Electronic Colloquium on Computational Complexity, Revision 2 of Report No. 171 (2024)

Condensing and Extracting Against Online Adversaries

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

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

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

Rocco Servedio‡
Columbia University
rocco@cs.columbia.edu

Abstract

We investigate the tasks of deterministically condensing and extracting randomness from Online Non-Oblivious Symbol Fixing (oNOSF) sources, a natural model of defective random sources for which it is known that extraction is impossible in many parameter regimes [AORSV, EUROCRYPT'20]. A (g, ℓ)-oNOSF source is a sequence of ℓ blocks $\mathbf{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_\ell) \sim (\{0,1\}^n)^\ell$, where at least g of the blocks are *good* (are independent and have some min-entropy), and the remaining *bad* blocks are controlled by an *online adversary* where each bad block can be arbitrarily correlated with any block that appears before it.

The existence of condensers (in regimes where extraction is impossible) was recently studied in [CGR, FOCS'24]. They proved condensing impossibility results for various values of g and ℓ, and they showed the existence of condensers matching the impossibility results in the special case when n is exponential in ℓ (i.e., the setting of few blocks of large length).

In this work, not only do we construct the first explicit condensers matching the existential results of [CGR, FOCS'24], but we make a doubly exponential improvement by handling the case when n is only polylogarithmic in ℓ. We also obtain a much improved explicit construction for transforming low-entropy oNOSF sources (where the good blocks only have min-entropy, as opposed to being uniform) into uniform oNOSF sources.

As our next result, we essentially resolve the question of the existence of condensers for oNOSF sources by showing the existence of condensers in almost all parameter regimes, even when n is a large enough constant and ℓ is growing.

We find interesting connections and applications of our results on condensers to collective coin flipping and collective sampling, problems that are well-studied in fault-tolerant distributed computing. We use our condensers to provide very simple protocols for these problems.

Next, we turn to understanding the possibility of extraction from oNOSF sources. For proving lower bounds, we introduce and initiate a systematic study of a new, natural notion of the influence of functions, which we call *online influence*, and establish tight bounds on the total online influence of functions, which imply extraction lower bounds. Lastly, we give explicit extractor constructions for oNOSF sources using novel connections to leader election protocols, and we further construct the required leader election protocols. These extractor constructions achieve parameters that go beyond the standard resilient functions of [AL, Combinatorica'93].

*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.
‡Supported by NSF Award CCF-2106429 and NSF Award CCF-2211238.

ISSN 1433-8092

Contents

1 Introduction

Randomness is extremely useful in computation with wide-ranging applications in algorithm design, cryptography, distributed computing protocols, machine learning, error-correcting codes, and much more [MR95, Vad12]. Most of these applications require access to high quality randomness. However in a lot of settings, especially arising in practice, algorithms only have access to low quality source of randomness. This motivates the notion of *condensers*: functions that transform weak random sources into strong random sources that are of *better quality*.

The standard way of measuring the amount of randomness is using min-entropy. Formally, for a source (distribution) $\mathbf{X}$ with support Ω, define its min-entropy as $H_\infty(\mathbf{X}) = \min_{x\in\Omega} \log_2(1/\Pr[\mathbf{X} = x])$. We will also need the notion of smooth min-entropy, which measures how close a distribution is to having high entropy. Formally, for a source $\mathbf{X}$, its smooth min-entropy with parameter ε is defined as $H^\varepsilon_\infty(\mathbf{X}) = \max_{\mathbf{Y}:|\mathbf{X}-\mathbf{Y}|\le\varepsilon}\{H_\infty(\mathbf{Y})\}$, where $|\cdot|$ denotes the statistical distance (Definition 4.1).

With this, we are ready to formally define *deterministic condensers*:

Definition 1.1. *A function* $\mathsf{Cond} : \{0,1\}^n \to \{0,1\}^m$ *is a* $(k_{in}, k_{out}, \varepsilon)$-condenser for a family of distributions $\mathcal{X}$ *if for all* $\mathbf{X} \in \mathcal{X}$ *with* $\mathbf{X} \sim \{0,1\}^n$ *and* $H_\infty(\mathsf{Cond}(\mathbf{X})) \ge k_{in}$*, we have that* $H^\varepsilon_\infty(\mathbf{X}) \ge k_{out}$*.*

We say $\frac{k_{in}}{n}$ *is the* input entropy rate*,* $\frac{k_{out}}{m}$ *is the* output entropy rate*, and* $m - k_{out}$ *is the* entropy gap *of* Cond*.*

The task of the condenser is to make the output entropy rate as high as possible compared to the input entropy rate; i.e., to make the output distribution more "condensed". Related to this, it is also desirable to have as small entropy gap as possible. Condensers with entropy gap 0 are known as *randomness extractors* and have been extensively studied in theoretical computer science.

When $\mathcal{X}$ is the family of all distributions, it is folklore that no non-trivial condensing is possible.[1] So, we additionally assume that $\mathcal{X}$ is a structured family of sources.[2] Since extractors are the highest quality condensers, a significant amount of work has focused on constructing extractors for many interesting families of sources [TV00, CZ19, DGW09, KZ07]. However, for many natural family of sources, one can provably show that no extractor can exist.

In this work, we focus on one natural family of sources where it is known that extraction is impossible (for many interesting parameter regimes): online non-oblivious symbol fixing sources (oNOSF sources).[3] Formally:

Definition 1.2. *A* (g, ℓ, n, k)-oNOSF source $\mathbf{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_\ell)$ *is such that each block* $\mathbf{X}_i$ *is over* $\{0,1\}^n$*,* g *of the blocks are independent sources with min-entropy* k *("good blocks"), and each "bad block" is an arbitrary function of the blocks with an index smaller than it. When* $k = n$*, we will call such sources uniform* (g, ℓ, n)*-oNOSF sources.*

Our results at a glance The previous work of [CGR24] gave a condenser impossibility result for oNOSF sources and showed the existence of a condenser matching that result as long as the block length n was exponential in the number of blocks ℓ. We construct explicit condensers for oNOSF sources that match the results of [CGR24]; in fact, we only require n to be polylogarithmic in ℓ, providing a doubly exponential improvement over the (existential) result of [CGR24]. Next, we essentially resolve the existence question for oNOSF source condensers by showing that good condensers exist even when n is a (large) constant and ℓ grows. To go with these results, we obtain an improved construction for transforming low-entropy oNOSF

[1] Assuming $m \le n$ (wlog this holds since $|\mathsf{Cond}(\{0,1\}^n)| \le 2^n$), $m - k_{out} \ge (n - k_{in}) - \log(1/(1-\varepsilon))$ and hence the output entropy rate cannot be more than the input entropy rate without incurring extremely large error (> 0.999).

[2] A different route, that has been widely studied, is to assume access to a short independent seed. In this work, we will limit ourselves to the *deterministic or seedless setting*.

[3] These sources are in contrast to non-oblivious symbol fixing (NOSF) sources where bad blocks can be arbitrary functions of all the good blocks. NOSF sources were introduced in [CGHFRS85] with applications in leakage-resilient cryptography, and have been well-studied.

sources into uniform oNOSF sources. Moreover, we find new applications of our results on condensers to collective coin flipping and collective sampling, and use these connections to provide simple protocols for these problems. In the complementary direction, we construct explicit extractors for uniform oNOSF sources by explicitly constructing the required leader election protocols, the results of which are summarized in Tables 1 and 2. Also in the context of extractors for oNOSF sources, we introduce the new, natural notion of *online influence* for Boolean functions and show extraction lower bounds for oNOSF sources by establishing tight bounds on the total online influence of functions.

On the Utility of Condensing for oNOSF sources

We note that condensers (and sources with high min-entropy rate) are very useful: the condensed distribution can be used to efficiently simulate randomized algorithms with small overhead, perform one-shot simulations for randomized protocols, cryptography, interactive proofs, and much more. For details on these applications and more, see [AORSV20, DMOZ23, CGR24, DPW14].

Practical applications to blockchains and cryptography oNOSF sources are inspired by real-time randomness generation settings such as in blockchains where the adversary has some probability of corrupting a block. Moreover, it is known that non-corrupted blocks have some amount of min-entropy [BCG15]. In fact, several works have attempted to use Bitcoin or Ethereum as a source of randomness in cryptographic protocols [BCG15, BGZ16, PW18, BGB17]. However, the authors of [BGZ16] showed that even when the adversary has a small, constant probability of corrupting a block, randomness extraction is impossible from Bitcoin.[4] Our results show that in this setting, it is still possible to get a condensed source with a high min-entropy rate. It is known that such sources are still useful for cryptographic protocols, such as hedged public-key enryption [BBNRSSY09]. Further, there are natural cryptographic settings, such as creating a Common Reference String, that are widely used in various cryptographic protocols where oNOSF sources naturally arise [AORSV20].

Practical applications to fault-tolerant distributed computing One common scenario in distributed computing is that of many agents (e.g., servers in a network) attempting to collectively take a decision using several rounds of communication over a common broadcast channel in the presence of computationally unbounded adversarial agents, which render cryptographic primitives ineffective. Protocols for collective coin flipping, leader election, and collective sampling are prime examples of this scenario that have been intensively studied ([BL89, GGL91, Dod06, AN93, Fei99] and many more). In Section 3, we explain how condensing or extracting from oNOSF sources can be viewed as a variant of these protocols. As a consequence, our new results on condensers provide a new protocol for collective sampling and impossibility results for these protocols can be translated into lower bounds against extractors and condensers for oNOSF sources.

Organization The remainder of our introduction is structured as follows. We give an overview of previous work in Section 1.1 before presenting our main existential and explicit condenser results in Section 1.2. In Section 1.3, we present our results on the limits of extraction from oNOSF sources. Later on, in Section 3, we explain how our results on condensers have implications for collective coin flipping and sampling protocols.

[4] This mirrors our extraction impossibility result for oNOSF sources in Section 10

1.1 Previous Work

Extractors The study of extractors for oNOSF sources was initiated by [AORSV20].[5] Their results include the following:

- It is impossible to extract from uniform oNOSF sources when the fraction of good blocks is 0.99.
- An explicit transformation from $(g, \ell, n, 0.9n)$-oNOSF source into a source over $(\{0,1\}^{O(n)})^{\ell-1}$ where $g-1$ of the blocks are uniform and independent.
- An explicit transformation from $(g, \ell, n, 0.1n)$-oNOSF source into a source over $(\{0,1\}^{O(n)})^{100\ell}$ where $g-1$ of the blocks are uniform and independent.

Even though the output entropy rate is only slightly more than the input-entropy rate in the second result and smaller in the third result, the fact that a lot of the blocks are truly uniform is very useful, and they find interesting cryptographic applications of these "somewhere-extractors".

Before our work, the best known extractors for oNOSF sources could be obtained by using resilient functions or equivalently, extractors for NOSF sources (non-online version of oNOSF sources) constructed by [AL93, CZ19, Mek17, IMV23, IV24] ; these require $g \geq \ell - \frac{\ell}{(\log \ell)^2}$.

Condensers oNOSF sources were further studied by [CGR24], where they obtained the following results regarding condensers:

- When $n \geq k \geq \ell$, there exist functions that can transform a (g, ℓ, n, k)-oNOSF source into a uniform $(g-1, \ell-1, O(k/\ell))$-oNOSF source (this function can be made explicit with slightly worse dependence on output length).
- When $n \geq 2^{\omega(\ell)}$ and $g > 0.5\ell$, there exists condenser $\mathsf{Cond} : (\{0,1\}^n)^\ell \to \{0,1\}^{m=O(n\cdot\ell/g)}$ such that for any uniform (g, ℓ, n)-oNOSF source $\mathbf{X}$, $H^\varepsilon_\infty(\mathsf{Cond}(\mathbf{X})) \geq m - O(\log(n/\varepsilon))$.[6] Their result is not explicit.
- It is impossible to condense from uniform $(0.5\ell, \ell, n)$-oNOSF sources with output entropy rate more than 0.5.[7]

We also mention a related family of sources, namely adversarial Chor-Goldreich sources. Uniform oNOSF sources can be seen as a special case of adversarial Chor-Goldreich sources where the good blocks are uniform. Constructing condensers where the output entropy rate is g/ℓ for adversarial Chor-Goldreich sources is already a challenging task, although such condensers in various parameter regimes have been recently constructed [DMOZ23, GLZ24]. The paper of [DMOZ25] recently constructed condensers for a related, more general model.

1.2 This Work: New Condenser Constructions

Previous works only showed the existence of condensers for oNOSF sources when $n \geq 2^{\omega(\ell)}$. We vastly improve on this result in two ways. First, we construct explicit condensers that work even when $n \geq \mathrm{polylog}(\ell)$ and provide an explicit transformation from low-entropy oNOSF sources to uniform oNOSF sources that works even when the min-entropy of a block k is only $\mathrm{polylog}(n)$. Second, we show that condensers for

[5] In [AORSV20], these sources were called SHELA (Somewhere Honest Entropic Look Ahead) sources.

[6] They get a tradeoff for $g \leq 0.5\ell$ as well

[7] They get impossibility for other smaller g as well

oNOSF sources exist when n is just a large constant, only leaving open the question of the existence of such condensers for when n is a very small constant (e.g., $n = 1$). We also discover surprising connections between condensers for oNOSF sources and protocols for natural problems in distributed computing, such as collective coin flipping and collective sampling. Lastly, we initiate the study of *online influence* of Boolean functions, a natural generalization of influence that captures the one-sided nature of our online adversary to help us analyze the setting of $n = 1$. We now discuss our results in detail below.

1.2.1 Explicit Condensers

We construct the first explicit condensers for oNOSF sources. Our ultimate result is founded on a baseline construction that itself is an explicit condenser for oNOSF sources that matches the existential results of [CGR24] and works for any block length $n = 2^{\Omega(\ell)} \log(1/\varepsilon)$ as long as at least 51% of blocks are good.

Theorem 1 (Theorem 2.4 restated)**.** *For all $\varepsilon > 0$ and $n, \ell \in \mathbb{N}$ where $n \geq 2^{\Omega(\ell)} \log(1/\varepsilon)$, there exists an explicit condenser $\mathsf{Cond} : (\{0,1\}^n)^\ell \to \{0,1\}^m$ such that for any $(g = 0.51\ell, \ell, n)$-oNOSF source $\mathbf{X}$, we have that $H^\varepsilon_\infty(\mathsf{Cond}(\mathbf{X})) \geq m - 2^{O(\ell)} \log(1/\varepsilon)$ where $m = 0.0001\ell n$.*

Surprisingly, we are able to improve upon this baseline to obtain explicit condensers that work for oNOSF sources where the block length n is only at least $\mathrm{poly}(\log(\ell)/\varepsilon)$.

Theorem 2 (Informal version of Theorem 2.1)**.** *For all $\varepsilon > 0$ and $n, \ell \in \mathbb{N}$ where $n \geq \mathrm{poly}(\log(\ell)/\varepsilon)$, there exists an explicit condenser $\mathsf{Cond} : (\{0,1\}^n)^\ell \to \{0,1\}^m$ such that for any $(g = 0.51\ell, \ell, n)$-oNOSF source $\mathbf{X}$, we have that $H^\varepsilon_\infty(\mathsf{Cond}(\mathbf{X})) \geq m - \mathrm{poly}(\log(\ell)/\varepsilon) \cdot \log(n)$ where $m = 0.001\ell n - O(\ell \log(\ell) \log(1/\varepsilon))$.*

Since condensing when $g = 0.5\ell$ is impossible, both results are tight. We note that neither result completely subsumes the other. Our baseline construction in Theorem 1 has an exponential dependence of n on ℓ instead of the polylogarithmic dependence achieved in Theorem 2; however, the latter result requires the dependence $n \geq \mathrm{poly}(1/\varepsilon)$ compared to $n \geq \log(1/\varepsilon)$ for the baseline construction.

Using our new results regarding transforming oNOSF sources to uniform oNOSF sources, we also obtain explicit condensers for $(0.51\ell, \ell, n, k)$-oNOSF sources for the same parameter regime:

Corollary 1.3 (Corollary 5.2, simplified)**.** *For all $n, \ell, k \in \mathbb{N}$ where $n \geq \mathrm{poly}(\ell)$ and $k \geq \mathrm{polylog}(n)$, there exists an explicit condenser $\mathsf{Cond} : (\{0,1\}^n)^\ell \to \{0,1\}^m$ such that for any $(g = 0.51\ell, \ell, n, k)$-oNOSF source $\mathbf{X}$, we have that $H^\varepsilon_\infty(\mathsf{Cond}(\mathbf{X})) \geq m - \mathrm{polylog}(\ell) \cdot \log(n)$ where $m = 0.001\ell n - O(\ell \log(\ell) \log\log(\ell))$ and $\varepsilon = \mathrm{poly}(1/\log(\ell))$.*

We can also extend our result to explicitly condense from uniform (g, ℓ, n)-oNOSF sources in the same parameter regime so that the output entropy rate is $1/\lfloor \ell/g \rfloor - o(1)$, which is tight according to the impossibility result of [CGR24].

Previously, [CGR24] showed how to existentially condense from uniform (g, ℓ, n)-oNOSF sources when $n = 2^{\Omega(\ell)}$. However, they relied on the existence of a very strong pseudorandom object: "output-light" low-error two-source extractors. Such extractors, even without the output-lightness requirement, are extremely hard to construct and it is a major open problem to obtain such extractors. We are able to construct explicit condensers by creating new tools that allow us to use an oNOSF source to sample indices within an oNOSF source, and stitching them together so that the base pseudorandom object we rely on are seeded extractors that we know how to explicitly construct with near optimal parameters.

1.2.2 Transforming Low-Entropy oNOSF sources to uniform oNOSF sources

We show how to existentially, as well as explicitly, with a slight loss in parameters, transform (g, ℓ, n, k)-oNOSF sources into uniform $(0.99g, \ell - 1, n)$-oNOSF sources. Formally, we show:

Theorem 1.4 (Informal version of Theorem 6.1)**.** *For all ℓ, n, k, ε where $n = \mathrm{poly}(\log(\ell)), k = O(\log(\ell/\varepsilon))$, there exists a function f such that f transforms $(0.51\ell, \ell, n, k)$-oNOSF sources into uniform $(0.509\ell, \ell, m)$-oNOSF sources with error ε where $m = \Omega(k)$.*

Our construction can also be made explicit with slightly worse dependence on m and ε. See Corollary 6.4 for the full tradeoff.

Previously, [CGR24] provided such a transformation only for $n \geq k \geq \Omega(\ell)$. Hence, our transformation makes a major improvement on their parameters. Such an improvement allows us to obtain better condensers for low-entropy oNOSF sources in the regime $n = \mathrm{poly}(\log(\ell/\varepsilon))$ (see Theorem 4).

1.2.3 Existential Condensers

We show how to condense from uniform (g, ℓ, n)-oNOSF sources for almost all settings of ℓ and n when $g \geq 0.51\ell$. In particular, we show:

Theorem 3 (Informal version of Theorem 7.1)**.** *For all ℓ, ε where $\ell \geq O(\log(1/\varepsilon))$, and $n = 10^4$, there exists a condenser $\mathsf{Cond} : (\{0,1\}^n)^\ell \to \{0,1\}^m$ such that for any uniform $(0.51\ell, \ell, n)$-oNOSF source $\mathbf{X}$, we have $H^\varepsilon_\infty(\mathsf{Cond}(\mathbf{X})) \geq 0.99m$ where $m = \Omega(\ell + \log(1/\varepsilon))$. Furthermore, when $n = \omega(1)$, the output entropy rate becomes $1 - o(1)$.*

This is tight since [CGR24] showed it is impossible to condense uniform $(0.5\ell, \ell, n)$-oNOSF sources beyond output entropy rate 0.5.

Using our new results regarding transforming oNOSF sources to uniform oNOSF sources, we also obtain condensers for $(0.51\ell, \ell, n, k)$-oNOSF sources when $n \geq \mathrm{poly}(\log(\ell))$,

Theorem 4. *For all ℓ, n, ε where $n = \mathrm{poly}(\log(\ell/\varepsilon)), k = \Omega(\log(\ell/\varepsilon))$, there exists a condenser $\mathsf{Cond} : (\{0,1\}^n)^\ell \to \{0,1\}^m$ such that for any $(0.51\ell, \ell, n, k)$-oNOSF source $\mathbf{X}$, we have $H^\varepsilon_\infty(\mathsf{Cond}(\mathbf{X})) \geq m - O(m/\log(m)) - O(\log(1/\varepsilon))$ where $m = \Omega(k)$.*

We sketch the proof of both of these theorems in Section 2.1 We can also extend our result to condense from uniform (g, ℓ, n)-oNOSF sources for all g, ℓ and constant n where the output entropy rate is $1/\lfloor \ell/g \rfloor - 0.001$. This is tight since [CGR24] showed it is impossible to condense such sources beyond output entropy rate $1/\lfloor \ell/g \rfloor$.

Previously, [CGR24] showed how to existentially condense from uniform (g, ℓ, n)-oNOSF sources when $g \geq 0.51\ell$, provided $n \geq 2^{\omega(\ell)}$. As n gets smaller, condensing becomes harder since a uniform (g, ℓ, n)-oNOSF source is also a uniform $(g \cdot n/1000, \ell \cdot n/1000, 1000)$-oNOSF source. Hence, we greatly improve the parameters while using different and much simpler techniques.

1.3 Extraction from oNOSF Sources

Next we discuss our positive and negative results on the limits of extraction from oNOSF sources. Our upper bound results (explicit extractors) are based on a novel connection to leader election and coin-flipping protocols; to instantiate this connection and give explicit extractors, we construct novel protocols for these distributed problems. Our lower bounds are based on a new notion of influence of functions, namely *online influence*, that we introduce and analyze.

Extraction Lower bounds via Online Influence

For simplicity, we focus on the case of $n = 1$, which leads to interesting new questions about Boolean functions. We refer to such uniform $(g, \ell, 1)$-oNOSF sources as (g, ℓ)-oNOBF sources; oNOBF stands for online non-oblivious bit-fixing sources. We ask what is the exact tradeoff between g, ℓ, and ε for extracting from oNOBF sources. Towards this, we introduce the notion of online influence.

Definition 1.5 (Online influence)**.** *For a function $f : \{0,1\}^\ell \to \{0,1\}$, the* online influence *of the i-th bit is*

$$\mathbf{oI}_i[f] = \mathop{\mathbb{E}}_{x \sim \mathbf{U}_{i-1}} \left[\left| \mathop{\mathbb{E}}_{y \sim \mathbf{U}_{\ell-i}} [f(x,1,y)] - \mathop{\mathbb{E}}_{y \sim \mathbf{U}_{\ell-i}} [f(x,0,y)] \right| \right]$$

and the total online influence *is* $\mathbf{oI}[f] = \sum_{i=1}^{\ell} \mathbf{oI}_i[f]$.

We establish new structural results on online influence, including a Poincaré-style inequality and use them to obtain the following extraction lower bound.

Theorem 1.6 (Informal version of Corollary 10.21)**.** *For $\varepsilon < 0.01$, there do not exist extractors for $(0.97\ell, \ell)$-oNOBF sources with error at most ε.*

A similar extraction lower bound was shown in [AORSV20] using different techniques.

Explicit Extractors via Leader Election Protocols

Here we present our explicit constructions of extractors for oNOBF and oNOSF sources. The following are our main results.

Theorem 5 (informal version of Theorem 8.3)**.** *There exists an explicit function $\mathsf{Ext} : \{0,1\}^\ell \to \{0,1\}$ such that for any (g, ℓ)-oNOBF source $\mathbf{X}$ where $g \geq \ell - \ell/(C \log(\ell))$, we have $\mathsf{Ext}(\mathbf{X}) \approx_{\varepsilon=1/100} \mathbf{U}_1$, where C is a large constant.*

Theorem 6 (informal version of Theorem 8.4)**.** *There exists an explicit function $\mathsf{Ext} : (\{0,1\}^n)^\ell \to \{0,1\}^n$ such that for any (g, ℓ, n)-oNOSF source $\mathbf{X}$ where $g \geq \ell - \ell/(C \log^*(\ell))$ and $n \geq \log(\ell)$, we have $\mathsf{Ext}(\mathbf{X}) \approx_{\varepsilon=1/100} \mathbf{U}_n$, where C is a large constant.*

It is instructive to contrast our results with the non-online setting (where adversarial bits may depend on any good bit), called NOSF sources and NOBF sources. For both these sources, the current best extractors require $g \geq \ell - \frac{\ell}{(\log \ell)^2}$, which is much more than what Theorem 5 and Theorem 6 require.

We contrast the results for both settings in Tables 1 and 2. In these tables, we are providing known upper and lower bounds on the value of $b(\ell)$, defined as the maximum number of bad symbols for which extraction is still possible with a small constant error — so lower bounds correspond to best known constructions of such functions and upper bounds refers to the best known limitation of such functions. We write "$O(\ell)$" to mean "$c\ell$ for some small universal constant $c < 1$".

To interpret our results in terms of (online) influence of coalitions, it will be useful to extend the definition of online influence to subsets of coordinates, which we do formally in Definition 8.1. Intuitively, we're measuring the influence of the exact same adversary as in an oNOSF source.

In Section 10.4, we note that online-resilient functions are equivalent to extractors for uniform oNOSF source sources (with one bit output). Thus, our explicit extractor results immediately imply explicit online-resilient functions.

Source	Lower bound	Upper bound
NOBF	$\Omega\left(\frac{\ell}{\log^2 \ell}\right)$, [AL93]	$O\left(\frac{\ell}{\log \ell}\right)$, [KKL88]
NOSF	$\Omega\left(\frac{\ell}{\log^2 \ell}\right)$, [AL93]	$O(\ell)$, [BKKKL92]

Table 1: $b(\ell)$ bounds in the non-online setting.

Source	Lower bound	Upper bound
oNOBF	$\Omega\left(\frac{\ell}{\log \ell}\right)$, [Theorem 8.3]	$O(\ell)$, Corollary 10.21 or [AORSV20]
oNOSF	$\Omega\left(\frac{\ell}{\log^* \ell}\right)$, [Theorem 8.4][8]	$O(\ell)$, [AORSV20]

Table 2: $b(\ell)$ bounds in the online setting.

Our main technique for Theorems 5 and 6 is a new generic way to transform leader election and coin flipping protocols (formally defined in Section 4.4) into extractors for oNOBF and oNOSF sources. This is given in Lemma 8.2; the general idea of constructing an extractor is to simulate an appropriate leader election protocol with the source at hand (oNOBF or oNOSF), and output according to the chosen leader. To instantiate this transformation, we revisit previous leader election protocols in Section 9. Our leader election protocols provide a slightly stronger than usual guarantee: a good player is elected as the leader with probability close to 1 (see Lemma 9.1 and Lemma 9.5). This contrasts with the usual guarantee in leader election protocols, where a good leader is chosen with only a non-trivial (constant) probability. We give more connections to distributed computing in Section 3 where we delineate applications of our results to collection coin flipping and collective sampling.

2 Proof Overview

Our proof overview begins by outlining our new explicit result for condensers in Section 2.1 that is able to handle polylogarithmic block length. Next, we present our transformation of low-entropy oNOSF sources to uniform oNOSF sources in Section 2.2 before discussing our existential results for condensers that can handle constant block length in Section 2.3. We present the main ideas behind our results regarding online influence and extractor lower bounds in Section 2.4. In Section 2.5, we overview our extractor constructions for oNOBF and oNOSF sources that are based on a general transformation from leader election protocols.

2.1 Explicit Condensers for Uniform oNOSF Sources

We sketch here our constructions (as well as proof ideas) of explicit condensers for uniform $(0.51\ell, \ell, n)$-oNOSF sources.[9] The goal will be to construct explicit condensers that work with as few good sources as possible while minimizing the block length n. In particular, we will show:

[8] Recall that this lower bound is for (g, ℓ, n)-oNOSF sources with $n \geq \log(\ell)$.

[9] Since all sources are uniform here, we will not explicitly mention this again.

Theorem 2.1 (Theorem 5.1, simplified). *For all $0 < \varepsilon$ and $n, \ell \in \mathbb{N}$ where $n \geq \left(\frac{\log(\ell)}{\varepsilon}\right)^{\Omega(1)}$, there exists an explicit condenser $\mathsf{2Cond} : (\{0,1\}^n)^\ell \to \{0,1\}^m$ such that for any $(g = 0.51\ell, \ell, n)$-oNOSF source $\mathbf{X}$, we have that $H^\varepsilon_\infty(\mathsf{Cond}(\mathbf{X})) \geq m - \left(\frac{\log(\ell)}{\varepsilon}\right)^{O(1)} \cdot \log(n)$ where $m = 0.001 \cdot \ell n - O\left(\ell \log(\ell) \log(1/\varepsilon)\right)$.*

We will require two main tools to show this. The first one uses ideas from the leader election literature and allows us to sample a $O(\log(\ell))$ sized committee starting from ℓ players while essentially maintaining the fraction of bad players. Formally:

Lemma 2.2 (Lemma 5.3, simplified). *For all $\varepsilon_s > 0$, $n, \ell \in \mathbb{N}$, and constant $\varepsilon_a > 0$ where $n \geq \Omega(\log(\ell)\log(1/\varepsilon_s))$, there exists an explicit function $\mathsf{oNOSFSamp} : (\{0,1\}^n)^\ell \to [\ell]^D$ where $D \leq O(\log(\ell/\varepsilon_s))$ with the following property.[10] For all $S \subset [\ell]$ and (g, ℓ, n)-oNOSF sources $\mathbf{X}$, we have that*

$$\Pr_{x \sim \mathbf{X}}\left[\left|\frac{|\mathsf{oNOSFSamp}(x) \cap S|}{D} - \frac{|S|}{\ell}\right| \geq \varepsilon_a\right] \leq \varepsilon_s$$

At a high level, to construct $\mathsf{oNOSFSamp}$, we slightly modify the committee selection procedure from [RZ01] and instantiate it with a seeded extractor with near optimal dependence on the seed [Zuc07].

Our second tool is a seeded condenser for general min-entropy sources $\mathbf{X}$ that uses an oNOSF source $\mathbf{Y}$ as the seed, where the bad bits in $\mathbf{Y}$ can depend on $\mathbf{X}$.

Lemma 2.3 (Lemma 5.4, simplified). *There exists a constant C_{2Cond} such that for all $n_x, k, n_y, t \in \mathbb{N}$ with $\varepsilon > 0$ and $n_y \geq (C_{\mathsf{2Cond}})^t \log(tn_x/\varepsilon)$, there exists an explicit condenser $\mathsf{Cond} : \{0,1\}^{n_x} \times (\{0,1\}^{n_y})^t \to \{0,1\}^m$ where $m = \frac{1}{3}(k - (C_{\mathsf{2Cond}})^t \log(tn_x/\varepsilon))$ so that the following holds: For all (n_x, k)-sources $\mathbf{X}$ and $(g = 1, \ell = t)$-oNOSF sources $\mathbf{Y} \sim (\{0,1\}^{n_y})^t$ such that the good blocks in $\mathbf{Y}$ are independent of $\mathbf{X}$ and the bad blocks in $\mathbf{Y}$ can depend on $\mathbf{X}$, we have that $H^\varepsilon_\infty(\mathsf{2Cond}(\mathbf{X}, \mathbf{Y})) \geq m - (C_{\mathsf{2Cond}})^t \log(tn_x/\varepsilon)$.*

We will sketch how to construct this in Section 2.1.6. Using these tools, we are ready to present our explicit condenser. In fact, we will provide sketches of five different constructions of explicit condensers with increasingly better parameter dependence; the fifth (and final) one is the construction given by Theorem 2.1. The parameters they will vary in are the fraction of good blocks (i.e., g/ℓ) and the block length n. As we will see, each construction will build on ideas from the previous construction.

2.1.1 Construction 1: 51% good and $n \geq 2^{\Omega(\ell)}$

We here construct the following condenser:

Theorem 2.4 (Theorem 5.6, simplified). *For all $0 < \varepsilon$ and $n, \ell \in \mathbb{N}$ where $n \geq 2^{\Omega(\ell)} \log(1/\varepsilon)$, there exists an explicit condenser $\mathsf{Cond} : (\{0,1\}^n)^\ell \to \{0,1\}^m$ such that for any $(g = 0.51\ell, \ell, n)$-oNOSF source $\mathbf{X}$, we have that $H^\varepsilon_\infty(\mathsf{Cond}(\mathbf{X})) \geq m - 2^{O(\ell)} \log(1/\varepsilon)$ where $m = 0.0001 \cdot n\ell$.*

Proof sketch. Let $\gamma = 0.01$ and $\ell' = \ell/2$. We decompose the input $\mathbf{X}$ into two equal sized parts so that $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2)$ where both $\mathbf{X}_1, \mathbf{X}_2 \sim (\{0,1\}^n)^{\ell/2} \equiv (\{0,1\}^n)^{\ell'}$. Since $\mathbf{X}$ has $(0.5 + \gamma)\ell$ good players, we infer that each of $\mathbf{X}_1, \mathbf{X}_2$ has at least $\gamma\ell = (2\gamma) \cdot \ell'$ good players. In particular, we use the fact that $H_\infty(\mathbf{X}_1) \geq (2\gamma) \cdot (\ell n/2)$ and that $\mathbf{X}_2$ is a $(g = (2\gamma)\ell/2, \ell/2, n)$-oNOSF source. With this, we let $\mathsf{2Cond}$ be the condenser from our second tool Lemma 2.3 and let out final output be $\mathsf{2Cond}(\mathbf{X}_1, \mathbf{X}_2)$. Note that here $t = \ell$ and so, this requires $n \geq (C_{\mathsf{2Cond}})^\ell \log(\ell n/\varepsilon)$, an inequality that we indeed satisfy. The guarantees from Lemma 2.3 provide us with the desired claim. □

[10] Even though the output domain of $\mathsf{oNOSFSamp}$ is a vector, we will abuse notation and often treat it as a set.

