A Recipe for Constructing Two-Source Extractors*

Eshan Chattopadhyay†

May 29, 2020

Abstract

There has been exciting recent progress on explicit constructions of two-source extractors leading to near optimal constructions. In this article, we survey key new notions and techniques that led to this progress. We pose some open problems along the way.

1 Introduction

Randomness is a valuable resource in computation. Randomness is used to run various Monte Carlo simulations of complex systems such as the stock market or weather prediction systems. Various randomized algorithms have been discovered that often vastly outperform known deterministic counterparts (see [MR10] for examples). Cryptography is another area that crucially relies on access to random bits, and it is known that various basic cryptographic primitives fail to be secure if the *quality* of the randomness used is poor [DOPS04]. However natural sources of randomness are typically defective. This leads to the following basic question:

"Can we efficiently produce truly random bits given access to defective sources of randomness?"

Modeling a weak source To answer the above question, of course one needs to work with a model for defective random sources. In the 1950's, von Neumann [vN51] considered the simple model of a weak source being a stream of independent bits, each bit following a Bernoulli distribution with (an unknown) parameter p. He devised an efficient algorithm to extract near uniform bits from such weak sources. In the 1980's, Blum [Blu86] generalized this model and studied the problem of extracting from weak sources that are generated by finite state Markov chains. Santha and Vazirani [SV86] investigated the model of weak sources as a stream of bits, where each bit brings in some fresh entropy conditioned on all the previous bits. By now, the most widely used model of a weak source is using the notion of min-entropy. This model was proposed by Chor and Goldreich [CG88] and Zuckerman [Zuc90].

Definition 1.1. *Let X be a distribution on some finite universe Ω. The min-entropy of a distribution X is defined as $H_\infty(X) = \min_{x \in \mathrm{supp}(X)}(\log(1/\Pr[X = x]))$, where $\mathrm{supp}(X) = \{x \in \Omega : \Pr[X = x] > 0\}$.*

*A version of this article will appear in the June 2020 issue of the *SIGACT News Complexity Theory Column*, edited by Lane Hemaspaandra.

†Department of Computer Science, Cornell University, NY 14853, USA. eshanc@cornell.edu. Supported by NSF grant CCF-1849899.

Note that for a distribution X supported on $\{0,1\}^n$, we have $0 \leq H_\infty(X) \leq n$. We define an (n,k)-source to be a distribution on $\{0,1\}^n$ with min-entropy at least k.

Randomness extractors Informally, a randomness extractor is a deterministic algorithm that produces nearly uniform bits given access to a weak random source. We measure the quality of the output of an extractor using the notion of statistical distance, defined as follows: let D_1, D_2 be two distributions on some universe Ω. The statistical distance between D_1 and D_2, denoted by $\Delta(D_1; D_2)$ is defined as $\Delta(D_1; D_2) = \frac{1}{2} \cdot \sum_{x \in \Omega} |D_1(x) - D_2(x)|$, where $D_i(x)$ denotes $\Pr[D_i = x]$. We will use the notation $D_1 \approx_\epsilon D_2$ to denote the fact that $\Delta(D_1; D_2) \leq \epsilon$.

We are now ready to define a randomness extractor for a family of distributions. Let U_m denote the uniform distribution on $\{0,1\}^m$.

Definition 1.2. *Let $\mathcal{X}$ be a family of distributions on universe $\{0,1\}^n$. A function* $\mathrm{Ext} : \{0,1\}^n \to \{0,1\}^m$ *is called an ϵ-extractor for $\mathcal{X}$ if for any distribution $X \in \mathcal{X}$, we have*

$$\Delta(\mathrm{Ext}(X); U_m) \leq \epsilon.$$

The parameter ϵ is called the error of the extractor.

Given our discussion on modeling weak sources, it is natural to ask if one can design an extractor for the family of (n,k)-sources, for some k. The following folklore lemma is a strong negative result in this direction.

Lemma 1.3. *There does not exist an ϵ-extractor* $\mathrm{Ext} : \{0,1\}^n \to \{0,1\}$ *for the family of $(n, n-1)$-sources, for any $\epsilon < 1/2$.*

Proof. Suppose there exists such an extractor $\mathrm{Ext} : \{0,1\}^n \to \{0,1\}$. For $b \in \{0,1\}$, define $S_b = \{x \in \{0,1\}^n : \mathrm{Ext}(x) = b\}$ and X_b to be the source that is uniformly distributed on the set S_b. Note that at least one of S_0 and S_1, say S_0, has cardinality $\geq 2^{n-1}$. Thus, $H_\infty(X_0) \geq n-1$ but the support of $\mathrm{Ext}(X_0)$ is $\{0\}$, yielding the required contradiction. □

Despite the impossibility of randomness extraction in such generality, intense research has been conducted on randomness extraction in more restricted settings over the last 4 decades, leading to a beautiful and rich theory of randomness extraction. It is well beyond the scope of this article to provide an exhaustive list of research undertaken on randomness extraction. Instead we will focus on the concrete problem of extracting from the class of sources where each weak source consists of two independent weak sources. We formally define *two-source extractors* as follows.

Definition 1.4 (Two-source extractor)**.** *A function* $2\mathrm{Ext} : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}^m$ *is called a (k,ϵ)-extractor if it satisfies the following: for any two independent (n,k)-sources X and Y, we have*

$$\Delta(2\mathrm{Ext}(X,Y); U_m) \leq \epsilon.$$

A simple probabilistic argument shows the existence of 2-source extractors for min-entropy $k = \log n + O(1)$ (setting m, ϵ to constants). Chor and Goldreich [CG88] asked the question of explicitly constructing two-source extractors. Using Lindsey's lemma, they constructed an explicit extractor that works for min-entropy more than $n/2$. After nearly two decades, Bourgain [Bou05] broke the "half entropy rate barrier", using techniques from additive combinatorics, and constructed a two-source extractor for min-entropy $(1/2 - \delta)n$, for some tiny constant $\delta > 0$. Based on exponential sum estimates of Karatsuba [Kar71, Kar91] it follows that the Paley graph extractor (introduced in [CG88]) is a two-source extractor requiring min-entropy $C \log n$ in one of the sources and min-entropy $(1/2+\delta)n$ in the other source. Raz [Raz05] gave a more general construction of a two-source extractor in this unbalanced entropy setting. However, it appeared to be a significant challenge to construct a two-source extractor for min-entropy much smaller than $n/2$ in both of the sources.

A successful line of work [BIW06, Rao09, Li11, Li13, Li15b] considered the relaxed setting of extracting with access to more than two sources. This has led to a near optimal three-source extractor that works for polylogarithmic min-entropy and has negligible error [Li15b].

The task of constructing a two-source extractor that works for min-entropy significantly smaller than $n/2$ was achieved by Chattopadhyay and Zuckerman [CZ19], using a new framework for constructing two-source extractors that they introduced. They constructed a two-source extractor that works for $\log^C n$ min-entropy, for some constant C. The extractor in [CZ19] outputs 1 bit and has error $1/n^{\Omega(1)}$. Li [Li16] soon improved the output length to $\Omega(k)$ bits.

It remains a challenging open problem to construct a two-source extractor with error $1/n^{\omega(1)}$ for min-entropy significantly smaller than $n/2$ (the extractor constructions in [CG88, Bou05] achieve exponentially small error). This is indeed important for many cryptographic applications that crucially require negligible error. Recently, Lewko [Lew19] used progress in additive combinatorics to improve the entropy requirement for low-error two-source extractors to roughly $4n/9$, which remains the state-of-art construction in the low-error regime. Ben-Aroya et al. [BACDTS19] constructed a weaker object known as a two-source condenser that works for polylogarithmic entropy and achieves negligible error. The output of a condenser is required to be close to a high-min-entropy distribution (instead of being close to uniform).

