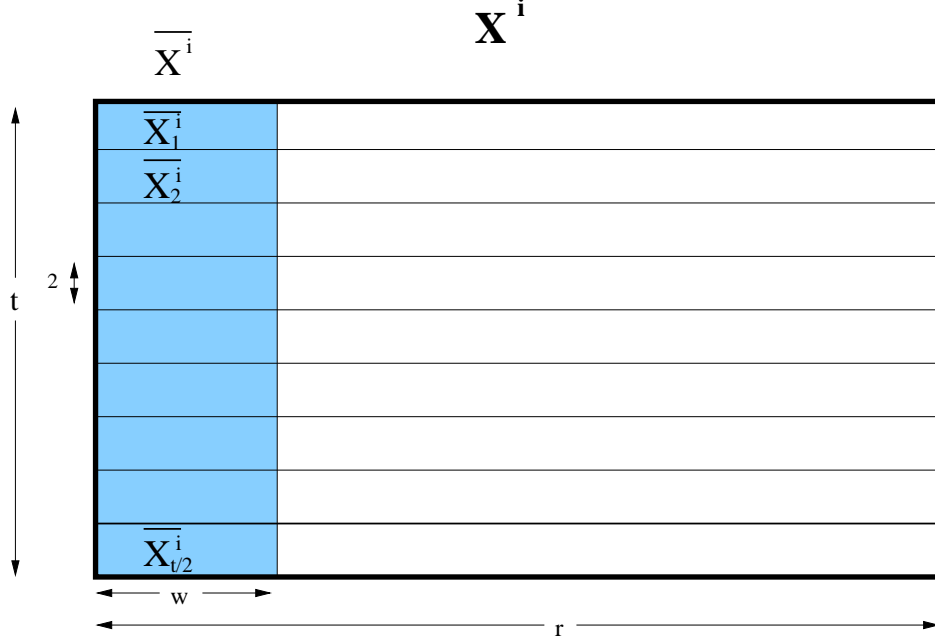


Figure 2: Notation in one source

1. For each source, partition its rows into pairs of rows.
2. For $i = 1, 2, \dots, u$ and $j = 1, 2, \dots, t/2$ let \bar{x}_j^i denote the first w bits of the j 'th pair of rows in the string x^i . Let \bar{x}^i denote the first w bits of every row of x^i . Let $\bar{x}_j^{\neq i}$ denote the concatenation of the first w bits of the j 'th pair rows of all sources except the i 'th source.
3. For every $i = 1, 2, \dots, u$, and $j = 1, 2, \dots, t/2$, let $z_j^i = \text{Ext}(x^i, \text{BGK}(x_j^{\neq i}))$.
4. For every $i = 1, 2, \dots, u$, let z^i be the concatenation of $z_1^i \circ \dots \circ z_{t/2}^i$

Lemma 5.13. *Let Cond be as above. If X^1, X^2, \dots, X^u are independent sources, with v of them being aligned $(t \times a)$ SR-sources, then Z^1, Z^2, \dots, Z^u are $v(\epsilon_1 + 2\sqrt{\epsilon_2} + 2^{-(l-tw)})$ -close to a convex combination of independent sources, v of which are aligned $(t/2 \times m)$ SR-sources.*

Proof. Let h be such that the h 'th pair of rows in X^1, \dots, X^u contains a random row. We will argue that the h 'th row of the output distribution is statistically close to uniform.

To see this, consider the random variable $\bar{X} = \bar{X}^1 \circ \dots \circ \bar{X}^u$, the concatenation of all the slices that are used to generate the various seeds.

We will partition the support of this variable into two sets, a *good* set and a *bad* set. We will then make the following two claims, which clearly imply the lemma.

Claim 5.14. *For good \bar{x} , $(Z^1 \circ \dots \circ Z^u) | \bar{X} = \bar{x}$ is the distribution of u independent sources, with v of them being $v\sqrt{\epsilon_2}$ close to aligned SR-sources.*

Claim 5.15. $\Pr[\bar{X} \text{ is not good}] < v\epsilon_1 + v\sqrt{\epsilon_2} + v2^{tw-l}$

To ensure these claims, the notion of good we will use is this one: call \bar{x} *good for source X^i* if

1. $X^i|\bar{X}=\bar{x}$ has min-entropy at least $r - l$
2. $\text{BGK}(\bar{x}_h^{\neq i})$ is a good seed to extract from $X^i|\bar{X}=\bar{x}$, i.e.

$$\| \text{Ext}(X^i|\bar{X}=\bar{x}, \text{BGK}(\bar{x}_h^{\neq i})) - U_m \| < \sqrt{\epsilon_2}$$

We will say that \bar{x} is *good* if it is good for all the v sources that contain a random row. [Claim 5.14](#) immediately follows from this notion of good. All we have left to prove is [Claim 5.15](#). The proof requires the following simple proposition.

Proposition 5.16. *Let X be a random variable with $H_\infty(X) = k$. Let A be any event in the same probability space. Then $H_\infty(X|A) < k' \Rightarrow \Pr[A] < 2^{k'-k}$.*

Proof. (of [Claim 5.15](#)) Fix an i so that X^i is one of the v aligned SR-sources that contains a random row. We will first argue that \bar{X} is good for X^i with high probability. Then we will use the union bound to claim that it is good with high probability.

\bar{X} is good for X^i when two events occur:

1. Event T : $X^i|\bar{X}=\bar{x}$ has min-entropy at least $a - l$. This event is equivalent to the event $X^i|\bar{X}^i=\bar{x}^i$ has min-entropy at least $a - l$, since X^i only depends on those bits of \bar{X} .
2. Event U : $\text{BGK}(\bar{x}_h^{\neq i})$ is a good seed to extract from $X^i|\bar{X}=\bar{x}$, i.e.

$$\| \text{Ext}(X^i|\bar{X}=\bar{x}, \text{BGK}(\bar{x}_h^{\neq i})) - U_m \| < \sqrt{\epsilon_2}$$

By [Proposition 5.16](#), the probability that event T does not occur is at most $2^{-l}2^{tw}$. This is because there are 2^{tw} possible settings for \bar{x}^i . Every bad setting occurs with probability at most 2^{-l} , thus by the union bound, the probability that any bad setting occurs is at most 2^{tw-l} .

Now given that T does occur, event U does not occur with probability at most $\sqrt{\epsilon_2} + \epsilon_1$. This is because the output of BGK is ϵ_1 -close to uniform and for a uniformly chosen seed the probability that Ext fails to extract from the source is at most $\sqrt{\epsilon_2}$ by the strong extractor property and Markov's inequality.

Thus by the union bound, the probability that either T or U do not occur is at most $2^{tw-l} + \sqrt{\epsilon_2} + \epsilon_1$.

Applying the union bound again, \bar{X} is good for all the X^i 's we care about with probability at least $1 - v(2^{tw-l} + \sqrt{\epsilon_2} + \epsilon_1)$. □

This concludes the proof of the lemma. □

Now we can prove the main theorem of this section.

Proof. (of [Theorem 5.4](#))

We will use the condenser Cond repeatedly. In each step we reduce the number of rows in each of the sources by a factor of 2. We need to repeat the condensation step at most $\lceil \gamma \log \ell \rceil$ times. By [Lemma 5.13](#) the error in each step is $v(\epsilon_1 + 2\sqrt{\epsilon_2} + 2^{-(l-tw)})$.

Recall that ϵ_1 is the error of BGK from [Corollary 4.5](#). Thus $\epsilon_1 = 2^{-\Omega(w)}$ in every step, since w is the length of the inputs to BGK. ϵ_2 was the error of Ext from [Theorem 2.11](#). Since the seed length $a = \Omega(w)$, the error ϵ_2 is at most $2^{-w^{\Omega(1)}}$ in every step.

Setting $l = 2\ell^{(1+\gamma)/2}$, $w = l/(2t) = \ell^{\Omega(1)}$, we get a total error of $2^{-\ell^{\Omega(1)}}$.

In each step we the length of the sources drops by $\ell^{\beta'}$ for some small β' . Thus the final output length is at least $\ell - \ell^\beta$ for some $\beta \in (0, 1)$. □

6 Better Extractors For Total-Rate Independent Sources With Many Short Subsources

Now we show how for sources consisting of many subsources of length ℓ we can do better than the constructions in the previous sections by generalizing earlier constructions for symbol-fixing sources. The base extractor simply takes the sum of the subsources modulo p for some prime $p > 2^\ell$. Then we divide the source into blocks, apply the base extractor to each block, and then use the result to take a random walk on an expander as in [KZ03].

We will need the following definition from [KZ03].

Definition 6.1. An independent source on $(\{0, 1\}^\ell)^r$ is a (k, ϵ) -approximate symbol-fixing source if k of the r subsources have distributions within an ℓ_2 distance ϵ of uniform.

These sources will be used as intermediate sources. We will transform the sources we wish to extract from into approximate symbol-fixing sources and then use the results of [KZ03] to extract from these sources.

6.1 Random Walks

Let $\lambda(P)$ be the second largest eigenvalue in absolute value of the transition matrix P for a random walk on a graph G . It is well known that the ℓ_2 distance from the uniform distribution decreases by a factor of $\lambda(P)$ for each uniform step of the random walk (see e.g. [Lov96]).

We will also need the following Lemma from [KZ03], which shows that we can use a random walk to extract from approximate symbol-fixing sources.

Lemma 6.2. [KZ03] *Let G be an undirected non-bipartite d -regular graph on M vertices with uniform transition matrix P . Suppose we take a walk on G for r steps, with the steps taken according to the symbols from a (k, ϵ) -approximate oblivious symbol-fixing sources on $[d]^r$. For any initial probability distribution, the variation distance from uniform at the end of the walk is at most $\frac{1}{2}(\lambda(P) + \epsilon\sqrt{d})^k \sqrt{M}$.*

Note that if $\lambda(P) + \epsilon\sqrt{d}$ is bounded above by a constant, as would happen if G were an expander and ϵ was small enough, then this immediately gives us a good extractor for approximate symbol-fixing sources. This is shown in the following proposition, which follows immediately from Lemma 6.2.

Proposition 6.3. *Let G be an undirected non-bipartite d -regular graph on 2^m vertices with uniform transition matrix P . Then we can construct a polynomial-time computable ϵ' -extractor for the set of (k, ϵ) -approximate oblivious symbol-fixing sources on $[d]^r$, where $\epsilon' = \frac{1}{2}(\lambda(P) + \epsilon\sqrt{d})^k 2^{m/2}$. This extractor simply uses the input from the source to take a random walk on G and outputs the label of the final vertex.*

6.2 Reducing to Flat Total-Rate Independent Sources

It will be simpler to analyze our extractor for flat total-rate independent sources. We show that any extractor that works for flat total-rate independent sources also works for general total-rate independent sources because any total-rate independent source is close to a convex combination of flat independent sources with high total-rate.

Lemma 6.4. *Any ϵ -extractor for the set of flat independent sources on $(\{0, 1\}^\ell)^r$ with total min-entropy $k/(2 \log 3)$ is also an $(\epsilon + e^{-k/9})$ -extractor for the set of independent sources on $(\{0, 1\}^\ell)^r$ with min-entropy k .*

This lemma follows directly from the following lemma.

