Electronic Colloquium on Computational Complexity, Report No. 210 (2023)

On the Existence of Seedless Condensers: Exploring the Terrain

Eshan Chattopadhyay*
Cornell University
eshan@cs.cornell.edu

Mohit Gurumukhani*
Cornell University
mgurumuk@cs.cornell.edu

Noam Ringach †
Cornell University
nomir@cs.cornell.edu

Abstract

While the existence of randomness extractors, both seeded and seedless, has been thoroughly studied for many sources of randomness, currently, very little is known regarding the existence of seedless condensers in many settings. Here, we prove several new results for seedless condensers in the context of three related classes of sources: Non-Oblivious Symbol Fixing (NOSF) sources, SHELA sources as defined by Aggarwal, Obremski, Ribeiro, Siniscalchi, and Visconti [AORSV, EUROCRYPT'20], and almost Chor-Goldreich (CG) sources as defined by Doron, Moshkovitz, Oh, and Zuckerman [DMOZ, STOC'23]. Here, we will think of these sources as a sequence of random variables $\mathbf{X} = \mathbf{X}_1, \ldots, \mathbf{X}_\ell$ on ℓ symbols where at least g out of these ℓ symbols are "good" (i.e., uniformly random), denoted as a (g, ℓ)-source, and the remaining "bad" $\ell - g$ symbols may adversarially depend on these g good blocks. The difference between each of these sources is realized by restrictions on the power of the adversary, with the adversary in NOSF sources having no restrictions.

Prior to our work, the only known seedless condenser upper or lower bound in these settings is due to [DMOZ, STOC'23] which explicitly constructs a seedless condenser for a restricted subset of (g, ℓ)-almost CG sources.

The following are our main results concerning seedless condensers for each of these three sources.

1. When $g \leq \ell/2$, we prove for all three classes of sources that condensing with error 0.99 above rate $\frac{1}{\lfloor \ell/g \rfloor}$ is impossible.
2. Next, we investigate the setting of $g > \ell/2$, and in particular focus on $g = 2$ and $\ell = 3$. We show that condensing with constant error above rate $\frac{2}{3}$ is impossible for uniform NOSF sources.
3. Quite surprisingly, we show the existence of excellent condensers for uniform $(2, 3)$-SHELA and uniform almost CG sources, thus proving a separation from NOSF sources. Further, we explicitly construct a condenser that outputs $m = \frac{n}{16}$ bits and condenses any uniform $(2, 3)$-SHELA source to entropy $m - O(\log(m/\varepsilon))$ (with error ε). Our construction is based on a new type of seeded extractor that we call *output-light*, which could be of independent interest.

   In contrast, we show that it is impossible to extract from uniform $(2, 3)$-SHELA sources.

These results extend seedless extractor lower bounds on NOSF sources (Bourgain, Kahn, Kalai, Katznelson, and Linial [BKKKL, Israel J. Math'92]) and make progress on several open question from [DMOZ, STOC'23], [AORSV, EUROCRYPT'20], and Kopparty and N [KN, RANDOM'23].

*Supported by a Sloan Research Fellowship and NSF CAREER Award 2045576.
†Supported by NSF GRFP grant DGE – 2139899, NSF CAREER Award 2045576 and a Sloan Research Fellowship.

ISSN 1433-8092

Contents

1 Introduction

One of the most fruitful lines of research in computer science has been that of randomness. From the traditionally more applied areas of algorithm design (e.g., Monte Carlo simulations), error-correcting codes and cryptography to the more theoretical areas of property testing, combinatorics, and circuit lower bounds, randomness has played a key role in seminal discoveries. In many of these works, the use of high-quality random bits, or alternatively, a way to convert low-quality randomness into high-quality randomness, is essential. In cryptography, the authors of [DOPS04] showed that high-quality randomness is essential for tasks such as bit commitment schemes and secure two-party computation. On the other hand, being able to extract uniform bits from low-quality randomness allows us to simulate randomized algorithms [Vad12, Zuc90, Zuc92].

In most use-cases, randomness takes the form of uniformly random bits. These motivated the construction of randomness extractors, functions that take low-quality randomness (which we often like to think of as natural processes) and convert it into uniformly random bits. There is a long line of works [DKSS13, DW11, LRVW03] that construct seeded extractors, functions that take in low-quality randomness along with a small amount of uniform bits, with close to optimal parameters. However, a number of works [CG88, RVW04, SV86, Zuc90] have shown that deterministic extraction is impossible for many classes of randomness sources.

Naturally, the question that arises for sources for which deterministic extraction is impossible is whether any improvement on their randomness can be made. That is, while it may not be possible to convert a source into uniform bits, maybe it is possible to condense a source into bits with a higher density of randomness than the source they came from. We now informally define these notions of randomness, extractors, and condensers.

The notion of randomness that is standard in this line of work is that of min-entropy. For a source $\mathbf{X}$ on n bits, which is just a discrete distribution, we define its *min-entropy* as $H_\infty(\mathbf{X}) = \min_{x\in\{0,1\}^n}\{-\log(\Pr[\mathbf{X} = x])\}$. Generally, we call a source $\mathbf{X}$ over n bits with min-entropy at least k an (n,k)-*source*. From the definition of min-entropy, we see that this requires that for any $x \in \{0,1\}^n$ we have $\Pr[X = x] \leq 2^{-k}$. Given any two distributions $\mathbf{X}$ and $\mathbf{Y}$ on $\{0,1\}^n$, we define their statistical distance or total-variation (TV) distance as $|\mathbf{X} - \mathbf{Y}| = \max_{Z\subseteq\{0,1\}^n} |\Pr_{x\sim\mathbf{X}}[x \in Z] - \Pr_{y\sim\mathbf{Y}}[y \in Z]|$.

These two definitions allow us to define the notion of *smooth min-entropy* which we will use extensively throughout this work. Conceptually, smooth min-entropy asks that the source we are looking at be close in TV-distance to some other source with the desired amount of min-entropy. We say that the *smooth min-entropy* of a source $\mathbf{X}$ on $\{0,1\}^n$ with smoothness parameter ε is $H^\varepsilon_\infty(\mathbf{X}) = \max_{\mathbf{Y}:|\mathbf{X}-\mathbf{Y}|\leq\varepsilon} H_\infty(\mathbf{Y})$.

We are now in a position to define randomness extraction and condensing.

Definition 1.1. *Let $\mathcal{X}$ be a family of distributions over $\{0,1\}^n$. A function $\mathsf{Ext} : \{0,1\}^n \to \{0,1\}^t$ is an* extractor *for $\mathcal{X}$ with error $\varepsilon > 0$ if $|\mathsf{Ext}(\mathbf{X}) - \mathbf{U}_t| \leq \varepsilon$. When $\mathcal{X}$ is the class of (n,k)-sources, we say that Ext is a (k,ε)-extractor.*

This definition formalizes the notion that the result of an extractor is close to uniform. This is generally too much to desire as, even for general (n,k)-sources, it is impossible to extract just one bit [Vad12]. Consequently, we turn to the looser requirements in condensing.

Definition 1.2. *For a family of distributions $\mathcal{X}$ over $\{0,1\}^n$, we say that a function $\mathsf{Cond} : \{0,1\}^n \to \{0,1\}^t$ is a* condenser *with error $\varepsilon \geq 0$ if for all $\mathbf{X} \in \mathcal{X}$ we have that $H_\infty(\mathbf{X})/n \leq H^\varepsilon_\infty(\mathsf{Cond}(\mathbf{X}))/t$. We say that Cond has* entropy gap *Δ if $H^\varepsilon_\infty(\mathsf{Cond}(\mathbf{X})) \geq t - \Delta$. When $\mathcal{X}$ is the class of (n,k)-sources and $k' = t - \Delta$, we say that Cond is a (k,k',ε)-condenser.*

In other words, the output of Cond is ε-close in statistical distance to some distribution with min-entropy rate higher than that of $\mathbf{X}$, as compared to extracting which requires that our output is close to uniform. One may wonder whether such output is still meaningfully useful, and we answer such doubts in the following section.

1.1 The utility of condensing

As insinuated before, one of the main distinguishing features of condensers from extractors is that there are random sources for which deterministic condensing is possible while deterministic extraction is not. Thus, they allow us to obtain randomness that is more useful than what we began with in cases where extracting uniform bits is impossible. One significant example is that of Santha-Vazirani (SV) sources [SV86] and their generalization, Chor-Goldreich (CG) sources [CG88].

Informally, an SV source is a string of random bits such that the conditional distribution of each bit on the bits that come before it still has some minimum amount of min-entropy, and a CG source generalizes this to allow each bit to instead be a symbol in $\{0,1\}^n$. For SV sources, Santha and Vazirani showed that no deterministic extractor is better than simply outputting the first bit [RVW04, SV86]. Similarly, for CG sources, Chor and Goldreich showed the strictly stronger result that deterministic extraction is impossible [CG88]. The recent result of [DMOZ23] with regards to condensing then stands in contrast to these impossibility results for extraction from CG sources since, not only do they show that it is possible to condense from CG sources, they construct condensers with constant entropy gap. In fact, their condensers still work under some relaxations to a class of sources they termed *almost CG* sources, which we shall elaborate more on later. Other examples of sources for which deterministic extraction is not possible while deterministic condensing is are the *somewhat dependent* sources of [BGM22] and the block sources of [BCDT19].

Another important property of condensers is that they can attain parameters that are unachievable for extractors and still be useful in simulating randomized algorithms with low overhead [DMOZ23]. In addition, while condensers are important in their own right, it is worth mentioning that they are useful in the construction of extractors, such as in [AORSV20, BDT17, GUV09, RSW06, TUZ07, Zuc07].

1.2 NOSF, SHELA, and Almost-CG Sources

The three randomness sources that we focus on in this work are all composed of blocks of bits, known as symbols, which vary in how they are permitted to relate to other symbols in the source. Nevertheless, one unifying characteristic of NOSF, SHELA, and almost-CG sources is that deterministic extraction is impossible for all of them. Indeed, [BKKKL92, KKL88], [AORSV20], and [CG88] showed that no deterministic function can extract from NOSF, SHELA, and almost-CG sources, respectively. To illustrate why it is not immediately obvious that deterministic extraction should not be possible from all of these sources, we briefly introduce each source and the properties of their adversaries.

In these definitions, we will consider sources $\mathbf{X} = \mathbf{X}_1, \ldots, \mathbf{X}_\ell$ of length ℓ where each $\mathbf{X}_i \in \{0,1\}^n$ is called a block. Generally, we will term blocks that have some minimum amount of randomness "good" and blocks that are chosen by an adversary as "bad". Such NOSF sources generalize the setting of non-oblivious bit-fixing (NOBF) sources [CGHFRS85], where each $\mathbf{X}_i$ is a bit (i.e., $n = 1$).

At a high level, Non-Oblivious Symbol Fixing (NOSF) sources are those in which the good blocks are independently sampled min-entropy sources and the bad blocks may depend arbitrarily on the good blocks.

Definition 1.3 (NOSF sources)**.** *A (g, ℓ, k)-*NOSF *source $\mathbf{X} = \mathbf{X}_1, \ldots, \mathbf{X}_\ell$ on $(\{0,1\}^n)^\ell$ is such that g out of the ℓ blocks are good, i.e., are independently sampled (n, k)-sources while the other $\ell - g$ bad blocks may depend arbitrarily on the good blocks.*

When $k = n$, we simply call $\mathbf{X}$ a *uniform* (g, ℓ)*-NOSF* source, and when $n = 1$, the term *non-oblivious bit fixing (NOBF)* source is often used. At first glance, the adversary in NOSF sources clearly has a significant amount of power. Every single good block is sampled before the adversary gets to decide what to place in the bad blocks. If we were to think of the blocks with smaller indices as coming first in time, then this essentially means that the adversary gets to look into the future to see what will be placed in the good blocks to decide what to place in the bad blocks. In contrast, SHELA sources limit the adversary by forcing it to only depend on blocks in the past. Instead of defining general SHELA sources here, we defer the formal definition to Definition 4.10 and define *fixed-index Somewhere Honest Entropic Look Ahead* (fiSHELA) sources.

Definition 1.4 (SHELA sources, [AORSV20])**.** *A* (g, ℓ, k)-fiSHELA *source* $\mathbf{X} = \mathbf{X}_1, \ldots, \mathbf{X}_\ell$ *on* $(\{0,1\}^n)^\ell$ *is such that* g *out of the* ℓ *blocks are good, i.e., are independently sampled* (n, k)*-sources. The remaining* $\ell - g$ *bad blocks may only depend on the blocks with a smaller index than it.*

Similarly as for NOSF sources, if $k = n$ we call $\mathbf{X}$ a *uniform* (g, ℓ)*-fiSHELA* source. Our final class of sources that we look at, almost Chor-Goldreich sources, share the motivation from SHELA sources that the adversary cannot see into the future. Rather than forcing the adversary to have its blocks only depend on blocks in the past (those with smaller indices), almost CG sources require that good blocks have some entropy conditioned on all blocks that came before them (i.e., bad blocks cannot expose all of the entropy of future good blocks).

Definition 1.5 (Almost CG sources, [CG88, DMOZ23])**.** *We define a* (g, ℓ, k)-almost CG *source* $\mathbf{X} = \mathbf{X}_1, \ldots, \mathbf{X}_\ell$ *to be a distribution on* $(\{0,1\}^n)^\ell$ *such that there exists a set of good indices* $\mathcal{G} \subseteq [\ell]$ *of size at least* g *for which* $H_\infty(\mathbf{X}_i \mid \mathbf{X}_1 = x_1, \ldots, \mathbf{X}_{i-1} = x_{i-1}) \geq k$ *for all* $i \in \mathcal{G}$ *and all prefixes* $x_1, \ldots, x_{i-1}$.

As before, if $k = n$ then we say that $\mathbf{X}$ is a uniform (g, ℓ)-almost CG source. A convenient fact that we later show and will rely on is that uniform (g, ℓ)-almost CG sources and uniform (g, ℓ)-fiSHELA sources are the same set of sources. The impossibility of extraction from these sources naturally raises the question of whether, with regards to randomness condensing, there is a distinction between them. The authors of [DMOZ23] constructed explicit condensers with a constant entropy gap for a subclass of almost CG sources, termed suffix-friendly CG sources, but in general achieved no non-trivial condensing.

1.3 Our Results

Here, we answer open questions for each of these classes of randomness sources and provide four main results. First, we show that condensing above a certain rate from uniform NOSF, SHELA, and almost CG sources when $g \leq \ell/2$ is impossible.

Theorem 1 (Theorem 5.2, restated)**.** *For any function* $f : (\{0,1\}^n)^\ell \to \{0,1\}^t$ *and for all* $\varepsilon > 0$ *there exists a constant* $\delta > 0$ *and uniform* (g, ℓ)*-NOSF/SHELA/almost CG source* $\mathbf{X}$ *with* $g \leq \ell/2$ *such that* $H^\varepsilon_\infty(f(\mathbf{X})) \leq \frac{1}{\lfloor \ell/g \rfloor} t + \delta$.

If we restrict ourselves to the cases where g divides ℓ, then this theorem exactly reduces to:

Corollary 1. *For any function* $f : (\{0,1\}^n)^\ell \to \{0,1\}^t$ *and for all* $\varepsilon > 0$ *there exists a constant* $\delta > 0$ *and uniform* (g, ℓ)*-NOSF/SHELA/almost CG source* $\mathbf{X}$ *(where* g *divides* ℓ*) such that* $H^\varepsilon_\infty(f(\mathbf{X})) \leq \frac{g}{\ell} t + \delta$.

Surprisingly, while we initially attempted to show these results for uniform NOSF sources, all of our lower bound results for condensing from uniform NOSF sources in regime of $g \leq \ell/2$ also hold for uniform

SHELA and uniform almost CG sources! We achieve this by constructing, for every such f, a uniform SHELA/almost CG source where f fails to condense.

Second, we demonstrate a condensing lower bound for uniform NOSF sources in the regime where $g > \ell/2$, specifically for uniform $(2,3)$-NOSF sources.

Theorem 2 (Theorem 5.8, restated)**.** *For any $f : (\{0,1\}^n)^3 \to \{0,1\}^t$ and $0 < \varepsilon < \frac{1}{4}$, there exists a constant $\delta > 0$ and uniform $(2,3)$-NOSF source $\mathbf{X}$ such that $H^{\varepsilon}_{\infty}(f(\mathbf{X})) < \frac{2}{3}t + \delta$.*

Third, we show a separation between NOSF and SHELA sources by exhibiting the existence of a condenser for uniform $(2,3)$-SHELA sources.

Theorem 3 (Informal version of Theorem 6.1)**.** *There exists a condenser $\mathsf{Cond} : (\{0,1\}^n)^3 \to \{0,1\}^t$ for any uniform $(2,3)$-SHELA source $\mathbf{X}$ so that $H^{\varepsilon}_{\infty}(\mathsf{Cond}(\mathbf{X})) \geq t - \Delta$ where $t = O(n - \log(1/\varepsilon))$ and the entropy gap is $\Delta = O(\log(t/\varepsilon))$.*

We also explicitly construct such a condenser, with asymptotically the same min entropy gap as the probabilistic construction:

Theorem 4 (Informal version of Theorem 6.10)**.** *For any $\varepsilon > 0$, we can explicitly construct a condenser $\mathsf{Cond} : (\{0,1\}^n)^3 \to \{0,1\}^t$ with $t = \frac{n}{16}$ such that for any uniform $(2,3)$-SHELA source $\mathbf{X}$, $H^{\varepsilon}_{\infty}(\mathsf{Cond}(\mathbf{X})) \geq t - O(\log(t/\varepsilon))$.*

We briefly mention that a key ingredient in the above condenser is a new type of seeded extractor, that we call an *output-light* seeded extractor. Informally, such a seeded extractor $\mathsf{Ext} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ satisfies the additional guarantee that each output $z \in \{0,1\}^m$ can be produced only by a bounded number of inputs $x \in \{0,1\}^n$. (See Definition 6.2 for the formal definition of such seeded extractors.) We explicitly construct such an output-light seeded extractor (Theorem 6.11) to obtain the above condenser result.

Lastly, we show that uniform $(2,3)$-SHELA sources form another class of sources for which condensing is possible but extraction is not.

