國立交通大學

資訊工程學系

博士論文

非種子萃取器之設計與分析

Seedless Extractors: Constructions and Analysis

研究生：李佳蓉

指導教授：蔡錫鈞 教授

中華民國九十九年五月

非種子萃取器之設計與分析

Seedless Extractors: Constructions and Analysis

研 究 生：李佳蓉　Student：Chia-Jung Lee

指導教授：蔡錫鈞　Advisor：Shi-Chun Tsai

國立交通大學資訊學院
資訊工程學系
博士論文

A dissertation is submitted to
Department of Computer Science
National Chiao Tung University
in partial Fulfillment of the Requirements
for the Degree of
Doctor of Philosophy
in
Computer Science
May 2010
Hsinchu, Taiwan, Republic of China

中華民國九十九年五月

摘　要

我們討論從各種隨機源中萃取出隨機元的問題。首先，我們考慮多個獨立的隨機源。當有兩個獨立的隨機源時，我們利用延伸的剩餘雜湊引理建造出一個萃取器。接著我們進而延伸此建造方法從更多個獨立的隨機源中萃取出隨機元。值得一提的是，即使除了一個隨機源外，所有的隨機源都被暴露，我們的萃取器依然可以萃取出隨機元。此外，我們可以將此萃取器應用到密碼學上，解決有一群人希望能透過一個不安全的管道且不使用當地的理想隨機元而能決定一個將用在群體通訊之秘密金鑰的問題。

我們也考慮獨立符號隨機源。一個獨立符號隨機源包含 n 個獨立的符號，其中每個符號都是屬於$\{0,1\}^d$，且獨立符號隨機源僅保證整個隨機源的 min-entropy 為 k 。我們提出一個對任何$n,d,k\in\mathbb{N}$，可萃取出大概$\Omega(\log k)$個隨機元的決定性萃取器。接著我們證明當$k\ge\log^c n$(對某個常數 c>0)時，我們幾乎可以萃取出所有包含於來源中的隨機元。更進一步地，我們證明只要滿足 min-entropy $k=\omega(d+\log(n/\varepsilon))$，則存在一個可萃取出 $m=k-\mathrm{O}(\log(1/\varepsilon))$ 個隨機元的決定性萃取器，其中ε為誤差。最後，我們亦證明任何針對固定某些位元的萃取器，不論其需要種子與否，其 min-entropy 損失為$k-m=\Omega(\log(1/\varepsilon))$，此為 Radhakrishnan 與 Ta-Shma 針對一般弱隨機源之結果的一個延伸。

然後，我們由另一角度出發去尋找一種更廣義的隨機源，其亦存在決定性萃取器。我們在這裡所考慮的是一種條件的來源(f(X)|X)，其中 X 是一個在$\{0,1\}^{n_1}$上的分布，而$f:\{0,1\}^{n_1}\to\{0,1\}^n$為某個函數。假設當輸入 x 是依照分布 X 所產生時，任何大小為2^k的電路最多只有2^{-k}的機率可以猜對 f(x)的值時，則我們說這種條件分布(f(X)|X)擁有計算的 min-entropy k。我們首先證明無法從單一個只擁有計算的 min-entropy 之來源中萃取出一個隨機元。接著我們證明可從一個只擁有計算的 min-entropy 與另一個擁有統計的 min-entropy 之隨機源中萃取出隨機元。更進一步地，我們亦證明可從兩個只擁有計算的 min-entropy 之隨機源中萃取出隨機元。這可看成是把原先在統計的環境中針對多隨機源萃取器的研究延伸到計算的環境。我們把建造此種萃取器的工作轉成是一個在學習理論上的問題：利用任何一個分布，在對手雜訊(adversarial noise)的模型下學習線性函數。針對此問題，我們亦提出一個學習演算法。

最後，我們考慮計算的獨立符號隨機源。如同獨立符號隨機源，計算的獨立符號隨機源亦包含 n 個獨立的符號 $(f_1(X_1)|X_1)$，…，$(f_n(X_n)|X_n)$，其中每個 $f_i(X_i)$ 都是分布在$\{0,1\}^d$，使得當輸入 x_i 是依照分布 X_i 所產生時，任何一個大小為 s 的電路最多只有 2^{-k_i}的機率可以猜對 $f_i(X_i)$ 對某個數 $k_i\le d$，且 $k_1+\cdots+k_n=k$。我們延伸核心集引理來證明我們針對獨立符號隨機源的萃取器亦可作用於計算的獨立符號隨機

源。事實上，此針對計算的獨立符號隨機源的萃取器之結果隱含延伸的 XOR 引理。我們亦提供一個在黑箱子建造法中，二元的核心集大小的上限。

Abstract

In this thesis, we consider the problem of extracting randomness from several classes of random sources. First, we consider multiple independent sources. With two independent sources, we have an explicit extractor, via generalized leftover hash lemma. We also extend our construction to extract randomness from more independent sources. One nice feature is that the extractor still works even with all but one source exposed. Moreover, we apply our extractor for a cryptographic task in which a group of parties want to agree on a secret key for group communication over an insecure channel, without using ideal local randomness.

We also consider the independent-symbol sources which consist of a sequence of n independent symbols from $\{0,1\}^d$, and the only randomness guarantee on such a source is that the whole source has min-entropy k. We give an explicit deterministic extractor which extracts about $\Omega(\log k)$ bits, for any $n, d, k \in \mathbb{N}$. When $k \geq \log^c n$, we can extract almost all randomness. Moreover, we show the existence of a non-explicit 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.

Then, we go to the other direction to look for a more general class of sources from which seedless extraction is still possible. The sources we consider have the form of a conditional distribution $(f(\mathcal{X})|\mathcal{X})$, for some function f and some distribution $\mathcal{X}$, and we say that such a source has computational min-entropy k if any circuit of size 2^k can only predict $f(x)$ correctly with probability at most 2^{-k} given input x sampled from $\mathcal{X}$. We first show that it is impossible to have a seedless extractor to extract from one single source of this kind. Then we show that it becomes possible if we are allowed a seed which is weakly random (instead of perfectly random) but contains some statistical min-entropy, or even a seed which is not random at all but contains some computational min-entropy. This can be seen as a step toward extending the study of multi-source extractors from the traditional, statistical setting to a computational

setting. We reduce the task of constructing such extractors to a problem in learning theory: learning linear functions under arbitrary distribution with adversarial noise. For this problem, we provide a learning algorithm, which may have interest of its own.

Finally, we consider computational independent-symbol sources, which consist of n mutually independent parts, $(f_1(\mathcal{X}_1)|\mathcal{X}_1), \cdots, (f_n(\mathcal{X}_n)|\mathcal{X}_n)$, each $f_i(\mathcal{X}_i)$ of length d such that for each i if given input x_i sampled from $\mathcal{X}_i$, any circuit of size s can only predict $f_i(x_i)$ with probability at most 2^{-k_i} for some $k_i \leq d$, and the sum of k_i's is k. We generalize the well-known hardcore set lemma to show that our extractor for independent-symbol sources still works for computational independent-symbol sources. In fact, the result of computational extractors for computational independent-symbol sources implies a generalization of the well-known XOR lemma. Besides, we provide a size upper bound on a binary hardcore set in any black-box construction.

Acknowledgements

I would like to thank my advisor, Dr. Shi-Chun Tsai, for his guidance and encouragement. I also thank Dr. Chi-Jen Lu for many useful discussions. I also want to thank my family and all members in CCIS lab for their support.

Seedless Extractors: Constructions and Analysis

Chia-Jung Lee

Contents

List of Figures

Chapter 1

Introduction

Randomness has become a useful tool in computer science. For many computational problems, the most efficient algorithms known are randomized. 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. From weakly random sources, we would like to extract almost perfect randomness, which can then be used for randomized algorithms.

The history of this extraction can be traced back to von Neumann [57] who showed how to use a biased coin (with unknown bias) to simulate a fair coin: Flip the biased coin twice; if the result is tail-head (respectively head-tail), the simulated coin outputs a head (respectively tail), otherwise repeat the process. We call the functions, which can extract almost perfect randomness from weakly random sources, *extractors* [61, 42]. Extractors turn out to have close connections to other fundamental objects such as **(a) pseudorandom generators, (b) hash functions, (c) error-correcting codes, (d) expander graphs, and (e) samplers,** and they have found a wide range of applications in areas such as **(a) complexity theory, (b) cryptography, (c) data structures, (d) coding theory, (e) distributed computing, and (f) combinatorics (e.g. [50, 42, 62, 63, 58, 54, 53, 36, 56, 15]).** A nice survey can be found in [48].

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 easy to see that one cannot deterministically extract even one bit from an n-bit source with min-entropy $n-1$ [10]. Assume that $\textsc{Ext} : \{0,1\}^n \to \{0,1\}$ is any such extractor, let S be the bigger set of the preimages $\textsc{Ext}^{-1}(0)$ and $\textsc{Ext}^{-1}(1)$. Clearly, the uniform distribution over S is a weak source with min-entropy at least $n-1$. However, since $\textsc{Ext}$ takes the same value for all $x \in S$, it must fail to extract from this weak source.

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. [42, 41, 46, 29, 47, 54, 52, 49]), and finally an optimal (up to constant factors) construction has been given recently [37].

However, there is an issue of using seeded extractors. Namely, we need a seed which is perfectly random and independent of the source we extract from. How do we get such a seed? For some applications, this issue can be taken care of (e.g. by enumerating through all possible seed values). The most important case is simulating BPP algorithms [62]. BPP is the set of languages L satisfying that there exists a polynomial-time TM $A_L(w,r)$ such that for every $w \in L$, $\Pr_r[A_L(w,r) = 1] \geq 2/3$ and for every $w \notin L$, $\Pr_r[A_L(w,r) = 0] \geq 2/3$. Consider a language $L \in \text{BPP}$, then given access to a weak random source with enough min-entropy, we can decide whether $w \in L$ as follows: On an input w and an element x sampled from the weak source, we output the majority of $A_L(w, \textsc{Ext}(x,y))$ over all seed values y. Unfortunately, the trick of enumerating over all possible seed values does not work for every application. Hence, for these applications, the issue seems to go back to the original problem which we try to solve using extractors. *Can we get rid of the need for a seed and have* **seedless** *extractors?*

When the sources are restricted and have special structures, it becomes possible to have seedless extractors. One example is samplable sources, which are generated by some efficient sampling algorithms [55]. Trevisan and Vadhan [55] showed that when extractors are allowed more computational resources (e.g. circuit size) than the sampling algorithms, seedless extracting becomes possible. Another example is bit-fixing sources, in which each bit is either fixed (containing no randomness) or random and is independent of others [11]. From such a source of length n with min-entropy $n^{1/2+\gamma}$, for any constant $\gamma \in (0, 1/2)$, Kamp and Zuckerman [33] 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. [17] 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.

On the other hand, it is also possible to have seedless extractors for multiple independent sources [10]. A recent breakthrough by Barak *et al.* [3] provides a seedless extractor for a constant number of independent sources each with some constant min-entropy rate (average min-entropy per bit). This has been improved subsequently [4, 45, 44], which now allows one to extract from three independent sources each with any constant min-entropy rate. For the case of extracting from two independent sources, Bourgain [8] gave a seedless extractor which lowers the requirement of min-entropy rate from a previous barrier of 1/2 to slightly below. Note that in the case of two independent sources, we can see one of them as the seed which now is only slightly random, instead of perfectly random as in the case for seeded extractors.

In this thesis, we will consider the task of deterministically extracting randomness from several classes of weak random sources. 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 statistical distance ε, which we call error, of its output distribution to the uniform one, where the statistical distance of two distributions $\mathcal{X}$ and $\mathcal{Y}$ are defined as $(1/2) \cdot \sum_z |\Pr[\mathcal{X} = z] - \Pr[\mathcal{Y} = z]|$.

1.1 Extracting Randomness from Multiple Independent Sources

We first work on deterministic extraction from multiple independent sources. Our first result is a simple extractor for two sources. One of our main technical contribution is a generalization of the well-known *leftover hash lemma* [28]. The leftover hash lemma says that if we sample a function h uniformly from a family H of pair-wise independent functions and apply it on an input x sampled from a source with enough min-entropy, the output $h(x)$ will look almost like random. This is usually applied in the setting of seeded extractors, in which the perfect random seed is used to sample uniformly from H. We generalize the leftover hash lemma to allow sampling from H according to any distribution with high enough min-entropy. This provides us a way to extract from two independent weakly random sources: one source to sample the input x while the other to sample the function h. More precisely, our extractor takes two input strings $v, w \in \{0,1\}^n$, sees them as vectors from $\mathcal{F}^\ell$, where $\mathcal{F} = GF(2^m)$ for some m with $n = m\ell$, and outputs their inner product $\langle v, w\rangle = \sum_{i=1}^{\ell} v_i w_i$ over $\mathcal{F}$. Then, from two independent sources of length n and of min-entropy k_1 and k_2, we can extract $k_1 + k_2 + 2 - n - 2\log\frac{1}{\varepsilon}$ bits with error ε. We also extend our construction for the case when there are $t \geq 3$ independent sources available. Our deterministic extractor can extract $k_1 + k_2 + 2 - n - 2\log\frac{1}{\varepsilon}$ bits, where k_1 and k_2 are the two largest min-entropies of the t sources. It has the following nice features. First, our extractor works as long as two sources have enough min-entropy; it can work even when only two sources contain randomness (thus with a very low average min-entropy rate). Second, as is in [14, 13], our extractor can still work even with all but one source exposed. In fact, the best result of [13] is a special case of ours. Finally, to construct our extractor, we do not need to know beforehand the specific min-entropy of each source.

Next, we introduce one possible application with strong multi-source extractors. We consider the following cryptographic task which generalizes the two-party case in [14]. Suppose a group of parties $P_1, \ldots, P_t$ are together initially and later go far away from each other, and then they want to establish a secret key for group communication over an insecure channel. Can this task be achieved without using ideal local randomness? We give one solution. Initially these parties share some $\mathcal{X}$ sampled from a weak source when they are together. After departing from each other, each party P_i samples $\mathcal{X}_i$ from his/her own local weak source, and sends it to the others. Once receiving all $\mathcal{X}_i$'s, each party computes the secret key $\textsc{Ext}(\mathcal{X}, \mathcal{X}_1, \ldots, \mathcal{X}_t)$ using our extractor $\textsc{Ext}$, which is secure even against an adversary who knows $\mathcal{X}_2, \ldots, \mathcal{X}_t$. This can be augmented with an authentication process to prevent an adversary from impersonating a legitimate party.

1.2 Deterministic Extractors for Independent-Symbol Sources

Note that the researches for multiple independent sources and bit-fixing sources 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.

1.2.1 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 we attempt to extract randomness from such sources.

Previously, [35, 34] 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. [32] 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.

1.2.2 Main Results

For independent-symbol sources, we first give 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 [33], 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 non-uniformly 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, 45] do not seem to work for such sources. Independent of our work, Kamp et al. [32] also gave the same construction but used a different analysis.

To extract more randomness, we borrow the technique of Gabizon et al. [17]. Now, as in [17], 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 $\varepsilon \geq 2^{-\Omega(n^{\gamma})}$. Second, when $k \geq \log^{c} n$, for some constant $c > 0$, we can extract $m = k - (1/\varepsilon)^{O(1)}$ bits with error $\varepsilon \geq k^{-\Omega(1)}$. 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 [17] 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 whenever $k = \omega(d + \log(n/\varepsilon))$. 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. Only for the case with $d = O(1)$, $k \geq n^{1/2+\gamma}$, and $\varepsilon \geq 2^{-\Omega(n^{\gamma})}$ do we have an explicit construction matching this bound.

On the other hand, one may also wonder whether this existential upper bound we derive on entropy loss is tight. We show 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 [43], which has the same bound on seeded extractors for *general* sources. The idea in [43] 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.

1.2.3 Techniques

Our first extractor for independent-symbol sources, which extracts about $\log k$ bits, was inspired by that of Kamp and Zuckerman [33], 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 $\mathcal{X} = \mathcal{X}_1 \circ \cdots \circ X_n$ where $\circ$ denotes concatenation, we see each $\mathcal{X}_i$ as an element of $\mathbb{Z}_M$ and output $\mathcal{X}_1 + \cdots + \mathcal{X}_n$ 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 [33], we will show that each step of our walk brings the distribution closer to uniform when the symbol from the source contains some randomness. We observe that the transition matrix of each step is a circulant matrix, in which each row is a cyclic shift of the previous row. Hence, we can use the properties of circulant matrices to show the progress we can make after each step. Our proof 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 [17]. Our constructions and proofs in this part follow very closely those in [17]. 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 [33] 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 [43]. Namely, given any function $\textsc{Ext} : \{0,1\}^n \times \{0,1\}^s \to \{0,1\}^m$ 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 $\textsc{Ext}$ 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 $\textsc{Ext}$ 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.

1.3 Extracting Computational Entropy and Learning Noisy Linear Functions

Next, we would like to go the other direction to look for a more general class of sources from which seedless extraction is still possible. In particular, we will consider sources which may contain no randomness at all in a statistical sense, but *look* slightly random to computational-bounded observers, such as small circuits. That is, we will go from a traditional, statistical setting to a computational one. It is conceivable that in many situations when we consider a source random, it may in fact only appear so to us, while its actual statistical min-entropy may be much smaller (or even zero) especially if we take into account some correlated information which we can observe. Another application of this notion is in cryptography, and in fact the idea of extracting computational randomness has appeared implicitly long ago [59, 19, 23], for the task of constructing pseudo-random generators from one-way functions. The idea is that given a one-way function g, it is hard to invert $g(y)$ to get y, and this means that given the (correlated) information $g(y)$, y still looks somewhat random, from which one can extract some bits that look almost random. However, while there is a natural and well-accepted definition for what we mean that a distribution looks almost random [59], it seems less clear for what we mean that a distribution looks slightly random and for how to measure the amount of randomness in it. In fact, there are several alternatives which all seem reasonable, but there are provable discrepancies among them [5, 25]. To extract randomness from a source with so-called HILL-entropy [5], the strongest among them, one can simply use any statistical extractor. Here we consider a weaker (more general) notion of computational randomness, which appears in [25], and we call it *computational min-entropy*.

1.3.1 Computational min-entropy

To model the more general situation that one may observe some correlated information about the sources, we consider sources of a conditional form $(\mathcal{V}|\mathcal{X})$, where $\mathcal{V}$ is the source from which we want to extract and $\mathcal{X}$ (could be empty) is some distribution which one can observe. The correlation between $\mathcal{V}$ and $\mathcal{X}$ is modeled by $\mathcal{V} = f(\mathcal{X})$ for some function f. In the example of the one-way function, f is the inverse function g^{-1}, which is hard to compute, and $\mathcal{X}$ is the distribution of $g(y)$ over a random y. Here in our definition, we allow f to be probabilistic and we even do not require it to have an efficient (or even computable) algorithm, and furthermore, we do not require $\mathcal{X}$ to be efficiently samplable either. We say that such a distribution $(f(\mathcal{X})|\mathcal{X})$ has computational min-entropy k if given input x sampled from $\mathcal{X}$, any circuit of size 2^k can only predict $f(x)$ correctly with probability at most 2^{-k} (a more general definition is to have the circuit size as a separate parameter, but our extractor construction does not seem to work for this general definition). From the distribution $f(\mathcal{X})$, we would like to extract randomness which when given $\mathcal{X}$ still looks random to circuits of a certain size. Note that a source $\mathcal{Y}$ with statistical min-entropy k can be seen as such a source $(f(\mathcal{X})|\mathcal{X})$ with computational min-entropy k, where we can simply have no $\mathcal{X}$ or just have $\mathcal{X}$ taking a fixed value, and let f be a probabilistic function with $\mathcal{Y}$ as its output distribution. This means that extractors for sources with computational min-entropy can immediately work for sources with statistical min-entropy, and thus results in the computational setting can be seen as a generalization of those in the traditional, statistical setting. On the other hand, for a deterministic function f, $f(x)$ has no statistical min-entropy at all when given x. Still, according to our definition, as long as f is hard to compute, $(f(\mathcal{X})|\mathcal{X})$ in fact can have high computational min-entropy.

Extractors for such sources were implicitly proposed before [19, 23], and they are seeded ones. In fact, any seeded statistical extractor with some additional *reconstruction* property (in the sense of [54]) gives a seeded extractor for such sources [5, 53, 25]. However, just as in the statistical setting, several natural questions arise in the computational setting too. *To extract from such sources, do we really need a seed? Can we use a weaker seed which is only slightly random, instead of perfectly random, but still in a statistical sense, or an even weaker seed which only looks slightly random in a computational sense but may contain no randomness at all in a statistical sense?* We will try to answer these questions. Seeing the seed as an additional independent source, a general question is: *Can we have seedless extractors for multiple independent sources, each with some computational min-entropy?* One can see this as a step toward extending the study of multi-source extractors from the traditional, statistical setting to a new, computational setting. One can also see this as providing a finer map for the landscape of statistical extractors, according to the degree of their reconstruction property.

1.3.2 Main Results

First, we show that it is impossible to have seedless extractors for one single source, even if the source of length n can have a computational min-entropy as high as $n-2$ and even if we only want to extract one bit.

Next, we show that with the help of a weak seed, it becomes possible to extract randomness from such sources. We use a two-source extractor of Lee *et al.* [35], denoted as Ext. As shown in [35], it works for any two independent sources both containing some statistical min-entropy. Moreover, it is also known to work when one source contains some computational min-entropy and the other, the seed, is perfectly random (in a statistical sense) [21]. Our second result shows that it even works when the seed only contains some statistical min-entropy. More precisely, we show that given any source $(f(\mathcal{X})|\mathcal{X})$ with computational min-entropy $k_1 = n - k + O(k/\log k)$ and another independent source $\mathcal{W}$ with statistical min-entropy k, $\textsc{Ext}(f(\mathcal{X}), \mathcal{W})$ given $\mathcal{X}$ cannot be distinguished from random with advantage $\varepsilon = 2^{-O(\sqrt{k/\log k})}$ by circuits of size $s = 2^{n-k+O(k/\log k)}$. Then we proceed to show that it works even when the seed only contains computational min-entropy. More precisely, for a source $(g(\mathcal{Y})|\mathcal{Y})$ with computational min-entropy k, $\textsc{Ext}(f(\mathcal{X}), g(\mathcal{Y}))$ given $(\mathcal{X}, \mathcal{Y})$ still cannot be distinguished with advantage ε by circuits of size about s. This can be seen as a seedless extractor for two independent sources, both with computational min-entropy.

We do not know if the statistical extractors of [3, 4, 45, 8, 44] for multiple independent sources can work in the computational setting, since to work in this setting, we need them to have some reconstruction property. For the extractors from [19, 21], this property can be translated to a task in learning theory, and the proofs there can be recast as providing an algorithm for learning linear functions under *uniform* distribution with adversarial noise. Our second result can be seen as a generalization of [19, 21], and we are facing a more challenging learning problem: learning linear functions under *arbitrary* distribution with adversarial noise. Our third result provides an algorithm for this problem, which, in addition to being used to prove our second result, may have interest of its own.

In the learning problem, there is an unknown linear function $v : \mathcal{F}^\ell \to \mathcal{F}$ which we want to learn, and a distribution $\mathcal{W}$ over $\mathcal{F}^\ell$ from which we can sample w to obtain a training example $(w, q(w))$, for some function $q : \mathcal{F}^\ell \to \mathcal{F}$. The function q can be seen as a noisy version of v with some noise rate α, and there are two noise models. In the adversarial-noise model, q is a deterministic function such that $\Pr_{w \in \mathcal{W}}[q(w) \neq v(w)] \leq \alpha$. In the random-noise model, q is a probabilistic function such that independently for any w, $\Pr[q(w) \neq v(w)] \leq \alpha$. We consider the more difficult adversarial-noise model, and our algorithm works for an arbitrary distribution $\mathcal{W}$, while its complexity depends on the min-entropy k of $\mathcal{W}$. More precisely, our algorithm samples $2^{O(k/\log k)}$ examples, runs in time $2^{n-k+O(k/\log k)}$, and with high probability outputs a list containing every linear function v satisfying $\Pr_{w \in \mathcal{W}}[q(w) \neq v(w)] \leq \alpha$, for $\alpha = 1 - 2^{-O(\sqrt{k/\log k})}$. The factor 2^{n-k}

in our running time is in fact unavoidable because one can easily find a distribution $\mathcal{W}$ for which the number of such v's, and thus the running time, is in fact at least 2^{n-k}. Note that when $\mathcal{W}$ is the uniform distribution (with $k = n$), our algorithm runs in time $2^{O(n/\log n)}$ and takes $2^{O(n/\log n)}$ samples.

Previously, the algorithm of Blum *et al.* [7] can learn under arbitrary distribution but in the random-noise model, while that of Feldman *et al.* [16] can learn in the adversarial-noise model but under the uniform distribution. Both algorithms learn the parity functions on n variables, tolerate a noise rate $\alpha \leq 1/2 - \Omega(1)$, run in time $2^{O(n/\log n)}$, and take $2^{O(n/\log n)}$ samples. Very recently, Kalai *et al.* [31] gave an algorithm which can learn the parity functions under arbitrary distribution in the adversarial-noise model, but the hypothesis they produce is not in the linear form, so it cannot be used for our extractors. Furthermore, they only produce one hypothesis instead of all the legitimate ones, and their technique does not seem to generalize from the parity functions to the linear functions over larger fields. Thus, to the best of our knowledge, the task our learning algorithm achieves has not been accomplished before. Finally, just as the result of [19] can yield a list-decoding algorithm for Hadamard codes, so can ours, while that of [31] can not. In fact, our list-decoding algorithm can work even when all but 2^k symbols from the codeword are erased and an α fraction of the remaining symbols are corrupted. It can also be seen as list-decoding a *punctured* Hadamard code, where a punctured code is obtained from a code by deleting all but a small number of symbols from the codeword.

1.3.3 Techniques.

For our impossibility result, we show that for any function $\text{Ext} : \{0,1\}^n \to \{0,1\}$, there exists a function $f : \{0,1\}^{3n} \to \{0,1\}^n$ such that $(f(\mathcal{X})|\mathcal{X})$ has computational min-entropy $n-2$, but $\text{Ext}(f(x))$ takes an identical value for all x. We show the existence of such a function f by a standard probabilistic argument: in fact, a random function from $\{0,1\}^{3n}$ to $\text{Ext}^{-1}(b)$ is likely to work, for $b \in \{0,1\}$ giving a larger $\text{Ext}^{-1}(b)$.

To show that our extractor works in the computational setting, we follow the approach of [19] and reduce it to the task of learning linear functions as we just discussed. Now more precisely, for the case when the source $(f(\mathcal{X})|\mathcal{X})$ has computational min-entropy and the seed $\mathcal{W}$ has statistical min-entropy, the reduction works as follows. Assume our extractor Ext does not work, then some efficient distinguisher can tell the distribution of $\text{Ext}(f(x), \mathcal{W}) = \langle f(x), \mathcal{W} \rangle$ from random given x, for a large fraction of x from $\mathcal{X}$. For any such x, we can then predict the value $\langle f(x), \mathcal{W} \rangle$ with a good probability, given the ability to sample from $\mathcal{W}$, which can then be used by the learning algorithm to learn $f(x)$. This would give us an efficient algorithm for predicting $f(x)$ for those x's, if we could in fact sample $\mathcal{W}$ efficiently, but this may not be the case in general as $\mathcal{W}$ could be any arbitrary distribution. Still, by an average argument, there exist some fixed samples which maintain the predicting probability, so we can hard-wire them in to get a circuit for predicting f well. If the function f is hard, this is impossible, so the assumed distinguisher cannot exit, and Ext indeed works. For the case that the seed comes from a distribution $(g(\mathcal{Y})|\mathcal{Y})$ with computational min-entropy, observe that $g(\mathcal{Y})$ alone (without conditioning on $\mathcal{Y}$) must have some statistical min-entropy, because otherwise it becomes easy to predict. Then a very similar argument as above can be used.

Note that our results on extractors still depend on the existence of a good learning algorithm, and our main technical contribution can be seen as providing such an algorithm. Our algorithm can be seen as extending that of [7] from the random-noise model to the adversarial-noise model.

Our learning algorithm works as follow. We start by sampling some number K of training examples $(w, q(w))$ from $(\mathcal{W}, q(\mathcal{W}))$. Note that each example $(w, q(w))$ gives us a linear equation $\langle v, w \rangle = q(w)$ for the v which we want to learn, so the K examples gives us a system of K linear equations, some of which may be wrong. We reduce the original problem of learning the unknown v to the problem of solving such a noisy system of learning equations, and to solve it, we proceed in two phases. In the forward phase, we start from the system, and use several iterations to produce smaller and smaller systems with fewer and fewer variables. When we have a small enough system which we can afford to solve using brute force, we enter the backward phase. In the backward phase, we start from the last system produced by the forward phase, and work backward on larger and larger systems produced in the forward phase to obtain solutions for more and more variables. Since the possible solutions may not be unique, we keep them all in a list in each iteration, and the list in the final iteration of the backward phase is our output, which we hope contains the correct v.

The forward phase is similar in spirit to an approach in [7]. The key is to guarantee that after each iteration, the new system still contains a good fraction of correct equations with respect to the solution v, so that v will not be lost when solving this new system. Using an argument similar to that in [7], we can show that this does hold with some significant probability. On the other hand, it is not clear whether or not some system produced in the forward phase could turn many originally-bad solutions into good ones for it (satisfying a good fraction of its equations). That is, not only is v a good solution for the system, there are in fact too many good solutions for it. If this happens, then in the backward when we try to solve this system, we cannot afford to keep all such solutions, and we have the risk of losing the actual solution v. This tricky situation does not arise in the random-noise model considered in [7], so a much simpler algorithm works there. In the adversarial-noise model, this seems unavoidable. Fortunately, we can show that with high probability, the systems we produce indeed do not have too many good solutions. This turns out to rely on the fact that our extractor is also a good *statistical* extractor, together with the property, which we will show, that each system is likely to be close to some good distribution with high statistical min-entropy.

1.4 Extracting Computational Entropy from Computational Independent-Symbol Sources

Finally, we consider computational independent-symbol sources, characterized by the parameters $n, d, k \in \mathbb{N}$, and s. Just as independent-symbol sources, each computational independent-symbol source consists of n mutually independent parts, $(f_1(\mathcal{X}_1)|\mathcal{X}_1), \cdots,$ $(f_n(\mathcal{X}_n)|\mathcal{X}_n)$, each $f_i(\mathcal{X}_i)$ of length d such that for each i if given input x_i sampled from $\mathcal{X}_i$, any circuit of size s can only predict $f_i(x_i)$ with probability at most 2^{-k_i} for some $k_i \leq d$, and the sum of k_i's is k. Note that we can allow to set circuit size as a separate parameter to define the computational independent-symbol sources.

