The Raymond and Beverly Sackler
Faculty of Exact Sciences
The Blavatnik School of Computer Science

On The Entropy Loss and Gap of Condensers

Thesis submitted in partial fulfillment of the requirements for the M.Sc. degree in the

School of Computer Science, Tel-Aviv University

by

Nir Aviv

This work has been carried out under the supervision of
Prof. Amnon Ta-Shma

.

August 2017

Contents

1 Introduction **1**

2 Preliminaries **4**

3 Samplers as condensers **8**

3.1 Samplers 8

3.2 The equivalence between samplers and condensers 10

4 An upper bound on the seed length and entropy loss of condensers with a small entropy gap **13**

4.1 Previous work 13

4.2 The upper bound: statement and discussion 15

4.3 Proof of the upper bound 17

5 Reducing the seed length of explicit constructions **20**

1 Introduction

Extractors and condensers are ubiquitous in theoretical computer science. Extractors are probabilistic hash functions that hash any given source with k min-entropy to a close to uniform source. Formally, $E : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ is a (k, ε) extractor, if for any distribution X on $\{0,1\}^n$ with k min-entropy it holds that the distribution $E(X, U_t)$ is ε-close to uniform. Condensers are probabilistic hash functions that hash any given source with k min-entropy to a shorter source with a lot of min-entropy. Formally, $C : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ is a $k \to_\varepsilon k'$ condenser, if for any distribution X on $\{0,1\}^n$ with k min-entropy it holds that the distribution $C(X, U_t)$ is ε-close to a distribution with k' min-entropy.

Often, the quality of an extractor is measured by its *seed length* d that we wish to minimize, and its output length m that we wish to maximize. The *entropy loss* of the extractor is $k - m$, i.e., the difference between the entropy k of the input and the length m of the output. Non-explicitly, there are extractors with seed length $d = \log n + 2\log\frac{1}{\varepsilon} + O(1)$ and entropy loss $2\log\frac{1}{\varepsilon} + O(1)$. There is a matching lower bound showing every non-trivial extractor must have seed length d at least $\log n + 2\log\frac{1}{\varepsilon} - O(1)$, and also must have entropy loss at least $2\log\frac{1}{\varepsilon} - O(1)$ [8].

While extractors have an unavoidable entropy loss, the GUV condenser [5] is lossless, i.e., it has *zero* entropy loss (see also [12, 10, 11]). Specifically, for every $\alpha > 0$, [5] show a

lossless $C : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ that is a $k \to_\varepsilon k$ condenser, with $m = (1+\alpha)k$ and $d = (1+\frac{1}{\alpha})(\log(nk) + \log\frac{1}{\varepsilon})$. Thus, the GUV condenser seed length dependence on the error may be approaching $\log\frac{1}{\varepsilon}$, whereas extractors must have dependence $2\log\frac{1}{\varepsilon}$.

As expected, the improvement in the reduced entropy loss and the shorter seed length in the GUV construction depends on how dense the output distribution is in the codomain of the condenser. Let us define the *entropy gap* of a $k \to_\varepsilon k'$ condenser $C : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ to be $m - k'$, i.e., the difference between the amount of entropy in the uniform distribution over the output domain and the amount of entropy guaranteed in the output source. Extractors are condensers with entropy gap 0.

When the entropy gap is zero the entropy loss must be $2\log\frac{1}{\varepsilon}$ and the seed length is at least $\log n + 2\log\frac{1}{\varepsilon}$. The GUV construction shows that when the entropy gap is $\Omega(k)$ the condenser may be lossless, and the seed length dependence may approach $O(\log n) + \log\frac{1}{\varepsilon}$. How fast does this dependence drop with the entropy gap?

Dodis et. al. [3] observe that even when the entropy gap is just a constant, the entropy loss may be $O(\log\log\frac{1}{\varepsilon})$ and the seed length dependence on the error is just $\log\frac{1}{\varepsilon}$. This is quite surprising, and shows that with even a minimal entropy gap one can substantially reduce the limitations imposed by the lower bound on extractors. Dodis et. al. [3] use this sudden drop in entropy loss for obtaining key derivation protocols with smaller entropy loss. Recently, Ben-Aroya et. al. used the drop in the seed length dependence on the error for obtaining better explicit two source extractors, and this is also used in the best explicit two source constructions to date [2, 7].

In this paper we study the exact dependence of the seed length and entropy loss on the entropy gap. We obtain the following non-explicit upper bounds:

Theorem 1.1 (detailed in Theorem 4.3)**.** *Let $0 < \varepsilon < \frac{1}{2}$, and let c be a constant. Let $g \geqslant c$, and $n \geqslant k > \log\log\frac{n+\log(e)+1}{\varepsilon g} + 3$. Then for $l \leqslant k$ such that*

$$l = \log(1 + \frac{\log\frac{e}{\varepsilon}}{g}) + O(1)$$

and for

$$d = \lceil \log\frac{n-k+\log(e)+1}{\varepsilon g} + 1 \rceil,$$

there exists a $k \to_\varepsilon k' = k - l$ strong condenser $C : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^{k'+g}$. The constant in the $O(1)$ notation depends on c.

This bound extends (with slightly improved parameters) the results in [3], which give condensers where the entropy loss at least $\log\log\frac{1}{\varepsilon}$. We show, in particular, that already

when the gap is $O(\log \frac{1}{\varepsilon})$ then the entropy loss is reduced to a constant. The following theorem addresses the case of zero entropy loss:

Theorem 1.2 (detailed in Theorem 4.4)**.** *Let* $0 < \varepsilon < \frac{1}{2}$*,* $g \geqslant \frac{\log(\frac{e}{\varepsilon})}{\varepsilon}$*, and* $n \geqslant k > \log\log \frac{n+\log(e)+1}{\varepsilon^2 g} + 3$*. Then for*

$$d = \lceil \log \frac{n - k + \log(e) + 1}{\varepsilon^2 g} + 1 \rceil$$

there exists a $k \to_{O(\varepsilon)} k$ *strong condenser* $C : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$.

This bound shows that to get to zero loss it suffices to have $O(\frac{\log \frac{1}{\varepsilon}}{\varepsilon})$ entropy gap. We note that in comparison with Theorem 1.1, the dependence of the seed length on the error in Theorem 1.2 is larger: here we have ε^2 rather than ε in the denominator. Since we use $g = \Omega(\frac{\log(1/\varepsilon)}{\varepsilon})$ we have that the degree of the condenser, $D = 2^d$, may depend on $\frac{1}{\varepsilon \log(1/\varepsilon)}$ whereas in Theorem 1.1 the dependence is at most $\frac{1}{\varepsilon}$.

We have not proved matching lower bounds, and we leave that for future work.

