Journal of Computer and System Sciences 77 (2011) 191–220

Contents lists available at ScienceDirect

Journal of Computer and System Sciences

www.elsevier.com/locate/jcss

Deterministic extractors for small-space sources ☆

Jesse Kamp [a,1], Anup Rao [b,2], Salil Vadhan [c,3], David Zuckerman [d,*,4]

[a] Oracle Corporation, 500 Oracle Parkway, Redwood Shores, CA 94065, United States
[b] Institute for Advanced Study, Princeton, NJ 08540, United States
[c] School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138, United States
[d] Department of Computer Science, University of Texas, Austin, TX 78712, United States

ARTICLE INFO

Article history:
Received 25 January 2009
Received in revised form 11 May 2010
Accepted 7 June 2010
Available online 11 June 2010

Keywords:
Randomness extractors
Pseudorandomness
Markov chains
Samplable sources
Bit-fixing sources
Independent sources

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. Specifically, 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 (in variation distance) from space s sources with min-entropy δn, where $s = \Omega(\zeta^3 n)$. Previously, nothing was known for $\delta \leqslant 1/2$, even for space 0. Our results are obtained by a reduction to the class of *total-entropy* 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 smaller sources over $\{0,1\}^\ell$, where the total min-entropy over all the smaller sources is k. We give deterministic extractors for such sources when k is as small as $\text{polylog}(r)$, for small enough ℓ.

Contents

☆ A preliminary version of this paper appeared in the *Proceedings of the 38th Annual ACM Symposium on Theory of Computing*, 2006, pages 691–700.

* Corresponding author.
E-mail addresses: jesse.kamp@oracle.com (J. Kamp), arao@ias.edu (A. Rao), salil@eecs.harvard.edu (S. Vadhan), diz@cs.utexas.edu (D. Zuckerman).

[1] Supported in part by NSF Grant CCR-0310960.
[2] Supported in part by NSF Grants CCR-0310960, CCF-0832797, and DMS-0835373.
[3] Supported by NSF grant CCF-0133096, ONR grant N00014-04-1-0478, and US-Israel BSF grant 2002246.
[4] Supported in part by a David and Lucile Packard Fellowship for Science and Engineering, NSF Grants CCR-0310960, CCF-0634811, and CCF-0916160, a Radcliffe Institute Fellowship, and a Guggenheim Fellowship.

0022-0000/$ – see front matter
doi:10.1016/j.jcss.2010.06.014

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. [19]) uses the extractor of von Neumann [42] 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. [6,33,12,41,11,2,7,23]. However, attempts to handle sources lacking a significant amount of independence were thwarted by results showing that it is impossible to devise a single function that extracts even one bit of randomness from sufficiently general classes of sources [33].

These impossibility results led researchers to focus on the weaker task of simulating probabilistic algorithms with weak random sources [43,12,39,14,46]. This line of work culminated in the introduction, by Nisan and Zuckerman [27], 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 [26,31] 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. [37,16]. 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 [10,16]. 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 [12,4,5,29,28]; *bit-fixing sources*, where some of the bits are perfectly random and the rest are fixed [11,14,22,18]; and *samplable sources*, where the source is generated by an efficient algorithm [37]. Our work relates to all of these models; indeed, we establish connections between them. However, our main motivation is a particular form of samplable sources — namely ones generated by algorithms that have small space.

Before proceeding, we recall a few standard definitions. A *source* is a probability distribution. 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 $|X_1 - X_2| = \max_{S\subseteq\Omega} |\Pr[X_1 \in S] - \Pr[X_2 \in S]| = \frac{1}{2}\sum_{s\in\Omega} |\Pr[X_1 = s] - \Pr[X_2 = s]|$.

Definition 1.1. A function $\mathrm{Ext}: \{0,1\}^n \to \{0,1\}^m$ is an *ϵ-extractor* for a class $\mathcal{X}$ of random sources if for every $X \in \mathcal{X}$, $\mathrm{Ext}(X)$ is ϵ-close to uniform in variation distance.

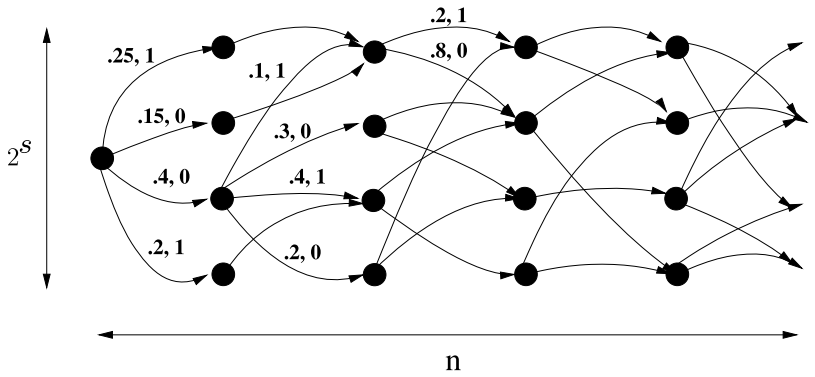

Fig. 1. Part of a space $s = 2$ source.

1.1. Small-space sources

Trevisan and Vadhan [37] 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 [42], Blum's finite Markov chain model [6], symbol-fixing sources [22], and sources that consist of many independent sources.[5] 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 [6], 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 [42] 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.

We study a generalization of the Markov chain model to time-dependent Markov chains. This yields a much richer class of sources, and is similar to models previously considered by Vazirani [40] and Koenig and Maurer [20,21]. Our model of a space s source is basically a source generated by a width 2^s branching program. More specifically, at each step the process generating the source is in one of 2^s states. We model this 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. (See Fig. 1.) Unlike in Blum's model [6], the transitions can be different at each time-step. Our model is also related to the trellis representation of error-correcting codes.

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 possible.

Vazirani [40] gave explicit extractors for space-bounded sources in which every bit has bounded bias conditioned on the previous state of the algorithm. (This is a space-bounded analogue of semi-random sources [34].) Koenig and Maurer [20,21] gave the first explicit constructions of extractors for space-bounded sources where we only assume a lower bound on the total min-entropy. 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 1. In fact, for space 0 sources, which are simply sources of independent bits each of which has a different and unknown bias, the only other extractor we know for low min-entropy is parity, which outputs just 1 bit.

[5] Any source consisting of t independent (flat) sources of min-entropy k can be computed in our model using space $s = k$. We show that (nonconstructive) extractors for small-space sources exist provided that the total min-entropy is greater than $2s + O(\log n)$, which in turn yield good extractors for t independent sources of min-entropy $k = s$ provided $t \geqslant 3$.

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
Theorem 1.2	$\delta \geqslant n^{-c}$	$c\delta^3 n$	$\exp(-n^c)$
Theorem 1.3	Any constant δ	cn	$\exp(-\widetilde{\Omega}(n))$
Theorem 1.4	$\delta \geqslant C/\log n$	$c\delta \log n$	$\exp(-n^{.99})$
Theorem 1.5 (nonconstructive)	$\delta \geqslant 2\log n/n$	$(\delta n)/2.01$	$\exp(-\Omega(\delta n))$

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 and stated more precisely below.

Our first extractor extracts whenever $\delta > n^{-\eta}$ for some fixed constant η and extracts almost all of the entropy.

Theorem 1.2. *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 $\mathsf{Ext} : \{0,1\}^n \to \{0,1\}^m$ for space s sources with min-entropy rate δ, where $s = \Omega(\zeta^3 n)$, $m = (\delta - \zeta)n$, and $\epsilon = 2^{-n^{\Omega(1)}}$.*

We also have a simpler construction for constant min-entropy rate, which achieves somewhat better error.

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

We give an alternate construction for min-entropy rate $\delta = \Omega(1/\log n)$ and space $O(\delta \log n)$, although for most parameters the previous constructions will dominate.

Theorem 1.4. *For every $n \in \mathbb{N}$ and $\delta > \zeta > 28/\log n$ and $s \leqslant (\zeta \log n)/28$, there is a polynomial-time computable ϵ-extractor $\mathsf{Ext} : \{0,1\}^n \to \{0,1\}^m$ 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. [20,21]) we have reduced the min-entropy required from $n/2$ to $n^{1-\Omega(1)}$ (in Theorem 1.2). However, we are still far from achieving what is possible nonconstructively:

Theorem 1.5. *For space s sources with min-entropy k, a function $f : \{0,1\}^n \to \{0,1\}^m$ chosen uniformly at random is an ϵ-extractor with output length $m = k - 2\log(1/\epsilon) - O(1)$ with probability at least $1 - \exp(-\Omega(2^k \epsilon^2))$, as long as $k \geqslant 2s + \log s + 2\log n + 3\log(1/\epsilon) + O(1)$.*

Note that here the min-entropy can be as small as $O(\log n)$, while our results require min-entropy nearly linear in n. In addition, 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 $\delta^3 n$.

The constant factor 2 in the min-entropy bound in Theorem 1.5 is tight. However, if we restrict to small-space sources where all transition probabilities are multiples of some fixed constant, e.g. $1/2$, then we can reduce the bound to $k \geqslant s + \log s + \log n + 2\log(1/\epsilon) + O(1)$.

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-entropy independent sources

Our extractors for small-space sources are all obtained via a reduction to a new model of sources we introduce called *total-entropy independent sources*. The reduction we use is based on one of Koenig and Maurer [20,21], who used it to show how reduce the task of extracting from two sources of "bounded dependency" to extracting from two independent sources. Total-entropy 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-entropy 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 [11,22] and independent sources [12,4], 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 [12]. 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 [4,5,29,28]. 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-entropy 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-entropy 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 [11,14] but required the source to have a large number of random positions. Kamp and Zuckerman [22] 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 [18] 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 [22].

Total-entropy independent sources can be seen as a generalization of symbol-fixing sources, where each symbol is viewed as a separate source.[6] The difference is that instead of each symbol being only fixed or uniformly random, the symbols (sources) in total-entropy 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-entropy independent sources are all based on generalizing various techniques from extractors for independent and symbol-fixing sources.

Koenig and Maurer [20,21] 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 technique generalizes to extractors for more than two sources. We show that it 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 [3] (see Bourgain [8]), and thereby obtain extractors for total-entropy 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 least significant bits of the product of these finite field elements.

We also show how to use ideas from the work of Rao [28] for extracting from several independent sources, together with recent constructions of randomness-efficient condensers [5,29], to get extractors for total-entropy independent sources that extract from sources of min-entropy $(r\ell)^{1-\Omega(1)}$.

When the smaller sources each have short length ℓ, we use ideas from the work of Kamp and Zuckerman [22] about bit-fixing sources to construct extractors for total-entropy independent sources with min-entropy k. We can extract many bits when $k > 2^{\ell}\sqrt{r\ell}$, and for $k = \Omega(2^{2\ell}\ell)$ 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 [22]. 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 [22].

