6

Randomness Extractors

Randomness extractors are functions that extract almost-uniform bits from sources of biased and correlated bits. The original motivation for extractors was to simulate randomized algorithms with weak random sources as might arise in nature. This motivation is still compelling, but extractors have taken on a much wider significance in the years since they were introduced. They have found numerous applications in theoretical computer science beyond this initial motivating one, in areas from cryptography to distributed algorithms to hardness of approximation. More importantly from the perspective of this survey, they have played a major unifying role in the theory of pseudorandomness. Indeed, the links between the various pseudorandom objects we are studying in this survey (expander graphs, randomness extractors, list-decodable codes, pseudorandom generators, samplers) were all discovered through work on extractors (even though now we find it more natural to present these links in the language of list decoding, as introduced in Section 5.3).

6.1 Motivation and Definition

6.1.1 Deterministic Extractors

Typically, when we design randomized algorithms or protocols, we assume that all algorithms/parties have access to sources of *perfect* randomness, i.e., bits that are unbiased and completely independent. However, when we implement these algorithms, the physical sources of randomness to which we have access may contain biases and correlations. For example, we may use low-order bits of the system clock, the user's mouse movements, or a noisy diode based on quantum effects. While these sources may have some randomness in them, the assumption that the source is perfect is a strong one, and thus it is of interest to try and relax it.

Ideally, what we would like is a compiler that takes any algorithm A that works correctly when fed perfectly random bits U_m, and produces a new algorithm A' that will work even if it is fed random bits $X \in \{0,1\}^n$ that come from a "weak" random source. For example, if A is a **BPP** algorithm, then we would like A' to also run in probabilistic polynomial time. One way to design such compilers is to design a *randomness extractor* $\mathrm{Ext} : \{0,1\}^n \to \{0,1\}^m$ such that $\mathrm{Ext}(X)$ is distributed uniformly in $\{0,1\}^m$.

IID-Bit Sources. A simple version of this question was already considered by von Neumann. He looked at sources that consist of boolean random variables $X_1, X_2, \ldots, X_n \in \{0,1\}$ that are independent but biased. That is, for every i, $\Pr[X_i = 1] = \delta$ for some unknown δ. How can such a source be converted into a source of independent, unbiased bits? Von Neumann proposed the following extractor: Break all the variables in pairs and for each pair output 0 if the outcome was 01, 1 if the outcome was 10, and skip the pair if the outcome was 00 or 11. This will yield an unbiased random bit after $1/(2\delta(1-\delta))$ pairs on average.

Independent-Bit Sources. Lets now look at a bit more interesting class of sources in which all the variables are still independent but the bias is no longer the same. Specifically, for every i, $\Pr[X_i = 1] = \delta_i$ and

$0 < \delta \leq \delta_i \leq 1 - \delta$ for some constant $\delta > 0$. How can we deal with such a source?

It can be shown that when we take a parity of ℓ bits from such an independent-bit source, the result approaches an unbiased coin flip exponentially fast in ℓ, i.e., $|\Pr\left[\oplus_{i=1}^{\ell} X_i = 1\right] - 1/2| = 2^{-\Omega(\ell)}$. The result is not a perfect coin flip but is as good as one for almost all purposes.

Let's be more precise about the problems we are studying. A *source* on $\{0,1\}^n$ is simply a random variable X taking values in $\{0,1\}^n$. In each of the above examples, there is an implicit *class* of sources being studied. For example, $\mathrm{IndBits}_{n,\delta}$ is the class of sources X on $\{0,1\}^n$ where the bits X_i are independent and satisfy $\delta \leq \Pr[X_i = 1] \leq 1 - \delta$. We could define $\mathrm{IIDBits}_{n,\delta}$ to be the same with the further restriction that all of the X_is are identically distributed, i.e., $\Pr[X_i = 1] = \Pr[X_j = 1]$ for all i, j, thereby capturing von Neumann sources.

Definition 6.1 (deterministic extractors). [1] Let $\mathcal{C}$ be a class of sources on $\{0,1\}^n$. An *ε-extractor* for $\mathcal{C}$ is a function $\mathrm{Ext} : \{0,1\}^n \to \{0,1\}^m$ such that for every $X \in \mathcal{C}$, $\mathrm{Ext}(X)$ is "ε-close" to U_m.

Note that we want a *single* function Ext that works for all sources in the class. This captures the idea that we do not want to assume we know the exact distribution of the physical source we are using, but only that it comes from some class. For example, for $\mathrm{IndBits}_{n,\delta}$, we know that the bits are independent and none are too biased, but not the specific bias of each bit. Note also that we only allow the extractor *one* sample from the source X. If we want to allow multiple independent samples, then this should be modelled explicitly in our class of sources; ideally we would like to minimize the independence assumptions used.

We still need to define what we mean for the output to be ε-close to U_m.

[1] Such extractors are called *deterministic* or *seedless* to contrast with the probabilistic or *seeded* randomness extractors we will see later.

Definition 6.2. For random variables X and Y taking values in $\mathcal{U}$, their *statistical difference* (also known as *variation distance*) is $\Delta(X,Y) = \max_{T\subset\mathcal{U}} |\Pr[X \in T] - \Pr[Y \in T]|$. We say that X and Y are *ε-close* if $\Delta(X,Y) \le \varepsilon$.

Intuitively, any event in X happens in Y with the same probability $\pm\varepsilon$. This is perhaps the most natural measure of distance for probability distributions (much more so than the ℓ_2 distance we used in the study of random walks). In particular, it satisfies the following natural properties.

Lemma 6.3 (properties of statistical difference). Let $X, Y, Z, X_1, X_2, Y_1, Y_2$ be random variables taking values in a universe $\mathcal{U}$. Then,

(1) $\Delta(X,Y) \ge 0$, with equality iff X and Y are identically distributed,
(2) $\Delta(X,Y) \le 1$, with equality iff X and Y have disjoint supports,
(3) $\Delta(X,Y) = \Delta(Y,X)$,
(4) $\Delta(X,Z) \le \Delta(X,Y) + \Delta(Y,Z)$,
(5) for every function f, we have $\Delta(f(X), f(Y)) \le \Delta(X,Y)$,
(6) $\Delta((X_1,X_2),(Y_1,Y_2)) \le \Delta(X_1,Y_1) + \Delta(X_2,Y_2)$ if X_1 and X_2, as well as Y_1 and Y_2, are independent, and
(7) $\Delta(X,Y) = \frac{1}{2} \cdot |X - Y|_1$, where $|\cdot|_1$ is the ℓ_1 distance. (Thus, X is ε-close to Y iff we can transform X into Y by "shifting" at most an ε fraction of probability mass.)

We now observe that extractors according to this definition give us the "compilers" we want.

Proposition 6.4. Let $A(w;r)$ be a randomized algorithm such that $A(w;U_m)$ has error probability at most γ, and let $\mathrm{Ext}: \{0,1\}^n \to \{0,1\}^m$ be an ε-extractor for a class $\mathcal{C}$ of sources on $\{0,1\}^n$. Define $A'(w;x) = A(w;\mathrm{Ext}(x))$. Then for every source $X \in \mathcal{C}$, $A'(w;X)$ has error probability at most $\gamma + \varepsilon$.

This application identifies some additional properties we'd like from our extractors. We'd like the extractor itself to be efficiently computable (e.g., polynomial time). In particular, to get m almost-uniform bits out, we should need at most $n = \mathrm{poly}(m)$ bits from the weak random source.

We can cast our earlier extractor for sources of independent bits in this language:

Proposition 6.5. For every constant $\delta > 0$, every $n, m \in \mathbb{N}$, there is a polynomial-time computable function $\mathrm{Ext} : \{0,1\}^n \to \{0,1\}^m$ that is an ε-extractor for $\mathrm{IndBits}_{n,\delta}$, with $\varepsilon = m \cdot 2^{-\Omega(n/m)}$.

In particular, taking $n = m^2$, we get exponentially small error with a source of polynomial length.

Proof. Ext breaks the source into m blocks of length $\lfloor n/m \rfloor$ and outputs the parity of each block. $\square$

Unpredictable-Bit Sources (aka Santha–Vazirani Sources). Another interesting class of sources, which looks similar to the previous example is the class $\mathrm{UnpredBits}_{n,\delta}$ of *unpredictable-bit sources.* These are the sources that for every i, every $x_1, \ldots, x_n \in \{0,1\}$ and some constant $\delta > 0$, satisfy

$$\delta \le \Pr[X_i = 1 \mid X_1 = x_1, X_2 = x_2, \ldots, X_{i-1} = x_{i-1}] \le 1 - \delta$$

The parity extractor used above will be of no help with this source since the next bit could be chosen in a way that the parity will be equal to 1 with probability δ. Indeed, there does not exist any nontrivial extractor for these sources — the best we can do is output the first bit:

Proposition 6.6. For every $n \in \mathbb{N}$, $\delta > 0$, and fixed extraction function $\mathrm{Ext} : \{0,1\}^n \to \{0,1\}$ there exists a source $X \in \mathrm{UnpredBits}_{n,\delta}$ such that either $\Pr[\mathrm{Ext}(X) = 1] \le \delta$ or $\Pr[\mathrm{Ext}(X) = 1] \ge 1 - \delta$. That is, there is no ε-extractor for $\mathrm{UnpredBits}_{n,\delta}$ for $\varepsilon < 1/2 - \delta$.

The proof is left as an exercise (Problem 6.6).

Nevertheless, as we will see, the answer to the question whether we can simulate **BPP** algorithms with unpredictable sources will be "yes"! Indeed, we will even be able to handle a much more general class of sources, introduced in the next section.

6.1.2 Entropy Measures and General Weak Sources

Intuitively, to extract m almost-uniform bits from a source, the source must have at least "m bits of randomness" in it. (In particular, its support cannot be much smaller than 2^m.) Ideally, this is all we would like to assume about a source. Thus, we need some measure of how much randomness is in a random variable; this can be done using various notions of *entropy* described below.

Definition 6.7 (entropy measures). Let X be a random variable. Then

- the *Shannon entropy* of X is:

$$\mathrm{H}_{Sh}(X) = \mathop{\mathrm{E}}_{x \overset{\mathrm{R}}{\leftarrow} X}\left[\log \frac{1}{\Pr[X = x]}\right].$$

- the *Rényi entropy* of X is:

$$\mathrm{H}_2(X) = \log\left(\frac{1}{\mathrm{E}_{x \overset{\mathrm{R}}{\leftarrow} X}[\Pr[X = x]]}\right) = \log \frac{1}{\mathrm{CP}(X)}, \text{ and}$$

- the *min-entropy* of X is:

$$\mathrm{H}_\infty(X) = \min_x \left\{\log \frac{1}{\Pr[X = x]}\right\},$$

where all logs are base 2.

Rényi entropy $\mathrm{H}_2(X)$ should not be confused with the binary entropy function $H_2(\delta)$ from Definition 5.6. Indeed, the q-ary entropy $H_q(\delta)$ is equal to the *Shannon* entropy of a random variable that equals 1 with probability $1 - \delta$ and is uniformly distributed in $\{2, \ldots, q\}$ with probability δ.

All the three measures satisfy the following properties we would expect from a measure of randomness:

Lemma 6.8 (properties of entropy). For each of the entropy measures $\mathrm{H} \in \{\mathrm{H}_{Sh}, \mathrm{H}_2, \mathrm{H}_\infty\}$ and random variables X, Y, we have:

- $\mathrm{H}(X) \geq 0$, with equality iff X is supported on a single element,
- $\mathrm{H}(X) \leq \log|\mathrm{Supp}(X)|$, with equality iff X is uniform on $\mathrm{Supp}(X)$,
- if X, Y are independent, then $\mathrm{H}((X,Y)) = \mathrm{H}(X) + \mathrm{H}(Y)$,
- for every deterministic function f, we have $\mathrm{H}(f(X)) \leq \mathrm{H}(X)$, and
- for every X, we have $\mathrm{H}_\infty(X) \leq \mathrm{H}_2(X) \leq \mathrm{H}_{Sh}(X)$.

To illustrate the differences between the three notions, consider a source X such that $X = 0^n$ with probability 0.99 and $X = U_n$ with probability 0.01. Then $\mathrm{H}_{Sh}(X) \geq 0.01n$ (contribution from the uniform distribution), $\mathrm{H}_2(X) \leq \log(1/0.99^2) < 1$ and $\mathrm{H}_\infty(X) \leq \log(1/0.99) < 1$ (contribution from 0^n). Note that even though X has Shannon entropy linear in n, we cannot expect to extract bits that are close to uniform or carry out any useful randomized computations with one sample from X, because it gives us nothing useful 99% of the time. Thus, we should use the stronger measures of entropy given by H_2 or H_∞.

