Electronic Colloquium on Computational Complexity, Report No. 106 (2016)

Low-error two-source extractors for polynomial min-entropy

Avraham Ben-Aroya* Dean Doron† Amnon Ta-Shma‡

Abstract

We construct explicit two-source extractors for n bit sources, requiring n^{α} min-entropy and having error $2^{-n^{\beta}}$, for some constants $0 < \alpha, \beta < 1$. Previously, constructions for exponentially small error required either min-entropy $0.49n$ [Bou05] or three sources [Li15b]. The construction combines somewhere-random condensers based on the Incidence Theorem [Zuc06, Li11], together with recent machinery surrounding non-malleable extractors.

*The Blavatnik School of Computer Science, Tel-Aviv University, Tel Aviv 69978, Israel. Supported by the Israel science Foundation grant no. 994/14 and by the United States – Israel Binational Science Foundation grant no. 2010120.

†The Blavatnik School of Computer Science, Tel-Aviv University, Tel Aviv 69978, Israel. Email: deandoron@mail.tau.ac.il. Supported by the Israel science Foundation grant no. 994/14 and by the United States – Israel Binational Science Foundation grant no. 2010120.

‡The Blavatnik School of Computer Science, Tel-Aviv University, Tel Aviv 69978, Israel. Email: amnon@tau.ac.il. Supported by the Israel science Foundation grant no. 994/14 and by the United States – Israel Binational Science Foundation grant no. 2010120.

ISSN 1433-8092

1 Introduction

A function $\mathsf{Ext} : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}^m$ is a *two-source k-extractor* with ε error if for every two independent distributions A, B over $\{0,1\}^n$ with min-entropy at least k, the output's distribution $\mathsf{Ext}(A,B)$ is ε-close to uniform.

It is well known that a two-source extractor with even just one output bit and any non-trivial error parameter $\varepsilon < 1$, that works for min-entropy k, implies a (bipartite) 2^k-Ramsey graph. A long line of research was devoted to the problem of explicitly constructing such a two-source extractor [Abb72, Nag75, Fra77, Chu81, FW81, Nao92, Alo98, Gro01, Bar06, BKS+10, BRSW12, Coh15b] culminating in the work of Chattopadhyay and Zuckerman [CZ15] who used non-malleable extractors to give a two-source extractor for $k = \mathrm{polylog}(n)$. Several improvements on the [CZ15] construction followed, including [Mek15, Li15a]. Currently, the best explicit construction achieves $k = \log^{1+o(1)}(n)$ [BADTS16].

The error parameter of an extractor should be measured with respect to its input entropy, as it can be easily seen that the error of a k-extractor must be at least $2^{-O(k)}$. The aforementioned papers construct two-source extractors with error parameter that is either a constant or polynomially small in the input entropy. While this suffices for Ramsey graph constructions, non-explicit constructions may have exponentially small error. Similarly, these constructions usually output few close-to-uniform bits, while non-explicitly, almost all of the entropy can be extracted.

There are several *explicit* two-source constructions with exponentially-small error. The inner-product function gives a simple solution when $k > n/2$ [CG88]. Bourgain [Bou05] gave a two-source extractor construction for $k = (\frac{1}{2} - \alpha)n$, for some small constant $\alpha > 0$. Raz [Raz05] constructed a two-source extractor that has an unbalanced entropy requirement; the first source should have more than $n/2$ min-entropy, while the second source's min-entropy can be as low as $c \cdot \log n$ (for some constant c).

The main result in this paper is a low-error two-source extractor that improves Bourgain's, and lowers the entropy requirement from $(\frac{1}{2} - \alpha)n$ to $n^{1-\alpha}$ for some small constant α. Specifically:

Theorem 1.1. *There exist constants $0 < \alpha, \beta, \gamma < 1$ such that for every large enough n and every $k \geq n^\alpha$, there exists an explicit $((n,k) \times (n,k) \to_\varepsilon m = k^\gamma)$ two-source extractor $\mathsf{2Ext} : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}^m$ with $\varepsilon = 2^{-k^\beta}$.*

We mention that if three sources are allowed, better explicit constructions are known. Specifically, [Li15b] constructs a three-source extractor with exponentially-small error for min-entropy $k = \mathrm{polylog}(n)$. Achieving the same with only two sources, is a challenging and important open problem.

1.1 The technique

The inner-product function over $\mathbb{F}_2$ maps $x, y \in \{0,1\}^n$ to $\mathsf{IP}_{\mathbb{F}_2}(x,y) = \sum x_i y_i \bmod 2$. It gives a two-source extractor with exponentially small error, if each of the two sources has min-entropy $k \gg n/2$. This entropy requirement is tight; $\mathsf{IP}_{\mathbb{F}_2}$ completely fails when X and Y are uniformly distributed over two orthogonal subspaces of dimension $n/2$ each.

Bourgain's two-source extractor applies the inner-product function on local encodings of the input. More precisely, the inputs x and y are viewed as coming from a low-dimensional vector space over a field $\mathbb{F}$ (e.g., $\mathbb{F}^2$). The extractor works as follows.

1. It *locally* encodes the inputs x and y to $C(x)$ and $C(y)$, where *local* here means that each source is encoded separately.

2. Then, it computes $\mathsf{IP}(C(x), C(y))$ where IP is the inner product function over $\mathbb{F}$.

3. Finally, it outputs some boolean function applied on $\mathsf{IP}(C(x), C(y))$.

One necessary (but not sufficient) role of the encoding C is to distort vector spaces. That is, the image $C(X)$ of any vector space X should be far away from any vector space in order to avoid the bad example described above. Bourgain ([Bou05], see also [Rao07]) considers several good functions C, and they all come from additive combinatorics and rely on the property that if a set does not expand much under addition (e.g., it is a vector space) then it must expand much under multiplication.

In a sense, our work abstracts the intuition behind Bourgain's construction and takes advantage of the new tools developed since then – most notably non-malleable extractors and advice correlation breakers. Abstractly, given x and y we do the following:

1. We locally encode one of the inputs. Say $y \sim Y$ and Y is an (n, k) source[1] for some $k \ll n/2$. We use a somewhere-random condenser that outputs a table with few rows $r_1, \ldots, r_t$, with the guarantee that one of the rows has density rate above half, i.e., it is an $(\ell, \alpha\ell)$-source for some $\alpha > \frac{1}{2}$.[2]

   The price of this step is that we get a somewhere-random source, i.e., we get t rows, one of which is good, and the rest are arbitrarily correlated with it. Fortunately, the number of rows can be small. In particular, [BKS+10, Zuc06] showed that for every constants $\alpha < \beta$, a constant number of rows suffices to improve the rate of the source from α to β.

2. We then use Raz's extractor. For each row r_i of the t rows in the local encoding of y we use the row as a seed in Raz's extractor applied on the other source, i.e., we output $\mathsf{Raz}(x, r_i)$. This step, conceptually, replaces the inner-product function in Bourgain's extractor.

3. At this stage we have a table with t rows, and one of which is exponentially-close to uniform. A *merger* now suffices in order to obtain an exponentially-close to uniform output.

We now face a problem. A merger requires a fresh seed, that can be obtained from another source, yet we want to use only two sources. This raises the question of whether we need to abandon the sources used so far, and the answer is a clear "No!".