Theorem 5 (Informal version of Theorem 5.12)**.** *For any $f : (\{0,1\}^n)^3 \to \{0,1\}$ there exists a uniform $(2,3)$-SHELA/almost CG source $\mathbf{X}$ such that $|f(\mathbf{X}) - \mathbf{U}_1| > 0.08$.*

We now delve into the results of prior work and how our work builds upon them.

1.4 Comparison to Previous Work

NOSF sources

As aforementioned, many papers have explored whether it is possible to extract from NOSF, SHELA, and CG sources. We can trace back study of extracting from NOBF sources the furthest, since the seminal work of Kahn, Kalai, and Linial in [KKL88] demonstrated that it is not possible to extract from uniform (g,ℓ)-NOBF when the number of bad blocks is $b = \ell - g = \Omega(\ell/\log \ell)$. The follow up work of [BKKKL92] then extended this extraction impossibility result to uniform (g,ℓ)-NOSF sources again with $b = \Omega(\ell/\log \ell)$.

To attempt to match these lower bounds on extraction, resilient functions, introduced by [BL85], have yielded the current best results. The resilient function of Ajtai and Linial in [AL93] and its explicit versions constructed by [CZ19, Mek17] achieve extractors for uniform (g,ℓ)-NOSF sources when $b = O(\ell/\log^2 \ell)$, leaving a $1/\log \ell$ gap between the lower and upper bounds.

In the context of our work, these results mean that for every ℓ there exists a g large enough so that extraction, and consequently condensing, is possible. Nevertheless, this still leaves open whether condensing is possible for all cases when $b = \omega(\ell/\log^2 \ell)$.

Also along the lines of NOSF sources, the work of [KN23] explores what they call extracting multi-mergers, which we may consider as extractors for uniform NOSF sources. Their main result shows that seeded extracting mergers, multimergers in the case of uniform $(1, \ell)$-NOSF sources, require $\log(n)$ seed length as is the case for seeded min-entropy extractors. For seedless extracting multimergers, their result implies extracting from uniform $(2, 3)$-NOSF sources is impossible.

Our contribution: In this work, we extend the lower bounds of extracting from NOSF sources to condensing from NOSF sources. Our Theorem 1 shows condensing lower bounds in the case that $g \leq \ell/2$ and Theorem 2 shows that condensing from uniform $(2, 3)$-NOSF sources is impossible. As NOSF sources are a special case of NOBF sources, all our lower bounds also apply in that setting.

SHELA sources

In [AORSV20], the authors were able to construct somewhere-extractors, functions that have a uniform NOSF as output, for SHELA sources and show that for any $\gamma \in (0, 1)$ there exists an ℓ such that extraction is not possible for $(\gamma\ell, \ell)$-SHELA sources. They conjectured that condensing is not possible for any uniform SHELA source (with parameters not meeting Ajtai-Linial, of course).

Our contribution: Our Theorem 1 proves their conjecture (mostly) true for the regime of $g \leq \ell/2$; however, we prove their conjecture false for uniform $(2, 3)$-SHELA sources in Theorem 3, demonstrating a possible threshold at $g = \ell/2$ for condensing from SHELA sources. As mentioned above, we in fact construct an explicit condenser for uniform $(2, 3)$-SHELA sources, using a new kind of seeded extractor that we introduce. We also prove that one cannot hope to extract from $(2, 3)$-SHELA sources.

Almost CG sources

For almost CG sources, [GP20] showed that errorless condensing is impossible. In contrast, [DMOZ23] proved several possibility results regarding condensing with error for CG sources. Their results are stated as assuming size of each block is very small (almost constant) and they have large number of such blocks. For a lot of cases, this setting is arguably harder to condense from as compared to the setting where you are given small number of large sized blocks since given a large sized blocks, we can partition them and get many small sized blocks. The most significant of their results for us is that they constructed an explicit deterministic condenser with exponentially small error using the constant-degree lossless expanders given by [CRVW02] for a subclass of almost CG sources they termed *suffix-friendly almost CG* sources. These suffix-friendly almost CG sources are like our (g, ℓ, k)-almost CG sources except with the requirement that the g good blocks be well-distributed among the ℓ total blocks. Their construction obtains a constant entropy gap with for suffix-friendly almost CG sources when the fraction of bad blocks $b = \ell - g$ is quite small $b \leq 10^{-8}\ell$, and otherwise they are not able to condense when $b \geq \ell/2$ or without the suffix-friendliness requirement.

Our contribution: We answer their open question of whether suffix-friendliness is required with our condensing lower bounds in Theorem 1 for the case of uniform (g, ℓ)-almost CG with $g \leq \ell/2$. Additionally, we show that suffix-friendliness is not required for condensing from uniform $(2, 3)$-almost CG sources in Theorem 3.

2 Proof Overview

2.1 Impossibility Results

In this subsection, we will go over the main techniques we used in proving the condensing impossibility result in the case that $g \leq \ell/2$ (Theorem 2.1), the condensing impossibility result for uniform $(2,3)$-NOSF sources (Theorem 2.2), and the extracting impossibility result for uniform $(2,3)$-SHELA sources (Theorem 2.3).

2.1.1 Impossibility of Condensing from Uniform (g,ℓ)-NOSF/SHELA/Almost CG Sources for $g \leq \ell/2$

Conceptually, our first result says that when the number of good blocks g is not more than half of the total number of blocks ℓ, then condensing beyond rate $\frac{1}{\lfloor \ell/g \rfloor}$ is impossible. Formally, we will prove the following statement.

Theorem 2.1 (Theorem 5.2, restated)**.** *For any function $f : (\{0,1\}^n)^\ell \to \{0,1\}^t$ and any $\varepsilon > 0$ there exists a $\delta > 0$ and uniform (g,ℓ)-NOSF/SHELA/almost CG source $\mathbf{X}$ such that $H^\varepsilon_\infty(f(\mathbf{X})) \leq \frac{1}{\lfloor \ell/g \rfloor} t + \delta$.*

The steps we take to achieve the result of Theorem 2.1 are, broadly, as follows:

1. We show that for a constant $0 < c_1 < 1$ and a bipartite graph $H = (U, V)$ with $|U| = N$ left vertices, $|V| = T$ right vertices, and $\deg(u) \geq T^\delta$ for all $u \in U$ and some $\delta > 0$, then there exists a subset $D \subseteq V$ of right vertices of size at most $\frac{c_1}{1-c_1} T^{1-\delta}$ where $|\mathcal{N}(D)| \geq c_1 N$.

2. We use this result to show that if condensing from a uniform (g,ℓ)-SHELA source is impossible, then condensing from a uniform $(g,\ell+g)$-SHELA source is impossible. Applying the base case of uniform $(1,1)$-SHELA sources (which is just $\mathbf{U}_n$) gives us that condensing above rate $\frac{1}{\ell}$ is impossible for uniform $(1,\ell)$-SHELA sources.

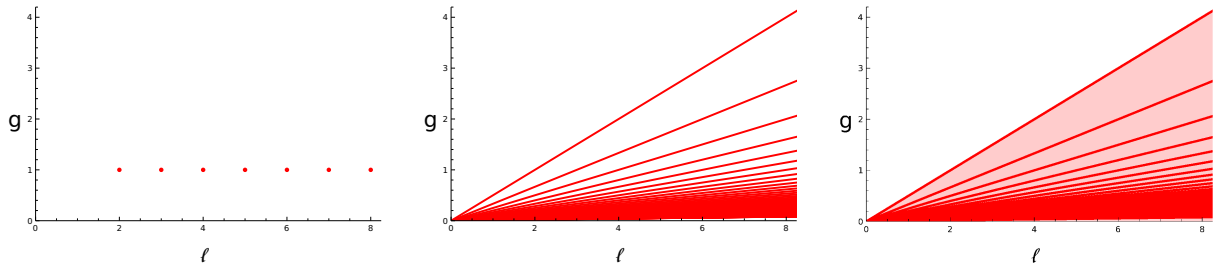
3. By simply chunking blocks together, we extend this result to uniform (g,ℓ) SHELA sources with rate $\frac{g}{\ell} = \frac{1}{c}$ for some $c \in \mathbb{N}$.

4. Finally, we generalize our result to all uniform (g,ℓ)-SHELA sources with $g \leq \ell/2$ to achieve Theorem 2.1.

If we consider the g and ℓ of a uniform (g,ℓ)-SHELA source as coordinates on a plane, then the progress of our results is actually quite visualizable. We delineate our process visually in this way in Figure 1.

Now, let us go through these four steps.

Step 1 Conveniently, a simple greedy algorithm that collects right vertices in V with highest degree first and adds them to D with the stopping condition that $|\mathcal{N}(D)| \geq c_1 N$ achieves our goal. Initially, we attempted to create D that would completely cover the left vertices in U. This strategy fails because after many steps of this greedy algorithm, the last few right vertices may have very low degree, forcing the cardinality of D to increase linearly in the number of left vertices we are covering. Our analysis is based on the observation that the initial vertices we collect must have very high degree. Thus, we can stop this greedy algorithm early once only a fraction $c_1 N$ of left vertices are covered without blowing up the size of D beyond $\frac{c_1}{1-c_1} T^{1-\delta}$.



(a) We show in Step 1 that condensing above rate $\frac{1}{\ell}$ is impossible for $(1, \ell)$-SHELA sources.

(b) Next, we show in Step 2 that condensing above rate $\frac{g}{\ell}$ is impossible for $(1, \ell)$-SHELA sources when $\ell = cg$ for some $c \in \mathbb{N}$.

(c) Finally, we prove in Step 3 the result of Theorem 2.1

Figure 1: We illustrate the progress of how we ultimately prove our Theorem 2.1. Solid red lines and points indicate strict condensing impossibility results (i.e., condensing above the initial rate is impossible), and shaded in regions indicate that condensing above the rate of the bounding line is impossible.

Step 2 This result may be thought of as an inductive step that constructs a uniform $(g, \ell + g)$-SHELA source $\mathbf{X} = \mathbf{X}_1, \ldots, \mathbf{X}_{\ell+g}$ from left to right for which f cannot condense from. We look at the behavior of f in its first g coordinates. In one case, we observe that if f has large support no matter the values input in its first g coordinates, then we can let the first g blocks $\mathbf{X}_1, \ldots, \mathbf{X}_g$ of $\mathbf{X}$ be the g good blocks and construct an explicit adversary in the last ℓ blocks. We construct this adversary by considering the bipartite graph $H = (U = (\{0,1\}^n)^g, V = \{0,1\}^t)$ where edges (u, v) are included if v is still in the support of f after fixing the first g coordinates to u. Our dominating set argument then gives us an adversary in $\mathbf{X}_{g+1}, \ldots, \mathbf{X}_\ell$ that restricts the support size of f with constant probability, which bounds condensing.

Otherwise, there must exist some fixed setting $x_1, \ldots, x_g$ of the first g coordinates of f that significantly limits its support size. Fixing these coordinates of f yields a function h in ℓ coordinates. By assumption, we can build a uniform (g, ℓ)-SHELA source $\mathbf{Y} = \mathbf{Y}_1, \ldots, \mathbf{Y}_\ell$ from which h cannot condense. Our final source is then simply the concatenation $\mathbf{X} = x_1, \ldots, x_g, \mathbf{Y}_1, \ldots, \mathbf{Y}_\ell$.

Finally, we consider the base case of uniform $(1, 1)$-SHELA sources to induct off of and yield Item 2. Conceptually, what is occurring in this inductive process is that, for a function f, we are creating a uniform $(1, \ell)$-fiSHELA source by placing blocks left to right. We begin by placing adversarial blocks with fixed values that limit the support of f — the second case we talk about here — and then when the support of f cannot be significantly limited by a single fixed value, we use our first case to place the uniform good block and fill the rest of the blocks to the right with an adversary that depends on that good block.

We apply this inductive logic to the base case that condensing is impossible for uniform $(1, 1)$-SHELA sources since they are already $\mathbf{U}_n$. We depict this construction in Figure 2.

Step 3 In this step, we make show that if we cannot condense from uniform (g, ℓ)-SHELA sources, then we cannot condense from uniform $(cg, c\ell)$-SHELA sources as well. We do this by observing that a uniform (g, ℓ)-SHELA source can be converted to a uniform $(cg, c\ell)$-SHELA source by splitting up each block into c sub-blocks. Thus, for a function $f : (\{0,1\}^n)^{c\ell} \to \{0,1\}^t$ on uniform $(cg, c\ell)$-SHELA sources, we use our ability to construct a uniform (g, ℓ)-SHELA source $\mathbf{X}$ with block length nc so that f cannot condense above rate $\frac{g}{\ell} = \frac{cg}{c\ell}$ from $\mathbf{X}$.

Step 4 Our final step is a generalization of the previous one. When $g \leq \ell/2$, we can divide ℓ by g as $\ell = cg + r$ where $c > 0$ and $r < g$. Then to show that no function f on uniform (g, ℓ)-SHELA sources can condense above rate $\frac{1}{c}$, we use the fact that an arbitrary uniform $(1, c)$-SHELA source $\mathbf{X}$ cannot be condensed above rate $\frac{1}{c} = \frac{1}{\lfloor \ell/g \rfloor}$. If such a condenser f for uniform (g, ℓ)-SHELA sources did exist, then dividing the blocks of $\mathbf{X}$ as evenly as possible (i.e., splitting up the first r blocks of $\mathbf{X}$ into $g + 1$ blocks and the last $c - r$ blocks into g blocks) would yield a uniform (g, ℓ)-SHELA source that we could pass to f. This gives us a contradiction as f would condense $\mathbf{X}$ above rate $\frac{1}{c}$.

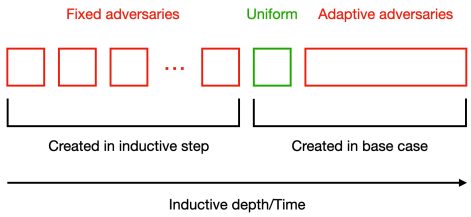


Figure 2: We show how we create a uniform $(1, \ell)$-fiSHELA adversary by placing fixed adversarial blocks via inductive steps first, then placing the one good block, and finally placing the remaining adaptive adversarial blocks last. We may also think of the inductive depth as time, so we clearly see that the adaptive adversarial blocks only depend on blocks in the past.

2.1.2 Impossibility of Condensing from Uniform $(2, 3)$-NOSF sources

A key ingredient in our impossibility result for the regime $g \leq \ell/2$ are the reductions we are able to perform between different parameters. In fact, Theorem 2.1 ultimately depends on the fact that condensing is not possible from a uniform $(1, 1)$-SHELA source. On the other hand, when $g > \ell/2$, we find such inductive arguments do not present themselves naturally, and instead we focus on the case when $g = 2$ and $\ell = 3$. For these parameters, we are able to show that a condenser for uniform $(2, 3)$-NOSF sources above rate $\frac{2}{3}$ does not exist.

Theorem 2.2 (Theorem 5.8, restated)**.** *For any $f : (\{0,1\}^n)^3 \to \{0,1\}^t$ and $0 < \varepsilon < \frac{1}{4}$, there exists a uniform $(2, 3)$-NOSF source $\mathbf{X}$ and a constant $\delta > 0$ such that $H^{\varepsilon}_{\infty}(f(\mathbf{X})) \leq \frac{2}{3}t + \delta$*

It is instructive to compare our construction of a uniform $(2, 3)$-NOSF source $\mathbf{X}$ for a given function f to our construction of uniform $(1, \ell)$-SHELA sources in Item 2. There, to create the appropriate adversary, we first placed fixed adversarial blocks, then placed our single uniform block, and finally placed our adaptive adversarial blocks that could depend on the uniform block. We were forced to place fixed and adaptive adversarial blocks in this order since, as we recall, the bad blocks blocks in a fiSHELA source may only depend on the blocks that come before it. This is no longer true for NOSF sources, giving us flexibility that we will use to our advantage.

Once we notice this difference, our construction of the appropriate uniform $(2, 3)$-NOSF source $\mathbf{X} = \mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3$ is not so unexpected. For ease of notation, let $T = 2^t$. We utilize three cases when constructing $\mathbf{X}$. First, we consider when there exists some value x_1 in the first coordinate that, with constant probability over the two other good blocks placed in the second and third coordinates, will limit the support of f to a linear fraction of $T^{2/3}$. As when constructing our uniform $(1, \ell)$-SHELA source before, we place a constant bad block in the first coordinate, setting $\mathbf{X}_1 = x_1$ to achieve our goal.

Second, we consider when a large fraction of fixings of $\mathbf{X}_1$ and $\mathbf{X}_2$ do not decrease the support of f below $O(T^{1/3})$. In this case, we can use a similar bipartite graph covering argument, but this time generalized to colored bipartite graphs, as before to generate an adversary in the third coordinate for $\mathbf{X}_3$ that limits the support of f to $O(T^{2/3})$ with constant probability.

Third, we take an analogous version of or previous case but for the first and third coordinates. In other words, that a large fraction of fixings of $\mathbf{X}_1$ and $\mathbf{X}_3$ do not decrease the support of f below $O(T^{1/3})$. We again use the same colored bipartite graph covering argument to create an adversary in the second

coordinate for $\mathbf{X}_2$ that depends on the first and third coordinates. Notice that this is exactly when we break the requirements of a SHELA source since $\mathbf{X}_2$ may depend on the block that comes after it!

Finally, we show that one of these three cases must always occur by showing that if our second and third cases did not occur, then our first case must have occurred. We again do this by a colored bipartite graph covering argument.

2.1.3 Impossibility of Extracting from Uniform $(2,3)$-SHELA sources

Overall, our method to show that extraction is not possible from uniform $(2,3)$-SHELA sources is conceptually similar to our strategy for proving condensing impossibility results when $g \leq \ell/2$. The main difference here is that we manage to fix the output of a function on a uniform $(2,3)$-SHELA source to a single value, not a small subset of values with constant probability. This strategy enables us to prove that no function can extract even one uniform bit.

Theorem 2.3 (Theorem 5.12, restated)**.** *For any $f : (\{0,1\}^n)^3 \to \{0,1\}$ there exists a uniform $(2,3)$-SHELA source $\mathbf{X}$ such that $|f(\mathbf{X}) - \mathbf{U}_1| \geq 0.08$.*

To prove this theorem, we consider a handful of cases as before that allow us to place our bad block in the first, second, or third coordinate. Our first two cases consider when there are a constant fraction of settings of $\mathbf{X}_1$ and $\mathbf{X}_2$ that allow f to be either 0 or 1. In these cases, we simply choose a value $x_3 = a(\mathbf{X}_1, \mathbf{X}_2)$ adversarially to force f to take on 0 or 1 with constant probability.

