On The Entropy Loss and Gap of Condensers*

NIR AVIV, Tel Aviv University
AMNON TA-SHMA, Tel Aviv University

Many algorithms are proven to work under the assumption that they have access to a source of random, uniformly distributed bits. However, in practice, sources of randomness are often imperfect, giving n random bits which have only $k < n$ min-entropy. The value $n - k$ is called the *entropy gap* of the source. *Randomness condensers* are hash functions which hash any such source to a shorter source with reduced entropy gap g. The goal is to lose as little entropy as possible in this process. Condensers also have an error parameter ε, and use a small *seed* of uniformly distributed bits whose length we desire to minimize as well.

In this work we study the exact dependencies between the different parameters of seeded randomness condensers. We obtain a non-explicit upper bound, showing the existence of condensers with entropy loss $\log(1+\frac{\log\frac{1}{\varepsilon}}{g})+O(1)$ and seed length $\log(\frac{n-k}{\varepsilon g})+O(1)$. In particular, this implies the existence of condensers with $O(\log\frac{1}{\varepsilon})$ entropy gap and constant entropy loss. This extends (with slightly improved parameters) the non-explicit upper bound for condensers presented in the work of Dodis et al. [Dodis et al. 2014], which gives condensers with entropy loss at least $\log\log\frac{1}{\varepsilon}$. We also give a non-explicit upper bound for *lossless* condensers, which have entropy gap $g \geq \frac{\log\frac{1}{\varepsilon}}{\varepsilon} + O(1)$ and seed length $\log(\frac{n-k}{\varepsilon^2 g}) + O(1)$.

Furthermore, we address an open question raised in [Dodis et al. 2014], where Dodis et al. showed an *explicit* construction of condensers with constant gap and $O(\log\log\frac{1}{\varepsilon})$ loss, using seed length $O(n\log\frac{1}{\varepsilon})$. In the same paper they improve the seed length to $O(k\log k)$, and ask whether it can be further improved. In this work, we reduce the seed length of their construction to $O(\log(\frac{n}{\varepsilon})\log(\frac{k}{\varepsilon}))$ by a simple concatenation.

In the analysis we use and prove a tight equivalence between condensers and extractors with multiplicative error. We note that a similar, but non-tight, equivalence was already proven by Dodis et al. [Dodis et al. 2014] using a weaker variant of extractors called *unpredictability extractors*. We also remark that this equivalence underlies the work of Ben-Aroya et al. [Ben-Aroya et al. 2016] and later work on explicit two source extractors, and we believe it is interesting on its own right.

ACM Reference format:
Nir Aviv and Amnon Ta-Shma. 2017. On The Entropy Loss and Gap of Condensers. *ACM Trans. Comput. Theory* 9, 4, Article 1 (December 2017), 14 pages.
https://doi.org/0000001.0000001

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

*Adapted from Nir Aviv's M.Sc thesis, supervised by Prof. Amnon Ta-Shma.

Manuscript submitted to ACM

ε-close to uniform. Condensers are probabilistic hash functions that hash any given source with k min-entropy to a shorter source while retaining much (or all) of the 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 $d + k - m$, i.e., the difference between the entropy $d + k$ of the input and seed combined, 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 $2\log\frac{1}{\varepsilon} - O(1)$ [Radhakrishnan and Ta-Shma 2000].

While extractors have an unavoidable entropy loss, the GUV condenser [Guruswami et al. 2009] is lossless, i.e., it has *zero* entropy loss (see also [Ta-Shma and Umans 2006, 2012; Ta-Shma et al. 2001]). Specifically, for every $\alpha > 0$, [Guruswami et al. 2009] show a lossless $C : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ that is a $k \to_\varepsilon k + d$ condenser, with $m \le 2d + (1+\alpha)k$ and $d = (1 + \frac{1}{\alpha})(\log(nk) + \log\frac{1}{\varepsilon}) + O(1)$. Notice also that in the GUV condenser the seed length dependence on the error may be approaching $\log\frac{1}{\varepsilon}$ whereas extractors must have dependence $2\log\frac{1}{\varepsilon}$. Thus, the GUV construction has reduced entropy loss and shorter seed length than the best possible extractor, and this is because the output distribution $E(X, U_d)$ of an extractor is dense in the co-domain of the condenser, and is sparse with condensers. 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. We define the *entropy loss* of the condenser to be $d + k - k'$. 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 $O(\log n) + \log\frac{1}{\varepsilon}$.