The *hierarchy of independence* (see Subsection 3.2 and references therein) gives sufficient conditions under which sources can be reused. Using this, Cohen [Coh15a] showed how to merge a table with t rows, one of which is *uniform*, using a weak random source X. We incorporate steps 2 and 3 together, and show that using Raz's extractor, Cohen's work extends to the case we need, as long as the good row has entropy rate at least $1 - \Omega\left(\frac{1}{t^3}\right)$. Specifically, there exists a merger that takes

- a table with t rows, one of which has rate at least $1 - \Omega\left(\frac{1}{t^3}\right)$, and,
- a weak (n, k) source X,

[1] An (n, k)-source is a random variable over $\{0, 1\}^n$ with k min-entropy.

[2] More accurately, we are in a convex combination of such sources.

and outputs nearly uniform bits with exponentially small error (see Section 3 for precise details).

This suggests the following construction. Suppose Y is a weak (n,k)-source. Condense Y using a somewhere-random condenser to a somewhere-random source with t rows, where one of the rows has rate at least $1-\Omega\left(\frac{1}{t^3}\right)$. Then apply the merger on the somewhere-random source and the weak source X.

The only question left is for which min-entropy k can we find a somewhere-random condenser with the required parameters. A-priori, it is not clear that such condensers exist. However, when $k=\alpha n$ and α is a constant, the somewhere random condensers of [Zuc06] easily achieve the above parameters.

When α is sub-constant, t becomes super-constant, and then the required rate $1-\Omega\left(\frac{1}{t^3}\right)$ becomes very close to 1. In this setting [Zuc06] does not guarantee anything. This stems from the fact that [Zuc06] relies on the Incidence Theorem of [BKT04], which does not support too large sets. Fortunately, Vinh [Vin11] extended the Incidence theorem to the high entropy regime, with a simple and elegant proof. This was used by Li [Li11] to give a somewhere-random condenser where one of the rows in the table has rate close to 1 (see Section 4) that is sufficient for our needs. In particular, working out the parameters and balancing the entropies of the two sources, we obtain Theorem 1.1.

2 Preliminaries

Throughout the paper we have the convention that lowercase variables are the logarithm (in base-2) of their corresponding uppercase variables, e.g., $n=\log N$, $d=\log D$, $a=\log A$, $r=\log R$, $r'=\log R'$, etc. We denote by $[t]$ the set $\{1,\ldots,t\}$.

2.1 Random variables, min-entropy

The *statistical distance* between two distributions X and Y on the same domain D is defined as $|X-Y|=\max_{A\subseteq D}(\Pr[X\in A]-\Pr[Y\in A])$. If $|X-Y|\leq\varepsilon$ we say that X is ε-close to Y and denote it by $X\approx_\varepsilon Y$. We will denote by U_n a random variable distributed uniformly over $\{0,1\}^n$ and which is independent of all other variables. We also say that a random variable is *flat* if it is uniform over its support.

For a function $f:D_1\to D_2$ and a random variable X distributed over D_1, $f(X)$ is the random variable distributed over D_2 which is obtained by choosing x according to X and computing $f(x)$. For a set $A\subseteq D_1$, we simply denote $f(A)=\{f(x)\mid x\in A\}$. It is well-known that for every $f:D_1\to D_2$ and two random variables X and Y, distributed over D_1, it holds that $|f(X)-f(Y)|\leq|X-Y|$.

The *min-entropy* of a random variable X is defined by

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

A random variable X distributed over $\{0,1\}^n$ with min-entropy at least k is called an (n,k)-*source*. Every distribution X with $H_\infty(X)\geq k$ can be expressed as a convex combination of flat distributions, each with min-entropy at least k.

For $\varepsilon\geq 0$, the *smooth min-entropy* $H_\infty^\varepsilon(X)$ is the supremum of $H_\infty(X')$ over all distributions X' such that $|X-X'|\leq\varepsilon$.

We now define average conditional min-entropy. Let X, Y be two random variables. The *average conditional min-entropy* of X given Y is

$$\tilde{H}_\infty(X|Y) = -\log\left(\mathbb{E}_{y\sim Y}\left[2^{-H_\infty(X|Y=y)}\right]\right).$$

We need the following standard lemmas regarding the conditional min-entropy (see, e.g., [DORS08])

Lemma 2.1. *Let X, Y, Z be random variables such that Y has support size at most 2^ℓ. Then, $\tilde{H}_\infty(X|(Y,Z)) \geq \tilde{H}_\infty(X|Z) - \ell$.*

Lemma 2.2. *For any two random variables X, Y and any $\varepsilon > 0$, it holds that*

$$\Pr_{y\sim Y}\left[H_\infty(X|Y=y) < \tilde{H}_\infty(X|Y) - \log\frac{1}{\varepsilon}\right] \leq \varepsilon.$$

2.2 Extractors

Definition 2.3 (extractor)**.** *A function* $\mathsf{Ext} : [N] \times [D] \to [M]$ *is a* (k, ε) strong extractor *if for every (n,k)-source X, and for Y that is uniform over $[D]$, $Y \circ \mathsf{Ext}(X,Y) \approx_\varepsilon Y \times U_m$.*

Theorem 2.4 (The GUV extractor, [GUV09])**.** *There exists a universal constant $c_{\mathsf{GUV}} > 0$ such that the following holds. For all positive integers n, k and $\varepsilon > 0$ there exists an efficiently-computable (k, ε) strong extractor $\mathsf{Ext} : [N] \times [D] \to [M]$ having seed length $d = c_{\mathsf{GUV}} \log\frac{n}{\varepsilon}$ and $m = \frac{k}{2}$ output bits.*

Definition 2.5 (two-source extractor)**.** *A function* $2\mathsf{Ext} : [N_1]\times[N_2] \to [M]$ *is an* $((n_1,k_1)\times(n_2,k_2) \to_\varepsilon m)$ two-source extractor *if for every two independent sources X_1 and X_2 where X_1 is an (n_1,k_1)-source and X_2 is an (n_2,k_2)-source, it holds that $2\mathsf{Ext}(X_1,X_2) \approx_\varepsilon U_m$.*

Theorem 2.6 (The Raz extractor, [Raz05, Theorem 4])**.** *There exists a universal constant $c_{\mathsf{RAZ}} > 0$ such that the following holds. For all positive integers n and $\varepsilon > 0$, set $d = c_{\mathsf{RAZ}} \log\frac{n}{\varepsilon}$. For all $k \geq c_{\mathsf{RAZ}} \cdot d$, there exists an efficiently-computable $((n,k)\times(d,0.6d) \to_\varepsilon k/2)$ two-source extractor $\mathsf{Raz} : [N]\times[D] \to [\sqrt{K}]$.*

3 An advice correlation breaker with a dense seed

In this section we construct the mergers needed for the construction. For that we use advice correlation breakers, first introduced in [Coh16].