Then why is Shannon entropy so widely used in information theory results? The reason is that such results typically study what happens when you have many independent samples from the source (whereas we only allow one sample). In the case of many samples, it turns out that the source is “close” to one where the min-entropy is roughly equal to the Shannon entropy. Thus the distinction between these entropy measures becomes less significant. Moreover, Shannon entropy satisfies many nice identities that make it quite easy to work with. Min-entropy and Rényi entropy are much more delicate.

We will consider the task of extracting randomness from sources where all we know is a lower bound on the min-entropy:

Definition 6.9. A random variable X is a *k-source* if $\mathrm{H}_\infty(X) \geq k$, i.e., if $\Pr[X = x] \leq 2^{-k}$.

A typical setting of parameters is $k = \delta n$ for some fixed δ, e.g., 0.01. We call δ the *min-entropy rate*. Some different ranges that are commonly studied (and are useful for different applications): $k = \mathrm{polylog}(n)$, $k = n^\gamma$ for a constant $\gamma \in (0,1)$, $k = \delta n$ for a constant $\delta \in (0,1)$, and $k = n - O(1)$. The middle two ($k = n^\gamma$ and $k = \delta n$) are the most natural for simulating randomized algorithms with weak random sources.

Examples of k-sources:

- k random and independent bits, together with $n - k$ fixed bits (in an arbitrary order). These are called *oblivious bit-fixing* sources.
- k random and independent bits, and $n - k$ bits that depend arbitrarily on the first k bits. These are called *adaptive bit-fixing sources.*
- Unpredictable-bit sources with bias parameter δ. These are k-sources with $k = \log(1/(1-\delta)^n) = \Theta(\delta n)$.
- Uniform distribution on a set $S \subset \{0,1\}^n$ with $|S| = 2^k$. These are called *flat k-sources.*

It turns out that flat k-sources are really representative of general k-sources.

Lemma 6.10. Every k-source is a convex combination of flat k-sources (provided that $2^k \in \mathbb{N}$), i.e., $X = \sum p_i X_i$ with $0 \leq p_i \leq 1$, $\sum p_i = 1$ and all the X_i are flat k-sources.

That is, we can think of any k-source as being obtained by first selecting a flat k-source X_i according to some distribution (given by the p_is) and then selecting a random sample from X_i. This means that if we can compile probabilistic algorithms to work with flat k-sources, then we can compile them to work with any k-source.

Proof. Let X be a k-source on $[N]$. We can view X as partitioning a circle of unit circumference into N (half-open) intervals, where the tth interval has length exactly $\Pr[X = t]$. (If we associate the points on the circle with $[0,1)$, then the tth interval is $[\Pr[X < t], \Pr[X \le t])$.) Now consider a set S of K points spaced evenly on the circle. Then since each interval is half-open and has length at most $1/K$, each interval contains at most one point from S, so the uniform distribution on the set $T(S) = \{t : S \cap I_t \neq \emptyset\}$ is a flat k-source. Moreover, if we perform a uniformly random rotation of S on the circle to obtain a rotated set R and then choose a uniformly random element of $T(R)$, the probability that we output any value $t \in [N]$ is exactly the length of I_t, which equals $\Pr[X = t]$. Thus we have decomposed X as a convex combination of flat k-sources. (Specifically, $X = \sum_T p_T U_T$, where the sum is over subsets $T \subseteq [N]$ of size K, and $p_T = \Pr_R[T(R) = T]$.) $\square$

We also sketch another proof of Lemma 6.10 that can be more easily generalized to other classes of sources. We can view a random variable X taking values in $[N]$ as an N-dimensional vector, where $X(i)$ is the probability mass of i. Then X is a k-source if and only if $X(i) \in [0, 2^{-k}]$ for every $i \in [N]$ and $\sum_i X(i) = 1$. The set of vectors X satisfying these linear inequalities is a convex polytope in $\mathbb{R}^N$. By basic linear programming theory, all of the points in the polytope are convex combinations of its *vertices*, which are defined to be the points that make a maximal subset of the inequalities tight. By inspection, the vertices of the polytope of k-sources are those sources where $X(i) = 2^{-k}$ for 2^k values of i and $X(i) = 0$ for the remaining values of i; these are simply the flat k-sources.

6.1.3 Seeded Extractors

Proposition 6.6 tells us that it is impossible to have deterministic extractors for unpredictable sources. Here we consider k-sources, which are more general than unpredictable sources, and hence it is also impossible to have deterministic extractors for them. The impossibility result for k-sources is stronger and simpler to prove.

Proposition 6.11. For any $\text{Ext} : \{0,1\}^n \to \{0,1\}$ there exists an $(n-1)$-source X so that $\text{Ext}(X)$ is constant.

Proof. There exists $b \in \{0,1\}$ so that $|\text{Ext}^{-1}(b)| \geq 2^n/2 = 2^{n-1}$. Then let X be the uniform distribution on $\text{Ext}^{-1}(b)$. □

On the other hand, if we reverse the order of quantifiers, allowing the extractor to depend on the source, it is easy to see that good extractors exist and in fact a randomly chosen function will be a good extractor with high probability.

Proposition 6.12. For every $n, k, m \in \mathbb{N}$, every $\varepsilon > 0$, and every flat k-source X, if we choose a random function $\text{Ext} : \{0,1\}^n \to \{0,1\}^m$ with $m = k - 2\log(1/\varepsilon) - O(1)$, then $\text{Ext}(X)$ will be ε-close to U_m with probability $1 - 2^{-\Omega(K\varepsilon^2)}$, where $K = 2^k$.

(In this section, we will make extensive use of the convention that capital variables are 2 raised to the power of the corresponding lower-case variable, such as $K = 2^k$ above.)

Proof. Choose our extractor randomly. We want it to have following property: for all $T \subset [M]$, $|\Pr[\text{Ext}(X) \in T] - \Pr[U_m \in T]| \leq \varepsilon$. Equivalently, $|\{x \in \text{Supp}(X) : \text{Ext}(x) \in T\}|/K$ differs from the density $\mu(T)$ by at most ε. For each point $x \in \text{Supp}(X)$, the probability that $\text{Ext}(x) \in T$ is $\mu(T)$, and these events are independent. By the Chernoff Bound (Theorem 2.21) for each fixed T, this condition holds with probability at least $1 - 2^{-\Omega(K\varepsilon^2)}$. Then the probability that condition is violated for at least one T is at most $2^M 2^{-\Omega(K\varepsilon^2)}$, which is less than 1 for $m = k - 2\log(1/\varepsilon) - O(1)$. □

Note that the failure probability is doubly-exponentially small in k. Naively, one might hope that we could get an extractor that's good for all flat k-sources by a union bound. But the number of flat k-sources is $\binom{N}{K} \approx N^K$ (where $N = 2^n$), which is unfortunately a larger double-exponential in k. We can overcome this gap by allowing the extractor to be "slightly" probabilistic, i.e., allowing the extractor a *seed* consisting of a small number of truly random bits in addition to the weak random source. We can think of this seed of truly random bits as a random choice of an extractor from family of extractors. This leads to the following crucial definition:

Definition 6.13 (seeded extractors). A function $\mathrm{Ext} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ is a (k,ε)-*extractor* if for every k-source X on $\{0,1\}^n$, $\mathrm{Ext}(X,U_d)$ is ε-close to U_m.

(Sometimes we will refer to extractors $\mathrm{Ext} : [N] \times [D] \to [M]$ whose domain and range do not consist of bit-strings. These are defined in the natural way, requiring that $\mathrm{Ext}(X,U_{[D]})$ is ε-close to $U_{[M]}$.)

The goal is to construct extractors that minimize d and maximize m. We prove the following theorem.

Theorem 6.14. For every $n \in \mathbb{N}$, $k \in [0,n]$ and $\varepsilon > 0$, there exists a (k,ε)-extractor $\mathrm{Ext} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ with $m = k + d - 2\log(1/\varepsilon) - O(1)$ and $d = \log(n-k) + 2\log(1/\varepsilon) + O(1)$.

One setting of parameters to keep in mind (for our application of simulating randomized algorithms with a weak source) is $k = \delta n$, with δ a fixed constant (e.g., $\delta = 0.01$), and ε a fixed constant (e.g., $\varepsilon = 0.01$).

Proof. We use the Probabilistic Method. By Lemma 6.10, it suffices for Ext to work for flat k-sources. Choose the extractor Ext at random. Then the probability that the extractor fails is at most the number of flat k-sources times the probability Ext fails for a fixed flat k-source. By the above proposition, the probability of failure for a fixed flat k-source is at most $2^{-\Omega(KD\varepsilon^2)}$, since (X,U_d) is a flat $(k+d)$-source) and $m = k + d - 2\log(\frac{1}{\varepsilon}) - O(1)$. Thus the total failure probability is at most

$$\binom{N}{K} \cdot 2^{-\Omega(KD\varepsilon^2)} \le \left(\frac{Ne}{K}\right)^K 2^{-\Omega(KD\varepsilon^2)}.$$

The latter expression is less than 1 if $D\varepsilon^2 \geq c\log(Ne/K) = c\cdot(n-k) + c'$ for constants c, c'. This is equivalent to $d \geq \log(n-k) + 2\log(\frac{1}{\varepsilon}) + O(1)$. □

It turns out that both bounds (on m and d) are individually tight up to the $O(1)$ terms.

Recall that our motivation for extractors was to simulate randomized algorithms given *only* a weak random source, so allowing a truly random seed may seem to defeat the purpose. However, if the seed is of logarithmic length as in Theorem 6.14, then instead of selecting it randomly, we can enumerate all possibilities for the seed and take a majority vote.

Proposition 6.15. Let $A(w;r)$ be a randomized algorithm for computing a function f such that $A(w;U_m)$ has error probability at most γ (i.e., $\Pr[A(w;U_m) \neq f(w)] \leq \gamma]$), and let $\mathrm{Ext}: \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ be a (k,ε)-extractor. Define

$$A'(w;x) = \underset{y\in\{0,1\}^d}{\mathrm{maj}} \{A(w;\mathrm{Ext}(x,y))\}.$$

Then for every k-source X on $\{0,1\}^n$, $A'(w;X)$ has error probability at most $2\cdot(\gamma+\varepsilon)$.

Proof. The probability that $A(w;\mathrm{Ext}(X,U_d))$ is incorrect is not more than the probability $A(w;U_m)$ is incorrect plus ε, i.e., $\gamma+\varepsilon$, by the definition of statistical difference. Then the probability that $\mathrm{maj}_y A(w,\mathrm{Ext}(X,y))$ is incorrect is at most $2\cdot(\gamma+\varepsilon)$, because each error of $\mathrm{maj}_y A(w;\mathrm{Ext}(x,y))$ corresponds to $A(w;\mathrm{Ext}(x,U_d))$ erring with probability at least 1/2. □

Note that the enumeration incurs a 2^d factor slowdown in the simulation. Thus, to retain running time $\mathrm{poly}(m)$, we want to construct extractors where (a) $d = O(\log n)$; (b) Ext is computable in polynomial time; and (c) $m = n^{\Omega(1)}$.

We remark that the error probability in Proposition 6.15 can actually be made exponentially small by using an extractor that is designed for slightly lower min-entropy. (See Problem 6.2.)

We note that even though seeded extractors suffice for simulating randomized algorithms with only a weak source, they do not suffice for all applications of randomness in theoretical computer science. The trick of eliminating the random seed by enumeration does not work, for example, in cryptographic applications of randomness. Thus the study of deterministic extractors for restricted classes of sources remains a very interesting and active research direction. We, however, will focus on seeded extractors, due to their many applications and their connections to the other pseudorandom objects we are studying.

6.2 Connections to Other Pseudorandom Objects

As mentioned earlier, extractors have played a unifying role in the theory of pseudorandomness, through their close connections with a variety of other pseudorandom objects. In this section, we will see two of these connections. Specifically, how by reinterpreting them appropriately, extractors can be viewed as providing families of hash functions, and as being a certain type of highly expanding graphs.

6.2.1 Extractors as Hash Functions

Proposition 6.12 says that for any subset $S \subset [N]$ of size K, if we choose a completely random hash function $h : [N] \to [M]$ for $M \ll K$, then h will map the elements of S almost-uniformly to $[M]$. Equivalently, if we let H be distributed uniformly over all functions $h : [N] \to [M]$ and X be uniform on the set S, then $(H, H(X))$ is statistically close to $(H, U_{[M]})$, where we use the notation U_T to denote the uniform distribution on a set T. Can we use a smaller family of hash functions than the set of all functions $h : [N] \to [M]$? This gives rise to the following variant of extractors.

Definition 6.16 (strong extractors). Extractor $\mathrm{Ext} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ is a *strong* (k,ε)*-extractor* if for every k-source X on $\{0,1\}^n$, $(U_d, \mathrm{Ext}(X, U_d))$ is ε-close to (U_d, U_m). Equivalently, $\mathrm{Ext}'(x,y) = (y, \mathrm{Ext}(x,y))$ is a standard (k,ε)-extractor.