A natural question that emerges is how fast does this dependence drop with the entropy gap? Dodis et. al. [Dodis et al. 2014] observe that when the entropy gap is just a constant, the entropy loss may be $O(\log\log\frac{1}{\varepsilon})$ (instead of $2\log\frac{1}{\varepsilon}$ when the entropy gap is 0) and the seed length dependence on the error is just $\log\frac{1}{\varepsilon}$ (instead of $2\log\frac{1}{\varepsilon}$ with no entropy gap). 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. [Dodis et al. 2014] 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 [Cohen 2016; Li 2016].

In this paper we study the exact dependence of the seed length and entropy loss on the entropy gap. We focus on *strong* $k \to_\varepsilon k'$ condensers, for which the output distribution is guaranteed to have k' entropy even conditioned on the seed. Thus no entropy is lost from the seed, and we define the entropy loss for strong condensers to be $k - k'$. We obtain the following non-explicit upper bounds:

Theorem 1.1 (see also Theorem 4.4). *Let $0 < \varepsilon < \frac{1}{2}$. Let $c > 0$ be any constant, $g \ge c$, and $n \ge k > \log\log\frac{n+\log(2e)}{\varepsilon g} + 3$. Then for $l \le k$ such that*

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

and for

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

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 [Dodis et al. 2014]. In particular, it implies that there exists a condenser with constant loss and $\log \frac{1}{\varepsilon}$ gap. In comparison, as we show in Lemma 4.2, the non-explicit bound in [Dodis et al. 2014] gives condensers with at least $\log\log \frac{1}{\varepsilon}$ loss. We note that the definition of entropy loss in [Dodis et al. 2014] is different than ours: they consider $k - m$ whereas we consider $k - k'$. According to their definition, they achieve zero loss. This does not contradict the results in Lemma 4.2, and is due to the different definition.

We further show that when the entropy gap is about $\frac{1}{\varepsilon}$ the entropy loss may be reduced to zero:

Theorem 1.2 (see also Theorem 4.5). *Let* $0 < \varepsilon < \frac{1}{2}$, $g \geq \frac{\log(\frac{e}{\varepsilon})}{\varepsilon}$, *and* $n \geq k > \log\log \frac{n+\log(2e)}{\varepsilon^2 g} + 3$. *Then for*

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

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

This bound shows that to get to zero loss it suffices to have $O(\frac{\log \frac{1}{\varepsilon}}{\varepsilon})$ entropy gap. Comparing Theorem 1.2 with Theorem 1.1 it seems that the dependence of the seed length on the error in Theorem 1.2 is larger, as it is $O(\log \frac{n-k}{\varepsilon^2 g})$, compared with $O(\log \frac{n-k}{\varepsilon g})$ in Theorem 1.1. This is, however, misleading, as in Theorem 1.1 g is at least a constant and $O(\log \frac{n-k}{\varepsilon g}) = O(\log \frac{n-k}{\varepsilon})$, whereas in Theorem 1.2 $g = \Omega(\frac{\log(1/\varepsilon)}{\varepsilon})$ and $O(\log \frac{n-k}{\varepsilon^2 g}) = O(\frac{n-k}{\varepsilon \log(1/\varepsilon)})$.

In the analysis, we use and prove a tight equivalence between condensers and extractors with multiplicative error (see Section 3). A similar, but non-tight, equivalence was already proved by Dodis et al. [Dodis et al. 2014], using somewhat different objects: they consider a weaker variant of extractors 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 extractors 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. [Ben-Aroya et al. 2016] and later work on explicit two source extractors, and we believe it is interesting on its own right.

Finally, we partially address an open question from [Dodis et al. 2014]. 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 further 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, is 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 in Section 5.