Lemma 6.5. *Any independent source $X = X_1, \dots, X_r$ on $(\{0, 1\}^\ell)^r$ with total min-entropy k is $e^{-k/9}$ -close to a convex combination of flat independent sources on $(\{0, 1\}^\ell)^r$ with total min-entropy $k/(2 \log 3)$.*

Proof. Let $H_\infty(X_i) = k_i$ for all i . If $k_i \geq 1$, we can write X_i as a convex combination of flat sources with support size $\lfloor 2^{k_i} \rfloor$. Each of these flat sources has min-entropy $\log \lfloor 2^{k_i} \rfloor > \frac{k_i}{\log 3}$, since we lose the largest fraction of min-entropy from taking the floor when 2^{k_i} is nearly 3.

If $k_i < 1$, then we must have constant sources in our convex combination, so if we did as above, we'd lose up to a bit of entropy for each such i . Instead, suppose k' of the total entropy is contained in X_i with less than a bit of entropy each. Call this set $S \subseteq [r]$. Now suppose $k' \leq k/2$. In this case, we can write X_S as a convex combination of constant sources and we are still left with $(k - k')/\log 3 \geq k/(2 \log 3)$ bits of entropy in each of our sources, as desired.

From now on we will assume $k' \geq k/2$. We will show we can write X_S as a convex combination of sources that with probability $1 - \epsilon$ have min-entropy $k'/3$. For each $i \in S$, we can write X_i as a convex combination of flat sources with one or zero bits of entropy. The one bit sources are obtained by choosing uniformly between the most probable value and each of the other values for X_i . Each of these sources occurs with probability equal to twice the probability of the less probable value. Since the most probable value occurs with probability 2^{-k_i} , we get one bit of entropy with probability $2(1 - 2^{-k_i})$. Otherwise, X_i is fixed to the most probable value.

Now we can use a Chernoff bound to bound the entropy in the sources in the overall convex combination of sources for X_S . Let Y_i be an indicator random variable for the i th source having one bit of entropy. Then $Y = \sum Y_i$ is a random variable representing the total entropy. Note that $\mathbb{E}[Y] = \sum \mathbb{E}[Y_i] = \sum 2(1 - 2^{-k_i}) \geq \sum k_i = k'$, where the inequality is true because $k_i < 1$. Now we are ready to apply the Chernoff bound (Theorem A.1.13 in Alon and Spencer [AS00]).

$$\Pr[Y < (1 - \lambda)k'] \leq \Pr[Y < (1 - \lambda)\mathbb{E}[Y]] < e^{-\lambda^2(\sum(1-2^{-k_i}))} \leq e^{-\lambda^2 \frac{k'}{2}} \leq e^{-\lambda^2 \frac{k}{4}}$$

Setting $\lambda = 2/3$ we get the desired error bound $\epsilon = e^{-\frac{k}{9}}$. Then with probability $1 - \epsilon$ we have at least $(k - k')/\log 3 + k'/3 \geq k/(2 \log 3)$ bits of entropy, as desired. \square

6.3 Extracting From Flat Total-Rate Independent Sources

Now we show how to extract from flat total-rate independent sources for small ℓ . Our initial extractor simply takes the sum modulo p of the individual sources, for some prime $p \geq 2^\ell$

Theorem 6.6. Let $\ell \geq 1$ and $p \geq 2^\ell$ a prime. Then $\text{Sum}_p : (\{0, 1\}^\ell)^r \rightarrow [p]$, where $\text{Sum}_p(x) = \sum_i x_i \pmod p$, is an ϵ -extractor for the set of flat independent sources on $(\{0, 1\}^\ell)^r$ with total min-entropy k , where $\epsilon = \frac{1}{2} 2^{-2k/p^2} \sqrt{p}$.

Combining [Theorem 6.6](#) with [Lemma 6.4](#) we get an extractor for total-rate independent sources.

Corollary 6.7. Suppose $p \geq 2^\ell$ is a prime. Then Sum_p is an ϵ -extractor for the set of independent sources on $(\{0, 1\}^\ell)^r$ with total min-entropy $k \geq \Omega(p^2 \log p)$, where $\epsilon = 2^{-\Omega(k/p^2)}$.

We will prove [Theorem 6.6](#) via the following lemma, which will be useful later.

Lemma 6.8. Let $\ell \geq 1$ and $p \geq 2^\ell$ a prime. Then for all sets of flat independent sources $X = X_1, \dots, X_r$ on $(\{0, 1\}^\ell)^r$ with min-entropy k , $\text{Sum}_p(x)$ has ℓ_2 distance from uniform at most $2^{-2k/p^2}$.

It is well known that if X and Y are both distributed over a universe of size p , then $|X - Y| \leq \frac{1}{2} \sqrt{p} \|X - Y\|_2$. [Theorem 6.6](#) then follows by combining this lemma with this relation between ℓ_2 and variation distance.

To analyze the distance from uniform of the sum modulo p , we use the following lemma that relates this distance to the additive characters of \mathbb{Z}_p . For \mathbb{Z}_p , where p is a prime, the i th additive character is defined as $\chi_j(a) = e^{\frac{2\pi i j a}{p}}$.

Lemma 6.9. For any function $f : \{0, 1\}^r \rightarrow \mathbb{Z}_p$ and random variable X over $\{0, 1\}^r$,

$$\|f(X) - U_p\|_2^2 = \frac{1}{p} \sum_{j=1}^{p-1} |\mathbb{E}[\chi_j(f(X))]|^2 < \max_{j \neq 0} |\mathbb{E}[\chi_j(f(X))]|^2,$$

where U_p denotes the uniform distribution over \mathbb{Z}_p .

Proof. Let $Y = f(X) - U_p$. The j th Fourier coefficient of Y is given by $\hat{Y}_j = \sum_{y=0}^{p-1} Y(y) \chi_j(y)$. By Parseval's Identity and using the fact that $\sum_{y=0}^{p-1} \chi_j(y) = 0$ when $j \neq 0$ we get

$$\begin{aligned} \|Y\|_2^2 &= \frac{1}{p} \sum_{j=0}^{p-1} |\hat{Y}_j|^2 = \frac{1}{p} \sum_{j=0}^{p-1} \left| \sum_{y=0}^{p-1} Y(y) \chi_j(y) \right|^2 \\ &= \frac{1}{p} \sum_{j=0}^{p-1} \left| \sum_{y=0}^{p-1} \Pr[f(X) = y] \chi_j(y) - \frac{1}{p} \sum_{y=0}^{p-1} \chi_j(y) \right|^2 \\ &= \frac{1}{p} \sum_{j=1}^{p-1} |\mathbb{E}[\chi_j(f(X))]|^2 \\ &< \max_{j \neq 0} |\mathbb{E}[\chi_j(f(X))]|^2. \end{aligned}$$

□

Using the previous lemma we can now prove [Theorem 6.6](#).

Proof. Let $f(X) = \sum_{i=1}^r X_i$ and fix $j \neq 0$. Then $|\mathbb{E}[\chi_j(f(X))]|^2 = \prod_{i=1}^r |\mathbb{E}[\chi_j(X_i)]|^2$. Suppose X_i has min-entropy k_i , so $k = \sum_i k_i$. Then since each X_i is a flat source, X_i is uniformly distributed over $K_i = 2^{k_i}$ values. Our goal is to upper bound $|\mathbb{E}[\chi_j(X_i)]|^2$ over all possible choices of X_i . Doing so, we get

$$\begin{aligned}
|\mathbb{E}[\chi_j(X_i)]|^2 &\leq \max_{X_i: \mathbb{Z}_p \rightarrow \{0, 1/K_i\}, \sum_x X_i(x)=1} |\mathbb{E}[\chi_j(X_i)]|^2 \\
&= \max_{X_i: \mathbb{Z}_p \rightarrow \{0, 1/K_i\}, \sum_x X_i(x)=1} \left| \sum_{x \in \mathbb{Z}_p} X_i(x) \chi_j(x) \right|^2 \\
&= \max_{y, |y|=1} \left(\max_{X_i: \mathbb{Z}_p \rightarrow \{0, 1/K_i\}, \sum_x X_i(x)=1} \left(\left(\sum_{x \in \mathbb{Z}_p} X_i(x) \chi_j(x) \right) \odot y \right)^2 \right) \\
&= \max_{X_i: \mathbb{Z}_p \rightarrow \{0, 1/K_i\}, \sum_x X_i(x)=1} \left(\max_{y, |y|=1} \left(\sum_{x \in \mathbb{Z}_p} X_i(x) (\chi_j(x) \odot y) \right)^2 \right),
\end{aligned}$$

where \odot denotes the complex dot product, where the complex numbers are viewed as two dimensional vectors, and the third line follows from the observation that the dot product is maximized when y is in the same direction as $(\sum_{x \in \mathbb{Z}_p} X_i(x) \chi_j(x))$, in which case we get exactly the square of the length. Now we further note that $\chi_j(x) \odot y$ is greatest for values of x for which $\chi_j(x)$ is closest to y . Thus we achieve the maximum when X_i is distributed over the K_i values closest to y . Without loss of generality we can assume these values correspond to $x = 0$ to $K_i - 1$ (since we only care about the magnitude). Thus

$$\begin{aligned}
|\mathbb{E}[\chi_j(X_i)]|^2 &\leq \left| \frac{1}{K_i} \sum_{j=0}^{K_i-1} e^{\frac{2\pi i j}{p}} \right|^2 \\
&= \left| \frac{1}{K_i} \frac{1 - e^{\frac{2\pi i j K_i}{p}}}{1 - e^{\frac{2\pi i}{p}}} \right|^2 \\
&= \left| \frac{1}{K_i} \frac{e^{\frac{\pi i K_i}{p}} (e^{-\frac{\pi i K_i}{p}} + e^{\frac{\pi i K_i}{p}})}{e^{\frac{\pi i}{p}} (e^{-\frac{\pi i}{p}} + e^{\frac{\pi i}{p}})} \right|^2 \\
&= \left(\frac{1}{K_i} \frac{\sin(\frac{\pi K_i}{p})}{\sin(\frac{\pi}{p})} \right)^2 \\
&= \left(\frac{1}{K_i} \frac{\frac{\pi K_i}{p} \prod_{m=1}^{\infty} (1 - \frac{K_i^2}{p^2 m^2})}{\frac{\pi}{p} \prod_{m=1}^{\infty} (1 - \frac{1}{p^2 m^2})} \right)^2 \\
&= \left(\prod_{m=1}^{\infty} \left(1 - \frac{K_i^2 - 1}{p^2 m^2 - 1} \right) \right)^2 \\
&< \left(1 - \frac{K_i^2 - 1}{p^2 - 1} \right)^2 \\
&< e^{-2(K_i^2 - 1)(p^2 - 1)},
\end{aligned}$$

where in the fifth line we use the infinite product representation of sine.