The nonconstructive existence proof of Theorem 6.14 can be extended to establish the existence of very good strong extractors:

Theorem 6.17. For every $n,k \in \mathbb{N}$ and $\varepsilon > 0$ there exists a strong (k,ε)-extractor $\mathrm{Ext} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ with $m = k - 2\log(1/\varepsilon) - O(1)$ and $d = \log(n-k) + 2\log(1/\varepsilon) + O(1)$.

Note that the output length is $m \approx k$ instead of $m \approx k + d$; intuitively a strong extractor needs to extract randomness that is *independent* of the seed and thus can only get the k bits from the source.

We see that strong extractors can be viewed as very small families of hash functions having the almost-uniform mapping property mentioned above. Indeed, our first explicit construction of extractors is obtained by using pairwise independent hash functions.

Theorem 6.18 (Leftover Hash Lemma). If $\mathcal{H} = \{h : \{0,1\}^n \to \{0,1\}^m\}$ is a pairwise independent (or even 2-universal) family of hash functions where $m = k - 2\log(1/\varepsilon)$, then $\mathrm{Ext}(x,h) \stackrel{\mathrm{def}}{=} h(x)$ is a strong (k,ε)-extractor. Equivalently, $\mathrm{Ext}(x,h) = (h,h(x))$ is a standard (k,ε)-extractor.

Note that the seed length equals the number of random bits required to choose $h \stackrel{\mathrm{R}}{\leftarrow} \mathcal{H}$, which is at least n by Problem 3.5.[2] This is far from optimal; for the purposes of simulating randomized algorithms we would like $d = O(\log n)$. However, the output length of the extractor is $m = k - 2\log(1/\varepsilon)$, which is optimal up to an additive constant.

Proof. Let X be an arbitrary k-source on $\{0,1\}^n$, $\mathcal{H}$ as above, and $H \stackrel{\mathrm{R}}{\leftarrow} \mathcal{H}$. Let d be the seed length. We show that $(H,H(X))$ is ε-close to $U_d \times U_m$ in the following three steps:

(1) We show that the collision probability of $(H,H(X))$ is close to that of $U_d \times U_m$.

[2] Problem 3.5 refers to pairwise independent families, but a similar argument shows that universal families require $\Omega(n)$ random bits. (Instead of constructing orthogonal vectors, we construct vectors that have nonpositive dot product.)

(2) We note that this is equivalent to saying that the ℓ_2 distance between $(H, H(X))$ and $U_d \times U_m$ is small.
(3) Then we deduce that the statistical difference is small, by recalling that the statistical difference equals half of the ℓ_1 distance, which can be (loosely) bounded by the ℓ_2 distance.

Proof of (1): By definition, $\mathrm{CP}(H, H(X)) = \Pr[(H, H(X)) = (H', H'(X'))]$, where (H', X') is independent of and identically distributed to (H, X). Note that $(H, H(X)) = (H', H'(X))$ if and only if $H = H'$ and either $X = X'$ or $X \neq X'$ but $H(X) = H(X')$. Thus

$$\begin{aligned}\mathrm{CP}(H, H(X)) &= \mathrm{CP}(H) \cdot (\mathrm{CP}(X) + \Pr[H(X) = H(X') \mid X \neq X'])\\ &\leq \frac{1}{D} \cdot \left(\frac{1}{K} + \frac{1}{M}\right) \leq \frac{1+\varepsilon^2}{DM}.\end{aligned}$$

To see the penultimate inequality, note that $\mathrm{CP}(H) = 1/D$ because there are D hash functions, $\mathrm{CP}(X) \leq 1/K$ because $\mathrm{H}_\infty(X) \geq k$, and $\Pr[H(X) = H(X') \,|X \neq X'] \leq 1/M$ by 2-universality.

Proof of (2):

$$\begin{aligned}\|(H, H(X)) - U_d \times U_m\|^2 &= \mathrm{CP}(H, H(X)) - \frac{1}{DM}\\ &\leq \frac{1+\varepsilon^2}{DM} - \frac{1}{DM} = \frac{\varepsilon^2}{DM}.\end{aligned}$$

Proof of (3): Recalling that the statistical difference between two random variables X and Y is equal to $\frac{1}{2}|X - Y|_1$, we have:

$$\begin{aligned}\Delta((H, H(X)), U_d \times U_m) &= \frac{1}{2} \cdot |(H, H(X)) - U_d \times U_m|_1\\ &\leq \frac{\sqrt{DM}}{2} \cdot \|(H, H(X)) - U_d \times U_m\|\\ &\leq \frac{\sqrt{DM}}{2} \cdot \sqrt{\frac{\varepsilon^2}{DM}}\\ &= \frac{\varepsilon}{2}.\end{aligned}$$

Thus, we have in fact obtained a strong $(k, \varepsilon/2)$-extractor. □

The proof above actually shows that $\mathrm{Ext}(x,h) = h(x)$ extracts with respect to collision probability, or equivalently, with respect to the ℓ_2-norm. This property may be expressed in terms of Rényi entropy $\mathrm{H}_2(Z) \overset{\mathrm{def}}{=} \log(1/\mathrm{CP}(Z))$. Indeed, we can define $\mathrm{Ext} : \{0,1\}^n \times \{0,1\}^d \longrightarrow \{0,1\}^m$ to be a (k,ε) *Rényi-entropy extractor* if $\mathrm{H}_2(X) \geq k$ implies $\mathrm{H}_2(\mathrm{Ext}(X,U_d)) \geq m - \varepsilon$ (or $\mathrm{H}_2(U_d,\mathrm{Ext}(X,U_d)) \geq m + d - \varepsilon$ for strong Rényi-entropy extractors). Then the above proof shows that pairwise-independent hash functions yield strong Rényi-entropy extractors.

In general, it turns out that an extractor with respect to Rényi entropy must have seed length $d \geq \min\{m/2, n-k\} - O(1)$ (as opposed to $d = O(\log n)$); this explains why the seed length in the above extractor is large. (See Problem 6.4.)

6.2.2 Extractors versus Expanders

Extractors have a natural interpretation as graphs. Specifically, we can interpret an extractor $\mathrm{Ext} : \{0,1\}^n \times \{0,1\}^d \longrightarrow \{0,1\}^m$ as the neighbor function of a bipartite multigraph $G = ([N],[M],E)$ with $N = 2^n$ left-vertices, $M = 2^m$ right-vertices, and left-degree $D = 2^d$,[3] where the rth neighbor of left-vertex u is $\mathrm{Ext}(u,r)$. Typically $n \gg m$, so the graph is very unbalanced. It turns out that the extraction property of Ext is related to various "expansion" properties of G. In this section, we explore this relationship.

Let $\mathrm{Ext} : \{0,1\}^n \times \{0,1\}^d \longrightarrow \{0,1\}^m$ be a (k,ε)-extractor and $G = ([N],[M],E)$ the associated graph. Recall that it suffices to examine Ext with respect to *flat* k-sources: in this case, the extractor property says that given a subset S of size $K = 2^k$ on the left, a random neighbor of a random element of S should be close to uniform on the right. In particular, if $S \subset [N]$ is a subset on the left of size K, then $|N(S)| \geq (1-\varepsilon)M$. This property is just like vertex expansion, except that it ensures expansion only for sets of size exactly K, not any size $\leq K$. Recall that we call such a graph an $(=K, A)$ *vertex expander* (Definition 5.32). Indeed, this gives rise to the following weaker variant of extractors.

[3] This connection is the reason we use d to denote the seed length of an extractor.

Definition 6.19 (dispersers). A function $\mathrm{Disp}: \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ is a (k,ε)-*disperser* if for every k-source X on $\{0,1\}^n$, $\mathrm{Disp}(X,U_d)$ has a support of size at least $(1-\varepsilon)\cdot 2^m$.

While extractors can be used to simulate **BPP** algorithms with a weak random source, dispersers can be used to simulate **RP** algorithms with a weak random source. (See Problem 6.2.)

Then, we have:

Proposition 6.20. Let $n,m,d \in \mathbb{N}$, $K = 2^k \in \mathbb{N}$, and $\varepsilon > 0$. A function $\mathrm{Disp}: \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ is a (k,ε)-disperser iff the corresponding bipartite multigraph $G = ([N],[M],E)$ with left-degree D is an $(=K,A)$ vertex expander for $A = (1-\varepsilon)\cdot M/K$.

Note that extractors and dispersers are interesting even when $M \ll K$, so the expansion parameter A may be less than 1. Indeed, $A < 1$ is interesting for vertex "expanders" when the graph is highly imbalanced. Still, for an *optimal* extractor, we have $M = \Theta(\varepsilon^2 KD)$ (because $m = k + d - 2\log(1/\varepsilon) - \Theta(1)$), which corresponds to expansion factor $A = \Theta(\varepsilon^2 D)$. (An optimal disperser actually gives $A = \Theta(D/\log(1/\varepsilon))$.) Note this is smaller than the expansion factor of $D/2$ in Ramanujan graphs and $D - O(1)$ in random graphs; the reason is that those expansion factors are for "small" sets, whereas here we are asking for sets to expand to almost the entire right-hand side.

Now let's look for a graph-theoretic property that is *equivalent* to the extraction property. Ext is a (k,ε)-extractor iff for every set $S \subset [N]$ of size K,

$$\Delta(\mathrm{Ext}(U_S,U_{[D]}),U_{[M]}) = \max_{T\subset[M]} |\Pr[\mathrm{Ext}(U_S,U_{[D]}) \in T] - \Pr[U_{[M]} \in T]| \le \varepsilon,$$

where U_S denotes the uniform distribution on S. This inequality may be expressed in graph-theoretic terms as follows. For every set $T \subset [M]$,

$$\begin{aligned} &\Big|\Pr\left[\mathrm{Ext}(U_S, U_{[D]}) \in T\right] - \Pr\left[U_{[M]} \in T\right]\Big| \le \varepsilon \\ \Leftrightarrow &\left|\frac{e(S,T)}{|S|D} - \frac{|T|}{M}\right| \le \varepsilon \\ \Leftrightarrow &\left|\frac{e(S,T)}{ND} - \mu(S)\mu(T)\right| \le \varepsilon\mu(S), \end{aligned}$$

where $e(S,T)$ is the number of edges from S to T (as in Definition 4.13).

Thus, we have:

Proposition 6.21. A function $\mathrm{Ext} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ is a (k,ε)-extractor iff the corresponding bipartite multigraph $G = ([N], [M], E)$ with left-degree D has the property that $|e(S,T)/ND - \mu(S)\mu(T)| \le \varepsilon\mu(S)$ for every $S \subset [N]$ of size K and every $T \subset [M]$.

Note that this is very similar to the Expander Mixing Lemma (Lemma 4.15), which states that if a graph G has spectral expansion λ, then for *all* sets $S,T \subset [N]$ we have

$$\left|\frac{e(S,T)}{ND} - \mu(S)\mu(T)\right| \le \lambda\sqrt{\mu(S)\mu(T)}.$$

It follows that if $\lambda\sqrt{\mu(S)\mu(T)} \le \varepsilon\mu(S)$ for all $S \subset [N]$ of size K and all $T \subset [N]$, then G gives rise to a (k,ε)-extractor (by turning G into a D-regular bipartite graph with N vertices on each side in the natural way). It suffices for $\lambda \le \varepsilon \cdot \sqrt{K/N}$ for this to work.

We can use this connection to turn our explicit construction of spectral expanders into an explicit construction of extractors. To achieve $\lambda \le \varepsilon \cdot \sqrt{K/N}$, we can take an appropriate power of a constant-degree expander. Specifically, if G_0 is a D_0-regular expander on N vertices with bounded second eigenvalue, we can consider the tth power of G_0, $G = G_0^t$, where $t = O(\log((1/\varepsilon)\sqrt{N/K})) = O(n - k + \log(1/\varepsilon))$. The degree of G is $D = D_0^t$, so $d = \log D = O(t)$. This yields the following result:

Theorem 6.22. For every $n,k \in \mathbb{N}$ and $\varepsilon > 0$, there is an explicit (k,ε)-extractor $\mathrm{Ext} : \{0,1\}^n \times \{0,1\}^d \longrightarrow \{0,1\}^n$ with $d = O(n - k + \log(1/\varepsilon))$.

Note that the seed length is significantly better than in the construction from pairwise-independent hashing when k is close to n, say $k = n - O(n)$ (i.e., $K = N^{1-o(1)}$). The output length is n, which is much larger than the typical output length for extractors (usually $m \ll n$). Using a Ramanujan graph (rather than an arbitrary constant-degree expander), the seed length can be improved to $d = n - k + 2\log(1/\varepsilon) + O(1)$, which yields an optimal output length $n = k + d - 2\log(1/\varepsilon) - O(1)$.

