Deterministic Extractors for Independent-Symbol Sources

Chia-Jung Lee, Chi-Jen Lu, and Shi-Chun Tsai

***Abstract*—In this paper, we consider the task of deterministically extracting randomness from sources consisting of a sequence of n independent symbols from $\{0,1\}^d$. The only randomness guarantee on such a source is that the whole source has min-entropy k. We give an explicit deterministic extractor which extract $\Omega(\log k - \log\log(1/\varepsilon))$ bits with error ε, for any n, d, $k \in \mathbb{N}$ and $\varepsilon \in (0,1)$. For sources with a larger min-entropy, we can extract even more randomness. When $k \geq n^{1/2+\gamma}$, for any constant $\gamma \in (0,1/2)$, we can extract $m = k - O(d\log(1/\varepsilon))$ bits with any error $\varepsilon \geq 2^{-\Omega(n^{\gamma})}$. When $k \geq \log^c n$, for some constant $c > 0$, we can extract $m = k - (1/\varepsilon)^{O(1)}$ bits with any error $\varepsilon \geq k^{-\Omega(1)}$. Our results generalize those of Kamp and Zuckerman and Gabizon *et al.* which only work for bit-fixing sources (with $d = 1$ and each bit of the source being either fixed or perfectly random). Moreover, we show the existence of a nonexplicit deterministic extractor which can extract $m = k - O(\log(1/\varepsilon))$ bits whenever $k = \omega(d + \log(n/\varepsilon))$. Finally, we show that even to extract from bit-fixing sources, any extractor, seeded or not, must suffer an entropy loss $k - m = \Omega(\log(1/\varepsilon))$. This generalizes a lower bound of Radhakrishnan and Ta-Shma on extracting from general sources.**

***Index Terms*—Independent-symbol sources, min-entropy, pseudo-randomness, randomness extractors.**

I. Introduction

RANDOMNESS has become a useful tool in computer science. For many computational problems, the most efficient algorithms known are randomized. For some tasks in distributed computing, only randomized solutions are possible. In cryptography, randomness is essential in generating secret keys. However, when using randomness in designing algorithms or protocols, people usually assume the randomness being perfect, and the performance guarantees are based on this assumption. In reality, the random sources we (or computers) have access to are typically not so perfect at all, but only contain some crude randomness. One approach to solve this problem is to construct so-called *extractors*, which can extract almost perfect randomness from weakly random sources [35], [22]. Extractors turn out to have close connections to other fundamental objects such as pseudorandom generators, hash functions, error-correcting codes, expander graphs, and samplers, and they have found a wide range of applications in areas such as complexity theory, cryptography, data structures, coding theory, distributed computing, and combinatorics (e.g., [29], [22], [36], [37], [34], [32], [31], [18], [33]). A nice survey can be found in [27].

We measure the amount of randomness in a source by its *min-entropy*; a source is said to have min-entropy k if every element occurs with probability at most 2^{-k}. Given sources with enough min-entropy, one would like to construct an extractor which can extract a string with distribution close to uniform. However, it is well known that one cannot deterministically extract even one bit from an n-bit source with min-entropy $n-1$ [6]. In contrast, it becomes possible if we are allowed a few random bits, called a seed, to aid the extraction. Such a procedure is called a *seeded extractor*. During the past decades, a long line of research has worked on using a shorter seed to extract more randomness (e.g., [22], [21], [24], [11], [26], [32], [30], [28]), and finally an optimal (up to constant factors) construction has been given recently [19].

The problem with a seeded extractor is again to get a seed which is perfectly (or almost) random. For some applications, this issue can be taken care of (for example, by enumerating all possible seed values when the seed is short), but for others, we are back to the same problem which extractors are originally asked to solve. This motivates one to consider the possibility of more restricted sources from which randomness can be extracted in a deterministic (seedless) way.

One line of research studies the case with multiple independent sources. The goal is to have a small number of independent sources with a low min-entropy requirement on sources, while still being able to extract randomness from them. With two independent sources, the requirement on the min-entropy rate (average min-entropy per bit) stayed slightly above $1/2$ for a long time [6], [8], [16], but this barrier has been broken by a recent construction which pushes the requirement slightly below $1/2$ [5]. The requirement on min-entropy rate can be lowered to any constant when there are a constant number of independent sources [3], and the number of sources has recently been reduced to three [4].

The other line of research considers the case of bit-fixing sources. In an oblivious bit-fixing source, each bit is either fixed (containing no randomness) or perfectly random, and is independent of other bits. From such a source of length n with min-entropy $n^{1/2+\gamma}$, for any constant $\gamma \in (0,1/2)$, Kamp and

Manuscript received June 09, 2008; revised July 16, 2010. Date of current version November 19, 2010. The work of C.-J. Lu was supported (in part) by the National Science Council of Taiwan under contract NSC-97-2221-E-001-012-MY3 and was performed while she was with the Department of Computer Science, National Chiao Tung University, Hsinchu, Taiwan. The work of S.-C. Tsai was supported (in part) by the National Science Council of Taiwan under Contracts NSC-97-2221-E-009-064-MY3 and NSC-98-2221-E-009-078-MY3. The material in this paper was presented (in part) at the 33rd International Colloquium on Automata, Languages, and Programming (ICALP 2006), Venice, Italy, July 2006.

C.-J. Lee and C.-J. Lu are with the Institute of Information Science, Academia Sinica, Taipei, Taiwan (e-mail: leecj@iis.sinica.edu.tw; cjlu@iis.sinica.edu.tw).

S.-C. Tsai is with the Department of Computer Science, National Chiao Tung University, Hsinchu, Taiwan (e-mail: sctsai@csie.nctu.edu.tw).

Communicated by T. Fujiwara, Associate Editor for Complexity and Cryptography.

Digital Object Identifier 10.1109/TIT.2010.2079012

Zuckerman [14] gave a seedless extractor which can extract $\Omega(n^{2\gamma})$ bits of randomness. Building on this result together with some new idea, Gabizon *et al.* [9] were able to extract even more randomness. In particular, when the source has min-entropy $k > n^{1/2+\gamma}$, they can extract $k - n^{1/2+\gamma}$ bits and when $k > \log^c n$ for some constant c, they can extract $k - k^{\Omega(1)}$ bits.

Note that the two lines of research discussed above can be seen as belonging to two extremes of a spectrum in the following sense. Sources in both cases consist of multiple parts which are mutually independent. In the first case, one usually has in mind sources with relatively few parts while each part is long and contains a substantial amount of randomness. In the second case, a bit-fixing source consists of many parts, while each part is only a single bit either random or fixed. We would like to put both cases in the same framework and study sources that lie in between these two extremes.

A. Independent-Symbol Sources

We consider the following more general class of sources, characterized by the parameters n, d, $k \in \mathbb{N}$, which we call independent-symbol sources. Each source in the class consists of n mutually independent parts, each of length d, and the whole source has min-entropy k. For small n and large d, this covers sources of the first type, while for large n and $d = 1$, this covers sources of the second type. For other ranges of n and d, very little is known, and the main focus of our paper is to extract randomness from such sources.

Previously, [16], [15] were able to extract randomness from such a source with the condition that there are two parts in it with a combined min-entropy slightly above d. Independent of our work, Kamp *et al.* [13] recently also considered the same class of sources as ours and obtained some similar results. Furthermore, they showed that extractors for such sources also work for a more general class of sources which can be generated in small space.

Note that for deterministic extractors, the goal is to maximize the number m of extracted bits (or equivalently to minimize the entropy loss $k - m$) and to minimize the distance ε, which we call error, of its output distribution to the uniform one.

B. Our Results

Our first result (Theorem 1 in Section III) gives an explicit extractor which works for any min-entropy k but extracts only about $\log k$ random bits. More precisely, for any $n, d, k \in \mathbb{N}$ and $\varepsilon \in (0, 1)$, our extractor can extract $\Omega(\log k - \log\log(1/\varepsilon))$ bits with error ε. This can be seen as a generalization of the extractor of Kamp and Zuckerman [14], but note that theirs only works for bit-fixing sources and does not seem to work for the case that allows each bit having arbitrary bias. In fact, our extractor works for sources in which randomness could be distributed very nonuniformly among the n parts (e.g., some may have no min-entropy at all, but we do not know which ones), while previous constructions such as [3], [4], [23] do not seem to work for such sources. Independent of our work, Kamp *et al.* [13] also gave the same construction but used a different analysis.

To extract more randomness, we borrow the technique of Gabizon *et al.* [9]. Now, as in [9], we need n to be at least some large enough constant, and we have two constructions, both built on our first construction mentioned above. First, when $k \geq n^{1/2+\gamma}$, for any constant $\gamma \in (0, 1/2)$, we can extract $m = k - O(d\log(1/\varepsilon))$ random bits with any error $\varepsilon \geq 2^{-\Omega(n^\gamma)}$ (Theorem 2 in Section IV). Second, when $k \geq \log^c n$, for some constant $c > 0$, we can extract $m = k - (1/\varepsilon)^{O(1)}$ bits with error $\varepsilon \geq k^{-\Omega(1)}$ (Theorem 3 in Section IV). That is, when the min-entropy k is high, we can have a small entropy loss and a small error, but when k is small, the loss and error become larger. Note that the two main results in [9] only work for bit-fixing sources (with $d = 1$) and follow from our two with $\varepsilon = 2^{-\Omega(n^\gamma)}$ and $m = k - O(n^\gamma)$, and $\varepsilon = k^{-\Omega(1)}$ and $m = k - k^{\Omega(1)}$, respectively. On the other hand, we cover a large range of d and ε, and capture the tradeoff between error and entropy loss. For example, for constant d and ε, we show that the entropy loss can be lowered to a constant.

One may wonder if the entropy loss can be further reduced. We show that this is indeed possible, by proving the existence of a seedless extractor which can extract $m = k - O(\log(1/\varepsilon))$ random bits for $k = \Omega(d + \log(n/\varepsilon))$ (Theorem 4 in Section V). However, the existence is not shown in an explicit way; we only know such an extractor exists but we do not know how to construct it. Still, this shows that better explicit constructions than ours may be possible. We only have an explicit construction matching this bound for the case with $d = O(1)$, $k \geq n^{1/2+\gamma}$, and $\varepsilon \geq 2^{-\Omega(n^\gamma)}$.

On the other hand, one may also wonder whether this existential upper bound we derive on entropy loss is tight. Our final result (Theorem 5 in Section VI) shows that this is indeed the case by giving a matching lower bound. In fact, we show that even for the case of bit-fixing sources and even allowing a seed of length s, any extractor can only extract $k + s - \Omega(\log(1/\varepsilon))$ random bits. That is, even to extract from bit-fixing sources, any extractor, seeded or not, must suffer an entropy loss of $\Omega(\log(1/\varepsilon))$. This generalizes the result of Radhakrishnan and Ta-Shma [25], which has the same bound on seeded extractors for *general* sources. The idea in [25] is to show that for any extractor with output longer than the bound, one can find a (general) source on which it fails, and our task is much harder because we need to find one from the much more restricted class of bit-fixing sources.