So

$$\begin{aligned} |\mathbb{E}[\chi_j(f(X))]|^2 &= \prod_{i=1}^r |\mathbb{E}[\chi_j(X_i)]|^2 \\ &< \prod_{i=1}^r e^{-2(K_i^2-1)/(p^2-1)} \\ &< e^{2r/p^2} e^{-2(\sum_i K_i^2)/p^2}. \end{aligned}$$

By the power mean inequality, $\sum_{i=1}^r K_i^2 \geq r \cdot (\prod_{i=1}^r K_i)^{2/r} = r2^{2k/r}$. Thus

$$|\mathbb{E}[\chi_j(f(X))]|^2 < e^{-\frac{2r(2^{2k/r}-1)}{p^2}}$$

Let $k = \delta r$. Then this quantity is $e^{-(2k/p^2)((2^{2\delta}-1)/\delta)}$. Since $(2^{2\delta}-1)/\delta$ is an increasing function of δ and goes to $2 \ln 2$ as δ goes to 0, we have

$$|\mathbb{E}[\chi_j(f(X))]|^2 < e^{-(2k/p^2)((2^{2\delta}-1)/\delta)} < e^{-4(\ln 2)k/p^2} = 2^{-4\frac{k}{p^2}}$$

Then since by [Lemma 6.9](#) $\|f(X) - U_p\|_2^2 < \max_{j \neq 0} |\mathbb{E}[\chi_j(f(X))]|^2$, $\|f(X) - U_p\|_2 < 2^{-2k/p^2}$. \square

Now we show that if we divide the source into blocks and take the sum modulo p for each block, we get a convex combination of approximate symbol-fixing sources, which we can then use an expander walk to extract from.

Lemma 6.10. *For any prime $p \geq 2^\ell$ and any t , any flat independent source X on $(\{0, 1\}^\ell)^r$ with total min-entropy k can be transformed in polynomial-time into a $(k', 1/p)$ -approximate oblivious symbol-fixing source $f(X)$ on $[p]^{r'}$, where $r' = k/(2p^2 \log p)$ and $k' = k^2/(4np^2 \log^2 p)$.*

Proof. First divide X into $\frac{k}{2t}$ blocks consisting of $\frac{2t}{k}r$ subsources, for $t = p^2 \log p$. Then for each block take the sum modulo p of the subsources in the block. Then $f(X)$ is the concatenation of the resulting symbols for each block.

By [Lemma 4.1](#), the number of blocks with min-entropy at least t is greater than $\frac{k^2}{4tr\ell} > \frac{k^2}{4tr \log p}$. For each of these blocks, by [Lemma 6.8](#), we mix within $2^{-t/p^2} = \frac{1}{p}$ of uniform. \square

Now, as in [\[KZ03\]](#), we use $f(X)$ as defined above to take a random walk on an expander graph, which will mix to uniform by [Lemma 6.2](#) and thus give us our extractor.

Theorem 6.11. *There exists an ϵ -extractor for the set of flat independent sources on $(\{0, 1\}^\ell)^r$ with total min-entropy k that outputs $m = \Omega(k^2/(r2^{2\ell}\ell))$ bits and has error $\epsilon = 2^{-m}$. This extractor is computable in time $\text{poly}(r, 2^\ell)$.*

Proof. Let p be the least prime greater than 2^ℓ . Since by Bertrand's Postulate $p < 2 \cdot 2^\ell$, this can easily be done in polynomial time in 2^ℓ by exhaustive search. Given a source X , first apply $f(X)$ from [Lemma 6.10](#) to get a $(k', 1/p)$ -approximate oblivious symbol-fixing source on $[p]^{r'}$, where $r' = k/(2p^2 \log p)$ and $k' = k^2/(4rp^2 \log^2 p)$. Then apply the extractor from [Proposition 6.3](#) to $f(X)$, taking the graph G to be a p regular expander graph on 2^m vertices (for m to be given later). Specifically, assume G has $\lambda(G) \leq \frac{1}{p^\alpha} - \frac{1}{\sqrt{p}}$

for some constant $\alpha < 1/2$. This can be achieved, for example, by taking G to be an $O(\log p)$ power of a constant degree expander with self loops added to make it degree p . Then by [Proposition 6.3](#) $f(X)$ is within

$$\begin{aligned} \epsilon &\leq \frac{1}{2} \left(\lambda(G) + \frac{1}{\sqrt{p}} \right)^{(k^2/4rp^2 \log^2 p)} 2^{m/2} \\ &< p^{-(\alpha k^2/4rp^2 \log^2 p)} 2^{m/2} \\ &= 2^{-((\alpha k^2/4rp^2 \log p) - (m/2))} \end{aligned}$$

of uniform. Then let $m = \alpha k^2/6rp^2 \log p$ so then $\epsilon < 2^{-m}$. \square

Combining this theorem with our reduction from general to flat sources, we get that this same extractor works for general total-rate independent sources.

Theorem 6.12. *There exists an ϵ -extractor for the set of independent sources on $(\{0, 1\}^\ell)^r$ with total min-entropy k that outputs $m = \Omega(k^2/r2^{2\ell})$ bits and has error $\epsilon = 2^{-m}$. This extractor is computable in time $\text{poly}(r, 2^\ell)$.*

Proof. Combine [Theorem 6.11](#) and [Lemma 6.4](#). \square

7 Extracting More Bits From Total-Rate Independent Sources

7.1 Seed Obtainers

Now that we have extractors for total-rate independent sources, we can extract even more bits using the techniques that Gabizon et al. [[GRS04](#)] used to extract more bits out of oblivious bit-fixing sources. Assuming the entropy is high enough to use the extractors from [Theorem 6.12](#), [Theorem 4.6](#), or [Corollary 5.2](#), we can extract almost all of the entropy. Their construction works by using an extractor for bit-fixing sources and a sampler to construct a seed obtainer. This seed obtainer outputs a source and a seed that is close to a convex combination of independent bit-fixing sources and uniform seeds. We generalize their definition of seed obtainer to total-rate independent sources.

Definition 7.1. A function $F : (\{0, 1\}^\ell)^r \rightarrow (\{0, 1\}^\ell)^r \times \{0, 1\}^d$ is a (k', ρ) -seed obtainer for all independent sources X on $(\{0, 1\}^\ell)^r$ with total min-entropy k if the distribution $R = F(X)$ can be expressed as a convex combination of distributions $R = \eta Q + \sum_a \alpha_a R_a$ (where the coefficients η and α_a are nonnegative and $\eta + \sum_a \alpha_a = 1$) such that $\eta \leq \rho$ and for every a there exists an independent source Z_a on $(\{0, 1\}^\ell)^r$ with min-entropy k' such that R_a is ρ -close to $Z_a \otimes U_d$.

Now, as in the bit-fixing case, we can use a seeded extractor for total-rate independent sources together with a seed obtainer to construct a deterministic extractor for total-rate independent sources. The proof for the following Theorem is the same as the proof for the bit-fixing case in [[GRS04](#)]. We include it here for the sake of completeness.

Theorem 7.2. *Let $F : (\{0, 1\}^\ell)^r \rightarrow (\{0, 1\}^\ell)^r \times \{0, 1\}^d$ be a (k', ρ) -seed obtainer for independent sources X on $(\{0, 1\}^\ell)^r$ with total min-entropy k . Let $E_1 : (\{0, 1\}^\ell)^r \times \{0, 1\}^d \rightarrow \{0, 1\}^m$ be a seeded ϵ -extractor for independent sources on $(\{0, 1\}^\ell)^r$ with total min-entropy k . Then $E : (\{0, 1\}^\ell)^r \rightarrow \{0, 1\}^m$ defined by: $E(x) = E_1(F(x))$ is a deterministic $(\epsilon + 2\rho)$ -extractor for independent sources on $(\{0, 1\}^\ell)^r$ with total min-entropy k .*

Proof. By the definition of a seed obtainer we have that $E(X) = \eta E_1(Q) + \sum_a \alpha_a E_1(R_a)$ for some $\eta \leq \rho$. For each a we have that R_a is ρ -close to $Z_a \otimes U_d$, so $E_1(R_a)$ is ρ -close to $E_1(Z_a \otimes U_d)$, which is itself ϵ -close to U_m since E_1 is an ϵ -extractor. Thus $E_1(R_a)$ is $(\epsilon + \rho)$ -close to U_m , which implies that $E(X)$ is $(\epsilon + \rho)$ -close to $\eta E_1(Q) + (1 - \eta)U_m$. Therefore by [Lemma 2.4](#) we have that $E(X)$ is $(\eta + \epsilon + \rho)$ -close to uniform. The lemma follows because $\eta \leq \rho$. \square

To construct seed obtainers, we need to extend the definition of averaging samplers from [\[GRS04\]](#) to general functions as follows. This definition is similar in spirit to that of Vadhan in [\[Vad04\]](#), except the sample size is not fixed and we both upper and lower bound the total value of the sample.

Definition 7.3. A function $Samp : \{0, 1\}^t \rightarrow P([r])$ is a $(\delta, \theta_1, \theta_2, \gamma)$ averaging sampler if for every function $f : [r] \rightarrow [0, 1]$ with average value $\frac{1}{r} \sum_i f(i) = \delta$, it holds that

$$\Pr_{w \leftarrow U_t} \left[\theta_1 \leq \sum_{i \in Samp(w)} f(i) \leq \theta_2 \right] \geq 1 - \gamma.$$

When applying these samplers to total-rate independent sources, we get the following lemma.

Lemma 7.4. Let $Samp : \{0, 1\}^t \rightarrow P([r])$ be a $(\delta, \delta_1 r, \delta_2 r, \gamma)$ averaging sampler. Then for any independent source X on $(\{0, 1\}^\ell)^r$ with total min-entropy $k = \delta r \ell$, we have

$$\Pr_{w \leftarrow U_t} [\delta_1 r \ell \leq H_\infty(X_{Samp(w)}) \leq \delta_2 r \ell] \geq 1 - \gamma.$$

Proof. Let $f(i) = H_\infty(X_i)/\ell$. \square

Given these definitions, we can show that essentially the same construction from Gabizon et al. [\[GRS04\]](#) for bit-fixing seed obtainers works for total-rate independent source seed obtainers.

Theorem 7.5. Let $Samp : \{0, 1\}^t \rightarrow P([r])$ be a $(\delta, \delta_1 r, \delta_2 r, \gamma)$ averaging sampler and $E : (\{0, 1\}^\ell)^r \rightarrow \{0, 1\}^m$ be an ϵ -extractor for independent sources on $(\{0, 1\}^\ell)^r$ with total min-entropy $k = \delta_1 r \ell$. Then $F : (\{0, 1\}^\ell)^r \rightarrow (\{0, 1\}^\ell)^r \times \{0, 1\}^{m-t}$ defined as follows is a (k', ρ) -seed obtainer for independent sources on $(\{0, 1\}^\ell)^r$ with total min-entropy $k = \delta r \ell$ with $k' = (\delta - \delta_2) r \ell$ and $\rho = \max(\epsilon + \gamma, \epsilon \cdot 2^{t+1})$.

The Construction of F :

- Given $x \in (\{0, 1\}^\ell)^r$ compute $z = E(x)$. Let $E_1(x)$ denote the first t bits of $E(x)$ and $E_2(x)$ denote the remaining $m - t$ bits.
- Let $T = Samp(E_1(x))$.
- Let $x' = x_{[r] \setminus T}$. If $|x'| < n$ we pad it with zeroes to get an r source long string.
- Let $y = E_2(x)$. Output x', y .

