Electronic Colloquium on Computational Complexity, Report No. 12 (2025)

Bit-Fixing Extractors for Almost-Logarithmic Entropy

Dean Doron*
Department of Computer Science
Ben Gurion University
deand@bgu.ac.il

Ori Fridman*
Department of Computer Science
Ben Gurion University
orifrid@post.bgu.ac.il

Abstract

An oblivious bit-fixing source is a distribution over $\{0,1\}^n$, where k bits are uniform and independent and the rest $n-k$ are fixed a priori to some constant value. Extracting (close to) true randomness from an oblivious bit-fixing source has been studied since the 1980s, with applications in cryptography and complexity theory.

We construct explicit extractors for oblivious bit-fixing source that support $k = \widetilde{O}(\log n)$, outputting almost all the entropy with low error. The previous state-of-the-art construction that outputs many bits is due to Rao [Rao, CCC '09], and require entropy $k \geq \log^c n$ for some large constant c. The two key components in our constructions are new low-error affine condensers for poly-logarithmic entropies (that we achieve using techniques from the nonmalleable extractors literature), and a dual use of linear condensers for OBF sources.

*Supported in part by NSF-BSF grant #2022644.

ISSN 1433-8092

1 Introduction

Whenever randomness is used in computations–whether because it is necessary or just because it is faster and simpler in practice, access to unbiased bits is assumed. Because sources of perfect randomness are notoriously hard to come by, ad hoc, very weak sources of randomness are used instead. It is essential, therefore, that the crude randomness generated by such sources be purified, thus driving the development of the beautiful theory of randomness extractors. This problem of extracting randomness from imperfect sources can be traced back to von Neumann [vN51]. A *randomness extractor* is a deterministic procedure that converts a weak random source into a random source that is close to uniform.

Definition 1.1 (seedless extractor)**.** *Let $\mathcal{X}$ be a family of distributions over $\{0,1\}^n$. We say that* $\mathsf{Ext}\colon \{0,1\}^n \to \{0,1\}^m$ *is an* extractor for $\mathcal{X}$ *with error ε, if for every $X \in \mathcal{X}$, $\mathsf{Ext}(X)$ is ε-close, in total variation distance, to U_m, the uniform distribution over m bits.*

One common assumption is that each weak source $X \in \mathcal{X}$ has min-entropy.[1] Unfortunately, if all we assume is an entropy guarantee, it is easy to show that such an Ext does not exist. However, seedless extraction *is* possible for some restricted classes of sources,[2] and indeed, studying the capabilities and limitations of extraction from *natural* families of structured sources has been a fruitful endeavor for the past four decades (see [Goo23, Section 1.3] for a comprehensive up-to-date survey of seedless extraction results).

One of the simplest and most natural families of weak sources of randomness is *oblivious bit fixing sources*, wherein k bits are uniform and independent, while the rest $n-k$ bits are fixed.

Definition 1.2 (OBF source)**.** *A distribution $X \sim \{0,1\}^n$ is an (n,k)* oblivious bit-fixing source *if there exists a subset $I \subseteq [n]$ of size k of "good indices", such that the $\{X_i\}_{i\in I}$-s are uniform and independent, and the rest are fixed.*[3]

In addition to simulating true randomness, extractors have been found to be useful, even essential, in myriad applications, and extractors for oblivious bit-fixing sources are no exception. In cryptography, a prominent application of OBF extractors is *exposure resilient cryptography*, where a malicious adversary learns some bits of a secret string (see, e.g., [BBR85, Dod00, CDH+00, DSS01]). Additional cryptographic applications include the related notion of all-or-nothing transforms [Riv97, CDH+00, DSS01], generation of block ciphers [JSY99, Bla96], and reducing the number of secret sharings in the distributed setting [DKM+06]. OBF extractors have also found applications in complexity theory, mainly for lower bounds [KRT13, CKK+15].

Previous constructions. Extractors for OBF sources were first studied by Chor, Goldreich, Håstad, Freidmann, Rudich, and Smolensky [CGH+85] in the zero-error setting (indeed, this is one of the rare examples in which outputting exactly uniform bits is possible). They observed that the simple XOR function outputs a uniform bit for any $k \geq 1$, but also proved that $k \geq n/3$ is necessary even

[1] We say that $X \sim \{0,1\}^n$ has min-entropy k, $H_\infty(X) \geq k$, if $\max_{x\sim X} \Pr[X = x] \leq 2^{-k}$. We call such X an (n,k) source.

[2] To extract from general (n,k) sources, one can use an additional short uniform seed, driving the rich theory of *seeded* extractors. See Definition 3.5 for the definition.

[3] To clarify the terminology, the word "oblivious" refers to the fact that the fixed bits are chosen before the random bits are tossed. On the other hand, in *nonoblivious* bit fixing sources, the non-uniform bits can depend arbitrarily on the values of the good ones (see Section 3.4).

for the $m = 2$ case. More generally, via establishing connections to error correcting codes, Chor et al. showed that when $k = n - t$, we can (explicitly) extract $m \approx n - t\log(n/t)$ bits, but cannot extract more than, roughly, $n - (t/2)\log(n/t)$ bits. Further lower bounds in this regime were obtained by Friedman [Fri92].

What if we allow error, and wish to support lower entropies? Kamp and Zuckerman were the first to go below linear entropy, achieving $k \geq n^{1/2+\gamma}$ and $m = \Omega(n^{2\gamma})$, with error $\varepsilon = 2^{-\Omega(m)}$. The entropy loss in the [KZ06] construction was later improved by Gabizon, Raz, and Shaltiel [GRS06], who achieved $m = (1 - o(1))k$ and exponentially-small error. Gabizon et al. were also able to reduce the entropy down to $k \geq \mathrm{poly}(\log n)$, but with a worse error of $k^{-\Omega(1)}$. As we will soon discuss, a large error is often prohibitive, especially in cryptographic applications. The first *low-error* OBF extractor for poly-logarithmic entropy was constructed by Rao [Rao09], outputting $m = (1 - o(1))k$ bits with error $2^{-k^{\Omega(1)}}$. We will further discuss Rao's construction in Section 2.1.

For even lower entropies, recent explicit constructions that support near-logarithmic entropy already work for the more general family of *affine sources*.[4] Using techniques from the nonmalleable extractors literature, Chattopadhyay, Goodman, and Liao [CGL22] achieved $k = \widetilde{O}(\log n)$, and this was later improved by a recent work of Li [Li23] that gets $k = O(\log n)$. Unfortunately, the latter two constructions work for *constant error* and *one output bit*, which is not relevant for OBF extractors. However, one can think of applying techniques similar to the ones in [Sha08, Li16]. But these, to the best of our knowledge, do not readily give a significant improvement on either m or ε.

Lastly, it is interesting to note that while a straightforward application of the probabilistic method gives a non-explicit construction for (roughly) $k \geq \log n + 2\log(1/\varepsilon)$ and $m = k - 2\log(1/\varepsilon)$, logarithmic entropy is *not* a lower bound. In [KZ06], they gave a construction that outputs only $\Omega(\log k)$ bits, has error $k^{-\Omega(1)}$, but works for *any* k. In [CS15], Cohen and Shinkar gave a lower bound for extraction, and showed that when k is a small enough function of n, $\Omega(\log k)$ output bits are all we can hope for. Moreover, they gave a "semi-explicit" construction, when the error is large enough, for sub-logarithmic entropy. Their construction supports $k = \Omega(\log\log n)$, outputs $m = k - O_\varepsilon(1)$ bits, and runs in time $2^{\widetilde{O}_\varepsilon(\log n)}$.

Our result. We give an explicit low-error construction for $k \geq \widetilde{O}(\log n)$ that outputs almost all the entropy.

Theorem 1.3 (see also Theorem 5.4)**.** *There exist constants $c > 1$ and $\beta \in (0,1)$ such that the following holds for any $n \in \mathbb{N}$ and $k \geq \log n \cdot (\log\log n)^c$. There exists an explicit function*

$$\mathsf{OBFExt}\colon \{0,1\}^n \to \{0,1\}^{m=(1-o(1))k}$$

that is a (k,ε) extractor for OBF sources, where $\varepsilon = 2^{-(k/\log n)^\beta}$.

While the error guarantee does not meet the optimal $2^{-\Omega(k)}$ (or even $2^{-k^{\Omega(1)}}$), both our construction and [Rao09] achieve $\varepsilon = n^{-\omega(1)}$ starting from $k = \mathrm{poly}(\log n)$. Importantly, our construction is the first to extract almost all the entropy with vanishing error starting from $k = \widetilde{O}(\log n)$. In particular, our error guarantee surpasses a $k^{-\Omega(1)}$-type dependence for all ranges of k.

[4] An (n,k) affine source is a distribution that is flat over some affine subspace of $\mathbb{F}_2^n$ of dimension k. Thus, an (n,k) OBF source is simply an affine source whose basis consists of only elementary vectors.

