Electronic Colloquium on Computational Complexity, Revision 1 of Report No. 65 (2018)

Near-Optimal Erasure List-Decodable Codes

Avraham Ben-Aroya[*†] Dean Doron[‡§†] Amnon Ta-Shma[¶†]

Abstract

A code $\mathcal{C}$ is $(1-\tau, L)$ erasure list-decodable if for every word w, after erasing any $1-\tau$ fraction of the symbols of w, the remaining τ-fraction of its symbols have at most L possible completions into codewords of $\mathcal{C}$. Non-explicitly, there exist binary $(1-\tau, L)$ erasure list-decodable codes having rate $O(\tau)$ and tiny list-size $L = O(\log \frac{1}{\tau})$. Achieving either of these parameters explicitly is a natural open problem (see, e.g., [GI02, Gur03, Gur04]). While partial progress on the problem has been achieved, no prior explicit construction achieved rate better than $\Omega(\tau^2)$ or list-size smaller than $\Omega(1/\tau)$. Furthermore, Guruswami showed no *linear* code can have list-size smaller than $\Omega(1/\tau)$ [Gur03]. We construct an explicit binary $(1-\tau, L)$ erasure list-decodable code having rate $\tau^{1+\gamma}$ (for any constant $\gamma > 0$ and small τ) and list-size $\mathrm{poly}(\log \frac{1}{\tau})$, answering simultaneously both questions, and exhibiting an explicit non-linear code that provably beats the best possible linear code.

The binary erasure list-decoding problem is equivalent to the construction of explicit, low-error, strong dispersers outputting one bit with minimal entropy-loss and seed-length. For error ε, no prior explicit construction achieved seed-length better than $2\log(\frac{1}{\varepsilon})$ or entropy-loss smaller than $2\log(\frac{1}{\varepsilon})$, which are the best possible parameters for extractors. We explicitly construct an ε-error one-bit strong disperser with near-optimal seed-length $(1+\gamma)\log(\frac{1}{\varepsilon})$ and entropy-loss $O(\log\log\frac{1}{\varepsilon})$.

The main ingredient in our construction is a new (and almost-optimal) *unbalanced* two-source extractor. The extractor extracts one bit with constant error from two independent sources, where one source has length n and tiny min-entropy $O(\log\log n)$ and the other source has length $O(\log n)$ and arbitrarily small constant min-entropy rate. When instantiated as a balanced two-source extractor, it improves upon Raz's extractor [Raz05] in the constant error regime. The construction incorporates recent components and ideas from extractor theory with a delicate and novel analysis needed in order to solve dependency and error issues that prevented previous papers (such as [Li15, CZ16, Coh16b]) from achieving the above results.

[*]The Blavatnik School of Computer Science, Tel-Aviv University.

[†]Supported by the Israel science Foundation grants no. 994/14 and 952/18 and by Len Blavatnik and the Blavatnik Family foundation.

[‡]Department of Computer Science, University of Texas at Austin. Email: deandoron@utexas.edu.

[§]This work was done while being at Tel-Aviv University.

[¶]The Blavatnik School of Computer Science, Tel-Aviv University. Email: amnon@tau.ac.il.

ISSN 1433-8092

1 Introduction

Extractors and dispersers are important derandomization tools with numerous applications (see, e.g., [Sha02, Wig09]). Both extractors and dispersers are hash functions $C : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ that take an input string $x \in \{0,1\}^n$ and an auxiliary seed $y \in \{0,1\}^d$, and output an element $C(x,y)$ in a smaller universe $\{0,1\}^m$ where $m \ll n$. Both extractors and dispersers are meant to hash any input distribution X that has some crude uniformity to a nearly uniform distribution. Also, the measure of crude uniformity is the same for both objects: We say a distribution X is a k-source if it has k min-entropy, i.e., the probability of each $x \sim X$ is at most 2^{-k}.

Extractors and dispersers differ in the way they measure the proximity of the output distribution to the uniform distribution: Extractors use the total-variation distance, whereas dispersers use support-size distance (that is, they count the number of elements not in the image of the hash function). Extractors are stronger objects, and, roughly speaking, extractors are needed to derandomize two-sided error algorithms whereas dispersers suffice for one-sided error derandomization.

More formally, a function $C : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ is a *strong* (k,ε) *extractor* if for any k-source X the output distribution $(U_d, C(X,U_d))$, containing the seed y along with output $C(x,y)$, is ε-close to the uniform distribution over $\{0,1\}^d \times \{0,1\}^m$. We say C is a *strong* (k,ε) *disperser* if for any k-source X, the support of $(U_d, C(X,U_d))$ covers at least $(1-\varepsilon)2^{d+m}$ elements from $\{0,1\}^d \times \{0,1\}^m$.

There are two natural parameters measuring the quality of extractors and dispersers:

1. **Seed length.** Both extractors and dispersers use an auxiliary uniform independent source to extract the entropy from the weak source X. The length d of the auxiliary source is called the *seed-length*. We would like the seed-length to be as small as possible.

2. **Entropy loss.** There is $k+d$ bits of entropy in the system: k bits coming from the k-source X, and d bits from the independent uniform seed. The *entropy-loss* is $k-m$, i.e., the difference between the entropy in the input system (including the seed) and the output system (of length $d+m$).

As noted, strong dispersers are weaker objects than strong extractors. The interest in dispersers stems from the fact that their parameters can outperform those of extractors. For extractors, [RTS00] showed that every strong extractor requires seed-length $d \geq 2\log(\frac{1}{\varepsilon}) + \log(n-k) - O(1)$ and has an unavoidable entropy-loss of $k - m \geq 2\log(\frac{1}{\varepsilon}) - O(1)$. Non-explicitly there exist strong extractors with seed-length $d \leq 2\log(\frac{1}{\varepsilon}) + \log(n-k) + O(1)$ and entropy-loss $k - m \leq 2\log(\frac{1}{\varepsilon}) + O(1)$. For strong dispersers, [RTS00] showed that every strong disperser requires seed-length $d \geq \log(\frac{1}{\varepsilon}) + \log(n-k) - O(1)$ and has an unavoidable entropy-loss $k - m \geq \log\log(\frac{1}{\varepsilon}) - O(1)$. Again, non-explicitly, there exist strong dispersers with seed-length $d \leq \log(\frac{1}{\varepsilon}) + \log(n-k) + O(1)$ and entropy-loss $k - m \leq \log\log(\frac{1}{\varepsilon}) + O(1)$ [RTS00, MRZ14].

For strong dispersers, even the case of outputting just one bit in a way that outperforms extractors constructions has been widely open. Indeed, Gradwohl et al. [GKRTS05] noticed that such strong dispersers imply good Ramsey graphs, another problem that withstood many attempts for many years, until the recent breakthrough result of Chattopadhyay and Zuckerman [CZ16].

In this paper we go in the reverse direction of that taken in [GKRTS05]. By using the recent machinery of *non-malleable* extractors and their connection to two-source extractors [CZ16, BADTS17, Coh16d, Li17, Li18], we construct near-optimal *unbalanced* two-source extractors (which imply near-optimal *unbalanced* Ramsey graphs). We use these extractors to obtain explicit strong dispersers that output a single bit, with near-optimal seed-length and near-optimal entropy-loss.

Theorem 1.1 (see also Theorem 5.2)**.** *For every constant $0 < \gamma < 1$ and $\varepsilon = n^{-\Omega(1)}$ there exists an explicit strong (k, ε) disperser* $\mathsf{Disp} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}$ *with $d = (1 + \gamma)\log(\frac{1}{\varepsilon})$ and $k = O(\log\log\frac{1}{\varepsilon})$, where the constant in the $O(\cdot)$ notation is independent of n but may depend on γ.*

We remark that the dependence of the seed-length on the error is $(1+\gamma)\log(\frac{1}{\varepsilon}) < 2\log(\frac{1}{\varepsilon})$, and the entropy-loss is $O(\log\log\frac{1}{\varepsilon}) < 2\log(\frac{1}{\varepsilon})$ and both these bounds are optimal for dispersers up to small factors and are *impossible* for extractors. Most previous disperser constructions have not obtained parameters better than the extractors lower bounds, and we are only aware of one exception: Meka et al. [MRZ14], extending the techniques in [GKRTS05], gave a strong disperser with optimal entropy-loss. However, their construction works only for extremely high min-entropy $k = n - \Theta(1)$ and has suboptimal seed-length.

1.1 Erasure List-Decodable Codes

We now turn our attention to binary list-decodable codes in the erasures model. A code $\mathcal{C}$ is a set $\mathcal{C} \subseteq \mathbb{F}_2^n$. We call elements in $\mathbb{F}_2^n$ *words* and elements in C *codewords*. Two interesting parameters of a code are its *redundancy* and its *noise-resiliency*. The redundancy is measured by the *rate* of the code, $\frac{\log|\mathcal{C}|}{n}$. The noise-resiliency is measured according to the model of noise.

In the errors model: A code $\mathcal{C}$ is $(\tau n, L)$ *list-decodable* if for every word $w \in \mathbb{F}_2^n$ there exist at most L codewords in the Hamming ball of radius τn around w.

In the erasures model: A code $\mathcal{C}$ is $(\tau n, L)$ *erasure list-decodable* if for every $z \in \mathbb{F}_2^{(1-\tau)n}$ and every set $T \subseteq [n]$ of size $(1-\tau)n$, the number of codewords that have z in the coordinates indexed by T is at most L.

If $\mathcal{C}$ is $(\tau n, L)$ list-decodable we can recover from τn errors in the following sense: Given a word $w \in \mathbb{F}_2^n$ that was obtained by corrupting at most τn entries of some codeword c, one can (perhaps non-efficiently) produce a small set of size L that necessarily contains c.

Similarly, if $\mathcal{C}$ is $(\tau n, L)$ erasure list-decodable we can recover from τn erasures in the following sense: Given a word $w \in \{0,1,?\}^n$ that was obtained by replacing at most τn entries of some codeword c with the erasure sign '?', one can (perhaps non-efficiently) produce a small set of size L that necessarily contains c.

A strong (k, ε) extractor with one output bit is roughly equivalent to a binary $(\frac{1-\varepsilon}{2}n, L = 2^k)$ list-decodable code [Tre01]. In the same spirit, Guruswami [Gur04] observed that strong dispersers with one output bit can be used to construct erasure list-decodable codes. In this paper we complement his argument with the converse statement, showing that erasure list-decodable codes are essentially *equivalent* to strong dispersers with one output bit. Specifically, $\mathsf{Disp} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}$ is a strong (k, ε) disperser if and only if the code $\mathcal{C} : \{0,1\}^n \to \{0,1\}^{2^d}$ defined by $\mathcal{C}(x)_i = \mathsf{Disp}(x, i)$ is $((1-2\varepsilon)2^d, 2^k)$ erasure list-decodable.

As we can see, for both extractors and dispersers, the seed-length corresponds to the rate of the code, whereas the entropy-loss corresponds to the list-size of the code. Thus, the gap between the seed-lengths of dispersers (which is $\log(\frac{1}{\varepsilon})$) and extractors (which is $2\log(\frac{1}{\varepsilon})$) translates to a difference between rate ε in the erasures model compared with rate ε^2 in the errors model. Similarly, the gap between the entropy-loss of dispersers (which is $\log\log(\frac{1}{\varepsilon})$) and extractors (which is $2\log(\frac{1}{\varepsilon})$) translates to a difference between list-size $\log(\frac{1}{\varepsilon})$ in the erasures model compared with list-size $\mathrm{poly}(\frac{1}{\varepsilon})$ in the errors model. Formally:

- Non-explicitly there exist binary codes having rate $\Omega(\varepsilon^2)$ that are $(\frac{1-\varepsilon}{2} \cdot n, \mathrm{poly}(\frac{1}{\varepsilon}))$ list-decodable and these parameters are tight.
- Non-explicitly there exist binary codes having rate $\Omega(\varepsilon)$ that are $((1-\varepsilon)n, O(\log \frac{1}{\varepsilon}))$ erasure list-decodable, and up to a constant multiplicative factor in the list-size these parameters are tight [Gur03].

Thus, erasure list-decodable codes can have quadratically better rate and exponentially smaller list-size than list-decodable codes. In fact, Guruswami proved that any *linear* erasure list-decodable codes must have $L = \Omega(1/\varepsilon)$ [Gur03], and so the exponential improvement (or any better than polynomial improvement) is necessarily only possible for non-linear constructions.

The state of affairs for *explicit* binary erasure list-decodable codes is similar to that of *explicit* dispersers. That is, only few explicit binary erasure list-decodable codes are known to have rate below $\Omega(\varepsilon^2)$ or list-size below $\Omega(1/\varepsilon)$. Guruswami and Indyk [GI02] gave a *probabilistic* polynomial-time algorithm that outputs with high probability an erasure list-decodable code of rate $\Omega\left(\frac{\varepsilon^2}{\log(1/\varepsilon)}\right)$ and optimal list-size (their construction can be explicitly derandomized when ε is constant). The natural open problem of obtaining erasure list-decodable codes having rate better than ε^2 was explicitly mentioned several times, e.g., in [GI02, Gur04]. More concretely, in [Gur01, Open Question 10.2], Guruswami posed the open problem of constructing efficient erasure list-decodable codes of rate ε^{2-a}.