C. Our Techniques

Our first extractor, which extracts about $\log k$ bits, was inspired by that of Kamp and Zuckerman [14], but our approach is quite different. Instead of taking a random walk on an odd cycle, we walk on the group $\mathbb{Z}_M$ for a prime M. More precisely, given a source $\mathcal{X} = (\mathcal{X}_1, \ldots, X_n)$, we see each $\mathcal{X}_i$ as an element of $\mathbb{Z}_M$ and outputs $\mathcal{X}_1 + \cdots + \mathcal{X}_n$ over Z_M. More precisely, after reading the i'th symbol $\mathcal{X}_i$, we walk from the state $S = \mathcal{X}_1 + \cdots + \mathcal{X}_{i-1}$ to the state $S + \mathcal{X}_i$. As in [14], we will show that each step of our walk brings the distribution closer to uniform when the symbol from the source contains some randomness. However, even for the case of $d = 1$, we cannot directly use the analysis from [14], which is based on bonding the second eigenvalue of the transition matrix for a perfectly random step on a cycle. This is because we may walk in a highly biased way as each bit of our source can have an arbitrary

bias. Our proof is very different and elementary, and has the following interesting point. The recent breakthrough construction of multi-source extractors [3] and its subsequent works all relied on using both sums and products to increase entropy. We show that in fact even doing sums alone can increase entropy. The increase, however, is slower, so we need a larger number of sources (as opposed to a constant number in [3]).

To extract more randomness, we apply the technique of [9]. Our constructions and proofs in this part follow very closely those in [9]. The only difference is that we deal with a more general classes of sources, do a more careful analysis, and use our first extractor instead of that in [14] as a building block.

Our existential upper bound on entropy loss is proved via a probabilistic argument. That is, we generate a seedless extractor randomly, and show that it works for all of our sources with a positive probability. For each source, we can show that it fails with a small probability. However, the number of all possible sources is in fact infinite. Instead, we show that it suffices to consider only a small set of sources, since any source is close to a convex combination of them. Sources in this set are those with the property that their distributions in each dimension are "almost flat" and have only a small number of possible min-entropy values.

Our lower bound proof of entropy loss follows the outline of that in [25]. Namely, given any function EXT: $\{0,1\}^n \times \{0,1\}^s \to \{0,1\}^m$ with $m \geq k + s - o(\log(1/\varepsilon))$, we show the existence of a bit-fixing source with min-entropy k on which the error of EXT exceeds ε, again using a probabilistic argument. We generate a source by randomly picking $n - k$ bits of the source and fixing them to some random values; the remaining k bits are left free and given a uniform distribution. The difficult part is to show that any such EXT fails on such a randomly chosen source with a positive probability. This probability turns out to be related to the size of some "almost" t-wise independent space, whose distribution is close to random on most sets of t dimensions. This can be seen as a relaxation of the standard notion of approximate t-wise independent space, in which the close-to-randomness property is required on *every* set of t dimensions. We prove a size lower bound on such a sample space, which seems to have an interest of its own. In particular, it immediately implies a size lower bound on any approximate t-wise independent space.

II. PRELIMINARIES

For $n \in \mathbb{N}$, let $[n]$ denote the set $\{1, \ldots, n\}$. For $x \in \{0,1\}^n$, $i \in [n]$ and $I \subseteq [n]$, let x_i denote the bit in the i'th dimension of x and x_I denote the projection of x onto those dimensions in I. For a set S, let $P(S)$ denote the collection of subsets of S, and let $P(S,t)$, for $t \in \mathbb{N}$, denote the collection of t-element subsets of S. All the logarithms in this paper will have base two.

When we sample from a finite set, the default distribution is the uniform one. For $n \in \mathbb{N}$, let $\mathcal{U}_n$ denote the uniform distribution over $\{0,1\}^n$. For a distribution $\mathcal{X}$ over a set S and an element $x \in S$, let $\mathcal{X}(x)$ denote the probability measure of x in the distribution $\mathcal{X}$. We say that a distribution $\mathcal{X}$ is a convex combination of distributions $\mathcal{X}^1, \ldots, \mathcal{X}^t$ over a set S, if there exist numbers $\alpha_1, \ldots, \alpha_t \geq 0$ with $\sum_{i \in [t]} \alpha_i = 1$ such that for every $x \in S$, $\mathcal{X}(x) = \sum_{i \in [t]} \alpha_i \mathcal{X}^i(x)$. We will sometimes see a distribution $\mathcal{X}$ over a set S as an $|S|$-dimensional vector, with $\mathcal{X}(x)$ at dimension $x \in S$. We will mainly measure the distance between two distributions $\mathcal{X}$, $\mathcal{X}'$ over S by their L_1-distance, defined as

$$\|\mathcal{X} - \mathcal{X}'\|_1 = \sum_{x \in S} |\mathcal{X}(x) - \mathcal{X}'(x)|.$$

Note that this distance is exactly twice the variational distance, defined as

$$\max_{A \subseteq S} \left| \sum_{x \in A} \mathcal{X}(x) - \sum_{x \in A} \mathcal{X}'(x) \right|.$$

Another distance measure that will be used sometimes is the L_2-distance, defined as

$$\|\mathcal{X} - \mathcal{X}'\|_2 = \sqrt{\sum_{x \in S} (\mathcal{X}(x) - \mathcal{X}'(x))^2}.$$

Call a distribution ε-random if its L_1-distance to the uniform distribution is at most ε. We will measure the amount of randomness in a distribution $\mathcal{X}$ over S by its min-entropy, defined as

$$\mathrm{H}_\infty(\mathcal{X}) = \min_{x \in S} \log\left(1/\mathcal{X}(x)\right).$$

In this paper, we will focus on a special kind of sources which consist of n independent symbols over some set $[D]$.

Definition 1: A distribution $\mathcal{X} = (\mathcal{X}_1, \ldots, \mathcal{X}_n)$ over the set $[D]^n$ is called an (n, D)-source if the n symbols $\mathcal{X}_1, \ldots, \mathcal{X}_n$ are distributed independently from each other. An (n, D)-source with min-entropy k is called an (n, D, k)-source. A bit-fixing source is an $(n, 2)$-source with the additional condition that each bit of the source has min-entropy either 0 or 1.

When we talk about an (n, D, k)-source, we always assume $k \leq n \log D$ since any (n, D)-source has min-entropy at most $n \log D$. The task of this paper is to extract randomness from such (n, D, k)-sources.

Definition 2: For $n, D, k, s, m \in \mathbb{N}$ and $\varepsilon \in [0,1]$, a function EXT: $[D]^n \times \{0,1\}^s \to \{0,1\}^m$ is called an (n, D, k, ε)-extractor if for any (n, D, k)-source $\mathcal{X}$

$$\|\mathrm{EXT}(\mathcal{X}, \mathcal{U}_s) - \mathcal{U}_m\|_1 \leq \varepsilon.$$

The second input, of s-bit long, to an extractor is called its seed. We allow the case of $s = 0$ (i.e., without a seed) and we call such an extractor a *seedless* (or *deterministic*) extractor. The *entropy loss* of an extractor is defined as the value $k + s - m$, which is the difference between the amount of randomness given to the extractor and the amount of randomness it can extract. Minimizing this entropy loss is one of the main goals of extractor construction. Moreover, one usually prefers constructions which are *explicit*, in the sense that given any input, one can compute the output in polynomial time.

III. EXTRACTOR FROM RANDOM WALK

In this section, we give an explicit seedless extractor for independent-symbol sources, which works for any min-entropy k but only extracts about $\log k$ bits.

Theorem 1: For any $n, k, D \in \mathbb{N}$ and any prime number $M \geq D$, there is an explicit (n, D, k, ε)-extractor $\mathrm{Ext}_0 : [D]^n \to [M]$, with $\varepsilon \leq \sqrt{M} \cdot e^{-k/(8M^2 \log D)}$.

Note that for $k \geq \Omega(M^2 \log^2 D)$, our extractor has $\varepsilon \leq 2^{-\Omega(k/(M^2 \log D))}$. Alternatively, for any $\varepsilon \in (0,1)$, our extractor can extract $\Omega(\log k - \log\log D - \log\log(1/\varepsilon))$ bits. This achieves the same asymptotic bound as the recent result in [13], but here we provide a different and completely elementary proof.

To extract randomness, we will work on the group $\mathbb{Z}_M$, for a prime M, and see any symbol $\mathcal{X}_i \in [D]$ of the source as an element in $\mathbb{Z}_M$. Throughout this section, operation $+$ or $-$ on elements in $\mathbb{Z}_M$ is understood as an operation over the group $\mathbb{Z}_M$. Our extractor $\mathrm{Ext}_0 : [D]^n \to [M]$ is then defined as

$$\mathrm{Ext}_0(\mathcal{X}) = \sum_{t \in [n]} \mathcal{X}_t$$

which can be seen as taking an n-step walk on the group $\mathbb{Z}_M$, using the n symbols from the source in the following way. Each time when we are at some state $v \in \mathbb{Z}_M$ (initially at $0 \in \mathbb{Z}_M$) and read a symbol a from the source, we go to the state $v + a \in \mathbb{Z}_M$. The extractor of Kamp and Zuckerman [14] for bit-fixing sources can be seen as a special case of ours, with $D = 2$ and $\mathcal{X}_t \in \{-1, 1\}$.

As in [14], we will show that each step of the walk brings the distribution closer to uniform if the symbol read from the source contains some randomness. See a distribution over $\mathbb{Z}_M$ as an M-dimensional vector in the natural way. Suppose the current distribution is $\mathcal{P} = (\mathcal{P}_1, \ldots, \mathcal{P}_M)$ and the next symbol in the source has a distribution $\beta = (\beta_1, \ldots, \beta_M)$ (let $\beta_i = 0$ for $D + 1 \leq i \leq M$). Then the next distribution is $\bar{\mathcal{P}} = (\bar{\mathcal{P}}_1, \ldots, \bar{\mathcal{P}}_M)$ with

$$\bar{\mathcal{P}}_i = \sum_{j \in \mathbb{Z}_M} \beta_j \mathcal{P}_{i-j}$$

for $i \in \mathbb{Z}_M$. Let $\mathcal{U}$ denote the uniform distribution over $\mathbb{Z}_M$. Let $\delta = \mathcal{P} - \mathcal{U}$ and $\bar{\delta} = \bar{\mathcal{P}} - \mathcal{U}$, i.e., $\delta_i = \mathcal{P}_i - 1/M$ and $\bar{\delta}_i = \bar{\mathcal{P}}_i - 1/M$ for $i \in \mathbb{Z}_M$. The following is our key lemma which shows the progress we can make after each step.

Lemma 1: $\|\bar{\delta}\|_2^2 \leq \|\delta\|_2^2 \cdot (1 - \mathrm{H}_\infty(\beta)/(4M^2 \log D))$.

We will prove this lemma in Section III-A. Now let us see how it can be used to prove the theorem.