Another way of proving Theorem 6.22 is to use the fact that a random step on an expander decreases the ℓ_2 distance to uniform, like in the proof of the Leftover Hash Lemma. This analysis shows that we actually get a Rényi-entropy extractor; and thus explains the large seed length $d \approx n - k$.

The following table summarizes the main differences between "classic" expanders and extractors.

Table 6.1. Differences between "classic" expanders and extractors.

Expanders	Extractors
Measured by vertex or spectral expansion	Measured by min-entropy/ statistical difference
Typically constant degree	Typically logarithmic or poly-logarithmic degree
All sets of size *at most* K expand	All sets of size *exactly* (or at least) K expand
Typically balanced	Typically unbalanced, bipartite graphs

6.2.3 List Decoding View of Extractors

In this section, we cast extractors into the same list-decoding framework that we used to capture list-decodable codes, samplers, and expanders (in Section 5.3). Recall that all of these objects could be syntactically described as functions $\Gamma : [N] \times [D] \to [M]$, and their properties could be captured by bounding the sizes of sets of the form $\mathrm{LIST}_\Gamma(T,\varepsilon) \stackrel{\text{def}}{=} \{x : \Pr_y[\Gamma(x,y) \in T] > \varepsilon\}$ for $T \subset [M]$. We also considered a generalization to functions $f:[M]\to[0,1]$ where we defined $\mathrm{LIST}_\Gamma(f,\varepsilon)\overset{\mathrm{def}}{=}\mathrm{LIST}_\Gamma(f,\varepsilon)=\{x:\mathrm{E}_y[f(\Gamma(x,y))]>\varepsilon\}$.

Conveniently, an extractor $\mathrm{Ext}:\{0,1\}^n\times\{0,1\}^d\to\{0,1\}^m$ already meets the syntax of a function $\Gamma:[N]\times[D]\to[M]$ (matching our convention that $N=2^n,\ D=2^d,\ M=2^m$). The extraction property can be described in the list-decoding framework as follows:

Proposition 6.23. Let $\Gamma=\mathrm{Ext}:[N]\times[D]\to[M]$, let $K=2^k\in\mathbb{N}$, and $\varepsilon\in[0,1]$.

(1) If Ext is a (k,ε) extractor, then for every $f:[M]\to[0,1]$, we have

$$|\mathrm{LIST}_\Gamma(f,\mu(f)+\varepsilon)|<K.$$

(2) Suppose that for every $T\subset[M]$, we have

$$|\mathrm{LIST}_\Gamma(T,\mu(T)+\varepsilon)|\le K.$$

Then Ext is a $(k+\log(1/\varepsilon),2\varepsilon)$ extractor.

Proof.

(1) Suppose for contradiction that $|\mathrm{LIST}_\Gamma(f,\mu(f)+\varepsilon)|\ge K$. Let X be uniformly distributed over $\mathrm{LIST}_\Gamma(f,\mu(f)+\varepsilon)$. Then X is a k-source, and

$$\begin{aligned}\mathrm{E}[f(\mathrm{Ext}(X,U_{[D]}))] &= \mathop{\mathrm{E}}_{x\overset{\mathrm{R}}{\leftarrow}X}[f(\mathrm{Ext}(x,U_{[D]}))]\\ &> \mu(f)+\varepsilon\\ &= \mathrm{E}[f(U_{[M]})]+\varepsilon.\end{aligned}$$

By Problem 6.1, this implies that $\mathrm{Ext}(X,U_{[D]})$ and $U_{[M]}$ are ε-far, contradicting the hypothesis that Ext is a (k,ε) extractor.

(2) Let X be any $(k+\log(1/\varepsilon))$-source. We need to show that $\mathrm{Ext}(X,U_{[D]})$ is 2ε-close to $U_{[M]}$. That is, we need to show that for every $T\subset[M]$, $\Pr[\mathrm{Ext}(X,U_{[D]})\in T]\le\mu(T)+2\varepsilon$.

So let T be any subset of $[M]$. Then

$$\begin{aligned}
&\Pr[\mathrm{Ext}(X,U_{[D]}) \in T] \\
&\quad \leq \Pr[X \in \mathrm{LIST}(T,\mu(T)+\varepsilon)] \\
&\qquad + \Pr[\mathrm{Ext}(X,U_{[D]}) \in T | X \notin \mathrm{LIST}(T,\mu(T)+\varepsilon)] \\
&\quad \leq |\mathrm{LIST}(T,\mu(T)+\varepsilon)| \cdot 2^{-(k+\log(1/\varepsilon))} + (\mu(T)+\varepsilon) \\
&\quad \leq K \cdot 2^{-(k+\log(1/\varepsilon))} + \mu(T) + \varepsilon \\
&\quad = \mu(T) + 2\varepsilon.
\end{aligned}$$

□

The proposition does not give an *exact* list-decoding characterization of extractors, as the two parts are not exactly converses of each other. One difference is the extra $\log(1/\varepsilon)$ bits of entropy and the factor of 2 in ε appearing in Part 6.23. These are typically insignificant differences for extractors; indeed even an optimal extractor loses $\Theta(\log(1/\varepsilon))$ bits of entropy (cf. Theorem 6.14). A second difference is that Part 6.23 shows that extractors imply bounds on $|\mathrm{LIST}(f,\mu(f)+\varepsilon)|$ even for fractional functions f, whereas Part 6.23 only requires bounds on $|\mathrm{LIST}(T,\mu(T)+\varepsilon)|$ to deduce the extractor property. This only makes the result stronger, and indeed we will utilize this below.

Notice that the conditions characterizing extractors here are *identical* to the ones characterizing averaging samplers in Proposition 5.30. Actually, the condition in Part 6.23 is the one characterizing averaging samplers, whereas the condition in Part 6.23 is the one characterizing boolean averaging samplers. Thus we have:

Corollary 6.24. Let $\mathrm{Ext} : [N] \times [D] \to [M]$ and $\mathrm{Samp} : [N] \to [M]^D$ be such that $\mathrm{Ext}(x,y) = \mathrm{Samp}(x)_y$. Then:

(1) If Ext is a (k,ε) extractor, then Samp is a $(K/N,\varepsilon)$ averaging sampler, where $K = 2^k$.
(2) If Samp is a $(K/N,\varepsilon)$ boolean averaging sampler, then Ext is a $(k+\log(1/\varepsilon),2\varepsilon)$ extractor.
(3) If Samp is a (δ,ε) boolean averaging sampler, then Samp is a $(\delta/\varepsilon,2\varepsilon)$ averaging sampler.

Thus, the only real difference between extractors and averaging samplers is one of perspective, and both perspectives can be useful. For example, in samplers, we measure the error probability $\delta = K/N = 2^k/2^n$, whereas in extractors we measure the min-entropy threshold k on its own. Thus, the sampler perspective can be more natural when δ is relatively large compared to $1/N$, and the extractor perspective when δ becomes quite close to $1/N$. Indeed, an extractor for min-entropy $k = o(n)$ corresponds to a sampler with error probability $\delta = 1/2^{(1-o(1))n}$, which means that each of the n bits of randomness used by the sampler reduces the error probability by almost a factor of 2!

We can now also describe the close connection between strong extractors and list-decodable codes when the alphabet size/output length is small.

Proposition 6.25. Let $\mathrm{Ext} : [N] \times [D] \to [M]$ and $\mathrm{Enc} : [N] \to [M]^D$ be such that $\mathrm{Ext}(x,y) = \mathrm{Enc}(x)_y$, and let $K = 2^k \in \mathbb{N}$.

(1) If Ext is a strong (k,ε) extractor, then Enc is $(1 - 1/M - \varepsilon, K)$ list-decodable.
(2) If Enc is $(1 - 1/M - \varepsilon, K)$ list-decodable, then Ext is a $(k + \log(1/\varepsilon), M \cdot \varepsilon)$ strong extractor.

Proof.

(1) Follows from Corollaries 5.31 and 6.24.
(2) Let X be a $(k + \log(1/\varepsilon))$-source and $Y = U_{[D]}$. Then the statistical difference between $(Y, \mathrm{Ext}(X,Y))$ and $Y \times U_{[M]}$ equals

$$\begin{aligned}
&\Delta((Y,\mathrm{Ext}(X,Y)), Y \times U_{[M]}) \\
&\quad = \mathop{\mathrm{E}}_{y \xleftarrow{\mathrm{R}} Y}[\Delta(\mathrm{Ext}(X,y), U_{[M]})] \\
&\quad = \mathop{\mathrm{E}}_{y \xleftarrow{\mathrm{R}} Y}[\Delta(\mathrm{Enc}(X)_y, U_{[M]})] \\
&\quad \le \frac{M}{2} \mathop{\mathrm{E}}_{y \xleftarrow{\mathrm{R}} Y}\left[\max_z \Pr[\mathrm{Enc}(X)_y = z] - 1/M\right],
\end{aligned}$$

where the last inequality follows from the ℓ_1 formulation of statistical difference.

So if we define $r \in [M]^D$ by setting r_y to be the value z maximizing $\Pr[\mathrm{Enc}(X)_y = z] - 1/M$, we have:

$$\begin{aligned}
&\Delta((Y,\mathrm{Ext}(X,Y)), Y \times U_{[M]}) \\
&\quad \le \frac{M}{2} \cdot (\Pr[(Y,\mathrm{Enc}(X)_Y) \in T_r] - 1/M), \\
&\quad \le \frac{M}{2} \cdot (\Pr[X \in \mathrm{LIST}(T_r, 1/M + \varepsilon)] + \varepsilon) \\
&\quad \le \frac{M}{2} \cdot (2^{-(k+\log(1/\varepsilon))} \cdot K + \varepsilon) \\
&\quad \le M \cdot \varepsilon. \qquad \square
\end{aligned}$$

Note that the quantitative relationship between extractors and list-decodable codes given by Proposition 6.25 deteriorates extremely fast as the output length/alphabet size increases. Nevertheless, the list-decoding view of extractors as given in Proposition 6.23 turns out to be quite useful.

6.3 Constructing Extractors

In the previous sections, we have seen that very good extractors exist — extracting almost all of the min-entropy from a source with only a logarithmic seed length. But the explicit constructions we have seen (via universal hashing and spectral expanders) are still quite far from optimal in seed length, and in particular cannot be used to give a polynomial-time simulation of **BPP** with a weak random source.

Fortunately, much better extractor constructions are known — ones that extract any constant fraction of the min-entropy using a logarithmic seed length, or extract all of the min-entropy using a polylogarithmic seed length. In this section, we will see how to construct such extractors.

6.3.1 Block Sources

We introduce a useful model of sources that has more structure than an arbitrary k-source:

Definition 6.26. A random variable $X = (X_1, X_2, \ldots, X_t)$ is a $(k_1, k_2, \ldots, k_t)$ *block source* if for every $x_1, \ldots, x_{i-1}$, $X_i|_{X_1=x_1,\ldots,X_{i-1}=x_{i-1}}$ is a k_i-source. If $k_1 = k_2 = \cdots = k_t = k$, then we call X a $t \times k$ *block source*.

Note that a $(k_1, k_2, \ldots, k_t)$ block source is also a $(k_1 + \cdots + k_t)$-source, but it comes with additional structure — each block is guaranteed to contribute some min-entropy. Thus, extracting randomness from block sources is an easier task than extracting from general sources.

The study of block sources has a couple of motivations.

- They are a natural and plausible model of sources in their own right. Indeed, they are more general than unpredictable-bit sources of Section 6.1.1: if $X \in \text{UnpredBits}_{n,\delta}$ is broken into t blocks of length $\ell = n/t$, then the result is a $t \times \delta'\ell$ block source, where $\delta' = \log(1/(1-\delta))$.
- We can construct extractors for general weak sources by converting a general weak source into a block source. We will see how to do this later in the section.

We now illustrate how extracting from block sources is easier than from general sources. The idea is that we can extract almost-uniform bits from later blocks that are essentially independent of earlier blocks, and hence use these as a seed to extract more bits from the earlier blocks. Specifically, for the case of two blocks we have the following:

Lemma 6.27. Let $\text{Ext}_1 : \{0,1\}^{n_1} \times \{0,1\}^{d_1} \to \{0,1\}^{m_1}$ be a (k_1, ε_1)-extractor, and $\text{Ext}_2 : \{0,1\}^{n_2} \times \{0,1\}^{d_2} \to \{0,1\}^{m_2}$ be a (k_2, ε_2)-extractor with $m_2 \geq d_1$. Define $\text{Ext}'((x_1, x_2), y_2) = (\text{Ext}_1(x_1, y_1), z_2)$, where (y_1, z_2) is obtained by partitioning $\text{Ext}_2(x_2, y_2)$ into a prefix y_1 of length d_1 and a suffix z_2 of length $m_2 - d_1$.

Then for every (k_1, k_2) block source $X = (X_1, X_2)$ taking values in $\{0,1\}^{n_1} \times \{0,1\}^{n_2}$, it holds that $\text{Ext}'(X, U_{d_2})$ is $(\varepsilon_1 + \varepsilon_2)$-close to $U_{m_1} \times U_{m_2 - d_1}$.