Incorporating the above discussion with Theorem 1.1, we get the best explicit construction to date:

Theorem 1.2 (see also Theorem 5.8)**.** *For every constant $0 < \gamma < 1$ and $\varepsilon = n^{-\Omega(1)}$ there exists an explicit $((1-\varepsilon)\bar{n}, L = \log^{O(1)} \frac{1}{\varepsilon})$ erasure list-decodable code $\mathcal{C} : \{0,1\}^n \to \{0,1\}^{\bar{n}}$ of rate $\varepsilon^{1+\gamma}$, where the asymptotic notation hides constants that may depend on γ.*

Thus, Theorem 1.2 solves Guruswami's problem for the interesting regime of polynomially small ε. We stress that the codes we present are explicit in the sense that they have explicit encoding, but we do not know whether the codes we construct admit *efficient* erasure list-decoding algorithms. We also mention that the list-size $\mathrm{poly}\log(\frac{1}{\varepsilon})$ achieved by our code is exponentially smaller than the best possible list-size by any *linear* code.

1.2 Two-Source Extractors

A function $\mathsf{2Ext} : \{0,1\}^{n_1} \times \{0,1\}^{n_2} \to \{0,1\}$ is an $((n_1,k_1),(n_2,k_2),\varepsilon)$ *two-source extractor* if for any two independent sources X and Y, where X is an (n_1,k_1) source and Y is an (n_2,k_2) source, the output distribution $\mathsf{2Ext}(X,Y)$ is ε-close to uniform.

Often, the two-source extractor terminology is more expressive than the extractor notation, as we explain now. Suppose $\mathsf{Ext} : \{0,1\}^k \times \{0,1\}^d \to \{0,1\}$ is a strong (k,ε) extractor. Fix an (n,k) source X, and let ε_i be the distance of the distribution $\mathsf{Ext}(X,i)$ from uniform. By the extractor definition we know that $\mathbb{E}[\varepsilon_i] \leq \varepsilon$. However, the extractor definition does not distinguish between the case where the ε error occurs because all seeds $y \in \mathsf{Supp}(Y)$ have the same error ε, and the case where ε fraction of the seeds have constant error and the rest have none. The situation is different with two-source extractors. Roughly speaking, in an $((n,k),(d,d'),\varepsilon)$ two-source extractor, there are at most $2^{d'-d}$ bad seeds y with distance $\varepsilon_y \geq \varepsilon$. Thus, the two-source extractor notation allows separating the fraction of bad seeds from the quality of good seeds.

We would like to explicitly construct a strong (k,ε) disperser $\mathsf{Disp} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ with parameters better than those of (k,ε) extractors. Thus, on the one hand, for almost every seed y, $\mathsf{Disp}(X,y)$ covers almost all of $\{0,1\}^m$, and, on the other hand, Disp is not a strong extractor, so for almost every seed y, $\mathsf{Disp}(X,y)$ is *far* from uniform. How can this happen?

The situation becomes clearer if we look at strong dispersers with only one additional output bit, i.e., when $m = 1$. As $\mathsf{Disp}(X,y)$ is distributed over one bit, for almost every seed y, $\mathsf{Supp}(\mathsf{Disp}(X,y)) = \{0,1\}$. Yet, it is possible (even necessary, since Disp is not an extractor) that for many seeds y, $\mathsf{Disp}(X,y)$ is ε_0 away from uniform for some constant $\varepsilon_0 \gg \varepsilon > 0$., e.g., when $\mathsf{Disp}(X,y)$ has much more weight on 0 than on 1.

One clean way of capturing this is by using the two-source extractor terminology. We are looking for a two-source extractor $\mathsf{2Ext}$ where almost all seeds (except for ε fraction) are "good" in the sense that y is good if $\mathsf{2Ext}(X,y)$ covers both 0 and 1. Roughly speaking, this amounts to an explicit construction of an $((n,k),(d,d'),\varepsilon_0)$ two-source extractor having $\varepsilon = 2^{d'-d}$ and any non-trivial error $\varepsilon_0 < 1$. Two-source extractors with arbitrary $\varepsilon_0 < 1$ are also called bipartite Ramsey graphs (see Claim 5.11).

Explicitly constructing two-source extractors (and Ramsey graphs) is a long standing and important challenge. A long line of research (e.g., [CG88, Raz05, Bou05, BKS+10,

BRSW12]) culminated in $((n,k),(n,k),\varepsilon_0)$ two-source extractors supporting poly-log min-entropy [Coh16c, CZ16, Mek17]. This was later improved to $k = O(\log n \frac{\log\log n}{\log\log\log n})$ [BADTS17, Coh16d, Li17, Li18]. However, using the latter two-source extractors gives dispersers with suboptimal entropy-loss and long seed, or, equivalently, erasure list-decodable codes with large list-size and low rate.

Another natural two-source extractor is Raz's two-source extractor [Raz05]. Raz's function is an $((n,k),(d = O(\log \frac{n}{\varepsilon}),d'),\varepsilon_{\mathsf{Raz}})$ two-source extractor that has an unbalanced entropy requirement; the first source is long and very weak (k can be as small as, roughly, $\log\log \frac{n}{\varepsilon_{\mathsf{Raz}}}$), the second source is short and somewhat dense with $d' \geq \delta d$, for any constant $\delta > \frac{1}{2}$. The fact that k can be very small corresponds to a disperser with small entropy-loss, which is good for us. Moreover, d is small, which is again what we want because the length of the corresponding erasure list-decodable code is 2^d. The error $\varepsilon_{\mathsf{Raz}}$ of Raz's extractor is exponentially-small in $\min\{k,d'\}$ which is much better than the mere non-trivial error that we need. However, the second source must be relatively dense, satisfying $\frac{d'}{d} \geq \frac{1}{2}$. This implies that the error ε of the disperser is given by $2^{-d+d'}$ and as a consequence $d \geq 2\log(\frac{1}{\varepsilon})$.

In this paper we show how to explicitly construct the necessary two-source extractor. We show:

Theorem 1.3 (see also Theorem 4.1)**.** *For every two constants $\delta,\varepsilon_0 > 0$ and every $k \geq \Omega(\log\log n)$ there exists an explicit $((n,k),(d,\delta d),\varepsilon_0)$ two-source extractor* $\mathsf{2Ext} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}$ *with $d = O(\log n)$.*

Theorem 1.3 is interesting on its own right. The entropy requirement in both sources is optimal up to constant factors, as both sources have entropy which is logarithmic in the length of the other source. This property is also true for Raz's extractor. On the negative side, Theorem 1.3 has a large error ε_0, whereas Raz's extractor has a very small error. On the positive side, Raz's extractor works only when $d' = \delta d > 0.5d$ whereas Theorem 1.3 works with $d' = \delta d$ for any $\delta > 0$, and it is this feature that gives a disperser construction with parameters better than those possible for extractors. Having Theorem 1.3 immediately gives the strong one output bit disperser and the non-linear near-optimal erasure list-decodable code discussed above.

We also obtain a variant of Theorem 1.3 that gives a new construction of *balanced* two-source extractors.

Theorem 1.4. *For every two constants $\delta,\varepsilon_0 > 0$ and every $k \geq \Omega(\log n)$ there exists an explicit $((n,k),(n,\delta n),\varepsilon_0)$ two-source extractor* $\mathsf{2Ext} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}$.

We see that one source has a minimal entropy requirement of $O(\log n)$ while the other has arbitrarily small constant entropy rate. Again, this improves upon [Raz05] in terms of entropy requirement but is worse in terms of error. Theorem 1.4 is also incomparable to [Li18] as there, both sources require min-entropy at least $O(\log n \frac{\log\log n}{\log\log\log n})$.

Both Theorem 1.3 and Theorem 1.4 follow directly from Theorem 4.1.

1.3 The Two-Source Extractor Construction

We now give an informal presentation of the two-source extractor construction. We try to keep the discussion intuitive, and for that we omit (or ignore) some technical details. We also assume some familiarity with the field, sometimes using notions that will be formally presented in Section 2.

The input to the $((n,k),(d,\delta d),\varepsilon_0)$ two-source extractor is an (n,k) source X and a $(d,\delta d)$ source Y, for some $0<\delta<\frac{1}{2}$. At a high level, we do the following:

1. Increase the entropy rate of Y from δ to, say, 0.7. For that, we use a constant-error *condenser*. We cannot do it deterministically (because the condenser needs a uniformly random seed) and we still want to keep X fresh. Therefore, we apply the condenser on Y and every possible seed, letting the output of this procedure be a table Y' in which each row corresponds to an application with a different seed. The table Y' has the guarantee that most of the rows of Y' are close to having entropy rate 0.7.

2. Next, we would like to transform the dense rows of Y' to uniformly random strings. For that, we use Raz's extractor with the first source X and the rows of Y' as (independent) seeds. Call the resulting table Y'' and note that it is a function of both X and Y. Also note that although it is now guaranteed that a constant fraction of the rows of Y'' are uniform (Raz's extractor works with entropy rate above half), it is *not* guaranteed (and also not true) that the rows of Y'' are independent of each other.

3. Now we wish to break the dependence between the rows of Y'' so that (ideally) every t of them are uniform and independent (think of t as being poly-logarithmic in the number of rows of Y''). For that, we use a *correlation-breaker* that outputs one bit. The correlation-breaker requires two independent sources, which we do *not* have. Instead, we apply it on Y and Y''. Call the output table Y'''. We shall prove that with high probability, Y''' has many good rows and every t good rows of Y''' are *very* close to being uniform and independent.

4. Finally, we apply a *resilient function* f on the bits of Y'''. The output of our construction is the function's output $f(Y''')$.

   The property that we want from f is that it is nearly balanced and that its output cannot be heavily influenced by any small set of bad bits (the bad rows of Y'''). We need these properties to hold not only when the "good" bits are perfectly uniform and independent, but also under weaker conditions (e.g., that the good players are t-wise independent).

Thus, the coarse structure of our construction is essentially the same as many previous two-source extractor constructions. Namely, we use the two-sources to get a non-oblivious bit-fixing source and then apply a resilient function. There are two known approaches how to implement the first step of getting a non-oblivious bit-fixing source from two independent sources. The first approach was developed by Li [Li15] and uses

alternating extraction (or equivalently, correlation breakers).[1] The second approach, used by Chattopadhyay and Zuckerman [CZ16], uses a non-malleable extractor combined with a sampler. The second approach is more modular, while the first is more flexible.

We now elaborate more on how Li obtained the above reduction and compare it with our work. The input to Li's protocol are samples from two independent $(n, k = \mathrm{polylog}\, n)$ sources X and Y. The protocol works by applying an extractor E_1 to Y, where we enumerate over every possible seed of E_1 and build a table. Then, each such row is fed as a seed to an extractor E_2, applied on the source X. Li then proceeds to use alternating extraction to get a non-oblivious bit fixing source. (Eventually, he also uses the lightest-bin protocol to obtain a three-source extractor.)

Li's reduction and our construction are very similar, except that:

- In step (1) we replace E_1 with a *constant degree condenser*, and as a result,
- In step (2) the role of E_2 in our construction is played by a *two-source extractor* (of Raz). This is necessary because the output of the condenser is only guaranteed to have high min-entropy. Finally,
- In step (3) a correlation breaker replaces alternating extraction.

While the change seems small it is essential, and the reason why the problem waited its solution for so long. Next, we elaborate on why we use condensers instead of the extractor E_1 and which condensers should be used.

First, we notice that the two-source extractor we are set to construct is different than that of [Li15] and [CZ16]. [Li15, CZ16] construct *balanced* two-source extractors, where each of the two sources is weak (an in particular might have densities well below linear) whereas we are set to construct a highly *non-balanced* extractor where one of the sources (the small one) has *linear* density.

The key observation of the paper is that in such a situation (where the density of one source is linear) the condenser of step (1) has a huge advantage over the extractor E_1 since the condenser might have *constant* seed length (hence a constant number of rows in the table) independent of the row length, and therefore also independent of n, which is totally impossible with extractors. The fact that such explicit condensers exist is a beautiful result of [DKSS13]. The analysis (done in Section 4.3) critically uses this fact (that the number of output rows of the condenser is a *constant*) in a delicate way to prove the correctness of the construction.

We also mention that our construction shares steps that are similar to Cohen's construction [Coh16a] of three-source extractors. The vital difference is that in [Coh16a], a third source is used to achieve complete independence between the rows of a table and

[1] In [Li15] Li uses *three* independent sources in his construction. However, Chattopadhyay and Zuckerman [CZ16] remark that their two source extractors could also have been obtained using Li's approach, once a low-depth, highly resilient function is constructed (as is done in [CZ16]). Thus, we view Li's construction as a reduction from *two* independent sources to a non-oblivious bit-fixing source.

then a simple parity can be applied, even if only one row is close to uniform. Here, we only use *two* sources.