In the analysis, we use and prove a tight equivalence between condensers with small entropy gap and samplers with multiplicative error, see Corollary 3.4. A similar, but non-tight, equivalence was already proved by Dodis et al. [3], using somewhat different objects: they consider a weaker variant of samplers called (γ, δ) unpredictability extractors, which are only promised to work for sets of at most a given density γ, hitting them with probability at most δ. Their reduction is not tight and has a small gap in parameters. We show that by slightly changing the statement to consider samplers that work for all sets, one can get an equivalence that is completely tight. We remark that this equivalence underlies the work of Ben-Aroya et al. [1] and later work on explicit two source extractors, and we believe it is interesting in its own right.

Finally, we partially address an open question from [3]. Dodis et al. raise the problem of finding an *explicit* condenser with a small seed, entropy gap and loss. They show an explicit construction using seed length $O(n \log \frac{1}{\varepsilon})$, which they reduce to $O(k \log k)$. Here we show how to reduce this to $O(\log \frac{n}{\varepsilon} \cdot \log \frac{k}{\varepsilon})$, using a simple composition. The idea is to first use a lossy extractor, and extract *all* the entropy of the source, except for the unavoidable entropy loss $2 \log \frac{1}{\varepsilon} + O(1)$. The point is that this entropy is not lost, but rather still present in the source (even conditioned on the output so far). We now first apply the lossless condenser of GUV, to bring the length down to $O(\log \frac{n}{\varepsilon})$, and on this output we apply the explicit small loss condenser found by Dodis et al. We give full details of this in Chapter 5.

2

Preliminaries

Throughout this work, the notation log represents the base-2 logarithm and ln the base-e logarithm. Capital letters such as N, K, M, G, D are base-2 exponents of the respective lowercase n, k, m, g, d. For example, $d = \log(D)$. The notation $[N]$ is equivalent to $\{1, 2, \ldots, N\}$ for any integer N.

The *min-entropy* of a distribution X is $H_\infty(X) = -\log(\max_{x \in S} \Pr[X = x])$. We say that X is an (n, k) source if X is distributed over $\{0,1\}^n$ and $H_\infty(X) \geqslant k$. U_d denotes the uniform distribution on $\{0,1\}^d$.

We use the *variational* (or *statistical*) distance between distributions:

Definition 2.1 (variational distance)**.** *For every two random variables X and Y defined on a common finite set S,*

$$\|X - Y\| = \frac{1}{2} \sum_{x \in S} |\Pr[X = s] - \Pr[Y = s]| \;=\; \max_{A \subseteq S} \{\Pr[X \subseteq A] - \Pr[Y \subseteq A]\}.$$

The following lemma states the simple fact that the min-entropy of a distribution can be increased by moving weight from the set of heavy elements:

Lemma 2.1. *Let $n \geqslant k \geqslant 0$, and $0 \leqslant \varepsilon \leqslant n - k$. Let X be a (n, k)-source. Then X is ε-close (in variational distance) to a $(n, k + \varepsilon)$-source X'.*

Proof. Every (n,k) source is a convex combination of flat (n,k) sources, i.e., distributions that are uniform over $K = 2^k$ elements. It therefore suffices to prove the claim for flat sources: if X is a convex combination of $X_1,\ldots,X_t$ and $X_1,\ldots,X_t$ are all ε-close to X', then X is ε-close to X'. Assume therefore that X is a flat (n,k) source. Let X' be a random variable obtained by "distributing the extra weight", i.e., taking $2^{-k} - 2^{-(k+\varepsilon)}$ weight from every element in the support of X, and distributing it evenly on the remaining elements. Hence, $\|X - X'\| \leqslant 2^k(2^{-k} - 2^{-(k+\varepsilon)}) = 1 - 2^{-\varepsilon}$. Note that $1 - 2^{-\varepsilon} \leqslant \varepsilon$, since both sides are equal for $\varepsilon = 0$, and the derivative of ε is larger for $\varepsilon > 0$. In summary, we have that X is ε-close to the $(n, k+\varepsilon)$-source X'. □

We use several inequalities regarding the probability that a sample from a given distribution is far from the mean of that distribution. These involve exponents of the *Kullback-Leibler divergence* (or *KL-divergence*) for biased coin flips, as defined below:

Definition 2.2 (KL-divergence)**.** *Let* $0 < p < 1$ *and* $0 < q < 1$*. Then*

$$D(p||q) = p\ln(\frac{p}{q}) + (1-p)\ln(\frac{1-p}{1-q}).$$

The well-known Chernoff-Hoeffding inequality bounds from above the probability that a sample from a binomial distribution is far from the mean:

Theorem 2.2 (Chernoff-Hoeffding inequality)**.** *Let* $X_1,\ldots,X_n$ *be independent random variables taking values in* $[0,1]$ *such that each variable has expectation* q*. Then, for every* $q < p < 1$*,*

$$\Pr[\sum_{i=1}^{n} X_i \geqslant pn] \leqslant e^{-D(p||q)n}.$$

.

Proof. The following theorem is proved in [6]: for every $0 < t < 1 - q$, it holds that $Pr[\frac{1}{n}X_i - t \geqslant q] \leqslant [(\frac{q}{q+t})^{q+t}(\frac{1-q}{1-q-t})^{1-q-t}]^n = e^{-D(q+t||q)n}$. Taking $t = p - q$, we obtain the required form of the inequality. □

We bound the KL-divergence from below to simplify this inequality. First, we note

Claim 2.1. *For any* $0 < q < p < 1$*, it holds that*

$$D(p||q) \geqslant p\ln(\frac{p}{q}) + q - p.$$

Proof. By definition, $D(p||q) = p\ln(\frac{p}{q}) + (1-p)\ln(\frac{1-p}{1-q})$. Since $\ln(x) \geqslant 1 - \frac{1}{x}$ for $x > 0$, it holds that $(1-p)\ln(\frac{1-p}{1-q}) \geqslant (1-p)(1 - \frac{1-q}{1-p}) = q - p$, proving the claim. □

We use the previous claim to further bound the KL-divergence from below, for the case $p = Aq + b$.