The low-error challenge. The recent decade has seen exciting progress in extractors that support small entropies, most notably in constructions of two-source extractors[5] ([CZ19, BADT19, Coh17, Li19, Li23]), and the use of similar techniques to construct affine extractors [Li16, CGL22, Li23]. While these extractors, which generally follow the celebrated Chattopadhyay–Zuckerman framework [CZ19], work for low entropies, their error is relatively high for reasons which will be discussed in Section 2.2. Low-error constructions are crucial for the security of cryptographic applications, as well as for good correlation bounds in lower bounds applications, but despite numerous attempts, recent constructions do not obtain negligible error in polynomial time.[6]

Our construction uses key components of the [CZ19] framework (mainly towards constructing new low-error affine condensers, see Section 2.2), and is able to outperform the error parameter in most related constructions. We thus view this work also as a proof-of-concept for utilizing more recent machinery to tackle low-error constructions.

2 Proof Overview

Our construction combines two new components: A low-error affine condenser with a small gap, and a dual use of linear condensers for OBF sources. In Section 2.1, we will revisit Rao's construction [Rao09] and his use of linear condensers for OBF sources. In Section 2.2, we will discuss the [CZ19] framework, the low-error adaptation of [BCDT19] to two-source condensers, and our construction of affine condensers. In Section 2.3, we will give the full construction, after discussing our use of linear condensers for OBF sources.

2.1 Rao's Construction

Given an (n, k) OBF source X, Rao's construction [Rao09] goes roughly as follows.[7]

1. Transform X into $X' = P \cdot X$, where P is a linear transformation, X' is much shorter than X, yet it preserves the entropy of X. Thus, this step can be seen as applying a *linear condenser for OBF sources*, and we discuss this step a bit more thoroughly below. Note that X' is an affine source.

2. Transform X' into an *affine somewhere-random source*. This is a source distributed over a table, where one row is uniform, and every other row depends on the uniform one in an affine way. Z has few rows, $k^{\Omega(1)}$, but the length of each row is almost k. This transformation is done via applying a (linear) seeded extractor on X' and every possible seed, i.e., $Z_i = \mathsf{Ext}(X', i)$.[8] This is the part that requires X to have poly-logarithmic entropy.

[5] A two-source extractor for entropy k extracts from the family of sources $\mathcal{X}$ in which each $X \in \mathcal{X}$ comprises two *independent* sources $(X_1, X_2) \sim \{0,1\}^n \times \{0,1\}^n$, each X_i satisfies $H_\infty(X_i) \geq k$.

[6] For two-source extractors, the best low-error constructions require $k \geq (1/2 - \alpha)n$ for a small constant $\alpha > 0$ [Bou05, Lew19], whereas the [CZ19] construction gets $k = \mathrm{poly}(\log n)$ with $\varepsilon = k^{-\Omega(1)}$, and followup constructions for even lower entropies only work for constant error. For affine extractors, the best low-error constructions require $k = \omega(n/\log\log n)$ [Yeh11, Li11, LZ24], whereas, again, recent constructions in the near-logarithmic regime only work for constant error.

[7] The [Rao09] construction works not only for OBF sources, but more generally to low-weight affine sources. Our result captures low-weight affine sources as well, and we discuss this in Section 5.3.

[8] More accurately, to make each row longer than $k^{\Omega(1)}$, Rao injects more entropy to the table by performing an additional extraction step, this time with X itself as the source, and Z_i as the seed.

3. Next, extract from Z. This step is done via repeatedly applying an *affine somewhere-random condenser*, that halves the number of rows in the table, while only shortening each row by a little bit.

Our construction makes use of Step (1) (in fact, more than once, see Section 2.3), and completely dispenses with Step (3). In our construction, once we have our table Z from Step (2), we condense it into a single string using a different mechanism. This is encapsulated under a new affine condenser, which we discuss soon in Section 2.2.

Linear Condensers for OBF Sources. In [Rao09], Rao observed that parity check matrices of binary error correcting codes are good linear condensers for OBF sources, and low-weight affine sources in general.[9] Specifically, let $P \in \mathbb{F}_2^{n \times t}$ be the parity check matrix of a linear code $\mathcal{C} \subseteq \mathbb{F}_2^n$ with co-dimension t and distance $k+1$ (see Section 3.5 for the relevant definitions). Then, given an (n,k) OBF source X, the affine source $P \cdot X \sim \mathbb{F}_2^t$ still has entropy k (see Lemma 5.2 for the easy proof).

This (lossless) condensing allows Rao, and us, to apply affine primitives on sources of much shorter lengths. Indeed, working with BCH codes, one can get $t = O(k \log n)$.[10] In our construction, we will also use a *lossy* instantiation of these linear OBF condensers, where $t \ll k$. We discuss this further in Section 2.3.

2.2 Low-Error Affine Condensers

The [CZ19] Framework. We start by briefly describing the [CZ19] framework, originally geared towards obtaining an extractor for two independent sources. The CZ construction uses two main ingredients: A "correlation breaking" primitive,[11] and a resilient function. For the former, we will consider here a *t-correlation breaker with advice*,[12] which is a function $\mathsf{CB}\colon \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ that takes inputs from a weak source X_1, a uniform seed Y, and a fixed advice string α, and the guarantee is that

$$\mathsf{CB}(X_1, Y, \alpha) \approx U_m \;\Big|\; \left\{\mathsf{CB}\left(X_1, Y^{(1)}, \alpha^{(1)}\right), \ldots, \mathsf{CB}\left(X_1, Y^{(t)}, \alpha^{(t)}\right)\right\},$$

where $Y^{(1)}, \ldots, Y^{(t)}$ are independent of X, and the $\alpha^{(i)}$-s are all different from α. Roughly speaking, for a typical $y \sim Y$, $\mathsf{CB}(X_1, y, \alpha)$ is close to uniform even given the value of CB on t other adversarially chosen $y^{(i)}$-s. Hence, if we were to build a table with $D = 2^d$ rows, and put, say, $\mathsf{CB}(X_1, i, i)$ in the i-th row, then rows of good seeds were not only close to uniform, but they're also close to being t-wise independent.

[9] It is interesting to note that relying on the distance property of error correcting codes alone cannot give optimal linear OBF condensers. One can show that there exist condensers with output length $t = O(k)$.

[10] [Rao09] uses an error correcting code that comes from small-bias sets, and achieve a worse distance-to-codimension tradeoff. In fact, for our construction, BCH codes are necessary.

[11] The idea of constructing extractors via correlation breaking appeared prior to [CZ19], notably in [Li13a, Li13b, Li15, Coh16].

[12] The actual definition offers much more flexibility than what we describe here, but it suffices for this discussion. Also, the original CZ construction uses non-malleable extractors as the correlation breaking primitive, but there are known reductions between the two objects.

Yet, there are no one-source extractors, and [CZ19] need to use a second weak source X_2, to subsample from the seeds of CB. Specifically, take a seeded extractor $\mathsf{Ext}\colon \{0,1\}^n \times \{0,1\}^{d_0} \to \{0,1\}^d$,[13] and consider the table Z with $r = 2^{d_0} \ll D$ rows, in which each row is given by

$$Z_i = \mathsf{CB}(X_1, \mathsf{Ext}(X_2, i), i). \tag{1}$$

An analogous way to view the construction of Z, which would be more beneficial towards our construction, is to first create the table with r rows wherein the i-th row is simply $\mathsf{Ext}(X_2, i)$. While most of the rows in that table are marginally close to uniform, they are arbitrarily correlated. We then use CB with an independent source X_1 to (partially) *break correlations*.[14]

It turns out that Z itself, when the parameters are set correctly, is close to a table in which $r - q$ "good" rows are truly t-wise independent, and there are q "bad" rows that can depend arbitrarily on the good ones. Indeed, many constructions following [CZ19] can also be divided into two steps, where the first one transforms X_1 and X_2 into $Z = Z(X_1, X_2)$ with the above structure, known as a *non-oblivious bit-fixing source*, and the second one, which we now discuss, is where resilient functions comes into play.

Concisely, a resilient function $f\colon \{0,1\}^r \to \{0,1\}$ is an extractor for non-oblivious bit fixing sources. That is, we want a nearly-balanced f whose output cannot be heavily influenced by any small set of q bad bits (we're considering $m = 1$ for now), and in our case f needs to be resilient even when the good bits are only t-wise independent (and uniform), unlike the more standard case in which the good bits are completely independent. The second step of the CZ framework then amounts to applying a resilient function on Z, outputting $W = f(Z)$.

While this beautiful approach does give us that $W \approx_\varepsilon U_1$, it is inherently bound to have runtime which is polynomial in $1/\varepsilon$. The Kahn–Kalai–Linial theorem [KKL88] tells us that no matter what f is, there will always be a single bad bit that has influence $p = \Omega(\log r/r)$, i.e., with probability p over the good bits, the single corrupt bit can fully determine the result.[15] Thus, the running time, which is at least r, is also at least $1/\varepsilon$. This is a common feature of almost all constructions that follow the CZ approach.