The proof of this theorem is almost exactly the same as the proof in [\[GRS04\]](#), except substituting independent sources and the associated sampler and extractor for bit-fixing sources, so we omit it here. This theorem also follows from the main theorem of [\[Sha05\]](#).

7.2 Constructing Samplers

In order to use the seed obtainer construction to extract more bits, we first need a good averaging sampler. We will show that the same sampler construction given in Gabizon et al. [GRS04] generalizes to our definition. Our sampler works by generating d -wise independent variables $Z_1, \dots, Z_r \in [b]$ and letting $\text{Samp}(U_t) = \{i | Z_i = 1\}$.

Lemma 7.6. *For all δ and integers r, b, t such that $b/r \leq \delta \leq 1$ and $6 \log r \leq t \leq \frac{\delta r \log r}{20b}$ there is a polynomial-time computable $(\delta, \frac{\delta r}{2b}, \frac{3\delta r}{b}, 2^{-\Omega(t/\log r)})$ averaging sampler $\text{Samp} : \{0, 1\}^t \rightarrow P([r])$*

The following tail inequality for d -wise independent variables is due to Bellare and Rompel [BR94].

Theorem 7.7. [BR94] *Let $d \geq 6$ be an even integer. Suppose that X_1, \dots, X_r are d -wise independent random variables taking values in $[0, 1]$. Let $Y = \sum_{1 \leq i \leq r} Y_i$, $\mu = \mathbb{E}[Y]$, and $A > 0$. Then*

$$\Pr[|Y - \mu| \geq A] \leq 8 \left(\frac{d\mu + d^2}{A^2} \right)^{d/2}$$

Proof. (of Lemma 7.6) Let d be the largest even integer such that $d \log r \leq t$ and let $q = \lfloor \log b \rfloor \leq \log r$. Use $d \log r$ random bits to generate r d -wise independent random variables $Z_1, \dots, Z_r \in \{0, 1\}^q$ using the construction from [CW79]. Fix $a \in \{0, 1\}^q$. Let the random variable denoting the output of the sampler be $\text{Samp}(U_t) = \{i | Z_i = a\}$. For $1 \leq i \leq r$, define a random variable Y_i that is set to $f(i)$ if $i \in \text{Samp}(U_t)$ and 0 otherwise. Let $Y = \sum_i Y_i$ (note that Y is exactly the sum we wish to bound). Note that $\mu = \mathbb{E}[Y] = \delta r / 2^q$ and that the random variables Y_1, \dots, Y_r are d -wise independent. Applying Theorem 7.7 with $A = \delta r / 2b$,

$$\Pr[|Y - \mu| \geq A] \leq 8 \left(\frac{d \frac{\delta r}{2^q} + d^2}{A^2} \right)^{d/2}.$$

Note that

$$\begin{aligned} \{|Y - \mu| < A\} &\subseteq \left\{ \frac{\delta r}{2^q} - A < Y < \frac{\delta r}{2^q} + A \right\} \subseteq \left\{ \frac{\delta r}{b} - A < Y < \frac{2\delta r}{b} + A \right\} \\ &\subseteq \left\{ \frac{\delta r}{2b} \leq Y \leq \frac{3\delta r}{b} \right\} = \left\{ \frac{\delta r}{2b} \leq \sum_{i \in \text{Samp}(w)} f(i) \leq \frac{3\delta r}{b} \right\}. \end{aligned}$$

Note that $d \leq \frac{t}{\log r} \leq \frac{\delta r}{20b}$ by assumption. We conclude that

$$\begin{aligned} \Pr_{w \leftarrow U_t} \left[\frac{\delta r}{2b} \leq \sum_{i \in \text{Samp}(w)} f(i) \leq \frac{3\delta r}{b} \right] &\geq 1 - 8 \left(\frac{d \frac{\delta r}{2^q} + d^2}{(\delta r / 2b)^2} \right)^{d/2} \geq 1 - 8 \left(\frac{4b^2}{(\delta r)^2} \left(\frac{2d\delta r}{b} + \frac{d\delta r}{20b} \right) \right)^{d/2} \\ &\geq 1 - 8 \left(\frac{10db}{\delta r} \right)^{d/2} \geq 1 - 2^{-(d/2+3)} \geq 1 - 2^{-\Omega(t/\log r)} \end{aligned}$$

□

7.3 Extractors From Seed Obtainers

As in [GRS04] it will be convenient to combine Theorem 7.2 and Theorem 7.5 to get the following theorem.

Theorem 7.8. *Assume we have the following:*

- A $(\delta, \delta_1 r, \delta_2 r, \gamma)$ averaging sampler $Samp : \{0, 1\}^t \rightarrow P([r])$.
- A deterministic ϵ^* -extractor for total-rate δ_1 independent sources $E^* : (\{0, 1\}^\ell)^r \rightarrow \{0, 1\}^{m'}$.
- A seeded ϵ_1 -extractor for total-rate $\delta - \delta_2$ independent sources $E_1 : (\{0, 1\}^\ell)^r \times \{0, 1\}^s \rightarrow \{0, 1\}^m$, where $m' \geq s + t$.

Then we get a deterministic ϵ -extractor for total-rate δ independent sources $E : (\{0, 1\}^\ell)^r \rightarrow \{0, 1\}^m$ where $\epsilon = \epsilon_1 + 3 \cdot \max(\epsilon^* + \gamma, \epsilon^* \cdot 2^{t+1})$.

We will use the following seeded extractor from Raz, Reingold, and Vadhan [RRV99].

Theorem 7.9. [RRV99] For any r, k , and $\epsilon > 0$, there exists a ϵ -extractor $Ext : \{0, 1\}^r \times \{0, 1\}^s \rightarrow \{0, 1\}^m$ for all sources with min-entropy k , where $m = k$ and $s = \Theta(\log^2 r \cdot \log(1/\epsilon) \cdot \log m)$.

Combining the extractor from [RRV99] with the sampler from the previous section, we get the following general corollary, which shows how to transform a deterministic extractor that extracts just some of the min-entropy into one that extracts almost all of the min-entropy.

Corollary 7.10. Let $\delta, \delta_1, \epsilon_1$ and integers r, t be such that $\delta_1 \geq 1/2r$ and $6 \log r \leq t \leq \frac{\delta_1 r \log r}{10}$. Also let $m = (\delta - 6\delta_1)r\ell$ and $s = \Theta(\log^2(r\ell) \cdot \log(1/\epsilon_1) \cdot \log m)$. Then given any deterministic ϵ^* -extractor for total-rate δ_1 independent sources $E^* : (\{0, 1\}^\ell)^r \rightarrow \{0, 1\}^{m'}$ with $m' \geq s + t$, we can construct an ϵ -extractor for total-rate δ independent sources $E : (\{0, 1\}^\ell)^r \rightarrow \{0, 1\}^m$ where $\epsilon = \epsilon_1 + 3 \cdot \max(\epsilon^* + 2^{-\Omega(t/\log r)}, \epsilon^* \cdot 2^{t+1})$.

Proof. Combine Lemma 7.6 with $b = \delta/2\delta_1$, Theorem 7.9, and Theorem 7.8. \square

Now we can use Corollary 7.10 together with our previous deterministic extractor construction from Theorem 6.12 to show how we can extract nearly all of the entropy from total-rate independent sources with sufficiently high min-entropy, proving Theorem 1.7.

Proof. (Of Theorem 1.7.) Use the construction from Corollary 7.10 with the extractor from Theorem 6.12 as E^* and let $\epsilon_1 = 2^{-\Omega((\delta_1^2 r \ell)/(2^{2\ell} \log^3 r))}$ and $t = \Omega(\frac{\delta_1^2}{2^{2\ell}} r \ell)$. Then it's not hard to see that (choosing appropriate constants) these values satisfy $6 \log r \leq t \leq \frac{\delta_1 r \log r}{10}$ and $m' \geq s + t$ for sufficiently large r . \square

The extractor for small-space sources from Theorem 1.4 is then obtained by combining Theorem 1.7 with Lemma 3.1.

We could also use a seed obtainer together with the extractor for constant rate sources from Theorem 4.6. This lets us extract any constant fraction of the entropy and proves Theorem 1.6.

Proof. (Of Theorem 1.6.) Use the construction from Corollary 7.10 with the extractor from Theorem 4.6 as E^* and let $\epsilon_1 = 2^{-\Omega((r\ell)/(\log^3(r\ell)))}$ and $t = \Theta(r \log(\min(2^\ell, r)))$. Then it's not hard to see that (choosing appropriate constants) these values satisfy $6 \log r \leq t \leq \frac{\delta_1 r \log r}{10}$ and $m' \geq s + t$ for sufficiently large r . \square

The extractor for small-space sources from Theorem 1.3 is then obtained by combining Theorem 1.7 with Lemma 3.1.

We can also apply this construction to the polynomial entropy rate extractor from Corollary 5.2, which proves Theorem 1.5.

Proof. (Of Theorem 1.5.) Use the construction from Corollary 7.10 with the extractor from Corollary 5.2 as E^* and let $\epsilon_1 = 2^{-\Omega((\delta_1^2 r \ell)^{\Omega(1)})/(\log^3(r\ell))}$ and $t = (\delta_1^2 r \ell)^{\Omega(1)}$. Then it's not hard to see that (choosing appropriate constants) these values satisfy $6 \log r \leq t \leq \frac{\delta_1 r \log r}{10}$ and $m' \geq s + t$ for sufficiently large r . \square

The extractor for small-space sources from [Theorem 1.1](#) is then obtained by combining [Theorem 1.5](#) with [Lemma 3.1](#).

7.4 Extractors For Smaller Entropy

Gabizon et. al [[GRS04](#)] also showed how to use seed obtainers to extract more bits even when the initial extractor only extracts $\Theta(\log k)$ bits, which they're able to get from the cycle walk extractor from [[KZ03](#)]. We can generalize their construction to work for total-rate independent sources, which together with our generalization of the cycle walk extractor allows us to extract more bits from smaller entropy rates.

In order to get a seed obtainer that can use only $\Theta(\log k)$ bits, we need both a sampler and a seeded extractor for total-rate independent sources. To do so, as in [[GRS04](#)], we use d -wise ϵ -dependent random variables to both sample and partition. The proofs of the following two lemmas easily generalize the construction from [[GRS04](#)] in a similar way to our earlier sampler construction.

Lemma 7.11. *For any constant $0 < \alpha < 1$, there exist constants $c > 0$ and $0 < b < 1/2$ (both depending on α) such that for any $r \geq 16$ and $k = \delta r \ell \geq \log^c r$, the following holds. There is a polynomial-time computable $(\delta, \delta r/2k^b, 3\delta r/k^b, O(k^{-b}))$ sampler $\text{Samp} : \{0, 1\}^t \rightarrow P([r])$ where $t = \alpha \cdot \log k$.*

Lemma 7.12. *Fix any constant $0 < \alpha < 1$. There exist constants $c > 0$ and $0 < b < 1/2$ (both depending on α) such that for any $r \geq 16$ and $k = \delta r \ell \geq \log^c r$, we can use $\alpha \cdot \log k$ random bits to explicitly partition $[r]$ into $m = \Theta(k^b)$ sets T_1, \dots, T_m such that for every function $f : [r] \rightarrow [0, 1]$ with average value $\frac{1}{r} \sum_i f(i) = \delta$,*

$$\Pr \left[\forall i, \delta r/2k^b \leq \sum_{j \in T_i} f(j) \leq 3\delta r/k^b \right] \geq 1 - O(k^{-b}).$$

As in [Lemma 7.6](#), this lemma implies that if we partition a total-rate δ independent source, with high probability each T_i has some min-entropy.