Lemma 2.3. *Let $0 \leqslant q \leqslant 1$. Let $A > 1$ and $b > 0$ such that $Aq + b \leqslant 1$. Then*

$$D(Aq + b||q) \geqslant b \ln(A).$$

Proof. By Claim 2.1, $D(Aq + b||q) \geqslant (Aq + b) \ln(A + \frac{b}{q}) - Aq - b + q$. Denote $t = \frac{A}{b}q$. It follows that

$$\begin{aligned}
&(Aq+b) \ln(A + \frac{b}{q}) - Aq - b + q \\
&\quad = b(1 + t) \ln((1 + \frac{1}{t})A) - (1 + t)b + \frac{bt}{A} \\
&\quad = (1 + t)(b \ln(A) + b \ln(1 + \frac{1}{t})) - (1 + t)b + \frac{bt}{A} \\
&\quad = b \ln(A) + b \ln(1 + \frac{1}{t}) + tb \ln(A) + tb \ln(1 + \frac{1}{t}) - (1 + t)b + \frac{bt}{A}.
\end{aligned}$$

To show $D(Aq + b||q) \geqslant b \ln(A)$ it therefore suffices to show that

$$1 + t \leqslant (1 + t) \ln(1 + \frac{1}{t}) + t(\ln(A) + \frac{1}{A}).$$

Dividing by t and setting $s = 1 + \frac{1}{t}$, the previous inequality is equivalent to

$$s \leqslant s \ln(s) + \ln(A) + \frac{1}{A}.$$

It holds for all $A > 1$ that $\ln(A) + \frac{1}{A} \geqslant 1$ and so it is enough to show $s \leqslant 1 + s \ln(s)$, which is true for all $s > 1$ in the same way. □

Finally, we use the entropy function defined to base e,

Definition 2.3. *Let $0 \leqslant p \leqslant 1$. Then*

$$H(p) = -p \ln(p) - (1 - p) \ln(1 - p).$$

We use the following fact:

Claim 2.2. *For every $0 < p < 1$, $H(p) \leqslant p \ln(\frac{e}{p})$.*

Proof. Since $\ln(x) \leqslant x - 1$ for every $x > 0$, we have that

$$H(p) \leqslant p \ln(\frac{1}{p}) + (1-p)(\frac{1}{1-p} - 1) \leqslant p \ln(\frac{1}{p}) + p,$$

as required. □

3

Samplers as condensers

Zuckerman [14] (see also [4]) showed an equivalence between extractors and samplers with *additive error*. In this chapter, following [3] we extend this to an equivalence between samplers with *multiplicative* factor and condensers with a small *entropy gap*.

3.1 Samplers

A sampler is a randomized algorithm that tosses n random coins, and based on these coins outputs a subset of $D = 2^d$ elements from $\{0,1\}^m$ as its sample set. We represent such a sampler as

$$S : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$$

where $S(x,i)$ is the i'th element in the sample set obtained when using the random coins x. We call n the *randomness complexity* of S, and D and the sample size of S.

Now fix a set $T \subseteq \{0,1\}^m$ and denote its density by

$$\rho(T) = \frac{|T|}{M}.$$

The sampler's estimate to the density of T, given the random coins $x \in \{0,1\}^n$, is

$$\Pr_{y \in [D]}[S(x,y) \in T].$$

Often, the sampler's quality is measured as the difference between these two values. I.e., a sampler is said to have at most error ε on T, if (w.h.p. over x) the sampler's estimate is within ε additive distance from the correct density. In such a scenario we do not allow large overestimates nor large underestimates of the density of the subset.

In this paper we are interested in estimating *very small* densities, and we only want to avoid (w.h.p. over x) grossly exaggerated density estimates. Formally, we say (S,x) has at most G *multiplicative* error and ε *additive* error for T, if

$$\Pr_{y \in [D]}[S(x,y) \in T] \leqslant G \cdot \rho(T) + \varepsilon.$$

The quality of the sampler is then measured on its worst performance for T:

Definition 3.1 (Sampler error)**.** *Let $S : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ be a sampler. We say S is a (k, G, ε) sampler if for every (n,k) source X and every $T \subseteq \{0,1\}^m$,*

$$\Pr_{x \in X, y \in [D]}[S(x,y) \in T] \quad \leqslant \quad G \cdot \rho(T) + \varepsilon.$$

We call G the multiplicative *error of S and ε the* additive *error of S.*

Notice that the definition implies that there are at most $K = 2^k$ values $x \in \{0,1\}^n$ for which S is wrong and $\Pr_{y \in [D]}[S(x,y) \in T] > G \cdot \rho(T) + \varepsilon$. We remark that Dodis et. al. [3] defined an almost equivalent object and called it an *unpredictability extractor*. We repeat their definition here:

Definition 3.2 (Strong unpredictability extractor)**.** *A function $S : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ is a (k, δ, δ') strong unpredictability extractor, if for every (n,k) source X and every $T \subseteq \{0,1\}^{d+m}$ of size at most $\delta 2^{d+m}$, it holds that $\Pr_{x \in X, s \in U_d}[(s, S(x,s)) \in T] \leqslant \delta'$.*

Note that if S is a (k, δ, δ') strong unpredictability extractor for *all* $\delta > 0$, then the function $S' : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^d \times \{0,1\}^m$ defined by $S'(x,y) = (y, S(x,y))$ is a sampler with total error δ'. We call such functions *strong samplers:*

Definition 3.3 (Strong sampler)**.** *A function $S : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ is a (k, G, ε)* strong sampler *if it is a $(k, \delta, G\delta + \varepsilon)$ strong unpredictability extractor for every $\delta > 0$.*

It is useful to look at S' rather than S as we consider *strong* condensers throughout the rest of this work (see next section). From now on we shall generally use the definitions of strong unpredictability extractors and strong samplers rather than usual definition of samplers.

3.2 The equivalence between samplers and condensers

Definition 3.4. *(Condenser): A function*

$$C : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$$

is a $k \to_\varepsilon k'$ condenser if for all (n,k)-sources X, and a uniformly random and independent seed S over $\{0,1\}^d$, the distribution $Cond(X,S)$ is ε-statistically-close to some distribution W' with min-entropy k'.

We also define:

- *The* entropy loss *of C is $d + k - k'$,*
- *The* entropy gap *is $m - k'$,*
- *$C : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ is a strong $k \to_\varepsilon k'$ condenser, if the function $C'(X,Y) = (Y, C(X,Y))$ is a $k \to_\varepsilon k' + d$ condenser.*

The entropy loss measures how much entropy we lose in the process of condensing. In contrast, the entropy gap measures how close the output is to uniform. If the entropy gap is 0 we are close to true uniform bits. If the entropy gap is 1, we are missing one bit of entropy to become truly uniform.

Zuckerman [14] showed an equivalence between samplers with additive error and extractors. In the following subsections, following [3] we show an equivalence between samplers with multiplicative error and condensers.

3.2.1 Condensers imply samplers

Dodis et. al. proved the following in [3]:

Lemma 3.1 ([3, Lemma 3.3])**.** *Let $n \geqslant k > k' > 0$, $d > 0$ and $m > k'$. Let $C : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ be a strong $k \to_\varepsilon k'$ condenser. Denote $g = m - k'$ and $G = 2^g$. Then C is a $(k, \delta, \delta' = G\delta + \varepsilon)$ strong sampler.*

In other words, strong condensers with entropy gap g and error ε are also strong samplers with multiplicative error $G = 2^g$ and additive error ε. We restate the proof of [3], put slightly differently:

Proof. (of Lemma 3.1). Assume by way of contradiction that there exists some $\delta > 0$ such that C is not a $(k, \delta, G\delta + \varepsilon)$ strong unpredictability extractor. Consider the distribution $W = (s, C(x, s))$ where x is sampled from X and s is sampled from U_s. By our assumption, there is some set $T \subseteq \{0,1\}^d \times \{0,1\}^m$ such that $\rho(T) \leqslant \delta$ and

$$\Pr[W \in T] > G\delta + \varepsilon \geqslant G \cdot \frac{|T|}{DM} + \varepsilon = 2^{-(d+m-g)}|T| + \varepsilon = 2^{-(k'+d)}|T| + \varepsilon.$$

Therefore, even if we move ε weight from elements in T to elements outside of T, there will be some element $t \in T$ such that has a strictly larger probability than $2^{k'+d}$. In other words, for every distribution $W' \approx_\varepsilon W$, we have that $H_\infty(W') < k' + d$, and therefore C is not a $k \to_\varepsilon k'$ strong condenser. □

A natural question is whether this implication also holds in the reverse direction, that is, if samplers with multiplicative error are also condensers. We discuss this in the following subsection.