The [BCDT19] Low-Error Adaptation. In [BCDT19], Ben-Aroya, Cohen, Doron, and Ta-Shma, evaded the above resilient functions barrier by aiming for a weaker object – a two source *condenser*. Namely, the output W is not ε-close to uniform, but ε-close to some random variable $W' \sim \{0,1\}^m$ that has min-entropy $m - g$, where we call g the *entropy gap* (note that an extractor has gap $g = 0$). While when $m = 1$, a single malicious bit can bias the result pretty significantly, the key observation in [BCDT19] is that when m becomes large, the probability that the bad bits can reduce g bits of entropy from the output can be exponentially-small in g. We call such a function $f_{\mathsf{s}}\colon \Sigma^r \to \Sigma$, $\Sigma = \{0,1\}^m$, an *entropy-resilient function*, which is essentially a condenser for non-oblivious *symbol*-fixing sources (that is, the natural extension of bit-fixing sources to arbitrary alphabets). In [BCDT19], they constructed an explicit such f_{s}, and were able to conclude that $W = f_{\mathsf{s}}(Z)$ is ε-close to having gap $g = o(\log(1/\varepsilon))$ provided that k is at least polynomial in $\log(n/\varepsilon)$, and $m = k^{\Omega(1)}$. Importantly, the construction's runtime is polynomial in n, and not in $1/\varepsilon$.

[13] See Definition 3.5 for the formal definition.

[14] This viewpoint was taken, e.g., in [Coh16, Li16, BADT20].

[15] In terms of explicit f-s, [CZ19] derandomized the Ajtai-Linial [AL93] randomized construction, supporting $q = r^{1-\delta}$ for any constant $\delta > 0$. In a subsequent work, Meka obtained a derandomized version of [AL93] that matches the randomized construction and can support $q = O\left(\frac{r}{\log^2 r}\right)$ bad bits.

Teleporting to the affine setting. Starting from the work of Li [Li16], and throughout a sequence of followup works [CL22, CGL22, CL22, Li23], the correlation breaking framework was used to extract from affine sources and other related families of weak sources such as sumset sources, small-space sources, and interleaved sources. While in the affine case we don't have two sources at our disposal, our single source has *structure* we can utilize.

Specifically, recall our table $\{\mathsf{Ext}(X,i)\}_{i\in[r]}$ above, and assume that Ext is *linear*, meaning that for any fixed seed, the output is a linear function of the source. Here too, our table has many good rows (in fact, the good rows are exactly uniform), but they are, again, arbitrarily correlated. It turns out that we can use the same source X to break the correlation and make the table close to a t-non-oblivious bit-fixing source. This heavily uses the (by now) standard "affine conditioning" technique (see Lemma 3.4), in which given a linear function $L\colon \{0,1\}^n \to \{0,1\}^m$, we can decompose $X = A + B$, where both A and B are affine, and there is a bijection between A and $L(X)$, and B is independent of $L(X)$.[16] Conveniently, the intricate correlation breaking constructions for linearly-correlated source and seed was made explicit in [CL22], coined *affine correlation breakers*. The state-of-the-art affine correlation breakers follow from a recent work of Li [Li23] (see Definition 3.10 and Theorem 3.11).

Our Low-Error Affine Condenser. Armed with a low-error two-source condenser, and a method to adapt the two-source setting to the affine world, a low-error affine condenser follows rather easily. Consider the table Z from Equation (1), but this time we form it as follows:

$$Z_i = \mathsf{AffCB}(X, \mathsf{LExt}(X,i), i), \tag{2}$$

where AffCB is an affine correlation breaker, and LExt is a linear seeded extractor. We then output

$$W = f_{\mathsf{s}}(Z),$$

where f_{s} is the above entropy resilient function. We can then show that if our affine source $X \sim \mathbb{F}_2^n$ has at least $k \geq \mathrm{poly}(\log(n/\varepsilon))$ entropy, $W \sim \{0,1\}^{k^{\Omega(1)}}$ is ε-close to having entropy gap $o(\log(1/\varepsilon))$. The formal construction is given in Section 4.

2.3 Our OBF Extractor

Equipped with a linear condenser for OBF sources P, and a low-error affine condenser for poly-logarithmic entropies AffCB, one natural attempt would be to try and output

$$\mathsf{LExt}(X, \mathsf{AffCB}(P(X))),$$

where LExt is a linear seeded extractor. Note that $P(X)$ is a *short* affine source with entropy k, so the entropy lower bound of AffCB is now much lower! There are three potential problems with the above construction:

1. We want AffCB to supply the seed for LExt, but $Y = \mathsf{AffCB}(P(X))$ is not uniform – it is only ε-close to having some entropy gap g. But since g is tiny, we *can* in fact use Y as a seed, suffering only a small loss in the error. (This observation is by now standard.)

[16] In our case, the function L is chosen according to our linear seeded extractor Ext, and in fact bundles up t of them. For a more detailed description of the correlation breaking mechanism, see, for example, [Li16].

2. X has length n, and we don't have linear seeded extractors outputting many bits with optimal seed length. In particular, the entropy in Y cannot be greater than k, but all known extractors require seed $\gg \log^2 n$.

   To try and resolve the issue, one may change the construction to $\mathsf{LExt}(P(X), \mathsf{AffCB}(P(X)))$, thereby only working with the short $P(X)$. Even then, we are still left with the challenge of handling the *correlations* between the source and the seed.

3. Indeed, Y depends on $P(X)$ (or X) – it's a deterministic function of it! As mentioned above, we will use the affine conditioning technique in order to handle the correlations. We decompose $X = A + B$, where A and B are affine and independent, A is a deterministic function of $P(X)$, and B is independent of $P(X)$.

   Morally, after an appropriate "fixing", the source we use for LExt is the affine source B. But we can only guarantee that $H_\infty(B) \geq k - |P(X)|$, which provides no useful bound if we want to retain most of the entropy of X.

In order to make X shorter *and* guarantee that some entropy is left even after the affine conditioning, we make use of linear OBF condensers in two ways:

$$\mathsf{OBFExt}(X) = \mathsf{LExt}(P_1(X), \mathsf{AffCond}(P_2(X)))$$

Above,

- P_1 is a *lossless* condenser for OBF sources, mapping n bits to $n' = O(k \log n)$ bits.
- P_2 is a *lossy* condenser for OBF sources, mapping n bits to γk bits for a small $\gamma \in (0, 1)$, while retaining $k' \approx \frac{k}{\log n}$ entropy. We construct P_2 using error correcting codes as well, only with different parameters.

  For AffCB to condense, we need $k' \geq \mathrm{poly}(\log(n'/\varepsilon))$, and this sets our lower bound on k and the bound on the error.

Finally, we need to show that we can apply LExt. Towards this end, we "affine condition" with respect to $P_2(X)$, and so we can write

$$\mathsf{OBFExt}(X) = \mathsf{LExt}(P_1(A), Y) + \mathsf{LExt}(P_1(B), Y), \tag{3}$$

where $Y = \mathsf{AffCB}(P_2(X))$, A and B are independent, and $P_2(X)$ is a deterministic function of A. Thus, B and Y are also independent, and furthermore, B still has enough entropy. This, in turn, implies that $P_1(B)$ has enough entropy, and we can safely fix a good seed $y \sim Y$, making the first term in Equation (3) fixed, and the second one close to uniform. The full details appear in Section 5.2.

3 Preliminaries

We use $\log(x)$ for $\log_2(x)$. For an integer n, we denote by $[n]$ the set $\{1, \ldots, n\}$. The density of a subset $B \subseteq A$ is denoted by $\mu(B) = \frac{|B|}{|A|}$. For a function $f \colon \Omega_1 \to \Omega_2$, we say that f is *explicit* if there exists a deterministic procedure that runs in time $\mathrm{poly}(\log(|\Omega_1|))$ and computes f.

Random variables, entropy. The *support* of a random variable X distributed over some domain Ω is the set $x \in \Omega$ for which $\Pr[X = x] \neq 0$, which we denote by $\mathrm{Supp}(X)$. The *total variation distance* (or, statistical distance) between two random variables X and Y over the same domain Ω is defined as $|X - Y| = \max_{A \subseteq \Omega}(\Pr[X \in A] - \Pr[Y \in A])$. Whenever $|X - Y| \leq \varepsilon$ we say that X is ε-close to Y and denote it by $X \approx_\varepsilon Y$. We denote by U_n the random variable distributed uniformly over $\{0,1\}^n$. We say a random variable is *flat* if it is uniform over its support. Whenever we write $x \sim A$ for A being a set, we mean x is sampled uniformly at random from the flat distribution over A.