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

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 lower-case n, k, m, g, d. For example, $d = \log(D)$. $[N]$ denotes the set $\{1, 2, \ldots, N\}$.

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

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

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 2.2. *Let $n \geq k \geq 0$, and $0 \leq \varepsilon \leq n - k$. 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 lemma for flat sources: if X is a convex combination of $X_1, \ldots, X_t$, $X = \sum \lambda_i X_i$, and each X_i is ε-close to X_i' with $k + \varepsilon$ min-entropy, then X is ε close $X' = \sum \lambda_i X_i'$, and X' has at least $k + \varepsilon$ min-entropy.

Thus, let us assume that X is a flat (n, k) source. Thus, each of the 2^k elements in the support of X has weight exactly 2^{-k}. We let X' be the random variable obtained by taking from each element in the support of X the excess weight, i.e., $2^{-k} - 2^{-(k+\varepsilon)}$ and distributing it evenly over elements in $\{0,1\}^n \setminus Supp(X)$. Clearly, $H_\infty(X') \geq k + \varepsilon$ (because $k + \varepsilon \leq n$) and $\|X - X'\| \leq 2^k(2^{-k} - 2^{-(k+\varepsilon)}) = 1 - 2^{-\varepsilon} \leq \varepsilon$. Thus X is ε-close to the $(n, k + \varepsilon)$-source X'. □

We use several concentration bounds. For this we need the *Kullback-Leibler divergence*:

Definition 2.3 (KL-divergence). Let $0 < p, q < 1$. Then

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

Theorem 2.4 (Chernoff-Hoeffding inequality). *[Hoeffding 1963] Let $X_1, \ldots, X_n$ be independent random variables taking values in $[0, 1]$ and let $q = \mathbb{E}\frac{\sum_{i=1}^n X_i}{n}$. Then, for every $q < p < 1$,*

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

.

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

Lemma 2.5. *Let $0 < q < p < 1$,*

(1) *$D(p||q) \geq p \ln(\frac{p}{q}) + q - p$, and,*
(2) *Let $A > 1$, $b > 0$ such that $p = Aq + b \leq 1$. Then $D(Aq + b||q) \geq b \ln(A)$.*

Proof. For the first item note that, by definition, $D(p||q) = p\ln(\frac{p}{q}) + (1-p)\ln(\frac{1-p}{1-q})$. Since $\ln(x) \geq 1 - \frac{1}{x}$ for $x > 0$, it holds that $(1-p)\ln(\frac{1-p}{1-q}) \geq (1-p)(1-\frac{1-q}{1-p}) = q - p$, proving the first item.

Next, we prove the second item. By the first item, $D(Aq+b||q) \geq (Aq+b)\ln(A+\frac{b}{q}) - Aq - b + q$. Denote $t = \frac{A}{b}q$. Then,

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

Thus, to show $D(Aq+b||q) \geq b\ln(A)$ it suffices to show that

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

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

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

However, for all $x \geq 1$, $\ln(x) + \frac{1}{x} \geq 1$. As $A > 1$, $\ln(A) + \frac{1}{A} \geq 1$. Similarly, as $s \geq 1$, $1 + s\ln s \geq s$. Together, $s\ln(s) + \ln(A) + \frac{1}{A} \geq s\ln(s) + 1 \geq s$ as desired. □

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

Definition 2.6. Let $0 \leq p \leq 1$. Then $H(p) = -p\ln(p) - (1-p)\ln(1-p)$.

We use the following fact:

Lemma 2.7. *For every* $0 < p < 1$, $H(p) \leq p\ln(\frac{e}{p})$.

Proof. Since $\ln(x) \leq x - 1$ for every $x > 0$, we have that $H(p) \leq p\ln(\frac{1}{p}) + (1-p)(\frac{1}{1-p} - 1) \leq p\ln(\frac{1}{p}) + p$ as required. □

3 CONDENSERS AS EXTRACTORS WITH MULTIPLICATIVE ERROR

Dodis et al. [Dodis et al. 2014] showed an equivalence between strong randomness condensers and strong unpredictability extractors. This equivalence has a gap in parameters. In this section, we show that condensers are precisely equivalent (with no gap in parameters) to *extractors with multiplicative error*.