Proof. Since X_2 is a k_2-source conditioned on any value of X_1 and Ext_2 is a (k_2,ε_2)-extractor, it follows that $(X_1,Y_1,Z_2) = (X_1,\text{Ext}_2(X_2,U_{d_2}))$ is ε_2-close to $(X_1,U_{m_2}) = (X_1,U_{d_1},U_{m_2-d_1})$.

Thus, $(\text{Ext}_1(X_1,Y_1),Z_2)$ is ε_2-close to $(\text{Ext}_1(X_1,U_{d_1}),U_{m_2-d_1})$, which is ε_1-close to $(U_{m_1},U_{m_2-d_1})$ because X_1 is a k_1-source and Ext_1 is a (k_1,ε_1)-extractor.

By the triangle inequality, $\text{Ext}'(X,U_{d_2}) = (\text{Ext}_1(X_1,Y_1),Z_2)$ is $(\varepsilon_1+\varepsilon_2)$-close to $(U_{m_1},U_{m_2-d_1})$. □

The benefit of this composition is that the seed length of Ext' depends only one of the extractors (namely Ext_2) rather than being the sum of the seed lengths. (If this is reminiscent of the zig–zag product, it is because they are closely related — see Section 6.3.5). Thus, we get to extract from multiple blocks at the “price of one.” Moreover, since we can take $d_1 = m_2$, which is typically much larger than d_2, the seed length of Ext' can even be smaller than that of Ext_1.

The lemma extends naturally to extracting from many blocks:

Lemma 6.28. For $i = 1,\ldots,t$, let $\text{Ext}_i : \{0,1\}^{n_i} \times \{0,1\}^{d_i} \to \{0,1\}^{m_i}$ be a (k_i,ε_i)-extractor, and suppose that $m_i \geq d_{i-1}$ for every $i = 1,\ldots,t$, where we define $d_0 = 0$. Define $\text{Ext}'((x_1,\ldots,x_t),y_t) = (z_1,\ldots,z_t)$, where for $i = t,\ldots,1$, we inductively define (y_{i-1},z_i) to be a partition of $\text{Ext}_i(x_i,y_i)$ into a d_{i-1}-bit prefix and a $(m_i - d_{i-1})$-bit suffix.

Then for every $(k_1,\ldots,k_t)$ block source $X = (X_1,\ldots,X_t)$ taking values in $\{0,1\}^{n_1} \times \cdots \{0,1\}^{n_t}$, it holds that $\text{Ext}'(X,U_{d_t})$ is ε-close to U_m for $\varepsilon = \sum_{i=1}^{t} \varepsilon_i$ and $m = \sum_{i=1}^{t}(m_i - d_{i-1})$.

We remark that this composition preserves “strongness.” If each of the Ext_is correspond to strong extractors in the sense that their seeds are prefixes of their outputs, then Ext' will also correspond to a strong extractor. If in addition $d_1 = d_2 = \cdots = d_t$, then this construction can be seen as simply using the same seed to extract from all blocks.

Already with this simple composition, we can simulate **BPP** with an unpredictable-bit source (even though deterministic extraction from such sources is impossible by Proposition 6.6). As noted above, by breaking an unpredictable-bit source X with parameter δ into blocks of length ℓ, we obtain a $t \times k$ block source for $t = n/\ell$, $k = \delta'\ell$, and $\delta' = \log(1/(1-\delta))$.

Suppose that δ is a constant. Set $\ell = (10/\delta')\log n$, so that X is a $t \times k$ block source for $k = 10\log n$, and define $\varepsilon = n^{-2}$. Letting $\mathrm{Ext} : \{0,1\}^{\ell} \times \{0,1\}^{d} \to \{0,1\}^{d+m}$ be the (k,ε) extractor using universal hash functions (Theorem 6.18), we have:

$$\begin{aligned} d &= O(\ell) = O(\log n) \quad \text{and} \\ m &= k - 2\log\frac{1}{\varepsilon} - O(1) > k/2 \end{aligned}$$

Composing Ext with itself t times as in Lemma 6.27, we obtain $\mathrm{Ext}' : \{0,1\}^{t\cdot\ell} \times \{0,1\}^{d} \to \{0,1\}^{d+t\cdot m}$ such that $\mathrm{Ext}'(X,U_d)$ is ε'-close to uniform, for $\varepsilon' = 1/n$. (Specifically, $\mathrm{Ext}'((x_1,\ldots,x_t),h) = (h,h(x_1),\ldots,h(x_t))$.) This tells us that Ext' essentially extracts half of the min-entropy from X, given a random seed of logarithmic length. Plugging this extractor into the construction of Proposition 6.15 gives us the following result.

Theorem 6.29. For every constant $\delta > 0$, we can simulate **BPP** with an unpredictable-bit source of parameter δ. More precisely, for every $L \in \mathbf{BPP}$ and every constant $\delta > 0$, there is a polynomial-time algorithm A and a polynomial q such that for every $w \in \{0,1\}^*$ and every source $X \in \mathrm{UnpredBits}_{q(|w|),\delta}$, the probability that $A(w;X)$ errs is at most $1/|w|$.

6.3.2 Reducing General Sources to Block Sources

Given the results of the previous section, a common approach to constructing extractors for general k-sources is to reduce the case of general k-sources to that of block sources.

One approach to doing this is as follows. Given a k-source X of length n, where $k = \delta n$, pick a (pseudo)random subset S of the bits of X, and let $W = X|_S$ be the bits of X in those positions. If the set S is of size ℓ, then we expect that W will have at least roughly $\delta\ell$ bits of min-entropy (with high probability over the choice of S). Moreover, W reveals at most ℓ bits of information, so if $\ell < \delta n$, intuitively there should still be min-entropy left in X. (This is justified by Lemma 6.30 below.) Thus, the pair (W, X) should be a block source. This approach can be shown to work for appropriate ways of sampling the set S, and recursive applications of it formed the original approach to constructing good extractors (and is still useful in various contexts today). The fact mentioned above, that conditioning on a string of length ℓ reduces min-entropy by at most ℓ bits, is given by the following lemma (which is very useful when working with min-entropy).

Lemma 6.30 (chain rule for min-entropy). If (W, X) are two jointly distributed random variables, where (W, X) is a k-source and $|\text{Supp}(W)| \leq 2^{\ell}$, then for every $\varepsilon > 0$, it holds that with probability at least $1 - \varepsilon$ over $w \overset{\text{R}}{\leftarrow} W$, $X|_{W=w}$ is a $(k - \ell - \log(1/\varepsilon))$-source.

The proof of this lemma is left as an exercise (Problem 6.1).

This is referred to as the "chain rule" for min-entropy by analogy with the chain rule for Shannon entropy, which states that $\mathrm{H}_{Sh}(X|W) = \mathrm{H}_{Sh}(W, X) - \mathrm{H}_{Sh}(W)$, where the conditional Shannon entropy is defined to be $\mathrm{H}_{Sh}(X|W) = \mathrm{E}_{w \overset{\text{R}}{\leftarrow} W}[\mathrm{H}_{Sh}(X|_{W=w})]$. Thus, if $\mathrm{H}_{Sh}(W, X) \geq k$ and W is of length at most ℓ, we have $\mathrm{H}_{Sh}(X|W) \geq k - \ell$. The chain rule for min-entropy is not quite as clean; we need to assume that W has small support (rather than just small min-entropy) and we lose $\log(1/\varepsilon)$ bits of additional min-entropy. (The $\log(1/\varepsilon)$ bits of entropy can be saved by using an appropriate notion of conditional min-entropy; see Problems 6.7 and 6.8.)

Another approach, which we will follow, is based on the observation that every source of high min-entropy rate (namely, greater than 1/2) is (close to) a block source, as shown by the lemma below. Thus, we will try to convert arbitrary sources into ones of high min-entropy rate.

Lemma 6.31. If X is an $(n - \Delta)$-source of length n, and $X = (X_1, X_2)$ is a partition of X into blocks of lengths n_1 and n_2, then for every $\varepsilon > 0$, (X_1, X_2) is ε-close to some $(n_1 - \Delta, n_2 - \Delta - \log(1/\varepsilon))$ block source.

Consider $\Delta = \alpha n$ for a constant $\alpha < 1/2$, and $n_1 = n_2 = n/2$. Then each block contributes min-entropy at least $(1/2 - \alpha)n$. The proofs of Lemmas 6.30 and 6.31 are left as exercises (Problem 6.1).

6.3.3 Condensers

The previous section left us with the problem of converting a general k-source into one of high min-entropy rate. We will do this via the following kind of object:

Definition 6.32. A function $\mathrm{Con}: \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ is a $k \to_\varepsilon k'$ *condenser* if for every k-source X on $\{0,1\}^n$, $\mathrm{Con}(X,U_d)$ is ε-close to some k'-source. Con is *lossless* if $k' = k + d$.

If $k'/m > k/n$, then the condenser increases the min-entropy rate, intuitively making extraction an easier task. Indeed, condensers with $k' = m$ are simply extractors themselves.

Like extractors, it is often useful to view condensers graph-theoretically. Specifically, we think of Con as the neighbor function of a bipartite multigraph G with $N = 2^n$ left vertices, left degree $D = 2^d$, and $M = 2^m$ right vertices.

The *lossless* condensing property of Con turns out to be equivalent to G having vertex expansion close to the degree:

Proposition 6.33. Let $n, d, m \in \mathbb{N}$, $K = 2^k \in \mathbb{N}$, and $\varepsilon > 0$. A function $\mathrm{Con}: \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ is a $k \to_\varepsilon k + d$ lossless condenser if and only if the corresponding bipartite multigraph $G = ([N],[M],E)$ of left degree D is an $(= K, (1-\varepsilon)D)$ vertex expander.

Proof. $\Rightarrow$: Let $S \subset [N]$ be any set of size K. Then U_S is a k-source, so $\mathrm{Con}(U_S,U_d)$ is ε-close to a $(k+d)$-source. This implies that $|\mathrm{Supp}(\mathrm{Con}(U_S,U_d))| \geq (1-\varepsilon)\cdot 2^{k+d}$. Noting that $\mathrm{Supp}(\mathrm{Con}(U_S,U_d)) = N(S)$ completes the proof.

$\Leftarrow$: By Lemma 6.10, it suffices to prove that for every subset $S \subset [N]$ of size K, it holds that $\mathrm{Con}(U_S,U_d)$ is ε-close to a $(k+d)$-source. By expansion, we know that $|N(S)| \geq (1-\varepsilon)\cdot KD$. Since there are only KD edges leaving S, by redirecting εKD of the edges, we can ensure that they all go to KD distinct vertices in $[M]$. The uniform distribution on these KD vertices is a $(k+d)$-source that is ε-close to $\mathrm{Con}(U_S,U_d)$. □

Recall that vertex expansion normally corresponds to the *disperser* property (see Proposition 6.20), which is weaker than extraction and condensing. Indeed, vertex expansion generally does not guarantee much about the distribution induced on a random neighbor of a set S, except that its support is large. However, in case the expansion is very close to the degree ($A = (1-\varepsilon)\cdot D$), then the distribution must be nearly flat (as noted in the above proof).

By applying Proposition 6.33 to the expanders based on Parvaresh–Vardy codes (Theorem 5.35), we get the following lossless condenser:

Theorem 6.34. For every constant $\alpha > 0$, for all positive integers $n \geq k$ and all $\varepsilon > 0$, there is an explicit

$$k \to_\varepsilon k + d$$

lossless condenser $\mathrm{Con} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ with $d = O(\log n + \log(1/\varepsilon))$ and $m = (1+\alpha)k + O(\log(n/\varepsilon))$.

Note that setting α to be a small constant, we obtain an output min-entropy rate arbitrarily close to 1.

Problem 6.5 gives a simple extractor Ext for high min-entropy rate when the error parameter ε is constant. Applying that extractor to the output of the above condenser, we obtain extractors with a seed length of $d = O(\log n)$ that extract $\Omega(k)$ almost-uniform bits (with constant error ε) from sources of any desired min-entropy k. In the next section, we will use the above condenser to give an efficient construction of extractors for arbitrary values of ε.

We remark that the fact that having output min-entropy rate bounded away from 1 is not inherent for lossless condensers. Nonconstructively, there exist lossless condensers with output length $m = k + d + \log(1/\varepsilon) + O(1)$, and Open Problem 5.36 about expanders can be restated in the language of lossless condensers as follows:

Open Problem 6.35. Give an explicit construction of a $k \to_\varepsilon k + d$ lossless condenser $\mathrm{Con} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ with $d = O(\log n)$, $m = k + d + O(1)$, and $\varepsilon = 0.01$.

If we had such a condenser, then we could get extractors that extract *all* but $O(1)$ bits of the min-entropy by then applying extractors based on spectral expanders (Theorem 6.22).

