How to Extract Useful Randomness from Unreliable Sources

Divesh Aggarwal* Maciej Obremski† João Ribeiro‡ Luisa Siniscalchi§
Ivan Visconti¶

Abstract

For more than 30 years, cryptographers have been looking for public sources of uniform randomness in order to use them as a set-up to run appealing cryptographic protocols without relying on trusted third parties. Unfortunately, nowadays it is fair to assess that assuming the existence of physical phenomena producing public uniform randomness is far from reality.

It is known that uniform randomness cannot be extracted from a single weak source. A well-studied way to overcome this is to consider several independent weak sources. However, this means we must trust the various sampling processes of weak randomness from physical processes.

Motivated by the above state of affairs, this work considers a set-up where players can access multiple *potential* sources of weak randomness, several of which may be jointly corrupted by a computationally unbounded adversary. We introduce *SHELA* (Somewhere Honest Entropic Look Ahead) sources to model this situation.

We show that there is no hope of extracting uniform randomness from a *SHELA* source. Instead, we focus on the task of *Somewhere-Extraction* (i.e., outputting several candidate strings, some of which are uniformly distributed – yet we do not know which). We give explicit constructions of *Somewhere-Extractors* for *SHELA* sources with good parameters.

Then, we present applications of the above somewhere-extractor where the public uniform randomness can be replaced by the output of such extraction from corruptible sources, greatly outperforming trivial solutions. The output of somewhere-extraction is also useful in other settings, such as a suitable source of random coins for many randomized algorithms.

In another front, we comprehensively study the problem of *Somewhere-Extraction* from a *weak* source, resulting in a series of bounds. Our bounds highlight the fact that, in most regimes of parameters (including those relevant for applications), *SHELA* sources significantly outperform *weak* sources of comparable parameters both when it comes to the process of *Somewhere-Extraction*, and in the task of amplification of success probability in randomized algorithms. Moreover, the low quality of somewhere-extraction from weak sources excludes its use in various efficient applications.

1 Introduction

Perfect (i.e., uniform) public randomness is an extremely valuable resource in computer science, and in cryptography in particular. For example, it can be used to create a Common Reference

*Centre for Quantum Technologies and National University of Singapore. dcsdiva@nus.edu.sg

†Centre for Quantum Technologies. obremski.math@gmail.com

‡Imperial College London. j.lourenco-ribeiro17@imperial.ac.uk

§Aarhus University. lsiniscalchi@cs.au.dk. Part of the work was done while at the University of Salerno. Research supported in part by the European Union's Horizon 2020 research and innovation programme under grant agreement No 780477 (project PRIViLEDGE) and in part by "GNCS - INdAM".

¶University of Salerno. visconti@unisa.it. Research supported by the European Union's Horizon 2020 research and innovation programme under grant agreement No 780477 (project PRIViLEDGE).

String (CRS) drawn from an uniform distribution, which is a widely used set-up for cryptographic protocols. However, the randomness that we can obtain from physical phenomena (such as solar radiation, temperature readings, and electricity fluctuations) is far from perfect (in particular when public randomness sources are taken into account). Such phenomena belong to the family of *weak* randomness sources [1]. These are sources that carry some min-entropy, but are still very far from uniformly distributed. As a result, in most applications a so-called randomness extractor must be applied to the weak sources in order to extract (close to) uniformly distributed bits. A basic result about randomness extraction dictates that deterministic extraction from one weak source is not possible. Nevertheless, deterministic extraction *is* possible if one has access to at least two independent weak sources.

Sampling from several independent physical weak sources presents serious security issues. For example, if different phenomena are being publicly measured (to ensure some kind of independence), then different instrumentation and potentially different entities must be involved in the sampling process. Not only that, but sampling may also be compromised by instrument failures. Going back to our CRS example, if we want to generate CRS from such sources, then we are assuming that every instrument and entity that took part in sampling the weak sources is trusted. This is not a desirable situation, and indeed it was previously noticed that generating a uniformly distributed CRS from such weak sources is complicated [2]. A natural question follows: *Which forms of common public set-up can we achieve (or, more generally, what kind of randomness can we extract) if some of the sources are maliciously corrupted, but some of them remain honest?*

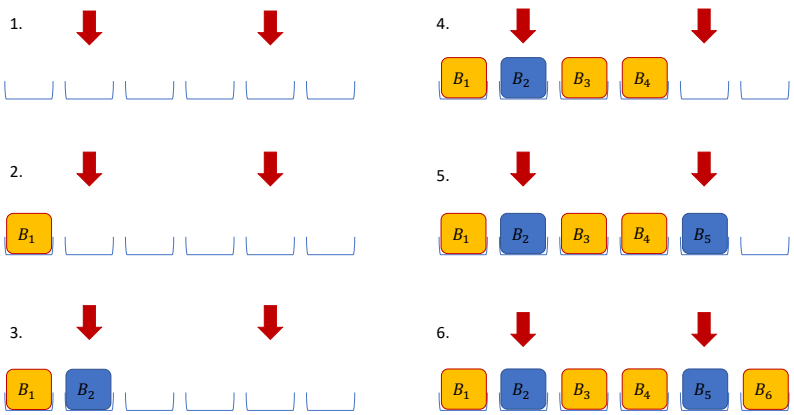

Figure 1: The procedure of sampling from a SHELA source. 1) Adversary chooses the positions of the honest blocks. 2) Adversary produces the first block. 3) Honest block is produced, it is independent of B_1. 4) Adversary fills out next blocks. Unlike B_1, blocks B_3, B_4 can depend on honest block B_2. 5) Honest block is produced, it is independent of B_1, B_2, B_3, B_4. 6) Adversary produces last block which can depend on all previous blocks.

Intuitively, this scenario leads us to define a structured weak source in an adversarial setting where a sample from the source is divided into multiple sub-parts, that we call *blocks*. One may imagine that each block corresponds to a different sampling process as per the previous paragraph. In this setting there is an ordered sequence of samplings from the sub-sources and some of them

are controlled by the adversary. More specifically, the adversary can decide the positions of the honest blocks since it can decide which sampling processes to corrupt. Honest blocks correspond to (correct) samples from independent weak sources (these sources are known to the adversary but are not controlled by the adversary). Given a sequence of blocks the sampling proceeds by obtaining blocks in chronological order. As a result, if the i-th block is to be corrupted, then the adversary is allowed to fix it to any value based on the (already determined) values from the first through $(i-1)$-th blocks.

We will call such source a "t-out-of-ℓ" Somewhere Honest Entropic Look-Ahead (SHELA) source, where ℓ indicates the total number of blocks, out of which t must be honest. We consider only the case $t \geq 2$, since the case $t = 1$ essentially reduces to the setting with a single weak source. Moreover, we assume without loss of generality[1] that each block has length n, and the honest blocks have min-entropy at least k for some decent parameter k. Observe that corrupted blocks are heavily correlated with previous honest blocks, and may even have zero min-entropy. Moreover, we allow the number of honest blocks t to be any function of ℓ, as long as $t \geq 2$.

There is a second real-world scenario that can be naturally modelled as a SHELA source. Some blockchains can be considered as sequences of blocks generated in chronological order, some of which contain high min-entropy strings. For instance, such strings could be the new wallet's identifier used to cash a reward when a new block is added to the chain, financial data containing some min-entropy [3], or a random nonce added for some security reasons. It is well-known [4, 5] that in a sequence of blocks of the blockchain there will be a fraction ν of them added by honest players. Moreover, we could assume that when a new block is added to the blockchain by a honest player, such a block (sometimes) contains high min-entropy strings that are independent of the previous ones already in the blockchain (we notice that a similar assumption has already been used in [6]). Therefore, if we consider ℓ consecutive blocks and for each of them we consider the part of the block that, in case the block is honest, could contain an independent weak source with decent min-entropy, we obtain a public SHELA source[2].

1.1 Our Contributions

Our main goal in this paper is to study SHELA sources and what kind of applications their availability enables.

The first natural question that arises when encountering SHELA sources is the following: *Are we able to extract independent and (close to) uniformly distributed bits from it?* We will prove in this work that the answer to this question is negative. Given this, we shift our focus from standard randomness extraction, and instead we investigate the possibility of constructing a deterministic *somewhere-extractor* SomeExt for SHELA sources. Intuitively, the somewhere-extractor SomeExt takes as input a SHELA source and outputs a distribution that is close (in statistical distance) to a convex combination of so-called "T-out-of-L" Somewhere-Random (SR) sources. SR sources are composed of L blocks, T of which (at fixed, unknown positions) are jointly independent and uniformly distributed. We call a convex combination of SR-sources a convSR-source for short.

It turns out that convSR sources are an extremely useful type of randomness. For example, armed with our somewhere-extractor, we show how to build non-interactive (and thus accepted by

[1] Given blocks of different sizes, one can always fill out the shorter blocks with zeros, similarly given blocks of different min-entropy we can assume k to be the minimum of min-entriopies of honest blocks.

[2] In this example we are assuming that when using a blockchain as a SHELA source, the adversary of the sampling procedure from a SHELA source has no control over the choices of the honest blocks posted permanently in the blockchain (i.e., the adversary does not decide which honest block is selected and remains permanently in the blockchain out of multiple candidates).

any receiver) commitments from one-way functions and non-interactive (and thus publicly verifiable) witness indistinguishable proofs from generic complexity assumptions[3] when both players (a sender and a receiver, or a prover and a verifier, respectively) have access to a public SHELA source. Remarkably, convSR-sources are also important intermediate objects used in the construction of multi-source and non-malleable extractors for weak sources (we discuss this in more detail later).

Parameters of the somewhere-extractor for SHELA sources. The computational complexity and security of our applications of convSR-sources will heavily depend on various parameters of the convSR-source: the number of total blocks L, the number of "good" (i.e., independent and uniformly distributed) blocks T, and the length m of each block. In turn, these depend on the parameters of the underlying SHELA source and the quality of the somewhere-extractor.

Ideally, we want our somewhere-extractor SomeExt to extract a convSR source with low error, small number of total blocks, and large block length from a SHELA source. More precisely, the error ε of SomeExt should satisfy $\varepsilon = 2^{-\Omega(n)}$, where n is the block length of the SHELA source, the total number L of blocks of the convSR source should be at most $O(\ell)$, where ℓ denotes the total number of blocks in the SHELA source, and the length m of each output block should satisfy $m = \Omega(n)$. We will comment later that these parameters ensure that the output of SomeExt can be used in our applications without compromising security, while ensuring that the efficiency and reliability of the application in question remain good enough.

Moreover, we do not want to assume that honest blocks in the SHELA source must have significant amounts of min-entropy for extraction to be successful. Instead, we aim to extract such high-quality convSR-sources from SHELA sources whose honest blocks have *arbitrary* constant min-entropy rate. In other words, we allow the min-entropy k of each honest n-bit block to satisfy $k = \delta n$ for an arbitrarily small constant $\delta > 0$.

A very first naive approach to designing a somewhere-extractor (that we will denote by NaiveSomeExt) is to apply a c-source extractor, for $c \geq 2$, to every subset of c blocks of a SHELA source. This immediately leads to a convSR-source. However, the total number of output blocks satisfies $L = \Theta(\ell^c)$ for $c \geq 2$, where ℓ denotes the total number of blocks of the SHELA source. This leads to a much worse efficiency blow-up for applications than what we aim to obtain, as detailed earlier. Another problem of the naive construction is that, if we wish to minimize the blowup of L with respect to ℓ by setting $c = 2$, we run into problems of explicitness. In fact, known explicit constructions of 2-source extractors require sources with high min-entropy to achieve exponentially small error [8, 9, 10]. We also note that, besides leading to worse efficiency, using a c-source extractor for $c > 2$ requires assuming that there are at least $c > 2$ honest blocks in the SHELA source, which might not be reasonable in some scenarios.

In this work, we design a non-trivial somewhere-extractor SomeExt that achieves our ideal goals put forth above. We begin by looking at the setting where the min-entropy rate k/n of honest blocks in the SHELA source is a large enough constant. In this case, if $X \in \{0,1\}^{n \cdot \ell}$ is a t-out-of-ℓ SHELA source with honest block min-entropy $k = \delta n$, then SomeExt(X) is ε-close to a T-out-of-L convSR-source $Y \in \{0,1\}^{m \cdot L}$ with $T = t - 1$, $L = \ell - 1$, $\varepsilon = 2^{-\Omega(n)}$, and output block length $m = \Omega(n)$. The only thing missing is that, as previously discussed, we wish to extract with similar parameters from SHELA sources whose honest blocks have arbitrarily small constant min-entropy rate (i.e., $k = \delta n$ for arbitrarily small constant $\delta > 0$). Notably, using a modified construction, we are able to transfer these ideal parameters to the "arbitrary constant min-entropy rate" setting. The only difference is that now $L = O(\ell)$.

Somewhere-extraction of SHELA source vs. weak source. We have already established

[3] We will show how to start from any public-coin 2-round WI proof system in the standard model which in turn means any non-interactive zero-knowledge proof system in the common random string model [7].

that we can deterministically extract high-quality convSR-sources from SHELA sources. However, an attentive reader might notice that deterministic somewhere extraction is also possible from *weak* sources. In fact, any strong seeded (k, ε)-extractor with seed length d yields a somewhere-extractor with error ε, $L = 2^d$ total output blocks, and $T = 1$ uniform blocks for weak sources with min-entropy at least k by considering a block for each possible fixing of the seed. This naive construction of a convSR-source is actually crucial in many constructions of multi-source extractors (we expand on this later in this section). However, it has strong limitations. In particular, even if we use an optimal strong seeded extractor, seed length lower bounds [11] imply that

$$L = \Omega\left(\frac{1}{\varepsilon^2}\right). \tag{1}$$

This means that if we require $\varepsilon = 2^{-\Omega(n)}$, then $L = 2^{\Omega(n)}$, which precludes any efficient cryptographic application of the resulting convSR-source.

Given the above shortcoming, one might wonder whether significantly better somewhere-extractors exist for weak sources. We dedicate part of our paper to the study of this problem. It turns out that the answer to this question is largely negative. In particular, a disperser-based lower bound shows that, similarly to the naive construction above, *every* somewhere-extractor for weak sources with error $\varepsilon = 2^{-\Omega(n)}$ and output block length $m = \Omega(n)$ must have $L = 2^{\Omega(n)}$ total output blocks.

In our work, we derive a set of lower bounds that complement each other and succeed in showing that somewhere-extractors for weak sources must perform significantly worse than the analogous objects for SHELA sources over various regimes of parameters. We are particularly interested in lower bounds on the total number of blocks of the output convSR-source, as this dictates the computational complexity blow-up suffered by a protocol when using this source. In the end, we put forth the conjecture that the above lower bound (1) actually holds for *every* somewhere-extractor (regardless of the output block length m), and we make some progress towards proving it.

Randomized algorithms and amplification of success probability using SHELA source vs. weak source. We remark that convSR-sources are well-suited for simulation of randomized algorithms whose outputs can be efficiently checked for correctness (e.g., searching for witnesses for the membership of some string in an NP language, or approximation algorithms for NP languages). In fact, one can simply run the algorithm using each block as its randomness. As a result, one obtains a few candidate solutions, and can efficiently check if at least one of them is correct. The success probability of the algorithm is thus amplified by the number of good (i.e., uniform) blocks.

It is well-known and easy to see that, in the procedure above, we do not need good blocks to be exactly uniformly distributed. Indeed, it is enough to rely on the weaker guarantee that good blocks are sufficiently close to uniform in statistical distance, say, $1/\mathrm{poly}(n)$-close, where n is some soundness parameter. We call this weaker family of sources *somewhere-amplifiable* (SA) sources, and denote the class of convex combinations of SA-sources as convSA-sources.

While weak sources can be used to efficiently produce convSA-sources, we show that this comes at a heavy price: Roughly speaking, if one wants to generate enough, and long enough, good blocks for appropriate and efficient success probability amplification, then the weak source needs to have very high min-entropy. Therefore, in many reasonable regimes of parameters, one is unable to extract suitable convSA-sources from weak sources, while one can extract high-quality convSR-sources (a stronger notion) from SHELA sources in those regimes. We refer to Section 6 for a more detailed discussion.

We conclude from the two discussions above that there is a fundamental separation between somewhere-extraction from SHELA and weak sources. Indeed, we are able to efficiently extract convSR-sources with much higher quality from a SHELA source than what we can obtain from a weak source.

Non-interactive witness indistinguishable proofs assuming public-coin ZAPs and relying on public SHELA sources. In a proof system, a prover proves to a verifier the veracity of some statement $x \in \mathcal{L}$ (where $\mathcal{L}$ is an NP-language). A soundness property guarantees that it is unlikely that an honest verifier accepts the proof of a false statement. When a proof system is non-interactive any verifier is able to check the validity of the proof. Non-interative proofs are therefore publicly verifiable and they are very appealing since the prover computes the proof once, while still it can be useful in many different cases (i.e., with many different verifiers). Non-interactive proofs are usually trivial to achieve since a prover could just send a witness proving membership in the language. The interesting case consists of offering some form of privacy for the secret (i.e., the witness) of the prover. We will in particular consider witness indistinguishability [12] that requires that the proof hides which witness has been used by the prover out of multiple witnesses. A special category of interactive proof systems is called "public coin" and refers to the role of the verifier that sends random strings only as messages. When there is only one message played by the verifier then a 2-round witness indistinguishable proof system is referred as ZAP[7]. The round of the verifier can be recycled among any polynomial number of proofs givens by provers. Since public-coin ZAPs exist, a natural question is whether the verifier can just be replaced by a sample from a high min-entropy source, therefore obtaining a non-interactive WI proof under the same computational assumptions of ZAPs and relying on the existence of SHELA sources. The answer is unfortunately negative. Indeed, consider the ZAP of [7]. The message of the prover consists of computing some non-interactive zero-knowledge (NIZK) proofs in the common random string model. In general, NIZK proofs (e.g., [12]) are not sound when the common random string is replaced by the output of high min-entropy sources. In turn, when trying to make a generic public-coin ZAP relying on a high min-entropy source non-interactive, soundness could be lost. Moreover, the issue with soundness remains also in case of parallel repetition since for some high min-entropy sources an accepting proof of a false statement can be produced with probability 1.

On the positive side, equipped with our constructive results about obtaining a convSR-source from a SHELA source, we show that assuming a public SHELA source, non-interactive witness indistinguishable proofs exist by just using a parallel repetition of any public-coin ZAP[4].

Non-interactive commitments from one-way functions and SHELA sources. In a commitment scheme, sender and receiver interact in a commitment phase so that the (even malicious) sender can later on show only one message consistent with such interaction, while the (even malicious) receiver has no specific advantage in detecting the message committed by the sender. The security property for the receiver is called "binding" while the security for the sender is called "hiding".

Non-interactive commitments guarantee that the sender has to work only once to produce a commitment of a message, while this commitment can be used to convince any receiver about the committed message. We focus on statistically binding commitments where, except with negligible probability, there is a unique message that is consistent with the transcript of the commitment phase, regardless of the computational power of the (even malicious) sender. A commitment scheme is "public coin" if the receiver sends only random strings.

Public-coin statistically binding commitment schemes in two rounds exist under the minimal assumption of the existence of any one-way function [13]. A natural question is whether, given any public-coin 2-round commitment scheme from one-way functions, the receiver can just be replaced by a sample from a high min-entropy source, therefore obtaining a non-interactive commitment scheme relying on the existence of SHELA sources[5]. We show that the answer is in general negative,

[4]Notice that we are considering generic weak sources and it is unknown whether such distributions can all be efficiently simulatable. Consequently we cannot obtain a non-interactive zero knowledge proof.

[5]We recall that obviously a SHELA source is also a high min-entropy source.

by providing a variation of the construction of [13] where the binding property breaks down when the first round is sampled from a specific SHELA source. Moreover, parallel repetitions do not help to obtain binding. The construction of [13] can become non-interactive using any SHELA source, however in this last case there is a price to pay in communication complexity since the size of the resulting non-interactive commitment scheme is equal to the size of the SHELA source X.

The real good news come from using our tool: a convSR-source extracted from a SHELA source (without adding any computational assumption). Indeed, in this case we can get a non-interactive statistically binding commitment scheme just by running a parallel repetition of any public-coin 2-round statistically binding commitment scheme. When applied to the scheme of [13], we can get better communication complexity compared to the previously described approach that consists of using a SHELA source directly. Indeed, consider a 2-round statistically binding commitment scheme where the first round of the receiver (in the commitment phase) consists of λ bits, and let us assume that in each high min-entropy honest block of a 2-out-of-ℓ SHELA there are k bits of min-entropy, where $k \gg \lambda$. If $Y = \mathsf{SomeExt}(X) \in \{0,1\}^{m \cdot L}$ for $L = \ell - 1$ and we set $m = \lambda$ (by truncation), then $|Y| = m \cdot L \ll n \cdot \ell = |X|$. Therefore, with the parameters discussed above, if we instantiate the scheme of [13] using X directly, the resulting non-interactive commitment scheme has significantly worse communication complexity than the one built from the convSR-source.

Additional contributions. In this work, we also consider an *online* variant of a SHELA source. This modified source admits a stronger adversarial model. Recall that in a standard SHELA source the adversary must decide the positions of the honest blocks a priori, i.e., before any honest block is sampled. In contrast, the adversary in an online SHELA source is allowed to decide whether the i-th block should be honest or corrupted based on the values of the first through $(i-1)$-th blocks that have already been generated, and constrained by the fact that there must be at least t honest blocks in the source. In particular, the position of the second honest block may depend on the value of the first honest block, which is not possible in the standard SHELA model. Notably, we show that under this stronger adversarial setting our extraction procedure still works and outputs a convSR-source.

1.2 Related Work

Applications of convSR-sources in pseudorandomness. We would like to point out that convSR-sources are also very useful in a context different than those already presented, and many previous works on pseudorandomess exploit their structure. Indeed, convSR-sources are key intermediate objects in several constructions of multi-source and non-malleable randomness extractors for weak sources. A central approach in such constructions is to reduce the task of extracting a uniform string from independent weak sources to that of extracting such a string from one or more independent convSR-sources potentially satisfying a few additional properties, sometimes coupled with additional independent weak sources or small uniform seeds.

The connection between multi-source extraction and convSR-sources has been known since they were first defined [14]. convSR-sources have also been used in early constructions of seeded extractors [15].

Barak et al. [16] and Raz [17] showed how to convert two independent weak sources into an convSR-source with few blocks. This reduction was then used directly to obtain 3- and 4-source extractors with constant error. Such an approach has also proved useful in the construction of dispersers [16, 18].

To obtain extractors for a constant number of sources with lower error and min-entropy requirement $n^{\Omega(1)}$, Rao [19] transforms independent input sources into several independent *aligned* convSR-sources, i.e., there is at least one position at which all convSR-sources have a uniform

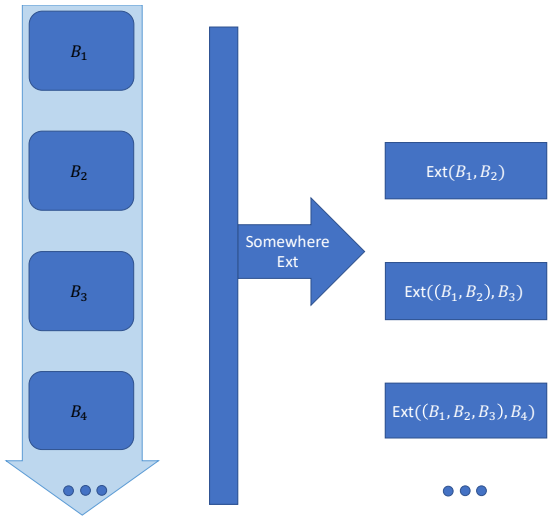

Figure 2: The somewhere-extraction procedure for a SHELA source with honest blocks with high min-entropy. In the diagram, $B_1, \ldots, B_4, \ldots$ are blocks of a single SHELA source, and $\mathsf{Ext}(X, S)$ is a strong seeded extractor with seed S and input source X.

block. If the number of blocks in each convSR-source is not too large, then an iterative procedure succeeds in extracting a uniform string from such independent aligned convSR-sources with small error. Li [20] also used a similar approach with aligned convSR-sources to obtain better 3-source extractors.