We show that our extractor for independent-symbol sources, defined as $\textsc{Ext}(\mathcal{V}_1, \cdots,$ $\mathcal{V}_n) = \sum_i \mathcal{V}_i$, still works for computational independent-symbol sources. One of our main technical contribution is a generalization of the well-known *hardcore set lemma* [26]. The well-known Impagliazzo's hardcore set lemma says that if a function $f : \{0,1\}^\ell \to \{0,1\}$ is *mildly hard*, that is, any small circuit must fail to compute it correctly on more than some fraction of inputs, then there exists a large enough *hardcore set* H, such that f is *extremely hard* on H, in the sense that any somewhat smaller circuit must fail to compute f correctly on more than a $\frac{1}{2} - \varepsilon$ fraction of inputs in H, for some small ε. We extend the case of $f : \{0,1\}^\ell \to \{0,1\}$ to $f : \{0,1\}^\ell \to \{0,1\}^d$. More precisely, we show that if any circuit of size s must fail to compute $f : \{0,1\}^\ell \to \{0,1\}^d$ correctly on more than a δ fraction of inputs, then there exist some *disjoint binary hardcore sets* $H_1, \cdots, H_r$ of total size at least $(\delta/2) \cdot 2^\ell$, where for any $i \in [r]$, $H_i \subseteq f^{-1}(I_i)$ for some $I_i \subseteq \{0,1\}^d$ with $|I_i| = 2$, such that any circuit of size $\Omega(\varepsilon^2\delta^2 s/2^{6d})$ must fail to compute f correctly on more than a $\frac{1}{2} - \varepsilon$ fraction of inputs in H, for some small ε. Moreover, using the generalized hardcore set lemma and the technique of Sudan et al. [51], we can find a source $\mathcal{Y}$ such that the two distributions $\mathcal{X} \circ f(\mathcal{X})$ and $\mathcal{X} \circ \mathcal{Y}$ cannot be distinguished with advantage $\varepsilon\delta$ by circuits of size $\Omega(\varepsilon^2\delta^2 s/2^{6d})$ and with a significant probability, $(\mathcal{Y}|\mathcal{X})$ has enough min-entropy. Hence, we can reduce the problem of constructing computational extractors for computational independent-symbol sources to that of constructing extractors for independent-symbol sources, and show that no circuit of size $\Omega(s(\log n/nk2^d)^2))$ can distinguish the distributions $\mathcal{X}_1 \circ \cdots \circ \mathcal{X}_n \circ \sum_{i=1}^n f_i(\mathcal{X}_i)$ and $\mathcal{X}_1 \circ \cdots \circ \mathcal{X}_n \circ \mathcal{U}_{[M]}$, where $\mathcal{U}_{[M]}$ is the uniform distribution over $\{1, 2, \cdots, M\}$, with advantage $O(M^2 \log n/k)$ for $k \geq \Omega(M^2d^2)$.

In fact, the result of computational extractors for computational independent-symbol sources implies a generalization of the well-known XOR lemma [59]. The XOR lemma says that if $f : \{0,1\}^\ell \to \{0,1\}$ is ”mildly hard” for small circuits, then $F(x_1, \cdots, x_t) \equiv \oplus_{i=1}^t f(x_i)$ for sufficiently large t, is ”extreme hard” for smaller size circuits. We show that if there are n functions $f_1, \cdots, f_n$ such that for each i, any circuit of size s must fail to compute $f_i : \{0,1\}^{\ell_i} \to \{0,1\}^d$ correctly for a δ_i fraction of inputs, and $\delta = \sum_{i=1}^n \delta_i \geq \Omega(M^2d)$, then any circuit of size $\Omega(s(\log n/n\delta2^d)^2))$ with input $x_1, \cdots, x_n$ can only compute $\sum_{i=1}^n f_i(x_i)$ correctly on no more than a $\frac{1}{M} + \frac{M^2 \log n}{\delta}$ fraction of inputs.

For the generalized hardcore set lemma, one may wonder if there exists a larger binary hardcore set, for example, with size $\delta \cdot 2^\ell/2^d$. We show that in any black-box construction, one can only prove the existence of a binary hardcore set with size $O(\delta 2^{\ell}/2^{2d})$. We say that an oracle algorithm $\textsc{Dec}^{(\cdot)}$ is a *black-box* (δ, ε, d)*-construction* of a hardcore set, if the following holds. Given any function $f : \{0,1\}^{\ell} \to \{0,1\}^{d}$, where $\ell = \Omega(d)$, and a family of functions $G = \{g_I | I \subseteq \{0,1\}^d \text{ with } |I| = 2\}$ satisfying that for each $g_I \in G$ and $H \subseteq f^{-1}(I)$ with size s, g_I must fail to compute $f(x)$ on at most a $(1-\varepsilon)/2$ fraction of inputs in H, then $\textsc{Dec}^{G}$ must fail to compute f on at most a δ fraction of inputs. We call s the size complexity of black-box construction. We use a probabilistic method to show that any black-box (δ, ε, d)-construction must have size complexity $O(\delta 2^{\ell}/2^{2d})$ where $\Omega(2^{-cd}) \leq \delta$ for some constant c, $\varepsilon \leq 1/5$ and $d \geq 2$.

1.5 The Rest of This Thesis

In Chapter 2, we give some preliminaries, and define several classes of sources which we will consider in this thesis. In Chapter 3, we consider the case of multiple independent sources, while in Chapter 4, we consider a new class of weak random sources, independent-symbol sources, which can be seen as sources that lie in between multiple independent sources and bit-fixing sources. In Chapter 5, we will go to a computational setting and consider a more general class of sources, weak random sources with computational min-entropy. In Chapter 6, we consider the computational independent-symbol sources. Finally, we will discuss the future work in Chapter 7.

Chapter 2

Preliminaries

In this chapter, we describe some notations, and define some classes of sources which we will consider in this thesis. Throughout this thesis, we will use the terms *random variable* and *distribution* interchangeably. All logarithms will have base two. Let $\mathsf{SIZE}(s)$ be the class of functions computable by Boolean circuits of size s. For $n \in \mathbb{N}$, let $[n]$ denote the set $\{1, \ldots, n\}$. For $i, j \in \mathbb{N}$ with $i \leq j$, let $[i, j]$ denote the set $\{i, i+1, \cdots, j\}$. For convenience, let $D = 2^d$, and $M = 2^m$. 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.

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 $\{0,1\}^n$. We will sometimes see a distribution $\mathcal{X}$ over a set S as an $|S|$-dimensional vector, with $\Pr[\mathcal{X} = x]$ at dimension $x \in S$, and 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 $x \in S$, $\Pr[\mathcal{X} = x] = \sum_{i \in [t]} \alpha_i \Pr[\mathcal{X}^i = x]$. A distribution is called *flat* if it is a uniform distribution over some set S.

2.1 Distances

We will mainly measure the distance between two distributions $\mathcal{X}, \mathcal{Y}$ over a set S by their *statistical distance* or *variational distance*, defined as

$$\Delta(\mathcal{X}, \mathcal{Y}) = \max_{T \subseteq S} \left| \Pr[\mathcal{X} \in T] - \Pr[\mathcal{Y} \in T] \right|.$$

Note that this distance is exactly half of the L_1-distance, defined as

$$\|\mathcal{X} - \mathcal{Y}\|_1 = \sum_{x \in S} \left| \Pr[\mathcal{X} = x] - \Pr[\mathcal{Y} = x] \right|.$$

Call a distribution ε-random if its statistical distance to the uniform distribution is at most ε.

The statistical distance has the following nice properties.

Proposition 2.1.1. For any two distributions $\mathcal{X}$ and $\mathcal{Y}$, $0 \leq \Delta(\mathcal{X}, \mathcal{Y}) \leq 1$.

Proposition 2.1.2. For any three distribution $\mathcal{X}, \mathcal{Y}$, and $\mathcal{Z}$, $\Delta(\mathcal{X}, \mathcal{Z}) \leq \Delta(\mathcal{X}, \mathcal{Y}) + \Delta(\mathcal{Y}, \mathcal{Z})$.

Proposition 2.1.3. If $\Delta(\mathcal{X}, \mathcal{Y}) \leq \varepsilon$ for two distributions $\mathcal{X}$ and $\mathcal{Y}$ over a set S, then for any function $f : S \to T$, $\Delta(f(\mathcal{X}), f(\mathcal{Y})) \leq \varepsilon$.

Proof.

$$\begin{aligned}
\Delta(f(\mathcal{X}), f(\mathcal{Y})) &= \frac{1}{2} \sum_{t \in T} \left| \Pr[f(\mathcal{X}) = t] - \Pr[f(\mathcal{Y}) = t] \right| \\
&= \frac{1}{2} \sum_{t \in T} \left| \sum_{s \in f^{-1}(t)} \Pr[\mathcal{X} = s] - \sum_{s \in f^{-1}(t)} \Pr[\mathcal{Y} = s] \right| \\
&\leq \frac{1}{2} \sum_{t \in T} \sum_{s \in f^{-1}(t)} \left| \Pr[\mathcal{X} = s] - \Pr[\mathcal{Y} = s] \right| \qquad (2.1) \\
&= \frac{1}{2} \sum_{s \in S} \left| \Pr[\mathcal{X} = s] - \Pr[\mathcal{Y} = s] \right| \\
&= \Delta(\mathcal{X}, \mathcal{Y}) \\
&\leq \varepsilon,
\end{aligned}$$

where (2.1) is due to the triangle inequality. □

Lemma 2.1.4. *Suppose that $\varepsilon_1, \varepsilon_2 > 0$. For any independent sources $\mathcal{X}_1, \mathcal{X}_2, \mathcal{Y}_1$, and $\mathcal{Y}_2$ satisfying that $\Delta(\mathcal{X}_1, \mathcal{Y}_1) = \varepsilon_1$ and $\Delta(\mathcal{X}_2, \mathcal{Y}_2) = \varepsilon_2$, then*

$$\Delta(\mathcal{X}_1 \circ \mathcal{X}_2, \mathcal{Y}_1 \circ \mathcal{Y}_2) \leq \varepsilon_1 + \varepsilon_2$$

where $\circ$ denotes concatenation.

Proof. Applying the triangle inequality and the fact that $\mathcal{X}_1, \mathcal{X}_2, \mathcal{Y}_1$, and $\mathcal{Y}_2$ are independent, we obtain that

$$\begin{aligned} \Delta(\mathcal{X}_1 \circ \mathcal{X}_2, \mathcal{Y}_1 \circ \mathcal{Y}_2) &\leq \Delta(\mathcal{X}_1 \circ \mathcal{X}_2, \mathcal{Y}_1 \circ \mathcal{X}_2) + \Delta(\mathcal{Y}_1 \circ \mathcal{X}_2, \mathcal{Y}_1 \circ \mathcal{Y}_2) \\ &= \Delta(\mathcal{X}_1, \mathcal{Y}_1) + \Delta(\mathcal{X}_2, \mathcal{Y}_2) \\ &= \varepsilon_1 + \varepsilon_2. \end{aligned}$$

□

We can easily generalize the above lemma to 2ℓ independent sources for any integer ℓ.

Corollary 2.1.5. *Suppose that $\ell \in \mathbb{N}$, and $\varepsilon_1, \cdots, \varepsilon_\ell > 0$. For any independent sources $\mathcal{X}_1, \cdots, \mathcal{X}_\ell, \mathcal{Y}_1, \cdots, \mathcal{Y}_\ell$ satisfying that for any $i \in [\ell]$, $\Delta(\mathcal{X}_i, \mathcal{Y}_i) = \varepsilon_i$, then*

$$\Delta(\mathcal{X}_1 \circ \cdots \circ \mathcal{X}_\ell, \mathcal{Y}_1 \circ \cdots \circ \mathcal{Y}_\ell) \leq \sum_{i=1}^{\ell} \varepsilon_i.$$

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

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

2.2 Min-entropy and Computational Min-entropy

We mainly measure the amount of randomness in a distribution $\mathcal{X}$ by its *(statistical) min-entropy.*

Definition 2.2.1. We say that a distribution $\mathcal{X}$ has (statistical) min-entropy k, denote by $\mathrm{H}_\infty(\mathcal{X}) = k$, if for any x, $\Pr[\mathcal{X} = x] \leq 2^{-k}$.

The notion of min-entropy has the following properties.

Proposition 2.2.2. Let $\mathcal{X}$ be a source over $\{0,1\}^n$. Then $0 \leq \mathrm{H}_\infty(\mathcal{X}) \leq n$.

Proposition 2.2.3. For any deterministic function f and any distribution $\mathcal{X}$, $\mathrm{H}_\infty(f(\mathcal{X})) \leq \mathrm{H}_\infty(\mathcal{X})$.

Proof. Note that for any z, $\Pr[f(\mathcal{X}) = z] = \sum_{x \in f^{-1}(z)} \Pr[\mathcal{X} = x]$. Hence, $\max_z \Pr[f(\mathcal{X}) = z] \geq \max_x \Pr[\mathcal{X} = x]$, which implies $\mathrm{H}_\infty(f(\mathcal{X})) \leq \mathrm{H}_\infty(\mathcal{X})$. □

In fact, if we can extract randomness from $\mathcal{X}$, then $\mathcal{X}$ is close to have high min-entropy [48].

Remark 2.2.4. The more familiar notion of entropy is the *Shannon entropy*, $H(\mathcal{X}) = -\sum_x \Pr[\mathcal{X} = x] \log(\Pr[\mathcal{X} = x])$. However, consider a source $\mathcal{X}$ which outputs 0^n with probability 0.9 and outputs a random value in $\{0,1\}^n$ with probability 0.1. Note that $H(\mathcal{X}) \geq 0.1n$, but any function will output the same value at least 0.9 of the times. Hence, we will not use the Shannon entropy to measure the amount of randomness.

Proposition 2.2.5. For any $k \in \mathbb{N}$, any source with min-entropy k is a convex combination of flat sources with min-entropy k.

Proof. Consider any $\mathcal{X}$ over $\{0,1\}^n$ with $\mathrm{H}_\infty(\mathcal{X}) = k$. Let $N = 2^n$, and see $\{0,1\}^n = [N]$. We can see any source over $\{0,1\}^n$ with min-entropy k as a vector $(p_1, \ldots, p_N)$ with the property that $\sum_{j \in [N]} p_j = 1$ and $0 \leq p_j \leq 2^{-k}$ for every $j \in [N]$. Since all these inequalities are linear, the set of such vectors forms a convex polytope. Hence, each vector in the set can be represented as a convex combination of vertices (corners) of the polytope, where a vertex of the polytope is a vector satisfying a maximal subset of these inequalities when treat these inequalities as equalities. Note that a vector is a vertex of the polytope if and only if it has 2^k values of j such that $p_j = 2^{-k}$, and for the rest of j, $p_j = 0$. That is, the vertices of the polytope correspond exactly to the vectors given by those flat sources over $[N]$, and we complete the proof. □

According to the above proposition, it suffices to work for flat distributions to analyze extractors.

Next, we define the notion of *computational min-entropy*. Here, we consider conditional distributions of the form $(\mathcal{V}|\mathcal{X})$, with $\mathcal{V} = f(\mathcal{X})$, for some distribution $\mathcal{X}$ and some function f, which could be either probabilistic or deterministic, and we say that it has some computational min-entropy if $v = f(x)$ is hard to predict given $x \in \mathcal{X}$.

Definition 2.2.6. We say that a conditional distribution $(\mathcal{V}|\mathcal{X})$ has *computational min-entropy* k, denoted by $\mathrm{H}_c(\mathcal{V}|\mathcal{X}) = k$, if for any $C \in \mathsf{SIZE}(2^k)$, $\Pr[C(\mathcal{X}) = \mathcal{V}] \leq 2^{-k}$.

Note that for a deterministic function f, when given $x \in \mathcal{X}$, $f(x)$ has no randomness left and thus has no statistical min-entropy, but it still can have high computational min-entropy if f is in fact hard to compute.

2.3 Multiple Independent Sources

In Chapter 3, we will study deterministic extractors for multiple independent sources.

Definition 2.3.1. For $t \in \mathbb{N}$, a function $\text{EXT} : (\{0,1\}^n)^t \to \{0,1\}^m$ is called a $(k_1, k_2, ..., k_t, \varepsilon)$-extractor if for any t *independent* random variables $\mathcal{X}_1, \ldots, \mathcal{X}_t$, with each $\mathcal{X}_i$ distributed over $\{0,1\}^n$ and $\mathrm{H}_\infty(\mathcal{X}_i) \geq k_i$, we have

$$\Delta(\text{EXT}(\mathcal{X}_1, \ldots, \mathcal{X}_t), \mathcal{U}_m) \leq \varepsilon.$$

EXT is called a $(k_1, \ldots, k_t, \varepsilon)$-strong-multi-source-extractor if it satisfies the stronger property that

$$\Delta\left(\mathcal{X}_{[2,t]} \circ \text{EXT}(\mathcal{X}_1, \ldots, \mathcal{X}_t), \mathcal{X}_{[2,t]} \circ \mathcal{U}_m\right) \leq \varepsilon.$$

2.4 Independent-Symbol Sources

In Chapter 4, we will focus on a special kind of sources which consist of n independent symbols over some set $[D]$.

Definition 2.4.1. A distribution $\mathcal{X} = \mathcal{X}_1 \circ \cdots \circ \mathcal{X}_n$ 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 Chapter 4 is to extract randomness from such (n, D, k)-sources.

Definition 2.4.2. For $n, D, k, s, m \in \mathbb{N}$ and $\varepsilon \in [0, 1]$, a function $\textsc{Ext} : [D]^n \times \{0, 1\}^s \to \{0, 1\}^m$ is called an (n, D, k, ε)-extractor if for any (n, D, k)-source $\mathcal{X}$,

$$\Delta(\textsc{Ext}(\mathcal{X}, \mathcal{U}_s), \mathcal{U}_m) \leq \varepsilon.$$

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.

2.5 Computational Extractors

We say that a function $D : \{0, 1\}^n \to \{0, 1\}$ is an ε-distinguisher for two distributions $\mathcal{X}$ and $\mathcal{Y}$ over $\{0, 1\}^n$ if

$$|\Pr[D(\mathcal{X}) = 1] - \Pr[D(\mathcal{Y}) = 1]| \geq \varepsilon.$$

In Chapter 5, we consider two kinds of extractors: *hybrid* extractors and *computational* extractors for extracting computational min-entropy.

Definition 2.5.1. A function $\textsc{Ext} : \{0, 1\}^n \times \{0, 1\}^n \to \{0, 1\}^m$ is called a

- $(k_1, k_2, \varepsilon, s)$*-hybrid-extractor* if for any source $(\mathcal{V}|\mathcal{X})$ with $\mathrm{H}_c(\mathcal{V}|\mathcal{X}) \geq k_1$ and any source $\mathcal{W}$, independent of $(\mathcal{V}|\mathcal{X})$, with $\mathrm{H}_\infty(\mathcal{W}) \geq k_2$, there is no ε-distinguisher in $\mathsf{SIZE}(s)$ for the distributions $\mathcal{X} \circ \mathcal{W} \circ \textsc{Ext}(\mathcal{V}, \mathcal{W})$ and $\mathcal{X} \circ \mathcal{W} \circ \mathcal{U}_m$.

- $(k_1, k_2, \varepsilon, s)$*-computational-extractor* if for any source $(\mathcal{V}|\mathcal{X})$ with $\mathrm{H}_c(\mathcal{V}|\mathcal{X}) \geq k_1$ and any source $(\mathcal{W}|\mathcal{Y})$, independent of $(\mathcal{V}|\mathcal{X})$, with $\mathrm{H}_c(\mathcal{W}|\mathcal{Y}) \geq k_2$, there is no ε-distinguisher in $\mathsf{SIZE}(s)$ for the distributions $\mathcal{X} \circ \mathcal{Y} \circ \mathcal{W} \circ \textsc{Ext}(\mathcal{V}, \mathcal{W})$ and $\mathcal{X} \circ \mathcal{Y} \circ \mathcal{W} \circ \mathcal{U}_m$.

Remark 2.5.2. Note that if $\textsc{Ext}$ is a (k_1, k_2, ε)-strong-two-source-extractor then for any source $\mathcal{V}$ with $\mathrm{H}_\infty(\mathcal{V}) \geq k_1$ and any source $\mathcal{W}$, independent of $\mathcal{V}$, with $\mathrm{H}_\infty(\mathcal{W}) \geq k_2$, there is no ε-distinguisher (without any complexity bound) for the distributions $\mathcal{W} \circ \textsc{Ext}(\mathcal{V}, \mathcal{W})$ and $\mathcal{W} \circ \mathcal{U}_m$.

We will need the following fact about strong-two-source extractor.

Lemma 2.5.3. *Let* $\textsc{Ext} : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}^m$ *be any* (k_1, k_2, ε)*-strong-two-source-extractor. Then for any source* $\mathcal{W}$ *over* $\{0,1\}^n$ *with* $\mathrm{H}_\infty(\mathcal{W}) = k_2$ *and any function* $q : \{0,1\}^n \to \{0,1\}^m$*, there are at most* 2^{k_1} *different* v*'s satisfying*

$$\Pr_{w \in \mathcal{W}}[q(w) = \textsc{Ext}(v, w)] \geq 1/M + \varepsilon.$$

Proof. Let V be the set consisting of such v's and let $\mathcal{V}$ be the uniform distribution over V. Consider the distinguisher D defined as $D(w \circ u) = 1$ if $q(w) = u$ and $D(w \circ u) = 0$ otherwise. Then, we have

$$\begin{aligned}
& \Pr_{v \in \mathcal{V}, w \in \mathcal{W}}[D(w \circ \textsc{Ext}(v, w)) = 1] - \Pr_{w \in \mathcal{W}, u \in \mathcal{U}_m}[D(w \circ u) = 1] \\
= & \Pr_{v \in \mathcal{V}, w \in \mathcal{W}}[q(w) = \textsc{Ext}(v, w)] - \Pr_{w \in \mathcal{W}, u \in \mathcal{U}_m}[q(w) = u] \\
\geq & \ 1/M + \varepsilon - 1/M \\
= & \ \varepsilon.
\end{aligned}$$

This implies that $\log |V| = \mathrm{H}_\infty(\mathcal{V}) \leq k_1$, because otherwise it would contradict the fact that $\textsc{Ext}$ is a good strong-two-source-extractor. □

Finally, we will need the following lemma about obtaining predictors from distinguishers. The Boolean case ($m = 1$) is well known, and a proof for general m can be found in [21].

Lemma 2.5.4. *For any source* $\mathcal{Z}$ *over* $\{0,1\}^n$ *and any function* $b : \{0,1\}^n \to \{0,1\}^m$*, if there is an* ε*-distinguisher* E *for the distributions* $\mathcal{Z} \circ b(\mathcal{Z})$ *and* $\mathcal{Z} \circ \mathcal{U}_m$*, then there is a predictor* P *with* E *as oracle which calls* E *once and runs in time* $O(m)$ *such that*

$$\Pr_{z \in \mathcal{Z}}\left[P^E(z) = b(z)\right] \geq (1 + \varepsilon)/M.$$

2.6 Computational Independent-Symbol Sources

We generalized the definition of independent-symbol sources to consider independent-symbol sources with computational entropy in Chapter 6. Unlike the definition of computational entropy, we can have the circuit size as a separate parameter to define computational independent-symbol sources.

Definition 2.6.1. A distribution $(\mathcal{V}|\mathcal{X}) = (\mathcal{V}_1|\mathcal{X}_1) \circ \cdots \circ (\mathcal{V}_n|\mathcal{X}_n)$, where for each i, $\mathcal{V}_i$ is over the set $[D]$ and $\mathcal{X}_i$ is over $\{0,1\}^{\ell_i}$, is called a computational (n, D, k, s)-source if the n symbols $(\mathcal{V}_1|\mathcal{X}_1), \ldots, (\mathcal{V}_n|\mathcal{X}_n)$ are distributed independently from each other and for every $i \in [n]$, any circuit $C \in \mathsf{SIZE}(s)$, $\Pr[C(\mathcal{X}_i) = \mathcal{V}_i] \leq 2^{-k_i}$ for some $0 \leq k_i \leq \log D$, and $\sum_{i=1}^{n} k_i = k$.

Then, our target is to extract randomness from such computational independent-symbol sources.

Definition 2.6.2. A function $\textsc{Ext} : [D]^n \to [M]$ is called an $(n, D, k, \varepsilon, s_1, s_2)$ - *computational - extractor* if for any computational (n, D, k, s_1)-source, there is no ε-distinguisher in $\mathsf{SIZE}(s_2)$ for the distributions $\mathcal{X}_{[1,n]} \circ \textsc{Ext}(\mathcal{V}_1, \cdots, \mathcal{V}_n)$ and $\mathcal{X}_{[1,n]} \circ \mathcal{U}_{[M]}$, where $\mathcal{U}_{[M]}$ denotes the uniform distribution over the set $[M]$.

2.7 Some Useful Tools

In this section, we introduce some useful tools which will be used in this thesis. The first two lemmas are the well-known Jensen's inequality and Cauchy-Schwartz inequality.

Lemma 2.7.1 (Jensen's inequality [39]). *Let X be a random variable. If f is a convex function, then*

$$\mathrm{E}[f(X)] \geq f(\mathrm{E}[X]).$$

Lemma 2.7.2 (Cauchy-Schwartz inequality). *For any $x, y \in \mathbb{R}^n$,*

$$|\langle x, y \rangle| \leq \|x\|_2 \|y\|_2,$$

where $\langle x, y \rangle$ is the standard inner product of x and y.

In this thesis, we often need to bound the *tail distribution* [39], which is the probability of a random variable taking values far from its expectation. One useful tool is the *Markov's inequality*, when we only know the expectation of an nonnegative random variable.

Lemma 2.7.3 (Markov's inequality [39])**.** *Let X be a random variable that assumes only nonnegative values. Then, for all $a > 0$,*

$$\Pr[X \geq a] \leq \frac{\mathrm{E}[X]}{a}.$$

Sometimes, we will apply the following corollary, which is an application of Lemma 2.7.3.

Corollary 2.7.4. *Let X be a random variable taking values in the interval $[0,1]$. Then for $0 < \delta < \mathrm{E}[X]$,*

$$\Pr\left[X \geq \delta\right] \geq \mathrm{E}[X] - \delta.$$

Proof. Here we give two different proofs. The first one is by the Markov's inequality of Lemma 2.7.3, while the second one proves the corollary directly.

1. Define a random variable $Y = 1 - X$. Note that $\mathrm{E}[Y] = 1 - \mathrm{E}[X]$. Then,

$$\begin{aligned}
\Pr[X < \delta] &\leq \Pr[Y \geq 1 - \delta] \\
&\leq \frac{\mathrm{E}[Y]}{1-\delta} \qquad (2.2) \\
&= \frac{1 - \mathrm{E}[X]}{1-\delta} \\
&< 1 - (\mathrm{E}[X] - \delta),
\end{aligned}$$

   where (2.2) is due to the Markov's inequality of Lemma 2.7.3. Hence,

$$\Pr\left[X \geq \delta\right] = 1 - \Pr[X < \delta] \geq \mathrm{E}[X] - \delta.$$

2. Note that

$$\begin{aligned}
\mathrm{E}[X] &= \sum_x x \cdot \Pr[X = x] \\
&\leq \Pr[X \geq \delta] \cdot 1 + \Pr[X < \delta] \cdot \delta \\
&\leq \Pr[X \geq \delta] + \delta,
\end{aligned}$$

   which implies that $\Pr[X \geq \delta] \geq \mathrm{E}[X] - \delta$.

□

When we can bound the variance of the random variable, the following Chebyshev's inequality is useful.

Lemma 2.7.5 (Chebyshev's inequality [39])**.** *Let X be a random variable. For any positive a,*

$$\Pr[|X - \mathrm{E}[X]| \geq a] \leq \frac{\mathrm{Var}[X]}{a^2}.$$

Moreover, if the random variables are independent, then we can obtain a much better bound on the tail distribution of the sum of these random variables by Chernoff bound.

Lemma 2.7.6 (Chernoff bound [27])**.** *Let $X_1, \cdots, X_t$ be independent random variables taking values in the interval $[0,1]$. Let $X = \sum_{i=1}^{t} X_i$, and $\mu = \mathrm{E}[X]$. For any $0 < \lambda \leq 1$,*

$$\Pr[|X - \mu| \geq \lambda\mu] \leq 2 \cdot e^{-\lambda^2\mu/3}.$$

Chapter 3

Extracting Randomness from Multiple Independent Sources

In this chapter, we study the problem of deterministically extracting almost perfect random bits from multiple weakly random sources that are mutually independent. First, we prove the generalized leftover hash lemma in Section 3.1. Then we consider extracting randomness from two independent sources in Section 3.2, and extend this result to $t \geq 3$ independent sources in Section 3.3. Finally, we give an application in Section 3.4.

3.1 Generalized Leftover Hash Lemma

Definition 3.1.1. We call a family H of functions from $\{0,1\}^n$ to $\{0,1\}^m$ *pair-wise independent* if

$$\forall x_1 \neq x_2 : \quad \Pr_{h \in H}[h(x_1) = h(x_2)] = \frac{1}{2^m}.$$

The well-known leftover hash lemma [28] says that if h is sampled uniformly from such a family H and x is sampled from a distribution with enough min-entropy, the distribution of $h \circ h(x)$ is close to uniform. We extend it to the case that h is sampled from a large enough subset G of H. Note that the original leftover hash lemma is a special case of our lemma with $G = H$.

Lemma 3.1.2. *(Generalized leftover hash lemma) Let H be any family of pair-wise independent functions from $\{0,1\}^n$ to $\{0,1\}^m$. Let $\mathcal{G}$ denote the uniform distribution over* *a set $G \subseteq H$ and let $\mathcal{X}$ denote the uniform distribution over a set $X \subseteq \{0,1\}^n$. Then*

$$\Delta\left(\mathcal{G} \circ \mathcal{G}(\mathcal{X}), \mathcal{G} \circ \mathcal{U}_m\right) \leq \frac{1}{2}\sqrt{\frac{2^m|H|}{|X||G|}}.$$

Proof.

$$\begin{aligned}
4\Delta\left(\mathcal{G} \circ \mathcal{G}(\mathcal{X}), \mathcal{G} \circ \mathcal{U}_m\right)^2 &= \left(\sum_{h \in G} \sum_{z \in \{0,1\}^m} \frac{1}{|G|}\left|\Pr_{x \in X}[h(x) = z] - \frac{1}{2^m}\right|\right)^2 \\
&\leq \frac{2^m}{|G|}\left[\sum_{h \in H} \sum_{z \in \{0,1\}^m} \left(\Pr_{x \in X}[h(x) = z] - \frac{1}{2^m}\right)^2\right] \qquad (3.1) \\
&= \frac{2^m}{|G|}\left[\left(\sum_{h \in H} \sum_{z \in \{0,1\}^m} \Pr_{x, x' \in X}[h(x) = h(x') = z]\right) - \frac{|H|}{2^m}\right] \\
&= \frac{2^m|H|}{|G|}\left(\Pr_{h \in H; x, x' \in X}[h(x) = h(x')] - \frac{1}{2^m}\right) \\
&\leq \frac{2^m|H|}{|G|}\left(\Pr[x = x'] + \Pr[h(x) = h(x') \mid x \neq x'] - 1/2^m\right) \\
&\leq \frac{2^m|H|}{|G|}\left(\frac{1}{|X|} + \frac{1}{2^m} - \frac{1}{2^m}\right) \qquad (3.2) \\
&= \frac{2^m|H|}{|X||G|},
\end{aligned}$$

where (3.1) is due to the Jensen's inequality of Lemma 2.7.1, and (3.2) holds because H is pair-wise independent. □

3.2 Extracting from Two Independent Sources

We apply the generalized leftover hash lemma to extract randomness from two independent sources $\mathcal{X}$ and $\mathcal{Y}$ over $\{0,1\}^n$ with $\mathrm{H}_\infty(\mathcal{X}) \geq k_1$ and $\mathrm{H}_\infty(\mathcal{Y}) \geq k_2$. For any $n, m \in \mathbb{N}$ with $m|n$, let $\ell = \frac{n}{m}$, and we treat any $v \in \{0,1\}^n$ as an ℓ-dimensional vector $v = (v_1, v_2, \ldots, v_\ell)$ with each $v_i \in \mathcal{F} = GF(2^m)$. Now define our extractor $\textsc{Ext}^2 : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}^m$ as

$$\textsc{Ext}^2(x, y) = \langle x, y \rangle_m = \sum_{i=1}^{\ell} x_i \cdot y_i \in \mathcal{F},$$

which is the inner product of x and y over $\mathcal{F}$.