Each of the subsequent constructions will use a similar construction idea as above. However, they will try to decrease t as much as possible, where t is the number of players in the oNOSF source when applying 2Cond from Lemma 2.3. Note that any reduction results in a decrease in the block length requirement n.

2.1.2 Construction 2: 67% good and $n \geq \mathrm{poly}(\ell)$

Our next construction requires a slightly larger fraction of good blocks. However, the block length required is exponentially improved.

Theorem 2.5 (Theorem 5.7, simplified). *For all $0 < \varepsilon$ and $n, \ell \in \mathbb{N}$ where $n \geq \left(\frac{\ell}{\varepsilon}\right)^{\Omega(1)}$, there exists an explicit condenser $\mathsf{Cond} : (\{0,1\}^n)^\ell \to \{0,1\}^m$ such that for any $(g = 0.67\ell, \ell, n)$-oNOSF source $\mathbf{X}$, we have that $H^\varepsilon_\infty(\mathsf{Cond}(\mathbf{X})) \geq m - \left(\frac{\ell}{\varepsilon}\right)^{O(1)} \log(n)$ where $m = 0.001 \cdot \ell n$.*

Proof sketch. Let $\gamma = 0.67 - (2/3)$. Let $\ell' = \ell/3$. We decompose the input $\mathbf{X}$ into three equal sized parts so that $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3)$ where all $\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3 \sim (\{0,1\}^n)^{\ell'}$. Since $\mathbf{X}$ has $((2/3) + \gamma)\ell$ good players, we easily see that each $\mathbf{X}_i$ is a $(g = 3\gamma\ell', \ell', n)$-oNOSF source.

We use $\mathbf{X}_1$ to sample a $O(\log(\ell))$ sized committee from $\mathbf{X}_3$. To do so, we use $\mathsf{oNOSFSamp}_{1\to 3} : (\{0,1\}^n)^{\ell'} \to (\{0,1\}^{\ell'})^D$ from Lemma 2.2 with $S \subset [\ell']$ being the set of good players from $\mathbf{X}_3$, the approximation factor $\varepsilon_a = \gamma$ and sampling error $\varepsilon_s = \varepsilon/3$. Let $\mathcal{C}_3 \subset [\ell'], |\mathcal{C}_3| \leq D_3 = O(\log(\ell'/\varepsilon)) = O(\log(\ell/\varepsilon))$ be the committee of players thus obtained. The approximation property of the sampler guarantees us that out of $\geq 3\gamma\ell'$ good players in $\mathbf{X}_3$, at least $2\gamma\,|\mathcal{C}_3|$ many good players will be in $\mathcal{C}_3$ with probability $1 - \varepsilon/3$. Let $\mathbf{Y}_3$ be the $(2\gamma D_3, D_3, n)$-oNOSF source obtained by restricting the players in $\mathbf{X}_3$ to the committee $\mathcal{C}_3$. We finally use 2Cond from Lemma 2.3 and output $\mathsf{2Cond}(\mathbf{X}_2, \mathbf{Y}_3)$. Here, the parameter t in Lemma 2.3 will be set to $D \leq O(\log(\ell/\varepsilon))$, and hence, Lemma 2.3 would only require that $n \geq (C_{2_{\mathsf{2Cond}}})^t \log(\ell n/\varepsilon) = \left(\frac{\ell}{\varepsilon}\right)^{O(1)}$, a condition that we do meet. We carefully compute the remaining parameters to infer the claim. □

2.1.3 Construction 3: 76% good and $n \geq \mathrm{poly}(\log(\ell))$

We build on our previous construction and show how to condense when the block length requirement is again exponentially decreased. This comes at a cost of slightly larger fraction of good blocks.

Theorem 2.6 (Theorem 5.9, simplified). *For all $0 < \varepsilon$ and $n, \ell \in \mathbb{N}$ where $n \geq \left(\frac{\log(\ell)}{\varepsilon}\right)^{\Omega(1)}$, there exists an explicit condenser $\mathsf{Cond} : (\{0,1\}^n)^\ell \to \{0,1\}^m$ such that for any $(g = 0.76\ell, \ell, n)$-oNOSF source $\mathbf{X}$, we have that $H^\varepsilon_\infty(\mathsf{Cond}(\mathbf{X})) \geq m - \left(\frac{\log(\ell)}{\varepsilon}\right)^{O(1)} \log(n)$ where $m = 0.001 \cdot \ell n$.*

Proof sketch. Let $\gamma = 0.01$. Let $\ell' = \ell/4$. We decompose the input $\mathbf{X}$ into four equal sized parts such that $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \mathbf{X}_4)$ and conclude that each $\mathbf{X}_i$ is a $(g = 4\gamma\ell', \ell', n)$-oNOSF source. However here, we claim something stronger. Call $i \in [\ell']$ a totally good index if it corresponds to a good player across each of the four blocks. Since $\mathbf{X}$ has $(3/4 + \gamma) \cdot (4\ell')$ good players, there must be $\geq 4\gamma\ell'$ totally good indices.

We first use $\mathbf{X}_1$ to sample a $D_2 = O(\log(\ell/\varepsilon))$ sized committee $\mathcal{C}_2 \subset [\ell']$ using the sampler from Lemma 2.2 such that $\mathcal{C}_2$ will have at least 3γ fraction of totally good indices. We let $\mathbf{Y}_2$ be the $(g = 3\gamma D_2, D_2, n)$-oNOSF source obtained by restricting $\mathbf{X}_2$ to indices from $\mathcal{C}_2$.

Second, we use $\mathbf{Y}_2$ to sample a $D_4 = O(\log(\log(\ell)/\varepsilon))$ sized committee $\mathcal{C}_4 \subset \mathcal{C}_2$ such that $\mathcal{C}_4$ has at least 2γ fraction of totally good indices. We let $\mathbf{Y}_4$ be the $(g = 2\gamma D_4, D_4, n)$-oNOSF source obtained by restricting $\mathbf{X}_4$ to indices from $\mathcal{C}_4$.

Third and last, we use 2Cond from Lemma 2.3 and output $\mathsf{2Cond}(\mathbf{X}_3, \mathbf{Y}_4)$. Here, the parameter t in Lemma 2.3 will be set to $D_4 \leq O(\log(\log(\ell)/\varepsilon))$, and hence, Lemma 2.3 would only require that $n \geq (C_{\mathsf{2Cond}})^t \log(\ell n/\varepsilon) = \left(\frac{\log(\ell)}{\varepsilon}\right)^{O(1)}$, a condition that we do meet. We carefully compute the remaining parameters to infer the claim. □

2.1.4 Construction 4: 67% good and $n \geq \mathrm{poly}(\log(\ell))$

Our next construction maintains a similar guarantee as before on the block length and decreases the requirement on the fraction of good blocks.

Theorem 2.7 (Theorem 5.10, simplified)**.** *For all $0 < \varepsilon$ and $n, \ell \in \mathbb{N}$ where $n \geq \left(\frac{\log(\ell)}{\varepsilon}\right)^{\Omega(1)}$, there exists an explicit condenser $\mathsf{Cond} : (\{0,1\}^n)^\ell \to \{0,1\}^m$ such that for any $(g = 0.67\ell, \ell, n)$-oNOSF source $\mathbf{X}$, we have that $H^\varepsilon_\infty(\mathsf{Cond}(\mathbf{X})) \geq m - \left(\frac{\log(\ell)}{\varepsilon}\right)^{O(1)} \log(n)$ where $m = 0.001 \cdot \ell n$.*

Proof sketch. Let $\gamma = 0.67 - (2/3)$. Let $\ell' = \ell/3$. We decompose $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3)$ so that each $\mathbf{X}_i$ is a $(g = 3\gamma\ell', \ell', n)$-oNOSF source with $3\gamma\ell'$ totally good indices.

For the first step, we again use $\mathbf{X}_1$ along with a sampler from Lemma 2.2 to obtain $\mathcal{C}_2 \subset [\ell']$ with $D_2 = |\mathcal{C}_2| \leq O(\log(\ell/\varepsilon)$ and a subsource of $\mathbf{X}_2$ restricted to $\mathcal{C}_2$ - Y_2 that is a $(g = 2\gamma D_2, D_2, n)$-oNOSF source. For the second step, we again use $\mathbf{Y}_2$ to sample a $D_3 = O(\log(\log(\ell)/\varepsilon))$ sized committee $\mathcal{C}_3 \subset \mathcal{C}_2$ such that $\mathcal{C}_3$ has $\geq \gamma$ fraction of totally good indices. We let $\mathbf{Y}_3$ be the $(g = \gamma D_3, D_3, n)$-oNOSF source obtained by restricting $\mathbf{X}_3$ to indices from $\mathcal{C}_3$. Third and last, we use 2Cond from Lemma 2.3 and output $\mathsf{2Cond}(\mathbf{X}_2, \mathbf{Y}_3)$. Again the parameter t in Lemma 2.3 will be set to $D_3 \leq O(\log(\log(\ell)/\varepsilon))$ and would only require that $n \geq \left(\frac{\log(\ell)}{\varepsilon}\right)^{O(1)}$.

Analyzing this construction requires more care since we use $\mathbf{X}_2$ to both sample from $\mathbf{X}_3$ and as a source for 2Cond. We use the chain rule for min-entropy (Lemma 4.3) to argue that most fixings of $\mathbf{Y}_2 = y_2$ will leave $\mathbf{X}_2$ with lot of entropy (since sampler requires few random bits) and observe that such a fixing still leaves $\mathbf{X}_3$ as oNOSF source with the same parameters. Also since for most fixings of $\mathbf{Y}_2 = y_2$, the committee $\mathcal{C}_3$ has γ fraction of good players, we obtain our claim. □

2.1.5 Construction 5: 51% good and $n \geq \mathrm{poly}(\log(\ell))$

We lastly construct the condenser promised in our main result - Theorem 2.1.

Proof sketch of Theorem 2.1. Let $\gamma = 0.01$. Let $\ell' = \ell/2$. We decompose $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2)$ so that each $\mathbf{X}_i$ is a $(g = 2\gamma\ell', \ell', n)$-oNOSF source with $2\gamma\ell'$ totally good indices.

For the first step, we let $\mathbf{X}'_1$ be the $(2\gamma\ell', \ell', n'_1)$-oNOSF source obtained by taking a prefix of length n'_1 from each source where we set $n'_1 \ll n$ but also n'_1 is long enough to be used by the sampler from Lemma 2.2. We use $\mathbf{X}'_1$ with such a sampler to obtain a committee $\mathcal{C}_{1\to2} \subset [\ell']$ with $D_{1\to2} = |\mathcal{C}_{1\to2}| \leq O(\log(\ell/\varepsilon)$ and a subsource of $\mathbf{X}_2$ restricted to $\mathcal{C}_{1\to2}$, which we call $\mathbf{Y}_{1\to2}$, that is a $(g = (3/2)\gamma D_{1\to2}, D_{1\to2}, n)$-oNOSF source.

We argue that: 1) most fixings of $\mathbf{X}'_1$ are such that they leave $\mathbf{X}_1$ with high entropy, and 2) that the committee $\mathcal{C}_{1\to2}$ obtained will have $3\gamma/2$ fraction of good players. We obtain this using the chain rule for min-entropy and by guarantees of the sampler. We condition on such a fixing from here on.

For the second step, we use $\mathbf{Y}_{1\to2}$ to sample a $D_{2\to2} = O(\log(\log(\ell)/\varepsilon))$ sized committee $\mathcal{C}_{2\to2} \subset \mathcal{C}_{1\to2}$ such that $\mathcal{C}_{2\to2}$ has $\geq \gamma$ fraction of totally good indices. We let $\mathbf{Y}_{2\to2}$ be the $(g = \gamma D_{2\to2}, D_{2\to2}, n)$-oNOSF source obtained by restricting $\mathbf{X}_2$ to indices from $\mathcal{C}_{2\to2}$.

Lastly, we use 2Cond from Lemma 2.3 and output $\mathsf{2Cond}(\mathbf{X}_1, \mathbf{Y}_{2\to 2})$. The parameter t in Lemma 2.3 will be set so that it would only require that $n \geq \left(\frac{\log(\ell)}{\varepsilon}\right)^{O(1)}$. We must be a bit careful about the error parameter $\varepsilon_{\mathsf{2Cond}}$ for 2Cond and we will choose it to be extremely small.

Analyzing this construction requires care since we use $\mathbf{X}_2$ both to sample from within $\mathbf{X}_2$ itself and also as a seed for 2Cond. We cannot use the chain rule since a fixing of $\mathbf{Y}_{1\to 2}$ will destroy the structure of the source $\mathbf{X}_2$. We first see that the sampler guarantees that no matter how the adversary behaves, with probability $1-\varepsilon/4$, the sampler will succeed in selecting a committee $\mathcal{C}_{2\to 2}$ with γ fraction of good players. We pay $\varepsilon/4$ in error and now assume that the sampler always succeeds in doing so. We then compare two scenarios: one scenario Opt where the bits in $\mathbf{Y}_{1\to 2}$ are all uniform and independent of all other bits and another scenario Adv where the bits in $\mathbf{Y}_{1\to 2}$ are all controlled by an adversary (guarantees on this latter scenario suffice for our claim). Let the total number of bits in $\mathbf{Y}_{1\to 2}$ be equal to $b_{1\to 2}$. Under scenario Opt, we easily see that we succeed and with error $\varepsilon_{\mathsf{2Cond}}$ will have high entropy - say k. To compare this to scenario Adv, we use Lemma 7.9 that lets us conclude that in scenario Adv, with error $\varepsilon_{\mathsf{2Cond}} \cdot 2^{b_{1\to 2}}$, the output will have entropy $k - b_{1\to 2}$. Since we carefully chose $\varepsilon_{\mathsf{2Cond}}$ to be small enough and $b_{1\to 2}$ is small since we only use $\mathbf{Y}_{1\to 2}$ as source for the sampler, the output will still have small error and will have high-entropy as desired. □

2.1.6 Construction of 2Cond

We now sketch how to construct our desired 2Cond.

Proof sketch of Lemma 2.3. Let $\mathbf{X}$ and $\mathbf{Y} = (\mathbf{Y}_1, \ldots, \mathbf{Y}_t)$ be the two sources. For $i \in [t]$, let $n_i \approx C^{t-i+1} \log(tn_x/\varepsilon)$ where C is a large constant. For $i \in [t]$, let $\mathbf{Z}_i$ be the length n_i prefix of the block $\mathbf{Y}_i$. Our final construction will be the parity of the outputs of seeded extractors applied with source $\mathbf{X}$ and seeds $\mathbf{Z}_i$. More formally, we output

$$\bigoplus_{i=1}^{t} \mathsf{sExt}_i(\mathbf{X}, \mathbf{Z}_i),$$

where sExt_i is any explicit near optimal seeded extractor (such as the extractor from Theorem 4.6).

We proceed to sketch the analysis. We are guaranteed that there exists at least one $j \in [t]$ from $\mathbf{Y}$ that is good. We first condition on fixing blocks $\mathbf{Z}_1, \ldots, \mathbf{Z}_{j-1}$. Since these blocks can depend on $\mathbf{X}$, we apply the chain rule for min-entropy (Lemma 4.3) and conclude $\mathbf{X}$ will only lose some small amounts of entropy (the amount will be very small since these blocks are tiny compared to the amount entropy in $\mathbf{X}$). Moreover, since the adversary is online, $\mathbf{Z}_j$ remains uniform even after this fixing. We now view our construction as

$$g(\mathbf{X}) \oplus \bigoplus_{i=j}^{t} \mathsf{sExt}_i(\mathbf{X}, \mathbf{Z}_i)$$

where g is the fixed function obtained by fixing $\mathbf{Y}_1, \ldots, \mathbf{Y}_{j-1}$.

We now compare two scenarios: (1) Where all of $\mathbf{Z}_j, \ldots, \mathbf{Z}_t$ are uniform (2) Only $\mathbf{Z}_j$ is uniform and $\mathbf{Z}_{j+1}, \ldots, \mathbf{Z}_t$ are arbitrarily controlled by an adversary and can even depend on $\mathbf{X}$:

In the first scenario, we further condition on fixing $\mathbf{Z}_{j+1}, \ldots, \mathbf{Z}_t$. Since in this scenario $\mathbf{Z}_i$ are independent and random, $\mathbf{X}$ retains the same entropy and $\mathbf{Y}_j$ remains uniform. So our overall output is of the form $h(\mathbf{X}) \oplus \mathsf{sExt}_j(\mathbf{X}, \mathbf{Z}_j)$ for some fixed function h. We condition on fixing output $h(\mathbf{X})$. Since the number of output bits $m \ll H_\infty(\mathbf{X})$, we apply the chain rule to infer that $\mathbf{X}$ still has lots of entropy when we do this fixing. Now the output is just $z \oplus \mathsf{sExt}_j(\mathbf{X}, \mathbf{Z}_j)$ where z is a fixed string, and hence is uniform.

The second scenario is more realistic and, in the worst case, this is what can actually happen. We then use the result that if an adversary controls few bits in the input distribution, then they cannot make the output of the condenser too bad (see Lemma 7.9 for the full statement). With this, since we carefully chose geometrically decreasing lengths of $\mathbf{Z}_i$ to help control the error, we indeed obtain that the output will be condensed. □

2.2 Converting Low-Entropy oNOSF Sources to Uniform oNOSF Sources

They key part of our proof for condensing from low-entropy oNOSF sources is a transformation from low-entropy oNOSF sources to uniform oNOSF sources. Here, we sketch the proof for our transformation in Theorem 1.4 and compare it to that of [CGR24]. Both these transformations rely on two-source extractors (see Definition 4.7 for definition) as a basic primitive.

Given a (g, ℓ, n, k)-oNOSF source $\mathbf{X} = \mathbf{X}_1, \ldots, \mathbf{X}_\ell$, [CGR24] uses excellent existential two-source extractors (such as from Lemma 6.6) to define output blocks $\mathbf{O}_i = 2\mathsf{Ext}(\mathbf{X}_1 \circ \cdots \circ \mathbf{X}_{i-1}, \mathbf{X}_i)$ for $i \in \{2, \ldots, \ell\}$ and define their transformation as $f(\mathbf{X}) = \mathbf{O}_2, \ldots, \mathbf{O}_\ell$. They show that $\mathbf{O}_i$ is a good block if: (1) $\mathbf{X}_i$ is a good block and (2) at least one block amongst $\mathbf{X}_1, \ldots, \mathbf{X}_{i-1}$ is a good block. They showed that such a good block will be uniform and independent of the blocks $\mathbf{O}_2, \ldots, \mathbf{O}_{i-1}$ and argued there will be $g - 1$ such good output blocks. This indeed shows their output is a uniform $(g - 1, \ell - 1, m)$-oNOSF source. However, each of their output blocks has length $m = O\left(\frac{k}{\ell}\right) \leq O\left(\frac{n}{\ell}\right)$, and so they were not able to handle the case of $n = o(\ell)$. We improve on their construction by using a "sliding window" based technique to obtain a much better transformation that can even handle $n = \mathrm{poly}(\log(\ell))$.

Theorem 2.8 (Theorem 6.1 restated)**.** *Let $d, g, g_{out}, \ell, n, m, k, \varepsilon$ be such that $g_{out} \leq g - \frac{\ell - g + 2}{d}$, $n \geq k \geq \log(nd - k) + md + 2\log(2g_{out}/\varepsilon)$. Then, there exists a function $f : (\{0,1\}^n)^\ell \to (\{0,1\}^m)^{\ell-1}$ such that for any (g, ℓ, n, k)-oNOSF source $\mathbf{X}$, there exists uniform $(g_{out}, \ell - 1, m)$-oNOSF source $\mathbf{Y}$ for which $|f(\mathbf{X}) - \mathbf{Y}| \leq \varepsilon$.*

The parameter d in our theorem statement above is the width of our sliding window. When we set $d = \ell$ we recover the analysis of [CGR24]. The true advantage of our transformation emerges when d is very small compared to ℓ. For instance, when $g = 0.51\ell, n = \mathrm{poly}(\log(\ell))$ and $k = \mathrm{poly}(\log(\ell))$, we set d to be a large constant and conclude that the output distribution is a uniform $(0.509\ell, \ell, \mathrm{poly}(\log(\ell))$-NOSF source.

Proof sketch of Theorem 2.8. Define $\mathbf{O}_i = 2\mathsf{Ext}(\mathbf{X}_{i-d} \circ \cdots \circ \mathbf{X}_{i-1}, \mathbf{X}_i)$. We call $\mathbf{O}_i$ to be a good output block when $\mathbf{X}_i$ is good and there's at least one good block amongst $\{\mathbf{X}_{i-d}, \ldots, \mathbf{X}_{i-1}\}$.

We first compute the number of good output blocks g_{out}. Let $j_1, \ldots, j_g$ be the indices of the good input blocks in $\mathbf{X}$ and $d_i = j_{i+1} - j_i$ be the gap between the i-th good block and the next $(i+1)$-th good block. If the gap d_i is at most d, then $\mathbf{O}_{i+1}$ must be a good output block. So, g_{out} is the number of i such that $d_i \leq d$. Since $g \geq 0.51\ell$, such large gaps can't appear too often and we compute that $g_{out} \geq g - \frac{\ell - g + 2}{d}$ as desired.

Next, we show that the good output blocks are indeed uniform conditioned on all previous output blocks. With this, we will obtain that the output distribution will be uniform $(g_{out}, \ell - 1, m)$-oNOSF source as desired. Let i be the index of a good output block. We want to show that $\mathbf{O}_i$ is uniform conditioned on $\mathbf{O}_1, \ldots, \mathbf{O}_{i-1}$. To do this, we first observe that any input block contributes to at most $d + 1$ good output blocks. This means that$(\mathbf{X}_{i-d} \circ \cdots \circ \mathbf{X}_{i-1})$, which has min-entropy at least k, loses at most $d \cdot m$ min-entropy conditioned on fixing $\mathbf{O}_1, \ldots, \mathbf{O}_{i-1}$. Moreover, $\mathbf{X}_i$ still remains uniform and independent of $(\mathbf{X}_{i-d} \circ \cdots \circ \mathbf{X}_{i-1})$ when fixing these previous output blocks. Hence, the output of the two-source extractor will indeed be uniform as desired. □

We can make Theorem 2.8 explicit by using the explicit two-source extractors of Theorem 6.7 at a slight cost of dependence on m and ε as seen in Corollary 6.4.

2.3 Existence of oNOSF Condensers for All ℓ and n

Here we sketch the proof of Theorem 3. This result states that when $g = 0.51\ell$ and $n = 1000$, there exists a condenser Cond for uniform (g, ℓ, n)-oNOSF sources so that the output entropy rate is 0.99, the number of output bits is $m = O(\ell + \log(1/\varepsilon))$, and the error of the condenser is ε where $\varepsilon \leq 2^{-\Omega(\ell)}$ is arbitrary.

Our construction uses amazing seeded condensers (see Definition 4.4) with $1 \cdot \log(1/\varepsilon)$ dependence on seed length. We slightly modify our source and then apply such seeded condenser. Here is a proof sketch:

Proof sketch for Theorem 3. Let $\mathbf{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_\ell)$ be such a source. Let $\mathbf{Y}_1 \sim (\{0,1\}^n)^{0.5\ell}$ be the source obtained by concatenating the first 0.5ℓ blocks of $\mathbf{X}$. Since 0.51ℓ blocks are good, there exist at least 0.01ℓ uniform blocks in $\mathbf{Y}_1$. We treat $\mathbf{Y}_1$ as a single distribution over $n\ell$ bits with min-entropy $\geq 0.01\ell n$. Let $\mathbf{Y}_2 \sim \{0,1\}^{0.5\ell}$ be the source obtained by concatenating 1 bit from each of the last 0.5ℓ blocks of $\mathbf{X}$. Once again, since 0.51ℓ blocks are good, there exist at least 0.01ℓ uniform bits in $\mathbf{Y}_2$. We will use the following seeded condenser:

Theorem 2.9 (Theorem 7.8, simplified)**.** *For all d, ε such that $d \geq \log(\ell n/\varepsilon) + O(1)$, there exists a seeded condenser $\mathsf{sCond} : \{0,1\}^{0.5\ell n} \times \{0,1\}^d \to \{0,1\}^m$ s.t. for all $\mathbf{X} \sim \{0,1\}^{0.5\ell n}$ with $H_\infty(\mathbf{X}) \geq 0.01\ell n$, we have $H^\varepsilon_\infty(\mathsf{sCond}(\mathbf{X}, \mathbf{U}_d)) \geq 0.01\ell n + d$ where $m = 0.01\ell n + d + \log(1/\varepsilon) + O(1)$.*

Our condenser Cond will output $\mathsf{sCond}(\mathbf{Y}_1, \mathbf{Y}_2)$. Observe that not only is $\mathbf{Y}_2$ not uniform, there could be as many as 0.49ℓ "bad bits" in $\mathbf{Y}_2$ that can depend on $\mathbf{Y}_1$. To remedy this, we use the well known fact that the behavior of such adversarial $\mathbf{Y}_2$ cannot be far worse than the behavior if $\mathbf{Y}_2$ were uniform. Concretely, suppose if $\mathbf{Y}_2$ were uniform and the output entropy and error were k and ε. Then for the actual $\mathbf{Y}_2$, the output entropy will be $k - 0.49\ell$ and error will be $\varepsilon \cdot 2^{0.49\ell}$. See Lemma 7.9 for the formal statement.

For us, it means the following: let $\varepsilon_{\mathsf{sCond}}, k_{\mathsf{sCond}}$ be such that $H^{\varepsilon_{\mathsf{sCond}}}_\infty(\mathsf{sCond}(\mathbf{Y}_1, \mathbf{U}_{0.5\ell})) \geq k_{\mathsf{sCond}}$. Then, it must be that $H^{2^{0.49\ell} \cdot \varepsilon_{\mathsf{sCond}}}_\infty(\mathsf{sCond}(\mathbf{Y}_1, \mathbf{Y}_2)) \geq k_{\mathsf{sCond}} - 0.49\ell$. So, for our final error to be some ε, we need to have $\varepsilon_{\mathsf{sCond}} = \varepsilon \cdot 2^{-0.49\ell}$. For seeded condensers to exist, we need $0.5\ell \geq \log(\ell n/\varepsilon_{\mathsf{sCond}}) + O(1)$ and we check that such an inequality can indeed be satisfied if $\varepsilon \geq 2^{-0.01\ell}$.

Hence, we finally obtain that our seeded condenser will output $0.01\ell n + O(\ell)$ bits and will have output entropy $m - \Delta$ where $\Delta = O(\ell)$. Hence, if n is a large enough constant, our output entropy rate, $\frac{m-\Delta}{m}$, will be ≥ 0.99 as desired. □

Remark 2.10. *Here (in the inequality $0.5\ell \geq 1 \cdot \log(\ell n/\varepsilon_{\mathsf{sCond}})$) we crucially used the fact that there exist seeded condensers with seed length dependence $1 \cdot \log(1/\varepsilon)$. Currently, we do not have explicit constructions with this dependence. We also couldn't have used a seeded extractor since for them, the seed length dependence is $2 \cdot \log(1/\varepsilon)$. For that to work, we would need to assume $g \geq 0.76\ell$.*

2.4 Online Influence and Extractor Lower Bounds

In this subsection, we provide a brief overview of our results regarding online influence and sketch how they imply extractor lower bounds against oNOBF sources. We also contrast this with the established notion of influence for Boolean functions. For any function $f : \{0,1\}^n \to \{0,1\}$, define the function $e(f)(x) = (-1)^{f(x)}$.

A Poincaré inequality and extractor lower bounds One fundamental inequality about regular influence is the Poincaré inequality which states that $\mathrm{Var}(f) \leq \mathbf{I}[f]$. We prove a similar result for online influence.

Theorem 2.11 (Theorem 10.5 restated)**.** *For any $f : \{0,1\}^\ell \to \{0,1\}$, we have $\mathrm{Var}(e(f)) \leq \mathbf{oI}[f] \leq \sqrt{\ell \, \mathrm{Var}(e(f))}$.*

It is not hard to derive extractor lower bounds for oNOBF sources from the above result. The high level idea is to collect bits with high online influence, which is guaranteed by the first inequality in the above theorem (using an averaging argument) to form a *coalition of coordinates* that has enough online influence to bias the claimed extractor. We refer the reader to Theorem 10.19 for more details.

The proof of Theorem 2.11 is based on techniques from the Fourier analysis of Boolean functions.[11] The following key result implies Theorem 2.11. We refer the reader to Section 10 for more details.

Lemma 2.12 (Lemma 10.7 restated)**.** *For any $f : \{0,1\}^\ell \to \{0,1\}$ and $i \in [\ell]$,*

$$\mathbf{oI}_i(f)^2 \leq \sum_{\substack{S \subseteq [i] \\ S \ni i}} \widehat{f}(S)^2 \leq \mathbf{oI}_i(f).$$

Influence vs Online Influence It is not hard to see that $\mathbf{oI}_i[f] \leq \mathbf{I}_i[f]$ for all $i \in [\ell]$, with equality always holding for $i = \ell$ as an adversarial online bit in the last index can see every good bit. Moreover, we observe that for monotone functions, the notion of online influence is equivalent to regular influence, so any separation between the two notions must come from non-monotone functions.

We exactly exhibit such a separation via the address function $\mathrm{Addr}_\ell : \{0,1\}^{\log \ell + \ell} \to \{0,1\}$ which considers its first $\log \ell$ bits as an index in $\{1, \ldots, \ell\}$ and then outputs the value of the chosen index. It is easy to show (as we do in Lemma 10.12) that the first $\log \ell$ bits of Addr_ℓ have no online influence, while the remaining bits have online influence of $O\left(\frac{1}{\ell}\right)$. This is in contrast to the well known result of [KKL88] showing that, for a balanced function such as Addr_ℓ, there must exist a bit with influence at least $\Omega\left(\frac{\log \ell}{\ell}\right)$.

2.5 Extractors via Leader Election Protocols

We sketch our main idea for constructing an extractor for oNOBF sources (Theorem 8.3). Similar ideas work more generally for extracting from oNOSF sources (Theorem 8.4). As mentioned above, we use a novel connection to leader election protocols to construct extractors. We refer the reader to Section 4.4 for a quick recap of the leader election protocols.

Suppose π is an $(r-1)$-round leader election protocol over ℓ players where in each round, each player sends 1 bit and with the guarantee that if there are at most $\delta\ell$ bad players, then a good player is chosen as leader with probability $1 - \epsilon$. Suppose $\mathbf{X}$ is an $(g, \ell r)$-oNOBF source, where $g \geq lr - \delta\ell$. We simply partition the bits of $\mathbf{X}$ into chunks $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_r$, where each X_i is on ℓ bits, and simulate the protocol π by using the j'th bit of $\mathbf{X}_i$ as the message of the j'th player in round i, for all $1 \leq j \leq \ell$ and $1 \leq i \leq r-1$. At the end of this simulation suppose $j^* \in [\ell]$ is the chosen leader. Then we output the j^*'th bit of $\mathbf{X}_r$ as the output of the extractor.

Briefly, the reason that the above is a valid simulation of π is the fact that the value of any bad bit in this online setting just depends on bits that appear before it, which is allowed in the leader election protocol (where in round i, the message of a bad player can be any function of the messages in the same round or previous rounds). The correctness of the extractor now follows from the fact that since the number of bad players (i.e., bad bits in $\mathbf{X}$) is at most $\delta\ell$, the guarantee of the protocol ensures that the chosen leader $j^* \in [\ell]$ is a good player with probability at least $1 - \epsilon$, and in this case the j^*'th bit of $\mathbf{X}_r$ must be uniform.

We note here that in the usual definition of leader election protocols, the requirement is to select a good leader with constant probability, which is a weaker guarantee than what we need to instantiate the above plan. It turns out that we can combine leader election protocols from prior works, in particular from [Fei99] and [AN93], to construct protocols with the stronger guarantee we require. We refer the reader to Section 9 for more details on the construction of our leader election protocols.

[11] We give a very brief recap of necessary notions from Fourier analysis of Boolean functions in Section 10.2.

2.6 Organization

We give some preliminaries in Section 4 before moving on to our core results. Section 5 provides proofs for our explicit condenser constructions with small block length, and Section 6 shows how to handle converting low-entropy oNOSF sources to uniform oNOSF source for a wider range of parameters. Next, Section 7 details our proof for the existence of seedless condensers for oNOSF sources for all regimes of ℓ and n. In Section 8, we present our explicit constructions of extractors for oNOBF and oNOSF sources using a connection to leader election protocols. Then, in Section 9, we explicitly construct the required leader election protocols. Finally, we introduce the notion of online influence in Section 10 and use it to provide an extraction lower bound for oNOBF sources. We discuss some open questions in Section 11.

In Appendix B, we consider a natural local variant of oNOSF sources and show that it is straightforward to extract from such sources using existing extractors for small-space sources.

3 Application to Collective Coin Flipping and Collective Sampling

We now discuss applications of our results on condensers for oNOSF sources to fault-tolerant distributed computing. Condensing from oNOSF sources can be viewed as a special case of coin flipping and collective sampling protocols in the full information model that arise in fault-tolerant distributed computing.

3.0.1 Background

Say there are ℓ players who have a common broadcast channel and want to jointly perform a task such as collectively flipping a coin. Some b players out of them are "bad" and want to deter the task. We assume the bad players are computationally unbounded so cryptographic primitives are of no use. We further assume that each player has private access to uniform randomness. [BL89] initiated the study of this model and aptly termed this task as "collective coin flipping."