An important step in many recent constructions of 2- and 3-source extractors [21, 22, 23, 10, 24, 25] consists in generating convSR-sources with many "good" blocks (i.e., blocks close to uniform) which additionally satisfy a notion of w-wise independence for an appropriate parameter w: Every set of w good blocks is also close to jointly uniformly distributed. convSR-sources are also used in other recent constructions of multi-source extractors [26, 27].

The usefulness of convSR-sources extends to more recent notions of randomness extraction. In fact, convSR-sources have been used in the construction of seedless non-malleable extractors [28] for weak sources, which are closely connected to non-malleable codes.

The ubiquity of convSR-sources (generated from weak sources) in extractor constructions provides one more compelling reason for our study of lower bounds for deterministic somewhere-extraction from weak sources.

Finally, we should mention that, because of the close connection between convSR-sources and randomness extraction from general weak sources, several works other than those already mentioned have focused directly on designing randomness extractors for the restricted class of convSR-sources [29, 30, 31, 32, 33]. Such extractors are usually called *mergers*.

Deterministic randomness extraction from restricted classes of sources. Our work is also related to the fundamental and well-studied problem of deterministic randomness extraction. Given the impossibility of deterministic extraction from general weak sources, the following natural question arises: *Under which conditions is deterministic randomness extraction possible from imperfect sources of randomness?*

Several works (some even predating the definition of weak sources [1]) have studied this question from various perspectives. Some works have considered deterministic randomness extraction from

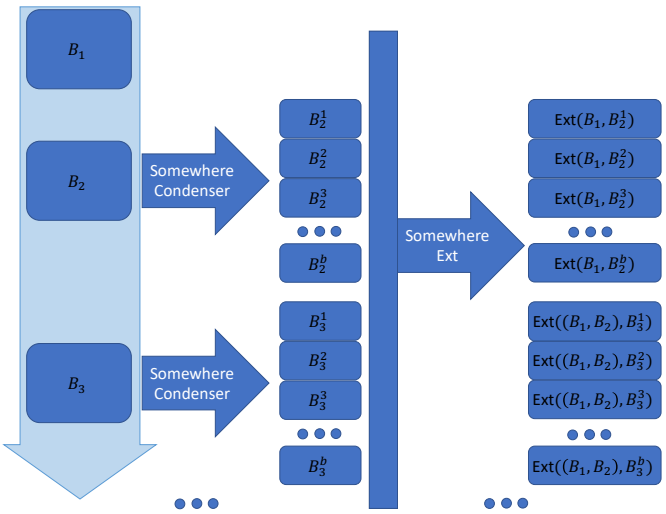

Figure 3: The simplified somewhere-extraction procedure for a SHELA source with honest blocks with low min-entropy. In the diagram, $B_1, \ldots, B_3, \ldots$ are blocks of a single SHELA source, and $\mathsf{Ext}(X, S)$ is a strong seeded extractor with seed S and input source X. Please note that, for the sake of clarity, the diagram is a slight simplification of our construction.

streams of bits generated i.i.d. with unknown bias [34, 35], or according to a Markov chain [36]. In a parallel line of research, settings where some input bits may be (adversarially or not) fixed, while the remaining ones are random, have also been considered [37, 38, 39, 40, 41, 42, 43, 44, 45, 46]. Other classes of sources considered in the context of deterministic extraction include sources with efficient sampling procedures [47, 48] or sampled in small space [49], sources defined over subspaces [50, 51, 44, 52, 53, 54, 55, 24], sources determined by zero sets of polynomials [56, 57], sources sampled by Turing machines [58] or small circuits [59], and sets of independent weak sources (already discussed in this section). Some works have constructed such extractors for subclasses of Santha-Vazirani sources [60, 61], which are known not to admit deterministic extraction in general. We note that Bentov, Gabizon, and Zuckerman [62] studied deterministic randomness extraction from the blockchain of Bitcoin, which has some connections to our model. However, their focus is on standard deterministic extraction, instead of somewhere-extraction. They show that standard deterministic extraction is impossible against an adversary with an unbounded budget, and then study the same problem against a "budget-constrained" adversary.

Although we are not dealing with standard randomness extraction like most of the works above, we present a result of a similar flavor: The restricted (and practically motivated) class of SHELA sources allows for deterministic *somewhere*-extraction with much better parameters than the class of weak sources.

Randomness extraction from adversarial sources. Subsequently to the announcement of our work, the problem of extracting randomness from adversarial sources (of which SHELA sources are an example) has received significant attention.

Chattopadhyay, Goodman, Goyal, and Li [63] study randomness extraction from an adversarial source model similar to SHELA sources. However, there are important distinctions between the two models, which we discuss next. In both cases, a source can be divided into blocks, some of which are

independently generated and contain appropriate min-entropy, while other blocks are adversarially controlled. However, in SHELA sources the adversarial block is allowed to depend arbitrarily on all previous blocks (but *not* on subsequent blocks), while in [63] is only allowed to depend on at most d other arbitrary blocks for a small "locality parameter" d. Deterministic randomness extraction turns out to be possible in the adversarial model from [63], while it is impossible in the SHELA model and we instead study deterministic *somewhere*-extraction and its applications. Based on this, the results in these two models are incomparable.

Dodis, Vaikuntanathan, and Wichs [64] study seeded randomness extraction from so-called *extractor-dependent* sources. This adversarial model differs significantly from SHELA sources. At a very high level, a source is sampled by an adversary that is first allowed to query the extractor on different inputs with the same seed, with the condition that the source contains enough min-entropy and other sensible constraints to make the problem non-trivial. Extractor-dependent sources aim to capture scenarios where a random seed may be re-used several times.

1.3 Technical Overview on Deterministic Somewhere-Extraction from SHELA and Weak Sources

Impossibility of deterministic extraction from SHELA sources. We show that if at most a γ-fraction of the ℓ blocks in a SHELA source are honest, where $\gamma \in [0, 1)$ is an *arbitrary* constant, and ℓ is a large enough constant depending on γ, then deterministic randomness extraction is impossible from this class of SHELA sources. Notably, this impossibility result holds even if we allow the honest blocks to be *uniformly distributed*, instead of only requiring them to have enough min-entropy.

This result is obtained by reducing the problem of deterministic extraction from SHELA sources to the problem of deterministic extraction from so-called *resettable* sources, introduced in [62]. In the same work, the latter problem has been shown to be closely related to deterministic extraction from Santha-Vazirani (SV) sources [65], which is widely known to be an impossible task. For more details we refer to Section 3.1.

Constructions of somewhere-extractors for SHELA sources. Our constructions of somewhere-extractors for SHELA sources are mainly based on the following trick, which we illustrate for a SHELA source with three blocks B_1, B_2, B_3, two of which are honest. If we applied the naive somewhere-extractor previously discussed with a 2-source extractor, we would obtain a convSR-source with three rows. Recall that one of our main goals is to reduce the total number of blocks in the resulting convSR-source as much as possible due to efficiency concerns. With this in mind, instead of applying the naive somewhere-extractor, we can notice that there are two cases:

- B_3 is honest. Then, B_3 and (B_1, B_2) are two independent weak sources. This means we can extract randomness from the two sources (B_1, B_2) and B_3;

- B_3 is not honest. Then, B_1 and B_2 are honest, and hence are independent weak sources. In this case, we can extract randomness from the two sources B_1 and B_2.

For the sake of this example, let Ext_1 and Ext_2 be two-source extractors, and compute $\mathsf{Ext}_1((B_1, B_2), B_3)$ and $\mathsf{Ext}_2(B_1, B_2)$.[6] The key observation, stemming from the two cases above, is that we are guaranteed that at least one of the two outputs is close to uniformly distributed. As a result, we obtain a convSR-source with two rows instead of three.

[6] In reality, we are able to use strong seeded extractors (for which we know much better explicit constructions) in place of two-source extractors. This is due to the disproportion in the size of the sources. In fact, the size of one of the sources given to the extractor grows linearly with the total number of blocks.

As already mentioned, we design explicit somewhere-extractors in two main settings. Our first, simpler, somewhere-extractor can be applied whenever the underlying SHELA source has $t \geq 2$ honest n-bit blocks with min-entropy $k = (1-\gamma)n$ for a small enough constant $\gamma > 0$. The construction is a generalization of the reasoning we presented for three blocks above, and a diagram is presented in Figure 2. It proceeds by iteratively using a strong seeded extractor to extract randomness from ever-growing sequences of blocks (using another block as a seed). A bit more precisely, if $X \in \{0,1\}^{n \cdot \ell}$ is a SHELA source and $X = (B_1, B_2, \ldots, B_\ell)$, then for every $i = 2, 3, \ldots, \ell$ we consider

$$B'_i = \mathsf{Ext}_i((B_1, \ldots, B_{i-1}), B_i), \tag{2}$$

where $(B_1, \ldots, B_{i-1})$ acts as the input weak source, B_i acts as the seed, and Ext_i is an appropriate strong seeded extractor. Then, we set $\mathsf{SomeExt}(X) = (B'_2, \ldots, B'_\ell)$. A diagram of the construction can be found in Figure 2. The first problem we run into is that in usual applications of seeded extractors, the seed is uniformly distributed. This is not the case here, since, even if B_i is an honest block, it is only guaranteed to have min-entropy $(1-\gamma)n$. However, it is not hard to show, using the strongness of the extractor, that using a source with high min-entropy as the seed is sufficient. Another issue we encounter is that we are reutilizing many SHELA blocks when computing output blocks via (2). This appears to be at odds with the requirement that good output blocks should be close (in statistical distance) to independent and uniformly distributed. A careful conditioning argument, again exploiting the strangeness of the extractor, shows that independence and uniformity are actually attained with small error. In fact, whenever B_i is honest and there is an honest block in $(B_1, \ldots, B_{i-1})$, we succeed in generating (with small error) a new good block of the output convSR-source. Instantiating this construction with the nearly-optimal GUV strong seeded extractor [66] and assuming the SHELA source $X \in \{0,1\}^{n \cdot \ell}$ has t honest blocks, we output a distribution $Y \in \{0,1\}^{m \cdot L}$ that is $(t \cdot 2^{-\Omega(n)})$-close to a T-out-of-L convSR-source with $m = \Omega(n)$. Moreover, from the discussion above it follows that $L = \ell - 1$ and $T = t - 1$.

In the second setting, we consider deterministic somewhere-extractors for SHELA sources with honest blocks having *arbitrary* constant min-entropy rate k/n. In other words, we allow the min-entropy requirement k of honest blocks to satisfy $k = \delta n$ for arbitrarily small $\delta > 0$. Notably, in this significantly harder setting we are able to obtain essentially the same parameters as the somewhere-extractor for the high min-entropy setting detailed above. In fact, all parameters remain unchanged, except that now we cannot guarantee that $L = \ell - 1$, and instead have the (still highly desirable) relationship $L = O(\ell)$. The main barrier towards making the previous construction work in this setting is that if honest blocks do not have high min-entropy, they can no longer be used as seeds for strong seeded extractors. This issue is surpassed by using the somewhere-condenser for weak sources from [17, 16]. Intuitively, a somewhere-condenser is to a randomness condenser as a deterministic somewhere-extractor is to an extractor. On input a weak source with low min-entropy, the somewhere-condenser $\mathsf{SomeCond}$ outputs (with small error) a constant number of (sufficiently long) blocks with the guarantee that at least one block has very high min-entropy rate. Because the focus is not on extraction of *perfect* randomness, somewhere-condensers for weak sources are allowed to have much better parameters than somewhere-extractors for the same class of sources. We modify the construction for honest blocks with high min-entropy above by adding a first step of somewhere-condensation for each block of the input SHELA source, as detailed in Figure 3. We show that our somewhere-extractors designed for SHELA sources can also be applied to *online* SHELA sources as is to extract convSR-sources (for full definitions and discussion please see Section 4).

Lower bounds for deterministic somewhere-extraction from weak sources. We consider the natural problem of understanding the performance of somewhere-extractors for weak sources,

and derive a set of lower bounds which show that, particularly for parameters relevant to cryptographic applications, *every* somewhere-extractor (regardless of efficiency) for weak sources must have significantly worse parameters than the somewhere-extractors we obtain for the class of SHELA sources. As previously discussed, these negative results for weak sources are strong enough that they preclude the use of convSR-sources generated from weak sources in efficient cryptographic protocols.

Suppose $\mathsf{SomeExt} : \{0,1\}^{\tilde{n}} \to \{0,1\}^{m \cdot L}$ is a somewhere-extractor for $(\tilde{n}, k)$-sources[7]. We begin by noting that a simple reasoning analogous to the proof of impossibility of deterministic extraction from weak sources immediately shows that $L = \Omega(\tilde{n} - k)$. Our first non-trivial lower bound is obtained by relating a somewhere-extractor to a *disperser* (for weak sources). Roughly speaking, a disperser is a fundamental pseudorandom object that transforms a weak source and a short uniform seed into an output distribution that hits every appropriately large subset of the output space with non-zero probability. Optimal seed length lower bounds are known for dispersers [11]. We show that if $\mathsf{SomeExt} : \{0,1\}^{\tilde{n}} \to \{0,1\}^{m \cdot L}$ is a somewhere-extractor for $(\tilde{n}, k)$-sources with error ε, then the function $G : \{0,1\}^{\tilde{n}} \times [L] \to \{0,1\}^m$ given by

$$G(x, i) = \mathsf{SomeExt}(X)_i$$

is a disperser with seed length $\log L$ and error ε. This immediately leads to a lower bound on the number L of output blocks of SomeExt (excluding a minor technicality that does not affect the quality of the lower bound),

$$L = \Omega\left(\frac{\tilde{n} - k}{\max(\varepsilon, 2^{-m})}\right). \tag{3}$$

This means, as discussed in more detail in Section 5, weak sources behave exponentially worse than comparable SHELA sources for somewhere-extraction in the linear output block length regime.

Note that the two lower bounds in the previous paragraph do not give anything when $k \approx \tilde{n}$ and m is small. This naturally leads us to consider lower bounds for L in an extreme 1-bit block setting with $k = \tilde{n} - 1$ and $m = 1$. Although we do not obtain a lower bound for extraction of convSR-sources in this extreme regime, we are able to prove a non-trivial lower bound that scales with the error for the harder, but related, task of extracting an SR-source from a weak source (*not* a convex combination of SR-sources as before). Note that, in particular, the naive somewhere-extractor obtained by enumerating the seed of a strong extractor satisfies this property. To be precise, we show that in this setting we must have

$$L = \Omega\left(\log\left(\frac{1}{\max(\varepsilon, 2^{-k})}\right)\right). \tag{4}$$

The lower bound in (4) is obtained by an adaptive version of the basic argument for the impossibility of deterministic extraction from weak sources. Given a candidate function $F : \{0,1\}^{\tilde{n}} \to \{0,1\}^L$, our goal is to show the existence of a weak source $X^\star$ with enough min-entropy such that *every* bit $F(X^\star)_i$ is sufficiently biased. We begin by setting $X_0^\star$ to be uniformly distributed over $\{0,1\}^{\tilde{n}}$, and analyze its performance w.r.t. F. If $F_i(X_0^\star)$ is the first bit close to uniform, we remove an appropriate set of elements from the support of $X_0^\star$ to obtain $X_1^\star$ such that $F_i(X^\star)$ biased enough. Then, we repeat the reasoning with the new source $X_1^\star$ and so on, until every bit is biased[8]. Then,

[7]The set of $(\tilde{n}, k)$-sources consists of all weak sources over $\{0,1\}^{\tilde{n}}$ with min-entropy at least k. We use $\tilde{n}$ to avoid confusion with the block length of SHELA sources.

[8]When biasing the next coordinates, we have to be careful not to 'spoil' biases of previous coordinates. This results in the log factor in the bound.

L must be large enough to ensure the outcome $X^\star$ of this process has too small support (and hence does not satisfy the min-entropy requirement of F), which yields the lower bound.

With these bounds in mind, it is natural to consider whether arguments that yield lower bounds of this type on the seed length of extractors, more precisely the granularity argument of Nisan and Zuckerman [67, Theorem 3] and the techniques due to Radhakrishnan and Ta-Shma [11, Section 2.2], could be extended to the setting of somewhere-extraction. Unfortunately, such arguments crucially rely on the ability of picking a seed at random: There, one is only worried about showing that the bias is large enough *on average*, while we must show that the bias is large enough *for every choice of the seed*[9].

1.4 Technical Overview on Non-Interactive Proof Systems and Commitments from Public SHELA Sources

Non-interactive (publicly verifiable) witness indistinguishable proof system. We will now describe how to construct a non-interactive (and therefore publicly verifiable) Witness Indistinguishable (WI) proof system Π_{pv} from a public SHELA source X and starting with the existence of a public-coin ZAP Π. Π_{pv} works as follows: The prover of Π_{pv} receives X and runs the somewhere-extractor $\mathsf{SomeExt}$ on X to obtain $(R_1, \ldots, R_L)$. Then, the prover on input the witness w for the statement x computes a second-round π_i from Π using R_i for $i = 1, \ldots, L$. The verifier of Π_{pv}, having access to X, also computes $(R_1, \ldots, R_L) = \mathsf{SomeExt}(X)$, and accepts the proof only if all pairs (R_i, π_i) are accepting by the verifier of Π w.r.t. the statement x. Observe that WI of Π is preserved under parallel composition and holds even when the first round of Π is chosen by a malicious verifier. Therefore, Π_{pv} also enjoys the WI property. The soundness of Π_{pv} is based on the observation that T blocks of $(R_1, \ldots, R_L)$ are negligibly close to a uniform distribution over $\{0,1\}^m$. Denote them by $R_{I_1}, \ldots, R_{I_T}$. Then, the soundness of Π ensures that a malicious prover could not cheat when the second round of Π is computed w.r.t. $R_{I_1}, \ldots, R_{I_T}$.

As a result, using known constructions of public-coin ZAPs, we are able to construct a non-interactive WI proof system from trapdoor permutations that requires as a set-up a SHELA source only. Notice that a SHELA source is a CRS that can be corrupted (in a natural, structured manner) by an unbounded adversary. Still, we assume that the adversarial verifier can run only in polynomial time to distinguish the witness, even though he does not have such restriction when affecting the sample from the public SHELA source. Previous constructions of non-interactive WI proof systems either require a common random string as set-up, or were based on specific number-theoretic hardness assumptions in bilinear groups [68, 69], or on indistinguishability obfuscation and one-way permutations [70].

From another point of view, one can see our result as a Non-Interactive (NI) WI proof system where the soundness and the WI property hold even when the set-up phase is partially generated by the adversary. We note that the work of [71] investigates if soundness and WI of a NIWI proof system hold even when the adversary takes complete control of the set-up phase. They achieve a positive result relying on some specific number-theoretic assumption in bilinear groups. Instead, our NIWI proof system can be instantiated from trapdoor permutations and the adversary has only a partial control over the set-up.

Notice that [2] studies cryptographic protocols with simulatable security by considering a simulatable CRS drawn from a high min-entropy distribution. In this work we do not assume that public sources of randomness are simulatable and we do not investigate simulatable security. Our CRS is not a generic min-entropy string but instead corresponds to a structured min-entropy source

[9] By *seed* we mean i in $F_i(X^\star)$.

that is partially controlled by an unbounded adversary.

Given the above construction of a non-interactive WI proof system Π_{pv}, one could argue that a convSA-source suffices for constructing Π_{pv}. Recall that a convSA-source is a convex combination of T-out-of-L SA-sources, which consist of L blocks, T of which are independent and $\frac{1}{\mathrm{poly}(n)}$-close to uniform in statistical distance, where n is some relevant security parameter. This is because the soundness of the protocol can be amplified by using the T "good" blocks, which correspond to independent parallel repetitions of the underlying protocol Π.

In order to adequately compare the performance of the protocol under convSA-extraction from weak sources and convSR-extraction from SHELA sources, we compare a t-out-of-ℓ SHELA source $X \in \{0,1\}^{n \cdot \ell}$ with honest blocks having linear min-entropy k' with an arbitrary weak $(\tilde{n} = n \cdot \ell, k = k' \cdot t)$-source $\tilde{X}$. We are able to show that convSR-sources extracted from X are much better suited for applications than convSA-sources generated from $\tilde{X}$ in two aspects:

1. **Efficiency:** The efficiency of Π_{pv} depends on L. It is not hard to see that every convSA-source extractor for weak sources $\tilde{X}$ must have $\Omega(\tilde{n}) = \Omega(n \cdot \ell)$ total output blocks (even if we only require constant error). On the other hand, we can extract convSR-sources from X with only $O(\ell)$ blocks.

2. **Security:** Let us assume that Π requires a first round of $m = \Omega(k')$ bits. Then, we show that every *efficient, low-error* convSA-source extractor for weak sources outputs at most $T = O(k/m) = O(k' \cdot t/m)$ *good* blocks of length m. As a result, if t is constant, it follows that such an extractor only outputs $T = O(1)$ good blocks. This is not enough to successfully amplify the soundness of the protocol. Finally, we note that if we build our Π_{pv} starting from a convSR-source extracted from a t-out-of-ℓ SHELA source with constant t, the analysis of soundness described in this subsection holds, and therefore Π_{pv} is sound.

Improving the efficiency of [6]. We note that the work of [6] constructs a publicly verifiable proof system from any blockchain under some assumptions on the min-entropy of honestly generated blocks. Notably, under the same assumptions the blockchain can be used to implement also an online SHELA source. In [6], the authors construct a publicly verifiable proof system by applying the naive somewhere-extractor NaiveSomeExt (that we discussed earlier) to extract a convSR-source from the blockchain. Therefore our somewhere-extractor SomeExt (instead of NaiveSomeExt) could be used in their work to immediately improve the efficiency of their proof system. More details are provided in Section 7.4.

Non-interactive statistically binding commitments. We introduce now a construction of non-interactive statistically binding commitments from a public SHELA source relying on one-way functions. This is achieved by making use of any two-round public-coin commitment scheme Π_{com} from one-way functions.

First of all we remark that one can not simply replace the first round of Π_{com} with a sample from a source with linear min-entropy (say, min-entropy $0.5n$). Indeed, start from Π_{com} and consider a scheme Π'_{com} where: a) the random string played as first round of Π_{com} must be twice in length, and b) the sender ignores the first half of the first round and continues as in Π_{com} using the second half. It is straightforward to see that Π'_{com} is a 2-round public-coin statistically binding commitment scheme from any one-way functions. If we replace the first round of Π'_{com} with the output of a linear min-entropy source we might have that the entire min-entropy is in the first half of the first round and is therefore wasted completely. The malicious sender could therefore violate binding since it would end up running Π_{com} on input a first round with zero min-entropy! Obviously in this case parallel repetition does not help.

We now proceed to describe how our scheme Π_{compv} works starting with any 2-round public-coin statistically binding commitment scheme (including the above Π'_{com}). Moreover, Π_{compv} can be run with efficient parameters because of the use of $\mathsf{SomeExt}$.