Corollary 7.13. *For any constant $0 < \alpha < 1$, there exist constants $c > 0$ and $0 < b < 1/2$ (both depending on α) such that for any $r \geq 16$ and $k \geq \log^c r$, the following holds. We can use $\alpha \cdot \log k$ random bits to explicitly partition $[r]$ into $m = \Theta(k^b)$ sets T_1, \dots, T_m such that for any independent sources X on $(\{0, 1\}^\ell)^r$ with total min-entropy k ,*

$$\Pr \left[\forall i, k^{1-b}/2 \leq H_\infty(X_{T_i}) \leq 3k^{1-b} \right] \geq 1 - O(k^{-b}).$$

Now we will use this partitioning to construct a seeded extractor for total-rate independent sources that uses a small seed. As in [[GRS04](#)] once we partition the source, we apply an extractor to each part. The extractor we will use is our sum mod p extractor.

Theorem 7.14. *For any constant $0 < \alpha < 1$, there exist constants $c > 0$ and $0 < b < 1/2$ (both depending on α) such that for any $r \geq 16$, $k \geq \log^c r$, $0 < \delta \leq 1$ and $\ell \leq \log(k^{(1-b)/2}/\sqrt{\log k^{2b}})$, the following holds. There is a polynomial-time computable seeded ϵ -extractor $E : (\{0, 1\}^\ell)^r \times \{0, 1\}^s \rightarrow \{0, 1\}^m$ for independent sources on $(\{0, 1\}^\ell)^r$ with total min-entropy $k = \delta r \ell$, with $s = \alpha \cdot \log k$, $m = \Theta(k^b \ell)$ and $\epsilon = O(k^{-b})$.*

Proof. As stated above, E works by first partitioning the input x into $m' = \Theta(k^b)$ parts $T_1, \dots, T_{m'}$ using [Corollary 7.13](#). Next we find the next largest prime $p \geq 2^\ell$, which by Bertrand's postulate is at most $2 \cdot 2^\ell$, so we can find it efficiently by brute force search. Then for each T_i we compute $z_i = \sum_{j \in T_i} x_j \pmod p$ and output $z = z_1, \dots, z_{m'}$.

Let Z be the distribution of the output string z . Let A be the “good” event that all sets T_i have entropy at least $k^{1-b}/2$. Then we decompose Z as

$$Z = \Pr[A^c] \cdot (Z|A^c) + \Pr[A] \cdot (Z|A).$$

Now by [Corollary 7.13](#), $\Pr[A] \geq 1 - O(k^{-b})$. By [Corollary 6.7](#), $(Z|A)$ is $m' \cdot 2^{-\Omega(k^{1-b}/2^{2^\ell})}$ close to uniform. Since $\ell \leq \log(k^{(1-b)/2}/\sqrt{\log k^{2b}})$, $(Z|A)$ is $O(k^{-b})$ close to uniform. Thus by [Lemma 2.4](#), Z is $O(k^{-b})$ close to uniform. \square

Now we are ready to combine these ingredients using [Theorem 7.8](#) to get an improved extractor.

Theorem 7.15. *There exist constants $c > 0$ and $0 < b < 1/2$ such that for $k \geq \log^c r$ and $2^\ell \leq O(k^{(1-b)/2}/\sqrt{\log k^{2b}})$, the following holds. There exists a polynomial-time computable ϵ -extractor $E : \{0, 1\}^{\ell r} \rightarrow \{0, 1\}^m$ for independent sources on $(\{0, 1\}^\ell)^r$ with min-entropy k , where $m = \Theta(k^b \ell)$ and $\epsilon = O(k^{-b})$.*

Proof. Use [Theorem 7.8](#) together with the sampler from [Lemma 7.11](#), the deterministic extractor from [Corollary 6.7](#), and the seeded extractor from [Theorem 7.14](#) \square

This still doesn't get all of the entropy out of the source, but now we have a long enough output that we can use the seeded extractor from [Theorem 7.9](#) to get the rest of the entropy, which proves [Theorem 1.8](#).

Proof. (Of [Theorem 1.8](#).) Use [Theorem 7.8](#) together with the sampler from [Lemma 7.11](#), the deterministic extractor from [Theorem 7.15](#), and the seeded extractor from [Theorem 7.9](#). \square

8 Nonconstructive Results

In this section, we describe nonconstructive results for both small-space and total-rate independent sources. These results make use of the probabilistic method. We show that a randomly chosen function is an extractor for each of these classes of sources with high probability, and is able to extract almost all of the entropy even when the entropy is logarithmically small.

The key result we use is the well-known fact that if we wish to extract from a class of sources, a random function is a good extractor as long as the number of sources is not too large.³ The only requirement on the sources is that they are close to having high min-entropy. The following theorem is similar to a theorem in [\[Dod00a\]](#), but we prove it here to get the exact parameters we need.

Theorem 8.1. *Suppose we have a set \mathcal{X} of random sources on $\{0, 1\}^n$ and $\epsilon > 0$ such that $\forall X \in \mathcal{X}$, there is a source X' with $|X' - X| \leq \frac{\epsilon}{2}$ and $H_\infty(X') \geq k$. Let $m = k - (2 \log \frac{1}{\epsilon} + \log k + \log \log |\mathcal{X}| + \log \log \frac{1}{\xi})$ for some $\xi \in (0, 1]$. Then a function $f : \{0, 1\}^n \rightarrow \{0, 1\}^m$ chosen uniformly at random is an ϵ -extractor for \mathcal{X} with probability at least $1 - \xi$.*

³In fact, if we wish to save randomness in selecting the function, then we can get nearly the same parameters by using a random d -wise independent function instead of a completely random function [\[Dod00a\]](#).

In other words, with probability $1 - \xi$, a random function extracts m bits from \mathcal{X} with error ϵ as long as $|\mathcal{X}| \leq 2^{(\epsilon^2/4 \log(1/\xi))2^{k-m}}$.

We need the following Chernoff bound to prove [Theorem 8.1](#).

Lemma 8.2. *Let Z_1, \dots, Z_r be independent indicator random variables such that $\Pr[Z_1 = 1] = p_i$. Let $Z = \sum_{i=1}^r a_i Z_i$ where $0 \leq a_i \leq 1$ for all i , and let $\mu = \mathbb{E}[Z]$. Then for any $0 < \epsilon \leq 1$*

$$\Pr[|Z - \mu| \geq \epsilon\mu] < 2e^{-\mu\epsilon^2/3}.$$

To prove [Theorem 8.1](#), we'll first use [Lemma 8.2](#) to show that a random function is a good extractor for a single source, and then apply the union bound.

Proof. (of [Theorem 8.1](#))

Let $f : \{0, 1\}^n \rightarrow \{0, 1\}^m$ be chosen uniformly random from all functions from n bits to m bits. Fix $X \in \mathcal{X}$ and $y \in \{0, 1\}^m$. Let X' be such that $|X' - X| \leq \epsilon/2$ and $H_\infty(X') \geq k$. Let Z_x be an indicator random variable for whether $f(x)$ is equal to y . Let $a_x = 2^k \Pr[X' = x]$ (so $0 \leq a_x \leq 1$) and $Z = \sum_{x \in \{0, 1\}^n} a_x Z_x = 2^k \Pr_{x \in X'}[f(x) = y]$. Since the function f is chosen uniformly at random, $\mathbb{E}[Z] = 2^{k-m}$. Thus we can apply [Lemma 8.2](#) to get

$$\Pr_f \left[\left| \Pr_{x \in X'}[f(x) = y] - \frac{1}{2^m} \right| \geq \frac{\epsilon}{2^m} \right] = \Pr_f \left[\left| Z - 2^{k-m} \right| \geq \epsilon 2^{k-m} \right] \leq 2e^{-\epsilon^2 2^{k-m}/3}$$

Suppose that for all y $|\Pr_{x \in X'}[f(x) = y] - \frac{1}{2^m}| \leq \frac{\epsilon}{2^m}$. By the union bound, this occurs with probability at least $1 - 2^{m+1}e^{-\epsilon^2 2^{k-m}/3}$. In this case, by the definition of variation distance

$$|f(X') - U_m| = \frac{1}{2} \sum_{y \in \{0, 1\}^m} \left| \Pr_x[f(x) = y] - \frac{1}{2^m} \right| \leq \frac{\epsilon}{2}.$$

Then

$$|f(X) - U_m| \leq |f(X) - f(X')| + |f(X') - U_m| \leq |X - X'| + \epsilon/2 \leq \epsilon.$$

Thus f is an ϵ -extractor for X with probability at least $1 - 2^{m+1}e^{-\epsilon^2 2^{k-m}/3}$. Taking the union bound over all sources $X \in \mathcal{X}$, f is an ϵ -extractor for \mathcal{X} with probability at least $1 - |\mathcal{X}|2^{m+1}e^{-\epsilon^2 2^{k-m}/3}$. When $m = k - (2 \log \frac{1}{\epsilon} + \log k + \log \log |\mathcal{X}| + \log \log \frac{1}{\xi})$ this probability is at least $1 - \xi$, as desired. \square

8.1 Small-Space Sources

Since the probabilities on the edges in small-space sources can be any real number in $[0, 1]$, there are an infinite number of such sources, and so we cannot directly apply [Theorem 8.1](#). We instead introduce a more restricted model to which we can apply [Theorem 8.1](#), and show that general small-space sources are close to convex combinations of this more restricted model. The more restricted model we consider restricts all probabilities to be a multiple of some α .

Definition 8.3. An α -approximate space s source is a space s source where the probabilities on all edges are multiples of α .

We'll show that any rate δ small-space source is a convex combination of α -approximate small-space sources, each of which is close to the original source. Thus any extractor that works on α -approximate sources that are close to having rate δ will also be an extractor for rate δ small-space sources.

Lemma 8.4. *Let X be a space s source on $\{0, 1\}^n$ with min-entropy rate δ . The source X is a convex combination of α -approximate space s sources, each of which has distance at most $\alpha n 2^s$ to X .*

Proof. We can write X as a convex combination of sources X_a such that each X_a is obtained from X by replacing each edge probability p with either $\lfloor \frac{p}{\alpha} \rfloor \alpha$ or $(\lfloor \frac{p}{\alpha} \rfloor + 1)\alpha$.

We will show that X_a is close to X via a hybrid argument. Let X_a^i be the hybrid obtained by the first i bits having probabilities from X_a and the rest of the bits having probabilities from X . So $X = X_a^0$ and $X_a = X_a^n$. Then $|X - X_a| = |\sum_{i=1}^n (X_a^{i-1} - X_a^i)| \leq \sum_{i=1}^n |X_a^{i-1} - X_a^i|$.