The simplest way to collectively flip a coin would be for all the players to initially agree on a function $f : \{0,1\}^\ell \to \{0,1\}$, then synchronously broadcast one random bit r_i, and to finally agree on the output being $f(r_1, \ldots, r_\ell)$. However, synchronizing broadcasts is hard, and it could be that the bad players set their output as function of the bits of the good players. [KKL88] showed that no function f can handle more than $O\left(\frac{\ell}{\log \ell}\right)$ corruptions.

One way to allow for more corruptions (almost linear) among players is to consider "protocols" that allow more rounds of communication. In particular, a protocol can be thought of as a tree where each vertex represents a "round" where in every round the following happens: all good players sends their bits, then all bad players send their bits as a function of the bits of the good players, and they jointly compute a function of these bits. Depending on the outcome of the function, everyone branches on one branch in this tree. Furthermore, every leaf is labeled with final outcomes (say 0 or 1) and, once a leaf is reached, that is the outcome that everybody agrees on. [GGL98] initiated the study of protocols where the outcomes are from a larger range and where the bad players are trying to minimize the largest probability of any outcome. They called this problem "collective sampling." For a formal definition, see Section 4.4.

3.0.2 Known Results

[BL89] showed that for protocols with outcomes $\{0,1\}$, b bad players can always ensure that some outcome occurs with probability at least $\frac{1}{2} + \frac{b}{2\ell}$. [AN93] first constructed a protocol that can handle a linear number of corruptions. Follow-up works tried to reduce the number of rounds in this protocol where, in some settings, players were allowed to send more than one bit per round [RZ01, Fei99].

[GGL98] showed that for all collective sampling protocols and all outcomes, there exists a way for b bad players to coordinate and ensure that an outcome that happens without corruption with probability p, now happens with probability $p^{1-(b/n)} \geq p\left(1 + \frac{b}{n}\log(1/p)\right)$. Nearly matching collective sampling protocols were constructed by [GGL98, SV08, GVZ06]. For an overview of further results and bounds, see [Dod06].

3.0.3 Connection to oNOSF Sources

The problem of extracting or condensing from oNOSF sources can be seen as special cases or variants of collective coin flipping and collective sampling that provide very simple protocols. For instance, suppose one has an extractor or condenser f for uniform (g, ℓ, n)-oNOSF sources. Then, consider a protocol where all ℓ players take turns and output n random bits. The agreed final outcome is f applied on these ℓn bits. This leads to protocols that are structurally much simpler since players don't have to carefully compute whose turn it is to go in various rounds and can obliviously prepare for their turn.

The above protocol can also be viewed as a relaxed version of a 1-round protocol where instead of everyone providing their output asynchronously, they take turns and provide outputs one after another in a simple sequential manner.

3.0.4 Previous Results Interpreted in oNOSF source context

Previous impossibility results can be interpreted in the context of extracting / condensing from uniform oNOSF sources. For instance, collective coin flipping impossibility results of [BL89] imply extraction impossibility results for uniform (g, ℓ, n)-oNOSF sources when $n = 1$. They imply:

Corollary 3.1. *There does not exist an $\frac{b}{2\ell}$-extractor for uniform $(g, \ell, 1)$-oNOSF sources.*

Similarly, we observe that the notion of collective sampling is equivalent to 0-error condensing. Hence, lower bounds of [GGL98] imply zero-error condensing lower bounds for uniform (g, ℓ, n)-oNOSF sources when $n = 1$. Formally:

Corollary 3.2. *There does not exist a condenser $\mathsf{Cond} : \{0,1\}^\ell \to \{0,1\}^m$ for uniform $(g, \ell, 1)$-oNOSF sources that can guarantee output smooth min-entropy (with parameter $\varepsilon = 0$) more than $k = \frac{g}{\ell} \cdot m$.*

3.0.5 ε-Collective Sampling

Since collective sampling lower bounds show that for any protocol, 0-error condensing beyond rate g/ℓ is impossible, one can naturally ask whether condensing with small error ε is possible. We call this problem *ε-collective sampling*, where the goal is to output a distribution which is ε-close to a distribution where every output has small probability.

Interpreted this way, this is exactly what protocols arising out of our condensers for uniform oNOSF sources provide: Using Theorem 3, when each player has access to 10^4 random bits, there exists a simple protocol that can handle 0.49ℓ corrupt players such that the players can collectively sample a distribution over $m = O(\ell)$ bits which is $2^{-\Omega(\ell)}$-close to having entropy $0.99m$. As far as we are aware, such a protocol is not implied by any other previous protocol. Most previous protocols are obtained through *leader election* protocols, which do not seem useful here since the leader has access to only constant number of bits.

We similarly obtain explicit protocols using Theorem 2 for the case when each player has access to $n \geq \mathrm{poly}(\log(\ell)/\varepsilon)$ many bits.

3.0.6 Collective Coin Flipping and Sampling with Weak Random Sources

A natural extension to collective coin flipping and sampling in the full information model is when all players only have access to weak source of randomness (that are independent from each other) instead of true uniform randomness. This question was first studied by [GSV05]. [KLRZ08] used network extractor protocol to transform weak random sources of each player into independent private random sources. This way, after using the network extraction protocol, players can follow the usual collective coin flipping / sampling protocol. [GSZ21] improved the network extraction protocol using two-source non-malleable extractors.

Using our (g, ℓ, n, k)-oNOSF source condensers, we obtain alternative, simple ε-collective sampling protocols in the setting where players have access to weak sources of randomness. We obtain such an existential protocol using Theorem 4, and explicit protocol using Corollary 1.3.

4 Preliminaries

In this section we give some basic background and facts used throughout our paper. We use boldfaced font to indicate a random variable such as $\mathbf{X}$. Often we will use $\circ$ or , to indicate concatenation of blocks. So if $\mathbf{X}_1 \sim \{0,1\}^n$ and $\mathbf{X}_2 \sim \{0,1\}^n$, then $\mathbf{X}_1, \mathbf{X}_2$ will be the concatenated random variable over $\{0,1\}^{2n}$. We will use the notation $[n]$ as shorthand for $\{1, \ldots, n\}$. All logs in this paper will have base 2 unless stated otherwise.

4.1 Basic Probability Notions

We measure the distance between two distributions via statistical distance:

Definition 4.1 (Statistical Distance)**.** *For any two distributions $\mathbf{X}, \mathbf{Y}$ over Ω, we define the statistical distance or total-variation distance (TV) distance as:*

$$|\mathbf{X} - \mathbf{Y}| = \max_{S \subset \Omega} |\Pr[\mathbf{X} \in S] - \Pr[\mathbf{Y} \in S]| = \frac{1}{2} \sum_{s \in \Omega} |\Pr[\mathbf{X} = s] - \Pr[\mathbf{Y} = s]|$$

We use the notation $\mathbf{X} \approx_\varepsilon \mathbf{Y}$ to denote the fact that $|\mathbf{X} - \mathbf{Y}| \leq \varepsilon$.

We also state the useful folklore result of the data processing inequality.

Fact 4.2. *For any two distributions $\mathbf{X}, \mathbf{Y}$ over Ω and function $f : \Omega \to \mathbb{R}$,*

$$|\mathbf{X} - \mathbf{Y}| \geq |f(\mathbf{X}) - f(\mathbf{Y})|.$$

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

Lemma 4.3 (Min-entropy chain rule, [MW97])**.** *For any random variables $\mathbf{X} \sim X$ and $\mathbf{Y} \sim Y$ and $\varepsilon > 0$,*

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

4.2 Condensers and Extractors

We recall the definition of a seeded condenser.

Definition 4.4. *A $(k_{in}, k_{out}, \varepsilon)$-seeded condenser* $\mathsf{sCond} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ *satisfies the following: for every source* $\mathbf{X} \sim \{0,1\}^n$ *with* $H_\infty(\mathbf{X}) \geq k_{in}$*, and* $\mathbf{Y} = \mathbf{U}_d$*,*

$$H^\varepsilon_\infty(\mathsf{Cond}(\mathbf{X}, \mathbf{Y})) \geq k_{out}.$$

Here, d is called the seed length *of* sCond*.*

A seeded extractor is the special case of seeded condenser where $k_{out} = m$. We record the full definition for completeness sake:

Definition 4.5. *A (k, ε)-seeded extractor* $\mathsf{sExt} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ *satisfies the following: for every source* $\mathbf{X} \sim \{0,1\}^n$ *with* $H_\infty(\mathbf{X}) \geq k$*, and* $\mathbf{Y} = \mathbf{U}_d$*,*

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

Here, d is called the seed length *of* sExt*.* sExt *is called strong if*

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

We will use the following near optimal explicit construction of seeded extractors:

Theorem 4.6 (Theorem 1.5 in [GUV09])**.** *For all constant $0 < \alpha < 1$, there exists a constant C such that for 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 $d = C \log(n/\varepsilon)$ and $m \geq (1-\alpha)k$.*

Next, we recall the definition of two-source extractors.

Definition 4.7. *A function* $\mathsf{2Ext} : \{0,1\}^{n_1} \times \{0,1\}^{n_2} \to \{0,1\}^m$ *is a* (k_1, k_2, ε)-two-source extractor *if for every source* $\mathbf{X}_1 \sim \{0,1\}^{n_1}$ *with* $H_\infty(\mathbf{X}_1) \geq k_1$ *and* $\mathbf{X}_2 \sim \{0,1\}^{n_2}$ *with* $H_\infty(\mathbf{X}_2) \geq k_2$ *where* $\mathbf{X}_1$ *and* $\mathbf{X}_2$ *are independent of each other, we have*

$$\mathsf{2Ext}(\mathbf{X}_1, \mathbf{X}_2) \approx_\varepsilon \mathbf{U}_m.$$

It is said to be strong in the first argument *if*

$$\mathsf{2Ext}(\mathbf{X}_1, \mathbf{X}_2), \mathbf{X}_1 \approx_\varepsilon \mathbf{U}_m, \mathbf{X}_1.$$

4.3 Averaging Samplers

Recall the definition of an averaging sampler

Definition 4.8. *A (k, δ, ε)-averaging sampler is a function* $\mathsf{Samp} : \{0,1\}^n \to (\{0,1\}^m)^D$ *such that for any function $f : \{0,1\}^m \to [0,1]$ and any (n,k)-source $\mathbf{X}$, we have that*

$$\Pr_{(x_1,\ldots,x_D)\sim\mathsf{Samp}(\mathbf{X})}\left[\left|\frac{1}{D}\sum_{i=1}^{D} f(x_i) - \mathop{\mathbb{E}}_{x\sim\mathbf{U}_m}[f(x)]\right| \geq \varepsilon\right] \leq \delta.$$

It was shown in [Zuc97] that strong extractors and averaging samplers are equivalent. We reproduce the proof here for completeness.

Lemma 4.9. *Let* $\mathsf{Ext} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ *be a (k, ε)-extractor. Define* $\mathsf{Samp} : \{0,1\}^n \to (\{0,1\}^m)^D$*, where $D = 2^d$ as $\mathsf{Samp}(x) = (\mathsf{Ext}(x,1), \ldots, \mathsf{Ext}(x,D))$ where we identify $[D]$ with $\{0,1\}^d$. Then* Samp *is a $(k + \log(1/\delta), \delta, \varepsilon)$-averaging sampler.*

Proof. Let $\mathbf{X}$ be an $(n, k + \log(1/\delta))$-source. For $x \in \mathrm{Supp}(\mathbf{X})$, call x *bad* if $|\mathsf{Ext}(x, \mathbf{U}_d) - \mathbf{U}_m| > \varepsilon$ and let $B \subseteq \{0,1\}^n$ be the subset of bad x's. Now, suppose for the sake of contradiction that $\Pr[\mathbf{X} \in B] > \delta$. Then, letting $\mathbf{X}_B = \mathbf{X} \mid (\mathbf{X} \in B)$, we can compute the min-entropy of $\mathbf{X}_B$ as

$$\begin{aligned} H_\infty(\mathbf{X}_B) &\geq H_\infty(\mathbf{X}) - \log\left(\frac{1}{\Pr[\mathbf{X} \in B]}\right) \\ &\geq k + \log(1/\delta) - \log(1/\delta) \\ &= k. \end{aligned}$$

Therefore, we can apply Ext to $\mathbf{X}_B$ and obtain that $|\mathsf{Ext}(\mathbf{X}_B, \mathbf{U}_d) - \mathbf{U}_n| \leq \varepsilon$. However, by assumption we know that for all $x \in B$, $|\mathsf{Ext}(x, \mathbf{U}_d) - \mathbf{U}_m| > \varepsilon$, meaning that $|\mathsf{Ext}(\mathbf{X}_b, \mathbf{U}_d) - \mathbf{U}_m| > \varepsilon$, giving us a contradiction. Thus, we have that $\Pr[\mathbf{X} \in B] \leq \delta$.

Now, we turn our attention to the good $x \notin B$. Using Fact 4.2, we know that for any $x \notin B$,

$$\begin{aligned} |f(\mathsf{Ext}(x, \mathbf{U}_d)) - f(\mathbf{U}_m)| &\leq |\mathsf{Ext}(x, \mathbf{U}_d) - \mathbf{U}_m| \\ &\leq \varepsilon. \end{aligned}$$

This is equivalent to saying that, for all good $x \in \{0,1\}^n$, $\left|\frac{1}{D}\sum_{s \in \{0,1\}^d} f(\mathsf{Ext}(x, s)) - f(\mathbf{U}_m)\right| \leq \varepsilon$. This is exactly the requirement of our sampler, and we have shown that this happens with probability $\Pr[\mathbf{X} \notin B] \geq 1 - \delta$, as required. □

In particular, we will use the following strong extractor from [Zuc07] to instantiate an averaging sampler.

Theorem 4.10 ([Zuc07])**.** *For all constant $\alpha, \delta, \varepsilon > 0$, there is an efficient family of strong $(k = \delta n, \varepsilon)$-extractors $\mathsf{Ext} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ with $m \leq (1-\alpha)\delta n$ and $D = 2^d = O(n)$.*

Using Lemma 4.9, we get the following averaging sampler that we shall use later.

Lemma 4.11. *For all constant $\alpha, \delta, \varepsilon > 0$, we can construct an explicit sampler $\mathsf{Samp} : \{0,1\}^t \to (\{0,1\}^m)^D$ with $m \leq (1-\alpha)\delta t$ and $D = O(t)$ such that for all sets $S \subseteq [M]$, where $M = 2^m$, and all (t, k)-sources X, we have that*

$$\Pr_{x \sim X}\left[\left|\frac{|\mathsf{Samp}(x) \cap S|}{D} - \frac{|S|}{M}\right| \geq \varepsilon\right] \leq 2^{\delta t - k}.$$

Proof. Simply apply Lemma 4.9 to Theorem 4.10 and consider the indicator function $f(x) = 1_{x \in S}$ of S for the resulting sampler. □

4.4 Leader Election, Collective Coin Flipping, and Sampling Protocols

We formalize the definition of protocols in the full information model. Collective coin flipping protocols, leader election protocols, and collective sampling protocols are special cases of such protocols where the output domain is $[\ell]$ and $\{0,1\}$ and $\{0,1\}^m$ for some m respectively.

Definition 4.12 (Protocol in the full information model)**.** *A k-round protocol with output domain Y over ℓ players where each player sends n random bits per round is a function*

$$\pi : \left(\left(\{0,1\}^n\right)^\ell\right)^k \to Y$$

that takes in the input of each of the players during each round and outputs an element from set Y which is the outcome of the protocol.

Here is how the protocol operates in the presence of a set $B \subset [\ell]$ of bad players: In round i, each of the players from $[\ell] \setminus B$ independently output a uniformly random element from $\{0,1\}^n$. Let their collective outputs be $\alpha_i \in (\{0,1\}^n)^{[\ell]\setminus B}$. Then, depending on $\alpha_1, \ldots, \alpha_i$, the players in B together output an element of $(\{0,1\}^n)^B$. Hence, we model the strategy of the bad players as a sequence of functions $\sigma = (\sigma_1, \ldots, \sigma_k)$, where

$$\sigma_i : \left((\{0,1\}^n)^{[\ell]\setminus B}\right)^i \to (\{0,1\}^n)^B,$$

where σ_i takes in the inputs of the good players from the first i rounds and maps it to the output of the bad players for round i. For a fixed strategy σ, the outcome of the protocol can be modeled as follows: uniform random strings $\alpha_1, \ldots, \alpha_k \in (\{0,1\}^n)^{[\ell]\setminus B}$ are chosen, and the outcome of the protocol is

$$\pi(\alpha_1 : \sigma_1(\alpha_1), \alpha_2 : \sigma_2(\alpha_1, \alpha_2), \ldots, \alpha_k : \sigma_k(\alpha_1, \ldots, \alpha_k)).$$

We now specialize this definition to define collective coin flipping protocols

Definition 4.13 (Collective coin flipping protocol)**.** *A* collective coin flipping protocol *π is a protocol in the full information model with output domain $Y = \{0,1\}$. Furthermore, we say π is (b,γ) resilient if in the presence of any set B of bad players with $|B| \le b$, we have that $\max_{o\in\{0,1\}} \Pr[\pi|_B = o] \le 1-\gamma$.*

Note that when $k = 1$, the protocol π just becomes a function over $\{0,1\}^\ell$; such 1-round coin flipping protocols which cannot be biased by any small set of bad players are also known as *resilient functions*.

We also specialize the definition of protocols to define leader election protocols:

Definition 4.14 (Leader election protocol)**.** *A* leader election protocol *π is a protocol in the full information model with output domain $Y = [\ell]$, the number of players the protocol is operating on. Furthermore, we say π is (b,γ) resilient if in the presence of any set B of bad players with $|B| \le b$, we have that $\Pr[\pi|_B \in B] \le 1-\gamma$.*

Remark 4.15. *The definition of resilience that we use, which is standard in the leader election and collective coin flipping literature, requires only that bad players can be elected as a leader with probability at most $1-\gamma$. Our leader election protocols satisfy (and need) the stronger measure of quality that is standard in the pseudorandomness literature: that bad players are chosen with probability at most ε for small ε.*

We lastly define collective sampling protocols:

Definition 4.16 (Collective sampling protocol)**.** *A* collective sampling protocol *π is a protocol in the full information model, typically with output domain $Y = \{0,1\}^m$ for some m which is a function of ℓ and n. The goal of collective sampling protocols is to ensure that for every output set $S \subset \{0,1\}^m$ with density μ, in the presence of b bad players, the probability that the output lies in S is at most ε, with the goal to make ε as close to μ as possible.*

5 Explicit Condensers for oNOSFs with Small Block Length

In this section we will give an explicit construction of condensers for uniform oNOSF sources with small block length. Our main theorem shows that, given an oNOSF source[12] with slightly more than a half fraction

[12] Unless stated otherwise, in this section we will use oNOSF source to refer to uniform oNOSF source

of good players, the block length only needs to be polylogarithmic in the length of the source to condense from it. This nearly matches the lower bound that it is impossible to condense above entropy rate $1/2$ when the fraction of good players is at most $1/2$.

Theorem 5.1. *There exists a universal constant C such that for any constant $\gamma > 0$ the following holds. For all $0 < \varepsilon < 1/2$ and $n, \ell \in \mathbb{N}$ where $n \geq \left(\frac{\log(\ell)}{\varepsilon}\right)^C$, there exists an explicit condenser $\mathsf{Cond} : (\{0,1\}^n)^\ell \to \{0,1\}^m$ satisfying: For any $(g = (1/2+\gamma)\ell, \ell, n)$-oNOSF source $\mathbf{X}$, we have that $H^\varepsilon_\infty(\mathsf{Cond}(\mathbf{X})) \geq m - \left(\frac{\log(\ell)}{\varepsilon}\right)^C \log(n)$ where $m = \frac{1}{3} \cdot \gamma\ell n - C\ell \log(\ell)\log(1/\varepsilon)$.*

We will prove this result in Section 5.5.

Using the transformation of low-entropy oNOSF sources to uniform oNOSF sources from Corollary 6.4, we get an explicit condenser for low-entropy oNOSF sources.

Corollary 5.2. *There exists universal constants C, C' such that for any constant $\gamma > 0$ the following holds. For all $n, \ell, k \in \mathbb{N}$ where $n \geq \ell^C, k \geq (\log(n))^C$, there exists an explicit condenser $\mathsf{Cond} : (\{0,1\}^n)^\ell \to \{0,1\}^m$ satisfying: For any $(g = (1/2+\gamma)\ell, \ell, n, k)$-oNOSF source $\mathbf{X}$, we have that $H^\varepsilon_\infty(\mathsf{Cond}(\mathbf{X})) \geq m - (\log(\ell))^C \log(n)$ where $m = \frac{1}{4} \cdot \gamma\ell n - C\ell \log(\ell)\log(\log(\ell))$ and $\varepsilon = \frac{1}{(\log(\ell))^{C'}}$.*

Proof. To do this, we apply Corollary 6.4 with d as a very large universal constant, and its number of output bits $m = (\log(\ell))^{C_0}$ for some large universal constant C_0. We obtain that with error at most $\varepsilon/2$, the resultant distribution is uniform $\left(g = (1/2 + 0.99\gamma)\ell', \ell', (\log(\ell'))^{C_1}\right)$-oNOSF source where $\ell' = \ell - 1$ and C_1 is a large universal constant. On that resultant source, we apply Theorem 5.1 with error $\varepsilon/2$ to obtain the desired result. □

We will need two main tools to prove Theorem 5.1. The first one will allow us to use an oNOSF source to sample a logarithmically sized 'committee' from any given subset of players that still has approximately the same fraction of good players.

Lemma 5.3. *There exists a universal constant C such that for all constant $0 < \gamma, \varepsilon_a < 1$, the following holds. For all $0 < \varepsilon_s < 1/2$, and $n, \ell \in \mathbb{N}$ where $n \geq 6\log(\ell)\log(1/\varepsilon_s)/\gamma$, there exists an explicit function $\mathsf{oNOSFSamp} : (\{0,1\}^n)^\ell \to [\ell]^D$ where $D \leq C\log(\ell/\varepsilon_s)$ with the following property.[13] For all $S \subset [\ell]$ and $(\gamma\ell, \ell, n)$-oNOSF sources $\mathbf{X}$, we have that*

$$\Pr_{x \sim \mathbf{X}}\left[\left|\frac{|\mathsf{oNOSFSamp}(x) \cap S|}{D} - \frac{|S|}{\ell}\right| \geq \varepsilon_a\right] \leq \varepsilon_s$$

We will prove this in Section 5.6.

Our second tool is a seeded condenser that works even when the seed is an oNOSF source. In fact, the bad bits in the seed are allowed to depend on the general min-entropy source that the condenser is acting on.

Lemma 5.4. *There exists a universal constant C such that for all $n_x, k, n_y, t \in \mathbb{N}$ with $\varepsilon > 0$ and $n_y \geq (C)^t \log(tn_x/\varepsilon)$, there exists an explicit condenser $\mathsf{2Cond} : \{0,1\}^{n_x} \times (\{0,1\}^{n_y})^t \to \{0,1\}^m$ where $m = \frac{1}{3}(k - (C)^t \log(tn_x/\varepsilon))$ so that the following holds: For all (n_x, k)-sources $\mathbf{X}$ and $(g = 1, \ell = t)$-oNOSF sources $\mathbf{Y} \sim (\{0,1\}^{n_y})^t$ such that the good blocks in $\mathbf{Y}$ are independent of $\mathbf{X}$ and the bad blocks in $\mathbf{Y}$ can depend on $\mathbf{X}$, we have that $H^\varepsilon_\infty(\mathsf{2Cond}(\mathbf{X}, \mathbf{Y})) \geq m - (C)^t \log(tn_x/\varepsilon)$.*

We prove this in Section 5.7.

[13] Even though the output domain of $\mathsf{oNOSFSamp}$ is a vector, we will abuse notation and often treat it as a set

Remark 5.5. *In Lemma 5.4, we have the requirement that $n_y \geq (C)^t \log(tn_x/\varepsilon)$ so that both $m > 0$ and* Cond *outputs non-zero min-entropy. Therefore, decreasing t will give a smaller bound on the length of each block of $\mathbf{Y}$ and, ultimately, the block length of the oNOSF source we are condensing from. Our trick here will be to get t down to $t \approx O(\log\log(\ell))$ when starting from a (g, ℓ, n)-oNOSF source so that the block length of this source only needs to be at least* $\mathrm{polylog}(\ell)$.

To give some intuition behind our explicit construction in Theorem 5.1, this section is organized in parts that each provide a step towards the proof. For base line construction, we observe that Lemma 5.13 already yields an explicit condenser for (g, ℓ)-oNOSF sources where $g > \ell/2$ with block length $n \geq \exp(\ell)$. This is formally proven in Section 5.1. In Section 5.2, we will first see how we use both of these main tools from above to handle the easier setting of condensing from a (g, ℓ)-oNOSF source with the number of good blocks $g > \frac{2}{3}\ell$ and we have polynomial block length $n \geq \mathrm{poly}(\ell)$, an exponential improvement over the base line construction. We build upon these ideas to further exponentially decrease our block length requirement in Section 5.3 to handle oNOSF sources with $g > \frac{3}{4}\ell$ and $n \geq \mathrm{polylog}(\ell)$. To then decrease the fraction of good blocks that we require, we introduce a correlated sampling trick in Section 5.4 so that we only require $g > \frac{2}{3}\ell$ while retaining the requirement $n \geq \mathrm{poly}(\log(\ell))$. Finally, we obtain Theorem 5.1, which only requires $g > \frac{1}{2}\ell$ and $n \geq \mathrm{poly}(\log(\ell))$ in Section 5.5 by repeating such a correlated sampling trick, carefully handling the fact that our sources lose some structure each time we do so. Each of these 4 sections is self contained and only depends only on Lemma 5.3 and Lemma 5.4.

5.1 Condensing from 51% good oNOSF sources with $n \geq \exp(\Omega(\ell))$

We first construction our baseline condenser that requires at least 51% good blocks but requires the block length $n \geq \exp(\Omega(\ell))$. This construction solely relies on Lemma 5.4:

Theorem 5.6. *There exists a universal constant C such that for any constant $\gamma > 0$, the following holds. For all $n, \ell \in \mathbb{N}$, and $0 < \varepsilon < 1/2$ where $n \geq (C)^\ell \log(1/\varepsilon)$, there exists an explicit condenser* $\mathsf{Cond} : (\{0,1\}^n)^\ell \to \{0,1\}^m$ *satisfying: for any uniform $(g = (1/2+\gamma)\ell, \ell)$-oNOSF source $\mathbf{X}$, we have $H^\varepsilon_\infty(\mathsf{Cond}(\mathbf{X})) \geq m - (C)^\ell \log(\ell n/\varepsilon)$ where $m = \frac{1}{6} \cdot \gamma\ell n$.*

Proof. If ℓ is odd, then we split each block of $\mathbf{X}$ into two contiguous blocks of length $n/2$ each and view $\mathbf{X}$ as $((1/2+\gamma)2\ell, \ell)$-oNOSF source. This allows us to without loss of generality assume ℓ is even since this transformation preserves the output guarantees required by our condenser.

We begin by decomposing $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2)$ where in this decomposition we are simply splitting $\mathbf{X}$ into two parts, so that each $\mathbf{X}_i$ is a $(g = (2\gamma) \cdot (\ell/2), (\ell/2), n)$-oNOSF source. We use $\mathsf{2Cond}$ from Lemma 5.4 with $n_x = \ell n/2, k = \gamma\ell n, n_y = n$, error parameter equal to ε and output $\mathsf{2Cond}(\mathbf{X}_1, \mathbf{X}_2)$. Let C_{2Cond} be the universal constant from Lemma 5.4. We let our universal constant C be much larger than C_{2Cond} so that we satisfy $n \geq (C_{\mathsf{2Cond}})^\ell \log(\ell^2 n/\varepsilon)$ as required by Lemma 5.4 and also so that in odd ℓ cases, $n/2$ is also sufficiently large. □

5.2 Condensing from 67% good oNOSF sources with $n \geq \mathrm{poly}(\ell)$

We begin by constructing condenser that requires at least 67% good blocks instead of just 51%, but allows for the block length n to be polynomial instead of super-exponential in ℓ. Formally, we will show that:

Theorem 5.7. *There exists a universal constant C such that for any constant $\gamma > 0$ the following holds. For all $0 < \varepsilon < 1/2$ and $n, \ell \in \mathbb{N}$ where $n \geq \left(\frac{\ell}{\varepsilon}\right)^C$, there exists an explicit condenser* $\mathsf{Cond} : (\{0,1\}^n)^\ell \to \{0,1\}^m$ *satisfying: For any $(g = (2/3+\gamma)\ell, \ell, n)$-oNOSF source $\mathbf{X}$, we have that $H^\varepsilon_\infty(\mathsf{Cond}(\mathbf{X})) \geq m - \left(\frac{\ell}{\varepsilon}\right)^C \log(n)$ where $m = \frac{1}{3} \cdot \gamma\ell n$.*

The key idea here will be to sample a few blocks in the last third of the source to take short prefixes from and then use these as the seed of a particular two source condenser. With this, we present the proof.

Proof of Theorem 5.7. Let $\ell' = \ell/3$. We begin by decomposing $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3)$ where in this decomposition we are simply splitting $\mathbf{X}$ into thirds, so each $\mathbf{X}_i$ is a $(g = 3\gamma\ell', \ell', n)$-oNOSF source.

Our first step is to use $\mathbf{X}_1$ to sample a logarithmically sized committee of players from $\mathbf{X}_3$. We do so by using $\mathsf{oNOSFSamp} : (\{0,1\}^n)^{\ell'} \to [\ell']^D$ from Lemma 5.3 with $\varepsilon_s = \varepsilon/2$, $\varepsilon_a = \gamma$, the corresponding γ equal to 3γ, and the S of Lemma 5.3 corresponding to the indices $B \subseteq [\ell]$ of the bad players in $\mathbf{X}_3$, so $\frac{|B|}{\ell'} \leq 1 - 3\gamma$. Note that $D \leq C_0 \log(\ell/\varepsilon_s) \leq C_1 \log(\ell/\varepsilon))$ where C_0, C_1 are some universal constants. Lemma 5.3 then allows us to conclude that

$$\begin{aligned}
&\Pr_{x\sim\mathbf{X}_1}\left[\left|\frac{|\mathsf{oNOSFSamp}(x)\cap B|}{D} - \frac{|B|}{\ell'}\right| \geq \varepsilon_a\right] \leq \varepsilon_s \\
\implies &\Pr_{x\sim\mathbf{X}_1}\left[\frac{|\mathsf{oNOSFSamp}(x)\cap B|}{D} \leq \varepsilon_a + \frac{|B|}{\ell'}\right] \geq 1 - \varepsilon_s \\
\implies &\Pr_{x\sim\mathbf{X}_1}\left[\left|\mathsf{oNOSFSamp}(x)\cap \overline{B}\right| \geq 2\gamma\cdot D\right] \geq 1 - \frac{\varepsilon}{2}.
\end{aligned}$$

Let $\mathbf{Y} = (\mathbf{X}_3)_{\mathsf{oNOSFSamp}(\mathbf{X}_1)}$ be the $D \leq C_1 \log(\ell'/\varepsilon)$-sized committee of players from $\mathbf{X}_3$ chosen by $\mathsf{oNOSFSamp}(\mathbf{X}_1)$. The above can then be interpreted as saying that, with at least a $1-\varepsilon/2$ probability over $\mathbf{X}_1$, we have that $\mathbf{Y}$ is a $(g = 2\gamma D, D, n)$-oNOSF source (we will only need the fact that $\mathbf{Y}$ contains at least one good block, though).

Our second step is to apply the condenser from Lemma 5.4 to $\mathbf{X}_2$ and $\mathbf{Y}$. We instantiate Lemma 5.4 with $n_x = \ell' n$, $n_y = \left(\frac{\ell}{\varepsilon}\right)^{C'} \log(n)$, $t = D \leq C_1 \log(\ell'/\varepsilon)$, $k = 3\gamma\ell' n$, and error equal to $\frac{\varepsilon}{2}$ where C' is a large enough constant. One can verify that this setting of parameters along with our assumption that $n \geq \left(\frac{\ell}{\varepsilon}\right)^C$ for a universal constant C (by setting it to be large enough) satisfies the requirements of Lemma 5.4. This yields a condenser that we shall call $\mathsf{2Cond} : (\{0,1\}^n)^{\ell'} \times (\{0,1\}^n)^D \to \{0,1\}^m$ (so as to avoid confusion with our ultimate condenser Cond).

Lastly, we let $\mathbf{Z} = \mathsf{2Cond}\left(\mathbf{X}_2, (\mathbf{X}_3)_{\mathsf{oNOSFSamp}(\mathbf{X}_1)}\right)$ and let m_z be the length of $\mathbf{Z}$. If $m_z \leq m = \frac{\gamma\ell n}{3}$, then we output $\mathbf{Z}$ followed by $m_z - m$ many zeros. Otherwise, we output a prefix of $\mathbf{Z}$ of length $m = \frac{\gamma\ell n}{3}$.