Our commitment scheme Π_{compv} works as follows: First, the sender runs the somewhere-extractor $\mathsf{SomeExt}$ on the public SHELA source X, obtaining $\mathsf{SomeExt}(X) = (R_1, \ldots, R_L)$. Then, the sender on input the message m and R_i (used as the receiver's first round) computes a commitment com_i and the opening information dec_i using the sender of Π_{com}, for $i = 1, \ldots, L$. In the opening phase, the receiver on input $\mathsf{dec}_1, \ldots, \mathsf{dec}_L$ having access to X computes $(R_1, \ldots, R_L) = \mathsf{SomeExt}(X)$, and outputs the message m only if it holds that for all $i = 1, \ldots, L$ the message committed in com_i is m. Hiding of our scheme holds from the observation that hiding is preserved under parallel composition and when the first round of Π_{com} is chosen by a malicious receiver. The binding of Π_{compv} is based on the observation that at least T blocks $R_{I_1}, \ldots, R_{I_T}$ are negligibly close to a uniform distribution over $\{0,1\}^m$. This implies that there are at least T commitments computed w.r.t. a good block R_{I_j} that is statistically close to a first round sent by a receiver of Π_{com}. Therefore, from the statistically binding of Π_{com} it follows that a malicious sender could not cheat when the commitment is computed w.r.t. $R_{I_1}, \ldots, R_{I_T}$.

1.5 Open Questions

We present some interesting directions for future research:

- Prove (or disprove) Conjecture 46.
- Given any SHELA or convSR source, we can define its *rate* as number of good[10] blocks divided by total number of blocks. Our constructions from Section 3.2 transform SHELA sources with rate t/ℓ into convSR-sources with rate $\frac{t-1}{\ell-1} \leq \frac{t}{\ell}$. We conjecture that the rate of the output convSR-source cannot be larger than t/ℓ.
- Find good bounds on the number of output blocks of convSA-source extractors for weak sources.

1.6 Organization of the Paper

We introduce relevant notation and definitions in Section 2. SHELA sources are defined in Section 3, and deterministic somewhere-extractors are presented in Section 3.2. Lower bounds for somewhere-extraction are studied in Section 5, and the limits of SA-source extraction are considered in Section 6. Detailed arguments, along with standard definitions and lemmas, have been deferred to the supplementary material.

2 Preliminaries and definitions

2.1 Notation

Sets are usually denoted by calligraphic letters such as $\mathcal{S}$ and $\mathcal{I}$. Random variables are usually denoted by uppercase letters such as X, Y, and Z. We may identify a random variable X with its distribution. The support of a distribution X is denoted by $\mathsf{supp}(X)$. We denote the uniform distribution over $\{0,1\}^m$ by U_m. We may write $X \sim Y$ to denote that X has the same distribution

[10]For a SHELA source, a good blocks correspond to honest blocks, while they correspond to jointly uniform blocks in convSR-sources.

as Y. All logarithms log are taken to base 2. The Shannon entropy of a distribution X is denoted by $H(X)$, and we denote the binary entropy function by h. The notation $\mathrm{poly}(n)$ denotes an arbitrary polynomial in n. We denote a negligible function of a parameter n by $\mathsf{negl}(n)$.

2.2 Statistical Distance and Min-Entropy

In this section, we define statistical distance and min-entropy, along with useful results.

Definition 1 (Statistical distance)**.** *Given two distributions X and Y over a set $\mathcal{X}$, the* statistical distance between X and Y*, denoted by $\Delta(X;Y)$, is defined as*

$$\Delta(X;Y) = \max_{\mathcal{S}\subseteq\mathcal{X}} |\Pr[X \in \mathcal{S}] - \Pr[Y \in \mathcal{S}]| = \frac{1}{2}\sum_{x\in\mathcal{X}} |\Pr[X = x] - \Pr[Y = x]|.$$

*We may write $\Delta(X;Y|Z)$ as shorthand for $\Delta(X,Z;Y,Z)$, and say that X and Y are ε-*close*, also written $X \approx_\varepsilon Y$, if $\Delta(X;Y) < \varepsilon$. For a random variable $X \in \{0,1\}$, we informally call $\Delta(X;U_1) = |\Pr[X = 1] - 1/2|$ the* bias *of X.*

The following is a well-known result about couplings and statistical distance.

Lemma 2. *Given two distributions X and Y, it holds that*

$$\Delta(X;Y) = \inf\{\Pr[P \neq Q] : P \sim X, Q \sim Y\}.$$

Moreover, there is a coupling (P,Q) with $P \sim X$ and $Q \sim Y$ such that $\Delta(X;Y) = \Pr[P \neq Q]$.

Definition 3 (Min-entropy)**.** *Given a distribution X over $\mathcal{X}$, the* min-entropy of X*, denoted by $\mathbf{H}_\infty(X)$, is defined as*

$$\mathbf{H}_\infty(X) = -\log\left(\max_{x\in\mathcal{X}} \Pr[X = x]\right).$$

Definition 4 (Conditional min-entropy)**.** *Given distributions X and Z, we define the* conditional min-entropy of X given Z*, denoted by $\widetilde{\mathbf{H}}_\infty(X|Z)$, as*

$$\widetilde{\mathbf{H}}_\infty(X|Z) = -\log\left(\mathbb{E}_{z\leftarrow Z}\left[\max_{x\in\mathcal{X}} \Pr[X = x|Z = z]\right]\right).$$

We state a fundamental property of the conditional min-entropy.

Lemma 5 ([72])**.** *Given distributions X, Y, and Z where $|\mathsf{supp}(Y)| \leq 2^\lambda$, we have that*

$$\widetilde{\mathbf{H}}_\infty(X|Y,Z) \geq \widetilde{\mathbf{H}}_\infty(X,Y|Z) - \lambda \geq \widetilde{\mathbf{H}}_\infty(X|Z) - \lambda.$$

The following connection between the statistical distance and the Shannon entropy, $H(\cdot)$, will be useful.

Lemma 6 ([73, Theorem 6])**.** *Fix $X, Y \in \{0,1\}^m$ such that $X \approx_\varepsilon Y$ for some $\varepsilon \leq 1 - 2^{-m}$. Then, it holds that*

$$|H(X) - H(Y)| \leq h(\varepsilon) + \varepsilon m,$$

where $h(\varepsilon) = -\varepsilon\log\varepsilon - (1-\varepsilon)\log(1-\varepsilon)$ is the binary entropy function.

2.3 Weak Sources and Extractors

Two basic objects in pseudorandomness, which have been extensively studied in the literature, are weak (n,k)-sources and (seeded) extractors. We proceed to define both.

Definition 7 ((n,k)-source)**.** *A distribution X over $\{0,1\}^n$ is said to be an* (n,k)-source *provided that $\mathbf{H}_\infty(X) \geq k$. Furthermore, we say an (n,k)-source is* flat *if it is uniformly distributed over a set of size at least 2^k.*

Definition 8 (Extractor)**.** *A function $\mathsf{Ext} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ is said to be a* strong (k,ε)-extractor *if*

$$\mathsf{Ext}(X,U_d),U_d \approx_\varepsilon U_m,U_d$$

for every (n,k)-source X and U_d independent of X.

Moreover, a function $\mathsf{Ext} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ is said to be an average-case strong (k,ε)-extractor *if*

$$\mathsf{Ext}(X,U_d),U_d,W \approx_\varepsilon U_m,U_d,W$$

for every X and W such that $\widetilde{\mathbf{H}}_\infty(X|W) \geq k$, and U_d independent of X and W.

Average-case strong and strong extractors are closely related, as evidenced by the following well-known result.

Lemma 9. *If Ext is a strong (k,ε)-extractor, then Ext is an average-case strong $(k+\log(1/\eta),\varepsilon+\eta)$-extractor for every $\eta > 0$.*

The following tight lower bound on the seed length of an extractor will be relevant when discussing lower bounds for somewhere-extraction from weak sources.

Lemma 10 ([11])**.** *Let $\mathsf{Ext} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ be a (k,ε)-extractor. If $\varepsilon \leq 1/2$, then $d \geq \log(n-k) + 2\log(1/\varepsilon) + O(1)$.*

We now describe an explicit strong extractor with near-optimal parameters that will be useful when instantiating our constructions of somewhere-extractors for SHELA sources.

Lemma 11 ([66])**.** *For every constant $\alpha > 0$, all $k \leq n$ and $\varepsilon > 0$ there exists an explicit strong seeded (k,ε)-extractor $\mathsf{Ext} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ with $d = \log n + O(\log(k/\varepsilon))$ and $m = (1-\alpha)k$. The hidden constant in the expression for d depends on α.*

We will need to handle cases where the seed for the strong extractor is not uniform, but rather it is only known to have high min-entropy. The following simple lemma states that strong extractors still work under such imperfect seeds, provided not much min-entropy is lost.

Lemma 12. *Let $\mathsf{Ext} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ be an average-case strong (k,ε)-extractor, X and W such that $\widetilde{\mathbf{H}}_\infty(X|W) \geq k$, and S an independent (d,k')-source. Then, we have*

$$\mathsf{Ext}(X,S),S,W \approx_{\varepsilon \cdot 2^{d-k'}} U_m,S,W.$$

Proof. Since every (d,k')-source is a convex combination of flat sources with min-entropy k', without loss of generality we may assume that S is uniformly distributed over a set $\mathcal{S} \subseteq \{0,1\}^d$ of size $2^{k'}$. Since Ext is an average-case strong (k,ε)-extractor, we have

$$\Delta(\mathsf{Ext}(X,U_d);U_m|U_d,W) = \sum_{s \in \{0,1\}^d} 2^{-d}\Delta(\mathsf{Ext}(X,s);U_m|W) \leq \varepsilon.$$

This means that

$$\begin{aligned}\varepsilon &\geq \sum_{s\in\mathcal{S}} 2^{-d}\Delta(\mathsf{Ext}(X,s);U_m|W)\\ &= 2^{k'-d}\sum_{s\in\mathcal{S}} 2^{-k'}\Delta(\mathsf{Ext}(X,s);U_m|W)\\ &= 2^{k'-d}\Delta(\mathsf{Ext}(X,S);U_m|S,W),\end{aligned}$$

and so the desired result follows. □ □

2.4 Dispersers

In this work we will also exploit properties of another pseudorandom object, called a *disperser*, when proving lower bounds for somewhere-extraction. The original motivation behind these objects is the simulation of randomized algorithms with one-sided error from weak randomness.

Definition 13 (Disperser)**.** *A function* $\mathsf{Disp} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ *is said to be a* (k,ε)-disperser *if*

$$\Pr[\mathsf{Disp}(X,U_d) \in \mathcal{S}] > 0$$

for every (n,k)*-source* X *independent of* U_d *and every set* $\mathcal{S} \subseteq \{0,1\}^m$ *of size* $|\mathcal{S}| \geq \varepsilon 2^m$*.*

The following tight lower bound on the seed length of dispersers will be useful.

Lemma 14 ([11])**.** *Let* $\mathsf{Disp} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ *be a* (k,ε)*-disperser. If* $2^{-m} \leq \varepsilon \leq 1/2$ *and* $2^d \leq \frac{(1-\varepsilon)2^m}{2}$ *(i.e.,* ε *is not trivial), then* $d \geq \log(n-k) + \log(1/\varepsilon) + O(1)$*.*

2.5 Somewhere-Random Sources and Somewhere-Extractors

In this section, we define SR- and convSR-sources, along with the notion of a deterministic somewhere-extractor and a basic result. Standard notions such as statistical distance, min-entropy, weak $(\tilde{n},k)$-sources, and extractors are defined in Section 2.2.

Definition 15 (Somewhere-random source)**.** *A distribution* $X = (X_1,\ldots,X_L)$ *over* $\{0,1\}^{m\cdot L}$ *is said to be a* (T,L,m)-somewhere-random source*, SR-source in short, if there exist indices* $i_1 < i_2 < \cdots < i_T$ *such that the tuple* $(X_{i_1},X_{i_2},\ldots,X_{i_T})$ *is uniformly distributed over* $\{0,1\}^{m\cdot T}$*. We denote the set of all* (T,L,m)*-somewhere-random sources by* $\mathsf{SR}_{T,L,m}$*, and the set of all convex combinations of sources in* $\mathsf{SR}_{T,L,m}$ *by* $\mathsf{convSR}_{T,L,m}$*.*

Definition 16 (Somewhere-extractor)**.** *Given a set of sources* $\mathcal{F}$ *over* $\{0,1\}^{\tilde{n}}$*, a function* $\mathsf{SomeExt} : \{0,1\}^{\tilde{n}} \to \{0,1\}^{m\cdot L}$ *is said to be a* (T,L,ε)-somewhere-extractor for $\mathcal{F}$ *if for every* $X \in \mathcal{F}$ *there exists* $Y \in \mathsf{convSR}_{T,L,m}$ *such that*

$$\mathsf{SomeExt}(X) \approx_\varepsilon Y.$$

A simple construction shows that strong (k,ε)-extractors imply the existence of deterministic somewhere-extractors for the class of general (n,k)-sources with the same error ε.

Lemma 17. *Let* $\mathsf{Ext} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ *be a strong* (k,ε)*-extractor, and set* $\{0,1\}^d = \{s_1,s_2,\ldots,s_{2^d}\}$*. Given* $x \in \{0,1\}^n$*, define* $\mathsf{SomeExt}(x) : \{0,1\}^n \to \{0,1\}^{m\cdot 2^d}$ *as*

$$\mathsf{SomeExt}(x) = (\mathsf{Ext}(x,s_1),\mathsf{Ext}(x,s_2),\ldots,\mathsf{Ext}(x,s_{2^d})).$$

Then, $\mathsf{SomeExt}$ *is a* $(1,2^d,\varepsilon)$*-somewhere-extractor for the class of* (n,k)*-sources.*

Proof. Fix an (n,k)-source X. By the strong property of the extractor, there is an index i such that

$$\mathsf{SomeExt}(X)_i = \mathsf{Ext}(X, s_i) \approx_\varepsilon U_m. \tag{5}$$

Consider $Y \in \mathsf{SR}_{1,2^d,m}$ defined as $Y_j = \mathsf{SomeExt}(X)_j$ for all $j \neq i$ and $Y_i = U_m$. Combining (5) with Lemma 2, it follows that $\mathsf{SomeExt}(X) \approx_\varepsilon Y$. □

The construction from Lemma 17 actually guarantees that a very large fraction of blocks of $Y = \mathsf{SomeExt}(X)$ will be close to uniform over $\{0,1\}^m$, provided X is an (n,k)-source. However, there is no guarantee that any pair of blocks (Y_{i_1}, Y_{i_2}) will be close to uniformly distributed over $\{0,1\}^{2m}$, as we cannot ensure that such blocks are close to being independent. Therefore, we only know that Y is ε-close to a $(1, 2^d, m)$-somewhere-random source.

2.6 Somewhere-Condensers

In this section, we introduce somewhere-condensers and related notions.

Definition 18 (Somewhere-entropic source)**.** *A distribution $X = (X_1, \ldots, X_L)$ over $\{0,1\}^{m \cdot L}$ is said to be a* (T, L, m, k)-somewhere-entropic source *if there exist indices $i_1 < i_2 < \cdots < i_T$ such that the random variables $X_{i_1}, X_{i_2}, \ldots, X_{i_T}$ are independently distributed and satisfy $\mathbf{H}_\infty(X_{i_j}) \geq k$ for all j. We denote the set of all (T, L, n, k)-somewhere-entropic sources by $\mathsf{SE}_{T,L,n,k}$, and the set of all convex combinations of sources in $\mathsf{SE}_{T,L,n,k}$ by $\mathsf{convSE}_{T,L,n,k}$.*

Definition 19 (Somewhere-condenser)**.** *A function $\mathsf{SomeCond} : \{0,1\}^n \to \{0,1\}^{m \cdot \ell}$ is said to be a* (k, k', L, ε)-somewhere condenser *if for every (n,k)-source X there exists $Y \in \mathsf{convSE}_{1,L,m,k'}$ such that*

$$\mathsf{SomeCond}(X) \approx_\varepsilon Y.$$

There exist explicit constructions of somewhere-condensers with a constant number of output blocks, linear output block length, and exponentially small error for arbitrarily low linear min-entropy.

Lemma 20 ([17])**.** *For all constants $\delta, \delta' > 0$ there exist constants $b, \beta, \rho > 0$ such that for large enough n there exists an explicit (k, k', b, ε)-somewhere condenser $\mathsf{SomeCond} : \{0,1\}^n \to \{0,1\}^{m \cdot b}$ with $k = \delta n$, $m = \beta n$, $k' = (1 - \delta')m$, and $\varepsilon = 2^{-\rho m}$.*

Remark 1. *The version of Lemma 20 presented in [17] is specialized for $\delta' = \delta$. However, inspection of [17, Lemmas 4.2 and 4.3] shows that the construction works for any constant $\delta' > 0$, as long as we allow the constants ℓ, β, ρ to depend simultaneously on δ and δ'. This observation is similar to the remark in [16] after Theorem 5.2.*

2.7 Proof Systems and Commitment Schemes

In this section, we present some definitions related to proof systems and commitment schemes.

Definition 21 (Computational indistinguishability)**.** *Let $X = \{X_m\}_{m \in \mathbb{N}}$ and $Y = \{Y_m\}_{m \in \mathbb{N}}$ be ensembles, where X_m's and Y_m's are probability distributions over $\{0,1\}^l$, for some $l = poly(m)$. We say that X and Y are* computationally indistinguishable*, denoted $X \approx Y$, if for every PPT distinguisher D there exists a negligible function ν such that for sufficiently large $m \in \mathbb{N}$,*

$$\Big| \Pr[t \leftarrow X_m : D(1^m, t) = 1] - \Pr[t \leftarrow Y_m : D(1^m, t) = 1] \Big| < \nu(m).$$

Definition 22 (Proof/argument system)**.** *A pair of PPT interactive algorithms* $\Pi = (\mathcal{P}, \mathcal{V})$ *constitute a* proof system *(resp., an* argument system*) for an NP-language* $\mathcal{L}$ *if the following conditions hold:*

Completeness: *For every* $x \in L$ *and* w *such that* $(x, w) \in \mathcal{R}_{\mathsf{L}}$*, it holds that:*

$$\Pr[\ \langle \mathcal{P}(w), \mathcal{V} \rangle(x) = 1\ ] = 1.$$

Soundness*: For every interactive (resp., PPT interactive) algorithm* $\mathcal{P}^\star$*, there exists a negligible function* ν *such that for every* $x \notin L$ *and every* z*:*

$$\Pr[\ \langle \mathcal{P}^\star(z), \mathcal{V} \rangle(x) = 1\ ] < \nu(|x|).$$

Definition 23 (Public coin protocol)**.** *An interactive protocol* $\Pi = (\mathcal{P}, \mathcal{V})$ *is* public coin *if, at every round,* $\mathcal{V}$ *simply tosses a predetermined number of coins (i.e., a random challenge) and sends the outcome to the prover. Moreover, we say that the transcript* τ *of an execution* $b = \langle \mathcal{P}(z), \mathcal{V} \rangle(x)$ *is* accepting *if* $b = 1$*.*

Definition 24 (Witness Indistinguishable (WI))**.** *An argument/proof system* $\Pi = (\mathcal{P}, \mathcal{V})$*, is* Witness Indistinguishable (WI) *for a relation* $\mathcal{R}$ *if, for every malicious PPT verifier* $\mathcal{V}^*$*, there exists a negligible function* ν *such that for all* x, w, w' *such that* $(x, w) \in \mathcal{R}$ *and* $(x, w') \in \mathcal{R}$ *it holds that:*

$$\Big|\Pr[\langle \mathcal{P}(w), \mathcal{V}^* \rangle(x) = 1] - \Pr[\langle \mathcal{P}(w'), \mathcal{V}^* \rangle(x) = 1]\Big| < \nu(|x|).$$

Definition 25 (Commitment Scheme)**.** *Given a security parameter* 1^m*, a commitment scheme* $\Pi_{\mathsf{com}} = (\mathcal{S}, \mathcal{R})$ *is a two-phase protocol between two PPT interactive algorithms, a sender* $\mathcal{S}$ *and a receiver* $\mathcal{R}$*. In the commitment phase* $\mathcal{S}$ *on input a message* m *interacts with* $\mathcal{R}$ *to produce a commitment* com*, and the private output* dec*.*

In the decommitment phase, $\mathcal{S}$ *sends to* $\mathcal{R}$ *a decommitment information* (m, dec) *such that* $\mathcal{R}$ *accepts* m *as the decommitment of* com*.*

Formally, we say that $\Pi_{\mathsf{com}} = (\mathcal{S}, \mathcal{R})$ *is a statistically binding commitment scheme if the following properties hold:*

Correctness:

- *Commitment phase. Let* com *be the commitment of the message* m *given as output of an execution of* $\Pi_{\mathsf{com}} = (\mathcal{S}, \mathcal{R})$ *where* $\mathcal{S}$ *runs on input a message* m*. Let* dec *be the private output of* $\mathcal{S}$ *in this phase.*
- *Decommitment phase*[11]*.* $\mathcal{R}$ *on input* m *and* dec *accepts* m *as decommitment of* com*.*

Computational Hiding : *for any PPT adversary* A *and a randomly chosen bit* $b \in \{0, 1\}$*, consider the following hiding experiment* $\mathtt{Exp}^b_{A,\Pi_{\mathsf{com}}}(m)$*:*

- *Upon input* 1^m*, the adversary* A *outputs a pair of messages* m_0, m_1 *that are of the same length.*
- $\mathcal{S}$ *on input the message* m_b *interacts with* A *to produce a commitment of* m_b*.*
- A *outputs a bit* b' *and this is the output of the experiment.*

[11] In this paper we consider a non-interactive decommitment phase only.