Definition 3.1. *An $((n,k)\times(\ell,\alpha\ell) \to_\varepsilon m)$ t-NM advice correlation breaker is a function*

$$\mathsf{AdvCB} : \{0,1\}^n \times \{0,1\}^\ell \times \{0,1\}^a \to \{0,1\}^m$$

such that for every random variables $\left\{X^{(j)}\right\}_{0\leq j\leq t}$, where $X^{(j)}$ is distributed over $\{0,1\}^n$, and for every random variables $\left\{Y^{(j)}\right\}_{0\leq j\leq t}$ that are distributed over $\{0,1\}^\ell$, and for every $t+1$ strings $adv^{(0)},\ldots,adv^{(t)} \in \{0,1\}^a$, the following holds. Denote $X = X^{(0)}, Y = Y^{(0)}$. If

- *$\left\{X^{(j)}\right\}_{0\leq j\leq t}$ are independent of $\left\{Y^{(j)}\right\}_{0\leq j\leq t}$,*
- *$H_\infty(X) \geq k$,*

- $H_\infty(Y) \geq \alpha\ell$, *and,*
- $adv^{(0)} \notin \left\{adv^{(j)}\right\}_{j\in[t]}$,

then

$$\mathsf{AdvCB}\left(X, Y, adv^{(0)}\right) \circ \left\{\mathsf{AdvCB}\left(X^{(j)}, Y^{(j)}, adv^{(j)}\right)\right\}_{j\in[t]} \approx_\varepsilon U_m \times \left\{\mathsf{AdvCB}\left(X^{(j)}, Y^{(j)}, adv^{(j)}\right)\right\}_{j\in[t]}.$$

Notice that if

- we have a table with t rows $r_1, \ldots, r_t \in \{0,1\}^\ell$, one of which coming from an $(\ell, \alpha\ell)$ source, and,
- if X is sampled from an (n,k) distribution X,

then $\mathsf{AdvCB}(x, r_1, 1) \oplus \mathsf{AdvCB}(x, r_2, 2) \oplus \ldots \oplus \mathsf{AdvCB}(x, r_t, t)$ is close to uniform, since $\mathsf{AdvCB}(x, r_i, i)$ is close to uniform for the good row r_i, and independent of all other values $\mathsf{AdvCB}(x, r_j, j)$ for $j \neq i$. Thus, in this section we focus on obtaining an advice correlation breaker where X is a weak (n,k)-source and Y is an $(\ell, \alpha\ell)$-source.

Cohen [Coh16] proved:

Theorem 3.2. *There exists a constant $c' > 1$ such that the following holds. For all integers n, m, a, for any $\varepsilon > 0$, and for any constant integer $c \geq 1$, there exists an explicit $((n,k) \times (\ell,\ell) \to_\varepsilon m)$ 1-NM advice correlation breaker*

$$\mathsf{AdvCB} : \{0,1\}^n \times \{0,1\}^\ell \times \{0,1\}^a \to \{0,1\}^m$$

with

$$\begin{aligned} \ell &= c' \log(an) + (c')^{\sqrt{\log(\log^{(c)}(a) + \log(1/\varepsilon))}} \cdot \left(\log^{(c)}(a) + \log(1\varepsilon)\right) \\ k &= 4m + c'\ell. \end{aligned}$$

Plugging in $c = 4$, we have that $\ell = O(\log a + \log n) + (\log(1/\varepsilon))^{1+o(1)}$ and $k = 3m + O(\ell)$.

We need a similar result, but with the following differences:

- Y is not a uniform source but rather a dense weak source.
- In our case a is small and we do not care much about the dependence on a.
- We make the dependence on t explicit (Cohen uses the theorem for a general constant t, but does not specify the dependence on t).

Our construction roughly follows the base correlation breaker with advice of [Coh16] using the independence-persevering mergers of [CL16]. We do not try to reduce the advice length, because $A = 2^a$ is the same order as t, which we pay anyway. The fact that $A = t$ also allows a simple and explicit treatment of the t shadows. We prove:

Theorem 3.3. *There exist constants $c > 1$, $c' > 0$ such that for every $n, a, t, \zeta > 0$, $k \geq ct^3A^2 \log \frac{n}{\zeta}$ and $\alpha \geq 1 - \frac{c'}{t^2 A}$, there exists a $((n,k) \times (\ell, \alpha\ell) \to_\zeta m)$ t-NM advice correlation breaker*

$$\mathsf{AdvCB} : \{0,1\}^n \times \{0,1\}^\ell \times \{0,1\}^a \to \{0,1\}^m$$

with seed length $\ell = ct^2 A \log \frac{nA}{\zeta}$ and output length $m \leq \frac{k}{ct^2 A} - cA \log \frac{k}{t\zeta}$.

Plugging in $A = t$, we have the following corollary:

Corollary 3.4. *There exist constants $c_{\mathsf{adv}} > 1$, $c_{\mathsf{gap}} > 0$ such that for every $n, t \leq n, \zeta > 0$, $k \geq c_{\mathsf{adv}} t^5 \log \frac{n}{\zeta}$ and $\alpha \geq 1 - \frac{c_{\mathsf{gap}}}{t^3}$, there exists a $((n,k) \times (\ell, \alpha\ell) \to_\zeta m)$ t-NM advice correlation breaker*

$$\mathsf{AdvCB} : \{0,1\}^n \times \{0,1\}^\ell \times \{0,1\}^{\log t} \to \{0,1\}^m$$

with seed length $\ell = c_{\mathsf{adv}} t^3 \log \frac{n}{\zeta}$ and output length $m \leq \frac{k}{c_{\mathsf{adv}} t^3} - c_{\mathsf{adv}} t \log \frac{n}{\zeta}$.

For the construction we need independence-preserving mergers and the hierarchy of independence, which we review in the next two subsections.

3.1 Independence-preserving mergers

Independence-preserving mergers were first introduced in [CS15]. We define:

Definition 3.5. *An $((L \times m) \times (d, k_2) \to_\varepsilon m')$ t-NM independence-preserving merger*

$$\mathsf{IPM} : (\{0,1\}^m)^L \times \{0,1\}^d \to \{0,1\}^{m'}$$

is a function such that for every random variables $\left\{X^{(j)}\right\}_{0 \leq j \leq t}$, where every $X^{(j)}$ is a boolean $L \times m$ matrix, and for every $\left\{Y^{(j)}\right\}_{0 \leq j \leq t}$ that are distributed over $\{0,1\}^d$ the following holds. Let $X = X^{(0)}$, $Y = Y^{(0)}$ and $X_i^{(j)}$ be the i-th row of $X^{(j)}$ (hence X_i is the i-th row of X). If

- *$\left\{X^{(j)}\right\}_{0 \leq j \leq t}$ are independent of $\left\{Y^{(j)}\right\}_{0 \leq j \leq t}$,*
- *for every $0 \leq i \leq L-1$, X_i is uniform,*
- *there exists $0 \leq i \leq L-1$ such that $\left(X_i, \left\{X_i^{(j)}\right\}_{j \in [t]}\right) = \left(U_m, \left\{X_i^{(j)}\right\}_{j \in [t]}\right)$, and,*
- *$H_\infty(Y) \geq k_2$,*

then

$$\mathsf{IPM}(X, Y) \circ \left\{\mathsf{IPM}\left(X^{(j)}, Y^{(j)}\right)\right\}_{j \in [t]} \approx_\varepsilon U_{m'} \times \left\{\mathsf{IPM}\left(X^{(j)}, Y^{(j)}\right)\right\}_{j \in [t]}.$$

Chattopadhyay and Li [CL16] proved:

Theorem 3.6. *There exists a constant $c_{\mathsf{IPM}} > 0$ such that for all integers $L, m, d, k_2, t > 0$ and all $\varepsilon > 0$, if*

$$m \geq d \geq k_2 > c_{\mathsf{IPM}}(t+1)\ell \log \frac{m}{\varepsilon},$$

then there exists an explicit $((L \times m) \times (d, k_2) \to_{c_{\mathsf{IPM}} L\varepsilon} m')$ t-NM independence-preserving merger function

$$\mathsf{IPM} : (\{0,1\}^m)^L \times \{0,1\}^d \to \{0,1\}^{m'},$$

with $m' = \frac{0.9}{t}(m - c_{\mathsf{IPM}}(t+1)\ell \log \frac{m}{\varepsilon})$.