We now analyze the guarantees of this condenser. We first observe that

$$\begin{aligned}
(C_{\mathsf{2Cond}})^t \log(2tn_x/\varepsilon) &\leq (C_{\mathsf{2Cond}})^{C_1(\log(\ell/3\varepsilon))} \log(4\cdot C_1 \log(\ell/\varepsilon)\ell n/\varepsilon) \\
&\leq \left(\frac{\ell}{\varepsilon}\right)^{C_2} \log(n) \qquad (*)
\end{aligned}$$

where C_{2Cond} is the universal constant from Lemma 5.4 (we will use this notation in subsequent proofs) and C_2 is a large enough universal constant. With this, we are guaranteed that the length of $\mathbf{Z} = m_z$ is such that

$$\begin{aligned}
m_z &= \frac{1}{3}\left(k - (C_{\mathsf{2Cond}})^t \log(2tn_x/\varepsilon)\right) \\
&= \frac{1}{3}\left(\gamma\ell n - (C_{\mathsf{2Cond}})^t \log(2tn_x/\varepsilon)\right) \\
&\geq \frac{\gamma\ell n}{3} - \left(\frac{\ell}{\varepsilon}\right)^{C_2} \log(n) \qquad \text{(by Equation } (*)\text{)} \\
&= m - \left(\frac{\ell}{\varepsilon}\right)^{C_2} \log(n)
\end{aligned}$$

Moreover, conditioned on the fact that $\mathbf{Y}$ has at least one good block, Lemma 5.4 guarantees that

$$\begin{aligned} H_\infty^{\varepsilon/2}(\mathbf{Z}) &= H_\infty^{\varepsilon/2}(2\mathsf{Cond}(\mathbf{X}_2, \mathbf{Y})) \\ &\geq m_z - (C_{2\mathsf{Cond}})^t \log(2tn_x/\varepsilon) \\ &\geq m_z - \left(\frac{\ell}{\varepsilon}\right)^{C_2} \log(n) \qquad \text{(by Equation } (*)\text{)} \end{aligned}$$

Thus, adding up the two $\varepsilon/2$ errors from both our steps, we see that $H_\infty^\varepsilon(\mathbf{Z}) \geq m_z - \left(\frac{\ell}{\varepsilon}\right)^{C_2} \log(n)$.

If $m_z > m$, then our final output inherits the smooth min-entropy gap of $\mathbf{Z}$ which is $\left(\frac{\ell}{\varepsilon}\right)^{C_2} \log(n)$. If $m_z \leq m$, then our output inherits not only the entropy gap of $\mathbf{Z}$ but also an entropy gap of $m - m_z$. Since $m_z \geq m - \left(\frac{\ell}{\varepsilon}\right)^{C_2} \log(n)$, our output will have smooth min-entropy gap at most $2\left(\frac{\ell}{\varepsilon}\right)^{C_2} \log(n)$. In either case, our gap will be at most $2\left(\frac{\ell}{\varepsilon}\right)^{C_2} \log(n)$. We let our final universal constant C be much larger than C_2 to obtain our claim. □

Remark 5.8. *The padding or truncating trick at the end of all our steps to meet the desired output length is standard and for next subsections and proofs, we will omit it and use it implicitly.*

5.3 Condensing from 76% good oNOSF sources with $n \geq \mathrm{polylog}(\ell)$

To decrease our block length requirement all the way down to $\mathrm{polylog}(\ell)$, we simply apply the idea in the previous section twice. We split up our oNOSF source $\mathbf{X}$ into four blocks $\mathbf{X} = \mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \mathbf{X}_4$ and use $\mathbf{X}_1$ to sample a logarithmically sized committee from $\mathbf{X}_2$, use this committee to sample a doubly logarithmically sized committee from $\mathbf{X}_4$, and finally apply the condenser from Lemma 5.4 to $\mathbf{X}_3$ and this final committee.

Theorem 5.9. *There exists a universal constant C such that for any constant $\gamma > 0$ the following holds. For all $0 < \varepsilon < 1/2$ and $n, \ell \in \mathbb{N}$ where $n \geq \left(\frac{\log(\ell)}{\varepsilon}\right)^C$, there exists an explicit condenser $\mathsf{Cond} : (\{0,1\}^n)^\ell \to \{0,1\}^m$ satisfying: For any $(g = (3/4+\gamma)\ell, \ell, n)$-oNOSF source $\mathbf{X}$, we have that $H_\infty^\varepsilon(\mathsf{Cond}(\mathbf{X})) \geq m - \left(\frac{\log(\ell)}{\varepsilon}\right)^C \log(n)$ where $m = \frac{1}{3} \cdot \gamma\ell n$.*

Proof. Let $\ell' = \ell/4$. We decompose $\mathbf{X}$ into quarters as $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \mathbf{X}_4)$, so each $\mathbf{X}_i$ is a $(g = 4\gamma\ell', \ell', n)$-oNOSF source. We in fact claim something stronger. Call an index $i \in [\ell']$ totally good if it is good in each of $\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \mathbf{X}_4$. For the rest of the $i \in [\ell']$ that are not totally good, we refer to them as somewhat bad. Since $\mathbf{X}$ has $(3+4\gamma)\ell'$ good indices out of $4\ell'$, we see that there are must be at least $4\gamma\ell'$ totally good indices, i.e., indices that are good across all of the 4 blocks.

Our first step is to use $\mathbf{X}_1$ to sample a logarithmically sized committee of players from $\mathbf{X}_2$. We obtain $\mathsf{oNOSFSamp}_2 : (\{0,1\}^n)^{\ell'} \to [\ell']^{D_2}$ from Lemma 5.3 with $\varepsilon_s = \varepsilon/3$, $\varepsilon_a = \gamma$, the corresponding γ equal to 4γ, and the set S of Lemma 5.3 corresponding to the indices $B_2 \subset [\ell]$ of the somewhat bad players in $\mathbf{X}_2$, so $\frac{|B_2|}{\ell'} \leq 1 - 4\gamma$. Note that $D_2 \leq C_0 \log(\ell/\varepsilon_s) \leq C_1(\log(\ell/\varepsilon))$ where C_0 and C_1 are some universal constants. Lemma 5.3 then guarantees that

$$\begin{aligned} &\Pr_{x\sim\mathbf{X}_1}\left[\left|\frac{|\mathsf{oNOSFSamp}_2(x) \cap B_2|}{D_2} - \frac{|B_2|}{\ell'}\right| \geq \varepsilon_a\right] \leq \varepsilon_s \\ \implies &\Pr_{x\sim\mathbf{X}_1}\left[\frac{|\mathsf{oNOSFSamp}_2(x) \cap B_2|}{D_2} \leq \varepsilon_a + \frac{|B_2|}{\ell'}\right] \geq 1 - \varepsilon_s \\ \implies &\Pr_{x\sim\mathbf{X}_1}\left[\left|\mathsf{oNOSFSamp}_2(x) \cap \overline{B_2}\right| \geq 3\gamma \cdot D_2\right] \geq 1 - \frac{\varepsilon}{3}. \end{aligned}$$

Let $\mathcal{C}_2 \subset [\ell']$ be the D_2 sized-committee of indices such sampled and let $\mathbf{Y}_2 = (\mathbf{X}_2)_{\mathsf{oNOSFSamp}_2(\mathbf{X}_1)}$ be the induced source by restricting $\mathbf{X}_2$ to indices from $\mathcal{C}_2$. The above then says that with at least a $1 - \frac{\varepsilon}{3}$ probability over $\mathbf{X}_1$, we have that $\mathbf{Y}_2$ is a $(g = 3\gamma D_2, D_2, n)$-oNOSF source and $\mathcal{C}_2$ contains $\geq 3\gamma D_2$ totally good indices.

Our second step is to use $\mathbf{Y}_2$ to sample from $\mathcal{C}_2$ and obtain a subsource over those indices from $\mathbf{X}_4$. We again do this by using $\mathsf{oNOSFSamp}_4 : (\{0,1\}^n)^{D_2} \to [D_2]^{D_4}$ from Lemma 5.3 with $\varepsilon_s = \varepsilon/3$, $\varepsilon_a = \gamma$, the corresponding γ equal to 3γ, and the set S of Lemma 5.3 corresponding to the indices $B_{\mathcal{C}_2} \subseteq \mathcal{C}_2$ of the weakly bad indices in $\mathcal{C}_2$ so that $\frac{|B_{\mathcal{C}_2}|}{D_2} \leq 1 - 3\gamma$. Here, $D_4 \leq C_3 \log(D_2/\varepsilon_s) \leq C_4(\log(\log(\ell)/\varepsilon))$ where C_3, C_4 are some universal constants. From Lemma 5.3, once again we are guaranteed that

$$\begin{aligned}
&\Pr_{y \sim \mathbf{Y}_2}\left[\left|\frac{|\mathsf{oNOSFSamp}_4(y) \cap B_{\mathcal{C}_2}|}{D_4} - \frac{|B_{\mathcal{C}_2}|}{D_2}\right| \geq \varepsilon_a\right] \leq \varepsilon_s \\
\implies &\Pr_{y \sim \mathbf{Y}_2}\left[\frac{|\mathsf{oNOSFSamp}_4(y) \cap B_{\mathcal{C}_2}|}{D_4} \leq \varepsilon_a + \frac{|B_{\mathcal{C}_2}|}{D_2}\right] \geq 1 - \varepsilon_s \\
\implies &\Pr_{y \sim \mathbf{Y}_2}\left[\left|\mathsf{oNOSFSamp}_4(y) \cap \overline{B_{\mathcal{C}_2}}\right| \geq 2\gamma \cdot D_4\right] \geq 1 - \frac{\varepsilon}{3}.
\end{aligned}$$

If we define $\mathbf{Y}_4 = (\mathbf{X}_4)_{\mathsf{oNOSFSamp}_4(\mathbf{Y}_2)}$, then the above guarantees that, with probability $1 - \frac{\varepsilon}{3}$ over $\mathbf{Y}_2$ (conditioned on $\mathcal{C}_2$ containing $\geq 3\gamma D_2$ totally good indices), we have that $\mathbf{Y}_4$ is a $(g = 2\gamma D_4, D_4, n)$-oNOSF source.

In our third and final step, we will use Lemma 5.4 to condense from $\mathbf{X}_3$ and $\mathbf{Y}_4$. We instantiate Lemma 5.4 with $n_x = \ell' n$, $n_y = \left(\frac{\log(\ell)}{\varepsilon}\right)^{C'} \log(n)$, $t = D_4 \leq C_4 \log(\log(\ell)/\varepsilon)$, $k = 4\gamma\ell' n$, and error equal to $\frac{\varepsilon}{3}$ where C' is a large enough universal constant. Given these parameters and our assumption that $n \geq (\log(\ell)/\varepsilon)^C$ for a universal constant C (that is large enough), as well as the fact that $\mathbf{Y}_4$ contains at least one good index, the requirements of Lemma 5.4 are satisfied. Consequently, we obtain the function $\mathsf{2Cond} : (\{0,1\}^n)^{\ell'} \times (\{0,1\}^n)^{D_4} \to \{0,1\}^{m_{\mathsf{2Cond}}}$. Finally the overall output of our explicit condenser is $\mathbf{Z} = \mathsf{2Cond}(\mathbf{X}_3, \mathbf{Y}_4)$.

We now analyze the guarantees of this condenser. We first observe that

$$\begin{aligned}
(C_{\mathsf{2Cond}})^t \log(3tn_x/\varepsilon) &\leq (C_{\mathsf{2Cond}})^{C_4(\log(\log(\ell)/\varepsilon))} \log(3 \cdot \ell n \cdot (\ell/\varepsilon)^{C'}/\varepsilon) \\
&\leq \left(\frac{\log(\ell)}{\varepsilon}\right)^{C_5} \log(n) \qquad (*)
\end{aligned}$$

where C_5 is a large enough universal constant. With this, we are guaranteed that the length of $\mathbf{Z} = m_z$ above is

$$\begin{aligned}
m_z &= \frac{1}{3}\left(k - (C_{\mathsf{2Cond}})^t \log(3tn_x/\varepsilon)\right) \\
&= \frac{1}{3}\left(\gamma\ell n - (C_{\mathsf{2Cond}})^t \log(3tn_x/\varepsilon)\right) \\
&\geq \frac{\gamma\ell n}{3} - \left(\frac{\log(\ell)}{\varepsilon}\right)^{C_5} \log(n) \qquad \text{(by Equation } (*)\text{)}
\end{aligned}$$

Moreover, conditioned on the fact that $\mathbf{Y}_4$ has at least one good block, Lemma 5.4 guarantees that

$$H_\infty^{\varepsilon/3}(\mathbf{Z}) = H_\infty^{\varepsilon/3}(\mathsf{2Cond}(\mathbf{X}_3, \mathbf{Y}_4))$$

$$\geq m_z - (C_{\mathsf{2Cond}})^t \log(3tn_x/\varepsilon)$$
$$\geq m_z - \left(\frac{\log(\ell)}{\varepsilon}\right)^{C_5} \log(n) \qquad \text{(by Equation } (*)\text{)}$$

Thus, adding up the three $\varepsilon/3$ errors from both our steps, we see that $H^{\varepsilon}_{\infty}(\mathbf{Z}) \geq m_z - \left(\frac{\ell}{\varepsilon}\right)^{C_5} \log(n)$.

We let our final universal constant C be much larger than C_5 to obtain our desired claim. □

5.4 Condensing from 67% good oNOSF sources with $n \geq \mathrm{polylog}(\ell)$

We now decrease not only our block length requirement all the way down to $\mathrm{polylog}(\ell)$ but also require that only 67% of the blocks are good. To do this, we apply the idea in the previous section twice; but this time at the end, we reuse one of the blocks we previously used to sample as a source. We show this by observing that sampling requires the usage of very few bits. By using the chain rule for min-entropy, we conclude that fixing those bits still leaves the source with lots of entropy.

Theorem 5.10. *There exists a universal constant C such that for any constant $\gamma > 0$ the following holds. For all $0 < \varepsilon < 1/2$ and $n, \ell \in \mathbb{N}$ where $n \geq \left(\frac{\log(\ell)}{\varepsilon}\right)^C$, there exists an explicit condenser $\mathsf{Cond} : (\{0,1\}^n)^\ell \to \{0,1\}^m$ satisfying: For any $(g = (2/3+\gamma)\ell, \ell, n)$-oNOSF source $\mathbf{X}$, we have that $H^{\varepsilon}_{\infty}(\mathsf{Cond}(\mathbf{X})) \geq m - \left(\frac{\log(\ell)}{\varepsilon}\right)^C \log(n)$ where $m = \frac{1}{3} \cdot \gamma\ell n$.*

Proof. Let $\ell' = \ell/3$. We decompose $\mathbf{X}$ into three parts as $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3)$, so each $\mathbf{X}_i$ is a $(g = 3\gamma\ell', \ell', n)$-oNOSF source. We in fact claim something stronger. Call an index $i \in [\ell']$ totally good if it is good in each of $\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3$. For the rest of the indices $i \in [\ell']$ that are not totally good, we refer to them as somewhat bad. Since $\mathbf{X}$ has $(2+3\gamma)\ell'$ good indices out of $3\ell'$, we see that there are must be at least $3\gamma\ell'$ totally good indices, i.e., indices that are good across each of the 3 blocks.

Our first step is to use $\mathbf{X}_1$ to sample a logarithmically sized committee of players from $\mathbf{X}_2$. We use $\mathsf{oNOSFSamp}_2 : (\{0,1\}^n)^{\ell'} \to [\ell']^{D_2}$ from Lemma 5.3 with $\varepsilon_s = \varepsilon/4$, $\varepsilon_a = \gamma$, the corresponding γ equal to 3γ, and the set S of Lemma 5.3 corresponding to the indices $B_2 \subset [\ell']$ of the somewhat bad indices in $\mathbf{X}_2$, so $\frac{|B_2|}{\ell'} \leq 1 - 3\gamma$. Note that $D_2 \leq C_0 \log(\ell/\varepsilon_s) \leq C_1(\log(\ell/\varepsilon))$ where C_0 and C_1 are some universal constants. Lemma 5.3 then guarantees that

$$\Pr_{x\sim\mathbf{X}_1}\left[\left|\frac{|\mathsf{oNOSFSamp}_2(x) \cap B_2|}{D_2} - \frac{|B_2|}{\ell'}\right| \geq \varepsilon_a\right] \leq \varepsilon_s$$
$$\implies \Pr_{x\sim\mathbf{X}_1}\left[\frac{|\mathsf{oNOSFSamp}_2(x) \cap B_2|}{D_2} \leq \varepsilon_a + \frac{|B_2|}{\ell'}\right] \geq 1 - \varepsilon_s$$
$$\implies \Pr_{x\sim\mathbf{X}_1}\left[\left|\mathsf{oNOSFSamp}_2(x) \cap \overline{B_2}\right| \geq 2\gamma \cdot D_2\right] \geq 1 - \frac{\varepsilon}{4}.$$

Let $\mathcal{C}_2 \subset [\ell']$ be the $D_2 \leq C_1(\log(\ell/\varepsilon))$ sized-committee of indices thus sampled and let $\mathbf{Y}_2 = (\mathbf{X}_2)_{\mathsf{oNOSFSamp}_2(\mathbf{X}_1)}$ be the source obtained by restricting $\mathbf{X}_2$ to indices from $\mathcal{C}_2$. However, when we do this, instead of each player in $\mathbf{Y}_2$ holding n bits, we take a prefix of length $n_2 = C_2 \cdot \log(\log(\ell)/\varepsilon) \cdot \log(1/\varepsilon)$ from each where C_2 is a sufficiently large universal constant. The above then says that with at least a $1 - \frac{\varepsilon}{4}$ probability over $\mathbf{X}_1$, we have that $\mathbf{Y}_2$ is a $(g = 2\gamma D_2, D_2, n_2)$-oNOSF source and $\mathcal{C}_2$ contains $\geq 2\gamma D_2$ totally good indices.

Our second step is to use $\mathbf{Y}_2$ to sample from $\mathcal{C}_2$ and obtain a subsource over those indices from $\mathbf{X}_3$. We again do this by using $\mathsf{oNOSFSamp}_3 : (\{0,1\}^n)^{D_2} \to [D_2]^{D_3}$ from Lemma 5.3 with $\varepsilon_s = \varepsilon/4$, $\varepsilon_a = \gamma$,

the corresponding γ equal to 2γ, and the set S of Lemma 5.3 corresponding to the indices $B_{\mathcal{C}_2} \subseteq \mathcal{C}_2$ of the weakly bad indices in $\mathcal{C}_2$ so that $\frac{|B_{\mathcal{C}_2}|}{D_2} \leq 1 - 2\gamma$. Here, $D_3 \leq C_3 \log(D_2/\varepsilon_s) \leq C_4(\log(\log(\ell)/\varepsilon))$ where C_3, C_4 are some universal constants. From Lemma 5.3, once again we are guaranteed that

$$\begin{aligned}
&\Pr_{y\sim\mathbf{Y}_2}\left[\left|\frac{|\mathsf{oNOSFSamp}_3(y)\cap B_{\mathcal{C}_2}|}{D_3} - \frac{|B_{\mathcal{C}_2}|}{D_2}\right| \geq \varepsilon_a\right] \leq \varepsilon_s \\
\implies &\Pr_{y\sim\mathbf{Y}_2}\left[\frac{|\mathsf{oNOSFSamp}_3(y)\cap B_{\mathcal{C}_2}|}{D_3} \leq \varepsilon_a + \frac{|B_{\mathcal{C}_2}|}{D_2}\right] \geq 1 - \varepsilon_s \\
\implies &\Pr_{y\sim\mathbf{Y}_2}\left[\left|\mathsf{oNOSFSamp}_3(y)\cap \overline{B_{\mathcal{C}_2}}\right| \geq \gamma\cdot D_3\right] \geq 1 - \frac{\varepsilon}{4}.
\end{aligned}$$

If we define $\mathbf{Y}_3 = (\mathbf{X}_3)_{\mathsf{oNOSFSamp}_3(\mathbf{Y}_2)}$, then the above guarantees that, with probability $1 - \frac{\varepsilon}{4}$ over $\mathbf{Y}_2$ (conditioned on $\mathcal{C}_2$ containing $\geq 2\gamma D_2$ totally good indices), we have that $\mathbf{Y}_3$ is a $(g = \gamma D_3, D_3, n)$-oNOSF source.

We will show that $\mathbf{X}_2$ has entropy conditioned on most fixings of $\mathbf{Y}_2$. Recall that $\mathbf{X}_2$ is a $(g = 3\gamma\ell', \ell', n)$-oNOSF source. We use the min-entropy chain rule (Lemma 4.3) to conclude that with probability $1 - \varepsilon/4$ over $y \sim \mathbf{Y}_2$, we have that

$$\begin{aligned}
H_\infty(\mathbf{X}_2|(\mathbf{Y}_2 = y)) &\geq 3\gamma\ell' n - \log(4/\varepsilon) - D_2\cdot n_2 \\
&\geq 3\gamma\ell' n - \log(4/\varepsilon) - C_1\log(\ell/\varepsilon)\cdot C_2\log(\log(\ell)/\varepsilon)\cdot\log(1/\varepsilon) \\
&\geq 3\gamma\ell' n - C_5\left(\log(\ell/\varepsilon)\right)^3
\end{aligned}$$

where C_5 is a large enough universal constant.

With this, we apply union bound to conclude that conditioned on $\mathcal{C}_2$ containing $\geq 2\gamma D_2$ totally good indices, with probability $1 - \varepsilon/2$ over $y \sim \mathbf{Y}_2$, we have that $\mathbf{Y}_3$ is a $(g = \gamma D_3, D_3, n)$-oNOSF source and $H_\infty(\mathbf{X}_2) \geq \gamma\ell n - C_5\left(\log(\ell/\varepsilon)\right)^3$. We refer to such a fixing of $\mathbf{Y}_2 = y_2$ as 'good.'

In our third and final step, we use Lemma 5.4 to condense from $\mathbf{X}_2$ and $\mathbf{Y}_3$. We instantiate Lemma 5.4 with $n_x = \ell' n$, $n_y = \left(\frac{\log(\ell)}{\varepsilon}\right)^{C'}\log(n)$, $t = D_3 \leq C_6\log(\log(\ell)/\varepsilon)$, $k = 3\gamma\ell' n - C_5\left(\log(\ell/\varepsilon)\right)^3$, and error equal to $\frac{\varepsilon}{4}$ where C' is a large enough universal constant. Using our assumption that $n \geq (\log(\ell)/\varepsilon)^C$ for a universal constant C (that is large enough), these parameters satisfy the requirements of Lemma 5.4,, and the lemma gives us the explicit condenser $\mathsf{2Cond} : (\{0,1\}^n)^{\ell'} \times (\{0,1\}^{n_y})^{D_3} \to \{0,1\}^{m_{\mathsf{2Cond}}}$. Let $\mathbf{Z} = \mathsf{2Cond}\left(\mathbf{X}_2, \mathbf{Y}_3\right)$ and let this be the final output of our own condenser.

We now analyze the guarantees of this condenser. We first observe that

$$\begin{aligned}
(C_{\mathsf{2Cond}})^t\log(4tn_x/\varepsilon) &\leq (C_{\mathsf{2Cond}})^{C_6(\log(\log(\ell)/\varepsilon))}\log(4\cdot C_6(\log(\log(\ell)/\varepsilon))\cdot\ell n/\varepsilon) \\
&\leq \left(\frac{\log(\ell)}{\varepsilon}\right)^{C_7}\log(n) \qquad (*)
\end{aligned}$$

where C_7 is a large enough universal constant.

Let m_z be the length of the source $\mathbf{Z}$. With this, we are guaranteed from Lemma 5.4 the following lower bound on m_z:

$$\begin{aligned}
m_z &= \frac{1}{3}\left(k - (C_{\mathsf{2Cond}})^t\log(4tn_x/\varepsilon)\right) \\
&= \frac{1}{3}\left(\gamma\ell n - C_5(\log(\ell/\varepsilon))^3 - (C_{\mathsf{2Cond}})^t\log(4tn_x/\varepsilon)\right)
\end{aligned}$$

$$\geq \frac{1}{3}\left(\gamma\ell n - C_5(\log(\ell/\varepsilon))^3 - \left(\frac{\log(\ell)}{\varepsilon}\right)^{C_7}\log(n)\right) \qquad \text{(by Equation } (*)\text{)}$$
$$\geq \frac{\gamma\ell n}{3} - \left(\frac{\log(\ell)}{\varepsilon}\right)^{C_8}\log(n)$$

where C_8 is a large enough universal constant.

We condition on both $\mathcal{C}_2$ containing $2\gamma D_2$ totally good indices and a 'good' fixing of $\mathbf{Y}_2$. Under this conditioning, Lemma 5.4 guarantees that

$$\begin{aligned} H_\infty^{\varepsilon/4}(\mathbf{Z}) &= H_\infty^{\varepsilon/4}(2\mathsf{Cond}(\mathbf{X}_2, \mathbf{Y}_3)) \\ &\geq m_z - (C_{2\mathsf{Cond}})^t \log(4tn_x/\varepsilon) \\ &\geq m_z - \left(\frac{\log(\ell)}{\varepsilon}\right)^{C_7}\log(n) \qquad \text{(by Equation } (*)\text{)} \\ &\geq m_z - \left(\frac{\log(\ell)}{\varepsilon}\right)^{C_8}\log(n). \end{aligned}$$

Thus, adding up the four $\varepsilon/4$ errors from both our steps, we see that $H_\infty^{\varepsilon}(\mathbf{Z}) \geq m_z - \left(\frac{\log(\ell)}{\varepsilon}\right)^{C_8}\log(n)$.

We let our final universal constant C be much larger than C_8 to obtain our final claim. □

5.5 Condensing from 51% good oNOSF sources with $n \geq \mathrm{polylog}(\ell)$

We now decrease not only our block length requirement all the way down to $\mathrm{polylog}(\ell)$ but also require that only 51% of the blocks are good. To do this, we build upon the previous ideas with one more 'self-sampling' idea - where we sample from within the blocks in the same source. This introduces correlations between the bits that are being used to sample and the source itself, in a way that makes us lose the structure of our sources. Nevertheless we rely on the fact that very few bits are required to do sampling, and that sampling succeeds regardless of the behavior of the bad players. To handle this situation, we use Lemma 7.9 that states if an adversary is allowed to arbitrarily control few bits of the source (that were previously uniform), then the damage they can do is not too much.

Proof of Theorem 5.1. Let $\ell' = \ell/2$. We begin by decomposing $\mathbf{X}$ into two parts as $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2)$, so each $\mathbf{X}_i$ is a $(g = 2\gamma\ell', \ell', n)$-oNOSF source. We in fact claim something stronger. Call an index $i \in [\ell']$ totally good if it is good in both of $\mathbf{X}_1, \mathbf{X}_2$. For the rest of the indices $i \in [\ell']$ that are not totally good, we refer to them as somewhat bad. Since $\mathbf{X}$ has $(1+2\gamma)\ell'$ good indices out of $2\ell'$, we see that there are must be at least $2\gamma\ell'$ totally good indices, i.e., indices that are good across both the blocks.

Let $\mathbf{X}_1'$ be the subsource obtained from $\mathbf{X}_1$ by taking prefixes of all blocks of length $n_1' = C_1\log(\ell)\log(1/\varepsilon)$ where C_1 is a large enough universal constant. Hence, $\mathbf{X}_1'$ is a $(g = 2\gamma\ell', \ell', n_1')$-oNOSF source.

Our first step is to use $\mathbf{X}_1'$ to sample a logarithmically sized committee of players from $\mathbf{X}_2$. We obtain $\mathsf{oNOSFSamp}_{1\to 2} : (\{0,1\}^{n_1'})^{\ell'} \to [\ell']^{D_2}$ from Lemma 5.3 with $\varepsilon_s = \varepsilon/4$, $\varepsilon_a = \gamma/2$, the corresponding γ equal to 2γ, and the set S of Lemma 5.3 corresponding to the indices $B_2 \subset [\ell']$ of the somewhat bad indices in $\mathbf{X}_2$, so $\frac{|B_2|}{\ell'} \leq 1 - 2\gamma$. Note that

$$D_2 \leq C_0\log(\ell/\varepsilon_s) \leq C_1(\log(\ell/\varepsilon)) \tag{1}$$

where C_0 and C_1 are some universal constants. Lemma 5.3 then guarantees that

$$\begin{aligned}
&\Pr_{x\sim\mathbf{X}'_1}\left[\left|\frac{|\mathsf{oNOSFSamp}_{1\to 2}(x)\cap B_2|}{D_2}-\frac{|B_2|}{\ell'}\right|\geq\varepsilon_a\right]\leq\varepsilon_s\\
\implies &\Pr_{x\sim\mathbf{X}'_1}\left[\frac{|\mathsf{oNOSFSamp}_{1\to 2}(x)\cap B_2|}{D_2}\leq\varepsilon_a+\frac{|B_2|}{\ell'}\right]\geq 1-\varepsilon_s\\
\implies &\Pr_{x\sim\mathbf{X}'_1}\left[\left|\mathsf{oNOSFSamp}_{1\to 2}(x)\cap\overline{B_2}\right|\geq(3\gamma/2)\cdot D_2\right]\geq 1-\frac{\varepsilon}{4}.
\end{aligned}$$

Let $\mathcal{C}_2\subset[\ell']$ be the $D_2\leq C_1(\log(\ell/\varepsilon))$ sized committee of indices thus sampled and let $\mathbf{Y}_2=(\mathbf{X}_2)_{\mathsf{oNOSFSamp}_2(\mathbf{X}_1)}$ be the source obtained by restricting $\mathbf{X}_2$ to indices from $\mathcal{C}_2$. Let $E_{1\to 2}$ be the event that the sampler $\mathsf{oNOSFSamp}_{1\to 2}$ above succeeds. We have that $\Pr[E_{1\to 2}]\geq 1-\varepsilon/4$ with the probability being over sampling from $\mathbf{X}'_1$. We see that when $E_{1\to 2}$ occurs, $\mathbf{Y}_2$ will be an $(g=(3\gamma/2)D_2,D_2,n)$-oNOSF source.

We will show that $\mathbf{X}_1$ has entropy conditioned on most fixings of $\mathbf{X}'_1$. We use the min-entropy chain rule (Lemma 4.3) to conclude that with probability $1-\varepsilon/4$ over $x\sim\mathbf{X}'_1$, we have that

$$\begin{aligned}
H_\infty(\mathbf{X}_1|(\mathbf{X}'_1=x))&\geq 2\gamma\ell' n-\ell' n'_1-\log(4/\varepsilon)\\
&=2\gamma\ell' n-\ell'\cdot C_1\log(\ell)\log(1/\varepsilon)-\log(4/\varepsilon)\\
&\geq 2\gamma\ell' n-C_3\ell'\log(\ell')\log(1/\varepsilon)
\end{aligned}$$

where C_3 is a large enough universal constant. Let

$$k_1=2\gamma\ell' n-C_3\ell'\log(\ell')\log(1/\varepsilon)\tag{2}$$

Let E_1 be the event that $x\sim\mathbf{X}'_1$ is such that $H_\infty(\mathbf{X}_1|(\mathbf{X}'_1=x_1))\geq k_1$. Then, we have that $\Pr[E_1]\geq 1-(\varepsilon/4)$ with the probability being over sampling from $\mathbf{X}'_1$.

By a union bound, we have that both E_1 and $E_{1\to 2}$ happen together with probability at least $1-\varepsilon/2$. Note that conditioning on both E_1 and $E_{1\to 2}$, the online structure of the source still remains intact, i.e., $\mathbf{Y}_2$ still remains an oNOSF source and the good bits in $\mathbf{Y}_2$ still are independent of $\mathbf{X}_1$. So, for instance we also satisfy the required independence conditions of Lemma 5.13 and if we also satisfied the parameter conditions for it, we could apply it. However, we do not satisfy the parameter conditions since our guarantees on n are too small. To remedy this, we will use a subsource of $\mathbf{Y}_2$ to sample from within itself. Doing so will shrink our source and let us satisfy the parameter conditions from Lemma 5.13. However, we will then no longer satisfy the independence requirements to apply it. Nevertheless we do this anyways and argue that we can so, while only sacrificing the final guarantees of the condenser by a tiny amount.