For a function $f\colon \Omega_1 \to \Omega_2$ and a random variable X distributed over Ω_1, $f(X)$ is the random variable distributed over Ω_2 obtained by choosing x according to X and computing $f(x)$. For a set $A \subseteq \Omega_1$, $f(A) = \{f(x) : x \in A\}$. For every $f\colon \Omega_1 \to \Omega_2$ and two random variables X and Y distributed over Ω_1 it holds that $|f(X) - f(Y)| \leq |X - Y|$, and is often referred to as a data-processing inequality. Another property of statistical distance is the triangle inequality, which states that for all distributions X, Y, Z, we have $|X - Y| \leq |X - Z| + |Y - Z|$.

The min-entropy of X is defined by

$$H_\infty(X) = \min_{x \in \mathrm{Supp}(X)} \log \frac{1}{\Pr[X = x]},$$

and it always holds that $H_\infty(X) \leq H(X)$, where $H()$ is Shannon's entropy. A random variable X is an (n,k) *source* if X is distributed over $\{0,1\}^n$ and has min-entropy at least k.

Limited Independence. We say that a distribution $X \sim \{0,1\}^n$ is (t,γ)-wise independent, if the restriction of X to any t coordinates is γ-close to U_t. When $\gamma = 0$, we simply say that X is t-wise independent.

Lemma 3.1 ([AGM03])**.** *Let $X \sim \{0,1\}^n$ be (t,γ)-wise independent. Then, X is $(n^\gamma t)$-close to a t-wise independent distribution.*

3.1 OBF and Affine Sources

We repeat the definition of affine and OBF sources.

Definition 3.2 (affine source)**.** *An (n,k)* affine source *is a distribution $X \sim \mathbb{F}_2^n$ that is flat over some (unknown) affine subspace of dimension k.*

In other words, for any such X there exist independent $v_0, v_1, \ldots, v_k \in \mathbb{F}_2^n$ such that sampling $x \sim X$ amounts to sampling $\alpha_1, \ldots, \alpha_k \sim \mathbb{F}_2$ uniformly at random, and outputting $x = v_0 + \sum_{i=1}^k \alpha_i v_i$. Notice that when X is affine, $H_\infty(X) = H(X)$.

Definition 3.3 (OBF source)**.** *An (n,k)* oblivious bit-fixing (OBF) source *is a distribution $X \sim \{0,1\}^n$ for which there exists an (unknown) subset $I \subseteq [n]$ of size k, and $c \in \{0,1\}^{n-k}$, such that $X_I = U_k$, and $X_{[n]\setminus I}$ is fixed to c.*

In other words, a bit-fixing source is a very structured affine source – one in which $v_1, \ldots, v_k$ are indicator vectors. The following lemma will let us use linearly-correlated source and seed when using linear seeded extractors.

Lemma 3.4 (affine conditioning, [GR08, Rao09, Li11])**.** *Let X be an (n,k) affine source, and let $L\colon \{0,1\}^n \to \{0,1\}^m$ be a linear function. Then, there exist independent affine sources A and B, over $\{0,1\}^n$, such that:*

- $X = A + B$.
- *There exists* $c \in \{0,1\}^m$ *such that for every* $b \in \mathrm{Supp}(B)$ *it holds that* $L(b) = c$.
- $H(A) = H(L(A))$ *and there exists an affine function* $L_{\mathsf{inv}} \colon \{0,1\}^n \to \{0,1\}^m$ *such that* $A = L_{\mathsf{inv}}(L(A))$. *In particular,* $H(A) \leq m$ *and* $H(B) \geq k - m$.
- *For every* $\ell \in \mathrm{Supp}(L(X))$, $H(X | \{L(X) = \ell\}) = H(B)$.

3.2 Seeded Extractors

Definition 3.5 (seeded extractor)**.** *A function* $\mathsf{Ext} \colon \{0,1\}^n \times \{0,1\}^n \to \{0,1\}^m$ *is a* (k, ε) (seeded) extractor *if for every* (n,k) *source and an independent and uniform* $Y \sim \{0,1\}^d$, *it holds that* $\mathsf{Ext}(X,Y) \approx_\varepsilon U_m$. *Furthermore, we say that* Ext *is* strong *if* $(\mathsf{Ext}(X,Y), Y) \approx_\varepsilon (U_m, Y)$.

We say that Ext *is* linear *if it is linear in the source, namely if for any* $x_1, x_2 \in \{0,1\}^n$ *and* $y \in \{0,1\}^d$, *it holds that* $\mathsf{Ext}(x_1 + x_2, y) = \mathsf{Ext}(x_1, y) + \mathsf{Ext}(x_2, y)$.

Combining a (linear variant) of the GUV seeded condenser [GUV09, CI17] with the Shaltiel–Umans extractor [SU05], we can get the following logarithmic-seed linear seeded extractor, that outputs a constant power of the entropy.

Theorem 3.6 (linear seeded extractor, I [Li16])**.** *There exists a constant* c_1 *such that the following holds, for any positive integers* n *and* $c_1 \log^8 n \leq k \leq n$, *and any* $\varepsilon \geq n^{-2}$. *There exists an explicit linear strong* (k, ε) *extractor* $\mathsf{LExt}_1 \colon \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ *with* $d = c_1 \cdot \log n$ *and* $m = \sqrt{k}$.

Replacing the SU extractor with Trevisan's extractor [Tre01], as analyzed in [RRV02], we can output almost all the entropy, at the cost of a larger seed.

Theorem 3.7 (linear seeded extractor, II)**.** *There exists a constant* c_2 *such that the following holds, for any positive integers* n *and* $k \leq n$, *and any* $\varepsilon, \gamma > 0$. *There exist an explicit linear strong* (k, ε) *extractor* $\mathsf{LExt}_2 \colon \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ *where* $m = (1 - \gamma)k$ *and*

$$d = c_2 \cdot \left(\log n + \log^2 k \cdot \log(1/\varepsilon) \cdot \log(1/\gamma) \right).$$

We omit the easy proof, which is similar to the proof of Theorem 3.6 in [Li16].

When the extractor is linear and the source is affine, a good seed already implies perfect uniformity, for any nontrivial error.

Lemma 3.8 ([Rao09])**.** *Let* $\mathsf{LExt} \colon \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ *be a linear strong* (k, ε) *extractor with* $\varepsilon < 1/2$. *Then, for every* (n,k) *affine source, it holds that*

$$\Pr_{y \sim U_d} [|\mathsf{LExt}(X,y) - U_m| = 0] \geq 1 - 2\varepsilon.$$

We will also need the following claim, that is often used when one wishes to use seeded extractors with a non perfect seed.

Claim 3.9 (strong extractors with weak seeds)**.** *Let* $\mathsf{Ext} \colon \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ *be a strong* (k, ε) *extractor, let* X *be an* (n,k) *source, and let* Y *be* δ*-close to a* $(d, d-g)$ *source which is independent of* X. *Then, with probability at least* $1 - 2^g \sqrt{\varepsilon} - \delta$ *over* $y \sim Y$, *it holds that* $\mathsf{Ext}(X,y) \approx_{\sqrt{\varepsilon}} U_m$.

Proof: By Markov's inequality, there exists a set $\mathrm{BAD} \subseteq \{0,1\}^d$ of density $\rho(\mathrm{BAD}) \leq \sqrt{\varepsilon}$ such that for every $z \notin \mathrm{BAD}$ it holds that $\mathsf{Ext}(X,z) \approx_{\sqrt{\varepsilon}} U_m$. Let Y' be a $(d, d-g)$ source that is δ-close to Y. Then,

$$\begin{aligned}\Pr[Y \in \mathrm{BAD}] &\leq \Pr[Y' \in \mathrm{BAD}] + \delta \\ &= \sum_{z \in \mathrm{BAD}} \Pr[Y' = z] + \delta \leq |\mathrm{BAD}| \cdot 2^{-(d-g)} + \delta \leq 2^g \sqrt{\varepsilon} + \delta.\end{aligned}$$

■

3.3 Affine Correlation Breakers

We proceed with formally defining affine correlation breakers (with advice).

Definition 3.10 (affine correlation breaker)**.** *We say that* $\mathsf{AffCB}\colon \{0,1\}^n \times \{0,1\}^d \times \{0,1\}^a \to \{0,1\}^m$ *is a* (t,k,ε) affine correlation breaker *if for all distributions* $X, X^{(1)}, \ldots, X^{(t)} \sim \{0,1\}^n$, $B, B^{(1)}, \ldots, B^{(t)} \sim \{0,1\}^n$, $A, A^{(1)}, \ldots, A^{(t)} \sim \{0,1\}^n$, $Y, Y^{(1)}, \ldots, Y^{(t)} \sim \{0,1\}^d$, *and all strings* $\alpha, \alpha^{(1)}, \ldots, \alpha^{(t)} \in \{0,1\}^a$ *such that:*

- $X = A + B$ *and* $X^{(i)} = A^{(i)} + B^{(i)}$ *for all* $i \in [t]$,
- $H_\infty(B) \geq k$ *and* Y *is uniform,*
- $\left(B, B^{(1)}, \ldots, B^{(t)}\right)$ *is independent of* $\left(A, A^{(1)}, \ldots, A^{(t)}, Y, Y^{(1)}, \ldots, Y^{(t)}\right)$*, and,*
- *For all* $i \in [t]$, $\alpha \neq \alpha^{(i)}$,

it holds that

$$\left(\mathsf{AffCB}(X,Y,\alpha), \left\{\mathsf{AffCB}\left(X^{(i)}, Y^{(i)}, \alpha^{(i)}\right)\right\}_{i\in[t]}\right) \approx_\varepsilon \left(U_m, \left\{\mathsf{AffCB}\left(X^{(i)}, Y^{(i)}, \alpha^{(i)}\right)\right\}_{i\in[t]}\right).$$

We say that AffCB *is* strong *if the above holds also when we add* $Y, Y^{(1)}, \ldots, Y^{(t)}$.