6.3.4 The Extractor

In this section, we will use the ideas outlined in the previous section — namely condensing and block-source extraction — to construct an extractor that is optimal up to constant factors.

Theorem 6.36. For all positive integers $n \geq k$ and all $\varepsilon > 0$, there is an explicit (k,ε) extractor $\mathrm{Ext} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ with $m \geq k/2$ and $d = O(\log(n/\varepsilon))$.

We will use the following building block, constructed in Problem 6.9.

Lemma 6.37. For every *constant* $t > 0$ and all positive integers $n \geq k$ and all $\varepsilon > 0$, there is an explicit (k,ε) extractor $\mathrm{Ext} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ with $m \geq k/2$ and $d = k/t + O(\log(n/\varepsilon))$.

The point is that this extractor has a seed length that is an arbitrarily large constant factor (approximately $t/2$) smaller than its output length. Thus, if we use it as Ext_2 in the block-source extraction of Lemma 6.27, the resulting seed length will be smaller than that of Ext_1 by an arbitrarily large constant factor. (The seed length of the composed extractor Ext' in Lemma 6.27 is the same of that as Ext_2, which will be a constant factor smaller than its output length m_2, which we can take to be equal to the seed length d_1 of Ext_1.)

Overview of the Construction. Note that for small min-entropies k, namely $k = O(\log(n/\varepsilon))$, the extractor we want is already given by Lemma 6.37 with seed length d smaller than the output length m by any constant factor. (If we allow $d \geq m$, then extraction is trivial — just output the seed.) Thus, our goal will be to recursively construct extractors for large min-entropies using extractors for smaller min-entropies. Of course, if $\mathrm{Ext} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ is a (k_0,ε)

extractor, say with $m = k_0/2$, then it is also a (k,ε) extractor for every $k \geq k_0$. The problem is that the output length is only $k_0/2$ rather than $k/2$. Thus, we need to increase the output length. This can be achieved by simply applying extractors for smaller min-entropies several times:

Lemma 6.38. Suppose $\mathrm{Ext}_1\colon \{0,1\}^n \times \{0,1\}^{d_1} \to \{0,1\}^{m_1}$ is a (k_1,ε_1) extractor and $\mathrm{Ext}_2 : \{0,1\}^n \times \{0,1\}^{d_2} \to \{0,1\}^{m_2}$ is a (k_2,ε_2) extractor for $k_2 = k_1 - m_1 - \log(1/\varepsilon_3)$. Then $\mathrm{Ext}' : \{0,1\}^n \times \{0,1\}^{d_1+d_2} \to \{0,1\}^{m_1+m_2}$ defined by $\mathrm{Ext}'(x,(y_1,y_2)) = (\mathrm{Ext}_1(x,y_1),\mathrm{Ext}_2(x,y_2))$ is a $(k_1,\varepsilon_1 + \varepsilon_2 + \varepsilon_3)$ extractor.

The proof of this lemma follows from Lemma 6.30. After conditioning a k_1-source X on $W = \mathrm{Ext}_1(X,U_{d_1})$, X still has min-entropy at least $k_1 - m_1 - \log(1/\varepsilon_3) = k_2$ (except with probability ε_3), and thus $\mathrm{Ext}_2(X,U_{d_2})$ can extract an additional m_2 almost-uniform bits.

To see how we might apply this, consider setting $k_1 = 0.8k$ and $m_1 = k_1/2$, $\varepsilon_1 = \varepsilon_2 = \varepsilon_3 = \varepsilon \geq 2^{-0.1k}$, $k_2 = k_1 - m_1 - \log(1/\varepsilon_3) \in [0.3k, 0.4k]$, and $m_2 = k_2/2$. Then we obtain a $(k,3\varepsilon)$ extractor Ext' with output length $m = m_1 + m_2 > k/2$ from two extractors for min-entropies k_1,k_2 that are smaller than k by a constant factor, and we can hope to construct the latter two extractors recursively via the same construction.

Now, however, the problem is that the seed length grows by a constant factor in each level of recursion (e.g., if $d_1 = d_2 = d$ in Lemma 6.38, we get seed length $2d$ rather than d). Fortunately, block source extraction using the extractor of Lemma 6.37 gives us a method to reduce the seed length by a constant factor. (The seed length of the composed extractor Ext' in Lemma 6.27 is the same of that as Ext_2, which will be a constant factor smaller than its output length m_2, which we can take to be equal to the seed length d_1 of Ext_1. Thus, the seed length of Ext' will be a constant factor smaller than that of Ext_1.) In order to apply block source extraction, we first need to convert our source to a block source; by Lemma 6.31, we can do this by using the condenser of Theorem 6.34 to make its entropy rate close to 1.

One remaining issue is that the error ε still grows by a constant factor in each level of recursion. However, we can start with polynomially small error at the base of the recursion and there are only logarithmically many levels of recursion, so we can afford this blow-up.

We now proceed with the proof details. It will be notationally convenient to do the steps in the reverse order from the description above — first we will reduce the seed length by a constant factor via block-source extraction, and then apply Lemma 6.38 to increase the output length.

Proof of Theorem 6.36. Fix $n \in \mathbb{N}$ and $\varepsilon_0 > 0$. Set $d = c\log(n/\varepsilon_0)$ for an error parameter ε_0 and a sufficiently large constant c to be determined in the proof below. (To avoid ambiguity, we will keep the dependence on c explicit throughout the proof, and all big-Oh notation hides universal constants independent of c.) For $k \in [0,n]$, let $i(k)$ be the smallest nonnegative integer i such that $k \le 2^i \cdot 8d$. This will be the level of recursion in which we handle min-entropy k; note that $i(k) \le \log k \le \log n$.

For every $k \in [0,n]$, we will construct an explicit $\mathrm{Ext}_k : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^{k/2}$ that is a $(k, \varepsilon_{i(k)})$ extractor, for an appropriate sequence $\varepsilon_0 \le \varepsilon_1 \le \varepsilon_2 \cdots$. Note that we require the seed length to remain d and the fraction of min-entropy extracted to remain $1/2$ for all values of k. The construction will be by induction on $i(k)$.

Base Case: $i(k) = 0$, i.e., $k \le 8d$. The construction of Ext_k follows from Lemma 6.37, setting $t = 9$ and taking c to be a sufficiently large constant.

Inductive Case: We construct Ext_k for $i(k) \ge 1$ from extractors $\mathrm{Ext}_{k'}$ with $i(k') < i(k)$ as follows. Given a k-source X of length n, Ext_k works as follows.

(1) We apply the condenser of Theorem 6.34 to convert X into a source X' that is ε_0-close to a k-source of length $(9/8)k + O(\log(n/\varepsilon_0))$. This requires a seed of length $O(\log(n/\varepsilon_0))$.
(2) We divide X' into two equal-sized halves (X_1, X_2). By Lemma 6.31, (X_1, X_2) is $2\varepsilon_0$-close to a $2 \times k'$ block source for

$$k' = k/2 - k/8 - O(\log(n/\varepsilon_0)).$$

Note that $i(k') < i(k)$. Since $i(k) \geq 1$, we also have $k' \geq 3d - O(\log(n/\varepsilon_0)) \geq 2d$, for a sufficiently large choice of the constant c.

(3) Now we apply block-source extraction as in Lemma 6.27. We take Ext_2 to be a $(2d, \varepsilon_0)$ extractor from Lemma 6.37 with parameter $t = 16$, which will give us $m_2 = d$ output bits using a seed of length $d_2 = (2d)/16 + O(\log(n/\varepsilon_0))$. For Ext_1, we use our recursively constructed $\mathrm{Ext}_{k'}$, which has seed length d, error $\varepsilon_{i(k')}$, and output length $k'/2 \geq k/6$ (where the latter inequality holds for a sufficiently large choice of the constant c, because $k > 8d > 8c\log(1/\varepsilon)$).

All in all, our extractor so far has seed length at most $d/8 + O(\log(n/\varepsilon_0))$, error at most $\varepsilon_{i(k)-1} + O(\varepsilon_0)$, and output length at least $k/6$. This would be sufficient for our induction except that the output length is only $k/6$ rather than $k/2$. We remedy this by applying Lemma 6.38.

With one application of the extractor above, we extract at least $m_1 = k/6$ bits of the source min-entropy. Then with another application of the extractor above for min-entropy threshold $k_2 = k - m_1 - \log(1/\varepsilon_0) = 5k/6 - \log(1/\varepsilon_0)$, by Lemma 6.38, we extract another $(5k/6 - \log(1/\varepsilon_0))/6$ bits and so on. After four applications, we have extracted all but $(5/6)^4 \cdot k + O(\log(1/\varepsilon_0)) \leq k/2$ bits of the min-entropy. Our seed length is then $4 \cdot (d/8 + O(\log(n/\varepsilon_0))) \leq d$ and the total error is $\varepsilon_{i(k)} = O(\varepsilon_{i(k)-1})$.

Solving the recurrence for the error, we get $\varepsilon_i = 2^{O(i)} \cdot \varepsilon_0 \leq \mathrm{poly}(n) \cdot \varepsilon_0$, so we can obtain error ε by setting $\varepsilon_0 = \varepsilon/\mathrm{poly}(n)$. As far as explicitness, we note that computing Ext_k consists of four evaluations of our condenser from Theorem 6.34, four evaluations of $\mathrm{Ext}_{k'}$ for values of k' such that $i(k') < (i(k) - 1)$, four evaluations of the explicit extractor from Lemma 6.37, and simple string manipulations that can be done in time $\mathrm{poly}(n,d)$. Thus, the total computation time is at most $4^{i(k)} \cdot \mathrm{poly}(n,d) = \mathrm{poly}(n,d)$. □

Repeatedly applying Lemma 6.38 using extractors from Theorem 6.36, we can extract any constant fraction of the min-entropy using a logarithmic seed length, and all the min-entropy using a polylogarithmic seed length.

Corollary 6.39. The following holds for every constant $\alpha > 0$. For every $n \in \mathbb{N}$, $k \in [0,n]$, and $\varepsilon > 0$, there is an explicit (k,ε) extractor $\mathrm{Ext} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ with $m \geq (1-\alpha)k$ and $d = O(\log(n/\varepsilon))$.

Corollary 6.40. For every $n \in \mathbb{N}$, $k \in [0,n]$, and $\varepsilon > 0$, there is an explicit (k,ε) extractor $\mathrm{Ext} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ with $m = k - O(\log(1/\varepsilon))$ and $d = O(\log k \cdot \log(n/\varepsilon))$.

We remark that the above construction can be modified to yield *strong* extractors achieving the same output lengths as above (so the entropy of the seed need not be lost in Corollary 6.40).

A summary of the extractor parameters we have seen is in the following table.

Table 6.2. Parameters for some constructions of $(k, 0.01)$ extractors.

Method	Seed Length d	Output Length m
Optimal and Nonconstructive	$\log(n-k) + O(1)$	$k + d - O(1)$
Necessary for **BPP** Simulation	$O(\log n)$	$k^{\Omega(1)}$
Spectral Expanders	$O(n-k)$	n
Pairwise Independent Hashing	$O(n)$	$k + d - O(1)$
Corollary 6.39	$O(\log n)$	$(1-\gamma)k$, any constant $\gamma > 0$
Corollary 6.40	$O(\log^2 n)$	$k - O(1)$

While Theorem 6.36 and Corollary 6.39 give extractors that are optimal up to constant factors in both the seed length and output length, it remains an important open problem to get one or both of these to be optimal to within an *additive* constants while keeping the other optimal to within a constant factor.

Open Problem 6.41. Give an explicit construction of $(k, 0.01)$ extractors $\mathrm{Ext} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ with seed length $d = O(\log n)$ and output length $m = k + d - O(1)$.

By using the condenser of Theorem 6.34, it suffices to achieve the above for high min-entropy rate, e.g., $k = 0.99n$. Alternatively, a better condenser construction would also resolve the problem. (See Open Problem 6.35.) We note that there is a recent construction of extractors with seed length $d = O(\log n)$ and output length $m = (1 - 1/\mathrm{polylog}(n))k$ (improving the output length of $m = \Omega(k)$ of Corollary 6.39, in the case of constant or slightly subconstant error).

Open Problem 6.42. Give an explicit construction of $(k, 0.01)$ extractors $\mathrm{Ext} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ with seed length $d = \log n + O(1)$ and $m = \Omega(k)$ (or even $m = k^{\Omega(1)}$).

One of the reasons that these open problems are significant is that, in many applications of extractors, the resulting complexity depends *exponentially* on the seed length d and/or the entropy loss $k + d - m$. (An example is the simulation of **BPP** with weak random sources given by Proposition 6.15.) Thus, additive constants in these parameters corresponds to constant multiplicative factors in complexity.