For any PPT adversary A, there exist a negligible function ν s.t.:

$$\Big|\Pr[\mathtt{Exp}^0_{A,\Pi_{\mathsf{com}}}(m)=1]-\Pr[\mathtt{Exp}^1_{A,\Pi_{\mathsf{com}}}(m)=1]\Big|<\nu(m).$$

Statistical Binding: *for every commitment* com *generated during the commitment phase by a possibly malicious unbounded sender $\mathcal{S}^\star$ there exists a negligible function ν such that $\mathcal{S}^\star$, with probability at most $\nu(m)$, outputs two decommitments (m_0,dec_0) and (m_1,dec_1), with $m_0\neq m_1$, such that $\mathcal{R}$ accepts both decommitments.*

Definition 26 (Non-Interactive Proof System.)**.** *A tuple of PPT algorithms $\Pi=(\mathcal{G},\mathcal{P},\mathcal{V})$ is a non-interactive proof system in the CRS model for the NP-language $\mathcal{L}$ with witness relation $\mathcal{R}$ if it satisfies the following properties:*

Completeness. $\forall\ x,w$ s.t. $(x,w)\in\mathcal{R}$:

$$\Pr\big[\sigma\leftarrow\mathcal{G}(1^m),\ \pi\leftarrow\mathcal{P}(\sigma,x,w)\ :\ \mathcal{V}(x,\pi,\sigma)=1\big]=1.$$

Soundness. $\forall\ x\notin\mathcal{L}$, $\forall$ adversary $\mathcal{P}^\star$, there exists a negligible function ν such that:

$$\Pr\big[\sigma\leftarrow\mathcal{G}(1^m),\ \pi\leftarrow\mathcal{P}^\star(\sigma,x)\ :\ \mathcal{V}(x,\pi,\sigma)=1\big]\leq\nu(m).$$

Definition 27 (Non-Interactive WI Proof System.)**.** *A non-interactive proof system $\Pi=(\mathcal{G},\mathcal{P},\mathcal{V})$ in the CRS model for the NP-language $\mathcal{L}$ with witness relation $\mathcal{R}$ is witness indistinguishable (WI) if it satisfies the following property:*

$\forall\ x,w_0,w_1$ *s.t.* $(x,w_0)\in\mathcal{R}$ *and* $(x,w_1)\in\mathcal{R}$*, the following two distributions are computationally indistinguishable:*

$$\Big\{\pi\ :\ \sigma\leftarrow\mathcal{G}(1^m),\pi\leftarrow\mathcal{P}(\sigma,x,w_0)\Big\},\Big\{\pi\ :\ \sigma\leftarrow\mathcal{G}(1^m),\pi\leftarrow\mathcal{P}(\sigma,x,w_1)\Big\}$$

Definition 28 (Non-interactive Commitment Scheme)**.** *Consider a message space M and PPT algorithms $\Pi=(\mathcal{G},\mathcal{S},\mathcal{R})$ where $\mathcal{G}$ on input 1^m outputs σ, $\mathcal{S}$ is the randomized commitment algorithm that takes as input security parameter 1^m, σ a message $\mathsf{msg}\in M$ and outputs commitment com and decommitment $(\mathsf{msg},\mathsf{dec})$; $\mathcal{R}$ is the verification algorithm that takes as input $(\sigma,\mathsf{com},\mathsf{dec},\mathsf{msg})$ and decides whether msg is the decommitment of com. Π is a non-interactive commitment scheme in the CRS model if it satisfies the following properties.*

Correctness. $\forall\ \mathsf{msg}\in M$ it holds that:

$$\Pr\big[\sigma\leftarrow\mathcal{G}(1^m),\ \mathsf{com}\leftarrow\mathcal{S}(1^m,\sigma,\mathsf{msg})\ :\mathcal{R}(\sigma,\mathsf{com},\mathsf{dec},\mathsf{msg})=\mathsf{msg}\big]=1.$$

Hiding. For every PPT adversary A there exists a negligible function ν such that, $\forall\ \mathsf{msg}_0,\mathsf{msg}_1\in M$ it holds that:

$$\Pr\begin{bmatrix}\sigma\leftarrow\mathcal{G}(1^m),\\ b\leftarrow\{0,1\},\\ \mathsf{com}\leftarrow\mathcal{S}(1^m,\sigma,\mathsf{msg_b})\end{bmatrix}\ :\ b=A(\sigma,\mathsf{com},\mathsf{msg}_0,\mathsf{msg}_1)\ \Bigg]\leq\frac{1}{2}+\nu(m).$$

Binding. For every commitment com *generated during the commitment phase by a possibly malicious unbounded sender $\mathcal{S}$ there exists a negligible function ν such that $\mathcal{S}$, with probability at most $\nu(m)$, outputs two decommitments $(\mathsf{msg}_0,\mathsf{dec}_0)$ and $(\mathsf{msg}_1,\mathsf{dec}_1)$, with $\mathsf{msg}_0\neq\mathsf{msg}_1$, such that $\mathcal{R}$ accepts both decommitments.*

3 SHELA Sources

In this section, we give a formal definition of Somewhere Honest Entropic Look Ahead (SHELA) sources, and present explicit constructions of somewhere-extractors with good parameters for this class of sources.

Definition 29 (SHELA source)**.** *A distribution $X \in \{0,1\}^{n \cdot \ell}$ is said to be an (n, k, t, ℓ)-*SHELA source *if there exist random variables $1 \leq I_1 < I_2 < \cdots < I_t \leq \ell$ with arbitrary joint distribution, t independent (n, k)-sources $Z_1, Z_2, \ldots, Z_t$, and a (possibly randomized) adversary $\mathcal{A}$ such that X is generated as follows:*

1. *Sample $(i_1, i_2, \ldots, i_t) \leftarrow (I_1, I_2, \ldots, I_t)$;*
2. *For each $j \in [t]$, set $B_{i_j} \leftarrow Z_j$;*
3. *For each $i \in [\ell] \setminus \{i_1, \ldots, i_t\}$, $\mathcal{A}$ sets $B_i = \mathcal{A}(B_1, \ldots, B_{i-1}, i_1, \ldots, i_t)$;*
4. *Set $X = (B_1, B_2, \ldots, B_\ell)$.*

We denote the set of all such SHELA sources by $\mathsf{SHELA}_{n,k,t,\ell}$.

We begin by showing that deterministic randomness extraction is impossible from SHELA sources.

3.1 Impossibility of Deterministic Extraction from SHELA Sources

In this section, we show that deterministic randomness extraction is impossible from SHELA sources. This is achieved by relating SHELA sources to resettable sources, for which such impossibility has already been shown [62].

As a starting point, we state the definition of resettable sources, originally introduced in [62].

Definition 30 (p-resettable source [62])**.** *A distribution $X \in [a]^\ell$ is said to be a p-*resettable source *if there exists a randomized and computationally unbounded adversary $\mathcal{A}$ such that X is generated as follows:*

1. *For each $i \in [n]$, a uniformly random $v \in [a]$ is chosen;*
2. *With probability $1-p$, X_i is set to v and the process moves to $i+1$. Else, the value v is given to the adversary $\mathcal{A}$, which chooses, based on v and $X_1, \ldots, X_{i-1}$, whether to set $X_i = v$ or sample a new value $v' \in [a]$ uniformly at random and set $X_i = v'$;*
3. *The output is $X = (X_1, \ldots, X_\ell) \in [a]^\ell$.*

The following lemma was proved in [62]. It states that deterministic extraction with subconstant error from p-resettable sources is impossible for any p, and it is obtained by relating resettable sources to Santha-Vazirani sources.

Lemma 31 ([62, Theorem 3.5])**.** *For every function $F : [a]^\ell \to \{0,1\}$ and any $0 < p \leq 1$ there exists a p-resettable source X such that $F(X) \not\approx_{p/12} U_1$.*

We prove the impossibility of deterministic extraction from SHELA sources with the help of Lemma 31.

Theorem 32. *For every function $F : \{0,1\}^{n\cdot\ell} \to \{0,1\}$, any constant $0 \le \gamma < 1$, and $\ell \ge c(\gamma)$ for $c(\gamma)$ a large enough constant depending on γ, there is an $(n, n, \lfloor \gamma \cdot \ell \rfloor, \ell)$-SHELA source X such that $F(X) \not\approx_{\frac{1-\gamma}{48}} U_1$.*

Proof. Fix a function F as in the theorem statement, write $\gamma = 1 - 2p$ for a constant $p \in (0, 1/2]$, and let Z be the p-resettable source guaranteed by Lemma 31 with $[a] = \{0,1\}^n$.

According to the process generating Z, we may picture that the blocks which cannot be resampled by the adversary $\mathcal{A}$ are chosen a priori and independently with probability $1-p$. Let $\mathcal{S}$ be the set of indices of blocks which cannot be resampled. Then, we have $\mathbb{E}[|\mathcal{S}|] = (1-p)\ell$. Therefore, by a Chernoff bound we conclude that

$$\begin{aligned} \Pr[|\mathcal{S}| \le (1-2p)\ell = \gamma \cdot \ell] &\le \exp\left(-\frac{\gamma^2 \ell}{4}\right) \\ &\le p/24, \end{aligned} \tag{6}$$

where the last inequality holds provided ℓ is large enough depending only on γ. Consider the event E_1 corresponding to $|\mathcal{S}| \ge \gamma \cdot \ell$ and its complement E_2. Let Z_i denote the random variable Z conditioned on E_i for $i = 1, 2$.

We now show that Z_1 is an $(n, n, t = \lfloor \gamma \cdot \ell \rfloor, \ell)$-SHELA source. Then, we will show that if Z is highly biased, then so is Z_1. This concludes our proof with $X = Z_1$. Let $\mathcal{A}$ denote the adversary associated to Z. Sample $\mathcal{S} \subseteq [\ell]$ by including each $i \in [\ell]$ independently with probability $1-p$, conditioned on E_1. Then, sample $I_1, \ldots, I_t$ by choosing t elements of $\mathcal{S}$ uniformly at random, and sample the blocks at those locations uniformly at random from $\{0,1\}^n$. To generate the remaining blocks, consider the following adversary $\mathcal{A}'$: If $i \in \mathcal{S}$, then $\mathcal{A}'$ samples the block uniformly at random. Else, $\mathcal{A}'$ behaves like $\mathcal{A}$ given the values of the previous blocks. The random variable generated by this process is an $(n, n, t = \lfloor \gamma \cdot \ell \rfloor, \ell)$-SHELA source, and is identically distributed to Z_1, as desired.

It remains to show that $F(Z_1) \not\approx_{p/24} U_1$. We have

$$\begin{aligned} p/12 &\le \Delta(F(Z); U_1) \\ &\le \Pr[E_1] \cdot \Delta(F(Z_1); U_1) + \Pr[E_2] \cdot \Delta(F(Z_2); U_1) \\ &\le \Delta(F(Z_1); U_1) + p/24, \end{aligned}$$

where the third inequality holds because $\Pr[E_2] \le p/24$ by (6). This implies $\Delta(F(Z_1); U_1) \ge p/12 - p/24 = p/24$. The desired result follows by noting that $p = \frac{1-\gamma}{2}$. □

3.2 Deterministic Somewhere-Extractors for SHELA Sources

In this section, we construct deterministic somewhere-extractors for regular SHELA sources. For deterministic somewhere-extraction from *online* SHELA sources, the reader is referred to Section 4.

3.2.1 Honest Blocks with High Min-Entropy

In this section, we consider the case where each honest block in a SHELA source has min-entropy $(1-\gamma)n$ for some sufficiently small constant $\beta > 0$. The following result states that an explicit somewhere-extractor with exponentially small error and linear output block length exists for such SHELA sources. Notably, it is also the case that if the number of honest input blocks is t and the total number of input blocks is ℓ, then the number of uniform output blocks is $T = t - 1$ and the number of total output blocks is $L = \ell - 1$.

Theorem 33. *There exists a small enough constant $\gamma > 0$ such that for n large enough and $2 \leq t \leq \ell \leq \mathrm{poly}(n)$ there exists an explicit $(t-1, \ell-1, \varepsilon')$-somewhere extractor $\mathsf{SomeExt} : \{0,1\}^{n\cdot\ell} \to \{0,1\}^{m\cdot(\ell-1)}$ for $\mathsf{SHELA}_{n,k',t,\ell}$ with $k' = (1-\gamma)n$, $m = \frac{(1-7\gamma)n}{3}$, and $\varepsilon' = 2(t-1)\cdot 2^{-\gamma n}$.*

The construction we use to prove Theorem 33 makes use of the following objects: For $i \in \{2, \ldots, \ell\}$, let $\mathsf{Ext}_i : \{0,1\}^{n\cdot(i-1)} \times \{0,1\}^n \to \{0,1\}^m$ be an average-case strong seeded (k, ε)-extractor with $k = 2k'/3$, $k' = (1-\gamma)n$, $m = \frac{(1-7\gamma)n}{3}$ and $\varepsilon = 2^{-2\gamma n}$ for a small enough constant $\gamma > 0$. These can be obtained by using the explicit GUV extractor [66] with appropriate parameters.

In fact, let Ext_i be the GUV extractor from Lemma 11 with $\alpha = 1/2$, $k = 2k'/3$, $k' = (1-\gamma)n$, $m = \frac{(1-7\gamma)n}{3}$ and $\varepsilon = 2^{-2\gamma n}$ for a constant $\gamma > 0$ small enough such that the required seed length

$$d = \log(n \cdot (i-1)) + C_\alpha \log(k/\varepsilon) \leq \log \ell + 2C_\alpha(\log n + \log(1/\varepsilon))$$

satisfies $d \leq n$ for all i and n large enough. This choice of parameters is possible because $m = \frac{(1-7\gamma)n}{3} < k/2 = \frac{(1-\gamma)n}{3}$ and since $\ell \leq \mathrm{poly}(n)$. By Lemma 9 with $\eta = \varepsilon$, it holds that Ext_i is an average-case strong $(\overline{k} = k + 2\gamma n, \overline{\varepsilon} = 2\varepsilon)$-extractor. Note that the extractor still works with a larger seed length than d by restricting it to use the first d bits of the seed only.

We are now ready to describe our construction of the somewhere-extractor $\mathsf{SomeExt} : \{0,1\}^{n\cdot\ell} \to \{0,1\}^{m\cdot(\ell-1)}$ for $X \in \mathsf{SHELA}_{n,k,t,\ell}$. First, write $X = (B_1, B_2, \ldots, B_\ell)$. Then, the output $\mathsf{SomeExt}(X)$ can be written as $\mathsf{SomeExt}(X) = (B'_2, B'_3, \ldots, B'_\ell)$, where each B'_i is obtained as

$$B'_i = \mathsf{Ext}_i((B_1, B_2, \ldots, B_{i-1}), B_i) \in \{0,1\}^m. \tag{7}$$

The following lemma is the key component of our proof of Theorem 33 using the construction of $\mathsf{SomeExt}$ detailed above.

Lemma 34. *Let $X \in \mathsf{SHELA}_{n,k,t,\ell}$, and recall $I_1 < I_2 < \cdots < I_t$ denote the positions of the honest blocks in X. Then, we have*

$$B'_{I_2}, \ldots, B'_{I_t}, I_2, \ldots, I_t \approx_{2(t-1)\cdot 2^{-\gamma n}} U_m^{t-1}, I_2, \ldots, I_t,$$

where B'_i is defined as in (7) and U_m^{t-1} denotes $t-1$ independent copies of U_m which are also independent of the remaining random variables.

Proof. It suffices to show that

$$B'_{I_2}, \ldots, B'_{I_{j-1}}, B'_{I_j}, I_2, \ldots, I_t \approx_{2\cdot 2^{-\gamma n}} B'_{I_2}, \ldots, B'_{I_{j-1}}, U_m, I_2, \ldots, I_t \tag{8}$$

for all $j \in \{2, \ldots, t\}$. The desired statement then follows by repeated application of the triangle inequality. We begin by noting that B_{I_j} is independent of all random variables in (8). Moreover, it holds that

$$\begin{aligned}
&\widetilde{\mathbf{H}}_\infty(B_1, \ldots, B_{I_j-1} | B'_{I_2}, \ldots, B'_{I_{j-1}}, I_2, \ldots, I_t) \\
&\geq \widetilde{\mathbf{H}}_\infty(B_{I_{j-1}} | B'_{I_2}, \ldots, B'_{I_{j-1}}, I_2, \ldots, I_t) \\
&\geq \widetilde{\mathbf{H}}_\infty(B_{I_{j-1}} | B'_{I_2}, \ldots, B'_{I_{j-2}}, I_2, \ldots, I_t) - m \\
&= \mathbf{H}_\infty(B_{I_{j-1}}) - m \\
&\geq k' - m \\
&= \overline{k}.
\end{aligned} \tag{9}$$

The second inequality holds by Lemma 5. The first equality is true because $B_{I_{j-1}}$ is independent of $B'_{I_2}, \ldots, B'_{I_{j-2}}, I_2, \ldots, I_t$. The third inequality follows from the min-entropy constraint on the honest blocks of X.

Finally, to see that (8) holds, it remains to observe that Ext_i is an average-case strong extractor with min-entropy requirement $\overline{k}$, B_{I_j} is independent of $B_1, \ldots, B_{I_j-1}, I_2, \ldots, I_t$, and $\mathbf{H}_\infty(B_{I_j}) \geq k'$. Therefore, if we use B_{I_j} as the seed for Ext_i, then Lemma 20 and the fact that Ext_i is an average-case strong $(\overline{k}, \overline{\varepsilon})$-extractor ensure that

$$\mathsf{Ext}_i(X, B_{I_j}), W \approx_{2^{\gamma n} \cdot \overline{\varepsilon} = 2 \cdot 2^{-\gamma n}} U_m, W \tag{10}$$

for all X and W independent of B_{I_j} such that $\widetilde{\mathbf{H}}_\infty(X|W) \geq \overline{k}$. Instantiating X and W in (10) as

$$X = B_1, \ldots, B_{I_j-1}$$
$$W = B'_{I_2}, \ldots, B'_{I_{j-1}}, I_2, \ldots, I_t,$$

and recalling (9) and that X and W are independent of B_{I_j} yields the desired result. □

To conclude the proof of Theorem 33, it suffices to combine Lemma 34 with the following result.

Lemma 35. *Let $X \in \{0,1\}^{\ell \cdot m}$ and I any random variable over subsets of size T of $[L]$ for some $T \leq L$. Suppose that*

$$X_I, I \approx_\varepsilon U_m^T, I.$$

Then, it holds that $X \approx_\varepsilon Y$ for some $Y \in \mathsf{convSR}_{T,L,m}$.

We prove a more general form of this result in Appendix A (Lemma 35 is obtained from Lemma 58 by setting $Z = U_m^T$). Such a result already appears in [16] in a different form, but we choose to state and prove it here taking into account our setting and notation for the sake of clarity.

3.2.2 Honest Blocks with Low Linear Min-Entropy

In this section, we construct somewhere-extractors for SHELA sources that have honest blocks with min-entropy δn for some arbitrarily small constant $\delta > 0$. We show that there is an explicit somewhere-extractor for such SHELA sources with exponentially small error and linear output block length. Moreover, if the number of input honest and total blocks are t and ℓ, respectively, then the number of output uniform and total blocks are $T = t - 1$ and $L = O(\ell)$, respectively.

Theorem 36. *For every constant $\delta > 0$ there exist constants $a_1, a_2, a_3 > 0$ such that for n large enough and all $2 \leq t \leq \ell \leq \mathrm{poly}(n)$ there exists an explicit (T, L, ε')-somewhere extractor $\mathsf{SomeExt} : \{0,1\}^{n \cdot \ell} \to \{0,1\}^{m \cdot L}$ for $\mathsf{SHELA}_{n,k',t,\ell}$ with $k' = \delta n$, $m = a_1 \cdot n$, $\varepsilon' = 2(t-1)2^{-a_2 \cdot n}$, $T = t - 1$, and $L = a_3 \cdot \ell$.*

We now turn to a precise description of our construction. Fix a constant $\delta \in (0,1)$ and consider the $(\delta n, (1-\gamma)n', b, 2^{-\rho n'})$-somewhere-condenser $\mathsf{SomeCond} : \{0,1\}^n \to \{0,1\}^{b \cdot n'}$ from Lemma 20, where $\gamma > 0$ is a small constant to be determined, $n' \geq \beta n$, and b, β, and ρ depend only on δ and γ. For each $i = 2, \ldots, \ell$, consider also the average-case strong (k, ε)-extractor

$$\mathsf{Ext}_i : \{0,1\}^{b \cdot n' (i-1)} \times \{0,1\}^{n'} \to \{0,1\}^m$$

with $\varepsilon = 2^{-2\gamma n'}$, $k = \frac{2(1-3\gamma)n'}{3}$, and $m = \frac{(1-3\gamma)n'}{3}$. These extractors can be instantiated using the strong GUV extractor [66] with appropriate parameters.

In fact, consider the strong GUV (k, ε)-extractor

$$\mathsf{Ext}_i : \{0,1\}^{b \cdot n'(i-1)} \times \{0,1\}^{n'} \to \{0,1\}^m$$

from Lemma 11 with $\alpha = 1/2$, $\varepsilon = 2^{-2\gamma n'}$, $k = \frac{2(1-3\gamma)n'}{3}$, and $m = \frac{(1-3\gamma)n'}{3}$. We set $\gamma > 0$ to be a small enough constant so that the required seed length

$$d = \log(b \cdot n'(i-1)) + C_\alpha \log(k/\varepsilon) \leq \log b + \log \ell + 2C_\alpha(\log n' + \log(1/\varepsilon)) \tag{11}$$

satisfies $d \leq n'$ for large enough n'. It suffices to set γ small enough so that, say, $4C_\alpha\gamma < 1/2$. Observe that b and β (which come from the somewhere-condenser $\mathsf{SomeCond}$ defined in Section 3.2.2) change with the choice of γ. However, after γ is fixed they become constants and so the right-hand side of (11) is indeed smaller than n' for n' large enough. Moreover, note that $m = k/2$, and so the choice of parameters above is possible according to Lemma 11. Finally, recalling Lemma 9 with $\eta = \varepsilon$, it follows that Ext_i is an average-case strong $(\overline{k}, \overline{\varepsilon})$-extractor with $\overline{k} = 2n'/3 = (1-\gamma)n' - m$ and $\overline{\varepsilon} = 2\varepsilon$.

We are now ready to define $\mathsf{SomeExt}(X)$ for $X = (B_1, \ldots, B_\ell) \in \mathsf{SHELA}_{n,k',t,\ell}$. We write

$$\mathsf{SomeCond}(B_i) = (B_{i1}, \ldots, B_{ib}) \in \{0,1\}^{n' \cdot b}.$$

Then, we have

$$\mathsf{SomeExt}(X) = (B'_{ij})_{i \in [\ell], j \in [b]} \in \{0,1\}^{m \cdot L}$$

for B'_{ij} defined as

$$B'_{ij} = \mathsf{Ext}_i((B_{i'j'})_{i'<i, j' \in [b]}, B_{ij}) \in \{0,1\}^m. \tag{12}$$

Before we proceed with the proof of Theorem 36 using the construction just described, we need the following observation. Let $Y \in \mathsf{convSE}_{1,b,n',k}$, which can be written as a convex combination $Y = \sum_j \pi_j Y_j$ for some $Y_j \in \mathsf{SE}_{1,b,n',k}$. Suppose Y_{j_1} and Y_{j_2} both have blocks with min-entropy at least k at position p. Then, any convex combination of Y_{j_1} and Y_{j_2} also has a block with min-entropy at least k at the same position p. As a result, it follows that we can always write Y as a convex combination $Y = \sum_{j=1}^b \pi_j Y_j$ where Y_j has a block of min-entropy at least k at position j. In particular, there is a random variable $J \in [b]$ with $\Pr[J = j] = \pi_j$ such that $\mathbf{H}_\infty(Y_J | J = j) \geq k$ for all j.

For $X \in \mathsf{SHELA}_{n,k',t,\ell}$ as above, let $Z_1, \ldots, Z_t$ denote the values of its honest blocks, and $I_1, \ldots, I_t$ their respective positions. For each $i \in [t]$, we have $\mathsf{SomeCond}(Z_i) \approx_{2^{-\rho n'}} Y^i$ for some $Y^i \in \mathsf{convSE}_{1,b,n',(1-\gamma)n'}$. Based on the previous paragraph, there exist independent random variables $J_1, \ldots, J_t \in [b]$ such that $\mathbf{H}_\infty(Y^i_{J_i} | J_i = j) \geq k$ for all j.

We begin by showing that, since

$$\mathsf{SomeCond}(B_{I_r}) = (B_{I_r 1}, \ldots, B_{I_r b}) \approx_{2^{-\rho n'}} Y^r \tag{13}$$

for some $Y^r \in \mathsf{convSE}_{1,b,n',(1-\gamma)n'}$ and $r = 2, \ldots, t$, we can essentially replace $\mathsf{SomeCond}(B_{I_r})$ by Y^r for all r by paying an additive $t \cdot 2^{-\rho n'}$ penalty in statistical distance.

Lemma 37. *Let*

$$X^1 = (\mathsf{SomeCond}(B_1), \ldots, \mathsf{SomeCond}(B_\ell)).$$

Then, there is $X^2 = (X^2_1, \ldots, X^2_\ell)$ satisfying:

1. *$X^2_{I_r}$ is distributed as Y^r;*

2. $X^2_1, X^2_2, \ldots, X^2_{I_r-1}$ *are independent of* $X^2_{I_r}, \ldots, X^2_{I_t}$;

3. $(I_1, \ldots, I_t)$ *are independent of* $(X^2_{I_1}, \ldots, X^2_{I_t})$;

4. $X^1 \approx_{t\cdot 2^{-\rho n'}} X^2$.