Theorem 3.2.1. *The function* $\textsc{Ext}^2$ *is a* (k_1, k_2, ε)*-strong-two-source-extractor with* $\varepsilon = 2^{-(k_1+k_2+2-n-m)/2}$.

Proof. Let $H = \{h_y \mid y \in \{0,1\}^n\}$, where $h_y(x) = \langle x, y\rangle_m$ for $x, y \in \{0,1\}^n$. It is easy to check that the family H is pair-wise independent. Then the theorem follows immediately from Lemma 3.1.2 and the fact that $\mathrm{H}_\infty(\mathcal{X}) \geq k_1$ implies $|X| \geq 2^{k_1}$. □

3.3 Extracting from t Independent Sources

Next, we show how to extract randomness from t independent sources $\mathcal{X}_1, \ldots, \mathcal{X}_t$. Define the extractor $\textsc{Ext}^t : (\{0,1\}^n)^t \to \{0,1\}^m$ as

$$\textsc{Ext}^t(x_1, \ldots, x_t) = \sum_{1\leq i<j\leq t} \langle x_i, x_j\rangle_m.$$

Theorem 3.3.1. *The function* $\textsc{Ext}^t$ *is a* $(k_1, \ldots, k_t, \varepsilon)$*-strong-multi-source-extractor with* $\varepsilon = 2^{-(k_1+k+2-n-m)/2}$*, where* $k = \max\{k_2, \ldots, k_t\}$.

Proof. Assume without loss of generality that $\mathcal{X}_2$ is the source with the largest min-entropy among $\mathcal{X}_2, \ldots, \mathcal{X}_t$. Fix any values $x_3, \ldots, x_t$, let $s = x_3 + \cdots + x_t$, and let $\alpha = \sum_{3\leq i<j\leq t}\langle x_i, x_j\rangle_m$. Then

$$\textsc{Ext}^t(\mathcal{X}_1, \mathcal{X}_2, x_3, \ldots, x_t) = \langle \mathcal{X}_1, \mathcal{X}_2\rangle_m + \langle \mathcal{X}_1, s\rangle_m + \langle \mathcal{X}_2, s\rangle_m + \alpha.$$

Consider the family of functions $H = \{h_y \mid y \in \{0,1\}^n\}$ where $h_y(x) = \langle x, y\rangle_m + \langle x, s\rangle_m + \langle y, s\rangle_m + \alpha$. It is pair-wise independent because for any $x \neq x'$,

$$\begin{aligned}
& \Pr_y[h_y(x) = h_y(x')] \\
= & \Pr_y[\langle x, y\rangle_m + \langle x, s\rangle_m = \langle x', y\rangle_m + \langle x', s\rangle_m] \\
= & \Pr_y[\langle x - x', y\rangle_m = \langle x' - x, s\rangle_m] \\
= & \frac{1}{2^m}.
\end{aligned}$$

Therefore, Theorem 3.2.1 implies for any $x_3, \ldots, x_t$,

$$\Delta(\mathcal{X}_2 \circ \textsc{Ext}^t(\mathcal{X}_1, \mathcal{X}_2, x_3, \ldots, x_t), \mathcal{X}_2 \circ \mathcal{U}_m) \leq 2^{-(k_1+k_2+2-n-m)/2}.$$

Thus

$$\begin{aligned}
& \Delta\left(\mathcal{X}_{[2,t]} \circ \textsc{Ext}^t(\mathcal{X}_1,\ldots,\mathcal{X}_t), \mathcal{X}_{[2,t]} \circ \mathcal{U}_m\right) \\
\leq & \sum_{x_{[3,t]}} \left\{ \Pr\left[X_{[3,t]} = x_{[3,t]}\right] \cdot \Delta\left(\mathcal{X}_2 \circ \textsc{Ext}^t(\mathcal{X}_1, \mathcal{X}_2, x_3, \ldots, x_t), \mathcal{X}_2 \circ \mathcal{U}_m\right) \right\} \\
\leq & \; 2^{-(k_1+k_2+2-n-m)/2}.
\end{aligned}$$

$\square$

If the sources are not exposed, we can have a slightly better result. Note that

$$\Delta\left(\textsc{Ext}^t(\mathcal{X}_1,\ldots,\mathcal{X}_t), \mathcal{U}_m\right) \leq \Delta\left(\mathcal{X}_{[2,t]} \circ \textsc{Ext}^t(\mathcal{X}_1,\ldots,\mathcal{X}_t), \mathcal{X}_{[2,t]} \circ \mathcal{U}_m\right),$$

so by taking $\mathcal{X}_1$ in Theorem 3.3.1 to be the source with the highest min-entropy, we have the following.

Corollary 3.3.2. *The function* $\textsc{Ext}^t$ *is a* $(k_1,\ldots,k_t,\varepsilon)$*-extractor with* $\varepsilon = 2^{-(b_1+b_2+2-n-m)/2}$, *where* b_1 *and* b_2 *are the two largest values among* $k_1,\ldots,k_t$.

Note that in the construction of our extractor $\textsc{Ext}^t$, we do not need to know beforehand the specific min-entropy of each source. It works as long as the sum of the two largest min-entropies is large enough.

3.4 Application

Consider the following cryptographic setting in which a group of parties $P_1, P_2, \ldots, P_u$ want to establish a secret key for group communication. Suppose initially these parties are together and can sample $\mathcal{A}_1, \mathcal{B}_1, \ldots, \mathcal{A}_u, \mathcal{B}_u, \mathcal{X}$ from some block-wise source [10], where each block ($\mathcal{A}_i$, $\mathcal{B}_i$, or $\mathcal{X}$) is n-bit long and has min-entropy at least k even given all the previous blocks. After that, all parties go far away from each other but are connected by an insecure network. If they want to communicate securely later on, they can execute the following protocol:

1. In the order of i from 1 to u, party P_i samples $\mathcal{X}_i$ from his/her own local source, computes $\mathcal{Y}_i = \mathcal{A}_i\mathcal{X}_i + \mathcal{B}_i$, and sends $(\mathcal{X}_i, \mathcal{Y}_i)$ to all other parties.

2. When receiving $(\tilde{\mathcal{X}}_j, \tilde{\mathcal{Y}}_j)$ from an alleged party P_j, each P_i verifies whether $\tilde{\mathcal{Y}}_j = \mathcal{A}_j \tilde{\mathcal{X}}_j + \mathcal{B}_j$. Let $T = \{P_{i_1}, P_{i_2}, \ldots, P_{i_{t-1}}\}$ be the set of parties who pass this authentication test.

3. Each party in T computes the secret key $\mathcal{K} = \textsc{Ext}^t(\mathcal{X}, \mathcal{X}_{i_1}, \mathcal{X}_{i_2}, \ldots, \mathcal{X}_{i_{t-1}})$, which can be used, for example, as the secret key of the one-time pad encryption.

We discuss two security issues. For authentication, we know that after seeing $(\mathcal{X}_i, \mathcal{Y}_i)$ for every $i < j$, $(\mathcal{A}_j, \mathcal{B}_j)$ still has min-entropy $2k$, so from [14], an adversary can only impersonate a party P_j with probability $2^{-(2k-n)}$. For the security of $\mathcal{K}$, note that $\mathcal{X}$ is assumed to have min-entropy k even given $\mathcal{A}_1, \mathcal{B}_1, \ldots, \mathcal{A}_u, \mathcal{B}_u$, and we can assume that $\mathcal{X}, \mathcal{X}_{i_1}, \mathcal{X}_{i_2}, \ldots, \mathcal{X}_{i_{t-1}}$ are mutually independent as they are generated in distant places. Thus, Theorem 3.3.1 implies $\Delta(\mathcal{X}_{i_1} \circ \cdots \circ \mathcal{X}_{i_{t-1}} \circ \mathcal{K}, \mathcal{X}_{i_1} \circ \cdots \circ \mathcal{X}_{i_{t-1}} \circ \mathcal{U}_m) \leq 2^{-(k+b+2-n-m)/2}$, where $b = \max\{\mathrm{H}_\infty(\mathcal{X}_{i_1}), \ldots, \mathrm{H}_\infty(\mathcal{X}_{i_{t-1}})\}$. That is, $\mathcal{K}$ is secure enough when $k + b \gg n + m$. Note that any strong extractor will also work.

Chapter 4

Deterministic Extractors for Independent-Symbol Sources

In this chapter, we consider (n, D, k)-sources which consist of a sequence of n independent symbols from $[D]$, and the only randomness guarantee on such a source is that the whole source has min-entropy k. First, we give an explicit deterministic extractor which extracts about $\log k$ bits in Section 4.1. Then, in Section 4.2, we use the technique of Gabizon et al. [17] to extract more random bits. Moreover, in Section 4.3, we show the existence of a non-explicit deterministic extractor which can extract $m = k - O(\log(1/\varepsilon))$ bits with error ε 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 $\Omega(\log(1/\varepsilon))$ in Section 4.4.

4.1 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[1].

Theorem 4.1.1. *For any $n, k, D \in \mathbb{N}$ and any prime number $M \geq D$, there is an explicit (n, D, k, ε)-extractor* $\textsc{Ext}_0 : [D]^n \to [M]$, *with* $\varepsilon \leq \frac{1}{2} \cdot \sqrt{M} \cdot 2^{-\Omega(k/M^2)}$.

[1] For this extractor, we provide a different and completely elementary proof in Appendix B, which achieves the same asymptotic bound as [32].

Note that for $k \geq \Omega(M^2 \log M)$, our extractor has $\varepsilon \leq 2^{-\Omega(k/M^2)}$. Moreover, Theorem 4.1.1 achieves the same bound as [32], but here we provide a proof from a different point of view.

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 $\textsc{Ext}_0 : [D]^n \to [M]$ is then defined as $\textsc{Ext}_0(\mathcal{X}) = \sum_i \mathcal{X}_i$, 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 [33] for bit-fixing sources can be seen as a special case of ours, with $D = 2$ and $\mathcal{X}_i \in \{-1, 1\}$.

To prove Theorem 4.1.1, we need the following lemma in [32], which implies that we only need to consider flat independent-symbol sources.

Lemma 4.1.2. *[32] Any (n, D, k)-source is $e^{-k/9}$-close to a convex combination of flat $(n, D, k/(2 \log 3))$-sources.*

Now, assume that $\mathcal{X}$ is a flat (n, D, k)-source. Then, as in [33], 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. More precisely, suppose the current distribution is $\mathcal{P} = [\mathcal{P}_1, \ldots, \mathcal{P}_M]^T$ and the next symbol in the source has a distribution $\beta = [\beta_1, \ldots, \beta_M]^T$ (let $\beta_i = 0$ for $D + 1 \leq i \leq M$). Then the next distribution is $\bar{\mathcal{P}} = [\bar{\mathcal{P}}_1, \ldots, \bar{\mathcal{P}}_M]^T$ 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 4.1.3. *If the symbol is a flat source with distribution β, we have*

$$\|\bar{\delta}\|_2 \leq \|\delta\|_2 \cdot e^{-\frac{2^{2\mathrm{H}_\infty(\beta)}-1}{M^2-1}}.$$

We will prove this lemma in Subsection 4.1.1. Now let us see how it can be used to prove the theorem.

Proof. (of Theorem 4.1.1) By Lemma 4.1.2, it is sufficient to consider flat (n, D, k)-sources. Given a flat (n, D, k)-source $\mathcal{X} = \mathcal{X}_1 \circ \cdots \circ \mathcal{X}_n$, for $t \in [n]$, let $K_t = 2^{\mathrm{H}_\infty(\mathcal{X}_t)}$. According to Lemma 4.1.3, we have

$$\|\textsc{Ext}_0(\mathcal{X}) - \mathcal{U}\|_2^2 \le \prod_{t\in[n]} e^{-2\cdot\frac{K_t^2-1}{M^2-1}} < e^{-\frac{2\sum_t(K_t^2-1)}{M^2}} < e^{-\frac{2n(2^{2k/n}-1)}{M^2}},$$

where the last inequality is due to the power mean inequality, $\sum_t K_t^2 \ge n2^{2k/n}$. Let $k = \delta n$. Then we obtain that

$$\|\textsc{Ext}_0(\mathcal{X}) - \mathcal{U}\|_2^2 < e^{-\frac{2n(2^{2k/n}-1)}{M^2}} = e^{\frac{2k}{M^2}\cdot\frac{2^{2\delta}-1}{\delta}}.$$

Since $(2^{2\delta} - 1)/\delta > 2\ln 2$ for $\delta > 0$, we have

$$\|\textsc{Ext}_0(\mathcal{X}) - \mathcal{U}\|_2^2 < e^{\frac{2k}{M^2}\cdot\frac{2^{2\delta}-1}{\delta}} < e^{-\frac{4(\ln 2)k}{M^2}} = 2^{-4k/M^2}.$$

Hence, after n steps, the L_2-distance from the uniform distribution is at most $2^{-2k/M^2}$. Then the theorem follows from the Cauchy-Schwartz inequality of Lemma 2.7.2 and Lemma 4.1.2. □

4.1.1 Independent-symbol Sources and Circulant Matrices

In this subsection, we will prove Lemma 4.1.3 according to the relation between independent-symbol sources and circulant matrices [12], where a circulant matrix is a square matrix with the following form

$$\begin{aligned} C &= \texttt{circ}[c_1, c_2, \cdots, c_n] \\ &= \begin{pmatrix} c_1 & c_2 & \cdots & c_n \\ c_n & c_1 & \cdots & c_{n-1} \\ \vdots & & \cdots & \vdots \\ c_2 & c_3 & \cdots & c_1 \end{pmatrix}, \end{aligned}$$

where each row is a cyclic shift of the previous row. We observe that the transition matrix of each step is a circulant matrix, where the transition matrix is an $M \times M$ matrix, in which the ith row and jth column element is the probability of going to the state i from state j. That is, let $B = \texttt{circ}[\beta_M, \beta_{M-1}, \ldots, \beta_1]$ be a circulant matrix. Then $\bar{\mathcal{P}} = B\mathcal{P}$. Due to the special structure of circulant matrices, we can obtain their eigenvalues and corresponding eigenvectors.

Theorem 4.1.4. *[12] For $j \in [n]$, the j-th eigenvalue of a circulant matrix $C = \texttt{circ}[c_0, c_1, \cdots, c_{n-1}]$ is*

$$\lambda_j(C) = \sum_{\ell=0}^{n-1} c_\ell \cdot \omega_n^{j\ell},$$

where $\omega_n = e^{\frac{2\pi i}{n}}$, and the corresponding eigenvector is $\mathcal{V}_n^{(j)} = [1, \omega_n^j, \omega_n^{2j}, \cdots, \omega_n^{(n-1)j}]^T$.

Note that the eigenvectors of a circulant matrix are orthogonal. Hence, we can obtain the following lemma which shows that the progress we can make after each step is corresponding to the second largest (in absolute value) eigenvalue of the transition matrix.

Lemma 4.1.5.

$$\|\bar{\delta}\|_2 \leq \lambda(B) \cdot \|\delta\|_2,$$

where $\lambda(B)$ is the second largest (in absolute value) eigenvalue of B.

Proof. Since the vector $\mathcal{U}$ is invariant under B in the sense that $B\mathcal{U} = \mathcal{U}$ and the vector δ is orthogonal to $\mathcal{U}$, we have

$$\|\bar{\delta}\|_2 = \|B\mathcal{P} - \mathcal{U}\|_2 = \|B(\mathcal{P} - \mathcal{U})\|_2 \leq \lambda(B) \cdot \|\mathcal{P} - \mathcal{U}\|_2 = \lambda(B) \cdot \|\delta\|_2.$$

□

Let $k_\beta = \mathrm{H}_\infty(\beta)$. Recall that the distribution β is flat. Hence, there are $K_\beta = 2^{k_\beta}$ nonzero entries in each row of B, and each nonzero entry is 2^{-k_β}. According to

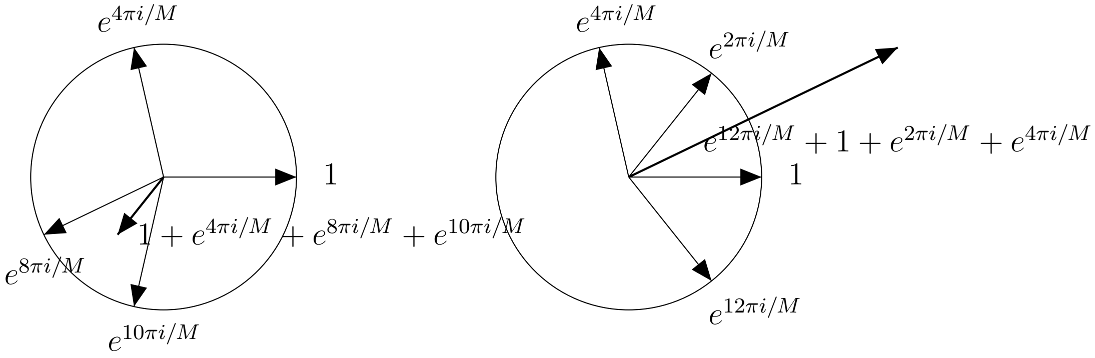
$e^{4\pi i/M}$
1
$1 + e^{4\pi i/M} + e^{8\pi i/M} + e^{10\pi i/M}$
$e^{8\pi i/M}$
$e^{10\pi i/M}$
$e^{4\pi i/M}$
$e^{2\pi i/M}$
$e^{12\pi i/M} + 1 + e^{2\pi i/M} + e^{4\pi i/M}$
1
$e^{12\pi i/M}$

Figure 4.1: Let $M = 7$, and $k_\beta = 2$. The left diagram shows the sum of 1, $e^{4\pi i/M}$, $e^{8\pi i/M}$, and $e^{10\pi i/M}$, while the right diagram shows the sum of $e^{12\pi i/M}$, 1, $e^{2\pi i/M}$, and $e^{4\pi i/M}$. In these diagrams, it is easy to see that the maximum occurs when these 4 nonzero entries are consecutive.

Theorem 4.1.4, the square of $\lambda(B)$ is

$$\begin{aligned}
\max_{j\neq 0} \lambda_j(B)^2 &\leq \left| 2^{-k_\beta} + 2^{-k_\beta} \cdot e^{\frac{2\cdot\pi i}{M}} + \cdots + 2^{-k_\beta} \cdot e^{\frac{2\cdot(K_\beta-1)\pi i}{M}} \right|^2 \\
&= \left| 2^{-k_\beta} \cdot \sum_{\ell=0}^{K_\beta-1} e^{\frac{2\pi i \ell}{M}} \right|^2 = \left| 2^{-k_\beta} \cdot \frac{1-e^{\frac{2\cdot K_\beta \pi i}{M}}}{1-e^{\frac{2\pi i}{M}}} \right|^2 = \left| 2^{-k_\beta} \cdot \frac{e^{\frac{K_\beta \pi i}{M}} \left(e^{-\frac{K_\beta \pi i}{M}} - e^{\frac{K_\beta \pi i}{M}} \right)}{e^{\frac{\pi i}{M}} \left(e^{-\frac{\pi i}{M}} - e^{\frac{\pi i}{M}} \right)} \right|^2 \\
&= \left(2^{-k_\beta} \cdot \frac{\sin\left(\frac{\pi K_\beta}{M}\right)}{\sin\left(\frac{\pi}{M}\right)} \right)^2 = \left(2^{-k_\beta} \frac{\frac{\pi K_\beta}{M} \prod_{\ell=1}^{\infty} \left(1 - \frac{K_\beta^2}{M^2 \ell^2}\right)}{\frac{\pi}{M} \prod_{\ell=1}^{\infty} \left(1 - \frac{1}{M^2\ell^2}\right)} \right)^2 \\
&= \left(\prod_{\ell=1}^{\infty} \left(1 - \frac{K_\beta^2 - 1}{M^2 \ell^2 - 1} \right) \right)^2 < \left(1 - \frac{K_\beta^2 - 1}{M^2 - 1} \right)^2 < e^{-2\left(2^{2k_\beta}-1\right)/\left(M^2-1\right)},
\end{aligned}$$

where the first inequality is because it is easy to see in the diagram (see Figure 4.1.1 as an example) that the maximum occurs when these K_β nonzero entries are consecutive, and the fourth equality is due to the infinite product representation of sine [22].

4.2 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 [17]. 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 4.2.1. *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* $\textsc{Ext} : [D]^n \to \{0,1\}^m$ *with $m \geq k - O(d \log(1/\varepsilon))$.*

Theorem 4.2.2. *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* $\textsc{Ext} : [D]^n \to \{0,1\}^m$ *with $m \geq k - O((1/\varepsilon)^{c_3})$.*

Note that the two main results in [17] only work for bit-fixing sources (with $D = 2$) and follow respectively from Theorem 4.2.1 with $\varepsilon = 2^{-cn^{\gamma}}$ and $m = k - O(n^{\gamma})$, and from Theorem 4.2.2 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 4.2.3 and 4.2.4 respectively, which follow closely the corresponding ones in [17]. The main difference is that we consider independent-symbol sources, so we cannot build on the extractor of [33] as [17] did, and instead, we build on our extractor in Theorem 4.1.1. 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.

4.2.1 Construction

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 [17] by constructing the so-called seed obtainer. The idea is to divide z into two parts $w \circ y$ 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 4.2.3. Suppose $n, d, k \in \mathbb{N}$, $\delta \in (0,1)$, and $k_{\text{min}}, k_{\text{max}} \in \mathbb{R}$, with $0 \le k_{\text{min}} \le k_{\text{max}} \le k \le n$. 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 $\sum_{i\in[n]} h(i) = k$,

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

Throughout this section, for an (n, D)-source $\mathcal{X} = \mathcal{X}_1 \circ \cdots \circ \mathcal{X}_n$, we will let h be the function such that $h(i) = \mathrm{H}_\infty(\mathcal{X}_i) \in [0, d]$. Note that the definition of samplers used in [17] 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 [17], 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[2].

Lemma 4.2.4. *Suppose there exist explicit constructions for the following three ingredients: (1) a seedless $(n, D, k_{\text{min}}, \varepsilon_1)$-extractor* $\text{E}\textsc{xt}_1 : [D]^n \to \{0,1\}^{t+s}$*, (2) an $(n, d, k, k_{\text{min}}, k_{\text{max}}, \delta)$-sampler* $\text{S}\textsc{amp} : \{0,1\}^t \to P([n])$*, and (3) a seeded $(n, D, k - k_{\text{max}}, \varepsilon_2)$-extractor* $\text{E}\textsc{xt}_2 : [D]^n \times \{0,1\}^s \to \{0,1\}^m$*. Then there exists an explicit seedless (n, D, k, ε_3)-extractor* $\text{E}\textsc{xt}_3 : [D]^n \to \{0,1\}^m$ *with $\varepsilon_3 = 3\max\{\varepsilon_1 + \delta, 2^{t+1}\varepsilon_1\} + \varepsilon_2$.*

To prove Lemma 4.2.4, we need the following two lemmas[3].

Lemma 4.2.5. *Let $\mathcal{A} \circ \mathcal{B}$ be a random variable over $[D]^u \times \{0,1\}^v$ and for every $a \in [D]^u$ with $\Pr[\mathcal{A} = a] > 0$, the distribution $(\mathcal{B}|\mathcal{A} = a)$ is ε-random. Then $\Delta(\mathcal{A} \circ \mathcal{B}, \mathcal{A} \circ \mathcal{U}_v) \le \varepsilon$, where $\mathcal{U}_v$ is independent of $\mathcal{A}$.*

[2] Note that this was proved in [17] for bit-fixing sources.

[3] Note that these two lemmas are generalizations of the corresponding ones in [17].

Proof.

$$\begin{aligned}
& \Delta(\mathcal{A}\circ\mathcal{B},\mathcal{A}\circ\mathcal{U}_v) \\
= \ & \frac{1}{2}\sum_{a\in[D]^u,b\in\{0,1\}^v} |\Pr[\mathcal{A}\circ\mathcal{B}=a\circ b]-\Pr[\mathcal{A}\circ\mathcal{U}_v=a\circ b]| \\
= \ & \frac{1}{2}\sum_{a\in[D]^u,b\in\{0,1\}^v} |\Pr[\mathcal{A}=a]\cdot\Pr[(\mathcal{B}|\mathcal{A}=a)=b]-\Pr[\mathcal{A}=a]\cdot\Pr[\mathcal{U}_v=b]| \\
= \ & \sum_{a\in[D]^u}\Pr[\mathcal{A}=a]\cdot\left(\frac{1}{2}\sum_{b\in\{0,1\}^v}|\Pr[(\mathcal{B}|\mathcal{A}=a)=b]-\Pr[\mathcal{U}_v=b]|\right) \\
= \ & \sum_{a\in[D]^u}\Pr[\mathcal{A}=a]\cdot\Delta((\mathcal{B}|\mathcal{A}=a),\mathcal{U}_v) \\
\leq \ & \varepsilon.
\end{aligned}$$

□

Lemma 4.2.6. *Let $\mathcal{A}\circ\mathcal{B}$ be a random variable over $S\times\{0,1\}^v$ for some set S. If $\Delta(\mathcal{A}\circ\mathcal{B},\mathcal{A}'\circ\mathcal{U}_v)\leq\varepsilon$ where $\mathcal{U}_v$ is independent of $\mathcal{A}'$, then, for every $b\in\{0,1\}^v$, $\Delta((\mathcal{A}|\mathcal{B}=b),\mathcal{A}')\leq\varepsilon\cdot 2^{v+1}$.*

Proof. By way of contradiction, assume that there exists $b\in\{0,1\}^v$ such that $\Delta((\mathcal{A}|\mathcal{B}=b),\mathcal{A}')>\varepsilon\cdot 2^{v+1}$, which implies that there exists $T\subseteq S$ such that

$$|\Pr[(\mathcal{A}|\mathcal{B}=b)\in T]-\Pr[\mathcal{A}'\in T]|>\varepsilon\cdot 2^{v+1}.$$

Without loss of generality, we can assume that

$$\Pr[(\mathcal{A}|\mathcal{B}=b)\in T]-\Pr[\mathcal{A}'\in T]>\varepsilon\cdot 2^{v+1},$$

because otherwise we can consider the set $S\setminus T$.

Since $\mathcal{U}_v$ and $\mathcal{A}'$ are independent, we have that

$$\Pr[\mathcal{A}'\circ\mathcal{U}_v\in T\circ b]=\Pr[\mathcal{U}_v=b]\cdot\Pr[\mathcal{A}'\in T]=2^{-v}\cdot\Pr[\mathcal{A}'\in T].$$

On the other hand, since $\mathcal{B}$ is ε-random, we know that $\Pr[\mathcal{B}=b]\geq 2^{-v}-\varepsilon$. Therefore, we have that

$$\begin{aligned}
& \Pr[\mathcal{A} \circ \mathcal{B} \in T \circ b] - \Pr[\mathcal{A}' \circ \mathcal{U}_v \in T \circ b] \\
= \; & \Pr[\mathcal{B} = b] \cdot \Pr[(\mathcal{A}|\mathcal{B} = b) \in T] - \Pr[\mathcal{U}_v = b] \cdot \Pr[\mathcal{A}' \in T] \\
\geq \; & (2^{-v} - \varepsilon) \Pr[(\mathcal{A}|\mathcal{B} = b) \in T] - 2^{-v} \Pr[\mathcal{A}' \in T] \\
\geq \; & 2^{-v}(\Pr[(\mathcal{A}|\mathcal{B} = b) \in T] - \Pr[\mathcal{A}' \in T]) - \varepsilon \\
> \; & 2^{-v} \cdot \varepsilon 2^{v+1} - \varepsilon \\
= \; & \varepsilon,
\end{aligned}$$

which contradicts the assumption that $\Delta(\mathcal{A} \circ \mathcal{B}, \mathcal{A}' \circ \mathcal{U}_v) \leq \varepsilon$. □

We also need the following lemma.

Lemma 4.2.7. *[17] Let $\mathcal{A}, \mathcal{B}$ and $\mathcal{C}$ be distributions over $\{0,1\}^n$ such that $\mathcal{A}$ is ε-random and $\mathcal{C} = \delta \cdot \mathcal{B} + (1 - \delta) \cdot \mathcal{A}$. Then $\mathcal{C}$ is $(2\delta + \varepsilon)$-random.*

Proof. For any $T \subseteq \{0,1\}^n$,

$$\begin{aligned}
|\Pr[\mathcal{C} \in T] - \Pr[\mathcal{U}_n \in T]| & = |\delta \cdot \Pr[\mathcal{B} \in T] + (1 - \delta) \cdot \Pr[\mathcal{A} \in T] - \Pr[\mathcal{U}_n \in T]| \\
& \leq 2\delta + |\Pr[\mathcal{A} \in T] - \Pr[\mathcal{U}_n \in T]| \\
& \leq 2\delta + \varepsilon.
\end{aligned}$$

□

Now we proceed to prove Lemma 4.2.4.

Proof. (of Lemma 4.2.4) Recall that we construct $\textsc{Ext}_3 : [D]^n \rightarrow \{0,1\}^m$ as follows:

1. Given an input source $\mathcal{X} = \mathcal{X}_1 \circ \cdots \circ \mathcal{X}_n$, compute $\mathcal{Z} = \textsc{Ext}_1(\mathcal{X})$.

2. Divide $\mathcal{Z}$ into two parts $\mathcal{W} \circ \mathcal{Y} \in \{0,1\}^t \times \{0,1\}^s$.

3. Compute $\textsc{Samp}(\mathcal{W})$.

4. Let $\mathcal{X}'$ be the source over $[D]^n$ obtained by padding $\mathcal{X}_{[n] \setminus \textsc{Samp}(\mathcal{W})}$ with zeros.

5. Output $\textsc{Ext}_2(\mathcal{X}', \mathcal{Y})$.

Next, we prove the correctness of the construction. For an (n, D, k)-source $\mathcal{X} = \mathcal{X}_1 \circ \cdots \circ \mathcal{X}_n$, define the function $h : [n] \to [0, d]$ by $h(i) = H_\infty(\mathcal{X}_i)$. Note that $\sum_{i\in[n]} h(i) = k$. Since Ext_1 is a seedless $(n, D, k_{\min}, \varepsilon_1)$-extractor, we have that $\mathcal{Z}$ is ε_1-random, which implies that both $\mathcal{W}$ and $\mathcal{Y}$ are ε_1-random.

We say that $w \in \{0,1\}^t$ is *good* if $k_{\min} \leq \sum_{i\in\text{Samp}(w)} h(i) \leq k_{\max}$ and *bad* otherwise. Since Samp is an $(n, d, k, k_{\min}, k_{\max}, \delta)$-sampler and $\mathcal{W}$ is ε_1-random, we obtain that

$$\Pr[\mathcal{W} \text{ is bad}] \leq \gamma,$$

for some $\gamma \leq \varepsilon_1 + \delta$.

Note that we can write the output source $\mathcal{R} = \mathcal{X}' \circ \mathcal{Y}$ as a convex combination

$$\mathcal{R} = \mathcal{B} \cdot \Pr[\mathcal{W} \text{ is bad}] + \sum_{w \text{ is good}} \mathcal{R}_w \cdot \Pr[\mathcal{W} = w]$$

where $\mathcal{B} = (\mathcal{R}|\mathcal{W} \text{ is bad})$ and $\mathcal{R}_w = (\mathcal{R}|\mathcal{W} = w)$. The following claim shows that for a good w, $\mathcal{R}_w$ is close to a pair of random variables $\mathcal{A}_w \circ \mathcal{U}_s$ where $\mathrm{H}_\infty(\mathcal{A}_w) \geq k - k_{\max}$ and $\mathcal{A}_w$ is independent of $\mathcal{U}_s$.