3.2 A hierarchy of independence

We state a variant of the *hierarchy of independence* [DP07, DW09, Li13, Li15b, Coh15a, CS15]. We first give a short informal explanation for those unfamiliar with it. A hierarchy of independence usually deals with two issues:

1. Suppose we start with two independent sources X and Y, and do some computation $f(X,Y)$ that involves both sources. Are the two sources independent given the computation result?
2. We do the computation on the two sources X and Y given to us. Suppose the adversary is given access to the same computation done on a shadow input, i.e., on X' and Y' where (X',Y') is correlated with (X,Y) in some restricted way. Can we say that the computation we do is *independent* of the computation on the shadow?

An example for a computation that correlates X and Y is the inner product function. One can check that $(X|\mathsf{IP}(X,Y))$ and $(Y|\mathsf{IP}(X,Y))$ are *not* independent. However, there are computations that preserve independence. One simple example is if $f(X,Y)$ is *local*. If $f(X,Y)$ depends on X alone (or on Y alone), then $(X|f(X,Y))$ is independent of $(Y|f(X,Y))$. As a result, if we have a computation $W_1 = f(X)$, $W_2 = f(Y,W_1)$, $W_3 = f(X,W_1,W_2)$, etc., then $(X|W_1,\ldots,W_t)$ is independent of $(Y|W_1,\ldots,W_t)$. This addresses the first issue. This kind of computation is called in [Coh15a] *an (X,Y)-history*.

To address the second issue we make the following assumptions. We assume X and all its shadows together are independent of Y and all its shadows together. We also assume some part X_1 of X is known to be uniform. We then notice that if Ext is a strong extractor, then $\mathsf{Ext}(Y,X_1)$ is uniform even when we fix X_1 (with high probability over X_1). In fact, it is an easy exercise that $\mathsf{Ext}(Y,X_1)$ is uniform even given the shadows of X. This phenomenon continues throughout the hierarchy, a value $W_i = \mathsf{Ext}(Source, W_{i-1})$ in the i-th level that depends on one source, is uniform even given the history $W_1,\ldots,W_{i-1}$ and the other source, and furthermore, this is true even when all these values are given for the shadows too.

In our case we assume we know a part X_1 of X that has high min-entropy, rather than the usual requirement that X_1 is uniform. Formally,

Parameters: We are given:

- n, k_1 and a sample x from an (n,k_1) source X,
- ℓ, k_2 and a sample y from an (ℓ,k_2) source Y,
- m – the desired number of bits in a row,
- A – the number of levels in the hierarchy, and,
- ε – the error parameter.

Ingredients: For the protocol we use the following ingredients:

- An extractor
$$\mathsf{Raz} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m ,$$
which is a $(2m,\varepsilon)$ strong extractor with 0.6-dense seed (see Theorem 2.6), and,
$$\mathsf{Ext}_x : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m ,$$

which is a $(2m, \varepsilon)$ strong extractor (see Theorem 2.4), for $d = \max\{c_{\mathsf{RAZ}}, c_{\mathsf{GUV}}\} \cdot \log \frac{n}{\varepsilon}$. In fact, the seed length for Raz is $c_{\mathsf{RAZ}} \log \frac{n}{\varepsilon}$ and for Ext_x is $c_{\mathsf{GUV}} \log \frac{n}{\varepsilon}$, but we combine them to save a parameter.

- An extractor
$$\mathsf{Ext}_y : \{0,1\}^{\ell} \times \{0,1\}^{d'} \to \{0,1\}^{d},$$
which is a strong $(2d, \varepsilon)$ extractor and $d' = c_{\mathsf{GUV}} \log \frac{\ell}{\varepsilon}$.

Protocol: The protocol is as follows. Given $x \in \{0,1\}^n$ and $y \in \{0,1\}^{\ell}$, let y_1 be a prefix of y of length d. Compute the following (X, Y)-history:

$$\begin{aligned} HY_0 &= y_1 \\ HX_0 &= \mathsf{Raz}(x, HY_0) \\ HY_1 &= \mathsf{Ext}_y(y, HX_0|_{[d']}) \\ HX_1 &= \mathsf{Ext}_x(x, HY_1) \\ &\vdots \\ HY_{A-1} &= \mathsf{Ext}_y(y, HX_{A-2}|_{[d']}) \\ HX_{A-1} &= \mathsf{Ext}_x(x, HY_{A-1}), \end{aligned}$$

and output $\mathcal{H}(x, y) = (HX_0, \ldots, HX_{A-1})$.

We claim:

Lemma 3.7 ([Li13], revised)**.** *Fix n, k_1, ℓ, k_2, t, A and ε. Set $d = \max\{c_{\mathsf{RAZ}}, c_{\mathsf{GUV}}\} \cdot \log \frac{n}{\varepsilon}$. Assume $m \geq d$. There exists a constant $c_{\mathsf{H}} > 1$ for which the following holds:*

- *Let X be an (n, k_1)-source and Y be an (ℓ, k_2)-source independent of X, where $k_1 \geq c_{\mathsf{H}} tAm + \log \frac{1}{\varepsilon}$, $k_2 \geq c_{\mathsf{H}} tAd + \log \frac{1}{\varepsilon}$ and $\ell \leq k_2 + 0.4d$.*
- *Let $X^{(1)}, \ldots, X^{(t)}$ be t random variables over n bits arbitrarily correlated with X. Let $Y^{(1)}, \ldots, Y^{(t)}$ be t random variables over ℓ bits arbitrarily correlated with Y. We denote $X^{(0)} = X$ and $Y^{(0)} = Y$. We assume the random variables $\left\{X^{(j)}\right\}_{0 \leq j \leq t}$ together are independent of $\left\{Y^{(j)}\right\}_{0 \leq j \leq t}$.*

Let $\mathcal{H}\left(X^{(j)}, Y^{(j)}\right) = \left(HX_0^{(j)}, \ldots, HX_{A-1}^{(j)}\right)$ for every $0 \leq j \leq t$. Then, we claim that for every $0 \leq i \leq A-1$, HX_i is uniform even conditioned on the values of the hierarchy (including those of the t shadows) in the first $i-1$ levels. Specifically,

$$\left(HX_i, \left\{Y^{(j)}, HX_b^{(j)}\right\}_{\substack{0 \leq j \leq t \\ 0 \leq b \leq i-1}}\right) \approx_{c_{\mathsf{H}} i\varepsilon} \left(U_m, \left\{Y^{(j)}, HX_b^{(j)}\right\}_{\substack{0 \leq j \leq t \\ 0 \leq b \leq i-1}}\right).$$

The proof is a standard application of (X, Y)-history reasoning (see [Coh15a]). We do not give the full proof, but we do comment on a few things that need to be checked:

- In the application of the Raz extractor, y_1 is dense enough.
- In all the extractor applications, the seed length is long enough.

- Throughout the execution of the protocol, X and Y retain (with high probability) enough min-entropy (i.e., at least $2m$ min-entropy).

For the first item, notice that by our choice of parameters, $0.4d \geq \ell - k_2$. Hence, $H_\infty(Y_1) \geq k_2 - H_\infty(\overline{Y_1}) \geq k_2 - (\ell - d) \geq 0.6d$, so $\frac{H_\infty(Y_1)}{d} \geq 0.6$. The second item follows because d is large enough (and $m \geq d$). For the third condition we have that $k_1 = \Omega(tAm)$ and $k_2 = \Omega(tAd)$.