Proof. By (13) and Lemma 2, for each $r = 1, \ldots, t$ there is a coupling $(X^1_{I_r}, Q_r)$ such that Q_r is distributed as Y^r and $\Pr[X^1_{I_r} \neq Q_r] \leq \varepsilon$. Moreover, the Q_r's are all independent of each other. Consider X^2 defined by setting, for each valid fixing $I_1 = i_1, \ldots, I_t = i_t$, $X^2_{i_r} = Q_r$ and $X^2_i = X^1_i$ for $i \neq i_1, \ldots, i_t$. We now check that X^2 satisfies the desired properties: First, since Q_r is distributed like Y^r, so is $X^2_{I_r}$ by definition. Second, since the Q_r's are independent and X^1_i for $i \neq i_1, \ldots, i_t$ only depends on $X^1_{i'}$ for $i' < i$, it holds that $X^2_1, \ldots, X^2_{I_r-1}$ are independent of $X^2_{I_r}, \ldots, X^2_{I_t}$. Property 3 holds because $(I_1, \ldots, I_t)$ are chosen a priori exactly as in X. Finally, since

$$\Pr[X^1 \neq X^2] \leq \sum_{r=1}^{t} \Pr[X^1_{I_r} \neq X^2_{I_r}] \leq t \cdot 2^{-\rho n'}$$

according to (13), it holds that $X^1 \approx_{t\cdot 2^{-\rho n'}} X^2$. □

From here onwards we work with X^2 instead of X^1. We define

$$X^2_i = (\overline{B}_{i1}, \ldots, \overline{B}_{ib}),$$

and

$$\overline{B}'_{ij} = \mathsf{Ext}_i((\overline{B}_{i'j'})_{i'<i, j' \in [b]}, \overline{B}_{ij}) \in \{0,1\}^m.$$

To finish the proof, it now suffices to show the following result.

Lemma 38. *We have*

$$\Delta(\overline{B}'_{I_rJ_r}; U_m | \overline{B}'_{I_2J_2}, \ldots, \overline{B}'_{I_{r-1}J_{r-1}}, I_2, \ldots, I_t, J_2, \ldots, J_t) \leq 2 \cdot 2^{-\gamma n'} \tag{14}$$

for all $r \in \{2, \ldots, t\}$.

In fact, by combining Lemma 38 and repeated application of the triangle inequality we then have

$$\Delta(\overline{B}'_{I_2J_2}, \ldots, \overline{B}'_{I_tJ_t}; U_m^{t-1} | I_2, \ldots, I_t, J_2, \ldots, J_t) \leq (t-1) \cdot 2 \cdot 2^{-\gamma n'}.$$

By Lemma 35, this implies that

$$(\overline{B}'_{i,j})_{i \in [\ell], j \in [b]} \approx_{2(t-1)\cdot 2^{-\gamma n'}} Y \tag{15}$$

for some $Y \in \mathsf{convSR}_{t',\ell',m}$. Therefore, since the left hand side of (15) is a deterministic function of X_2 and $X^1 \approx_{t\cdot 2^{-\rho n'}} X^2$ by Lemma 37, we conclude that

$$\mathsf{SomeExt}(X) = (B'_{i,j})_{i \in [\ell], j \in [b]} \approx_{2(t-1)\cdot 2^{-\gamma n'} + t \cdot 2^{-\rho n'}} Y.$$

Note that the positions of the uniform blocks in Y are given by (I_r, J_r) for $r = 2, \ldots, t$. Finally, setting, say, $a_1 = \frac{\beta\cdot(1-3\gamma)}{3}$, $a_2 = \frac{\beta\cdot\min(\gamma,\rho)}{2}$, and $a_3 = b$ yields Theorem 36 (recall that β, ρ, and b are constants after γ is fixed). It remains to prove Lemma 38.

Proof of Lemma 38. We have

$$\begin{aligned}
&\widetilde{\mathbf{H}}_\infty((\overline{B}_{ij})_{i<I_r,j\in[b]}|\overline{B}'_{I_2J_2},\ldots,\overline{B}'_{I_{r-1}J_{r-1}},I_2,\ldots,I_t,J_2,\ldots,J_t)\\
&\geq \widetilde{\mathbf{H}}_\infty(\overline{B}_{I_{r-1}J_{r-1}}|\overline{B}'_{I_2J_2},\ldots,\overline{B}'_{I_{r-1}J_{r-1}},I_2,\ldots,I_t,J_2,\ldots,J_t)\\
&\geq \widetilde{\mathbf{H}}_\infty(\overline{B}_{I_{r-1}J_{r-1}}|\overline{B}'_{I_2J_2},\ldots,\overline{B}'_{I_{r-2}J_{r-2}},I_2,\ldots,I_t,J_2,\ldots,J_t)-m\\
&= \widetilde{\mathbf{H}}_\infty(\overline{B}_{I_{r-1}J_{r-1}}|J_{r-1})-m\\
&\geq (1-\gamma)n'-m\\
&= \overline{k}. \qquad (16)
\end{aligned}$$

The second inequality follows from Lemma 5. The first equality holds because $(\overline{B}_{I_{r-1}J_{r-1}},J_{r-1})$ is independent of the remaining random variables in the expression by the properties of X^2. The third inequality is true because $\mathbf{H}_\infty(\overline{B}_{I_{r-1}J_{r-1}}|J_{r-1}=j)=\mathbf{H}_\infty(Y^{r-1}_{J_{r-1}}|J_{r-1}=j)\geq(1-\gamma)n'$ for every j, since $X^2_{I_r}$ is distributed like Y^r.

Define the random variables

$$\begin{aligned}
Z &= (\overline{B}_{ij})_{i<I_r,j\in[b]}\\
W &= \overline{B}'_{I_2J_2},\ldots,\overline{B}'_{I_{r-1}J_{r-1}},I_2,\ldots,I_t,J_2,\ldots,J_{r-1},J_{r+1},\ldots,J_t\\
S &= \overline{B}_{I_rJ_r}.
\end{aligned}$$

Then, the statement in (14) is equivalent to

$$\mathsf{Ext}_r(Z,S),W,J_r\approx_{2\cdot 2^{-\gamma n}}U_m,W,J_r. \qquad (17)$$

Since J_r is independent of W and Z and only affects $\mathsf{Ext}_r(Z,S)$ through S, it is enough to show that

$$\mathsf{Ext}_r(Z,S),W,S\approx_{2\cdot 2^{-\gamma n}}U_m,W,S.$$

This statement follows from Lemma 12 by noting that Ext_i is an average-case strong $(\overline{k},\overline{\varepsilon}=2\cdot 2^{-\gamma n})$-extractor with seed length n', Z has enough min-entropy by (16), both Z and the side information W are independent of S, and $\mathbf{H}_\infty(S)\geq(1-\gamma)n'$ since S is distributed like $Y^r_{J_r}$. □

4 Somewhere-Extraction from Online SHELA Sources

In this section, we consider a different type of SHELA source under a stronger adversarial model. In Definition 29, the adversary must choose the location of the honest blocks before any blocks are generated. However, one might imagine that the adversary might be able to corrupt blocks during the generation process.

Motivated by this, we will define another type of source, which we call an *Online* SHELA source, that captures this behavior. Specifically, the locations of the honest blocks are no longer chosen a priori (as in Definition 29), but the adversary is allowed to choose whether a given block is honest or malicious given the values of all previous blocks. Our only requirement is that exactly t blocks must be honest at the end of the process. A formal definition follows.

Definition 39 (Online SHELA source)**.** *A distribution $X\in\{0,1\}^{n\cdot\ell}$ is said to be an* online (n,k,t,ℓ)-SHELA source *if there exist t independent (n,k)-sources $Z_1,Z_2,\ldots,Z_t$ and a randomized, computationally unbounded adversary $\mathcal{A}$ such that X is generated as follows:*

1. *Set $j=0$ and $\mathsf{Hon}=\{\}$;*

2. *For $i \in [\ell]$, the adversary $\mathcal{A}$ decides whether B_i is honest or malicious based on $B_1, \ldots, B_{i-1}$ and* Hon*:*

   If B_i is honest, set $j = j + 1$, $B_i \leftarrow Z_j$, and $\mathsf{Hon} = \mathsf{Hon} \cup \{i\}$;

   Else, set $B_i = \mathcal{A}(B_1, \ldots, B_{i-1}, \mathsf{Hon})$.

3. *The process finishes successfully if $j = t$, in which case we set*

$$X = (B_1, B_2, \ldots, B_\ell).$$

We denote the set of all such online SHELA sources by $\mathsf{onSHELA}_{n,k,t,\ell}$*.*

We showed that deterministic randomness extraction from SHELA sources is impossible in Section 3.1. We proceed to show that the somewhere-extractors designed for standard SHELA sources in Section 3.2 can also be applied to online SHELA sources.

When dealing with *online* SHELA sources, we are only able to extract a special type of 1-out-of-L convSR-sources, which we call *as-you-go* somewhere-random sources (in short, $\mathsf{AYG\text{-}SR}$-sources). In general, for T-out-of-L convSR-sources it is the case that a subset of T good blocks is jointly uniform. However, in $\mathsf{AYG\text{-}SR}$-sources we may picture blocks being generated in chronological fashion, and we only require that the j-th "good" block is uniformly distributed given the values and positions of all previous good blocks. In particular, the position of a given good block is not fixed and may depend on the values and positions of other blocks, and that good block may not be uniformly distributed conditioned on the values and positions of future good blocks. A precise definition follows.

Definition 40 (As-you-go somewhere-random source)**.** *A source*

$$X = (X_1, X_2, \ldots, X_L) \in \{0,1\}^{m \cdot L}$$

is said to be a (T, L, m) as-you-go somewhere-random source*,* $\mathsf{AYG\text{-}SR}$*-source in short, if there exist random variables $1 \leq I_1 < I_2 < \cdots < I_T \leq m$ such that*

$$X_{I_j}, X_{I_1}, X_{I_2}, \ldots, X_{I_{j-1}}, I_1, \ldots, I_j = U_m, X_{I_1}, X_{I_2}, \ldots, X_{I_{j-1}}, I_1, \ldots, I_j$$

for all $j = 1, \ldots, T$, where U_m is independent of the remaining random variables on the right hand side.

Similarly to somewhere-random extractors, we can define *as-you-go somewhere-extractors* as functions that extract as-you-go somewhere-random sources from a given family of input sources.

Definition 41 (As-you-go somewhere-extractor)**.** *Given a set of sources $\mathcal{F}$ over $\{0,1\}^{\tilde{n}}$, a function $\mathsf{SomeExt} : \{0,1\}^{\tilde{n}} \to \{0,1\}^{m \cdot L}$ is said to be a (T, L, ε)*-as-you-go somewhere-extractor for $\mathcal{F}$*, or* AYG*-somewhere-extractor in short, if for every $X \in \mathcal{F}$ there exists a (T, L, m)* $\mathsf{AYG\text{-}SR}$*-source Y such that*

$$\mathsf{SomeExt}(X) \approx_\varepsilon Y.$$

We emphasize that $\mathsf{AYG\text{-}SR}$-sources *are* convSR-sources with one good block. In particular, this means we can also achieve somewhere-extraction of 1-out-of-L convSR-sources from online SHELA sources. These sources are already well-suited for several applications.

In this section, we consider the application of the deterministic somewhere-extractors designed in Section 3.2 to extract $\mathsf{AYG\text{-}SR}$-sources (and hence 1-out-of-L convSR-sources) from online SHELA sources.

We proceed to state the results we obtain for online SHELA sources. As before, we consider two main settings: Honest blocks with high enough min-entropy, or honest blocks with arbitrarily low linear min-entropy. We begin with the former.

Theorem 42. *There exists a small enough constant $\gamma > 0$ such that for n large enough and all $2 \leq t \leq \ell \leq \mathrm{poly}(n)$ there exists an explicit $(t-1, \ell-1, \varepsilon')$-AYG-somewhere-extractor $\mathsf{SomeExt} : \{0,1\}^{n \cdot \ell} \to \{0,1\}^{m \cdot (\ell-1)}$ for $\mathsf{onSHELA}_{n,k',t,\ell}$ with $k' = (1-\gamma)n$, $m = \frac{(1-7\gamma)n}{3} - \log \ell$, and $\varepsilon' = 2(t-1) \cdot 2^{-\gamma n}$.*

Proof. In order to prove the theorem, we make use of the function $\mathsf{SomeExt} : \{0,1\}^{n \cdot \ell} \to \{0,1\}^{m \cdot (\ell-1)}$ defined in Section 3.2.1. All the parameters stay the same except for the output length m, which we set to be $m = \frac{(1-7\gamma)n}{3} - \log \ell$ instead. This is still a valid choice of parameters since m became smaller.

Let $X \in \mathsf{onSHELA}_{n,k',t,\ell}$ with $1 \leq I_1 < I_2 < \cdots < I_t \leq \ell$ denoting the positions of the t honest blocks in X. Define $(B'_2, B'_3, \ldots, B'_\ell) = \mathsf{SomeExt}(X)$. We begin by showing that

$$B'_{I_2}, \ldots, B'_{I_{j-1}}, B'_{I_j}, I_1, \ldots, I_j \approx_{\overline{\varepsilon}=2 \cdot 2^{-\gamma n}} B'_{I_2}, \ldots, B'_{I_{j-1}}, U_m, I_1, \ldots, I_j \tag{18}$$

for all $j = 2, \ldots, t$. The proof of (18) follows the same steps as the proof of (8). First, we have

$$\begin{aligned}
&\widetilde{\mathbf{H}}_\infty(B_1, \ldots, B_{I_j - 1} | B'_{I_2}, \ldots, B'_{I_{j-1}}, I_1, \ldots, I_j) \\
&\geq \widetilde{\mathbf{H}}_\infty(B_{I_{j-1}} | B'_{I_2}, \ldots, B'_{I_{j-1}}, I_1, \ldots, I_j) \\
&\geq \widetilde{\mathbf{H}}_\infty(B_{I_{j-1}} | B'_{I_2}, \ldots, B'_{I_{j-2}}, I_1, \ldots, I_j) - m \\
&\geq \widetilde{\mathbf{H}}_\infty(B_{I_{j-1}} | B'_{I_2}, \ldots, B'_{I_{j-2}}, I_1, \ldots, I_{j-1}) - m - \log \ell \\
&= \mathbf{H}_\infty(B_{I_{j-1}}) - m - \log \ell \\
&\geq k' - m - \log \ell \\
&= \overline{k}.
\end{aligned} \tag{19}$$

The second and third inequalities follow from Lemma 5, the first equality holds because $B_{I_{j-1}}$ is independent of $B_1, \ldots, B_{I_{j-1}-1}$ and $I_1, \ldots, I_{j-1}$, and the fourth inequality follows from the min-entropy constraint on $B_{I_{j-1}}$. Combining (19), the definition of B'_{I_j}, the fact that B_{I_j} is independent of $B'_{I_2}, \ldots, B'_{I_{j-1}}, I_1, \ldots, I_j$ and satisfies $\mathbf{H}_\infty(B_{I_j}) \geq k'$, and Lemma 12 yields (18).

To wrap up the proof, we construct the desired AYG-SR-source Y by using (18) to replace the B'_{I_j}'s by uniformly distributed random variables. From (18) and Lemma 2, we conclude that for every j there exists Q_j coupled with $B'_{I_2}, \ldots, B'_{I_j}, I_2, \ldots, I_j$ such that

$$(Q_j | B'_{I_2} = b_2, \ldots, B'_{I_{j-1}} = b_{j-1}, I_2 = i_2, \ldots, I_j = i_j)$$

is distributed as U_m for every valid fixing, and $\Pr[B'_{I_j} \neq Q_j] \leq \overline{\varepsilon}$. Consider $Y = (Y_2, \ldots, Y_\ell)$ obtained from $(B'_2, \ldots, B'_\ell)$ by replacing each B'_{I_j} by Q_j (and leaving the remaining random variables as is). Since

$$B'_{I_2}, \ldots, B'_{I_{j-1}}, Q_j, I_1, \ldots, I_j = B'_{I_2}, \ldots, B'_{I_{j-1}}, U_m, I_1, \ldots, I_j$$

by definition of the Q_j's, it holds that Y is a $(T = t-1, L = \ell-1, m)$-AYG-SR-source. Moreover, by construction of Y, we have that

$$\Pr[\mathsf{SomeExt}(X) \neq Y] \leq \sum_{j=2}^{t} \Pr[B'_{I_j} \neq Q_j] \leq (t-1) \cdot \overline{\varepsilon},$$

and hence $\mathsf{SomeExt}(X) \approx_{(t-1) \cdot \overline{\varepsilon}} Y$, as desired. □

We now present the result for honest blocks with arbitrarily low linear min-entropy. The proof is very similar to that of Theorem 36.

Theorem 43. *For every constant $\delta > 0$ there exist constants $a_1, a_2, a_3 > 0$ such that for n large enough and all $2 \leq t \leq \ell \leq \mathrm{poly}(n)$ there exists an explicit (T, L, ε')-*AYG*-somewhere-extractor $\mathsf{SomeExt} : \{0,1\}^{n \cdot \ell} \to \{0,1\}^{m \cdot L}$ for $\mathsf{onSHELA}_{n,k',t,\ell}$ with $k' = \delta n$, $m = a_1 \cdot n - \log \ell$, $\varepsilon' = 2(t-1)2^{-a_2 \cdot n}$, $T = t - 1$, and $L = a_3 \cdot \ell$.*

Proof. In order to prove the theorem, we make use of the function $\mathsf{SomeExt} : \{0,1\}^{n \cdot \ell} \to \{0,1\}^{m \cdot L}$ defined in Section 3.2.2. All the parameters stay the same except for the output length m, which we set to be $m = \frac{(1-3\gamma)n'}{3} - \log \ell$ instead. This is still a valid choice of parameters since m became smaller.

Let $X^1 = (\mathsf{SomeCond}(B_1), \mathsf{SomeCond}(B_2), \ldots, \mathsf{SomeCond}(B_\ell))$, and recall that we have

$$\mathsf{SomeCond}(B_{I_j}) \approx_{2^{-\rho n'}} Y^j$$

for $j = 1, \ldots, t$, where $Y^j \in \mathsf{convSE}_{1,b,n',(1-\gamma)n'}$. Furthermore, it holds that B_{I_j} is independent of $B_1, \ldots, B_{I_j - 1}, I_1, \ldots, I_j$. Therefore, by a reasoning analogous to the proof of Lemma 37, we can assume that $X^1_{I_j}$ is distributed as Y^j for $j = 1, \ldots, t$ by paying an additive $t \cdot 2^{-\rho n'}$ penalty in statistical distance.

Define $X^1_i = (B_{i1}, \ldots, B_{ib})$. Our goal now is to show that

$$\Delta(B'_{I_r J_r}; U_m | B'_{I_2 J_2}, \ldots, B'_{I_{r-1} J_{r-1}}, I_2, \ldots, I_r, J_2, \ldots, J_r) \leq \overline{\varepsilon} = 2 \cdot 2^{-\gamma n'} \tag{20}$$

for $r = 2, \ldots, t$, where $B'_{ij} = \mathsf{Ext}_i((B_{i'j'})_{i'<i, j' \in [b]}, B_{ij})$ as in (12). If (20) holds for all r, then the reasoning from the final part of the proof of Theorem 42 shows that there exists a (T, L, m)-*AYG*-*SR*-source Y such that

$$\mathsf{SomeExt}(X) \approx_{(t-1)\cdot\overline{\varepsilon} + t \cdot 2^{-\rho n'}} Y.$$

Note that the positions of the uniform blocks in Y are given by (I_r, J_r) for $r = 2, \ldots, t$.

We proceed similarly to the proof of Lemma 38. We have

$$\begin{aligned}
&\widetilde{\mathbf{H}}_\infty((B_{ij})_{i<I_r, j\in[b]} | B'_{I_2 J_2}, \ldots, B'_{I_{r-1} J_{r-1}}, I_2, \ldots, I_r, J_2, \ldots, J_r) \\
&\geq \widetilde{\mathbf{H}}_\infty(B_{I_{r-1} J_{r-1}} | B'_{I_2 J_2}, \ldots, B'_{I_{r-1} J_{r-1}}, I_2, \ldots, I_r, J_2, \ldots, J_r) \\
&\geq \widetilde{\mathbf{H}}_\infty(B_{I_{r-1} J_{r-1}} | B'_{I_2 J_2}, \ldots, B'_{I_{r-2} J_{r-2}}, I_2, \ldots, I_r, J_2, \ldots, J_r) - m \\
&\geq \widetilde{\mathbf{H}}_\infty(B_{I_{r-1} J_{r-1}} | B'_{I_2 J_2}, \ldots, B'_{I_{r-2} J_{r-2}}, I_2, \ldots, I_{r-1}, J_2, \ldots, J_r) - m - \log \ell \\
&= \widetilde{\mathbf{H}}_\infty(B_{I_{r-1} J_{r-1}} | J_{r-1}) - m \\
&\geq (1-\gamma)n' - m \\
&= \overline{k}.
\end{aligned} \tag{21}$$

The second and third inequalities follow from Lemma 5. The first equality holds because the pair $(B_{I_{r-1} J_{r-1}}, J_{r-1})$ is independent of the remaining random variables in the expression on the line above. The fourth inequality is a consequence of our prior assumption that $X^1_{I_{r-1}}$ is distributed as Y^{r-1}, and $\widetilde{\mathbf{H}}_\infty(Y^{r-1}_{J_{r-1}} | J_{r-1}) \geq (1-\gamma)n'$.

Define

$$\begin{aligned}
Z &= (B_{ij})_{i<I_r, j\in[b]} \\
W &= B'_{I_2 J_2}, \ldots, B'_{I_{r-1} J_{r-1}}, I_2, \ldots, I_r, J_2, \ldots, J_{r-1}
\end{aligned}$$

$$S = B_{I_r J_r}.$$

Then, we conclude that (20) is equivalent to

$$\mathsf{Ext}_r(Z, S), W, J_r \approx_{\bar{\varepsilon}} U_m, W, J_r,$$

which can be seen to hold by (21) and the properties of Ext_r analogously to the proof of (17). Finally, setting, say, $a_1 = \frac{\beta \cdot (1-3\gamma)}{3}$, $a_2 = \frac{\beta \cdot \min(\gamma, \rho)}{2}$, and $a_3 = b$ as in Section 3.2.2 yields the desired result. □

5 Lower Bounds for Deterministic Somewhere-Extraction from Weak Sources

In this section, we study lower bounds for somewhere-extractors that work for the general class of weak $(\tilde{n}, k)$-sources (we use $\tilde{n}$ to avoid confusion with the block length n of a SHELA source). Here, we are mostly interested in lower bounds on the number of output blocks generated by such somewhere-extractors with respect to the length $\tilde{n}$ of a source, the length m of an output block, and the error ε of the somewhere-extractor.

The only known construction of a somewhere-extractor for general $(\tilde{n}, k)$-sources described in Lemma 17 requires 2^d blocks, where d is the seed length of the underlying strong extractor/non-malleable extractor. As stated in Lemma 10, it holds that $d \geq \log(\tilde{n} - k) + 2\log(1/\varepsilon) + O(1)$ for every extractor, and so the somewhere-random source output by the somewhere-extractor from Lemma 17 has

$$L = \Omega\left(\frac{\tilde{n} - k}{\varepsilon^2}\right)$$

blocks. We remark that a probabilistic argument with a random function yields somewhere-extraction with the same number of output blocks.

The discussion in the previous paragraph leads to the following natural questions: *Is it possible to do better than Lemma 17 for $(\tilde{n}, k)$-sources? In particular, is it possible to obtain a number of output blocks comparable to that obtained from SHELA sources?*

We present some results that aim to answer this question in several parameter regimes. The first result comes from the observation that the basic argument for impossibility of deterministic extraction yields a non-trivial lower bound on the number of output blocks whenever the min-entropy requirement k is not very large.