Proof: (of Theorem 1)

From Lemma 1, we know that after reading the t'th symbol $\mathcal{X}_t$ from the source, the L_2-distance between the resulting distribution and the uniform one decreases by a factor

$$1 - \mathrm{H}_\infty(\mathcal{X}_t)/(4M^2 \log D) \leq e^{-\mathrm{H}_\infty(\mathcal{X}_t)/(4M^2 \log D)}.$$

Therefore, we have

$$\begin{aligned} \|\mathrm{Ext}_0(\mathcal{X}) - \mathcal{U}\|_2^2 &\leq \prod_{t \in [n]} e^{-\mathrm{H}_\infty(\mathcal{X}_t)/(4M^2 \log D)} \\ &= e^{-\sum_{t \in [n]} \mathrm{H}_\infty(\mathcal{X}_t)/(4M^2 \log D)}. \end{aligned}$$

Since the n symbols of the source are independent of each other, we have $\sum_{t \in [n]} \mathrm{H}_\infty(\mathcal{X}_t) = \mathrm{H}_\infty(\mathcal{X}) = k$, so the bound above becomes $e^{-k/(4M^2 \log D)}$. Then by Cauchy-Schwartz inequality

$$\begin{aligned} \|\mathrm{Ext}_0(\mathcal{X}) - \mathcal{U}\|_1 &\leq \sqrt{M} \cdot \|\mathrm{Ext}_0(\mathcal{X}) - \mathcal{U}\|_2 \\ &\leq \sqrt{M} \cdot e^{-k/(8M^2 \log D)}. \end{aligned}$$

■

A. Proof of Lemma 1

Note that for $i \in \mathbb{Z}_M$, $\bar{\delta}_i = \sum_{j \in \mathbb{Z}_M} \beta_j \delta_{i-j}$. So

$$\begin{aligned} \|\bar{\delta}\|_2^2 &= \sum_i \left(\sum_j \beta_j \delta_{i-j} \right)^2 \\ &= \sum_i \sum_j \beta_j^2 \delta_{i-j}^2 + \sum_i \sum_{j \neq \ell} \beta_j \beta_\ell \delta_{i-j} \delta_{i-\ell} \end{aligned}$$

which, using the equality $ab = (a^2 + b^2 - (a-b)^2)/2$ on the second term, equals

$$\begin{aligned} &\sum_j \beta_j^2 \sum_i \delta_{i-j}^2 \\ &+ \sum_{j \neq \ell} \beta_j \beta_\ell \sum_i \left(\delta_{i-j}^2 + \delta_{i-\ell}^2 - (\delta_{i-j} - \delta_{i-\ell})^2 \right)/2 \\ &= \sum_j \beta_j^2 \|\delta\|_2^2 + \sum_{j \neq \ell} \beta_j \beta_\ell \|\delta\|_2^2 \\ &\quad - \sum_{j \neq \ell} \beta_j \beta_\ell \sum_i (\delta_{i-j} - \delta_{i-\ell})^2/2 \\ &= \|\delta\|_2^2 - \sum_{j \neq \ell} \beta_j \beta_\ell \sum_i (\delta_i - \delta_{i+j-\ell})^2/2 \end{aligned}$$

where the last line follows from the fact that $\sum_j \beta_j^2 + \sum_{j \neq \ell} \beta_j \beta_\ell = (\sum_j \beta_j)^2 = 1$. Then we need the following two claims.

Claim 1: For any nonzero $s \in \mathbb{Z}_M$, $\sum_{i \in \mathbb{Z}_M} (\delta_i - \delta_{i+s})^2 \geq \|\delta\|_2^2 / M^2$.

Proof: First, by an average argument, there exists some $i_0 \in \mathbb{Z}_M$ such that $\delta_{i_0}^2 \geq \|\delta\|_2^2 / M$. Next, since $\sum_i \delta_i = 0$, there exists some $i_1 \in \mathbb{Z}_M$ such that δ_{i_1} and δ_{i_0} have different signs, so $|\delta_{i_0} - \delta_{i_1}|^2 \geq \delta_{i_0}^2 \geq \|\delta\|_2^2 / M$. Since M and s are relatively prime, the sequence of elements $i_0, i_0 + s, i_0 + 2s, \ldots$ in $\mathbb{Z}_M$ must have period M and contain every element of $\mathbb{Z}_M$. Thus, there exists an integer $t \in [1, M-1]$ such that $i_1 = i_0 + ts$ over $\mathbb{Z}_M$. By a triangle inequality, $\sum_{1 \leq j \leq t} |\delta_{i_0+(j-1)s} - \delta_{i_0+js}| \geq |\delta_{i_0} - \delta_{i_0+ts}| = |\delta_{i_0} - \delta_{i_1}|$. Finally

$$\sum_{i \in \mathbb{Z}_M} (\delta_i - \delta_{i+s})^2 \geq \sum_{1 \leq j \leq t} (\delta_{i_0+(j-1)s} - \delta_{i_0+js})^2$$

which by Cauchy-Schwartz inequality is at least

$$\begin{aligned} \left(\sum_{1 \leq j \leq t} |\delta_{i_0+(j-1)s} - \delta_{i_0+js}| \right)^2 / t &\geq |\delta_{i_0} - \delta_{i_1}|^2 / t \\ &\geq \|\delta\|_2^2 / M^2. \end{aligned}$$

■

Claim 2: $\sum_{j\neq\ell}\beta_j\beta_\ell \geq \mathrm{H}_\infty(\beta)/(2\log D)$.

Proof: Let $\hat{\beta} = \max\{\beta_i : i \in [M]\}$, so $\mathrm{H}_\infty(\beta) = \log(1/\hat{\beta})$. Then we have

$$\sum_{j\neq\ell}\beta_j\beta_\ell = \sum_j \beta_j \sum_{\ell\neq j}\beta_\ell \geq \sum_j \beta_j(1-\hat{\beta}) = 1-\hat{\beta}.$$

Note that β is a distribution over $[D]$, so $\hat{\beta}\in[1/D,1]$. For $\hat{\beta}$ in this range, we have

$$1-\hat{\beta} \geq (\log(1/\hat{\beta}))(1-1/D)/\log D \geq \mathrm{H}_\infty(\beta)/(2\log D).$$

■

Using the bounds of the claims in our derivation before, we have

$$\begin{aligned}\|\bar{\delta}\|_2^2 &\leq \|\delta\|_2^2 \cdot \left(1-\sum_{j\neq\ell}\beta_j\beta_\ell/(2M^2)\right)\\ &\leq \|\delta\|_2^2\cdot(1-\mathrm{H}_\infty(\beta)/(4M^2\log D))\end{aligned}$$

which proves the lemma.

IV. Extracting More Randomness

The extractor in the previous section can extract about $\log k$ bits of randomness. Building on this, we show how to extract more randomness in this section. More precisely, we have the following two extractors, which generalize the corresponding ones in [9]. The first one works for the case of large min-entropy and can achieve a smaller error and a smaller entropy loss, while the second can work for the case of smaller min-entropy but has a larger error and a larger entropy loss.

Theorem 2: For any constant $\gamma\in(0,1/2)$, and $D=2^d\in\mathbb{N}$, there exist constants $n_0>0$, $c>0$ such that for any $n\geq n_0$, $k\geq n^{1/2+\gamma}$, and $\varepsilon\geq 2^{-cn^\gamma}$, there exists an explicit seedless (n,D,k,ε)-extractor $\mathrm{Ext}:[D]^n\to\{0,1\}^m$ with $m\geq k-O(d\log(1/\varepsilon))$.

Theorem 3: There exist constants $n_0>0$, $c_0\in(0,1)$, $c_1>0$, $c_2\in(0,1)$, $c_3\in(0,1/c_2)$ such that for any $n\geq n_0$, $D=2^d$ with $d\leq k^{c_0}$, $k\geq\log^{c_1}n$, and $\varepsilon\geq k^{-c_2}$, there exists an explicit seedless (n,D,k,ε)-extractor $\mathrm{Ext}:[D]^n\to\{0,1\}^m$ with $m\geq k-O((1/\varepsilon)^{c_3})$.

Note that the two main results in [9] only work for bit-fixing sources (with $D=2$) and follow respectively from Theorem 2 with $\varepsilon=2^{-cn^\gamma}$ and $m=k-O(n^\gamma)$, and from Theorem 3 with $\varepsilon=k^{-c_2}$ and $m=k-k^{\Omega(1)}$. On the other hand, our two theorems above cover a large range of the parameters D and ε, and capture the tradeoff between error and entropy loss. In particular, for a small d, if we allow a large ε, the entropy loss can become very small.

We will give the proofs of the two theorems in Sections IV-B and IV-C respectively, which follow closely the corresponding ones in [9]. The main difference is that we consider independent-symbol sources, so we cannot build on the extractor of [14] as [9] did, and instead, we build on our extractor in Theorem 1. Furthermore, we do a more careful analysis in order to identify the relationship between error and entropy loss. Before giving the proofs, let us first describe some basic ideas and useful tools.

Suppose we have extracted a short random string z from the source $\mathcal{X}$. One may think about using z as a seed for a seeded extractor to extract more randomness from $\mathcal{X}$, but the problem is that z may have dependence on $\mathcal{X}$. This issue was taken care of in [9] by constructing the so-called seed obtainer. The idea is to divide z into two parts (w,y) and use w to sample a set $S(w)\subseteq[n]$ of positions from the source so that $\mathcal{X}_{[n]\setminus S(w)}$ still has enough min-entropy but becomes independent of y. To guarantee this, we would like the set $S(w)$ to have the property that the min-entropy of the sampled bits is within a certain range, which can be achieved by using the so-called averaging sampler.

Definition 3: Suppose n, d, $k\in\mathbb{N}$, $\delta\in(0,1)$, and $k_{\min},k_{\max}\in\mathbb{R}$, with $0\leq k_{\min}\leq k_{\max}\leq k\leq n$. An $(n,d,k,k_{min},k_{max},\delta)$-sampler $S:\{0,1\}^t\to P([n])$ is a function such that for every function $h:[n]\to[0,d]$ with $\sum_{i\in[n]}h(i)=k$

$$\Pr_{w\in\mathcal{U}_t}\left[k_{\min}\leq\sum_{i\in S(w)}h(i)\leq k_{\max}\right]\geq 1-\delta.$$

Throughout this section, we will let $d=\log D$ for an (n,D)-source $\mathcal{X}=(\mathcal{X}_1,\ldots,\mathcal{X}_n)$, and h will be the function such that $h(i)=\mathrm{H}_\infty(\mathcal{X}_i)\in[0,d]$. Note that the definition of samplers used in [9] is a special case of ours, as it only deals with Boolean functions $h:[n]\to\{0,1\}$, which arise from bit-fixing sources considered there. As shown in [9], after obtaining $\mathcal{X}_{[n]\setminus S(w)}$ of enough min-entropy together with an independent seed y, one can then apply a seeded extractor to extract more randomness. This is guaranteed by the following lemma. Note that this was proved in [9] for bit-fixing sources, but it is easy to check that the same proof indeed works for our independent-symbol sources.

Lemma 2: [9] Suppose there exist explicit constructions for the following three ingredients: 1) a seedless $(n,D,k_{\min},\varepsilon_1)$-extractor $\mathrm{Ext}_1:[D]^n\to\{0,1\}^{t+s}$; 2) an $(n,k,k_{\min},k_{\max},\delta)$-sampler $\mathrm{Samp}:\{0,1\}^t\to P([n])$; and 3) a seeded $(n,D,k-k_{\max},\varepsilon_2)$-extractor $\mathrm{Ext}_2:[D]^n\times\{0,1\}^s\to\{0,1\}^m$. Then there exists an explicit seedless (n,D,k,ε_3)-extractor $\mathrm{Ext}_3:[D]^n\to\{0,1\}^m$ with $\varepsilon_3=3\max(\varepsilon_1+\delta,2^{t+1}\varepsilon_1)+\varepsilon_2$.

A. Sampling and Partitioning

For our two extractors, we need the following two samplers respectively. Both constructions basically come from [9], and the proofs are very similar. The first sampler uses a longer seed and achieves a smaller error probability, while the second one uses a shorter seed but has a higher error probability.

Lemma 3: There exist constants n_0, c_1, c_2 such that for any $n\geq n_0$, k, $d\in\mathbb{N}$, $\delta\geq 2^{-c_1k}$, and $k_{\min}\geq c_2d\log(1/\delta)$, there exists an explicit $(n,d,k,k_{\min},6k_{\min},\delta)$-sampler $\mathrm{Samp}:\{0,1\}^t\to P([n])$ with $t=O(\log n\cdot\log(1/\delta))$.