An impressive recent line of works by several researchers [BADTS16, CL16, Coh16b, Mek17, Coh17, Li17, Li19] built on the [CZ19] framework, to lower the min-entropy requirement of the 2-source extractor in the constant error regime. The state-of art construction by Li [Li19] requires min-entropy $C \log n(\log\log n)/\log\log\log n$, for some constant $C > 0$.

Ramsey graphs A major motivation for the line of work focusing on constructing near optimal two-source extractors in terms of min-entropy (with constant error) is that such extractors directly imply explicit Ramsey graphs, a major open problem raised by Erdös [Erd47] in extremal combinatorics. Recall that an undirected graph on N vertices is called a K-*Ramsey graph* if it does not contain any independent set or clique of size K. In 1930, Ramsey [Ram30] showed that there cannot exist a $(\log N)/2$-Ramsey graph on N nodes. In 1947, Erdös [Erd47] used the probabilistic method to prove the existence of $2\log N$-Ramsey graphs and posed it as a challenging open problem to explicitly construct such graphs.

It turns out that a (k,ϵ)-two-source extractor $\mathrm{2Ext} : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ with error $\epsilon < 1/2$ implies a $K/2$-Ramsey graph on N nodes, where $N = 2^n$ and $K = 2^k$. Thus, plugging in the two-source extractor from [Li19] implies a $(\log N)^{o(\log\log\log N)}$-Ramsey graph on N nodes. We refer the interested reader to [Coh19, CZ19] for more references and discussion on explicit constructions of Ramsey graphs.

Outline The main goal of this article is to provide an accessible introduction to the various techniques and notions that play a key role in the recent developments of two-source extractors. We use Section 2 to briefly discuss seeded extractors, a key component in all recent progress on explicit extractor constructions. In Section 3, we introduce seeded non-malleable extractors and give detailed sketches of the new ideas that go into their recent explicit constructions. In Section 4, we discuss resilient functions and their use in extracting from bit-fixing sources. Finally in Section 5, we sketch the construction of the two-source extractor from [CZ19] that relies on the all the ingredients discussed in previous sections.

2 Seeded extractors

Informally, a seeded extractor uses a short independent and uniform string, called a *seed*, to extract randomness from a weak source. This notion was introduced by Nisan and Zuckerman [NZ96] in the context of derandomizing space bounded computation.

Definition 2.1 (seeded extractor)**.** *A function* $\mathrm{Ext} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ *is called a* (k,ϵ)*-seeded extractor if the following holds: for any* (n,k)*-source* X*, we have*

$$\Delta(\mathrm{Ext}(X, U_m); U_m) \leq \epsilon.$$

Using the probabilistic method, it is possible to show that a random function is a seeded extractor with $d = \log(n-k) + 2\log(1/\epsilon) + O(1)$ and $m = k + d - 2\log(1/\epsilon) - O(1)$. Around three decades of research on seeded extractors have led to optimal constructions (up to constants) [LRVW03, GUV09, DKSS13] and some remarkable connections to other areas of theoretical computer science and mathematics [WZ93, Zuc96, Uma99, Tre01, MU02, Zuc06, GUV09, DW11]. We refer the reader to excellent surveys of Shaltiel [Sha04] and Vadhan [Vad12] (and references therein) for more details on explicit constructions of seeded extractors and their applications.

A strengthened notion is that of a *strong seeded* extractor, which can be informally described as requiring the output of the extractor and the seed to be uncorrelated. Before formally defining this, we introduce a couple of convenient notations.

Notation For arbitrary random variables A, B, C, we use the notation $\Delta((A;B)|C)$ to denote the quantity $\Delta((A,C);(B,C))$. For a sequence of random variables $A_1, \ldots, A_t$, we use the notation $\{A_i\}_{i=1}^{t}$ to denote the joint random variable $(A_1, \ldots, A_t)$.

Definition 2.2 (strong seeded extractor)**.** *A function* $\mathrm{Ext} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ *is called a* (k,ϵ)*-strong seeded extractor if the following holds: for any* (n,k)*-source* X*, we have*

$$\Delta((\mathrm{Ext}(X, U_d); U_m)|U_d) \leq \epsilon.$$

Many of the above mentioned constructions yield strong seeded extractors. In particular, [DKSS13] constructs a strong seeded extractor that has seed length $d = O(\log(n/\epsilon))$ and output length $m = (1 - o(1))k$.

Alternate view It is sometimes useful to view a seeded extractor as a collection of functions indexed by the seed. The following lemma presents this alternate view and reframes a strong seeded extractor in this view.

Claim 2.3. *Let* $\mathrm{Ext} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ *be a* (k,ϵ)*-strong seeded extractor. For each seed* $y \in \{0,1\}^d$*, define the function* $h_y : \{0,1\}^n \to \{0,1\}^m$ *as* $h_y(x) = \mathrm{Ext}(x,y)$*. Let* $D = 2^d$*.*

For any (n,k)*-source* X*, there exists a subset of seeds* $S_X \subset \{0,1\}^d$*,* $|S_X|/D \geq 1 - \sqrt{\epsilon}$ *such that for all* $y \in S_X$*, we have*

$$\Delta(h_y(X); U_m) \leq \sqrt{\epsilon}.$$

Proof. It follows from the definition of a strong seeded extractor that

$$\mathbb{E}_{y \sim U_d}[\Delta(h_y(X); U_m)] \leq \epsilon.$$

Thus, by a Markov argument, it follows that there exists $S_X \subset \{0,1\}^d$, $|S_X| \geq (1-\sqrt{\epsilon})D$ such that $\Delta(h_y(X); U_m) \leq \sqrt{\epsilon}$ for all $y \in S_X$. □

3 Non-malleable extractors

Dodis and Wichs [DW09] introduced the notion of a non-malleable extractor motivated by applications to a well-studied problem in cryptography, known as privacy amplification [BBR88, Mau92, BBCM95, MW97]. Informally, the output of a seeded non-malleable extractor looks uniform even conditioned on its output on a "correlated seed", where the correlated seed can be thought of as being produced by an adversary who has access to the seed. We present a more general definition of non-malleable extractors that was first studied by Cohen, Raz and Segev [CRS14].

Definition 3.1. *A function* $\text{Ext} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ *is called a* (t, k, ϵ)*-non-malleable extractor if the following holds: for any* (n, k)*-source* X*, any* t*-tuple of functions* $(f_1, \ldots, f_t)$*, where each* $f_i : \{0,1\}^d \to \{0,1\}^d$ *has no fixed points*[1]*, we have*

$$\Delta\left((\text{nmExt}(X, U_d); U_m) | \{\text{nmExt}(X, f_i(U_d))\}_{i=1}^t, U_d\right) \leq \epsilon.$$

When the parameters k, ϵ are clear from context, we sometimes drop these parameters from the notation and simply write t-non-malleable extractor. The case of $t = 1$ is the standard definition of a non-malleable extractor, as introduced in [DW09]. Further note that the degenerate setting of $t = 0$ recovers the definition of a strong seeded extractor.

Dodis and Wichs [DW09] used the probabilistic technique in a clever way to prove the existence of $(1, n, k)$-non-malleable extractors. This argument was extended in [BACD$^+$18] to prove the existence of (t, n, k)-non-malleable extractors with $k \geq (t+1)m + 2\log(1/\epsilon) + \log d + 4\log t + O(1)$ and $d \leq 2\log(1/\epsilon) + \log(n-k) + 2\log(t+1) + O(1)$.

Alternate view As in the case of seeded extractors, we can view non-malleable extractors as a collection of functions, indexed by the seed. We record an analogue of Claim 2.3 for non-malleable extractors that was proved in [CZ19]. Informally, it states that for any source X, there exists a large fraction of the functions in this collection that are almost t-wise independent.