3.1 Basic Definitions

We begin with the standard definition of condensers:

Definition 3.1 (Randomness condenser). A function $C : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ is a $k \to_\varepsilon k'$ condenser if for any (n,k)-source X, and a uniformly random and independent seed S over $\{0,1\}^d$, the distribution $C(X,S)$ is ε-statistically-close to some distribution 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 $k \to_\varepsilon k'$ strong 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 1, the output distribution is ε-close to some distribution with min-entropy $m-1$. There can be many such distributions. If the entropy gap is 0, then the output distribution is ε-close to the distribution U_m. In this case, the function is called a *randomness extractor*:

Definition 3.2 (Randomness extractor). A function $E : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ is a (k,ε) extractor if for every (n,k)-source X and every $T \subseteq \{0,1\}^m$, it holds that

$$\Pr_{x \in X, s \in U_d}[E(x,s) \in T] \leq \frac{|T|}{2^m} + \varepsilon.$$

E is a (k,ε) strong extractor, if the function $E'(X,Y) = (Y, E(X,Y))$ is a (k,ε) extractor.

We immediately see from the definitions that $E : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ is a (k,ε) extractor if and only if E is a $k \to_\varepsilon m$ condenser.

In Definition 3.2 E is an extractor if it passes every "test" T with at most *additive* error ε. We now define extractors which have *multiplicative* error as well as additive error:

Definition 3.3 (Extractor with multiplicative error). A function $E : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ is a (k,G,ε) extractor with multiplicative error if for every (n,k)-source X and every $T \subseteq \{0,1\}^m$, it holds that

$$\Pr_{x \in X, s \in U_d}[E(x,s) \in T] \leq G\frac{|T|}{2^m} + \varepsilon.$$

E is a (k,G,ε) strong extractor with multiplicative error if the function $E'(X,Y) = (Y, E(X,Y))$ is a (k,G,ε) extractor with multiplicative error.

This definition is motivated by the work of [Dodis et al. 2014] who gave a similar definition:

Definition 3.4 (Strong unpredictability extractor). A function $S : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ is a strong (k,δ,δ') 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 $\Pr_{x \in X, s \in U_d}[(s, S(x,s)) \in T] \leq \delta'$.

Comparing the two definitions we see that E is a strong (k,G,ε) extractor with multiplicative error if and only if E is a strong $(k,\delta,G\delta+\varepsilon)$ unpredictability extractor for every $\delta > 0$.

3.2 Condensers Imply Extractors With Multiplicative Error

Dodis et al. [Dodis et al. 2014] proved the following:

Lemma 3.5 ([Dodis et al. 2014, Lemma 3.3]). *Let $n \geq k > k' > 0$, $d > 0$ 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,G,ε) strong extractor with multiplicative error.*

Proof. 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 $|T| \leq \delta 2^{d+m}$ and

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

Therefore, even if we move ε weight from elements in T to elements outside of T, there will be some element $t \in T$ 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 extractors with multiplicative error are also condensers. We discuss this in the following section.

3.3 Extractors With Multiplicative Error Imply Condensers

We begin with a result of [Dodis et al. 2014] which shows that strong unpredictability extractors are strong condensers:

Lemma 3.6 ([Dodis et al. 2014, Section 3, Summary]). *Let $n \geq k > 0$, $d > 0$, $m > k$ 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. Denote $g = \log \frac{\delta'}{\delta}$ and $\varepsilon = \delta'$. Then S is a $k \to_\varepsilon m - g$ strong condenser.*

Dodis et al. [Dodis et al. 2014] give an indirect proof of Lemma 3.6, going through *balanced hash functions*. Note that in the terms of Lemma 3.6, $\delta' = \frac{G\delta+\varepsilon}{2}$, whereas in Lemma 3.5 $\delta' = G\delta + \varepsilon$, so there is a small gap in the parameters of Lemma 3.6 and Lemma 3.5. We show a direct reduction from strong extractors with multiplicative error to strong condensers. Along with Lemma 3.5, this reduction shows that strong condensers with entropy gap g and error ε are precisely equivalent to strong extractors with multiplicative error $G = 2^g$ and additive error ε. Note that a strong extractor with multiplicative error is a stronger object than an unpredictability extractor, as it is one function that is a (k, δ, δ') unpredictability extractor for *all* $\delta > 0$ and $\delta' = G\delta + \varepsilon$.