Otherwise, we find ourselves in a case where there are many pairs of (x_1, x_2) that force f to take on 0 or 1 regardless of the value of $\mathbf{X}_3$, so we place our adversary in the first or second coordinate. To decide in which coordinate to place our adversary, we consider a bipartite graph on the $[N]$ possible values of $\mathbf{X}_1$ and the $[N]$ possible values of $\mathbf{X}_2$ with edges labeled by 0 or 1 if a pair (x_1, x_2) fixes the output of f to that value regardless of $\mathbf{X}_3$. We restrict ourselves to looking at left vertices of high degree and notice that either one of these left vertices has edges only labeled by 0's or 1's, in which case we can place our bad block in the first coordinate and fix it to this value, or all left vertices have at least one edge labeled by 0 and one edge labeled by a 1. In this latter case we can place our adversary in the second coordinate to force f to output 0 for all of these high degree vertices.

In all cases, we manage to restrict f to a single output value with constant probability, preventing extraction beyond constant error.

2.2 Possibility Results

In this subsection, we will present our probabilistic as well as explicit constructions of condensers for uniform $(2,3)$-SHELA sources.

2.2.1 Probabilistic construction

Before we dive into the actual proof, it is instructive to see why does a random function fail to be a condenser for uniform $(2,3)$-SHELA sources. For a random function $f : \{0,1\}^{3n} \to \{0,1\}^t$, with high probability over $x_1, x_2 \in \{0,1\}^n$, we have $|f(x_1, x_2, \cdot)| = 2^t$. Hence, if the adversary is in position 3, then it can depend on x_1 and x_2 to ensure the output of f always lies in a small set. To overcome this, one can consider restricting the number of choices adversary has when it is in position 3. This intuition indeed works out and we give further details:

Theorem 2.4 (Theorem 6.1, restated)**.** *There exists a condenser* $\mathsf{Cond} : (\{0,1\}^n)^3 \to \{0,1\}^t$ *where* $t = n + \log(n) - \log(1/\varepsilon)$ *and for any uniform* $(2,3)$*-SHELA source* $\mathbf{X}$*,* $H^{\varepsilon}_{\infty}(\mathsf{Cond}(\mathbf{X})) \geq t - \log(t) - 4\log(1/\varepsilon) - O(1)$.

We generate Cond by a random process. In this process, we sample sets $S_{i,j} \subset \{0,1\}^t$ for $i,j \in [N]$ where each $z \in \{0,1\}^t$ is included in $S_{i,j}$ with very small probability p. To simplify matters, we set $t = n$. Based on these sets, define Cond as follows: on input (x_1, x_2, x_3), use x_3 to index and output an element from S_{x_1,x_2}. We claim that with high probability such a Cond will be a condenser for uniform $(2,3)$-SHELA sources. Let $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3)$ be such an arbitrary source in this family. If the adversary controls position 3 in $\mathbf{X}$, then Cond fails to condense if and only if the adversary can find a small set of elements $D \subset \{0,1\}^t$ such that at least ε fraction of the sets $S_{i,j}$ have an element that lies in D. This implies there exists an element $z \in \{0,1\}^t$ that appears in at least $\frac{\varepsilon N^2}{|D|}$ such sets $S_{i,j}$. By the random process, for each $z \in \{0,1\}^t$ and set $S_{i,j}$, we include $z \in S_{i,j}$ independently with probability p. So, the expected number of sets that contain z is pN^2. We set p to be small enough so that pN^2 is much smaller than $\frac{\varepsilon N^2}{|D|}$. A Chernoff bound based argument combined with a union bound lets us argue that with high probability, no z can appear in so many sets and hence, Cond will be a condenser for sources where adversary controls position 3. Consider sources $\mathbf{X}$ where the adversary controls position 1 or position 2. Then, the adversary restricts Cond to some N sets out of the N^2 sets and the distribution $\mathsf{Cond}(\mathbf{X})$ becomes equivalent to the process of randomly picking a set out of these fixed N sets and outputting a random element from it. In fact, as these N sets were initially randomly sampled, we are able to show that $\mathsf{Cond}(\mathbf{X})$ will be close to the uniform distribution. Thus, $\mathsf{Cond}(\mathbf{X})$ will indeed be a condenser for uniform $(2,3)$-SHELA sources.

Careful examination of the above argument reveals that we are only using a very small prefix of x_3. This is because sets $S_{i,j}$ are small sets of size about $pT = pN \leq o(N)$ while we have access to n bits of x_3. It turns out, the above probabilistic construction actually yields a seeded extractor[1] $\mathsf{sExt} : \{0,1\}^{2n} \times \{0,1\}^{O(\log n/\varepsilon)} \to \{0,1\}^t$. This extractor works for very small min entropy but for our purposes, it suffices for it to handle min entropy n (out of $2n$). Moreover, it is "output-light", meaning for every $z \in \{0,1\}^t$, $\left|\{x \in \{0,1\}^{2n} : \exists y \in \{0,1\}^{O(\log n/\varepsilon)}(\mathsf{sExt}(x,y) = z)\}\right|$ is small (see Definition 6.2 for formal definition). Given such an extractor, one can easily construct the required condenser and this is indeed the proof strategy we use in Section 6.

2.2.2 An Explicit Condenser for Uniform (2,3)-SHELA Sources

We will prove that we can get explicit condensers for Uniform $(2,3)$ SHELA sources.

Theorem 2.5 (Informal version of Theorem 6.10)**.** *For any* $\varepsilon > 0$*, we can explicitly construct a condenser* $\mathsf{Cond} : (\{0,1\}^n)^3 \to \{0,1\}^t$ *with* $t = \frac{n}{16}$ *such that for any uniform* $(2,3)$*-SHELA source* $\mathbf{X}$*,* $H^{\varepsilon}_{\infty}(\mathsf{Cond}(\mathbf{X})) \geq t - O(\log(t/\varepsilon))$.

As discussed above, we observe that for condensing from uniform $(2,3)$-SHELA sources, it suffices to construct an output-light seeded extractor (Definition 6.2) for somewhere-random sources, i.e., sources of the type $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2)$ where both $\mathbf{X}_1, \mathbf{X}_2$ are over n bits and at least one of $\mathbf{X}_1$ or $\mathbf{X}_2$ is guaranteed to be uniform. We explicitly construct such a seeded extractor and get our explicit condenser.

[1] see Definition 4.6 for a definition of seeded extractors

2.2.3 An Explicit Output-Light Seeded Extractor for Somewhere-Random Sources

We here describe an output-light seeded extractor sExt for somewhere random sources with seed length $d = O(\log(n/\varepsilon))$ and output length $m = O(n)$. Moreover it will be very "output-light", i.e., for every $z \in \{0,1\}^m$, $\left|\{x \in \{0,1\}^{2n} : \exists y \in \{0,1\}^{O(\log n/\varepsilon)}(\mathsf{sExt}(x,y) = z)\}\right| \leq 2^{2n-m+d}$. This suffices to get the desired explicit condenser.

We transform the input distribution $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2)$ into a block source $\mathbf{Y} = (\mathbf{Y}_1, \mathbf{Y}_2)$. This means we want that for all fixings of $\mathbf{Y}_1 = Y_1$, $\mathbf{Y}_2$ will have high min-entropy. For some choice of n_1, we let $\mathbf{Y}_1$ be the source obtained by concatenating the first n_1 bits of $\mathbf{X}_1$ and the first n_1 bits of $\mathbf{X}_2$. Then, as at least one of $\mathbf{X}_1$ or $\mathbf{X}_2$ is uniform, $H_\infty(\mathbf{Y}_1) \geq n_1$. We set $n_2 = n - n_1$ and let $\mathbf{Y}_2$ be the source obtained by picking bits at indices $n_1 + 1, \ldots, n$ from both $\mathbf{X}_1$ and $\mathbf{X}_2$ and concatenating them. Then by min entropy chain rule, with high probability, $\mathbf{Y}_2$ will have high min entropy conditioned on fixing of $\mathbf{Y}_1$.

We then take a good seeded extractor sExt and use our input seed S to get a large number of random bits from $\mathbf{Y}_2$, i.e., $\mathbf{R}_2 = \mathsf{sExt}(\mathbf{Y}_2, S)$. We then compute the inner product of $\mathbf{R}_2$ and $\mathbf{Y}_1$ over a large field. We are allowed to use $\mathbf{R}_2$ as a source of randomness for $\mathbf{Y}_1$ because $\mathbf{Y}$ is a block source. Moreover, as inner product is a good two source extractor (Theorem 4.8), the resultant output will be uniform. Call this m bit output (for say $m = n_1/4$) distribution $\mathbf{R}_1$.

In this inner product, we bypass the case of $\mathbf{R}_2 = 0$ (0 of the field) by artificially changing last bit of $\mathbf{R}_2$ to be 1. For a fixed seed S, a fixing R_1 of $\mathbf{R}_1$, and a fixing Y_2 of $\mathbf{Y}_2$, the number of fixings of $\mathbf{Y}_1$ that map to R_1 are about $2^{n_1 - m}$. This is because given $n_1 - m$ bits in addition to R_1, and having access to Y_2, we can invert this inner product operation. Thus, for fixed seed S and output R_1, the number of inputs that map to it are 2^{2n-m} and consequently, for fixed output R_1, the number of inputs that map to it are 2^{2n-m+d} where d is the seed length we require. The seed length we require is $O(\log(n/\varepsilon))$ and, hence, we get a sufficiently output-light seeded extractor as desired.

3 Open Questions

Our research simultaneously breaks ground in several new directions while raising a host of open questions. Now that we have obtained both lower and upper bounds for condensing from uniform NOSF, SHELA, and almost CG sources for various settings of the g and ℓ, exploring the terrain for new settings of parameters is natural. A few immediate open questions brought up by our work are:

1. Is our result of Theorem 5.2 tight? That is, when we write $\ell = cg + r$ for $c, r \in \mathbb{N}$ and we have $r > 0$, then is it possible to condense up to rate $\frac{1}{c}$ or just to $\frac{g}{\ell}$?

2. Can we generalize our impossibility result of Theorem 5.8? In particular, is condensing from uniform (g,ℓ)-NOSF sources generally impossible when $g > \ell/2$? It would be exciting to see whether this result can extend all the way to that of [BKKKL92].

3. Can we generalize our condenser in Theorem 6.1? In general, we are unsure what condensing looks like for SHELA sources when $g > \ell/2$.

4. Can we improve and generalize our construction of the output-light seeded extractor of Theorem 6.10?

4 Preliminaries

We will generally denote distributions or sources in a bold font, such as $\mathbf{X}$, and reserve $\mathbf{U}_t$ to be the uniform distribution on t bits. When these sources are actually a sequence of sources, we use subscripts to denote blocks of that source as $\mathbf{X} = \mathbf{X}_1, \ldots, \mathbf{X}_\ell$. In addition, since we often consider binary strings of length n and t, we let $N = 2^n$ and $T = 2^t$. Often it is convenient to consider strings as labels, in which case we use the notation $[N] = \{1, 2, \ldots, N\}$.

4.1 Basic probability lemmas

Here, we first state a few basic probability facts that will be useful to us throughout. Our first one is a direct reverse Markov style inquality.

Claim 4.1 (Reverse Markov)**.** *Let $\mathbf{X}$ be a random variable taking values in $[0, 1]$. Then, for $0 \leq d < \mathbb{E}[X]$, it holds that*

$$\Pr[\mathbf{X} > d] \geq \frac{\mathbb{E}[\mathbf{X}] - d}{1 - d}$$

Proof. Let $\mathbf{Y} = 1 - \mathbf{X}$. Then, by Markov's inequality,

$$\Pr[\mathbf{Y} \geq 1 - d] \leq \frac{\mathbb{E}[\mathbf{Y}]}{1 - d} = \frac{1 - \mathbb{E}[\mathbf{X}]}{1 - d}$$

So,

$$\Pr[\mathbf{X} > d] = \Pr[\mathbf{Y} < 1 - d] = 1 - \frac{1 - \mathbb{E}[\mathbf{X}]}{1 - d} = \frac{\mathbb{E}[\mathbf{X}] - d}{1 - d}$$

□

We will also use a few versions of the classic Chernoff bound.

Claim 4.2 (Chernoff Bound)**.** *Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be independent random variables taking values in $\{0, 1\}$. Let $\mathbf{X} = \sum_i \mathbf{X}_i$. Let $\mu = \mathbb{E}[\mathbf{X}]$. Then, for all $\delta \geq 0$, the following holds:*

$$\Pr[\mathbf{X} \geq (1 + \delta)\mu] \leq \left(\frac{e^\delta}{(1 + \delta)^{1+\delta}} \right)^\mu$$

We will sometimes use the following two weaker versions of this:

1. $$\Pr[\mathbf{X} \geq (1 + \delta)\mu] \leq \left(\frac{e}{(1 + \delta)} \right)^{(1+\delta)\mu} = \exp(-\mu(1 + \delta) \log((1 + \delta)/e)$$

2. $$\Pr[\mathbf{X} \geq (1 + \delta)\mu] \leq \exp(-\delta^2 \mu/(2 + \delta))$$

We also utilize the following version of the Chernoff bound for $0 < \delta < 1$:

$$\Pr[|\mathbf{X} - \mu| \geq \delta\mu] \leq \exp(-\delta^2 \mu/3)$$

Several of our impossibility results rely on a simple TV distance bound.

Claim 4.3 (TV distance lower bound)**.** *Let $\mathbf{X} \sim \{0,1\}^t$ and $S \subset \{0,1\}^t$ be such that $\Pr_{x\sim\mathbf{X}}[x \in S] \geq p$. Then, for $0 < \varepsilon < p$, it holds that $H^{\varepsilon}_{\infty}(\mathbf{X}) \leq \log\left(\frac{|S|}{p-\varepsilon}\right)$.*

Proof. Let $k = \log\left(\frac{|S|}{p-\varepsilon}\right)$. Let $\mathbf{Y} \sim \{0,1\}^t$ be an arbitrary distribution with $H_{\infty}(\mathbf{Y}) \geq k$. By the min entropy condition, for all $s \in S$, it holds that $\Pr[\mathbf{Y} = s] \leq 2^{-k}$. Hence,

$$|\mathbf{X} - \mathbf{Y}| \geq \Pr_{x\in\mathbf{X}}[x \in S] - \Pr_{y\in\mathbf{Y}}[y \in S] = p - 2^{-k} \cdot |S| = \varepsilon$$

□

We will utilize the very useful min entropy chain rule in our constructions.

Lemma 4.4 (Min-entropy chain rule)**.** *For any random variables $\mathbf{X} \sim X$ and $\mathbf{Y} \sim Y$ and $\varepsilon > 0$,*

$$\Pr_{y\sim\mathbf{Y}}[H_{\infty}(\mathbf{X} \mid \mathbf{Y} = y) \geq H_{\infty}(\mathbf{X}) - \log|\mathit{support}(\mathbf{Y})| - \log(1/\varepsilon)] \geq 1 - \varepsilon.$$

Lastly, we will later utilize a consequence of upper bounds on smooth min-entropy.

Claim 4.5 (Lemma 8.8 from [Zuc07])**.** *Let $\mathbf{X} \sim \{0,1\}^t$ be such that $H^{\varepsilon}_{\infty}(\mathbf{X}) < k$. Then, there exists $D \subset \mathit{support}(\mathbf{X})$ such that $|D| < 2^k$ and $\Pr[\mathbf{X} \in D] \geq \varepsilon$.*

4.2 Extractors

Let $\mathbf{A} \approx_{\varepsilon} \mathbf{B}$ mean that $\mathbf{A}$ and $\mathbf{B}$ are ε close in statistical distance. We recall the definition of a seeded extractor.

Definition 4.6. *A (k,ε)-seeded extractor $\mathsf{Ext} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ satisfies the following: for every (n,k)-source $\mathbf{X}$, and every $\mathbf{Y} = \mathbf{U}_d$,*

$$\mathsf{Ext}(\mathbf{X}, \mathbf{Y}) \approx_{\varepsilon} \mathbf{U}_m.$$

We call d the seed length *of* Ext*. Ext is called strong if*

$$\mathsf{Ext}(\mathbf{X}, \mathbf{Y}), \mathbf{Y} \approx_{\varepsilon} \mathbf{U}_m, \mathbf{Y}.$$

We will use the following construction of seeded extractors:

Theorem 4.7 (Theorem 1.5 in [GUV09])**.** *For all constant $\alpha > 0$ and all n, k, ε, there exists an explicit (k,ε)-seeded extractor $\mathsf{sExt} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ with $d = O(\log(n/\varepsilon)$ and $m \geq (1-\alpha)k$.*

We use fact that inner product function over finite fields is a good two source extractor:

Theorem 4.8. *[Cha16, CG88, ILL89, Vaz85] Let $\mathbf{X}, \mathbf{Y} \sim \{0,1\}^n$ with $H_{\infty}(\mathbf{X}) = k_1, H_{\infty}(\mathbf{Y}) = k_2$. Let $m = \frac{n}{r}$ for some $r \in \mathbb{N}$. Let $\mathsf{IP}(x,y) : \{0,1\}^{2n} \to \{0,1\}^m$ be the function that interprets x, y as elements of $\mathbb{F}^r_{2^m}$ and outputs the m bit string corresponding to $x \cdot y$. Then, $|\mathsf{IP}(\mathbf{X}, \mathbf{Y}) - \mathbf{U}_m| \leq 2^{(n+m-k_1-k_2)/2}$.*

4.3 Randomness sources relevant to our work

We now formally introduce the randomness sources we will be working with. We begin with NOSF sources, which have no restrictions on the adversary producing the bad blocks.

Definition 4.9 (NOSF source)**.** *A (g, ℓ, n, k)-non-oblivious symbol fixing source (NOSF) $\mathbf{X}$ with symbols in $\Sigma = \{0,1\}^n$ and length ℓ is over Σ^ℓ, written as $\mathbf{X} = \mathbf{X_1}, \ldots, \mathbf{X}_\ell$, and has the following property: There exists a set of good blocks $\mathcal{G} \subseteq [\ell]$ such that $|\mathcal{G}| \geq g$ and the random variables in $\{\mathbf{X}_i\}_{i \in \mathcal{G}}$ are each independently sampled (n, k)-sources.*

We say that a block $\mathbf{X}_i$ is good *if $i \in \mathcal{G}$ and* bad *otherwise.*

Note that we have no restrictions on how bad blocks may depend on the good blocks. If $k = n$, we say that $\mathbf{X}$ is a *uniform (g, ℓ, n)-NOSF* source. When n is implicit or not relevant, we simply call $\mathbf{X}$ a uniform (g, ℓ)-NOSF source.