Claim 3.2. *Let* $\text{nmExt} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ *be a* (t, k, ϵ)*-non-malleable extractor. For each seed* $y \in \{0,1\}^d$*, define the function* $h_y : \{0,1\}^n \to \{0,1\}^m$ *as* $h_y(x) = \text{nmExt}(x, y)$*.*

For any (n, k)*-source* X*, there exists a subset of seeds* $S_X \subset \{0,1\}^d$*,* $|S_X|/D \geq 1 - \sqrt{\epsilon}$ *such that for all such that for any distinct* $y_1, \ldots, y_t \in S_X$*, we have*

$$\Delta\left(\{h_{y_i}(X)\}_{i=1}^t; U_{mt}\right) \leq O(2^m \cdot t \cdot \sqrt{\epsilon}).$$

Proof sketch. Define a "bad" set of seeds as follows:

$$BAD = \{y \in \{0,1\}^d : \exists \text{ distinct } y_1, \ldots, y_t \in \{0,1\}^d \setminus \{y\}, \Delta\left((h_y(X); U_m) | \{h_{y_i}(X)\}_{i=1}^t\right) > \sqrt{\epsilon}\}$$

[1] For a function $f : \Omega \to \Omega$, we say that $x \in \Omega$ is a fixed point (of f) if $f(x) = x$.

The idea is to bound the size of BAD using the fact that nmExt is a (t, k, ϵ)-non-malleable extractor. In particular, define t adversarial functions $f_1, \ldots, f_t$ as follows: for any $y \in BAD$, and $i \in [t]$, set $f_i(y) = y_i$. It now follows that

$$\Delta((\text{nmExt}(X, U_d); U_m) | \{\text{nmExt}(X, f_i(U_d))\}_{i=1}^t, U_d) \geq \sqrt{\epsilon} \cdot |BAD|/D,$$

and thus $|BAD| \leq \sqrt{\epsilon}D$. Setting $S_X = \{0,1\}^d \backslash BAD$, the lemma can now be proved using standard probability lemmas. We skip the details here and refer the reader to Lemma 2.17 in [CZ19]. □

3.1 Explicit constructions

The task of constructing non-malleable extractors seemed quite challenging even for the simple case of $t = 1$ and $k = 0.99n$. The first explicit (t, k, ϵ)-non-malleable extractor was constructed by Dodis, Li, Wooley and Zuckerman [DLWZ14]. Their construction worked for $t = 1$, $k \geq (1/2 + \delta)n$ and $\epsilon = 2^{-\Omega(n)}$, for any constant $\delta > 0$. Subsequently, Cohen, Raz and Segev [CRS14] constructed a non-malleable extractor for general t but still required $k \geq (1/2 + \delta)n$. For the case of $t = 1$, Li [Li17] improved the entropy requirement to $k \geq (1/2 - \gamma)n$ for some tiny constant $\gamma > 0$.

A common feature of the non-malleable extractor constructions in [DLWZ14, CRS14, Li17] is that they are based on existing constructions of two-source extractors. For instance, Dodis et al. [DLWZ14] show that the Paley graph extractor, that was introduced in [CG88], is a non-malleable extractor. Cohen et al. [CRS14] proved that the two-source extractor constructed by Raz [Raz05] is a non-malleable extractor, and Li [Li17] constructed a non-malleable extractor by adapting Bourgain's two-source extractor [Bou05]. The best available explicit two-source extractor at that time was due to Bourgain [Bou05] that required min-entropy rate $\geq (1/2 - \delta)$ in each source, and hence it appeared to be a dead-end to pursue this line of attack (of proving non-malleability properties of an existing two-source extractor) to construct better non-malleable extractors.

Chattopadhyay, Goyal, and Li [CGL16] introduced a new framework (which, for the rest of this article, we call as the CGL framework) for constructing non-malleable extractors, and gave explicit t-non-malleable extractors that work for $k \geq c \cdot t \cdot (\log(n/\epsilon))^2$, for some constant $c > 0$. In particular, for $t = 1$, this provided the first explicit non-malleable extractor that could handle polylogarithmic min-entropy, providing an exponential improvement over prior work (described in the previous paragraph). Subsequent refinements and modifications of this framework have led to near optimal non-malleable extractors. In particular, the state-of-art explicit non-malleable extractor [Li19] works for $k \geq t \cdot (\log \log n + \log(1/\epsilon) \cdot o(\log \log(1/\epsilon)))$. We focus on presenting the CGL framework.

3.1.1 The CGL framework for constructing non-malleable extractors

The CGL framework relies on two new pseudorandom objects that we introduce in this section. We note that [CGL16] used these pseudorandom objects implicitly, and Cohen [Coh16c] distilled the ideas of this framework and explicitly defined these objects.

Advice correlation breakers Informally, a correlation breaker uses some independent randomness to "destroy correlation" that may exist between a sequence of random variables. More formally, the task of a correlation breaker CB : $\{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ can be formalized as follows: Let $X_1, \ldots, X_t$ be a sequence of (possibly) correlated random variables, with each X_i supported on

$\{0,1\}^n$. Further suppose there is an $\ell \in [t]$ such that X_ℓ is "good", i.e., X_ℓ is an (n,k)-source. Let Y be a uniform independent seed. We then require

$$\Delta\left((\mathrm{CB}(X_\ell, Y); U_m) | \{\mathrm{CB}(X_j, Y)\}_{j\in[t]\setminus\{\ell\}}, Y\right) < \epsilon.$$

It is not hard to see that such a function CB cannot exist in this generality, with a simple counterexample being that all the X_i's are the same random variable. However, it turns out that one can fix this problem by additionally supplying the correlation breaker with some "advice".

We now record a formal definition of an *advice correlation breaker.*

Definition 3.3 (Advice correlation breaker)**.** *A function* $\mathrm{ACB} : \{0,1\}^n \times \{0,1\}^d \times \{0,1\}^a \to \{0,1\}^m$ *is called a* (t,k,ϵ)*-ACB if the following holds:*

- *Correlated variables: let* $\{X_i\}_{i=1}^t$ *be any sequence of (possibly correlated) random variables, each supported on* $\{0,1\}^n$*. Suppose that there exists* $\ell \in [t]$ *such that* X_ℓ *is an* (n,k)*-source.*

- *Independent randomness: let* $\{Y_i\}_{i=1}^t$ *be another sequence of random variables such that* Y_ℓ *is uniform (on* $\{0,1\}^d$*). Further suppose* $\{X_i\}_{i=1}^t$ *and* $\{Y_i\}_{i=1}^t$ *are independent random variables.*

- *Advice strings: let* $\alpha_1, \dots \alpha_t \in \{0,1\}^a$ *be such that* $\alpha_\ell \neq \alpha_j$ *for all* $j \in [t] \setminus \{\ell\}$*.*

Then,

$$\Delta\left((\mathrm{ACB}(X_\ell, Y_\ell, \alpha_\ell); U_m) | \{\mathrm{ACB}(X_j, Y_j, \alpha_j)\}_{j\in[t]\setminus\{\ell\}}, \{Y_i\}_{i=1}^t\right) < \epsilon.$$

The first construction of an advice correlation breaker was given in [CGL16], relying on a beautiful construction known as the *flip flop* construction, introduced by Cohen [Coh16a]. Indeed a crucial observation in [CGL16] was that the flip flop construction is just an advice correlation breaker that works for one bit of advice (i.e., $a = 1$). We note that techniques introduced by Li [Li13] can also be adapted to give an alternate construction of an advice correlation breaker that works for one bit of advice. We think that the flip flop construction is easier to digest, and will focus on it here.