Let $\mathbf{Y}_{2,\mathsf{Samp}}$ be the subsource obtained from $\mathbf{Y}_2$ by taking prefixes of all blocks of length

$$n'_2=C_2\cdot\log(\log(\ell)/\varepsilon)\cdot\log(1/\varepsilon)\tag{3}$$

where C_2 is a large enough universal constant. So, when $E_{1\to 2}$ occurs, we have that $\mathbf{Y}_{2,\mathsf{Samp}}$ is $(g=(3\gamma/2)D_2,D_2,n'_2)$-oNOSF source

In the second step, we will use $\mathbf{Y}_{2,\mathsf{Samp}}$ to sample from $\mathcal{C}_2$ and obtain a subsource over those indices from $\mathbf{Y}_2$. We again do this by using $\mathsf{oNOSFSamp}_{2\to 2}:(\{0,1\}^n)^{D_2}\to[D_2]^{D_{2,\mathsf{Cond}}}$ from Lemma 5.3 with $\varepsilon_s=\varepsilon/4$, $\varepsilon_a=\gamma/2$, the corresponding γ equal to $3\gamma/2$, and the set S of Lemma 5.3 corresponding to the indices $B_{\mathcal{C}_2}\subseteq\mathcal{C}_2$ of the weakly bad indices in $\mathcal{C}_2$ so that $\frac{|B_{\mathcal{C}_2}|}{D_2}\leq 1-(3\gamma/2)$. Here, $D_{2,\mathsf{Cond}}\leq$

$C_4 \log(D_2/\varepsilon_s) \le C_5(\log(\log(\ell)/\varepsilon))$ where C_4, C_5 are some universal constants. From Lemma 5.3, once again we are guaranteed that

$$\begin{aligned}
&\Pr_{y\sim \mathbf{Y}_{2,\mathsf{Samp}}}\left[\left|\frac{|\mathsf{oNOSFSamp}_{2\to 2}(y)\cap B_{\mathcal{C}_2}|}{D_{2,\mathsf{Cond}}} - \frac{|B_{\mathcal{C}_2}|}{D_2}\right| \ge \varepsilon_a\right] \le \varepsilon_s \\
\implies &\Pr_{y\sim \mathbf{Y}_{2,\mathsf{Samp}}}\left[\frac{|\mathsf{oNOSFSamp}_{2\to 2}(y)\cap B_{\mathcal{C}_2}|}{D_{2,\mathsf{Cond}}} \le \varepsilon_a + \frac{|B_{\mathcal{C}_2}|}{D_2}\right] \ge 1-\varepsilon_s \\
\implies &\Pr_{y\sim \mathbf{Y}_{2,\mathsf{Samp}}}\left[\left|\mathsf{oNOSFSamp}_{2\to 2}(y)\cap \overline{B_{\mathcal{C}_2}}\right| \ge \gamma\cdot D_{2,\mathsf{Cond}}\right] \ge 1-\frac{\varepsilon}{4}. \qquad (**)
\end{aligned}$$

For $y_{2,\mathsf{Samp}} \in (\{0,1\}^n)^{D_2}$, let $\mathcal{C}_{2,\mathsf{Cond}}(y_{2,\mathsf{Samp}}) = \mathsf{oNOSFSamp}_{2\to 2}(y_{2,\mathsf{Samp}})$ so that the number of players in the resultant committee is $|\mathcal{C}_{2,\mathsf{Cond}}(y_{2,\mathsf{Samp}})| = D_{2,\mathsf{Cond}} \le C_5(\log(\log(\ell)/\varepsilon))$. We let $\mathbf{Y}_{2,\mathsf{Cond}}(y_{2,\mathsf{Samp}})$ be the subsource obtained from $\mathbf{Y}_2$ by taking suffix of all blocks of length $n-n_2'$ where n_2' is as above. Then Equation $(**)$ guarantees that, with probability $1-\frac{\varepsilon}{4}$ over sampling $y_{2,\mathsf{Samp}} \sim \mathbf{Y}_{2,\mathsf{Samp}}$ (conditioned on E_1 and $E_{1\to 2}$), we have that $\mathbf{Y}_{2,\mathsf{Cond}}(y_{2,\mathsf{Samp}})$ is a $(g = \gamma D_{2,\mathsf{Cond}}, D_{2,\mathsf{Cond}}, n)$-oNOSF source. We refer to such a $y_{2,\mathsf{Samp}}$ as 'good.'

In our third and final step, we use Lemma 5.4 to condense from $\mathbf{X}_1$ and $\mathbf{Y}_{2,\mathsf{Cond}}(y_{2,\mathsf{Samp}})$. We condition on events E_1 and $E_{1\to 2}$ here. We also pay additional $\varepsilon/4$ in error and assume that all $y_{2,\mathsf{Samp}}$ are good, i.e., the sampler always succeeds. This brings the total error we have incurred so far to $3\varepsilon/4$. Note that since we used $y_{2,\mathsf{Samp}} \sim \mathbf{Y}_{2,\mathsf{Samp}}$ to obtain $\mathbf{Y}_{2,\mathsf{Cond}}$, any fixing of the output of $\mathbf{Y}_{2,\mathsf{Samp}} = y_{2,\mathsf{Samp}}$ can create correlations between $\mathbf{X}_1$ and $\mathbf{Y}_{2,\mathsf{Cond}}$ and it may not even preserve the structure of $\mathbf{Y}_{2,\mathsf{Cond}}$. Formally for any fixed $y_{2,\mathsf{Samp}}$, conditioned on $\mathbf{Y}_{2,\mathsf{Samp}} = y_{2,\mathsf{Samp}}$, 1) it is not necessarily true that $\mathbf{Y}_{2,\mathsf{Cond}}$ still remains an oNOSF source, and 2) the good bits in $\mathbf{Y}_{2,\mathsf{Cond}}$ may not necessarily be independent of $\mathbf{X}_1$. We address these concerns by using Lemma 7.9 and paying with more error and more entropy gap at the end.

Let Opt (short for optimistic) be the assumption that all the bits (including the bad ones) in $\mathbf{Y}_{2,\mathsf{Samp}}$ were truly uniform and independent of $\mathbf{X}_1$ and independent of all length $n-n_2'$ suffices of the bits of good players in $\mathbf{X}_2$ (these bits in the suffixes are the ones that potentially can be used to form $\mathbf{Y}_{2,\mathsf{Cond}}$ above). We use this to assume we do meet the preconditions of Lemma 5.4. Let Actual be the realistic scenario where the above does not happen and $\mathbf{Y}_{2,\mathsf{Samp}}$ is allowed to have bad bits.

Let $\varepsilon_{\mathsf{Cond}} = 2^{-C_6(\log(\ell/\varepsilon))^3}$ where C_6 is a large universal constant. We then instantiate Lemma 5.4 with $n_x = \ell' n$, $n_y = \left(\frac{\log(\ell)}{\varepsilon}\right)^{C'} \log(n)$, $t = D_3 \le C_6 \log(\log(\ell)/\varepsilon)$, $k = k_1$ (from Equation (2)), and error equal to $\varepsilon_{\mathsf{Cond}}$ where C' is a large enough universal constant. Given these parameters and our assumption that $n \ge (\log(\ell)/\varepsilon)^C$ for a universal constant C (that is large enough), we indeed satisfy the requirements of Lemma 5.4 under Opt. Consequently, we obtain the function $\mathsf{2Cond} : (\{0,1\}^n)^{\ell'} \times (\{0,1\}^n)^{D_{2,\mathsf{Cond}}} \to \{0,1\}^{m_{\mathsf{2Cond}}}$ and our final output will be $\mathsf{2Cond}(\mathbf{X}_1, \mathbf{Y}_{2,\mathsf{Cond}})$.

Let $\mathbf{Z}_{\mathsf{Opt}} = \mathsf{2Cond}(\mathbf{X}_1, \mathbf{Y}_{2,\mathsf{Cond}})$ be the distribution under the assumption Opt. Let $\mathbf{Z}_{\mathsf{Actual}}$ be the actual output distribution that we obtain. i.e., $\mathbf{Z}_{\mathsf{Actual}} := \mathsf{Cond}(\mathbf{X})$, the output distribution of our condenser.

We first analyze the guarantees of this condenser under the assumption Opt. We first observe that

$$\begin{aligned}
(C_{\mathsf{2Cond}})^t \log(tn_x/\varepsilon_{\mathsf{Cond}}) &\le (C_{\mathsf{2Cond}})^{C_6(\log(\log(\ell)/\varepsilon))} \log(C_6(\log(\log(\ell)/\varepsilon))\cdot \ell n/\varepsilon_{\mathsf{Cond}}) \\
&= (C_{\mathsf{2Cond}})^{C_6(\log(\log(\ell)/\varepsilon))} \log(C_6(\log(\log(\ell)/\varepsilon))\cdot \ell n \cdot 2^{C_6(\log(\ell/\varepsilon))^3}) \\
&\le \left(\frac{\log(\ell)}{\varepsilon}\right)^{C_7} \log(n) \qquad (*)
\end{aligned}$$

where C_7 is a large enough universal constant.

Since $n \geq (\log(\ell)/\varepsilon)^C$ for a very large universal constant C, our parameter setting indeed satisfies all the requirements of Lemma 5.4 (and the independence based requirements hold because we are arguing under the assumption $\mathbf{Z}_{\mathsf{Opt}}$.

With this, Lemma 5.4 provides us with the following guarantee on the length m_z of our final output (recall that k_1 from Equation (2) is the entropy of $\mathbf{X}_1$ conditioned on event E_1 occurring):

$$\begin{aligned}
m_z &= \frac{1}{3}\left(k_1 - (C_{\mathsf{2Cond}})^t \log(tn_x/\varepsilon_{\mathsf{Cond}})\right) \\
&= \frac{1}{3}\left(2\gamma\ell' n - C_3\ell' \log(\ell')\log(1/\varepsilon) - (C_{\mathsf{2Cond}})^t \log(tn_x/\varepsilon_{\mathsf{Cond}})\right) && \text{(by Equation (2))} \\
&\geq \frac{1}{3}\left(2\gamma\ell' n - C_3\ell' \log(\ell')\log(1/\varepsilon) - \left(\frac{\log(\ell)}{\varepsilon}\right)^{C_7}\log(n)\right) && \text{(by Equation } (*)\text{)} \\
&= \frac{1}{3}\left(\gamma\ell n - C_3(\ell/2)\log(\ell/2)\log(1/\varepsilon) - \left(\frac{\log(\ell)}{\varepsilon}\right)^{C_7}\log(n)\right) && \text{(by definition of } \ell'\text{)} \\
&\geq \frac{\gamma\ell n}{3} - C_8\left(\ell\log(\ell)\log(1/\varepsilon)\right) - \left(\frac{\log(\ell)}{\varepsilon}\right)^{C_7}\log(n)
\end{aligned}$$

where C_8 is some large enough universal constant.

Then, Lemma 5.4 (under assumption Opt) guarantees that

$$\begin{aligned}
H_\infty^{\varepsilon_{\mathsf{Cond}}}(\mathbf{Z}_{\mathsf{Opt}}) &\geq m_z - (C_{\mathsf{2Cond}})^t \log(tn_x/\varepsilon_{\mathsf{Cond}}) \\
&\geq m_z - \left(\frac{\log(\ell)}{\varepsilon}\right)^{C_7}\log(n) && \text{(by Equation } (*)\text{)}
\end{aligned}$$

We are not yet done since in reality, the assumption Opt does not hold and we need to argue under Actual. To handle this situation, instead of using $\mathbf{Y}_{2,\mathsf{Samp}}$ above, we consider a distribution $\mathbf{X}_{\mathsf{Adv}} \sim (\{0,1\}^n)^\ell$ which is same as the distribution $\mathbf{X}$ but the $n_2' \cdot D_2$ many bits in $\mathbf{Y}_{2,\mathsf{Samp}}$ are instead controlled by an adversary ; we allow those bits to depend on any other bits from $\mathbf{X}$. Let $\mathbf{Z}_{\mathsf{Adv}}$ be the resulting output distribution when we do this. Since the adversary Adv is arbitrary, this adversarial assumption is stronger than Actual scenario, so it suffices to argue about $\mathbf{Z}_{\mathsf{Adv}}$. To argue regarding $\mathbf{Z}_{\mathsf{Adv}}$, we apply Lemma 7.9 to infer that

$$\begin{aligned}
H_\infty^{\varepsilon_{\mathsf{Cond}}\cdot 2^{n_2'\cdot D_2}}(\mathbf{Z}_{\mathsf{Adv}}) &\geq H_\infty^{\varepsilon_{\mathsf{Cond}}}(\mathbf{Z}_{\mathsf{Opt}}) - n_2' \cdot D_2 \\
&\geq m_z - \left(\frac{\log(\ell)}{\varepsilon}\right)^{C_7}\log(n) - \left(C_2 \cdot \log(\log(\ell)/\varepsilon)\cdot\log(1/\varepsilon)\right)\cdot\left(C_1\log(\ell/\varepsilon)\right) \\
&\quad \text{(by Equation (3) and Equation (1))} \\
&\geq m_z - \left(\frac{\log(\ell)}{\varepsilon}\right)^{C_9}\log(n)
\end{aligned}$$

where C_9 is a large enough universal constant. We also see that

$$\begin{aligned}
\varepsilon_{\mathsf{Cond}} \cdot 2^{n_2'\cdot D_2} &= 2^{-C_6(\log(\ell/\varepsilon))^3} \cdot 2^{(C_2\cdot\log(\log(\ell)/\varepsilon)\cdot\log(1/\varepsilon))\cdot(C_1\log(\ell/\varepsilon))} && \text{(by Equation (3) and Equation (1))} \\
&\leq \varepsilon/4
\end{aligned}$$

The last inequality follows since we will pick C_6 to be much larger than C_1 and C_2.

We note that $\mathbf{X}$, even after conditioning on $E_1, E_{1\to 2}$ can be assumed to be a flat source. This is because we can express $\mathbf{X}_1$ as a convex combination of sources with the same min-entropy and doing so retains the structure of our source $\mathbf{X}$. Hence, we do meet all the preconditions of Lemma 7.9.

Lastly, since we incurred $3\varepsilon/4$ error at the beginning of the fourth step, and we incur additional $\varepsilon/4$ here, we obtain that our final output distribution will be ε-close to having min-entropy $m_z - \left(\frac{\log(\ell)}{\varepsilon}\right)^{C_9}\log(n)$. By taking C to be a large enough constant for our actual claimed parameters, we infer the claim.

□

5.6 Constructing oNOSFSamp

In this subsection, we will construct oNOSFSamp and prove Lemma 5.3. Our construction of oNOSFSamp itself requires two ingredients: (1) A Reduce function that reduces an oNOSF source of length ℓ to an $O(\log(\ell))$ length source, and (2) a good averaging sampler with linear sample complexity from Lemma 4.11.

Let's formally define this Reduce function:

Lemma 5.11. *There exists a universal constant C such that the following holds. For all $0 < \gamma \le 1, 0 < \varepsilon < 1/2$ and all $\ell, n \in \mathbb{N}$ such that $n \ge 6\log(\ell)\log(1/\varepsilon)/\gamma$, there exists an explicit function $\mathsf{Reduce} : (\{0,1\}^n)^\ell \to \{0,1\}^t$ such that for all $(\gamma\ell, \ell)$-oNOSF sources $\mathbf{X}$, we have that $\mathsf{Reduce}(\mathbf{X})$ is a (t, k)-source where $t \le C\log(\ell/\varepsilon)$ and $k \ge 3\log(\ell/\varepsilon)$.*

We construct this function in Section 5.6.1. Let's see how using it we can construct oNOSFSamp.

Proof of Lemma 5.3. We use the given parameters γ, n, ℓ to instantiate $\mathsf{Reduce} : [N]^\ell \to \{0,1\}^t$ from Lemma 5.11 with $\varepsilon = \varepsilon_s$. This gives us a constant C_0 such that $\mathsf{Reduce}(\mathbf{X})$ is a $(t \le C_0\log(\ell/\varepsilon_s), k \ge 6\log(\ell/\varepsilon_s))$-source. We then let $\delta = \frac{k}{2t} = \frac{3}{C_0}$ and $\alpha = \frac{1}{3}$ which we use to instantiate Lemma 4.11 with $\varepsilon = \varepsilon_a$ to get $\mathsf{Samp} : \{0,1\}^t \to (\{0,1\}^m)^D$.

Define $\mathsf{oNOSFSamp}(\mathbf{X}) = \mathsf{Samp}(\mathsf{Reduce}(X)) : (\{0,1\}^n)^\ell \to (\{0,1\}^m)^D$. Since $(1-\alpha)\delta t = \frac{2}{3}\cdot\frac{3}{2}\log(\ell/\varepsilon_s) \ge \log(\ell)$, Lemma 4.11 allows us to take $m = \log(\ell)$. Moreover, because $k - \delta t = k/2 \ge 3\log(\ell/\varepsilon_s) \ge \log(1/\varepsilon_s)$, we have that $2^{\delta t - k} \le \varepsilon_s$, giving us the desired error bound. Finally, Lemma 4.11 also gives us that $D = O(t) = O(\log(\ell/\varepsilon_s))$, as claimed. □

5.6.1 Constructing Reduce

We here construct Reduce function as required by Lemma 5.11. Our construction is based on the construction from [RZ01] that utilizes hitting sets for combinatorial rectangles. We call their general constructed function as Reduce'.

This function Reduce' has the following guarantee:

Lemma 5.12 ([RZ01])**.** *There exists a universal constant C' such that for any $\gamma > 0$ and $a, d \in \mathbb{N}$, there exists an efficient function $\mathsf{Reduce}' : [a]^d \to \{0,1\}^t$ such that for any $(g = \gamma d, d, \log(a))$-oNOSF source $\mathbf{X}$, we have that $\mathsf{Reduce}'(\mathbf{X})$ is a (t, k)-source with $t \le C'(\log(a) + \log\log(d) + d/a)$ and $k \ge \gamma d/a$.*

We construct our desired function Reduce' in Appendix A. Let's see first how by carefully choosing a and d in Lemma 5.12, we get the Reduce function we require.

Proof of Lemma 5.11. We will consider two cases for the parameters of our oNOSF source and apply Lemma 5.12 with different parameters in each case. Recall that, in Lemma 5.12, d represents the number of blocks in our oNOSF source and $\log(a)$ represents the number of bits in each block. However, our given $(g = \gamma\ell, \ell, n)$-oNOSF source $\mathbf{X}$ is on $[N]^\ell$, so we must make these parameters match. We take cases on the relative size of $\log(\ell)$ and $\log(1/\varepsilon)$.

Case 1 First, if $\log(1/\varepsilon) \leq \log(\ell)$, then let $d = \ell$ and $a = \frac{\gamma\ell}{6\log(\ell)}$ in Lemma 5.12. In this setting, since we are guaranteed that $n \geq 6\log(\ell)\log(1/\varepsilon)/\gamma$, so $N \geq \ell > a$, we can simply truncate each block to $\log(a)$ bits and take $\mathbf{X}$ to be a $(g = \gamma\ell, \ell)$-oNOSF source on $[a]^d$. Lemma 5.12 gives us that there exists some C' such that

$$\begin{aligned} t &\leq C'(\log(a) + \log\log(d) + d/a) \\ &= C'(\log(\gamma) + \log(\ell) - \log(6) - \log\log(\ell) + \log\log(\ell) + 6\log(\ell)/\gamma) \\ &\leq C_2 \log(\ell) \\ &\leq C_2 \log(\ell/\varepsilon) \end{aligned}$$

for some constant C_2. Then, we compute the min-entropy of $\mathsf{Reduce}(\mathbf{X})$ as

$$\begin{aligned} k &\geq \gamma d/a \\ &= \gamma\ell \cdot 6\log(\ell)/(\gamma\ell)) \\ &= 6\log(\ell). \end{aligned}$$

Recall that the assumption in this case is that $\log(1/\varepsilon) \leq \log(\ell)$, which we can rearrange into $\log(\ell) \geq \log(\ell/\varepsilon)/2$. Applying this yields that $k \geq 6\log(\ell) \geq 3\log(\ell/\varepsilon)$, as desired.

Case 2 Second, if $\log(1/\varepsilon) > \log(\ell)$, then let $d = 6\ell\log(1/\varepsilon)/\gamma$ and $a = \ell$ in Lemma 5.12. In order to convert $\mathbf{X}$ to a source over $[a]^d$, we split each n length block of $\mathbf{X}$ into length $n' = \frac{\gamma n}{6\log(1/\varepsilon)}$ blocks. This gives us $\ell' = \ell\log(1/\varepsilon) \cdot \frac{6}{\gamma} = d$ total blocks with $g' = \gamma\ell'$ total good blocks. Thus, we now view $\mathbf{X}$ as a source $\mathbf{X}'$ over $[N']^{\ell'}$ where $N' = 2^{n'}$. To finish the conversion, we recall that $n \geq 6\log(\ell)\log(1/\varepsilon)/\gamma$, so $n' \geq \log(\ell) = \log(a)$, allowing us to just take a length $\log(a)$ prefix of each length n' block to create a new source $\mathbf{X}''$ over $[a]^d$, as required. Finally, we can analyze t and k in this setting. We begin with t using Lemma 5.12 to infer that there exists some C' such that

$$\begin{aligned} t &\leq C'(\log(a) + \log\log(d) + d/a) \\ &= C'(\log(\ell) + \log\log(6\ell\log(1/\varepsilon)/\gamma) + 6\log(1/\varepsilon)/\gamma) \\ &\leq C_3 \log(\ell/\varepsilon) \end{aligned}$$

for some constant C_3. We compute k as

$$\begin{aligned} k &\geq \gamma d/a \\ &= \gamma(6\ell\log(1/\varepsilon)/\gamma)/\ell \\ &= 6\log(1/\varepsilon). \end{aligned}$$

showing that $k \geq 6\log(1/\varepsilon)$. Finally, recall that in this case $\log(1/\varepsilon) > \log(\ell)$, so $\log(1/\varepsilon) > \log(\ell/\varepsilon)/2$, which we can apply to get that $k \geq 6\log(1/\varepsilon) > 3\log(\ell/\varepsilon)$.

Let $C = \max(C_2, C_3)$. In either case, we have that the number of output bits is $t \leq C(\log(\ell/\varepsilon))$ and the min-entropy k of $\mathsf{Reduce}(\mathbf{X})$ is $\geq 3\log(\ell/\varepsilon)$, as claimed. □

5.7 Constructing 2Cond

In this subsection we will prove our remaining helper lemma - Lemma 5.4. First, we will require the following result:

Lemma 5.13. *There exists universal constant C' such that for all $n_x, k_x, n_{y,1}, \ldots, n_{y,t}, m, 0 < \varepsilon_1 \leq \cdots \leq \varepsilon_t < 1$ satisfying $n_{y,i} \geq C' \log(2n_x/\varepsilon_i)$ and $m = \frac{k_x - \log(2/\varepsilon_1)}{3}$, the following holds: There exists an explicit extractor $\mathsf{Ext} : \{0,1\}^{n_x} \times \{0,1\}^{n_{y,i}} \times \cdots \times \{0,1\}^{n_{y,t}} \to \{0,1\}^m$ satisfying: For all $1 \leq j \leq t$ and all independent sources $\mathbf{X} \sim \{0,1\}^{n_x}, \mathbf{Y}_1 \sim \{0,1\}^{n_{y,1}}, \ldots, \mathbf{Y}_t \sim \{0,1\}^{n_{y,t}}$ where $H_\infty(\mathbf{X}) = k_x$, each of $\mathbf{Y}_1, \ldots, \mathbf{Y}_{j-1}$ are fixed constants and all $\mathbf{Y}_j, \ldots, \mathbf{Y}_t$ are uniform, we have that $\mathsf{Ext}(\mathbf{X}, \mathbf{Y}_1, \ldots, \mathbf{Y}_t)$ is ε_j-close to $\mathbf{U}_m$.*

Proof of Lemma 5.13. For $1 \leq i \leq t$, let $\mathsf{sExt}_i : \{0,1\}^{n_x} \times \{0,1\}^{n_{y,i}} \to \{0,1\}^m$ be explicit $(\varepsilon_i/2)$-seeded-extractor guaranteed by Theorem 4.6 - where we assume the universal constant from Theorem 4.6 is C' and check that our parameters meet the requirements. Our extractor construction is:

$$\mathsf{Ext}(x, y_1, \ldots, y_t) = \bigoplus_{i=1}^{t} \mathsf{sExt}_i(x, y_i).$$

Let $\mathbf{Z}_{good} = \mathsf{sExt}_j(\mathbf{X}, \mathbf{Y}_j)$ and let $\mathbf{Z}_{rest} = \bigoplus_{1 \leq i \leq t, i \neq j} \mathsf{sExt}_i(\mathbf{X}, \mathbf{Y}_i)$. Notice that our final output distribution is $\mathbf{Z}_{good} \oplus \mathbf{Z}_{rest}$. We will argue that on most fixings of $\mathbf{Z}_{rest}$, the output will be close to uniform.

By Lemma 4.3, we have the following (where the probability below is over sampling from $\mathbf{Z}_{rest}$)

$$\Pr[H_\infty(\mathbf{X}|\mathbf{Z}_{rest} = z_{rest}) \geq k_x - m - \log(2/\varepsilon_j)] \geq 1 - \varepsilon_j/2.$$

Call the fixings z_{rest} of $\mathbf{Z}_{rest}$ that satisfy the above property of leaving $\mathbf{X}$ with a lot of entropy when conditioning on them, as the "good fixings." As $\mathbf{Z}_{rest}$ is independent of $\mathbf{Y}_j$ and $\mathbf{X}$ is left with a lot of entropy conditioning on a good fixing z_{rest}, we have that

$$\mathsf{sExt}_j((\mathbf{X}|\mathbf{Z}_{rest} = z_{rest}), (\mathbf{Y}_j|\mathbf{Z}_{rest} = z_{rest})) \approx_{\varepsilon_j/2} \mathbf{U}_m.$$

As $1 - \varepsilon_j/2$ fraction of fixings of $\mathbf{Z}_{rest}$ are good, we conclude that $\mathsf{Ext}(\mathbf{X}, \mathbf{Y}_1, \ldots, \mathbf{Y}_t) \approx_{\varepsilon_j} \mathbf{U}_m$ as desired. □

With this, we finally proved the proof of our lemma:

Proof of Lemma 5.4. Let C' be a universal constant that we set later. For $1 \leq i \leq t$, let $n_{z,i} = 2C'(3C')^{t-i} \log(2tn_x/\varepsilon)$ and let $n_z = \sum_{i=1}^{t} n_{z,i}$. Let $\mathbf{Z}_i$ be the length $n_{z,i}$ prefix of the block $\mathbf{Y}_i$, and let $\mathbf{Z} = \mathbf{Z}_1, \ldots, \mathbf{Z}_t$ be the concatenation of these prefixes. Note that by our lower bound guarantee on n_y, each block is long enough to take such prefixes. We use the extractor Ext from Lemma 5.13 with $k_x = k - n_z - \log(2/\varepsilon)$, $m = \frac{1}{3}(k - (3C')^t \log(2tn_x/\varepsilon))$ as in the lemma statement, and for $1 \leq i \leq t$, we set $\varepsilon_i = \left(\frac{\varepsilon}{2tn_x}\right)^{(3C')^{t-i}}$. With this, we define our condenser as:

$$2\mathsf{Cond}(\mathbf{X}, \mathbf{Y}) = \mathsf{Ext}(\mathbf{X}, \mathbf{Z}) = \mathsf{Ext}(\mathbf{X}, \mathbf{Z}_1, \ldots, \mathbf{Z}_t).$$

We easily compute and check that our parameter settings satisfy the requirements of Lemma 5.13. We will show that the output entropy (with error ε) is at least $m - n_z$. We compute that $n_z \leq (3C')^t \log(2tn_x/\varepsilon)$, the output entropy gap. Hence if we show this, then our condenser will indeed have the claimed property.

We now show that our condenser construction is correct. Since $\mathbf{Y}$ is guaranteed to have at least one good block by assumption, let $j \in [t]$ be the index of this good block. Now, let $\mathbf{A} = \mathbf{Z}_1, \ldots, \mathbf{Z}_{j-1}$ and let $\mathbf{B} = \mathbf{Z}_{j+1}, \ldots, \mathbf{Z}_t$ so that $\mathbf{Z} = (\mathbf{A}, \mathbf{Z}_j, \mathbf{B})$. We will show that $H_\infty^\varepsilon(\mathsf{Cond}(\mathbf{X}, \mathbf{Y})) \geq m - n_z$.

We will now consider fixings of $\mathbf{A}$. We say a fixing of $\mathbf{A} = a$ is good if $H_\infty(\mathbf{X}|\mathbf{A} = a) \geq k - n_z - \log(2/\varepsilon) = k_x$. By the min-entropy chain rule (Lemma 4.3), at least $1 - \varepsilon/2$ fraction of fixings of $\mathbf{A}$ are good. Since $\mathbf{Y}$ is an oNOSF source, $\mathbf{Z}_j$ remains independent and uniform of $\mathbf{X}$ for every fixing of $\mathbf{A}$.

We will show that, conditioned on a good fixing a of $\mathbf{A}$, we have $H_\infty^{\varepsilon/2}(\mathsf{Cond}(\mathbf{X}, \mathbf{Y})) \geq m - \sum_{i=j+1}^{t} n_{z,i} \geq m - n_z$. This will prove our result as our total error will be $\varepsilon/2 + \varepsilon/2 = \varepsilon$ and the min-entropy guarantee will be $m - n_z$, as desired.

Consider the best case scenario when $(\mathbf{B}|\mathbf{A} = a) = \mathbf{U}_{|\mathbf{B}|}$. This is unrealistic since it is possible that all bits in $\mathbf{B}$ are bad and arbitrarily depend on the remaining bits. Nevertheless, it is instructive to see what happens in this scenario. In this case, $\mathbf{X}, \mathbf{Y}$ are independent distributions, and we can infer that $\mathsf{Cond}(\mathbf{X}, \mathbf{Y}) = \mathsf{Ext}(\mathbf{X}, \mathbf{Z}) \approx_{\varepsilon_j} \mathbf{U}_m$. However, as alluded before, all bits in $\mathbf{B}$ can be adversarially set. To overcome this, we invoke Lemma 7.9 that allows us to compare how worse off our output distribution can be compared to the best case scenario. We conclude that even when $\mathbf{B}$ is completely adversarially controlled, $H_\infty^{\varepsilon'}(\mathsf{Cond}(\mathbf{X}, \mathbf{Y})) \geq m - |\mathbf{B}| = m - \sum_{i=j+1}^{t} n_{z,i}$ where

$$
\begin{aligned}
\varepsilon' &= \varepsilon_j \cdot 2^{|\mathbf{B}|} \\
&= \left(\frac{\varepsilon}{2tn_x}\right)^{(3C')^{t-j}} \cdot 2^{\sum_{i=j+1}^{t} n_{z,i}} \\
&= \left(\frac{\varepsilon}{2tn_x}\right)^{(3C')^{t-j}} \cdot 2^{2C' \log(2tn_x/\varepsilon) \sum_{i=j+1}^{t} (3C')^{t-i}} \\
&= \left(\frac{\varepsilon}{2tn_x}\right)^{(3C')^{t-j}} \cdot \left(\frac{2tn_x}{\varepsilon}\right)^{2C' \frac{(3C')^{t-j}-1}{3C'-1}} \\
&\leq \left(\frac{\varepsilon}{2tn_x}\right)^{(3C')^{t-j}} \cdot \left(\frac{2tn_x}{\varepsilon}\right)^{(3C')^{t-j}-1} \\
&\leq \frac{\varepsilon}{2tn_x} \\
&\leq \varepsilon/2
\end{aligned}
$$

This proves our claim, showing that for all good fixings, our output is highly condensed. We set our final universal constant C to be $4 \cdot C'$ and see that doing so only weakens the promise of our condenser.

We also need to be careful when invoking Lemma 7.9 since it requires that $(\mathbf{X}, \mathbf{A}, \mathbf{Z}_j, \mathbf{U}_{|\mathbf{B}|})$ should be a flat distribution. While that may not be true, we can express $\mathbf{X}$ as a convex combination of flat sources with the same min-entropy and since $\mathbf{A}$ is fixed and $\mathbf{Z}_j$ and $\mathbf{U}_{|\mathbf{B}|}$ are independent and uniform, we can express the joint distribution as a convex combination of flat sources, for each of them invoke the lemma, and conclude that the original distribution will be condensed as well. □

6 Transforming Low-Entropy oNOSF Sources to Uniform oNOSF Sources

In this section, we show how to transform low-entropy oNOSF sources into uniform oNOSF sources. Such a transformation was also provided in [CGR24]. Here, we obtain improved bounds using a generalized construction that allows us to obtain better tradeoffs and parameters in many more regimes of n, ℓ. Our main theorem is:

Theorem 6.1. *Let $d, g, g_{out}, \ell, n, m, k, \varepsilon$ be such that $g_{out} \leq g - \frac{\ell-g+2}{d}, n \geq k \geq \log(nd-k) + md + 2\log(2g_{out}/\varepsilon)$. Then, there exists a function $f : (\{0,1\}^n)^\ell \to (\{0,1\}^m)^{\ell-1}$ such that for any (g, ℓ, n, k)-oNOSF source $\mathbf{X}$, there exists uniform $(g_{out}, \ell-1, m)$-oNOSF source $\mathbf{Y}$ for which $|f(\mathbf{X}) - \mathbf{Y}| \leq \varepsilon$.*

The flexibility of setting d to any desired value allows us to obtain stronger results. For instance by setting d to be a large constant, we can get the following transformation that works even when n is very small compared to ℓ:

Corollary 6.2 (Transformation for small n)**.** *Let $g, \ell, n, m, k, \varepsilon, \delta$ be such that $\delta \leq 0.99, g = \delta\ell, n = \mathrm{poly}(\log(\delta\ell/\varepsilon)), k = \Omega(\log(\delta\ell/\varepsilon)), m = \Omega(k)$. Then, we can construct a function $f : (\{0,1\}^n)^\ell \to (\{0,1\}^m)^{\ell-1}$ such that: for any (g, ℓ, n, k)-oNOSF source $\mathbf{X}$, there exists uniform $(0.99\delta\ell, \ell-1, m)$-oNOSF source $\mathbf{Y}$ such that $|f(\mathbf{X}) - \mathbf{Y}| \leq \varepsilon$.*

We additionally note that when we set $d = \ell$, we recover the same construction as in [CGR24], matching its parameters. This is most interesting in the regime when say $\ell = O(1)$ and n is arbitrarily growing.

Corollary 6.3 (similar parameters as Theorem 5.2 from [CGR24])**.** *Let $g, \ell, n, m, k, \varepsilon$ be such that $k \geq 1.01(\log(n\ell) + 2\log(2(g-1)/\varepsilon)), m = k/200\ell$. Then, we can construct a function $f : (\{0,1\}^n)^\ell \to (\{0,1\}^m)^{\ell-1}$ such that for any (g, ℓ, n, k)-oNOSF source $\mathbf{X}$, there exists uniform $(g-1, \ell-1, m)$-oNOSF source $\mathbf{Y}$ such that $|f(\mathbf{X}) - \mathbf{Y}| \leq \varepsilon$.*

To obtain these transformations, we will use two-source extractors. In fact, using explicit construction of two-source-extractors, we also obtain an explicit transformation:

Corollary 6.4 (Explicit Transformation)**.** *There exists a universal constant C such that for all $d, g, g_{out}, \ell, n, m, k, \varepsilon$ satisfying $g_{out} \leq g - \frac{\ell-g+2}{d}, k \geq \mathrm{poly}(\log(n)) + md + 2\log(2g_{out}/\varepsilon) + O(1), m \leq \mathrm{poly}(\log n), \varepsilon \geq n^{-\Omega(1)}/2g_{out}$. the following holds: There exists an explicit function $f : (\{0,1\}^n)^\ell \to (\{0,1\}^m)^{\ell-1}$ such that for any (g, ℓ, n, k)-oNOSF source $\mathbf{X}$, there exists uniform $(g_{out}, \ell-1, m)$-oNOSF source $\mathbf{Y}$ for which $|f(\mathbf{X}) - \mathbf{Y}| \leq \varepsilon$.*

We can instantiate this lemma even in the case of constant d and get an explicit transformation similar to Corollary 6.2 with fewer output bits per block.