Next, we introduce SHELA sources in their full generality.

Definition 4.10 (SHELA source [AORSV20])**.** *A distribution $\mathbf{X}$ over $(\{0,1\}^n)^\ell$ is a (g, ℓ, n, k)-*Somewhere Honest Entropic Look Ahead (SHELA) source *if there exists a (possibly randomized) adversary $\mathcal{A}$ such that $\mathbf{X}$ is produced by sampling g out of ℓ indices to place independently sampled (n, k)-sources and then placing adversarial blocks in the other $\ell - g$ indices that may depend arbitrarily on any block that comes before it.*

Concretely, there must exist random variables $1 \leq \mathbf{I}_1 < \mathbf{I}_2 < \cdots < \mathbf{I}_g \leq \ell$ with arbitrary joint distribution, denoting the indices of the independent (n, k)-sources, and g independent (n, k)-sources $\mathbf{Z}_1, \mathbf{Z}_2, \ldots, \mathbf{Z}_g$ such that $\mathbf{X}$ is generated in the following manner:

1. *Sample $(i_1, i_2, \ldots, i_g) \sim (\mathbf{I}_1, \mathbf{I}_2, \ldots, \mathbf{I}_g)$.*
2. *For all $j \in [g]$ set $\mathbf{B}_{i_j} = \mathbf{Z}_j$.*
3. *For all $i \in [\ell] \setminus \{i_1, i_2, \ldots, i_g\}$, the adversary sets $\mathbf{B}_i = \mathcal{A}(\mathbf{B}_1, \ldots, \mathbf{B}_{i-1}, i_1, \ldots, i_g\}$.*
4. *Finally, let $\mathbf{X} = (\mathbf{B}_1, \ldots, \mathbf{B}_\ell)$.*

We will generally call the blocks $\mathbf{Z}_1, \ldots, \mathbf{Z}_g$ the "good" blocks and the remaining blocks "bad" blocks.

Similar to NOSF sources, when $k = n$ we will simply say $\mathbf{X}$ is a (g, ℓ, n)-*uniform SHELA* source, and when n is implicit we will simplify further to a uniform (g, ℓ)-SHELA source.

In our construction of lower and upper bounds for SHELA sources, we will often think of *fixed-index SHELA* sources instead since they are easier to reason about. We define them now.

Definition 4.11 (Fixed-index SHELA source)**.** *A distribution $\mathbf{X}$ over $(\{0,1\}^n)^\ell$ is a (g, ℓ, n, k)-*fixed-index SHELA (fiSHELA) *source if there exists a (possibly randomized) adversary $\mathcal{A}$ such that $\mathbf{X}$ is produced by the adversary choosing g out of ℓ indices to place independently sampled (n, k)-sources and then placing adversarial blocks in the other $\ell - g$ indices that may depend arbitrarily on any block that comes before it.*

Concretely, the adversary $\mathcal{A}$ chooses $1 \leq i_1 < i_2 < \cdots < i_g \leq \ell$, denoting the indices of the independent (n, k)-sources, and g independent (n, k)-sources $\mathbf{Z}_1, \mathbf{Z}_2, \ldots, \mathbf{Z}_g$ such that $\mathbf{X}$ is generated in the following manner:

1. *For all $j \in [g]$ set $\mathbf{B}_{i_j} = \mathbf{Z}_j$.*
2. *For all $i \in [\ell] \setminus \{i_1, i_2, \ldots, i_g\}$, the adversary sets $\mathbf{B}_i = \mathcal{A}(\mathbf{B}_1, \ldots, \mathbf{B}_{i-1}, i_1, \ldots, i_g\}$.*

3. *Finally, let* $\mathbf{X} = (\mathbf{B}_1, \ldots, \mathbf{B}_\ell)$.

While working over fixed-index SHELA sources is easier than working over general SHELA sources, all of our results still apply to general SHELA sources since SHELA sources are convex combinations of fixed-index SHELA sources.

Proposition 4.12. *Every (g, ℓ, n, k)-SHELA source $\mathbf{X}$ is a convex combination of (g, ℓ, n, k)-fixed-index SHELA sources.*

Proof. Let $\mathbf{I} = \mathbf{I}_1, \mathbf{I}_2, \ldots, \mathbf{I}_g$ be the distribution of indices used in the construction of $\mathbf{X}$. For a sample $\mathcal{I} \sim \mathbf{I}$, let $\mathbf{X}_\mathcal{I}$ the (g, ℓ, n, k)-fixed-index SHELA source in the construction of which the adversary chose the good blocks to be at indices $\mathcal{I}$ and the functions describing the bad blocks to be identical to those of $\mathbf{X}$ when the sample of indices from I is $\mathcal{I}$. That is, when $\mathcal{I}$ is sampled in the construction of $\mathbf{X}$ we have for all $j \in [\ell] \setminus \mathcal{I}$ that $\mathbf{X}_j = (\mathbf{X}_\mathcal{I})_j$ as functions.

With this setup, we directly have that $\mathbf{X} = \mathbb{E}_{\mathcal{I} \sim \mathbf{I}}[\mathbf{X}_\mathcal{I}]$, so $\mathbf{X}$ is a convex combination of $\mathbf{X}_\mathcal{I}$'s. □

Lastly, we define almost Chor-Goldreich (CG) sources, which have an adversary like that of fiSHELA sources that can depend arbitrarily on past blocks, but the adversary of almost CG sources can have some effect on future blocks, unlike that of fiSHELA sources. Almost CG sources are more easily defined in two steps, the first of which tells us what "good" means in this context.

Definition 4.13 (Good CG block [DMOZ23])**.** *Let $0 \le \gamma \le 1$, $0 \le k \le n$, and $\mathbf{X} = \mathbf{X}_1, \ldots, \mathbf{X}_\ell$ be a source with each $\mathbf{X}_i$ over $\{0,1\}^n$. We say that $i \in [\ell]$ is* (γ, k)-good *for $\mathbf{X}$ if for all prefixes $(a_1, \ldots, a_{i-1}) \in (\{0,1\}^n)^{i-1}$ we have that*

$$H_\infty^\gamma(\mathbf{X}_i \mid \mathbf{X}_1, \ldots, \mathbf{X}_{i-1} = a_1, \ldots, a_{i-1}) \ge k.$$

In the trivial case that $i = 1$ we have $H_\infty^\gamma(\mathbf{X}_1) \ge k$. When γ, k, and $\mathbf{X}$ are clear from context, we will simply call a block i "good", otherwise we will call it "bad".

We can now succinctly state the definition of an almost CG source.

Definition 4.14 (Almost-CG source [DMOZ23])**.** *A (g, ℓ, n, k, γ)-*almost CG source *$\mathbf{X}$ is a sequence of random variables $\mathbf{X} = \mathbf{X}_1, \ldots, \mathbf{X}_\ell$ with $\mathbf{X}_i$ taking values in $\{0,1\}^n$. We require that at least g of the blocks of $\mathbf{X}$ are (γ, k)-good.*

As before, if $k = n$ and $\gamma = 0$, then we simply call $\mathbf{X}$ a uniform (g, ℓ, n)-almost CG source, and we omit n when it is implicit.

We have introduced all of these definitions since our results resolve open questions for each. The relationship between all these definitions is necessary to clearly see how our lower and upper bounds apply, so we provide two propositions to help elucidate the relationship between them.

Proposition 4.15. *Uniform fixed-index SHELA sources are also uniform NOSF sources.*

Proof. A uniform (g, ℓ, n)-fixed-index SHELA source $\mathbf{X}$ is also a uniform (g, ℓ, n)-NOSF source since the $\ell - g$ bad blocks in $\mathbf{X}$ are strictly more restricted than the bad blocks of a uniform NOSF source. □

Next, we show an equivalence between uniform fiSHELA and uniform almost CG sources

Proposition 4.16. *A source $\mathbf{X}$ is a uniform fixed-index SHELA source if and only if it is a uniform almost CG source.*

Proof. Say $\mathbf{X}$ is a uniform (g, ℓ, n)-fixed-index SHELA source. Then, because bad blocks may only depend on the good blocks that have a lower index than them and all the good blocks are sampled independently, the good blocks satisfy the prefix condition in Definition 4.13 to give us that $\mathbf{X}$ is a uniform (g, ℓ, n)-almost CG source.

On the other hand, say that $\mathbf{X}$ is a uniform (g, ℓ, n)-almost CG source. Then the fact that for a good block $\mathbf{X}_i$ we have for all $(a_1, \ldots, a_{i-1}) \in (\{0,1\}^n)^{i-1}$ that $H_\infty(\mathbf{X}_i \mid \mathbf{X}_1, \ldots, \mathbf{X}_{i-1} = a_1, \ldots, a_{i-1}) = n$, so $\mathbf{X}_i$ is uniform given any prefix, means that $\mathbf{X}_i$ is independent of all blocks that come before it. In particular, this means that bad blocks may only depend on the good blocks that come before them. In addition, the good blocks being uniform clearly means that they are independent from each other. Hence, $\mathbf{X}$ is a uniform (g, ℓ, n)-fixed-index SHELA source as well. □

Putting both of these propositions together yields Figure 3 which depicts how our definitions interact. Therefore, when we prove a lower bound by constructing a uniform fiSHELA source, that same lower bound

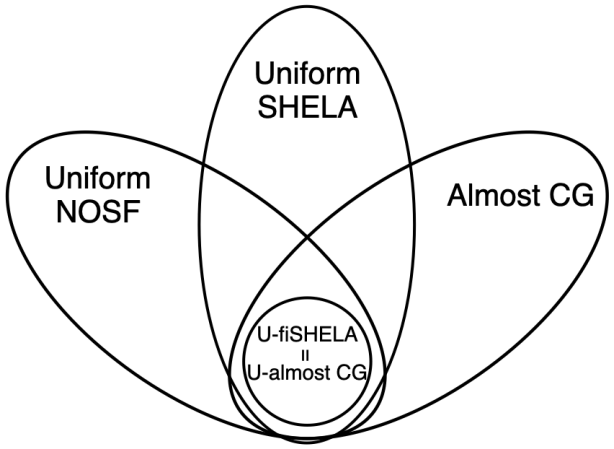


Figure 3: Here we illustrate the containments between our sources along with the equivalence between uniform fiSHELA (U-fiSHELA) and uniform almost CG (U-almost CG) sources.

applies to uniform NOSF and uniform almost CG sources as well.

5 Impossibility Results

In this section, we will first prove condensing impossibility results for uniform (g, ℓ)-SHELA sources when $g \leq \ell/2$ and for uniform $(2, 3)$-NOSF sources. Then we will show an extraction impossibility result for uniform $(2, 3)$-SHELA sources.

Remark 5.1. *Except for the impossibility result in Section 5.2, the rest of our results are accomplished by constructing a uniform fiSHELA adversary. Because uniform fiSHELA sources are contained in all of the other sources we use in this work (Figure 3), these results also apply to uniform NOSF and uniform almost CG sources as well. Thus, to be succinct, we will only state our results in terms of uniform SHELA sources.*

5.1 Impossibility of Condensing When $g \leq \ell/2$

We begin by proving that for $g \leq \ell/2$, it is impossible to condense from a (g, ℓ)-SHELA source $\mathbf{X}$ to rate more than $\frac{1}{\lfloor \ell/g \rfloor}$.

Theorem 5.2. *For any $f : (\{0,1\}^n)^\ell \to \{0,1\}^t$ and for all $\varepsilon \geq 0$, there exists a $\delta > 0$ and a uniform (g, ℓ)-SHELA source $\mathbf{X}$ with $g \leq \ell/2$ such that $H^\varepsilon_\infty(f(\mathbf{X})) \leq \frac{1}{\lfloor \ell/g \rfloor} \cdot t + \delta$.*

An immediate corollary of this theorem is for the special case where g divides ℓ.

Corollary 5.3. *For any $\varepsilon > 0$ and any g and ℓ such that $g \mid \ell$ and $f : (\{0,1\}^n)^\ell \to \{0,1\}^t$, there exists a uniform (g, ℓ)-SHELA source $\mathbf{X}$ a $\delta > 0$ such that $H^\varepsilon_\infty(f(\mathbf{X})) < \frac{g}{\ell} \cdot t + \delta$.*

The proof of Theorem 5.2 involves two ingredients. First, we must show that condensing above rate $\frac{1}{\ell}$ is impossible for uniform $(1, \ell)$-SHELA sources.

Lemma 5.4. *For any $f : (\{0,1\}^n)^\ell \to \{0,1\}^t$ and $\varepsilon \geq 0$ there exists a uniform $(1, \ell)$-SHELA source $\mathbf{X}$ and $\delta > 0$ such that $H^\varepsilon_\infty(f(\mathbf{X})) < \frac{1}{\ell} \cdot t + \delta$.*

Second, we show that these results extend to any uniform (g, ℓ)-SHELA source with $g \leq \ell/2$. That is, if it is impossible to condense from the class of uniform $(1, \ell')$ SHELA sources, then it is impossible to condense above rate $\frac{1}{\ell'}$ from the class of (g, ℓ) SHELA sources for any g and ℓ when $\ell \geq \ell'$ and $\frac{g}{\ell} \leq \frac{1}{\ell'}$.

Lemma 5.5. *Let $\ell \geq \ell'$ and $\frac{g}{\ell} \leq \frac{1}{\ell'}$. If for any function $f : (\{0,1\}^n)^{\ell'} \to \{0,1\}^t$ and fixed $\varepsilon > 0$ there exists a uniform $(1, \ell', n)$-fiSHELA source $\mathbf{Y}$ and constant $\delta > 0$ such that $H^\varepsilon_\infty(f(\mathbf{Y})) < \frac{g'}{\ell'} \cdot t + \delta$, then for any integer m such that $\lceil \ell/\ell' \rceil m < n$ and any function $h : (\{0,1\}^m)^\ell \to \{0,1\}^t$, there exists a uniform (g, ℓ, m)-fiSHELA source $\mathbf{X}$ such that $H^\varepsilon_\infty(h(\mathbf{X})) \leq \frac{1}{\ell'} \cdot t + \delta$.*

With these lemmas, our main theorem follows naturally.

Proof of Theorem 5.2. Let us divide ℓ by g as $\ell = c \cdot g + r$ for some $c, r \in \mathbb{N}$ with $r < g$ so $c > 0$ and $r < g$. Notice that $\lfloor \ell/g \rfloor = c$. We can derive our desired impossibility result by applying Lemma 5.5 to the result of Lemma 5.4 for uniform $(1, c)$-fiSHELA sources.

□

Now we will prove both of these lemmas. We begin with the self-contained Lemma 5.5.

Proof of Lemma 5.5. For the sake of contradiction, assume there exists a non-trivial condenser $h : (\{0,1\}^m)^\ell \to \{0,1\}^t$ that condenses above rate $\frac{1}{\ell'}$, meaning that there exists some $\varepsilon > 0$ so that for any uniform (g, ℓ, m)-fiSHELA source $\mathbf{X}$ we have $H^\varepsilon_\infty(h(\mathbf{X})) \geq \frac{1}{\ell'} \cdot t + \omega(1)$. We will now use h to construct a condenser $f : (\{0,1\}^n)^{\ell'} \to \{0,1\}^t$ for any uniform $(1, \ell', n)$-fiSHELA source $\mathbf{Y}$ to derive a contradiction to our assumption in the theorem statement.

Let us divide ℓ by ℓ' with remainder to get $\ell = c\ell' + r$ for some integers c and r where $r < \ell'$. Note that $c > 0$ since $\ell \geq \ell'$. Our construction of f is quite simple: on input $\mathbf{Y} = \mathbf{Y}_1, \ldots, \mathbf{Y}_{\ell'}$, we define $f(\mathbf{Y}_1, \ldots, \mathbf{Y}_{\ell'}) = h(\mathbf{X}_1, \ldots, \mathbf{X}_{\ell = c\ell' + r})$ where the $\mathbf{X}_i$ are constructed by splitting up the $\mathbf{Y}_j$ as evenly as possible. Concretely, from each $\mathbf{Y}_1, \ldots, \mathbf{Y}_r$ we will take $c + 1$ blocks of size m to form $r(c + 1)$ of the $\mathbf{X}_i$'s (this is where our requirement that $n > m(c + 1)$ comes from), and from each of the $\ell' - r$ remaining $\mathbf{Y}_{r+1}, \ldots, \mathbf{Y}'_\ell$ we will take c blocks of size m to form $(\ell' - r)c$ of the $\mathbf{X}_i$'s. In total, this gives us the desired $r(c + 1) + (\ell' - r)c = rc + r + c\ell' - rc = c\ell' + r = \ell$ blocks of $\mathbf{X}$.

Finally, to see that indeed g of these blocks in $\mathbf{X}$ are good so that we may apply h, recall that $\mathbf{Y}$ is a $(1, \ell')$-fiSHELA source, so there exists some index $j \in [\ell]$ such that $\mathbf{Y}_j$ is good. If $j \leq r$ we will get $c+1$ good blocks in $\mathbf{X}$ and otherwise we will get c good blocks in $\mathbf{X}$. Thus, we must show that $c \geq g$. By using the fact that $\frac{1}{\ell'} \geq \frac{g}{\ell}$ and that $\ell = c\ell' + r$ we can compute $c\ell + r = \ell \geq g\ell'$, meaning that $g - c \leq \frac{r}{\ell'}$. But since $r < \ell'$ and both g and c are integers, so their difference is an integer, it must be that $g - c \leq 0$, so $c \geq g$.

As promised, our constructed $\mathbf{X}$ meets the requirements for h to condense from it, yielding $H^{\varepsilon}_{\infty}(f(\mathbf{Y})) \geq \frac{g'}{\ell'} \cdot t + \omega(1)$, a contradiction to our assumption in the statement of the lemma. □

Next, we prove Lemma 5.4, which we obtain as a corollary to the fact that if we cannot condense from uniform (g, ℓ)-SHELA sources, then we cannot condense from uniform $(g, \ell + g)$-SHELA sources.