For each term $|X_a^{i-1} - X_a^i|$ the only difference is in the probabilities on the edges in the i th layer, which each differ by at most α . We fix i and calculate this distance. Let $v_{i,j}$ denote the j th vertex in the i th layer. Let $q_{i-1,j}$ denote the probability of reaching $v_{i-1,j}$ in X_a and $p_{j,j'}^0$ ($p_{j,j'}^1$) denote the probability on the 0 (1) edge from $v_{i-1,j}$ to $v_{i,j'}$ in X . Then

$$|X_a^{i-1} - X_a^i| \leq \frac{1}{2} \sum_{j,j'} q_{i-1,j} ((p_{j,j'}^0 + \alpha - p_{j,j'}^1) + (p_{j,j'}^1 + \alpha - p_{j,j'}^0)) \leq \alpha \sum_{j'} \sum_j q_{i-1,j} = \alpha \sum_{j'} 1 = \alpha 2^s.$$

So the overall error is bounded by $|X - X_a| \leq \sum_{i=1}^n \alpha 2^s = \alpha n 2^s$. \square

Lemma 8.5. *The number of α -approximate space s sources on $\{0, 1\}^n$ is less than $2^{(s+1)2^s n/\alpha}$.*

Proof. First count the number of possible edge configurations from any given vertex. There are 2^{s+1} possible edges, since there is a 0 edge and a 1 edge for each of the 2^s vertices in the next layer. Since all probabilities are multiples of α , there are less than $2^{(s+1)/\alpha}$ ways to allocate probabilities to these edges. Since there are n layers and 2^s vertices at each layer, the total number of possible sources is $2^{(s+1)2^s n/\alpha}$. \square

Now we invoke [Theorem 8.1](#) to show that a random function is a good extractor for small-space sources.

Theorem 8.6. *A function $f : \{0, 1\}^n \rightarrow \{0, 1\}^m$ chosen uniformly at random is an ϵ -extractor for space s sources with min-entropy rate δ with probability at least $1 - \xi$ when $m = \delta n - (3 \log \frac{1}{\epsilon} + \log(\delta n) + 2s + 1 + \log(s+1) + 2 \log n + \log \log \frac{1}{\xi})$.*

This theorem says that extractors exist for sources with space almost as large as $\delta n/2$ and with min-entropy as low as $\Omega(\log n)$.

Proof. First apply [Lemma 8.4](#) with $\alpha = \epsilon/n2^{s+1}$ to show that each small-space source X is a convex combination of α -approximate sources that are $\epsilon/2$ close to X . Then apply [Theorem 8.1](#) to the set of α -approximate sources that are $\epsilon/2$ close to having min-entropy rate δ , using [Lemma 8.5](#) as the bound on the number of such sources (since this set is a subset of all α -approximate space s sources). Since each rate δ space s source is a convex combination of these α -approximate sources, the extractors given by [Theorem 8.1](#) also work to these sources. \square

8.2 Total-Rate Independent Sources

We can also apply [Theorem 8.1](#) to total-rate independent sources. Similarly to the small-space case, we define an intermediate model to reduce the number of sources.

Definition 8.7. An α -approximate independent source X_1, \dots, X_r on $(\{0, 1\}^\ell)^r$ is an independent source such that $\forall y \in \{0, 1\}^\ell$ and $\forall i$, $\Pr[X_i = y]$ is a multiple of α .

We use this model rather than flat independent sources because as we saw in [Lemma 6.5](#), we can lose a constant fraction of the min-entropy when viewing an independent source as a convex combination of flat independent sources.

This lemma allows us to restrict our attention to α -approximate independent sources. We'll show that any total-rate δ independent-symbol source is a convex combination of α -approximate independent sources, each of which is close to the original source.

Lemma 8.8. *Let $X = X_1, \dots, X_r$ be an total-rate δ independent source on $(\{0, 1\}^\ell)^r$. The source X is a convex combination of α -approximate independent sources, each of which has distance at most $\frac{1}{2}\alpha r 2^\ell$ to X .*

Proof. We can write X as a convex combination of sources $X' = X'_1, \dots, X'_r$ such that each X'_i is obtained from X_i by replacing each output probability $\Pr[X_i = y]$ with either $\lfloor \frac{p}{\alpha} \rfloor \alpha$ or $(\lfloor \frac{p}{\alpha} \rfloor + 1)\alpha$.

Now the distance

$$\begin{aligned} |X' - X| &= \sum_{i=1}^r |X'_i - X_i| = \frac{1}{2} \sum_{i=1}^r \sum_{x \in (\{0,1\}^\ell)} |\Pr[X'_i = x] - \Pr[X_i = x]| \\ &\leq \frac{1}{2} \sum_{i=1}^r \alpha 2^\ell = \frac{1}{2} \alpha r 2^\ell, \end{aligned}$$

where the first inequality is because each string $x \in \{0, 1\}^\ell$ contributes at most α error for each X_i . □

Lemma 8.9. *The number of α -approximate independent sources on $(\{0, 1\}^\ell)^r$ is less than $2^{\frac{1}{\alpha} r \ell}$.*

Proof. Let $X = X_1, \dots, X_r$ be an α -approximate total-rate δ independent source on $(\{0, 1\}^\ell)^r$. Since there are 2^ℓ possible values for each X_i , each of which has a probability that is a multiple of α , there are less than $2^{\frac{\ell}{\alpha}}$ possible distributions for X_i . Thus there are less than $(2^{\frac{\ell}{\alpha}})^r = 2^{\frac{1}{\alpha} r \ell}$ possible distributions for X . □

Now we can apply [Theorem 8.1](#) to show that a random function is a good extractor for total-rate δ independent sources.

Theorem 8.10. *A function $f : (\{0, 1\}^\ell)^r \rightarrow \{0, 1\}^m$ chosen uniformly at random is an ϵ -extractor for total rate δ independent sources with probability at least $1 - \xi$ when $m = \delta r \ell - (2 \log r + \ell + \log \ell + \log(\delta r \ell) + 3 \log \frac{1}{\epsilon} + \log \log \frac{1}{\xi})$.*

This theorem says that we can extract when the entropy is as small as $\Omega(\log r + \ell)$. Note that the ℓ bound is necessary because otherwise all of the entropy could be contained within a single source, which we know is impossible to extract from.

Proof. First apply [Lemma 8.8](#) with $\alpha = \epsilon/r 2^\ell$ to show that the each total-rate δ independent source X is a convex combination of α -approximate sources total-rate δ independent sources that are $\epsilon/2$ close to X . Then apply [Theorem 8.1](#) to the set of α -approximate sources total-rate δ independent sources that are ϵ' close to having min-entropy rate δ , using [Lemma 8.9](#) as the bound on the number of such sources (since this set is a subset of all α -approximate independent sources). Since each total-rate δ independent source is a convex combination of these α -approximate sources, the extractors given by [Theorem 8.1](#) also work to these sources. □

9 Doing Better For Width Two

We consider the case of space 1 (width 2) sources where the output bit is restricted to be the same as the label of the next state, which we will call *restricted width two sources*. For such sources, we can improve our results by decreasing the alphabet size in the total-rate independent sources. This will allow us to extract from smaller entropy rates. We will need the following class of sources.

Definition 9.1. A previous-bit source on $\{0, 1\}^n$ with min-entropy k has at least k uniformly random bits and the rest of the bits are functions of the previous bit.

We will show that restricted width two sources are close to a convex combination of previous-bit sources, and then show that these previous bit sources can be converted into total-rate independent sources with small alphabet size.

9.1 Extracting From Previous-Bit Sources

To convert a previous-bit source to a total-rate independent source, we first divide the source into blocks as before, but instead of simply viewing each block as a binary number, we apply a function to reduce the alphabet size while still maintaining some of the entropy. Specifically, we will show that if a block has at least one random bit, then the output symbol will have at least one bit of entropy. The main lemma is as follows.

Lemma 9.2. Any length n previous-bit source X with min-entropy k can be converted in polynomial time to a convex combination of flat independent sources on $(\{0, 1\}^\ell)^r$ with min-entropy k' , where $r = \frac{k}{2}$, $k' = k^2/4n$ and $\ell = \lceil \log(\frac{2n}{k} + 1) \rceil$.

The following lemma shows that any block that contains at least one random bit will give a random source.

Lemma 9.3. We can construct a function $f : \{0, 1\}^t \rightarrow \{0, 1\}^{\lceil \log(t+1) \rceil}$ so that for any previous-bit source Y on $\{0, 1\}^t$ with exactly one random bit, f attains different values depending on whether the random bit in Y is set to 0 or 1.

Proof. For $0 \leq i \leq t$, let $z_i \in \mathbb{Z}_2^{\lceil \log(t+1) \rceil}$ be the standard representation of i as a vector over \mathbb{Z}_2 . (More generally, we only require the z_i to be distinct vectors.) Then $f(y) = \sum_{i=1}^t y_i(z_i - z_{i-1})$.

Let y_0 (y_1) be Y with the random bit set to 0 (1). Now we show that $f(y_0) \neq f(y_1)$. We see that

$$f(y_0) - f(y_1) = \sum_{i=1}^t (y_{0i} - y_{1i})(z_i - z_{i-1}).$$

It's easy to see that $y_{0i} - y_{1i}$ will be 0 for all fixed bits and 1 whenever the random bit or its negation appears. For our sources, all appearances of the random bit must appear consecutively. This means that if the random bit appears from positions j through k , $f(y_0) - f(y_1) = z_k - z_{j-1}$, since all of the other terms cancel. Thus since $z_k \neq z_{j-1}$, $f(y_0) - f(y_1) \neq 0$. \square

Now we can prove [Lemma 9.2](#).

Proof. Divide X into $r = k/2$ blocks of size $n/r = 2n/k$. Then apply the function f from [Lemma 9.3](#) to each block to get Y .

To see that this works, fix all of the random bits that cross between blocks. Also, for each block fix all but one of the random bits that are contained within the block. Now X is a convex combination of all of the sources given by every possible such fixing. Let X' be a source corresponding to one particular fixing. We will show that if we apply f to every block of X' , we will get a source with enough random blocks. Any block of X' with a random source is a previous-bit source with one random bit, so we can apply [Lemma 9.3](#) to see that the output of f on this block is uniformly chosen from among two different sources, as desired.

Now we just need to see how many blocks with at least one random bit there are. There can be at most r random bits that cross between blocks. So removing those bits we are left with at least $k - r = k/2$ random bits. These $k/2$ random bits must be contained in at least $k' = (k/2)/(n/r) = k^2/4n$ different blocks, which gives us the desired bound. \square

Now we can combine [Theorem 1.7](#) and [Lemma 9.2](#) to get an extractor for previous-bit sources.