Lemma 3.7. *Let $n \geq k > 0$, $d > 0$, $m > k$, $\varepsilon > 0$ and $g > 0$. Let $C : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ be a strong (k, G, ε) extractor with multiplicative error. Denote $g = \log G$. Then, C is a $k \to_\varepsilon k' = m - g$ strong condenser.*

Proof. Let X be a (n,k)-source. Consider the distribution $W = (s, C(x,s))$ where x is sampled from X and s is sampled from U_d. Let $H = \left\{h \in \{0,1\}^d \times \{0,1\}^m \mid \Pr[W = h] > 2^{k'+d}\right\}$. 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, C is a $(k, \delta, G\delta + \varepsilon)$ strong unpredictability extractor for all $\delta > 0$, and in particular for $\delta = \rho(H)$, and therefore

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

as required. □

The following table summarizes the error ε and entropy gap g of the condensers that are implied by the existence of a strong (k, G', ε') extractor with multiplicative error in Lemmas 3.6 and 3.7. Note that in Lemma 3.6 there is an additional parameter $\delta > 0$ which encapsulates a tradeoff between error and entropy gap: the lower the desired entropy gap, the larger the statistical distance becomes, and vice versa.

	Error	Entropy gap
Lemma 3.7	$\varepsilon = \varepsilon'$	$g = g' = \log G'$
Lemma 3.6 [Dodis et al. 2014]	$\varepsilon = G'\delta + \varepsilon'$	$g = \log(G' + \frac{\varepsilon'}{\delta})$

We see that Lemma 3.7 simultaneously obtains error ε' and gap g'. In contrast, with Lemma 3.6 one can obtain error ε' if δ tends to 0, and gap g' if δ is close to 1, but cannot obtain both simultaneously.

To summarize the equivalence between strong condensers and strong extractors with multiplicative error, the results of Lemma 3.7 proved in this subsection along with the results of Lemma 3.5 proved in [Dodis et al. 2014] give rise to the following corollary:

Corollary 3.8. *Let $n \geq k > 0$, $d > 0$, $m > k$, $\varepsilon > 0$ and $g > 0$. Any function $C : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ is a $k \to_\varepsilon k' = m - g$ strong condenser if and only if C is a (k, G, ε) strong extractor with multiplicative error.*

4 AN UPPER BOUND ON THE SEED LENGTH AND ENTROPY LOSS OF CONDENSERS WITH A SMALL ENTROPY GAP

The results of the previous section raise the question whether extractors that allow multiplicative error have substantially better parameters than those that only allow additive error. As mentioned in the introduction, the answer is yes, and this was already observed in [Dodis et al. 2014]. We begin this section recalling the results from [Dodis et al. 2014].

4.1 Previous work

In [Dodis et al. 2014], Dodis et al. gave a probabilistic proof for the existence of (k, δ, δ') unpredictablity extractors:

Theorem 4.1 ([Dodis et al. 2014, 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:*

$$\delta' > \max\left\{2e\delta, (n-k-2)2^{-d} + \log(e/\delta)\delta 2^{m-k}\right\} \tag{1}$$

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

Dodis et al. [Dodis et al. 2014] also gave a way to construct condensers from unpredictability extractors (see also Section 3.3, Lemma 3.6). Applying this to Theorem 4.1 gives an existence proof for the following condensers:

Lemma 4.2. *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 3.6 to the unpredictability extractors 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 3.6, 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 1 holds and the case where Equation 2 holds separately.