Lemma 5.6. *Assume that for every $s \in \mathbb{N}$ and function $f : \{0,1\}^{\ell n} \to \{0,1\}^s$, there exists a uniform (g, ℓ)-SHELA source $\mathbf{X}$ such that $H^{\varepsilon}_{\infty}(f(\mathbf{X})) \leq \frac{g}{\ell} \cdot s + \delta$. Let $c_0, c_1 \in \mathbb{R}$ be such that $0 < c_0 < 1$ and $\varepsilon < c_1 < 1$. Then, for every function $h : \{0,1\}^{\ell+g} \to \{0,1\}^t$, there exists a uniform $(g, \ell + g)$-SHELA source $\mathbf{Y}$ such that $H^{\varepsilon}_{\infty}(h(\mathbf{Y})) \leq \frac{g}{\ell+g} \cdot t + \delta'$ where $\delta' = \max\left(\log\left(\frac{c_1}{(1-c_1)c_0(c_1-\varepsilon)}\right), \delta + \frac{\log(c_0)g}{\ell}\right)$.*

We remark that Lemma 5.6 is exactly an inductive argument which we instantiate to get Lemma 5.4.

Proof of Lemma 5.4. All we must do is provide a base case for Theorem Lemma 5.6 to work off of and give us the result we desire. Notice that a uniform $(g = 1, \ell = 1)$-SHELA source $\mathbf{W}$ is just the uniform source on n bits. Thus, for any $f : \{0,1\}^n \to \{0,1\}^t$ it must be that for all $\varepsilon \geq 0$ we have $H^{\varepsilon}_{\infty}(f(\mathbf{W})) \leq t$.

We apply Lemma 5.6 by setting $c_0 = 1, c_1 = \frac{1+\varepsilon}{2}, \delta = \log\left(\frac{2(1+\varepsilon)}{(1-\varepsilon)^2}\right)$ to infer the claim! □

Before proving Lemma 5.6, we require a dominating set argument on bipartite graphs which we will use to construct the adversary in our SHELA sources.

Lemma 5.7 (Greedy Covering Argument)**.** *Let $c_0 > 0$ and $0 < c_1 < 1$. For a bipartite graph $H = (U, V, E)$ where $|U| = N$, $|V| = T$, and $\deg(u) \geq c_0 T^{\delta}$ for all $u \in U$ and some $\delta \in (0, 1)$, there exists a subset $D \subseteq V$ of size at most $\frac{c_1}{(1-c_1)c_0} T^{1-\delta}$ such that $|\mathcal{N}(D)| \geq c_1 N$.*

Proof. We will construct D via the greedy algorithm in Algorithm 1. This algorithm greedily chooses right vertices in V with highest degree first, adds them to D, and stops once the neighborhood of D, denoted $\mathcal{N}(D)$, is large enough. To analyze this algorithm, we can use loose bounds on the number of edges and vertices at any one step. Notice that since the algorithm stops once we have removed at least $c_1 N$ vertices from U, we have for all i that $|U_i| \geq (1 - c_1)N$. In addition, because left vertices are only removed when one of their neighbors in V is added to D, we see that remaining vertices in U still have their original degree since otherwise they would have been removed. In other words, at each time step i we have for all $u \in U_i$ that $\deg(u) \geq c_0 T^{\delta}$. Putting these two facts together gives us that at each time step the number of edges is

$$|E_i| \geq |U_i|\, d \geq (1 - c_1) N c_0 T^{\delta}.$$

Therefore, because we are never adding vertices to V so $|V_i| \leq V = |V| = T$ for all i, we see that there must exist a vertex in V_i of degree at least

$$\frac{|E_i|}{|V_i|} \geq \frac{(1 - c_1) N c_0 T^{\delta}}{T} = \frac{(1 - c_1) c_0 N}{T^{1-\delta}}.$$

Algorithm 1: Greedy covering algorithm

$i \leftarrow 0$
$D \leftarrow \varnothing$
$H_0 = (U_0, V_0, E_0) \leftarrow H = (U, V, E)$
while $|\mathcal{N}(D)| < c_1 N$ **do**
  Let $v_i \in V_i$ be the vertex of maximum degree in H_i
  $D \leftarrow D \cup \{v_i\}$
  $V_{i+1} \leftarrow V_i \setminus \{v_i\}$
  $U_{i+1} \leftarrow U_i \setminus \mathcal{N}(v_i)$
  $E_{i+1} \leftarrow E_i \setminus \{(u, v) \in E \mid v = v_i \text{ or } u \in \mathcal{N}(v_i)\}$
  $H_{i+1} \leftarrow (U_{i+1}, V_{i+1}, E_{i+1})$
end

Finally, since we stop exactly once we get at least $c_1 N$ vertices in D, we will stop in at most

$$\frac{c_1 N}{\frac{(1-c_1)c_0 N}{T^{1-\delta}}} = \frac{c_1}{(1-c_1)c_0} T^{1-\delta}$$

steps. But since each step adds exactly one vertex to D, we have that this is a bound on the size of D as well. □

With this greedy covering argument for bipartite graphs in hand, we are ready to show Lemma 5.6.

Proof of Lemma 5.6. Fix a function $h : \{0,1\}^{\ell+g} \to \{0,1\}^t$. We will construct a $(g, \ell+g)$ uniform SHELA source $\mathbf{Y}$ such that $H^{\varepsilon}_{\infty}(f(Y)) < \frac{g}{\ell+g} \cdot t + \delta'$. Let $N = 2^n, T = 2^t$. We consider two cases:

Case 1. For all $x_1, \ldots, x_g \in (\{0,1\}^n)^g$ it holds that $|\text{support}(h(x_1, \ldots, x_g, \mathbf{U}_\ell))| \geq c_0 T^{\ell/(\ell+g)}$.
Consider an undirected bipartite graph $G = (U, V)$ where $U = (\{0,1\}^n)^g$ and $V = \{0,1\}^t$ with edge from $u = (x_1, \ldots, x_g) \in U$ to $v \in V$ if there exist $x_{g+1}, \ldots, x_{\ell+g}$ such that $h(x_1, \ldots, x_{\ell+g}) = v$. We apply Lemma 5.7 to G and infer that there exists $D \subset \{0,1\}^t$ such that $|D| \leq \frac{c_1}{(1-c_1)c_0} T^{g/(\ell+g)}$ and for $c_1 N^g$ many tuples $(x_1, \ldots, x_g) \in (\{0,1\}^n)^g$, there exist $y_1, \ldots, y_\ell \in (\{0,1\}^n)^\ell$ such that $h(x_1, \ldots, x_g, y_1, \ldots, y_\ell) \in D$. Let $a : (\{0,1\}^n)^g \to (\{0,1\}^n)^\ell$ be defined as:

$$a(x_1, \ldots, x_g) = \begin{cases} (y_1, \ldots, y_\ell) & \text{if there exist } y_1, \ldots, y_\ell \text{ such that } h(x_1, \ldots, x_g, y_1, \ldots, y_\ell) \in D \\ (0^n)^\ell & \text{otherwise} \end{cases}$$

Consider the uniform $(g, \ell+g)$-SHELA source $\mathbf{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_{\ell+g})$ such that $\mathbf{X}_1, \ldots, \mathbf{X}_g$ are uniform independent distributions and $\mathbf{X}_{g+1}, \ldots, \mathbf{X}_{g+\ell} = a(\mathbf{X}_1, \ldots, \mathbf{X}_g)$. Then, we infer that with probability c_1, $h(\mathbf{X}) \in D$. Applying Claim 4.3, we infer that $H^{\varepsilon}_{\infty}(\mathbf{X}) \leq \log\left(\frac{c_1 T^{g/(\ell+g)}}{(1-c_1)c_0(c_1-\varepsilon)}\right) = \frac{g}{\ell+g} \cdot t + \log\left(\frac{c_1}{(1-c_1)c_0(c_1-\varepsilon)}\right) \leq \frac{g}{\ell+g} \cdot t + \delta'$.

Case 2. There exist $x_1, \ldots, x_g \in (\{0,1\}^n)^g$ such that $|\text{support}(h(x_1, \ldots, x_g, \mathbf{U}_\ell))| \leq c_0 T^{\ell/(\ell+g)}$.
Let $S = |\text{support}(h(x_1, \ldots, x_g, \mathbf{U}_\ell))|$. Let $b_1 : \{0,1\}^s \to \text{support}(h(x_1, \ldots, x_g, \mathbf{U}_\ell))$ be an arbitrary bijection with inverse function b_2. Let $f : \{0,1\}^\ell \to \{0,1\}^s$ be defined by $f(y_1, \ldots, y_\ell) = b_2(h(x_1, \ldots, x_g, y_1, \ldots, y_\ell))$. Then, by assumption, there exists some uniform (g, ℓ)-SHELA source $\mathbf{Y}$

such that $f(\mathbf{Y})$ is not ε close to min entropy $\frac{g}{\ell} \cdot s + \delta$. Consider uniform $(g, \ell+g)$-SHELA source $\mathbf{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_{\ell+g})$ where distributions $\mathbf{X}_1, \ldots, \mathbf{X}_g$ always output $x_1, \ldots, x_g$ and $\mathbf{X}_{g+1}, \ldots, \mathbf{X}_{\ell+g}$ are distributed as $\mathbf{Y}$. Then, we infer that with $H^{\varepsilon}_{\infty}(f(\mathbf{X})) \leq \frac{g}{\ell+g} \cdot t + \delta + \frac{\log(c_0)g}{\ell}$.

□

5.2 Impossibility of Condensing from Uniform $(2,3)$-NOSF Sources

Here, we will show that it is impossible to condense from uniform $(2,3)$-NOSF sources.

Theorem 5.8. *For any $f : (\{0,1\}^n)^3 \to \{0,1\}^t$ and $0 < \varepsilon < \frac{1}{4}$, there exists a uniform $(2,3)$-NOSF source $\mathbf{X}$ and a constant $\delta > 0$ such that $H^{\varepsilon}_{\infty}(f(\mathbf{X})) \leq \frac{2}{3} \cdot t + \delta$.*

We can extend Theorem 5.8 using techniques from Section 5.1 to show that condensing from uniform (g,ℓ)-NOSF sources is impossible whenever $\frac{g}{\ell} = \frac{2}{c}$ for some $c \in \mathbb{N}$.

Corollary 5.9. *If $\frac{g}{\ell} = \frac{2}{c}$ for some $c \in \mathbb{N}$ then there exists a $0 \leq \varepsilon < \frac{1}{4}$ and $\delta > 0$ such that for any $f : (\{0,1\}^n)^{\ell} \to \{0,1\}^t$ there exists a uniform (g,ℓ)-NOSF source $\mathbf{X}$ for which $H^{\varepsilon}_{\infty}(f(\mathbf{X})) \leq \frac{g}{\ell} \cdot t + \delta$.*

The only new method we need here to prove Corollary 5.9 is that condensing impossibility results scale up. In other words, let g_0, ℓ_0 be such that it is not possible to condense from uniform (g_0, ℓ_0)-SHELA sources. Then for all $c \in \mathbb{N}$, it holds that it is not possible to condense from uniform $(c \cdot g_0, c \cdot \ell_0)$-SHELA sources.

Lemma 5.10. *Assume there exist $0 < \varepsilon < 1, 0 < k \leq n$ such that for any $f : (\{0,1\}^n)^{\ell_0} \to \{0,1\}^t$, there exists a uniform (g_0, ℓ_0)-SHELA source $\mathbf{X}$ such that $H^{\varepsilon}_{\infty}(f(\mathbf{X})) < k$. Then, for any $f : (\{0,1\}^{n/c})^{c \cdot \ell_0} \to \{0,1\}^t$, there exists a uniform $(c \cdot g_0, c \cdot \ell_0)$-SHELA source $\mathbf{X}$ such that $H^{\varepsilon}_{\infty}(f(X)) < k$.*

Proof of Lemma 5.10. Proceed by contradiction and assume such f existed. Consider $h : (\{0,1\}^n)^{\ell_0} \to \{0,1\}^t$ where on input $(x_1, \ldots, x_{\ell})$, h partitions each x_i into c blocks of length ℓ/c each and calls f on the resultant input. Observe that if $\mathbf{X}$ is a uniform (g_0, ℓ_0)-SHELA source, then this partitioning operation turns it into a uniform $(c \cdot g_0, c \cdot \ell_0)$-SHELA source. As f is a (k, ε)-condenser for such sources, h will also be a (k, ε)-condenser for uniform (g_0, ℓ_0)-SHELA sources, which is a contradiction. □

Corollary 5.9 now follows directly.

Proof of Corollary 5.9. We first show that it is not possible to condense from uniform $(2,\ell)$-NOSF sources when ℓ is odd. Using Theorem 5.8 as the base case for Lemma 5.6 with $c_0 = 1$, $c_1 = \frac{1+\varepsilon}{2}$ and $\delta = \log\left(\frac{2(1+\varepsilon)}{(1-\varepsilon)^2}\right)$, we get that there exists a $0 \leq \varepsilon < \frac{1}{4}$ and $\delta > 0$ such that for any $f : (\{0,1\}^n)^{\ell} \to \{0,1\}^t$ there exists a uniform $(2,\ell)$-NOSF source $\mathbf{X}$ such that $H^{\varepsilon}_{\infty}(f(X)) \leq \frac{2}{\ell} \cdot t + \delta$.

Next, we extend this to any g and ℓ such that $\frac{g}{\ell} = \frac{2}{c}$ for odd c by using Lemma 5.10.

To get our final claim, we notice that if $\frac{g}{\ell} = \frac{2}{c}$ for some even $c \in \mathbb{N}$, then $\frac{g}{\ell} = \frac{1}{c'}$ for some $c' \in \mathbb{N}$, so Theorem 5.2 shows that for all $\varepsilon \geq 0$ there exists a $\delta > 0$ and uniform (g,ℓ)-NOSF source $\mathbf{X}$ such that for any $f : (\{0,1\}^n)^{\ell} \to \{0,1\}^t$ there exists a uniform (g,ℓ)-NOSF source $\mathbf{X}$ such that $H^{\varepsilon}_{\infty}(f(X)) \leq \frac{1}{c'} \cdot t + \delta = \frac{g}{\ell} \cdot t + \delta$. In particular, this holds for $0 \leq \varepsilon < \frac{1}{4}$. □

We can now move on to proving Theorem 5.8. As before, we will build our adversary via a covering set argument on bipartite graphs. The difference this time is that we will consider complete bipartite graphs with colored edges where we want our dominating set to be a set of colors that are used on many edges.

Lemma 5.11 (Greedy Covering Argument Again)**.** *Let $0 < c_0, 0 < c_1 < 1, 0 < c_2 < 1$ be such that $1 - c_0c_2 - c_1 > 0$. Consider a complete bipartite graph $H = (U, V, E)$ whose edges are colored in T colors with $|U| = N$, $|V| = N$. Moreover, assume that for every vertex $x \in H$, the number of distinct colors edges incident on x is $\leq c_0 T^{\delta}$ for some $\delta \in (0, 1)$ and some constant $c_0 > 0$. Then, there exists $D \subseteq [T]$ such that $|D| \leq \frac{c_0 c_1}{(1-c_0c_2-c_1)(c_2)} T^{2\delta}$ and $c_1 N^2$ edges in H are colored in one of the colors from D.*

We can now use this to prove Theorem 5.8.

Proof of Theorem 5.8. Fix a function $h : \{0,1\}^{3n} \to \{0,1\}^t$. We will construct a uniform $(2,3)$-NOSF source $\mathbf{Y}$ such that $H^{\varepsilon}_{\infty}(f(Y)) < \frac{2}{3} \cdot t + \delta$. Let $N = 2^n, T = 2^t$. Let $\alpha = \frac{1}{4} - \varepsilon$. Set $c_0 = \frac{1}{4} - \frac{\alpha}{2}, c_2 = \frac{1}{4} + \frac{\alpha}{8}, c_4 = 1 - \frac{\alpha}{4}, c_5 = \frac{\frac{1}{4} - \frac{\alpha}{2}}{\frac{1}{4} + \frac{\alpha^2}{16} - \frac{\alpha}{4}}$. Set $c_3 = c_6 = 1 - c_5$. Set $c_1 = \frac{c_3 c_5}{(1-c_3c_6-c_5)(c_6)}$. Then $c_0, c_1, c_2, c_3, c_4, c_5, c_5, c_6 > 0$ and they satisfy the following inequalities:

1. $\varepsilon < c_0 \leq 1$.
2. $\varepsilon < c_2 c_4$.
3. $c_4 < 1$.
4. $c_2 < \frac{1}{2}$.
5. $c_0 \leq c_5(1 - 2c_2)^2$.
6. $c_3 c_6 + c_5 < 1$.

We consider various cases:

Case 1. There exists $x_1 \in \{0,1\}^n, P_{23} \subset (\{0,1\}^n)^2$, and $D \subset \{0,1\}^t$ such that $|P_{23}| \geq c_0 N^2, |D| \leq c_1 T^{2/3}$ and for all $(x_2, x_3) \in P_{23}$, it holds that $h(x_1, x_2, x_3) \in D$. In this case, consider the uniform $(2,3)$-NOSF source $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3)$ where $\mathbf{X}_1$ always outputs x_1 and $\mathbf{X}_2, \mathbf{X}_3$ are independent uniform distributions over $\{0,1\}^n$. Then with probability c_0, $h(\mathbf{X}) \in D$. Applying Claim 4.3, we infer that

$$H^{\varepsilon}_{\infty}(\mathbf{X}) \leq \log\left(\frac{c_1 T^{2/3}}{c_0 - \varepsilon}\right) = \frac{2}{3} \cdot t + \log(c_1) - \log(c_0 - \varepsilon).$$

Case 2. There exists $P_{12} \subset (\{0,1\}^n)^2$ such that $|P_{12}| \geq c_2 N^2$ and for all $(x_1, x_2) \in P_{12}$, it holds that

$$|\{h(x_1, x_2, y_3) : y_3 \in \{0,1\}^n\}| \geq c_3 T^{1/3}.$$

In this case, consider the bipartite graph $H = (U, V, E)$ where $U = P_{12}, V = \{0,1\}^t$ and edge $e = (u, v) = ((x_1, x_2), t) \in E$ if and only if there exists $y_3 \in \{0,1\}^n$ such that $h(x_1, x_2, y_3) = t$. Then by assumption, for all $u \in U$, it holds that $\deg(u) \geq c_3 T^{1/3}$.