3.3 The construction and proof

Parameters: We are given $n, k_1 = k, a, t$ and $\zeta > 0$. Set the following parameters:

- $\varepsilon = \frac{\zeta}{2c_{\mathsf{IPM}}A}$.
- r is such that $k_1 = c_{\mathsf{H}}tAr + \log\frac{1}{\varepsilon}$.
- $m = \frac{0.9}{t}(r - c_{\mathsf{IPM}}(t+1)A\log\frac{r}{c_{\mathsf{H}}\varepsilon})$.

We apply the hierarchy of independence on X and a prefix of Y of length ℓ_{pref}. Then, we apply an independence-preserving merger on the table and the whole length-ℓ string y. For that we fix:

- $d = \max\{c_{\mathsf{RAZ}}, c_{\mathsf{GUV}}\} \cdot \log\frac{n}{\varepsilon}$.
- $\ell_{pref} = c_{\mathsf{H}}tAd + \log\frac{1}{\varepsilon} + 0.4d$.
- $\ell = (t+1)(\ell_{pref} + c_{\mathsf{IPM}}A\log\frac{r}{c_{\mathsf{H}}\varepsilon}) + \log\frac{1}{\varepsilon}$. We require k_1 to be large enough so that $r \geq \ell$.

Ingredients: For the construction we use the following ingredients:

- $\mathcal{H} : \{0,1\}^n \times \{0,1\}^{\ell_{pref}} \to (\{0,1\}^r)^A$ is the hierarchy of independence with A levels and error ε.
- $\mathsf{IPM} : (\{0,1\}^r)^A \times \{0,1\}^\ell \to \{0,1\}^m$ is an $((A \times r) \times (\ell, \alpha\ell) \to_\varepsilon m)$ t-NM independence-preserving merger, for a large enough α that will be determined soon.

Construction: The construction is as follows. Given $x \in \{0,1\}^n$, $y \in \{0,1\}^\ell$ and $adv \in [A]$, let y_{pref} be a prefix of y of length ℓ_{pref} and compute the following:

1. Let $\mathcal{H}(x, y_{pref}) = (HX_0, \ldots, HX_{A-1})$.
2. Construct the table
$$\mathcal{T} = (HX_{adv \bmod A}, HX_{(adv+1) \bmod A}, \ldots, HX_{(adv+A-1) \bmod A}).$$
3. Output $\mathsf{AdvCB}(x, y, adv) = \mathsf{IPM}(\mathcal{T}, y)$.

We now turn to the analysis. The analysis is standard except for the way we treat the advice (and therefore the t shadows). Usually, one works on the advice string bit by bit, and argues that in each bit, the good source becomes independent with all the shadow sources that have in their advice string a different bit (though one has to be careful with such claims). Here, we work on the advice string as a whole, and we use it as a cyclic shift on the the t levels in our hierarchy of independence. This means that the number of levels in our hierarchy of independence is A, rather than the usual $a = \log A$. However, we are willing to pay this price because $A = t$ and we pay it anyway. We claim:

Lemma 3.8. *Suppose $\alpha \geq 1 - \frac{c'}{t^2 A}$ for some constant $c' > 0$. Then,*

$$\mathsf{AdvCB} : \{0,1\}^n \times \{0,1\}^\ell \times \{0,1\}^a \to \{0,1\}^m$$

is an $((n,k) \times (\ell, \alpha\ell) \to_\zeta m)$ t-NM advice correlation breaker.

Proof: Let X be an (n, k_1)-source and let Y be an ℓ-bit random variable such that $H_\infty(Y) \geq \alpha\ell \geq \ell - \frac{c'\ell}{t^2 A} \geq \ell - 0.4d$, for $c' = \frac{0.4dt^2 A}{\ell}$. Notice that since $\ell \geq c_{\mathsf{H}} dt^2 A$, c' is indeed a constant.

Now, fix shadows $\left\{X^{(j)}, Y^{(j)}\right\}_{j\in[t]}$. Let $adv^{(j)}$ be the advice associated with shadow j, denote $adv = adv^{(0)}$, and assume $adv \notin \left\{adv^j\right\}_{j\in[t]}$. Denote $\mathcal{T}^{(j)} = \mathcal{H}(X^{(j)}, Y^{(j)})$ and $\mathcal{T} = \mathcal{T}^{(0)}$. Recall that for every $0 \leq j \leq t$, $\mathcal{T}^{(j)}$ is an $A \times r$ table.

By assumption, we have that $H_\infty(X) \geq c_{\mathsf{H}} tAr + \log\frac{1}{\varepsilon}$. Also, we have $H_\infty(Y) \geq \ell - 0.4d$, so $H_\infty(Y_{pref}) \geq \ell_{pref} - 0.4d$. By Lemma 3.7,

- (All the rows in $\mathcal{T}$ are close to uniform) For every $0 \leq i \leq A-1$,

$$\left(\mathcal{T}_i, \left\{Y^{(j)}_{pref}\right\}_{0\leq j\leq t}\right) \approx_{c_{\mathsf{H}}\varepsilon} \left(U_r, \left\{Y^{(j)}_{pref}\right\}_{0\leq j\leq t}\right).$$

- (There exists a row that is close to uniform even conditioned on the shadow) Since each player shifts the rows by $adv^{(j)} \in \{0,\ldots,A-1\}$, and the advice $adv^{(0)}$ given to the good player is different from the advices $adv^{(j)}$ given to all the shadows, there exists a row b, such that the good player outputs on that row $\mathcal{T}_b = \mathcal{H}(X, Y_{pref})_{A-1}$ and the j-th shadow player outputs on that row $\mathcal{T}^{(j)}_b = \mathcal{H}\left(X^j, Y^{(j)}_{pref}\right)_i$ for some $i < A-1$. Hence, again by Lemma 3.7,

$$\left(\mathcal{T}_b, Y_{pref}, \left\{Y^{(j)}_{pref}, \mathcal{T}^{(j)}_b\right\}_{j\in[t]}\right) \approx_{c_{\mathsf{H}}\varepsilon} \left(U_r, Y_{pref}, \left\{Y^{(j)}_{pref}, \mathcal{T}^{j}_b\right\}_{j\in[t]}\right).$$

Now, conditioned on the values $\left\{Y^{(j)}_{pref}\right\}_{0\leq j\leq t}$,

- (All the rows in $\mathcal{T}$ are close to uniform) For every $0 \leq i \leq A-1$,

$$\mathcal{T}_i \approx_{c_{\mathsf{H}}\varepsilon} U_r.$$

- (There exists a row that is close to uniform even conditioned on the shadow) There exists a row b such that

$$\left(\mathcal{T}_b, \left\{\mathcal{T}^{(j)}_b\right\}_{j\in[t]}\right) \approx_{c_{\mathsf{H}}\varepsilon} \left(U_r, \left\{\mathcal{T}^{(j)}_b\right\}_{j\in[t]}\right).$$

- $\left\{X^{(j)}\right\}_{0\leq j\leq t}$ and $\left\{Y^{(j)}\right\}_{0\leq j\leq t}$ are independent. Also, given the conditioning, $\left\{\mathcal{T}^{(j)}\right\}_{0\leq j\leq t}$ is a function of $\left\{X^{(j)}\right\}_{0\leq j\leq t}$ alone, so $\left\{\mathcal{T}^{(j)}\right\}_{0\leq j\leq t}$ is independent of $\left\{Y^{(j)}\right\}_{0\leq j\leq t}$.