3.2.2 Samplers imply condensers

We begin with a result of Dodis et al. [3] that shows how one can construct strong condensers from strong unpredictability extractors:

Lemma 3.2 ([3, Section 3, Summary])**.** *Let $n \geqslant k > 0$, $d > 0$, $m > k$ and $\delta' > \delta > 0$. Let $S : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ be a (k, δ, δ') strong unpredictablity extractor. Let $g = \log \frac{\delta'}{\delta}$ and $\varepsilon = \delta'$. Then, given S, one can efficiently construct a function $C : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ that is a $k \to_\varepsilon m - g$ strong condenser.*

Dodis et al. give in [3] an indirect proof for Lemma 3.2, going through a definition called *balanced hash functions*. Note that in the terms of Lemma 3.2, $\delta' = \frac{G\delta+\varepsilon}{2}$, whereas in Lemma 3.1 we have $\delta' = G\delta + \varepsilon$, and so there is a small gap in parameters between Lemma 3.2 and Lemma 3.1.

We show a direct reduction from strong samplers to strong condensers. Along with Lemma 3.1, this reduction shows that strong condensers with entropy gap g and error ε are precisely equivalent to strong samplers with multiplicative error $G = 2^g$ and additive error ε.

Note that here we have a requirement that Lemma 3.2 does not have: that we have a single function S that is a (k, δ, δ') unpredictability extractor for *all* $\delta > 0$, where $\delta' = G\delta + \varepsilon$.

Lemma 3.3. *Let $n \geqslant k > 0$, $d > 0$, $m > k$, $\varepsilon > 0$ and $g > 0$. Denote $G = 2^g$. Let $S : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ be a (k, G, ε) strong sampler. Then, S is a $k \to_\varepsilon k' = m - g$ strong condenser.*

Proof. Let X be a (n, k)-source. Consider the distribution $W = (s, S(x, s))$ where x is sampled from X and s is sampled from U_d. Let $H = \left\{ h \in \{0,1\}^d \times \{0,1\}^m \,|\Pr[W = h] > 2^{k'+d} \right\}$. This is the set of "heavy" elements: if there is at most $2^{-(k'+d)}|H| + \varepsilon$ weight on H, then by moving ε weight to elements outside of H, we can obtain a distribution W' which has $k' + d$ min-entropy, as required. Indeed, S is a $(k, \delta, G\delta + \varepsilon)$ strong unpredictability extractor for all $\delta > 0$, and in particular for $\delta = \rho(H)$, and therefore

$$\Pr[W \in H] \leqslant G \cdot \frac{|H|}{DM} + \varepsilon = 2^{-(d+m-g)}|H| + \varepsilon = 2^{-(k'+d)}|H| + \varepsilon,$$

as required. □

The following table summarizes the error ε and entropy gap g of the condensers that are implied by the existence of a (k, G', ε') strong sampler in Lemmas 3.2 and 3.3. Note that in Lemma 3.2, we first choose some $\delta > 0$, and then efficiently construct a condenser with the given parameters from the strong sampler.

	Error	Entropy gap
Lemma 3.3	$\varepsilon = \varepsilon'$	$g = g' = \log G'$
Lemma 3.2 [3]	$\varepsilon = G'\delta + \varepsilon'$	$g = \log(G' + \frac{\varepsilon'}{\delta})$

Note that in Lemma 3.3 we obtain error ε' and gap g' simultaneously. In Lemma 3.2 one can obtain error ε' if δ tends to 0, and gap g' if δ is close to 1.

To summarize the equivalence between strong condensers and strong samplers, the results of Lemma 3.3 proved in this subsection along with the results of Lemma 3.1 proved in [3] give rise to the following corollary:

Corollary 3.4. *Let $n \geqslant k > 0$, $d > 0$, $m > k$, $\varepsilon > 0$ and $g > 0$. Denote $G = 2^g$. Then a function $S : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ is a (k, G, ε) strong sampler if and only if S is a $k \to_\varepsilon k' = m - g$ strong condenser.*

4

An upper bound on the seed length and entropy loss of condensers with a small entropy gap

The results of the previous chapter raise the question whether samplers that allow multiplicative error have substantially better parameters than those only allowing additive error. As mentioned in the introduction, the answer is yes, and this was already observed in [3]. We first revisit the proof of [3] in the following section.

4.1 Previous work

In [3], Dodis et al. gave a probabilistic proof for the existence of (k, δ, δ') unpredictablity extractors:

Theorem 4.1 ([3, Theorem 4.15])**.** *There exists a function $S : \{0,1\}^m \ times \{0,1\}^d \to \{0,1\}^m$ which is a (k, δ, δ') unpredictability extractor if one of the following conditions hold:*

$$\delta' > \max\left\{2e\delta, (n-k-2)2^{-d} + \log(e/\delta)\delta 2^{m-k}\right\} \tag{4.1}$$

$$2e\delta \geqslant \delta' \geqslant \delta + 2\delta\sqrt{(1/\delta)(n-k+2)2^{-d} + \log(e/\delta)2^{m-k}}. \tag{4.2}$$

We are interested in condensers, and so recall that Dodis et al. also gave in [3] a way to construct condensers from unpredictability extractors (see also Section 3.2.2, Lemma 3.2). Applying this to Theorem 4.1 gives an existence proof for condensers whose parameters satisfy equations that correspond to Equations 4.1 and 4.2. This is detailed in the following claim:

Claim 4.1. *Let $C : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ be a $k \to_\varepsilon m - g$ condenser obtained by applying the reduction given in Lemma 3.2 to an unpredictability extractor S given by Theorem 4.1. Denote by $l = k - m + g$ the entropy loss of this condenser. Denote $D = 2^d$ and $L = 2^l$. Then the parameters of C satisfy*

$$1 > \frac{n-k+2}{\varepsilon D} + \frac{g + \log\frac{e}{\varepsilon}}{L}.$$

Proof. Recall that by the reduction of Lemma 3.2, the existence of S implies the existence of a $k \to_\varepsilon m - g$ condenser where $\varepsilon = \delta'$ and $g = \log\frac{\delta'}{\delta}$. By definition, the entropy loss of this condenser is $l = k - m + g$. We consider the case where Equation 4.1 holds and the case where Equation 4.2 hold separately.