Claim 4.2.8. For any good $w \in \{0,1\}^t$, there exists a pair of random variables $\mathcal{A}_w \circ \mathcal{U}_s$ where $\mathrm{H}_\infty(\mathcal{A}_w) \geq k - k_{\max}$ and $\mathcal{A}_w$ is independent of $\mathcal{U}_s$, such that

$$\Delta(\mathcal{R}_w, \mathcal{A}_w \circ \mathcal{U}_s) \leq \varepsilon_1 \cdot 2^{t+1}.$$

Proof. Fix a good $w \in \{0,1\}^t$. Set $\ell = |\text{Samp}(w)|$. For two strings $\rho_1 \in [D]^\ell$ and $\rho_2 \in [D]^{n-\ell}$, let $[\rho_1; \rho_2] \in [D]^n$ denote the unique string obtained by putting ρ_1 in the indices of $\text{Samp}(w)$ and ρ_2 in the indices of $[n] \setminus \text{Samp}(w)$. Let $\mathcal{X}_w = \mathcal{X}_{\text{Samp}(w)}$ and $\overline{\mathcal{X}_w} = \mathcal{X}_{[n]\setminus\text{Samp}(w)}$. Note that $\mathcal{X} = [\mathcal{X}_w; \overline{\mathcal{X}_w}]$. Then for every $b \in [D]^{n-\ell}$ such that $\Pr[\overline{\mathcal{X}_w} = b] > 0$,

$$\left(\mathcal{Z} \,\middle|\, \overline{\mathcal{X}_w} = b\right) = \left(\text{Ext}_1(\mathcal{X}) \,\middle|\, \overline{\mathcal{X}_w} = b\right) = \left(\text{Ext}_1\left([\mathcal{X}_w; \overline{\mathcal{X}_w}]\right) \middle|\, \overline{\mathcal{X}_w} = b\right) = \text{Ext}_1\left([\mathcal{X}_w; b]\right),$$

where the last equality is because $\mathcal{X}$ is an (n, D)-source which implies that $\mathcal{X}_w$ and $\overline{\mathcal{X}_w}$ are independent. Since w is good, we have that $H_\infty(\mathcal{X}_w) = \sum_{i\in\text{Samp}(w)} h(i) \geq k_{\min}$ and $\mathrm{H}_\infty(\overline{\mathcal{X}_w}) \geq k - k_{\max}$. Hence, the source $[\mathcal{X}_w; b]$ is an (n, D)-source with min-entropy

$\mathrm{H}_\infty(\mathcal{X}_w) \geq k_{\min}$, and then $\Delta(\textsc{Ext}_1([\mathcal{X}_w; b]), \mathcal{U}_{t+s}) \leq \varepsilon_1$, which implies that $\Delta((\mathcal{Z}|\overline{\mathcal{X}_w} = b), \mathcal{U}_{t+s}) \leq \varepsilon_1$. By Lemma 4.2.5, we have that for a good w,

$$\Delta(\overline{\mathcal{X}_w} \circ \mathcal{Z}, \overline{\mathcal{X}_w} \circ \mathcal{U}_{t+s}) \leq \varepsilon_1.$$

Then by Lemma 4.2.6, we obtain that for a good w,

$$\Delta((\overline{\mathcal{X}_w}, \mathcal{Y})|\mathcal{W} = w), (\overline{\mathcal{X}_w}, \mathcal{U}_s)) \leq \varepsilon_1 \cdot 2^{t+1},$$

which implies that $\Delta(\mathcal{R}_w, \mathcal{A}_w \circ \mathcal{U}_s) \leq \varepsilon_1 \cdot 2^{t+1}$, where $\mathcal{A}_w$ is the source over $[D]^n$ obtained by padding $\overline{X_w}$ with zeros. □

Recall that $\mathcal{R} = \mathcal{B} \cdot \Pr[\mathcal{W} \text{ is bad}] + \sum_{w \text{ is good}} \mathcal{R}_w \cdot \Pr[\mathcal{W} = w]$. Then,

$$\textsc{Ext}_2(\mathcal{R}) = \Pr[\mathcal{W} \text{ is bad}] \cdot \textsc{Ext}_2(\mathcal{B}) + \sum_{w \text{ is good}} \Pr[\mathcal{W} = w] \cdot \textsc{Ext}_2(\mathcal{R}_w).$$

For any good w, by Claim 4.2.8, there exists $\mathcal{A}_w \circ \mathcal{U}_s$ where $\mathrm{H}_\infty(\mathcal{A}_w) \geq k - k_{\max}$ and $\mathcal{A}_w$ is independent of $\mathcal{U}_s$, such that $\Delta(\mathcal{R}_w, \mathcal{A}_w \circ \mathcal{U}_s) \leq \varepsilon_1 \cdot 2^{t+1}$. Then by Proposition 2.1.3, we have

$$\Delta(\textsc{Ext}_2(\mathcal{R}_w), \textsc{Ext}_2(\mathcal{A}_w, \mathcal{U}_s)) \leq \varepsilon_1 \cdot 2^{t+1}.$$

Moreover, since $\mathrm{H}_\infty(\mathcal{A}_w) \geq k - k_{\max}$ and $\mathcal{A}_w$ is independent of $\mathcal{U}_s$, we have that

$$\Delta(\textsc{Ext}_2(\mathcal{A}_w, \mathcal{U}_s), \mathcal{U}_m) \leq \varepsilon_2.$$

By Proposition 2.1.2, we obtain that

$$\begin{aligned} \Delta(\textsc{Ext}_2(\mathcal{R}_w), \mathcal{U}_m) &\leq \Delta(\textsc{Ext}_2(\mathcal{R}_w), \textsc{Ext}_2(\mathcal{A}_w, \mathcal{U}_s)) + \Delta(\textsc{Ext}_2(\mathcal{A}_w, \mathcal{U}_s), \mathcal{U}_m) \\ &\leq \varepsilon_1 \cdot 2^{t+1} + \varepsilon_2. \end{aligned}$$

Hence, by Lemma 4.2.7, we have that

$$\begin{aligned} \Delta(\textsc{Ext}_2(\mathcal{R}), \mathcal{U}_m) &\leq 2\Pr[\mathcal{W} \text{ is bad}] + \varepsilon_1 \cdot 2^{t+1} + \varepsilon_2 \\ &\leq 2\gamma + \varepsilon_1 \cdot 2^{t+1} + \varepsilon_2 \\ &\leq 3\max\{\varepsilon_1 + \delta, 2^{t+1}\varepsilon_1\} + \varepsilon_2. \end{aligned}$$

□

4.2.2 Sampling and Partitioning

For our two extractors, we need the following two samplers respectively. Both constructions basically come from [17], 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 4.2.9. *There exist constants n_0, c_1, c_2 such that for any $n \geq n_0$, $k, d \in \mathbb{N}$, $\delta \geq 2^{-c_1 k}$, and $k_{\min} \geq c_2 d \log(1/\delta)$, there exists an explicit $(n, d, k, k_{\min}, 6k_{\min}, \delta)$-sampler* $\textsc{Samp} : \{0,1\}^t \to P([n])$ *with $t = O(\log n \cdot \log(1/\delta))$.*

Lemma 4.2.10. *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* $\textsc{Samp} :$ $\{0,1\}^t \to P([n])$ *with $t = \alpha \log k$.*

The proof of Lemma 4.2.9 is given in Section 4.2.2.1, while Lemma 4.2.10 is given in Section 4.2.2.2.

4.2.2.1 The Sampler Using a Longer Seed

In this subsection, we prove Lemma 4.2.9, which shows the existence of a sampler using a longer seed. More precisely, we will use the following lemma to prove Lemma 4.2.9.

Lemma 4.2.11. *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.*

To prove Lemma 4.2.11, we use the following ℓ-wise independent tail inequality.

Theorem 4.2.12. *[6] Let $\ell \geq 6$ be an even integer. Suppose that $X_1, \ldots, X_n$ are ℓ-wise independent random variables taking values in $[0,1]$. Let $X = \sum_{i=1}^n X_i$ and $\mu = E(X)$, and let $\varepsilon > 0$. Then*

$$\Pr[|X - \mu| \geq \varepsilon] \leq 8 \left(\frac{\ell\mu + \ell^2}{\varepsilon^2} \right)^{\ell/2}.$$

Proof. (of Lemma 4.2.11) First we show that for any t with $6\log n \le t \le (k\log n)/(20r)$, there exists an explicit function $S : \{0,1\}^t \to P([n])$ such that for any real-valued function[4] $h' : [n] \to [0,1]$ with $\sum_{i\in[n]} h'(i) = k$,

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

for $\delta' = 2^{-\Omega(t/\log n)}$. Set $q = \lfloor \log r\rfloor \le \log n$. Let ℓ be the largest even integer such that $\ell\log n \le t$.

We construct the sampler as follows. First, we use t random bits to generate n ℓ-wise independent random variables $Z_1,\ldots,Z_n \in \{0,1\}^q$ [9]. Fix some value $a\in\{0,1\}^q$, and the sampler outputs the index set $T = \{i|Z_i = a\}$. Then, we show that this sampler satisfies the requirement.

Define n random variables $R_1,\ldots,R_n$ such that for $1\le i\le n$, $R_i = h'(i)$ if $i\in T$ and $R_i = 0$ otherwise. Note that R_i's are ℓ-wise independent. Let $R = \sum_{i=1}^n R_i = \sum_{i\in T} h'(i)$, and $\mu = \mathrm{E}[R] = \sum_{i=1}^n \mathrm{E}[R_i] = \sum_{i=1}^n h'(i)\cdot \Pr[Z_i = a] = k/2^q \le 2k/r$.

Then by Theorem 4.2.12 with $\varepsilon = k/(2r)$ and the fact that $6\le \ell\le t/\log n \le k/(20r)$, we have that

$$\begin{aligned}
\Pr[|R-\mu|\ge\varepsilon] &\le 8\left(\frac{\ell\mu+\ell^2}{\varepsilon^2}\right)^{\ell/2}\\
&\le 8\left(\frac{\frac{k}{20r}\left(\frac{2k}{r}+\frac{k}{20r}\right)}{\left(\frac{k}{2r}\right)^2}\right)^{\ell/2}\\
&\le 8\cdot 2^{-\ell/2}\\
&\le 2^{-\Omega(t/\log n)}.
\end{aligned}$$

Moreover, since $r/2 < 2^q \le r$, we have

$$\begin{aligned}
|R-\mu|\le k/(2r) &\Rightarrow k/2^q - k/(2r) \le R \le k/2^q + k/(2r)\\
&\Rightarrow k/r - k/(2r) \le R \le 2k/r + k/(2r)\\
&\Rightarrow k/(2r) \le R \le 3k/r
\end{aligned}$$

Thus,

$$\Pr_{w\in\mathcal{U}_t}\left[\frac{k}{2r} \le \sum_{i\in T} h'(i) \le \frac{3k}{r}\right] \ge \Pr\left[|R-\mu| \le \frac{k}{2r}\right] \ge 1-2^{-\Omega(t/\log n)}.$$

[4] In [17], they showed the case of $h' : [n] \to \{0,1\}$.

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

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

Thus, S is also an $(n, d, k, kd/(2r), 3kd/r, \delta')$-sampler, which proves Lemma 4.2.11. □

Now we proceed to prove Lemma 4.2.9.

Proof. (of Lemma 4.2.9) Suppose $\delta \ge 2^{-c_1 k}$ for a small enough constant c_1, and $k_{\min} \ge c_2 d \log(1/\delta)$ for a large enough constant c_2. Let us choose $r = kd/(2k_{\min}) \le k$, so that $kd/(2r) = k_{\min}$ and

$$(k \log n)/(20r) \ge (k_{\min} \log n)/(10d) \ge (c_2/10) \cdot \log n \cdot \log(1/\delta).$$

Thus, we can choose $t = (c_2/10) \cdot \log n \cdot \log(1/\delta)$ and have $6 \log n \le t \le (k \log n)/(20r)$. From Lemma 4.2.11, we have an $(n, d, k, k_{\min}, 6k_{\min}, \delta')$-sampler, with $\delta' = 2^{-\Omega(t/\log n)} \le 2^{-\Omega(c_2 \log(1/\delta))} \le \delta$. This completes the proof of Lemma 4.2.9. □

4.2.2.2 The Sampler Using a Shorter Seed

In this subsection, we prove Lemma 4.2.10, which shows the existence of a sampler using a shorter seed. Lemma 4.2.10 follows immediately from Lemma 4.2.13 below by using T_1 as the output of the sampler Samp.

Lemma 4.2.13. *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 $n \ge n_0$, $d \le k^{c_0}$, $k \ge \log^{c_1} n$, and $\delta = O(k^{-\beta})$, one can use $\alpha \log k$ random bits to explicitly partition $[n]$ into $r' = \Omega(k^{\beta})$ sets $T_1, \dots, T_{r'}$ such that for any function $h : [n] \to [0, d]$ with $\sum_{i=1}^{n} h(i) = k$,*

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

In addition to proving Lemma 4.2.10, Lemma 4.2.13 will also be used to prove Theorem 4.2.2. The proof of Lemma 4.2.13 can be seen as a derandomization of Lemma 4.2.9, using approximate pair-wise independent variables.

Definition 4.2.14. [40] 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 4.2.15. *[2] Let $r' < n$ be a power of 2. For any $n \geq 16$ and $0 < \varepsilon < 1/2$, one can use $7 \log r' + 3(\log \log n + \log(1/\varepsilon))$ random bits to generate n random variables $Z_1, \ldots, Z_n \in [r']$ that are pair-wise ε-dependent.*

Proof. (of Lemma 4.2.13) For any given constant $\alpha \in (0, 1)$, let $\beta = \alpha/38$, $r = k^{\beta}$, and $\varepsilon = k^{-4\beta}$. Let r' be a power of 2 such that $r/2 < r' \leq r$. 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 4.2.15, the number of random bits needed to generate them is at most

$$7 \log r' + 3(\log \log n + \log(1/\varepsilon)) \leq 19\beta \log k + 3 \log \log n \leq \alpha/2 \log k + 3 \log \log n.$$

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 $R_1, \ldots, R_n$ such that for $i \in [n]$, $R_i = h(i)$ if $i \in T_v$ and $R_i = 0$ otherwise. Let $R = \sum_{i=1}^{n} R_i = \sum_{i \in T_v} h(i)$, and we would like to bound the probability $\Pr[|R - k/r'| > k/(2r')]$. Since the expected value of R is close to k/r', with

$$|\operatorname{E}[R] - k/r'| = \left| \sum_{i=1}^{n} h(i) \cdot \Pr[Z_i = v] - k/r' \right| \leq \sum_{i=1}^{n} h(i) \cdot |\Pr[Z_i = v] - 1/r'| \leq k\varepsilon,$$

we have $\Pr[|R - k/r'| > k/(2r')] \leq \Pr[|R - \operatorname{E}[R]| > k/(2r') - k\varepsilon]$. Since $k\varepsilon \leq k/(6r')$ for some large enough n and $k \geq \log^{c_1} n$, and thus $k/(2r') - k\varepsilon \geq k/(3r')$, it suffices to bound the probability $\Pr[|R - \operatorname{E}[R]| > k/(3r')]$. We would like to apply the Chebyshev's inequality of Lemma 2.7.5. Therefore, we need to bound the variance of R, which is $Var(R) = \sum_{i=1}^{n} Var(R_i) + \sum_{i \neq j} cov(R_i, R_j)$. For any $i \in [n]$, $Var(R_i)$ is

$$\operatorname{E}[R_i^2] - \operatorname{E}[R_i]^2 \leq \operatorname{E}[R_i^2] = h(i)^2 \cdot \Pr[Z_i = v] \leq h(i)^2 \cdot (1/r' + \varepsilon).$$

For any distinct $i, j \in [n]$, $cov(R_i, R_j)$ is

$$\begin{aligned}
\mathrm{E}[R_i \cdot R_j] - \mathrm{E}[R_i] \cdot \mathrm{E}[R_j] &= h(i)h(j) \cdot (\Pr[Z_i = Z_j = v] - \Pr[Z_i = v] \Pr[Z_j = v]) \\
&\leq h(i)h(j) \cdot \left((1/r'^2 + \varepsilon) - (1/r' - \varepsilon)^2\right) \\
&= h(i)h(j) \cdot (1 + 2/r' - \varepsilon)\, \varepsilon \\
&\leq h(i)h(j) \cdot 2\varepsilon,
\end{aligned}$$

as $r' \geq 2$. Therefore, $Var(R)$ is at most

$$\sum_i h(i)^2 (1/r' + \varepsilon) + \sum_{i \neq j} h(i)h(j) 2\varepsilon \leq (1/r') \sum_i h(i)^2 + 2\varepsilon \left(\sum_i h(i)\right)^2 \leq dk/r' + 2k^2\varepsilon,$$

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 the Chebyshev's inequality of Lemma 2.7.5, we have

$$\Pr[|R - \mathrm{E}[R]| > k/(3r')] < \frac{dk/r' + 2\varepsilon k^2}{(k/3r')^2} = 9dr'/k + 18\varepsilon r'^2 = O(\varepsilon r'^2),$$

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 $v \in [r']$, $\Pr[k^\tau/2 \leq \sum_{i \in T_v} h(i) \leq 3k^\tau] \geq 1 - O(k^{-2\beta})$. Then, Lemma 4.2.13 follows from the union bound. □

4.2.3 The Extractor for Sources with Large Min-entropy

In this subsection, we prove Theorem 4.2.1, which shows the existence of an extractor for independent-symbol sources with large min-entropy. The construction is very similar to that in [17]. First, as in [33], we have the following seedless extractor[5] for the case of large min-entropy.

Lemma 4.2.16. *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* $\textsc{Ext}_1 : [D]^n \to \{0,1\}^{m_1}$ *where $m_1 = \Omega(n^{2\gamma}/(D^2 d))$ and $\varepsilon_1 = 2^{-m_1}$.*

We need the following lemma to prove Lemma 4.2.16.

[5] This lemma is a generalization of the main result in [33] for bit-fixing sources.

Lemma 4.2.17. *[33] Let G be a p-regular graph with 2^{m_1} vertices, and the second largest eigenvalue (in absolute value) of its adjacency matrix is $p\lambda$. Suppose we take a walk on G for b steps according to b symbols independently chosen in $[p]$, in which ℓ symbols have distributions within a L_2-distance of ε from the uniform distribution. Then the statistical distance from the uniform distance at the end of the walk is bounded by $\frac{1}{2}(\lambda+\varepsilon\sqrt{p})^{\ell}\cdot 2^{m_1/2}$.*

Proof. (of Lemma 4.2.16) The extractor works as follows. Let p be the smallest prime greater than D. Set $k_0 = c_0 p^2 \log p$, 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 4.1.1 to extract a symbol in $\mathbb{Z}_p$. Then use the b extracted symbols (one per block) to take a b-step walk on a p-regular graph with 2^{m_1} nodes, for some m_1 to be determined later, and the second largest eigenvalue of its adjacency matrix is $p\lambda \le p^{1-c_1}$, for some constant $c_1 < 1/2$ [24, 38]. 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. Let ℓ denote the number of good blocks. Since each symbol has min-entropy at most $\log D \le \log p$, then counting the min-entropy, we have that

$$k_1 \le \ell\cdot(2nk_0/k_1)\log p + (b-\ell)\cdot k_0 \le \ell\cdot(2nk_0/k_1)\log p + b\cdot k_0 = \ell\cdot(2nk_0/k_1)\log p + k_1/2,$$

which implies that

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

Recall that in the proof of Theorem 4.1.1, we show that our extractor can extract a symbol which has L_2-distance from the uniform distance at most $2^{-\Omega(k/p^2)}$ from an independent-symbol source with min-entropy k. Since we can view each good block as a $(2nk_0/k_1, p, k_0)$-source, the L_2-distance between the extracted symbol and the uniform distance is at most ε for some $\varepsilon \le 1/p$. Then according to Lemma 4.2.17, after the b-step walk on the expander, the distribution of the final node is ε_1-random, for

$$\varepsilon_1 \le \frac{1}{2}\cdot(\lambda+\varepsilon\sqrt{p})^{\ell}\cdot 2^{m_1/2} \le 2^{-\Omega(k_1^2/(k_0 n))}\cdot 2^{m_1/2} \le 2^{-\Omega(n^{2\gamma}/(D^2 d))+m_1/2}.$$

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

Following [17], to extract more randomness, we apply Lemma 4.2.4 with $\textsc{Ext}_1$ above together with two additional ingredients: (1) an $(n, d, k, k_{\min}, k_{\max}, \varepsilon/4)$-sampler $\textsc{Samp} : \{0,1\}^t \to P([n])$ from Lemma 4.2.9, 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 $\textsc{Ext}_2 : [D]^n \times \{0,1\}^s \to \{0,1\}^m$ from [46], 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 4.2.4, we get an (n, D, k, ε_3)-extractor $\textsc{Ext}_3 : [D]^n \times \{0,1\}^m$ 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 4.2.1.

4.2.4 The Extractor for Sources with Smaller Min-entorpy

In this subsection, we prove Theorem 4.2.2, which shows the existence of an extractor for independent-symbol sources with smaller min-entropy. The construction is again very similar to the corresponding one in [17]. Suppose $k \geq \log^{c_1} n$ for a large enough constant c_1. We first use the seedless extractor in Theorem 4.1.1 to extract $O(\log k)$ bits of randomness. To apply Lemma 4.2.4 to extract more randomness, we need a seeded extractor with such a short seed. Similar to [17], the existence of such an extractor is guaranteed by the following.

Lemma 4.2.18. *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 $\textsc{Ext}' : [D]^n \times \{0,1\}^s \to \{0,1\}^m$ 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 4.2.13 to partition the source into several parts and then apply our seedless extractor in Theorem 4.1.1 on each part.

Fix any constant $\alpha \in (0,1)$. Consider any (n, D, k)-source $\mathcal{X} = \mathcal{X}_1 \circ \cdots \circ \mathcal{X}_n$. Let $h : [n] \to [0, d]$ be the function $h(i) = \mathrm{H}_\infty(\mathcal{X}_i)$. Note that $\sum_{i=1}^n h(i) = k$. According to Lemma 4.2.13, 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 $\textsc{Ext}'$ as $\textsc{Ext}'(x) = z_1 \circ \cdots \circ z_{r'}$, where $z_v = \textsc{Ext}_0(x_{T_v} \circ 0^{n-|T_v|})$ for $v \in [r']$ and $\textsc{Ext}_0 : [D]^n \to \{0,1\}^\ell$ is our $(n, D, k_0, \varepsilon_0)$-extractor in Theorem 4.1.1, with $k_0 = k^\tau/2$, $\ell = O(\log k)$, and $\varepsilon_0 = k^{-\omega(1)}$. We want to prove that $\textsc{Ext}'(\mathcal{X})$ 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 Corollary 2.1.5, 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,

$$\Delta(\textsc{Ext}'(\mathcal{X}), \mathcal{U}_m) \leq \Pr\left[(T_1, \ldots, T_{r'}) \text{ is not good}\right] + k^{-\Omega(1)} \leq k^{-\Omega(1)}.$$

□

Then we can apply Lemma 4.2.4 with the following ingredients: (1) a seedless (n, D, k, ε_1)-extractor $\textsc{Ext}_1 : [D]^n \to [M]$ from Theorem 4.1.1, 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 $\textsc{Samp} : \{0,1\}^t \to P([n])$ from Lemma 4.2.10, 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 $\textsc{Ext}_2 : [D]^n \times \{0,1\}^s \to \{0,1\}^{m_2}$ from Lemma 4.2.18 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 $\textsc{Ext}_3 : [D]^n \to \{0,1\}^{m_2}$, with $\varepsilon_3 = k^{-\Omega(1)} + O(2^t\varepsilon_1) = k^{-\Omega(1)}$.

To extract even more random bits, we again apply Lemma 4.2.4, but now using the above extractor $\textsc{Ext}_3$ together with the following two ingredients: (1) an $(n, d, k, k_{\min}, k_{\max}, \varepsilon/4)$-sampler $\textsc{Samp} : \{0,1\}^t \to P([n])$ from Lemma 4.2.10 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 $\textsc{Ext}_2 : \{0,1\}^n \times \{0,1\}^s \to \{0,1\}^m$ from [46], with $s \leq m_2/2$ and $m = k - k_{\max}$. As a result, we obtain a seedless (n, D, k, ε)-extractor $\textsc{Ext} : [D]^n \to \{0,1\}^m$, 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 4.2.2.

4.3 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 (non-explicit) seedless extractor for independent-symbol sources with entropy loss $O(\log(1/\varepsilon))$. More precisely, we have the following theorem.

Theorem 4.3.1. *Suppose $k \geq c\log(Dn/\varepsilon)$ for a large enough constant c. Then there exists an (n, D, k, ε)-extractor* $\textsc{Ext} : [D]^n \to \{0,1\}^m$ *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 $\textsc{Ext}$, then we succeed with a positive probability.

Proof. (of Theorem 4.3.1) Let F denote the set of all functions $f : [D]^n \to \{0,1\}^m$. We say that a function $f \in F$ fails on an (n, D, k)-source $\mathcal{X}$ if $\Delta(f(\mathcal{X}), \mathcal{U}_m) > \varepsilon/2$. We have the following.

Claim 4.3.2. For any (n, D, k)-source $\mathcal{X}$, $\Pr_{f\in F}[f \text{ fails on } \mathcal{X}] \leq 2^{2^m} \cdot 2^{-\Omega(\varepsilon^2 2^k)}$.

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

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

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

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

which by the Chernoff bound of Lemma 2.7.6 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 claim. □

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

Claim 4.3.3. For any (n, D, k)-source $\mathcal{X}$, there exists an (n, D, k)-source $\mathcal{X}'$ such that $\Delta(\mathcal{X}, \mathcal{X}') \leq \varepsilon/2$ and $\mathrm{H}_\infty(\mathcal{X}'_i)$ is a multiple of α for any $i \in [n]$.

Proof. For $i \in [n]$, let $k_i = \mathrm{H}_\infty(\mathcal{X}_i)$. 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 $\mathrm{H}_\infty(\mathcal{X}'_i) = k'_i$ 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 $i \in [n]$, $\Delta(\mathcal{X}_i, \mathcal{X}'_i) \leq D \cdot |k'_i - k_i|/2 < D \cdot \alpha/2 \leq \varepsilon/(2n)$. Then by Corollary 2.1.5, we have $\Delta(\mathcal{X}, \mathcal{X}') \leq \sum_{i \in [n]} \Delta(\mathcal{X}_i, \mathcal{X}'_i) \leq \varepsilon/2$. Since $\mathrm{H}_\infty(\mathcal{X}') = k'_1 + \cdots + k'_n = k = \mathrm{H}_\infty(\mathcal{X})$, we have the claim. □

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 $\mathrm{H}_\infty(\mathcal{Y}_i) = k_i$. 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.

Claim 4.3.4. 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 $i \in [n]$, $\mathrm{H}_\infty(\mathcal{Y}_i) = \mathrm{H}_\infty(\mathcal{X}'_i)$ 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.

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 $j \in [D]$. Since all these inequalities are linear, the set of such vectors forms a convex polytope. Hence, each vector in the set is expressible as a convex combination of vertices (corners) of the polytope, where a vertex of the polytope is a vector satisfying a maximal subset of these inequalities when treat these inequalities as equalities. Hence, a vector is a vertex of the polytope if and only if it has $\lfloor 2^{k_i} \rfloor$ values of j such that $p_j = 2^{-k_i}$, one value of j such that $p_j = 1 - \lfloor 2^{k_i} \rfloor 2^{-k_i}$, and for the rest j, $p_j = 0$. That is, 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 $\mathcal{S}$.

Claim 4.3.5. $|\mathcal{S}| \le 2^{2^{O(\log(Dn))}}$.

Proof. Recall that $\alpha = 1/\lceil Dn/\varepsilon \rceil \le \varepsilon/(Dn)$. 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

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

□

From Claim 4.3.2 and Claim 4.3.5 and using a union bound, we have

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

for some $m = k - O(\log(1/\varepsilon))$ when $k \geq c\log(Dn/\varepsilon)$ for a large enough constant c. This implies the existence of some $\textsc{Ext} \in F$ such that $\Delta(\textsc{Ext}(\mathcal{Y}), \mathcal{U}_m) \leq \varepsilon/2$ 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 Claim 4.3.3 and Claim 4.3.4, 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 by Proposition 2.1.2 and 2.1.3,

$$\begin{aligned}\Delta(\textsc{Ext}(\mathcal{X}), \mathcal{U}_m) &\leq \Delta(\textsc{Ext}(\mathcal{X}), \textsc{Ext}(\mathcal{Y})) + \Delta(\textsc{Ext}(\mathcal{Y}), \mathcal{U}_m)\\ &\leq \Delta(\mathcal{X}, \mathcal{Y}) + \Delta(\textsc{Ext}(\mathcal{Y}), \mathcal{U}_m)\\ &\leq \varepsilon.\end{aligned}$$

That is, $\textsc{Ext}$ is an (n, D, k, ε)-extractor, which proves Theorem 4.3.1. □

4.4 Lower Bound on Entropy Loss

In this section, we show that the existential upper bound on the entropy loss in Section 4.3 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 4.4.1. *Let* $\textsc{Ext} : \{0,1\}^n \times \{0,1\}^s \to \{0,1\}^m$ *be an* $(n, 2, k, \varepsilon)$*-extractor for bit-fixing sources, with* $n, s, m \in \mathbb{N}$, $\log(1/\varepsilon) \leq k \leq n - \log(1/\varepsilon)$, *and* $0 < \varepsilon < 1/c_1$, *for some large enough constants* c_1. *Then* $m \leq k + s - \Omega(\log(1/\varepsilon))$.

We will basically follow the proof idea in [43]. Briefly speaking, given any $\textsc{Ext} : \{0,1\}^n \times \{0,1\}^s \to \{0,1\}^m$ with m exceeding the bound, we will show the existence of a bit-fixing source of min-entropy k on which $\textsc{Ext}$ 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

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

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

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

We will rely on the following lemma from [43].[6]

Lemma 4.4.2. *Suppose $\mathcal{X}$ is the uniform distribution over a set $X \subseteq \{0,1\}^n$ with $|X| = 2^k$, and $\Delta(\textsc{Ext}(\mathcal{X}, \mathcal{U}_s), \mathcal{U}_m) \leq \varepsilon$. 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 4.4.3. *Suppose $n, t \in \mathbb{N}$ and $\delta \in (0,1)$, with $t \leq n - \Omega(\log(1/\delta))$ 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.*[7] *Then $|S| \geq 2^t(1/\delta)^{\Omega(1)}$.*

Note that a set S satisfying the property in Lemma 4.4.3 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 4.4.3 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 4.4.1. With this lemma, we can now prove Theorem 4.4.1.

[6] Note that this lemma does not appear explicitly in [43] 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".