Unlike the first two extractors, the extractors obtained using this technique use the full generality of the total-entropy independent sources. In the first two constructions, using a Markov argument we can essentially first reduce the total-entropy 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. [18] to total-entropy independent sources to enable us to extract more of the entropy. Note that we use it to improve not only the extractors based on [22] (analogous to what was done in [18] for bit-fixing sources), but also our extractors based on techniques developed for independent sources. Independently of our work, Shaltiel [32] has recently generalized the ideas in [18] 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 techniques based on [18] to our extractors that use the independent sources techniques of Rao [28], the results of [3], and two different bit-fixing source extractors from [22], respectively, we get the following four theorems. The

[6] Though for ease of presentation we define total-entropy 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 more general case.

Table 2
Total-entropy 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, and γ is a variable parameter that can be set to any desired value in $(0,1)$.

Reference	Min-entropy rate	Error
Theorem 1.6	$\delta \geqslant 1/(r\ell)^c$	$\exp(-(r\ell)^c)$
Theorem 1.7	Any constant δ	$\exp(-\tilde{\Omega}(r\ell))$
Theorem 1.8 ($\ell = o(\log r)$)	$\delta \geqslant 1/(r\ell)^{(1-\gamma-o(1))/2}$	$\exp(-(r\ell)^\gamma)$
Theorem 1.9	$\delta = (2^\ell \log r)^C / r$	$(\delta r\ell)^{-c}$
Theorem 1.10 (nonconstructive)	$\delta \geqslant 1.01(\ell + \log r)/(\ell r)$	$\exp(-\Omega(\delta \ell r))$

first three of these theorems are directly used to obtain the small-space extractors from Theorems 1.2, 1.3, and 1.4. Table 2 presents a summary of these extractors.

Theorem 1.6. *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* $\mathsf{Ext} : (\{0,1\}^\ell)^r \to \{0,1\}^m$ *for sets of r independent sources over $\{0,1\}^\ell$ with total min-entropy rate δ, 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 [4] as a corollary.

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

Theorem 1.7. *For any constants $\delta > \zeta > 0$ and every $r, \ell \in \mathbb{N}$, there is a polynomial-time computable ϵ-extractor* $\mathsf{Ext} : (\{0,1\}^\ell)^r \to \{0,1\}^m$ *for sets of r independent sources over $\{0,1\}^\ell$ with total min-entropy rate δ, where $m = (\delta - \zeta) r\ell$ and $\epsilon = \exp(-\Omega((r\ell)/\log^3(r\ell)))$.*

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 1.8. *For every $r, \ell \in \mathbb{N}$ such that $1 \leqslant \ell \leqslant \frac{1}{2}\log r$ and $\delta > \zeta > \sqrt{2^{2\ell}\log^3 r / r\ell}$ there is a polynomial-time computable ϵ-extractor* $\mathsf{Ext} : (\{0,1\}^\ell)^r \to \{0,1\}^m$ *for r independent sources over $\{0,1\}^\ell$ of total min-entropy rate δ, where $m = (\delta - \zeta) r\ell$ and $\epsilon = \exp(-\Omega((\zeta^2 r\ell)/(2^{2\ell}\log^3 r)))$.*

Our last extractor for total-entropy sources works even for polylogarithmic min-entropy k, provided ℓ is small enough:

Theorem 1.9. *There exists a constant $C > 0$ such that for every $r, \ell, k \in \mathbb{N}$ such that $k \geqslant (2^\ell \log r)^C$, there exists a polynomial-time computable ϵ-extractor* $\mathsf{Ext} : (\{0,1\}^\ell)^r \to \{0,1\}^m$ *for r independent sources 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 $\ell + \log r$:

Theorem 1.10. *For total-entropy k independent sources, a function $f : (\{0,1\}^\ell)^r \to \{0,1\}^m$ chosen uniformly at random is an ϵ-extractor with output length $m = k - 2\log(1/\epsilon) - O(1)$ with probability $1 - \exp(-\Omega(2^k \epsilon^2))$ as long as $k \geqslant \max\{\ell, \log\log(r/\epsilon)\} + \log r + 2\log(1/\epsilon) + O(1)$.*

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 1.9 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-entropy independent sources. We then restrict our focus to extracting from total-entropy independent sources, starting with the basic extractors. In Section 4 we describe the extractor that provides the basis for the extractor from Theorem 1.7. In Section 5 we describe the extractor that provides the basis for the extractor from Theorem 1.6. In Section 6 we describe the extractors that provide the basis for the extractors from Theorem 1.8 and Theorem 1.9. Then in Section 7, we describe how to generalize the techniques of Gabizon et al. [18] 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-entropy 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 are random variables $\{X_i\}$ and nonnegative real numbers γ_i such that $\sum_i \gamma_i = 1$, $X = \sum_i \gamma_i X_i$ (where we identify random variables with the probability mass vectors), and each random variable X_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.*

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

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

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 [20,21]. This is not quite the most general model of such space-bounded 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, it is not hard to transform the sources in the more general model 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 smaller sources on $\{0,1\}^\ell$ is an *independent source* on $(\{0,1\}^\ell)^r$ if each of the r smaller sources 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 $\delta r\ell$ and total-entropy k if the total min-entropy is k.

Definition 2.7. A source on $\{0,1\}^\ell$ is *flat* if it is uniformly distributed over a non-empty subset of $\{0,1\}^\ell$. In particular, a *flate independent source* is uniform on a cross product of sets.

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*[7] (*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 $X_1, \ldots, X_m$ 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 $X_1, \ldots, X_m$ 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.

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 $\mathsf{Ext} : \{0,1\}^n \times \{0,1\}^s \to \{0,1\}^m$ is a *seeded ϵ-extractor* for a set of random sources $\mathcal{X}$, if for every $X \in \mathcal{X}$, $\mathsf{Ext}(X, U_s)$ is ϵ-close to uniform. The extractor is called *strong* if for Y chosen according to U_s, $Y \circ \mathsf{Ext}(X, Y)$ 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. *(See [35,30].) For every $n, k \in \mathbb{N}$, $\epsilon > 0$, there is a polynomial-time computable strong seeded ϵ-extractor* $\mathsf{Ext} : \{0,1\}^n \times \{0,1\}^t \to \{0,1\}^{k - O(\log^3(n/\epsilon))}$ *for sources with min-entropy k, with $t = O(\log^3(n/\epsilon))$.*

3. Small-space sources as convex combinations of independent sources

Following Koenig and Maurer [20,21], we show how small-space sources can be decomposed 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 smaller sources on $\{0,1\}^t$. It can be shown that this conditioning reduces the min-entropy of the source by at most roughly $s \cdot (n/t)$ (with high probability), and thus we obtain a total-entropy source.

Koenig and Maurer [20,21] applied this reduction for partitioning into 2 blocks and thereby reduced extraction from small-space sources of min-entropy rate greater than 1/2 to the well-studied problem of extracting from two independent sources, each of which has some min-entropy. (Min-entropy rate greater than 1/2 is needed, or else all of the min-entropy may be contained in just one of the blocks and deterministic extraction is impossible.) In this paper, we handle lower min-entropy rates by partitioning into *many* shorter blocks; although this reduces the min-entropy by more, it ensures that the total min-entropy is spread among several of the independent blocks and thus deterministic extraction is possible.

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-entropy 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 [25], although it has been used implicitly earlier in, e.g., [45]).

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

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

[7] This definition is slightly different from the original one used by Ta-Shma [36]. 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.

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 for each y, $(X|Y=y)$ is a set of independent smaller sources with each block viewed as a smaller source 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 $(1-\alpha)\delta$. □

4. Extracting from total-entropy independent sources by reducing to standard independent sources

In this section, we show how to construct extractors for total-entropy 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 smaller 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 smaller sources on $\{0,1\}^{\ell r/t}$ where at least $\delta t/2$ of the smaller sources 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, after dividing by the length of the source we get $\delta \leqslant (b+(t-b)(\delta/2))/t$. A simple calculation then gives the desired result. □

Once we are in this model, we can generalize the result from Koenig and Maurer [20,21] 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 $\mathsf{Ext}(x_1,x_2,\ldots,x_b)$ be an extractor for b independent sources over $\mathcal{G}$, each of which has min-entropy rate at least δ. Suppose Ext has the form $\mathsf{Ext}(x_1,x_2,\ldots,x_b) := f(x_1 * x_2 * \cdots * x_b)$ for some f. Then $F(x_1,\ldots,x_r) := f(x_1 * \cdots * x_r)$ is an extractor for r independent sources over $\mathcal{G}$, b of which have min-entropy rate at least δ.*

The proof is simple and is the same as in [20,21]. 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, since the group operation cannot lower the entropy.

Bourgain, Glibichuk, and Konyagin [3] gave bounds on the exponential sums of the function $\prod_{i=1}^b x_i$ over large subsets of fields without large subfields, in particular $\mathrm{GF}(p)$ and $\mathrm{GF}(2^p)$ for p prime. This estimate gives an extractor for b independent sources where each source has high entropy via Vazirani's XOR lemma [39].

Theorem 4.3. *(See [3].) For every $\delta > 0$, there exist $b = b(\delta)$, $c = c(\delta) \in \mathbb{N}$ such that the following holds. Let K be a finite field of the form $\mathrm{GF}(p)$ or $\mathrm{GF}(2^p)$ for a prime p. Then the function $\mathsf{BGK}(x_1,\ldots,x_b)$ that outputs the m least significant bits[8] of the product $\prod_i x_i$ is an ϵ-extractor for b independent sources over K with min-entropy rate δ, for some $m = \Omega(c\log|K|)$ and $\epsilon = 2^{-\Omega(m)}$.*

Note that for constant δ, we can extract $\Theta(\log|K|)$ bits with only a constant number of sources. Using the explicit relationship between δ and the number of sources and entropy from [3], we can handle slightly subconstant δ, down to $\delta = \Omega(1/(\log\log|K|)^{(1/C)})$ for some constant C.

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. *For every $\delta > 0$, there exist $b = b(\delta)$, $c = c(\delta) \in \mathbb{N}$ such that the following holds. Let K be a finite field of the form $\mathrm{GF}(p)$ or $\mathrm{GF}(2^p)$ for a prime p, let $r \in \mathbb{N}$, and define $f : K^r \to \{0,\ldots,|K|-1\}$ by setting $f(x_1,\ldots,x_r)$ to equal $\prod_i x_i$, viewed as an integer from 0 to $|K|-1$. Then the function $\mathsf{BGK}(x_1,\ldots,x_r) = \lfloor (2^m f(x_1,\ldots,x_r))/|K| \rfloor$ is an ϵ-extractor for r independent sources over K, at least b of which have min-entropy rate δ, for some $m = \Omega(c\log|K|)$ 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 $\nu = \nu(\delta)$, such that for every $\ell, r \in \mathbb{N}$, there is a polynomial-time computable function $\mathsf{BGK} : (\{0,1\}^\ell)^r \to \{0,1\}^m$ that is an ϵ-extractor for r independent sources on $\{0,1\}^\ell$, at least ν of which have min-entropy rate δ, for some $m = \Omega(\ell)$ and $\epsilon = 2^{-\Omega(\ell)}$.*

[8] Here the least significant bits of an element in $\mathrm{GF}(2^p)$ are simply the coefficients of the low degree terms when the element is viewed as a polynomial of degree smaller than p in $\mathrm{GF}(2)[X]$.