Lemma 4: For any constant $\alpha\in(0,1)$, there exist constants $n_0>0$, $c_0\in(0,1)$, $c_1>0$, $\beta\in(0,1)$, $\tau\in(1/2,1)$

such that the following holds. For any $n \geq n_0$, $d \leq k^{c_0}$, $k \geq \log^{c_1} n$, and $\delta = O(k^{-\beta})$, there exists an explicit $(n, d, k, k^\tau/2, 3k^\tau, \delta)$-sampler $\text{SAMP} : \{0,1\}^t \to P([n])$ with $t = \alpha \log k$.

The proof of Lemma 3 is given in Appendix A. Lemma 4 follows immediately from Lemma 5 below by using T_1 as the output of the sampler SAMP.

Lemma 5: For any constant $\alpha \in (0,1)$, there exist constants $n_0 > 0$, $c_0 \in (0,1)$, $c_1 > 0$, $\beta \in (0,1)$, $\tau \in (1/2, 1)$ such that the following holds. For any $n \geq n_0$, $d \leq k^{c_0}$, $k \geq \log^{c_1} n$, and $\delta = O(k^{-\beta})$, one can use $\alpha \log k$ random bits to explicitly partition $[n]$ into $r' = \Omega(k^\beta)$ sets $T_1, \ldots, T_{r'}$ such that for any function $h : [n] \to [0, d]$ with $\sum_{i=1}^n h(i) = k$

$$\Pr\left[\forall v \in [r'], k^\tau/2 \leq \sum_{i \in T_v} h(i) \leq 3k^\tau\right] \geq 1 - \delta.$$

In addition to proving Lemma 4, Lemma 5 will also be used to prove Theorem 3. We give the proof of Lemma 5 in Appendix B.

B. Proof of Theorem 2

The construction is very similar to that in [9]. First, as in [14], we have the following seedless extractor for the case of large min-entropy.

Lemma 6: For any large enough $n \in \mathbb{N}$ and any $k_1 \geq n^{1/2+\gamma}$ with $\gamma \in (0, 1/2)$, there exists an explicit seedless $(n, D, k_1, \varepsilon_1)$-extractor $\text{EXT}_1 : [D]^n \to \{0,1\}^{m_1}$ where $m_1 = \Omega(n^{2\gamma}/(D^2 d^2))$ and $\varepsilon_1 = 2^{-m_1}$.

Proof: This lemma is a generalization of the main result in [14] for bit-fixing sources. The proof is very similar, so we only give a sketch here.

The extractor works as follows. Let p be the smallest prime greater than D. Set $k_0 = c_0 p^2 \log^2 p$, for some large enough constant c_0. Partition the n symbols of the source into $b = k_1/(2k_0)$ blocks, each consisting of $2nk_0/k_1$ symbols (assuming for simplicity that k_1 is a multiple of $2k_0$ and $2nk_0$ is a multiple of k_1). Within each block, use our extractor in Theorem 1 to extract a symbol in $\mathbb{Z}_p$. Then use the b extracted symbols (one per block) to take a b-step walk on an expander which has 2^{m_1} nodes, for some m_1 to be determined later, and has its second eigenvalue $\lambda \leq 1/p^{c_1}$, for some constant $c_1 < 1/2$. The final node of the walk is the output of the extractor.

Call a block *good* if it has min-entropy at least k_0. By a Markov inequality, the number of good blocks is at least

$$\ell = k_1^2/(4k_0 n \log p).$$

For each good block, the extracted symbol is ε-random for some $\varepsilon \leq 1/p$ by Theorem 1. Then according to Lemma 3.8. of [14], after the b-step walk on the expander, the distribution of the final node is ε_1-random, for

$$\begin{aligned}\varepsilon_1 &\leq (\lambda + \varepsilon\sqrt{p})^\ell \cdot 2^{m_1/2} \\ &\leq 2^{-\Omega(k_1^2/(k_0 n))} \cdot 2^{m_1/2} \\ &\leq 2^{-\Omega(n^{2\gamma}/(D^2 d^2)) + m_1/2}.\end{aligned}$$

Then for some $m_1 = \Theta(n^{2\gamma}/(D^2 d^2))$, we have $\varepsilon_1 \leq 2^{-m_1}$, which proves the lemma. ∎

Following [9], to extract more randomness, we apply Lemma 2 with EXT_1 above together with two additional ingredients: (1) an $(n, d, k, k_{\min}, k_{\max}, \varepsilon/4)$-sampler $\text{SAMP} : \{0,1\}^t \to P([n])$ from Lemma 3, with $t = m_1/2$, $k_{\min} = O(d \log(1/\varepsilon))$, and $k_{\max} = 6k_{\min}$, and (2) a seeded $(n, D, k - k_{\max}, \varepsilon_2)$-extractor $\text{EXT}_2 : [D]^n \times \{0,1\}^s \to \{0,1\}^m$ from [24], with $s = m_1/2$, $m = k - k_{\max}$, and $\varepsilon_2 = 2^{-\Omega(n^\gamma)}$. Note that the above three ingredients exist for large enough n. From Lemma 2, we get an (n, D, k, ε_3)-extractor $\text{EXT}_3 : [D]^n \times \{0,1\}^m$ with $\varepsilon_3 \leq 2^{-\Omega(n^\gamma)} + 3\varepsilon/4 + O(2^t \cdot \varepsilon_1) \leq 2^{-\Omega(n^\gamma)} + 3\varepsilon/4 \leq \varepsilon$, when $\varepsilon \geq 2^{-cn^\gamma}$ for a small enough constant c. This proves Theorem 2.

C. Proof of Theorem 3

The construction is again very similar to the corresponding one in [10]. Suppose $k \geq \log^{c_1} n$ for a large enough constant c_1. We first use the seedless extractor in Theorem 1 to extract $O(\log k)$ bits of randomness. To apply Lemma 2 to extract more randomness, we need a seeded extractor with such a short seed. Similar to [9], the existence of such an extractor is guaranteed by the following.

Lemma 7: For any constant $\alpha \in (0,1)$, there exist constants $n_0 > 0$, $c_0 \in (0,1)$, $c_1 > 0$ such that the following holds. For any $n \geq n_0$, $D = 2^d$ with $d \leq k^{c_0}$, and $k \geq \log^{c_1} n$, there exists an explicit seeded (n, D, k, ε')-extractor $\text{EXT}' : [D]^n \times \{0,1\}^s \to \{0,1\}^m$ with $s = \alpha \log k$, $m = k^{\Omega(1)}$, and $\varepsilon' = k^{-\Omega(1)}$.

Proof: The idea is to use the short seed for the partitioner in Lemma 5 to partition the source into several parts and then apply our seedless extractor in Theorem 1 on each part.

Fix any constant $\alpha \in (0,1)$. Consider any (n, D, k)-source $\mathcal{X} = (\mathcal{X}_1, \ldots, \mathcal{X}_n)$. Let $h : [n] \to [0, d]$ be the function $h(i) = \text{H}_\infty(\mathcal{X}_i)$. Note that $\sum_{i=1}^n h(i) = k$. According to Lemma 5, using $\alpha \log k$ random bits, one can partition the set $[n]$ into r' subsets $T_1, \ldots, T_{r'}$ such that the probability that $\sum_{i \in T_v} h(i) \geq k^\tau/2$ for every $v \in [r']$ (call such $(T_1, \ldots, T_{r'})$ *good*) is at least $1 - k^{-\Omega(1)}$.

Define our extractor EXT' as $\text{EXT}'(x) = z_1 \circ \cdots \circ z_{r'}$, where $z_v = \text{EXT}_0(x_{T_v} \circ 0^{n - |T_v|})$ for $v \in [r']$ and $\text{EXT}_0 : [D]^n \to \{0,1\}^\ell$ is our $(n, D, k_0, \varepsilon_0)$-extractor in Theorem 1, with $k_0 = k^\tau/2$, $\ell = O(\log k)$, and $\varepsilon_0 = k^{-\omega(1)}$. We want to prove that $\text{EXT}'(\mathcal{X})$ is close to $\mathcal{U}_m$ where $m = \ell \cdot r' = k^{\Omega(1)}$. Note that when $(T_1, \ldots, T_{r'})$ is good, the distribution of each z_j is ε_0-random, and, by a standard hybrid argument (see e.g., [10]), the distribution of $z_1 \circ \cdots \circ z_{r'}$ is $(r'\varepsilon_0)$-random, with $r'\varepsilon_0 = k^{O(1)} k^{-\omega(1)} \leq k^{-\Omega(1)}$. Thus

$$\begin{aligned}&\|\text{EXT}'(\mathcal{X}) - \mathcal{U}_m\|_1 \\ &\leq \Pr\left[(T_1, \ldots, T_{r'}) \text{ is not good}\right] + k^{-\Omega(1)} \\ &\leq k^{-\Omega(1)}.\end{aligned}$$

∎

Then we can apply Lemma 2 with the following ingredients: 1) a seedless (n, D, k, ε_1)-extractor $\text{EXT}_1 : [D]^n \to [M]$ from

Theorem 1, with $\varepsilon_1 = k^{-\Omega(1)}$ and $\log M \geq 2\alpha \log k$ for a small enough constant $\alpha \in (0,1)$; 2) an $(n, d, k, k_{\min}, k_{\max}, \delta)$-sampler $\text{SAMP} : \{0,1\}^t \to P([n])$ from Lemma 4, with $t = \alpha \log k$, $k_{\min} \leq k^c$ for a constant $c \in (0,1)$, $k_{\max} = 6k_{\min}$, and $\delta = k^{-\Omega(1)}$; and 3) a seeded $(n, D, k - k_{\max}, \varepsilon_2)$-extractor $\text{EXT}_2 : [D]^n \times \{0,1\}^s \to \{0,1\}^{m_2}$ from Lemma 7 with $s = \alpha \log k$, $m_2 = k^{\Omega(1)}$, and $\varepsilon_2 = k^{-\Omega(1)}$. As a result, we obtain a seedless (n, D, k, ε_3)-extractor $\text{EXT}_3 : [D]^n \to \{0,1\}^{m_2}$, with $\varepsilon_3 = k^{-\Omega(1)} + O(2^t \varepsilon_1) = k^{-\Omega(1)}$.

To extract even more random bits, we again apply Lemma 2, but now using the above extractor EXT_3 together with the following two ingredients: 1) an $(n, d, k, k_{\min}, k_{\max}, \varepsilon/4)$-sampler $\text{SAMP} : \{0,1\}^t \to P([n])$ from Lemma 4 with $t = \alpha \log k \leq m_2/2$, $k_{\min} = O((1/\varepsilon)^{c_3})$, and $k_{\max} = 6k_{\min}$ and 2) a seeded $(n, D, k - k_{\max}, 1/n)$-extractor $\text{EXT}_2 : \{0,1\}^n \times \{0,1\}^s \to \{0,1\}^m$ from [24], with $s \leq m_2/2$ and $m = k - k_{\max}$. As a result, we obtain a seedless (n, D, k, ε)-extractor $\text{EXT} : [D]^n \to \{0,1\}^m$, since $k^{-\Omega(1)} + 3\varepsilon/4 + O(2^t/n) \leq k^{-\Omega(1)} + 3\varepsilon/4 \leq \varepsilon$, when $\varepsilon \geq k^{-c_2}$ for a small enough constant c_2. This proves Theorem 3.