To conclude this part, we discuss the dependence problem (to be explained soon), and what in our solution of this problem is different than previous solutions:

- First, there is the issue of lack of independence between the source Y and the seed Y'' in item (3) of the construction. To overcome this, we show a conditioning under which Y'' is still good, Y is independent of Y'' and even after the conditioning the two sources have enough min-entropy. In recent years, such conditioning methods were very successful in constructing an abundance of primitives (e.g., correlation breakers, independence-preserving mergers and non-malleable extractors, etc.).

- Next, there is a delicate issue with the errors. The error $\varepsilon_{\mathsf{cond}}$ of the condenser is high (think of it as a constant). In a naive analysis we would argue that each t good rows are $\varepsilon' > \varepsilon_{\mathsf{cond}}$ close to uniform, and therefore the whole table Y''' is $A^t\varepsilon'$-close to a table where the good rows are perfectly t-wise independent, where A is the number of rows in the table Y'''. However, such an approach is doomed to fail, as necessarily $A\varepsilon_{\mathsf{cond}} > 1$.

  Our solution for this problem is the heart of the argument. We observe that some of the errors in the construction depend on A, the number of rows in the table, while others depend on the *row length*. In the construction we make sure that A is small (think of it as a fixed constant) while the row length is unbounded (and, e.g., grows to infinity as n grows to infinity). Thus, we have a natural separation between *large* errors that depend on the number of rows A, and *small* errors that depend on the row length. A similar distinction between large and small errors appears in [Li15].

  The condenser of step (1) and the resilient function of step (4) incur large errors. Raz's extractor (step (2)) and the correlation breaker with advice (step (3)) incur small errors that are exponentially-small in the row length. We show that with some constant probability we succeed in step (1), and that once we have succeed, the errors δ in steps (2) and (3) are so small that $A^t\delta$ is still small, hence Y''' is close to a table with t-wise independent good players, and so the resilient function in step (4) works (and incurs another constant error). Thus, while the failure probability is high, when we succeed we are exponentially-close to uniform.

  Notice that the fact that A is a constant (independent of the row length) is crucial for the argument to work, and this is why we resort to using condensers rather than extractors as in previous solutions.

- Finally, the argument used in the last bullet raises a difficulty treating the set of good rows. Specifically, in [CZ16], the set of good rows is a function of one of the sources. In our analysis the set of good rows is not just a function of the *sources* X and Y, but also depends on the specific *sample* $y \sim Y$.

Indeed, as we said before, this strategy leads to better unbalanced two-source constructions, and consequently to constructions of near-optimal erasure list-decodable codes (with high rate and small list-size), and one output-bit strong dispersers (with almost optimal seed length and entropy requirement) overcoming barriers that stood open for many years without seeing any progress.

1.4 Non-Strong Dispersers

Strong dispersers are the focal point of this paper. One may wonder why we insist on the strongness property, and whether the problem becomes easier when the strongness property is dropped.

- The answer to the first question is that the strongness property is essential. The equivalence between erasure list-decodable codes and dispersers requires the dispersers to be strong (see Lemma 5.6, and also notice the correspondence between code coordinates and seeds). Similarly, the connection to Ramsey graphs also requires the disperser to be strong, as already observed by Gradwohl et al. [GKRTS05]. [GKRTS05] constructed dispersers that are strong in almost all of the seed, but not strong in some part of the seed, and this drawback is severe enough that none of the applications go through.

- The answer to the second question is that it is easier to construct non-strong dispersers with good parameters. In the paper we prove that it is possible to output more bits from the source at the expense of being strong in only most of the bits (we are non-strong in only $O(1)$ bits of the seed). We prove:

  Theorem 1.5. *For every constant $0 < \gamma < 1$ and $\varepsilon = n^{-\Omega(1)}$ there exists an explicit (k, ε) disperser* $\mathsf{Disp} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ *with $d = (1+\gamma)\log(\frac{1}{\varepsilon})$, $k \geq \Omega(\log\log\frac{1}{\varepsilon})$ and $m = d + \Omega(k)$, where the constant in the $O(\cdot)$ notation may depend on γ. The disperser is strong in $d - O(1)$ bits of the seed.*

  We sketch a proof of the above theorem in Section 5.2.

1.5 Organization

The rest of the paper is organized as follows. Section 2 covers the preliminaries and notations we use. Section 3 describes the *constant degree* condenser that is used in step (1). Following the above discussion, it is important for us that A, the number of rows in the table, and equivalently the seed-length of the condenser, is a constant independent of the row length. In that section we show one can combine existing constructions of somewhere-random condensers and mergers to achieve that. Next, in Section 4, we describe and analyze the new unbalanced two-source extractor. In Section 5 we use the new two-source extractor to obtain near-optimal strong seeded dispersers, erasure list-decodable codes and unbalanced Ramsey graphs. We conclude with a few open problems in Section 6.

2 Preliminaries

Throughout the paper we use the convention that lowercase variables are the logarithm (in base 2) of their corresponding uppercase variables, e.g., $n = \log N$, $d = \log D$. We denote by $[A]$ the set $\{1, \ldots, A\}$. The density of a set $B \subseteq A$ is $\rho(B) = \frac{|B|}{|A|}$. We say a function $f : A \to B$ is *explicit* if there exists a deterministic polynomial algorithm that runs in time $\mathrm{poly}(\log |A|)$ and computes f.

2.1 Random Variables and Min-Entropy

The *statistical distance* between two distributions X and Y on the same domain Ω is defined as $|X - Y| = \max_{A \subseteq \Omega}(\Pr[X \in A] - \Pr[Y \in A])$. If $|X - Y| \leq \varepsilon$ we say 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.

For a function $f : \Omega_1 \to \Omega_2$ 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 $A \subseteq \Omega_1$, $f(A) = \{f(x) \mid x \in A\}$. For every $f : \Omega_1 \to \Omega_2$ and two random variables X and Y distributed over Ω_1, it holds that $|f(X) - f(Y)| \leq |X - Y|$.

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

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

A random variable X is an (n, k) source if X is distributed over $\{0,1\}^n$ and has min-entropy at least k. When n is clear from the context we sometimes omit it and simply say that X is a k-source. Every k-source X can be expressed as a convex combination of *flat* distributions each with min-entropy at least k.

Definition 2.1 (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 will use the following simple claim about average conditional min-entropy:

Claim 2.2. *For any random variables X, Y,*

$$\tilde{H}_\infty(X|Y) \;\geq\; H_\infty(X) - \log |\mathsf{Supp}(Y)|.$$

2.2 Condensers

Definition 2.3 (condenser)**.** *$C : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ is an $(n, k) \to_{\varepsilon_{\mathsf{cond}}} (m, k')$ condenser, if for every (n, k) source X, $C(X, U_d)$ is $\varepsilon_{\mathsf{cond}}$-close to an (m, k') source. If $k = \delta n$ and $k' = \delta' m$ we say C is a $\delta \to_{\varepsilon_{\mathsf{cond}}} \delta'$ condenser.*

Lemma 2.4. *Suppose $C : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ is an $(n,k) \to_{\varepsilon_{\mathsf{cond}}} (m, k'+d)$ condenser. Let X be an (n,k) source. Let ε_i be the minimal distance of $C(X,i)$ to an (m,k') source. Then, $\mathbb{E}_{i\in\{0,1\}^d}[\varepsilon_i] \leq \varepsilon_{\mathsf{cond}}$.*

Proof: Fix an (n,k) source X. For $i \in \{0,1\}^d$, let $H_i \subseteq \{0,1\}^m$ be the set of elements $w \in \{0,1\}^m$ such that $\Pr_{x\in X}[C(x,i) = w] \geq 2^{-k'}$. The distance of $C(X,i)$ from a k'-source is $\varepsilon_i = \Pr_{x\in X}[C(x,i) \in H_i] - 2^{-k'}|H_i|$. Let $H = \bigcup_{i\in\{0,1\}^d} H_i$. Then,

- For every $w \in H$, $\Pr_{x\in X, i\in\{0,1\}^d}[C(x,i) = w] \geq 2^{-d}2^{-k'} = 2^{-(k'+d)}$, and,
- it holds that

$$\begin{aligned}
\varepsilon_{\mathsf{cond}} &\geq \Pr_{x\in X, i\in\{0,1\}^d}[C(x,i) \in H] - |H|2^{-(k'+d)} \\
&= \sum_{i\in\{0,1\}^d} 2^{-d} \Pr_x[C(x,i) \in H] - |H|2^{-(k'+d)} \\
&\geq \sum_{i\in\{0,1\}^d} 2^{-d} \Pr_x[C(x,i) \in H_i] - 2^{-(k'+d)} \sum_{i\in\{0,1\}^d} |H_i| \\
&= \sum_{i\in\{0,1\}^d} 2^{-d} \left(\Pr_x[C(x,i) \in H_i] - 2^{-k'}|H_i|\right) \\
&= \sum_{i\in\{0,1\}^d} 2^{-d}\varepsilon_i \;=\; \mathbb{E}_{i\in\{0,1\}^d}[\varepsilon_i].
\end{aligned}$$

■

2.3 Two-Source Extractors

Definition 2.5 (two-source extractor)**.** $\mathsf{2Ext} : \{0,1\}^{n_1} \times \{0,1\}^{n_2} \to \{0,1\}^m$ *is an* $((n_1,k_1),(n_2,k_2),\varepsilon)$ *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* $\mathsf{2Ext}(X_1,X_2) \approx_\varepsilon U_m$*. We say that* 2Ext *is* strong *if*

$$(\mathsf{2Ext}(X_1,X_2), X_1) \approx_\varepsilon (U_m, X_1)$$

and

$$(\mathsf{2Ext}(X_1,X_2), X_2) \approx_\varepsilon (U_m, X_2).$$

In our construction, we will use the following two-source extractor:

Theorem 2.6 ([Raz05])**.** *For every constant $\delta_{\mathsf{Raz}} > \frac{1}{2}$ there exist constants $c_1 = c_1(\delta_{\mathsf{Raz}}), c_2 = c_2(\delta_{\mathsf{Raz}}) > 1$ such that for every n_1, k_1, n_2, k_2 satisfying*

- $k_1 \geq c_1 \log n_2$,
- $k_2 \geq c_2 \log n_1$,

there exists an explicit strong $((n_1,k_1),(n_2,k_2=\delta_{\mathsf{Raz}}n_2),\varepsilon_{\mathsf{Raz}})$ *two-source extractor*

$$\mathsf{Raz}:\{0,1\}^{n_1}\times\{0,1\}^{n_2}\to\{0,1\}^m$$

with $m=\Omega(\min\{k_1,k_2\})$ *and* $\varepsilon_{\mathsf{Raz}}=2^{-\Omega(m)}$*, where the constants hiding in the asymptotic notation may depend on* δ_{Raz}*.*

Claim 2.7. *Suppose*

$$\mathsf{2Ext}:\{0,1\}^{n_1}\times\{0,1\}^{n_2}\to\{0,1\}^m$$

is a strong $((n_1,k_1),(n_2,k_2),\varepsilon)$ *two-source extractor. Let* X *be an* (n,k_1) *source. Call an element* $y\in\{0,1\}^{n_2}$ *is* bad *if* $|\mathsf{2Ext}(X,y)-U_m|>\varepsilon$*, and let* BY *denote the set of all bad elements. Then,* $|BY|<2^{k_2}$*.*

Proof: Assume towards contradiction that $|BY|\geq 2^{k_2}$ and let Y be the uniform distribution over the set BY. Then, $H_\infty(Y)\geq k_2$ and so $(\mathsf{2Ext}(X,Y),Y)\approx_\varepsilon(U_m,Y)$ which implies that

$$\frac{1}{|BY|}\sum_{y\in BZ}|\mathsf{2Ext}(X,y)-U_m|\;\leq\;\varepsilon.$$

However, $|\mathsf{2Ext}(X,y)-U_m|>\varepsilon$ for every $y\in BY$, in contradiction. ∎

2.4 Mergers

A *merger* takes as input a list of possibly correlated random variables along with a short uniform seed and outputs one random variable which is close to having high min-entropy, provided at least one of the input variables has high min-entropy. Formally:

Definition 2.8 (somewhere-random source)**.** *A source* $X=X_1\circ\ldots\circ X_A$ *is an* $(n,k,(\alpha,\beta))$ somewhere-random (s.r.) source *if there is a random variable* $I\in\{0,\ldots,A\}$ *such that for every* $i\in[A]$*,* $(X_i|I=i)$ *is* α*-close to an* (n,k) *source and* $\Pr[I=0]\leq\beta$*. The variable* I *is called the* indicator *of source. If* $\alpha=\beta=0$ *we say* X *is a* (n,k) s.r. source*.*

Definition 2.9 (merger)**.** $B:(\{0,1\}^n)^D\times\{0,1\}^t\to\{0,1\}^m$ *is a* (k,k',ε)*-merger, if for every* (n,k) *s.r. source* $X=X_1\circ\ldots\circ X_A$*, the output* $M(X,U_t)$ *is* ε*-close to a* k'*-source.*