Another open problem is more aesthetic in nature. The construction of Theorem 6.36 makes use of the condenser of Theorem 6.36, the Leftover Hash Lemma (Theorem 6.18) and the composition techniques of Lemmas 6.27 and 6.38 in a somewhat complex recursion. It is of interest to have a construction that is more direct. In addition to the aesthetic appeal, such a construction would likely be more practical to implement and provide more insight into extractors. In Chapter 7, we will see a very direct construction based on a connection between extractors and pseudorandom generators, but its parameters will be somewhat worse than Theorem 6.36. Thus the following remains open:

Open Problem 6.43. Give a “direct” construction of (k, ε) extractors $\mathrm{Ext} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ with seed length $d = O(\log(n/\varepsilon))$ and $m = \Omega(k)$.

6.3.5 Block-Source Extraction versus the Zig–Zag Product

To further highlight the connections between extractors and expanders, here we describe how the block-source extraction method of Section 6.3.1 (which we used in the main extractor construction of Theorem 6.36) is closely related to the zig–zag graph product of Section 4.3.2.3 (which we used in the expander construction of Theorem 4.39).

Recall the block-source extraction method of Lemma 6.27: We define $\mathrm{Ext}' : \{0,1\}^{n_1+n_2} \times \{0,1\}^{d_2} \to \{0,1\}^{m_1}$ by $\mathrm{Ext}'((x_1,x_2),y_2) = \mathrm{Ext}_1(x_1,\mathrm{Ext}_2(x_2,y_2))$. (Here we consider the special case that $m_2 = d_1$.)

Viewing the extractors as bipartite graphs, the left-vertex set is $[N_1] \times [N_2]$ and the left-degree is D_2. A random step from a vertex $(x_1,x_2) \in [N_1] \times [N_2]$ corresponds to taking a random step from x_2 in G_2 to obtain a right-hand vertex $y_1 \in \{0,1\}^{m_2}$, which we view as an edge label y for G_1. We then move to the yth neighbor of x_1.

This is just like the first two steps of the zig–zag graph product. Why do we need a third step in the zig–zag product? It is because of the slightly different goals in the two settings. In a (spectral) expander, we consider an arbitrary initial distribution that does not have too much (Rényi) entropy, and need to add entropy to it. In a block-source extractor, our initial distribution is constrained to be a block source (so both blocks have a certain amount of min-entropy), and our goal is to produce an almost-uniform output (even if we end up with less bits than the initial entropy).

Thus, in the zig–zag setting, we must consider the following extreme cases (that are ruled out for block sources):

- The second block has no entropy given the first. Here, the step using G_2 will add entropy, but not enough to make y_1 close to uniform. Thus, we have no guarantees on the behavior of the G_1-step, and we may lose entropy with it. For this reason, we keep track of the edge used in the G_1-step — that is, we remember b_1 such that x_1 is the b_1th neighbor of $z_1 = \mathrm{Ext}(x_1,y_1)$. This ensures that the (edge-rotation) mapping $(x_1,y_1) \mapsto (z_1,b_1)$ is a permutation and does not lose any entropy. We can think of b_1 as a "buffer" that retains any extra entropy in (x_1,y_1) that did not get extracted into z_1. So a natural idea is to just do block source extraction, but output (z_1,b_1) rather than just z_1. However, this runs into trouble with the next case.

- The first block has no entropy but the second block is completely uniform given the first. In this case, the G_2 step cannot add any entropy and the G_1 step does not add any entropy because it is a permutation. However, the G_1 step transfers entropy into z_1. So if we add another expander-step from b_1 at the end, we can argue that it will add entropy. This gives rise to the 3-step definition of the zig–zag product.

While we analyzed the zig–zag product with respect to spectral expansion (i.e., Rényi entropy), it is also possible to analyze it in terms of a condenser-like definition (i.e., outputting distributions ε-close to having some min-entropy). It turns out that a variant of the zig–zag product for condensers leads to a construction of constant-degree bipartite expanders with expansion $(1-\varepsilon)\cdot D$ for the balanced $(M=N)$ or nearly balanced (e.g., $M=\Omega(N)$) case. However, as mentioned in Open Problems 4.43, 4.44, and 5.36, there are still several significant open problems concerning the explicit construction of expanders with vertex expansion close to the degree, achieving expansion $D-O(1)$ in the nonbipartite case, and achieving a near-optimal number of right-hand vertices.

6.4 Exercises

Problem 6.1 (Min-entropy and Statistical Difference).

(1) Prove that for every two random variables X and Y,

$$\Delta(X,Y)=\max_f|\mathrm{E}[f(X)]-\mathrm{E}[f(Y)]|=\frac{1}{2}\cdot|X-Y|_1,$$

where the maximum is over all $[0,1]$-valued functions f. (Hint: first identify the functions f that maximize $|\mathrm{E}[f(X)]-\mathrm{E}[f(Y)]|$.)

(2) Suppose that (W, X) are jointly distributed random variables where (W, X) is a k-source and $|\mathrm{Supp}(W)| \leq 2^{\ell}$. Show that for every $\varepsilon > 0$, with probability at least $1 - \varepsilon$ over $w \overset{\mathrm{R}}{\leftarrow} W$, we have $X|_{W=w}$ is a $(k - \ell - \log(1/\varepsilon))$-source.

(3) Suppose that X is an $(n - \Delta)$-source taking values in $\{0,1\}^n$, and we let X_1 consist of the first n_1 bits of X and X_2 the remaining $n_2 = n - n_1$ bits. Show that for every $\varepsilon > 0$, (X_1, X_2) is ε-close to some $(n_1 - \Delta, n_2 - \Delta - \log(1/\varepsilon))$ block source.

Problem 6.2 (Simulating Randomized Algorithms with Weak Sources).

(1) Let $A(w;r)$ be a randomized algorithm for computing a function f using m random bits such that $A(w;U_m)$ has error probability at most $1/3$ (i.e., for every w, $\Pr[A(w;U_m) \neq f(w)] \leq 1/3$) and let $\mathrm{Ext}\colon \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ be a $(k, 1/7)$-extractor. Define $A'(w;x) = \mathrm{maj}_{y \in \{0,1\}^d}\{A(w;\mathrm{Ext}(x;y))\}$ (breaking ties arbitrarily). Show that for every $(k+t)$-source X, $A'(w;X)$ has error probability at most 2^{-t}.

(2) Let $A(w;r)$ be a randomized algorithm for deciding a language L using m random bits such that $A(w;U_m)$ has one-sided error probability at most $1/2$ (i.e., if $w \in L$, then $\Pr[A(w;U_m) = 1] \geq 1/2$ and if $w \notin L$, then $\Pr[A(w;U_m) = 1] = 0$) and let Disp: $\{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ be a $(k, 1/3)$-disperser. Show how to decide L with one-sided error at most 2^{-t} given a single sample from a $(k+t)$-source X (with no other randomness) and running A and Disp each 2^d times.

Problem 6.3 (Almost-Universal Hashing). A family $\mathcal{H}$ of functions mapping domain $[N]$ to $[M]$ is said to have *collision probability* at most δ if for every $x_1 \neq x_2 \in [N]$, we have

$$\Pr_{H \overset{\mathrm{R}}{\leftarrow} \mathcal{H}} [H(x_1) = H(x_2)] \leq \delta.$$

$\mathcal{H}$ has is ε-*almost universal* if it has collision probability at most $(1+\varepsilon)/M$. (Note that this is a relaxation of the notion of ε-almost pairwise independence from Problem 3.4.)

(1) Show that if a family $\mathcal{H} = \{h : [N] \to [M]\}$ is ε^2-almost universal, $\mathrm{Ext}(x,h) \overset{\mathrm{def}}{=} h(x)$ is a (k,ε) strong extractor for $k = m + 2\log(1/\varepsilon) + O(1)$, where $m = \log M$.
(2) Use Problem 3.4 to deduce that for every $n \in \mathbb{N}$, $k \leq n$, and $\varepsilon > 0$, there is a (k,ε) strong extractor $\mathrm{Ext} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ with $d = O(k + \log(n/\varepsilon))$ and $m = k - 2\log(1/\varepsilon) - O(1)$.
(3) Given a family $\mathcal{H}$ of functions mapping $[N]$ to $[M]$, we can obtain a code $\mathrm{Enc} : [N] \to [M]^{|\mathcal{H}|}$ by $\mathrm{Enc}(x)_h = h(x)$, and conversely. Show that $\mathcal{H}$ has collision probability at most δ iff Enc has minimum distance at least $1 - \delta$.
(4) Use the above connections and the list-decoding view of extractors (Proposition 6.23) to prove the Johnson Bound for small alphabets: if a code $\mathrm{Enc} : [N] \to [M]^{\hat{n}}$ has minimum distance at least $1 - 1/M - \gamma/M$, then it is $(1 - 1/M - \sqrt{\gamma}, O(M/\gamma))$ list-decodable.

Problem 6.4 (Rényi extractors). Call a function $\mathrm{Ext} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ a (k,ε) *Rényi extractor* if for every source X on $\{0,1\}^n$ of Rényi entropy at least k, it holds that $\mathrm{Ext}(X,U_d)$ has Rényi entropy at least $m - \varepsilon$.

(1) Prove that a (k,ε) Rényi extractor is also a $(k,O(\sqrt{\varepsilon}))$ extractor.
(2) Show that for every $n,k,m \in \mathbb{N}$ with $m \leq n$ and $\varepsilon > 0$, there exists a (k,ε) Rényi extractor $\mathrm{Ext} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ with $d = O(\min\{n-k+\log(1/\varepsilon), m/2+\log(n/\varepsilon)\})$.

(3) Show that if $\mathrm{Ext} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ is a $(k,1)$ Rényi extractor, then $d \geq \min\{n-k, m/2\} - O(1)$. (Hint: consider a k-source that is uniform over $\{x : \exists y \mathrm{Ext}(x,y) \in T\}$ for an appropriately chosen set T.)

Problem 6.5 (Extractors versus Samplers). Use the connection between extractors and averaging samplers to do the following:

(1) Prove that for all constants $\varepsilon, \alpha > 0$, there is a constant $\beta < 1$ such that for all n, there is an explicit $(\beta n, \varepsilon)$ extractor $\mathrm{Ext} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ with $d \leq \log n$ and $m \geq (1-\alpha)n$.

(2) Prove that for every $m \in \mathbb{N}$, $\varepsilon, \delta > 0$, there exists a (non-constructive) (δ, ε) averaging sampler $\mathrm{Samp} : \{0,1\}^n \to (\{0,1\}^m)^t$ using $n = m + 2\log(1/\varepsilon) + \log(1/\delta) + O(1)$ random bits and $t = O((1/\varepsilon^2) \cdot \log(1/\delta))$ samples.

(1) Suppose we are given a constant-error **BPP** algorithm that uses $r = r(n)$ random bits on inputs of length n. Show how, using the explicit extractor of Theorem 6.36, we can reduce its error probability to $2^{-\ell}$ using $O(r) + \ell$ random bits, for any polynomial $\ell = \ell(n)$. (Note that this improves the $r + O(\ell)$ given by expander walks for $\ell \gg r$.) Conclude that every problem in **BPP** has a randomized polynomial-time algorithm that only errs for $2^{q^{0.01}}$ choices of its $q = q(n)$ random bits.

Problem 6.6 (Extracting from Unpredictable-Bit Sources).

(1) Let X be a source taking values in $\{0,1\}^n$ such that for all x, y, $\Pr[X = x]/\Pr[X = y] \leq (1-\delta)/\delta$. Show that $X \in \mathrm{UnpredBits}_{n,\delta}$.

(2) Prove that for every function $\mathrm{Ext} \colon \{0,1\}^n \to \{0,1\}$ and every $\delta > 0$, there exists a source $X \in \mathrm{UnpredBits}_{n,\delta}$ with parameter δ such that $\Pr[\mathrm{Ext}(X) = 1] \leq \delta$ or

$\Pr[\mathrm{Ext}(X) = 0] \geq 1 - \delta$. (Hint: for $b \in \{0,1\}$, consider X that is uniform on $\mathrm{Ext}^{-1}(b)$ with probability $1 - \delta$ and is uniform on $\mathrm{Ext}^{-1}(\bar{b})$ with probability δ.)

(3) (*) Show how to extract from sources in $\mathrm{UnpredBits}_{n,\delta}$ using a seeded extractor with a seed of *constant* length. That is, the seed length should not depend on the length n of the source, but only on the bias parameter δ and the statistical difference ε from uniform desired. The number of bits extracted should be $\Omega(\delta n)$.

Problem 6.7 (Average Min-Entropy). While there is no single "correct" definition of conditional min-entropy, in this problem you will explore one definition that is useful and has nice properties. For two jointly distributed random variables (X,Y) define

$$\mathrm{H}_\infty(Y|X) = \log\left(\frac{1}{\mathrm{E}_{x \overset{\mathrm{R}}{\leftarrow} X}\left[\max_y \Pr[Y = y | X = x]\right]}\right).$$

(1) Prove that $\mathrm{H}_\infty(Y|X) = \log(1/\max_P \Pr[P(X) = Y])$, where the maximum is over functions P from the domain of X to the domain of Y. So average min-entropy captures the unpredictability of Y from X.