V. Existential Upper Bound on Entropy Loss

In the previous section, we obtain two explicit extractors for independent-symbol sources. One may wonder if it is possible to extract more randomness and achieve a smaller entropy loss for such sources. In this section, we prove the existence of a (nonexplicit) seedless extractor for independent-symbol sources with entropy loss $O(\log(1/\varepsilon))$. More precisely, we have the following theorem, whose proof is given in Section V-A.

Theorem 4: Suppose $k \geq c \log(Dn/\varepsilon)$ for a large enough constant c. Then there exists an (n, D, k, ε)-extractor $\text{EXT} : [D]^n \to \{0,1\}^m$ with $m \geq k - O(\log(1/\varepsilon))$.

We will show the existence of such an extractor by a probabilistic argument. More precisely, we will show that if we choose a random function as our extractor EXT, then we succeed with a positive probability.

A. Proof of Theorem 4

Let $\mathcal{F}$ denote the set of all functions $f : [D]^n \to \{0,1\}^m$. We say that a function $f \in \mathcal{F}$ fails on an (n, D, k)-source $\mathcal{X}$ if $\|f(\mathcal{X}) - \mathcal{U}_m\|_1 > \varepsilon/2$. We have the following.

Lemma 8: For any (n, D, k)-source $\mathcal{X}$, we have

$$\Pr_{f \in \mathcal{F}}[f \text{ fails on } \mathcal{X}] \leq 2^{2^m} \cdot 2^{-\Omega(\varepsilon^2 2^k)}.$$

Proof: Consider any (n, D, k)-source $\mathcal{X}$. For a test $T \subseteq \{0,1\}^m$, we say that f fails on $(\mathcal{X}, T)$ if $|\Pr_{x \in \mathcal{X}}[f(x) \in T] - |T|/2^m| > \varepsilon/2$. Clearly, f fails on $\mathcal{X}$ if and only if f fails on $(\mathcal{X}, T)$ for some $T \subseteq \{0,1\}^m$. Now consider any test $T \subseteq \{0,1\}^m$, and we would like to bound the probability that a random f fails on $(\mathcal{X}, T)$.

Suppose $|T|/2^m = p$. For $x \in [D]^n$, let Y_x be the indicator random variable for the event $f(x) \in T$. Then

$$\Pr_{f \in \mathcal{F}}[f \text{ fails on } (\mathcal{X}, T)] = \Pr_{f \in \mathcal{F}}\left[\left|\sum_x \mathcal{X}(x) Y_x - p\right| > \varepsilon/2\right].$$

Note that the probability is a weighted sum of the random variables Y_x's, with each weight $\mathcal{X}(x)$ being at most 2^{-k}. Let us consider instead the random variable $Z_x = (\mathcal{X}(x)2^k)Y_x$, which now takes its value in the interval $[0,1]$, and note that $\text{E}_{f \in \mathcal{F}}[\sum_x Z_x] \leq 2^k p$. Then

$$\Pr_{f \in \mathcal{F}}[f \text{ fails on } (\mathcal{X}, T)] = \Pr_{f \in \mathcal{F}}\left[\left|\sum_x Z_x - 2^k p\right| > 2^k \varepsilon/2\right]$$

which by a Chernoff bound is at most

$$2^{-\Omega((\varepsilon/p)^2 2^k p)} \leq 2^{-\Omega(\varepsilon^2 2^k)}.$$

Since there are 2^{2^m} possible T's, a union bound gives the lemma. ∎

The lemma says that a random f fails on each source with a small probability. However, there are infinitely many sources, since for any $i \in [n]$, $\text{H}_\infty(\mathcal{X}_i)$ can have an arbitrary value in the interval $[0, k]$. The following shows that it suffices to consider sources $\mathcal{X}'$ with $\text{H}_\infty(\mathcal{X}'_i)$, for each $i \in [n]$, being an (integral) multiple of $\alpha = 1/\lceil 2Dn/\varepsilon \rceil$.

Lemma 9: For any (n, D, k)-source $\mathcal{X}$, there exists an (n, D, k)-source $\mathcal{X}'$ such that $\|\mathcal{X} - \mathcal{X}'\|_1 \leq \varepsilon/2$ and $\text{H}_\infty(\mathcal{X}'_i)$ is a multiple of α for any $i \in [n]$.

Proof: For $i \in [n]$, let $k_i = \text{H}_\infty(\mathcal{X}_i)$. It is not hard to see that there exists $(k'_1, \ldots, k'_n)$ such that $k'_1 + \cdots + k'_n = k$ and for each $i \in [n]$, k'_i is a multiple of α and $|k'_i - k_i| < \alpha$, by rounding each k_i up or down to its nearest multiple of α.

Next, we construct a source $\mathcal{X}'$ from $\mathcal{X}$ with $\text{H}_\infty(\mathcal{X}'_i) = k'_i$ for every $i \in [n]$. As we consider (n, D)-sources, we can deal with the n dimensions of the sources separately. For $i \in [n]$ with $k'_i < k_i$, we keep shifting measure into a fixed element until its measure reaches $2^{-k'_i}$. For $i \in [n]$ with $''k_i > k_i$, we keep shifting measure away from an element while its measure exceeds $2^{-k'_i}$. Clearly, we can do this while keeping the measures of any element in $\mathcal{X}_i$ and $\mathcal{X}'_i$ within a distance $|2^{-k'_i} - 2^{-k_i}|$. Note that for the function $f(x) = 2^{-x}$, its derivative at any $x \geq 0$ has an absolute value at most 1, which implies $|2^{-k'_i} - 2^{-k_i}| \leq |k'_i - k_i|$ by the mean value theorem in calculus. Thus for any $i \in [n]$, $\|\mathcal{X}_i - \mathcal{X}'_i\|_1 \leq D \cdot |k'_i - k_i| < D \cdot \alpha \leq \varepsilon/(2n)$. Then by a standard hybrid argument (see, e.g., [10]), we have $\|\mathcal{X} - \mathcal{X}'\|_1 \leq \sum_{i \in [n]} \|\mathcal{X}_i - \mathcal{X}'_i\|_1 \leq \varepsilon/2$. Since $\text{H}_\infty(\mathcal{X}') = k'_1 + \cdots + k'_n = k = \text{H}_\infty(\mathcal{X})$, we have the lemma. ∎

The other issue is that when $D > 2$, given any $k_i \in [0, k]$, for $i \in [n]$, there are still infinitely many $\mathcal{Y}_i$ over $[D]$ that can have $\text{H}_\infty(\mathcal{Y}_i) = k_i$. The following shows that it suffices to consider (n, D, k)-sources $\mathcal{Y}$'s with each $\mathcal{Y}_i$ being "almost flat" in the sense that $\lfloor 2^{k_i} \rfloor$ elements in $[D]$ have measure 2^{-k_i}, one element has measure $1 - \lfloor 2^{k_i} \rfloor 2^{-k_i}$, and the rest have measure 0.

Lemma 10: Any (n, D, k)-source $\mathcal{X}'$ can be expressed as a convex combination of (n, D, k)-sources $\mathcal{Y}$ with the property that for any $i \in [n]$, $\text{H}_\infty(\mathcal{Y}_i) = \text{H}_\infty(\mathcal{X}'_i)$ and $\mathcal{Y}_i$ is almost flat.

Proof: This is a generalization of the well known fact that any source with an integer min-entropy can be expressed as a convex combination of flat sources. Here, we need to deal with

real-valued min-entropy. The proof is similar so we will only give a sketch.

Consider any (n, D, k)-source $\mathcal{X}'$, with $\mathrm{H}_\infty(\mathcal{X}'_i) = k_i$ for $i \in [n]$. We claim that for each $i \in [n]$, the source $\mathcal{X}'_i$ can be expressed as a convex combination of almost-flat sources over $[D]$ with min-entropy k_i. The reason is the following. See any source over $[D]$ with min-entropy k_i as a vector $(p_1, \ldots, p_D)$ with the property that $\sum_{j\in[D]} p_j = 1$ and $0 \le p_j \le 2^{-k_i}$ for every $j \in [D]$. The set of such vectors forms a convex polytope, and each vector in the set is expressible as a convex combination of vertices (corners) of the polytope. The claim follows from the fact that the vertices of the polytope correspond exactly to the vectors given by those almost flat sources over $[D]$. Now as $\mathcal{X}'_i$ is a convex combination of almost-flat sources of min-entropy k_i for each $i \in [n]$, the source $\mathcal{X}'$ is a convex combination of (n, D, k)-sources $\mathcal{Y}$ in which $\mathcal{Y}_i$ is almost flat and has min-entropy k_i for $i \in [n]$. ∎

Let $\mathcal{S}$ denote the set of (n, D, k)-sources $\mathcal{Y}$ with the property that for every $i \in [n]$, $\mathcal{Y}_i$ is almost flat and $\mathrm{H}_\infty(\mathcal{Y}_i)$ is a multiple of α. The following gives a bound on the size of $\mathcal{S}$.

Lemma 11: $|\mathcal{S}| \le 2^{2^{O(\log(Dn))}}$.

Proof: Recall that $\alpha = 1/\lceil 2Dn/\varepsilon \rceil \le \varepsilon/(2Dn)$. Let us first bound the number of $(k_1, \ldots, k_n)$ such that $k_1 + \cdots + k_n = k$ and each $k_i \in [0, k]$ is a multiple of α for $i \in [n]$. Note that this is the same as the number of $(z_1, \ldots, z_n)$ such that $z_1 + \cdots + z_n = k/\alpha$ and z_i is an integer in $[0, k/\alpha]$ for $i \in [n]$. This number is exactly

$$\binom{k/\alpha + n - 1}{n - 1} \le 2^{O(n \log(Dn/\varepsilon))}.$$

Now for any $(k_1, \ldots, k_n)$, the number of (n, D, k)-sources $\mathcal{Y}$ such that each $\mathcal{Y}_i$, for $i \in [n]$, is almost flat with min-entropy k_i is at most $(2^D \cdot D)^n = 2^{O(Dn)}$. As a result, we have

$$|\mathcal{S}| \le 2^{O(n \log(Dn/\varepsilon))} \cdot 2^{O(Dn)} \le 2^{2^{O(\log(Dn))}}.$$

∎

From Lemma 8 and Lemma 11 and using a union bound, we have

$$\Pr_{f\in\mathcal{F}}[\exists \mathcal{Y} \in \mathcal{S}, f \text{ fails on } \mathcal{Y}] \le 2^{2^{O(\log(Dn))}} \cdot 2^{2^m} \cdot 2^{-\Omega(\varepsilon^2 2^k)} < 1$$

for some $m = k - O(\log(1/\varepsilon))$ when $k \ge c \log(Dn/\varepsilon)$ for a large enough constant c. This implies the existence of some $\mathrm{Ext} \in \mathcal{F}$ such that $\|\mathrm{Ext}(\mathcal{Y}) - \mathcal{U}_m\|_1 \le \varepsilon/2$ for any $\mathcal{Y} \in \mathcal{S}$, and thus for any $\mathcal{Y}$ which is a convex combination of sources in $\mathcal{S}$. According to Lemma 9 and Lemma 10, any (n, D, k)-source $\mathcal{X}$ has distance at most $\varepsilon/2$ to some source $\mathcal{Y}$ which is a convex combination of sources in $\mathcal{S}$, so

$$\|\mathrm{Ext}(\mathcal{X}) - \mathcal{U}_m\|_1 \le \|\mathcal{X} - \mathcal{Y}\|_1 + \|\mathrm{Ext}(\mathcal{Y}) - \mathcal{U}_m\|_1 \le \varepsilon.$$

That is, Ext is an (n, D, k, ε)-extractor, which proves Theorem 4.