- Finally, by Lemma 2.1,

$$\tilde{H}_\infty\left(Y \middle| \left\{Y^{(j)}_{pref}\right\}\right) \geq H_\infty(Y) - (t+1)\ell_{pref} \geq c_{\mathsf{IPM}}(t+1)A\log\frac{r}{c_{\mathsf{H}}\varepsilon} + \log\frac{1}{\varepsilon}.$$

By Lemma 2.2, except for error ε over the fixing of $Y^{(j)}_{pref}$, $H_\infty(Y) \geq c_{\mathsf{IPM}}(t+1)A \log \frac{r}{c_{\mathsf{H}}\varepsilon}$.

Let $W^{(j)} = \mathsf{AdvCB}\left(X^{(j)}, Y^{(j)}, adv^{(j)}\right) = \mathsf{IPM}\left(\mathcal{T}^{(j)}, Y^{(j)}\right)$. By Theorem 3.6,

$$\left(W, \left\{W^{(j)}\right\}_{j\in[t]}\right) \approx_{c_{\mathsf{IPM}}A\varepsilon+2\varepsilon} \left(U_m, \left\{W^{(j)}\right\}_{j\in[t]}\right).$$

Note that $c_{\mathsf{IPM}}A\varepsilon + 2\varepsilon \leq \zeta$, so we are done. ∎

Keeping track of the parameters, we can verify that Theorem 3.3 indeed follows.

4 A somewhere-random condenser with a dense output

Definition 4.1 (somewhere-random source)**.** *A source $X = X_1 \circ \ldots \circ X_A$ is a $(k, (\alpha, \beta))$* somewhere-random (s.r.) source *if there is a random variable $I \in \{0, \ldots, A\}$ such that for every $i \in [A]$, $H^\alpha_\infty(X_i|I = i) \geq k$ and $\Pr[I = 0] \leq \beta$. The variable I is called the* indicator *of source. If $\alpha = \beta = 0$ we say X is a k* s.r. source.

Claim 4.2. *Let X be a (k, α, β) s.r. source. Then, X is $(\alpha + \beta)$-close to a k s.r. source.*

Definition 4.3 (s.r. condenser)**.** *A function $C : [N] \to [M]^A$ is a $(k \to_\varepsilon k')$ s.r. (seedless) condenser if for every (n, k)-source X it holds that $C(X) = C(X, 1) \circ \ldots \circ C(X, A)$ is ε-close to a k'- s.r. source. We say that C is a rate-$(\delta \to_\varepsilon \mu)$ s.r. condenser if it is a $(\delta n \to_\varepsilon \mu m)$ s.r. condenser.*

4.1 A basic s.r. condenser from the Incidence theorem

In [Zuc06], Zuckerman showed:

Theorem 4.4. *For every constant $0 < c < 1$ there exists a constant $\alpha = \alpha(c)$[3] such that for every $\delta \leq c$ and N for which $N^\delta = \omega(1)$, there exists a rate-$(\delta \to_\varepsilon (1 + \alpha/2)\delta)$ s.r. condenser $C : [q^3] \to [q^2]^2$, where $\varepsilon = N^{-\alpha\delta/60}$.*

Li [Li11] gave a high density variant of this s.r. condenser. The construction uses the following incidence theorem of [Vin11] that has the advantage that it also works when the sets are dense.

Theorem 4.5. *Let $\mathbb{F}_q$ be a field. Let P be a set of points of $\mathbb{F}_q^2$ and L a set of lines. Let $I(P, L) = \{(p, \ell) \mid p \in P, \ell \in L, p \in \ell\}$. Then,*

$$|I(P, L)| \leq \frac{|P| \cdot |L|}{q} + \sqrt{q \cdot |P| \cdot |L|}.$$

We also use we use the following lemma to convert the statistical problem of constructing a s.r source into a counting problem.

Lemma 4.6 ([Zuc06, Lemma 8.6])**.** *If (X, Y) is not ε-close to a k-s.r. source, then there exist sets $S \subseteq \mathrm{supp}(X), T \subseteq \mathrm{supp}(Y), |S|, |T| < 2^{k+1}/\varepsilon$, such that $\Pr[X \in S \wedge Y \in T] > \varepsilon/2$.*

[3] The constant α is the constant guaranteed by the Incidence Theorem (see, e.g., [BKT04, BGK06]).

Li used the above two lemmas to give the high density version of the s.r. condensers, and we repeat the argument using the notation introduced above.

Theorem 4.7. *Let $\mathbb{F}_q$ be a field, and let n denote $\log q$. For every $g = g(n)$ and $0.6g + 2 \leq g' = g'(n) < \frac{n}{2}$ there exists a $((3n, 3n - g) \to_\varepsilon (2n, 2n - g'))$ s.r. condenser $C : \mathbb{F}_q^3 \to (\mathbb{F}_q^2)^2$ with $\varepsilon = 2^{-g/15}$.*

Proof: We use Zuckerman's condenser $C : \mathbb{F}^3 \to (\mathbb{F}^2)^2$, defined by

$$C(a, b, c) = ((b, ab + c), (a, c)).$$

The condenser connects an incidence of point-line to the point $(b, \ell(b))$ and line $\ell(x) = ax + c$ that defines it. The intuition is then simply that, if C is applied to a set of many incidences, then, by the Incidence Theorem this set must intersect many lines or many points.

Formally, suppose $X \subseteq \mathbb{F}^3$ has entropy gap g, i.e., $|X| \geq \frac{q^3}{G}$. Assume that $(C(X)_1, C(X)_2)$ is not a ε-close to a $(2n, 2n - g')$ s.r source. By Lemma 4.6 there exist sets P, L of size less than $\frac{2}{\varepsilon}\frac{q^2}{G'}$ of points and lines respectively, that contain at least $\varepsilon/2$ of the incidences, i.e.,

$$|I(P, L)| \geq |X| \cdot \frac{\varepsilon}{2} = \frac{\varepsilon}{2}\frac{q^3}{G}.$$

By Theorem 4.5,

$$I(P, L) \quad \leq \quad \left(\frac{2}{\varepsilon}\right)^2 \frac{q^3}{G'^2} + \sqrt{q}\frac{2}{\varepsilon}\frac{q^2}{G'} \quad < \quad 2\left(\frac{2}{\varepsilon}\right)^2 \frac{q^3}{G'^2},$$

because $g' < \frac{n}{2}$. Combining the two bounds, we get

$$G' < \frac{4}{\varepsilon^{1.5}}\sqrt{G},$$

a contradiction to $g' \geq 0.6g + 2$. ∎

Corollary 4.8. *Let $\mathbb{F}_q$ be a field, and let n denote $\log q$. For every $\delta = \delta(n) \geq \frac{3}{4}$ there exists a $((3n, \delta) \to_\varepsilon (2n, 1 - \delta'))$ s.r. condenser $C : \mathbb{F}_q^3 \to (\mathbb{F}_q^2)^2$ with $\varepsilon = 2^{-\Omega((1-\delta)n)}$ and $1 - \delta' = 0.9(1 - \delta)$.*

4.2 Composing s.r. condensers

We the following result about compositions of s.r. condensers.