Theorem 44. *Suppose $F : \{0,1\}^{\tilde{n}} \to \{0,1\}^{m \cdot L}$ is a $(1, L, \varepsilon)$-somewhere extractor for $(\tilde{n}, k)$-sources with $\varepsilon \leq 1 - 2^{-c}$ for some $1 \leq c \leq m$ (i.e., ε is not trivial). Then, it holds that*

$$L \geq \frac{\tilde{n} - k}{c}.$$

Proof. Without loss of generality, we may assume that $m = c$. By an averaging argument, there exists $y^\star \in \{0,1\}^{c \cdot L}$ such that $|F^{-1}(y^\star)| \geq 2^{\tilde{n} - c \cdot L}$. Let X be uniformly distributed over $F^{-1}(y^\star)$.

Fix an arbitrary $Z \in \mathsf{convSR}_{1,L,m}$. We may write $Z = \sum_{i=1}^{L} \pi_i Z^{(i)}$ for some $Z^{(i)} = (Z_1^{(i)}, \ldots, Z_L^{(i)}) \in \mathsf{SR}_{1,L,m}$ with $Z_i^{(i)} = U_c$ and $\pi_i \geq 0$ such that $\sum_{i=1}^{L} \pi_i = 1$. This follows from the fact that a convex combination of somewhere-random sources with uniform blocks in position i is also a somewhere-random source with a uniform block in position i. We show that $F(X) \not\approx_\varepsilon Z$. In fact, we have

$$\Delta(F(X); Z) = \frac{1}{2} \sum_{y \in \{0,1\}^{c \cdot L}} |\Pr[F(X) = y] - \Pr[Z = y]|$$

$$
\begin{aligned}
&= \frac{1}{2}\left(1 - \Pr[Z = y^\star] + \sum_{y \neq y^\star} \Pr[Z = y]\right) \\
&= 1 - \Pr[Z = y^\star] \\
&= 1 - \sum_{i=1}^{L} \pi_i \Pr[Z^{(i)} = y^\star] \\
&\geq 1 - \sum_{i=1}^{L} \pi_i \Pr[Z_i^{(i)} = y_i^\star] \\
&= 1 - 2^{-c} \\
&\geq \varepsilon,
\end{aligned}
$$

where the first inequality follows from the fact that $\Pr[Z^{(i)} = y^\star] \leq \Pr[Z_i^{(i)} = y_i^\star]$, and the fifth equality holds because $Z_i^{(i)} = U_c$. Since Z was arbitrary and $\mathbf{H}_\infty(X) \geq \tilde{n} - c \cdot L$, it must be the case that $k \geq \tilde{n} - c \cdot L$. Rearranging the expression yields the desired result. □

The lower bound from Theorem 44 is already enough to yield a separation between somewhere-extraction of SHELA and comparable $(\tilde{n}, k)$-sources whenever the min-entropy requirement k is not extremely large. Consider a SHELA source with constant entropy rate and ℓ blocks, each of length $n = \tilde{n}/\ell$ (so that the total length of the source is $\tilde{n}$). The constructions from Theorems 33 and 36 applied to the SHELA source lead to convSR-sources with $L = O(\ell)$ blocks with small error and large output block length if honest blocks have some constant entropy rate. In particular, L *does not depend directly on the input block length* n. On the other hand, the lower bound from Theorem 44 forces that $L = \Omega(\tilde{n} - k) = \Omega(n \cdot \ell)$ for convSR-sources extracted from $(\tilde{n}, k)$-sources, even with error $\varepsilon = 1/2$ (assuming $k/\tilde{n}$ is constant).

The second result is a disperser-based lower bound on the number of output blocks L. This bound is considerably stronger than the one in Theorem 44 whenever the output block length m is not very small and the error ε is small.

Theorem 45. *Suppose $F : \{0,1\}^{\tilde{n}} \to \{0,1\}^{m \cdot L}$ is a $(1, L, \varepsilon)$-somewhere extractor for $(\tilde{n}, k)$-sources with $\varepsilon \leq 1/2$ and $L \leq \frac{(1 - \max(\varepsilon, 2^{-m}))2^m}{2}$. Then, it holds that*

$$
L = \Omega\left(\frac{\tilde{n} - k}{\max(\varepsilon, 2^{-m})}\right).
$$

Proof. First, note that we may assume without loss of generality that $\varepsilon \geq 2^{-m}$. The result follows by relating F to a (k, ε)-disperser and employing Lemma 14.

We can write F as $F = (F_1, \ldots, F_L)$ for some functions $F_i : \{0,1\}^{\tilde{n}} \to \{0,1\}^m$. Define a function $G : \{0,1\}^{\tilde{n}} \times [L] \to \{0,1\}^m$ as $G(x, i) = F_i(x)$. We claim that G is a (k, ε)-disperser with seed length $\log L$. In fact, we know that for every $(\tilde{n}, k)$-source X it holds that $Y = F(X)$ is ε-close to some $Z = (Z_1, \ldots, Z_L) \in \mathsf{convSR}_{1,L,m}$. As in the proof of Theorem 44, we may write $Z = \sum_{i=1}^{L} \pi_i Z^{(i)}$ for some $Z^{(i)} = (Z_1^{(i)}, \ldots, Z_L^{(i)}) \in \mathsf{SR}_{1,L,m}$ with $Z_i^{(i)} = U_m$ and $\pi_i \geq 0$ such that $\sum_{i=1}^{L} \pi_i = 1$. This follows from the fact that a convex combination of somewhere-random sources with uniform blocks in position i is also a somewhere-random source with a uniform block in position i.

Fix $\mathcal{S} \subseteq \{0,1\}^m$ such that $|\mathcal{S}| \geq \varepsilon 2^m$. We now proceed to show $\Pr[G(X, U_L) \in \mathcal{S}] > 0$. We have

$$
\Pr[G(X, U_L) \in \mathcal{S}] = \frac{1}{L} \sum_{i=1}^{L} \Pr[F_i(X) \in \mathcal{S}]
$$

$$\geq \frac{1}{L} \cdot \Pr[\exists i : F_i(X) \in \mathcal{S}]$$
$$> \frac{1}{L} \cdot (\Pr[\exists i : Z_i \in \mathcal{S}] - \varepsilon),$$

where the first inequality follows from the union bound, and the second inequality holds because $F(X) \approx_\varepsilon Z$.[12] The desired property now follows if we show that $\Pr[\exists i : Z_i \in \mathcal{S}] \geq \varepsilon$. In fact, we have

$$\begin{aligned}
\Pr[\exists i : Z_i \in \mathcal{S}] &= \sum_{j=1}^{L} \pi_j \Pr[\exists i : Z_i^{(j)} \in \mathcal{S}] \\
&\geq \sum_{j=1}^{L} \pi_j \Pr[Z_j^{(j)} \in \mathcal{S}] \\
&\geq \sum_{j=1}^{L} \pi_j \cdot \varepsilon \\
&= \varepsilon,
\end{aligned}$$

as desired. The second inequality follows from the fact that $Z_j^{(j)} = U_m$ and $|\mathcal{S}| \geq \varepsilon \cdot 2^m$. Since X and $\mathcal{S}$ were arbitrary, this shows that G is a (k, ε)-disperser. □

Referring again to the comparison between SHELA and weak $(\tilde{n}, k)$-sources above, if we want to extract a 1-out-of-L convSR-source with block length $\Omega(n)$ from the weak source with error $2^{-\Omega(n)}$, as is possible for the relevant SHELA source, then Theorem 45 forces that $L = \tilde{n} \cdot 2^{\Omega(n)} = \ell \cdot n 2^{\Omega(n)}$. On the other hand, the convSR-source we extract from the relevant t-out-of-ℓ SHELA source only has $O(\ell)$ blocks.

While Theorems 44 and 45 imply strong separation between SHELA and weak sources for any conceivable application, they do not yield useful lower bounds for some regimes of parameters. For example, in the easiest setting for somewhere-extraction, when the min-entropy requirement k is very large (say, $k = \tilde{n} - 1$) and the output block length is very small (say, $m = 1$), both theorems only give a trivial $\Omega(1)$ lower bound on L, *even when ε is exponentially small in $\tilde{n}$.* On the other hand, the number of output blocks in the somewhere-extractor obtained from Lemma 17 instantiated with an optimal strong extractor scales as $1/\varepsilon^2$ even when $k = \tilde{n} - 1$ and $m = 1$. We believe it is not possible to improve significantly on the basic construction from Lemma 17, and so we put forth the following conjecture.

Conjecture 46. *Suppose $F : \{0,1\}^{\tilde{n}} \to \{0,1\}^{m \cdot L}$ is a (T, L, ε)-somewhere extractor for $(\tilde{n}, k)$-sources. Then, there exists a constant $c > 0$ such that if $\varepsilon \leq c$, we have*

$$L = \Omega\left(\frac{\tilde{n} - k}{\varepsilon^2}\right). \tag{22}$$

We do not prove Conjecture 46 and leave it as an open problem. Nevertheless, we prove a weaker lower bound on L in a similar spirit to (22) under a stronger property than somewhere-extraction, which is still satisfied by the construction from Lemma 17. This result can be regarded both as a first step towards a full proof of Conjecture 46, and a non-trivial lower bound on L (under this stronger property) that scales with ε and holds even when k is large and m is small. Before we

[12] Recall that $W \approx_\varepsilon Z$ is equivalent to $\Delta(W; Z) < \varepsilon$.

state our result, we must first define the alternative notion of somewhere-extraction. Observe that the construction of F from Lemma 17 actually ensures that for every $(\tilde{n}, k)$-source X it holds that $F(X)$ is ε-close to an element of $\mathsf{SR}_{T,L,m}$, instead of only a convex combination of such elements. We call a function that satisfies this for all $(\tilde{n}, k)$-sources a *strong* (T, L, ε, k)-somewhere extractor.

We may think of a strong $(1, L, \varepsilon, k)$-somewhere-extractor $F : \{0,1\}^{\tilde{n}} \to \{0,1\}^L$ as a family of L functions $F_1, \ldots, F_L$ such that for every $(\tilde{n}, k)$-source X, there is F_i such that $F_i(X) \approx_\varepsilon U_1$. Therefore, in order to show such a function F is not a strong somewhere-extractor, we must show the existence of an $(\tilde{n}, k)$-source X that is "bad" for all F_i's, in the sense that $F_i(X) \not\approx_\varepsilon U_1$ for every i. As previously discussed, existing techniques used in proving lower bounds for extractors cannot be applied to obtain similar lower bounds for strong somewhere-extractors. We use a fundamentally different technique to prove the following lower bound on L for strong somewhere-extractors.

Theorem 47. *Suppose $F : \{0,1\}^{\tilde{n}} \to \{0,1\}^{m \cdot L}$ is a strong $(1, L, \varepsilon, k)$-somewhere extractor for $k \leq \tilde{n} - 1$. Then, there exists an absolute constant $c > 0$ such that if $\varepsilon < c$, we have*

$$L = \Omega\left(\log\left(\frac{1}{\max(\varepsilon, 2^{-k})}\right)\right). \tag{23}$$

We require the following auxiliary lemmas in the proof of Theorem 47.

Lemma 48. *Fix $\delta \in (0, 1/10)$, a set $\mathcal{S} \subseteq \{0,1\}^{\tilde{n}}$ such that $|\mathcal{S}| \geq 4$, X uniformly distributed over $\mathcal{S}$, and a function $f : \{0,1\}^{\tilde{n}} \to \{0,1\}$. Suppose that $f(X) \approx_{\delta/4} U_1$. Let $\mathcal{S}'$ be obtained from $\mathcal{S}$ by choosing $\mathcal{B} \subseteq f^{-1}(1) \cap \mathcal{S}$ of size $|\mathcal{B}| = \lceil \delta \cdot |\mathcal{S}| \rceil$ and setting $\mathcal{S}' = \mathcal{S} \setminus \mathcal{B}$. Then, if X' is uniformly distributed over $\mathcal{S}'$, we have*

$$f(X') \not\approx_{\delta/4} U_1.$$

Lemma 49. *Fix $\alpha \in (0,1)$, a function $f : \{0,1\}^{\tilde{n}} \to \{0,1\}$ and a source X uniform over a set $\mathcal{S} \subseteq \{0,1\}^{\tilde{n}}$ such that $|\mathcal{S}| \geq 4$ and $f(X) \not\approx_\alpha U_1$. Let $\mathcal{S}' = \mathcal{S} \setminus \mathcal{B}$ for some $\mathcal{B} \subseteq \mathcal{S}$ satisfying $|\mathcal{B}| = \lceil \beta \cdot |\mathcal{S}| \rceil$ with $\alpha - 2\beta - 2/|\mathcal{S}| > 0$. Then, if X' is uniformly distributed over $\mathcal{S}'$, we have*

$$f(X') \not\approx_{\alpha - 2\beta - 2/|\mathcal{S}|} U_1.$$

We now use Lemmas 48 and 49 to prove Theorem 47.

Proof of Theorem 47. Without loss of generality we may assume that $m = 1$ and $\varepsilon \geq 2^{-k}$. We can also assume that $k = \Omega(\tilde{n})$. In fact, if $k = o(\tilde{n})$, Lemma 14 already gives an $\Omega(\tilde{n})$ lower bound for L, which is the best we can obtain with Theorem 47 in that setting too.

Fix some function $F : \{0,1\}^{\tilde{n}} \to \{0,1\}^L$. Then, we can write $F = (F_1, \ldots, F_L)$ for some functions $F_i : \{0,1\}^{\tilde{n}} \to \{0,1\}$. Suppose that

$$L < \frac{1}{100} \log(1/\varepsilon). \tag{24}$$

We show that if L satisfies (24), then F cannot be a strong $(1, L, \varepsilon, k)$-somewhere extractor. In order to do this, we will iteratively define sets $\mathcal{S}_0 \supseteq \mathcal{S}_1 \supseteq \cdots \supseteq \mathcal{S}_L$ with associated sources $X_0, X_1, \ldots, X_L$ such that X_i is uniformly distributed over $\mathcal{S}_i$. Our goal is to ensure that $|\mathcal{S}_L| \geq 2^k$ and $F_i(X_L) \not\approx_\varepsilon U_1$ for every $i = 1, \ldots, L$.

We now describe how to define the sets $\mathcal{S}_i$, and hence the corresponding sources X_i. With some hindsight, consider positive real numbers $\delta_1, \delta_2, \ldots, \delta_L$ such that $\delta_1 = \varepsilon \cdot 17^L$ and $\delta_i = \delta_{i-1}/17$ for $i = 2, \ldots, L$. We begin by setting $\mathcal{S}_0 = \{0,1\}^{\tilde{n}}$. Then, iteratively for each $i = 1, \ldots, L$ we proceed as follows:

- If $F_i(X_{i-1}) \not\approx_{\delta_i/4} U_1$, set $\mathcal{S}_i = \mathcal{S}_{i-1}$;
- Else, choose $\mathcal{B} \subseteq F_i^{-1}(1) \cap \mathcal{S}_{i-1}$ of size $|\mathcal{B}| = \lceil \delta_i \cdot |\mathcal{S}_{i-1}| \rceil$, and set $\mathcal{S}_i = \mathcal{S}_{i-1} \setminus \mathcal{B}$.

All that remains to show is that $\mathcal{S}_L$ is large enough and X_L is appropriately biased against $F_1, \ldots, F_L$, provided that ε is smaller than some absolute constant. We begin by lower bounding $|\mathcal{S}_L|$. First, observe that (24) implies that $\delta_i < 0.1$ for all i if $\varepsilon < c$ for a small enough constant $c > 0$. Also, note that

$$|\mathcal{S}_i| \geq |\mathcal{S}_{i-1}| - \lceil \delta_i \cdot |\mathcal{S}_{i-1}| \rceil \geq (1 - \delta_i)|\mathcal{S}_{i-1}| - 1 \tag{25}$$

for all $i \geq 1$. Using this, for large enough n we obtain

$$\begin{aligned}
|\mathcal{S}_L| &\geq |\mathcal{S}_0| \prod_{i=1}^{L} (1 - \delta_i) - L \\
&= 2^n \prod_{i=1}^{L} (1 - \delta_i) - L \\
&\geq 2^n \cdot \exp\left(-4 \sum_{i=1}^{L} \delta_i \right) - L \\
&= 2^n \cdot \exp\left(-4\delta_1 \sum_{i=0}^{L-1} 17^{-i} \right) - L \\
&\geq 2^n \cdot \exp\left(-4\delta_1 \sum_{i=0}^{\infty} 17^{-i} \right) - L \\
&\geq 2^n \cdot \exp(-0.5) - L \\
&\geq 2^k.
\end{aligned}$$

In the derivation above, the first inequality follows by repeated application of (25), the first equality holds by the definition of $\mathcal{S}_0$, the second inequality is a consequence of the fact that $1 - x \geq \exp(-4x)$ for $x \leq 0.9$ and $\delta_i < 0.1$ for all i, the second equality follows from the definition of δ_i, the fourth inequality holds because $\delta_1 < 0.1$, and the final inequality follows from the fact that $L = O(\tilde{n})$ (which is a consequence of (24) and $\varepsilon \geq 2^{-k}$), and that $k \leq \tilde{n} - 1$. As a result, we have that X_L is an (n, k)-source.

To conclude the proof, we show that

$$F_i(X_L) \not\approx_{\varepsilon} U_1 \tag{26}$$

for all $i = 1, \ldots, L$. First, we argue that

$$F_i(X_i) \not\approx_{\delta_i/4} U_1 \tag{27}$$

for every $i \geq 1$. To see that (27) holds, note that either $F_i(X_{i-1}) \not\approx_{\delta_i/4} U_1$, in which case we are done since then $\mathcal{S}_i = \mathcal{S}_{i-1}$ (and hence $X_i = X_{i-1}$), or $F_i(X_{i-1}) \approx_{\delta_i/4} U_1$. In the second case, we are in a condition to apply Lemma 48 with $\mathcal{S}_{i-1}$ and $\mathcal{S}_i$ in place of $\mathcal{S}$ and $\mathcal{S}'$, respectively, by definition of $\mathcal{S}_i$ and since $\delta_i < 0.1$, which immediately implies (27).

Next, we repeatedly apply Lemma 49 with the help of (27) to prove (26). As a starting point, we show that $F_{i-1}(X_i) \not\approx_{\varepsilon} U_1$. To do this, we apply Lemma 49 with $\mathcal{S}_{i-1}$, $\mathcal{S}_i$, f_{i-1}, $\delta_{i-1}/4$, and δ_i

in place of $\mathcal{S}$, $\mathcal{S}'$, f, α, and β. We claim that the conditions of Lemma 49 are satisfied for large enough $\tilde{n}$. In fact, note that $|\mathcal{S}_i| \geq |\mathcal{S}_L| \geq 2^k \geq 4$ for all i, and $\delta_i = \delta_{i-1}/17 < \delta_{i-1}/8 - 1/|\mathcal{S}_{i-1}|$, since $\delta_i \geq 17\varepsilon$, $\varepsilon \geq 2^{-k}$, and $|\mathcal{S}_i| \geq 2^k$ for all i. As a result, we conclude that

$$F_{i-1}(X_i) \not\approx_{\delta_{i-1}/4 - 2\delta_i - 2^{-k+1}} U_1 \tag{28}$$

for all $i = 2, \ldots, L$. In particular, this implies that

$$F_{L-1}(X_L) \not\approx_{\varepsilon} U_1,$$

since $\delta_{i-1}/4 - 2\delta_i - 2^{-k+1} > \varepsilon$, by the choice of δ_i's and the constraint $\varepsilon \geq 2^{-k/2}$. A further application of Lemma 49 using (28) with $\mathcal{S}_{i-1}$, $\mathcal{S}_i$, F_{i-2}, $\delta_{i-2}/4 - 2\delta_{i-1} - 2^{-k+1}$, and δ_i in place of $\mathcal{S}$, $\mathcal{S}'$, f, α, and β, respectively, leads to

$$f_{i-2}(X_i) \not\approx_{\delta_{i-2}/4 - 2(\delta_{i-1}+\delta_i) - 2\cdot 2^{-k+1}} U_1$$

for $i = 3, \ldots, L$. Similarly to what was observed before, this implies that

$$F_{L-2}(X_L) \not\approx_{\varepsilon} U_1.$$

Continuing in this fashion, from Lemma 49 we obtain

$$F_j(X_i) \not\approx_{\delta_j/4 - 2\sum_{r=j+1}^{i} \delta_r - (i-j)2^{-k+1}} U_1 \tag{29}$$

for $1 \leq j \leq i \leq L$. Such applications of Lemma 49 are valid because $|\mathcal{S}_i| \geq |\mathcal{S}_L| \geq 2^k \geq 4$ for all i and large enough $\tilde{n}$, and since $\delta_j/4 - 2\sum_{r=j+1}^{i} \delta_r - (i-j)2^{-k+1} > 0$. In fact, we have

$$\begin{aligned}
\delta_j/4 - 2\sum_{r=j+1}^{i} \delta_r - (i-j)2^{-k+1} &\geq \delta_j/4 - 2\delta_j \sum_{r=1}^{\infty} 17^{-r} - (i-j)2^{-k+1} \\
&= \delta_j/8 - (i-j)2^{-k+1} \\
&= \frac{17\varepsilon}{8} \cdot 17^{L-j} - (i-j)2^{-k+1} \\
&\geq \varepsilon\left(\frac{17}{8} \cdot 17^{L-j} - 2(L-j)\right) \\
&> \varepsilon,
\end{aligned} \tag{30}$$

where the first inequality and second equality follow by definition of the δ_i's, the second inequality holds because $\varepsilon \geq 2^{-k}$ and $j \leq i \leq L$, and the third inequality is true for all $j \leq L$.

Finally, combining (29) with (30) implies (26). Since X_L is an $(\tilde{n}, k)$-source, it follows that F cannot be a strong $(1, L, \varepsilon, k)$-somewhere extractor. This means any such strong somewhere-extractor must have $L \geq \frac{1}{100}\log(1/\varepsilon)$, as desired. □

Remark 2. *The condition $k \leq \tilde{n} - 1$ in the statement of Theorem 47 can be replaced by $k \leq \tilde{n} - c'$ for any real constant $c' > 0$, with the caveat that the hidden constant in* (23) *depends on c'.*

5.1 Proof of Lemma 48

In this section, we prove Lemma 48, which we restate here for convenience.