VI. Lower Bound on Entropy Loss

In this section, we show that the existential upper bound on the entropy loss in Section V is tight by giving a matching lower bound. In fact, we show that even for bit-fixing sources and even allowing a seed, any extractor must suffer an entropy loss of $\Omega(\log(1/\varepsilon))$.

Theorem 5: Let $\mathrm{Ext} : \{0,1\}^n \times \{0,1\}^s \to \{0,1\}^m$ be an $(n, 2, k, \varepsilon)$-extractor for bit-fixing sources, with n, s, $m \in \mathbb{N}$, $\log(1/\varepsilon) \le k \le n - \log(1/\varepsilon)$, and $0 < \varepsilon < 1/c_1$, for some large enough constants c_1. Then $m \le k + s - \Omega(\log(1/\varepsilon))$.

We will basically follow the proof idea in [25]. Briefly speaking, given any $\mathrm{Ext} : \{0,1\}^n \times \{0,1\}^s \to \{0,1\}^m$ with m exceeding the bound, we will show the existence of a bit-fixing source of min-entropy k on which Ext fails, using a probabilistic argument. Before giving the proof, let us first state some definitions and lemmas which will be needed. For any $z \in \{0,1\}^m$, consider the set

$$S^{(z)} = \{x \in \{0,1\}^n : \exists y \in \{0,1\}^s \text{ s.t. } z = \mathrm{Ext}(x, y)\},$$

and we say that z is δ-missed by $X \subseteq \{0,1\}^n$ if

$$\left| \Pr_{x\in S^{(z)}}[x \in X] - \Pr_{x\in\mathcal{U}_n}[x \in X] \right| \ge \delta.$$

We will rely on the following lemma from [25].[1]

Lemma 12: Suppose $\mathcal{X}$ is the uniform distribution over a set $X \subseteq \{0,1\}^n$ with $|X| = 2^k$, and $\|\mathrm{Ext}(\mathcal{X}, \mathcal{U}_s) - \mathcal{U}_m\|_1 \le \varepsilon$. Then at most $4\sqrt{\varepsilon}$ fraction of $z \in \{0,1\}^m$ can be $(2^{-(n-k)}\sqrt{\varepsilon})$-missed by X.

For n, $t \in \mathbb{N}$, $\beta \in (0,1)$, $I \in P([n], t)$, $u \in \{0,1\}^t$, and $S \subseteq \{0,1\}^n$, we say that u is (I, β)-biased in S if

$$\left| \Pr_{x\in S}[x_I = u] - 2^{-t} \right| > \beta.$$

Our key lemma is the following.

Lemma 13: Suppose n, $t \in \mathbb{N}$ and $\delta \in (0,1)$, with $t \le n - \Omega(\log(1/\delta))$ and $1/(8\binom{n}{t}2^t) < \delta < 1/c_2$ for some large enough constant c_2. Consider any $S \subseteq \{0,1\}^n$ satisfying the property that over a random $I \in P([n], t)$ and a random $u \in \{0,1\}^t$, u is $(I, 2^{-t}\delta)$-biased in S with probability at most 8δ.[2] Then $|S| \ge 2^t (1/\delta)^{\Omega(1)}$.

Note that a set S satisfying the property in Lemma 13 can be seen as an "almost" t-wise independent space, in the sense that the uniform distribution over S looks random on most sets of t dimensions. This can be seen as a relaxation of the standard notion of approximate t-wise independent space. Lemma 13 gives a size lower bound on such a set, which seems to have an interest of its own. We will prove the lemma in Section VI-A. With this lemma, we can now prove Theorem 5.

Proof: (of Theorem 5)

Assume for the sake of contradiction that $m \ge k + s - c\log(1/\varepsilon)$ for some small enough constant c. We will show that in this case Ext fails on some bit-fixing source of min-entropy k. As in [25], the existence of such a source will be shown

[1]Note that this lemma does not appear explicitly in [25] but corresponds to Claim 2.7 there, which is stated in a graph-theoretical term and says that any extractor gives rise to some kind of "slice-extractor".

[2]This justifies the condition $1/(8\binom{n}{t}2^t) < \delta$ assumed at the beginning of the lemma.

using a probabilistic argument. The difference is that [25] had the luxury of having all possible sources of min-entropy k to search through, while we are limited to the much smaller class of bit-fixing sources, which makes our task much harder. We randomly generate such a bit-fixing source in the following way.

- Randomly pick a set $I \in P([n], n-k)$ and a string $u \in \{0,1\}^{n-k}$. Generate the source $\mathcal{X}_I^u$ which is uniform over the set $X_I^u = \{x \in \{0,1\}^n : x_I = u\}$.

Next, we will show that EXT fails with a positive probability over such a randomly generated source $\mathcal{X}_I^u$. As in [25], the idea is to show that when m is large, most z's in $\{0,1\}^m$ can only have a small set $S^{(z)}$, and such z's are $(2^{-(n-k)}\sqrt{\varepsilon})$-missed by X_I^u with a nonnegligible probability. As we will show next, this probability is guaranteed by Lemma 13, by observing that the condition that z is $(2^{-(n-k)}\sqrt{\varepsilon})$-missed by X_I^u is exactly the condition that u is $(I, 2^{-(n-k)}\sqrt{\varepsilon})$-biased in $S^{(z)}$, because

$$\left| \Pr_{x \in S^{(z)}} [x \in X_I^u] - \Pr_{x \in \mathcal{U}_n} [x \in X_I^u] \right| = \left| \Pr_{x \in S^{(z)}} [x_I = u] - 2^{-(n-k)} \right|.$$

Let $t = n - k$ and $\delta = \sqrt{\varepsilon}$, and note that the conditions on the parameters in the theorem imply those in Lemma 13 (in particular, the condition $k \le n - \log(1/\varepsilon)$ implies the condition $\delta \ge 1/(8\binom{n}{t}2^t)$). Now the average of $|S^{(z)}|$ over z is

$$2^{n+s}/2^m \le 2^{n+s}/2^{k+s-c\log(1/\varepsilon)} = 2^t(1/\delta)^{2c}.$$

Call z *small* if $|S^{(z)}| < 2^t(1/\delta)^{c'}$ for a small enough constant c'. By Markov inequality, at least $1/2$ fraction of z's are small. From Lemma 13, for any small z, with $|S^{(z)}| < 2^t(1/\delta)^{c'}$, the probability over $I \in P([n], t)$ and $u \in \{0,1\}^t$ that z is $(2^{-t}\delta)$-missed by X_I^u is more than 8δ. By an average argument, there must exist $I \in P([n], t)$ and $u \in \{0,1\}^t$ such that more than 8δ fraction of small z's are $(2^{-t}\sqrt{\varepsilon})$-missed by X_I^u. Thus, for this I and u, more than

$$(1/2)8\delta = 4\sqrt{\varepsilon}$$

fractions of all possible $z \in \{0,1\}^m$ are $(2^{-t}\sqrt{\varepsilon})$-missed by X_I^u. From Lemma 12, this implies that $\|\mathrm{EXT}(\mathcal{X}_I^u) - \mathcal{U}_m\|_1 > \varepsilon$, a contradiction. Therefore, one must have $m \le k + s - \Omega(\log(1/\varepsilon))$, which proves the theorem. ∎

A. Proof of Lemma 13

Consider any set S satisfying the property stated in the lemma. Our goal is to show a lower bound on the size of such a set. We can assume without loss of generality that $|S| < 2^t/(2\delta)$, because otherwise we are done. From [2], [7], we know that for an even r, any r-wise independent space over $\{0,1\}^n$ must have a size at least $\binom{n}{r/2}$, and we would like to apply it to get our bound. However, there are two difficulties in front of us. One is that S only guarantees some randomness property on most, instead of all, collections of t dimensions. The other is that the randomness property only guarantees being close to random instead of perfectly random. We get around these by showing that for some appropriate $r < t$ to be chosen later, there exists some set $J \in P([n], t-r)$ such that when we partition S into subsets

$$S_{J,v} = \{x \in S : x_J = v\},$$

for $v \in \{0,1\}^{t-r}$, many of these subsets will embed an r-wise independent space.

From the property of S, an average argument shows the existence of some $J \in P([n], t-r)$ such that over a random $R \in P([n] \setminus J, r)$ and a random $u \in \{0,1\}^t$, u is $(J \cup R, 2^{-t}\delta)$-biased in S with probability at most 8δ. Fix one such set J, and let $\bar{J} = [n] \setminus J$. Call $v \in \{0,1\}^{t-r}$ *nice* for $R \in P(\bar{J}, r)$ if for every $w \in \{0,1\}^r$, (v, w) is not $(J \cup R, 2^{-t}\delta)$-biased in S. The following shows that most v are nice for most R.

Claim 3: At least $1 - 2^{r+3}\sqrt{\delta}$ fraction of $v \in \{0,1\}^{t-r}$ are nice for all but $\alpha = 2^r\sqrt{\delta}$ fraction of $R \in P(\bar{J}, r)$.

Proof: By a Markov inequality, there are at most $8\sqrt{\delta}$ fraction of $u = (v, w) \in \{0,1\}^t$ which are $(J \cup R, 2^{-t}\delta)$-biased in S for at least $\sqrt{\delta}$ fraction of $R \in P(\bar{J}, r)$. Thus, at most $2^r 8\sqrt{\delta}$ fraction of $v \in \{0,1\}^{t-r}$ can have some bad $w \in \{0,1\}^r$ (depending on v) which is bad for at least $\sqrt{\delta}$ fraction of $R \in P(\bar{J}, r)$ in the sense that (v, w) is $(J \cup R, 2^{-t}\delta)$-biased in S. As a result, at least $1 - 2^r 8\sqrt{\delta}$ fraction of $v \in \{0,1\}^{t-r}$ do not have such a bad w, and each such v is nice for all but $2^r\sqrt{\delta}$ fraction of $R \in P(\bar{J}, r)$, as any w now is bad for at most $\sqrt{\delta}$ fraction of R. ∎

Fix any $v \in \{0,1\}^{t-r}$ which is nice for all but $\alpha = 2^r\sqrt{\delta}$ fraction of $R \in P(\bar{J}, r)$. Next, we will show that $S_{J,v}$ embeds an r-wise independent space. For this, we need the following lemma which shows that if v is nice for $R \in P(\bar{J}, r)$, the space $S_{J,v}$ projected to dimensions in R gives a uniform distribution.

Claim 4: Suppose v is nice for $R \in P(\bar{J}, r)$. Then for any $w \in \{0,1\}^r$, $\Pr_{x \in S_{J,v}}[x_R = w] = 2^{-r}$.

Proof: Suppose v is nice for R, so for every $w \in \{0,1\}^r$

$$\left| \Pr_{x \in S} [(x_J, x_R) = (v, w)] - 2^{-t} \right| \le 2^{-t}\delta.$$

As we assume that $|S| < 2^t/(2\delta)$, this means that all the 2^r probabilities $\Pr_{x \in S}[(x_J, x_R) = (v, w)]$, for $w \in \{0,1\}^r$, have a distance less than $1/(2|S|)$ to the value 2^{-t}, so any two of the probabilities can only have a distance less than $1/|S|$ from each other. This implies that all these 2^r probabilities must all be equal, because they are all multiples of $1/|S|$. Then note that

$$\Pr_{x \in S_{J,v}} [x_R = w] = \Pr_{x \in S} [(x_J, x_R) = (v, w)] \,/\, \Pr_{x \in S} [x_J = v]$$

which is the same for every $w \in \{0,1\}^r$. As a result, all these 2^r probabilities $\Pr_{x \in S_{J,v}}[x_R = w]$, for $w \in \{0,1\}^r$, must all equal 2^{-r}. ∎

Then we consider the following two cases according to the range of δ. In each case, we will choose a proper r and show that $|S_{J,v}| \ge 2^r(1/\delta)^{\Omega(1)}$. Let $k = n - t$, so $|\bar{J}| = k + r$.