Large gap: Assume that Equation 4.1 holds. First, we have $\frac{\delta'}{\delta} > 2e$. Since $g = \log\frac{\delta'}{\delta}$, we have that $g > \log(e) + 1$. We divide the remaining inequality by δ', substitute $\delta' = \varepsilon$, $\delta = \frac{\varepsilon}{G}$, and $m - k = g - l$, and so we have that if Equation 4.1 holds then:

$$1 > \frac{n-k+2}{\varepsilon D} + \frac{g + \log\frac{e}{\varepsilon}}{L}.$$

Small gap: Assume that Equation 4.2 holds. We have $\frac{\delta'}{\delta} \leqslant 2e$, and so $g < \log(e) + 1$. We divide the remaining inequality by δ and, as in the previous case, we substitute $\delta' = \varepsilon$, $\delta = \frac{\varepsilon}{G}$, and $m - k = g - l$. We obtain that if Equation 4.2 holds then:

$$G > 1 + 2\sqrt{G}\sqrt{\frac{n-k+2}{\varepsilon D} + \frac{g + \log\frac{e}{\varepsilon}}{L}}.$$

In other words,

$$\frac{(G-1)^2}{4G} > \frac{n-k+2}{\varepsilon D} + \frac{g + \log\frac{e}{\varepsilon}}{L}.$$

Since $G \leqslant 2e$, we have that $\frac{(G-1)^2}{4G} \leqslant \frac{(2e-1)^2}{8e} < 1$.

□

Note that Claim 4.1 implies that the condensers given in [3] have seed length $d \geqslant \log \frac{n-k+2}{\varepsilon}$ and entropy loss $l \geqslant \log(g + \log \frac{e}{\varepsilon})$. We know by e.g. [5] that there exist *lossless* condensers with $l = 0$. In the next sections, we show a probabilistic argument for the existence of condensers that extends the results of [3] to constant entropy loss, with slightly improved seed length and entropy gap. We also show the existence of lossless condensers, at the cost of larger seed length.

4.2 The upper bound: statement and discussion

In this section we generally consider the entropy loss and seed length as a function of the entropy gap and error.

We first state a generalized theorem that has a parameter ζ in addition to the usual parameters of randomness condensers. This parameter captures a trade-off between entropy loss and seed length.

Theorem 4.2 (Generalized upper bound)**.** *Fix $0 < \zeta < 1$, $0 < \varepsilon < \frac{1}{2}$, $g > 0$, and $n \geqslant k \geqslant \log\log \frac{n+\log(e)+1}{\zeta\varepsilon g} + 2$. Let $l \leqslant k$ be such that*

$$l \geqslant \log(1 + \frac{\log \frac{e}{\varepsilon}}{g}) + \log \frac{1}{1 - \frac{1-(1/G)}{\ln G}} + \log \frac{1}{1-\zeta}$$

and let d be such that

$$d \geqslant \log \frac{n - k + \log(e) + 1}{\zeta\varepsilon g}.$$

Then, there exists a $k \to_\varepsilon k' = k - l$ strong condenser $C : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^{k'+g}$.

We prove Theorem 4.2 in the next section. The various parameters in Theorem 4.2 can be set in different ways to achieve specific objectives. We discuss two such instantiations. In the first, we simply set $\zeta = \frac{1}{2}$.

Theorem 4.3 (Upper bound for condensers with at least one bit of entropy loss)**.** *Fix $0 < \varepsilon < \frac{1}{2}$, $g > 0$, and $n \geqslant k \geqslant \log\log \frac{n+\log(e)+1}{\varepsilon g} + 3$. Let $l \leqslant k$ be such that*

$$l \geqslant \log(1 + \frac{\log \frac{e}{\varepsilon}}{g}) + \log \frac{1}{1 - \frac{1-(1/G)}{\ln G}} + 1$$

and let d be such that

$$d \geqslant \log \frac{n - k + \log(e) + 1}{\varepsilon g} + 1.$$

Then, there exists a $k \to_\varepsilon k' = k - l$ *strong condenser* $C : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^{k'+g}$.

We note the fact that with only one bit of entropy gap we achieve $\log\log\frac{1}{\varepsilon}$ entropy loss, as in the results of [3] which previously showed this is possible. This is remarkable in contrast with the unavoidable $2\log\frac{1}{\varepsilon} - O(1)$ entropy loss of extractors (which are condensers with zero entropy gap).

Further comparing Theorem 4.3 to the parameters of condensers given by [3] (see Claim 4.1), we see the following:

- For entropy gap $g > 1$, the parameters in Theorem 4.3 slightly improve over the parameters shown in Claim 4.1: Theorem 4.3 has $d = \frac{n-k+\log(e)+1}{\varepsilon g} + 1$ instead of $d \geqslant \frac{n-k+2}{\varepsilon}$ and $l = \log(1 + \frac{\log\frac{e}{\varepsilon}}{g}) + O(1)$ rather than $l \geqslant \log(g + \log\frac{e}{\varepsilon})$.
- There is no restriction that $l \geqslant \log\log\frac{1}{\varepsilon}$. Indeed, we see that with $O(\log\frac{1}{\varepsilon})$ entropy gap we achieve constant entropy loss.

The entropy loss in Theorem 4.3 does not drop below 1, since we have set $\zeta = \frac{1}{2}$. The second instantiation we discuss shows the existence of lossless condensers. Here we set ζ such that $\log(\frac{1}{1-\zeta})$ is close to ε. We also require that $g \geqslant \frac{\log\frac{1}{\varepsilon}}{\varepsilon}$. This results in the following theorem:

Theorem 4.4 (Upper bound for lossless condensers)**.** *Fix* $0 < \varepsilon < \frac{1}{2}$, $g \geqslant \frac{\log(\frac{e}{\varepsilon})}{\varepsilon}$, *and* $n \geqslant k \geqslant \log\log\frac{n+\log(e)+1}{\varepsilon^2 g} + 3$. *Let*

$$d \geqslant \log\frac{n-k+\log(e)+1}{\varepsilon^2 g} + 1.$$

Then, there exists a $k \to_{O(\varepsilon)} k$ *strong condenser* $C : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^{k+g}$.

Proof. First note that, by the definition of g, it holds that $\log(1 + \frac{\log(\frac{e}{\varepsilon})}{g}) \leqslant \log(1+\varepsilon) < 2\varepsilon$.

Set $\zeta = \frac{\varepsilon}{2}$. We show that $\log\frac{1}{1-\zeta} \leqslant \varepsilon$, or equivalently $\frac{\varepsilon}{2} \leqslant 1 - \frac{1}{2^\varepsilon}$. This holds since both sides are equal for $\varepsilon = 0$ and the derivative of $\frac{\varepsilon}{2}$ is smaller for $0 < \varepsilon < \frac{1}{2}$.