The flip flop construction makes clever use of a powerful technique introduced by Dziembowski and Pietrzak [DP07] known as *alternating extraction.* We note that all existing constructions of advice correlation breakers rely on alternating extraction (which, as we will see below, composes seeded extractors in an interesting way). We believe it should be possible to construct such objects from more elementary techniques, and record it as an open question.

Open Question 3.4. *Find a construction of an advice correlation breaker, even for* $a = 1$ *and* $t = 2$*, that does not rely on alternating extraction.*

We now describe the method of alternating extraction and then sketch the flip flop construction.

Definition 3.5 (Alternating extraction)**.** *The setup is the following:*

- *There are two parties, Quentin with access to a* (n_q, k_q)*-source* X*, and Wendy with access to a* (n_w, k_w)*-source* Y *and a seed* S_0 *that is uniform on* $\{0,1\}^d$*. The distributions* X *and* (Y, S_0) *are independent.*

- *Quentin and Wendy are equipped with* (k,ϵ)*-strong seeded extractors* $\mathrm{Ext}_q : \{0,1\}^{n_q} \times \{0,1\}^d \to \{0,1\}^d$ *and* $\mathrm{Ext}_w : \{0,1\}^{n_w} \times \{0,1\}^d \to \{0,1\}^d$ *respectively.*

Given a parameter h, the alternating extraction protocol consists of the following interactive protocol between Quentin and Wendy:

- *Wendy starts the interaction by sending her seed S_0 to Quentin. Quentin uses the seed S_0 to extract a new seed R_0 from the source X using the strong seeded extractor Ext_q, i.e., $R_0 = \mathrm{Ext}_q(X, S_0)$. Quentin now sends back the seed R_0 to Wendy to end this round of interaction.*

- *The next round starts with Wendy creating the seed $S_1 = \mathrm{Ext}_w(Y, R_0)$, and the interaction continues in this way. The number of rounds of interaction is given by the parameter h.*

Thus, the transcript of the communication between Quentin and Wendy is the following sequence of random variables:

$$S_0, R_0 = \mathrm{Ext}_q(X, S_0), S_1 = \mathrm{Ext}_w(Y, R_0), \ldots, S_h = \mathrm{Ext}_w(Y, R_{h-1}), R_h = \mathrm{Ext}_q(X, S_h).$$

Informally, the alternating extraction protocol enjoys the property that at any point during the interaction, a newly created seed (i.e, S_i or R_i) is close to uniform on a typical fixing of the interaction up to this point. We record this in the following claim.

Claim 3.6. *Assume that $k_w, k_q \geq k + 10hd + 2\log(1/\epsilon)$. Then, for all $i \leq h$, we have*

$$\Delta\left((R_i; U_d) | \{S_j\}_{j \leq i}, \{R_j\}_{j<i}, Y\right) < O(h\epsilon), \quad \textit{and} \;\; \Delta\left((S_i; U_d) | \{S_j\}_{j<i}, \{R_j\}_{j<i}, X\right) < O(h\epsilon).$$

We only briefly sketch the main ideas here and refer the reader to Appendix E in [DW09] for a formal proof. The proof goes via induction on i. First consider the base case of $i = 0$. Clearly S_0 is uniform and is independent of X. Thus, even conditioned on X, the random variable S_0 remains uniform. Further, since Ext_q is a strong seeded extractor, it follows that for a typical fixing of $S_0 = s_0$, we have that $R_0 = \mathrm{Ext}_q(X, s_0)$ is ϵ-close to uniform. Further note that on fixing S_0, the random variable R_0 is now a deterministic function of X. Thus, we can fix Y as well without affecting the distribution of R_0. This completes the base case. The inductive step can be proved using similar arguments and we skip it here.

It turns out that the alternating extraction protocol satisfies a much stronger property. To state this, we consider the more general setup, some variant of which has been considered in various works [DW09, Li13, Li15a, Coh16a, CGL16].

- Let $\{X_i\}_{i=1}^t$, $\{Y_i\}_{i=1}^t$, $\{S_{i,0}\}_{i=1}^t$ be sequences of random variables, each X_i on $\{0,1\}^{n_q}$, each Y_i on $\{0,1\}^{n_w}$ and each $S_{i,0}$ on $\{0,1\}^d$ such that for some $\ell \in [t]$, X_ℓ is an (n_q, k_q)-source, Y_ℓ is an (n_w, k_w)-source and $S_{\ell,0}$ is uniform on $\{0,1\}^d$. Assume that the random variables $\{X_i\}_{i=1}^t$ and $\left(\{Y_i\}_{i=1}^t, \{S_{i,0}\}_{i=1}^t\right)$ are independent.

- As before, Quentin and Wendy are equipped with (k, ϵ)-strong seeded extractors $\mathrm{Ext}_q : \{0,1\}^{n_q} \times \{0,1\}^d \to \{0,1\}^d$ and $\mathrm{Ext}_w \{0,1\}^{n_w} \times \{0,1\}^d \to \{0,1\}^d$, respectively.

- For $i \in [t]$, Quentin and Wendy produce the following transcript while executing the alternating extraction protocol for h rounds with access X_i and $(Y_i, S_{i,0})$, respectively:

$$S_{i,0}, R_{i,0} = \mathrm{Ext}_q(X_i, S_{i,0}), S_{i,1} = \mathrm{Ext}_w(Y_i, R_{i,0}), \ldots, S_{i,h} = \mathrm{Ext}_w(Y_i, R_{i,h-1}), R_{i,h} = \mathrm{Ext}_q(X_i, S_{i,h}).$$

Claim 3.7. *Assume that $k_w, k_q \geq k + 10htd + 2\log(1/\epsilon)$. Then, for all $i \leq h$, we have*

$$\Delta\left((R_{\ell,i}; U_d)|\{S_{e,j}\}_{e\in[t]\setminus\{\ell\},j\leq i}, \{R_{e,j}\}_{e\in[t]\setminus\{\ell\},j<i}, \{Y_j\}_{j\in[t]}\right) < O(h\epsilon),$$

and

$$\Delta\left((S_{\ell,i}; U_d)|\{R_{e,j}\}_{e\in[t]\setminus\{\ell\},j<i}, \{S_{e,j}\}_{e\in[t]\setminus\{\ell\},j<i}, \{X_j\}_{j\in[t]}\right) < O(h\epsilon).$$

The proof of the above claim uses inductive arguments that are similar to that sketched for Claim 3.6. We skip the proof and refer the interested reader to Lemma 4.1 in [Li13] for more details.

We are now finally ready to describe the flip flop construction. Instead of directly presenting the construction, we try to motivate it in a natural way, given the above properties of alternating extraction protocols. Recall that the flip flop construction is an advice correlation breaker that uses one bit of advice. Thus, we want to construct a function $\mathrm{FF} : \{0,1\}^n \times \{0,1\}^d \times \{0,1\} \to \{0,1\}^m$ such that if:

- $\{X_i\}_{i=1}^t$ is any sequence of (possibly correlated) random variables, each supported on $\{0,1\}^n$ and there exists $\ell \in [t]$ such that X_ℓ is an (n,k)-source.
- $\{Y_i\}_{i=1}^t$ is another sequence of random variables such that Y_ℓ is uniform (on $\{0,1\}^d$), and such that $\{X_i\}_{i=1}^t$ and $\{Y_i\}_{i=1}^t$ are independent random variables,
- $\alpha_1, \ldots \alpha_t \in \{0,1\}$ are such that $\alpha_\ell \neq \alpha_j$ for all $j \in [t] \setminus \{\ell\}$.

then

$$\Delta\left((\mathrm{FF}(X_\ell, Y_\ell, \alpha_\ell); U_m)|\{\mathrm{FF}(X_j, Y_j, \alpha_j)\}_{j\in[t]\setminus\{\ell\}}, \{Y_i\}_{i=1}^t\right) < \epsilon.$$

Here is an initial attempt to construct FF: On input (X_i, Y_i, α_i), we let $S_{i,0}$ denote the prefix Y_i of length $d' = d/10$. Now let Quentin and Wendy play two rounds of alternating extraction using X_i and $(Y_i, S_{i,0})$ respectively to produce random variables $S_{i,0}, R_{i,0}, S_{i,1}, R_{i,1}$. Finally, define the output of FF as R_{i,α_i}.