Large gap: Assume that Equation 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 1 holds then:

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

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

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

Equivalently,

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

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

□

Note that Lemma 4.2 implies that the condensers given in [Dodis et al. 2014] have seed length $d \geq \log\frac{n-k+2}{\varepsilon}$ and entropy loss $l \geq \log(g + \log\frac{e}{\varepsilon})$. We know that there exist *lossless* condensers with $l = 0$ (e.g., [Guruswami et al. 2009]). In the next subsections, we show a probabilistic argument for the existence of condensers that extends the results of [Dodis et al. 2014] to constant entropy loss, with slightly improved seed length and entropy gap. We also show the existence of lossless condensers when the entropy gap is large.

4.2 The upper bound: statement and discussion

In this subsection we study the entropy loss and seed length as a function of the entropy gap and error. We first state a 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 4.3 (Generalized upper bound). *Fix $0 < \zeta < 1$, $0 < \varepsilon < \frac{1}{2}$, $g > 0$, and $n \geq k \geq \log\log\frac{n+\log(e)+1}{\zeta\varepsilon g} + 2$. Let $l \leq k$ be such that*

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

and let d be such that

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

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 4.3 in the next subsection. The various parameters in Theorem 4.3 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 4.4 (Upper bound for condensers with at least one bit of entropy loss). *Fix $0 < \varepsilon < \frac{1}{2}$, $g > 0$, and $n \geq k \geq \log\log\frac{n+\log(e)+1}{\varepsilon g} + 3$. Let $l \leq k$ be such that*

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

and let d be such that

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

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

Thus with only one bit of entropy gap we achieve $\log\log\frac{1}{\varepsilon} + O(1)$ entropy loss, as in the results of [Dodis et al. 2014] which previously showed this is possible. Further comparing Theorem 4.4 to the parameters of condensers given by [Dodis et al. 2014] (see Lemma 4.2), we see the following:

- For entropy gap $g > 1$, the parameters in Theorem 4.4 slightly improve over the parameters shown in Lemma 4.2: Theorem 4.4 has $d = \frac{n-k+\log(e)+1}{\varepsilon g} + 1$ instead of $d \geq \frac{n-k+2}{\varepsilon}$ and $l = \log(1 + \frac{\log\frac{e}{\varepsilon}}{g}) + O(1)$ rather than $l \geq \log(g + \log\frac{e}{\varepsilon})$.
- There is no restriction that $l \geq \log\log\frac{1}{\varepsilon}$. Indeed, we see that with $\log\frac{1}{\varepsilon} + O(1)$ entropy gap we achieve constant entropy loss.

The entropy loss in Theorem 4.4 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 $g \geq \frac{\log\frac{1}{\varepsilon}}{\varepsilon}$. This results in the following theorem:

Theorem 4.5 (Upper bound for lossless condensers). *Fix $0 < \varepsilon < \frac{1}{2}, g \geq \frac{\log(\frac{e}{\varepsilon})}{\varepsilon}$, and $n \geq k \geq \log\log\frac{n+\log(e)+1}{\varepsilon^2 g} + 3$. Let*

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

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

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

Set $\zeta = \frac{\varepsilon}{2}$. We show that $\log\frac{1}{1-\zeta} \leq \varepsilon$, or equivalently $\frac{\varepsilon}{2} \leq 1 - \frac{1}{2^\varepsilon}$. 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 $\log\frac{1}{1-\frac{1-1/G}{\ln G}} \leq \varepsilon$. Since $\ln G > 1$, it suffices to show that $\log\frac{1}{1-\frac{1}{\ln G}} = \log(1+\ln\frac{1}{\varepsilon}) - \log(1+\ln\frac{1}{\varepsilon}-\varepsilon) \leq \varepsilon$. By the mean value theorem, for every pair of positive real numbers $x > t$, it holds that $\frac{\log(x)-\log(x-t)}{t} \leq \frac{\log(e)}{x-t}$. In particular,

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

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

By Theorem 4.3 and what we have shown so far, there exists a $k \rightarrow_{\varepsilon} k' = k - O(\varepsilon)$ strong condenser $C : \{0,1\}^n \times \{0,1\}^d \rightarrow \{0,1\}^{k'+g}$. By Lemma 2.2 this implies that C is a randomness condenser with error $O(\varepsilon)$ and zero entropy loss. □

4.3 Proof of the upper bound

It remains to prove the generalized upper bound of Theorem 4.3.