We will use the following recent affine correlation breaker.

Theorem 3.11 ([Li23])**.** *For any positive integers* n, m, t, a, *and any* $\varepsilon > 0$, *there exists an explicit strong* (t,k,ε) *affine correlation breaker* $\mathsf{AffCB}\colon \{0,1\}^n \times \{0,1\}^d \times \{0,1\}^a \to \{0,1\}^m$, *where* $k = O\big(tm + ta + t^2 \log^2(t+1)\log(nt/\varepsilon)\big)$ *and* $d = O\big(m + ta + t\log^3(t+1)\log(nt/\varepsilon)\big)$.

3.4 Entropy-Resilient Functions

We summarize the required definitions and results given in [BCDT19].

Definition 3.12 (NOSF sources)**.** *Let* $\Sigma = \{0,1\}^m$. *A* (q,t) non-oblivious Σ-fixing source $X = (X_1, \ldots, X_D)$ *is a random variable over* $\Sigma^D = \{0,1\}^{Dm}$ *for which there exists a set* $\mathrm{BAD} \subseteq [D]$ *of cardinality* $q' \leq q$ *such that:*

- *The joint distribution of* $\{(X_i)_j : i \in [D] \setminus \mathrm{BAD}, j \in [m]\}$, *denoted by* G_X, *is* t*-wise independent over* $\{0,1\}^{(D-q')m}$; *and*

- *Each of the random variables in* $B_X \triangleq \{(X_i)_j\}$ *with* $i \in \mathrm{BAD}$ *and* $j \in [m]$ *may depend arbitrarily on all other random variables in* G_X *and* B_X.

If $t = (D - q')m$, *we say that* X *is a* q*-non-oblivious* Σ*-fixing source. If* $m = 1$, *the definition coincides with the standard definition of non-oblivious bit-fixing sources.*

In [BCDT19], Ben-Aroya et al. constructed seedless condensers for NOSF sources, also known as *entropy-resilient functions*.

Definition 3.13 (entropy-resilient functions)**.** *Let* $\Sigma = \{0,1\}^m$. *A function* $f\colon \Sigma^D \to \Sigma$ *is a* (q,t,g,ε) entropy-resilient function *if for every* (q,t) *non-oblivious* Σ*-fixing source* X *over* Σ^D, *the output* $f(X)$ *is* ε*-close to an* $(m, m-g)$*-source.*

Theorem 3.14 ([BCDT19])**.** *For every constant* $0 < \gamma < 1$ *there exist constants* $0 < \alpha < 1$ *and* $c' \geq 1$ *such that the following holds. For all integers* D, $m \leq D^{\alpha/2}$, *every* $\varepsilon > 0$, *and for every integer* $t \geq m \cdot (\log D)^{c'}$, *there exists an explicit function* $\mathsf{EntRes}\colon \Sigma^D \to \Sigma$, *for* $\Sigma = \{0,1\}^m$, *that is* $(q = D^{1-\gamma}, t, g, \varepsilon)$ *entropy-resilient, with entropy gap* $g = o(\log(1/\varepsilon))$.

3.5 Error Correcting Codes

A binary code $\mathcal{C} \subseteq \mathbb{F}_2^n$ is *linear* if $\mathcal{C}$ is a linear subspace of $\mathbb{F}_2^n$. The dimension of $\mathcal{C}$ as a subspace is called the *dimension* of the code. We will identify a linear code $\mathcal{C} \subseteq \mathbb{F}_2^n$ of dimension k with the image of its encoding function $\mathcal{C}\colon \mathbb{F}_2^k \to \mathbb{F}_2^n$, which is given by a *generator matrix* $A \in \mathbb{F}_2^{n \times k}$. A *parity check matrix* of $\mathcal{C}$ is a generator matrix $P \in \mathbb{F}_2^{(n-k)\times n}$ of the dual code $\mathcal{C}^\perp = \{x \in \mathbb{F}_2^n : \langle x, c\rangle = 0 \ \forall c \in \mathcal{C}\}$, and thus $Px = 0$ if and only if $x \in \mathcal{C}$.

We say that an error correcting code $\mathcal{C} \subseteq \mathbb{F}_2^n$ has *distance* d if for any distinct codewords $x, y \in \mathcal{C}$ it holds that $x_i \neq y_i$ in at least d i-s. We say that $\mathcal{C}$ is an $[n,k,d]$ code, if $\mathcal{C} \subseteq \mathbb{F}_2^n$ is a linear code of dimension k and distance d. We will use the following explicit code, the BCH code.

Theorem 3.15 ([BRC60, Hoc59])**.** *There exists a constant* $c > 0$ *such that the following holds. For every positive integers* n *and* $d < n$, *there exists an* $[n, r, d]$ *code* $\mathcal{C}_{\mathsf{BCH}}$ *with co-dimension* $n - r = \lfloor c \log n \rfloor \cdot d$.[17] *Moreover, the parity-check matrix of* $\mathcal{C}_{\mathsf{BCH}}$ *can be constructed in time* $\mathrm{poly}(n)$.

4 Low-Error Affine Condensers

In this section, we give our construction of low-error affine condensers with tiny entropy gap.

Theorem 4.1. *There exist universal constants* $c \geq 1$ *and* $\alpha \in (0,1)$ *such that the following holds. For any positive integers* n, k, *and any* $\varepsilon > 0$ *such that* $n \geq k \geq \log^c\left(\frac{n}{\varepsilon}\right)$, *there exists an explicit function* $\mathsf{AffCond}\colon \{0,1\}^n \to \{0,1\}^m$ *where* $m \geq k^\alpha$ *such that for any* (n,k) *affine source* X, *it holds that* $\mathsf{AffCond}(X)$ *is* ε*-close to an* $(m, m-g)$ *source, where* $g = o(\log(1/\varepsilon))$.

We will start by reducing an affine source to a non-oblivious symbol-fixing (NOSF) source.

[17] The original BCH code assumes $n = 2^m - 1$ for some integer m, and then $n - r \leq \frac{md}{2}$. Standard manipulations let us work for any n.

Lemma 4.2. *There exist constants $c, c_1 > 1$ such that the following holds, for any positive integers n, t, m, k, and any $\varepsilon > 0$, satisfying $k \geq (mt \log(n/\varepsilon))^c$. There exists an explicit function $\mathsf{AffToNOSF}\colon \{0,1\}^n \to \Sigma^D$ with $\Sigma = \{0,1\}^m$ and $D = n^{c_1}$ such that for every (n,k) affine source X, $\mathsf{AffToNOSF}(X)$ is ε-close to a $(D^{1-1/c_1}, t)$ NOSF source.*

Proof: We use the following two building blocks.

- Let $\mathsf{LExt}\colon \{0,1\}^n \times \{0,1\}^{d_0} \to \{0,1\}^d$ be the linear strong $(k_{\mathsf{Ext}}, \varepsilon_{\mathsf{Ext}} = 1/2n)$ extractor given by Theorem 3.6, where $k_{\mathsf{Ext}} = d^2 \geq c_1 \log^8 n$, for c_1 being the constant given in Theorem 3.6. Note that we can take $d_0 = c_1 \cdot \log n$.
- Let $\mathsf{AffCB}\ \{0,1\}^n \times \{0,1\}^d \times \{0,1\}^{d_0} \to \{0,1\}^m$ be the $(t-1, k_{\mathsf{CB}} = k - td, \varepsilon_{\mathsf{CB}})$ affine correlation breaker of Theorem 3.11 with $\varepsilon_{\mathsf{CB}} = \frac{\varepsilon}{t(mD)^{tm}}$, and

  $$d = O\big(m + td_0 + t\log^3(t+1)\log(nt/\varepsilon_{\mathsf{CB}})\big) = O\big(t^5 m \log n + t^4 \log(1/\varepsilon)\big).$$

  We need to make sure that $k \geq k_{\mathsf{CB}} + td$ where $k_{\mathsf{CB}} = O(tm + td_0 + t^2\log^2(t+1)\log(nt/\varepsilon_{\mathsf{CB}}))$, and indeed, this is satisfied by taking $k \geq C \cdot \big(t^6 m \log n + t^5 \log(1/\varepsilon)\big)$ for a large enough constant C. We also need to make sure that $k \geq d^2$ and $k \geq c\log^8 n$, and both inequalities, together with our previous lower bound on k, indeed holds whenever

  $$k \geq C \cdot t^{10} m^2 \log^8 n \log^2(1/\varepsilon)$$

  for a large enough constant C.