Case 1: $\delta < 1/(4(k+2)^4)$. In this case, we choose r to be that guaranteed in the following claim.

Claim 5: There exists an even integer r such that $2 \le r \le \min\{t, k/24\}$ and $(1/\delta)^{\Omega(1)} \le \binom{k+r}{r}2^r < \sqrt{1/\delta}$.

Proof: Note that the value of $\binom{k+r}{r}2^r$ increases smoothly as we increase t from 2 to $\min\{t, k/24\}$. For $r = 2$, the value is

$$\binom{k+2}{2}2^2 < 2(k+2)^2 < \sqrt{1/\delta}.$$

On the other hand, for $r = k/24$, we have

$$\binom{k+r}{r}2^r \geq 2^{\Omega(k)} \geq (1/\delta)^{\Omega(1)}$$

according to the assumption that $k \geq \Omega(\log(1/\delta))$, while for $r = t$, we also have

$$\binom{k+t}{t}2^t = \binom{n}{t}2^t \geq 8/\delta \geq (1/\delta)^{\Omega(1)}$$

according to the assumption that $\delta \geq 1/(8\binom{n}{t}2^t)$. Thus, when increasing r from 2 to $\min\{t, k/24\}$, we will encounter an even integer r such that $(1/\delta)^{\Omega(1)} \leq \binom{k+r}{r}2^r < \sqrt{1/\delta}$. ∎

With this choice of r, we have

$$|P(\bar{J}, r)| \cdot \alpha = \binom{k+r}{r} \cdot 2^r\sqrt{\delta} < 1$$

which implies that v is nice for every $R \in P(\bar{J}, r)$. By Claim 4, this means that the set $S_{J,v}$ projected to dimensions in $\bar{J}$ forms an r-wise independent space. From [2] and [7], such a set must have size at least

$$\begin{aligned}\binom{|\bar{J}|}{r/2} &\geq \left(\frac{2(k+r)}{r}\right)^{r/2} \\ &= 2^r\left(\frac{k+r}{2r}\right)^{r/2} \\ &= 2^r\left(\frac{k+r}{4r}\cdot\frac{k+r}{r}\right)^{r/4} \\ &\geq 2^r\left(6\cdot\frac{k+r}{r}\right)^{r/4}\end{aligned}$$

where the last inequality follows from the condition $r \leq k/24$. As a result, we have

$$\begin{aligned}|S_{J,v}| &\geq 2^r\left(\left(3\cdot\frac{k+r}{r}\right)^r 2^r\right)^{1/4} \\ &\geq 2^r\left(\binom{k+r}{r}2^r\right)^{1/4} \\ &\geq 2^r\,(1/\delta)^{\Omega(1)}.\end{aligned}$$

Case 2: $\delta \geq 1/(4(k+2)^4)$. In this case, we choose $r = 2$, and now $\alpha = 4\sqrt{\delta}$. Then the following claim, together with Claim 4, implies that the set $S_{J,v}$ projected to dimensions in A gives a pair-wise independent space, so by [2] and [7], we have

$$|S_{J,v}| \geq |A| \geq (1/\delta)^{\Omega(1)} = 2^r(1/\delta)^{\Omega(1)}.$$

Claim 6: There exists a subset $A \subseteq \bar{J}$ of size $(1/\delta)^{\Omega(1)}$ such that v is nice for every $R \in P(A, 2)$.

Proof: Consider the undirected graph G with vertex set $V = \bar{J}$ and edge set $E = \{R \in P(\bar{J}, 2) : v \text{ is nice for } R\}$. Note that $|E|$ is at least

$$\begin{aligned}(1-\alpha)\binom{|V|}{2} &= (1-\alpha)\left(1-\frac{1}{|V|}\right)\frac{|V|^2}{2} \\ &> \left(1-\alpha-\frac{1}{|V|}\right)\frac{|V|^2}{2} \\ &\geq \left(1-\delta^{\Omega(1)}\right)\frac{|V|^2}{2}.\end{aligned}$$

Then by the well-known Turan's theorem in graph theory (e.g., see [13, Theorem 4.7]), G must contain a clique A of size at least $(1/\delta)^{\Omega(1)}$. By the definition of E, v is nice for every $R \in P(A, 2)$, which proves the claim. ∎

In both cases, we have shown that $|S_{J,v}| \geq 2^r(1/\delta)^{\Omega(1)}$, for any v which is nice for all but α fraction of $R \in P(\bar{J}, r)$. Since the number of such v's is at least

$$\left(1 - 2^{r+3}\sqrt{\delta}\right)2^{t-r} \geq (1/2)2^{t-r},$$

and the corresponding sets $S_{J,v}$'s are all disjoint subsets of S, we conclude that

$$|S| \geq (1/2)2^{t-r}2^r(1/\delta)^{\Omega(1)} = 2^t(1/\delta)^{\Omega(1)}.$$

This proves Lemma 13.

APPENDIX

A. Proof of Lemma 3

Similar to [9], we will use the following lemma to prove Lemma 3. It is very similar to a lemma in [9] showing the existence of an analogous sampler with respect to their definition, and our proof is based on theirs.

Lemma 14: For any $n, k, r, t \in \mathbb{N}$ such that $r \leq k \leq n$ and $6\log n \leq t \leq (k\log n)/(20r)$, there is an explicit $(n, d, k, kd/(2r), 3kd/r, 2^{-\Omega(t/\log n)})$-sampler which uses a seed of t random bits.

Proof: It is shown in [9] that for any t with $6\log n \leq t \leq (k\log n)/(20r)$, there exists an explicit function $S : \{0,1\}^t \to P([n])$ such that for any Boolean function $h' : [n] \to \{0,1\}$,

$$\Pr_{w\in\mathcal{U}_t}\left[k/(2r) \leq \sum_{i\in S(w)} h'(i) \leq 3k/r\right] \geq 1-\delta',$$

for $\delta' = 2^{-\Omega(t/\log n)}$. A closer inspection of their proof shows that it actually works for any real-valued function $h' : [n] \to [0,1]$. Now given any real-valued function $h : [n] \to [0, d]$, consider the function $h' : [n] \to [0,1]$ defined as $h'(i) = h(i)/d$ for $i \in [n]$, and note that

$$\begin{aligned}&\Pr_{w\in\mathcal{U}_t}\left[kd/(2r) \leq \sum_{i\in S(w)} h(i) \leq 3kd/r\right] \\ &\qquad = \Pr_{w\in\mathcal{U}_t}\left[k/(2r) \leq \sum_{i\in S(w)} h'(i) \leq 3k/r\right].\end{aligned}$$

Thus, S is also an $(n, d, k, kd/(2r), 3kd/r, \delta')$-sampler, which proves Lemma 14. ∎

Now we proceed to prove Lemma 3. Suppose $\delta \geq 2^{-c_1 k}$ for a small enough constant c_1, and $k_{\min} \geq c_2 d \log(1/\delta)$ for a large enough constant c_2. Let us choose $r = kd/(2k_{\min}) \leq k$, so that $kd/(2r) = k_{\min}$ and

$$\begin{aligned}(k \log n)/(20r) &\geq (k_{\min} \log n)/(10d) \\ &\geq (c_2/10) \cdot \log n \cdot \log(1/\delta).\end{aligned}$$

Thus, we can choose $t = (c_2/10) \cdot \log n \cdot \log(1/\delta)$ and have $6 \log n \leq t \leq (k \log n)/(20r)$. From Lemma 14, we have an $(n, d, k, k_{\min}, 6k_{\min}, \delta')$-sampler, with $\delta' = 2^{-\Omega(t/\log n)} \leq 2^{-\Omega(c_2 \log(1/\delta))} \leq \delta$. This completes the proof of Lemma 3.

B. Proof of Lemma 5

The proof is very similar to that for an analogous lemma in [9], which can be seen as a derandomization of Lemma 3, using approximate pair-wise independent variables.

Definition 4: [20] We say that the random variables $Z_1, \ldots, Z_n$ are pair-wise ϵ-dependent if the joint distribution of any two of them is ϵ-random.

Lemma 15: [1] Let $r' < n$ be a power of 2. For any $n \geq 16$ and $0 < \epsilon < 1/2$, one can use $7 \log r' + 3(\log \log n + \log(1/\epsilon))$ random bits to generate n random variables $Z_1, \ldots, Z_n \in [r']$ that are pair-wise ϵ-dependent.

Let r' be a power of 2 such that $r/2 < r' \leq r$. For any given constant $\alpha \in (0,1)$, let $\beta = \alpha/38$, $r = k^{\beta}$, and $\epsilon = k^{-4\beta}$. We use pair-wise ϵ-dependent random variables $Z_1, \ldots, Z_n \in [r']$ to partition the set $[n]$ into r' sets: $T_1, \ldots, T_{r'}$ where $T_v = \{i | Z_i = v\}$ for $v \in [r']$. By Lemma 15, the number of random bits needed to generate them is at most

$$\begin{aligned}&7 \log r' + 3(\log \log n + \log(1/\epsilon)) \\ &\leq 19\beta \log k + 3 \log \log n \\ &\leq \alpha/2 \log k + 3 \log \log n.\end{aligned}$$

Let $c_1 = 6/\alpha$. Then we have for all $k \geq \log^{c_1} n$, $3 \log \log n \leq \alpha/2 \log k$. This shows that one can generate such random variables $Z_1, \ldots, Z_n$ using $\alpha \log k$ random bits.

Now consider any function $h : [n] \to [0, d]$ satisfying $\sum_{i=1}^{n} h(i) = k$. For now, let us fix an $v \in [r']$, and define n random variables $B_1, \ldots, B_n$ such that for $i \in [n]$, $B_i = h(i)$ if $i \in T_v$ and $B_i = 0$ otherwise. Let $B = \sum_{i=1}^{n} B_i = \sum_{i \in T_v} h(i)$, and we would like to bound the probability $\Pr[|B - k/r'| > k/(2r')]$. Since the expected value of B is close to k/r', with

$$\begin{aligned}|\operatorname{E}[B] - k/r'| &= \left| \sum_{i=1}^{n} h(i) \cdot \Pr[Z_i = v] - k/r' \right| \\ &\leq \sum_{i=1}^{n} h(i) \cdot |\Pr[Z_i = v] - 1/r'| \\ &\leq k\epsilon\end{aligned}$$

we have $\Pr[|B - k/r'| > k/(2r')] \leq \Pr[|B - \operatorname{E}[B]| > k/(2r') - k\epsilon]$. Since $k\epsilon \leq k/(6r')$ for some large enough n and $k \geq \log^{c_1} n$, and thus $k/(2r') - k\epsilon \geq k/(3r')$, it suffices to bound the probability $\Pr[|B - \operatorname{E}[B]| > k/(3r')]$.