We apply Lemma 5.7, to H and infer that there exists $D \subset \{0,1\}^t$ such that $|D| \leq \frac{c_4}{(1-c_4)c_3} T^{2/3}$ and $\mathrm{Nbr}(D) \geq c_4 N^2$. Let $a : \{0,1\}^{2n} \to \{0,1\}^t$ be defined as:

$$a(x_1, \ldots, x_2) = \begin{cases} y_3 & \text{if there exists } y_3 \text{ such that } h(x_1, x_2, y_3) \in D \\ 0^n & \text{otherwise} \end{cases}$$

Consider the uniform $(2,3)$-NOSF source $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3)$ such that $\mathbf{X}_1, \mathbf{X}_2$ are uniform independent distributions over $\{0,1\}^n$ and $\mathbf{X}_3 = a(\mathbf{X}_1, \mathbf{X}_2)$. Then, we infer that with probability $c_2, (\mathbf{X}_1, \mathbf{X}_2) \in P_{12}$ and hence, with probability $c_2 c_4, h(\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3) \in D$. Applying Claim 4.3, we infer that

$$H_\infty^\varepsilon(h(\mathbf{X})) \leq \log\left(\frac{c_4 T^{2/3}}{(1-c_4)c_3(c_2c_4-\varepsilon)}\right) = \frac{2}{3}\cdot t + \log\left(\frac{c_4}{(1-c_4)c_3(c_2c_4-\varepsilon)}\right).$$

Case 3. There exists $P_{13} \subset (\{0,1\}^n)^2$ such that $|P_{13}| \geq c_2 N^2$ and for all $(x_1, x_3) \in P_{13}$, it holds that

$$|\{h(x_1, y_2, x_3) \mid y_2 \in \{0,1\}^n\}| \geq c_3 T^{1/3}.$$

By the exact same argument as in Case 2, we infer that there exist a uniform $(2,3)$-NOSF source $\mathbf{X}$ such that:

$$H_\infty^\varepsilon(h(\mathbf{X})) \leq \log\left(\frac{c_4 T^{2/3}}{(1-c_4)c_3(c_2c_4-\varepsilon)}\right) = \frac{2}{3}\cdot t + \log\left(\frac{c_4}{(1-c_4)c_3(c_2c_4-\varepsilon)}\right).$$

Case 4. None of the other cases happen. We prove that this case cannot occur and hence, we must be in one of the three cases above. We do this by showing that if Case 2 and Case 3 did not occur, then Case 1 must have occurred. As Case 2 did not occur, there exists $Q_{12} \subset (\{0,1\}^n)^2$ such that $|Q_{12}| \geq (1-c_2)N^2$ and for all $(x_1, x_2) \in Q_{12}$, it holds that

$$|\{h(x_1, x_2, y_3) \mid y_3 \in \{0,1\}^n\}| < c_3 T^{1/3}.$$

As Case 3 did not occur, there exists $Q_{13} \subset (\{0,1\}^n)^2$ such that $|Q_13| \geq (1-c_2)N^2$ and for all $(x_1, x_3) \in Q_{13}$, it holds that

$$|\{h(x_1, y_2, x_3) \mid y_2 \in \{0,1\}^n\}| < c_3 T^{1/3}.$$

This implies there exists $z_1 \in \{0,1\}^n, P_2 \subset \{0,1\}^n, P_3 \subset \{0,1\}^n$ such that $|P_2| \geq (1-2c_2)N, |P_3| \geq (1-2c_2)N$ and for all $x_2 \in P_2, x_3 \in P_3$, it holds that

$$|\{h(z_1, y_2, x_3) \mid y_2 \in P_2\}| < c_3 T^{1/3}$$

and

$$|\{h(z_1, x_2, y_3) \mid y_3 \in P_3\}| < c_3 T^{1/3}.$$

Consider the complete bipartite graph $H = (U, V, E)$ whose edges are colored in T colors with $U = P_2, V = P_3$ and edge $e = (u, v)$ is colored with color t iff $h(z_1, u, v) = t$. We apply Lemma 5.11 to H and infer that there exists $D \subset T$ such that $|D| \leq \frac{c_3 c_5}{(1-c_3c_6-c_5)(c_6)} T^{2t/3}$ and $c_5(1-2c_2)^2 N^2$ edges in H are colored in one of the colors from D. As $c_5(1-2c_2)^2 \geq c_0$ and $\frac{c_3 c_5}{(1-c_3c_6-c_5)(c_6)} \leq c_1$, we indeed satisfy the conditions to be in Case 1 with $x_1 = z_1$.

□

The proof of Lemma 5.11 is similar to that of Algorithm 1 where instead of picking vertices with highest degree first, our algorithm picks colors that are assigned to the most edges first.

Proof of Lemma 5.11. For any $e \in E$, let $\chi(e)$ denote the color of e in H. For any vertex $x \in H$, we define

$$\text{Nbr}_H(x) = \{y \in H : (x, y) \in E\}.$$

Similarly, for any vertex $x \in H$, and color $c \in [T]$, we define

$$\text{Nbr}_H(x, c) = \{y \in H : (x, y) \in E \wedge \chi((x, y)) = c\}.$$

For a color $c \in [T]$ and a graph H, we define

$$\mathsf{count}_H(c) = |\{e \in H : \chi(e) = c\}|.$$

For $C \subset [T]$, we can similarly define

$$\mathsf{count}_H(C) = \sum_{c \in C} \mathsf{count}_H(c).$$

We will construct D via the greedy algorithm in Algorithm 2. Notice that the number of steps that

Algorithm 2: Greedy covering algorithm

```
i ← 0
D ← ∅
H_0 = (U_0, V_0, E_0) ← H = (U, V, E)
while count_H(D) ≤ c_1 N^2 do
    Let d_i ∈ [T] be the color that maximizes count_{H_i}(d_i).
    D ← D ∪ {d_i}
    E_{i+1} ← E_i \ {e ∈ E | χ(e) = d_i}
    H_{i+1} ← (U, V, E_{i+1})
end
```

the loop in the algorithm runs for equals $|D|$. We will carefully delete some edges from H and call the resultant graph H'. We will bound the runtime of the algorithm when ran over the input graph H'. For notational convenience, let the graph considered and the color chosen at each step i of the algorithm be H'_i, and d'_i respectively. Let D' be the resultant set of colors chosen. We observe that at each step i, $\mathsf{count}_H(d_i) \geq \mathsf{count}_{H'}(d'_i)$. Hence, $|D| \leq |D'|$ and so it suffices to upper bound $|D'|$.

Let

$$E' = \{e = (u, v) \in E \mid \text{Nbr}_H(v, \chi(e)) \geq c_2 N / T^\delta\}$$

Let $H' = (U, V, E')$. Fix arbitrary $v \in V$. By assumption, we know that $|\{c \in [T] : |\text{Nbr}_H(v, c)| > 0\}| \leq c_0 T^\delta$ and that $|\text{Nbr}_H(v) = N|$. In H', we remove all edges incident to v with color c such that $|\text{Nbr}_H(v, c)| \leq c_2 N / T^\delta$. Hence, we can remove at most $c_0 c_2 N$ such edges. Hence, we infer that:

$$|\text{Nbr}_{H'}(u| \geq (1 - c_0 c_2) N$$

And so, $|E'| \geq (1 - c_0 c_2) N^2$

Consider the last step j before the loop terminated. At that point, $\left|E'_j\right| \geq (1 - c_0 c_2 - c_1) N^2$. Let $d_j \in [T]$ be the color chosen at that step. Let $u \in U$ be such that $\left|\text{Nbr}_{H'_j}(u)\right|$ is maximized. As there are

$(1-c_0c_2-c_1)N^2$ edges in H'_j, we can find such a u with $\left|\text{Nbr}_{H'_j}(u)\right| \geq (1-c_0c_2-c_1)N$. As these edges are colored in at most c_0T^δ colors, there must exist a color c such that $\left|\text{Nbr}_{H'_j}(u)(u,c)\right| \geq \frac{(1-c_0c_2-c_1)N}{c_0T^\delta}$. For every $v \in \text{Nbr}_{H'_j}(u)(u,c)$, using the degree property of H' and the fact that we remove all instances of only a single color in the algorithm at each step, it must be that $\left|\text{Nbr}_{H'_j}(v,c)\right| \geq c_2N/T^\delta$. Hence, $\mathsf{count}_{H'_j}(c) \geq \frac{(1-c_0c_2-c_1)c_2N^2}{c_0T^{2\delta}}$. As d_j is chosen so that $\mathsf{count}_{H_j}(d_j)$ maximized, we infer that

$$\mathsf{count}_{H_j}(d_j) \geq \mathsf{count}_{H_j}(c) \geq \frac{(1-c_0c_2-c_1)c_2N^2}{c_0T^{2\delta}}.$$

Recall that at each step i we choose the color d such that $\mathsf{count}_{H_i}(d_i)$ is maximized and then remove all edges which have that color. Hence, it must be that at every step i, $\mathsf{count}_{H_i}(d_i) \geq \mathsf{count}_{H_j}(d_j) \geq \frac{(1-c_0c_2-c_1)c_2N^2}{c_0T^\delta}$. As we stop once $\mathsf{count}(D) \geq c_1N^2$ and at each step, this quantity increases by at least $\frac{(1-c_0c_2-c_1)c_2N^2}{c_0T^\delta}$, the loop terminates in at most $\frac{c_0c_1}{(1-c_0c_2-c_1)(c_2)}T^{2\delta}$ steps. □

5.3 Impossibility of Extracting from Uniform $(2,3)$-SHELA Sources

Our last impossibility result is that it is impossible to extract even one bit from uniform $(2,3)$-SHELA sources.

Theorem 5.12. *For any function $f : (\{0,1\}^n)^3 \to \{0,1\}$ there exists a uniform $(2,3)$-SHELA source $\mathbf{X}$ such that $|f(\mathbf{X}) - \mathbf{U}_1| \geq 0.08$.*

Proof. To show that extraction is impossible, we will attempt to fix the output of f with constant probability over its inputs. We begin by classifying the points in the first two coordinates of f as follows.

$$\begin{aligned}
S_0 &= \{(x_1,x_2) \in [N]^2 \mid \forall x_3 \in [N],\ f(x_1,x_2,x_3) = 0\} \\
S_1 &= \{(x_1,x_2) \in [N]^2 \mid \forall x_3 \in [N],\ f(x_1,x_2,x_3) = 1\} \\
S_{0,1} &= \{(x_1,x_2) \in [N]^2 \mid \exists x_3, x'_3 \in [N],\ f(x_1,x_2,x_3) = 0 \text{ and } f(x_1,x_2,x'_3) = 1\}.
\end{aligned}$$

Note that we can write $S_{0,1} = [N]^2 \setminus (S_0 \cup S_1)$. In order, these are the sets of points in $\mathbf{X}_1$ and $\mathbf{X}_2$ that fix the output of f to 0, to 1, and that do not fix the output of f. We now take constants $0.5 \leq c_0, c_1 \leq 1$ and look at two cases that allow us to fix the output of f by putting an adversary in the third coordinate, $\mathbf{X}_3$.

Case 1. We have $|S_0| + |S_{0,1}| \geq c_0N^2$. Here, we know that for $(x_1,x_2) \in S_0 \cup S_{0,1}$ there exists some x_3 such that $f(x_1,x_2,x_3) = 0$. Define $a(x_1,x_2)$ be this x_3 for $(x_1,x_2) \in S_0 \cup S_{0,1}$ and 0 otherwise. Consequently, if we let $\mathbf{X}_1$ and $\mathbf{X}_2$ be random and define our uniform $(2,3)$-SHELA source as $\mathbf{X} = \mathbf{X}_1, \mathbf{X}_2, a(\mathbf{X}_1,\mathbf{X}_2)$, then we have that $\Pr[f(\mathbf{X}) = 0] \geq c_0$. It follows that $|f(\mathbf{X}) - \mathbf{U}_1| \geq c_0 - \frac{1}{2}$.

Case 2. We have $|S_1| + |S_{0,1}| \geq c_0N^2$. This case follows similarly since for $(x_1,x_2) \in S_1 \cup S_{0,1}$ there exists some x_3 such that $f(x_1,x_2,x_3) = 1$. Therefore, we can define an adversary $a(x_1,x_2)$ such that when $\mathbf{X} = \mathbf{X}_1, \mathbf{X}_2, a(\mathbf{X}_1,\mathbf{X}_2)$ with $\mathbf{X}_1$ and $\mathbf{X}_2$ uniform we have $|f(\mathbf{X}) - \mathbf{U}_1| \geq c_0 - \frac{1}{2}$.

Case 3. We are in neither of the previous two cases. Thus, because $|S_0| + |S_1| + |S_{0,1}| = N^2$ we have that $(1-c_0)N^2 < |S_0|, |S_1| < c_0N^2$ and $(2c_0-1)N^2 < |S_{0,1}| < c_0N^2$. To proceed, we will set up two sub-cases in which we either make $\mathbf{X}_1$ our bad block or $\mathbf{X}_2$ our bad block.

Consider the bipartite graph $H = (U, V)$ with $|U| = N$ left vertices representing the values of $\mathbf{X}_1$ and $|V| = N$ vertices representing the values of $\mathbf{X}_2$. We place an edge (u, v) with label t if $(u, v) \in S_t$ and do not place an edge otherwise. Consequently, the number of edges E in H is at least $E = |S_0| + |S_1| \geq 2(1 - c_0)N^2$. For any $u \in U$, define its normalized degree (counting edges with either label) as $d_u = \deg(u)/N$. We then see that $\mathbb{E}_{u \sim U}[d_u] = E/|U| \geq 2(1 - c_0)$. To split into our two sub-cases, we will consider the set of heavy vertices $U_H = \{u \in U \mid d_u > c_1\}$ in U. By Claim 4.1, we get that $\Pr_{u \in U}[d_u > c_1] \geq \frac{\mathbb{E}_u[d_u] - c_1}{1 - c_1} \geq \frac{2(1-c_0)-c_1}{1-c_1} =: c_2$, meaning that $|U_H| \geq c_2 N$.

Case i. For all $u \in U_H$ we have $u \in S_0 \cap S_1$ (i.e., u has at least one edge labeled with a 0 and another with a 1). this means that for any $u \in U_H$ there exists an $x_2 \in [N]$ such that for all $x_3 \in [N]$ we have that $f(u, x_2, x_3) = 0$. Let $a(x_1)$ be defined as outputting this x_2 that fixes f to 0 for $x_1 \in U_H$ and to be 0 otherwise. Defining $\mathbf{X} = \mathbf{X}_1, a(\mathbf{X}_1), \mathbf{X}_3$ with $\mathbf{X}_1$ and $\mathbf{X}_3$ uniform gives us a uniform $(2, 3)$-SHELA source for which $\Pr[f(\mathbf{X}) = 0] \geq |U_H|/N \geq c_2$, so $|f(\mathbf{X}) - U_1| \geq c_2 - \frac{1}{2}$.

Case ii. There exists a $u \in U_H$ such that $u \notin S_0 \cap S_1$. Without loss of generality, say $u \in S_0$, so all of the edges of u are labeled 0, meaning that for all $x_2 \in \mathcal{N}(u)$ and any $x_3 \in [N]$ we have that $f(u, x_2, x_3) = 0$. Because $u \in U_H$, we have that $d_u > c_1$, so defining $\mathbf{X} = u, \mathbf{X}_2, \mathbf{X}_3$ with $\mathbf{X}_2$ and $\mathbf{X}_3$ uniform gives us that $\Pr[f(\mathbf{X}) = 0] \geq c_1$. Therefore, $|f(\mathbf{X}) - U_1| \geq c_1 - \frac{1}{2}$.

Combining all of our cases and recalling that $c_2 = \frac{2(1-c_0)-c_1}{1-c_1}$, we have that we can construct a uniform $(2, 3)$-SHELA source $\mathbf{X}$ such that $|f(\mathbf{X}) - U_1| \geq \varepsilon$ where $\varepsilon = \min(c_0, c_1, c_2) - \frac{1}{2}$. Setting $c_0 = 0.58$ and $c_1 = 0.6$ gives us $\varepsilon = 0.58 - 0.5 = 0.08$. □

6 Possibility Results

In this section, we will show that it is possible to condense from uniform $(2, 3)$-SHELA sources. In Section 6.1, we show the existence of excellent condensers for uniform $(2, 3)$-SHELA sources. In Section 6.2, we will explicitly construct such condensers.

6.1 Probabilistic construction

Theorem 6.1. *There exists a condenser* $\mathsf{Cond} : (\{0,1\}^n)^3 \to \{0,1\}^t$ *where* $t = n + \log(n) - \log(1/\varepsilon)$ *such that for any uniform* $(2, 3)$*-SHELA source* $\mathbf{X}$*,* $H^{\varepsilon}_{\infty}(\mathsf{Cond}(\mathbf{X})) \geq t - \log(t) - 4\log(1/\varepsilon) - O(1)$.

We show that a carefully chosen random process will generate a seeded extractor[2] with some additional guarantees. We will use this seeded extractor to get our condenser.

Definition 6.2 (Output-light Seeded Extractor)**.** *Let* $\mathsf{Ext} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ *be a* (k, ϵ)*-seeded extractor. Further suppose that for any* $z \in \{0,1\}^m$*, for every* $z \in \{0,1\}^m$*, it holds that* $\left|\{x \in \{0,1\}^n : \exists y \in \{0,1\}^d (\mathsf{Ext}(x, y) = z)\}\right| < R$*. We call* Ext *to be an* R*-output-light* (k, ε)*-seeded extractor.*

Lemma 6.3. *For all* k, ε*, there exists an* R*-output-light* (k, ε)*-seeded extractor* $\mathsf{Ext} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ *with* $d = \log(n) + 3\log(1/\varepsilon) + O(1), m = k + \log(n) - \log(1/\varepsilon)$ *and* $R = O\left(\frac{N}{K\varepsilon^4} + \sqrt{\frac{N}{K\varepsilon^4}} \cdot (k + \log n - \log(1/\varepsilon))\right)$*, where* $N = 2^n$ *and* $K = 2^k$.

[2] see Definition 4.6 for a definition

Before proving this lemma, we first show how to use this to plugin such extractors to get a condenser.