We claim that this works in the case when $\alpha_\ell = 1$. Indeed note that since $\alpha_\ell \neq \alpha_j$ for all $j \in [t] \setminus \{\ell\}$, it must be that $\alpha_j = 0$. Thus, $\mathrm{FF}(X_\ell, Y_\ell, \alpha_\ell) = R_{\ell,1}$ and for all $j \in [t] \setminus \{\ell\}$, $\mathrm{FF}(X_j, Y_j, \alpha_j) = R_{j,0}$. Thus, we arrive at the desired conclusion using Claim 3.7.

However, it may be the case that $\alpha_\ell = 0$, in which case this construction fails to work. It turns out that this can be fixed by doing two additional rounds of alternating extraction, leading to the actual flip flop primitive.

We sketch the final construction of FF, and refer the reader to [Coh16a,CGL16] for more details of the proof. On input (X_i, Y_i, α_i), as described above, we produce the random variable R_{i,α_i}. Now, define $Y_i' = \mathrm{Ext}(Y, R_{i,\alpha_i})$ for an appropriately chosen strong seeded extractor Ext, and let $S_{i,0}'$ be the prefix of $Y_{i,0}'$ of length d. Now Quentin and Wendy runs two more rounds of alternating extraction using X_i and $(Y_i', S_{i,0}')$ respectively to produce random variables $S_{i,0}', R_{i,0}', S_{i,1}', R_{i,1}'$. Define the output of FF to be $R_{i,1-\alpha_i}'$.

The intuition for why this works is the following: in the case when $\alpha_\ell = 1$, the first two rounds of alternating extraction leads to the breaking of correlation. One can then show that additional rounds of alternating extraction does not affect this outcome. Further in the case when $\alpha_\ell = 0$, the final two rounds of alternating extraction leads to the desired outcome.

Theorem 3.8. *There exist constants $c_1, c_2 > 0$ and an explicit (k,t,ϵ)-advice correlation breaker $\mathrm{FF} : \{0,1\}^n \times \{0,1\}^d \times \{0,1\} \to \{0,1\}^m$ for $k \geq c_1 t \cdot (m + \log(n/\epsilon))$ and $d = c_2 t \cdot \log(n/\epsilon)$.*

Recall that our goal was to construct an advice correlation breaker that works for advice strings of length a, for any integer $a > 0$. It was shown in [CGL16] that this can be achieved by composing the FF construction in a natural way. We now sketch this construction.

The setup is exactly as described in the case of FF, except that the advice strings $\alpha_1, \ldots, \alpha_t$ are now bit strings of length a. We want to construct $\mathrm{ACB} : \{0,1\}^n \times \{0,1\}^d \times \{0,1\}^a \to \{0,1\}^m$ such that

$$\Delta\left((\mathrm{ACB}(X_\ell, Y_\ell, \alpha_\ell); U_m) | \{\mathrm{ACB}(X_j, Y_j, s_j)\}_{j \in [t] \setminus \{\ell\}}, \{Y_i\}_{i=1}^t\right) < \epsilon.$$

For any string z, let $(z)_i$ denote the i'th bit of z. The following is a sketch of the ACB construction:

- Let $Z_{i,1}$ be the prefix of Y_i of length $n_1 = c' \cdot t \log(n/\epsilon)$ for some large enough constant c'. Now let $W_{i,1} = \mathrm{FF}(X_i, Z_{i,1}, (\alpha_i)_1)$, where FF is the function from Theorem 3.8 with output length $m_1 = O(\log(n/\epsilon))$.
- Using $W_{i,1}$, extract $Z_{i,2} = \mathrm{Ext}(Y_i, W_{i,1})$, for a suitably chosen Ext with output length n_1. Now, define $W_{i,2} = \mathrm{FF}(X_i, Z_{i,2}, (\alpha_i)_2)$ and $Z_{i,3} = \mathrm{Ext}(Y_i, W_{i,2})$. Continuing this way, we create the sequence of random variables $\{W_{i,j}\}_{j=1}^a$. Finally let $\mathrm{Ext1}(X_i, W_{i,a})$ be the output of ACB, for a suitably chosen strong seeded extractor Ext'.

We sketch some intuition for the correctness of the above construction, and refer to [CGL16] for more details. The idea is the following: since the advice string $\alpha_\ell \neq \alpha_j$ for any $j \in [t] \setminus \{\ell\}$, there exists index $f(j) \in [a]$ such that α_ℓ and α_j differ on index $f(j)$. Informally, we expect any correlation between X_ℓ and X_j to be broken at the $f(j)$'th round of applying FF, i.e., the random variable $W_{\ell,f(j)}$ is close to uniform for a typical fixing of $W_{j,f(j)}$ (since FF is an advice correlation breaker that works for one bit of advice). This intuition turns out to be true, and can be formalized as follows: for $i \in [a]$, define $\mathrm{IND}_i = \{j \in [t] \setminus \{\ell\} : f(j) \leq i\}$. Then, for any $i \leq a$,

$$\Delta((W_{\ell,i}; U_{m_1}) | \{W_{i,j}\}_{j \in \mathrm{IND}_i}, \{Y_j\}_{j \in [t]}) < O(i \cdot \epsilon).$$

Noting that $\mathrm{IND}_a = [t] \setminus \{\ell\}$, the correctness of the ACB construction follows in a straightforward way from the above guarantee. By appropriate choice of parameters and seeded extractors, one obtains the following explicit advice correlation breaker.

Theorem 3.9. *There exist constants $c_1, c_2 > 0$ and an explicit (k, t, ϵ)-advice correlation breaker* $\mathrm{ACB} : \{0,1\}^n \times \{0,1\}^d \times \{0,1\}^a \to \{0,1\}^m$ *for $k \geq c_1 \cdot a \cdot t \cdot (m + \log(n/\epsilon))$ and $d = c_2 \cdot a \cdot t^2 \cdot \log(n/\epsilon)$.*

Subsequent works [Coh16b, CL16, Coh17, Li17, Li19] further improved parameters of the above theorem. However the dependence of seed length on t is at least linear in all these constructions, and it seems to be a bottleneck in current techniques (that are based on alternating extraction). On the other hand, an application of the probabilistic method shows the existence of such advice correlation breakers with seed length that is logarithmic in t. As shown in [BACD+18], progress in this direction will have applications in constructing low-error two-source extractors—a major open question in pseudorandomness. We record this as an open problem.

Open Question 3.10. *Construct an explicit (t, k, ϵ)-advice correlation breaker* $\mathrm{ACB} : \{0,1\}^n \times \{0,1\}^d \times \{0,1\}^a \to \{0,1\}^m$ *with $d = o(t)$.*

We believe that a good starting point for the above question is to make progress on Open Question 3.4.

We now introduce the second pseudorandom object that plays a key role in the CGL framework. Informally, the motivation for this object is supply ACB with the necessary "advice" that it requires to function.

Advice generators This object can be viewed as a weakening of a non-malleable extractor. Consider the following setting: let X be an (n, k)-source and let Y, Y' be arbitrarily correlated random variables on $\{0,1\}^d$ such that

- X and (Y, Y') are independent,
- Y is uniformly distributed on $\{0,1\}^d$,
- For any y, the support of $Y'|Y = y$ does not contain y.