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

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 every constant $\delta > 0$, we can construct a polynomial-time computable ϵ-extractor $\mathsf{Ext} : (\{0,1\}^\ell)^r \to \{0,1\}^m$ for total-rate δ independent sources on $(\{0,1\}^\ell)^r$, with $m = \Omega(r\ell)$ and $\epsilon = 2^{-\Omega(m)}$. This extractor can be computed in time* $\mathrm{poly}(r,\ell)$.

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

5. Extracting from polynomial entropy rate

In this section we will show how to extract from total-entropy 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. *There is a constant $\beta > 0$ such that for every $\ell \in \mathbb{N}$ and $\delta \geqslant \ell^{-\beta}$, there exists a polynomial-time computable ϵ-extractor $\mathsf{Ext} : (\{0,1\}^\ell)^r \to \{0,1\}^m$ for total-rate δ independent sources on $(\{0,1\}^\ell)^r$, with $r = \Omega(1/\delta^2)$, $m = \ell^{\Omega(1)}$ and $\epsilon = 2^{-\ell^{\Omega(1)}}$.*

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

Corollary 5.2. *There exists a constant $\eta > 0$ such that for every $r, \ell \in \mathbb{N}$, $\delta \geqslant (r\ell)^{-\eta}$, there exists a polynomial-time computable ϵ-extractor $\mathsf{Ext} : (\{0,1\}^\ell)^r \to \{0,1\}^m$ for total-rate δ independent sources on $(\{0,1\}^\ell)^r$, with $m = (\delta^2 r\ell)^{\Omega(1)}$ and $\epsilon = 2^{-(\delta^2 r\ell)^{\Omega(1)}}$.*

Proof. Let $r' = \Omega(1/\delta^2)$ be the number of sources that the extractor of Theorem 5.1 can handle. Divide the source into r' blocks of $r/r' = O(\delta^2 r)$ smaller sources each and apply Theorem 5.1. □

In this section we will describe a generic technique to turn any extractor for the model where a few smaller sources 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 [5,29]. 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. It won't actually be a set of aligned sources; instead, it will be a convex combination of sets of aligned sources, which will be good enough. 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. *For every constant $B \in \mathbb{N}$ and every sufficiently small constant α, there exist constants γ and $\mu > 2\gamma$ for which the following holds. For every $\ell \in \mathbb{N}$, there is a polynomial-time computable function $\mathsf{ACond} : \{0,1\}^\ell \to (\{0,1\}^{\ell^\mu})^{\ell^\gamma}$ such that if $X_1, \ldots, X_B$ are independent sources on $\{0,1\}^\ell$ of min-entropy rate $\delta = \ell^{-\alpha}$, then*

$$\mathsf{ACond}(X_1), \mathsf{ACond}(X_2), \ldots, \mathsf{ACond}(X_B)$$

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 [28]. 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. *There exists a constant $v \in \mathbb{N}$ such that the following holds for every constant $\gamma < 1$. For every $\ell, u \in \mathbb{N}$, there is a $2^{-\ell^{\Omega(1)}}$-extractor $\mathsf{SRExt} : (\{0,1\}^{\ell^\gamma \times \ell})^u \to \{0,1\}^m$ for u independent sources, of which v are independent aligned $(\ell^\gamma \times \ell)$ SR-sources, where $m = \ell - \ell^{1-\Omega(1)}$.*

We will first describe how to use these two ingredients to extract from an intermediate model. Then we will see that total-entropy 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 smaller sources $X^1, \ldots, X^{u^2}$, each on $\{0,1\}^\ell$. These smaller sources are partitioned into u sets $S_1, \ldots, S_u$ of u sources each, such that v of the sets have the property that at least v of their sources 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. *There are constants $v \in \mathbb{N}$ and $\alpha > 0$ such that for every $\ell \in \mathbb{N}$ there exists a polynomial-time computable $2^{-\ell^{\Omega(1)}}$-extractor* IExt *for (u, v, α) intermediate sources, with $m = \ell^{\Omega(1)}$.*

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

Proof of Theorem 5.1. Let $X = X_1, \ldots, X_r$ be an independent source on $(\{0,1\}^\ell)^r$ with total min-entropy rate $\delta \geqslant 4\ell^{-\alpha}$, where α is the constant from Theorem 5.6 and $r = u^2$ where u will be chosen later. Divide the source into u blocks with u smaller sources 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 smaller source 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 $\delta/4 \geqslant \ell^{-\alpha}$, for large enough ℓ. 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. □

Now we prove Theorem 5.6:

Proof of Theorem 5.6. We begin by describing the extractor. Let v be the constant that we will pick later. We use the following ingredients:

- Let BGK be as in Corollary 4.5 — an extractor for independent sources when $v - 1$ of the smaller sources have min-entropy.
- Let ACond be as in Theorem 5.3, letting $B = v^2$ — a condenser that converts sources with min-entropy rate $\ell^{-\alpha}$ into a convex combination of aligned sources consisting of ℓ^γ sources of length ℓ^μ, with somewhere-min-entropy rate 0.9, for appropriate constants α, γ, and μ, where $\mu > 2\gamma$.
- 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.

The extractor works as follows:

Construction. $\mathsf{IExt}(x^1, \ldots, x^{u^2})$
Input: $x^1, \ldots, x^{u^2}$ partitioned into sets $S_1, \ldots, S_u$
Output: z.

1. Compute $y^i = \mathsf{ACond}(x^i)$ for every source in the input. Let y^i_j denote the jth row of y^i.
2. For every $l \in [u]$, and every $j \in [2^{\ell^\gamma}]$, let b^l_j be the string obtained by applying BGK using the y^i_j from all $i \in S_l$ as input. We think of b^l as a sample from an SR-source with ℓ^γ rows.
3. Output $\mathsf{SRExt}(b^1, \ldots, b^u)$.

Now we analyze the extractor. If we restrict our attention to the v^2 high min-entropy smaller sources, from Theorem 5.3 we know that from the first step from these smaller sources 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 that contains v high min-entropy smaller sources.

Thus the result of the first two steps in the algorithm is a distribution that is $2^{-\ell^{\Omega(1)}}$-close to a convex combination of collections of u independent sources, 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. □

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 [47], which improves on the condenser from [5].

Lemma 5.7. *(See [47].) There is a constant $\alpha > 0$ such that for every $t, \ell \in \mathbb{N}$, there exists a polynomial-time computable condenser $\mathsf{Zuck} : \{0,1\}^\ell \to (\{0,1\}^{(2/3)^t \ell})^{2^t}$ such that if X has min-entropy rate δ, $\mathsf{Zuck}(X)$ is $t2^{-\Omega(\alpha\delta\ell)}$-close to somewhere-min-entropy rate $\min((1+\alpha)^t\delta, 0.9)$.*

We'll also need to use the condenser from Raz's work [29] with the improved analysis of Dvir and Raz (Lemma 3.2 in [17]), which shows that most of the output rows are statistically close to having high min-entropy.

Lemma 5.8. *(See [17].) For any constant $c > 0$ and every $\ell, r \in \mathbb{N}$, there is a polynomial-time computable function $\mathsf{Raz} : (\{0,1\}^\ell)^r \to (\{0,1\}^{\Omega(\ell)})^{2^{O(r)}}$ such that the following holds. If the input source X has somewhere-min-entropy rate δ, the output $\mathsf{Raz}(X)$ 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 $2^{O(r)}$ 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 5.9. *For every constant $c > 0$, there is a constant $C \in \mathbb{N}$, such that for every $\ell \in \mathbb{N}$ there exists a polynomial-time computable function $\mathsf{Cond} : \{0,1\}^\ell \to (\{0,1\}^{\Omega(\ell)})^C$ with the following property. If the input source X has min-entropy rate at least δ, the output $\mathsf{Cond}(X)$ 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 C rows have min-entropy rate at least $\min(2\delta, 0.9)$.*

Proof. Let $\mathsf{Cond}(x) = \mathsf{Raz}(\mathsf{Zuck}(x))$, picking t large enough in Lemma 5.7 so that $0.9(1+\alpha)^t \geqslant 2$. □

Now we can use this basic condenser to help prove Theorem 5.3. To do this, we apply this condenser to our input smaller sources 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 we will be applying the condenser many more than $1/c$ times. 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 the analysis of the recursion on that one aligned row.

Proof of Theorem 5.3. First, apply the function Cond from Lemma 5.9 to each X_i, choosing $c < 1/B$. Then the output $\langle \mathsf{Cond}(X_1), \mathsf{Cond}(X_2), \ldots, \mathsf{Cond}(X_B)\rangle$ is $2^{-\Omega(\delta\ell)}$-close to a convex combination of distributions $Y = \sum_j \beta_j Y^{(j)}$, where $Y^{(j)} = \langle Y_1^{(j)}, Y_2^{(j)}, \ldots, Y_B^{(j)}\rangle$ and $\sum_j \beta_j = 1$. Each smaller source $Y_i^{(j)}$ 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 $Y^{(j)}$ in the convex combination. In this source, at most a $cB < 1$ fraction of the rows have a smaller source $Y_i^{(j)}$ 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 smaller sources is at least 2δ, i.e., the output is aligned with somewhere-min-entropy rate $\min(2\delta, 0.9)$. 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 $t = \log(0.9/\delta) = \log(0.9\ell^\alpha)$, then these sources have somewhere-min-entropy rate 0.9. The total number of sources we ultimately construct is $C^t = \ell^\gamma$ for $\gamma = O(\alpha)$, and the length of each source is $\ell/2^{O(t)} = \ell^\mu$ for $\mu = 1 - O(\alpha)$. If we choose α small enough, then we can achieve $\mu > 2\gamma$, as desired. □

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 [28]. The additional challenge we face is that whereas in [28] 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.