Lemma 6.4. *Assume there exists an R-output-light (k,ε)-seeded extractor* $\mathsf{Ext} : \{0,1\}^{2n} \times \{0,1\}^d \to \{0,1\}^t$. *There exists a condenser* $\mathsf{Cond} : (\{0,1\}^n)^3 \to \{0,1\}^t$ *such that for any uniform $(2,3)$-SHELA source $\mathbf{X}$, $H^{\varepsilon}_{\infty}(\mathsf{Cond}(\mathbf{X})) \geq 2n - \log(R) - \log(1/\varepsilon) - O(1)$.*

Assuming these lemmas, our proof of main theorem is straightforward:

Proof of Theorem 6.1. We consider the extractor from Lemma 6.3 and use that to get the condenser from Lemma 6.4. We note that for the setting of Lemma 6.4, $R = \frac{CN}{\varepsilon^4}$ for a large constant C. □

We now show that we indeed can get a condenser from such an output-light extractor.

Proof of Lemma 6.4. By Proposition 4.12, SHELA sources are convex combination of fixed-index SHELA sources and so it suffices to show we can condense from the latter. Let k be the output min entropy we will guarantee. Let $N = 2^n, T = 2^t, K = 2^k$. We identify $\{0,1\}^n, \{0,1\}^t$ with $[N], [T]$ respectively and use them interchangeably in this proof.

Let $g : \{0,1\}^{3n} \to \{0,1\}^t$ be a function that on input (x_1, x_2, x_3), takes an appropriate length prefix $x_{3,pre}$ of x_3 and outputs $\mathsf{Ext}(x_1 \circ x_2, x_{3,pre})$. We claim that g will be a (k,ε)-condenser for uniform $(2,3)$-SHELA sources. By the property of the extractor, we are guaranteed that for each $z \in [T]$, $\left|\{(x_1,x_2) \in \{0,1\}^{2n} : \exists y \in \{0,1\}^d (g(x_1,x_2,y) = z)\}\right| < R$. We set $K = \frac{\varepsilon N^2}{2R}$. We take cases on positions $1,2,3$ and show show that g will indeed be a condenser for uniform $(2,3)$-SHELA sources with adversary at that position.

Case 1. The adversary is in position 3. We proceed by contradiction and assume there exists $\mathbf{X}$ with adversary in position 3 so that $H^{\varepsilon}_{\infty}(g(\mathbf{X})) < k$. By Claim 4.5, there exists $D \subset \text{support}(g(\mathbf{X}))$ such that $|D| \leq K$ and $\Pr[g(\mathbf{X}) \in D] \geq \varepsilon$. This implies there exists $z \in [T]$ and $P \subset [N]^2$ such that $|P| \geq \frac{\varepsilon N^2}{K}$ and for all $(x_1,x_2) \in P$ it holds that there exists $y \in \{0,1\}^n$ such that $g(x_1,x_2,y) = z$. As $\frac{\varepsilon N^2}{K} > R$. However, this is a contradiction to the fact that Ext is R-output-light.

Case 2. The adversary is in position 1 or 2. We prove the stronger claim that the output will be uniform. Let $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3)$ be a uniform $(2,3)$-SHELA source with adversary in position 1 or 2. Let $\mathbf{Y} = (\mathbf{X}_1, \mathbf{X}_2)$. Then, $H_{\infty}(\mathbf{Y}) = n$. As, $\mathbf{X}_3$ is uniform, the seed passed to f will also be uniform. As f is an extractor, $|f(\mathbf{Y}, \mathbf{U}_d) - U_t| \leq \varepsilon$ as desired.

□

Remark 6.5. *We note that in the above proof, the fact that* Ext *is a seeded extractor is utilized in Case 2. In particular, we just need that Ext works for sources $(\mathbf{X}_1 \circ \mathbf{X}_2)$, where one of them is guaranteed to be uniform. Such sources are called as* somewhere-random sources.

We are now ready to prove that a random process generates the desired extractor:

Proof of Lemma 6.3. Let $N = 2^n, M = 2^m, K = 2^k, D = 2^d$. We also identify $\{0,1\}^n, \{0,1\}^m, \{0,1\}^d$ with $[N], [M], [D]$ respectively and use these interchangeably in this proof. We will set D, M, R as follows:

$$D = \frac{10^4 \log(N)}{\varepsilon^3}, M = \varepsilon K \log(N), R = \frac{10^6 N}{\varepsilon^4 K} + \frac{500\sqrt{N \log(\varepsilon K \log(N))}}{\varepsilon^2 \sqrt{K}}$$

Algorithm 3: Random process to generate f

for $1 \leq i \leq N$ **do**
  $S_i = \emptyset$
  for $m \in [M]$ **do**
    Add m to S_i with probability p
  end
end
Define $f : [N] \times [D] \to [M]$ as follows:
On input x_1, x_2: output element at index $(x_2 \mod |S_{x_1}|)$ from S_{x_1}.

$p \leq \frac{R}{100N}$	$p \geq \frac{\log\left(\frac{2eN}{\varepsilon K}\right)}{\varepsilon M \log\left(\frac{\varepsilon M}{4eL_{big}}\right)}$	$p \geq \frac{6\log N}{\gamma^2 M}$	$p \geq \frac{16}{K\varepsilon^2 \log\left(\frac{\varepsilon M}{4eL_{big}}\right)}$
$p \leq \frac{R^2}{24N\log(M)}$	$p \geq \frac{192M}{\varepsilon^3 K L_{big}}$	$p \geq \frac{96\log\left(\frac{2eN}{\varepsilon K}\right)}{\varepsilon^2 L_{big}}$	$\gamma \leq \frac{\varepsilon}{5}$

Table 1: Constraints satisfied by the parameters

We introduce the quantity $p = \frac{6000}{\varepsilon^4 K}$ that we will also use in Algorithm 3. We consider the random process in Algorithm 3 that generates function f as our candidate extractor.

Notice that the function f itself is deterministic as the sets S_i are only sampled once to define f. We claim that $1-o(1)$ fraction of functions f generated by Algorithm 3 will be such extractors. For the analysis, we introduce two quantities:

$$\gamma = \frac{\varepsilon}{10}, L_{big} = \frac{\varepsilon M}{4e^2}$$

We observe that our setting of parameters has ensured the constraints in Table 1 are met.

We first show that that with $1 - o(1)$ probability, the seed length of f will indeed be small:

Claim 6.6. *With probability $o(1)$ over the random process above, there exists $i \in [N] : ||S_i| - pM| \leq \gamma pM$. Hence, with $1 - o(1)$ probability, the sizes of all these sets will be at most $(1+\gamma)pM \leq D$, and so, the seed length will be as desired.*

Proof. For $i \in [N]$, let E_i be the event that $||S_i| - pM| \geq \gamma pM$. We show that for all $i \in [N]$, the probability that E_i occurs is small. Indeed, applying the Chernoff bound from Claim 4.2 with $\delta = \gamma$ and $\mu = pM$, we infer that this event happens with probability at most

$$\exp\left(-\frac{\gamma^2 pM}{3}\right) \leq \exp(-2\log(N)) \leq \frac{1}{N^2}$$

where the first inequality follows because $p \geq \frac{6\log N}{\gamma^2 M}$. We finally do a union bound over all $(i) \in [N]$ to infer the claim. □

We now show the moreover part of the extractor, i.e., there is no popular output element:

Claim 6.7. *With at most $o(1)$ probability over the sampling process to generate f, there exists $z \in [M]$ and $P \subset [N]$ such that $|P| \geq R$ and for all $i \in P, z \in S_i$.*

Proof. For $z \in [M]$, let E_z be the event that there exists $P \subset [N]$ such that $|P| \geq R$ and for all $i \in P$, it holds that $z \in S_i$. We show that for all $z \in [M]$, the probability that E_z occurs is small. We will then do a union bound over all $z \in [M]$ to infer the claim. Applying the Chernoff bound from Claim 4.2 with $1 + \delta = R$ and $\mu = pN$, we infer that:

$$\begin{aligned}
\Pr[E_t] &\leq \exp\left(-\frac{\left(\frac{R}{pN} - 1\right)^2 pN}{3}\right) \\
&\leq \exp\left(-\frac{\left(\frac{R}{2pN}\right)^2 pN}{3}\right) \\
&\leq \exp\left(\frac{-R^2}{12pN}\right) \\
&\leq \frac{1}{M^2}
\end{aligned}$$

where the first inequality follows because $R \geq 2pN$, and the fourth inequality follows because $R^2 \geq 24pN\log(M)$. Hence, we can indeed do a union bound over all M elements to infer the claim. □

We now show that with $1 - o(1)$ probability, the f generated will indeed be an extractor. To do this, we introduce the following two bad events and show that if the function is not an extractor, then one of the following two bad cases must occur. We later show that each of these events occur with small probability.

1. Let E_1 be the event that there exist $D \subset [M]$ with $|D| = L \leq L_{big}$, and $P \subset [N]$ with $|P| \geq \frac{\varepsilon K}{2}$ such that for all $i \in P : |S_i \cap D| \geq \frac{\varepsilon pM}{4}$.
2. Let E_2 be the event that there exist $D \subset [M]$ with $|D| = L \geq L_{big}$, and $P \subset [N]$ with $|P| \geq \frac{\varepsilon K}{2}$ such that for all $i \in P : |S_i \cap D| \geq \left(1 + \frac{\varepsilon}{4}\right) pL$.

We will show that all these bad events happen with very small probability:

Claim 6.8. $\Pr[E_1] \leq o(1)$.

Claim 6.9. $\Pr[E_2] \leq o(1)$.

Assuming these claims, we show how to prove that a random f will indeed be an extractor. It suffices to show that with $1 - o(1)$ probability over sampling S_i the following holds: For every $I \subset [N]$ with $|I| = K$, if we sample a random i from I and output a random element from S_i, then the resultant distribution will be ε close to the uniform distribution over $[M]$. Consider arbitrary $I \subset [N]$ with $|I| = K$ and let the corresponding K sets in I be $R_1, \ldots, R_K$. We proceed by contradiction and assume there exists $D \subset [M]$ such that $\Pr[f(\mathbf{X}) \in D] \geq \varepsilon + \frac{|D|}{M}$. For $1 \leq i \leq K$, let $Y_i = \frac{|D \cap R_i|}{|R_i|}$. Let $\mathbf{Y}$ be the random variable which samples random $k \in [K]$ and outputs Y_k. Then, by assumption, $\mathbb{E}[\mathbf{Y}] = \varepsilon + \frac{|D|}{M}$. Applying Claim 4.1, we infer that $\Pr\left[\mathbf{Y} > \frac{|D|}{M} + \frac{\varepsilon}{2}\right] \geq \frac{\varepsilon}{2}$. Hence, there exists $B \subset [N]$ with $|B| \geq \frac{\varepsilon K}{2}$ such that for all $i \in B$, $\frac{|S_i \cap D|}{|S_i|} \geq \frac{\varepsilon}{2} + \frac{|D|}{M}$. We apply Claim 6.6, to infer that with $1 - o(1)$ probability, it will be that for all $i \in B : |S_i \cap D| \geq \frac{\varepsilon}{2}(1-\gamma)pM + (1-\gamma)p\,|D|$. We consider cases on $|D| = L$:

Case 1. $L \leq L_{big}$. We see that

$$\frac{\varepsilon(1-\gamma)pM}{2} + (1-\gamma)pL \geq \frac{\varepsilon pM}{4}$$

where the inequality follows because $\gamma \leq \frac{1}{2}$. Hence, for all $i \in B : |S_i \cap D| \geq \frac{\varepsilon pM}{4}$. Thus, the event E_1 must have occurred. As this happens with $o(1)$ probability, we indeed infer the claim.

Case 2. $L \geq L_{big}$. We see that

$$\frac{\varepsilon(1-\gamma)pM}{2} + (1-\gamma)pL \geq \left(1+\frac{\varepsilon}{2}\right)(1-\gamma)pL \geq \left(1+\frac{\varepsilon}{4}\right)pL$$

where the first inequality follows trivially as $M \geq L$ and second inequality follows because $\gamma \leq \frac{\varepsilon}{5}$. Hence, for all $i \in B : |S_i \cap D| \geq \left(1+\frac{\varepsilon}{4}\right)pL$. Thus, the event E_2 must have occurred. As this happens with $o(1)$ probability, we indeed infer the claim.

We now prove our various claims that the bad events indeed occur with $o(1)$ probability:

Proof of Claim 6.8. For $i \in [N]$ and $D \subset [M]$ with $|D| = L \leq L_{big}$, let $E_{1,D,i}$ be the event that $|S_i \cap D| \geq \frac{\varepsilon}{4}pM$. For fixed $D \subset [M]$ with $|D| = L \leq L_{big}$, let $E_{1,D}$ be the event that there exists $P \subset [N]$ with $|P| \geq \frac{\varepsilon K}{2}$ such that for all $i \in P : |S_i \cap D| \geq \frac{\varepsilon}{4}pM$. We first show that for all $i, D : \Pr[E_{1,D,i}]$ is small. Indeed, applying the Chernoff bound from Claim 4.2 with $1+\delta = \frac{\varepsilon T}{4L}$ and $\mu = pL$, we infer that:

$$\Pr[E_{1,D,i}] \leq \exp\left(-\frac{\varepsilon pM}{4}\log\left(\frac{\varepsilon M}{4eL}\right)\right) \leq \exp\left(-\frac{\varepsilon pM}{4}\log\left(\frac{\varepsilon M}{4eL_{big}}\right)\right)$$

where the last inequality follows because $L \leq L_{big}$. We use this to show that for all $D : \Pr[E_{1,D}]$ is small. Then, we will do a union bound over $D \subset [M]$ with $|D| \leq L_{big}$ to infer that $\Pr[E_1]$ is small. Let $q = \exp\left(-\frac{\varepsilon pM}{4}\log\left(\frac{\varepsilon M}{4eL_{big}}\right)\right)$. Applying the Chernoff bound from Claim 4.2 with $1+\delta = \frac{\varepsilon K}{2qN}$ and $\mu = qN$, we infer that:

$$\begin{aligned}
\Pr[E_{1,D}] &\leq \exp\left(-\frac{\varepsilon K}{2}\log\left(\frac{\varepsilon K}{2eqN}\right)\right) \\
&= \exp\left(-\frac{\varepsilon^2 pKM}{8}\log\left(\frac{\varepsilon M}{4eL_{big}}\right) + \frac{\varepsilon K}{2}\log\left(\frac{2eN}{\varepsilon K}\right)\right) \\
&\leq \exp\left(-\frac{\varepsilon^2 pKM}{16}\log\left(\frac{\varepsilon M}{4eL_{big}}\right)\right) \\
&\leq \exp\left(-M\right)
\end{aligned}$$

where the third inequality follows because $p \geq \frac{\log\left(\frac{2eN}{\varepsilon K}\right)}{\varepsilon M \log\left(\frac{\varepsilon M}{4eL_{big}}\right)}$, and the fourth inequality follows because $p \geq \frac{16}{K\varepsilon^2 \log\left(\frac{\varepsilon M}{4eL_{big}}\right)}$. We finally do a union bound over all $D \subset [M]$ with $|D| \leq L_{big}$. As there are at most 2^M such sets, the union bound indeed succeeds and $\Pr[E_1] \leq o(1)$ as desired. □

Proof of Claim 6.9. For $i \in [N]$ and $D \subset [M]$ with $|D| = L \geq L_{big}$, let $E_{2,D,i}$ be the event that $|S_i \cap D| \geq \left(1 + \frac{\varepsilon}{4}\right) pL$. For fixed $D \subset [M]$ with $|D| = L \geq L_{big}$, let $E_{2,D}$ be the event that there exists $P \subset [N]$ with $|P| \geq \frac{\varepsilon K}{2}$ such that for all $i \in P : |S_i \cap D| \geq \left(1 + \frac{\varepsilon}{4}\right) pL$.

We apply the Chernoff bound from Claim 4.2 with $\mu = pL, 1 + \delta = 1 + \frac{\varepsilon}{4}$ to infer that

$$\Pr[E_{2,D,i}] \leq \exp\left(-\frac{\varepsilon^2 pL}{48}\right) \leq \exp\left(-\frac{\varepsilon^2 pL_{big}}{48}\right)$$

We use this to show that, for all D, $\Pr[E_{2,D}]$ is small. Then, we will do a union bound over $D \subset [M]$ with $|D| \geq L_{big}$ to infer that $\Pr[E_2]$ is small. Let $q = \exp\left(-\frac{\varepsilon^2 pL_{big}}{48}\right)$. Applying the Chernoff bound from Claim 4.2 with $\mu = qN$ and $1 + \delta = \frac{\varepsilon K}{qN}$, we infer that:

$$\begin{aligned}
\Pr[E_{2,D}] &\leq \exp\left(-\frac{\varepsilon K}{2} \log\left(\frac{\varepsilon K}{2eqN}\right)\right) \\
&= \exp\left(-\frac{\varepsilon^3 KpL_{big}}{96} + \frac{\varepsilon K}{2} \log\left(\frac{2eN}{\varepsilon K}\right)\right) \\
&\leq \exp\left(-\frac{\varepsilon^3 KpL_{big}}{192}\right) \\
&\leq \exp\left(-M\right)
\end{aligned}$$

where the third inequality follows because $p \geq \frac{96}{\varepsilon^2 L_{big}} \log\left(\frac{2eN}{\varepsilon K}\right)$, and the fourth inequality follows because $p \geq \frac{192M}{\varepsilon^3 KL_{big}}$. We finally do a union bound over all $D \subset [M]$ with $|D| \geq L_{big}$. As there are at most 2^M such sets, the union bound indeed succeeds and $\Pr[E_2] \leq o(1)$ as desired. □

□

6.2 An Explicit Condenser for Uniform (2, 3)-SHELA Sources

In this section, we construct a condenser for Uniform $(2, 3)$ SHELA sources. The following is our main result.

Theorem 6.10. *There exists constant $0 < c_0 < 1$ such that for all $\varepsilon > 2^{-c_0 n}$, we can explicitly construct a condenser $\mathsf{Cond} : (\{0,1\}^n)^3 \to \{0,1\}^t$, where $t = \frac{n}{16}$ such that for any uniform $(2,3)$-SHELA source $\mathbf{X}$, $H^{\varepsilon}_{\infty}(\mathsf{Cond}(\mathbf{X})) \geq t - O(\log(t/\varepsilon))$.*

To prove this, we construct an explicit output-light seeded extractor (see Definition 6.2) that works for somewhere-random sources. We note that by Remark 6.5, this is sufficient to use Lemma 6.4 to get the claimed condenser in Theorem 6.10.