We will use the following main technical lemma that shows how to use two-source extractors to obtain these transformations:

Lemma 6.5 (Main Lemma)**.** *Let $d, g, g_{out}, \ell, n, m, k_{2\mathsf{Ext}}, k, \varepsilon_{2\mathsf{Ext}}$ be such that $k \geq k_{2\mathsf{Ext}} + m \cdot d + \log(1/\varepsilon_{2\mathsf{Ext}}), g_{out} \leq \frac{g(d+1)-\ell-2}{d}$. Let $2\mathsf{Ext} : \{0,1\}^{d \cdot n} \times \{0,1\}^n \to \{0,1\}^m$ be $(k_{2\mathsf{Ext}}, \varepsilon_{2\mathsf{Ext}})$-average-case-strong two-source extractor. Then, we can construct a function $f : (\{0,1\}^n)^\ell \to (\{0,1\}^m)^{\ell-1}$ such that for any (g, ℓ, n, k)-oNOSF source $\mathbf{X}$, there exists $(g_{out}, \ell-1, m)$-oNOSF source $\mathbf{Y}$ such that $|f(\mathbf{X}) - \mathbf{Y}| \leq \varepsilon$ where $\varepsilon = 2g_{out} \cdot \varepsilon_{2\mathsf{Ext}}$.*

Existentially, two-source-extractors with following parameters exist:

Lemma 6.6 (Lemma 5.4 from [CGR24])**.** *Let $n_1, n_2, k_1, k_2, m, \varepsilon$ be such that $k_1 \leq n_1, k_2 \leq n_2, m = k_1 + k_2 - 2\log(1/\varepsilon) - O(1)$, $k_2 \geq \log(n_1 - k_1) + 2\log(1/\varepsilon) + O(1)$, and $k_1 \geq \log(n_2 - k_2) + 2\log(1/\varepsilon) + O(1)$. Then, a random function $2\mathsf{Ext} : \{0,1\}^{n_1} \times \{0,1\}^{n_2} \to \{0,1\}^m$ is a (k_1, k_2, ε)-two source extractor with probability $1 - o(1)$.*

Using this, our main result follows:

Proof of Theorem 6.1. We use the two-source-extractors from Theorem 6.1 and apply it in Lemma 6.5. □

To make this transformation explicit, we can use the following construction of a two-source-extractor:

Theorem 6.7 ([CZ19, Mek17, Li16])**.** *There exists a universal constant $C \geq 1$ such that for all n, k, m, ε with $k \geq \log^C(n), m \leq n^{1/C}, \varepsilon \geq n^{-1/C}$, the following holds: There exists an explicit (n, k) two-source-extractor $2\mathsf{Ext} : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}^m$.*

With this our explicit transformation follows:

Proof of Corollary 6.4. We use the explicit two-source-extractors from Theorem 6.7 and apply it in Lemma 6.5. □

6.1 Low-Entropy oNOSF Source to Uniform Using Two-Source-Extractors

In this subsection, we will prove Lemma 6.5. To do this, we will use two-source-extractors and average-case two-source-extractors. Let's first define them:

Definition 6.8. *We say that* $2\mathsf{Ext}$ *is* (k_1, k_2, ε) average-case *strong if*

$$2\mathsf{Ext}(\mathbf{X}_1, \mathbf{X}_2), \mathbf{W} \approx_\varepsilon \mathbf{U}_m, \mathbf{W}$$

for every $\mathbf{X}_1$ and $\mathbf{W}$ such that $\widetilde{H}_\infty(\mathbf{X}_1 \mid \mathbf{W}) \geq k_1$ with $\mathbf{X}_2$ independent of $\mathbf{X}_1$ and $H_\infty(\mathbf{X}_2) \geq k_2$ and $\mathbf{W}$.

This notion of average-case two-source-extractors allows us obtain a simpler chain rule:

Lemma 6.9. *[DORS08] Let $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ be distributions such that $\mathsf{Supp}(\mathbf{B}) \leq 2^\lambda$. Then $\widetilde{H}_\infty(\mathbf{A} \mid \mathbf{B}, \mathbf{C}) \geq \widetilde{H}_\infty(\mathbf{A}, \mathbf{B} \mid \mathbf{C}) - \lambda \geq \widetilde{H}_\infty(\mathbf{A} \mid \mathbf{C}) - \lambda$.*

Lemma 2.3 of [DORS08] shows that all two-source extractors are average-case-two-source extractors with similar parameters.

Lemma 6.10. *[DORS08] For any $\eta > 0$, if $2\mathsf{Ext}$ is a (k_1, k_2, ε)-two-source extractor, then $2\mathsf{Ext}$ is a $(k_1 + \log(1/\eta), k_2, \varepsilon + \eta))$-average-case-two-source extractor.*

With this, we will finally prove our main lemma that shows how to use two-source-extractors to obtain our transformation:

Proof of Lemma 6.5. For $-d \leq i \leq 0$, define $\mathbf{X}_i$ to be the random variable that always outputs 0^n. For $2 \leq i \leq \ell$, we output $\mathbf{O}_i = 2\mathsf{Ext}(\mathbf{X}_{i-d} \circ \cdots \circ \mathbf{X}_{i-1}, \mathbf{X}_i)$.

For $2 \leq i \leq \ell$, we say that $\mathbf{O}_i$ is good if (1) $\mathbf{X}_i$ is good and (2) there exists a block amongst $\mathbf{X}_{i-d}, \ldots, \mathbf{X}_{i-1}$ that is good. We observe that if $\mathbf{O}_i$ is good, then $|\mathbf{O}_i - \mathbf{U}_m| \leq \varepsilon_{2\mathsf{Ext}}$. Let g' be the number of such good $\mathbf{O}_i$. Let $j_1, \ldots, j_g$ be the indices of the good blocks in $\mathbf{X}$. For $1 \leq i \leq g-1$, let $d_i = j_{i+1} - j_i$. We observe that g' equals number of i such that $d_i \leq d$. As $\sum_{i=1}^{g-1} d_i \leq \ell$ and $d_i \geq 1$, we infer that $g' \geq \frac{(g-1)(d+1)-\ell}{d}$. Hence, as long as $g_{out} \leq \lceil g' \rceil$, we can guarantee the desired number of good blocks in the output. This holds as long as $g_{out} \leq \frac{g(d+1)-\ell-2}{d}$.

Using Lemma 6.10, we infer that $2\mathsf{Ext}$ is $(k_{2\mathsf{Ext}} + \log(1/\varepsilon_{2\mathsf{Ext}}), 2\varepsilon_{2\mathsf{Ext}})$-average-case-two-source extractor. We will use this property below.

Now, using a hybrid argument we will show that

$$(\mathbf{O}_2, \ldots, \mathbf{O}_\ell) \approx_{2g_{out} \cdot \varepsilon_{2\mathsf{Ext}}} (\mathbf{Y}_2, \ldots, \mathbf{Y}_\ell)$$

where $\mathbf{Y} = (\mathbf{Y}_2, \ldots, \mathbf{Y}_\ell)$ is a uniform (g_{out}, ℓ, m)-oNOSF source that we will define as the proof goes. Let $\mathbf{Y}^{(1)} = (\mathbf{O}_2, \ldots, \mathbf{O}_\ell)$ and for $2 \le i \le \ell$, let $\mathbf{Y}^{(i)} = (\mathbf{O}_2, \ldots, \mathbf{O}_i, \mathbf{Y}_{i+1}, \ldots, \mathbf{Y}_\ell)$. Hence, $\mathbf{Y}^{(\ell)} = \mathbf{Y}$. We proceed by induction. We will show that for $2 \le i \le \ell$,

$$\left|\mathbf{Y}^{(i)} - \mathbf{Y}^{(i-1)}\right| \le 2\varepsilon_{\mathsf{2Ext}}$$

whenever $\mathbf{O}_i$ is good and

$$\mathbf{Y}^{(i)} = \mathbf{Y}^{(i-1)}$$

whenever $\mathbf{O}_i$ is bad. By repeated applications of the triangle inequality, we will have shown that our output is indeed close to some uniform oNOSF source with desired parameters.

We proceed by induction and let $i \ge 2$ be arbitrary. If $\mathbf{O}_i$ is bad, then we let $\mathbf{Y}_i = \mathbf{O}_i$. Then, we indeed have that $\mathbf{Y}^{(i)} = \mathbf{Y}^{(i-1)}$ as desired. Otherwise, we assume $\mathbf{O}_i$ is good. Then, it must be that $\mathbf{X}_i$ is good. Let i_{prev} be the index of the good block before $\mathbf{X}_i$ in $\mathbf{X}$. Then, we know that $i - i_{prev} \le d$. We first claim that

$$\widetilde{H}_\infty(\mathbf{X}_{i_{prev}} | \mathbf{O}_1, \ldots, \mathbf{O}_{i-1}) \ge k_{\mathsf{2Ext}} = k - m \cdot d$$

Firstly, by construction, blocks $\mathbf{O}_2, \mathbf{O}_{i_{prev}-1}$ are functions of blocks $\mathbf{X}_1, \ldots, \mathbf{X}_{i_{prev}-1}$. As $\mathbf{X}_{i_{prev}}$ is independent of $\mathbf{X}_1, \ldots, \mathbf{X}_{i_{prev}-1}$, we infer that $\mathbf{X}_{i_{prev}}$ is independent of $\mathbf{O}_2, \mathbf{O}_{i_{prev}-1}$. As 2Ext is average-case-strong, we apply Lemma 6.9 to get that

$$\widetilde{H}_\infty(\mathbf{X}_{i_{prev}} | \mathbf{O}_2, \ldots, \mathbf{O}_{i-1}) \ge k - m \cdot (i - i_{prev}) \ge k - m \cdot d = k_{\mathsf{2Ext}} + \log(1/\varepsilon)$$

where for the second last inequality, we used the fact that $i - i_{prev} \le d$. Moreover, as $\mathbf{X}_i$ is independent of $\mathbf{X}_1, \ldots, \mathbf{X}_{i-1}$ and $\mathbf{O}_2, \ldots, \mathbf{O}_{i-1}$ are solely functions of $\mathbf{X}_1, \ldots, \mathbf{X}_{i-1}$, we infer that $\mathbf{X}_i$ is independent of $\mathbf{O}_2, \ldots, \mathbf{O}_{i-1}$. Hence, conditioned on fixing $\mathbf{O}_2, \ldots, \mathbf{O}_{i-1}$, $\mathbf{O}_i$ will be $2\varepsilon_{\mathsf{2Ext}}$ close to $\mathbf{U}_m$. This implies $\mathbf{Y}^{(i-1)} \approx_{2\varepsilon_{\mathsf{2Ext}}} \mathbf{Y}^{(i)}$ as desired. This shows that a good block in $\mathbf{Y}$ is uniform conditioned on all previous blocks, .i.e., it is independent of all the blocks before it. This shows all bad blocks can only depend on good blocks appearing before them and that good blocks are independent of each other. This implies $\mathbf{Y}$ is indeed a uniform oNOSF source as desired. □

7 Existence of Condensers for All Values of ℓ, n

We will show that there exist condensers for uniform (g, ℓ, n)-oNOSF sources for almost all settings of ℓ, n, provided $g > 0.5\ell$. Observe that a uniform (g, ℓ, n)-oNOSF source is also a uniform $(g \cdot s, \ell \cdot s, n/s)$-oNOSF source by simply dividing up all blocks into s parts. This implies that as n becomes smaller (relative to ℓ), it gets harder to condense with the hardest case being $n = 1$. Our condenser will also be able to handle the case of $n = O(1)$ and ℓ arbitrarily growing:

Theorem 7.1 (Simplified version of Corollary 7.7)**.** *For all $g, \ell, n, \varepsilon, \delta$ where $g = 0.51\ell$, and $0.01\ell n \ge 2\log(\ell n/2\varepsilon) + O(1)$, there exists a condenser $\mathsf{Cond} : (\{0,1\}^n)^\ell \to \{0,1\}^m$ such that for any uniform (g, ℓ, n)-oNOSF source $\mathbf{X}$, we have $H^\varepsilon_\infty(\mathsf{Cond}(\mathbf{X})) \ge m - \Delta$ where $m = 0.005\ell n + 200(\ell + \log(\ell n/2\varepsilon)) + O(1)$ and $\Delta = 200(\ell + \log(\ell n/2\varepsilon)) + O(1)$.*

Note that when n is a large enough constant, $m \ge 100\Delta$ and hence, the output entropy rate is at least 0.99.

In fact, we obtain a general result for all values of n, ℓ and when $g = 0.5\ell + e$ where $e \in \mathbb{N}$ is arbitrary. See Lemma 7.4 for the full tradeoff; to get slightly better parameters for small n, see Corollary 7.6.

We combine the above condenser for uniform oNOSF sources with the transformation for low-entropy oNOSF sources to uniform oNOSF sources from Corollary 6.2 to obtain the following condenser for low-entropy oNOSF sources:

Corollary 7.2. *Let $g, \ell, n, m, k, \varepsilon$ be such that $g = 0.51\ell, n = \text{poly}(\log(\ell/\varepsilon)), k = \Omega(\log(\ell/\varepsilon)), m = \Omega(\ell \log(\ell/\varepsilon))$. Then, we can construct condenser $\mathsf{Cond} : (\{0,1\}^n)^\ell \to \{0,1\}^m$ such that for any (g, ℓ, n, k)-oNOSF source $\mathbf{X}$, we have $H^\varepsilon_\infty(\mathsf{Cond}(\mathbf{X})) \geq m - \Delta$ where $\Delta = O(\ell + \log(1/\varepsilon))$.*

Remark 7.3. *Previous condensers from [CGR24] could only show that condensers exist for uniform oNOSF sources when $\ell = o(\log n)$. They relied on existence of low-error two source extractors equipped with an additional "regularity" property. Our constructions are much simpler, recover all their results with even better parameters, and work for all values of n and ℓ, including the hardest case of $n = O(1)$.*

We provide our general construction of condensers in Section 7.1. To do that, we will require another type of condenser for two uniform oNOSF sources where the bad bits of the second block are allowed to depend on the bits of the first block. We provide this construction in Section 7.2.

7.1 Constructing Condensers for Uniform oNOSF Sources

In this subsection, we will construct the following general condenser for uniform oNOSF sources:

Lemma 7.4 (General uniform oNOSF source condensing)**.** *For all $g, \ell, n, \varepsilon, e$ where $g \geq (\ell/2) + e$, and $en \geq 2\log(\ell n/2\varepsilon) + O(1)$, there exists a condenser $\mathsf{Cond} : (\{0,1\}^n)^\ell \to \{0,1\}^m$ such that for any uniform (g, ℓ, n)-oNOSF source $\mathbf{X}$, we have $H^\varepsilon_\infty(\mathsf{Cond}(\mathbf{X})) \geq m - \Delta$ where $m = \frac{en}{2} + (2\ell - e)\left\lceil \frac{\log(\ell n/2\varepsilon)+O(1)}{e} \right\rceil + \log(1/\varepsilon) + O(1)$ and $\Delta = (2\ell - 2e)\left\lceil \frac{\log(\ell n/2\varepsilon)+O(1)}{e} \right\rceil + \log(1/\varepsilon) + O(1)$.*

To do this, we will use a condenser for two distinct uniform oNOSF sources where one source can depend on the other:

Lemma 7.5. *For all $g, \ell, n_x, n_y, \varepsilon$ where $n_x \geq n_y$ and $gn_y \geq \log(\ell n_x/\varepsilon) + O(1)$, there exists a condenser $\mathsf{Cond} : (\{0,1\}^{n_x})^\ell \times (\{0,1\}^{n_y})^\ell \to \{0,1\}^m$ such that: For any uniform (g, ℓ, n_x)-oNOSF source $\mathbf{X}$ and uniform (g, ℓ, n_y)-oNOSF source $\mathbf{Y}$ with the additional property that bad blocks in $\mathbf{Y}$ can depend on $\mathbf{X}$ as well, we have that $H^\varepsilon_\infty(\mathsf{Cond}(\mathbf{X}, \mathbf{Y})) \geq m - \Delta$ where $m = gn_x + (2\ell - g)n_y + \log(1/\varepsilon) + O(1)$ and $\Delta = (2\ell - 2g)n_y + \log(1/\varepsilon) + O(1)$.*

We construct this condenser in Section 7.2. Using this, our main general condenser can be constructed as follows:

Proof of Lemma 7.4. We split each block in $\mathbf{X}$ into 2 parts to obtain a uniform $(2g, 2\ell, n/2)$-oNOSF source. We call this resultant source $\mathbf{X}$ as well since it is the same distribution, just viewed differently. Let $\mathbf{U} = (\mathbf{U}_1, \ldots, \mathbf{U}_\ell)$ and where for $1 \leq i \leq \ell$, $\mathbf{U}_i = \mathbf{X}_i$. Let $\mathbf{V} = (\mathbf{V}_1, \ldots, \mathbf{V}_\ell)$ where for $1 \leq i \leq \ell$, we define $\mathbf{V}_i$ to be prefix of length n_v of $\mathbf{X}_{\ell+i}$ where $n_v = \left\lceil \frac{\log(\ell n/2\varepsilon)+O(1)}{e} \right\rceil$.

We observe that $\mathbf{U}$ is a uniform $(e, \ell, n/2)$-oNOSF source and $\mathbf{V}$ is a uniform (e, ℓ, n_v)-oNOSF source where bad bits in $\mathbf{V}$ can depend on $\mathbf{U}$ and the good bits in both sources are independent. We now define our condenser Cond to be the condenser from Lemma 7.5 applied to sources $\mathbf{U}, \mathbf{V}$. Hence, we will have that $H^\varepsilon_\infty(\mathsf{Cond}(\mathbf{U}, \mathbf{V})) \geq m - \Delta$ where $m = en/2 + (2\ell - e)n_y + \log(1/\varepsilon) + O(1)$ and $\Delta = (2\ell - 2e)n_y + \log(1/\varepsilon) + O(1)$ as desired. □

Our first corollary will apply to the regime that his the hardest to condense from, namely when n is very small compared to ℓ, even when $n = O(1)$ and ℓ is arbitrarily growing:

Corollary 7.6 (Small n)**.** *For all $g, \ell, n, \varepsilon, \delta$ where $g \geq (0.5+\delta)\ell, \varepsilon \geq 2^{-\delta\ell+O(1)}$, and $n \leq 2^{\delta\ell/2}$, there exists a condenser $\mathsf{Cond} : (\{0,1\}^n)^\ell \to \{0,1\}^m$ such that for any uniform (g, ℓ, n)-oNOSF source $\mathbf{X}$, we have $H^\varepsilon_\infty(\mathsf{Cond}(\mathbf{X})) \geq m-\Delta$ where $m = \delta\ell n/2+(2-\delta)\ell+\log(1/\varepsilon)+O(1)$ and $\Delta = (2-\delta)\ell+\log(1/\varepsilon)+O(1)$.*

Proof. We observe that $\left\lceil \frac{\log(\ell n/2\varepsilon)+O(1)}{e} \right\rceil = 1$ and directly apply Lemma 7.4. □

We also obtain the following general tradeoff for larger n that may be growing with ℓ or even when $\ell = O(1)$ and n growing alone (this applies to all n but is most interesting when n is large since Corollary 7.6 provides better tradeoff for small n).

Corollary 7.7 (Larger n)**.** *For all $g, \ell, n, \varepsilon, \delta$ where $g \geq (0.5 + \delta)\ell$, and $\delta\ell n \geq 2\log(\ell n/2\varepsilon) + O(1)$, there exists a condenser $\mathsf{Cond} : (\{0,1\}^n)^\ell \to \{0,1\}^m$ such that for any uniform (g, ℓ, n)-oNOSF source $\mathbf{X}$, we have $H^\varepsilon_\infty(\mathsf{Cond}(\mathbf{X})) \geq m-\Delta$ where $m = \frac{\delta\ell n}{2}+(2/\delta-1)(\log(\ell n/2\varepsilon)+O(1))+(2-\delta)\ell+\log(1/\varepsilon)+O(1)$ and $\Delta = (2/\delta - 1)(\log(\ell n/2\varepsilon) + O(1)) + 2(2 - \delta)\ell + \log(1/\varepsilon) + O(1)$.*

Proof. We observe that $\left\lceil \frac{\log(\ell n/2\varepsilon)+O(1)}{e} \right\rceil \leq 1 + \frac{\log(\ell n/2\varepsilon)+O(1)}{e}$ and apply that to the condenser from Lemma 7.4. □

7.2 Condenser for Two Uniform oNOSF Sources

In this subsection, we will prove Lemma 7.5. To construct the claimed condenser, we will use the following folklore result regarding existence of excellent seeded condensers (e.g., see Corollary 3 of [GLZ24]).

Theorem 7.8. *For all n, k, d, ε such that $d \geq \log(n/\varepsilon) + O(1)$, there exists a seeded condenser $\mathsf{sCond} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ such that for all $\mathbf{X} \sim \{0,1\}^n$ with $H_\infty(\mathbf{X}) = k$, we have $H^\varepsilon_\infty(\mathsf{Cond}(\mathbf{X})) \geq k + d$ where $m = k + d + \log(1/\varepsilon) + O(1)$.*

We will also use the following result from [CGR24] that states an adversary can't make things too bad if it controls very few bits. We note that similar lemmas have been useful in previous construction of condensers [BCDT19, BGM22, GLZ24]:

Lemma 7.9 (Lemma 6.18 in [CGR24])**.** *Let $\mathbf{X} \sim \{0,1\}^n$ be an arbitrary flat distribution and let $\mathsf{Cond} : \{0,1\}^n \to \{0,1\}^m$ be such that $H^\varepsilon_\infty(\mathsf{Cond}(\mathbf{X})) \geq k$. Let $G \subset [n]$ with $|G| = n - b$ be arbitrary. Let $\mathbf{X}_G \sim \{0,1\}^{n-b}$ be the projection of $\mathbf{X}$ onto G. Let $\mathbf{X}' \sim \{0,1\}^n$ be the distribution where the output bits defined by G equal $\mathbf{X}_G$ and remaining b bits are deterministic functions of the $n - b$ bits defined by G under the restriction that $\mathrm{Supp}(\mathbf{X}') \subset \mathrm{Supp}(\mathbf{X})$. Then, $H^{\varepsilon'}_\infty(\mathsf{Cond}(\mathbf{X}')) \geq k - b$ where $\varepsilon' = \varepsilon \cdot 2^b$.*

With this, we are ready to provide the construction of condensers for two uniform oNOSF sources:

Proof of Lemma 7.5. Let $\mathsf{sCond} : (\{0,1\}^{n_x})^\ell \times (\{0,1\}^{n_y})^\ell \to \{0,1\}^m$ be lossless condenser guaranteed from Theorem 7.8 with $\varepsilon_{\mathsf{sCond}} = \varepsilon \cdot 2^{-(\ell-g)n_y}$. We define $\mathsf{Cond}(x, y) = \mathsf{sCond}(x, y)$.

Let $\mathbf{O}_{unif} = \mathsf{Cond}(\mathbf{X}, \mathbf{U}_{\ell n_y})$ and $\mathbf{O}_{adv} = \mathsf{Cond}(\mathbf{X}, \mathbf{Y})$. We argue that $\mathbf{O}_{unif}$ will be highly condensed and since the adversary controls so few bits in $\mathbf{Y}$, $\mathbf{O}_{adv}$ will be condensed as well.

We first see that by the property of the seeded condenser, $H^{\varepsilon_{\mathsf{sCond}}}_\infty(\mathbf{O}_{unif}) \geq gn_x+\ell n_y$. Next we observe that $\mathbf{O}_{adv}$ can be obtained from $\mathbf{O}_{unif}$ by an adversary controlling $b = (\ell - g)n_y$ bits from $(\mathbf{X}, \mathbf{U}_{\ell n_y})$ to

obtain $(\mathbf{X}, \mathbf{Y})$ and considering the output of sCond. We apply Lemma 7.9 which allows us to compare output entropy in such scenarios and obtain that

$$H_\infty^{\varepsilon_{\mathsf{sCond}} \cdot 2^b}(\mathbf{O}_{adv}) \geq H_\infty^{\varepsilon}(\mathbf{O}_{unif}) - b \geq (gn_x + \ell n_y) - ((\ell - g)n_y) = m - \Delta.$$

As $\varepsilon_{\mathsf{sCond}} \cdot 2^b = \varepsilon$, we indeed have that $H_\infty^{\varepsilon}(\mathbf{O}_{adv}) \geq m - \Delta$ as desired. □

8 Extractors for oNOSF and oNOBF Sources via Leader Election Protocols

In this section, we provide a generic way to transform leader election and coin flipping protocols into extractors for oNOSF sources and oNOBF sources. To do so, we must formally define the online influence of coalitions.

Definition 8.1. *For any function $f : \Sigma^\ell \to \{0,1\}$, and any $B \subset [\ell]$, where $B = \{i_1 < i_2 < \ldots < i_k\}$, define $\mathbf{oI}_B(f)$ as follows: an online adversary $\mathcal{A}$ samples a distribution $\mathbf{X}$ in online manner. It starts by sampling the variables $x_1, x_2, \ldots, x_{i_1 - 1}$ independently and uniformly from Σ, then picking the value of x_{i_1} depending on $x_{<i_1}$. Next, the variables $x_{i_1+1}, \ldots, x_{i_2-1}$ are sampled independently and uniformly from Σ, and $\mathcal{A}$ sets the value of x_{i_2} based on all variables set so far, and so on. Define the advantage of $\mathcal{A}$ to be $adv_{f,B}(\mathcal{A}) = |\mathbb{E}[f(\mathbf{X})] - \mathbb{E}[f(\mathbf{U}_\ell)]|$. Then $\mathbf{oI}_B(f)$ is defined to be $\max_{\mathcal{A}}\{adv_{f,B}(\mathcal{A})\}$, where the maximum is taken over all online adversaries $\mathcal{A}$ that control the bits in B.*

We say a function f is (b, ε)-online-resilient if $\mathbf{oI}_B(f) \leq \varepsilon$ for every set $B \subset [\ell]$ of size at most b.

We note that Definition 8.1 is a special case of Definition 1.5, for $\Sigma = \{0,1\}$ and $|B| = 1$.

Now we return to our transformation from leader election and coin flipping protocols into extractors for oNOSF sources and oNOBF sources. Conceptually, given a leader election protocol, we can use an oNOSF source to simulate the protocol and then have the elected leader output its last block. We formalize this below.

Lemma 8.2. *For any integers $r > 1, \ell > 0$ and any $\delta > 0$, let π be an $(r-1)$-round protocol over ℓ players that send n bits per round such that for any $\delta\ell$ bad players, the protocol elects a good leader with probability $1 - \varepsilon$.*

Then, there exists an explicit function $\mathsf{Ext} : (\{0,1\}^n)^{\ell r} \to \{0,1\}^n$ such that for any $(g, \ell r, n)$-oNOSF source $\mathbf{X}$ where $g \geq \ell r - \delta\ell$, we have $\mathsf{Ext}(\mathbf{X}) \approx_\varepsilon \mathbf{U}_n$.

Instantiating our lemmas with the leader election protocols from Section 9, we construct explicit extractors for oNOBF sources and uniform oNOSF sources:

Theorem 8.3. *There exists an explicit function $\mathsf{Ext} : \{0,1\}^\ell \to \{0,1\}$ such that for any δ and any (g, ℓ)-oNOBF source $\mathbf{X}$ where $g \geq \ell - \delta\ell/\log(\ell)$, we have $\mathsf{Ext}(\mathbf{X}) \approx_\varepsilon \mathbf{U}_1$ where $\varepsilon = C\delta + 12\,(C\delta)^{3/2} + \log(\ell)^{-1/3}$ where C is a large universal constant.*

Proof. This directly follows by instantiating Lemma 8.2 with the protocol guaranteed from Lemma 9.1. □

By using a the leader election protocol of Lemma 9.5 with multiple bits per round, we construct extractors for oNOSF sources:

Theorem 8.4. *There exists an explicit function $\mathsf{Ext} : (\{0,1\}^n)^\ell \to \{0,1\}^n$ such that for any constant δ and any (g, ℓ, n)-oNOSF source $\mathbf{X}$ where $g \geq \ell - \delta\ell/\log^*(\ell)$ and $n \geq \log(\ell)$, we have $\mathsf{Ext}(\mathbf{X}) \approx_\varepsilon \mathbf{U}_n$ where $\varepsilon = C\delta + 13\,(C\delta)^{3/2}$.*

Proof. This directly follows by instantiating Lemma 8.2 with the protocol guaranteed from Lemma 9.5. □

We finally prove our lemma regarding obtaining extractors for oNOSF Sources from leader election protocols:

Proof of Lemma 8.2. Define function Ext as follows: On input $(y_1, \ldots, y_r)$ where $y_i \in (\{0,1\}^n)^\ell$, let $y_{i,j} \in \{0,1\}^n$ denote the j'th block of y_i. Simulate the protocol π with the message of the j'th player in round i being $y_{i,j}$, where $1 \leq i \leq r-1$ and $1 \leq j \leq \ell$. Let $j^* \in [\ell]$ be the leader that is elected by π; then output y_{r,j^*}

Let us analyze Ext on some source $\mathbf{Y} \sim (\{0,1\}^n)^{\ell r}$. Let the bad symbols in $\mathbf{Y}$ be given by $A \subset [\ell] \times [r]$ where $|A| \leq \delta\ell$. Let $\mathbf{X} \sim ((\{0,1\}^n)^\ell)^r$ be the exact same source as $\mathbf{Y}$. We write $\mathbf{X} = \{\mathbf{X}_{i,j}\}_{1 \leq i \leq r, 1 \leq j \leq \ell}$ and interpret it as the distribution where $\mathbf{X}_{i,j}$ denotes the random bits of player j in round i. Call $\mathbf{X}_{i,j}$ a bad block if the corresponding index (i,j) is in A, i.e., the block is bad in $\mathbf{Y}$. Since a bad block in $\mathbf{Y}$ can only depend on blocks before it, the corresponding bad block in $\mathbf{X}$ satisfies the criteria for being bad in $\mathbf{X}$; this is because a bad block in the protocol setting is allowed to depend on all blocks in the same or previous rounds. Thus $\mathbf{X}$ has at most $\delta\ell$ bad blocks as well. By declaring the player corresponding to the bad block in $\mathbf{X}$ as bad, we obtain that the distribution $\mathbf{X}$ can be simulated by at most $\delta\ell$ bad players. Formally, for $1 \leq i \leq r$, let $B_i \subset [\ell]$ be the set of bad blocks in $\mathbf{X}$ among all blocks in round i. Let $B = \cup_{i=1}^r B_i$. We declare all players in B as bad players. Finally, observe that

$$|B| \leq \sum_{i=1}^{r} |B_i| = |A| = \delta\ell$$

as desired. Thus the correctness of π implies that after $(r-1)$ rounds, the chosen leader j^* does not belong to B with probability at least $1-\varepsilon$. By construction, it follows that $(r, j^*) \notin A$ whenever $j^* \notin B$. Thus, the output of the extractor, $\mathbf{Y}_{r,j*}$ is uniform on n bits, with probability at least $1-\varepsilon$. □

9 High Probability Leader Election Protocols

We use this section to provide the leader election protocols that are used in Section 8. In Section 9.1, we present leader protocols where each player is allowed to send one bit per round. We tackle the case where players can send multiple bits per round in Section 9.2.

9.1 One Bit per Round

We will construct leader election protocols with the following guarantees:

Lemma 9.1. *There exists a universal constant C and an explicit protocol over ℓ players, where each player sends $n = 1$ bit per round, that lasts for $C\log(\ell)$ rounds such that for any $\delta > 0$, if $\delta\ell$ players are bad, then a good leader is chosen with probability $\geq 1-\varepsilon$ where $\varepsilon = \delta + 12\delta^{3/2} + \log(\ell)^{-1/3}$.*

We will use the following protocol from [AN93]:

Lemma 9.2. *There exists a protocol π over ℓ players where each player sends at most 1 bit per round, that lasts for $O(\ell)$ rounds such that if $\delta\ell$ players are bad for $\delta \leq 1/4$, then a good leader is chosen with probability $\geq 1-\varepsilon$ where $\varepsilon = \delta + 12\delta^{3/2}$. Furthermore, this protocol can be explicitly constructed in time $2^{O(\ell)}$.*

We will also need the Chernoff bound:

Lemma 9.3. *Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be independent random variables taking values in $\{0,1\}$. Let $\mathbf{X} = \sum_{i=1}^{n} \mathbf{X}_i$ and let $\mu = \mathbb{E}[\mathbf{X}]$. Then, $\Pr[\mathbf{X} \leq (1-\delta)\mu] \leq e^{-\delta^2\mu/2}$.*

Proof of Lemma 9.1. Our protocol will have two stages. In the first stage, we will use the lightest bin protocol from [Fei99] until the number of players is small enough, and then in the second stage we use the protocol from Lemma 9.2. Let C_0 be a large constant that we set later. In particular, our final protocol will be:

1. Let $P_1 = [\ell]$.
2. In round i of stage 1, all players in P_i will present their value in $\{0,1\}$ and based on that, they will be divided into P_i^0, P_i^1.
3. Set P_{i+1} equal to the smaller set among P_i^0, P_i^1 (breaking ties arbitrarily).
4. Repeat this until the number of players becomes at most $C_0 \log \ell$. Let this happens after r rounds. This marks the end of the first stage.
5. In the second stage, apply the protocol from Lemma 9.2 to P_{r+1} and output the leader from that protocol.