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

Proof. (of Theorem 4.4.1)

Assume for the sake of contradiction that $m \geq k+s-c\log(1/\varepsilon)$ 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 [43], the existence of such a source will be shown using a probabilistic argument. The difference is that [43] 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 [43], 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 non-negligible probability. As we will show next, this probability is guaranteed by Lemma 4.4.3, 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 4.4.3 (in particular, the condition $k \leq n - \log(1/\varepsilon)$ implies the condition $\delta \geq 1/(8\binom{n}{t}2^t)$). Now the average of $|S^{(z)}|$ over z is

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

Call z *small* if $|S^{(z)}| < 2^t(1/\delta)^{c'}$ for a small enough constant c'. By Markov inequality of Lemma 2.7.3, at least 1/2 fraction of z's are small. From Lemma 4.4.3, 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 4.4.2, this implies that $\Delta(\textsc{Ext}(\mathcal{X}_I^u), \mathcal{U}_m) > \varepsilon$, a contradiction. Therefore, one must have $m \le k + s - \Omega(\log(1/\varepsilon))$, which proves the theorem. $\square$

4.4.1 Size Lower Bound on Almost k-wise Independent Spaces

In this subsection, we prove Lemma 4.4.3, which shows a size lower bound on almost k-wise independent spaces.

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 [1, 11], 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 4.4.4. 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 the Markov inequality of Lemma 2.7.3, 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 4.4.5. 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 $w \in \{0,1\}^r$,

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

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

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

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 $|S_{J,v}| \geq 2^r(1/\delta)^{\Omega(1)}$. 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 4.4.6. There exists an even integer r such that $2 \leq r \leq \min\{t, k/24\}$ and $(1/\delta)^{\Omega(1)} \leq \binom{k+r}{r}2^r < \sqrt{1/\delta}$.

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 4.4.5, this means that the set $S_{J,v}$ projected to dimensions in $\bar{J}$ forms an r-wise independent space. From [1, 11], such a set must have size at least

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

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

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

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 4.4.5, implies that the set $S_{J,v}$ projected to dimensions in A gives a pair-wise independent space, so by [1, 11] we have

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

Claim 4.4.7. 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

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

Then by the well-known Turan's theorem in graph theory (e.g., see Theorem 4.7 in [30]), 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 4.4.3.

4.5 Open Problems

In Section 4.2, we provide two deterministic extractors for (n, D, k)-sources. 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 $\varepsilon \geq 2^{-\Omega(n^\gamma)}$, while when $k \geq \log^c n$, for some constant $c > 0$, we can extract $m = k - (1/\varepsilon)^{O(1)}$ bits with error $\varepsilon \geq k^{-\Omega(1)}$. However, in Section 4.3, we show the existence of a seedless extractor for (n, D, k)-sources which can extract $m = k - O(\log(1/\varepsilon))$ bits whenever $k = \omega(d + \log(n/\varepsilon))$. This means that a better explicit extractor than ours may exist. Hence, we intend to give a better explicit construction.

In Section 4.4, we provide a lower bound on entropy loss for bit-fixing sources. We mean to provide a better lower bound on entropy loss for independent-symbol sources. In addition, in the proof of lower bound, we show a size lower bound on an ”almost” t-wise independent space, which can be seen as a relaxation of the standard notation of approximation t-wise independent space. Hence, the size lower bound on an ”almost” t-wise independent space immediately implies a size lower bound on an approximation

t-wise independent space. We aim to show a better size lower bound on an approximation t-wise independent space.

Chapter 5

Extracting Computational Entropy and Learning Noisy Linear Functions

In this chapter, we study the task of deterministically extracting randomness from sources containing computational entropy. The sources we consider have the form of a conditional distribution $(f(\mathcal{X})|\mathcal{X})$, for some function f and some distribution $\mathcal{X}$, and we say that such a source has computational min-entropy k if any circuit of size 2^k can only predict $f(x)$ correctly with probability at most 2^{-k} given input x sampled from $\mathcal{X}$. First, in Section 5.1, we describe the Goldreich-Levin Theorem, which implicitly provides a seeded computational extractor. In Section 5.2, we show that it is impossible to have a seedless extractor to extract from one single source of this kind. Then, in Section 5.3, we show that it becomes possible if we are allowed a seed which is weakly random (instead of perfectly random) but contains some statistical min-entropy, or even a seed which is not random at all but contains some computational min-entropy. This can be seen as a step toward extending the study of multi-source extractors from the traditional, statistical setting to a computational setting. We reduce the task of constructing such extractors to a problem in learning theory: learning linear functions under arbitrary distribution with adversarial noise. In Section 5.4, we provide a learning algorithm for this problem, which may have interest of its own.

5.1 The Goldreich-Levin Theorem

The idea of extracting computational randomness has appeared implicitly long ago [59, 19, 23], for the task of constructing pseudo-random generators from one-way functions. Moreover, our extractor is motivated by the Goldreich-Levin Theorem. For completeness, we give a proof here.

Theorem 5.1.1. *[19, 60] Let g be a one-way permutation on n bits, and let A be an algorithm with running time* $\mathrm{Time}(A)$ *such that*

$$\Pr_{y,r}[A(g(y),r) = \langle y,r\rangle] > 1/2+\varepsilon.$$

Then there is an algorithm P such that

$$\Pr[P(g(y)) = y] > \varepsilon/2,$$

and runs in time $O((n^3/\varepsilon^4)\mathrm{Time}(A))$.

Proof. It is sufficient to show that P outputs a list Y containing the correct y. This is because g is easy to compute, we can compute $g(z)$ for all $z \in Y$ until we find $z \in Y$ such that $g(z) = g(y)$.

Since A is an algorithm such that

$$\Pr_{y,r}[A(g(y),r) = \langle y,r\rangle] > 1/2+\varepsilon,$$

by the Markov's inequality of Corollary 2.7.4, we have that at least an $\varepsilon/2$ fraction of y's satisfying that

$$\Pr_{r}[A(g(y),r) = \langle y,r\rangle] > (1+\varepsilon)/2.$$

We call such y's *good*. Then we prove that for every good y, we can output a list containing y.

Fix a good y. First note that when $A(g(y),r) = \langle y,r\rangle$ for all $r \in \{0,1\}^n$, we can obtain the ith bit of y by

$$y_i = \langle y, e_i\rangle = A(g(y), e_i),$$

where $e_i \in \{0,1\}^n$ is the string whose bits are all zero except the ith bit. Moreover, if $A(g(y),r) = \langle y,r\rangle$ with high probability, then we can compute

$$y_i = \langle A(g(y),r), A(g(y), r\oplus e_i)\rangle,$$

where $\oplus$ denotes the bitwise exclusive or, with high probability. However, for a good y, we only can guarantee that $\Pr_r[A(g(y),r) = \langle y,r\rangle] > (1+\varepsilon)/2$, which is too small to apply the above strategy. In fact, if we know one of $A(g(y),r)$ and $A(g(y), r\oplus e_i)$ for sure, it suffices to do that. Hence, we consider the following algorithm in Figure 5.1.

```
• Input: $z \in \{0,1\}^n$.

• Procedure:

    1. Randomly choose $r^1, \cdots, r^\ell \in \{0,1\}^n$ for $\ell = \log((12n/\varepsilon)+1)$.

    2. For every $w = w_1 \circ w_2 \circ \cdots \circ w_\ell \in \{0,1\}^\ell$

        (a) We guess that $\langle y, r^j\rangle = w_j$ for every $j \in [\ell]$. Then, for any nonempty
            set $J \subseteq [\ell]$, compute $s^J = \sum_{j\in J} r^j$ and $t^J = \sum_{j \in J} w_j$.

        (b) For every $i \in [n]$, set $y_i$ to be the majority value among $\{\langle t^J, A(z, s^J \oplus
            e_i)\rangle\}$ over all nonempty set $J \subseteq [\ell]$.

        (c) Include $y = y_1 \circ \cdots \circ y_n$ into $Y$.

    3. Output $Y$.
```

Figure 5.1: The construction of the algorithm P

Note that there exists $w \in \{0,1\}^\ell$, such that $\langle y, r^j\rangle = w_j$ for every $j \in [\ell]$. Fix such w. Consider any $i \in [n]$. For every nonempty set $J \subseteq [\ell]$, let Z_J be the indicator random variable for the event of $A(g(y), s^J \oplus e_i) \neq \langle y, s^J \oplus e_i\rangle$. Set $Z = \sum_{J\neq\emptyset} Z_J$. Then $\mathrm{E}[Z] = \sum_{J \neq \emptyset} \mathrm{E}[Z_J] \leq (2^\ell - 1)\cdot(1-\varepsilon)/2$. Moreover, since Z_J's are pairwise independent, we have that $Var[Z] = \sum_{J\neq\emptyset} Var[Z_J] < 2^\ell - 1$. Hence, by the Chebyshev's inequality of Lemma 2.7.5, we obtain that

$$\begin{aligned}
\Pr[P(g(y))_i \neq y_i] &= \Pr\left[Z \geq 1/2 \cdot \left(2^\ell - 1\right)\right] \\
&\leq \Pr\left[|Z - \mathrm{E}[Z]| \geq (2^\ell - 1)\cdot \varepsilon/2\right] \\
&< \frac{4}{\varepsilon^2 \cdot (2^\ell - 1)} \\
&< \frac{1}{3n}.
\end{aligned}$$

Hence, a union bound shows that for any good y, $\Pr[P(g(y)) \neq y] = \Pr[\exists i, P(g(y))_i \neq y_i] < 1/3$. Moreover, since P calls A at most $O((n^3/\varepsilon^4)$ times, the running time of P is $O((n^3/\varepsilon^4)\mathrm{Time}(A))$. □

Remark 5.1.2. The Goldreich-Levin Theorem implies that $\textsc{Ext}(v, w) = \langle v, w\rangle$ is a seeded extractor for sources $(\mathcal{V}|\mathcal{X})$ with computational min-entropy. The reason is that if there exists a distinguisher E for the two distributions $\mathcal{X}\circ\mathcal{U}_n\circ\textsc{Ext}(\mathcal{V},\mathcal{U}_n)$ and $\mathcal{X}\circ\mathcal{U}_n\circ\mathcal{U}_1$ where $\mathcal{U}_1$ and $\mathcal{U}_n$ are independent, then Lemma 2.5.4 shows that there exists a predictor A which predicts $\textsc{Ext}(\mathcal{V},\mathcal{U}_n)$ well with input $\mathcal{X}\circ\mathcal{U}_n$. Hence, by Goldreich-Levin Theorem of Theorem 5.1.1, we can get an algorithm P which can predict $\mathcal{V}$ well with input $\mathcal{X}$, and this contradicts the assumption that $(\mathcal{V}|\mathcal{X})$ has computational min-entropy.

5.2 An Impossibility Result

Just as in the statistical setting [10], we show that seedless extractors do not exist either in the computational setting. In fact, we show the impossibility result even for sources with a computational min-entropy as high as $n - 2$.

Theorem 5.2.1. *For any $n_1, n \in \mathbb{N}$ with $n_1 \geq 3n$ and for any function $\textsc{Ext} : \{0,1\}^n \to \{0,1\}$, there exists a deterministic function $f : \{0,1\}^{n_1} \to \{0,1\}^n$ such that $\mathrm{H}_c(f(\mathcal{X})|\mathcal{X}) = n - 2$ for $\mathcal{X} = \mathcal{U}_{n_1}$ but $\textsc{Ext}(f(x))$ takes the same value for all x (so can be easily distinguished from random).*

Proof. Consider any function $\textsc{Ext} : \{0,1\}^n \to \{0,1\}$. Assume without loss of generality that $|\textsc{Ext}^{-1}(1)| \geq 2^{n-1}$. Then we will show the existence of a function f such that $\mathrm{H}_c(f(\mathcal{X})|\mathcal{X}) = n - 2$ but $\textsc{Ext}(f(x)) = 1$ for all x. In fact, a standard argument can show that a random function is likely to work, as we will describe next.

Consider a random function $f : \{0,1\}^{n_1} \to \textsc{Ext}^{-1}(1)$. Fix any $C : \{0,1\}^{n_1} \to \{0,1\}^n \in \mathsf{SIZE}(2^{n-2})$, and for each $x \in \{0,1\}^{n_1}$, define a binary random variable C_x such that $C_x = 1$ if and only if $C(x) = f(x)$. Observe that $\sum_x C_x$ is the number of x satisfying $C(x) = f(x)$. Note that

$$\mathop{\mathrm{E}}_f\left[\sum_x C_x\right] = \sum_x \mathop{\mathrm{E}}_f[C_x] = \sum_x \Pr_f[C(x) = f(x)] \leq 2^{n_1-(n-1)},$$

and let $\mu = 2^{n_1-(n-1)}$. Then by the Chernoff bound of Lemma 2.7.6, we have

$$\Pr_f\left[\sum_x C_x \geq 2\mu\right] \leq 2^{-\Omega(\mu)} = 2^{-\Omega(2^{n_1-n})}.$$

Since $|\mathsf{SIZE}(2^{n-2})| \leq 2^{O(n2^n)}$ and $n_1 \geq 3n$, a union bound gives

$$\Pr_f\left[\exists C \in \mathsf{SIZE}(2^{n-2}) \text{ s.t. } \sum_x C_x \geq 2\mu\right] \leq 2^{O(n2^n)} \cdot 2^{-\Omega(2^{n_1-n})} < 1.$$

Hence, there exists some f, such that $\Pr_x[C(x) = f(x)] < 2\mu \cdot 2^{-n_1} = 2^{-(n-2)}$ for any $C \in \mathsf{SIZE}(2^{n-2})$, but $\textsc{Ext}(f(x)) = 1$ for any x. This completes the proof. □

5.3 Hybrid and Computational Extractors

In this section, we show that the function $\textsc{Ext}^2 : \mathcal{F}^\ell \times \mathcal{F}^\ell \to \mathcal{F}$ defined in Theorem 3.2.1 as

$$\textsc{Ext}^2(v, w) = \langle v, w\rangle_m,$$

which is known to be a good stron-two-source-extractor, is also a good hybrid extractor and a good computational extractor.

Theorem 5.3.1. *For any $k \geq \Omega(\log^2 n)$, any $m \leq O(\sqrt{k/\log k})$ dividing n, any $\varepsilon \geq 2^{-O(\sqrt{k/\log k})}$, any $s \leq 2^{n-k+O(k/\log k)}$, and for some $k_1 = n - k + O(k/\log k)$, the function $\textsc{Ext}^2 : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}^m$ defined above is both a (k_1, k, ε, s)-hybrid-extractor and a (k_1, k, ε, s)-computational-extractor.*

The proof for Theorem 5.3.1 relies on the following result, which gives an algorithm for the problem of learning linear functions under arbitrary distribution with adversarial noise.

Theorem 5.3.2. *For any $k \geq \Omega(\log^2 n)$, any $m \leq O(k/\log k)$ dividing n, and any $\delta \geq 2^{-O(\sqrt{k/\log k})}$, there exists a learning algorithm A with the following property. Given any source $\mathcal{W}$ over $\{0,1\}^n = \mathcal{F}^\ell$ with $\mathrm{H}_\infty(\mathcal{W}) = k$ and any function $q : \mathcal{F}^\ell \to \mathcal{F}$, the algorithm A samples $2^{O(k/\log k)}$ training examples from the distribution $(\mathcal{W}, q(\mathcal{W}))$ and then runs in time $2^{n-k+O(k/\log k)}$ to output a list of size $2^{n-k+O(k/\log k)}$ which with probability $1 - o(1)$ contains every $v \in \mathcal{F}^\ell$ satisfying*

$$\Pr_{w\in\mathcal{W}}[q(w) = \langle v, w\rangle_m] \geq 1/2^m + \delta.$$

Note that as in a standard learning-theoretical setting, we do not count the complexity of sampling the training examples (or just count each sampling as unit cost) in Theorem 5.3.2. We will prove the theorem in the next section, and now let us see how it is used to show Theorem 5.3.1.

Proof. (of Theorem 5.3.1)

First, we prove that the function $\textsc{Ext}^2$ is a good hybrid extractor. Consider any source $(\mathcal{V}|\mathcal{X})$ with $\mathrm{H}_c(\mathcal{V}|\mathcal{X}) = k_1$ and any source $\mathcal{W}$, which is independent of $(\mathcal{V}|\mathcal{X})$, with $\mathrm{H}_\infty(\mathcal{W}) = k$. Assume for the sake of contradiction that there exists an ε-distinguisher $E \in \mathsf{SIZE}(s)$ for the distributions $\mathcal{X} \circ \mathcal{W} \circ \langle \mathcal{V}, \mathcal{W} \rangle_m$ and $\mathcal{X} \circ \mathcal{W} \circ \mathcal{U}_m$. By Lemma 2.5.4, this implies the existence of a predictor $Q \in \mathsf{SIZE}(s + O(m))$ with

$$\Pr_{x \in \mathcal{X}, v \in \mathcal{V}, w \in \mathcal{W}}[Q(x \circ w) = \langle v, w \rangle_m] \geq (1 + \varepsilon)/2^m.$$

Let $\delta = \varepsilon/2^{m+1} \geq 2^{-O(\sqrt{k/\log k})}$, and call (x, v) heavy if

$$\Pr_{w \in \mathcal{W}}[Q(x \circ w) = \langle v, w \rangle_m] \geq 1/2^m + \delta.$$

Then the Markov inequality of Corollary 2.7.4 shows that $\Pr_{x \in \mathcal{X}, v \in \mathcal{V}}[(x, v) \text{ is heavy}] \geq \delta$.

Given any heavy (x, v), we want to predict v from x with a good probability. This can be reduced to the task of learning the linear function $\langle v, \cdot \rangle_m$, through noisy training examples $(w, q(w))$, with $q(w) = Q(x \circ w)$, under the distribution $w \in \mathcal{W}$. Consider the algorithm C which on input x calls the algorithm A in Theorem 5.3.2 using the function $q(\cdot) = Q(x \circ \cdot)$, and outputs a random element in the list produced by A. It samples $2^{O(k/\log k)}$ independent elements, denoted as W, from $\mathcal{W}$, makes $2^{O(k/\log k)}$ calls to Q, and for any heavy (x, v) it outputs v with probability $(1 - o(1)) \cdot 2^{-(n-k+O(k/\log k))}$. Then we have

$$\begin{aligned} \Pr_{x,v,W}[C(x) = v] &\geq \Pr_{x,v}[(x,v) \text{ is heavy}] \cdot \Pr_{x,v,W}[C(x) = v \mid (x,v) \text{ is heavy}] \\ &\geq \delta \cdot (1 - o(1)) \cdot 2^{-(n-k+O(k/\log k))} \\ &\geq 2^{-(n-k+O(k/\log k))}. \end{aligned}$$

We are almost done except that we still can not bound the complexity of the algorithm C because it needs a way to sample elements from the source $\mathcal{W}$ which may not have an efficient sampling algorithm, unlike in the learning setting where one does not count the complexity of sampling. Fortunately, by an average argument, the bound above still holds for some fixed W, and we can simply hard-wire it into C. Similarly, we can do this for other random choices of C, and it is not hard to show that one can have a resulting circuit of size

$$|W|^{O(1)} + 2^{O(k/\log k)} \cdot (s + O(m)) + 2^{n-k+O(k/\log k)} \leq 2^{n-k+O(k/\log k)}.$$

Thus, for some large enough $k_1 = n - k + O(k/\log k)$, we have a circuit of size smaller than 2^{k_1} which can predict v correctly with probability at least $2^{-(n-k+O(k/\log k))} > 2^{-k_1}$. This contradicts the assumption that $\mathrm{H}_c(\mathcal{V}|\mathcal{X}) = k_1$, which means that the distinguisher E assumed at the beginning cannot exist, so $\textsc{Ext}^2$ is a good hybrid extractor as claimed.

Next, we prove that $\textsc{Ext}^2$ is also a good computational extractor, and the proof is almost identical. Consider two independent sources $(\mathcal{V}|\mathcal{X})$ and $(\mathcal{W}|\mathcal{Y})$, with $\mathrm{H}_c(\mathcal{V}|\mathcal{X}) = k_1$ and $\mathrm{H}_c(\mathcal{W}|\mathcal{Y}) = k$. Observe that the distribution of $\mathcal{W}$ must have statistical min-entropy at least k, because otherwise the predictor which always outputs the value with the largest measure can predict $\mathcal{W}$ correctly with probability larger than 2^{-k}, a violation of the hardness assumption of g. Then we can follow the proof above: assuming the existence of a distinguisher for $\textsc{Ext}^2$, we can obtain a predictor of size smaller than 2^{k_1}, with some $2^{O(k/\log k)}$ elements from $(\mathcal{W}, \mathcal{Y})$ hard-wired in it, which can predict $\mathcal{V}$ correctly with probability larger than 2^{-k_1}. This contradicts the fact that $\mathrm{H}_c(\mathcal{V}|\mathcal{X}) = k_1$, so $\textsc{Ext}^2$ is a good computational extractor. □

5.4 Learning Noisy Linear Functions

In this section, we prove Theorem 5.3.2. Recall that given any source $\mathcal{W}$ over $\{0,1\}^n = \mathcal{F}^\ell$ with $\mathrm{H}_\infty(\mathcal{W}) = k$, any $\delta \geq 2^{-O(\sqrt{k/\log k})}$, and any function $q : \mathcal{F}^\ell \to \mathcal{F}$, we would like to learn some unknown $v \in \mathcal{F}^\ell$ such that

$$\Pr_{w \in \mathcal{W}} [q(w) = \langle v, w \rangle_m] \geq 1/2^m + \delta. \tag{5.1}$$

Since such v may not be unique, we will list them all. Let us first imagine one such fixed v.

We start by randomly choosing $K = 2^{c(k/\log k)}$ independent training examples (with replacement) from the distribution $(\mathcal{W}, q(\mathcal{W}))$, for some large enough constant c (depending on δ). Let $W^{(0)}$ denote the $K \times \ell$ matrix and $q^{(0)}$ the K-dimensional vector, both over $\mathcal{F}$, such that for each training example $(w, q(w))$, $W^{(0)}$ has $w \in \mathcal{F}^{\ell}$ as a row and $q^{(0)}$ has $q(w) \in \mathcal{F}$ as an entry. Note that each training example $(w, q(w))$, with $w = (w_1, w_2, \ldots, w_\ell)$, gives us a linear equation

$$w_1 v_1 + w_2 v_2 + \cdots + w_\ell v_\ell = q(w)$$

for $v = (v_1, v_2, \ldots, v_\ell) \in \mathcal{F}^{\ell}$. Thus from these K training examples, we obtain a system of K linear equations, denoted as $[W^{(0)}|q^{(0)}]$, and we would like to reduce the task of learning v to that of solving this system of linear equations. However, this system is highly noisy as about $1 - 1/2^m$ fraction of the equations are likely to be wrong, according to (5.1). We will roughly follow the approach of Gaussian Elimination (which works for noiseless systems of linear equations), but will make substantial changes in order to deal with our noisy case.

Our algorithm consists of two phases: the forward phase, shown in Figure 5.2, and the backward phase, shown in Figure 5.3. The forward phase works as follows, which is similar to an approach of Blum *et al.* [7]. Starting from the system $[W^{(0)}|q^{(0)}]$ of linear equations, we use several iterations to produce smaller and smaller systems with fewer and fewer variables, until we have a small enough system which we can afford to solve using brute force. More precisely, we choose the parameters

$$T = \log \sqrt{k/\log k} \;\text{ and }\; d = k/(mT),$$

divide each row of $W^{(0)}$ into ℓ/d blocks, with each block containing d elements in $\mathcal{F}$, and proceed in T iterations, as shown in Figure 5.2. Note that after iteration t, we have the system $[W^{(t)}|q^{(t)}]$ which has $\ell - dt$ variables and $K^{(t)}$ equations, with

$$K^{(t)} \geq K - t2^{md} = 2^{c(k/\log k)} - t2^{k/T} \geq 2^{c(k/\log k)}/2 = K/2,$$

for a large enough constant c. The key is to guarantee that the system still contains a good fraction of correct equations. Let

$$\delta_0 = \delta/2 \;\text{ and }\; \delta_t = (\delta_{t-1}/2)^2 \;\text{ for } t \geq 1,$$

```
1. For $t$ from 1 to $T$ do
    (a) Partition the equations of $[W^{(t-1)}|q^{(t-1)}]$ into at most $2^{md}$ groups according
        to their first blocks in $W^{(t)}$ (same block value in the same group).
    (b) Within each group, randomly select an equation which we call pivot.
    (c) Within each group, subtract each equation by the pivot.
    (d) Remove the pivots and delete the first block from each equation. Let
        $[W^{(t)}|q^{(t)}]$ be the resulting system of equations.
```

Figure 5.2: Forward Phase

```
1. Set $V^{(T)} = \mathcal{F}^{(n-k)/m}$, and set $V^{(t)} = \emptyset$ for $0 \le t \le T-1$.
2. For $t$ from $T-1$ down to 0 do
    (a) For any $z \in \mathcal{F}^d \times V^{(t+1)}$ which is $\delta_t$-good for $[W^{(t)}|q^{(t)}]$, include $z$ into $V^{(t)}$
        if $|V^{(t)}| \le L$, and break otherwise.
3. Output $V^{(0)}$.
```

Figure 5.3: Backward Phase

and a simple induction shows that

$$\delta_t \ge (\delta/8)^{2^t} \ge 2^{-0.1c(k/\log k)} = K^{-0.1},$$

for a large enough constant c. We say that any $z \in \mathcal{F}^{\ell-dt}$ is δ_t-*good* for the system $[W^{(t)}|q^{(t)}]$ if it satisfies at least $1/2^m + \delta_t$ fraction of equations in the system. Let $v^{(t)} \in \mathcal{F}^{\ell-dt}$ denote v without its first t blocks, and we call the forward phase *good* if for every t, $v^{(t)}$ is δ_t-good for $[W^{(t)}|q^{(t)}]$. Lemma 5.4.1 below, which will be proved in Subsection 5.4.1, guarantees that the forward phase is good with a significant probability.

Lemma 5.4.1. *The forward phase is good with probability at least* $2^{-O(k/\log k)}$.

For the backward phase, we start from the last system $[W^{(T)}|q^{(T)}]$ produced by the forward phase, and work backward on larger and larger systems produced in the forward phase to obtain solutions for more and more variables. More precisely, we go from $t = T-1$ down to $t = 0$, and while in iteration t, we try to find all possible solutions which extend solutions from iteration $t+1$ and are δ_t-good for $[W^{(t)}|q^{(t)}]$, as shown in Figure 5.3. However, in order to bound the running time, we will stop including the solutions once their number grows beyond the threshold

$$L = 2^{n-k+m+T+2\log(1/\delta_T)} = 2^{n-k+O(k/\log k)}.$$

If this happens, we may fail to include the actual solution v in our final list. Call the backward phase *good* if for every t, the number of such δ_t-good solutions for $[W^{(t)}|q^{(t)}]$ is at most L. Lemma 5.4.2 below, which will be proved in Subsection 5.4.2, guarantees that the backward phase is indeed good with a high probability.

Lemma 5.4.2. *The backward phase is not good with probability at most* $2^{-\Omega(k)}$.

From Lemma 5.4.1 and Lemma 5.4.2, the probability that both the forward and backward phases are good is at least

$$2^{-O(k/\log k)} - 2^{-\Omega(k)} = 2^{-O(k/\log k)}.$$

Assuming that both phases are good, a simply induction shows that $v^{(t)} \in V^{(t)}$ for any t and hence $v \in V^{(0)}$. Thus, we have shown that any fixed v satisfying the bound in (5.1) is contained in the list $V^{(0)}$ of size at most L with probability $2^{-O(k/\log k)}$. We can further reduce the probability of missing this v to $2^{-\omega(n)}$ by repeating the process $2^{O(k/\log k)}$ times, and take the union of the produced lists. Then a union bound shows that some v satisfying (5.1) is not included in the final output with probability only $o(1)$.

Finally, let us measure the complexity of our algorithm. First, $K \leq 2^{O(k/\log k)}$ training examples are sampled from the distribution $(\mathcal{W}, q(\mathcal{W}))$. Then the forward phase runs in time

$$T \cdot \mathrm{poly}(K) \leq 2^{O(k/\log k)},$$

and the backward phase runs in time $T \cdot 2^{md} \cdot L \cdot K$, which is at most

$$O(\log(k/\log k)) \cdot 2^{O(k/\log k)} \cdot 2^{n-k+O(k/\log k)} \cdot 2^{O(k/\log k)} \leq 2^{n-k+O(k/\log k)}.$$

The process is repeated for $2^{O(k/\log k)}$ times, and the total running time is

$$2^{O(k/\log k)} \cdot \left(2^{O(k/\log k)} + 2^{n-k+O(k/\log k)}\right) \leq 2^{n-k+O(k/\log k)}.$$

Thus, we have Theorem 5.3.2. To complete the proof, it remains to prove Lemma 5.4.1 and Lemma 5.4.2, which we do next.

5.4.1 Analysis on the Forward Phase

In this subsection, we prove Lemma 5.4.1, which shows that the forward phase is good with a significant probability.

First, by the Chernoff bound of Lemma 2.7.6, we know that $v = v^{(0)}$ satisfies less than $1/2^m + \delta_0$ fraction of equations in $[W^{(0)}|q^{(0)}]$ with probability at most $2^{-\Omega(\delta_0^2 K)} = o(1)$. That is, $v^{(0)}$ is δ_0-good for $[W^{(0)}|q^{(0)}]$ with probability $1-o(1)$. Next, we need the following lemma.

Lemma 5.4.3. *In the forward phase, if $v^{(t-1)}$ is δ_{t-1}-good for $[W^{(t-1)}|q^{(t-1)}]$, then $v^{(t)}$ is δ_t-good for $[W^{(t)}|q^{(t)}]$ with probability at least δ_t.*

Proof. Let $M = 2^m$ and $\tau = \delta_{t-1}$. Assume that $v^{(t-1)}$ is τ-good, so it satisfies at least $\frac{1}{M} + \tau$ fraction of equations in the system $[W^{(t-1)}|q^{(t-1)}]$. Partition equations in the system $[W^{(t-1)}|q^{(t-1)}]$ into groups according to their first blocks, as in Step 1(a) of the forward phase. Suppose group i contains p_i fraction of equations in $[W^{(t-1)}|q^{(t-1)}]$ and $v^{(t-1)}$ satisfies $\frac{1}{M} + \tau_i$ fraction of equations in the group, for some $\tau_i \in [-\frac{1}{M}, 1 - \frac{1}{M}]$. Then we have

$$\sum_i p_i \cdot \left(\frac{1}{M} + \tau_i\right) \geq \frac{1}{M} + \tau. \tag{5.2}$$

We would like to count the expected fraction of new equations satisfied by $v^{(t-1)}$. Before doing that, let us first count the fraction with respect to the system obtained before Step 1(d) (before removing pivots). Fix any group i. For $u \in \mathcal{F}$, let α_u denote the fraction of equations in the group which are off by a value u in the sense that

$$q^{(t-1)}(w^{(t-1)}) = \langle v^{(t-1)}, w^{(t-1)} \rangle_m + u.$$

Note that for $v^{(t-1)}$ to satisfy a new equation, which is the difference between two equations, these two involved equations must be off by the same value. Therefore, the expected fraction of new satisfied equations in this group is $\sum_u \alpha_u^2$, which under the constraint $\alpha_0 = \frac{1}{M} + \tau_i$ achieves its minimum when $\alpha_u = \frac{1}{M} - \frac{\tau_i}{M-1}$ for all other $u \neq 0$. Hence, after one iteration, the expected fraction of new equations in group i (before removing pivots) satisfied by $v^{(t-1)}$ is at least

$$\begin{aligned}\left(\frac{1}{M}+\tau_i\right)^2+(M-1)\cdot\left(\frac{1}{M}-\frac{\tau_i}{M-1}\right)^2 &= M\cdot\left(\frac{1}{M}\right)^2+\frac{2-2}{M}\cdot\tau_i+\frac{M-1+1}{M-1}\cdot\tau_i^2\\ &\geq \frac{1}{M}+\tau_i^2.\end{aligned}$$

Combing all groups together, the expected fraction of satisfied equations overall (before removing the pivots) is at least

$$\sum_i p_i\left(\frac{1}{M}+\tau_i^2\right)=\frac{1}{M}+\sum_i p_i\tau_i^2\geq\frac{1}{M}+\left(\sum_i p_i\tau_i\right)^2\geq\frac{1}{M}+\tau^2,$$

where the first inequality is due to Jensen inequality of Lemma 2.7.1, and the second inequality uses the bound $\sum_i p_i\tau_i\geq\tau$ implied by that in (5.2).