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

Definition 4.6. 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 3.8, 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 lemma bounds the chance of a given pair to be C-bad.

Lemma 4.7. *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 independent and $E[\sum_{a,s} R_{a,s}] = \delta KD$. By Theorem 2.4, $\Pr[\sum_{a,d} X_{a,d} \geq (G\delta + \varepsilon)KD] \leq e^{-D(G\delta+\varepsilon||\delta)KD}$. □

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 $\rho(B) \geq \frac{1-\varepsilon}{G}$. Therefore, by Lemma 4.7 and the union bound,

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

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

$$\binom{N}{K} e^{-\zeta D(G\delta+\varepsilon||\delta)KD} < \frac{G}{DM}, \tag{3}$$

and also that

$$\binom{DM}{\delta DM} e^{-(1-\zeta)D(G\delta+\varepsilon||\delta)KD} < 1 \tag{4}$$

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

To prove Eq. (3), using $\binom{N}{K} \leq (\frac{eN}{K})^K = 2^{K(n-k+\log(e))}$, it suffices to show that

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

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

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

By Lemma 2.5, we have that $D(G\delta + \varepsilon||\delta) \geq \varepsilon \ln(G)$, and the required equation then follows from the statement of the theorem.

It remains to prove Eq. (4). Using the fact that $\binom{DM}{\delta DM} \leq e^{H(\delta)DM}$, 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

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

We divide into cases:

Small δ: assume that $\delta \leq \frac{\varepsilon}{G}$. Since $\varepsilon < \frac{1}{2}$ and $H(\delta)$ is monotone increasing for $\delta \in (0, \frac{1}{2})$, we have that $H(\delta) \leq H(\frac{\varepsilon}{G})$. By Lemma 2.7, $H(\frac{\varepsilon}{G}) \leq \frac{\varepsilon}{G}\ln(\frac{Ge}{\varepsilon})$. By Lemma 2.5, $D(G\delta + \varepsilon||\delta) \geq \varepsilon\ln(G)$. It therefore suffices to prove

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

which is implied by the statement of the theorem.

Large δ: assume that $\frac{\varepsilon}{G} < \delta$. By Lemma 2.7, $H(\delta) \leq \delta\ln(\frac{e}{\delta}) \leq \delta\ln(\frac{eG}{\varepsilon})$. By Lemma 2.5, $D(G\delta + \varepsilon||\delta) \geq (G\delta+\varepsilon)(\ln(G+\frac{\varepsilon}{\delta}) + \frac{1}{G+\frac{\varepsilon}{\delta}} - 1)$. Using that $\varepsilon > 0$ and that $\ln(x) + \frac{1}{x}$ is monotone increasing for all $x > 1$, we have that $D(G\delta + \varepsilon||\delta) \geq G\delta(\ln G + \frac{1}{G} - 1)$. 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 Section 4, Dodis et al. [Dodis et al. 2014] show that there exist condensers with only a single bit of entropy gap which have $\log\log\frac{1}{\varepsilon} + O(1)$ entropy loss. Furthermore, they show an explicit construction for such condensers:

Theorem 5.1 ([Dodis et al. 2014, Corollary 4.5]). *For $0 < \varepsilon < 2^{-5}$ and $k = m + \log\log\frac{1}{\varepsilon} + 4$, there exists an explicit $k \to_\varepsilon k' = m - 1$ 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 [Dodis et al. 2014] for details.

As explained in Lemma 3.5 one can construct unpredictability extractors from condensers. With that we get:

Theorem 5.2 ([Dodis et al. 2014, Theorem 4.1, item 3]). *Fix $\delta > 0$ and $m \geq k - \log\log\frac{1}{\delta} - 4$. 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 3.5, 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? Dodis et al. [Dodis et al. 2014] show that the seed length can be reduced to $O(n\log k)$ by a more intricate construction and then to $O(k\log k)$.