There are explicit constructions of good mergers. Dvir and Wigderson [DW11] constructed the *curve merger* and proved that it works with $t=O(\log\frac{n}{\varepsilon})$. This was further improved in [DKSS13] who proved that $t=O(\log\frac{D}{\varepsilon})$ suffices. Notice that now t only depends on the number of sources D and the requested error ε, but *not* on the source length n, and this remarkable property will turn crucial for us. Formally,

Theorem 2.10 ([DW11, DKSS13])**.** *There exists a constant* $c_{\mathsf{DKSS}}\geq 1$ *such that the following holds. Fix* $\beta,\delta,\varepsilon>0$*. There exists an explicit function* $B:(\{0,1\}^n)^D\times\{0,1\}^t\to\{0,1\}^n$ *that is a* $(k=\delta n,k'=(1-\beta)\delta n,\varepsilon)$ *merger, with* $t=c_{\mathsf{DKSS}}\cdot\frac{1}{\beta}\log\frac{D}{\varepsilon}$*.*

2.5 Correlation Breakers with Advice

A correlation-breaker with advice is a function $\mathsf{CBA} : \{0,1\}^n \times \{0,1\}^\ell \times [A] \to \{0,1\}^m$ where we think of the first input as a weak source, the second as an independent short seed and the last as an advice string. Roughly speaking, applying CBA on t possibly correlated seeds with t distinct advice strings results in independent random variables. For example, $\mathsf{CBA}(X, Y, \alpha)$ is (nearly) independent of $\mathsf{CBA}(X, Y, \alpha')$ for any $\alpha \neq \alpha'$. Formally,

Definition 2.11. *A function* $\mathsf{CBA} : \{0,1\}^n \times \{0,1\}^\ell \times [A] \to \{0,1\}^m$ *is a* $(t, k, \varepsilon_{\mathsf{CBA}})$ *correlation-breaker with advice if the following holds. If* Y *is a distribution over* $\{0,1\}^n$*,* $Z = (Z_1, \ldots, Z_t)$ *is a distribution on* $(\{0,1\}^\ell)^t$*,* $\mathcal{H}$ *is a random variable and* $\delta > 0$ *satisfying:*

- Y *and* Z *are independent, conditioned on* $\mathcal{H}$*,*
- $\tilde{H}_\infty(Y|H) \geq k + \log(1/\varepsilon_{\mathsf{CBA}})$*,*
- $(Z_1, \mathcal{H}) \approx_\delta (U_\ell, \mathcal{H})$*, and,*
- $\alpha_1, \ldots, \alpha_t \in [A]$ *are distinct strings.*

Then,

$$\left(\mathsf{CBA}(Y, Z_1, \alpha_1), \left(\mathsf{CBA}(Y, Z_i, \alpha_i)\right)_{i=2}^t, \mathcal{H}\right) \approx_{\delta + 2\varepsilon_{\mathsf{CBA}}} \left(U_m, \left(\mathsf{CBA}(Y, Z_i, \alpha_i)\right)_{i=2}^t, \mathcal{H}\right).$$

We use the following result:

Theorem 2.12 ([Coh16b, Theorem 4.12])**.** *There exists a constant* $c_{\mathsf{CBA}} \geq 1$ *such that the following holds. Let* n, a *be integers and* $\varepsilon_{\mathsf{CBA}} > 0$*. Then, there exists an explicit* $(t, k_{\mathsf{CBA}}, \varepsilon_{\mathsf{CBA}})$ *correlation-breaker with advice*

$$\mathsf{CBA} : \{0,1\}^n \times \{0,1\}^\ell \times [A] \to \{0,1\}$$

with $\ell = c_{\mathsf{CBA}} \cdot at \cdot \log \frac{n}{\varepsilon_{\mathsf{CBA}}}$ *and* $k_{\mathsf{CBA}} \geq \ell$*.*

In our setting, the number of rows A is a constant independent of n. For this reason we work with a "basic" correlation-breaker, where there is no attempt to optimize the dependence of ℓ on a. This gives a seed-length which is optimal up to constant multiplicative factors.

We also need the following lemma.

Lemma 2.13. *Let* $X_1, \ldots, X_t$ *be random variables over* $\{0,1\}^m$*. Further suppose that for any* $i \in [t]$*,*

$$\left(X_i, \{X_j\}_{j \neq i}\right) \approx_\varepsilon \left(U_m, \{X_j\}_{j \neq i}\right).$$

Then, $(X_1, \ldots, X_t) \approx_{t\varepsilon} U_{tm}$*.*

2.6 Limited Independence and Non-Oblivious Bit-Fixing Sources

Definition 2.14. *A distribution X over $\{0,1\}^A$ is called* (t,γ)-wise independent *if the restriction of X to every t coordinates is γ-close to U_t. A source X over $\{0,1\}^A$ is called a* (q,t,γ) non-oblivious bit-fixing source *if there exists a subset $Q \subseteq A$ of size at most q such that the joint distribution of the bits in $A \setminus Q$ is (t,γ)-wise independent. The bits in Q are allowed to arbitrarily depend on the bits in $A \setminus Q$. If $\gamma = 0$ we often say that X is a (q,t) non-oblivious bit-fixing source.*

Lemma 2.15 ([AGM03])**.** *A (t,γ)-wise distribution over A bits is $(A^t\gamma)$-close to some t-wise independent distribution.*

Definition 2.16. *Let $f : \{0,1\}^A \to \{0,1\}$, $\mathcal{D}$ a distribution over $\{0,1\}^A$ and $Q \subseteq A$. Let $I_{Q,\mathcal{D}}(f)$ denote the probability that f is undetermined when the variables outside Q are sampled from $\mathcal{D}$. We define $I_{q,t,\gamma}(f)$ to be the maximum of $I_{Q,\mathcal{D}}(f)$ over all $Q \subseteq A$ of size q and all $\mathcal{D}$ that is a (t,γ) independent distribution. We say that f is* (t,γ)-independent (q,ε)-resilient *if $I_{q,t,\gamma}(f) \leq \varepsilon$.*

Theorem 2.17 ([CZ16, Mek17])**.** *For every $0 < \gamma < 1$ there exists a constant $c_\gamma \geq 1$ such that for all $A > 0$ there exists an explicit function $f : \{0,1\}^A \to \{0,1\}$ with the following property:*

For every $t \geq c_\gamma \log^4 A$,

- *f is almost balanced: For any t-wise independent distribution $\mathcal{D}$ on $\{0,1\}^A$,*

$$\Pr_{x\sim\mathcal{D}}[f(x)=1] = 1/2 \pm A^{-1/c_\gamma}, \text{ and,}$$

- *f is resilient: $I_{q,t,\gamma}(f) \leq c_\gamma \cdot \frac{q}{A^{1-\gamma}}$.*

3 Constant Degree Condensers

In this section we prove:

Theorem 3.1. *For every constant $0 < \delta_1 < \delta_2 = 0.7$, every $s \geq s_0(\delta_1)$ and every integer n_1 and $\varepsilon_{\mathsf{cond}} \geq 2^{-\Omega(n_1)}$ there exists an explicit $\delta_1 \to_{\varepsilon_{\mathsf{cond}}} \delta_2$ condenser $C : \{0,1\}^{n_1} \times \{0,1\}^d \to \{0,1\}^{n_2}$ with $n_2 = (\frac{2}{3})^s n_1$ and $d = 4c_{\mathsf{DKSS}}\left(s + \log\frac{1}{\varepsilon_{\mathsf{cond}}}\right)$, where c_{DKSS} is the constant from Theorem 2.10. Note that d is independent of n_1.*

Note that, in particular, for every $\delta_1 > 0$ there exists an explicit $\delta_1 \to_{\varepsilon_{\mathsf{cond}}} \delta_2 = 0.7$ condenser $C : \{0,1\}^{n_1} \times \{0,1\}^d \to \{0,1\}^{n_2}$ with $n_2 = \Omega(n_1)$ and $d = O(\log\frac{1}{\varepsilon_{\mathsf{cond}}})$. However, we will need the more precise version that appears in Theorem 3.1.

The proof goes through *somewhere-random condensers*, so let us first discuss the similarities and differences between condensers and somewhere-random condensers. We begin with the necessary definitions:

Definition 3.2 (s.r. condenser). *A function $C : \{0,1\}^n \to (\{0,1\}^m)^A$ is an $(n,k) \to_\varepsilon (m,k')$ s.r. 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. If $k = \delta n$ and $k' = \delta m$ we say C is $\delta \to_\varepsilon \delta'$ s.r. condenser.*

We may take a condenser $C : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ and expand it to a table with the outputs of all possible seeds, i.e., define $S : \{0,1\}^n \to (\{0,1\}^m)^D$, with $D = 2^d$, where $S(x)_i = C(x,i)$. The condenser property guarantees that for every k-source X, *most* rows in the table are close to having k' min-entropy. In contrast, a s.r. condenser is a weaker object, because it only guarantees that *one* row has k' entropy (or more precisely that we are in a convex combination of such cases).

The major question we consider now is the dependence of the degree (2^d for condensers and A for s.r. condensers) on n, m, k, k' and ε. We focus on the case where $m = \Omega(n)$, $k = \delta n$, $k' = \delta' m$ and $\delta < \delta'$ are constants. A-priori, we could have expected the degree to depend on n and ε, as is indeed the case when m might be arbitrarily small. However, remarkably, things are drastically different when $m = \Omega(n)$. In this case both condensers and s.r. condensers may be of degree that is independent of n and this will be crucial for us. If we consider the dependence on the error, then s.r. condensers may have exponentially-small error and constant D, whereas the degree of a condenser is at least $d \geq \log(\frac{1}{\varepsilon})$. Remarkably, all of that can be explicitly achieved, as we now explain.

The basic building block we use is the following beautiful result of Zuckerman, which is based on additive combinatorics:

Theorem 3.3 ([Zuc06, Theorem 8.3]). *For every constant $0 < c < 1$ there exists a constant $\alpha = \alpha(c)$ such that for every constant $\delta \leq c$ and integer n there exists an explicit function $C : \{0,1\}^n \to (\{0,1\}^{\frac{2}{3}n})^2$ that is a $\delta \to_\varepsilon (1+\alpha)\delta$ s.r. condenser with $\varepsilon = 2^{-\Omega(\alpha\delta n)}$.*

Somewhere-random condensers can be easily composed. Specifically, Barak et al. [BKS+10] showed that if $C_1 : \{0,1\}^{n_1} \to (\{0,1\}^{n_2})^{\ell_1}$ is a $\delta_1 \to_{\varepsilon_1} \delta_2$ s.r. condenser and $C_2 : \{0,1\}^{n_2} \to (\{0,1\}^{n_3})^{\ell_2}$ a $\delta_2 \to_{\varepsilon_2} \delta_3$ s.r. condenser then $C_2 \circ C_1 : \{0,1\}^{n_1} \to (\{0,1\}^{n_3})^{\ell_1 \cdot \ell_2}$ defined by $C_2 \circ C_1(x)_{(i_1,i_2)} = C_2(C_1(x)_{i_1})_{i_2}$ is a $\delta_1 \to_{\varepsilon_1+\varepsilon_2} \delta_3$ s.r. condenser.

Composing the s.r condenser of Theorem 3.3 with itself s times we get an explicit function $C : \{0,1\}^n \to (\{0,1\}^m)^D$ with $D = 2^s$ and $m = (\frac{2}{3})^s n$ that is a $\delta \to_\varepsilon \delta'$ s.r. condenser with $\varepsilon = \sum_{i=1}^{s} 2^{-\Omega((1+\alpha)^i \delta (\frac{2}{3})^i n)} = 2^{-\Omega(m)}$ and $\delta' \geq (1+\alpha(\delta'))^s \delta$. Therefore:

Lemma 3.4. *For every constants $0 < \delta_1 < \delta_2 < 1$ there exists a constant $s = s(\delta_1, \delta_2)$ and an explicit function $C : \{0,1\}^{n_1} \to (\{0,1\}^{n_2})^D$ that is a $\delta_1 \to_\varepsilon \delta_2$ s.r. condenser with $D = 2^s$, $n_2 = (\frac{2}{3})^s n_1$ and $\varepsilon = 2^{-\Omega(n_2)}$. Note that D is independent of n and ε.*

Right now, if X is a k-source, the table $C(X)$ has D rows, and, roughly speaking, the guarantee is that one of these rows has density δ'. We want to change this to get a condenser, i.e., we are willing to invest a short seed (that is independent of n) and we want to get *one* output which is close to uniform. (Alternatively, we can write the condenser as a table with one row per seed, the number of rows is independent of n

and most rows are close to uniform.) This is exactly what a *merger* does and applying the merger of Theorem 2.10 with $\beta = \frac{1}{4}$ on the s.r. condenser of Lemma 3.4 (with δ_2 close to 1) gives Theorem 3.1

4 The Unbalanced Two-Source Extractor Construction

The main result of this section is the following two-source extractor.

Theorem 4.1. *For every integer n and two constants $\delta_0, \varepsilon_0 > 0$ there exists a constant c such that for $d \geq c \log n$ and $k \geq c \log d$ there exists an explicit $((n,k),(d,\delta_0 d),\varepsilon_0)$ two-source extractor* $\mathsf{2Ext} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}$.