(2) Show that if $\mathrm{H}_\infty(Y|X) \geq k$, then with probability at least $1 - \varepsilon$ over $x \overset{\mathrm{R}}{\leftarrow} X$, we have $\mathrm{H}_\infty(Y|_{X=x}) \geq k - \log(1/\varepsilon)$.

(3) Show that $\mathrm{H}_\infty(Y) \geq \mathrm{H}_\infty(Y|X) \geq \mathrm{H}_\infty(X,Y) - \log|\mathrm{Supp}(X)|$.

(4) Use the previous two items to deduce Problem 6.1, Part 6.1.

(5) Give an example showing that $\mathrm{H}_\infty(Y|X)$ can be much smaller than $\mathrm{H}_\infty(Y) - \mathrm{H}_\infty(X)$. Specifically, construct n-bit random variables where $\mathrm{H}_\infty(Y) = n$ but $\mathrm{H}_\infty(Y|X)$ and $\mathrm{H}_\infty(X)$ are both $O(1)$.

Problem 6.8 (Average Min-Entropy Extractors). A function $\mathrm{Ext} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ is a (k,ε) *average min-entropy* *extractor* if for every two jointly distributed random variables (X,Y) such that X takes values in $\{0,1\}^n$ and $\mathrm{H}_\infty(X|Y) \geq k$, we have $(\mathrm{Ext}(X,U_d),Y)$ is ε-close to (U_m,Y), where U_d and U_m are taken to be independent of (X,Y). By Problem 6.7, Part 6.7, it follows that if Ext is a (k,ε) extractor, then it is a $(k+\log(1/\varepsilon),2\varepsilon)$ average min-entropy extractor. Below, you will show how to avoid the $\log(1/\varepsilon)$ entropy loss from this reduction.

(1) (*) Show that if $\mathrm{Ext}: \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ is a (k,ε)-extractor with $k \leq n-1$, then for every $t > 0$, Ext is a $(k-t, 2^{t+1}\cdot\varepsilon)$-extractor. (Hint: for a statistical test $T \subseteq \{0,1\}^m$, relate the $(k-t)$-sources X maximizing the distinguishing advantage $|\Pr[\mathrm{Ext}(X,U_d) \in T] - \Pr[U_m \in T]|$ to the k-sources maximizing the distinguishing advantage.)
(2) Show that if $\mathrm{Ext}: \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ is a (k,ε)-extractor, then Ext is a $(k,3\varepsilon)$ average min-entropy extractor.
(3) Use these results and Problem 6.3 to improve Corollary‘6.40, and construct, for every $n \in \mathbb{N}$, $k \leq n$, and $\varepsilon > 0$, an explicit (k,ε) extractor $\mathrm{Ext}: \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ with seed length $d = O((\log k)\cdot(\log(n/\varepsilon)))$ and output length $m = k + d - 2\log(1/\varepsilon) - O(1)$.

Problem 6.9 (The Building-Block Extractor). Prove Lemma 6.37: Show that for every *constant* $t > 0$ and all positive integers $n \geq k$ and all $\varepsilon > 0$, there is an *explicit* (k,ε)-extractor $\mathrm{Ext}\colon \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ with $m \geq k/2$ and $d = k/t + O(\log(n/\varepsilon))$. (Hint: convert the source into a block source with blocks of length $k/O(t) + O(\log(n/\varepsilon))$.)

Problem 6.10 (Encryption and Deterministic Extraction). A (one-time) *encryption scheme* with key length n and message length m consists of an encryption function $\mathrm{Enc}\colon \{0,1\}^n \times \{0,1\}^m \to \{0,1\}^\ell$ and a decryption function $\mathrm{Dec}\colon \{0,1\}^n \times \{0,1\}^\ell \to \{0,1\}^m$ such that $\mathrm{Dec}(k,\mathrm{Enc}(k,u)) = u$ for every $k \in \{0,1\}^n$ and $u \in \{0,1\}^m$. Let K be a random variable taking values in $\{0,1\}^n$. We say that (Enc, Dec) is *(statistically) ε-secure with respect to K* if for every two messages $u,v \in \{0,1\}^m$, we have $\Delta(\mathrm{Enc}(K,u),\mathrm{Enc}(K,v)) \leq \varepsilon$. For example, the *one-time pad*, where $n = m = \ell$ and $\mathrm{Enc}(k,u) = k \oplus u = \mathrm{Dec}(k,u)$ is 0-secure (a.k.a perfectly secure) with respect to the uniform distribution $K = U_m$. For a class $\mathcal{C}$ of sources on $\{0,1\}^n$, we say that the encryption scheme (Enc, Dec) is *ε-secure with respect to $\mathcal{C}$* if Enc is ε-secure with respect to every $K \in \mathcal{C}$.

(1) Show that if there exists a deterministic ε-extractor $\mathrm{Ext}\colon \{0,1\}^n \to \{0,1\}^m$ for $\mathcal{C}$, then there exists an 2ε-secure encryption scheme with respect to $\mathcal{C}$.

(2) Conversely, use the following steps to show that if there exists an ε-secure encryption scheme (Enc, Dec) with respect to $\mathcal{C}$, where $\mathrm{Enc}\colon \{0,1\}^n \times \{0,1\}^m \to \{0,1\}^\ell$, then there exists a deterministic 2ε-extractor $\mathrm{Ext}\colon\{0,1\}^n \to \{0,1\}^{m-2\log(1/\varepsilon)-O(1)}$ for $\mathcal{C}$, provided $m \geq \log n + 2\log(1/\varepsilon) + O(1)$.

   (a) For each fixed key $k \in \{0,1\}^n$, define a source X_k on $\{0,1\}^\ell$ by $X_k = \mathrm{Enc}(k,U_m)$, and let $\mathcal{C}'$ be the class of all these sources (i.e., $\mathcal{C}' = \{X_k : k \in \{0,1\}^n\}$). Show that there exists a deterministic ε-extractor $\mathrm{Ext}'\colon \{0,1\}^\ell \to \{0,1\}^{m-2\log(1/\varepsilon)-O(1)}$ for $\mathcal{C}'$, provided $m \geq \log n + 2\log(1/\varepsilon) + O(1)$.

   (b) Show that if Ext' is a deterministic ε-extractor for $\mathcal{C}'$ and Enc is ε-secure with respect to $\mathcal{C}$, then $\mathrm{Ext}(k) = \mathrm{Ext}'(\mathrm{Enc}(k,0^m))$ is a deterministic 2ε-extractor for $\mathcal{C}$.

Thus, a class of sources can be used for secure encryption iff it is deterministically extractable.

Problem 6.11 (Extracting from Symbol-Fixing Sources*). A generalization of a bit-fixing source is a *symbol-fixing source* X taking values in Σ^n for some alphabet Σ, where subset of the coordinates of X are fixed and the rest are uniformly distributed and independent elements of Σ. For $\Sigma = \{0,1,2\}$ and $k \in [0,n]$, give an explicit ε-extractor

Ext : $\Sigma^n \to \{0,1\}^m$ for the class of symbol-fixing sources on Σ^n with min-entropy at least k, with $m = \Omega(k)$ and $\varepsilon = 2^{-\Omega(k)}$. (Hint: use a random walk on a consistently labelled 3-regular expander graph.)

6.5 Chapter Notes and References

Surveys on randomness extractors and their applications are given by Nisan and Ta-Shma [301] and Shaltiel [352, 354].

The study of randomness extraction goes back to von Neumann [416], who gave the deterministic extractor described in the text for IID-bit sources. (See [130, 307] for improvements in the extraction rate.) The study of extraction was renewed by Blum [69], who showed how to extend von Neumann's method to extract perfectly uniform bits from a source generated by a finite-state Markov chain. Santha and Vazirani [346] proposed the model of unpredictable-bit sources, suggested that an extractor need only produce an output that is statistically close to uniform (rather than perfectly uniform), and proved that there is no deterministic extractor for a single unpredictable-bit source even under this relaxation (Proposition 6.6; the proof we give in Problem 6.6 is from [334]). Vazirani and Vazirani [407] showed that nevertheless every problem in **BPP** can be solved in polynomial time given a single unpredictable-bit source (Theorem 6.29.) Chor and Goldreich [96] generalized this result to block sources, in the process introducing min-entropy to the literature on randomness extraction. Cohen and Wigderson [107] showed how to simulate **BPP** with any source of sufficiently high min-entropy rate. This was strengthened to any constant entropy rate by Zuckerman [426], and polynomially small entropy rate in [343, 28].

The notion of seeded extractor we focus on in this section (Definitions 6.13,6.16) was introduced by Nisan and Zuckerman [303], who also gave the first construction of extractors with polylogarithmic seed length (building on [426]). A form of the Leftover Hash Lemma (for flat distributions, and where the quality of the output distribution is measured with entropy rather than statistical difference) was first proven by Bennett, Brassard, and Robert [63]. The version in Theorem 6.18 is due to Impagliazzo, Levin, and Luby [197], and the proof we give is due to Rackoff [217]. The connection between the Leftover Hash Lemma and the Johnson Bound in Problem 3.4 is due to Ta-Shma and Zuckerman [383], with the equivalence between almost-universal hashing and minimum distance being from [67]. The extractor parameters of Problem 3.4 were first obtained in [155, 372].

The study of deterministic extractors has also remained active, motivated by applications in cryptography and other settings where enumerating seeds does not work. Bit-fixing sources, first studied in [63, 98, 409], received particular attention for their applications to maintaining security when an adversary learns some bits of a cryptographic secret key [90]. (There has been work on protecting against even more severe forms of leakage, surveyed in [175].) The deterministic extractor for symbol-fixing sources in Problem 6.11 is from [232]. Problem 6.10 (encryption requires deterministic extraction) is due to Bosley and Dodis [77]. For a discussion of deterministic extraction for other models of sources, see Section 8.2.

The relevance of bipartite expanders to simulating randomized algorithms with weak sources is due to Cohen and Wigderson [107], who studied dispersers and generalizations of them (as graphs rather than functions). The expander-based extractor of Theorem 6.22 is due to Goldreich and Wigderson [174]. The connection between extractors and list-decodable codes emerged in the work of Trevisan [389], and the list-decoding view of extractors (Proposition 6.23) was crystallized by Ta-Shma and Zuckerman [383]. The equivalence between extractors and averaging samplers (Corollary 6.24) is due to Zuckerman [427] (building on previous connections between other types of samplers and expanders/dispersers [365, 107]).

The notion of entropy was introduced by Shannon in his seminal paper [361] that gave rise to the field of information theory. The many properties of Shannon entropy and related quantities and their applications to communications theory are covered in the textbook [110]. Therefore, the fact that Shannon entropy converges to min-entropy (up to a vanishing statistical distance) when we take independent samples is called the Asymptotic Equipartition Property. Rényi entropy (indeed the entropies H_α for all $\alpha \in (0, \infty)$) were introduced in [335].

The fact that every k-source is a convex combination of flat k-sources (Lemma 6.10) is implicit in [96]. Our proof is due to Kalai [230].

The optimal nonconstructive bounds for extractor parameters (Theorem 6.14) were identified by Radhakrishnan and Ta-Shma [315], who proved a matching lower bound on seed length and upper bound on output length.

The method for extracting from block sources given by Lemmas 6.27 and 6.28 was developed over a series of works [96, 410, 426]; the form we use is from [303]. The first method described in Section 6.3.2 for converting general weak sources to block sources is from the original papers of Nisan and Zuckerman [426, 303] (see [402] for a tighter analysis). The observation that high min-entropy sources can be partitioned into block sources, and the benefits of this for constructing extractors, is due to Goldreich and Wigderson [174]. Lossless condensers were first introduced by Raz and Reingold [319] (formulated in a slightly different manner), and the general definition of condensers is from Reingold, Shaltiel, and Wigderson [328]. The equivalence of lossless condensing and vertex expansion close to the degree is from Ta-Shma, Umans, and Zuckerman [382]. The extractor construction of Section 6.3.4 is due to Guruswami, Umans, and Vadhan [192]. (Earlier papers had results similar to Theorem 6.36 with some restrictions on the parameters, such as $k = \Omega(n)$ [427] and $1/\varepsilon$ subpolynomial [272].) Recently, Dvir, Kopparty, Saraf, and Sudan [124] constructed extractors that have logarithmic seed length and extract a $1 - O(1)$ fraction of the min-entropy (for constant or slightly subconstant error).

The zig–zag product for extractors and condensers described in Section 6.3.5 is from [332, 92].

Problem 6.2 (achieving exponentially small error when simulating **BPP** with a weak source) is from [426].