Theorem 9.4. *There exists a polynomial-time computable ϵ -extractor for the set of previous-bit sources of length n with min-entropy k that outputs $m = \frac{k^2}{8n}$ bits and has error $\epsilon = \exp(-\Omega(k^5/(n^4 \log(n/k) \log^3 k)))$.*

Proof. Given a source X , apply [Lemma 9.2](#) to convert X into a convex combination of flat independent sources on $(\{0, 1\}^{\ell})^r$ with total min-entropy k' , where $r = \frac{k}{2}$, $k' = \frac{k^2}{4n}$, and $\ell' = \lceil \log(\frac{2n}{k} + 1) \rceil$. Then apply the extractor from [Theorem 1.7](#) with $\zeta = k^2/(48n \cdot r\ell)$. \square

9.2 Restricted Width Two Sources As Convex Combinations Of Previous-Bit Sources

To show we can extract from restricted width two sources, we will prove that these sources can be viewed as convex combinations of previous bit sources. With high probability, these previous-bit sources will have sufficient entropy so that our extractor from the previous section will work.

Lemma 9.5. *Any length n restricted width two source X with min-entropy k is a convex combination of length n previous bit sources Z_j so that with probability at least $1 - 2^{-k/4} - e^{-9k^2/2n}$, the sources Z_j have at least $k' = \min(k/48 \log(n/k), k/96)$ random bits.*

To get our extractor, we just combine this lemma with the extractor from [Theorem 9.4](#).

Theorem 9.6. *There exists a polynomial-time computable ϵ -extractor for the set of length n restricted width two sources with min-entropy k that outputs $m = \Omega(k^2/n(\max(\log(n/k), 1))^2)$ bits and has error $\epsilon = 2^{-\Omega((k')^5/(n^4 \log(n/k') \log^3 k'))}$, where $k' = \min(k/48 \log(n/k), k/96)$.*

Proof. By [Lemma 9.5](#) our source X is $2^{-k/4} + e^{-9k^2/2n}$ close to a convex combination of length n previous-bit sources with $k' = \min(k/48 \log(n/k), k/96)$ random bits. We can then apply the extractor from [Theorem 9.4](#) to get out $m = \frac{(k')^2}{8n} = \Omega(k^2/n(\max(\log(n/k), 1))^2)$ bits. \square

Notice that here we only need $k \gg n^{4/5}$ whereas before we required $k \gg n^{1-\eta}$ for some small constant η .

Now we describe how we express the restricted width two source X as a convex combination of previous-bit sources Z_j . This is done recursively on the layers of the branching program for the source. We say we are in a given state at each layer; either “open”, “closed at 0”, or “closed at 1”. Each sequence of states corresponds to a previous-bit source. The way we divide the next layer up depends on the state we

are in. The high level picture is that each random bit corresponds to going into the open state, which we are in until we get a fixed bit, which takes us to the corresponding closed state. We stay closed until another random bit occurs. An example is shown in Figure 9.2.

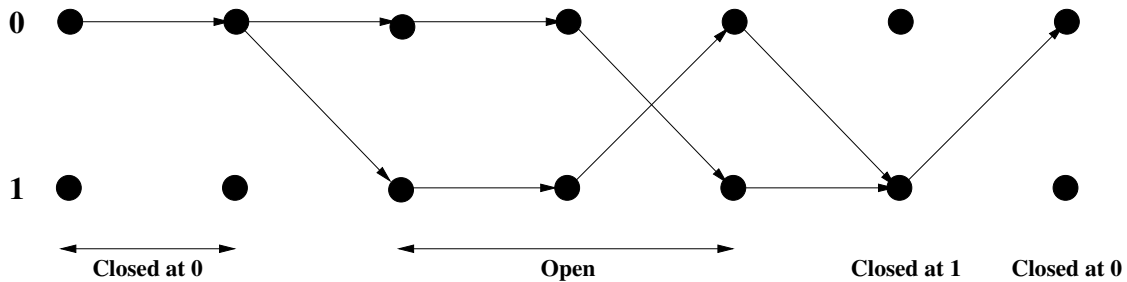


Figure 3: A previous-bit source viewed as a restricted width two source. This source consists of the bits $0, 0, r, r, \bar{r}, 1, 0$, where r is a random bit.

More formally, we define the following probabilities, shown in Figure 9.2.

$$p_{i0} = \Pr[X_i = 0 | X_{i-1} = 0]$$

$$p_{i1} = \Pr[X_i = 1 | X_{i-1} = 0]$$

$$q_{i0} = \Pr[X_i = 0 | X_{i-1} = 1]$$

$$q_{i1} = \Pr[X_i = 1 | X_{i-1} = 1]$$

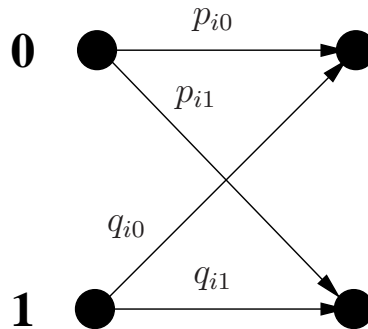


Figure 4: The probabilities for a single bit of a restricted width two source.

First, we describe what happens if we are currently in the open state. The next bit is fixed to 0 (resp. 1) and the state becomes closed at 0 (1) with probability $\min(p_{i0}, q_{i0})$ ($\min(p_{i1}, q_{i1})$). Else we stay in the open state and the next bit is either equal to the previous bit or the negation of the previous bit depending on which edges have the remaining probability.

If we are closed at 0, the next bit is random and we go into the open state with probability $2 \min(p_{i0}, p_{i1})$. If $p_{i0} < p_{i1}$, the next bit is fixed to 1 and we go into the closed at 1 state with probability $1 - 2p_{i0}$. Else the next bit is fixed to 0 and we go into the closed at 0 state with probability $1 - 2p_{i1}$.

If we are closed at 1, the next bit is random and we go into the open state with probability $2 \min(q_{i0}, q_{i1})$. If $q_{i0} < q_{i1}$, the next bit is fixed to 1 and we go into the closed at 1 state with probability $1 - 2q_{i0}$. Else the next bit is fixed to 0 and we go into the closed at 0 state with probability $1 - 2q_{i1}$.

Now we show that with high probability, the sources in the convex combination have sufficient min-entropy. We do this by looking at the relationships between paths in the original source X and the min-entropy of the Z_j . First, note that each path in the branching program corresponds to an output value of X , so each path has probability at most 2^{-k} . Note that the min-entropy of Z_j is equal to the number of openings in Z_j .

Each path can be divided into edges that are the most probable edge coming out of a node and those that are the least probable. We will show how the number of least probable edges on a path in X relates to the min-entropy of a Z_j that contains this path. First note that every least probable edge corresponds to either an opening, a closing, or what we call a “false closing”. A false closing is defined as transitioning from the open state to the open state yet still taking a least probable edge. Let $C(Z_j)$ denote the number of closings in Z_j , $A(Z_j)$ denote the number of openings, and $B(Z_j)$ denote the number of false closings.

If we could ignore the false closings, showing that with high probability we take the least probable edge a large number of times would be enough. Since $C(Z_j) \leq A(Z_j)$, this would imply that with high probability $A(Z_j)$ is large, and thus the Z_j have large min-entropy with high probability. To take account of the false closings, we also have to show that there aren’t too many of them, which we will do by a martingale argument.

First, we show that with high probability over all paths in X , we take the least probable edge a large number of times.

Lemma 9.7. *For any length n restricted width two source with min-entropy k , the total probability of all paths that have at most $t = \min(k/8 \log(n/k), k/16)$ least probable edges is less than $2^{-k/4}$.*

Proof. Since the source has min-entropy k , each path has probability at most 2^{-k} . There are $\binom{n}{i}$ paths that have i least probable edges. Thus the total probability of all paths that have at most t least probable edges is at most

$$2^{-k} \sum_{i=0}^t \binom{n}{i} \leq 2^{-k} 2^{nH(t/n)} < 2^{-k+2t \log(n/t)}$$

where $H(t/n)$ is the standard Shannon entropy $H(p) = -p \log p - (1-p) \log(1-p)$.

Suppose $k \leq n/4$. Then $s = k/8 \log(n/k)$, so

$$2s \log \frac{n}{t} = \frac{k}{4} \left(1 + \frac{\log(8 \log \frac{n}{k})}{\log \frac{n}{k}} \right) \leq \frac{3k}{4}.$$

If $k > n/4$, then $t = \frac{k}{16}$, so

$$2t \log \frac{n}{t} = \frac{k}{8} \left(4 + \log \frac{n}{k} \right) \leq \frac{3k}{4}.$$

Thus the probability of taking at most t least probable edges is at most $2^{-k+2t \log(n/t)} \leq 2^{-k/4}$. \square

To show that the number of false closings is small, we first define a submartingale that is equal to the number of closings minus the number of false closings after the first i bits. Then we use the following simple variant of Azuma’s inequality for submartingales (see [Wor99] for a proof).

Definition 9.8. A submartingale with respect to a random process G_0, G_1, \dots , with G_0 fixed, is a sequence Y_0, Y_1, \dots of random variable defined on the random process such that

$$\mathbb{E}[Y_{i+1} | G_0, G_1, \dots, G_i] \geq Y_i$$

for all $i \geq 0$.

Lemma 9.9. Let Y_0, Y_1, \dots, Y_n be a submartingale with respect to G_0, G_1, \dots, G_n where $Y_0 = 0$ and $|Y_i - Y_{i-1}| \leq 1$ for $i \geq 1$. Then for all $\alpha > 0$,

$$\Pr[Y_n \leq -\alpha] \leq e^{-\frac{\alpha^2}{2n}}.$$

Now we are ready to prove that with high probability the number of false closings can't be too large.

Lemma 9.10. For all $\alpha > 0$,

$$\Pr[B(Z_j) \geq C(Z_j) + \alpha] \leq e^{-\frac{\alpha^2}{2n}}.$$

Proof. Let Y_i be the number of closings from X_1, \dots, X_i minus the number of false closings from X_1, \dots, X_i and let $Y_0 = 0$. Let G_0, G_1, \dots, G_n be the random process for dividing X into previous-bit sources, so G_i is the state after the first i bits have been divided.

Now we show that Y_0, \dots, Y_n is a submartingale with respect to G_0, G_1, \dots, G_n . If we are in a closed state after i bits, then we have no closings or false closings at $i + 1$, so $\mathbb{E}[Y_{i+1}|G_0, G_1, \dots, G_i] = Y_i$. If we are in an open state at i , we show that if we have the possibility of a false closing at $i + 1$, then the probability of closing is greater than $1/2$, and in particular is greater than the probability of a false closing. This would imply that $\mathbb{E}[Y_{i+1}|G_0, G_1, \dots, G_i] \geq Y_i$, as desired. First, note that the probability of closing at $i + 1$ is

$$\min(p_{i+1,0}, q_{i+1,0}) + \min(p_{i+1,1}, q_{i+1,1}) = \min(p_{i+1,0} + q_{i+1,1}, q_{i+1,0} + p_{i+1,1}).$$

Suppose without loss of generality that $p_{i+1,0} + q_{i+1,1} \geq q_{i+1,0} + p_{i+1,1}$, so we close with probability $q_{i+1,0} + p_{i+1,1}$. In this case, the edges we would take in a false closing are the 00 and 11 edges. So if we have a false closing, either $p_{i+1,0} \leq 1/2$ or $q_{i+1,1} \leq 1/2$, which implies either $p_{i+1,1} \geq 1/2$ or $q_{i+1,0} \geq 1/2$, and thus the probability of closing is at least $1/2$.