The extractor in the above theorem has constant error, and works when:

1. Each source's entropy is in the order of the logarithm of the length of the other source.
2. The shorter source, of length d, has an arbitrarily small constant density δ_0.

We think of n and $d = d(n)$ as growing parameters while ε_0 and δ_0 are constants. We use asymptotic notations (such as $\Omega(\cdot)$) to hide constants that are independent of n and d (but may depend on ε_0 and δ_0).

4.1 The Construction

Recall that ε_0 is the target error of the extractor 2Ext. The input to 2Ext is a pair (x,y) where x is drawn from an (n,k) source X, and y is drawn from an independent $(d,\delta_0 d)$ source Y. Our problem is that the y comes from a $\delta_0 d$-source for some $\delta_0 < \frac{1}{2}$. To overcome this, we do the following:

- We apply the condenser of Theorem 3.1 on y to get a table y' that is 1-wise 0.7-dense. Notice that the output of this step is a table rather than a single output.
- We apply Raz's extractor (Theorem 2.6) on the table and the input x from the other source to convert the table y' to another table y'' that is 1-wise uniform.
- We apply the t correlation-breaker with advice of Theorem 2.12 on y, using the table y'' as the seed, to get a table y''' that is t-wise uniform.
- Finally, we apply the resilient function f of Theorem 2.17 on the table y''' to collapse the many rows of the table to a single, close to uniform, output.

Formally, these steps work as follows:

Condense the short source: We are given $\delta_0 < \frac{1}{2}$. Set $\delta' = 0.69$ and $\delta_2 = 0.7$.

By Theorem 3.1 there exists a constant $s_0 = s_0(\delta_0)$ such that for every $s \geq s_0$ there exists an explicit

$$C : \{0,1\}^d \times \{0,1\}^a \to \{0,1\}^{d'}$$

that is a $\delta_0 \to_{\varepsilon_{\mathsf{cond}}} \delta_2 = 0.7$ condenser with $a = 4c_{\mathsf{DKSS}}(s + \log \frac{1}{\varepsilon_{\mathsf{cond}}})$ and $d' = (\frac{2}{3})^s d$.

We set

$$\gamma = \frac{1}{2^5 c_{\mathsf{DKSS}}},$$

and this also fixes c_γ as in Theorem 2.17. Notice that γ and c_γ are fixed constants independent of all other parameters in our system.

Now, choose $\varepsilon_{\mathsf{cond}}$ so that

$$\left(\frac{1}{\varepsilon_{\mathsf{cond}}}\right)^{\log(3/2)} \geq \frac{4}{\delta_0} 2^{12} c_\gamma c_{\mathsf{DKSS}}^4 \log^4 \frac{1}{\varepsilon_{\mathsf{cond}}}, \tag{1}$$

and also so that $\varepsilon_{\mathsf{cond}} \leq \xi(\varepsilon_0, \delta_0)$, where

$$\xi(\varepsilon_0, \delta_0) = \min\left\{2^{-s_0}, \left(\frac{\varepsilon_0}{8}\right)^2, \left(\frac{\varepsilon_0}{5}\right)^{c_\gamma}, \left(\frac{\varepsilon_0}{5c_\gamma}\right)^{1/\gamma}\right\}. \tag{2}$$

Given $\varepsilon_{\mathsf{cond}}$, we set

$$s = \log \frac{1}{\varepsilon_{\mathsf{cond}}} \geq s_0,$$

giving $a = 8c_{\mathsf{DKSS}} \log \frac{1}{\varepsilon_{\mathsf{cond}}}$. Note that the degree of the condenser, $A = 2^a$, satisfies

$$\sqrt{\varepsilon_{\mathsf{cond}}} A = 2^{-\frac{1}{2}\log \frac{1}{\varepsilon_{\mathsf{cond}}} + a} = 2^{-\frac{a}{2^4 c_{\mathsf{DKSS}}} + a} = A^{1-2\gamma}.$$

Observe that $s \geq s_0$ and that $d' = \Omega(d)$. Also, notice that $(\delta_2 - \delta')d' = d'/100 \geq a = \log A$ for large enough d. Thus, C is a $(d, \delta d) \to_{\varepsilon_{\mathsf{cond}}} (d', \log(A) + \delta' d')$ condenser.

Define an $A \times d'$ table Y' where

$$Y'_i = C(Y, i) \in \{0,1\}^{d'}$$

for $i = 1, \ldots, A$.

1-wise uniformity: Let c_1, c_2 be the constants from Theorem 2.6 for $\delta_{\mathsf{Raz}} = 0.6$.

Notice that $\delta_{\mathsf{Raz}} d' = \Omega(d') = \Omega(d)$. Therefore, for a constant c large enough, $d \geq c \log n$ is large enough so that $\delta_{\mathsf{Raz}} d' \geq c_2 \log n$. We can, in particular, choose c such that in addition $c \geq c_1$. Recalling that $k \geq c \log d$, we have $k \geq c_1 \log d'$. By Theorem 2.6, there exists an explicit function

$$\mathsf{Raz} : \{0,1\}^n \times \{0,1\}^{d'} \to \{0,1\}^{d''}$$

that is a strong $((n,k),(d',\delta_{\mathsf{Raz}}d'),\varepsilon_{\mathsf{Raz}}=2^{-\Omega(d'')})$ two-source extractor with

$$d''=\Omega(\min\{k,\delta_{\mathsf{Raz}}d'\})=\Omega(k).^{2}$$

Define an $A\times d''$ table Y'' where

$$Y''_i=\mathsf{Raz}(X,Y'_i)$$

for $i=1,\ldots,A$.

t-wise uniformity: Let $k_{\mathsf{CBA}}=\frac{\delta_0 d}{8}$ and $\varepsilon_{\mathsf{CBA}}=\frac{1}{d}$. Set

$$t\;=\;\frac{\delta_0}{4}\left(\frac{3}{2}\right)^{s}.$$

Notice that for a large enough constant c we have $d''=\Omega(k)=\Omega(c\log d)\geq c_{\mathsf{CBA}}at\log\frac{d}{\varepsilon_{\mathsf{CBA}}}$, where the latter is the seed-length required by the correlation-breaker from Theorem 2.12. Also, $k_{\mathsf{CBA}}=\frac{\delta_0 d}{8}\geq d''$ for large enough d, as $d''=\Omega(k)=\Omega(\log d)$. Hence, by Theorem 2.12 there exists an explicit function

$$\mathsf{CBA}:\{0,1\}^{d}\times\{0,1\}^{d''}\to\{0,1\}$$

that is a $(t,k_{\mathsf{CBA}},\varepsilon_{\mathsf{CBA}})$ correlation-breaker with advice.

Define an $A\times 1$ table Y''' where

$$Y'''_i=\mathsf{CBA}(Y,Y''_i,i)$$

for $i=1,\ldots,A$.

Keep in mind that the entropy in Y suffices for CBA since $H_\infty(Y)=8k_{\mathsf{CBA}}$.

Collapse: Take $f:\{0,1\}^{A}\to\{0,1\}$ to be the $(q=A^{1-2\gamma},t,\varepsilon_f=c_\gamma A^{-\gamma})$ resilient function of Theorem 2.17 and output $f(y'''_1,\ldots,y'''_A)$.

4.2 Two Subtleties

As mentioned in the introduction, there are several delicate issues in the analysis:

1. Circular dependence: Y'' depends on both X and Y, and is used as a seed in the application of the correlation-breaker with advice on Y.
2. We need Y''' to be close to a perfect t-wise independent table, while the correlation-breaker with advice only guarantees that every t good rows are close to uniform. To bridge the gap we need the error to be at least polynomially-small in the number of rows, but some of the steps incur a large constant error.

[2] Although $k\geq c\log d$ we can always assume w.l.o.g. that $k=c\log d$ and so $k\leq\delta_{\mathsf{Raz}}d'=\Omega(d)$.

To overcome the first issue we show a conditioning under which Y'' is still good, Y is independent of Y'' and even after the conditioning the two sources have enough min-entropy.

To overcome the second issue we distinguish between large errors that depend on the number of rows A, and small errors that depend on the row length (see Section 1.3 in the introduction). In particular, the errors are of three types:

- The probability p_1 that a value we condition upon is bad. This error is incurred by the condenser and is high (think of it as being a constant).
- We show that when we condition on a good value, every t good rows in Y''' are p_2-close to uniform. We then claim that Y''' as a table is $A^t p_2$-close to a table where the good rows are truly t-wise independent (where A is the number of rows in the table Y'''). The error p_2 is incurred by Raz's extractor and by the correlation-breaker, and can be made very small if we deal with a source X having enough min-entropy. We make p_2 small enough so that $A^t p_2$ is also small.
- A third error p_3 is incurred by the resilient function f. This error is large, say, a constant, and we are fine with that.

Note that we cannot just accumulate all errors as $A^t p_1$ is way larger than 1. Instead, we argue that with a constant probability $1 - p_1$, we get extremely close to perfect behavior, and then we get such a small error p_2 so that $A^t p_2$ is also small.

4.3 The Analysis

Proof of Theorem 4.1: Fix an (n, k) source X and an independent $(d, \delta d)$ source Y. We decompose the proof into three parts:

- In the first part we prove that very often (except for a small constant probability) the table Y'' contains many rows that are marginally close to uniform.
- Next, we prove that every set of t rows $\{i, j_1, \ldots, j_{t-1}\}$ in Y''' are *product* in the sense that if i is a good row (intuitively meaning that Y''_i is close to uniform) and $j_1, \ldots, j_{t-1}$ are $t-1$ other rows, then in Y''', Y'''_i is close to uniform and *independent* of $Y'''_{j_1}, \ldots, Y'''_{j_{t-1}}$. This part involves applying a correlation-breaker with advice on Y and Y''. In order to ensure that Y and Y'' are independent, we condition on the values of Y' in the t rows $\{i, j_1, \ldots, j_{t-1}\}$.
- Together, except for a small constant probability, there are many good rows, and every t rows of Y''' are product, hence the table Y''' is close to a (q, t) non-oblivious bit-fixing source, where every good row is a good bit in the bit-fixing source. Hence, $f(Y''')$ is close to uniform.

Part 1: Often, many rows in Y' are good

Let ε_i be the minimal distance of $C(Y,i)$ from a $\delta' d'$-source. According to Lemma 2.4,

$$\mathbb{E}_{i\in[A]}[\varepsilon_i] \leq \varepsilon_{\mathsf{cond}}.$$

Definition 4.2. *We say $z \in \{0,1\}^{d'}$ is* good *if $\mathsf{Raz}(X,z)$ is $\varepsilon_{\mathsf{Raz}}$-close to uniform. Let GZ be the set of all good z-s, and BZ the rest. We say $i \in [A]$ is* good for $y \in \{0,1\}^d$ *if $C(y,i) \in GZ$ and* bad *otherwise. We define a random variable B_i, where the sample space is Y, and $B_i(y) = 1$ if i is bad for y and 0 otherwise.*

By Claim 2.7, $|BZ| \leq 2^{\delta_{\mathsf{Raz}} d'}$. Therefore, in expectation, the number of bad rows for y is small:

Claim 4.3. $\mathbb{E}_{y\in Y}\big[\sum_{i\in[A]} B_i(y)\big] \leq 2\varepsilon_{\mathsf{cond}} A$.

Proof: Fix an $i \in [A]$. We have that $C(Y,i)$ is ε_i-close to some $\delta' d' = 0.69d'$-source R. Hence:

$$\mathbb{E}_y[B_i(y)] \;=\; \Pr_{y\in Y}[C(y,i)\in BZ] \;\leq\; \varepsilon_i + \Pr_{r\in R}[r\in BZ] \;\leq\; \varepsilon_i + \frac{|BZ|}{2^{\delta' d'}} \;=\; \varepsilon_i + 2^{-0.09d'}.$$

Thus, for d large enough,

$$\mathbb{E}_y\Big[\sum_{i\in[A]} B_i(y)\Big] \;=\; \sum_{i\in[A]} \mathbb{E}_y[B_i(y)] \;\leq\; \sum_{i\in[A]} \Big(\varepsilon_i + 2^{-0.09d'}\Big) \;\leq\; \varepsilon_{\mathsf{cond}} A + 2^{-0.09d'} A \;\leq\; 2\varepsilon_{\mathsf{cond}} A.$$

■

Definition 4.4. *We say $y \in \mathsf{Supp}(Y)$* has many bad rows *if $\sum_{i\in[A]} B_i(y) \geq \sqrt{\varepsilon_{\mathsf{cond}}} A$.*

Denote $p_{1,1} = \frac{\varepsilon_0}{4}$.

Claim 4.5. $\Pr_{y\in Y}[y \text{ has many bad rows}] \leq p_{1,1}$.