Theorem 6.11. *There exists constant $0 < c_0 < 1$ such that for all $\varepsilon > 2^{-c_0 n}$, there exists a R-output-light strong linear seeded ε-extractor $\mathsf{Ext} : \{0,1\}^{2n} \times \{0,1\}^d \to \{0,1\}^m$ for the class of distributions $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2)$, each $\mathbf{X}_i$ being a r.v. on n bits and at least one of $\mathbf{X}_1$ or $\mathbf{X}_2$ is guaranteed to be uniform, with $d = O(\log n/\varepsilon), m = \frac{n}{16}$ and $R = \frac{2^{2n-m}}{\mathrm{poly}(m,1/\varepsilon)}$.*

We note that this construction matches the probabilistic bounds (Lemma 6.3) as the t bit output is condensed to entropy $t - O(\log(t))$ with $t = O(n)$. We also remark that we have not tried to optimize the constant appearing in the output length of the extractor.

6.2.1 An Explicit Output-Light Seeded Extractor for Somewhere-Random Sources

We prove Theorem 6.11 in this section and show

Algorithm 4: Ext (Output-light Somewhere-extractor)

Input: source $X = (X_1, X_2) \in \{0,1\}^n \times \{0,1\}^n$, seed $S \in \{0,1\}^d$

Let $\mathsf{Ext}_{GUV} : \{0,1\}^{7n/4} \times \{0,1\}^d \to \{0,1\}^{n/2-\log(1/\varepsilon_0)}$ be the GUV extractor from Theorem 4.7 instantiated for entropy $3n/4$ and error $\varepsilon_0 = \varepsilon/4$.

Let $U = X_1, V = X_2$.
Let $n_1 = \frac{n}{4}, n_2 = \frac{7n}{4}$.
Let $Y = (Y_1, Y_2)$ where $Y_1 = \left(U_{[1,n_1/2]}, V_{[1,n_1/2]}\right), Y_2 = \left(U_{[(n_1/2)+1,n]}, V_{[(n_1/2)+1,n]}\right)$.
Let $R_2 = \mathsf{Ext}_{GUV}(Y_2, S)$.
Let R_2' be a length $n/4$ prefix of R_2 with last bit set to 1. Let $R_1 \in \{0,1\}^{n/16} = R_2' \cdot Y_1$ where the operation is done over the finite field $\mathbb{F}^4_{2^{n/16}}$.
Output R_1.

Proof of Theorem 6.11. We claim that Ext computed by Algorithm 4 computes the desired extractor.

Let $\mathbf{Y} = (\mathbf{Y}_1, \mathbf{Y}_2)$ be the distribution of the variable Y above. Let $\mathbf{R}_1, \mathbf{R}_2, \mathbf{R}_2'$ be the distribution of the variables R_1, R_2, R_2' above. We will show that $\mathbf{Y}$ is ε_0 close to being a block source. As either $\mathbf{X}_1$ or $\mathbf{X}_2$ is guaranteed to be uniform, $H_\infty(\mathbf{Y}_i) \geq \frac{n_i}{2}$. By the min-entropy chain rule Lemma 4.4, with probability $1-\varepsilon_0$ over fixings of $\mathbf{Y}_1 = \alpha$, it holds that $H_\infty(\mathbf{Y}_2 \restriction_{\mathbf{Y}_1=\alpha}) \geq \frac{n_2}{2} - n_1 - \log(1/\varepsilon_0) = \frac{3n}{4} - \log(1/\varepsilon_0)$. We will add ε_0 to our total error and assume this property about $\mathbf{Y}$ from here on. By property of Ext_{GUV}, it holds that, $\left|\mathbf{R}_2 - \mathbf{U}_{|\mathbf{R}_2|}\right| \leq \varepsilon_0$. We will add ε_0 to our total error and assume $\mathbf{R}_2$ is uniform from here on. So, $\mathbf{R}_2'$ is a distribution over $\{0,1\}^{n/4}$ with min entropy $\frac{n}{4} - 1$. As $\mathbf{Y}_1 \sim \{0,1\}^{n/4}$ is such that $H_\infty(\mathbf{Y}_1) \geq \frac{n}{8}$, by Theorem 4.8, it holds that $\left|\mathbf{R}_1 - \mathbf{U}_{|\mathbf{R}_1|}\right| \leq 2^{-n/32+1}$. As $\mathbf{Y}$ is a block source, for each fixing α of $\mathbf{Y}_1$, it holds that:

$$\left|\mathsf{Ext}_{GUV}(\mathbf{Y}_2, S) - \mathbf{U}_{|\mathbf{R}_2|}\right| \leq \varepsilon_0$$

Hence, it must be that

$$\left|(\mathbf{Y}_1, \mathsf{Ext}_{GUV}(\mathbf{Y}_2, S)) - (\mathbf{Y}_1, \mathbf{U}_{|\mathbf{R}_2|})\right| \leq \varepsilon_0$$

and thus,

$$\left|\mathbf{R}_1 - \mathbf{U}_{|\mathbf{R}_1|}\right| \leq 2\varepsilon_0 + 2^{-n/32+1} \leq 3\varepsilon_0,$$

using the fact that $\varepsilon \geq 2^{-c_0 n}$, for some small $c_0 > 0$. The total error of the extractor on input $\mathbf{X}$ is thus bounded by $4\varepsilon_0 = \varepsilon$, as desired.

We now prove that this extractor is indeed output-light. For every fixing of the output R_1 of $\mathbf{R}_1$, β of $\mathbf{Y}_2$ and the seed S, we can uniquely recover R_2'. Given $\frac{3n}{16}$ bits corresponding to first three out of the 4 intermediate outputs of the inner product, we can use R_1 to compute the fourth intermediate outer product and then use R_2' to invert each of the products and recover R_1. Thus for a fixed seed S and output R_1, there can be at most $2^{3n/16+7n/4} = 2^{31n/16}$ such $x \in \{0,1\}^{2n}$ so that $\mathsf{Ext}(x,s) = z$. As there are at most 2^d seeds, for a fixed output $R_1 \in \{0,1\}^{n/16}$, $\left|\{x \in \{0,1\}^{2n} : \exists y(\mathsf{Ext}(x,y)) = z\}\right| \leq 2^{2n-n/16-\log(n/\varepsilon)} = \frac{2^{2n-m}}{\mathrm{poly}(n,1/\varepsilon)} = \frac{2^{2n-m}}{\mathrm{poly}(m,1/\varepsilon)}$. □

References

[AORSV20] Divesh Aggarwal, Maciej Obremski, João Ribeiro, Luisa Siniscalchi, and Ivan Visconti. "How to Extract Useful Randomness from Unreliable Sources". en. In: *Advances in Cryptology – EUROCRYPT 2020*. Ed. by Anne Canteaut and Yuval Ishai. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2020, pp. 343–372. ISBN: 978-3-030-45721-1. DOI: 10.1007/978-3-030-45721-1_13 (cit. on pp. 4, 5, 7, 16).

[AL93] Miklós Ajtai and Nathan Linial. "The influence of large coalitions". en. In: *Combinatorica* 13.2 (June 1993), pp. 129–145. ISSN: 1439-6912. DOI: 10.1007/BF01303199 (cit. on p. 6).

[BGM22] Marshall Ball, Oded Goldreich, and Tal Malkin. "Randomness Extraction from Somewhat Dependent Sources". In: *13th Innovations in Theoretical Computer Science Conference (ITCS 2022)*. Ed. by Mark Braverman. Vol. 215. Leibniz International Proceedings in Informatics (LIPIcs). ISSN: 1868-8969. Dagstuhl, Germany: Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2022, 12:1–12:14. ISBN: 978-3-95977-217-4. DOI: 10.4230/LIPIcs.ITCS.2022.12 (cit. on p. 4).

[BCDT19] Avraham Ben-Aroya, Gil Cohen, Dean Doron, and Amnon Ta-Shma. "Two-Source Condensers with Low Error and Small Entropy Gap via Entropy-Resilient Functions". en. In: (2019). Artwork Size: 20 pages Medium: application/pdf Publisher: Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik GmbH, Wadern/Saarbruecken, Germany Version Number: 1.0, 20 pages. DOI: 10.4230/LIPICS.APPROX-RANDOM.2019.43 (cit. on p. 4).

[BDT17] Avraham Ben-Aroya, Dean Doron, and Amnon Ta-Shma. "An efficient reduction from two-source to non-malleable extractors: achieving near-logarithmic min-entropy". In: *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*. STOC 2017. New York, NY, USA: Association for Computing Machinery, June 2017, pp. 1185–1194. ISBN: 978-1-4503-4528-6. DOI: 10.1145/3055399.3055423 (cit. on p. 4).

[BL85] Michael Ben-Or and Nathan Linial. "Collective coin flipping, robust voting schemes and minima of Banzhaf values". In: *26th Annual Symposium on Foundations of Computer Science (sfcs 1985)*. ISSN: 0272-5428. Oct. 1985, pp. 408–416. DOI: 10.1109/SFCS.1985.15 (cit. on p. 6).

[BKKKL92] Jean Bourgain, Jeff Kahn, Gil Kalai, Yitzhak Katznelson, and Nathan Linial. "The influence of variables in product spaces". en. In: *Israel Journal of Mathematics* 77.1 (Feb. 1992), pp. 55–64. ISSN: 1565-8511. DOI: 10.1007/BF02808010 (cit. on pp. 4, 6, 13).

[CRVW02] Michael Capalbo, Omer Reingold, Salil Vadhan, and Avi Wigderson. "Randomness conductors and constant-degree lossless expanders". In: *Proceedings of the thiry-fourth annual ACM symposium on Theory of computing*. STOC '02. New York, NY, USA: Association for Computing Machinery, May 2002, pp. 659–668. ISBN: 978-1-58113-495-7. DOI: 10.1145/509907.510003 (cit. on p. 7).

[Cha16] Eshan Chattopadhyay. "Explicit two-source extractors and more". en. In: (May 2016) (cit. on p. 15).

[CZ19] Eshan Chattopadhyay and David Zuckerman. "Explicit two-source extractors and resilient functions". In: *Annals of Mathematics* 189.3 (May 2019). Publisher: Department of Mathematics of Princeton University, pp. 653–705. ISSN: 0003-486X, 1939-8980. DOI: 10.4007/annals.2019.189.3.1 (cit. on p. 6).

[CG88] Benny Chor and Oded Goldreich. "Unbiased Bits from Sources of Weak Randomness and Probabilistic Communication Complexity". In: *SIAM Journal on Computing* 17.2 (Apr. 1988). Publisher: Society for Industrial and Applied Mathematics, pp. 230–261. ISSN: 0097-5397. DOI: 10.1137/0217015 (cit. on pp. 3–5, 15).

[CGHFRS85] Benny Chor, Oded Goldreich, Johan Hasted, Joel Freidmann, Steven Rudich, and Roman Smolensky. "The bit extraction problem or t-resilient functions". In: *26th Annual Symposium on Foundations of Computer Science (sfcs 1985)*. ISSN: 0272-5428. Oct. 1985, pp. 396–407. DOI: 10.1109/SFCS.1985.55 (cit. on p. 4).

[DOPS04] Y. Dodis, Shien Jin Ong, M. Prabhakaran, and A. Sahai. "On the (im)possibility of cryptography with imperfect randomness". In: *45th Annual IEEE Symposium on Foundations of Computer Science*. ISSN: 0272-5428. Oct. 2004, pp. 196–205. DOI: 10.1109/FOCS.2004.44 (cit. on p. 3).

[DMOZ23] Dean Doron, Dana Moshkovitz, Justin Oh, and David Zuckerman. "Almost Chor-Goldreich Sources and Adversarial Random Walks". In: *Proceedings of the 55th Annual ACM Symposium on Theory of Computing*. STOC 2023. New York, NY, USA: Association for Computing Machinery, June 2023, pp. 1–9. ISBN: 978-1-4503-9913-5. DOI: 10.1145/3564246.3585134 (cit. on pp. 4, 5, 7, 17).

[DKSS13] Zeev Dvir, Swastik Kopparty, Shubhangi Saraf, and Madhu Sudan. "Extensions to the Method of Multiplicities, with Applications to Kakeya Sets and Mergers". In: *SIAM Journal on Computing* 42.6 (Jan. 2013). Publisher: Society for Industrial and Applied Mathematics, pp. 2305–2328. ISSN: 0097-5397. DOI: 10.1137/100783704 (cit. on p. 3).

[DW11] Zeev Dvir and Avi Wigderson. "Kakeya Sets, New Mergers, and Old Extractors". In: *SIAM J. Comput.* 40.3 (2011), pp. 778–792. DOI: 10.1137/090748731 (cit. on p. 3).

[GP20] Dmitry Gavinsky and Pavel Pudlák. "Santha-Vazirani sources, deterministic condensers and very strong extractors". en. In: *Theory of Computing Systems* 64.6 (Aug. 2020), pp. 1140–1154. ISSN: 1433-0490. DOI: 10.1007/s00224-020-09975-8 (cit. on p. 7).

[GUV09] Venkatesan Guruswami, Christopher Umans, and Salil Vadhan. "Unbalanced expanders and randomness extractors from Parvaresh–Vardy codes". In: *Journal of the ACM* 56.4 (July 2009), 20:1–20:34. ISSN: 0004-5411. DOI: 10.1145/1538902.1538904 (cit. on pp. 4, 15).

[ILL89] Russell Impagliazzo, Leonid A. Levin, and Michael Luby. "Pseudo-random Generation from one-way functions (Extended Abstracts)". In: *Proceedings of the 21st Annual ACM Symposium on Theory of Computing, May 14-17, 1989, Seattle, Washington, USA*. Ed. by David S. Johnson. ACM, 1989, pp. 12–24. DOI: 10.1145/73007.73009 (cit. on p. 15).

[KKL88] J. Kahn, G. Kalai, and N. Linial. "The influence of variables on Boolean functions". In: *[Proceedings 1988] 29th Annual Symposium on Foundations of Computer Science*. Oct. 1988, pp. 68–80. DOI: 10.1109/SFCS.1988.21923 (cit. on pp. 4, 6).

[KN23] Swastik Kopparty and Vishvajeet N. “Extracting Mergers and Projections of Partitions”. In: *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, APPROX/RANDOM 2023, September 11-13, 2023, Atlanta, Georgia, USA*. Ed. by Nicole Megow and Adam D. Smith. Vol. 275. LIPIcs. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2023, 52:1–52:22. DOI: 10.4230/LIPICS.APPROX/RANDOM.2023.52 (cit. on p. 7).

[LRVW03] Chi-Jen Lu, Omer Reingold, Salil Vadhan, and Avi Wigderson. “Extractors: optimal up to constant factors”. In: *Proceedings of the thirty-fifth annual ACM symposium on Theory of computing*. STOC ’03. New York, NY, USA: Association for Computing Machinery, June 2003, pp. 602–611. ISBN: 978-1-58113-674-6. DOI: 10.1145/780542.780630 (cit. on p. 3).

[Mek17] Raghu Meka. “Explicit Resilient Functions Matching Ajtai-Linial”. In: *Proceedings of the 2017 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*. Proceedings. Society for Industrial and Applied Mathematics, Jan. 2017, pp. 1132–1148. DOI: 10.1137/1.9781611974782.73 (cit. on p. 6).

[RSW06] Omer Reingold, Ronen Shaltiel, and Avi Wigderson. “Extracting Randomness via Repeated Condensing”. In: *SIAM Journal on Computing* 35.5 (Jan. 2006). Publisher: Society for Industrial and Applied Mathematics, pp. 1185–1209. ISSN: 0097-5397. DOI: 10.1137/S0097539703431032 (cit. on p. 4).

[RVW04] Omer Reingold, Salil Vadhan, and Avi Wigderson. *A Note on Extracting Randomness from Santha-Vazirani Sources*. en. Tech. rep. 2004 (cit. on pp. 3, 4).

[SV86] Miklos Santha and Umesh V Vazirani. “Generating quasi-random sequences from semi-random sources”. In: *Journal of Computer and System Sciences* 33.1 (Aug. 1986), pp. 75–87. ISSN: 0022-0000. DOI: 10.1016/0022-0000(86)90044-9 (cit. on pp. 3, 4).

[TUZ07] Amnon Ta-Shma, Christopher Umans, and David Zuckerman. “Lossless Condensers, Unbalanced Expanders, And Extractors”. en. In: *Combinatorica* 27.2 (Mar. 2007), pp. 213–240. ISSN: 1439-6912. DOI: 10.1007/s00493-007-0053-2 (cit. on p. 4).

[Vad12] Salil P. Vadhan. “Pseudorandomness”. English. In: *Foundations and Trends® in Theoretical Computer Science* 7.1–3 (Dec. 2012). Publisher: Now Publishers, Inc., pp. 1–336. ISSN: 1551-305X, 1551-3068. DOI: 10.1561/0400000010 (cit. on p. 3).

[Vaz85] Umesh V. Vazirani. “Towards a Strong Communication Complexity Theory or Generating Quasi-Random Sequences from Two Communicating Slightly-random Sources (Extended Abstract)”. In: *Proceedings of the 17th Annual ACM Symposium on Theory of Computing, May 6-8, 1985, Providence, Rhode Island, USA*. Ed. by Robert Sedgewick. ACM, 1985, pp. 366–378. DOI: 10.1145/22145.22186 (cit. on p. 15).

[Zuc90] David Zuckerman. “General weak random sources”. In: *Proceedings of the 31st Annual Symposium on Foundations of Computer Science*. SFCS ’90. USA: IEEE Computer Society, Oct. 1990, 534–543 vol.2. ISBN: 978-0-8186-2082-9. DOI: 10.1109/FSCS.1990.89574 (cit. on p. 3).

[Zuc92] David Zuckerman. “Computing efficiently using weak random sources”. UMI Order No. GAX92-28930. phd. USA: University of California at Berkeley, 1992 (cit. on p. 3).

[Zuc07] David Zuckerman. “Linear Degree Extractors and the Inapproximability of Max Clique and Chromatic Number”. In: *Theory of Computing* 3 (Aug. 2007). Number: 6 Publisher: Theory of Computing, pp. 103–128. DOI: 10.4086/toc.2007.v003a006 (cit. on pp. 4, 15).

ECCC ISSN 1433-8092

https://eccc.weizmann.ac.il