A $(1, k, \epsilon)$-non-malleable $\mathrm{nmExt} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ then has the property that $\mathrm{nmExt}(X, Y)$ looks close to uniform on a typical fixing of $\mathrm{nmExt}(X, Y')$. An advice generator $\mathrm{Adv} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^a$ instead requires the weaker property that for most $\alpha \in \{0,1\}^a$, the support of $\mathrm{Adv}(X, Y)|\mathrm{Adv}(X, Y') = \alpha$ does not contain α.

Given the above (informal) definition, it is trivial to construct an advice generator if we allow $a \geq d$. Indeed, one can just set to output of $\mathrm{Adv}(x, y) = y$. Thus, we are interested in achieving output length a that is much smaller than d.

We now define advice generators more generally and then sketch a construction from [CGL16].

Definition 3.11. *A function* $\mathrm{Adv} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ *is called a* (t, k, ϵ)*-advice generator if the following holds:*

- *Let* $\{X_i\}_{i\in[t]}$, $\{Y_i\}_{i\in[t]}$ *be two sequences of random variables on* $\{0,1\}^n$ *and* $\{0,1\}^d$ *respectively.*
- *Suppose for some* $\ell \in [t]$, X_ℓ *is an* (n, k)*-source and* Y_ℓ *is uniform on* $\{0,1\}^d$.
- *Suppose for any* $y_\ell \in \{0,1\}^d$, *and any* $j \in [t] \setminus \{\ell\}$, *the support of* $Y_j|Y_\ell = y_\ell$ *does not contain* y_ℓ.

Then, with probability at least $1 - \epsilon$ *over fixing* $\{\mathrm{Adv}(X_i, Y_i)\}_{i\in[t]\setminus\{\ell\}}, \{Y_i\}_{i\in[t]}$, *we have, for any* $i \in [t] \setminus \{\ell\}$,

$$\mathrm{Adv}(X_\ell, Y_\ell) \neq \mathrm{Adv}(X_i, Y_i).$$

The main idea in [CGL16] for constructing an advice generator Adv is quite simple, and involves the following two steps: On input X_i, Y_i,

1. encode Y_i to Y_i' using a good error correcting code (i.e., constant rate and distance), and
2. sample a small subset of coordinates in Y_i' using X.

Since by assumption we have that $Y_\ell \neq Y_j$ for any $j \in [t] \setminus \{\ell\}$, Step 1 ensures that the Hamming distance between Y_ℓ' and Y_j' is large. Thus, even sampling a small set of coordinates in Step 2 ensures that with high probability we sample a coordinate on which Y_ℓ' and Y_j' are distinct.

It turns out that for using advice generators to construct non-malleable extractors in the CGL framework, one needs the additional property that X_ℓ contains enough min-entropy even conditioned on a typical fixing of $\{\mathrm{Adv}(X_i, Y_i)\}_{i\in[t]}, \{Y_i\}_{i\in[t]}$. This leads to some additional subtlety in executing Step 2 and we refer the interested reader to the actual construction in [CGL16] for more details.

We record the parameters achieved by the construction in [CGL16].

Theorem 3.12. *There exist constants $c_1, c_2 > 0$ and an explicit (k, t, ϵ)-advice generator* $\mathrm{Adv} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^a$ *with $k \geq c_1 t \log(n/\epsilon)$, $d = c_2 t \log(n/\epsilon)$, and $a = O(\log(n/\epsilon))$.*

Subsequent work of Cohen [Coh16b] improved the advice length a to $O(\log(1/\epsilon))$ for the case of $t = O(1)$, which is optimal up to constants.

Explicit non-malleable extractors We are now ready to present the CGL framework [CGL16] for constructing a (t, k, ϵ)-non-malleable extractor $\mathrm{nmExt} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$. We use the following ingredients:

- Let $\mathrm{ACB} : \{0,1\}^n \times \{0,1\}^d \times \{0,1\}^a \to \{0,1\}^m$ be a $(t+1, k/2, , \epsilon)$-advice correlation breaker
- Let $\mathrm{Adv} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^a$ be an $(t+1, k, \epsilon)$-advice generator.

Define

$$\mathrm{nmExt}(X, Y) = \mathrm{ACB}(X, Y, \mathrm{Adv}(X, Y)).$$

Assume X is an (n, k)-source and Y is a uniform independent seed of length d. Further, for $i \in [t]$, let $f_i : \{0,1\}^d \to \{0,1\}^d$ be a tampering function with no fixed points.

We want to prove that

$$\Delta((\mathrm{nmExt}(X, Y); U_m) | \{\mathrm{nmExt}(X, f_i(Y))\}_{i=1}^t, Y) < C \cdot \epsilon,$$

for some constant $C > 0$.

For ease of notation, define $Y_i = f_i(Y)$. Note that by assumption, $Y \neq Y_i$ for any $i \in [t]$. Thus, using the fact that Adv is a $(t+1, k, \epsilon)$-advice generator, we have that with probability $1 - O(\epsilon)$ over fixing $\{\mathrm{Adv}(X, Y_i)\}_{i\in[t]}$, it holds that for any $i \in [t]$, $\mathrm{Adv}(X, Y) \neq \mathrm{Adv}(X, Y_i)$. We also require the stronger property of Adv that with probability $1 - O(\epsilon)$ over fixing $\big(\mathrm{Adv}(X, Y), \{\mathrm{Adv}(X, Y_i)\}_{i\in[t]}\big)$,

- X remains independent of $Y, \{Y_i\}_{i\in[t]}$, and
- X is $O(\epsilon)$-close to a distribution with min-entropy at least $k/2$ and Y is $O(\epsilon)$-close to a distribution with min-entropy $d(1 - o(1))$.

It turns out the construction of Adv in [CGL16] (see Theorem 3.12) indeed satisfies the above properties assuming $k > c_1 at$ and $d = c_2 at^2$, for large enough constants $c_1, c_2 > 0$. Thus, assume a fixing of $\mathrm{Adv}(X, Y) = \alpha$ and $\{\mathrm{Adv}(X, Y_i) = \alpha_i\}_{i\in[t]}$ such that for all $i \in [t]$, we have $\alpha \neq \alpha_i$. The proof now follows almost directly from the fact that ACB is an advice correlation breaker.

The attentive reader may notice the following problem: for the ACB to work, we require Y to be uniform, but we are only guaranteed that it has min-entropy rate $1 - o(1)$. It turns out that this is not much of a problem, and one can get around this by paying a small price in the error of the ACB. We refer to [CGL16] for more details of the proof.

By composing the advice correlation breaker and advice generator constructed in [CGL16], they obtained the following theorem.

Theorem 3.13. *There exist constants $c_1, c_2 > 0$ and an explicit (t, k, ϵ)-non-malleable extractor $\mathrm{nmExt} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ for $k \geq c_1 t(m + \log(n/\epsilon))$ and $d = c_2 t^2 \log^2(n/\epsilon)$.*

Subsequent work [Coh16b, CL16, Coh17, Li17, Li19] gave better constructions of advice correlation breakers and advice generators, with the state-of-art construction [Li19] yielding a (t, k, ϵ)-non-malleable extractor for $k \geq c_1 t(m + \log\log(n) + \log(1/\epsilon) \cdot o(\log\log(1/\epsilon)))$ and $d = c_2 t^2(\log\log(n) + \log(1/\epsilon) \cdot o(\log\log(1/\epsilon)))$.

Similar to the case of advice correlation breakers, we do not have explicit t-non-malleable extractors with seed length with sublinear dependence on t. Non-explicitly it is known that logarithmic dependence on t suffice [BACD+18], while it appears to be a fundamental bottleneck of techniques that are based on alternating extraction to break the linear barrier. It was proved in [BACD+18] that progress in this direction will lead to better low-error two-source extractors.