Proof: By Markov,

$$\Pr_{y\in Y}\Big[\sum_i B_i(y) \;\geq\; \sqrt{\varepsilon_{\mathsf{cond}}} A\Big] \;\leq\; \frac{\mathbb{E}\big[\sum_i B_i(y)\big]}{\sqrt{\varepsilon_{\mathsf{cond}}} A} \;\leq\; \frac{2\varepsilon_{\mathsf{cond}} A}{\sqrt{\varepsilon_{\mathsf{cond}}} A} = 2\sqrt{\varepsilon_{\mathsf{cond}}} \;\leq\; \frac{\varepsilon_0}{4},$$

where the last inequality follows from the fact that $\varepsilon_{\mathsf{cond}} \leq (\frac{\varepsilon_0}{8})^2$. ■

Part 2: The good rows are t-wise independent

We introduce some notations to simplify the expressions in the proof. For $y_0 \in \{0,1\}^d$ and $k \in [A]$, let $Y_k'''(y_0)$ denote $(Y_k'''|Y = y_0)$. Also, for a set $S \subseteq [A]$, define $Y_S'''(y_0) = \left\{Y_j'''(y_0)\right\}_{j\in S}$. Denote $p_2 = \varepsilon_{\mathsf{Raz}} + 2\varepsilon_{\mathsf{CBA}}$.

Definition 4.6. *Let $y_0 \in \{0,1\}^d$ (not necessarily in the support of Y). Let $i \in [A]$ and $S \subseteq [A] \setminus \{i\}$ of cardinality $t-1$. We say y_0* violates the product rule *for (i,S) if $B_i(y_0) = 0$ and*

$$(Y_i'''(y_0), Y_S'''(y_0)) \;\not\approx_{p_2}\; U_1 \times Y_S'''(y_0).$$

Definition 4.7. *Let $y_0 \in \{0,1\}^d$ (not necessarily in the support of Y). Let $i \in [A]$ and $S \subseteq [A] \setminus \{i\}$ of cardinality $t-1$. We say y_0* violates the product rule with distinguisher *$\Delta : \{0,1\}^t \to \{0,1\}$ for (i,S) if $B_i(y_0) = 0$ and*

$$\left| \Pr[\Delta\left(Y_i'''(y_0), Y_S'''(y_0)\right) = 1] - \Pr[\Delta(U_1, Y_S'''(y_0)) = 1] \right| > p_2.$$

Observe that if y_0 violates the product rule then there exists some Δ such that y_0 violates the product rule with distinguisher Δ.

Lemma 4.8. *For every i and S as above, the number of $y \in \{0,1\}^d$ that violate the product rule for (i,S) is at most $2^{\delta_0 d/2 + 2^t}$.*[3]

Proof: Suppose the lemma is false for some (i,S). Then, by the pigeonhole principle there exists some Δ such that the number of elements $y \in \{0,1\}^d$ that violate the product rule for (i,S) with distinguisher Δ is at least $2^{\delta_0 d/2}$. Let BY denote the set of these elements. Identify BY with the uniform distribution over the set BY.

Let $BY_i' = C(BY, i)$, $BY_i'' = \mathsf{Raz}(X, BY_i')$ and $BY_i''' = \mathsf{CBA}(BY, BY_i'', i)$. For a subset $T \subseteq [A]$ Let BY_T' denote the sub-table of BY' corresponding to the rows of T, and similarly BY_T'' and BY_T'''. Since for every $y \in BY$, we have that

$$\Delta\left(BY_i'''(y), BY_S'''(y)\right) \;\not\approx_{p_2}\; \Delta(U_1, BY_S'''(y)),$$

this holds also on average, that is

$$\Delta\left(BY_i''', BY_S'''\right) \;\not\approx_{p_2}\; \Delta(U_1, BY_S''').$$

Thus, it follows that

$$BY_{S\cup\{i\}}''' \;\not\approx_{p_2}\; U_1 \times BY_S'''. \tag{3}$$

On the other hand, when we condition on the values of

$$\mathcal{H} \;=\; BY_{S\cup\{i\}}'$$

the conditions for the correlation-breaker with advice hold:

[3] We could have used an alternative argument that avoids the 2^{2^t} factor here by a minor deterioration in the error of the CBA. However, since the t we use is constant the 2^{2^t} factor is negligible.

- BY and $BY''_{S\cup\{i\}}$ are independent given $\mathcal{H} = BY'_{S\cup\{i\}}$, since $\mathcal{H}$ is a function of BY alone, and given that $\mathcal{H} = BY'_{S\cup\{i\}} = h$ for some h, $BY''_{S\cup\{i\}}$ is a function of X alone.

- $$\begin{aligned}\tilde{H}_\infty(BY|\mathcal{H}) &\geq H_\infty(BY) - \log(|\mathsf{Supp}(\mathcal{H})|) \\ &= H_\infty(BY) - td' \geq \frac{\delta_0 d}{2} - td' \geq \frac{\delta_0 d}{4},\end{aligned}$$

  because

  $$\frac{td'}{d} = t \cdot \left(\frac{2}{3}\right)^s = \frac{\delta_0}{4}.$$

  Now, since $k_{\mathsf{CBA}} = \frac{\delta_0 d}{8}$ and $\varepsilon_{\mathsf{CBA}} = \frac{1}{d}$ we also have for d large enough,

  $$\tilde{H}_\infty(BY|\mathcal{H}) \;\geq\; \frac{\delta_0 d}{4} \;\geq\; k_{\mathsf{CBA}} + \log\frac{1}{\varepsilon_{\mathsf{CBA}}}.$$

- $B_i(y) = 0$, hence $BY'_i \in GZ$ and $BY''_i = \mathsf{Raz}(X, BY'_i)$ is $\varepsilon_{\mathsf{Raz}}$-close to uniform.

Thus, by the correlation-breaker with advice property,

$$\left(\mathsf{CBA}(BY, BY''_i, i), \left\{\mathsf{CBA}(BY, BY''_j, j)\right\}_{j\in S}\right) \;\approx_{\varepsilon_{\mathsf{Raz}}+2\varepsilon_{\mathsf{CBA}}}\; \left(U_1, \left\{\mathsf{CBA}(BY, BY''_j, j)\right\}_{j\in S}\right),$$

or, equivalently,

$$(BY'''_i, BY'''_S) \;\approx_{p_2}\; U_1 \times BY'''_S,$$

in contradiction to Equation (3). ∎

Definition 4.9. *Say $y \in \{0,1\}^d$* violates the product rule *if it violates it for some $i \in [A]$ and $S \subseteq [A] \setminus \{i\}$ of cardinality $t-1$.*

As $H_\infty(Y) \geq \delta_0 d$, the probability $y \in Y$ violates the product rule for a specific (i, S) is at most $2^{\delta_0 d/2 + 2^t - \delta_0 d} = 2^{2^t - \delta_0 d/2}$. Let $p_{1,2} = \frac{\varepsilon_0}{10}$. Then, by the union bound, for d large enough:

Corollary 4.10. $\Pr_{y\in Y}[y \text{ violates the product rule}] \leq 2^{2^t - \delta_0 d/2} \cdot A^t \leq p_{1,2}$.

Part 3: Completing the proof

Definition 4.11. *We say y is* bad *if it has many bad rows or if it violates the product rule. If y is not bad we say it is* good.

Let $p_1 = p_{1,1} + p_{1,2}$. Clearly, by Claim 4.5 and Corollary 4.10, $\Pr_{y\in Y}[y \text{ is bad}] \leq p_1 = \left(\frac{1}{4} + \frac{1}{10}\right)\varepsilon_0$.

Claim 4.12. *Fix any good $y \in Y$. Then, $Y'''(y)$ is a (q, t, tp_2) non-oblivious bit-fixing source, for $q = \sqrt{\varepsilon_{\mathsf{cond}}}A$.*

Proof: Let $Q(y) \subseteq [A]$ be the set of bad rows for y. As y does not have many bad rows, $|Q(y)| = \sum_{i\in[A]} B_i(y) \leq \sqrt{\varepsilon_{\mathsf{cond}}}A = q$.

Now, fix any set $S \subseteq [A] \setminus Q(y)$ of cardinality t. Let $i \in S$. As $S \subseteq [A] \setminus Q(y)$ and $i \in S$ we have $i \notin Q(y)$ and therefore $B_i(y) = 0$. Also, y does not violate the product rule, hence,

$$\left(Y'''_i(y), Y'''_{S\setminus\{i\}}(y)\right) \approx_{p_2} U_1 \times Y'''_{S\setminus\{i\}}(y).$$

As this is true for any $i \in S$, by Lemma 2.13,

$$Y'''_S(y) \approx_{tp_2} U_t.$$

Thus, $Y'''(y)$ is a (q, t, tp_2) non-oblivious bit-fixing source. ■

In particular, By Lemma 2.15, for every good y, $Y'''(y)$ is $tA^t p_2$-close to a (q, t) non-oblivious bit-fixing source. By the choices we have made above $q = \sqrt{\varepsilon_{\mathsf{cond}}}A \leq A^{1-2\gamma}$. Equation (1) implies that

$$t = \frac{\delta_0}{4}\left(\frac{1}{\varepsilon_{\mathsf{cond}}}\right)^{\log(3/2)} \geq c_\gamma \log^4 A.$$

Using the resiliency of f from Theorem 2.17 (and the fact that it is almost balanced), the output when y is good is p_3-close to uniform for $p_3 = tA^t p_2 + \varepsilon_f + A^{-1/c_\gamma}$, where the first term is due to the distance from a t-wise distribution, the second is due to the resiliency and the third is due to the bias of f (see, e.g., Lemma 2.11 in [CZ16]). To that we also have to add the probability p_1 that y is not good. To finish the proof we notice that:

- It holds that

$$\varepsilon_f \leq c_\gamma \frac{q}{A^{1-\gamma}} \leq c_\gamma \frac{A^{1-2\gamma}}{A^{1-\gamma}} = c_\gamma A^{-\gamma} \leq c_\gamma 2^{-\gamma \log \frac{1}{\varepsilon_{\mathsf{cond}}}} \leq \frac{\varepsilon_0}{5},$$

  because $\varepsilon_{\mathsf{cond}} \leq (\frac{\varepsilon_0}{5c_\gamma})^{1/\gamma}$.

- Also,

$$A^{-1/c_\gamma} \leq 2^{-\frac{1}{c_\gamma}\log\frac{1}{\varepsilon_{\mathsf{cond}}}} = {\varepsilon_{\mathsf{cond}}}^{1/c_\gamma} \leq \frac{\varepsilon_0}{5},$$

  because $\varepsilon_{\mathsf{cond}} \leq (\frac{\varepsilon_0}{5})^{c_\gamma}$.

- Finally, $tp_2 = t(\varepsilon_{\mathsf{Raz}} + 2\varepsilon_{\mathsf{CBA}})$, $\varepsilon_{\mathsf{Raz}} = 2^{-\Omega(k)} = d^{-\Omega(1)}$, $\varepsilon_{\mathsf{CBA}} = \frac{1}{d}$. Thus, $tp_2 \leq 4td^{-\Omega(1)}$. A and t are constants, so for d large enough, $tA^t p_2 \leq \frac{\varepsilon_0}{5}$.

Together, the error is at most $p_1 + p_3 \leq \varepsilon_0$ completing the proof of the theorem. ■

5 Strong Seeded Dispersers and Friends

5.1 Strong Seeded Dispersers

Definition 5.1 (strong disperser)**.** $\mathsf{Disp} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ *is a strong* (k,ε) *disperser, if for every* (n,k) *source* X*,*

$$|\mathsf{Supp}((Y, \mathsf{Disp}(X,Y)))| > (1-\varepsilon)DM.$$

We say Disp *is (source) linear if for every* $y \in \{0,1\}^d$ *and every* $x_1, x_2 \in \mathbb{F}_2^n$, $\mathsf{Disp}(x_1+x_2, y) = \mathsf{Disp}(x_1, y) + \mathsf{Disp}(x_2, y)$.

We are interested in the important special case where $m = 1$. In this case, non-explicitly, a random function is (w.h.p.) a strong (k,ε) disperser with $d = \log n + \log(\frac{1}{\varepsilon}) + O(1)$ provided that $k \geq \log\log(\frac{1}{\varepsilon}) + O(1)$ [RTS00, MRZ14]. A matching lower bound, up to additive constant factors, was given by [RTS00].

Using the translation between strong seeded dispersers and erasure list-decodable codes which we discuss in Section 5.3, Guruswami and Indyk's result [GI02] gives a probabilistic polynomial time algorithm that outputs with high probability a strong seeded disperser with seed-length $d = 2\log(\frac{1}{\varepsilon}) + \log n + \log\log(\frac{1}{\varepsilon})$ and optimal entropy-loss. The construction can be made deterministic, but with running time exponential in $1/\varepsilon$. See Table 1 for a summary of previous results.

	Required entropy k	Seed length d	
Lower-bound, non-explicit	$\log\log(\frac{1}{\varepsilon})$	$\log(\frac{1}{\varepsilon}) + \log n$	[RTS00, MRZ14]
[GI02]	$\log\log(\frac{1}{\varepsilon})$	$(2+\gamma)\log(\frac{1}{\varepsilon}) + \log n$	Constant ε, or prob. construction
This work (Theorem 5.2)	$O(\log\log\frac{1}{\varepsilon})$	$(1+\gamma)\log(\frac{1}{\varepsilon})$	$\mathrm{poly}(1/n)$ error

Table 1: Parameters of strong (k,ε) one-bit dispersers, up to additive $O(1)$ terms. γ is an arbitrarily small positive constant.