We now analyze this protocol. We argue that at the end of the first stage, with high probability, the fraction of good players in P_{r+1} will be at least $(1-\delta) - o(1)$. For the second stage, the correctness of the protocol follows from Lemma 9.2.

For $1 \leq i \leq r+1$, let g_i be the number of good players in P_i and let $p_i = |P_i|$. As we always choose the lightest bin at each stage, $p_{i+1} \leq p_i/2$. Hence, we infer that $p_i \leq 2^{-i+1} \cdot \ell$. Let $g_1 = g$. We next lower bound g_i:

Claim 9.4. *With probability at least $1 - \exp(-(1/10) \cdot (g/2^r))$, it holds that for all $1 \leq i \leq r+1$, $g_i \geq \frac{g}{2^i} - 5\left(\frac{g}{2^i}\right)^{2/3}$.*

We prove this claim using concentration bounds later. Using this claim, we see that in P_{r+1}, the number of good players will be at least

$$\frac{(1-\delta)\ell}{2^r} - 5\left(\frac{(1-\delta)\ell}{2^r}\right)^{2/3}$$

out of $p_{r+1} \leq \frac{\ell}{2^r}$ many surviving players. In particular, $g_{r+1} \geq (1-\delta)p_{r+1} - 5p_{r+1}^{2/3}$. So, in stage 2, we have p_{r+1} many players remaining where the fraction of bad players is $\delta' = \delta + 5p_{r+1}^{-1/3}$. Applying Lemma 9.2 with these parameters, we infer that probability of electing a good leader is at least

$$1 - \left(\delta + 5p_{r+1}^{-1/3} + 12\left(\delta + 5p_{r+1}^{-1/3}\right)^{3/2}\right) \geq 1 - \delta - 12\delta^{3/2} - 6p_{r+1}^{-1/3}$$

where the last inequality follows because $p_{r+1} \geq \omega(1)$. Hence, our overall probability of electing a good leader is at least

$$1 - \delta - 12\delta^{3/2} - 6p_{r+1}^{-1/3} - \exp(-(1/10) \cdot (1-\delta)p_{r+1}) \geq 1 - \delta - 12\delta^{3/2} - \log(\ell)^{-1/3} = 1 - \varepsilon$$

where the last inequality follows because we let $p_{r+1} = C_0 \log(\ell)$ for a large constant C_0. We check that the number of rounds in the first stage is no more than $\log(\ell)$ and in stage 2, as guaranteed by Lemma 9.2, the number of rounds is no more than $O(\log(\ell))$. These together give us our universal constant C that we use in the claim.

Proof of Claim 9.4. Fix either of the two bins. We apply Lemma 9.3 with $\delta = \mu^{-1/3}$ to infer that with probability at least $1 - \exp(-(g_i/2)^{1/3}/2)$, it holds that the number of good players in that bin is $\geq g_i/2 - (g_i/2)^{2/3}$. Applying this to both bins, we infer that with probability at least $1 - 2\exp(-(g_i/2)^{1/3}/2)$, it holds that $g_{i+1} \geq g_i/2 - (g_i/2)^{2/3}$. By unravelling this recurrence and lower bounding, we see that

$$g_{i+1} \geq \frac{g}{2^i} - \sum_{j=1}^{i} \frac{(g/2^j)^{2/3}}{2^{i-j}}$$

Hence,

$$\begin{aligned}
g_{i+1} &\geq \frac{g}{2^i} - g^{2/3} \sum_{j=1}^{i} 2^{j/3-i} \\
&= \frac{g}{2^i} - \left(\frac{g}{2^i}\right)^{2/3} \sum_{j=1}^{i} (2^{1/3})^{j-i} \\
&= \frac{g}{2^i} - \left(\frac{g}{2^i}\right)^{2/3} \sum_{j=0}^{i-1} (2^{-1/3})^{j} \\
&\geq \frac{g}{2^i} - \left(\frac{g}{2^i}\right)^{2/3} \frac{1}{1-2^{-1/3}} \\
&\geq \frac{g}{2^i} - 5\left(\frac{g}{2^i}\right)^{2/3}.
\end{aligned}$$

□

By union bound, the overall probability that the claim holds is at least

$$\begin{aligned}
1 - \sum_{i=1}^{r+1} 2\exp(-(g_i/2)^{1/3}/2) &\geq 1 - \exp(-g_{r+1}/6) \\
&\geq 1 - \exp(-(1/10)\cdot(g/2^r)).
\end{aligned}$$

□

9.2 Multiple Bits per Round

If the players are allowed to send $O(\log \ell)$ bits per round, then the number of rounds can be significantly improved.

Lemma 9.5. *There exists a universal constant C and an explicit protocol over ℓ players where each player sends $n = \log \ell$ bits per round, that lasts for $C \cdot \log^* \ell$ rounds such that for any constant $\delta > 0$, if $\delta\ell$ players are bad, then a good leader is chosen with probability $1 - \varepsilon$ where $\varepsilon = \delta + 13\delta^{3/2}$.*

Proof. Our protocol and proof is similar to Lemma 9.1 with the key difference being that the larger value of n allows us to increase the number of bins and simplify our analysis. Here, we end up being verbose and repeating ourselves for clarity. Just like earlier, our protocol will have two stages, one using the lightest bin protocol from [Fei99] until the number of players is small enough and then resorting to the protocol from Lemma 9.2. Let C_0, C_1 be large constants that we set later. Our final protocol will be:

1. Let $P_1 = [\ell]$.
2. In round i of stage 1, all players in P_i will present a number between 1 and $b_i = |P_i| / \log(|P_i|)^{C_0}$. Based on this value, they will be divided into setsP_i^j where $j \in [b_i]$.

3. Set P_{i+1} equal to the smallest set amongst $P_i^1, \ldots, P_i^{b_i}$ (breaking ties arbitrarily).

4. Repeat this until the number of players becomes at most $\exp\left((\log(1/\delta))^{C_1}\right)$ (stop right before it goes below this value). Let this happens after r rounds. This marks the end of the first stage.

5. In the second stage, apply the protocol from Lemma 9.2 to P_{r+1} and output the leader from that protocol.

We now analyze this protocol. We argue that at the end of the first stage, with high probability, the fraction of good players in P_{r+1} will be at least $(1-\delta) - o(1)$. For the second stage, the correctness of the protocol follows from Lemma 9.2.

For $1 \le i \le r+1$, let g_i be the number of good players in P_i and let $p_i = |P_i|$. As we always choose the lightest bin at each stage, $p_{i+1} \le p_i/b_i$. Hence, we infer that $p_{r+1} \le \ell/\prod_{i=1}^r b_i$. Let $g = g_1$. We first bound g_i:

Claim 9.6. *For any constant C_1, with probability at least $1 - \exp(-\log(p_{r+1})^{1/5})$, it holds that for all $1 \le i \le r+1$, $g_i \ge \frac{g}{\prod_{j=1}^{i-1} b_j} - 2\left(\frac{g}{\prod_{j=1}^{i-1} b_j}\right)^{2/3}$.*

We prove this claim using concentration bounds later, and we remark that C_0 will be a growing function of C_1. Using this claim, we see that in P_{r+1}, the number of good players will be at least

$$\frac{(1-\delta)\ell}{\prod_{i=1}^r b_i} - 2\left(\frac{(1-\delta)\ell}{\prod_{j=1}^r b_i}\right)^{2/3}$$

out of $p_{r+1} \le \frac{\ell}{\prod_{i=1}^r b_i}$ many surviving players. In particular, $g_{r+1} \ge (1-\delta)p_{r+1} - 2p_{r+1}^{2/3}$.

So, in stage 2, we have p_{r+1} many players remaining where the fraction of bad players is $\delta' = \delta + 2p_{r+1}^{-1/3}$. Applying Lemma 9.2 with these parameters, we infer that probability of electing a good leader is at least

$$1 - \left(\delta + 2p_{r+1}^{-1/3} + 12\left(\delta + 2p_{r+1}^{-1/3}\right)^{3/2}\right) \ge 1 - \delta - 12\delta^{3/2} - 3p_{r+1}^{-1/3}$$

where the last inequality follows because $p_{r+1} \ge \omega(1)$. Hence, our overall probability of electing a good leader is at least

$$1-\delta-12\delta^{3/2}-3p_{r+1}^{-1/3}-\exp(-\log(p_{r+1})^{1/5}) \ge 1-\delta-12\delta^{3/2}-\exp(-\log(p_{r+1})^{1/6}) \ge 1-\delta-13\delta^{3/2} = 1-\varepsilon$$

where the first inequality follows because C_1 is a large enough universal constant, and $\delta < 1/4$. We check that the number of rounds in the first stage is no more than $O(\log^*(\ell))$ and in stage 2, as guaranteed by Lemma 9.2, the number of rounds is no more than $c\log^*(\ell)$, where c is a constant that just depends on δ and C_1 (and is independent of ℓ). These together give us our universal constant C of Lemma 9.5.

Proof of Claim 9.6. Fix any of the b_i bins in round i. We apply Lemma 9.3 with $\delta = \mu^{-1/3}$ to infer that with probability at least $1 - \exp(-(g_i/b_i)^{1/3}/2)$, it holds that the number of good players in that bin is $\ge g_i/b_i - (g_i/b_i)^{2/3}$. Applying this to all b_i bins, we infer that with probability at least $1 - b_i\exp(-(g_i/b_i)^{1/3}/2)$, it holds that $g_{i+1} \ge g_i/b_i - (g_i/b_i)^{2/3}$. By unraveling this recurrence and lower bounding, we see that

$$g_{i+1} \ge \frac{g}{\prod_{j=1}^i b_j} - \sum_{j=1}^i \frac{(g/\prod_{k=1}^j b_k)^{2/3}}{\prod_{k=j+1}^i b_k}$$

For ease of notation, let $\alpha(u,v) = \prod_{j=u}^{v} b_j$. Hence,

$$\begin{aligned} g_{i+1} &\geq \frac{g}{\alpha(1,i)} - g^{2/3} \sum_{j=1}^{i} \frac{(1/\alpha(1,j))^{2/3}}{\alpha(j+1,i)} \\ &= \frac{g}{\alpha(1,i)} - \left(\frac{g}{\alpha(1,i)}\right)^{2/3} \sum_{j=1}^{i} \frac{(\alpha(1,i)/\alpha(1,j))^{2/3}}{\alpha(j+1,i)} \\ &= \frac{g}{\alpha(1,i)} - \left(\frac{g}{\alpha(1,i)}\right)^{2/3} \sum_{j=1}^{i} \alpha(j+1,i)^{-1/3}. \end{aligned}$$

□

We observe that each term in the summand is exponentially decreasing. Hence, we can upper bound the the sum by $2\left(\frac{g}{\alpha(1,i)}\right)^{2/3}$.

This means

$$g_{i+1} \geq \frac{g}{\alpha(1,i)} - 2\left(\frac{g}{\alpha(1,i)}\right)^{2/3}.$$

By union bound, the overall probability that the claim holds is at least

$$1 - \sum_{i=1}^{r+1} b_i \exp(-(g_i/b_i)^{1/3}/2) = 1 - \sum_{i=1}^{r+1} \exp(-(g_i/b_i)^{1/3}/2 + \log(b_i)).$$

By our choice of parameters, in particular by letting C_0 to be a large enough constant, we can ensure that $g_i/b_i \geq \mathrm{poly}(b_i)$. Thus, we can ensure that the probability that the claim holds is at least

$$1 - \sum_{i=1}^{r+1} \exp(-(g_i/b_i)^{1/3}/2 + \log(b_i)) \geq 1 - \sum_{i=1}^{r+1} \exp(-\log(p_i)^{1/4})$$

where we get the constant $1/4$ by appropriately increasing C_0 and we used the fact that $\delta < 1/4$. As p_i is exponentially decreasing, we infer that the overall probability that the desired conclusion holds is at least

$$1 - \exp(-\log(p_{r+1})^{1/5}).$$

□

10 Online Influence and Extraction Lower Bounds

Towards proving lower bounds on the possiblity of extraction from oNOSF sources, we introduce a new, natural notion of influence of Boolean functions, which we call *online influence*. For simplicity, we first start by considering the class of oNOBF sources, which corresponds to uniform $(g, \ell, n = 1)$-oNOSF sources.

We believe this is an interesting new measure and is worth studying in its own right, and we refer the reader to Example 10.6 for a couple of interesting examples. For monotone functions (and more generally, unate functions), it is not hard to see that online influence equals the usual notion of influence (see Lemma 10.4 for a proof). Thus, to find interesting properties of online influence (compared to standard influence, Definition 10.1), one must look at non-monotone (in fact, non-unate) Boolean functions.

The following natural question arises towards our goal of proving extractor lower bounds: for a function f, what is the maximum online influence out of all n bits? For the usual notion of influence, this question was resolved by the well-known theorem of [KKL88], who showed there always exists a bit with influence at least $\mathrm{Var}(f) \cdot \Omega\left(\frac{\log \ell}{\ell}\right)$.

We show that surprisingly, there exists a balanced function, namely the address function, where every bit has online influence at most $O\left(\frac{1}{\ell}\right)$ (see Lemma 10.12 for a proof). This provides a separation between the usual notion of influence and online influence.

Organization We formally define the notion for Boolean functions and discuss some basic properties in Section 10.1. We establish tight bounds on the online influence for general functions, including a Poincaré style inequality, in Section 10.2. We provide an example exhibiting a separation between maximum (standard) influence and online influence in Section 10.3. Finally, in Section 10.4, we extend the definition of online influence to subsets of coordinates (and functions from $\Sigma^n \to \{0,1\}^m$, for arbitrary alphabet Σ). This allows us to prove the required lower bounds on extraction (and condensing) from oNOSF sources.

Notation For convenience, we introduce some notation that we use for the rest of this section. For any bit $b \in \{0,1\}$, let $e(b) = (-1)^b$. For any Boolean function $f : \{0,1\}^\ell \to \{0,1\}$, let $e(f)$ denote the function $e(f)(x) = (-1)^{f(x)}$.

10.1 Basic Properties

In this section, for a function $f : \{0,1\}^\ell \to \{0,1\}$, we will freely use commas to indicate concatenation in its input. For example, for $x \in \{0,1\}^{i-1}$ and $y \in \{0,1\}^{\ell-i}$, we write $f(x,1,y)$ to indicate f applied to the tuple $(x_1,\ldots,x_{i-1},1,y_1,\ldots,y_{\ell-i})$.

When asking about the influence of a single bit, such as the i-th bit, previous work has specifically looked at whether the i-th bit still has the ability to change the output of some function $f : \{0,1\}^\ell \to \{0,1\}$ after all other $\ell - 1$ bits have been set. In other words, if the i-th bit is a non-oblivious adversary (that is, it can look at the values of all the other bits before setting its own value), how much power does it have? This has led to a standard notion of influence defined below.

Definition 10.1 (Influence)**.** *For a function $f : \{0,1\}^\ell \to \{0,1\}$, the* influence *of the i-th bit is*

$$\mathbf{I}_i[f] = \mathop{\mathbb{E}}_{\substack{x \sim \mathbf{U}_{i-1} \\ y \sim \mathbf{U}_{n-i}}} [|f(x,1,y) - f(x,0,y)|]$$

and the total influence *is*

$$\mathbf{I}[f] = \sum_{i=1}^{\ell} \mathbf{I}_i[f].$$

However, in our setting of oNOSF sources and oNOBF sources, an adversarial bit can only depend on the bits that come before it. This motivates our new definition of online influence, where we prevent the i-th bit from depending on bits that come after it by independently sampling subsequent bits.

Definition 10.2 (Online influence)**.** *For a function $f : \{0,1\}^\ell \to \{0,1\}$, the* online influence *of the i-th bit is*

$$\mathbf{oI}_i[f] = \mathop{\mathbb{E}}_{x \sim \mathbf{U}_{i-1}} \left[\left| \mathop{\mathbb{E}}_{y \sim \mathbf{U}_{\ell-i}} [f(x,1,y)] - \mathop{\mathbb{E}}_{y \sim \mathbf{U}_{\ell-i}} [f(x,0,y)] \right| \right]$$

and the total online influence *is*

$$\mathbf{oI}[f] = \sum_{i=1}^{\ell} \mathbf{oI}_i[f].$$

Remark 10.3. *It is easy to see that for any $f : \{0,1\}^\ell \to \{0,1\}$, and any $i \in \ell$, we have $\mathbf{oI}_i(f) \le \mathbf{I}_i(f)$. Further, they are the same for the last bit: $\mathbf{I}_\ell[f] = \mathbf{oI}_\ell[f]$.*

Many results for the influence of a function are based on working with monotone functions. In contrast, it turns out that monotone functions are not very interesting for online influence as the definition collapses to that of regular influence.

Lemma 10.4. *If $f : \{0,1\}^\ell \to \{0,1\}$ is monotone, then $\mathbf{oI}_i[f] = \mathbf{I}_i[f]$ for all $i \in [\ell]$.*

Proof. Using the monotonicity of f, note that for any $x \in \{0,1\}^{i-1}$ and any $y \in \{0,1\}^{\ell-i}$, $f(x,1,y) \ge f(x,0,y)$. Thus, $\mathbf{oI}_i[f] = \mathbb{E}_{x\sim\mathbf{U}_{i-1},y\sim\mathbf{U}_{\ell-i}}[f(x,1,y) - f(x,0,y)] = \mathbf{I}_i(f)$. □

Thus, any difference between influence and online influence can only be demonstrated by non-monotone functions.

10.2 A Poincaré Inequality for Online Influence

Similar to regular influence, we prove a Poincaré-style inequality holds for online influence, and also provide an upper bound on online influence. The following is the main result of this subsection.

Theorem 10.5. *For any $f : \{0,1\}^\ell \to \{0,1\}$, we have $\mathrm{Var}(e(f)) \le \mathbf{oI}[f] \le \sqrt{\ell \,\mathrm{Var}(e(f))}$.*

Before proving the above result, we observe that the MAJORITY and PARITY functions provide tight examples for the upper and lower bound respectively for Theorem 10.5.

Example 10.6. *The majority function on ℓ bits $\mathrm{Maj}_\ell : \{0,1\}^\ell \to \{0,1\}$, is monotone, and hence by by Lemma 10.4, has total online influence $\mathbf{oI}[\mathrm{Maj}_\ell] = \mathbf{I}[\mathrm{Maj}_\ell] = \sqrt{2\ell/\pi} + O(1/\sqrt{\ell})$, achieving the upper bound (up to constants).*

The PARITY function on ℓ bits $\bigoplus_\ell : \{0,1\}^\ell \to \{0,1\}$ for $i \in [\ell-1]$ has online influence $\mathbf{oI}_i[\bigoplus_\ell] = 0$, while $\mathbf{oI}_\ell[\bigoplus_\ell] = 1$. Thus, PARITY meets the lower bound of Theorem 10.5. We note that this is starkly different from regular influence where $\mathbf{I}_i[\bigoplus_\ell] = 1$ for all i.

To prove Theorem 10.5, we will use Boolean Fourier analysis. For any $f : \{0,1\}^n \to \{0,1\}$, $e(f)$ has a unique Fourier expansion given by: $e(f(x)) = \sum_{S\subseteq[\ell]} \widehat{f}(S)\chi_S(x)$, where $\chi_S(x) = (-1)^{\sum_{i\in S} x_i}$ and $\widehat{f}(S) = \mathbb{E}_{y\sim\mathbf{U}_\ell}[e(f)(y)\chi_S(y)]$.[14] Also recall that $\widehat{f}(\emptyset) = \mathbb{E}_{x\sim\mathbf{U}_n}[e(f)(x)]$, $\mathrm{Var}(e(f)) = \sum_{S\subseteq[\ell],S\neq\emptyset} \widehat{f}(S)^2$, and for any $S \neq T$, $\mathbb{E}_{x\sim\mathbf{U}_\ell}[\chi_S(x)\chi_T(x)] = 0$. For more background, we refer the reader to the excellent book by O'Donnell [ODo14].

The following is our key lemma, from which Theorem 10.5 is easy to derive.

Lemma 10.7. *For any $f : \{0,1\}^\ell \to \{0,1\}$ and $i \in [\ell]$, $\mathbf{oI}_i(f)^2 \le \sum_{\substack{S\subseteq[i]\\ S\ni i}} \widehat{f}(S)^2 \le \mathbf{oI}_i(f)$.*

We first derive Theorem 10.5 using Lemma 10.7.

Proof of Theorem 10.5. We start with the lower bound. We have,

$$\mathbf{oI}[f] = \sum_{i=1}^{\ell} \mathbf{oI}_i[f] \ge \sum_{i=1}^{\ell} \sum_{\substack{S\subseteq[i]\\ S\ni i}} \widehat{f}(S)^2 = \sum_{\substack{S\subseteq[\ell]\\ S\neq\varnothing}} \widehat{f}(S)^2 = \mathrm{Var}(e(f)),$$

[14] For simplicity of notation, we use $\widehat{f}(S)$ for $\widehat{e(f)}(S)$.

where the inequality uses Lemma 10.7.

The upper bound is easy to derive as well.

$$\begin{aligned}
\mathbf{oI}[f] &= \sum_{i=1}^{\ell} \mathbf{oI}_i[f] \\
&\le \sqrt{\ell \sum_{i=1}^{\ell} (\mathbf{oI}_i[f])^2} && \text{(Cauchy-Schwarz inequality)} \\
&\le \sqrt{\ell \sum_{i=1}^{\ell} \sum_{\substack{S \subseteq [i] \\ S \ni i}} \widehat{f}(S)^2} && \text{(Lemma 10.7)} \\
&= \sqrt{\ell \operatorname{Var}(e(f))}.
\end{aligned}$$

This completes the proof. □

We now focus on proving Lemma 10.7. We need the following useful characterization of $\mathbf{oI}_i(f)$.

Claim 10.8. *For any $f : \{0,1\}^\ell \to \{0,1\}$, we can write the online influence of its i-th bit as*

$$\mathbf{oI}_i[f] = \mathop{\mathbb{E}}_{x \sim \mathbf{U}_{i-1}} \left[\left| \sum_{\substack{T \subseteq [i] \\ T \ni i}} \widehat{f}(T) \chi_{T \setminus \{i\}}(x) \right| \right].$$

Assuming the above claim, let us prove Lemma 10.7. We supply the proof of Claim 10.8 below.

Proof of Lemma 10.7. We first prove the inequality $\mathbf{oI}_i(f) \ge \sum_{\substack{S \subseteq [i] \\ S \ni i}} \widehat{f}(S)^2$. Since for any $x \in \{0,1\}^{i-1}$ we have $\left| \mathbb{E}_{y \sim \mathbf{U}_{\ell-i}}[e(f|_{x,1})(y)] - \mathbb{E}_{y \sim \mathbf{U}_{\ell-i}}[e(f|_{x,0})(y)] \right| = 2 \left| \sum_{\substack{T \subseteq [i] \\ T \ni i}} \widehat{f}(T) \chi_{T \setminus \{i\}}(x) \right|$ by Claim 10.8, and the fact that $\mathbb{E}_{y \sim \mathbf{U}_{\ell-i}}[e(f|_{x,b})(y)]$ is in $[-1,1]$ for all $x \in \{0,1\}^{i-1}, b \in \{0,1\}$, it follows that $\left| \sum_{\substack{T \subseteq [i] \\ T \ni i}} \widehat{f}(T) \chi_{T \setminus \{i\}}(x) \right|$ is in $[0,1]$.

Thus,

$$\begin{aligned}
\mathbf{oI}_i[f] &= \mathop{\mathbb{E}}_{x \sim \mathbf{U}_{i-1}} \left[\left| \sum_{\substack{T \subseteq [i] \\ T \ni i}} \widehat{f}(T) \chi_{T \setminus \{i\}}(x) \right| \right] \\
&\ge \mathop{\mathbb{E}}_{x \sim \mathbf{U}_{i-1}} \left[\left(\sum_{\substack{T \subseteq [i] \\ T \ni i}} \widehat{f}(T) \chi_{T \setminus \{i\}}(x) \right)^2 \right] \\
&= \sum_{\substack{T \subseteq [i] \\ T \ni i}} \sum_{\substack{S \subseteq [i] \\ S \ni i}} \widehat{f}(T) \widehat{f}(S) \cdot \mathop{\mathbb{E}}_{x \sim \mathbf{U}_{i-1}} [\chi_{T \setminus \{i\}}(x) \chi_{S \setminus \{i\}}(x)]
\end{aligned}$$

$$= \sum_{\substack{S \subseteq [i] \\ S \ni i}} \widehat{f}(S)^2.$$

Next, we prove $\mathbf{oI}_i(f)^2 \leq \sum_{\substack{S \subseteq [i] \\ S \ni i}} \widehat{f}(S)^2$. We have,

$$\begin{aligned}
\mathbf{oI}_i[f]^2 &= \left(\mathop{\mathbb{E}}_{x \sim \mathbf{U}_{i-1}} \left[\left| \sum_{\substack{T \subseteq [i] \\ T \ni i}} \widehat{f}(T) \chi_{T \setminus \{i\}}(x) \right| \right] \right)^2 && \text{(Claim 10.8)} \\
&\leq \mathop{\mathbb{E}}_{x \sim \mathbf{U}_{i-1}} \left[\left(\sum_{\substack{T \subseteq [i] \\ T \ni i}} \widehat{f}(T) \chi_{T \setminus \{i\}}(x) \right)^2 \right] \\
&= \sum_{\substack{S \subseteq [i] \\ S \ni i}} \widehat{f}(S)^2 && \text{(derived above).} \qquad \square
\end{aligned}$$

Next, we show how to rewrite $\mathbf{oI}_i[f]$ in terms of the Fourier coefficients of f.

Proof of Claim 10.8. We begin by defining the restriction $f|_{x,b}(y) = f(x,b,y)$ for $x \in \{0,1\}^{i-1}$, $b \in \{0,1\}$, and $y \in \{0,1\}^{\ell - i}$. Thus, we can rewrite $\mathbf{oI}_i[f]$ as

$$\mathbf{oI}_i[f] = \frac{1}{2} \cdot \mathop{\mathbb{E}}_{x \sim \mathbf{U}_{i-1}} \left[\left| \mathop{\mathbb{E}}_{y \sim \mathbf{U}_{\ell - i}} [e(f|_{x,1})(y)] - \mathop{\mathbb{E}}_{y \sim \mathbf{U}_{\ell - i}} [e(f|_{x,0})(y)] \right| \right]. \tag{4}$$

We would like to put the above expression in terms of Fourier coefficients of f. This motivates us to find the Fourier coefficients of $f|_{x,b}(y)$ in terms of those of f, which we do via computation. We manipulate the Fourier expansion of $f(z)$ for $z = (x,b,y) \in \{0,1\}^\ell$ to get

$$\begin{aligned}
e(f)(z) &= \sum_{S \subseteq [\ell]} \widehat{f}(S) \chi_S(z) \\
&= \sum_{S \subseteq [\ell]} \widehat{f}(S) \chi_S(x,b,y) \\
&= \sum_{S \subseteq [\ell]} \widehat{f}(S) \chi_{S \cap [i]}(x,b) \chi_{S \setminus [i]}(y) \\
&= \sum_{S \subseteq \{i+1, \ldots, \ell\}} \left(\sum_{T \subseteq [i]} \widehat{f}(S \cup T) \chi_T(x,b) \right) \chi_S(y).
\end{aligned} \tag{5}$$

We also have that

$$\begin{aligned}
e(f)(z) &= e(f)(x,b,y) \\
&= e(f|_{x,b})(y) \\
&= \sum_{S \subseteq \{i+1, \ldots, \ell\}} \widehat{f|_{x,b}}(S) \chi_S(y).
\end{aligned} \tag{6}$$

Therefore, Equation (5) and Equation (6) allow us to conclude that

$$\widehat{f|_{x,b}}(S) = \sum_{T \subseteq [i]} \widehat{f}(S \cup T)\chi_T(x,b).$$

Thus, we have

$$\begin{aligned}
\mathop{\mathbb{E}}_{y \sim \mathbf{U}_{\ell-i}}[e(f|_{x,b})(y)] &= \widehat{f|_{x,b}}(\varnothing) \\
&= \sum_{T \subseteq [i]} \widehat{f}(T)\chi_T(x,b) \\
&= \sum_{\substack{T \subseteq [i] \\ T \ni i}} \widehat{f}(T)\chi_{T \setminus \{i\}}(x)b + \sum_{T \subseteq [i-1]} \widehat{f}(T)\chi_T(x).
\end{aligned}$$

We now plug this in to our definition of $\mathbf{oI}_i[f]$ in Equation (4) to get a simplified expression. Recalling the fact that for any $x \in \{0,1\}^n$, $f(x) = (1 - e(f)(x))/2$, we have

$$\begin{aligned}
\mathbf{oI}_i[f] &= \frac{1}{2} \mathop{\mathbb{E}}_{x \sim \mathbf{U}_{i-1}} \left[\left| \mathop{\mathbb{E}}_{y \sim \mathbf{U}_{\ell-i}}[e(f|_{x,1})(y)] - \mathop{\mathbb{E}}_{y \sim \mathbf{U}_{\ell-i}}[e(f|_{x,0})(y)] \right| \right] \\
&= \frac{1}{2} \mathop{\mathbb{E}}_{x \sim \mathbf{U}_{i-1}} \left[\left| \left(-\sum_{\substack{T \subseteq [i] \\ T \ni i}} \widehat{f}(T)\chi_{T \setminus \{i\}}(x) + \sum_{T \subseteq [i-1]} \widehat{f}(T)\chi_T(x) \right) - \left(\sum_{\substack{T \subseteq [i] \\ T \ni i}} \widehat{f}(T)\chi_{T \setminus \{b\}}(x) + \sum_{T \subseteq [i-1]} \widehat{f}(T)\chi_T(x) \right) \right| \right] \\
&= \mathop{\mathbb{E}}_{x \sim \mathbf{U}_{i-1}} \left[\left| \sum_{\substack{T \subseteq [i] \\ T \ni i}} \widehat{f}(T)\chi_{T \setminus \{i\}}(x) \right| \right].
\end{aligned}$$

□

10.3 A Tight Example for Maximum Online Influence

The lower bound on total online influence from Theorem 10.5 allows us to conclude that for balanced functions, there must be at least one bit with online influence $\Omega(1/\ell)$. We can phrase this in terms of maximum influence.

Definition 10.9 (Maximum influence)**.** *For a function $f : \{0,1\}^\ell \to \{0,1\}$, we define its* maximum influence *as $\mathbf{I}_{\max}[f] = \max_{i \in [\ell]} \mathbf{I}_i[f]$ and its* maximum online influence *as $\mathbf{oI}_{\max}[f] = \max_{i \in [\ell]} \mathbf{oI}_i[f]$.*

In terms of maximum online influence, we get the following corollary from Theorem 10.5.

Corollary 10.10. *For a function $f : \{0,1\}^\ell \to \{0,1\}$, we have $\mathbf{oI}_{\max}[f] \geq \mathrm{Var}(e(f))/\ell$.*

Proof. By Theorem 10.5 we have that $\mathbf{oI}[f] = \sum_{i=1}^{\ell} \mathbf{oI}_i[f] \geq \mathrm{Var}(e(f))$, and the conclusion follows via an averaging argument. □

We show that the bound in Corollary 10.10 is in fact tight (up to constants), as witnessed by the address function.

Definition 10.11. *We define the* address function $\mathrm{Addr}_\ell : \{0,1\}^{\log(\ell)+\ell} \to \{0,1\}$ *as follows: For* $z \in \{0,1\}^{\log(\ell)+\ell}$*, split* z *up as* $z = (x, y)$ *with* x *of length* $\log(\ell)$ *and* y *of length* ℓ*. Then interpret* x *as a binary number which gives us an index* $i(x) \in [\ell]$*. The output of* Addr_ℓ *is the* $i(x)$*-th bit of* y*, so* $\mathrm{Addr}_\ell(x, y) = y_{i(x)}$.

Lemma 10.12. *Let* $m = \ell + \log \ell$ *and* Addr_ℓ *be the function defined above. Then,*

- *for* $1 \le i \le \log \ell$*,* $\mathbf{oI}_i[\mathrm{Addr}_\ell] = 0$.
- *for* $\log \ell < i \le m$*,* $\mathbf{oI}_i[\mathrm{Addr}_\ell] = 1/\ell$.

Thus, $\mathbf{oI}_{\max}(\mathrm{Addr}_\ell) = \Theta(1/m)$.

Proof. For $i \in [\log \ell]$, no matter what the value of the i-th bit of Addr_ℓ is set to, the output bit will be a uniform bit, so we immediately get that $\mathbf{oI}_i[f] = 0$. For $i \in \{\log \ell + 1, \ldots, m\}$, the i-th bit only has control if it's selected by the first $\log \ell$ address bits, meaning it has a $1/\ell$ chance of controlling the output (and otherwise the output is uniform). Hence, $\mathbf{oI}_i[f] = \frac{1}{\ell}$. □

Compared with the result of [KKL88] that $\mathbf{I}_{\max}[f] \ge \mathrm{Var}(f) \cdot \Omega\left(\frac{\log \ell}{\ell}\right)$, this exhibits a separation between maximum (standard) influence and the online influence (of balanced functions).