To get the expected fraction of satisfied equations in the final system $[W^{(t)}|q^{(t)}]$, after performing Step 1(d), observe that we only need to discard at most $2^{md}=2^{O(k/\log k)}$ equations, each with measure $\frac{1}{K^{(t)}}\leq\frac{2}{K}$, so the total discarded measure, denoted as μ, is at most

$$2^{md}\cdot\frac{2}{K}\leq 2^{O(k/\log k)}\cdot 2\cdot 2^{-c(k/\log k)}\leq\frac{\tau^2}{2},$$

for a large enough constant c. As a result, the expected fraction of equations in $[W^{(t)}|q^{(t)}]$ satisfied by $v^{(t)}$ is at least

$$\frac{1}{1-\mu}\cdot\left(\frac{1}{M}+\tau^2-\mu\right)\geq\frac{1}{M}+\tau^2-\mu\geq\frac{1}{M}+\frac{\tau^2}{2}=\frac{1}{M}+2\delta_t,$$

by recalling that $\tau=\delta_{t-1}$ and $\delta_t=(\delta_{t-1}/2)^2$. Finally, by the Markov inequality of Corollary 2.7.4, we have the lemma. □

Then by Lemma 5.4.3 and an induction, the forward phase is good with probability at least

$$(1-o(1))\prod_{t=1}^{T}\delta_t\geq(1-o(1))\prod_{t=1}^{T}(\delta/8)^{2^t}\geq(1-o(1))\,(\delta/8)^{2^{T+1}}\geq 2^{-O(k/\log k)}.$$

This proves Lemma 5.4.1.

5.4.2 Analysis on the Backward Phase

In this subsection, we will prove Lemma 5.4.2, which shows that the backward phase is indeed good with a high probability. Recall that a solution is δ_t-good for the system

$[W^{(t)}|q^{(t)}]$ if it satisfies at least $1/2^m + \delta_t$ fraction of the equations. For any t such that $0 \leq t \leq T-1$, consider the following event

- $B^{(t)}$: the number of δ_t-good solutions for $[W^{(t)}|q^{(t)}]$ exceeds L.

Thus, our goal is to show that

$$\Pr\left[\bigvee_{t=0}^{T-1} B^{(t)}\right] \leq 2^{-\Omega(k)}.$$

We will prove this by a union bound, so our goal is reduced to bounding each $\Pr[B^{(t)}]$ for $0 \leq t \leq T-1$.

To get a quick idea, let us first consider how to bound $\Pr[B^{(0)}]$. Note that since Ext^2 is a good *strong-two-source-extractor* and $\mathcal{W}$ has a high min-entropy, Lemma 2.5.3 guarantees that the number of z satisfying the probability bound $\Pr_{w \in \mathcal{W}}[q(w) = \langle z, w\rangle_m] \geq 1/2^m + \delta_0/2$ is at most L. Any other z is very unlike to be δ_0-good for $[W^{(0)}|q^{(0)}]$ by a Chernoff bound because each row of $W^{(0)}$ is sampled independently from $\mathcal{W}$. Since $B^{(0)}$ happens only when any such z (not satisfying that probability bound) is δ_0-good, a union bound shows that $\Pr[B^{(0)}]$ is indeed small.

Now for $t \geq 1$, to follow this idea to bound $\Pr[B^{(t)}]$, we would also like the distribution of $W^{(t)}$ to have the nice property that each of its rows comes independently from a high min-entropy source. Unfortunately, this is not true in general,[1] and a much more involved analysis is needed. Our approach is to consider the condition of restricting pivots in the first t iterations and to show that the distribution of $W^{(t)}$ conditioned on most restrictions is close to a distribution with the nice property. More precisely, a restriction of the pivots in an iteration includes fixing the indices and the values of some rows as pivots while leaving other rows free, and we say that two distributions are γ-close if the probabilities of any event according to the two distributions are within a multiplicative factor of γ from each other.

Observe that the distribution of $W^{(t)}$ can be generated alternatively in two passes as follows. In the first pass, we select a restriction of pivots in the first t iterations, denoted

[1] This is true in the simple case considered by [7] that one has $\mathcal{W} = \mathcal{U}_n$ to start with. In this case, for each t, one can easily show that each row of $W^{(t)}$ does come independently from the uniform distribution $\mathcal{U}_{n-tmd}$.

as $R^{(1)}, \ldots, R^{(t)}$, by running the forward phase on the matrix $W^{(0)}$ sampled from $\mathcal{W}$ and collecting the pivots in each iteration. In the second pass, we sample a matrix $W^{(0)}$ from $\mathcal{W}$ and then run the forward phase accordingly for t iterations to derive the matrix $W^{(t)}$, under the condition that the pivots selected in the t iterations match $R^{(1)}, \ldots, R^{(t)}$. Let $\tilde{\mathcal{D}}^{(t)}$ denote such a conditional distribution of $W^{(t)}$ with respect to $R^{(1)}, \ldots, R^{(t)}$. Now consider the following event about $\tilde{\mathcal{D}}^{(t)}$, over the distribution of $R^{(1)}, \ldots, R^{(t)}$ selected in the first pass.

- $E^{(t)}$: the distribution $\tilde{\mathcal{D}}^{(t)}$ is γ_t-close to some distribution $\mathcal{D}^{(t)}$ which has $K^{(t)}$ rows, each coming independently from a distribution $\mathcal{W}^{(t)}$ with $\mathrm{H}_\infty(\mathcal{W}^{(t)}) \geq k - t(md+1)$, for some $\gamma_t \leq K^{2^{md}(2^t-1)} \leq 2^{\sqrt{K}}$.

The following lemma, which will be proved later, shows that when conditioned on $E^{(t)}$, the probability of $B^{(t)}$ is indeed small.

Lemma 5.4.4. *For any t such that $0 \leq t \leq T-1$, $\Pr[B^{(t)} \mid E^{(t)}] \leq 2^{-\Omega(k)}$.*

Proof. Let us first count the number of solution z such that

$$\Pr_{w \in \mathcal{W}^{(t)}} \left[q^{(t)}(w) = \langle z, w \rangle_m \right] \geq 1/2^m + \delta_t/2.$$

Let Z denote the set of such z's. Note that $\mathcal{W}^{(t)}$ is a source over $\mathcal{F}^{\ell - td} = \{0,1\}^{(\ell - td)m}$ with $\mathrm{H}_\infty(\mathcal{W}^{(t)}) \geq k - t(md+1)$. Thus by Theorem 3.2.1 and Lemma 2.5.3, we have

$$|Z| \leq 2^{(\ell - td)m + m + 2\log(2/\delta_t) - 2 - (k - t(md+1))} = 2^{n-k+m+t+2\log(1/\delta_t)} \leq L.$$

This means that for the event $B^{(t)}$ to happen, some $z \notin Z$ must be δ_t-good.

Consider any restriction $R^{(1)}, \ldots, R^{(t)}$ such that the event $E^{(t)}$ happens. If we sample the matrix $W^{(t)}$ according to the distribution $\mathcal{D}^{(t)}$, which has each row coming independently from $\mathcal{W}^{(t)}$, then any fixed $z \notin Z$ is δ_t-good (satisfying at least $1/2^m + \delta_t$ fraction of equations in $[W^{(t)} | q^{(t)}]$) with probability at most $2^{-\Omega(\delta_t^2 K^{(t)})}$ by the Chernoff bound of Lemma 2.7.6, and a union bound shows that

$$\Pr_{\mathcal{D}^{(t)}} \left[B^{(t)} \right] \leq \Pr_{\mathcal{D}^{(t)}} \left[\exists z \notin Z : z \text{ is } \delta_t\text{-good} \right] \leq 2^n \cdot 2^{-\Omega(\delta_t^2 K^{(t)})} \leq 2^{-\Omega(K^{0.8})}.$$

Now if we sample $W^{(t)}$ according to the distribution $\tilde{\mathcal{D}}^{(t)}$, which is γ_t-close to $\mathcal{D}^{(t)}$ (given that $E^{(t)}$ happens), the probability is only scaled up by a factor γ_t. Thus, we have

$$\Pr_{\tilde{\mathcal{D}}^{(t)}}\left[B^{(t)}\right] \leq \gamma_t \cdot 2^{-\Omega(K^{0.8})} \leq 2^{\sqrt{K}} \cdot 2^{-\Omega(K^{0.8})} \leq 2^{-\Omega(k)}.$$

Since the bound holds for any restriction $R^{(1)}, \ldots, R^{(t)}$ such that the event $E^{(t)}$ happens, we have the lemma. □

Next, we would like to show that $E^{(t)}$ happens with high probability. Note that for $t = 0$, the event $E^{(0)}$ always happens because the initial distribution $\tilde{\mathcal{D}}^{(0)}$ has the nice property itself, so we have $\mathcal{D}^{(0)} = \tilde{\mathcal{D}}^{(0)}$ and $\gamma_0 = 1$. For $1 \leq t \leq T-1$, we use induction to show that

$$\Pr\left[\neg E^{(t)}\right] \leq \Pr\left[\neg E^{(t)} \mid E^{(t-1)}\right] + \Pr\left[\neg E^{(t-1)}\right] \leq \sum_{\tau=1}^{t} \Pr\left[\neg E^{(\tau)} \mid E^{(\tau-1)}\right],$$

and then we rely on the following lemma.

Lemma 5.4.5. *For any t such that $1 \leq t \leq T-1$, $\Pr[\neg E^{(t)} \mid E^{(t-1)}] \leq 2^{-\Omega(K)}$.*

Proof. Let us consider any restriction $R^{(1)}, \ldots, R^{(t-1)}$ such that the event $E^{(t-1)}$ happens, and we will show that $E^{(t)}$ happens with high probability, over the selection of $R^{(t)}$. More precisely, the assumption that $E^{(t-1)}$ happens means that we start iteration t from the distribution $\tilde{\mathcal{D}}^{(t-1)}$ which is close to some nice distribution $\mathcal{D}^{(t-1)}$, and our task is to show that with high probability over the selection of $R^{(t)}$, the resulting conditional distribution $\tilde{\mathcal{D}}^{(t)}$ after iteration t is close to another nice distribution $\mathcal{D}^{(t)}$, so that $E^{(t)}$ happens. For this, we need to figure out which of these $R^{(t)}$'s make $E^{(t)}$ happen.

Note that for a restriction $R^{(t)}$, the corresponding distribution $\tilde{\mathcal{D}}^{(t)}$ is obtained by applying Steps 1(c) and 1(d) on the matrix $W^{(t-1)}$ sampled from $\tilde{\mathcal{D}}^{(t-1)}$ under the condition that it is consistent with $R^{(t)}$. The restriction $R^{(t)}$ fixes some $r \leq 2^{md}$ rows of the matrix $W^{(t-1)}$ as pivots and it has the effect on the distribution $\tilde{\mathcal{D}}^{(t-1)}$ that all the rows of $W^{(t-1)}$ must belong to the r groups of those r rows. Consider the following event, over the selection of $R^{(t)}$.

- $G^{(t)}$: those r groups have a combined measure of $\rho \geq 1/2$ in the distribution $\mathcal{W}^{(t-1)}$.

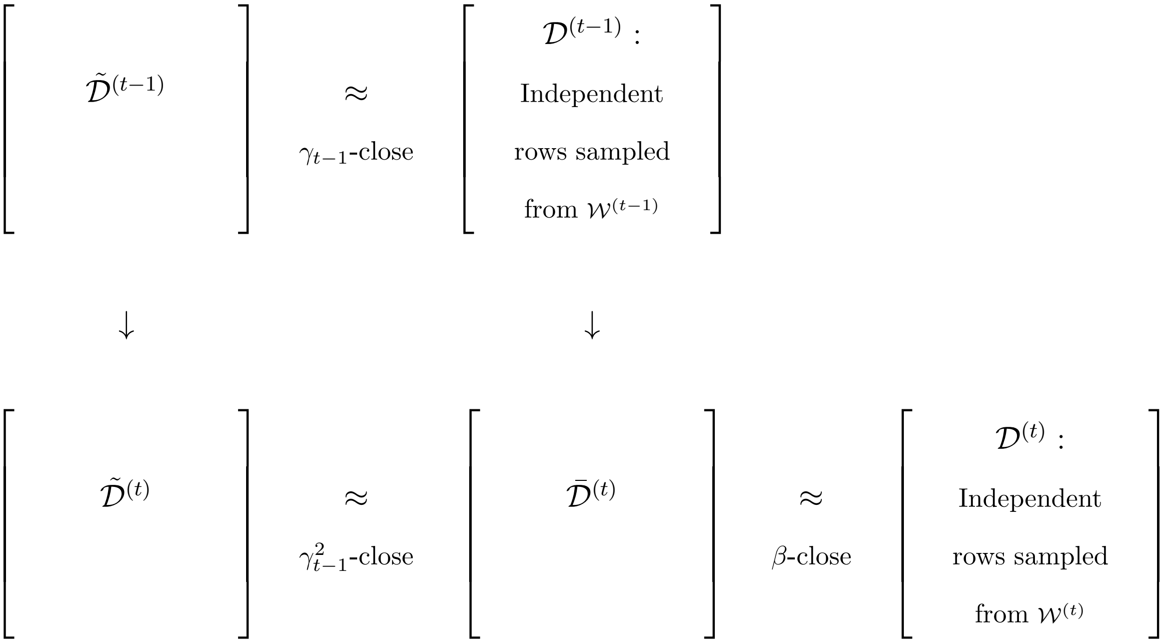

Figure 5.4: If $\tilde{\mathcal{D}}^{(t-1)}$ is close to $\mathcal{D}^{(t-1)}$, then $\tilde{\mathcal{D}}^{(t)}$ is close to $\mathcal{D}^{(t)}$, conditioned on $R^{(t)}$

We will show that if $G^{(t)}$ happens then $E^{(t)}$ happens. For this, let us consider any fixed restriction $R^{(t)}$ such that $G^{(t)}$ happens, and let us also use $R^{(t)}$ to denote the event that the pivots chosen in iteration t match those in $R^{(t)}$. Our approach is illustrated in Figure 5.4.

First, let us consider the case of starting iteration t from the nice distribution $\mathcal{D}^{(t-1)}$, instead of $\tilde{\mathcal{D}}^{(t-1)}$, conditioned on $R^{(t)}$, and let $\bar{\mathcal{D}}^{(t)}$ be the resulting distribution after iteration t. The following claim shows that $\bar{\mathcal{D}}^{(t)}$ is in fact close to a nice distribution.

Claim 5.4.6. For some $\beta \leq K^{2^{md}}$, the distribution $\bar{\mathcal{D}}^{(t)}$ is β-close to some nice distribution $\mathcal{D}^{(t)}$ which satisfies the condition in the event $E^{(t)}$.

Proof. Recall that we have fixed a restriction $R^{(t)}$ which fixes some r rows as pivots such that the event $G^{(t)}$ happens, and we also use $R^{(t)}$ to denote the event that the pivots selected during iteration t match those in the restriction $R^{(t)}$. In this claim, we consider the situation of starting iteration t from the nice distribution $\mathcal{D}^{(t-1)}$ conditioned on the event $R^{(t)}$.

First, let us see how the distribution $\mathcal{D}^{(t-1)}$ is affected by the conditioning on $R^{(t)}$. Consider any fixed matrix M of $K^{(t)} = K^{(t-1)} - r$ rows, insert the rows of $R^{(t)}$ at the proper places to get a fixed matrix $W^{(t-1)}$ of $K^{(t-1)}$ rows, and let us also use $W^{(t-1)}$ to denote the event that a randomly sampled matrix from $\mathcal{D}^{(t-1)}$ equals this matrix $W^{(t-1)}$. If the matrix has a row not in the r groups of $R^{(t)}$, then $\Pr_{\mathcal{D}^{(t-1)}}\left[W^{(t-1)} \mid R^{(t)}\right] = 0$. Otherwise, we have

$$\Pr_{\mathcal{D}^{(t-1)}}\left[W^{(t-1)} \mid R^{(t)}\right] = \frac{\left(\prod_{j=1}^{K^{(t)}} \mathcal{W}^{(t-1)}(M_j)\right)\cdot\left(\prod_{i=1}^{r}\frac{1}{\ell_i'+1}\right)}{\sum_{\ell_1+\cdots+\ell_r=K^{(t)};\ell_i\geq 0}\binom{K^{(t)}}{\ell_1,\cdots,\ell_r}\cdot\left(\prod_{i=1}^{r}\rho_i^{\ell_i}\right)\cdot\left(\prod_{i=1}^{r}\frac{1}{\ell_i+1}\right)},$$

where $\mathcal{W}^{(t-1)}(M_j)$ is the measure of the j'th row of M in $\mathcal{W}^{(t-1)}$, ℓ_i' is the number of rows of M in group i, and ρ_i is the measure of group i in $\mathcal{W}^{(t-1)}$. Note that for some $\alpha_1, \alpha_2 \in [K^{-r}, 1]$, the numerator equals

$$\left(\prod_{j=1}^{K^{(t)}} \mathcal{W}^{(t-1)}(M_j)\right)\cdot\alpha_1,$$

while the denominator equals

$$\sum_{\ell_1+\cdots+\ell_r=K^{(t)};\ell_i\geq 0}\binom{K^{(t)}}{\ell_1,\cdots,\ell_r}\cdot\left(\prod_{i=1}^{r}\rho_i^{\ell_i}\right)\cdot\alpha_2 = \left(\sum_{i=1}^{r}\rho_i\right)^{K^{(t)}}\cdot\alpha_2 = \rho^{K^{(t)}}\cdot\alpha_2,$$

where $\sum_{i=1}^{r}\rho_i = \rho \geq 1/2$ as we assume that the event $G^{(t)}$ happens. As a result, for $\beta = \frac{\alpha_1}{\alpha_2} \in [K^{-r}, K^{r}]$, we have

$$\Pr_{\mathcal{D}^{(t-1)}}\left[W^{(t-1)} \mid R^{(t)}\right] = \left(\prod_{j=1}^{K^{(t)}}\frac{\mathcal{W}^{(t-1)}(M_j)}{\rho}\right)\cdot\beta.$$

Note that the first factor above can be seen as the probability when we sample each row of the matrix independently according a new distribution $\tilde{\mathcal{W}}^{(t-1)}$, which is the distribution $\mathcal{W}^{(t-1)}$ restricted to those r groups of $R^{(t)}$ and normalized by their measure ρ. Thus, although the conditioning on the event $R^{(t)}$ may destroy the independence so that we can no longer see each row as coming independently from $\mathcal{W}^{(t-1)}$, we can somehow have the independence restored by considering another distribution $\tilde{\mathcal{W}}^{(t-1)}$ with some distortion factor β. More precisely, we have shown that the distribution $\mathcal{D}^{(t-1)}$ conditioned on the event $R^{(t)}$ is β-close to a nice distribution, denoted as $\hat{D}^{(t-1)}$, which has each of its remaining row (not fixed by $R^{(t)}$) coming independently from $\tilde{\mathcal{W}}^{(t-1)}$, with

$$\mathrm{H}_\infty(\tilde{\mathcal{W}}^{(t-1)}) \geq \mathrm{H}_\infty(\mathcal{W}^{(t-1)}) - \log(1/\rho) \geq k - (t-1)(md+1) - 1.$$

Next, let us see what the resulting distribution $\bar{\mathcal{D}}^{(t)}$ will be when Steps 1(c) and 1(d) are performed on the distribution $\mathcal{D}^{(t-1)}$ conditioned on $R^{(t)}$. Again, we first consider the case of applying the two steps on the nice distribution $\hat{\mathcal{D}}^{(t-1)}$ instead. When we perform Step 1(c) to subtract from each row its corresponding pivot, which is a fixed value, each resulting row still remains independent from others. However, the distribution of each resulting row is now changed to another distribution which may have a smaller min-entropy than that of $\tilde{\mathcal{W}}^{(t-1)}$, because different initial rows after subtracting their corresponding pivots may result in the same value. Still, the number of such initial rows can be at most 2^{md} since no two such rows can come from the same group, which implies that the min-entropy only decreases by at most md. Then after performing Step 1(d) to remove the pivots and delete the first blocks, the resulting matrix has each row coming independently from some distribution $\mathcal{W}^{(t)}$ with min-entropy at least

$$\mathrm{H}_\infty(\tilde{\mathcal{W}}^{(t-1)}) - md \geq k - t(md+1).$$

That is, after performing Steps 1(c) and 1(d) on the distribution $\hat{\mathcal{D}}^{(t-1)}$, the resulting distribution, denoted as $\mathcal{D}^{(t)}$, satisfies the condition in event $E^{(t)}$. Finally, let us get back to the actual case of starting with the distribution $\mathcal{D}^{(t-1)}$ conditioned on $R^{(t)}$. Since it is β-close to $\hat{\mathcal{D}}^{(t-1)}$, the resulting distribution $\bar{\mathcal{D}}^{(t)}$ after applying the two steps is β-close to the corresponding resulting distribution $\mathcal{D}^{(t)}$, which proves the claim.

□

Next, let us go back to the actual situation of starting iteration t from the distribution $\tilde{\mathcal{D}}^{(t-1)}$, instead of $\mathcal{D}^{(t-1)}$ as we did in the above claim. Using the assumption that $\tilde{\mathcal{D}}^{(t-1)}$ is close to $\mathcal{D}^{(t-1)}$, our next claim shows that when we start iteration t from the distribution $\tilde{\mathcal{D}}^{(t-1)}$ conditioned on $R^{(t)}$, the resulting distribution $\tilde{\mathcal{D}}^{(t)}$ is close to the distribution $\bar{\mathcal{D}}^{(t)}$.

Claim 5.4.7. The distribution $\tilde{\mathcal{D}}^{(t)}$ is γ_{t-1}^2-close to the distribution $\bar{\mathcal{D}}^{(t)}$.

Proof. In this claim, we go back to the actual situation of starting iteration t from the distribution $\tilde{\mathcal{D}}^{(t-1)}$, instead of $\mathcal{D}^{(t-1)}$ as we just did. We would like to show that the resulting distribution $\tilde{\mathcal{D}}^{(t)}$ when starting from $\tilde{\mathcal{D}}^{(t-1)}$ is γ_{t-1}^2-close to the distribution $\bar{\mathcal{D}}^{(t)}$ when starting from $\mathcal{D}^{(t-1)}$. For this, it suffices to show that for any event A, the probabilities of $\Pr_{\mathcal{D}^{(t-1)}}\left[A \mid R^{(t)}\right]$ and $\Pr_{\tilde{\mathcal{D}}^{(t-1)}}\left[A \mid R^{(t)}\right]$ are within a multiplicative factor of γ_{t-1}^2. This is true because from the fact that $\mathcal{D}^{(t-1)}$ and $\tilde{\mathcal{D}}^{(t-1)}$ are γ_{t-1}-close, we know that $\Pr_{\mathcal{D}^{(t-1)}}\left[R^{(t)}\right]$ and $\Pr_{\tilde{\mathcal{D}}^{(t-1)}}\left[R^{(t)}\right]$ are within a multiplicative factor of γ_{t-1}, and so are $\Pr_{\mathcal{D}^{(t-1)}}\left[A \wedge R^{(t)}\right]$ and $\Pr_{\tilde{\mathcal{D}}^{(t-1)}}\left[A \wedge R^{(t)}\right]$. □

From these two claims, we can conclude that $\tilde{\mathcal{D}}^{(t)}$ is γ_t-close to $\mathcal{D}^{(t)}$, for $\gamma_t = \gamma_{t-1}^2 \beta \le \gamma_{t-1}^2 K^{2^{md}}$, which by induction is at most

$$K^{2^{md}(2^t-2)} K^{2^{md}} \le K^{2^{md}(2^t-1)} \le 2^{\sqrt{K}}.$$

This implies that for any restriction $R^{(t)}$ such that the event $G^{(t)}$ happens, the event $E^{(t)}$ must happen as well. Therefore, the probability that $E^{(t)}$ does not happen is at most the probability that $G^{(t)}$ does not happen, which we bound by the following claim.

Claim 5.4.8. The probability over the selection of $R^{(t)}$ that $G^{(t)}$ does not happen is at most $2^{-\Omega(K)}$.

Proof. Note that the restriction $R^{(t)}$ can be selected by sampling a matrix $W^{(t-1)}$ according to the distribution $\tilde{\mathcal{D}}^{(t-1)}$ and then applying Steps 1(a) and 1(b) to select the pivots. Thus, the probability that $G^{(t)}$ does not happen is at most the probability that all the $K^{(t-1)}$ rows of $W^{(t-1)}$ lie in some r groups with a combined measure of $\rho \le 1/2$ in the distribution $\mathcal{W}^{(t-1)}$.

Again, let us first consider the case of sampling $W^{(t-1)}$ according to the distribution $\mathcal{D}^{(t-1)}$, instead of $\tilde{\mathcal{D}}^{(t-1)}$. Note that there are at most $2^{2^{md}}$ ways of choosing the r groups with a combined measure of $\rho \le 1/2$ in $\mathcal{W}^{(t-1)}$, and the probability that all the $K^{(t-1)} \ge K/2$ independent rows lie in any particular choice of such r groups is at most $(1/2)^{K/2}$. Then a union bound shows that the probability of having $\rho \le 1/2$ is at most

$$2^{2^{md}} \cdot (1/2)^{K/2} \le 2^{-\Omega(K)}.$$

Next, let us go back to actual case of sampling $W^{(t-1)}$ according to the distribution $\tilde{\mathcal{D}}^{(t-1)}$. Note that the probability of having $\rho \le 1/2$ according to $\tilde{\mathcal{D}}^{(t-1)}$ can only be larger than that according to $\mathcal{D}^{(t-1)}$ by at most a factor of γ_{t-1}, and hence it is still at most

$$\gamma_{t-1} \cdot 2^{-\Omega(K)} \le 2^{\sqrt{K}} \cdot 2^{-\Omega(K)} \le 2^{-\Omega(K)}.$$

□

We have shown that for any restriction $R^{(1)}, \ldots, R^{(t-1)}$ such that the event $E^{(t-1)}$ happens, the probability, over the selection of $R^{(t)}$, that the event $E^{(t)}$ does not happen is at most $2^{-\Omega(K)}$. This implies that $\Pr[\neg E^{(t)} \mid E^{(t-1)}] \leq 2^{-\Omega(K)}$, which proves Lemma 5.4.5.

□

From these two lemmas, we have that for any t such that $1 \leq t \leq T-1$,

$$\begin{aligned}
\Pr\left[B^{(t)}\right] &\leq \Pr\left[B^{(t)} \mid E^{(t)}\right] + \Pr\left[\neg E^{(t)}\right] \\
&\leq \Pr\left[B^{(t)} \mid E^{(t)}\right] + \sum_{\tau=1}^{t} \Pr\left[\neg E^{(\tau)} \mid E^{(\tau-1)}\right] \\
&\leq 2^{-\Omega(k)}.
\end{aligned}$$

For $t = 0$, we have

$$\Pr\left[B^{(0)}\right] = \Pr\left[B^{(0)} \mid E^{(0)}\right] \leq 2^{-\Omega(k)}.$$

As a result, a union bound gives us

$$\Pr\left[\bigvee_{t=0}^{T-1} B^{(t)}\right] \leq \sum_{t=0}^{T-1} \Pr\left[B^{(t)}\right] \leq T \cdot 2^{-\Omega(k)} = 2^{-\Omega(k)},$$

which proves Lemma 5.4.2.

Chapter 6

Extracting Computational Entropy from Computational Independent-Symbol Sources

In this chapter, we consider computational independent-symbol sources. In Section 6.1, we generalized the well-known Impagliazzo's hardcore set lemma. Then, in Section 6.2, we use the generalized hardcore set lemma to show that the extractor described in Section 4.1 also works for computational independent-symbol sources. Using the result of extractors for computational independent-symbol sources, we can generalize the well-known XOR lemma in Section 6.3. Finally, we prove the size upper bound on a binary hardcore set in any black-box construction in Section 6.4.

6.1 Generalized Hardcore Set Lemma

The well-known Impagliazzo's hardcore set lemma [26] says that if a function $f : \{0,1\}^\ell \to \{0,1\}$ is *mildly hard*, that is, any small circuits must fail to compute it correctly on more than a δ fraction of inputs, then there exists a *hardcore set* $H \subseteq \{0,1\}^\ell$ of density roughly δ, where the density of H is defined as $\rho(H) = |H|/2^\ell$, such that f is *extremely hard* on H, in the sense that any somewhat smaller circuits must fail to compute f correctly on more than a $\frac{1}{2} - \varepsilon$ fraction of inputs in H, for some small ε.

Lemma 6.1.1. *[26] Let $f : \{0,1\}^\ell \to \{0,1\}$ be a function such that for any $h \in \mathsf{SIZE}(s)$,*

$$\Pr_{x\in\{0,1\}^\ell}[h(x) \neq f(x)] > \delta.$$

Then for any $\varepsilon > 0$, there exists $H \subseteq \{0,1\}^\ell$ with $|H| = \delta \cdot 2^\ell$ such that for any $C \in \mathsf{SIZE}(\Omega(\varepsilon^2\delta^2 s))$,

$$\Pr_{x\in H}[C(x) \neq f(x)] > \frac{1}{2} - \varepsilon.$$

We extend it to the case that $f : \{0,1\}^\ell \to [D]$ is *mildly hard*, that is, any small circuit must fail to compute it correctly on more than a δ fraction of inputs, and show that there exist some *disjoint binary hardcore sets* $T_1, \cdots, T_r$ of total size at least $(\delta/2) \cdot 2^\ell$, where for every $i \in [r]$, $T_i \subseteq f^{-1}(I_i)$ for some $I_i \subseteq [D]$ with $|I_i| = 2$, such that f is *extremely hard* on $H = \cup_{i=1}^r T_i$, in the sense that any smaller circuit must fail to compute f correctly on more than a $\frac{1}{2} - \varepsilon$ fraction of inputs in H, for some small ε.