Note that as we discuss the one output bit case, the required entropy is essentially the *entropy-loss*. Form Theorem 4.1 we can derive a better explicit construction of a strong disperser with small error.

Theorem 5.2. *For every constant* $0 < \gamma < 1$ *there exists a constant* $c \geq 1$ *such that for every integer* n *and* $\varepsilon \leq n^{-\frac{c}{1-\gamma}}$ *there exists an explicit strong* (k,ε) *disperser* $\mathsf{Disp} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}$ *where* $d = (1+\gamma)\log(\frac{1}{\varepsilon})$ *and* $k = c\log d$*.*

Proof: Set $\varepsilon_0 = \frac{1}{4}$ and $\delta_0 = \frac{\gamma}{1+\gamma}$. Let c be the constant from Theorem 4.1 for δ_0 and ε_0 and let $\mathsf{2Ext} : [N] \times [D] \to \{0,1\}$ be the $((n,k),(d,k_2 = \delta_0 d),\varepsilon_0)$ two-source extractor

where $d = (1+\gamma)\log(\frac{1}{\varepsilon})$ and $k = c\log d$. Notice that $d \geq c\log n$ (because $\varepsilon \leq n^{-\frac{c}{1+\gamma}}$) as required. Let $\mathsf{Disp}(x,y) = \mathsf{2Ext}(x,y)$.

Let $X \subseteq [N]$ be a set of size K and call a value $y \in [D]$ *b-bad* if $\mathsf{Disp}(X,y) = \{b\}$. It follows that the sets of 0-bad y-s and 1-bad y-s are each of size less than K_2. Therefore,

$$|\mathsf{Supp}\big((U_d, \mathsf{Disp}(X,U_d))\big)| \;>\; 2K_2 + 2(D-2K_2) \;=\; 2D - 2K_2 \;=\; \left(1-\frac{K_2}{D}\right)2D \;=\; (1-\varepsilon)2D,$$

because $\frac{K_2}{D} = 2^{-(1-\delta_0)d} = 2^{-\frac{1}{1+\gamma}d} = 2^{-\log(\frac{1}{\varepsilon})} = \varepsilon$. ∎

5.2 Non-strong dispersers

We now prove Theorem 1.5 and output more bits from the source at the expense of being strong in only most of the seed. We construct

$$\mathsf{Disp} : \{0,1\}^n \times \{0,1\}^{d_1} \times \{0,1\}^{d_2} \to \{0,1\}^m,$$

where we think of d_1 and d_2 as two parts of the seed. Disp will be strong in the first d_1 bits of the seed. Using the notations of Section 4 we let

$$\mathsf{Disp}(x,y,i) \;\;=\;\; \big(y, \mathsf{Raz}(x, C(y,i))\big).$$

We now prove (in sketch) Theorem 1.5.

Proof: We adopt the notations of Section 4. In those notations, $\mathsf{Disp}(X,Y,I) = (Y, Y''_I)$. First note that the length of $i \in \{0,1\}^{d_2}$ is the logarithm of the number of rows in the table Y'' which is $a = O(1)$. By Claim 4.5 we know that for nearly every $y \in \{0,1\}^{d_1}$ we have many values $i \in \{0,1\}^{d_2}$ such that $\mathsf{Raz}(X, C(y,i))$ is $\varepsilon_{\mathsf{Raz}}$-close to uniform. In particular, for every y that has many good rows, let i_y be any such row. Then,

$$\begin{aligned} |\mathsf{Supp}(\mathsf{Disp}(X, U_{d_1}, U_{d_2}))| \;&\geq \sum_{y \text{ has many good rows}} |\mathsf{Supp}(\mathsf{Disp}(X,y,i_y))| \\ &\geq \sum_{y \text{ has many good rows}} (1-\varepsilon_{\mathsf{Raz}})2^{d''}. \end{aligned}$$

The theorem now follows since $d'' = \Omega(k)$ and $\varepsilon_{\mathsf{Raz}}$ is smaller than $2^{-\Omega(d'')}$, which implies that we can truncate the output of Raz such that when $\mathsf{Raz}(X, C(y,i))$ is $\varepsilon_{\mathsf{Raz}}$-close to uniform it covers its entire support. ∎

5.3 Erasure List-Decodable Codes

An $(\bar{n}, n)$ (binary) code is a mapping $\mathcal{C} : \{0,1\}^n \to \{0,1\}^{\bar{n}}$. The code $\mathcal{C}$ is *linear* if $\mathcal{C}$ is linear, and is denoted by $[\bar{n}, n]$. We identify a code with the image of $\mathcal{C}$. For a linear $\mathcal{C}$ this image is a linear subspace of $\mathbb{F}_2^{\bar{n}}$ of dimension n. A generator matrix for an $[\bar{n}, n]$ code $\mathcal{C}$ is any matrix whose columns form a basis for $\mathcal{C}$. In the *erasures* noise model, an adversarially chosen subset of the codeword's symbols are erased and the positions where erasures have occurred are known.

Definition 5.3 (erasure list-decodable code)**.** *A code $\mathcal{C} \subseteq \{0,1\}^{\bar{n}}$ is (s, L) erasure list-decodable if for every $r \in \{0,1\}^{\bar{n}-s}$ and every set $T \subseteq [\bar{n}]$ of size $\bar{n} - s$,*

$$\left|\left\{c \in \mathcal{C} \;\middle|\; c|_T = r\right\}\right| \;<\; L,$$

where $c|_T$ denotes the projection of c to the coordinates in T.

The following folklore lemma (see, e.g., [Gur03, Lemma 1]) gives an alternative characterization of *linear* erasure list-decodable codes.

Lemma 5.4. *An $[\bar{n}, n]_2$ linear code $\mathcal{C}$ is $((1-\varepsilon)\bar{n}, L)$ erasure list-decodable if and only if its $\bar{n} \times n$ generator matrix G has the property that every $\varepsilon\bar{n} \times n$ sub-matrix of G has rank greater than $n - \log L$.*

Non-explicitly, we have:

Theorem 5.5 ([Gur03])**.** *For every n and $\varepsilon > 0$, there exists an $(\bar{n}, n)$ binary code that is $((1-\varepsilon)\bar{n}, L)$-erasure list-decodable of rate $\frac{n}{\bar{n}} = \Omega(\varepsilon)$ and $L = O(\log(1/\varepsilon))$.*

See Table 2 for a summary of previous results.

	Rate $R = n/\bar{n}$	List size L	
Lower-bound, non-explicit	ε	$\log(\frac{1}{\varepsilon})$	[Gur03]
[GI02]	$\frac{\varepsilon^2}{\log(1/\varepsilon)}$	$\log(\frac{1}{\varepsilon})$	Constant ε, or prob. construction
This work (Theorem 5.8)	$\varepsilon^{1+\gamma}$	$\log^{O(1)}(\frac{1}{\varepsilon})$	$\mathrm{poly}(1/n)$ error

Table 2: Parameters of $(\bar{n}, N)_2$ codes, $((1-\varepsilon)\bar{n}, L)$ erasure list-decodable, up to constant multiplicative factors. γ is an arbitrarily small positive constant.

Guruswami [Gur04] observed that strong dispersers can be used to construct erasure list-decodable codes. Here we complement his argument, and note that strong dispersers are *equivalent* to erasure list-decodable codes. Given a function $\mathsf{Disp} : [N] \times [D] \to \{0,1\}$, we consider the (D, n) code $\mathcal{C}_{\mathsf{Disp}} : \{0,1\}^n \to \{0,1\}^D$ defined by $\mathcal{C}_{\mathsf{Disp}}(x)_i = \mathsf{Disp}(x, i)$. Note that the code is linear if and only if Disp is linear.

Lemma 5.6 (following [Gur04, Lemma 12])**.** *The function $\mathsf{Disp} : [N] \times [D] \to \{0,1\}$ is a strong (k, ε) disperser if and only if $\mathcal{C}_{\mathsf{Disp}}$ is $((1-2\varepsilon)D, K)$ erasure list-decodable.*

Proof: For one direction, assume Disp is a strong (k, ε) disperser. We wish to prove that $\mathcal{C}_{\mathsf{Disp}}$ is $((1-2\varepsilon)D, K)$ erasure list-decodable. Let $T = \{t_1, \ldots, t_{2\varepsilon D}\} \subseteq [D]$ be an arbitrary set of size $2\varepsilon D$ and $r \in \{0,1\}^{2\varepsilon D}$ an arbitrary string. Let $X_{T,r} \subseteq \{0,1\}^n$ denote the set of all the messages x for which $\mathcal{C}_{\mathsf{Disp}}(x)|_T = r$. Then,

$$|\mathsf{Supp}\big((U_d, \mathsf{Disp}(X_{T,r}, U_d))\big)| \;\leq\; |T| \cdot 1 + (D - |T|) \cdot 2 \;\leq\; (1-\varepsilon)2D,$$

where the first inequality follows by considering seeds in T and seeds in $[D] \setminus T$. For a seed $t_i \in T$ we have that $\mathsf{Disp}(X_{T,r}, t_i)$ is fixed, hence each such seed contributes 1 to the support size. For any other seed y, the support size of $\mathsf{Disp}(X_{T,r}, y)$ is at most 2. As Disp is a strong (k, ε) disperser, we conclude that $|X_{T,r}| \leq K$ as desired.

For the other direction assume Disp is a not a strong (k, ε) disperser. Then, there exists a set $X \subseteq \{0,1\}^n$ such that $|X| \geq K$ and $|\mathsf{Supp}\big((U_d, \mathsf{Disp}(X, U_d))\big)| \leq (1-2\varepsilon)2D$. Note that for every $y \in [D]$ we have $|\mathsf{Supp}(\mathsf{Disp}(X, y))| \in \{1, 2\}$. Therefore, following the above calculation, there exists a set $T \subseteq D$ of size at least $2\varepsilon D$ such that for each $y \in T$, $|\mathsf{Supp}(\mathsf{Disp}(X, y))| = 1$. But this means that for every $x \in X$, $\mathcal{C}_{\mathsf{Disp}}(x)|_T$ is the same (punctured) codeword. It follows that $\mathcal{C}_{\mathsf{Disp}}$ is not $((1-2\varepsilon)D, K)$ erasure list-decodable. ■

Corollary 5.7. *If* $\mathsf{Disp} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}$ *is a strong* (k, ε) *disperser with seed-length* $d = a_1 \log n + a_2 \log(\frac{1}{\varepsilon}) + a_3$ *(for some* $a_1 \geq 1, a_2 \geq 1$ *and* a_3*) then* $\mathcal{C}_{\mathsf{Disp}}$ *is a* $((1-2\varepsilon)D, K)$ *erasure list-decodable code of rate* $2^{-a_3} \cdot n^{1-a_1} \cdot \varepsilon^{a_2}$.

When ε is much smaller than $\frac{1}{n}$ the dominant factor is determined by a_2. As we mentioned earlier (and as Guruswami also notes in [Gur04]) previous explicit constructions for binary codes had $a_2 \geq 2$ (usually inherited from extractor constructions). Our construction is the first to get arbitrary close to $a_2 = 1$ and small list-size. Combining Corollary 5.7 and Theorem 5.2, we obtain:

Theorem 5.8. *For every constant* $0 < \gamma < 1$ *there exists a constant* $c \geq 1$ *such that for every integer* n *and* $\varepsilon \leq n^{-\frac{c}{1-\gamma}}$ *there exists an explicit code* $\mathcal{C} : \{0,1\}^n \to \{0,1\}^{(\frac{1}{\varepsilon})^{1+\gamma}}$ *that is*

$$\left((1-2\varepsilon)\left(\frac{1}{\varepsilon}\right)^{1+\gamma}\ ,\ \left((1+\gamma)\log\frac{1}{\varepsilon}\right)^c\right)$$

erasure list-decodable of rate $n\varepsilon^{1+\gamma}$.

5.4 Ramsey Graphs

Ramsey theory studies inevitable order that appears in large structures. It was initiated by Ramsey [Ram30], who showed that any graph over $N = 2^n$ vertices must contain a clique or an independent set of size $n/2$. A graph over N vertices is called *K-Ramsey* if it contains neither a clique nor an independent set of size K. Inaugurating the probabilistic method, Erdős [Erd47] showed that there are $2n$-Ramsey graphs. He also offered a bounty of \$100 for an *explicit* construction of an $O(n)$-Ramsey graphs.

Erdős's challenge initiated a line of beautiful constructions of Ramsey graphs [Abb72, Nag75, Fra77, Chu81, FW81]. The study of pseudorandomness gave a new perspective on Ramsey graphs. Specifically, any two-source disperser or extractor gives rise to a bipartite Ramsey graph (and hence, also to a non-bipartite Ramsey graph [Sha11]). This connection led to to new constructions of Ramsey graph [CG88, Nao92, Alo98,

Raz05, Bou05, Bar06, Gop06, BKS+10, BRSW12, Coh16c, CZ16, Mek17, Coh16d, Li17] culminating in $(N, n^{O(\log\log n/\log\log\log n)})$-Ramsey graphs [BADTS17, Li18].