We now show that $\log\frac{1}{1-\frac{1-1/G}{\ln G}} \leqslant \varepsilon$. Since $\ln G > 1$, it suffices to show that $\log\frac{1}{1-\frac{1}{\ln G}} = \log(1+\ln\frac{1}{\varepsilon}) - \log(1+\ln\frac{1}{\varepsilon} - \varepsilon) \leqslant \varepsilon$. By the mean value theorem, for every pair of positive real numbers $x > t$, it holds that $\frac{\log(x)-\log(x-t)}{t} \leqslant \frac{\log(e)}{x-t}$. In particular,

$$\log(1+\ln\frac{1}{\varepsilon}) - \log(1+\ln\frac{1}{\varepsilon} - \varepsilon) \leqslant \frac{\log(e)}{1+\ln\frac{1}{\varepsilon} - \varepsilon} \cdot \varepsilon.$$

It therefore suffices to show $\ln\frac{1}{\varepsilon} \geqslant \log(e)$, which holds since $\varepsilon < \frac{1}{2}$.

By Theorem 4.2 and what we've shown so far, there exists a $k \to_\varepsilon k' = k - O(\varepsilon)$ strong condenser $C : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^{k'+g}$. This implies that C is also a randomness condenser with error $O(\varepsilon)$ and zero entropy loss (and $g - O(\varepsilon)$ entropy gap), as detailed in Lemma 2.1. □

It is interesting to note that in comparison with Theorem 4.3, the dependence of the seed length on the error in Theorem 4.4 is larger: here we have ε^2 rather than ε in the denominator. Since we use $g = \Omega(\frac{\log(1/\varepsilon)}{\varepsilon})$, we have that $D = 2^d$ may depend on $\frac{1}{\varepsilon \log(1/\varepsilon)}$ whereas in Theorem 4.3, if $g = \Omega(1)$, the dependence is at most $\frac{1}{\varepsilon}$.

4.3 Proof of the upper bound

It remains to prove the generalized upper bound of Theorem 4.2. We give a proof using the probabilistic method.

Proof. Let $C : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ be chosen uniformly at random from the set of such functions. We show that, with positive probability, C is a $k \to_\varepsilon m - g$ strong condenser, and therefore such a condenser must exist. We use the following definition:

Definition 4.1. *Let $A \subseteq \{0,1\}^n$ such that $|A| = K$, and let $B \subseteq \{0,1\}^{d+m}$ be any subset. We say the pair (A, B) is C-*bad *if $\Pr_{x \in A, s \in U_d}[(s, C(x,s)) \in B] > G\rho(B) + \varepsilon$.*

By Corollary 3.4, and since every (n,k) source is a convex combination of flat (n,k) sources, i.e., distributions that are uniform over some set $A \subseteq \{0,1\}^n$ of size $|A| = K$, we have that C is a $k \to_\varepsilon m - g$ strong condenser if and only if no C-bad pairs (A, B) exist. The following claim bounds the chance of a given pair to be C-bad.

Claim 4.2. *Let $A \subseteq \{0,1\}^n$ such that $|A| = K$, and let $B \subseteq \{0,1\}^{d+m}$ with $\rho(B) = \delta$. The probability (over the choice of C) that the pair (A, B) is C-bad is at most $e^{-D(G\delta+\varepsilon||\delta)KD}$.*

Proof. For every $a \in A$ and $s \in \{0,1\}^d$, let $X_{a,s}$ denote the random variable that is 1 if $(s, C(a,s)) \in B$ and zero otherwise. Then, the boolean random variables $\{X_{a,s}\}_{a \in A, s \in \{0,1\}^d}$ are i.i.d. with expectation δ. By Theorem 2.2, $\Pr[\sum_{a,d} X_{a,d} \geqslant (G\delta+\varepsilon)KD] \leqslant e^{-D(G\delta+\varepsilon||\delta)KD}$. □

Next we use the union bound to get an upper bound on the probability that a randomly chosen C is not a condenser. Note that a pair (A, B) cannot be C-bad if $\rho(B) \geqslant \frac{1-\varepsilon}{G}$.

Therefore, by Claim 4.2 and the union bound,

$$\Pr_C[\text{C is not a } k \to_\varepsilon m-g \text{ strong condenser}] \leqslant \sum_{i=0}^{\frac{DM}{G}} \binom{N}{K}\binom{DM}{i} e^{-D(G\cdot\delta(i)+\varepsilon||\delta(i))KD},$$

where $\delta(i) = \frac{i}{DM}$. Splitting the term $e^{-D(G\cdot\delta(i)+\varepsilon||\delta(i))KD}$ into two terms: $e^{-\zeta D(G\cdot\delta(i)+\varepsilon||\delta(i))KD}$ and $e^{-(1-\zeta)D(G\cdot\delta(i)+\varepsilon||\delta(i))KD}$, it suffices to prove for every fixed $\delta \in (0, \frac{1}{G}]$ that

$$\binom{N}{K} e^{-\zeta D(G\delta+\varepsilon||\delta)KD} < \frac{G}{DM}, \tag{4.3}$$

and also that

$$\binom{DM}{\delta DM} e^{-(1-\zeta)D(G\delta+\varepsilon||\delta)KD} < 1 \tag{4.4}$$

since then the above probability is bounded from above by $\frac{DM}{G}\cdot\frac{G}{DM} = 1$.

To prove Eq. (4.3), using $\binom{N}{K} \leqslant (\frac{eN}{K})^K = 2^{K(n-k+\log(e))}$, it suffices to show that

$$\zeta\log(e)D(G\delta+\varepsilon||\delta)D - (n-k+\log(e)) \geqslant \frac{d+m-g}{K}.$$

We have that $\frac{d+m-g}{K} = \frac{d+k'}{K} \leqslant \frac{d+k}{K} \leqslant \frac{d}{K} + \frac{\log(e)}{e}$. By the statement of the theorem we have that $\frac{d}{K} \leqslant \frac{1}{4}$, and so $\frac{d+m-g}{K} < 1$, and it suffices to show

$$D \geqslant \frac{n-k+\log(e)+1}{\zeta\log(e)D(G\delta+\varepsilon||\delta)}.$$

By Lemma 2.3, we have that $D(G\delta+\varepsilon||\delta) \geqslant \varepsilon\ln(G)$, and the required equation then follows from the statement of the theorem.