4 Resilient functions and bit-fixing extractors

An important ingredient in the construction of two-source extractors in [CZ19] is the seemingly unrelated notion of a resilient function, which arises in distributed computing [BOL85]. Informally, a (q, δ)-resilient function $f : \{0,1\}^n \to \{0,1\}$ has the property that no subset of coordinates of size at most q can "influence" the outcome of the function by more than δ. The influence of a subset of coordinates S (on f) is defined to be the probability that on randomly fixing values of coordinates outside S, the value of the function f is still undetermined (and thus the coordinates in S determine the outcome of f).

We record more general definitions of influence and resilient functions. First, we recall that a distribution X on $\{0,1\}^n$ is called (t, γ)-wise independent if for any set $S \subset [n]$, $|S| = t$, we have $\Delta(\{X_i\}_{i \in S}, U_t) \leq \gamma$, where X_i denotes the i'th bit of X. Such distributions are referred to as almost t-wise independent distributions.

Definition 4.1. *Let $I_{Q,D}(f)$ denote the probability that f is undetermined when the variables outside Q are set by sampling from the distribution D. Now, define $I_{Q,t,\gamma}(f) = \max_{D \in D_{t,\gamma}} I_{Q,D}(f)$, where $D_{t,\gamma}$ denotes the family of all (t, γ)-wise independent distributions. Finally, define $I_{q,t,\gamma}(f)$ as the maximum value of $I_{Q,t,\gamma}(f)$ over all subsets of coordinates Q of size q.*

When $\gamma = 0$, we simply drop γ from the notations and use $I_{Q,t}(f)$ and $I_{q,t}$. Further, when $t = n$, (i.e., all the bits outside Q are uniform and independent), we drop the parameter t and use the notations $I_Q(f)$ and $I_q(f)$.

Definition 4.2. *A function $f : \{0,1\}^n \to \{0,1\}$ is (q, t, γ, ϵ)-resilient if $I_{q,t,\gamma}(f) \leq \epsilon$. Similarly, f is (q, t, ϵ)-resilient if $I_{q,t}(f) \leq \epsilon$ and is (q, ϵ)-resilient if $I_q(f) \leq \epsilon$.*

We record a useful claim that lets us bound $I_{Q,t,\gamma}$ from a bound on $I_{Q,t}$. This follows from a result in [AGM03] which states that every almost t-wise independent distribution is close to some t-wise independent distribution.

Claim 4.3. *Suppose that $f : \{0,1\}^n \to \{0,1\}$ is (q, t, ϵ)-resilient function. Then, for any $\gamma > 0$, f is a $(q, t, \gamma, \epsilon + \gamma \cdot n^t)$-resilient function.*

Explicit resilient functions The usual setting in which resilient functions are studied in distributed computing assume that the "good bits" (i.e., bits outside Q) are completely uniform and independent. Thus, the most well studied notion is that of (q,ϵ)-resilient functions. It is known that the MAJORITY function (which we will denote by Maj) is a $(q, O(q/\sqrt{n}))$-resilient function. It turns out that there are much better resilient functions than Maj. Ajtai and Linial [AL93] gave a probabilistic construction of Boolean functions that are $(q, O(q \cdot (\log n)^2/n))$-resilient, for all $q \leq n/\log^2 n$. (In fact, their construction is a distribution over constant depth circuits.) This is close to optimal, since by a result on Boolean functions [KKL89] it is known that for any for any Boolean function f, there exists a set of coordinates Q of size $cn/\log n$ with $I_Q(f) = \Omega(1)$.

However much less was known about (q,t,ϵ)-resilient functions. Viola [Vio14] proved that Maj is a $\left(q, t, O\left(\frac{q}{\sqrt{n}} + \frac{\log t}{t}\right)\right)$-resilient function for any t. Chattopadhyay and Zuckerman [CZ19] derandomized the probabilistic construction of Ajtai and Linial [AL93] to obtain the following result.

Theorem 4.4. *There exist constants $C, \delta > 0$ and an explicit function $f : \{0,1\}^n \to \{0,1\}$ that is $(q, \log^C n, O(q/n^{1-\delta}))$-resilient. Further, f is a monotone constant depth circuit and its bias is $1/n^{\Omega(1)}$.*

The above theorem also relied on the breakthrough result of Braverman [Bra10] that polylogarithmic independence fools constant depth circuits. Subsequently Meka [Mek17] improved the above result to exactly match the bound obtained in [AL93].

Non-oblivious bit-fixing sources The primary motivation for obtaining explicit resilient functions in [CZ19] was that they were trying to extract from non-oblivious bit-fixing (NOBF) sources. Informally, these are distributions which have hidden random coordinates, and the remaining coordinates arbitrarily depend on the values of the random coordinates. We define this more formally.

Definition 4.5 (NOBF sources). *A distribution V on $\{0,1\}^\ell$ is called a (q,t,γ)-NOBF source if there exists a subset $S \subset [\ell]$, $|S| \geq q$ such that for any $T \subset S$, $|T| = t$, we have $\Delta(\{V_i\}_{i \in T}, U_t) \leq \gamma$, where V_i is the i'th bit of V.*

It turns out that (almost) unbiased (q,t,γ,ϵ)-resilient functions are extractors for (q,t,γ)-NOBF sources. The intuition for this is that no set of coordinates of cardinality $\leq q$ are influential, and thus the set of "bad" coordinates in the NOBF source cannot bias the resilient function by a lot. The following claim formalizes this.

Lemma 4.6. *Let $f : \{0,1\}^n \to \{0,1\}$ be a Boolean function that is $(q,t,\gamma,\epsilon_1,)$-resilient. Further suppose that for any (t,γ)-wise independent distribution $\mathcal{D}$, $|\mathbb{E}_{\mathbf{x}\sim\mathcal{D}}[f(\mathbf{x})] - \frac{1}{2}| \leq \epsilon_2$. Then f is an extractor for (q,t,γ)-NOBF sources with error $\epsilon_1 + \epsilon_2$.*

We skip the proof here and refer to Lemma 2.9 in [CZ19] for more details.

5 An explicit two-source extractor

In this section we sketch the two-source extractor construction of Chattopadhyay and Zuckerman [CZ19] using ingredients developed in the previous sections. We recall their main result.

Theorem 5.1. *There exists a constant $C > 0$ and an explicit (k, ϵ)-two-source extractor* $2\mathrm{Ext} : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ *for $k \geq \log^C n$ and $\epsilon = 1/n^{\Omega(1)}$.*

Let X, Y be independent (n, k)-sources. Let $\mathrm{nmExt} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}$ be an explicit (t, k, ϵ_1)-non-malleable extractor (e.g., the explicit construction from Theorem 3.13 or any of the subsequently improved constructions). Let $D = 2^d$. The first idea is to use nmExt on X by "brute forcing" over all the seeds of the non-malleable extractor, i.e., define $Z_w = \mathrm{nmExt}(X, w)$ for $w \in \{0,1\}^d$. It follows from the alternate view of non-malleable extractors (Claim 3.2) that there exists $S_X \subset \{0,1\}^d$, $|S_X| \geq (1 - \sqrt{\epsilon_1})D$ such that for any $T \subset S_X$, $|T| = t$, we have

$$\Delta(\{Z_w : w \in T\}, U_t) \leq O(t\sqrt{\epsilon_1}).$$

Define Z to be the concatenation of all the Z_w's. It follows from the above discussion that Z is a $(\sqrt{\epsilon_1} \cdot D, t, O(t\sqrt{\epsilon_1}))$-NOBF source (see Definition 4.5) on $\{0,1\}^D$. Thus, a natural idea is to use an appropriate (q, t, ϵ_1)-resilient function, which as recorded in Lemma 4.6, is exactly an extractor for NOBF sources. This almost works except for the following problem: recall that by Claim 4.3, a (q, t, ϵ_1)-resilient function $f : \{0,1\}^D \to \{0,1\}$ is a $(q, t, \gamma, \gamma D^t + \epsilon_1)$-resilient function for any $\gamma > 0$. Hence if the bias of f is bounded by ϵ_1, then by Lemma 4.6 f can extract (one bit) from Z with error $\gamma D^t + 2\epsilon_1$, where $\gamma = O(t\sqrt{\epsilon_1})$. All this is fine, except that $D = 2^d$, where d is the seed length of nmExt and hence grows with ϵ_1. In fact, it can be shown that through known existing lower bounds on the seed length of non-malleable extractors, the term γD^t is always larger than 1.