Lemma 6.1.2. *Let $f : \{0,1\}^\ell \to [D]$ be a function such that for any $h \in \mathsf{SIZE}(s)$,*

$$\Pr_{x\in\{0,1\}^\ell}[h(x) \neq f(x)] > \delta.$$

Then for any $\varepsilon > 0$, there exists some integer r satisfying that there are $I_1, \cdots, I_r \subseteq [D]$ with $|I_i| = 2$, and disjoint binary hardcore sets $T_1, \cdots, T_r$ where $T_i \subseteq f^{-1}(I_i)$, such that the size of $H = \cup_{i=1}^r T_i$ is $|H| \geq (\delta/2) \cdot 2^\ell$ and for any $C \in \mathsf{SIZE}(\Omega(\varepsilon^2\delta^2 s/D^6))$,

$$\Pr_{x\in H}[C(x) \neq f(x)] > \frac{1}{2} - \varepsilon.$$

Proof. First, we show that there exists one binary hardcore set. For every $I \subseteq [D]$ with $|I| = 2$, let

$$\begin{aligned}
\alpha_I &= \min_{h\in\mathsf{SIZE}(s/D^2)} \left\{ \Pr_{x\in\{0,1\}^\ell}[h(x) \neq f(x) | f(x) \in I] \right\} \\
\beta_I &= \Pr_{x\in\{0,1\}^\ell}[f(x) \in I] \\
g_I &= \arg \min_{h\in\mathsf{SIZE}(s/D^2)} \left\{ \Pr_{x\in\{0,1\}^\ell}[h(x) \neq f(x) | f(x) \in I] \right\}
\end{aligned}$$

Then using these g_I's, we construct a function $g : \{0,1\}^\ell \to [D]$ as shown in Figure 6.1, and to give an example, we show the decision tree of function g for the case of $D = 4$ in Figure 6.2.

```
• Input: $x \in \{0,1\}^\ell$

• Procedure:

    1. Set $s = 1$.
    2. For $t = 2$ to $D$ do
        (a) Compute $z = g_{\{s,t\}}(x)$.
        (b) If $z = t$, set $s = t$.
    3. Output $z$.
```

Figure 6.1: THE CONSTRUCTION OF THE FUNCTION $g : \{0,1\}^\ell \to [D]$

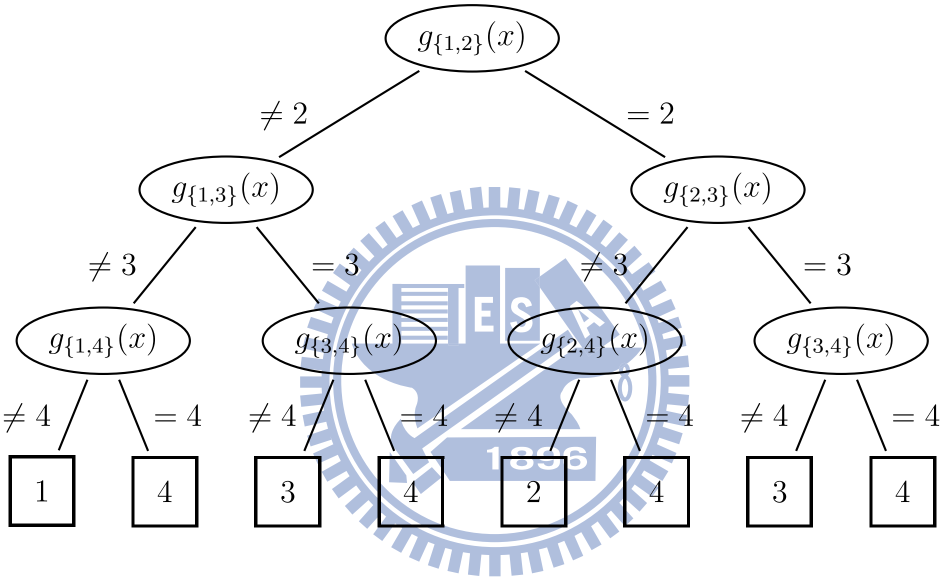

Figure 6.2: The decision tree of function g for $D = 4$

Since for every I, $g_I \in \mathsf{SIZE}(s/D^2)$, we obtain that $g \in \mathsf{SIZE}(s)$. Hence, we have that

$$\Pr_{x \in \{0,1\}^\ell}[g(x) \neq f(x)] > \delta. \tag{6.1}$$

Next, we claim that for any x, if $g_I(x) = f(x)$ for all I with $f(x) \in I$, then the function g can output the correct value $f(x)$. First suppose that $f(x) = 1$. To compute $g(x)$, we first compute $g_{\{1,2\}}(x)$, which is 1 according to our assumption that $g_I(x) = f(x)$ for all I with $f(x) \in I$. Next, we will consider $g_{\{1,3\}}(x)$, which is also 1. That is, in the procedure of computing $g(x)$, we will consider a sequence of values $g_{\{1,j\}}(x)$ for all $j \in \{2, 3, \cdots, D\}$. Since $g_{\{1,j\}}(x) = 1$ for all $j \in \{2, 3, \cdots, D\}$, the function g will output the correct value

$f(x) = 1$. Similarly, suppose that $f(x) = i$ for some $i \in \{2, 3, \cdots, D\}$. We must consider $g_{\{i_1,i\}}(x)$ for some $i_1 < i$ in the step 2 of Figure 6.1 with $t = i$. Since $g_{\{i_1,i\}}(x) = i$ according to our assumption, we will go on to consider $g_{\{i,i_2\}}(x)$ for all $i_2 > i$ and the function g will output the correct value $f(x) = i$.

Hence, we can bound the error probability of g as

$$\Pr_{x\in\{0,1\}^\ell}[g(x) \neq f(x)] \leq \Pr_{x\in\{0,1\}^\ell}[\exists I \text{ with } f(x) \in I, \text{ and } g_I(x) \neq f(x)]. \tag{6.2}$$

According to equation (6.1) and (6.2), we obtain that

$$\begin{aligned}
\delta &< \Pr_{x\in\{0,1\}^\ell}[g(x) \neq f(x)] \\
&\leq \Pr_{x\in\{0,1\}^\ell}[\exists I \text{ with } f(x) \in I, \text{ and } g_I(x) \neq f(x)] \\
&\leq \sum_I \Pr[f(x) \in I \text{ and } g_I(x) \neq f(x)] \\
&= \sum_I \Pr[g_I(x) \neq f(x) | f(x) \in I] \cdot \Pr[f(x) \in I] \\
&= \sum_I \alpha_I \beta_I.
\end{aligned}$$

Then, by an average argument, there exists some I, such that $\alpha_I\beta_I > \delta/D^2$, which implies that $\alpha_I > \delta/D^2$. Fix such I, and by the definition of α_I, we have that for any $h \in \mathsf{SIZE}(s/D^2)$, $\Pr_{x\in f^{-1}(I)}[h(x) \neq f(x)] \geq \alpha_I > \delta/D^2$. Then by Lemma 6.1.1, there exists a hardcore set $T_1 \subseteq f^{-1}(I)$ with size at least $\alpha_I\beta_I \cdot 2^\ell \geq \delta \cdot 2^\ell/D^2$ such that for any circuit $C \in \mathsf{SIZE}(\Omega(\varepsilon^2\delta^2 s/D^6))$,

$$\Pr_{x\in T_1}[C(x) \neq f(x)] > \frac{1}{2} - \varepsilon.$$

If T_1 is indeed much larger, say with $\rho(T_1) \geq \delta/2$, then we are done. Otherwise, we can continue the process on the remaining inputs as follows to amplify the size of hardcore. By excluding T_1, the remaining input must still have hardness at least $\delta - \rho(T_1)$ in the sense that for any circuit $h \in \mathsf{SIZE}(s)$, we must have $\Pr_{x\in\{0,1\}^\ell}[h(x) \neq f(x) \text{ and } x \notin T_1] \geq \delta - \rho(T_1)$. The reason is that otherwise using that h, we have

$$\Pr_{x\in\{0,1\}^\ell}[h(x) \neq f(x)] < \Pr_{x\in\{0,1\}^\ell}[h(x) \neq f(x) \text{ and } x \notin T_1] + \Pr_{x\in\{0,1\}^\ell}[x \in T_1] \leq \delta.$$

Therefore, if $\rho(T_1)$ is small, the remaining hardness is still large, so we can find a new binary hardcore set T_2, disjoint from T_1. Continuing in this way, we can find a sequence of hardcore sets $T_1, T_2, T_3, \cdots$, as long as the sum of density (in the whole space $\{0,1\}^\ell$) is small, say at most $\delta/2$. This means that when we stop, we have found a sequence of disjoint binary hardcore sets $T_1, \cdots, T_r$ for some integer r such that the density of $H = \cup_{i=1}^r T_r \geq \delta/2$. Moreover, since for any circuit $C \in \mathsf{SIZE}(\Omega(\varepsilon^2\delta^2 s/D^6))$, for each i, $\Pr_{x\in T_i}[C(x) \neq f(x)] > \frac{1}{2} - \varepsilon$, we obtain that for any circuit $C \in \mathsf{SIZE}(\Omega(\varepsilon^2\delta^2 s/D^6))$,

$$\Pr_{x\in H}[C(x) \neq f(x)] > \frac{1}{2} - \varepsilon.$$

□

6.2 Computational Extractors

In this section, we show that the extractor $\textsc{Ext}_0 : [D]^n \to [M]$ in Theorem 4.1.1, which is shown to be a good extractor for independent-symbol sources, also works for computational independent-symbol sources. Throughout this section, we see any symbol $\mathcal{V}_i \in [D]$ of the source as an element in $\mathbb{Z}_M$, and operation $+$ on elements in $\mathbb{Z}_M$ is understood as an operation over the group $\mathbb{Z}_M$. Recall that our extractor $\textsc{Ext}_0 : [D]^n \to [M]$ is defined as

$$\textsc{Ext}_0(\mathcal{V}_1, \cdots, \mathcal{V}_n) = \sum_{t\in[n]} \mathcal{V}_t.$$

Theorem 6.2.1. *Suppose that $M \geq D$ is a prime. For any $n, k \in \mathbb{N}$ with $k \geq \Omega(M^2 \log^2 D)$, the function $\textsc{Ext}_0 : [D]^n \to [M]$ defined above is an $(n, D, k, \varepsilon, s_1, s_2)$-computational-extractor, where $\varepsilon \leq O(M^2 \log n/k)$ and $s_2 = \Omega(s_1(\log n/nkD)^2)$.*

Proof. Let $(\mathcal{V}|\mathcal{X}) = (\mathcal{V}_1|\mathcal{X}_1) \circ \cdots \circ (\mathcal{V}_n|\mathcal{X}_n)$ be a computational (n, D, k, s_1)-source with for every $i \in [n]$, $\mathcal{V}_i = f_i(\mathcal{X}_i)$ for some function $f_i : \{0,1\}^{\ell_i} \to [D]$, satisfying that for any circuit $C \in \mathsf{SIZE}(s_1)$, $\Pr[C(\mathcal{X}_i) = \mathcal{V}_i] \leq 2^{-k_i}$ for some $0 \leq k_i \leq \log D$, and $\sum_{i=1}^n k_i = k$. Fix $\xi = M^2 \log n/nk > 0$.

First, we use the analysis in [51] to show that for any $i \in [n]$ with $k_i > 0$, there exists a source $\mathcal{Y}_i$ over $[D]$ such that no small circuit can distinguish the distributions $\mathcal{X}_i \circ f_i(\mathcal{X}_i)$ and $\mathcal{X}_i \circ \mathcal{Y}_i$.

Fix any $i \in [n]$ with $k_i > 0$. Recall that for any $C \in \mathsf{SIZE}(s_1)$,

$$\Pr[C(\mathcal{X}_i) \neq f_i(\mathcal{X}_i)] > 1 - 2^{-k_i}.$$

1. Sample $x \in \{0,1\}^{\ell_i}$.
2. If $x \in T_j^i$ for some $j \in [r_i]$, output a random value $y \in I_j^i$; otherwise, output $f_i(x)$.

Figure 6.3: The construction of the source $\mathcal{Y}_i$

For simplicity, let $\delta_i = 1 - 2^{-k_i}$. By Lemma 6.1.2, there exit some integer r_i, and $I_1^i, \cdots, I_{r_i}^i \subseteq [D]$ with $|I_j^i| = 2$ for any $j \in [r_i]$ such that there are disjoint binary hardcore sets $T_1^i, \cdots, T_{r_i}^i$ where $T_j^i \subseteq f_i^{-1}(I_j^i)$ satisfying that the size of $H^i = \cup_{j=1}^{r_i} T_j^i$ is $|H^i| = (\delta_i/2) \cdot 2^{\ell_i}$. Then we define a source $\mathcal{Y}_i$ as in Figure 6.3, and the following claim shows that no small circuit can distinguish the distributions $\mathcal{X}_i \circ f_i(\mathcal{X}_i)$ and $\mathcal{X}_i \circ \mathcal{Y}_i$.

Claim 6.2.2. For any $i \in [n]$ with $k_i > 0$, the source $\mathcal{Y}_i$ defined as in Figure 6.3 satisfying that for any $\varepsilon > 0$, no $(\varepsilon\delta_i)$-distinguisher in $\mathsf{SIZE}(\Omega(\varepsilon^2\delta_i^2 s_1/D^6))$ for the distributions $\mathcal{X}_i \circ f_i(\mathcal{X}_i)$ and $\mathcal{X}_i \circ \mathcal{Y}_i$.

Proof. Fix any $i \in [n]$ with $k_i > 0$. By way of contradiction, suppose that there exists a circuit $C \in \mathsf{SIZE}(\Omega(\varepsilon^2\delta_i^2 s_1/D^6))$ such that

$$\Pr[C(\mathcal{X}_i \circ f_i(\mathcal{X}_i)) = 1] - \Pr[C(\mathcal{X}_i \circ \mathcal{Y}_i) = 1] \geq \varepsilon\delta_i,$$

which implies that

$$\begin{aligned}
\varepsilon\delta_i &\leq \Pr\left[C(\mathcal{X}_i \circ f_i(\mathcal{X}_i)) = 1\right] - \Pr\left[C(\mathcal{X}_i \circ \mathcal{Y}_i) = 1\right] \\
&= \sum_{j=1}^{r_i} \left\{\Pr\left[C(\mathcal{X}_i \circ f_i(\mathcal{X}_i)) = 1 | \mathcal{X}_i \in T_j^i\right] - \Pr\left[C(\mathcal{X}_i \circ \mathcal{Y}_i) = 1 | \mathcal{X}_i \in T_j^i\right]\right\} \Pr\left[\mathcal{X}_i \in T_j^i\right] \\
&\quad + \left\{\Pr\left[C(\mathcal{X}_i \circ f_i(\mathcal{X}_i)) = 1 | \mathcal{X}_i \notin H^i\right] - \Pr\left[C(\mathcal{X}_i \circ \mathcal{Y}_i) = 1 | \mathcal{X}_i \notin H^i\right]\right\} \Pr\left[\mathcal{X}_i \notin H^i\right] \\
&= \sum_{j=1}^{r_i} \Pr\left[\mathcal{X}_i \in T_j^i\right] \left\{\Pr_{x \in T_j^i}\left[C(x \circ f_i(x)) = 1\right] - \Pr_{x \in T_j^i, y \in I_j^i}\left[C(x \circ y) = 1\right]\right\}. \qquad (6.3)
\end{aligned}$$

Let

$$t = \arg\max_j \left\{\Pr_{x \in T_j^i}\left[C(x \circ f_i(x)) = 1\right] - \Pr_{x \in T_j^i, y \in I_j^i}\left[C(x \circ y) = 1\right]\right\}.$$

Then by equation (6.3), we obtain that

$$\begin{aligned}
\varepsilon\delta_i &\le \sum_{j=1}^{r_i} \Pr[\mathcal{X}_i \in T_j^i] \left\{ \Pr_{x \in T_j^i}[C(x \circ f_i(x)) = 1] - \Pr_{x \in T_j^i, y \in I_j^i}[C(x \circ y) = 1] \right\} \\
&\le \left\{ \Pr_{x \in T_t^i}[C(x \circ f_i(x)) = 1] - \Pr_{x \in T_t^i, y \in I_t^i}[C(x \circ y) = 1] \right\} \cdot \sum_{j=1}^{r_i} \Pr[\mathcal{X}_i \in T_j^i] \\
&= \left\{ \Pr_{x \in T_t^i}[C(x \circ f_i(x)) = 1] - \Pr_{x \in T_t^i, y \in I_t^i}[C(x \circ y) = 1] \right\} \cdot \Pr[\mathcal{X}_i \in H^i],
\end{aligned}$$

which implies that

$$\Pr_{x \in T_t^i}[C(x \circ f_i(x)) = 1] - \Pr_{x \in T_t^i, y \in I_t^i}[C(x \circ y) = 1] \ge \frac{\varepsilon\delta_i}{\delta_i/2} = 2\varepsilon.$$

By the standard approach of obtaining predictors from distinguishers (or Lemma 2.5.4 with $m = 1$), we obtain that there is a predictor $P \in \mathsf{SIZE}(\Omega(\varepsilon^2\delta_i^2 s_1/D^6))$ with

$$\Pr_{x \in T_t^i}[P(x) = f_i(x)] \ge \frac{1 + 2\varepsilon}{2} = \frac{1}{2} + \varepsilon,$$

which contradicts the fact that T_t^i is a hardcore set. Therefore, we conclude that no $(\varepsilon\delta_i)$-distinguisher in $\mathsf{SIZE}(\Omega(\varepsilon^2\delta_i^2 s_1/D^6))$ for the distributions $\mathcal{X}_i \circ f_i(\mathcal{X}_i)$ and $\mathcal{X}_i \circ \mathcal{Y}_i$.

□

Hence, for any $i \in [n]$ with $k_i > 0$, by claim 6.2.2 with $\varepsilon = \xi/(1 - 2^{-k_i}) > 0$, we have that there exists a source $\mathcal{Y}_i$ such that no ξ-distinguisher in $\mathsf{SIZE}(\Omega(\xi^2 s_1/D^6))$ for the distributions $\mathcal{X}_i \circ \mathcal{V}_i$ and $\mathcal{X}_i \circ \mathcal{Y}_i$.

On the other hand, for $i \in [n]$ such that $k_i = 0$, we can set $\mathcal{Y}_i = \mathcal{V}_i$. Clearly, no ξ-distinguisher for the distributions $\mathcal{X}_i \circ \mathcal{V}_i$ and $\mathcal{X}_i \circ \mathcal{Y}_i$.

Claim 6.2.3. There is no $(n\xi)$-distinguisher in $\mathsf{SIZE}(\Omega(\xi^2 s_1/D^6))$ for the distributions $\mathcal{X}_{[1,n]} \circ \mathcal{V}_{[1,n]}$ and $\mathcal{X}_{[1,n]} \circ \mathcal{Y}_{[1,n]}$.

Proof. We prove it by the standard hybrid argument (see e.g. [18]). Let, for every $i = 0, 1, \cdots, n$,

$$\mathcal{H}_i = \mathcal{X}_{[1,n]} \circ \mathcal{V}_{[1,i]} \circ \mathcal{Y}_{[i+1,n]}.$$

Note that $\mathcal{H}_n = \mathcal{X}_{[1,n]} \circ \mathcal{V}_{[1,n]}$ and $\mathcal{H}_0 = \mathcal{X}_{[1,n]} \circ \mathcal{Y}_{[1,n]}$. Suppose that there exists an $(n\xi)$-distinguisher C in $\mathsf{SIZE}(\Omega(\xi^2 s_1/D^6))$ for the distributions $\mathcal{X}_{[1,n]} \circ \mathcal{V}_{[1,n]}$ and $\mathcal{X}_{[1,n]} \circ \mathcal{Y}_{[1,n]}$, that is

$$n\xi \leq \Pr[C(\mathcal{H}_n) = 1] - \Pr[C(\mathcal{H}_0) = 1] = \sum_{i=0}^{n-1} \Pr[C(\mathcal{H}_{i+1}) = 1] - \Pr[C(\mathcal{H}_i) = 1],$$

which implies that there exists $j \in \{0, 1, \cdots, n-1\}$ such that

$$\Pr[C(\mathcal{H}_{j+1}) = 1] - \Pr[C(\mathcal{H}_j) = 1] \geq \xi.$$

Then, we construct the distinguisher $C' \in \mathsf{SIZE}(\Omega(\xi^2 s_1/D^6))$ for the distributions $\mathcal{X}_{j+1} \circ \mathcal{V}_{j+1}$ and $\mathcal{X}_{j+1} \circ \mathcal{Y}_{j+1}$ as in Figure 6.4.

- Input: $x \circ \alpha$
- Procedure:
  1. Sample $x_i \circ v_i \in \mathcal{X}_i \circ \mathcal{V}_i$, for all $i \leq j$.
  2. Sample $x_i \circ y_i \in \mathcal{X}_i \circ \mathcal{Y}_i$ for all $i > j+1$.
  3. Output $C(x_{[1,j]} \circ x \circ x_{[j+2,n]} \circ v_{[1,j]} \circ \alpha \circ y_{[j+2,n]})$.

Figure 6.4: THE CONSTRUCTION OF THE DISTINGUISHER C'

By the definition of C', we obtain that

$$\Pr[C'(\mathcal{X}_{j+1} \circ \mathcal{V}_{j+1}) = 1] - \Pr[C'(\mathcal{X}_{j+1} \circ \mathcal{Y}_{j+1}) = 1] = \Pr[C(\mathcal{H}_{j+1}) = 1] - \Pr[C(\mathcal{H}_j) = 1] \geq \xi,$$

that is, C' is a ξ-distinguisher for the distributions $\mathcal{X}_{j+1} \circ \mathcal{V}_{j+1}$ and $\mathcal{X}_{j+1} \circ \mathcal{Y}_{j+1}$, and we have a contradiction. □

Claim 6.2.4. There exists no $(n\xi)$-distinguisher in $\mathsf{SIZE}(\Omega(\xi^2 s_1/D^6))$ for the distributions $\mathcal{X}_{[1,n]} \circ \sum_{i=1}^{n} \mathcal{V}_i$ and $\mathcal{X}_{[1,n]} \circ \sum_{i=1}^{n} \mathcal{Y}_i$.

Proof. By way of contradiction, assume that there exists a circuit $E \in \mathsf{SIZE}(\Omega(\xi^2 s_1/D^6))$ such that

$$\Pr\left[E\left(\mathcal{X}_{[1,n]} \circ \sum_{i=1}^{n} \mathcal{V}_i\right) = 1\right] - \Pr\left[E\left(\mathcal{X}_{[1,n]} \circ \sum_{i=1}^{n} \mathcal{Y}_i\right) = 1\right] \geq n\xi.$$

Then we can construct $E' \in \mathsf{SIZE}(\Omega(\xi^2 s_1/D^6))$ to distinguish the distributions $\mathcal{X}_{[1,n]} \circ \mathcal{V}_{[1,n]}$ and $\mathcal{X}_{[1,n]} \circ \mathcal{Y}_{[1,n]}$ as in Figure 6.5.

```
• Input: $x_{[1,n]} \circ \alpha_{[1,n]}$

• Procedure:

    1. Compute $\alpha = \sum_{i=1}^{n} \alpha_i$.

    2. Output $E(x_{[1,n]} \circ \alpha)$.
```

Figure 6.5: The construction of the distinguisher E'

Hence, we have that

$$\begin{aligned} & \Pr[E'(\mathcal{X}_{[1,n]} \circ \mathcal{V}_{[1,n]}) = 1] - \Pr[E'(\mathcal{X}_{[1,n]} \circ \mathcal{Y}_{[1,n]}) = 1] \\ = \; & \Pr\left[E\left(\mathcal{X}_{[1,n]} \circ \sum_{i=1}^{n} \mathcal{V}_i\right) = 1\right] - \Pr\left[E\left(\mathcal{X}_{[1,n]} \circ \sum_{i=1}^{n} \mathcal{Y}_i\right) = 1\right] \\ \geq \; & n\xi, \end{aligned}$$

which means that E' is an $(n\xi)$-distinguisher for the distributions $\mathcal{X}_{[1,n]} \circ \mathcal{V}_{[1,n]}$ and $\mathcal{X}_{[1,n]} \circ \mathcal{Y}_{[1,n]}$. This contradicts Claim 6.2.3. □

Next, we show that the sources $\mathcal{X}_{[1,n]} \circ \sum_{i=1}^{n} \mathcal{Y}_i$ and $\mathcal{X}_{[1,n]} \circ \mathcal{U}_{[M]}$ are close.

Claim 6.2.5. $\Delta(\mathcal{X}_{[1,n]} \circ \sum_{i=1}^{n} \mathcal{Y}_i, \mathcal{X}_{[1,n]} \circ \mathcal{U}_{[M]}) \leq e^{-\Omega(k/M^2 \log D)}$, for $k \geq \Omega(M^2 \log^2 D)$.

Proof. For each $i \in [n]$ with $k_i > 0$, let Z_i be the indicator random variable for the event of $\mathcal{X}_i \in H^i$. For the rest of $i \in [n]$, let $Z_i = 0$. Define $Z = \sum_{i \in [n]} Z_i$.

Note that for any $i \in [n]$ with $k_i > 0$, $\mathrm{H}_\infty(\mathcal{Y}_i | Z_i = 1) = 1$. Recall that for each $i \in [n]$ with $k_i > 0$, $\Pr[\mathcal{X}_i \in H^i] = \delta_i/2 = (1 - 2^{-k_i})/2$. Then

$$\begin{aligned} \mathrm{E}[Z] &= \sum_{i=1}^{n} (1 - 2^{-k_i})/2 \\ &\geq \frac{1}{2} \cdot \left[n - \left\lfloor \frac{k}{\log D} \right\rfloor \cdot \frac{1}{D} - \left(n - \left\lfloor \frac{k}{\log D} \right\rfloor\right) \cdot 1\right] \\ &= \frac{1}{2} \cdot \left[\left\lfloor \frac{k}{\log D} \right\rfloor \left(1 - \frac{1}{D}\right)\right]. \end{aligned}$$

Since Z_i's are independent, by the Chernoff bound of Lemma 2.7.6, we have that

$$\Pr\left[Z \leq \frac{\mathrm{E}[Z]}{2}\right] \leq \Pr\left[|Z - \mathrm{E}[Z]| \geq \frac{\mathrm{E}[Z]}{2}\right] \leq e^{-\Omega(\mathrm{E}[Z])} \leq e^{-\Omega(k/\log D)}.$$

On the other hand, $\mathrm{H}_\infty(\mathcal{Y}_{[1,n]}|Z > \mathrm{E}[Z]/2) \geq \sum_{i=1}^{n}(1-2^{-k_i})/4$. Therefore, by Theorem 4.1.1,

$$\Delta\left(\left(\mathcal{X}_{[1,n]} \circ \sum_{i=1}^{n} \mathcal{Y}_i \,\middle|\, Z > \frac{\mathrm{E}[Z]}{2}\right), \left(\mathcal{X}_{[1,n]} \circ \mathcal{U}_{[M]} \,\middle|\, Z > \frac{\mathrm{E}[Z]}{2}\right)\right) \leq e^{-\Omega(k/M^2 \log D)},$$

for $k \geq \Omega(M^2 \log^2 D)$. Hence, we have

$$\begin{aligned} & \Delta\left(\mathcal{X}_{[1,n]} \circ \sum_{i=1}^{n} \mathcal{Y}_i, \mathcal{X}_{[1,n]} \circ \mathcal{U}_{[M]}\right) \\ \leq\ & \Pr\left[Z \leq \frac{\mathrm{E}[Z]}{2}\right] + \Delta\left(\left(\mathcal{X}_{[1,n]} \circ \sum_{i=1}^{n} \mathcal{Y}_i \,\middle|\, Z > \frac{\mathrm{E}[Z]}{2}\right), \left(\mathcal{X}_{[1,n]} \circ \mathcal{U}_{[M]} \,\middle|\, Z > \frac{\mathrm{E}[Z]}{2}\right)\right) \\ \leq\ & e^{-\Omega(k/M^2 \log D)}, \end{aligned}$$

for $k \geq \Omega(M^2 \log^2 D)$. □

The above claim implies that no $e^{-\Omega(k/M^2 \log D)}$-distinguisher (without any complexity bound) for the distributions $\mathcal{X}_{[1,n]} \circ \sum_{i=1}^{n} \mathcal{Y}_i$ and $\mathcal{X}_{[1,n]} \circ \mathcal{U}_{[M]}$. Then by Claim 6.2.4 with $\xi = M^2 \log n/nk > 0$, we have that there exists no $((M^2 \log n/k) + e^{-\Omega(k/M^2 \log D)})$-distinguisher in $\mathsf{SIZE}(\Omega(s_1(\log n/nkD)^2))$ for the distributions $\mathcal{X}_{[1,n]} \circ \textsc{Ext}(\mathcal{V}_1, \cdots, \mathcal{V}_n)$ and $\mathcal{X}_{[1,n]} \circ \mathcal{U}_{[M]}$. □

6.3 Generalized XOR Lemma

In this section, we will show that the above result about extractors for computational independent-symbol sources can generalize the well-known XOR lemma [59], which says that if $f : \{0,1\}^\ell \to \{0,1\}$ is "mildly hard" for small circuits, then $F(x_1, \cdots, x_t) \equiv \oplus_{i=1}^{t} f(x_i)$ for sufficiently large t, is "extremely hard" for smaller size circuits.

To prove the generalized XOR lemma, we will need the following lemma about obtaining distinguishers from predictors.

Lemma 6.3.1. *For any source $\mathcal{Z}$ over $\{0,1\}^\ell$ and any function $b : \{0,1\}^\ell \to [M]$, if there is a predictor P such that $\Pr_{z \in \mathcal{Z}}[P(z) = b(z)] \geq 1/M + \varepsilon$, then there is an ε-distinguisher E with P as oracle which calls P once and runs in time $O(m)$ for the distributions $\mathcal{Z} \circ b(\mathcal{Z})$ and $\mathcal{Z} \circ \mathcal{U}_{[M]}$.*

- Input: $z \circ \alpha$
- Procedure:
  1. Compute $P(z)$.
  2. If $P(z) = \alpha$, then output 1; otherwise, output 0.

Figure 6.6: The construction of the distinguisher E^P

Proof. Consider the distinguisher E^P as in Figure 6.6.

Then we have

$$\begin{aligned} & \Pr[E^P(\mathcal{Z} \circ b(\mathcal{Z})) = 1] - \Pr\left[E^P\left(\mathcal{Z} \circ \mathcal{U}_{[M]}\right) = 1\right] \\ = \; & \Pr[P(\mathcal{Z}) = b(\mathcal{Z})] - \Pr\left[P(\mathcal{Z}) = \mathcal{U}_{[M]}\right] \\ \geq \; & \frac{1}{M} + \varepsilon - \frac{1}{M} = \varepsilon. \end{aligned}$$

Hence, E^P is an ε-distinguisher for the distributions $\mathcal{Z} \circ b(\mathcal{Z})$ and $\mathcal{Z} \circ \mathcal{U}_{[M]}$. □

Then, we see any symbol $\mathcal{V}_i \in [D]$ of the source as an element in $\mathbb{Z}_M$, and operation $+$ on elements in $\mathbb{Z}_M$ is understood as an operation over the group $\mathbb{Z}_M$.

Theorem 6.3.2. *Suppose that $s \geq \Omega\left(m \cdot (n^2 D/\log n)^2\right)$. For $i \in [n]$, let $f_i : \{0,1\}^{\ell_i} \to [D]$ be a function such that for any circuit $h \in \mathsf{SIZE}(s)$,*

$$\Pr_{x_i \in \{0,1\}^{\ell_i}}[h(x_i) \neq f_i(x_i)] > \delta_i.$$

If $\delta = \sum_{i=1}^n \delta_i \geq \Omega(M^2 \log D)$, then for any circuit $C \in \mathsf{SIZE}(\Omega(s(\log n/nD\delta)^2))$,

$$\Pr_{x_1 \in \{0,1\}^{\ell_1}, \cdots, x_n \in \{0,1\}^{\ell_n}}\left[C\left(x_{[1,n]}\right) = \sum_{i=1}^n f_i(x_i)\right] < \frac{1}{M} + O\left(\frac{M^2 \log n}{\delta}\right).$$

Proof. First, note that by the similar argument in Theorem 6.2.1 with $\xi = \log nM^2/\delta n$, we have that for $\delta = \sum_{i=1}^n \delta_i \geq \Omega(M^2 \log D)$, there exists no $O(M^2 \log n/\delta)$-distinguisher in $\mathsf{SIZE}(\Omega(s(\log n/nD\delta)^2))$ for the distributions $\mathcal{X}_{[1,n]} \circ \sum_{i=1}^n f_i(\mathcal{X}_i)$ and $\mathcal{X}_{[1,n]} \circ \mathcal{U}_{[M]}$.