It remains to prove Eq. (4.4). Using the fact that $\binom{DM}{\delta DM} \leqslant e^{H(\delta)DM}$, it suffices to show that $e^{H(\delta)DM-(1-\zeta)D(G\delta+\varepsilon||\delta)KD} < 1$, or equivalently that $H(\delta) < (1-\zeta)D(G\delta+\varepsilon||\delta)\frac{K}{M}$. Since $\frac{K}{M} = \frac{L}{G}$, we need to show

$$L \geqslant \frac{H(\delta)G}{(1-\zeta)D(G\delta+\varepsilon||\delta)}.$$

We divide into cases:

Small δ: assume that $\delta \leqslant \frac{\varepsilon}{G}$. Since $\varepsilon < \frac{1}{2}$ and $H(\delta)$ is monotone increasing for $\delta \in (0, \frac{1}{2})$, we have that $H(\delta) \leqslant H(\frac{\varepsilon}{G})$. By Claim 2.2, $H(\frac{\varepsilon}{G}) \leqslant \frac{\varepsilon}{G}\ln(\frac{Ge}{\varepsilon})$. By Lemma 2.3, $D(G\delta+\varepsilon||\delta) \geqslant$

$\varepsilon \ln(G)$. It therefore suffices to prove

$$L \geqslant \frac{\ln \frac{eG}{\varepsilon}}{(1-\zeta)\ln G}$$

which is implied by the statement of the theorem.

Large δ: assume that $\frac{\varepsilon}{G} < \delta$. By Claim 2.2, $H(\delta) \leqslant \delta \ln(\frac{e}{\delta}) \leqslant \delta \ln(\frac{eG}{\varepsilon})$. By Claim 2.1, $D(G\delta + \ \varepsilon||\delta) \geqslant (G\delta+\varepsilon)(\ln(G+\frac{\varepsilon}{\delta}) + \frac{1}{G+\frac{\varepsilon}{\delta}} - 1)$. Using that $\varepsilon > 0$ and that $\ln(x) + \frac{1}{x}$ is monotone increasing for all $x > 1$, we have that $D(G\delta + \ \varepsilon||\delta) \geqslant G\delta(\ln G + \frac{1}{G} - 1)$. It therefore suffices to prove

$$L > \frac{\ln \frac{eG}{\varepsilon}}{(1-\zeta)(\ln G + \frac{1}{G} - 1)} = \frac{\ln \frac{eG}{\varepsilon}}{(1-\zeta)\ln G} \cdot \frac{1}{1 - \frac{1-1/G}{\ln G}},$$

which is given by the statement of the theorem.

□

5

Reducing the seed length of explicit constructions

As mentioned in Chapter 4, Dodis et al. show in [3] that there exist condensers with only a single bit of entropy gap which have $O(\log\log\frac{1}{\varepsilon})$ entropy loss. Furthermore, they show an explicit construction for such condensers:

Theorem 5.1 ([3, Corollary 4.5])**.** *For* $0 < \varepsilon < 2^{-5}$ *and* $k = m + \log\log\frac{1}{\varepsilon} + 4$*, there exists a* $k \to_\varepsilon k' = m - 1$ *strong condenser* $C : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ *with seed length* $d = O(n \log\frac{1}{\varepsilon})$.

In this construction, the seed is used to choose a random hash function from a $(\lceil\log\frac{1}{\varepsilon}\rceil+6)$-independent hash family. The chosen function is then applied to the weak random source. See [3] for details.

A main goal in [3] is to construct unpredictability extractors. These are useful for key derivation, i.e., using weak random sources to generate keys for certain cryptographic protocols. One way to achieve unpredictability extractors that is detailed in [3] is by using randomness condensers. We restate this in Lemma 3.1. Dodis et al. consider unpredictability extractors with various output lengths. We note that, given a construction of C as in Theorem 5.1, one can construct unpredictability extractors for any output length, including

those that are larger than $k - \log\log\frac{1}{\varepsilon} - 4$, in a way that asymptotically matches the results in [3]. This is done by simply padding the output with (for example) zeroes until we reach the desired output length. We restate the results of [3] regarding the construction of unpredictability extractors for general output lengths. The statement and proof are slightly reformulated to emphasize this possible use of C:

Theorem 5.2 ([3, Theorem 4.1, item 3])**.** *Fix $\delta > 0$ and $m \geqslant k - \log\log\frac{1}{\delta} - 4$. Then, there exists an efficient construction of a (k, δ, δ') strong unpredictability extractor $S : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ for $\delta' = O(1 + 2^{m-k}\log\frac{1}{\delta})\delta$, with seed length $d = O(n\log\frac{1}{\delta})$.*

Proof. Denote $m' = k - \log\log\frac{1}{\delta} - 4$. Let $C : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^{m'}$ be a $k \to_\delta k' = m' - 1$ strong condenser as in Theorem 5.1. Let $S(x,s)$ be a binary string obtained by appending $C(x,s)$ with $m - m' = m - k + \log\log\frac{1}{\delta} + 4$ zeroes. S is a strong $k \to_\delta k'$ condenser, i.e., it has error δ and entropy gap $m - k' = m - k + \log\log\frac{1}{\delta} + 5$. By Lemma 3.1, for all $t > 0$, S is a $(k, t, 2^{m-k+5}\log(\frac{1}{\delta})t + \delta)$ unpredictability extractor. In particular, taking $t = \delta$, we have that S is a $(k, \delta, (1 + 2^{m-k+5}\log\frac{1}{\delta})\delta)$ unpredictability extractor, as required. □

Dodis et al. ask the following question: can the seed length in the construction of unpredictability extractors be further reduced? As we note, reducing the seed length in the construction of condensers with constant entropy gap and $O(\log\log\frac{1}{\varepsilon})$ entropy loss would imply the same reduction in seed length for unpredictability extractors for the entire range of output lengths. It was already shown in [3] that the seed length can in fact be reduced, where the seed length was lowered to $O(n\log k)$ by a more intricate construction and then to $O(k\log k)$. Here we show that a seed length of $O(\log\frac{n}{\varepsilon}\cdot\log\frac{k}{\varepsilon})$ can be achieved by concatenating the simpler construction from Theorem 5.1 with an existing explicit extractor and condenser. This improves upon the previous construction in some cases, for example, when ε is constant and k is asymptotically larger than $\log n$.

Theorem 5.3 (Seed reduction via extracting and then condensing)**.** *For $0 < \varepsilon < 2^{-5}$ and $k = m + \log\log\frac{1}{\varepsilon} - 4$, there exists a a $k \to_\varepsilon k' = m - 2$ strong condenser $C : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ with seed length $d = O(\log\frac{n}{\varepsilon}\log\frac{k}{\varepsilon})$.*

Proof. **Construction.** The construction uses the following components:

1. A strong extractor with minimal entropy loss. Let $E : \{0,1\}^n \times \{0,1\}^{d_1} \to \{0,1\}^{m_1}$ be an explicit (k, ε) strong extractor with entropy loss $l_1 = O(\log\frac{1}{\varepsilon})$ and seed length $d_1 = O(\log k \cdot \log\frac{n}{\varepsilon})$. Note that $m_1 = k - l_1$. It is shown in [13, Corollary 6.40] that E exists.