Moreover, this analysis of the address function also shows us that it is an extractor for uniform $(\ell - 1, \ell)$-oNOSF sources.

Lemma 10.13. *For all* ℓ, n *where* $\ell \ge 2$ *and* $n \ge \log(\ell - 1)$*, there exists an explicit extractor* $\mathsf{Ext} : (\{0,1\}^n)^\ell \to \{0,1\}^n$ *such that for any uniform* $(\ell - 1, \ell, n)$*-oNOSF source* $\mathbf{X}$*, we have* $\mathsf{Ext}(\mathbf{X}) \approx_\varepsilon \mathbf{U}_n$ *where* $\varepsilon = \frac{1}{\ell - 1}$.

Proof. Let Ext be defined as follows: From the first block, use the first $\log(\ell - 1)$ bits and interpret them as an index $j \in [\ell - 1]$. Then, output the block with index $j + 1$. For a source $\mathbf{X}$ with first block controlled by an adversary, the output will be truly uniform and for a source $\mathbf{X}$ with adversary controlling one of the last $\ell - 1$ blocks, that block will be outputted with probability $\frac{1}{\ell - 1}$ while a uniform block will be outputted otherwise. This makes our total error at most $\frac{1}{\ell - 1}$ as desired. □

10.4 Online Influence of Sets and Extraction Lower Bounds

For convenience we restate the definition of online influence of sets of coordinates.

Definition 10.14 (Online influence, Definition 8.1 restated)**.** *For any function* $f : \Sigma^\ell \to \{0,1\}$*, and any* $B \subset [\ell]$*, where* $B = \{i_1 < i_2 < \ldots < i_k\}$*, define* $\mathbf{oI}_B(f)$ *as follows: an online adversary* $\mathcal{A}$ *samples a distribution* $\mathbf{X}$ *in online manner. It starts by sampling the variables* $x_1, x_2, \ldots, x_{i_1 - 1}$ *independently and uniformly from* Σ*, then picking the value of* x_{i_1} *depending on* $x_{<i_1}$*. Next, the variables* $x_{i_1+1}, \ldots, x_{i_2-1}$ *are sampled independently and uniformly from* Σ*, and* $\mathcal{A}$ *sets the value of* x_{i_2} *based on all variables set so far, and so on. Define the advantage of* $\mathcal{A}$ *to be* $adv_{f,B}(\mathcal{A}) = |\,\mathbb{E}[f(\mathbf{X})] - \mathbb{E}[f(\mathbf{U}_\ell)]|$*. Then* $\mathbf{oI}_B(f)$ *is defined to be* $\max_{\mathcal{A}}\{adv_{f,B}(\mathcal{A})\}$*, where the maximum is taken over all online adversaries* $\mathcal{A}$ *that control the bits in* B.

We say a function f *is* (b, ε)*-online-resilient if* $\mathbf{oI}_B(f) \le \varepsilon$ *for every* B *of size at most* b.

In the special case where $\Sigma = \{0,1\}$ and we are considering the online influence of a single coordinate, the definition simplifies nicely.

Definition 10.15. *For a function $f : \{0,1\}^\ell \to \{0,1\}$, the* online influence *of the i-th bit is*

$$\mathbf{oI}_i[f] = \mathop{\mathbb{E}}_{x\sim\mathbf{U}_{i-1}}\left[\left|\mathop{\mathbb{E}}_{y\sim\mathbf{U}_{\ell-i}}[f(x,1,y)] - \mathop{\mathbb{E}}_{y\sim\mathbf{U}_{\ell-i}}[f(x,0,y)]\right|\right]$$

and the total online influence *is*

$$\mathbf{oI}[f] = \sum_{i=1}^{\ell} \mathbf{oI}_i[f].$$

Online-resilient functions are equivalent to extractors (with 1 output bit) for oNOSF sources.

Lemma 10.16 (online-resilient functions yield extractors)**.** *Let $f : \Sigma^\ell \to \{0,1\}$ be a (b,ε_1)-online-resilient function with the property that $|f(\mathbf{U}_\ell) - \mathbf{U}_1| \le \varepsilon_2$. Then f can extract from $(g = \ell - b, \ell)$-oNOSF sources with error at most $\varepsilon_1 + \varepsilon_2$.*

Proof. Consider a $(g = \ell - b, \ell)$-oNOSF source $\mathbf{X}$. Recall that $\mathbf{X}$ is created by choosing some set of bad indices B of size b, letting the symbols in $\overline{B}$ be uniform, and finally setting the symbols in B adversarially while only depending on uniform symbols to the left of them. Using the triangle inequality for total variation distance, we get that

$$\begin{aligned}|f(\mathbf{X}) - \mathbf{U}_1| &\le |f(\mathbf{X}) - f(\mathbf{U}_\ell)| + |f(\mathbf{U}_\ell) - \mathbf{U}_1| \\ &\le \varepsilon_1 + \varepsilon_2,\end{aligned}$$

as claimed. □

Remark 10.17. *We note that the other direction is immediate from definitions. If $\mathsf{Ext} : \Sigma^\ell \to \{0,1\}$ is an extractor with error ε for $(g = \ell - b, \ell)$-oNOSF sources, then Ext is a $(b, 2\varepsilon)$-online-resilient function.*

Remark 10.18. *Our results below on oNOBF extraction impossibility can be interpreted as a limit on online-resilience of balanced Boolean functions.*

For $B \subset [\ell]$, we use the notation $f|_{\overline{B}}$ to indicate the function obtained from f by letting an online adversary control the indices in B.

Theorem 10.19. *Let $f : \{0,1\}^\ell \to \{0,1\}$ be such that $\mathbb{E}_{x\sim\mathbf{U}_\ell}[f(x) = 1] = \alpha$. Then for any $1 \ge \beta > \alpha$, there exists a coalition $B \subseteq [\ell]$ such that $\mathbf{oI}_B(f) \ge \beta - \alpha$, where $|B| \le \gamma\ell$ and $\gamma = \frac{\beta-\alpha}{4\alpha(1-\beta)}$.*

Proof. We greedily collect the bits with the most online influence and add them to B until our goal of $\mathbb{E}_{x\sim\mathbf{U}_\ell|_{\overline{B}}}[f|_{\overline{B}}(x) = 1] \ge \beta$ is achieved. Our first step is as follows: let $B_0 = \varnothing$, $f_0 = f$, and $i_1 = \operatorname{argmax}_{i\in[\ell]}\{\mathbf{oI}_i[f]\}$. Corollary 10.10 tells us that $\mathbf{oI}_{i_1} \ge \mathrm{Var}(e(f_0))/\ell$. Recall that if $\mathbb{E}_{x\sim\mathbf{U}_\ell}[f(x) = 1] = p$ then $\mathrm{Var}(e(f)) = 4p(1-p)$. Because we have not yet achieved our goal of $\mathbb{E}_{x\sim\mathbf{U}_\ell|_{\overline{B}}}[f|_{\overline{B}}(x) = 1] \ge \beta$, we have that $\mathrm{Var}(f_0) \ge 4\alpha(1-\beta)$. Thus, we collect i_1 as $B_1 = \{i_1\}$, let $f_1 = f_0|_{\overline{B_1}}$ and see that $\mathbb{E}_x[f_1(x)] \ge \mathbb{E}_x[f_0(x)] + \mathbf{oI}_{i_1}[f_0] \ge \alpha + \frac{4\alpha(1-\beta)}{\ell}$.

We now repeat this process t times to get $B_t = \{i_1, \ldots, i_t\}$ until our goal is achieved. For general t, let $f_t = f|_{B_t}$ where $B_t = B_{t-1} \cup \{i_t\}$ and $i_t = \operatorname{argmax}_{i\in[n]\setminus B_{t-1}}\{\mathbf{oI}_i[f_{t-1}]\}$. At the $(t-1)$-th step, since we have not stopped, it means that $\mathbb{E}_x[f_{t-1}(x) = 1] < \beta$, but we of course have $\mathbb{E}_x[f_{t-1}(x) = 1] \ge \alpha$ as well. Thus, by Corollary 10.10, collecting i_t as a bad bit gives us that

$$\mathop{\mathbb{E}}_x[f_t(x)] \ge \mathop{\mathbb{E}}_x[f_{t-1}(x)] + \mathbf{oI}_{i_t}[f_{t-1}]$$

$$\geq \alpha + \frac{4\alpha(1-\beta)}{\ell}(t-1) + \frac{4\alpha(1-\beta)}{\ell}$$
$$= \alpha + \frac{4\alpha(1-\beta)}{\ell} \cdot t.$$

We repeat this process until $\Pr_x[f_t(x) = 1] \geq \beta$. Therefore, the number of steps is the smallest b such that $\alpha + \frac{4\alpha(1-\beta)}{\ell} \cdot b \geq \beta$, meaning that the number of steps is at most $b \leq \ell \cdot \frac{\beta-\alpha}{4\alpha(1-\beta)}$. We let $B = B_b$ and get the desired coalition. □

We can also ask the dual question of how large we are able to make β given some budget b of bad bits.

Corollary 10.20. *Let $f : \{0,1\}^\ell \to \{0,1\}$ be such that $\Pr_{x\sim\mathbf{U}_\ell}[f(x) = 1] \geq \alpha$. If we are able to control b bits in an online adversarial manner, then there exists a set $B \subseteq [\ell]$ of indices of size $|B| = b$ such that $\Pr_{x\sim\mathbf{U}_{\ell|_{\overline{B}}}}[f|_{\overline{B}}(x) = 1] \geq \beta$ where $\beta \geq \frac{\alpha(\ell+4b)}{\ell+4\alpha b}$.*

Proof. For a fixed β, Theorem 10.19 tells us that $b \leq \ell \cdot \frac{\beta-\alpha}{4\alpha(1-\beta)}$. Solving for β gives the desired bound. □

We now immediately obtain our oNOBF extraction impossibility result.

Corollary 10.21. *For any balanced function $f : \{0,1\}^\ell \to \{0,1\}$ and $0 < \varepsilon < 1/3$, there exists a $(g = \ell - b, \ell)$-oNOBF source $\mathbf{X}$ with $b \leq 3\varepsilon\ell$ such that $|f(\mathbf{X}) - \mathbf{U}_1| \geq \varepsilon$.*

Proof. It is enough to find a set B of indices such that $\mathbf{oI}_B(f) \geq \beta$. By Theorem 10.19, there exists such a set B of size $b = |B| \leq \ell \cdot \frac{\varepsilon}{1-2\varepsilon}$. The bound on $|B|$ follows since $\varepsilon \leq \frac{1}{3}$. □

Remark 10.22. *By essentially following our Fourier analytic proof, one can similarly obtain a Poincaré inequality for functions $f : \Sigma^n \to \{0,1\}$, for arbitrary alphabet Σ. To obtain extraction impossibility for such uniform oNOSF sources with constant δ fraction of corrupt blocks, we do the following: Let f be a candidate extractor for uniform $((1-\delta)\ell, \ell, n)$-oNOSF sources. Then, f also extracts from uniform $(\lceil 1/\delta \rceil - 1, \lceil 1/\delta \rceil, \ell n/\lceil 1/\delta \rceil)$-oNOSF source. Since there exists an influential coordinate with influence $O(\delta)$, we let the adversary control that coordinate and infer that there exists constant $\varepsilon = O(\delta)$ for which it is impossible to extract with error less than ε.*

11 Open Problems

We list here some interesting open problems left by our work:

- While we obtain explicit condensers for almost all parameter regimes, it remains open to construct them when the bock length is constant, matching the parameters of our existential results. As we show, one way of achieving this would be to explicitly construct a seeded condenser with dependence on seed length being $1 \cdot \log(1/\varepsilon)$.
- All our condensers have entropy gap much larger than a constant. It will be interesting to show there exist condensers with constant entropy gap (for any values of n, ℓ) for uniform oNOSF sources. A slightly weaker but equally interesting question is to construct seeded extractors for uniform oNOSF sources with constant seed length.
- Show that there exist non-trivial condensers for oNOBF sources or show no such condenser exists. We conjecture that no condenser exists with output entropy rate larger than the input entropy rate for such sources.

- Construct ε-collective sampling protocols with fewer rounds than the ones obtained using uniform oNOSF source condensers. It will also be interesting to explicitly construct such protocols when the number of players are very large compared to the number of bits each player has access to. Further, proving lower bounds for ε-collective sampling protocols is a natural direction to explore.

- Determine the exact threshold for extracting from oNOBF sources and oNOSF sources. Our lower bounds show extraction is impossible when $g \leq 0.99\ell$ while our constructions using leader election protocols require $g \geq \ell - \Omega\left(\frac{\ell}{\log \ell}\right)$ for oNOBF sources and $g \geq \ell - \Omega\left(\frac{\ell}{\log^*(\ell)}\right)$ for (g, ℓ, n)-oNOSF sources where $n \geq \log(\ell)$. Using the connection between extractors and leader election protocols, lower bounds for extraction imply lower bounds for leader election protocols. In particular, matching lower bounds for extraction would imply all current leader election protocols are tight, a long standing open problem.

Acknowledgements

We thank Madhur Tulsiani for asking a question that motivated us to consider the model of local oNOSF sources in Appendix B. We thank the organizers of the Dagstuhl Seminar on Algebraic and Analytic Methods in Computational Complexity and Schloss Dagstuhl for providing a stimulating research environment, where discussions between R.S. and E.C. contributed to this collaboration.

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. 2, 3, 6, 7).

[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 pp. 3, 7).

[AN93] Noga Alon and Moni Naor. "Coin-Flipping Games Immune Against Linear-Sized Coalitions". In: *SIAM J. Comput.* 22.2 (1993), pp. 403–417. DOI: 10.1137/0222030 (cit. on pp. 2, 14, 15, 42).

[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. 40).

[BBNRSSY09] Mihir Bellare, Zvika Brakerski, Moni Naor, Thomas Ristenpart, Gil Segev, Hovav Shacham, and Scott Yilek. "Hedged Public-Key Encryption: How to Protect against Bad Randomness". en. In: *Advances in Cryptology – ASIACRYPT 2009*. Ed. by Mitsuru Matsui. Berlin, Heidelberg: Springer, 2009, pp. 232–249. ISBN: 978-3-642-10366-7. DOI: 10.1007/978-3-642-10366-7_14 (cit. on p. 2).

[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: *DROPS-IDN/v2/document/10.4230/LIPIcs.APPROX-RANDOM.2019.43*. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2019. DOI: 10.4230/LIPIcs.APPROX-RANDOM.2019.43 (cit. on p. 40).

[BL89] Michael Ben-Or and Nathan Linial. "Collective Coin Flipping". In: *Advances In Computing Research* 5 (1989), pp. 91–115 (cit. on pp. 2, 15, 16).

[BGZ16] Iddo Bentov, Ariel Gabizon, and David Zuckerman. *Bitcoin Beacon*. arXiv:1605.04559 [cs]. May 2016. DOI: 10.48550/arXiv.1605.04559 (cit. on p. 2).

[BCG15] Joseph Bonneau, Jeremy Clark, and Steven Goldfeder. *On Bitcoin as a public randomness source*. Publication info: Preprint. MINOR revision. 2015 (cit. on p. 2).

[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 p. 7).

[BGB17] Benedikt Bünz, Steven Goldfeder, and Joseph Bonneau. "Proofs-of-delay and randomness beacons in ethereum". In: *IEEE Security and Privacy on the blockchain (IEEE S&B)* (2017) (cit. on p. 2).

[CGR24] Eshan Chattopadhyay, Mohit Gurumukhani, and Noam Ringach. "On the Existence of Seedless Condensers: Exploring the Terrain". In: *Proceedings of the 65th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*. To appear. 2024 (cit. on pp. 1–5, 12, 35, 36, 39, 40).

[CL22] Eshan Chattopadhyay and Jyun-Jie Liao. "Extractors for sum of two sources". In: *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing*. STOC 2022. New York, NY, USA: Association for Computing Machinery, June 2022, pp. 1584–1597. ISBN: 978-1-4503-9264-8. DOI: 10.1145/3519935.3519963 (cit. on p. 61).

[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 pp. 1, 3, 37).

[CGHFRS85] Benny Chor, Oded Goldreich, Johan Hasted, Joel Freidmann, Steven Rudich, and Roman Smolensky. "The bit extraction problem or t-resilient functions". In: *Proceedings of the 26th Annual Symposium on Foundations of Computer Science*. SFCS '85. USA: IEEE Computer Society, Oct. 1985, pp. 396–407. DOI: 10.1109/SFCS.1985.55 (cit. on p. 1).

[Dod06] Yevgeniy Dodis. *Fault-tolerant leader election and collective coin-flipping in the full information model*. 2006 (cit. on pp. 2, 16).

[DORS08] Yevgeniy Dodis, Rafail Ostrovsky, Leonid Reyzin, and Adam Smith. "Fuzzy Extractors: How to Generate Strong Keys from Biometrics and Other Noisy Data". In: *SIAM Journal on Computing* 38.1 (Jan. 2008). Publisher: Society for Industrial and Applied Mathematics, pp. 97–139. ISSN: 0097-5397. DOI: 10.1137/060651380 (cit. on p. 37).

[DPW14] Yevgeniy Dodis, Krzysztof Pietrzak, and Daniel Wichs. "Key Derivation without Entropy Waste". In: *Advances in Cryptology - EUROCRYPT 2014 - 33rd Annual International Conference on the Theory and Applications of Cryptographic Techniques, Copenhagen, Denmark, May 11-15, 2014. Proceedings*. Ed. by Phong Q. Nguyen and Elisabeth Oswald. Vol. 8441. Lecture Notes in Computer Science. Springer, 2014, pp. 93–110. DOI: 10.1007/978-3-642-55220-5_6 (cit. on p. 2).

[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. 2, 3).

[DMOZ25] Dean Doron, Dana Moshkovitz, Justin Oh, and David Zuckerman. "Online Condensing of Unpredictable Sources via Random Walks". In: *40th Computational Complexity Conference (CCC 2025)*. Ed. by Srikanth Srinivasan. Vol. 339. Leibniz International Proceedings in Informatics (LIPIcs). ISSN: 1868-8969. Dagstuhl, Germany: Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2025, 30:1–30:17. ISBN: 978-3-95977-379-9. DOI: 10.4230/LIPIcs.CCC.2025.30 (cit. on p. 3).

[DGW09] Zeev Dvir, Ariel Gabizon, and Avi Wigderson. "Extractors And Rank Extractors For Polynomial Sources". In: *Comput. Complex.* 18.1 (2009), pp. 1–58. DOI: 10.1007/S00037-009-0258-4 (cit. on p. 1).

[Fei99] Uriel Feige. "Noncryptographic Selection Protocols". In: *40th Annual Symposium on Foundations of Computer Science, FOCS '99, 17-18 October, 1999, New York, NY, USA*. IEEE Computer Society, 1999, pp. 142–153. DOI: 10.1109/SFFCS.1999.814586 (cit. on pp. 2, 14, 15, 43, 44).

[GGL91] O. Goldreich, S. Goldwasser, and N. Linial. "Fault-tolerant computation in the full information model". In: *[1991] Proceedings 32nd Annual Symposium of Foundations of Computer Science*. Oct. 1991, pp. 447–457. DOI: 10.1109/SFCS.1991.185405 (cit. on p. 2).

[GGL98] Oded Goldreich, Shafi Goldwasser, and Nathan Linial. "Fault-Tolerant Computation in the Full Information Model". In: *SIAM J. Comput.* 27.2 (1998), pp. 506–544. DOI: 10.1137/S0097539793246689 (cit. on pp. 15, 16).

[GSV05] Shafi Goldwasser, Madhu Sudan, and Vinod Vaikuntanathan. "Distributed Computing with Imperfect Randomness". In: *Distributed Computing, 19th International Conference, DISC 2005, Cracow, Poland, September 26-29, 2005, Proceedings*. Ed. by Pierre Fraigniaud. Vol. 3724. Lecture Notes in Computer Science. Springer, 2005, pp. 288–302. DOI: 10.1007/11561927_22 (cit. on p. 17).

[GLZ24] Jesse Goodman, Xin Li, and David Zuckerman. "Improved Condensers for Chor-Goldreich Sources". In: *Proceedings of the 65th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*. To appear. 2024 (cit. on pp. 3, 40).

[GSZ21] Vipul Goyal, Akshayaram Srinivasan, and Chenzhi Zhu. "Multi-source Non-malleable Extractors and Applications". In: *Advances in Cryptology - EUROCRYPT 2021 - 40th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Zagreb, Croatia, October 17-21, 2021, Proceedings, Part II*. Ed. by Anne Canteaut

and François-Xavier Standaert. Vol. 12697. Lecture Notes in Computer Science. Springer, 2021, pp. 468–497. DOI: 10.1007/978-3-030-77886-6_16 (cit. on p. 17).

[GVZ06] Ronen Gradwohl, Salil P. Vadhan, and David Zuckerman. "Random Selection with an Adversarial Majority". In: *Advances in Cryptology - CRYPTO 2006, 26th Annual International Cryptology Conference, Santa Barbara, California, USA, August 20-24, 2006, Proceedings*. Ed. by Cynthia Dwork. Vol. 4117. Lecture Notes in Computer Science. Springer, 2006, pp. 409–426. DOI: 10.1007/11818175_25 (cit. on p. 16).

[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 p. 18).

[IMV23] Peter Ivanov, Raghu Meka, and Emanuele Viola. "Efficient resilient functions". In: *Proceedings of the 2023 ACM-SIAM Symposium on Discrete Algorithms, SODA 2023, Florence, Italy, January 22-25, 2023*. Ed. by Nikhil Bansal and Viswanath Nagarajan. SIAM, 2023, pp. 2867–2874. DOI: 10.1137/1.9781611977554.CH108 (cit. on p. 3).

[IV24] Peter Ivanov and Emanuele Viola. "Resilient functions: Optimized, simplified, and generalized". In: *CoRR* abs/2406.19467 (2024). DOI: 10.48550/ARXIV.2406.19467. arXiv: 2406.19467 (cit. on p. 3).

[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. 7, 14, 15, 46, 52).

[KLRZ08] Yael Tauman Kalai, Xin Li, Anup Rao, and David Zuckerman. "Network Extractor Protocols". In: *49th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2008, October 25-28, 2008, Philadelphia, PA, USA*. IEEE Computer Society, 2008, pp. 654–663. DOI: 10.1109/FOCS.2008.73 (cit. on p. 17).

[KRVZ11] Jesse Kamp, Anup Rao, Salil Vadhan, and David Zuckerman. "Deterministic extractors for small-space sources". In: *Journal of Computer and System Sciences*. Celebrating Karp's Kyoto Prize 77.1 (Jan. 2011), pp. 191–220. ISSN: 0022-0000. DOI: 10.1016/j.jcss.2010.06.014 (cit. on p. 61).

[KZ07] Jesse Kamp and David Zuckerman. "Deterministic Extractors for Bit-Fixing Sources and Exposure-Resilient Cryptography". en. In: *SIAM Journal on Computing* 36.5 (Jan. 2007), pp. 1231–1247. ISSN: 0097-5397, 1095-7111. DOI: 10.1137/S0097539705446846 (cit. on p. 1).

[Li16] Xin Li. "Improved Two-Source Extractors, and Affine Extractors for Polylogarithmic Entropy". en. In: *2016 IEEE 57th Annual Symposium on Foundations of Computer Science (FOCS)*. New Brunswick, NJ, USA: IEEE, Oct. 2016, pp. 168–177. ISBN: 978-1-5090-3933-3. DOI: 10.1109/FOCS.2016.26 (cit. on p. 37).

[Li23] Xin Li. "Two Source Extractors for Asymptotically Optimal Entropy, and (Many) More". In: *64th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2023, Santa Cruz, CA, USA, November 6-9, 2023*. IEEE, 2023, pp. 1271–1281. DOI: 10.1109/FOCS57990.2023.00075 (cit. on p. 61).

[LLSZ97] Nathan Linial, Michael Luby, Michael Saks, and David Zuckerman. “Efficient construction of a small hitting set for combinatorial rectangles in high dimension”. In: *Combinatorica* 17.2 (1997), pp. 215–234 (cit. on p. 60).

[MW97] Ueli Maurer and Stefan Wolf. “Privacy amplification secure against active adversaries”. In: *Advances in Cryptology—CRYPTO'97: 17th Annual International Cryptology Conference Santa Barbara, California, USA August 17–21, 1997 Proceedings 17*. Springer. 1997, pp. 307–321 (cit. on p. 17).

[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 pp. 3, 37).

[MR95] Rajeev Motwani and Prabhakar Raghavan. *Randomized algorithms*. Cambridge University Press, 1995 (cit. on p. 1).

[ODo14] Ryan O'Donnell. *Analysis of boolean functions*. Cambridge University Press, 2014 (cit. on p. 48).

[PW18] Cécile Pierrot and Benjamin Wesolowski. “Malleability of the blockchain's entropy”. In: *Cryptography Commun.* 10.1 (Jan. 2018), pp. 211–233. ISSN: 1936-2447. DOI: 10.1007/s12095-017-0264-3 (cit. on p. 2).

[RZ01] Alexander Russell and David Zuckerman. “Perfect Information Leader Election in log* n+O (1) Rounds”. In: *J. Comput. Syst. Sci.* 63.4 (2001), pp. 612–626. DOI: 10.1006/JCSS.2001.1776 (cit. on pp. 8, 15, 32, 60).

[SV08] Saurabh Sanghvi and Salil P. Vadhan. “The Round Complexity of Two-Party Random Selection”. In: *SIAM J. Comput.* 38.2 (2008), pp. 523–550. DOI: 10.1137/050641715 (cit. on p. 16).

[TV00] Luca Trevisan and Salil P. Vadhan. “Extracting Randomness from Samplable Distributions”. In: *41st Annual Symposium on Foundations of Computer Science, FOCS 2000, 12-14 November 2000, Redondo Beach, California, USA*. IEEE Computer Society, 2000, pp. 32–42 (cit. on p. 1).

[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. 1).

[Zuc97] David Zuckerman. “Randomness-optimal oblivious sampling”. In: *Random Structures & Algorithms* 11.4 (1997). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/%28SICI%291098-2418%28199712%2911%3A4%3C345%3A%3AAID-RSA4%3E3.0.CO%3B2-Z, pp. 345–367. DOI: https://doi.org/10.1002/(SICI)1098-2418(199712)11:4<345::AID-RSA4>3.0.CO;2-Z (cit. on p. 18).

[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. 8, 19).

A Constructing Reduce'

In this section we construct Reduce' which has the properties as guaranteed by Lemma 5.12. In [RZ01], the authors use hitting sets for combinatorial rectangles to reduce ℓ-length oNOSF sources to shorter min-entropy sources. We provide a proof of their lemma for completeness here.

Let's first define combinatorial rectangles.

Definition A.1 (Combinatorial rectangle)**.** *Let $a, d \in \mathbb{N}$. We say that a set $R \subseteq [a]^d$ is a* combinatorial rectangle *if $R = R_1 \times R_2 \times \cdots \times R_d$ for some sets $R_i \subseteq [a]$ for $i \in [d]$. The* density *of R is* $\mathrm{Density}(R) = \frac{1}{a^d} \prod_{i=1}^{d} |R_d|$.

A hitting set for a family of combinatorial rectangles is a subset of $[a]^d$ such that it has an intersection with every combinatorial rectangle in the family. Formally:

Definition A.2 (Hitting sets for combinatorial rectangles)**.** *A set $\mathcal{H} \subseteq [a]^d$ is a (a, d, δ)*-hitting set for combinatorial rectangles *if for every combinatorial rectangle $R \subseteq [a]^d$ with $\mathrm{Density}(R) \geq \delta$ we have that $R \cap \mathcal{H} \neq \varnothing$.*

Of course, taking $\mathcal{H} = [a]^d$ is a trivial hitting set for any combinatorial rectangle, so the difficulty lies in decreasing the cardinality of $\mathcal{H}$ while keeping the density requirement δ of the combinatorial rectangle low. In [LLSZ97], the authors create a small enough hitting set for our use.

Lemma A.3 ([LLSZ97])**.** *There exists a universal constant C such that for any $\delta > 0$ and $a, d \in \mathbb{N}$, there exists an explicit construction of an (a, d, δ)-hitting set $\mathcal{H} \subseteq [a]^d$ such that $|\mathcal{H}| \leq \left(\frac{a \log(d)}{\delta}\right)^C$.*

Let's see how using all of these ingredients we can construct Reduce'.

Proof of Lemma 5.12. To construct Reduce', we begin by defining a family of functions $\mathcal{F} \subseteq \{f : [a]^d \to \{0,1\}\}$ and a hitting set $\mathcal{H} \subseteq [a]^d$ of size $|\mathcal{H}| = 2^t = T$. For every $x \in \mathrm{Supp}(\mathbf{X})$, we will select a $f_x \in \mathcal{F}$ and output the smallest $y \in \mathcal{H}$ such that $f_x(y) = 1$, where we consider our output as an element of $[|\mathcal{H}|] = \{0,1\}^t$. Then, for all $y \in \mathcal{H}$, we will show that $\Pr_{x \sim \mathbf{X}}[f_x(y) = 1] \leq 2^{-k}$, meaning that $\mathsf{Reduce}'(\mathbf{X})$ is a (t, k)-source.

Formally, we let $\mathcal{F}$ be the following family of combinatorial rectangles on $[a]^d$. Given an $x \in [a]^d$, we define the combinatorial rectangle $\mathsf{Rect}_x = \{y \in [a]^d \mid \forall i \in [d], y_i \neq x_i\}$ and the associated function $f_x : [a]^d \to \{0,1\}$ for this rectangle as $f_x(y) = 1_{y \in \mathsf{Rect}_x}$. Then, we let $\mathcal{F} = \{f_x \mid x \in [a]^d\}$.

Note that the density of any particular rectangle Rect_x is $\delta = \mathrm{Density}(\mathsf{Rect}_x) = \frac{\mathsf{Rect}_x}{a^d} = \left(1 - \frac{1}{a}\right)^d$. We can lower bound δ as $\delta \geq (\exp(-d/a))^{C_\delta}$ for some universal constant C_δ. Rearranging then gives us that $\log(1/\delta) \leq C_\delta \cdot \frac{d}{a}$. With this in mind, we set up our hitting set $\mathcal{H}$ for $\mathcal{F}$. From [LLSZ97], we know that there exists a universal constant C_1 and an explicit hitting set $\mathcal{H}$ such that $T = |\mathcal{H}| \leq \left(\frac{a \log(d)}{\delta}\right)^{C_1}$. Simplifying this expression yields

$$
\begin{aligned}
T &\leq \left(\frac{a \log(d)}{\delta}\right)^{C_1} \\
t &\leq C_1(\log(a) + \log\log(d) + \log(1/\delta)) \\
t &\leq C_1(\log(a) + \log\log(d) + C_\delta d/a) \\
t &\leq C'(\log(a) + \log\log(d) + d/a), \qquad (7)
\end{aligned}
$$

where C' is a sufficiently large universal constant, depending only on C_1, C_δ. To analyze the min-entropy of $\mathsf{Reduce}'(\mathbf{X})$, we note that for all $y \in [a]^d$

$$\begin{aligned}\Pr_{x\sim\mathbf{X}}[f_x(y)=1] &\le \left(1-\frac{1}{a}\right)^{\gamma d}\\ &\le \exp(-\gamma d/a) \le 2^{-\gamma d/a},\end{aligned}$$

which directly implies that the min-entropy k of $\mathsf{Reduce}'(\mathbf{X})$ is $\ge \gamma d/a$, as desired. □

B Extracting from Local oNOSF Sources

A natural variation on our definition of oNOSF sources is to consider the case where the adversary cannot remember the value of every good block in the past; rather, it can only remember the value of the most recent s blocks. Arguably, this is a realistic assumption in the setting of many short blocks, where it could be difficult to introduce long range correlation.

Definition B.1 (Local oNOSF sources)**.** *We call a (g,ℓ,n,k)-oNOSF source $\mathbf{X}=(\mathbf{X}_1,\dots,\mathbf{X}_\ell)$ an* s-local (g,ℓ,n,k)-oNOSF source *if each bad block $\mathbf{X}_i$ can only depend on at most s blocks $\mathbf{X}_{i-s},\dots,\mathbf{X}_{i-1}$ that come before it.*

Interestingly, weakening the adversary in this way converts our oNOSF source into a small-space source. These sources were first studied by [KRVZ11] and we refer the reader to them for a definition and background. Since the adversarial blocks of an s-local (g,ℓ,n,k)-oNOSF source can only depend on the binary string of length at most sn to its left, we easily see that an s-local (g,ℓ,n,k)-oNOSF source is samplable by a space-sn source.

Using recent explicit extractors for low-space sources provided by [CL22, Li23] and the fact that a (g,ℓ,n,k)-oNOSF source has entropy at least gk, we get the following extraction result for these local online sources.

Theorem B.2 (Using the explicit extractor of [CL22])**.** *There exists a universal constant C such that for every s and $k \ge \frac{2sn+\log^C(n\ell)}{g}$ there is an explicit extractor $\mathsf{Ext}:(\{0,1\}^n)^\ell \to \{0,1\}^m$ with error $\varepsilon = (n\ell)^{-\Omega(1)}$ and output length $m=(gk-2sn)^{\Omega(1)}$ for every s-local (g,ℓ,n,k)-oNOSF source.*

A similar result with slightly better entropy requirement, but constant error, can be obtained using the small-space extractor from [Li23].

ECCC ISSN 1433-8092
https://eccc.weizmann.ac.il