Our construction goes as follows. For simplicity, identify $\{0,1\}^{d_0}$ with $[D = n^{c_1}]$. Given $x \in \{0,1\}^n$:

1. For every $i \in [D]$, compute $y_i = \mathsf{LExt}(x, i)$.
2. For every $i \in [D]$, compute $z_i = \mathsf{AffCB}(x, y_i, i)$.

We'll show that the table $Z = (Z_1, \ldots, Z_D)$ satisfies the requirement of the lemma. First, by Lemma 3.8, we have a set $\mathrm{BAD} \subseteq [D]$ of density $\rho(\mathrm{BAD}) \leq 1/n$ such that for any $i \notin \mathrm{BAD}$ it holds that $Y_i \sim \{0,1\}^d$ is uniform. Next, fix t good rows $\{i_1, \ldots, i_t\} \subseteq [D] \setminus \mathrm{BAD}$. Consider the linear function $L\colon \{0,1\}^n \to \{0,1\}^{t\cdot d}$ that is given by $(\mathsf{LExt}(\cdot, i_1), \ldots, \mathsf{LExt}(\cdot, i_t))$. Via the affine conditioning lemma, Lemma 3.4, we can write $X = A + B$, where:

- A and B are affine and independent.
- $H(B) \geq H(X) - td = k_{\mathsf{CB}}$.
- $L(B)$ is constant, so $Y_{i_1}, \ldots, Y_{i_t}$ is independent of B, and moreover, B is independent of $(A, Y_{i_1}, \ldots, Y_{i_t})$.

Recalling that each Y_{i_j} is uniform, all conditions are met to apply are affine correlation breaker, which gives us

$$\left(Z_{i_j}, \{Z_{i_k}\}_{k\in[t]\setminus\{j\}}\right) \approx_{\varepsilon_{\mathsf{CB}}} \left(U_m, \{Z_{i_k}\}_{k\in[t]\setminus\{j\}}\right)$$

for every $j \in [t]$. Therefore,

$$(Z_{i_1}, \ldots, Z_{i_t}) \approx_{t\cdot\varepsilon_{\mathsf{CB}}} U_{tm},$$

and so by Lemma 3.1, Z itself is $(mD)^{tm} \cdot t\varepsilon_{\mathsf{CB}} = \varepsilon$ close to a (q,t) non-oblivious $\{0,1\}^m$-fixing source for $q = \frac{D}{n} = D^{1-\frac{1}{c_1}}$. ∎

Next, we can simply apply the low-error condenser for NOSF sources given by [BCDT19].

Proof of Theorem 4.1: Given n and $\varepsilon > 0$, let m be the number of bits we eventually output, and the entropy of X will be determined according to that later on. Set $\gamma = \frac{1}{c_1}$, and let $\alpha \in (0,1)$ and $c' \geq 1$ be the constants that are set according to γ, guaranteed to us by Theorem 3.14.

Let $\mathsf{AffToNOSF}\colon \{0,1\}^n \to \Sigma^D$ be the function from Lemma 4.2, where $\Sigma = \{0,1\}^m$ and $D = n^{c_1}$, set with $t \leftarrow m \cdot d_0^{c'}$ and $\varepsilon \leftarrow \varepsilon/2$. By that lemma, there exists a constant $C > 1$ such that whenever $k \geq (m \log(n/\varepsilon))^C$ and X is an (n,k) affine source, it holds that $\mathsf{AffToNOSF}(X)$ is $\frac{\varepsilon}{2}$-close to an $(D^{1-\gamma}, t)$ NOSF source.

Now, let $\mathsf{EntRes}\colon \Sigma^D \to \Sigma$ be the $(D^{1-\gamma}, t, g, \varepsilon/2)$ entropy resilient function from Theorem 3.14. Whenever $m \leq D^{\alpha/2}$, we are guaranteed that $\mathsf{EntRes}(\mathsf{AffToNOSF}(X))$ is ε-close to an $(m, m-g)$ source with $g = o(\log(1/\varepsilon))$. To conclude, we have that there exist constants $C \geq 1$ and $\delta \in (0,1)$ such that whenever $m \leq n^\delta$ and $k \geq (m \log(n/\varepsilon))^C$, the function

$$\mathsf{AffCond}(x) = \mathsf{EntRes}(\mathsf{AffToNOSF}(x))$$

is an affine condenser with entropy gap $g = o(\log(1/\varepsilon))$. Putting it differently, there exist constants $C' \geq 1$ and $\delta' \in (0,1)$ such that whenever $k \geq (\log(n/\varepsilon))^{C'}$, our condenser outputs $k^{\delta'}$ bits. ■

5 Extractors for OBF Sources

5.1 Linear OBF Condensers

We begin by defining these condensers, and focus on the linear setting.

Definition 5.1 (OBF condenser). *A matrix $L \in \mathbb{F}_2^{m \times n}$ is an $(n,k) \to (m,t)$* linear condenser for OBF sources *if $H_\infty(L(X)) \geq t$ for any (n,k) OBF source X.*

Rao showed that linear condensers for OBF sources can be obtained by parity check matrices of binary codes. The next statement is a bit different than the one in [Rao09], so we provide the proof for completeness.

Lemma 5.2 (following [Rao09]). *Let X be an (n,k) OBF source, let $k' \leq k$ be any positive integer, and let $\mathcal{C} \subseteq \mathbb{F}_2^n$ be an $[n,r,d]$ code with a parity check matrix $P \in \mathbb{F}_2^{m \times n}$, where $m = n - r$ and $d \geq k' + 1$. Then, $Z = P \cdot X \sim \{0,1\}^m$ satisfies $H_\infty(Z) \geq k'$.*

Proof: Letting $i_1, \ldots, i_k \in [n]$ be the entropic coordinates of X, and let $c \in \mathbb{F}_2^n$ be its constant shift.[18] Consider

$$V_{k'} = \mathrm{Span}\left(\left\{e_{i_1}, \ldots, e_{i_{k'}}\right\}\right) \subseteq \mathrm{Supp}(X) + c,$$

observing that each element of $V_{k'}$ has Hamming weight at most k'. Since $\mathcal{C}$ has distance greater than k', $V_{k'} \cap \mathcal{C} = \{0\}$ and so P is injective on $V_{k'}$ and thus also on $V_{k'} + c$. Let W be the subspace for which $\mathrm{Supp}(X) + c = V_{k'} \oplus W$, and also refer to $V_{k'}$ and W as the corresponding flat distributions. Thus, for each $z \in \{0,1\}^m$,

$$\Pr[Z = z] = \Pr[P \cdot V_{k'} = z + P \cdot W + c] = \mathop{\mathbb{E}}_{w \sim W}[\Pr[P \cdot V_{k'} = z + P \cdot w + c]] \leq 2^{-k'},$$

and we are done.

■

[18] We can assume that c is zero on $i_1, \ldots, i_k$.

Using the BCH codes from Theorem 3.15, we get the following linear condensers for OBF sources.

Corollary 5.3. *There exists a constant $c > 0$ such that the following holds. For every positive integers n, $k \leq n$, and $k' \leq k$, there exists an explicit linear function $P\colon \{0,1\}^n \to \{0,1\}^m$ for $m = ck' \log n$, such that the following holds. For every (n,k) OBF source X, it holds that $H_\infty(P(X)) \geq k'$. That is, P is an $(n,k) \to (ck' \log n, k')$ linear condenser for OBF sources.*

5.2 OBF Extractors for Almost-Logarithmic Entropy

We are given $n, k \in \mathbb{N}$, and $\gamma > 0$. Let ε be our error guarantee, to be determined later on. We use the following ingredients:

- Let $P_1\colon \{0,1\}^n \to \{0,1\}^{n_1}$ be the linear condenser for OBF sources given in Corollary 5.3, set with $k' \leftarrow k$, so $n_1 = c \cdot k \log n$ for some universal constant $c > 0$.
- Let $P_2\colon \{0,1\}^n \to \{0,1\}^{n_2}$ be the linear condenser for OBF sources given in Corollary 5.3, set with $k' \leftarrow \frac{\gamma k}{2c \log n} \triangleq k_{\mathsf{Cond}}$, so $n_2 = \gamma k/2$.
- Let $\mathsf{AffCond}\colon \{0,1\}^{n_2} \to \{0,1\}^d$ be the affine condenser from Theorem 4.1, set with $k \leftarrow k_{\mathsf{Cond}}$ and $\varepsilon \leftarrow \varepsilon/2$.
- Let $\mathsf{LExt}\colon \{0,1\}^{n_1} \times \{0,1\}^d \to \{0,1\}^m$ be the $(k_{\mathsf{Ext}} = (1-\gamma/2)k, \varepsilon_{\mathsf{Ext}})$ extractor of Theorem 3.7, where $\varepsilon_{\mathsf{Ext}} = \varepsilon^4/16$ and $m = (1-\gamma/2)k_{\mathsf{Ext}} \geq (1-\gamma)k$. There exists a constant c_2 such that a seed of length $d = c_2\big(\log(n_1) + \log^2(k/\varepsilon) \cdot \log(1/\gamma)\big)$ suffices.