Lemma 50 (Lemma 48, restated)**.** *Fix $\delta \in (0, 1/10)$, a set $\mathcal{S} \subseteq \{0,1\}^n$ such that $|\mathcal{S}| \geq 4$, X uniformly distributed over $\mathcal{S}$, and a function $f : \{0,1\}^n \to \{0,1\}$. Suppose that $f(X) \approx_{\delta/4} U_1$. Let $\mathcal{S}'$ be obtained from $\mathcal{S}$ by choosing $\mathcal{B} \subseteq f^{-1}(1) \cap \mathcal{S}$ of size $|\mathcal{B}| = \lceil \delta \cdot |\mathcal{S}| \rceil$ and setting $\mathcal{S}' = \mathcal{S} \setminus \mathcal{B}$. Then, if X' is uniformly distributed over $\mathcal{S}'$, we have*

$$f(X') \not\approx_{\delta/4} U_1.$$

Proof. First, note that we can indeed pick such a set $\mathcal{B} \subseteq f^{-1}(1) \cap \mathcal{S}$. This is because $|f^{-1}(1) \cap \mathcal{S}| > |\mathcal{S}|/2 - \delta \cdot |\mathcal{S}|/4 > |\mathcal{S}|/3$ since $f(X) \approx_{\delta/4} U_1$, and $|\mathcal{B}| = \lceil \delta \cdot |\mathcal{S}| \rceil \leq \lceil |\mathcal{S}|/10 \rceil < |\mathcal{S}|/3$. In order to prove the lemma, it now suffices to lower bound $|\Pr[f(X') = 1] - 1/2|$ appropriately. We have

$$\begin{aligned}
|\Pr[f(X') = 1] - 1/2| &= \left| \frac{|f^{-1}(1) \cap \mathcal{S}'|}{|\mathcal{S}'|} - 1/2 \right| \\
&= \left| \frac{|f^{-1}(1) \cap \mathcal{S}| - \lceil \delta \cdot |\mathcal{S}| \rceil}{|\mathcal{S}| - \lceil \delta \cdot |\mathcal{S}| \rceil} - 1/2 \right| && (31) \\
&= \left| \frac{\frac{|f^{-1}(1) \cap \mathcal{S}|}{|\mathcal{S}|} - \frac{\lceil \delta \cdot |\mathcal{S}| \rceil}{|\mathcal{S}|}}{1 - \frac{\lceil \delta \cdot |\mathcal{S}| \rceil}{|\mathcal{S}|}} - 1/2 \right| \\
&= \left| \frac{\frac{|f^{-1}(1) \cap \mathcal{S}|}{|\mathcal{S}|} - 1/2 - \frac{\lceil \delta \cdot |\mathcal{S}| \rceil}{2|\mathcal{S}|}}{1 - \frac{\lceil \delta \cdot |\mathcal{S}| \rceil}{|\mathcal{S}|}} \right| \\
&\geq \left| \frac{|f^{-1}(1) \cap \mathcal{S}|}{|\mathcal{S}|} - 1/2 - \frac{\lceil \delta \cdot |\mathcal{S}| \rceil}{2|\mathcal{S}|} \right| && (32) \\
&\geq \delta/2 - \delta/4 && (33) \\
&= \delta/4,
\end{aligned}$$

as desired. In the derivation above, we have that (31) follows from the fact that $\mathcal{B} \subseteq f^{-1}(1) \cap \mathcal{S}$ and $\mathcal{S}' = \mathcal{S} \setminus \mathcal{B}$, and (32) holds because

$$0 < \frac{\lceil \delta \cdot |\mathcal{S}| \rceil}{|\mathcal{S}|} < \delta + \frac{1}{|\mathcal{S}|} \leq 3\delta/2 < 1.$$

Finally, (33) holds because

$$\frac{\lceil \delta \cdot |\mathcal{S}| \rceil}{2 \cdot |\mathcal{S}|} \geq \delta/2$$

and

$$\left| \frac{|f^{-1}(1) \cap \mathcal{S}|}{|\mathcal{S}|} - 1/2 \right| < \delta/4,$$

since $f(X) \approx_{\delta/4} U_1$ by hypothesis. □

5.2 Proof of Lemma 49

In this section, we prove Lemma 49, which we restate here for convenience.

Lemma 51 (Lemma 49, restated)**.** *Fix $\alpha \in (0,1)$, a function $f : \{0,1\}^n \to \{0,1\}$ and a source X uniform over a set $\mathcal{S} \subseteq \{0,1\}^n$ such that $|\mathcal{S}| \geq 4$ and $f(X) \not\approx_\alpha U_1$. Let $\mathcal{S}' = \mathcal{S} \setminus \mathcal{B}$ for some $\mathcal{B} \subseteq \mathcal{S}$ satisfying $|\mathcal{B}| = \lceil \beta |\mathcal{S}| \rceil$ with $\beta < \alpha/2 - 1/|\mathcal{S}|$. Then, if X' is uniformly distributed over $\mathcal{S}'$, we have*

$$f(X') \not\approx_{\alpha - 2\beta - 2/|\mathcal{S}|} U_1.$$

Proof. Without loss of generality, suppose that $|f^{-1}(1) \cap \mathcal{S}| \geq |\mathcal{S}|/2$. Define $\mathcal{B}_1 = f^{-1}(1) \cap \mathcal{B}$. Similarly to the proof of Lemma 48, it suffices to lower bound the quantity $|\Pr[f(X') = 1] - 1/2|$ appropriately. We have

$$\begin{aligned} |\Pr[f(X') = 1] - 1/2| &= \left| \frac{|f^{-1}(1) \cap \mathcal{S}'|}{|\mathcal{S}'|} - 1/2 \right| \\ &= \left| \frac{|f^{-1}(1) \cap \mathcal{S}| - |\mathcal{B}_1|}{|\mathcal{S}| - \lceil \beta |\mathcal{S}| \rceil} - 1/2 \right| \qquad (34) \\ &= \left| \frac{|f^{-1}(1) \cap \mathcal{S}|}{|\mathcal{S}| - \lceil \beta |\mathcal{S}| \rceil} - 1/2 - \frac{|\mathcal{B}_1|}{|\mathcal{S}| - \lceil \beta |\mathcal{S}| \rceil} \right|, \qquad (35) \end{aligned}$$

where (34) follows from the definition of $\mathcal{B}_1$ and the fact that $\mathcal{B} \subseteq \mathcal{S}$ with $|\mathcal{B}| = \lceil \beta |\mathcal{S}| \rceil$. To finalize the proof, we first recall that we assumed $|f^{-1}(1) \cap \mathcal{S}| \geq |\mathcal{S}|/2$. This means that

$$\frac{|f^{-1}(1) \cap \mathcal{S}|}{|\mathcal{S}| - \lceil \beta |\mathcal{S}| \rceil} \geq \frac{|f^{-1}(1) \cap \mathcal{S}|}{|\mathcal{S}|} \geq 1/2. \qquad (36)$$

Combining (36) with the hypothesis that $f(X) \not\approx_\alpha U_1$ allows us to conclude that

$$\frac{|f^{-1}(1) \cap \mathcal{S}|}{|\mathcal{S}| - \lceil \beta |\mathcal{S}| \rceil} - 1/2 \geq \alpha. \qquad (37)$$

On the other hand, we have

$$\begin{aligned} \frac{|\mathcal{B}_1|}{|\mathcal{S}| - \lceil \beta |\mathcal{S}| \rceil} &\leq \frac{\beta |\mathcal{S}| + 1}{|\mathcal{S}| - \lceil \beta |\mathcal{S}| \rceil} \\ &\leq 2\beta + \frac{2}{|\mathcal{S}|}, \qquad (38) \end{aligned}$$

where (38) holds because $\lceil \beta |\mathcal{S}| \rceil < |\mathcal{S}|/2$ since $\beta < 1/2 - 1/|\mathcal{S}|$. The desired result now follows by combining (35) with (37), (38), and the fact that $\alpha > 2\beta + 2/|\mathcal{S}|$ by hypothesis. □

6 Bounds for Somewhere-Amplifiable-Source Extraction from Weak Sources

The lower bounds obtained in Section 5 show that convSR-sources extracted from SHELA sources are much better (in terms of number of blocks with respect to desired extraction error) than convSR-sources extracted from weak sources. This has direct consequences in the time complexity blowup incurred when using convSR-sources in several applications, as discussed in Section 1. However, as discussed in that same section, it is possible in some scenarios to use a weaker object than convSR-sources, which we call *somewhere-amplifiable sources*, where the good independent blocks are not required to be exactly uniformly distributed. A precise definition follows.

Definition 52 (Somewhere-amplifiable source)**.** *We say $Y = (Y_1, \ldots, Y_L)$ over $\{0,1\}^{m \cdot L}$ is a (T, L, ε)*-somewhere-amplifiable source *if there exist distinct indices $i_1, \ldots, i_T$ such that $Y_{i_1}, \ldots, Y_{i_T}$ are independent and $Y_{i_j} \approx_\varepsilon U_m$ for all $j = 1, \ldots, T$. The set of all such SA sources is denoted by $\mathsf{SA}_{T,L,\varepsilon}$, and the set of all convex combinations of sources in $\mathsf{SA}_{T,L,\varepsilon}$ is denoted by $\mathsf{convSA}_{T,L,\varepsilon}$.*

Since the error required from each good block in a convSA-source is not that small (in fact, it can even be constant), one may hope to transform weak sources into convSA-sources whose number of blocks is much closer to that of convSR-sources obtained from SHELA sources, and which have blocks long enough to be used in the applications already discussed in Section 1 and later in Section 7. To this end, we define *somewhere-amplifiable source extractors* (convSA-source extractors).

Definition 53 (Somewhere-amplifiable source extractor)**.** *A function* $\mathsf{SomeExt} : \{0,1\}^{\tilde{n}} \to \{0,1\}^{m \cdot L}$ *is said to be a* $(T, L, k, \varepsilon_1, \varepsilon_2)$-somewhere-amplifiable extractor *if for every* $(\tilde{n}, k)$*-source* X *there exists* $Y \in \mathsf{convSA}_{T,L,\varepsilon_2}$ *such that*

$$\mathsf{SomeExt}(X) \approx_{\varepsilon_1} Y.$$

We begin by noting that Theorem 44 also applies to convSA-source extractors for weak sources. This shows that every such extractor (even with constant error) must have $L = \Omega(\tilde{n} - k)$. As discussed in Section 1, this already provides an efficiency separation between convSA-source extraction from weak sources and convSR-source extraction from SHELA sources.

The main result we prove in this section is a different type of separation between convSA-source extraction from weak sources and convSR-source extraction from SHELA sources. Roughly speaking, we show that if we want to extract a convSA-source with many good blocks (necessary to obtain good final error) from an $(\tilde{n}, k)$-source, then either the resulting convSA-source has too many blocks to allow for efficient construction of the publicly verifiable protocols, or the length of each block is very small, and so they may not be usable in some protocols. This is discussed for the particular case of our publicly verifiable proof system in Section 1.4. A precise statement follows.

Theorem 54. *Suppose* $F : \{0,1\}^{\tilde{n}} \to \{0,1\}^{m \cdot L}$ *is a* $(T, L, k, \varepsilon_1, \varepsilon_2)$*-somewhere-amplifiable extractor for* $\varepsilon_1 = \mathsf{negl}(\tilde{n})$*, and* $\varepsilon_2 \leq c_2$ *for some arbitrary constant* $c_2 \leq 1 - 2^{-m}$ *(so that* ε_1 *is useful for applications and* ε_2 *is non-trivial). Then, either the number of blocks* L *is superpolynomial in* $\tilde{n}$ *(and hence amplification is inefficient), or we have* $m = O(k/T)$*.*

Proof. Suppose $L \leq \mathrm{poly}(n)$ (otherwise, we are done). Fix a flat $(\tilde{n}, k)$-source X. By hypothesis, we have

$$F(X) \approx_{\varepsilon_1} Y \tag{39}$$

for some $Y \in \mathsf{convSA}_{T,L,\varepsilon_2}$. In particular, Y can be written as $Y = \sum_i \pi_i Y^{(i)}$ for $Y^{(i)} \in \mathsf{SA}_{t,L,\varepsilon_2}$.

We prove that $m = O(k/T)$ by relating $H(X)$ with $H(Y)$, and estimating the latter. First, note that

$$\begin{aligned} k &= H(X) \\ &\geq H(F(X)) \\ &\geq H(Y') - (h(\varepsilon_1) + \varepsilon_1 \cdot m \cdot L) \\ &= H(Y') - o(m). \end{aligned} \tag{40}$$

The first equality stems from the fact that X is flat, the second inequality follows from Lemma 6 applied to $F(X)$ and Y using (39), and the second equality holds because $\varepsilon_1 \cdot L = o(1)$ by hypothesis. We proceed to lower bound $H(Y)$ appropriately. Exploiting the concavity of $H(\cdot)$, it follows that

$$H(Y) \geq \sum_i \pi_i H(Y^{(i)}), \tag{41}$$

so it remains to lower bound each $H(Y^{(i)})$ term. Fix some $Y^{(i)}$, and let $i_1, \ldots, i_T$ be distinct indices such that $Y^{(i)}_{i_1}, \ldots, Y^{(i)}_{i_T}$ are independent and $Y^{(i)}_{i_j} \approx_{\varepsilon_2} U_m$ for $j = 1, \ldots, T$. Then,

$$\begin{aligned} H(Y^{(i)}) &\geq H(Y^{(i)}_{i_1}, \ldots, Y^{(i)}_{i_T}) \\ &= \sum_{j=1}^{T} H(Y^{(i)}_{i_j}) \\ &\geq \sum_{j=1}^{T} [m - (h(\varepsilon_2) + \varepsilon_2 \cdot m)] \\ &\geq \frac{(1-\varepsilon_2)mT}{2}, \end{aligned} \tag{42}$$

provided that $m \geq c'$ for some large enough constant depending only on c_2. The first equality holds because $Y^{(i)}_{i_1}, \ldots, Y^{(i)}_{i_T}$ are independent, and the second inequality follows from Lemma 6 applied to $Y^{(i)}_{i_j}$ and U_m. Combining (41) with (42), we conclude that

$$H(Y) \geq \frac{(1-\varepsilon_2)mT}{2}.$$

Therefore, recalling (40) and that $\varepsilon_2 \leq c_2$, it follows that

$$\begin{aligned} k &\geq \frac{(1-c_2)mT}{2} - o(m) \\ &\geq \frac{(1-c_2)mT}{3} \end{aligned} \tag{43}$$

for n large enough. Noting that c_2 is a constant and rearranging (43) yields the desired result. □

Some comments are due about Theorem 54. First, Theorem 54 provides a strong separation between convSA-source extraction from weak sources and convSR-source extraction from SHELA sources, as already evidenced in Section 1.4. Consider a SHELA source with ℓ blocks of length n, $\ell = \text{poly}(n)$, $t = 2$ of which are honest with arbitrary linear min-entropy. Then, Theorem 36 shows we can efficiently extract (to within error $2^{-\Omega(\text{poly}(n))}$) a convSR-source with $\text{poly}(n)$ number of blocks each of length $\Omega(n)$ and at least one good block from the SHELA source. Such SHELA source can be compared with an arbitrary weak $(\tilde{n} = n \cdot \ell, k = O(n))$-source. In this case, Theorem 54 shows that if we want to obtain a T-out-of-L convSA-source with block length $\Omega(n)$ from the weak source, then T must be constant. This precludes many applications of the resulting convSA-source as discussed in Section 1. Finally, note that Theorem 54 also applies to the extraction of convSR-sources with several uniform blocks from weak sources.

7 Non-Interactive Protocols from Public SHELA Sources

7.1 CRS Generation through a SHELA Sample

The definitions of proof systems and commitment schemes in the plain model and in the CRS model are standard and can be found in Section 2.7.

Such definitions assume the existence of an efficient CRS generation procedure $\mathcal{G}$ that, however, will instead be realized in our protocols through a sample from a public SHELA source. Our

NON-INTERACTIVE WI PROOF SYSTEM $\Pi_{\mathsf{pv}} = (\mathcal{G}, \mathcal{P}_{\mathsf{pv}}, \mathcal{V}_{\mathsf{pv}})$

CRS GENERATION: $\mathcal{G}$ on input 1^m outputs $\sigma \leftarrow \mathsf{SHELA}_{n,k,t,\ell}$.

PROVER PROCEDURE: $\mathcal{P}_{\mathsf{pv}}$. Input: instance x, witness w s.t. $(x, w) \in \mathcal{R}$ and $\sigma \in \mathsf{SHELA}_{n,k,t,\ell}$.

1. Run $\mathsf{SomeExt}(\sigma)$ obtaining $R_1, \ldots R_L$.
2. For $i = 1, \ldots, L$: Run $\pi_i \leftarrow \mathcal{P}(1^m, x, w, R_i)$.
3. Set $\pi = (\pi_1, \ldots, \pi_L)$, output π.

VERIFIER PROCEDURE: $\mathcal{V}_{\mathsf{pv}}$. Input: instance x and $\sigma \in \mathsf{SHELA}_{n,k,t,\ell}$.

1. Run $\mathsf{SomeExt}(\sigma)$ obtaining $R_1, \ldots, R_L$.
2. If $\mathcal{V}(x, w, R_i, \pi_i) = 1 \ \forall i = 1, \ldots, L$ accept, otherwise reject.

Figure 4: Non-Interactive WI Proof System $\Pi_{\mathsf{pv}} = (\mathcal{G}, \mathcal{P}_{\mathsf{pv}}, \mathcal{V}_{\mathsf{pv}})$.

constructions will convert 2-round public-coin protocols into non-interactive protocols by using a SHELA source and the somewhere-extrator to replace the first round. Therefore, following the notation in the CRS model, when running $\mathcal{G}$ on input 1^m to generate a sufficiently long CRS, we assume that the CRS is generated through a sample $\sigma \leftarrow \mathsf{SHELA}_{n,k,t,\ell}$ from a SHELA source such that when running $\mathsf{SomeExt}(\sigma)$ and obtaining blocks $R_1, \ldots, R_L$ we have that the size of each R_i is equal to the size of the first round of the 2-round public-coin protocol. We recall that $\mathcal{G}$ is not supposed to be efficient and neither simulatable. Moreover, this procedure allows an unbounded adversary to partially control the sampling process. We obviously require that the output of $\mathcal{G}$ be available to all players. In our protocols, some adversaries are restricted to run in polynomial-time only, but still can affect the outcome of the SHELA sample without such restriction.

7.2 Non-Interactive WI Proof System Π_{pv}

Here we present our construction of NIWI proof system from SHELA sources assuming public-coin ZAPs. In order to describe our proof system $\Pi_{\mathsf{pv}} = (\mathcal{G}, \mathcal{P}_{\mathsf{pv}}, \mathcal{V}_{\mathsf{pv}})$ for the NP-language $\mathcal{L}$, we will make use of the following tools: 1) A somewhere extractor $\mathsf{SomeExt} : \{0,1\}^{n \cdot \ell} \to \{0,1\}^{m \cdot L}$ defined in Section 3.2[13]. 2) A 2-round public-coin WI proof system $\Pi = (\mathcal{P}, \mathcal{V})$. Our Non-Interactive WI proof system $\Pi_{\mathsf{pv}} = (\mathcal{G}, \mathcal{P}_{\mathsf{pv}}, \mathcal{V}_{\mathsf{pv}})$ with a CRS generated through a sample from a SHELA source is described in Figure 4. We stress that our protocol can be instantiated using doubly enhanced trapdoor permutations.

Theorem 55. *Assuming the existence of public SHELA sources, if public-coin ZAPs exist, then Π_{pv} is a non-interactive proof system for all NP-languages.*

Proof. **Completeness.** Completeness follows by inspection. We observe also that it is possible to instantiate Π from (doubly) enhanced trapdoor permutations using the construction of [7].
Statistical soundness. Let us fix $x \notin \mathcal{L}$ and $\sigma \in \mathsf{SHELA}_{n,k,t,\ell}$. The statistical soundness of Π implies that for a uniformly chosen string over $\{0,1\}^m$ (which corresponds to the first round of Π) it is infeasible for a malicious prover $\mathcal{P}^\star$ to compute an accepting proof of Π for the instance x. The next observation is that from Theorems 33 and 36 it follows that the procedure $\mathsf{SomeExt}(\sigma)$ on input $\sigma \in \mathsf{SHELA}_{n,k,t,\ell}$ outputs $t-1$ good strings, namely $R_{I_1}, \ldots, R_{I_{t-1}}$, that are ϵ'-close to uniform distribution over $\{0,1\}^m$, where ϵ' is a negligible function. We conclude that a malicious prover $\mathcal{P}^\star_{\mathsf{pv}}$ is able to compute an accepting proof for x w.r.t. R_{I_j} for all $j \in [t-1]$ only with negligible probability.

[13] With high min-entropy we set $L = \ell - 1$, while with low min-entropy we set $L = O(\ell)$.

NON-INTERACTIVE COMMITMENT SCHEME $\Pi_{\mathsf{pvcom}} = (\mathcal{G}, \mathcal{S}_{\mathsf{pvcom}}, \mathcal{R}_{\mathsf{pvcom}})$

CRS GENERATION: $\mathcal{G}$ on input 1^m outputs $\sigma \leftarrow \mathsf{SHELA}_{n,k,t,\ell}$.

SENDER PROCEDURE: $\mathcal{S}_{\mathsf{pvcom}}$. Input: message msg and $\sigma \in \mathsf{SHELA}_{n,k,t,\ell}$.

1. Run $\mathsf{SomeExt}(\sigma)$ obtaining $R_1, \dots R_L$.
2. For $i = 1, \dots, L$: Run $\mathsf{com}_i, \mathsf{dec}_i \leftarrow \mathcal{S}(1^m, \mathsf{msg}, R_i)$.
3. Set $\mathsf{com} = (\mathsf{com}_1, \dots, \mathsf{com}_L)$, $\mathsf{dec} = (\mathsf{dec}_1, \dots, \mathsf{dec}_L)$ and output com.

RECEIVER PROCEDURE: $\mathcal{R}_{\mathsf{pvcom}}$. Input: commitment com, decommitment $\mathsf{dec}, \mathsf{msg}$ and $\sigma \in \mathsf{SHELA}_{n,k,t,\ell}$.

1. Run $\mathsf{SomeExt}(\sigma)$ obtaining $R_1, \dots, R_L$.
2. If $\mathcal{R}(\mathsf{msg}, \mathsf{com}_i, R_i, \mathsf{dec}_i) = 1 \ \forall i = 1, \dots, L$ outputs msg, otherwise reject.

Figure 5: Non-Interactive Commitment Scheme from OWFs $\Pi_{\mathsf{pvcom}} = (\mathcal{G}, \mathcal{S}_{\mathsf{pvcom}}, \mathcal{R}_{\mathsf{pvcom}})$.

Witness indistinguishability. Suppose by contradiction that there exists an adversary A against the WI property of Π_{pv}. Then, we devise an adversary A_{WI} against the WI property of Π as follows.

Let us fix $\sigma \in \mathsf{SHELA}_{n,k,t,\ell}$, and let $\mathcal{CH}$ be the challenger of the WI game. First, we recall that the WI property of Π holds also in the parallel setting. The new adversary A_{WI} will act as a prover of Π_{pv} and as a verifier of Π with $\mathcal{CH}$. The original adversary A on input σ will output an instance x and witnesses w_0, w_1 such that $(x, w_b) \in \mathcal{R}$ with $b \in \{0, 1\}$. First, A_{WI} sends x, w_0, w_1 to $\mathcal{CH}$. Second, acting as a prover of Π_{pv}, A_{WI} obtains σ and runs $\mathsf{SomeExt}(\sigma)$, obtaining $R_1, \dots, R_L$. Third, A_{WI} starts L parallel executions of Π with $\mathcal{CH}$ sending R_i as the first round of the i-th execution, for $i = 1, \dots, L$. Fourth, A_{WI} sends the proof $\pi = (\pi_1, \dots, \pi_L)$ for x to A. Here, π_i is sent by the challenger $\mathcal{CH}$ for the instance x w.r.t. the first round R_i, for $i = 1, \dots, L$. Finally A_{WI} outputs whatever A outputs. The proof is concluded observing that if $\mathcal{CH}$ uses the witness w_b to compute $\pi_1, \dots, \pi_L$ then the reduction is distributed as an honest prover of Π, that is using the witness w_b to compute $\pi_1, \dots, \pi_L$. □

7.3 Non-Interactive Commitment Scheme Π_{pvcom}

Here we present our construction of non-interactive statistically binding commitment scheme from SHELA sources assuming 2-round public-coin statistically binding commitments. In order to describe our commitment scheme $\Pi_{\mathsf{pvcom}} = (\mathcal{G}, \mathcal{P}_{\mathsf{pvcom}}, \mathcal{V}_{\mathsf{pvcom}})$ for the message space M, we will make use of the following tools: 1) a somewhere extractor $\mathsf{SomeExt} : \{0,1\}^{n \cdot \ell} \to \{0,1\}^{m \cdot L}$ defined in Section 3.2[14]; 2) a 2-round public-coin statistically binding commitment scheme $\Pi_{\mathsf{com}} = (\mathcal{S}, \mathcal{R})$. Our Non-Interactive Commitment Scheme $\Pi_{\mathsf{pvcom}} = (\mathcal{G}, \mathcal{P}_{\mathsf{pvcom}}, \mathcal{V}_{\mathsf{pvcom}})$ using a public SHELA source is described in in Figure 5. We stress that our protocol can be instantiated through a black-box use of any one-way function.