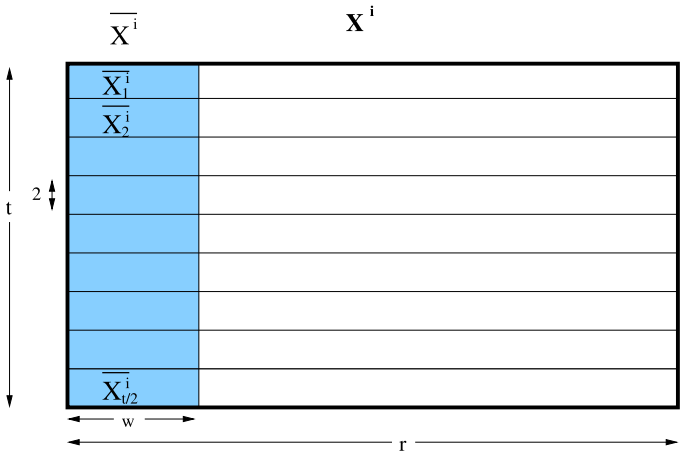

Fig. 2. Notation in one source.

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 5.10. 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 [8]. 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. The ingredients are the following:

- 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 $\{0,1\}^{t\ell}$ with min-entropy at least $\ell - l$ and to have output length m, using seed length a. Let ϵ_2 be the error of Ext.

Construction. $\mathsf{Cond}(x^1, \ldots, x^u)$
Input: $x^1, \ldots, x^u$, strings each divided into t rows of length r.
Output: $z^1, \ldots, z^u$.

1. For each source, group its rows into pairs of rows.
2. For $i = 1, 2, \ldots, u$, and $j = 1, 2, \ldots, t/2$ let $\bar{x}^i_j$ 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 $\bar{x}^{\neq i}_j$ denote the first w bits of the j'th pair rows of all sources except the i'th source. (See Fig. 2.)
3. For every $i = 1, 2, \ldots, u$, and $j = 1, 2, \ldots, t/2$, let $z^i_j = \mathsf{Ext}(x^i, \mathsf{BGK}(x^{\neq i}_j))$.
4. For every $i = 1, 2, \ldots, u$, let z^i consist of rows $(z^i_1, \ldots, z^i_{t/2})$.

Lemma 5.11. *Let* Cond *be as above. If $X^1, X^2, \ldots, X^u$ are independent sources, with v of them being aligned $(t \times a)$ SR-sources, then $Z^1, Z^2, \ldots, Z^u$ 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 $X^{i_1}, \ldots, X^{i_v}$ contains a random row for some distinct sources $i_1, \ldots, i_v \in [u]$. We will argue that the h'th row of the output sources $Zi_1, \ldots, Z^{i_v}$ is statistically close to uniform.

To see this, consider the random variable $\overline{X} = \overline{X}^1 \circ \cdots \circ \overline{X}^u$, 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 5.12. *For good $\bar{x}$, $(Z^1 \circ \cdots \circ Z^u)|\overline{X} = \bar{x}$ is the distribution of u independent sources, with v of them being $v\sqrt{\epsilon_2}$-close to aligned SR-sources.*

Claim 5.13. $\Pr[\overline{X}$ *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. $X^i|\overline{X} = \bar{x}$ has min-entropy at least $r - l$.
2. $\mathsf{BGK}(\bar{x}_h^{\neq i})$ is a good seed to extract from $X^i|\overline{X} = \bar{x}$, i.e.

$$\left\|\mathsf{Ext}\left(X^i|\overline{X} = \bar{x}, \mathsf{BGK}\left(\bar{x}_h^{\neq i}\right)\right) - U_m\right\| \leqslant \sqrt{\epsilon_2}.$$

We will say that $\bar{x}$ is *good* if it is good for all the v sources $X^{i_1}, \ldots, X^{i_v}$ whose h'th row is random. Claim 5.12 immediately follows from this notion of good. All we have left to prove is Claim 5.13. The proof requires the following simple proposition.

Proposition 5.14. *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 5.13. Fix an i so that X^i is one of the v aligned SR-sources $X^{i_1}, \ldots, X^{i_v}$ whose h'th row is random. We will first argue that $\overline{X}$ is good for X^i with high probability. Then we will use the union bound to claim that $\overline{X}$ is good with high probability.

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

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

$$\left\|\mathsf{Ext}\left(X^i\middle|\overline{X} = \bar{x}, \mathsf{BGK}\left(\bar{x}_h^{\neq i}\right)\right) - U_m\right\| \leqslant \sqrt{\epsilon_2}.$$

The probability that event T does not occur is at most $2^{-l}2^{tw}$. This is because by Proposition 5.14, 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 has 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, $\overline{X}$ is good for $X^{i_1}, \ldots, X^{i_v}$ whose h'th row is random with probability at least $1 - v \cdot (2^{tw-l} + \sqrt{\epsilon_2} + \epsilon_1)$. □

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 to obtain a single row, at which point we XOR the sources together to obtain an almost-uniform output. By Lemma 5.11 the error in each step is $v \cdot (\epsilon_1 + 2\sqrt{\epsilon_2} + 2^{-(l-tw)})$.

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 is $a = \Omega(w)$, 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, the length r of the sources drops additively by $O(l)$. Thus the final output length is at least $\ell - \ell^\beta$ for some $\beta \in (0, 1)$. □

6. Better extractors for total-entropy independent sources with many short smaller sources

Now we show how for sources consisting of many smaller sources 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 smaller sources 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 [22].

We will need the following definition from [22].

Definition 6.1. An independent source on $(\{0,1\}^\ell)^r$ is a (k,ϵ)-*approximate symbol-fixing source* if k of the r smaller sources 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 [22] 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. [24]).

We will also need the following lemma from [22], which shows that we can use a random walk to extract from approximate symbol-fixing sources.

Lemma 6.2. *(See [22].) 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 starting from an arbitrary vertex, and outputs the label of the final vertex.*

6.2. Reducing to flat total-entropy independent sources

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

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 $X = X_1, \ldots, X_r$ 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 $k_i \geqslant 1$, 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 $k' \leqslant k/2$. In this case, we can write X_S as a convex combination of constant sources and we are still left with $(k - k')/\log 3 \geqslant k/(2\log 3)$ bits of entropy in each of our sources, as desired.

From now on we will assume $k' \geqslant k/2$. 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 ith source having one bit of entropy. Then $Y = \sum Y_i$ is a random variable representing the total entropy. Note that $\mathbb{E}[Y] = \sum \mathbb{E}[Y_i] = \sum 2(1-2^{-k_i}) \geqslant \sum k_i = k'$, 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 [1]).

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

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

6.3. Extracting from flat total-entropy independent sources

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

Theorem 6.6. *Let $\ell \geqslant 1$ and $p \geqslant 2^\ell$ a prime. Then $Sum_p : (\{0,1\}^\ell)^r \to [p]$, where $Sum_p(x) = \sum_i x_i \bmod p$ (viewing each ℓ-bit string x_i as a number in $\{0,1,\ldots,2^\ell-1\}$), 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-entropy independent sources.

Corollary 6.7. *Suppose $p \geqslant 2^\ell$ 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 $k \geqslant \Omega(p^2 \log p)$, 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 $\ell \geqslant 1$ and $p \geqslant 2^\ell$ a prime. Then for all sets of flat independent sources $X = X_1, \ldots, X_r$ on $(\{0,1\}^\ell)^r$ with min-entropy k, $Sum_p(x)$ 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 $|X-Y| \leqslant \frac{1}{2}\sqrt{p}\|X-Y\|_2$. 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 $\mathbb{Z}_p$, the *jth additive character* is defined as $\chi_j(a) = e^{2\pi ija/p}$.

Lemma 6.9. *For any random variable W over $\mathbb{Z}_p$,*

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

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

Proof. Let $Y = W - U_p$. Thus Y is a vector with p coordinates, with $\Pr[W=i] - 1/p$ in the ith coordinate. The jth 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

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

Here we used the fact that $\chi_0(y) = 1$, for every y. □

Using the previous lemma we can now prove Theorem 6.6.

Proof. Let $(X_1, \ldots, X_n)$ be a flat independent source on $(\{0,1\}^\ell)^r$ with total min-entropy k, and let $W = \sum_t X_t \bmod p$. Let $W = \sum_{t=1}^r X_t$ and fix $j \neq 0$. Then $|\mathbb{E}[\chi_j(W)]|^2 = \prod_{t=1}^r |\mathbb{E}[\chi_j(X_t)]|^2$. Suppose X_t has min-entropy k_t, so $k = \sum_t k_t$. Then since each X_t is a flat source, X_t is uniformly distributed over $K_t = 2^{k_t}$ values. Our goal is to upper bound $|\mathbb{E}[\chi_j(X_t)]|^2$ over all possible choices of X_t. Doing so, we get

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

where $\odot : \mathbb{C}\times\mathbb{C}\to\mathbb{R}$ denotes the dot product, where the complex numbers are viewed as vectors $\mathbb{R}^2$, and the third line follows from the observation that the dot product is maximized when y is in the same direction as $(\sum_{x\in\mathbb{Z}_p} X_t(x)\chi_j(x))$, in which case we get exactly 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_t is distributed over the K_t values closest to y. Without loss of generality we can assume these values correspond to $x = 0$ to $K_t - 1$ (since we only care about the magnitude). Thus

$$\begin{aligned}
\left|\mathbb{E}\left[\chi_j(X_t)\right]\right|^2 &\leqslant \left|\frac{1}{K_t}\cdot\left(\sum_{x=0}^{K_t-1} e^{2\pi i x/p}\right)\right|^2 = \left|\frac{1}{K_t}\cdot\frac{1-e^{2\pi i K_t/p}}{1-e^{2\pi i/p}}\right|^2 = \left|\frac{1}{K_t}\cdot\frac{e^{\pi i K_t/p}\cdot(e^{-\pi i K_t/p}-e^{\pi i K_t/p})}{e^{\pi i/p}\cdot(e^{-\pi i/p}-e^{\pi i/p})}\right|^2 \\
&= \left(\frac{1}{K_t}\cdot\frac{\sin(\frac{\pi K_t}{p})}{\sin(\frac{\pi}{p})}\right)^2 = \left(\frac{1}{K_t}\cdot\frac{(\pi K_t/p)\cdot\prod_{m=1}^\infty(1-\frac{K_t^2}{p^2m^2})}{(\pi/p)\cdot\prod_{m=1}^\infty(1-\frac{1}{p^2m^2})}\right)^2 = \left(\prod_{m=1}^\infty\left(1-\frac{K_t^2-1}{p^2m^2-1}\right)\right)^2 \\
&< \left(1-\frac{K_t^2-1}{p^2-1}\right)^2 < e^{-2(K_t^2-1)/(p^2-1)} < e^{-(4\ln 2)k_t/(p^2-1)},
\end{aligned}$$

where in the fifth line we use the infinite product representation of sine and in the last line we use $2^x \geqslant 1 + (\ln 2)x$. So

$$\left|\mathbb{E}\left[\chi_j(W)\right]\right|^2 = \prod_{t=1}^r \left|\mathbb{E}\left[\chi_j(X_t)\right]\right|^2 < \prod_{t=1}^r e^{-(4\ln 2)k_t/(p^2-1)} = e^{-(4\ln 2)k/(p^2-1)} < e^{-2k/p^2}.$$

Thus,

$$|X-Y| \leqslant \frac{\sqrt{p}}{2}\cdot\|X-Y\|_2 \leqslant \frac{\sqrt{p}}{2}\cdot\max_{j\neq 0}\left|\mathbb{E}\left[\chi_j(W)\right]\right|^2 \leqslant \frac{\sqrt{p}}{2}\cdot e^{-2k/p^2}. \qquad \square$$

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 $p \geqslant 2^\ell$ 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 $(k', 1/p^{\Omega(1)})$-approximate oblivious symbol-fixing source $f(X)$ on $[p]^{r'}$, where $r' = k/(2p^2\log p)$ and $k' = k^2/(4trp^2\log^2 p)$.*

Proof. First divide X into $\frac{k}{2t}$ blocks consisting of $\frac{2t}{k}r$ smaller sources, for $t = p^2\log p$. Then for each block take the sum modulo p of the smaller sources 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 Corollary 6.7, we mix within $2^{-\Omega(t/p^2)} = \frac{1}{p}$ of uniform. $\square$

Now, as in [22], 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* $\mathrm{poly}(r, 2^\ell)$.