2. A lossless condenser. Let $GUV : \{0,1\}^n \times \{0,1\}^{d_2} \to \{0,1\}^{m_2}$ be an explicit $l_1 \to_\varepsilon l_1$ strong condenser with seed length $d_2 = 2(\log(n) + \log(l_1) + \log(\frac{1}{\varepsilon})) = O(\log \frac{n}{\varepsilon})$ and output length $m_2 = 2(d_2 + l_1) = O(\log \frac{n}{\varepsilon})$. It is shown in [5, Theorem 1.7] that GUV exists.
3. A previous construction of a condenser with one bit of entropy gap and $O(\log\log \frac{1}{\varepsilon})$ entropy loss. Let $DPW : \{0,1\}^{m_2} \times \{0,1\}^{d_3} \to \{0,1\}^{m_3}$ be a $l_1 \to_\varepsilon m_3 - 1$ strong condenser, where $m_3 = l_1 - \log\log \frac{1}{\varepsilon} - 4$ and $d_3 = O(m_2 \log \frac{1}{\varepsilon}) = O(\log \frac{n}{\varepsilon} \log \frac{1}{\varepsilon})$. The existence of DPW follows from Theorem 5.1.

We define $d = d_1 + d_2 + d_3 = O(\log \frac{n}{\varepsilon} \log k) + O(\log \frac{n}{\varepsilon} \log \frac{1}{\varepsilon}) = O(\log \frac{n}{\varepsilon} \log \frac{k}{\varepsilon})$ and $m = m_1 + m_3 = k - \log\log \frac{1}{\varepsilon} - 4$.

Now, given $x \in \{0,1\}^n$ and $s_i \in \{0,1\}^{d_i}$ for $i \in [3]$, we define $C : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ in the following way:

$$C(x, s_1, s_2, s_3) = (E(x, s_1), DPW(GUV(x, s_2), s_3)).$$

Claim: C is a $k \to_\varepsilon m - 2$ strong condenser.

Proof: Let X be a (n, k) source. Let x be sampled from X and let s_i be sampled from U_{d_i} for $i \in [3]$. Denote $y_1 = E(x, s_1)$, $y_2 = GUV(x, s_2)$ and $y_3 = DPW(y_2, s_3)$. By the definition of E, the distribution of $y_1|s_1$ is ε-close to U_{m_1}. It holds with probability $1 - \varepsilon$ that the distribution $X|(s_1, y_1)$ has at least $l_1 - 1$ min-entropy (see [9, Claims 29 and 30]). We assume this is the case (this adds ε error to the final condenser). We therefore have by the definition of GUV that $y_2|(s_1, s_2, y_1)$ is ε-close to a distribution with at least $l_1 - 1$ min-entropy. Therefore, by the definition of DPW, $y_3|(s_1, s_2, s_3, y_1)$ is 2ε-close to a distribution with $m_3 - 2$ min-entropy.

To summarize, we want to show that the distribution of

$$(s_1, s_2, s_3, y_1, y_3),$$

has at least $d + m - 2 = d + k - \log\log \frac{1}{\varepsilon} - 6$ min-entropy and we have that

- s_1, s_2, s_3 are uniformly distributed over $\{0,1\}^d$,
- $y_1|(s_1, s_2, s_3)$ is ε-close to uniform over $\{0,1\}^{k-l_1}$,
- $y_3|(s_1, s_2, s_3, y_1)$ is 2ε-close to a distribution with $l_1 - \log\log \frac{1}{\varepsilon} - 6$ min-entropy.

Indeed we see that with probability $1 - O(\varepsilon)$, every output string has weight at most $2^d \cdot 2^{k-l_1} \cdot 2^{l_1 - \log\log \frac{1}{\varepsilon} - 6}$, as required.

□

We conclude by stating the result for unpredictability extractors which follows from Theorem 5.2 and Theorem 5.3:

Corollary 5.4. *Fix $\delta > 0$ and $m \geqslant k - \log\log\frac{1}{\delta} - 4$. Then, there exists an efficient construction of a (k, δ, δ') strong unpredictability extractor $S : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ for $\delta' = O(1 + 2^{m-k}\log\frac{1}{\delta})\delta$, with seed length $d = O(\log\frac{n}{\delta}\log\frac{k}{\delta})$.*

Proof. This proof is precisely the same as the preceding proof of Theorem 5.2, except that here C is the condenser given by Theorem 5.3. □

Bibliography

[1] Avraham Ben-Aroya, Dean Doron, and Amnon Ta-Shma. Explicit two-source extractors for near-logarithmic min-entropy. In *Electronic Colloquium on Computational Complexity (ECCC)*, 2016.

[2] Gil Cohen. Two-source extractors for quasi-logarithmic min-entropy and improved privacy amplification protocols. *ECCC*, 2016.

[3] Yevgeniy Dodis, Krzysztof Pietrzak, and Daniel Wichs. Key derivation without entropy waste. In *Annual International Conference on the Theory and Applications of Cryptographic Techniques*, pages 93–110. Springer, 2014.

[4] Oded Goldreich. A sample of samplers: A computational perspective on sampling. In *Studies in Complexity and Cryptography. Miscellanea on the Interplay between Randomness and Computation*, pages 302–332. Springer, 2011.

[5] Venkatesan Guruswami, Christopher Umans, and Salil Vadhan. Unbalanced expanders and randomness extractors from parvaresh–vardy codes. *Journal of the ACM (JACM)*, 56(4):20, 2009.

[6] Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, 1963.

[7] Xin Li. Improved non-malleable extractors, non-malleable codes and independent source extractors. *arXiv preprint arXiv:1608.00127*, 2016.

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

[9] Ran Raz, Omer Reingold, and Salil Vadhan. Extracting all the randomness and reducing the error in trevisan's extractors. *Journal of Computer and System Sciences*, 65(1):97 – 128.

[10] Amnon Ta-Shma and Christopher Umans. Better lossless condensers through derandomized curve samplers. In *Foundations of Computer Science, 2006. FOCS'06. 47th Annual IEEE Symposium on*, pages 177–186. IEEE, 2006.

[11] Amnon Ta-Shma and Christopher Umans. Better condensers and new extractors from parvaresh-vardy codes. In *Computational Complexity (CCC), 2012 IEEE 27th Annual Conference on*, pages 309–315. IEEE, 2012.

[12] Amnon Ta-Shma, Christopher Umans, and David Zuckerman. Loss-less condensers, unbalanced expanders, and extractors. In *Proceedings of the thirty-third annual ACM symposium on Theory of computing*, pages 143–152. ACM, 2001.

[13] Salil P Vadhan et al. Pseudorandomness. *Foundations and Trends® in Theoretical Computer Science*, 7(1–3):1–336, 2012.

[14] David Zuckerman. Randomness-optimal oblivious sampling. *Random Structures and Algorithms*, 11(4):345–367, 1997.