Lemma 4.9 ([BKS+10])**.** *Let $C_1 : [N_1] \to [N_2]^{\ell_1}$ be a rate $(\delta_1 \to_{\varepsilon_1} \delta_2)$ s.r. condenser. Let $C_2 : [N_2] \to [M]^{\ell_2}$ be a rate $(\delta_2 \to_{\varepsilon_2} \delta_3)$ s.r. condenser. We define $C_2 \circ C_1 : [N] \to [M]^{\ell_1 \cdot \ell_2}$ as follows: Identify an index $i \in [\ell_1 \cdot \ell_2]$ as a pair $(i_1, i_2) \in [\ell_1] \times [\ell_2]$ and let $C_2 \circ C_1(x)_{(i_1,i_2)} = C_2(C_1(x)_{i_1})_{i_2}$. Then, $C_2 \circ C_1$ is a rate-$(\delta_1 \to_{\varepsilon_1+\varepsilon_2} \delta_3)$ s.r. condenser.*

Composing the s.r. condenser from Theorem 4.4 with itself s times with repeated application of Lemma 4.9, we get:

Theorem 4.10. *Let α be the constant from the Incidence Theorem. Fix $\delta < 3/4$ and N for which $N^\delta = \omega(1)$. Fix any constant $c \geq c_1 = \frac{\log_{1+\alpha/2}\frac{3}{4\delta}}{\log(1/\delta)}$. Let $s = c \log 1/\delta$. There exists a rate-$(\delta \to 3/4, \varepsilon)$ s.r. condenser $C : [N] \to [M]^D$ with:*

- $D = 2^s = (1/\delta)^c$,
- $m = (\frac{2}{3})^s n = (1/\delta)^{-c \log \frac{3}{2}} n$, *and,*
- $\varepsilon = \sum_{i=1}^{s} 2^{-\frac{\alpha}{60}(\frac{2}{3})^i n (1+\frac{\alpha}{2})^i \delta} = 2^{-\Omega(m)}$.

Theorem 4.4 is sufficient when we want to achieve a constant rate s.r source. Composing the s.r. condenser of Theorem 4.7 gives:

Theorem 4.11. *There exists a constant $c_2 > 1$ such that for every $c \geq c_2$ and every $T > 0$ the there exists a rate-$(3/4 \to 1 - \frac{1}{T}, \varepsilon)$ s.r. condenser $C : [N] \to [M]^D$ with:*

- $D = T^c$,
- $m = T^{-c \log(3/2)} n$, *and,*
- $\varepsilon = \sum_{i=1}^{c \log T} 2^{-\Omega(0.9^i \cdot \frac{3}{4} \cdot (\frac{2}{3})^i n)} = 2^{-\Omega(m/T)}$.

Proof: We start with a rate-$3/4$ source, and repeatedly (somewhere) condense it. Let μ_i denote the entropy gap in the i-th step (that is, $\mu_0 = 1/4$). Corollary 4.8 guarantees us that $\mu_i \leq 0.9\mu_{i-1}$. Let ℓ denote the number of iterations. We require that in the last step $\mu_\ell = 0.9^\ell \cdot \frac{3}{4} \leq 1/T$, hence, $\ell = C \log T$ for some constant $C > 1$ suffices. The length of the source in the i-th step is $m_i = (\frac{2}{3})^i n$ and so the second item holds. The error of the condenser is the accumulation of the errors in each step. The error in the i-th step is $2^{-\mu_i m_i} = 2^{0.9^i \cdot \frac{3}{4} \cdot (\frac{2}{3})^i n}$. ■

5 A two-source extractor with a small error

5.1 The construction

Given parameters: n, k_1, k_2 and ε.

Fixed constants: $c = \max\{c_1, c_2, 2\}$, where c_1, c_2 are the constants from Theorems 4.10 and 4.11 respectively. $c_{\mathsf{out}} = \log(3/2)c$. $c_{\mathsf{gap}}, c_{\mathsf{adv}}$ are the constants from Corollary 3.4.

Input: A sample x_1 from an (n, k_1)-source X_1 and a sample x_2 from an (n, k_2)-source X_2.

The ingredients:

- $\mathsf{Cond}_1 : [N] \to [M_1]^{(k_1/n)^c}$, a rate-$(k_1/n \to_\varepsilon \frac{3}{4})$ s.r. condenser from Theorem 4.10, where $m_1 = (\frac{n}{k_1})^{-c_{\mathsf{out}}} \cdot n$.
- $\mathsf{Cond}_2 : [M_1] \to [L]^{T^c}$, a rate-$(3/4 \to_\varepsilon 1 - 1/T)$ s.r. condenser from Theorem 4.11, where $l = T^{-c_{\mathsf{out}}} \cdot m_1$, $T = \frac{t^3}{c_{\mathsf{gap}}}$, and $t = (\frac{c_{\mathsf{gap}} n}{k_1})^{\frac{c}{3c-1}}$. Note that $t = (k_1/n)^c \cdot T^c$, the multiplication of the number of rows of Cond_1 and Cond_2.
- $\mathsf{AdvCB} : [N] \times [L] \times [t] \to [M]$, a $((n, k_2) \times (l, \alpha l))$ t-NM advice correlation breaker for dense seeds, where $m = \frac{k_2}{c_{\mathsf{adv}} t^3} - t \log \frac{t}{\varepsilon}$ and $\alpha = 1 - 1/T$.

The construction: We construct $\mathsf{2Ext} : [N] \times [N] \to [M]$ defined by:

$$\mathsf{2Ext}(x_1, x_2) = \bigoplus_{i=1}^{t} \mathsf{AdvCB}(x_2, (\mathsf{Cond}_2 \circ \mathsf{Cond}_1(x_1))_i, i).$$

5.2 The analysis

Theorem 5.1. *There exists a constant $0 < c_{\mathsf{Ext}} < 1$ such that the following holds. For every large enough n, $k_1 \geq n^{1-\frac{1}{4}c_{\mathsf{Ext}}}$ and $k_2 \geq n^{c_{\mathsf{Ext}}}$ there exists an explicit $((n,k_1)\times(n,k_2) \to_\varepsilon m)$ two-source extractor, where $\varepsilon = 2^{-k_2^{\Omega(1)}}$ and $m = k_2^{\Omega(1)}$.*

Proof: Fix such sources X_1 and X_2. Consider the $t \times M$ table defined by $\mathsf{Cond}_1 \circ \mathsf{Cond}_2(X_1)$. By our choice of Cond_1 and Cond_2 we have that it is 2ε close to a $(1-1/T)l$ s.r. source.

We now need to check that the requirements for the correlation breaker hold.

- We first need that $l \geq c_{\mathsf{adv}} t^3 \log\frac{n}{\varepsilon}$. Assume that $\frac{1}{n} \geq \varepsilon \geq 2^{-\frac{1}{2}c_{\mathsf{gap}}^{1/3} n^{\frac{1}{12c_{\mathsf{out}}}}}$ and $k_1 \geq n^{1-\frac{1}{4c_{\mathsf{out}}}}$. Then, it holds that $\log\frac{n}{\varepsilon} \leq t$, and one can check that indeed

$$l \;=\; \left(\frac{Tn}{k_1}\right)^{-c_{\mathsf{out}}} \cdot n \;\geq\; c_{\mathsf{adv}} t^4 \;\geq\; c_{\mathsf{adv}} t^3 \log\frac{n}{\varepsilon}.$$

- We also need that $k_2 \geq c_{\mathsf{adv}} t^5 \log\frac{n}{\varepsilon}$, or, alternatively, $k_2 \geq c_{\mathsf{adv}} t^6$. This is satisfied by taking $k_2 \geq c_{\mathsf{adv}} c_{\mathsf{gap}}^3 n^{\frac{3}{4c_{\mathsf{out}}}}$. For a large enough n, the requirement $k_2 \geq n^{\frac{1}{c_{\mathsf{out}}}}$ suffices, and we can now see that we can set $c_{\mathsf{Ext}} = \frac{1}{c_{\mathsf{out}}}$.