Proof. Let p be the least prime greater than 2^ℓ. Since by Bertrand's Postulate $p < 2 \cdot 2^\ell$, p can easily be found 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 $\lambda(G) \leqslant \frac{1}{p^\alpha} - \frac{1}{\sqrt{p}}$ 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

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

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

Combining this theorem with our reduction from general to flat sources, we get that this same extractor works for general total-entropy 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 $m = \Omega(k^2/r2^{2\ell}\ell)$ bits and has error $\epsilon = 2^{-m}$. This extractor is computable in time* $\mathrm{poly}(r, 2^\ell)$.

Proof. Combine Theorem 6.11 and Lemma 6.4. □

7. Extracting more bits from total-entropy independent sources

7.1. Seed obtainers

Now that we have extractors for total-entropy independent sources, we can extract even more bits using the techniques that Gabizon et al. [18] used to extract more bits out of oblivious bit-fixing sources. The results in this section may be simplified by the ideas of Shaltiel [32]. Assuming the entropy is high enough to use the extractors from Theorems 6.12, 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-entropy independent sources.

Definition 7.1. A function $F : (\{0,1\}^\ell)^r \to (\{0,1\}^\ell)^r \times \{0,1\}^d$ 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 $\eta \leqslant \rho$ 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-entropy independent sources together with a seed obtainer to construct a deterministic extractor for total-entropy independent sources. The proof for the following theorem is the same as the proof for the bit-fixing case in [18].

Theorem 7.2. *Let $F : (\{0,1\}^\ell)^r \to (\{0,1\}^\ell)^r \times \{0,1\}^t$ be a (k', ρ)-seed obtainer for independent sources X on $(\{0,1\}^\ell)^r$ with total min-entropy k. Let $E_1 : (\{0,1\}^\ell)^r \times \{0,1\}^d \to \{0,1\}^m$ be a seeded ϵ-extractor for independent sources on $(\{0,1\}^\ell)^r$ with total min-entropy k. Then $E : (\{0,1\}^\ell)^r \to \{0,1\}^m$ 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.*

To construct seed obtainers, we need to extend the definition of averaging samplers from [18] to general functions as follows. This definition is similar in spirit to that of [38], 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 $Samp : \{0,1\}^t \to P([r])$ is a $(\delta, \theta_1, \theta_2, \gamma)$ averaging sampler if for every function $f : [r] \to [0,1]$ with average value $\frac{1}{r}\sum_i f(i) = \delta$, it holds that

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

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

Lemma 7.4. *Let $Samp : \{0,1\}^t \to P([r])$ 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*

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

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

Given these definitions, we can show that essentially the same construction from Gabizon et al. [18] for bit-fixing seed obtainers works for total-entropy independent source seed obtainers.

Theorem 7.5. *Let $Samp : \{0,1\}^t \to P([r])$ be a $(\delta, \delta_1 r, \delta_2 r, \gamma)$ averaging sampler and $E : (\{0,1\}^\ell)^r \to \{0,1\}^m$ be an ϵ-extractor for independent sources on $(\{0,1\}^\ell)^r$ with total min-entropy $k = \delta_1 r\ell$. Then $F : (\{0,1\}^\ell)^r \to (\{0,1\}^\ell)^r \times \{0,1\}^{m-t}$ 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 $x' = x_{[r]\setminus T}$, padded with $\ell \cdot |T|$ zeroes to get a string in $(\{0,1\}^\ell)^r$.*
- *Let $y = E_2(x)$. Output (x', y).*

The proof of this theorem is almost exactly the same as the proof in [18], 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 [32].

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. [18] generalizes to our definition. Our sampler works by generating d-wise independent variables $Z_1, \ldots, Z_r \in [b]$ and letting $Samp(U_t) = \{i | Z_i = 1\}$.

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

We use the following tail inequality for d-wise independent variables due to Bellare and Rompel [9].

Theorem 7.7. *(See [9].) Let $d \geqslant 6$ be an even integer. Suppose that $X_1, \ldots, X_r$ are d-wise independent random variables taking values in $[0,1]$. Let $Y = \sum_{1 \leqslant i \leqslant r} Y_i$, $\mu = \mathbb{E}[Y]$, and $A > 0$. Then*

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

Proof of Lemma 7.6. Let d be the largest even integer such that $d\log r \leqslant t$ and let $q = \lfloor \log b \rfloor \leqslant \log r$. Use $d\log r$ random bits to generate r d-wise independent random variables $Z_1, \ldots, Z_r \in \{0,1\}^q$ using the construction from [13]. Fix $a \in \{0,1\}^q$. Let the random variable denoting the output of the sampler be $Samp(U_t) = \{i | Z_i = a\}$. For $1 \leqslant i \leqslant r$, define a random variable Y_i that is set to $f(i)$ if $i \in Samp(U_t)$ 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 $Y_1, \ldots, Y_r$ are d-wise independent. Applying Theorem 7.7 with $A = \delta r/2b$,

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

Note that

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

Note that $d \leqslant t/\log r \leqslant \delta r/20b$ by assumption. We conclude that:

$$\Pr_{w \leftarrow U_t}\left[\frac{\delta r}{2b} \leqslant \sum_{i \in Samp(w)} f(i) \leqslant \frac{3\delta r}{b}\right] \geqslant 1 - 8\left(\frac{d\frac{\delta r}{2q} + d^2}{(\delta r/2b)^2}\right)^{d/2} \geqslant 1 - 8\left(\frac{4b^2}{(\delta r)^2}\left(\frac{2d\delta r}{b} + \frac{d\delta r}{20b}\right)\right)^{d/2}$$
$$\geqslant 1 - 8\left(\frac{10db}{\delta r}\right)^{d/2} \geqslant 1 - 2^{-(d/2+3)} \geqslant 1 - 2^{-\Omega(t/\log r)}. \qquad \square$$

7.3. Extractors from seed obtainers

As in [18] 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* $Samp : \{0,1\}^t \to P([r])$.
- *A deterministic* ϵ^**-extractor for total-rate* δ_1 *independent sources* $E^* : (\{0,1\}^\ell)^r \to \{0,1\}^{m'}$.
- *A seeded* ϵ_1*-extractor for total-rate* $\delta - \delta_2$ *independent sources* $E_1 : (\{0,1\}^\ell)^r \times \{0,1\}^s \to \{0,1\}^m$, *where* $m' \geqslant s + t$.

Then we get a deterministic ϵ*-extractor for total-rate* δ *independent sources* $E : (\{0,1\}^\ell)^r \to \{0,1\}^m$ *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 [30].

Theorem 7.9. *(See [30].) For any* $r, k \in \mathbb{N}$, *and* $\epsilon > 0$, *there exists a* ϵ*-extractor* $Ext : \{0,1\}^r \times \{0,1\}^s \to \{0,1\}^m$ *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 [30] 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* $\delta, \delta_1, \epsilon_1 \in (0,1)$ *and* $r, t \in \mathbb{N}$ *be such that* $\delta_1 \geqslant 1/2r$ *and* $6\log r \leqslant t \leqslant \frac{\delta_1 r \log r}{10}$. *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* $E^* : (\{0,1\}^\ell)^r \to \{0,1\}^{m'}$ *with* $m' \geqslant s + t$, *we can construct an* ϵ*-extractor for total-rate* δ *independent sources* $E : (\{0,1\}^\ell)^r \to \{0,1\}^m$ *where* $\epsilon = \epsilon_1 + 3 \cdot \max(\epsilon^* + 2^{-\Omega(t/\log r)}, \epsilon^* \cdot 2^{t+1})$.

Proof. Combine Lemma 7.6 with $b = \delta/2\delta_1$, Theorems 7.9 and 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-entropy independent sources with sufficiently high min-entropy, proving Theorem 1.8.

Proof of Theorem 1.8. 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 $6\log r \leqslant t \leqslant \frac{\delta_1 r \log r}{10}$ and $m' \geqslant s + t$ for sufficiently large r. $\square$

The extractor for small-space sources from Theorem 1.4 is then obtained by combining Theorem 1.8 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 1.7.

Proof of Theorem 1.7. 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 $6\log r \leqslant t \leqslant \frac{\delta_1 r \log r}{10}$ and $m' \geqslant s + t$ for sufficiently large r. $\square$

The extractor for small-space sources from Theorem 1.3 is then obtained by combining Theorem 1.8 with Lemma 3.1. We can also apply this construction to the polynomial entropy rate extractor from Corollary 5.2, which proves Theorem 1.6.

Proof of Theorem 1.6. Use the construction from Corollary 7.10 with the extractor from Corollary 5.2 as E^* and let $\epsilon_1 = 2^{-(\delta_1^2 r\ell)^{\Omega(1)}/(\log^3(r\ell))}$ and $t = (\delta_1^2 r\ell)^{\Omega(1)}$. Then it's not hard to see that (choosing appropriate constants) these values satisfy $6\log r \leqslant t \leqslant \frac{\delta_1 r \log r}{10}$ and $m' \geqslant s + t$ for sufficiently large r. □

The extractor for small-space sources from Theorem 1.2 is then obtained by combining Theorem 1.6 with Lemma 3.1.