Theorem 56. *Assuming the existence of public SHELA sources, if 2-round public-coin statistically binding commitment schemes exist then Π_{pvcom} is a non-interactive commitment scheme.*

Proof. **Completeness.** Completeness follows by inspection. We observe also that it is possible to instantiate Π_{pv} from one-way functions using the construction of [13].

Statistical Binding. Let us fix $\sigma \in \mathsf{SHELA}_{n,k,t,\ell}$. The statistical binding of Π_{pv} implies that for a uniformly chosen string over $\{0,1\}^m$ (which corresponds to the first round of $\mathcal{R}_{\mathsf{pv}}$ in the commitment phase) it is infeasible for a malicious sender $\mathcal{S}^{\star}_{\mathsf{pvcom}}$ to compute a commitment com and two decommitments $(\mathsf{msg}_0, \mathsf{dec}_0)$ and $(\mathsf{msg}_1, \mathsf{dec}_1)$, with $\mathsf{msg}_0 \neq \mathsf{msg}_1$, such that $\mathcal{R}$ accepts

[14] We set L precisely as specified in the previous footnote.

both decommitments w.r.t. com. The next observation is that from Theorems 33 and 36 it follows that $\mathsf{SomeExt}(\sigma)$ with $\sigma \in \mathsf{SHELA}_{n,k,t,\ell}$ outputs $t-1$ good strings, denote them by $R_{I_1}, \ldots, R_{I_{t-1}}$, that are independent and ϵ'-close to the uniform distribution over $\{0,1\}^m$, where ϵ' is a negligible function. We conclude that a malicious sender $\mathcal{S}^{\star}_{\mathsf{pvcom}}$ is able to compute a commitment com and two accepting decommitments w.r.t. R_{I_j} for all $j \in [t-1]$ only with negligible probability.

Computational Hiding. Suppose by contradiction that there exists an adversary A against the computational hiding property of Π_{pvcom}. Then, we devise an adversary A_H against the computational hiding property of Π_{pv} as follows.

Fix $\sigma \in \mathsf{SHELA}_{n,k,t,\ell}$, and let $\mathcal{CH}$ be the challenger of the hiding game. First, we recall that the hiding property of Π_{com} holds also in the parallel setting. The new adversary A_H will act as a receiver of Π_{pvcom} and as a sender of Π_{pv} with $\mathcal{CH}$. The original adversary A on input σ outputs messages $\mathsf{msg}_0, \mathsf{msg}_1$. First, A_H sends $\mathsf{msg}_0, \mathsf{msg}_1$ to $\mathcal{CH}$. Second, A_H, acting as a receiver of Π_{pvcom}, obtains σ and runs $\mathsf{SomeExt}(\sigma)$ to obtain $R_1, \ldots, R_L$. Third, A_H starts L parallel executions of Π_{pv} with $\mathcal{CH}$, using R_i as the first round of the i-th execution, for $i = 1, \ldots, L$. Fourth, A_H sends the commitment $\mathsf{com} = (\mathsf{com}_1, \ldots, \mathsf{com}_L)$ to A, where com_i was received by $\mathcal{CH}$ w.r.t. the first round R_i, for $i = 1, \ldots, L$. The proof is concluded observing that if $\mathcal{CH}$ uses the message m_b to compute $\mathsf{com}_1, \ldots, \mathsf{com}_L$ then the reduction is distributed as an honest sender of Π_{com}, that is using the message m_b to compute $\mathsf{com}_1, \ldots, \mathsf{com}_L$. Finally, A_H outputs whatever A outputs. □

7.4 Improving the Efficiency of [6]

In this section, we will briefly discuss how the somewhere-extractors $\mathsf{SomeExt}$ described in Section 4 can be used to improve the computational efficiency and communication efficiency of the publicly verifiable witness indistinguishable argument of knowledge $\Pi_{\mathsf{SSV}} = (\mathcal{P}_{\mathsf{SSV}}, \mathcal{V}_{\mathsf{SSV}})$ over generic blockchains constructed in [6]. In more details, Π_{SSV} can be built from any blockchain that satisfies the following assumption: In a long sequence ℓ of blocks there will be blocks generated by honest players, and some of these blocks (at least 2) contain a high min-entropy string that is independent from the rest of the content of the blockchain. Moreover, the chunk of each block to consider is well defined. In other words, there exists a deterministic a function s that on input a block of the blockchain B parses B and outputs the chunk of the block that could contain the high min-entropy string. As we discussed in the introduction, if we consider blocks $B_1, \ldots, B_\ell$ of the blockchain, then $s(B_1), \ldots, s(B_\ell)$ constitute a source in $\mathsf{onSHELA}_{n,k,t,\ell}$, where n, k, and t depend on the blockchain and $t \geq 2$.

Π_{SSV} makes use of a 3-round public-coin WI proof system $\Pi_\Sigma = (\mathcal{P}_\Sigma, \mathcal{V}_\Sigma)$ for the relation $\mathcal{R}$. At a very high-level, Π_{SSV} works as follows. $\mathcal{P}_{\mathsf{SSV}}$ computes τ first rounds $\Sigma_1^1, \ldots, \Sigma_\tau^1$ of Π_Σ and publishes them on the blockchain. Then, $\mathcal{P}_{\mathsf{SSV}}$ waits for ℓ new blocks added to the blockchain after the first message was posted. Let us denote such blocks as $B_1, \ldots, B_\ell$. The prover obtains challenges $\Sigma_1^2, \ldots, \Sigma_\tau^2$ by applying an efficient procedure $\mathtt{Extract}$ on input $B_1, \ldots, B_\ell$. Finally, $\mathcal{P}_{\mathsf{SSV}}$ publishes the third rounds $\Sigma_1^3, \ldots, \Sigma_\tau^3$ of Π_Σ on the blockchain. In [6], the efficient procedure $\mathtt{Extract}$ (i.e., the procedure used to compute $\Sigma_1^2, \ldots, \Sigma_\tau^2$) takes as input ℓ blocks $B_1, \ldots, B_\ell$, the deterministic function s and outputs τ strings $s_1, \ldots, s_\tau$ such that at least one string s_i is distributed statistically close to the uniform distribution over $\{0,1\}^m$. $\mathtt{Extract}$ is implemented in [6] as the naive somewhere-extractor $\mathsf{NaiveSomeExt}$ described in Section 1. In other words, $\mathtt{Extract}$, on inputs ℓ blocks $B_1, \ldots, B_\ell$ and the deterministic function s, considers all the possible $\binom{\ell}{c}$ combinations of c-tuples of the ℓ blocks, and runs a c-source extractor on each c-tuple, for some $c \geq 2$.

It is easy to see that the number of times that $\mathcal{P}_{\mathsf{SSV}}$ executes Π_Σ depends on τ, which is the number of strings that $\mathtt{Extract}$ outputs. In the case of [6], we have $\tau = O(\ell^c)$ for some $c \geq 2$.

The somewhere-extractors $\mathsf{SomeExt} : \{0,1\}^{n \cdot \ell} \to \{0,1\}^{m \cdot L}$ defined in Section 4 could help

to implement `Extract` with a much smaller value of τ. Indeed `Extract` can be implemented as follows. `Extract`, on inputs ℓ blocks $B_1, \ldots, B_\ell$ and the deterministic function s, computes $\mathsf{SomeExt}(s(B_1), \ldots, s(B_\ell))$, obtaining blocks $R_1, \ldots, R_L$ as output. By Theorems 42 and 43, we are guaranteed that at least one output block is statistically close to uniform over $\{0,1\}^m$, thus fulfilling the requirement specified for `Extract`. Notably, in this case we have $\tau = L$, where $L = O(\ell)$ in the worst case. Therefore, the proof size of Π is only $O(\ell)$ times (instead of $O(\ell^c)$-times) larger then the proof size of Π. Moreover, $\mathcal{P}_{\mathsf{SSV}}$ executes the prover of Π only $O(\ell)$ times (instead of $O(\ell^c)$ times).

References

[1] B. Chor and O. Goldreich, "Unbiased bits from sources of weak randomness and probabilistic communication complexity," in *FOCS 1985*, pp. 429–442.

[2] R. Canetti, R. Pass, and A. Shelat, "Cryptography from sunspots: How to use an imperfect reference string," in *FOCS 2007*, pp. 249–259.

[3] J. Clark and U. Hengartner, "On the use of financial data as a random beacon," in *EVT/WOTE 2010*, pp. 1–8.

[4] J. A. Garay, A. Kiayias, and N. Leonardos, "The bitcoin backbone protocol: Analysis and applications," in *EUROCRYPT 2015*, pp. 281–310.

[5] R. Pass, L. Seeman, and A. Shelat, "Analysis of the blockchain protocol in asynchronous networks," in *EUROCRYPT 2017*, pp. 643–673.

[6] A. Scafuro, L. Siniscalchi, and I. Visconti, "Publicly verifiable proofs from blockchains," in *PKC 2019*, pp. 374–401.

[7] C. Dwork and M. Naor, "Zaps and their applications," in *FOCS 2000*, pp. 283–293.

[8] J. Bourgain, "More on the sum-product phenomenon in prime fields and its applications," *International Journal of Number Theory*, vol. 01, no. 01, pp. 1–32, 2005.

[9] M. Lewko, "An explicit two-source extractor with min-entropy rate near 4/9," *Mathematika*, vol. 65, no. 4, p. 950–957, 2019.

[10] E. Chattopadhyay and D. Zuckerman, "Explicit two-source extractors and resilient functions," *Annals of Mathematics*, vol. 189, no. 3, pp. 653–705, 2019.

[11] J. Radhakrishnan and A. Ta-Shma, "Bounds for dispersers, extractors, and depth-two superconcentrators," *SIAM Journal on Discrete Mathematics*, vol. 13, no. 1, pp. 2–24, 2000.

[12] U. Feige, D. Lapidot, and A. Shamir, "Multiple noninteractive zero knowledge proofs under general assumptions," *SIAM J. Comput.*, vol. 29, no. 1, pp. 1–28, 1999.

[13] M. Naor, "Bit commitment using pseudorandomness," *J. Cryptology*, vol. 4, no. 2, pp. 151–158, 1991.

[14] A. Ta-Shma, "On extracting randomness from weak random sources (extended abstract)," in *STOC 1996*, pp. 276–285.

[15] C.-J. Lu, O. Reingold, S. Vadhan, and A. Wigderson, "Extractors: Optimal up to constant factors," in *STOC 2003*, pp. 602–611.

[16] B. Barak, G. Kindler, R. Shaltiel, B. Sudakov, and A. Wigderson, "Simulating independence: New constructions of condensers, Ramsey graphs, dispersers, and extractors," *Journal of ACM, April 2010, 20:1–20:52.*

[17] R. Raz, "Extractors with weak random seeds," in *STOC 2005*, pp. 11–20.

[18] B. Barak, A. Rao, R. Shaltiel, and A. Wigderson, "2-source dispersers for $n^{o(1)}$ entropy, and Ramsey graphs beating the Frankl-Wilson construction," *Annals of Mathematics*, vol. 176, no. 3, pp. 1483–1543, 2012.

[19] A. Rao, "Extractors for a constant number of polynomially small min-entropy independent sources," *SIAM Journal on Computing*, vol. 39, no. 1, pp. 168–194, 2009.

[20] X. Li, "Improved constructions of three source extractors," in *CCC 2011*, pp. 126–136.

[21] ——, "New independent source extractors with exponential improvement," in *STOC 2013*, pp. 783–792.

[22] ——, "Extractors for a constant number of independent sources with polylogarithmic min-entropy," in *FOCS 2013*, pp. 100–109.

[23] ——, "Three-source extractors for polylogarithmic min-entropy," in *FOCS 2015*, pp. 863–882.

[24] ——, "Improved two-source extractors, and affine extractors for polylogarithmic entropy," in *FOCS 2016*, pp. 168–177.

[25] A. Ben-Aroya, E. Chattopadhyay, D. Doron, X. Li, and A. Ta-Shma, "A new approach for constructing low-error, two-source extractors," in *CCC 2018, pp. 3:1–3:19.*

[26] G. Cohen, "Local correlation breakers and applications to three-source extractors and mergers," in *FOCS 2015*, pp. 845–862.

[27] G. Cohen and L. J. Schulman, "Extractors for near logarithmic min-entropy," in *FOCS 2016*, pp. 178–187.

[28] E. Chattopadhyay, V. Goyal, and X. Li, "Non-malleable extractors and codes, with their many tampered extensions," in *STOC 2016*, pp. 285–298.

[29] D. Zuckerman, "Linear degree extractors and the inapproximability of max clique and chromatic number," in *STOC 2006*, pp. 681–690.

[30] Z. Dvir and A. Shpilka, "An improved analysis of linear mergers," *computational complexity*, vol. 16, no. 1, pp. 34–59, May 2007.

[31] Z. Dvir and R. Raz, "Analyzing linear mergers," *Random Structures & Algorithms*, vol. 32, no. 3, pp. 334–345, 2008.

[32] Z. Dvir and A. Wigderson, "Kakeya sets, new mergers and old extractors," in *FOCS 2008.*

[33] Z. Dvir, S. Kopparty, S. Saraf, and M. Sudan, "Extensions to the method of multiplicities, with applications to Kakeya sets and mergers," *SIAM J. on Computing*, vol. 42, no. 6, pp. 2305–2328, 2013.

[34] J. von Neumann, "Various techniques used in connection with random digits," in *Monte Carlo Method*, ser. National Bureau of Standards Applied Mathematics Series, 1951, vol. 12, ch. 13, pp. 36–38.

[35] P. Elias, "The efficient construction of an unbiased random sequence," *Ann. Math. Statist.*, vol. 43, no. 3, pp. 865–870, 06 1972.

[36] M. Blum, "Independent unbiased coin flips from a correlated biased source—a finite state Markov chain," *Combinatorica*, vol. 6, no. 2, pp. 97–108, Jun 1986.

[37] B. Chor, O. Goldreich, J. Håstad, J. Freidmann, S. Rudich, and R. Smolensky, "The bit extraction problem or t-resilient functions," in *FOCS 1985*, pp. 396–407.

[38] U. V. Vazirani, "Towards a strong communication complexity theory or generating quasi-random sequences from two communicating slightly-random sources," in *STOC 1985*, pp. 366–378.

[39] C. H. Bennett, G. Brassard, and J.-M. Robert, "How to reduce your enemy's information (extended abstract)," in *CRYPTO 1985, pp. 468–476.*

[40] D. Lichtenstein, N. Linial, and M. Saks, "Some extremal problems arising from discrete control processes," *Combinatorica*, vol. 9, no. 3, pp. 269–287, Sep 1989.

[41] Y. Dodis, "New imperfect random source with applications to coin-flipping," in *ICALP 2001*, pp. 297–309.

[42] A. Gabizon, R. Raz, and R. Shaltiel, "Deterministic extractors for bit-fixing sources by obtaining an independent seed," *SIAM J. on Computing*, vol. 36, no. 4, pp. 1072–1094, 2006.

[43] J. Kamp and D. Zuckerman, "Deterministic extractors for bit-fixing sources and exposure–resilient cryptography," *SIAM J. on Computing*, vol. 36, no. 5, pp. 1231–1247, 2007.

[44] A. Rao, "Extractors for low-weight affine sources," in *CCC 2009*, pp. 95–101.

[45] G. Cohen and I. Shinkar, "Zero-fixing extractors for sub-logarithmic entropy," in *ICALP 2015*, pp. 343–354.

[46] P. Pudlak and V. Rodl, "Extractors for small zero-fixing sources," *arXiv e-prints*, p. arXiv:1904.07949, April 2019.

[47] L. Trevisan and S. Vadhan, "Extracting randomness from samplable distributions," in *FOCS 2000*, pp. 32–42.

[48] A. De and T. Watson, "Extractors and lower bounds for locally samplable sources," in *APPROX/RANDOM 2011*, pp. 483–494.

[49] J. Kamp, A. Rao, S. Vadhan, and D. Zuckerman, "Deterministic extractors for small-space sources," *Journal of Computer and System Sciences*, vol. 77, no. 1, pp. 191 – 220, 2011.

[50] A. Gabizon and R. Raz, "Deterministic extractors for affine sources over large fields," in *FOCS 2005*, pp. 407–416.

[51] J. Bourgain, "On the construction of affine extractors," *GAFA Geometric And Functional Analysis*, vol. 17, no. 1, pp. 33–57, Apr 2007.

[52] M. DeVos and A. Gabizon, "Simple affine extractors using dimension expansion," in *CCC 2010*, pp. 50–57.

[53] A. Yehudayoff, "Affine extractors over prime fields," *Combinatorica*, vol. 31, no. 2, p. 245, Aug 2011.

[54] X. Li, "A new approach to affine extractors and dispersers," in *CCC 2011*, pp. 137–147.

[55] J. Bourgain, Z. Dvir, and E. Leeman, "Affine extractors over large fields with exponential error," *computational complexity*, vol. 25, no. 4, pp. 921–931, Dec 2016.

[56] Z. Dvir, A. Gabizon, and A. Wigderson, "Extractors and rank extractors for polynomial sources," *Computational Complexity*, vol. 18, no. 1, pp. 1–58, Apr 2009.

[57] F. Li and D. Zuckerman, "Improved extractors for recognizable and algebraic sources," *Electronic Colloquium on Computational Complexity (ECCC)*, vol. 25, p. 110, 2018.

[58] E. Viola, "Extractors for Turing-machine sources," in *APPROX/RANDOM 2012*, pp. 663–671.

[59] ——, "Extractors for circuit sources," *SIAM Journal on Computing*, vol. 43, no. 2, pp. 655–672, 2014.

[60] S. Beigi, O. Etesami, and A. Gohari, "Deterministic randomness extraction from generalized and distributed Santha-Vazirani sources," in *ICALP 2015*.

[61] S. Beigi, A. Bogdanov, O. Etesami, and S. Guo, "Optimal deterministic extractors for generalized Santha-Vazirani sources," in *APPROX/RANDOM 2018*.

[62] I. Bentov, A. Gabizon, and D. Zuckerman, "Bitcoin beacon," *arXiv e-prints*, p. arXiv:1605.04559, May 2016.

[63] E. Chattopadhyay, J. Goodman, V. Goyal, and X. Li, "Extractors for adversarial sources via extremal hypergraphs," Cryptology ePrint Archive, Report 2019/1450, 2019, https://eprint.iacr.org/2019/1450.

[64] Y. Dodis, V. Vaikuntanathan, and D. Wichs, "Extracting randomness from extractor-dependent sources," Cryptology ePrint Archive, Report 2019/1339, 2019, https://eprint.iacr.org/2019/1339.

[65] M. Santha and U. V. Vazirani, "Generating quasi-random sequences from slightly-random sources," in *FOCS 1984*, pp. 434–440.

[66] V. Guruswami, C. Umans, and S. Vadhan, "Unbalanced expanders and randomness extractors from Parvaresh–Vardy codes," *J. ACM*, vol. 56, no. 4, pp. 20:1–20:34, Jul. 2009.

[67] N. Nisan and D. Zuckerman, "Randomness is linear in space," *Journal of Computer and System Sciences*, vol. 52, no. 1, pp. 43 – 52, 1996.

[68] J. Groth, R. Ostrovsky, and A. Sahai, "Non-interactive zaps and new techniques for NIZK," in *CRYPTO 2006*, pp. 97–111.

[69] G. Fuchsbauer and M. Orrù, "Non-interactive zaps of knowledge," in *ACNS 2018*, pp. 44–62.

[70] N. Bitansky and O. Paneth, "Point obfuscation and 3-round zero-knowledge," in *TCC 2012*, pp. 190–208.

[71] M. Bellare, G. Fuchsbauer, and A. Scafuro, "Nizks with an untrusted CRS: security in the face of parameter subversion," in *ASIACRYPT 2016*, pp. 777–804.

[72] Y. Dodis, R. Ostrovsky, L. Reyzin, and A. Smith, "Fuzzy extractors: How to generate strong keys from biometrics and other noisy data," *SIAM J. Comput.*, pp. 97–139.

[73] S. Ho and R. W. Yeung, "The interplay between entropy and variational distance," *IEEE Transactions on Information Theory*, vol. 56, no. 12, pp. 5906–5929, Dec 2010.

A Somewhere-Z Sources and Convex Combinations

In this section, we prove a more general version of Lemma 35. Before we state it, we need the following definition.

Definition 57. *Given some fixed source Z, a source $X = (X_1, \ldots, X_L)$ is said to be* somewhere-Z *if there is $\mathcal{I} \subseteq [L]$ such that $X_{\mathcal{I}} \sim Z$. The set of all convex combinations of somewhere-Z sources is denoted by $\mathsf{conv}Z$.*

Lemma 58 (Lemma 35, generalized)**.** *Let $X \in \{0,1\}^{m \cdot L}$ and I denote any random variable over subsets of $[L]$. Suppose that*

$$X_I, I \approx_\varepsilon Z, I, \tag{44}$$

where Z is independent of I. Then, it holds that $X \approx_\varepsilon Y$ for some $Y \in \mathsf{conv}Z$.

Proof. Fix X and I as in the lemma statement. For each fixing $I = \mathcal{I}$ in the support of I, let $X^{\mathcal{I}}$ denote $(X|I = \mathcal{I})$ and $\varepsilon_{\mathcal{I}} = \Delta(X^{\mathcal{I}}_{\mathcal{I}}; Z)$. By (44), we know that

$$\sum_{\mathcal{I}} \Pr[I = \mathcal{I}] \cdot \varepsilon_{\mathcal{I}} < \varepsilon. \tag{45}$$

Furthermore, Lemma 2 guarantees that for each $\mathcal{I}$ there is $Q^{\mathcal{I}}$ such that $Q^{\mathcal{I}} \sim Z$ and $\Pr[X^{\mathcal{I}}_{\mathcal{I}} \neq Q^{\mathcal{I}}] \leq \varepsilon_{\mathcal{I}}$.

Consider now $Y^{\mathcal{I}}$ coupled with $X^{\mathcal{I}}$, defined as $Y^{\mathcal{I}}_j = X^{\mathcal{I}}_j$ for all $j \notin \mathcal{I}$ and $Y^{\mathcal{I}}_{\mathcal{I}} = Q^{\mathcal{I}}$. Observe that $Y^{\mathcal{I}}$ is somewhere-Z since $Y^{\mathcal{I}}_{\mathcal{I}} = Q^{\mathcal{I}} \sim Z$, and

$$\Pr[X^{\mathcal{I}} \neq Y^{\mathcal{I}}] = \Pr[X^{\mathcal{I}}_{\mathcal{I}} \neq Q^{\mathcal{I}}] \leq \varepsilon_{\mathcal{I}}. \tag{46}$$

With this in mind, we define Y by setting $(Y|I = \mathcal{I}) = Y^{\mathcal{I}}$ (here I still denotes the indicator of X). It follows that $Y = \sum_{\mathcal{I}} \Pr[I = \mathcal{I}] \cdot Y^{\mathcal{I}} \in \mathsf{conv}Z$, and, by Lemma 2,

$$\begin{aligned}
\Delta(X;Y) &\leq \Pr[X \neq Y] \\
&= \sum_{\mathcal{I}} \Pr[I = \mathcal{I}] \cdot \Pr[X^{\mathcal{I}} \neq Y^{\mathcal{I}}] \\
&\leq \sum_{\mathcal{I}} \Pr[I = \mathcal{I}] \cdot \varepsilon_{\mathcal{I}} \\
&< \varepsilon.
\end{aligned}$$

The second inequality follows from (46), and the third inequality from (45). This yields the desired result.

□