Now, given $x \in \{0,1\}^n$, our extractor $\mathsf{OBFExt}\colon \{0,1\}^n \to \{0,1\}^m$ is given by

$$\mathsf{OBFExt}(x) = \mathsf{LExt}(P_1(x), \mathsf{AffCond}(P_2(x)))$$

The fact that OBFExt is explicit readily follows from the explicitness of its components. The correctness is established in the following theorem.

Theorem 5.4. *There exist universal constants $c_k, c_\gamma > 1$ and $\beta \in (0,1)$ such that the following holds, assuming $k \geq \frac{\log n}{\gamma} \cdot (\log \frac{\log n}{\gamma})^{c_k}$ and $\gamma \geq \frac{\log^{c_\gamma} k}{k}$. For every (n,k) OBF source X it holds that $\mathsf{OBFExt}(X) \approx_\varepsilon U_m$ for $\varepsilon = 2^{-((\gamma k)/\log n)^\beta}$, recalling that $m = (1-\gamma)k$.*

Proof: Denote $X_1 = P_1(X)$, $X_2 = P_2(X)$, $Y = \mathsf{AffCond}(X_2)$, and $Z = \mathsf{LExt}(X_1, Y)$. We "affine condition" (Theorem 4.1) according to P_2, and so we can write X as $X = A + B$, where A and B are affine and independent and there exists a linear bijection between A and X_2 (so B is independent of X_2 and thus also of Y). We can then write

$$Z = \mathsf{LExt}(X_1, Y) = \mathsf{LExt}(P_1(A), Y) + \mathsf{LExt}(P_1(B), Y).$$

We know that $H(X_1) \geq k$, $H(A) \leq n_2$, and $H(B) \geq H(X) - n_2 \geq (1-\gamma/2)k = k_{\mathsf{Ext}}$. Also, $X_1 = P_1(A) + P_1(B)$, and $P_1(A)$ and $P_1(B)$ are independent. Thus, $H(P_1(B)) \geq k_{\mathsf{Ext}}$ as well,[19] and $P_1(B)$ is independent of Y.

[19] To see this, recall that B is a subspace of X (up to a shift), and P_1 is injective on X.

Claim 5.5. *With probability at least* $1 - \varepsilon/2$ *over* $y \sim Y$ *it holds that* $\mathsf{LExt}(P_1(B), y) \approx_{\varepsilon/2} U_m$.

Proof: To apply our affine condenser, Theorem 4.1, we need to make sure that:

1. $k_{\mathsf{Cond}} \geq \log^{c'}(n_2/\varepsilon)$, where $c' > 1$ is the constant guaranteed to us by Theorem 4.1. Recall that

$$\log^{c'}\left(\frac{n_2}{\varepsilon}\right) = O\left(\log^{c'}\left(\frac{k}{\varepsilon}\right)\right),$$

   and that

$$k_{\mathsf{Cond}} = \Omega\left(\frac{\gamma k}{\log n}\right),$$

   so the inequality is met as long as $k \geq \frac{\log n}{\gamma} \cdot (\log \frac{\log n}{\gamma})^{c_k}$ and $\varepsilon \geq 2^{-((\gamma k)/\log n)^{\beta}}$ for some constants $c_k > 1$ and $\beta \in (0,1)$.

2. $d \leq k_{\mathsf{Cond}}^{\alpha}$, where $\alpha \in (0,1)$ is the constant guaranteed to us by Theorem 4.1. Plugging in the expression for d above, we get that similarly, we need to satisfy:
   (a) $k \geq \frac{\log n}{\gamma} \cdot (\log \frac{\log n}{\gamma})^{c_k}$ for some constants $c_k > 1$,
   (b) $\varepsilon \geq 2^{-((\gamma k)/\log n)^{\beta}}$ for some constant $\beta \in (0,1)$, and,
   (c) $\gamma \geq \frac{\log^{c_\gamma} k}{k}$ for some constant $c_\gamma > 1$.

So indeed, Theorem 4.1 tells us that Y itself is $(\varepsilon/4)$-close to an $(d, d-g)$-source for $g = o(\log(1/\varepsilon))$, and thus by Claim 3.9 we have that $\mathsf{LExt}(P_1(B), y) \approx_{\sqrt{\varepsilon_{\mathsf{Ext}}}} U_m$ except for probability

$$2^g\sqrt{\varepsilon_{\mathsf{Ext}}} + \frac{\varepsilon}{4} \leq \frac{1}{\varepsilon} \cdot \frac{\varepsilon^2}{4} + \frac{\varepsilon}{4} \leq \frac{\varepsilon}{2}.$$

■

Fix a good y (in the sense of the above claim), and observe that $\mathsf{LExt}(P_1(A), y)$ is independent from $\mathsf{LExt}(P_1(B), y)$, making

$$\mathsf{LExt}(X_1, y) = \mathsf{LExt}(P_1(A), y) + \mathsf{LExt}(P_1(B), y)$$

close to uniform as well. Overall, $Z \approx_{2\cdot(\varepsilon/2)} U_m$, and we are done. ■

Choosing $\gamma = \frac{1}{\log^c k}$ for any constant c, we obtain our Theorem 1.3.

5.3 Handling Low-Weight Affine Sources

We say that an (n,k) affine source X has *weight* w if each basis element of X has weight at most w. This generalizes OBF sources (that have $w = 1$), and [Rao09] can handle $w = k^{\alpha}$ for a small constant $\alpha > 0$. Note that the only place in the proof that we used the fact that X is an OBF source (rather than a full-fledges affine source) is Lemma 5.2, where we argued that codes with sufficiently large distance give linear condensers. This can easily be extended to the $w > 1$ case.

Lemma 5.6 (following [Rao09])**.** *Let* X *be an* (n,k) *affine source of weight* w*, let* $k' \leq k$ *be any positive integer, and let* $\mathcal{C} \subseteq \mathbb{F}_2^n$ *be an* $[n, r, d]$ *code with a parity check matrix* $P \in \mathbb{F}_2^{m \times n}$*, where* $m = n - r$ *and* $d \geq wk' + 1$*. Then,* $Z = P \cdot X \sim \{0,1\}^m$ *satisfies* $H_{\infty}(Z) \geq k'$.

Thus, as w grows, the co-dimension of the code we need to take grows too, and so the condensing quality decreases. To modify our construction to handle $w > 1$, we need to set $n_1 = c \cdot wk \log n$ and $k_{\mathsf{Cond}} = \frac{\gamma k}{2cw \log n}$. Repeating the same analysis as in Section 5.2, we see that one can take $w = \mathrm{poly}(\log k)$ without significant loss in parameters.

Theorem 5.7. *For any constant $\ell > 0$ there exist constants $c > 1$ and $\beta \in (0, 1)$ such that the following holds for any $n \in \mathbb{N}$ and $k \geq \log n \cdot (\log \log n)^c$. There exists an explicit function*

$$\mathsf{OBFExt}\colon \{0,1\}^n \to \{0,1\}^{m=(1-o(1))k}$$

that is a (k, ε) extractor for affine sources of weight $w = \log^\ell k$, where $\varepsilon = 2^{-(k/\log n)^\beta}$.

Acknowledgment

We thank Jesse Goodman for many helpful and interesting conversations.

References

[AGM03] Noga Alon, Oded Goldreich, and Yishay Mansour. Almost k-wise independence versus k-wise independence. *Information Processing Letters*, 88(3):107–110, 2003.

[AL93] Miklós Ajtai and Nathan Linial. The influence of large coalitions. *Combinatorica*, 13(2):129–145, 1993.

[BADT19] Avraham Ben-Aroya, Dean Doron, and Amnon Ta-Shma. An efficient reduction from two-source to nonmalleable extractors:: Achieving near-logarithmic min-entropy. *SIAM Journal on Computing*, 51(2):STOC17–31, 2019.

[BADT20] Avraham Ben-Aroya, Dean Doron, and Amnon Ta-Shma. Near-optimal erasure list-decodable codes. In *Computational Complexity Conference (CCC)*, pages 1:1–1:27. Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2020.

[BBR85] Charles H. Bennett, Gilles Brassard, and Jean-Marc Robert. How to reduce your enemy's information. In *Advances in Cryptology (CRYPTO)*, volume 218 of *Lecture Notes in Computer Science*, pages 468–476. Springer, 1985.