As both conditions hold, we can apply the correlation breaker. By Corollary 3.4, $\mathsf{Cond}_1 \circ \mathsf{Cond}_2(X_1)$ is 2ε-close to a $(1-1/T)l$-s.r. source Λ. The fact that Λ is a s.r. source guarantees that it is a convex combination of sources, such that in each source, there exists a row with $(1-1/T)l$ min-entropy. It suffices to analyze each such source separately. Henceforth, we shall assume the i-th row, Λ_i, has $(1-1/T)l$ min-entropy. Using also the t-wise independence property of the correlation breaker,

$$\left(\mathsf{AdvCB}(X_2,\Lambda_i,i), \{\mathsf{AdvCB}(X_2,\Lambda_j,j)\}_{j\neq i}\right) \approx_\varepsilon \left(U_m, \{\mathsf{AdvCB}(X_2,\Lambda_j,j)\}_{j\neq i}\right),$$

so overall the XOR of these t variables is 3ε-close to uniform. The value of m is set by Corollary 3.4, noticing that for a large enough n, $\frac{k_2}{c_{\mathsf{adv}}t^3} \geq n^{\frac{2}{3}c_{\mathsf{Ext}}}$ and $c_{\mathsf{adv}} t \log\frac{n}{\varepsilon} \leq c_{\mathsf{adv}} t^2 \leq n^{\frac{1}{3}c_{\mathsf{Ext}}}$. ■

References

[Abb72] HL Abbott. Lower bounds for some ramsey numbers. *Discrete Mathematics*, 2(4):289–293, 1972.

[Alo98] Noga Alon. The shannon capacity of a union. *Combinatorica*, 18(3):301–310, 1998.

[BADTS16] Avraham Ben-Aroya, Dean Doron, and Amnon Ta-Shma. Explicit two-source extractors for near-logarithmic min-entropy. *ECCC*, 2016.

[Bar06] Boaz Barak. A simple explicit construction of an $n^{\tilde{o}(\log n)}$-ramsey graph. *arXiv preprint math/0601651*, 2006.

[BGK06] Jean Bourgain, AA Glibichuk, and Sergei V Konyagin. Estimates for the number of sums and products and for exponential sums in fields of prime order. *Journal of the London Mathematical Society*, 73(2):380–398, 2006.

[BKS+10] Boaz Barak, Guy Kindler, Ronen Shaltiel, Benny Sudakov, and Avi Wigderson. Simulating independence: New constructions of condensers, ramsey graphs, dispersers, and extractors. *Journal of the ACM (JACM)*, 57(4):20, 2010.

[BKT04] Jean Bourgain, Nets Katz, and Terence Tao. A sum-product estimate in finite fields, and applications. *Geometric & Functional Analysis GAFA*, 14(1):27–57, 2004.

[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.

[BRSW12] Boaz Barak, Anup Rao, Ronen Shaltiel, and Avi Wigderson. 2-source dispersers for $n^{o(1)}$ entropy, and ramsey graphs beating the frankl-wilson construction. *Annals of Mathematics*, 176(3):1483–1544, 2012.

[CG88] Benny Chor and Oded Goldreich. Unbiased bits from sources of weak randomness and probabilistic communication complexity. *SIAM Journal on Computing*, 17(2):230–261, 1988.

[Chu81] Fan RK Chung. A note on constructive methods for ramsey numbers. *Journal of Graph Theory*, 5(1):109–113, 1981.

[CL16] Eshan Chattopadhyay and Xin Li. Explicit non-malleable extractors, multi-source extractors and almost optimal privacy amplification protocols. *ECCC*, 2016.

[Coh15a] Gil Cohen. Local correlation breakers and applications to three-source extractors and mergers. *ECCC*, 2015.

[Coh15b] Gil Cohen. Two-source dispersers for polylogarithmic entropy and improved ramsey graphs. *arXiv preprint arXiv:1506.04428*, 2015.

[Coh16] Gil Cohen. Making the most of advice: New correlation breakers and their applications. *ECCC*, 2016.

[CS15] Gil Cohen and Leonard Schulman. Extractors for near logarithmic min-entropy. *ECCC*, 2015.

[CZ15] Eshan Chattopadhyay and David Zuckerman. Explicit two-source extractors and resilient functions. In *Electronic Colloquium on Computational Complexity (ECCC)*, volume 22, page 119, 2015.

[DORS08] Yevgeniy Dodis, Rafail Ostrovsky, Leonid Reyzin, and Adam Smith. Fuzzy extractors: How to generate strong keys from biometrics and other noisy data. *SIAM journal on computing*, 38(1):97–139, 2008.

[DP07] Stefan Dziembowski and Krzysztof Pietrzak. Intrusion-resilient secret sharing. In *Foundations of Computer Science, 2007. FOCS'07. 48th Annual IEEE Symposium on*, pages 227–237. IEEE, 2007.

[DW09] Yevgeniy Dodis and Daniel Wichs. Non-malleable extractors and symmetric key cryptography from weak secrets. In *Proceedings of the forty-first annual ACM symposium on Theory of computing*, pages 601–610. ACM, 2009.

[Fra77] Peter Frankl. A constructive lower bound for ramsey numbers. *Ars Combinatoria*, 3(297-302):28, 1977.

[FW81] Peter Frankl and Richard M. Wilson. Intersection theorems with geometric consequences. *Combinatorica*, 1(4):357–368, 1981.

[Gro01] Vince Grolmusz. Low rank co-diagonal matrices and ramsey graphs. *Journal of combinatorics*, 7(1):R15–R15, 2001.

[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.

[Li11] Xin Li. A new approach to affine extractors and dispersers. In *Computational Complexity (CCC), 2011 IEEE 26th Annual Conference on*, pages 137–147. IEEE, 2011.

[Li13] Xin Li. Extractors for a constant number of independent sources with polylogarithmic min-entropy. In *Foundations of Computer Science (FOCS), 2013 IEEE 54th Annual Symposium on*, pages 100–109. IEEE, 2013.

[Li15a] Xin Li. Improved constructions of two-source extractors. *arXiv preprint arXiv:1508.01115*, 2015.

[Li15b] Xin Li. Three-source extractors for polylogarithmic min-entropy. In *Foundations of Computer Science (FOCS), 2015 IEEE 56th Annual Symposium on*, pages 863–882. IEEE, 2015.

[Mek15] Raghu Meka. Explicit resilient functions matching Ajtai-Linial. *CoRR*, abs/1509.00092, 2015.

[Nag75] Zs Nagy. A constructive estimation of the ramsey numbers. *Mat. Lapok*, 23:301–302, 1975.

[Nao92] Moni Naor. Constructing ramsey graphs from small probability spaces. *IBM Research Report RJ*, 8810, 1992.

[Rao07] Anup Rao. An exposition of bourgains 2-source extractor. In *Electronic Colloquium on Computational Complexity (ECCC)*, volume 14, 2007.

[Raz05] Ran Raz. Extractors with weak random seeds. In *Proceedings of the thirty-seventh annual ACM symposium on Theory of computing*, pages 11–20. ACM, 2005.

[Vin11] Le Anh Vinh. The Szemerédi–Trotter type theorem and the sum-product estimate in finite fields. *European Journal of Combinatorics*, 32(8):1177 – 1181, 2011.

[Zuc06] David Zuckerman. Linear degree extractors and the inapproximability of max clique and chromatic number. In *Proceedings of the thirty-eighth annual ACM symposium on Theory of computing*, pages 681–690. ACM, 2006.

ECCC ISSN 1433-8092

http://eccc.hpi-web.de