We would like to apply Chebyshev inequality, so we need to bound the variance of B, which is $Var(B) = \sum_{i=1}^{n} Var(B_i) + \sum_{i \neq j} cov(B_i, B_j)$. For any $i \in [n]$

$$\begin{aligned}Var(B_i) &= \operatorname{E}[B_i^2] - \operatorname{E}[B_i]^2 \\ &\leq \operatorname{E}[B_i^2] \\ &= h(i)^2 \cdot \Pr[Z_i = v] \\ &\leq h(i)^2 \cdot (1/r' + \epsilon).\end{aligned}$$

For any distinct $i, j \in [n]$

$$\begin{aligned}cov(B_i, B_j) &= \operatorname{E}[B_i \cdot B_j] - \operatorname{E}[B_i] \cdot \operatorname{E}[B_j] \\ &= h(i)h(j) \cdot \Pr[Z_i = Z_j = v] \\ &\quad - h(i)h(j) \cdot \Pr[Z_i = v] \cdot \Pr[Z_j = v] \\ &\leq h(i)h(j) \cdot \left((1/r'^2 + \epsilon) - (1/r' - \epsilon)^2 \right) \\ &= h(i)h(j) \cdot (1 + 2/r' - \epsilon)\,\epsilon \\ &\leq h(i)h(j) \cdot 2\epsilon\end{aligned}$$

as $r' \geq 2$. Therefore

$$\begin{aligned}Var(B) &\leq \sum_i h(i)^2 (1/r' + \epsilon) + \sum_{i \neq j} h(i)h(j) 2\epsilon \\ &\leq (1/r') \sum_i h(i)^2 + 2\epsilon \left(\sum_i h(i) \right)^2 \\ &\leq dk/r' + 2k^2\epsilon\end{aligned}$$

where the last inequality follows from the fact that $h(i) \leq d$ for every $i \in [n]$ and $\sum_i h(i) = k$.

Now by Chebyshev inequality, we have

$$\begin{aligned}\Pr[|B - \operatorname{E}[B]| > k/(3r')] &< \frac{dk/r' + 2\epsilon k^2}{(k/3r')2} \\ &= 9dr'/k + 18\epsilon r'^2 \\ &= O(\epsilon r'^2)\end{aligned}$$

for some small enough constant c_0 and $d \leq k^{c_0}$. Thus, setting $\tau = 1 - \beta \geq 1/2$, we have for any $v \in [r']$

$$\Pr\left[k^{\tau}/2 \leq \sum_{i \in T_v} h(i) \leq 3k^{\tau} \right] \geq 1 - O(k^{-2\beta}).$$

Then, Lemma 5 follows from the union bound.

References

[1] N. Alon, O. Goldreich, J. Håstad, and R. Peralta, "Simple constructions of almost k-wise independent random variables," in *Proc. IEEE 31st Annu. IEEE Symp. Foundations of Computer Science (FOCS'90)*, 1990, pp. 544–553.

[2] N. Alon, L. Babai, and A. Itai, "A fast and simple randomized parallel algorithm for the maximal independent set problem," *J. Algorith.*, vol. 7, no. 4, pp. 567–583, 1986.

[3] B. Barak, R. Impagliazzo, and A. Wigderson, "Extracting randomness using few independent sources," in *Proc. IEEE 45th Annu. IEEE Symp. Foundations of Computer Science (FOCS'04)*, 2004, pp. 384–393.

[4] B. Barak, G. Kindler, R. Shaltiel, B. Sudakov, and A. Wigderson, "Simulating independence: New constructions of condensers, Ramsey graphs, dispersers, and extractors," in *Proc. 37th Annu. ACM Symp. Theory of Computing (STOC'05)*, 2005, pp. 1–10.

[5] J. Bourgain, "More on the sum-product phenomenon in prime fields and its applications," *Int. J. Numb. Theory*, vol. 1, no. 1, pp. 1–32, 2005.
[6] B. Chor and O. Goldreich, "Unbiased bits from sources of weak randomness and probabilistic communication complexity," *SIAM J. Comput.*, vol. 17, no. 2, pp. 230–261, Apr. 1988.
[7] B. Chor, O. Goldreich, J. Håstad, J. Friedman, S. Rudich, and R. Smolensky, "The bit extraction problem of t-resilient functions," in *Proc. 26th Annu. IEEE Symp. Foundations of Computer Science*, 1985, pp. 396–407.
[8] Y. Dodis, A. Elbaz, R. Oliveira, and R. Raz, "Improved randomness extraction from two independent sources," in *Proc. 8th Int. Workshop on Randomization and Computation (RANDOM'04)*, 2004, pp. 334–344.
[9] A. Gabizon, R. Raz, and R. Shaltiel, "Deterministic extractors for bit-fixing sources by obtaining an independent seed," *SIAM J. Comput.*, vol. 36, no. 4, pp. 1072–1094, 2006.
[10] O. Goldreich, *Foundations of Cryptography: Volume 1, Basic Tools*. Cambridge, U.K.: Cambridge University Press, 2001.
[11] R. Impagliazzo, R. Shaltiel, and A. Wigderson, "Extractors and pseudo-random generators with optimal seed length," in *Proc. 32nd Annu. ACM Symp. the Theory of Computing*, 2000, pp. 1–10.
[12] S. Jukna, *Extremal Combinatorics*. New York: Springer-Verlag, 2001.
[13] J. Kamp, A. Rao, S. P. Vadhan, and D. Zuckerman, "Deterministic extractors for small-space sources," in *Proc. 38rd Annu. ACM Symp. Theory of Computing*, 2006, pp. 691–700.
[14] J. Kamp and D. Zuckerman, "Deterministic extractors for bit-fixing sources and exposure-resilient cryptography," *SIAM J. Comput.*, vol. 36, no. 5, pp. 1231–1247, 2007.
[15] R. König and U. Maurer, "Generalized strong extractors and deterministic privacy amplification," in *Proc. Cryptography and Coding*, 2005, pp. 322–339.
[16] C.-J. Lee, C.-J. Lu, S.-C. Tsai, and W.-G. Tzeng, "Extracting randomness from multiple independent sources," *IEEE Trans. Inform. Theory*, vol. 51, no. 6, pp. 2224–2227, Jun. 2005.
[17] C.-J. Lee, C.-J. Lu, and S.-C. Tsai, "Deterministic extractors for independent-symbol sources," in *Proc. 33rd Int. Colloq. Automata, Languages and Programming (ICALP 2006)*, Venice, Italy, Jul. 9–16, 2006.
[18] C.-J. Lu, "Encryption against storage-bounded adversaries from on-line strong extractors," *J. Cryptol.*, vol. 17, no. 1, pp. 27–42, 2004.
[19] C.-J. Lu, O. Reingold, S. Vadhan, and A. Wigderson, "Extractors: Optimal up to constant factors," in *Proc. 35th Annu. ACM Symp. Theory of Computing (STOC'03)*, 2003, pp. 602–611.
[20] J. Naor and M. Naor, "Small-bias probability spaces: Efficient constructions and applications," *SIAM J. Comput.*, vol. 22, no. 4, pp. 838–856, 1993.
[21] N. Nisan and A. Ta-Shma, "Extracting randomness: A survey and new constructions," *J. Comput. Syst. Sci.*, vol. 58, no. 1, pp. 148–173, 1999.
[22] N. Nisan and D. Zuckerman, "Randomness is linear in space," *J. Comput. and Syst. Sci.*, vol. 52, no. 1, pp. 43–52, 1996.
[23] R. Raz, "Extractors with weak random seeds," in *Proc. 37th Annu. ACM Symp. Theory of Computing (STOC'05)*, 2005, pp. 11–20.
[24] R. Raz, O. Reingold, and S. P. Vadhan, "Extracting all the randomness and reducing the error in Trevisan's extractors," in *Proc. 31st Annu. ACM Symp. Theory of Computing (STOC'99)*, 1999, pp. 149–158.
[25] J. Radhakrishnan and A. Ta-Shma, "Bounds for dispersers, extractors, and depth-two superconcentrators," *SIAM J. Discrete Math.*, vol. 13, no. 1, pp. 2–24, 2000.
[26] O. Reingold, R. Shaltiel, and A. Wigderson, "Extracting randomness via repeated condensing," in *Proc. 41st Annu. IEEE Symp. Foundations of Computer Science*, Redondo Beach, CA, Nov. 12–14, 2000.
[27] R. Shaltiel, "Recent developments in explicit constructions of extractors," *Bull. Eur. Assoc. Theoret. Comput. Sci.*, vol. 77, pp. 67–95, 2002.
[28] R. Shaltiel and C. Umans, "Simple extractors for all min-entropies and a new pseudo-random generator," in *Proc. 42nd Annu. IEEE Symp. Foundations of Computer Science*, 2001, pp. 648–657.
[29] M. Sipser, "Expanders, randomness, or time versus space," *J. Comput. Syst. Sci.*, vol. 36, no. 3, pp. 379–383, 1988.
[30] A. Ta-Shma, C. Umans, and D. Zuckerman, "Loss-less condensers, unbalanced expanders, and extractors," in *Proc. 33rd Annu. ACM Symp. Theory of Computing*, 2001, pp. 143–152.
[31] A. Ta-Shma and D. Zuckerman, "Extractor codes," in *Proc. 33rd Annu. ACM Symp. Theory of Computing (STOC'01)*, 2001, pp. 193–199.
[32] L. Trevisan, "Extractors and pseudorandom generators," *J. ACM*, vol. 48, no. 4, pp. 860–879, 2001.
[33] S. P. Vadhan, "Constructing locally computable extractors and cryptosystems in the bounded-storage model," *J. Cryptol.*, vol. 17, no. 1, pp. 43–77, 2004.
[34] A. Wigderson and D. Zuckerman, "Expanders that beat the eigenvalue bound: Explicit construction and applications," *Combinatorica*, vol. 19, no. 1, pp. 125–138, 1999.
[35] D. Zuckerman, "General weak random sources," in *Proc. IEEE 31st Annu. IEEE Symp. Foundations of Computer Science (FOCS'90)*, 1990, pp. 534–543.
[36] D. Zuckerman, "Simulating BPP using a general weak random source," *Algorithmica*, vol. 16, no. 4/5, pp. 367–391, 1996.
[37] D. Zuckerman, "Randomness-optimal oblivious sampling," *Random Struct. and Algorith.*, vol. 11, pp. 345–367, 1997.

Chia-Jung Lee received the B.S. degree from National Taiwan Normal University, Taipei, Taiwan, R.O.C., in 2000 and the Ph.D. degree in computer science from National Chiao-Tung University, Hsinchu, Taiwan, in 2010.

She is currently a Postdoctoral Researcher at the Institute of Information Science, Academia Sinica, Taipei. Her research interests include randomness in computation, cryptography, and theoretical computer science.

Chi-Jen Lu received the B.S. and M.S. degrees from National Taiwan University, Taipei, Taiwan, in 1988 and 1990, respectively, and the Ph.D. degree from the University of Massachusetts at Amherst 1999, all in computer science.

He is currently a Research Fellow with the Institute of Information Science, Academia Sinica, Taiwan. His research interests include randomness in computation, computational complexity, cryptography, game theory, and machine learning.

Shi-Chun Tsai received the B.S. and M.S. degrees in computer science and information engineering from National Taiwan University, Taipei, Taiwan, in 1984 and 1988, respectively, and the Ph.D. degree in computer science from the University of Chicago, Chicago, IL, in 1996.

During 1993–1996, he served as a Lecturer in the Computer Science Department, University of Chicago. During 1996–2001, he was an Associate Professor in the Information Management Department and the Computer Science and Information Engineering Department, National Chi Nan University, Taiwan. He has been with the Department of Computer Science, National Chiao-Tung University, Taiwan, since 2001, where he was promoted to Full Professor in 2007. He has served as the Deputy Director of the Information Technology Service Center since 2009. His research interests include computational complexity, algorithms, coding theory and combinatorics.