[BCDT19] Avraham Ben-Aroya, Gil Cohen, Dean Doron, and Amnon Ta-Shma. Two-source condensers with low error and small entropy gap via entropy-resilient functions. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM)*, pages 43:1–43:20. Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2019.

[Bla96] Matt Blaze. High-bandwidth encryption with low-bandwidth smartcards. In *Fast Software Encryption: Third International Workshop Cambridge, UK, February 21–23 1996 Proceedings 3*, pages 33–40. Springer, 1996.

[Bou05] Jean Bourgain. More on the sum-product phenomenon in prime fields and its applications. *International Journal of Number Theory*, 1(01):1–32, 2005.

[BRC60] Raj Chandra Bose and Dwijendra K. Ray-Chaudhuri. On a class of error correcting binary group codes. *Information and control*, 3(1):68–79, 1960.

[CDH+00] Ran Canetti, Yevgeniy Dodis, Shai Halevi, Eyal Kushilevitz, and Amit Sahai. Exposure-resilient functions and all-or-nothing transforms. In *Advances in Cryptology (EUROCRYPT)*, pages 453–469. Springer, 2000.

[CGH+85] Benny Chor, Oded Goldreich, Johan Håstad, Joel Freidmann, Steven Rudich, and Roman Smolensky. The bit extraction problem or t-resilient functions. In *Annual Symposium on Foundations of Computer Science (FOCS)*, pages 396–407. IEEE, 1985.

[CGL22] Eshan Chattopadhyay, Jesse Goodman, and Jyun-Jie Liao. Affine extractors for almost logarithmic entropy. In *Annual Symposium on Foundations of Computer Science (FOCS)*, pages 622–633. IEEE, 2022.

[CI17] Mahdi Cheraghchi and Piotr Indyk. Nearly optimal deterministic algorithm for sparse Walsh-Hadamard transform. *ACM Transactions on Algorithms (TALG)*, 13(3):1–36, 2017.

[CKK+15] Ruiwen Chen, Valentine Kabanets, Antonina Kolokolova, Ronen Shaltiel, and David Zuckerman. Mining circuit lower bound proofs for meta-algorithms. *computational complexity*, 24:333–392, 2015.

[CL22] Eshan Chattopadhyay and Jyun-Jie Liao. Extractors for sum of two sources. In *Annual Symposium on Theory of Computing (STOC)*, pages 1584–1597. ACM, 2022.

[Coh16] Gil Cohen. Local correlation breakers and applications to three-source extractors and mergers. *SIAM Journal on Computing*, 45(4):1297–1338, 2016.

[Coh17] Gil Cohen. Towards optimal two-source extractors and Ramsey graphs. In *Annual Symposium on Theory of Computing (STOC)*, pages 1157–1170. ACM, 2017.

[CS15] Gil Cohen and Igor Shinkar. Zero-fixing extractors for sub-logarithmic entropy. In *International Colloquium on Automata, Languages, and Programming (ICALP)*, pages 343–354. Springer, 2015.

[CZ19] Eshan Chattopadhyay and David Zuckerman. Explicit two-source extractors and resilient functions. *Annals of Mathematics*, 189(3):653–705, 2019.

[DKM+06] Cynthia Dwork, Krishnaram Kenthapadi, Frank McSherry, Ilya Mironov, and Moni Naor. Our data, ourselves: Privacy via distributed noise generation. In *Advances in Cryptology (EUROCRYPT)*, pages 486–503. Springer, 2006.

[Dod00] Yevgeniy Dodis. *Exposure-Resilient Cryptography*. Phd thesis, MIT, 2000.

[DSS01] Yevgeniy Dodis, Amit Sahai, and Adam Smith. On perfect and adaptive security in exposure-resilient cryptography. In *Advances in Cryptology (EUROCRYPT)*, pages 301–324. Springer, 2001.

[Fri92] Joel Friedman. On the bit extraction problem. In *Annual Symposium on Foundations of Computer Science (FOCS)*, volume 33, pages 314–314. IEEE, 1992.

[Goo23] Jesse Patrick McGrenra Goodman. *Seedless Extractors*. Phd thesis, Cornell University, 2023. Available at https://jpmgoodman.com/thesis.pdf.

[GR08] Ariel Gabizon and Ran Raz. Deterministic extractors for affine sources over large fields. *Combinatorica*, 28(4):415–440, 2008.

[GRS06] Ariel Gabizon, Ran Raz, and Ronen Shaltiel. Deterministic extractors for bit-fixing sources by obtaining an independent seed. *SIAM Journal on Computing*, 36(4):1072–1094, 2006.

[GUV09] Venkatesan Guruswami, Christopher Umans, and Salil Vadhan. Unbalanced expanders and randomness extractors from parvaresh–vardy codes. *Journal of the ACM (JACM)*, 56(4):20, 2009.

[Hoc59] Alexis Hocquenghem. Codes correcteurs d'erreurs. *Chiffers*, 2:147–156, 1959.

[JSY99] Markus Jakobsson, Julien P. Stern, and Moti Yung. Scramble all, encrypt small. In *International Workshop on Fast Software Encryption*, pages 95–111. Springer, 1999.

[KKL88] Jeff Kahn, Gil Kalai, and Nathan Linial. The influence of variables on boolean functions. In *Annual Symposium on Foundations of Computer Science (FOCS)*, pages 68–80. IEEE, 1988.

[KRT13] Ilan Komargodski, Ran Raz, and Avishay Tal. Improved average-case lower bounds for demorgan formula size. In *Annual Symposium on Foundations of Computer Science (FOCS)*, pages 588–597. IEEE, 2013.

[KZ06] Jesse Kamp and David Zuckerman. Deterministic extractors for bit-fixing sources and exposure-resilient cryptography. *SIAM Journal on Computing*, 36(5):1231–1247, 2006.

[Lew19] Mark Lewko. An explicit two-source extractor with min-entropy rate near 4/9. *Mathematika*, 65(4):950–957, 2019.

[Li11] Xin Li. A new approach to affine extractors and dispersers. In *Computational Complexity Conference (CCC)*, pages 137–147. IEEE, 2011.

[Li13a] Xin Li. Extractors for a constant number of independent sources with polylogarithmic min-entropy. In *Annual Symposium on the Foundations of Computer Science (FOCS)*, pages 100–109. IEEE, 2013.

[Li13b] Xin Li. New independent source extractors with exponential improvement. In *Annual Symposium on Theory of Computing (STOC)*, pages 783–792. ACM, 2013.

[Li15] Xin Li. Three-source extractors for polylogarithmic min-entropy. In *2015 IEEE 56th Annual Symposium on Foundations of Computer Science*, pages 863–882. IEEE, 2015.

[Li16] Xin Li. Improved two-source extractors, and affine extractors for polylogarithmic entropy. In *Annual Symposium on Foundations of Computer Science (FOCS)*, pages 168–177. IEEE, 2016.

[Li19] Xin Li. Non-malleable extractors and non-malleable codes: Partially optimal constructions. In *Computational Complexity Conference (CCC)*, pages 28:1–28:49. Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2019.

[Li23] Xin Li. Two source extractors for asymptotically optimal entropy, and (many) more. In *Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1271–1281. IEEE, 2023.

[LZ24] Xin Li and Yan Zhong. Explicit directional affine extractors and improved hardness for linear branching programs. In *Computational Complexity Conference (CCC)*, pages 10:1–10:14. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2024.

[Rao09] Anup Rao. Extractors for low-weight affine sources. In *Computational Complexity Conference (CCC)*, pages 95–101. IEEE, 2009.

[Riv97] Ronald L Rivest. All-or-nothing encryption and the package transform. In *Fast Software Encryption: 4th International Workshop, FSE'97 Haifa, Israel, January 20–22 1997 Proceedings 4*, pages 210–218. Springer, 1997.

[RRV02] Ran Raz, Omer Reingold, and Salil Vadhan. Extracting all the randomness and reducing the error in trevisan's extractors. *Journal of Computer and System Sciences*, 65(1):97–128, 2002.

[Sha08] Ronen Shaltiel. How to get more mileage from randomness extractors. *Random Structures & Algorithms*, 33(2):157–186, 2008.

[SU05] Ronen Shaltiel and Christopher Umans. Simple extractors for all min-entropies and a new pseudorandom generator. *Journal of the ACM (JACM)*, 52(2):172–216, 2005.

[Tre01] Luca Trevisan. Extractors and pseudorandom generators. *Journal of the ACM (JACM)*, 48(4):860–879, 2001.

[vN51] John von Neumann. Various techniques used in connection with random digits. *Applied Math Series*, 12(36-38):5, 1951.

[Yeh11] Amir Yehudayoff. Affine extractors over prime fields. *Combinatorica*, 31(2):245–256, 2011.

ECCC ISSN 1433-8092

https://eccc.weizmann.ac.il