By the definition of Y_i , $|Y_i - Y_{i-1}| \leq 1$, so we can apply [Lemma 9.9](#) to get

$$\Pr[Y_n \leq -\alpha] \leq e^{-\frac{\alpha^2}{2n}},$$

which implies the desired result. □

Now we are finally ready to prove [Lemma 9.5](#).

Proof. (Of [Lemma 9.5](#).)

First, express the restricted width two source X as a convex combination of previous-bit sources Z_j as described previously, so $X = \sum_j \alpha_j Z_j$. Now look at a randomly chosen Z_j , chosen with probability α_j . The number of random bits in Z_j is equal to the number of openings $A(Z_j)$. Since the number of closings is either equal to or one less than the number of openings, either $C(Z_j) = A(Z_j)$ or $C(Z_j) = A(Z_j) - 1$. So if we can prove with high probability that $C(Z_j)$ is large, then with high probability the number of random bits in Z_j is also large. For every path in Z_j , every least probable edge on the path corresponds to either an opening, a closing, or a false closing. Thus the probability that $A(Z_j) + B(Z_j) + C(Z_j) \geq s$ is at least the probability over all paths that the path has at least s least probable edges. Thus we can apply [Lemma 9.7](#) and get

$$\Pr[B(Z_j) + 2C(Z_j) \geq s - 1] \geq \Pr[A(Z_j) + B(Z_j) + C(Z_j) \geq s] > 1 - 2^{-k/4}$$

for $s = \min(k/8 \log(n/k), k/16)$.

By Lemma 9.10,

$$\Pr[B(Z_j) < C(Z_j) + \frac{s}{2}] \geq 1 - e^{-\frac{s^2}{8n}}.$$

With high probability both of these events occur, so

$$\Pr[C(Z_j) \geq \frac{s}{6}] \geq 1 - 2^{-k/4} - e^{-\frac{s^2}{8n}}.$$

□

References

- [AKS04] Manindra Agrawal, Neeraj Kayal, and Nitin Saxena. Primes is in p. *Annals of Mathematics*, 160(2):781–793, 2004.
- [AS00] N. Alon and J. H. Spencer. *The Probabilistic Method*. Wiley–Interscience Series, John Wiley & Sons, Inc., New York, 2000.
- [BBR88] Charles H. Bennett, Gilles Brassard, and Jean-Marc Robert. Privacy amplification by public discussion. *SIAM Journal on Computing*, 17(2):210–229, April 1988.
- [BGK06] J. Bourgain, A. Glibichuk, and S. Konyagin. Estimates for the number of sums and products and for exponential sums in fields of prime order. *J. London Math. Soc.*, 73(2):380–398, 2006.
- [BIW04] Boaz Barak, Russell Impagliazzo, and Avi Wigderson. Extracting randomness using few independent sources. In *Proceedings of the 45th Annual IEEE Symposium on Foundations of Computer Science*, pages 384–393, 2004.
- [BKS⁺05] Boaz Barak, Guy Kindler, Ronen Shaltiel, Benny Sudakov, and Avi Wigderson. Simulating independence: new constructions of condensers, ramsey graphs, dispersers, and extractors. In *Proceedings of the 37th Annual ACM Symposium on Theory of Computing*, pages 1–10, 2005.
- [Blu86] M. Blum. Independent unbiased coin flips from a correlated biased source: a finite Markov chain. *Combinatorica*, 6(2):97–108, 1986.
- [BOL90] M. Ben-Or and N. Linial. Collective coin flipping. In S. Micali, editor, *Randomness and Computation*, pages 91–115. Academic Press, New York, 1990.
- [Bou05] J. Bourgain. More on the sum-product phenomenon in prime fields and its applications. *International Journal of Number Theory*, 1:1–32, 2005.
- [BR94] M. Bellare and J. Rompel. Randomness-efficient oblivious sampling. In *Proceedings of the 35th Annual IEEE Symposium on Foundations of Computer Science*, pages 276–287, 1994.
- [CDH⁺00] Ran Canetti, Yevgeniy Dodis, Shai Halevi, Eyal Kushilevitz, and Amit Sahai. Exposure-resilient functions and all-or-nothing transforms. In Bart Preneel, editor, *Advances in Cryptology — EUROCRYPT 2000*, volume 1807 of *Lecture Notes in Computer Science*, pages 453–469. Springer-Verlag, May 2000.

- [CFG⁺85] B. Chor, J. Friedman, O. Goldreich, J. Håstad, S. Rudich, and R. Smolensky. The bit extraction problem or t -resilient functions. In *Proceedings of the 26th Annual IEEE Symposium on Foundations of Computer Science*, pages 396–407, 1985.
- [CG88] B. Chor and O. Goldreich. Unbiased bits from sources of weak randomness and probabilistic communication complexity. *SIAM Journal on Computing*, 17(2):230–261, 1988.
- [Cra37] H. Cramer. On the order of magnitude of the difference between consecutive prime numbers. *Acta Arithmetica*, pages 23–46, 1937.
- [CW79] J. L. Carter and M. N. Wegman. Universal classes of hash functions. *Journal of Computer and System Sciences*, 18:143–154, 1979.
- [CW89] A. Cohen and A. Wigderson. Dispersers, deterministic amplification, and weak random sources. In *Proceedings of the 30th Annual IEEE Symposium on Foundations of Computer Science*, pages 14–19, 1989.
- [Dod00a] Yevgeniy Dodis. *Exposure-Resilient Cryptography*. PhD thesis, MIT, 2000.
- [Dod00b] Yevgeniy Dodis. Impossibility of black-box reduction from non-adaptively to adaptively secure coin-flipping. Unpublished manuscript, April 2000.
- [DR05] Z. Dvir and R. Raz. Analyzing linear mergers. Technical Report TR05-25, ECCCC: Electronic Colloquium on Computational Complexity, 2005.
- [GRS04] A. Gabizon, R. Raz, and R. Shaltiel. Deterministic extractors for bit-fixing sources by obtaining an independent seed. In *Proceedings of the 45th Annual IEEE Symposium on Foundations of Computer Science*, pages 394–403, 2004.
- [JK99] B. Jun and P. Kocher. The intel random number generator, 1999. <http://www.cryptography.com/resources/whitepapers/IntelRNG.pdf>.
- [KM04] Robert Koenig and Ueli Maurer. Extracting randomness from generalized symbol-fixing and markov sources. In *Proceedings of 2004 IEEE International Symposium on Information Theory*, page 232, June 2004.
- [KM05] Robert Koenig and Ueli Maurer. Generalized strong extractors and deterministic privacy amplification. In Nigel Smart, editor, *Cryptography and Coding 2005*, volume 3796 of *Lecture Notes in Computer Science*, pages 322–339. Springer-Verlag, December 2005.
- [KZ03] Jesse Kamp and David Zuckerman. Deterministic extractors for bit-fixing sources and exposure-resilient cryptography. In *Proceedings of the 44th Annual IEEE Symposium on Foundations of Computer Science*, pages 92–101, 2003.
- [LLS89] D. Lichtenstein, N. Linial, and M. Saks. Some extremal problems arising from discrete control processes. *Combinatorica*, 9(3):269–287, 1989.
- [Lov96] L. Lovász. Random walks on graphs: A survey. In D. Miklós, V. T. Sós, and T. Szőnyi, editors, *Combinatorics, Paul Erdős is Eighty, Vol. 2*, pages 353–398. J. Bolyai Math. Soc., Budapest, 1996.

- [MW97] Ueli Maurer and Stefan Wolf. Privacy amplification secure against active adversaries. In Burton S. Kaliski Jr., editor, *Advances in Cryptology — CRYPTO '97*, volume 1294 of *Lecture Notes in Computer Science*, pages 307–321. Springer-Verlag, August 1997.
- [NTS99] N. Nisan and A. Ta-Shma. Extracting randomness: A survey and new constructions. *Journal of Computer and System Sciences*, 58:148–173, 1999.
- [NZ96] N. Nisan and D. Zuckerman. Randomness is linear in space. *Journal of Computer and System Sciences*, 52(1):43–52, 1996.
- [Rao06] A. Rao. Extractors for a constant number of polynomial min-entropy independent sources. In *Proceedings of the 38th Annual ACM Symposium on Theory of Computing*, 2006.
- [Raz05] Ran Raz. Extractors with weak random seeds. In *Proceedings of the 37th Annual ACM Symposium on Theory of Computing*, pages 11–20, 2005.
- [RRV99] R. Raz, O. Reingold, and S. Vadhan. Extracting all the randomness and reducing the error in Trevisan’s extractors. In *Proceedings of the 31st Annual ACM Symposium on Theory of Computing*, pages 149–158, 1999.
- [RRV02] R. Raz, O. Reingold, and S. Vadhan. Extracting all the randomness and reducing the error in trevisan’s extractors. *Journal of Computer and System Sciences*, 65(1):97–128, 2002.
- [Sha02] Ronen Shaltiel. Recent developments in explicit constructions of extractors. *Bulletin of the European Association for Theoretical Computer Science*, (77):67–95, June 2002.
- [Sha05] R. Shaltiel. How to get more mileage from randomness extractors. Technical Report TR05-145, ECCC: Electronic Colloquium on Computational Complexity, 2005.
- [SV86] M. Santha and U. V. Vazirani. Generating quasi-random sequences from semi-random sources. *Journal of Computer and System Sciences*, 33:75–87, 1986.
- [Tre01] L. Trevisan. Extractors and pseudorandom generators. *Journal of the ACM*, pages 860–879, 2001.
- [TS96] A. Ta-Shma. On extracting randomness from weak random sources. In *Proceedings of the 28th Annual ACM Symposium on Theory of Computing*, pages 276–285, 1996.
- [TV00] Luca Trevisan and Salil P. Vadhan. Extracting randomness from samplable distributions. In *Proceedings of the 41st Annual IEEE Symposium on Foundations of Computer Science*, pages 32–42, 2000.
- [Vad04] S. Vadhan. On constructing locally computable extractors and cryptosystems in the bounded-storage model. *Journal of Cryptology*, 17(1):43–77, Winter 2004.
- [Vaz86] U. V. Vazirani. *Randomness, Adversaries and Computation*. PhD thesis, EECS, University of California at Berkeley, 1986.
- [Vaz87] Umesh V. Vazirani. Strong communication complexity or generating quasirandom sequences from two communicating semirandom sources. *Combinatorica*, 7(4):375–392, 1987.

- [vN51] J. von Neumann. Various techniques used in connection with random digits. *National Bureau of Standards, Applied Mathematics Series*, 12:36–38, 1951.
- [VV85] U. V. Vazirani and V. V. Vazirani. Random polynomial time is equal to slightly-random polynomial time. In *Proceedings of the 26th Annual IEEE Symposium on Foundations of Computer Science*, pages 417–428, 1985.
- [Wor99] N. C. Wormald. *The differential equation method for random graph processes and greedy algorithms*, pages 73–155. PWN, Warsaw, 1999.
- [WZ99] A. Wigderson and D. Zuckerman. Expanders that beat the eigenvalue bound: Explicit construction and applications. *Combinatorica*, 19(1):125–138, 1999.
- [Zuc96] D. Zuckerman. Simulating BPP using a general weak random source. *Algorithmica*, 16:367–391, 1996.