Suppose that there exists a circuit $C \in \mathsf{SIZE}(\Omega(s(\log n/nD\delta)^2))$ such that

$$\Pr_{x_1 \in \{0,1\}^{\ell_1}, \cdots, x_n \in \{0,1\}^{\ell_n}}\left[C(x_1, \cdots, x_n) = \sum_{i=1}^n f_i(x_i)\right] \geq \frac{1}{M} + O\left(\frac{M^2 \log n}{\delta}\right).$$

Then by Lemma 6.3.1, there is a $O(M^2 \log n/\delta)$-distinguisher for the distributions $\mathcal{X}_{[1,n]} \circ \sum_{i=1}^n f_i(x_i)$ and $\mathcal{X}_{[1,n]} \circ \mathcal{U}_{[M]}$, which is a contradiction. □

6.4 Hardcore Set Size in Black-Box Constructions

For the generalized hardcore set lemma in Section 6.1, one may wonder whether we can find a larger binary hardcore set, for example, a set with density δ/D. In this section, we give an upper bound on the size of binary hardcore sets of black-box construction. First, we introduce a black-box construction of a hardcore set.

Definition 6.4.1. We say that an oracle algorithm $\textsc{Dec}^{(\cdot)}$ is a *black-box* (δ, ε, D) *- construction* of a hardcore set, if the following holds. Given any function $f : \{0,1\}^\ell \to [D]$, where $\ell = \Omega(\log D)$, and a family of functions $G = \{g_I | I \subseteq [D] \text{ with } |I| = 2\}$ satisfying that for each $g_I \in G$ and $H \subseteq f^{-1}(I)$ with size s, $\Pr_{x \in H}[g_I(x) \neq f(x)] \leq (1-\varepsilon)/2$, then $\Pr_{x \in \{0,1\}^\ell}[\textsc{Dec}^G(x) \neq f(x)] \leq \delta$. We call s the size complexity of black-box construction.

Now, we give an upper bound on the size of binary hardcore sets.

Theorem 6.4.2. *Suppose that $\Omega(1/D^{c_1}) \leq \delta$ for some constant c_1, $\varepsilon \leq 1/5$ and $D \geq 4$. Then, any black-box (δ, ε, D)-construction must have size complexity $O(\delta 2^\ell / D^2)$.*

Proof. We use a probabilistic method to show that there exist a function f and a family of functions $G = \{g_I | I \subseteq [D] \text{ with } |I| = 2\}$ satisfying that for each $g_I \in G$ and $H \subseteq f^{-1}(I)$ with size $10\delta 2^\ell / D(D-1)$, $\Pr_{x \in H}[g_I(x) \neq f(x)] \leq (1-\varepsilon)/2$, but $\Pr_{x \in \{0,1\}^\ell}[\textsc{Dec}^G(x) \neq f(x)] \geq \delta/2$.

Suppose that $\textsc{Dec}$ is a (δ, ε, D)-black-box construction. We choose a random function f and a random family of functions $G = \{g_I | I \subseteq [D] \text{ and } |I| = 2\}$ as follows. First, we pick $\binom{D}{2}$ disjoint sets $A_{\{1,2\}}, A_{\{1,3\}}, \cdots, A_{\{D-1,D\}} \subseteq \{0,1\}^\ell$, each with size $4\delta \cdot 2^\ell / D(D-1)$, and then partition $\{0,1\}^\ell \setminus \left(\bigcup_{i_1 < i_2} A_{\{i_1,i_2\}}\right)$ into D sets: $B_1, B_2 \cdots, B_D$ (note that B_i could be empty). We define f as in Figure 6.7, and g_I for each $I = \{i_1, i_2\} \subseteq [D]$ as in Figure 6.8.

Next, we claim that for each $g_I \in G$ and $H \subseteq f^{-1}(I)$ with size $10\delta 2^\ell / D(D-1)$, g_I predicts f well in H.

Claim 6.4.3. For each $g_I \in G$, $H \subseteq f^{-1}(I)$ with size $10\delta 2^\ell / D(D-1)$, and $\varepsilon \leq 1/5$,

$$\Pr_{x \in H}[g_I(x) \neq f(x)] \leq (1-\varepsilon)/2.$$

Proof. Fix any $I = \{i_1, i_2\} \subseteq [D]$, and $H \subseteq f^{-1}(I)$ with size $10\delta 2^\ell / D(D-1)$. For each $x \in H$, let $Z_{x,f,G}$ be the indicator random variable for the event of $g_I(x) \neq f(x)$. Note

- Input: $x \in \{0,1\}^\ell$
- Procedure:
  1. If $x \in A_{\{i_1,i_2\}}$ for some $i_1 < i_2 \in [D]$, let $\Pr[f(x) = i] = 1/2$ for $i \in \{i_1, i_2\}$.
  2. If $x \in B_j$ for some $j \in [D]$, set $f(x) = j$.

Figure 6.7: The function f

- Input: $x \in \{0,1\}^\ell$
- Procedure:
  1. If $x \in A_I$, let $\Pr[g_I(x) = i] = 1/2$ for $i \in \{i_1, i_2\}$.
  2. If $x \in B_{i_1}$ or $A_{\{i_1,j\}}$ for some $j \in [D] \setminus I$, set $g_I(x) = i_1$.
  3. If $x \in B_{i_2}$ or $A_{\{i_2,j\}}$ for some $j \in [D] \setminus I$, set $g_I(x) = i_2$.
  4. Otherwise, let $\Pr[g_I(x) = i] = 1/D$ for any $i \in [D]$.

Figure 6.8: The function g_I for $I = \{i_1, i_2\}$

that for $x \in H \cap A_I$, $\Pr_{f,g_I}[Z_{x,f,G} = 1] = 1/2$. On the other hand, consider $x \in H \setminus A_I$. If $x \in f^{-1}(i_1) \setminus A_I \subseteq B_{i_1} \cup \left(\bigcup_{j \notin I} A_{\{i_1,j\}}\right)$, then we have $g_I(x) = i_1$, and similarly for $x \in f^{-1}(i_2) \setminus A_I$. That is, for $x \in H \setminus A_I$, $Z_{x,f,G} = 0$. Hence,

$$\sum_{x \in H} Z_{x,f,G} = \sum_{x \in H \cap A_I} Z_{x,f,G} \le \sum_{x \in A_I} Z_{x,f,G} \le \frac{4\delta 2^\ell}{D(D-1)}.$$

Therefore, for $\varepsilon \le 1/5$,

$$\Pr_{x \in H}[g_I(x) \neq f(x)] = \frac{1}{|H|} \sum_{x \in H} Z_{x,f,G} \le \frac{D(D-1)}{10\delta 2^\ell} \cdot \frac{4\delta 2^\ell}{D(D-1)} = \frac{2}{5} \le \frac{1-\varepsilon}{2}.$$

□

Then, we claim that f and G are likely to satisfy that $\Pr_{x \in \{0,1\}^\ell}[\textsc{Dec}^G(x) \neq f(x)] \ge \delta/2$. For each $x \in \{0,1\}^\ell$, let $B_{x,f,G}$ be the indicator random variable for the event of $\textsc{Dec}^G(x) \neq f(x)$.

Claim 6.4.4. $\Pr_{f,G}\left[\sum_{x \in \{0,1\}^\ell} B_{x,f,G} < \delta 2^\ell/2\right] = o(1)$.

Proof. Note that for any $x \in A_I$ for some $I \subseteq [D]$ with $|I| = 2$, we have that

$$\Pr_{f,G}[B_{x,f,G} = 1] \geq 1/2.$$

Hence, by the Chernoff bound of Lemma 2.7.6, we get

$$\Pr_{f,G}\left[\sum_{x \in \{0,1\}^\ell} B_{x,f,G} < \frac{\delta 2^\ell}{2}\right] \leq \Pr_{f,G}\left[\sum_{x \in \cup_I A_I} B_{x,f,G} < \frac{\delta 2^\ell}{2}\right] < e^{-\Omega(\delta 2^\ell)} = o(1),$$

for $\ell = \Omega(\log D)$ and $\delta > 1/D^{c_1}$ for some constant c_1. □

From Claim 6.4.3, and 6.4.4, we conclude that there exist f and $G = \{g_I | I \subseteq [D] \text{ with } |I| = 2\}$ such that for each $g_I \in G$ and $H \subseteq f^{-1}(I)$ with size $10\delta 2^\ell / D(D-1)$, $\Pr_{x \in H}[g_I(x) \neq f(x)] \leq (1-\varepsilon)/2$, but $\Pr_{x \in \{0,1\}^\ell}[\textsc{Dec}^G(x) \neq f(x)] \geq \delta/2$.

□

6.5 Open Problems

In section 6.2, we show that our extractor for independent-symbol sources still works for computational independent-symbol sources. We would like to find a better extractor for computational independent-symbol sources or show that our extractor is optimal.

On the other hand, there are several ways to prove the well-known XOR lemma [20], and one is through the hardcore set lemma. In Section 6.3, we show how to prove the generalized XOR lemma using the generalized hardcore set lemma. It would be interesting to consider other proofs for the XOR lemma to prove the generalized XOR lemma.

Chapter 7

Conclusion and Future Works

In this thesis, we consider the problem of deterministically extracting almost perfect random bits from several classes of random sources. First, we consider multiple weakly random sources that are mutually independent. We generalize the well-known leftover hash lemma, and this lemma gives us a way to extract randomness from two independent sources as long as two sources have enough min-entropy. We also extend our construction to extract randomness from $t \geq 3$ independent sources as long as two of them have enough min-entropy. One nice feature is that the extractor still works even with all but one source exposed. Moreover, we apply our extractor for a cryptographic task in which a group of parties want to agree on a secret key for group communication over an insecure channel, without using ideal local randomness.

We also consider the independent-symbol sources which are the sources lie in between multiple independent sources and bit-fixing sources. Each independent-symbol source consists of a sequence of n independent symbols from $\{0,1\}^d$, and the only randomness guarantee on such a source is that the whole source has min-entropy k. We give an explicit deterministic extractor which extracts about $\Omega(\log k)$ bits, for any $n, d, k \in \mathbb{N}$. 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 [33] and Gabizon et al. [17] 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 non-explicit 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.

Then, we go to the other direction to look for a more general class of sources from which seedless extraction is still possible. The sources we consider have the form of a conditional distribution $(f(\mathcal{X})|\mathcal{X})$, for some function f and some distribution $\mathcal{X}$, and we say that such a source has computational min-entropy k if any circuit of size 2^k can only predict $f(x)$ correctly with probability at most 2^{-k} given input x sampled from $\mathcal{X}$. We first show that it is impossible to have a seedless extractor for one single source of this kind. Then we show that it becomes possible if we are allowed a seed which is weakly random (instead of perfectly random) but contains some statistical min-entropy, or even a seed which is not random at all but contains some computational min-entropy. This can be seen as a step toward extending the study of multi-source extractors from the traditional, statistical setting to a computational setting. We reduce the task of constructing such extractors to a problem in learning theory: learning linear functions under arbitrary distribution with adversarial noise. For this problem, we provide a learning algorithm, which may have interest of its own.

Finally, we consider computational (n, D, k, s)-sources, which, just as (n, D, k)-sources, consist of n mutually independent parts, $(f_1(\mathcal{X}_1)|\mathcal{X}_1), \cdots, (f_n(\mathcal{X}_n)|\mathcal{X}_n)$, each $f_i(\mathcal{X}_i)$ of length d such that for each i if given input x_i sampled from $\mathcal{X}_i$, any circuit of size s can only predict $f_i(x_i)$ with probability at most 2^{-k_i} for some $k_i \leq d$, and the sum of k_i's is k. Note that we can set the circuit size as a separate parameter to define the computational independent-symbol sources. We generalize the well-known hardcore set lemma to show that our extractor for independent-symbol sources still works for computational independent-symbol sources. In addition, the result of extractors for computational independent-symbol sources implies a generalization of the well-known XOR lemma. Finally, we show an upper bound on the size of a binary hardcore set in any black-box construction.

Since the proofs of Lemma 4.4.3 and 5.4.2 are complicated, in the future, we would like to simplify these proofs. Moreover, we will go on to construct better extractors for these classes of sources or prove that these extractors are optimal.

In addition, in the proof of the lower bound on entropy loss for independent-symbol sources, we provide a size lower bound on an ”almost” t-wise independent space, and this immediately implies a size lower bound on any approximate t-wise independent space. It may be interesting to find a better size lower bound on an approximate t-wise independent space.

On the other hand, there are several ways to prove the well-known XOR lemma [20], and one is through the hardcore set lemma. In Chapter 6, we generalize the hardcore set lemma to prove the generalized XOR lemma. It may be interesting to consider other proofs for the XOR lemma to prove the generalized XOR lemma.

Bibliography

[1] Noga Alon, László Babai, and Alon Itai. A fast and simple randomized parallel algorithm for the maximal independent set problem. *Journal of Algorithms*, 7(4):567–583, 1986.

[2] Noga Alon, Oded Goldreich, Johan Håstad, and René Peralta. Simple constructions of almost k-wise independent random variables. In *Proc. IEEE 31st Annual IEEE Symposium on Foundations of Computer Science (FOCS'90)*, pages 544–553, 1990.

[3] Boaz Barak, Russell Impagliazzo, and Avi Wigderson. Extracting randomness using few independent sources. In *Proc. IEEE 45th Annual IEEE Symposium on Foundations of Computer Science (FOCS'04)*, pages 384–393, Rome, Italy, October 2004.

[4] Boaz Barak, Guy Kindler, Ronen Shaltiel, Benny Sudakov, and Avi Wigderson. Simulating Independence: New Constructions of Condensers, Ramsey graphs, Dispersers, and Extractors. In *Proc. 37th Annual ACM Symposium on Theory of Computing (STOC'05)*, pages 1–10, Baltimore, MD, May 2005.

[5] Boaz Barak, Ronen Shaltiel, and Avi Wigderson. Computational analogues of entropy. In *Proc. of APPROX 2003 and RANDOM 2003*, pages 200–215, 2003.

[6] Mihir Bellare and John Rompel. Randomness-efficient oblivious sampling. In *Proc. IEEE 35th Annual Symposium on Foundations of Computer Science (FOCS'94)*, pages 276–287, Santa Fe, New Mexico, November 1994.

[7] Avrim Blum, Adam Kalai, and Hal Wasserman. Noise-tolerant learning, the parity problem, and the statistical query model. *J. ACM*, 50(4):506–519, 2003.

[8] Jean Bourgainu. More on the sum-product phenomenon in prime fields and its applications. *International Journal of Number Theory*, 1(1):1–32, 2005.

[9] J. Lawrence Carter and Mark N. Wegman. Universal classes of hash functions. In *Proc. 9th Annual ACM Symposium on Theory of Computing (STOC'77)*, pages 106–112, 1977.

[10] Benny Chor and Oded Goldreich. Unbiased bits from sources of weak randomness and probabilistic communication complexity. *SIAM J. Comput.*, 17(2):230–261, April 1988.

[11] Benny Chor, Oded Goldreich, Johan Håstad, Joel Friedman, Steven Rudich, and Roman Smolensky. The bit extraction problem of t-resilient functions. In *Proc. IEEE 26th Annual Symposium on Foundations of Computer Science (FOCS'85)*, pages 396–407, 1985.

[12] Philip J. Davis. *Circulant Matrices*. John Wiley, 1979.

[13] Yevgeniy Dodis, Ariel Elbaz, Roberto Oliveira, and Ran Raz. Improved randomness extraction from two independent sources. In *APPROX-RANDOM*, pages 334–344, Cambridge, MA, USA, August 2004.

[14] Yevgeniy Dodis and Roberto Oliveira. On extracting private randomness over a public channel. In *APPROX-RANDOM*, pages 252–263, Princeton, NY, USA, August 2003.

[15] Yevgeniy Dodis, Rafail Ostrovsky, Leonid Reyzin, and Adam Smith. Fuzzy extractors: How to generate strong keys from biometrics and other noisy data. *SIAM J. Comput.*, 38(1):97–139, 2008.

[16] Vitaly Feldman, Parikshit Gopalan, Subhash Khot, and Ashok Kumar Ponnuswami. New results for learning noisy parities and halfspaces. In *Proc. 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS'06)*, pages 563–574, 2006.

[17] Ariel Gabizon, Ran Raz, and Ronen Shaltiel. Deterministic extractors for bit-fixing

sources by obtaining an independent seed. *SIAM Journal on Computing*, 36(4):1072–1094, 2006.

[18] Oded Goldreich. *Foundations of Cryptography: Volume 1, Basic Tools.* Cambridge University Press, Cambridge, 2001.

[19] Oded Goldreich and Leonid A. Levin. A hard-core predicate for all one-way functions. In *Proc. 21st Annual ACM Symposium on Theory of Computing (STOC'89)*, pages 25–32, 1989.

[20] Oded Goldreich, Noam Nisan, and Avi Wigderson. On Yao's **XOR** lemma. *Electronic Colloquium on Computational Complexity (ECCC)*, 2(50), 1995.

[21] Oded Goldreich, Ronitt Rubinfeld, and Madhu Sudan. Learning polynomials with queries: the highly noisy case. *SIAM J. Disc. Math.*, 13(4):535–570, 2000.

[22] Eldon R. Hansen. *A Table of Series and Products.* Prentice-Hall, Englewood Cliffs, N.J., 1975.

[23] Johan Håstad, Russell Impagliazzo, Leonid A. Levin, and Michael Luby. A pseudorandom generator from any one-way function. *SIAM J. Comput.*, 28(4):1364–1396, 1999.

[24] Shlomo Hoory, Nathan Linial, and Avi Wigderson. Expander graphs and their applications. *Bulletin (New series) of the American Mathematical Society*, 43(4).

[25] Chun-Yuan Hsiao, Chi-Jen Lu, and Leonid Reyzin. Conditional computational entropy, or toward separating pseudoentropy from compressibility. In *Proc. Advances in Cryptology - EUROCRYPT*, pages 169–186, 2007.

[26] Russell Impagliazzo. Hard-core distributions for somewhat hard problems. In *Proc. 36th Annual IEEE Symposium on Foundations of Computer Science (FOCS'95)*, pages 538–545, 1995.

[27] Russell Impagliazzo, Ragesh Jaiswal, Valentine Kabanets, and Avi Wigderson. Uniform direct product theorems: simplified, optimized, and derandomized. In *Proc.*

40th Annual ACM Symposium on Theory of Computing (STOC'08), pages 579–588, 2008.

[28] Russell Impagliazzo, Leonid A. Levin, and Michael Luby. Pseudo-random generation from one-way functions. In *Proc. 21st Annual ACM Symposium on Theory of Computing (STOC'89)*, pages 12–24, 1989.

[29] Russell Impagliazzo, Ronen Shaltiel, and Avi Wigderson. Extractors and pseudo-random generators with optimal seed length. In *Proc. 32nd Annual ACM Symposium on Theory of Computing (STOC'00)*, pages 1–10, 2000.

[30] Stasys Jukna. *Extremal Combinatorics.* Springer-Verlag, 2001.

[31] Adam Tauman Kalai, Yishay Mansour, and Elad Verbin. On agnostic boosting and parity learning. In *Proc. 40th Annual ACM Symposium on Theory of Computing (STOC'08)*, pages 629–638, 2008.

[32] Jesse Kamp, Anup Rao, Salil P. Vadhan, and David Zuckerman. Deterministic extractors for small-space sources. In *Proc. 38rd Annual ACM Symposium on Theory of Computing (STOC'06)*, pages 691–700, 2006.

[33] Jesse Kamp and David Zuckerman. Deterministic extractors for bit-fixing sources and exposure-resilient cryptography. *SIAM Journal on Computing*, 36(5):1231–1247, 2007.

[34] Robert König and Ueli M. Maurer. Generalized strong extractors and deterministic privacy amplification. In *Proc. Cryptography and Coding*, pages 322–339, 2005.

[35] Chia-Jung Lee, Chi-Jen Lu, Shi-Chun Tsai, and Wen-Guey Tzeng. Extracting randomness from multiple independent sources. *IEEE Transactions on Information Theory*, 51(6):2224–2227, 2005.

[36] Chi-Jen Lu. Encryption against storage-bounded adversaries from on-line strong extractors. *Journal of Cryptology*, 17(1):27–42, 2004.

[37] Chi-Jen Lu, Omer Reingold, Salil P. Vadhan, and Avi Wigderson. Extractors: Optimal up to constant factors. In *Proc. 35th Annual ACM Symposium on Theory of Computing (STOC'03)*, pages 602–611, San Diego, California, June 2003.

[38] Alexander Lubotzky, R. Phillips, and Peter Sarnak. Ramanujan graphs. *Combinatorica*, 8(3):261–277, 1988.

[39] Michael Mitzenmacher and Eli Upfal. *Probability and Computing: Randomized Algorithms and Probabilistic Analysis.* Cambridge University Press, 2005.

[40] Joseph Naor and Moni Naor. Small-bias probability spaces: efficient constructions and applications. *SIAM Journal on Computing*, 22(4):838–856, 1993.

[41] Noam Nisan and Amnon Ta-Shma. Extracting randomness: A survey and new constructions. *J. Comput. Syst. Sci.*, 58(1):148–173, 1999.

[42] Noam Nisan and David Zuckerman. Randomness is linear in space. *Journal of Computer and System Sciences*, 52(1):43–52, February 1996.

[43] Jaikumar Radhakrishnan and Amnon Ta-Shma. Bounds for dispersers, extractors, and depth-two superconcentrators. *SIAM Journal on Discrete Mathematics*, 13(1):2–24, 2000.

[44] Anup Rao. Extractors for a constant number of polynomially small min-entropy independent sources. In *Proc. 38th Annual ACM Symposium on Theory of Computing (STOC'06)*, pages 497–506, 2006.

[45] Ran Raz. Extractors with weak random seeds. In *Proc. 37th Annual ACM Symposium on Theory of Computing (STOC'05)*, pages 11–20, Baltimore, MD, USA, May 2005.

[46] Ran Raz, Omer Reingold, and Salil P. Vadhan. Extracting all the randomness and reducing the error in Trevisan's extractors. In *Proc. 31st Annual ACM Symposium on Theory of Computing (STOC'99)*, pages 149–158, Atlanta, Georgia, USA, May 1999.

[47] Omer Reingold, Ronen Shaltiel, and Avi Wigderson. Extracting randomness via repeated condensing. In *Proc. IEEE 41st Annual Symposium on Foundations of Computer Science (FOCS'00)*, pages 12–14, 2000.

[48] Ronen Shaltiel. Recent developments in explicit constructions of extractors. *Bulletin of the EATCS*, 77:67–95, 2002.

[49] Ronen Shaltiel and Christopher Umans. Simple extractors for all min-entropies and a new pseudo-random generator. In *Proc. IEEE 42nd Annual IEEE Symposium on Foundations of Computer Science (FOCS'01)*, pages 648–657, 2001.

[50] Michael Sipser. Expanders, randomness, or time versus space. *Journal of Computer and System Sciences*, 36(3):379–383, 1988.

[51] Madhu Sudan, Luca Trevisan, and Salil Vadhan. Pseudorandom generators without the **XOR** lemma. *Journal of Computer and System Sciences*, 62(2):236–266, 2001.

[52] Amnon Ta-Shma, Christopher Umans, and David Zuckerman. Loss-less condensers, unbalanced expanders, and extractors. In *Proc. 33rd Annual ACM Symposium on Theory of Computing (STOC'01)*, pages 143–152, Crete, Greece, July 2001.

[53] Amnon Ta-Shma and David Zuckerman. Extractor codes. *IEEE Trans. Info. Theory*, 50(12):3015–3025, 2004.

[54] Luca Trevisan. Extractors and pseudorandom generators. *Journal of ACM*, 48(4):860–879, 2001.

[55] Luca Trevisan and Salil P. Vadhan. Extracting randomness from samplable distributions. In *Proc. IEEE 41st Annual IEEE Symposium on Foundations of Computer Science (FOCS'00)*, pages 32–42, 2000.

[56] Salil P. Vadhan. Constructing locally computable extractors and cryptosystems in the bounded-storage model. *J. Cryptology*, 17(1):43–77, 2004.

[57] John von Neumann. Various techniques used in connection with random digits. *National Bureau of Standards Applied Mathematics Series*, 12:36–38, 1951.

[58] Avi Wigderson and David Zuckerman. Expanders that beat the eigenvalue bound: Explicit construction and applications. *Combinatorica*, 19(1):125–138, 1999.

[59] Andrew Chi-Chih Yao. Theory and applications of trapdoor functions. In *Proc. IEEE 23rd Annual IEEE Symposium on Foundations of Computer Science (FOCS'82)*, pages 80–91, 1982.

[60] David Zuckerman. Lecture notes for `CS 395T` - pseudorandomness and combinatorial constructions. http://userweb.cs.utexas.edu/users/diz/.

[61] David Zuckerman. General weak random sources. In *Proc. IEEE 31st Annual Symposium on Foundations of Computer Science (FOCS'90)*, pages 534–543, 1990.

[62] David Zuckerman. Simulating **BPP** using a general weak random source. *Algorithmica*, 16(4/5):367–391, 1996.

[63] David Zuckerman. Randomness-optimal oblivious sampling. *Random Structures and Algorithms*, 11:345–367, 1997.

Appendix A

An Example of Pair-wise Independent Hash Family

We claim that the best result of [13] is a special case of generalized leftover hash lemma. Let $A^1, \ldots, A^m$ be $n \times n$ matrices over $GF[2]$, such that, $\forall S \subseteq [m], S \neq \emptyset$, the rank of $A^S \overset{def}{=} \sum_{i \in S} A^i$ is n. Define a family of hash functions $H = \{f_x | f_x : \{0,1\}^n \rightarrow \{0,1\}^m\}$ by

$$f_x(y) = \langle A^1 x, y \rangle \circ \langle A^2 x, y \rangle \circ \cdots \circ \langle A^m x, y \rangle$$

where $A^i x$ is a matrix-vector multiplication over $GF[2]$.

We show that the family H is pair-wise independent. For any $y, y' \in \{0,1\}^n, y \neq y'$, define $z = z_1 \circ \cdots \circ z_n = y - y' \neq 0$. Let $wt(z)$ denote the *Hamming weight* of z. Since $z \neq 0$, $wt(z) = k$ for some $1 \leq k \leq n$. W.L.O.G., suppose that $z_1 = z_2 = \cdots = z_k = 1$, and $z_{k+1} = z_{k+2} = z_n = 0$. Let A^i_j denote the transpose of the jth row of A^i, and let $x = x_1 \circ \cdots \circ x_n$. Then we obtain

$$\begin{aligned}
\Pr_{f_x \in H}[f_x(y) = f_x(y')] &= \Pr_{f_x \in H}[\forall i, \langle A^i x, y \rangle = \langle A^i x, y' \rangle] \\
&= \Pr_{f_x \in H}[\forall i, \langle A^i x, z \rangle = 0] \\
&= \Pr_{f_x \in H}[\forall i, \langle A^i_1, x \rangle z_1 + \cdots + \langle A^i_n, x \rangle z_n = 0] \\
&= \Pr_{f_x \in H}[\forall i, \langle A^i_1, x \rangle + \langle A^i_2, x \rangle + \cdots + \langle A^i_k, x \rangle = 0] \\
&= \Pr_{f_x \in H}[\forall i, \langle A^i_1 + A^i_2 + \cdots + A^i_k, x \rangle = 0]
\end{aligned}$$

where $+$ is the addition over $GF[2]$.

Now we evaluate the number of $x = x_1 \circ \cdots \circ x_n$ satisfying the following system of equations:

$$\begin{gathered}
\langle A_1^1 + A_2^1 + \cdots + A_k^1, x \rangle = 0 \\
\langle A_1^2 + A_2^2 + \cdots + A_k^2, x \rangle = 0 \\
\vdots \\
\langle A_1^m + A_2^m + \cdots + A_k^m, x \rangle = 0
\end{gathered}$$

Suppose that some of the above m equations are dependent, then there exist $b_1, \ldots, b_t$ for some $2 \leq t \leq m$ such that

$$(A_1^{b_1} + A_2^{b_1} + \cdots + A_k^{b_1}) + (A_1^{b_2} + A_2^{b_2} + \cdots + A_k^{b_2}) + \cdots + (A_1^{b_t} + A_2^{b_t} + \cdots + A_k^{b_t}) = 0$$

It means that the sum of the first k rows of $A^{b_1} + A^{b_2} + \cdots + A^{b_t}$ is 0, contradict $A^{b_1} + A^{b_2} + \cdots + A^{b_t}$ having full rank. Hence these m equations are independent. There are 2^{n-m} different values of x to satisfy the above system of m different equations and n variables, hence

$$\Pr_{f_x \in_R H}[f_x(y) = f_x(y')] = \Pr_{f_x \in_R H}[\forall i, \langle A^i x, y - y' \rangle = 0] = \frac{1}{2^m}.$$

We complete the proof.

Appendix B

An Elementary Proof of Extractors for Independent-Symbol Sources

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 B.0.1. *For any $n, k, D \in \mathbb{N}$ and any prime number $M \geq D$, there is an explicit (n, D, k, ε)-extractor* $\textsc{Ext}_0 : [D]^n \to [M]$, *with* $\varepsilon \leq \frac{1}{2} \cdot \sqrt{M} \cdot e^{-k/(8M^2 \log D)}$.

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 [32], 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 $\textsc{Ext}_0 : [D]^n \to [M]$ is then defined as

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

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 [33] 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 [33], 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 B.0.2. $\|\bar{\delta}\|_2^2 \leq \|\delta\|_2^2 \cdot (1 - \mathrm{H}_\infty(\beta)/(4M^2 \log D))$.

Proof. Note that for $i \in \mathbb{Z}_M$, $\bar{\delta}_i = \sum_{j \in \mathbb{Z}_M} \beta_j \delta_{i-j}$. So $\|\bar{\delta}\|_2^2 = \sum_i (\sum_j \beta_j \delta_{i-j})^2 = \sum_i \sum_j \beta_j^2 \delta_{i-j}^2 + \sum_i \sum_{j \neq \ell} \beta_j \beta_\ell \delta_{i-j} \delta_{i-\ell}$ which, using the equality $ab = (a^2 + b^2 - (a-b)^2)/2$ on the second term, equals

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

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 B.0.3. 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, $\sum_{1 \le j \le t} |\delta_{i_0+(j-1)s} - \delta_{i_0+js}| \ge |\delta_{i_0} - \delta_{i_0+ts}| = |\delta_{i_0} - \delta_{i_1}|$. Finally,

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

which by the Cauchy-Schwartz inequality of Lemma 2.7.2 is at least

$$\left(\sum_{1 \le j \le t} |\delta_{i_0+(j-1)s} - \delta_{i_0+js}| \right)^2 /t \ge |\delta_{i_0} - \delta_{i_1}|^2/t \ge \|\delta\|_2^2/M^2.$$

□

Claim B.0.4. $\sum_{j \ne \ell} \beta_j \beta_\ell \ge \mathrm{H}_\infty(\beta)/(2 \log D)$.

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

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

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

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

□

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

$$\|\bar{\delta}\|_2^2 \le \|\delta\|_2^2 \cdot \left(1 - \sum_{j \ne \ell} \beta_j \beta_\ell/(2M^2)\right) \le \|\delta\|_2^2 \cdot (1 - \mathrm{H}_\infty(\beta)/(4M^2 \log D)),$$

which proves the lemma. □

Now let us see how it can be used to prove the theorem.

Proof. (of Theorem B.0.1)

From Lemma B.0.2, 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

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

Therefore, we have

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

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 the Cauchy-Schwartz inequality of Lemma 2.7.2,

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

□