The proof we gave to Theorem 5.2 shows that reducing the seed length of condensers with constant entropy gap and $\log\log\frac{1}{\varepsilon} + O(1)$ entropy loss would imply the same reduction in seed length for unpredictability extractors for the entire range of output lengths. We now 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.

Theorem 5.3 (Seed reduction via extracting and then condensing). *For $0 < \varepsilon < 2^{-5}$ and $k = m - \log\log\frac{1}{\varepsilon} - 5$, 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 $d = O(\log\frac{n}{\varepsilon}\log\frac{k}{\varepsilon})$.*

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 [Vadhan et al. 2012, 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 $(l_1-1) \to_\varepsilon (l_1-1)$ 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 [Guruswami et al. 2009, 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 $(l_1 - 1) \to_\varepsilon m_3 - 1$ strong condenser, where $m_3 = l_1 - \log\log\frac{1}{\varepsilon} - 5$ 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 $m = m_1 + m_3 = k - \log\log\frac{1}{\varepsilon} - 5$. 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 $k \to_\varepsilon m - 1$ 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

$$\begin{aligned} y_1 &= E(x, s_1), \\ y_2 &= GUV(x, s_2), \text{ and,} \\ y_3 &= DPW(y_2, s_3). \end{aligned}$$

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 $k - m - 1 = l_1 - 1$ min-entropy (see [Raz et al. 2002, Claims 29 and 30]). We assume this is the case (this adds ε error to the final condenser).

We therefore have by the properties 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. Thus, by the properties of DPW, $y_3|(s_1, s_2, s_3, y_1)$ is 2ε-close to a distribution with $m_3 - 1$ 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 $d + m - 1 = d + k - \log\log\frac{1}{\varepsilon} - 6$ 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.

Together 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.

□

Note that the extractor E used in the proof of Theorem 5.3 is not optimal: non-explicitly, there exist strong extractors with $l = 2\log\frac{1}{\varepsilon} - O(1)$ entropy loss and $d = \log\frac{n-k}{\varepsilon^2} + O(1)$ seed length. If we had an explicit construction of such an extractor, then the seed length in our construction would have been reduced to $O(\log\frac{n}{\varepsilon}\log\frac{1}{\varepsilon})$. The proof of this is precisely the same as the proof of Theorem 5.3, but with $d_1 = \log\frac{n-k}{\varepsilon^2} + O(1)$.

Finally, we state the result for unpredictability extractors which follows from Theorem 5.2 and Theorem 5.3:

Corollary 5.4. *Fix $\delta > 0$ and $m \geq k - \log\log\frac{1}{\delta} - 5$. 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. □

REFERENCES

Avraham Ben-Aroya, Dean Doron, and Amnon Ta-Shma. 2016. Explicit two-source extractors for near-logarithmic min-entropy.. In *Electronic Colloquium on Computational Complexity (ECCC)*, Vol. 23. 88.

Gil Cohen. 2016. Two-Source Extractors for Quasi-Logarithmic Min-Entropy and Improved Privacy Amplification Protocols.. In *Electronic Colloquium on Computational Complexity (ECCC)*, Vol. 23. 114.

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

Venkatesan Guruswami, Christopher Umans, and Salil Vadhan. 2009. Unbalanced expanders and randomness extractors from Parvaresh–Vardy codes. *Journal of the ACM (JACM)* 56, 4 (2009), 20.

Wassily Hoeffding. 1963. Probability Inequalities for Sums of Bounded Random Variables. *J. Amer. Statist. Assoc.* 58, 301 (1963), 13–30. http://www.jstor.org/stable/2282952

Xin Li. 2016. Improved non-malleable extractors, non-malleable codes and independent source extractors. *arXiv preprint arXiv:1608.00127* (2016).

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

Ran Raz, Omer Reingold, and Salil Vadhan. 2002. Extracting all the randomness and reducing the error in Trevisan's extractors. *J. Comput. System Sci.* 65, 1 (2002), 97–128.

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

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

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

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