7.4. Extractors for smaller entropy

Notice that the method given by Corollary 7.10 requires $m > s = \text{polylog}(r, \ell)$. Gabizon et al. [18] also showed how to use seed obtainers to extract more bits even when the initial extractor only extracts a small logarithmic number of bits, which they're able to get from the cycle walk extractor from [22]. We can generalize their construction to work for total-entropy 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 a small logarithmic number of bits, we need both a sampler and a seeded extractor for total-entropy independent sources. To do so, as in [18], 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 [18] 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 $r \geqslant 16$ and $k = \delta r\ell \geqslant \log^c r$, the following holds. There is a polynomial-time computable $(\delta, \delta r/2k^b, 3\delta r/k^b, O(k^{-b}))$ sampler Samp : $\{0,1\}^t \to P([r])$ 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 $r \geqslant 16$ and $k = \delta r\ell \geqslant \log^c r$, we can use $\alpha \cdot \log k$ random bits to explicitly partition $[r]$ into $m = O(k^b)$ sets $T_1, \ldots, T_m$ such that for every function $f : [r] \to [0,1]$ with average value $\frac{1}{r}\sum_i f(i) = \delta$,*

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

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 $r \geqslant 16$ and $k \geqslant \log^c r$, the following holds. We can use $\alpha \cdot \log k$ random bits to explicitly partition $[r]$ into $m = \Theta(k^b)$ sets $T_1, \ldots, T_m$ such that for any independent sources X on $(\{0,1\}^\ell)^r$ with total min-entropy k,*

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

Now we will use this partitioning to construct a seeded extractor for total-entropy independent sources that uses a small seed. As in [18] 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 $r \geqslant 16$, $k \geqslant \log^c r$, $0 < \delta \leqslant 1$ and $2^\ell \leqslant k^{(1-b)/2}/(c\sqrt{\log k^{2b}})$, the following holds. There is a polynomial-time computable seeded ϵ-extractor $E : (\{0,1\}^\ell)^r \times \{0,1\}^s \to \{0,1\}^m$ for independent sources on $(\{0,1\}^\ell)^r$ with total min-entropy k, 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 $T_1, \ldots, T_{m'}$ using Corollary 7.13. Next we find the next largest prime $p \geqslant 2^\ell$, 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 $z_i = \sum_{j \in T_i} x_j \bmod p$ and output $z = z_1, \ldots, z_{m'}$.

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

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

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

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 $k \geqslant \log^c r$ and $2^\ell \leqslant k^{(1-b)/2}/(c\sqrt{\log k^{2b}})$, the following holds. There exists a polynomial-time computable ϵ-extractor $E : (\{0,1\}^\ell)^r \to \{0,1\}^m$ for independent sources on $(\{0,1\}^\ell)^r$ with total 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 □

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 1.9.

Proof of Theorem 1.9. 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. □

8. Nonconstructive results

In this section, we describe nonconstructive results for both small-space and total-entropy independent sources. 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. In particular, this argument shows that a function achieving these parameters exists. To do so we use a standard argument that shows that a randomly chosen function is an extractor for any class of sources that is not too large, as long as the sources in the class are close to having high min-entropy.[9]

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 $|X' - X| \leqslant \frac{\epsilon}{2}$ and $H_\infty(X') \geqslant k$. Then, with probability $1 - \exp(-\Omega(2^k\epsilon^2))$ a function chosen uniformly at random is an extractor for $\mathcal{X}$ as long as $k \geqslant \log(2^m + \log|\mathcal{X}|) + 2\log(1/\epsilon) + O(1)$. In particular, as long as $k \geqslant \log\log|\mathcal{X}| + 2\log(1/\epsilon) + O(1)$, we can extract $m = k - 2\log(1/\epsilon) - O(1)$ bits.*

We need the following Chernoff bound to prove Theorem 8.1.

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

$$\Pr\big[|Z - \mu| \geqslant \epsilon\mu\big] < 2\exp\big(-\mu\epsilon^2/3\big).$$

Proof of 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.

Let $f : \{0,1\}^n \to \{0,1\}^m$ be chosen uniformly at random from all functions from n bits to m bits. Fix $X \in \mathcal{X}$ and $S \subset \{0,1\}^m$. Let X' be such that $|X' - X| \leqslant \epsilon/2$ and $H_\infty(X') \geqslant k$. Let Z_x be the indicator random variable for whether $f(x) \in S$. Let

$$Z = 2^k \Pr_{x \leftarrow_R X'}\big[f(x) \in S\big] = \sum_{x \in \mathrm{supp}(X')} \big(2^k \Pr\big[X' = x\big]\big) Z_x.$$

Note that the coefficients $2^k \Pr[X' = x]$ are in the interval $[0,1]$. Since the function f is chosen uniformly at random, the random variables Z_x are independent, and $E[Z] = 2^k|S|/2^m$. Thus we can apply Lemma 8.2 to get

$$\Pr_f\left[\left|\Pr_{x \in X'}\big[f(x) \in S\big] - \frac{|S|}{2^m}\right| \geqslant \epsilon'\frac{|S|}{2^m}\right] = \Pr_f\left[\left|Z - \frac{2^k|S|}{2^m}\right| \geqslant \epsilon'\frac{2^k|S|}{2^m}\right] \leqslant 2\exp\left(-\epsilon'^2\frac{2^k|S|}{3\cdot 2^m}\right).$$

Making the change of variables $\epsilon' = \epsilon 2^m/|S|$, we get that for any fixed set S, we proved that

$$\Pr_f\big[\big|\Pr\big[f\big(X'\big) \in S\big] - \Pr[U_m \in S]\big| \geqslant \epsilon/2\big] \leqslant 2\exp\left(-\left(\frac{\epsilon 2^m}{2|S|}\right)^2\frac{2^k|S|}{3\cdot 2^m}\right) = 2\exp\left(-\frac{\epsilon^2 2^k 2^m}{12|S|}\right).$$

Recall that $|f(X') - U_m| = \max_S\{|\Pr[f(X') \in S] - |S|/2^m|\}$. By the union bound over all sets $S \subset \{0,1\}^m$ and all $X \in \mathcal{X}$, and since $2^m/|S| \geqslant 1$,

[9] In fact, if we wish to save randomness in selecting the function, then [37,15] showed that we can get a similar result by using a random d-wise independent function instead of a completely random function. However, the parameters proved there are not quite as good as in Theorem 8.1.

$$\Pr_f\Big[\max_S\big\{\big|f(X')-U_m\big|\geqslant\epsilon/2\big\}\Big]\leqslant 2\exp\big(-\epsilon^2 2^k/12\big)2^{2^m}|\mathcal{X}|.$$

Now whenever f does satisfy $|f(X')-U_m|<\epsilon/2$, we have that $|f(X)-U_m|<\epsilon/2+\epsilon/2=\epsilon$. Setting the above error to $1/2^{2^m}|\mathcal{X}|$ and solving for k, we get that a function chosen uniformly at random is an extractor for $|\mathcal{X}|$ with probability $1-1/2^{2^m}|\mathcal{X}|$ as long as $k\geqslant\log(2^m+\log|\mathcal{X}|)+2\log(1/\epsilon)+O(1)$. In particular, as long as $k\geqslant\log\log|\mathcal{X}|+2\log(1/\epsilon)+O(1)$, we can extract $m=k-2\log(1/\epsilon)-O(1)$ bits. □

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.

Note that α must be a reciprocal of an integer for the above definition to be achievable.

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, and let $\alpha=1/d$ for some $d\in\mathbb{N}$. Then 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 in the branching program for X 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 generated by the branching program whose first i layers are as in the branching program for X and the rest are is in the branching program for X_s. So $X=X_a^0$ and $X_a=X_a^n$. Then $|X-X_a|=|\sum_{i=1}^n(X_a^{i-1}-X_a^i)|\leqslant\sum_{i=1}^n|X_a^{i-1}-X_a^i|$.

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

$$\big|X_a^{i-1}-X_a^i\big|\leqslant\frac{1}{2}\sum_{j,j'}q_{i-1,j}\big(\big(p^0_{j,j'}+\alpha-p^0_{j,j'}\big)+\big(p^1_{j,j'}+\alpha-p^1_{j,j'}\big)\big)\leqslant\alpha\sum_{j'}\sum_{j}q_{i-1,j}=\alpha\sum_{j'}1=\alpha 2^s.$$

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

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 $(2^{s+1})^{1/\alpha}$ ways to allocate probabilities to these edges. (For each of the $1/\alpha$ "units" of probability, we can assign it to one of the 2^{s+1} edges.) Since there are n layers and 2^s vertices at each layer, the total number of possible sources is $2^{(s+1)2^s n/\alpha}$. □

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

Theorem 8.6 *(Theorem 1.5, restated). For space s sources with min-entropy k, a function $f:\{0,1\}^n\to\{0,1\}^m$ chosen uniformly at random is an ϵ-extractor with output length $m=k-2\log(1/\epsilon)-O(1)$ with probability at least $1-\exp(-\Omega(2^k\epsilon^2))$, as long as $k\geqslant 2s+\log s+2\log n+3\log(1/\epsilon)+O(1)$.*

This theorem says that extractors exist for sources with space almost as large as $k/2$ and with min-entropy as low as $\Theta(\log n)$. This factor of 2 in the relationship between space and min-entropy is necessary, as we'll see shortly. On the other hand, note that if we restrict to α-approximate space s sources for a fixed constant α (e.g. $\alpha=1/2$), then we can reduce the bound to $k\geqslant s+\log s+\log n+2\log(1/\epsilon)+O(1)$.

Proof. First apply Lemma 8.4 with $\alpha=\epsilon/n2^{s+1}$ to show that the 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 k, 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 min-entropy k space s source is a convex combination of these α-approximate sources, the extractors given by Theorem 8.1 also work with these sources. □

To see that the factor of 2 is necessary, we show that our model of a space s source can sample an arbitrary distribution of length $\ell = 2s$ (actually even $2s+1$). It is known [12] that there is no deterministic extractor that works for all sources of length ℓ and min-entropy $\ell - 1$. (Indeed, for every $\mathsf{Ext} : \{0,1\}^\ell \to \{0,1\}^m$, there is a source Z of min-entropy $\ell - 1$ on which the first bit of $\mathsf{Ext}(Z)$ is constant.) The following space s source samples an arbitrary source (X, Y), where X and Y are each of length s.

1. In the first layer, choose x according to X, output the first bit of x, and move to state x.
2. In the next $s-1$ steps, output the remaining bits of x, and remain in state x.
3. In the next layer, choose y according to the distribution $(Y|X=x)$, output the first bit of y, and move to state y.
4. In the next $s-1$ steps, output the remaining bits of y and remain in state y.

8.2. Total-entropy independent sources

We can also apply Theorem 8.1 to total-entropy independent sources. Similarly to the small-space case, we define an intermediate model to reduce the number of sources.

Definition 8.7. An *approximate flat source* X is a source in which all elements of $\mathrm{supp}(X)$ have the same probability, except for at most one *exceptional* string x^*. If the probability of x^* is an integer multiple of α, we call X an *α-approximate flat source*.

An *α-approximate flat independent source* $X_1, \ldots, X_r$ on $(\{0,1\}^\ell)^r$ is an independent source such that for every i, X_i is an α-approximate flat source.