In this section we tackle the problem of constructing *unbalanced* Ramsey graphs.

Definition 5.9 (Ramsey graph)**.** *A bipartite graph* $\mathsf{Ram} : [N_1] \times [N_2] \to \{0,1\}$ *is a* (K_1, K_2) *bipartite Ramsey graph if every* $K_1 \times K_2$ *induced subgraph of* Ram *is neither a bipartite clique nor a bipartite independent set.*

While it is possible to interpret some pseudorandom objects as unbalanced Ramsey graphs, they were less studied explicitly. See Table 3 for a summary of previous results.

	$K_1 : N_1$	$K_2 : N_2$	
Lower-bound	$(c-1)\log n : 2^n$	$n : n^c$	By [RTS00] and Claim 5.11
Non-explicit	$O(c\log n) : 2^n$	$n : n^c$	Probabilistic method
[Raz05]	$\log^{O(1)} n : 2^n$	$N_2^{0.5+\gamma} : n^{O(1)}$	$O(1)$ terms depend on γ
This work (Theorem 4.1)	$\log^{O(1)} n : 2^n$	$N_2^{\gamma} : n^{O(1)}$	$O(1)$ terms depend on γ

Table 3: Parameters of (K_1, K_2) Ramsey graphs in the unbalanced case, $[N_1 = 2^n] \times [N_2]$. c is any large enough constant and γ is an arbitrarily small positive constant.

It is easy to see that a two-source extractor with any nontrivial error is, in fact, a bipartite Ramsey graph, so as a corollary of Theorem 4.1, we obtain:

Corollary 5.10. *For every integer* N_1 *and a constant* $0 < \delta < 1$ *there exists a constant* $c = c(\delta) \geq 1$ *and an explicit function* $\mathsf{Ram} : [N_1] \times [N_2] \to \{0,1\}$ *that is a bipartite* $(K_1, K_2 = N_2^{\delta})$ *Ramsey graph, for* $N_2 = \log^c N_1$ *and* $K_1 = \log^c N_2$*.*

We start with the easy claim that bipartite Ramsey graphs are equivalent to strong one-bit dispersers.

Claim 5.11. *If* $\mathsf{Ram} : [N_1] \times [N_2] \to \{0,1\}$ *is a* (K_1, K_2) *bipartite Ramsey graph then* Ram *is a strong* $(k_1, \varepsilon \geq \frac{K_2}{N_2})$ *disperser with seed-length* $n_2 = k_2 + \log(\frac{1}{\varepsilon})$*. Also, if* Ram *is a strong* $(k_1, \varepsilon = \frac{K_2}{2N_2})$ *disperser then it is a* (K_1, K_2) *bipartite Ramsey graph.*

Proof: The first claim follows from the proof of Theorem 5.2.

For the other claim, which was already observed in [GKRTS05], assume Ram is a $(k_1, \varepsilon = \frac{K_2}{2N_2})$ disperser and assume towards contradiction that it is not a $(K_1, K_2 = 2\varepsilon N_2)$ bipartite Ramsey graph. Hence, there exist some $S \subseteq [N_1]$ and $T \subseteq [N_2]$ so that $|S| \geq K_1$ and $|T| \geq K_2$ such that either $\mathsf{Ram}(S,T) = \{0\}$ or $\mathsf{Ram}(S,T) = \{1\}$. Assume w.l.o.g. that $\mathsf{Ram}(S,T) = \{0\}$, so for every $t \in T$, $(t,1) \notin \mathsf{Supp}\big((U_{n_2}, \mathsf{Ram}(S, U_{n_2}))\big)$. But then,

$$\left|\mathsf{Supp}\big((U_{n_2}, \mathsf{Ram}(S, U_{n_2}))\big)\right| \;\leq\; 2(N_2 - |T|) + |T| \;\leq\; (1-\varepsilon)2N_2,$$

a contradiction. ∎

As observed in [GKRTS05], the quality of the Ramsey graph implied by the above theorem crucially depends on the seed-length of the given disperser. Specifically, if the seed-length dependence on the error ε is $2 \cdot \log(\frac{1}{\varepsilon})$ then $K_2 = 2\varepsilon N_2 > \sqrt{N_2}$ and if it is $1 \cdot \log(\frac{1}{\varepsilon})$ then K_2 can be very small.

We mention a more frugal way of obtaining Ramsey graphs from *linear* dispersers. The argument is a straightforward adaptation of an argument of Alon [Gur01, Proposition 10.15].[4] The parameters we obtain are identical to the above claim (and [GKRTS05]), except that one side of the graph is scaled down (from N to n) as is its entropy (from K to k).

Theorem 5.12. *Suppose* $\mathsf{Disp} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}$ *is a linear strong* (K, ε) *disperser. Let* G *be the* $D \times n$ *generating matrix of the* $[D, n]_2$ *linear code* $\mathcal{C}_{\mathsf{Disp}}$*. Then,* G *is a* $(2\varepsilon D, k+1)$ *bipartite-Ramsey-graph.*

Proof: Assume Disp is a linear strong (K, ε) disperser. By Lemma 5.6, $\mathcal{C}_{\mathsf{Disp}}$ is a $((1-2\varepsilon)D, K)$ erasure list-decodable code. Assume towards contradiction that G is not a $(2\varepsilon D, k+1)$ bipartite Ramsey graph. Let M' be a monochromatic $2\varepsilon D \times k+1$ sub-matrix of G. Assume that M' is the all-ones matrix (a similar argument handles the all-zeros matrix). Denote by M the $2\varepsilon D \times n$ sub-matrix of G that is formed by taking the rows of M' and all columns of G. On the one hand, by Lemma 5.6 and Lemma 5.4, $\mathrm{rank}(M) > n - \log K = n - k$. On the other hand, as M contains $k+1$ columns of rank 1, $\mathrm{rank}(M) \leq n - k$, a contradiction. ∎

It is natural to ask whether the other direction also holds, namely whether an adjacency matrix of a bipartite Ramsey graph is in fact a generating matrix of a linear, erasure list-decodable code. Stated differently, whether a low-rank matrix must contain large monochromatic rectangles. That question received much attention, as it is tightly related to the famous "log-rank conjecture" in communication complexity [Lov14, NW95]. Unfortunately, the acclaimed unconditional upper bound of Lovett [Lov16] still does not give us a meaningful result.

6 Concluding Remarks and Open Problems

- The strong disperser we construct in this paper outputs one bit, and for $k = O(\log\log\frac{1}{\varepsilon})$,
  - has $O(\log\log\frac{1}{\varepsilon})$ entropy-loss, and,
  - $(1+\gamma)\cdot\log(\frac{1}{\varepsilon})$ dependence of the seed-length on the error.

  It is natural to ask to extend the results of the paper to arbitrarily large values of k, matching (up to multiplicative factors) the non-explicit results.

[4] Alon's argument is aimed at obtaining *balanced* Ramsey graphs, while we are more concerned with the entropy they can handle.

- Our dispersers are inherently non-linear, and therefore we also get non-linear erasure list-decodable codes. How can we obtain near optimal *linear* codes?
- The erasure list-decodable code we construct is explicit in the sense that the code can be efficiently encoded. Does it also admit an efficient erasure list-*decoding* algorithm?
- The seed-length of our strong disperser is $c \log n + \log(\frac{1}{\varepsilon})$. Pushing c closer to 1 is an important open problem. In particular it would imply erasure list-decodable codes of near-optimal rate even for relatively large ε. Such a disperser with many output bits can also be used for simulating one-sided error randomized algorithms using weak random sources with nearly linear overhead [Zuc96].

References

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

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

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

[BADTS17] Avraham Ben-Aroya, Dean Doron, and Amnon Ta-Shma. An efficient reduction from two-source to non-malleable extractors: achieving near-logarithmic min-entropy. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, pages 1185–1194. ACM, 2017.

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

[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.

[Coh16a] Gil Cohen. Local correlation breakers and applications to three-source extractors and mergers. *SIAM Journal on Computing*, 45(4):1297–1338, 2016.

[Coh16b] Gil Cohen. Non-malleable extractors-new tools and improved constructions. In *LIPIcs-Leibniz International Proceedings in Informatics*, volume 50. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2016.

[Coh16c] Gil Cohen. Two-source dispersers for polylogarithmic entropy and improved ramsey graphs. In *Proceedings of the forty-eighth annual ACM symposium on Theory of Computing*, pages 278–284. ACM, 2016.

[Coh16d] Gil Cohen. Two-source extractors for quasi-logarithmic min-entropy and improved privacy amplification protocols. In *Electronic Colloquium on Computational Complexity (ECCC)*, volume 23, page 114, 2016.

[CZ16] Eshan Chattopadhyay and David Zuckerman. Explicit two-source extractors and resilient functions. In *Proceedings of the 48th Annual ACM SIGACT Symposium on Theory of Computing*, pages 670–683. ACM, 2016.

[DKSS13] Zeev Dvir, Swastik Kopparty, Shubhangi Saraf, and Madhu Sudan. Extensions to the method of multiplicities, with applications to kakeya sets and mergers. *SIAM Journal on Computing*, 42(6):2305–2328, 2013.

[DW11] Zeev Dvir and Avi Wigderson. Kakeya sets, new mergers, and old extractors. *SIAM Journal on Computing*, 40(3):778–792, 2011.

[Erd47] Paul Erdös. Some remarks on the theory of graphs. *Bulletin of the American Mathematical Society*, 53(4):292–294, 1947.

[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.

[GI02] Venkatesan Guruswami and Piotr Indyk. Near-optimal linear-time codes for unique decoding and new list-decodable codes over smaller alphabets. In *Proceedings of the thiry-fourth annual ACM symposium on Theory of computing*, pages 812–821. ACM, 2002.

[GKRTS05] Ronen Gradwohl, Guy Kindler, Omer Reingold, and Amnon Ta-Shma. On the error parameter of dispersers. In *Approximation, Randomization and Combinatorial Optimization. Algorithms and Techniques*, pages 294–305. Springer, 2005.

[Gop06] Parikshit Gopalan. Constructing ramsey graphs from boolean function representations. In *Computational Complexity, 2006. CCC 2006. Twenty-First Annual IEEE Conference on*, pages 14–pp. IEEE, 2006.

[Gur01] Venkatesan Guruswami. *List decoding of error-correcting codes*. PhD thesis, Massachusetts Institute of Technology, 2001.

[Gur03] Venkatesan Guruswami. List decoding from erasures: Bounds and code constructions. *IEEE Transactions on Information Theory*, 49(11):2826–2833, 2003.

[Gur04] Venkatesan Guruswami. Better extractors for better codes? In *Proceedings of the thirty-sixth annual ACM symposium on Theory of computing*, pages 436–444. ACM, 2004.

[Li15] 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.

[Li17] Xin Li. Improved non-malleable extractors, non-malleable codes and independent source extractors. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2017, pages 1144–1156, New York, NY, USA, 2017. ACM.

[Li18] Xin Li. Pseudorandom correlation breakers, independence preserving mergers and their applications. *Electronic Colloquium on Computational Complexity (ECCC)*, 2018.

[Lov14] Shachar Lovett. Recent advances on the log-rank conjecture in communication complexity. *Bulletin of EATCS*, 1(112), 2014.

[Lov16] Shachar Lovett. Communication is bounded by root of rank. *Journal of the ACM (JACM)*, 63(1):1, 2016.

[Mek17] Raghu Meka. Explicit resilient functions matching Ajtai-Linial. In *Proceedings of the Twenty-Eighth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1132–1148. SIAM, 2017.

[MRZ14] Raghu Meka, Omer Reingold, and Yuan Zhou. Deterministic coupon collection and better strong dispersers. In *LIPIcs-Leibniz International Proceedings in Informatics*, volume 28. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2014.

[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.

[NW95] Noam Nisan and Avi Wigderson. On rank vs. communication complexity. *Combinatorica*, 15(4):557–565, 1995.

[Ram30] FP Ramsey. On a problem of formal logic. *Proceedings of the London Mathematical Society*, 2(1):264–286, 1930.

[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.

[RTS00] Jaikumar Radhakrishnan and Amnon Ta-Shma. Bounds for dispersers, extractors, and depth-two superconcentrators. *SIAM Journal on Discrete Mathematics*, 13(1):2–24, 2000.

[Sha02] Ronen Shaltiel. Recent developments in explicit constructions of extractors. *Bulletin of the EATCS*, 77(67-95):10, 2002.

[Sha11] Ronen Shaltiel. An introduction to randomness extractors. In *International Colloquium on Automata, Languages, and Programming*, pages 21–41. Springer, 2011.

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

[Wig09] Avi Wigderson. Randomness extractors–applications and constructions. In *LIPIcs-Leibniz International Proceedings in Informatics*, volume 4. Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2009.

[Zuc96] David Zuckerman. Simulating BPP using a general weak random source. *Algorithmica*, 16(4):367–391, 1996.

[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

https://eccc.weizmann.ac.il