The following 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 $X = X_1, \ldots, X_r$ be an independent source on $(\{0,1\}^\ell)^r$ of total-entropy k. For every $\alpha > 0$, X is $r\alpha$-close to a convex combination of α-approximate flat independent sources, each of which is $r\alpha$-close to some independent source of total-entropy k.*

Proof. For each i, let k_i be the min-entropy of X_i, so $\sum_i k_i = k$. X_i can be written as a convex combination of approximate flat sources of min-entropy k_i.[10] This induces a decomposition of X as a convex combination of approximate flat independent sources X' of min-entropy k. For each such $X' = (X'_1, \ldots, X'_r)$, we can round the probabilities of the r exceptional strings to integer multiples of α while paying αr in statistical distance. □

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

Proof. To specify an α-approximate flat independent sources on $(\{0,1\}^\ell)^r$, we can specify each of its r components, each of which is specified by the exceptional string (2^ℓ possibilities), the probability mass of the exceptional string (at most $1/\alpha$ possibilities) and the support of the distribution (at most 2^{2^ℓ} possibilities). □

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 *(Theorem 1.10, restated). For total-entropy k independent sources, a function $f : (\{0,1\}^\ell)^r \to \{0,1\}^m$ chosen uniformly at random is an ϵ-extractor with output length $m = k - 2\log(1/\epsilon) - O(1)$ with probability $1 - \exp(-\Omega(2^k\epsilon^2))$ as long as $k \geqslant \max\{\ell, \log\log(r/\epsilon)\} + \log r + 2\log(1/\epsilon) + O(1)$.*

Note that the $k > \ell$ is necessary because otherwise all of the entropy could be contained within a single source, which we know is impossible to extract from. Thus, the bound in this theorem is close to the best we could hope for.

[10] It is well known that if 2^{k_i} is an integer, then X_i is a convex combination of standard flat sources (with no exceptional string). The general case is proven in the same way: the set of sources of min-entropy at least k_i is a convex polytope defined by the inequalities $\forall x\ 0 \leqslant p_x \leqslant 2^{-k_i}$ and $\sum_x p_x = 1$. Every element of the polytope is a convex combination of the vertices of the polytope, which are the points that make a maximal set of inequalities tight, which in turn correspond to the approximate flat sources of min-entropy k_i. We note that rounding 2^{k_i} down to the nearest integer to get standard flat sources may cost too much entropy (e.g. in the case when the sources are of length 1, so $k_i \in [0,1]$).

Proof. First apply Lemma 8.8 with $\alpha = \epsilon/(2r)$ to show that the each total-entropy k independent source X is a convex combination of α-approximate flat independent sources of total-entropy k that are $\epsilon/2$ close to having min-entropy k. Then apply Theorem 8.1 to the set of α-approximate total-entropy k independent sources that are $\epsilon/2$ close to having min-entropy k, 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-entropy k independent source is a convex combination of these α-approximate sources, the extractors given by Theorem 8.1 also work with 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-entropy 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* X on $\{0,1\}^n$ with min-entropy k has at least k uniformly random bits X_i and the rest of the bits X_j are functions of the previous bit (i.e. $X_j = X_{j-1}$, $X_j = \neg X_{j-1}$, $X_j = 0$, or $X_j = 1$).

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-entropy independent sources with small alphabet size.

9.1. Extracting from previous-bit sources

To convert a previous-bit source to a total-entropy 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 $r = k/2$, $k' = k^2/4n$ and $\ell = \lceil \log(2n/k+1) \rceil$.*

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

Lemma 9.3. *For every $t \in \mathbb{N}$, there is a polynomial-time computable function $f : \{0,1\}^t \to \{0,1\}^{\lceil \log(t+1) \rceil}$ 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 $0 \leqslant i \leqslant t$, 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 $f(y) = \sum_{i=1}^{t} y_i (z_i - z_{i-1}) \in \mathbb{Z}_2^{\lceil \log(t+1) \rceil}$.

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 (as addition is modulo 2). 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 $z_k \neq z_{j-1}$, $f(y_0) - f(y_1) \neq 0$. □

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 strings, 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. □

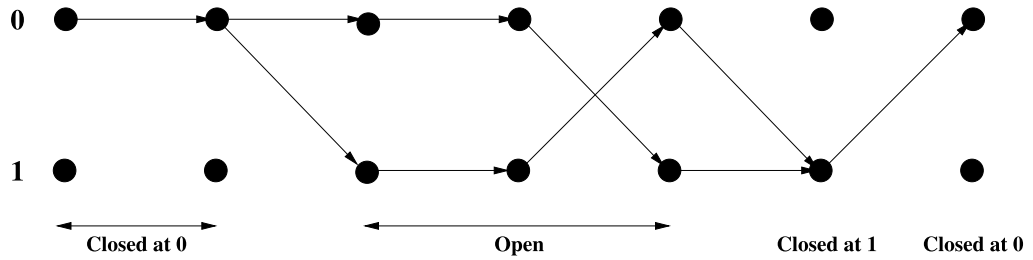

Fig. 3. A previous-bit source viewed as a restricted width two source. This source consists of the bits $0, 0, r, r, \bar{r}, 1, 0$, where r is a random bit.

Now we can combine Theorem 1.8 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 $m \geqslant k^2/8n$ bits and has error $\epsilon = \exp(-\Omega(\lfloor k^5/(n^4 \log(n/k) \log^3 k \rfloor)))$.*

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 $r = k/2$, $k' = k^2/4n$, and $\ell = \lceil \log(2n/k + 1) \rceil$. Then apply the extractor from Theorem 1.8 with $\zeta = k^2/(8n \cdot r\ell)$. □

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 at least a $1 - 2^{-k/4} - e^{-9(k')^2/2n}$ fraction of 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 $\epsilon = \exp(-\Omega((k')^5/(n^4 \log(n/k') \log^3 k'))$, where $k' = \min(k/48 \log(n/k), k/96)$.*

Proof. By Lemma 9.5 our source X is $(2^{-k/4} + e^{-9(k')^2/2n})$-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. □

Notice that here we only need $k \gg n^{4/5}$ whereas all of our extractors for general small-space sources require $k \gg n^{1-\eta}$ for some small constant η.

In order to prove Lemma 9.5, we now describe how to 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 Fig. 3.

Let $X = (X_1, \ldots, X_n)$ be the bits of our restricted width 2 source. We will define (correlated) random variables $G = (G_1, \ldots, G_n) \in \{0, 1, *\}^n$ (to represent the states) and $X' = (X'_1, \ldots, X'_n) \in \{0,1\}^n$ such that:

1. X' is identically distributed to X.
2. For every $g = (g_1, \ldots, g_n)$ in the support of G, $X'|_{G=g}$ is a previous-bit source.
3. For every $g = (g_1, \ldots, g_n)$, if $g_i \in \{0,1\}$, then $X'_i|_{G=g}$ is always equal to g_i. In such a case, we say "X_i is closed at g_i".
4. For every $g = (g_1, \ldots, g_n)$, if $g_i = *$, then $X'_i|_{G=g}$ is a uniformly random bit (possibly equal to the previous bit or its negation). In such a case, we say "X_i is open".

Then it follows that X is a convex combination of the random variables $X'|_{G=g}$, where these are weighted according to $\Pr[G = g]$.

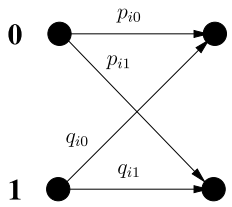

Fig. 4. The probabilities for a single bit of a restricted width two source.

We construct X'_i and G_i inductively conditioned on the values of $X'_{i-1} = x'_{i-1}$ and $G_i = g_{i-1}$. To do this, we consider the following transition probabilities, shown in Fig. 4.

$$p_{i0} = \Pr[X_i = 0 | X_{i-1} = 0],$$
$$p_{i1} = \Pr[X_i = 1 | X_{i-1} = 0],$$
$$q_{i0} = \Pr[X_i = 0 | X_{i-1} = 1],$$
$$q_{i1} = \Pr[X_i = 1 | X_{i-1} = 1].$$

First, we describe what happens if we are currently in the open state (i.e. $g_{i-1} = *$). We become closed at 0 (i.e. we set $X'_i = G_i = 0$) with probability $\min(p_{i0}, q_{i0})$. We become closed at 1 (i.e. we set $X'_i = G_i = 1$) with probability $\min(p_{i1}, q_{i1})$. Otherwise, we stay open (i.e. set $G_i = *$), and consider the remaining probabilities, namely $p'_{ib} = p_{ib} - \min\{p_{ib}, q_{ib}\}$ and $q'_{ib} = q_{ib} - \min\{p_{ib}, q_{ib}\}$ for $b \in \{0, 1\}$. Then we have either $p'_{i0} = q'_{i1} = 0$, in which case we set $X'_i = \neg x_{i-1}$, or we have $p'_{i1} = q'_{i0} = 0$, in which case we set $X'_i = x_{i-1}$.

If we are closed at 0 (i.e. $g_{i-1} = 0$), then with probability $2\min(p_{i0}, p_{i1})$, we go into the open state (i.e. set $G_i = *$ and X'_i to be a uniformly random bit). If $p_{i0} < p_{i1}$, then with probability $1 - 2p_{i0}$, we go into the closed at 1 state (i.e. we set $X'_i = G_i = 1$). Otherwise, with probability $1 - 2p_{i1}$, we go into the closed at 0 state (i.e. set $X'_i = G_i = 0$).

If we are closed at 1 (i.e. $g_{i-1} = 1$), then with probability $2\min(q_{i0}, q_{i1})$, we go into the open state (i.e. set $G_i = *$ and X'_i to be a uniformly random bit). If $q_{i0} < q_{i1}$, then with probability $1 - 2q_{i0}$, we go into the closed at 1 state (i.e. set $X'_i = G_i = 1$). Otherwise, with probability $1 - 2q_{i1}$, we go into the closed at 0 state (i.e. set $X'_i = G_i = 0$).

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.

Every node has a *more probable edge* and a *less probable edge* exiting it (breaking ties arbitrarily), where the probabilities are according to distribution X. We will show how the number of less probable edges on a path in X relates to the min-entropy of a Z_j that contains this path. First note that every less 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 less 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, then it would suffice to show that with high probability, we take the less probable edge a large number of times. Since $C(Z_j) \leqslant A(Z_j)$, 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 less 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* $t = \min(k/(8\log(n/k)), k/16)$ *less 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 less probable edges is at most

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

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

Suppose $k \leqslant n/4$. Then t, as defined in the lemma is equal to $k/(8\log(n/k))$, so

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

If $k > n/4$, then $t = k/16$, so

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

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

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 [44] for a proof).

Definition 9.8. A stochastic process $Y_0, Y_1, \ldots$ is a *submartingale* with respect to a stochastic process $G_0, G_1, \ldots$ if

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

for all $i \geqslant 0$.

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

$$\Pr[Y_n \leqslant -\alpha] \leqslant e^{-\alpha^2/2n}.$$

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 $\alpha > 0$,*

$$\Pr\big[B(Z_j) \geqslant C(Z_j) + \alpha\big] \leqslant e^{-\alpha^2/2n}.$$

Proof. Let Y_i be the number of closings from $X_1, \ldots, X_i$ minus the number of false closings from $X_1, \ldots, X_i$ and let $Y_0 = 0$. Let $G_0, G_1, \ldots, G_n$ be the states as defined earlier.

Now we show that $Y_0, \ldots, Y_n$ is a submartingale with respect to $G_0, G_1, \ldots, G_n$. If $G_i = 0$ or 1, then we have no closings or false closings at $i+1$, so $\mathbb{E}[Y_{i+1} | G_0, G_1, \ldots, G_i] = Y_i$. We show that if $G_i = *$, 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 $\mathbb{E}[Y_{i+1} | G_0, G_1, \ldots, G_i] \geqslant Y_i$, 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 $p_{i+1,0} + q_{i+1,1} \geqslant q_{i+1,0} + p_{i+1,1}$, 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 $p_{i+1,0} \leqslant 1/2$ or $q_{i+1,1} \leqslant 1/2$, which implies either $p_{i+1,1} \geqslant 1/2$ or $q_{i+1,0} \geqslant 1/2$, and thus the probability of closing is at least $1/2$.

By the definition of Y_i, $|Y_i - Y_{i-1}| \leqslant 1$, so we can apply Lemma 9.9 to get

$$\Pr[Y_n \leqslant -\alpha] \leqslant e^{-\alpha^2/2n},$$

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 less probable edge on the path corresponds to either an opening, a closing, or a false closing. Thus the probability that $A(Z_j) + B(Z_j) + C(Z_j) \geqslant s$ 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

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

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

By Lemma 9.10,

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

With high probability both of these events occur, so

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

Acknowledgments

In this special issue in honor of Dick Karp's Kyoto Prize, we would like to recognize Dick's tremendous contributions to theoretical computer science. His notable contributions to probabilistic analysis and randomized algorithms are most relevant for this paper. The last author in particular is grateful for learning a lot of probability from Dick's excellent courses.

We thank the anonymous referee for many helpful comments.

References

[1] N. Alon, J.H. Spencer, The Probabilistic Method, Wiley–Interscience Series, John Wiley & Sons, Inc., New York, 2000.
[2] Charles H. Bennett, Gilles Brassard, Robert Jean-Marc, Privacy amplification by public discussion, SIAM J. Comput. 17 (2) (April 1988) 210–229.
[3] J. Bourgain, A. Glibichuk, S. Konyagin, Estimates for the number of sums and products and for exponential sums in fields of prime order, J. Lond. Math. Soc. 73 (2) (2006) 380–398.
[4] B. Barak, R. Impagliazzo, A. Wigderson, Extracting randomness using few independent sources, SIAM J. Comput. 36 (2006) 1095–1118.
[5] Boaz Barak, Guy Kindler, Ronen Shaltiel, Benny Sudakov, 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, 2005, pp. 1–10.
[6] M. Blum, Independent unbiased coin flips from a correlated biased source: A finite Markov chain, Combinatorica 6 (2) (1986) 97–108.
[7] M. Ben-Or, N. Linial, Collective coin flipping, in: S. Micali (Ed.), Randomness and Computation, Academic Press, New York, 1990, pp. 91–115.
[8] J. Bourgain, More on the sum-product phenomenon in prime fields and its applications, Int. J. Number Theory 1 (2005) 1–32.
[9] M. Bellare, J. Rompel, Randomness-efficient oblivious sampling, in: Proceedings of the 35th Annual IEEE Symposium on Foundations of Computer Science, 1994, pp. 276–287.
[10] Ran Canetti, Yevgeniy Dodis, Shai Halevi, Eyal Kushilevitz, Amit Sahai, Exposure-resilient functions and all-or-nothing transforms, in: Bart Preneel (Ed.), Advances in Cryptology — EUROCRYPT 2000, in: Lecture Notes in Comput. Sci., vol. 1807, Springer-Verlag, May 2000, pp. 453–469.
[11] B. Chor, J. Friedman, O. Goldreich, J. Håstad, S. Rudich, R. Smolensky, The bit extraction problem or t-resilient functions, in: Proceedings of the 26th Annual IEEE Symposium on Foundations of Computer Science, 1985, pp. 396–407.
[12] B. Chor, O. Goldreich, Unbiased bits from sources of weak randomness and probabilistic communication complexity, SIAM J. Comput. 17 (2) (1988) 230–261.
[13] J.L. Carter, M.N. Wegman, Universal classes of hash functions, J. Comput. System Sci. 18 (1979) 143–154.
[14] A. Cohen, A. Wigderson, Dispersers, deterministic amplification, and weak random sources, in: Proceedings of the 30th Annual IEEE Symposium on Foundations of Computer Science, 1989, pp. 14–19.
[15] Yevgeniy Dodis, Exposure-resilient cryptography, PhD thesis, MIT, 2000.
[16] Yevgeniy Dodis, Impossibility of black-box reduction from non-adaptively to adaptively secure coin-flipping, Technical Report 039, Electronic Colloquium on Computational Complexity, 2000.
[17] Z. Dvir, R. Raz, Analyzing linear mergers, Random Structures Algorithms 32 (2008) 334–345.
[18] A. Gabizon, R. Raz, R. Shaltiel, Deterministic extractors for bit-fixing sources by obtaining an independent seed, SIAM J. Comput. 36 (2006) 1072–1094.
[19] B. Jun, P. Kocher, The Intel random number generator, http://www.cryptography.com/resources/whitepapers/IntelRNG.pdf, 1999.
[20] Robert Koenig, Ueli Maurer, Extracting randomness from generalized symbol-fixing and Markov sources, in: Proceedings of 2004 IEEE International Symposium on Information Theory, June 2004, p. 232.
[21] Robert Koenig, Ueli Maurer, Generalized strong extractors and deterministic privacy amplification, in: Nigel Smart (Ed.), Cryptography and Coding 2005, in: Lecture Notes in Comput. Sci., vol. 3796, Springer-Verlag, December 2005, pp. 322–339.
[22] J. Kamp, D. Zuckerman, Deterministic extractors for bit-fixing sources and exposure-resilient cryptography, SIAM J. Comput. 36 (2006) 1231–1247.
[23] D. Lichtenstein, N. Linial, M. Saks, Some extremal problems arising from discrete control processes, Combinatorica 9 (3) (1989) 269–287.
[24] L. Lovász, Random walks on graphs: A survey, in: D. Miklós, V.T. Sós, T. Szőnyi (Eds.), Combinatorics, in: Paul Erdős is Eighty, vol. 2, J. Bolyai Math. Soc., Budapest, 1996, pp. 353–398.
[25] Ueli Maurer, Stefan Wolf, Privacy amplification secure against active adversaries, in: Burton S. Kaliski Jr. (Ed.), Advances in Cryptology — CRYPTO '97, in: Lecture Notes in Comput. Sci., vol. 1294, Springer-Verlag, August 1997, pp. 307–321.
[26] N. Nisan, A. Ta-Shma, Extracting randomness: A survey and new constructions, J. Comput. System Sci. 58 (1999) 148–173.
[27] N. Nisan, D. Zuckerman, Randomness is linear in space, J. Comput. System Sci. 52 (1) (1996) 43–52.
[28] A. Rao, Extractors for a constant number of polynomially small min-entropy independent sources, in: Proceedings of the 38th Annual ACM Symposium on Theory of Computing, 2006, pp. 497–506.
[29] Ran Raz, Extractors with weak random seeds, in: Proceedings of the 37th Annual ACM Symposium on Theory of Computing, 2005, pp. 11–20.
[30] R. Raz, O. Reingold, S. Vadhan, Extracting all the randomness and reducing the error in Trevisan's extractors, J. Comput. System Sci. 65 (1) (2002) 97–128.
[31] Ronen Shaltiel, Recent developments in explicit constructions of extractors, Bull. Eur. Assoc. Theor. Comput. Sci. 77 (June 2002) 67–95.
[32] R. Shaltiel, How to get more mileage from randomness extractors, in: Proceedings of the 21st Annual IEEE Conference on Computational Complexity, 2006, pp. 49–60.
[33] M. Santha, U.V. Vazirani, Generating quasi-random sequences from semi-random sources, J. Comput. System Sci. 33 (1986) 75–87.
[34] Miklós Sántha, Umesh V. Vazirani, Generating quasirandom sequences from semirandom sources, in: Twenty-fifth Annual Symposium on Foundations of Computer Science, Singer Island, Fla., 1984, J. Comput. System Sci. 33 (1) (1986) 75–87.
[35] L. Trevisan, Extractors and pseudorandom generators, J. ACM (2001) 860–879.
[36] A. Ta-Shma, On extracting randomness from weak random sources, in: Proceedings of the 28th Annual ACM Symposium on Theory of Computing, 1996, pp. 276–285.
[37] Luca Trevisan, Salil P. Vadhan, Extracting randomness from samplable distributions, in: Proceedings of the 41st Annual IEEE Symposium on Foundations of Computer Science, 2000, pp. 32–42.
[38] S. Vadhan, On constructing locally computable extractors and cryptosystems in the bounded-storage model, J. Cryptology 17 (1) (Winter 2004) 43–77.
[39] U.V. Vazirani, Randomness, adversaries and computation, PhD thesis, EECS, University of California at Berkeley, 1986.

[40] Umesh V. Vazirani, Efficiency considerations in using semi-random sources (extended abstract), in: Proceedings of the Nineteenth Annual ACM Symposium on Theory of Computing, New York City, 25–27 May 1987, pp. 160–168.
[41] Umesh V. Vazirani, Strong communication complexity or generating quasirandom sequences from two communicating semirandom sources, Combinatorica 7 (4) (1987) 375–392.
[42] J. von Neumann, Various techniques used in connection with random digits, National Bureau of Standards, Applied Mathematics Series 12 (1951) 36–38.
[43] U.V. Vazirani, 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, 1985, pp. 417–428.
[44] N.C. Wormald, The differential equation method for random graph processes and greedy algorithms, in: M. Karonski, H.J. Proemel (Eds.), Lectures on Approximation and Randomized Algorithms, PWN, Warsaw, 1999, pp. 73–155.
[45] A. Wigderson, D. Zuckerman, Expanders that beat the eigenvalue bound: Explicit construction and applications, Combinatorica 19 (1) (1999) 125–138.
[46] D. Zuckerman, Simulating BPP using a general weak random source, Algorithmica 16 (1996) 367–391.
[47] D. Zuckerman, Linear degree extractors and the inapproximability of max clique and chromatic number, Theory Comput. 3 (2007) 103–128.