It is in fact reassuring that the above does not work since then we would have constructed a 1-source extractor! This is where the second source Y comes in. The idea is to use Y to sample a small set (with cardinality that is polynomial in n) of pseudorandom coordinates in Z. This can be accomplished using standard techniques introduced by Zuckerman [Zuc97] of sampling using weak sources. We skip the details of the sampling step here and refer the interested reader to [CZ19].

As a result of this sampling, we obtain a new source Z' that is a $(\epsilon' D_1, t, \gamma = O(t\sqrt{\epsilon_1}))$-NOBF source on $D_1 = \mathrm{poly}(n)$ bits, where ϵ' is still of the same order as ϵ_1. Thus, now we can control the term γD_1^t since D_1 is now disentangled from the error parameter of nmExt (i.e., ϵ_1). By appropriate choice of parameters one obtains that Z' is a $(D_1^{1-\eta}, t = \mathrm{poly}(\log n), 1/D_1^{t+2})$-NOBF, for some small $\eta > 0$. Thus plugging in the explicit resilient function from Theorem 4.4, we obtain Theorem 5.1. This completes the sketch of the construction.

It can be shown that using the sampling technique from [Zuc97], the constant η is the above paragraph is smaller than $1/2$. Thus one cannot use Maj as the resilient function in this construction, and has to rely on the derandomization of the Ajtai-Linial function. Subsequent work of Ben-Aroya, Doron and Ta-Shma [BADTS16] improved on the framework of constructing two-source extractors described here and indeed uses the Maj function as their resilient function. Using Maj as the resilient function has the advantage that the parameter t can even be set to a constant (where as in [CZ19], it is required that $t = \mathrm{poly}(\log n)$). This added flexibility has been a crucial ingredient in the recent line of work, described in the introduction, of obtaining two-source extractors with near optimal parameters in the constant error regime.

We conclude by recording a couple of natural open questions. We would consider any progress on these questions to be very interesting.

Open Question 5.2. *Construct an explicit $(o(n), 1/n^{\omega(1)})$-two-source extractor.*

Recall that currently the best known explicit constructions of negligible error two-source constructions [Bou05, Lew19] require min-entropy close to $n/2$. In fact these extractor constructions are much simpler (though the analysis is based on sophisticated techniques from additive combinatorics) compared to the construction in [CZ19] and follow-up works. The following question is posed in hope of continuing the spirit of exhibiting extraction properties of functions that are simple to describe.

Open Question 5.3. *Give a simpler construction of a two-source extractor (even for constant error) that works for entropy* $0.1n$.

6 Acknowledgements

I thank Jesse Goodman and Jyun-Jie Liao for carefully reading an earlier version of this article, and sharing many thoughtful comments. I thank Lane Hemaspaandra for helpful comments and for the opportunity to write this article.

References

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

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

[BACD$^+$18] Avraham Ben-Aroya, Eshan Chattopadhyay, Dean Doron, Xin Li, and Amnon Ta-Shma. A new approach for constructing low-error, two-source extractors. In *33rd Computational Complexity Conference (CCC 2018)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2018.

[BACDTS19] Avraham Ben-Aroya, Gil Cohen, Dean Doron, and Amnon Ta-Shma. Two-source condensers with low error and small entropy gap via entropy-resilient functions. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2019)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2019.

[BADTS16] Avraham Ben-Aroya, Dean Doron, and Amnon Ta-Shma. Explicit two-source extractors for near-logarithmic min-entropy. In *Electronic Colloquium on Computational Complexity (ECCC)*, volume 23, page 88, 2016.

[BBCM95] Charles H Bennett, Gilles Brassard, Claude Crépeau, and Ueli M Maurer. Generalized privacy amplification. *IEEE Transactions on Information Theory*, 41(6):1915–1923, 1995.

[BBR88] Charles H Bennett, Gilles Brassard, and Jean-Marc Robert. Privacy amplification by public discussion. *SIAM journal on Computing*, 17(2):210–229, 1988.

[BIW06] Boaz Barak, Russell Impagliazzo, and Avi Wigderson. Extracting randomness using few independent sources. *SIAM Journal on Computing*, 36(4):1095–1118, 2006.

[Blu86] Manuel Blum. Independent unbiased coin flips from a correlated biased source—a finite state Markov chain. *Combinatorica*, 6(2):97–108, 1986.

[BOL85] Michael Ben-Or and Nathan Linial. Collective coin flipping, robust voting schemes and minima of Banzhaf values. In *26th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 408–416. IEEE, 1985.

[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.

[Bra10] Mark Braverman. Polylogarithmic independence fools AC^0 circuits. *Journal of the ACM (JACM)*, 57(5):1–10, 2010.

[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.

[CGL16] Eshan Chattopadhyay, Vipul Goyal, and Xin Li. Non-malleable extractors and codes, with their many tampered extensions. In *Proceedings of the forty-eighth annual ACM symposium on Theory of Computing*, pages 285–298, 2016.

[CL16] Eshan Chattopadhyay and Xin Li. Explicit non-malleable extractors, multi-source extractors, and almost optimal privacy amplification protocols. In *2016 IEEE 57th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 158–167. IEEE, 2016.

[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. Making the most of advice: New correlation breakers and their applications. In *2016 IEEE 57th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 188–196. IEEE, 2016.

[Coh16c] Gil Cohen. Non-malleable extractors – New tools and improved constructions. In *31st Conference on Computational Complexity (CCC 2016)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2016.

[Coh17] Gil Cohen. Towards optimal two-source extractors and ramsey graphs. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, pages 1157–1170, 2017.

[Coh19] Gil Cohen. Two-source dispersers for polylogarithmic entropy and improved Ramsey graphs. *SIAM Journal on Computing*, pages STOC16–30–STOC16–67, 2019.

[CRS14] Gil Cohen, Ran Raz, and Gil Segev. Nonmalleable extractors with short seeds and applications to privacy amplification. *SIAM Journal on Computing*, 43(2):450–476, 2014.

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

[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.

[DLWZ14] Yevgeniy Dodis, Xin Li, Trevor D Wooley, and David Zuckerman. Privacy amplification and nonmalleable extractors via character sums. *SIAM Journal on Computing*, 43(2):800–830, 2014.

[DOPS04] Yevgeniy Dodis, Shien Jin Ong, Manoj Prabhakaran, and Amit Sahai. On the (im)possibility of cryptography with imperfect randomness. In *45th Annual IEEE Symposium on Foundations of Computer Science*, pages 196–205. IEEE, 2004.

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

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

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

[Erd47] P. Erdős. Some remarks on the theory of graphs. *Bull. Amer. Math. Soc.*, 53(4):292–294, 04 1947.

[GUV09] Venkatesan Guruswami, Christopher Umans, and Salil Vadhan. Unbalanced expanders and randomness extractors from Parvaresh–Vardy codes. *Journal of the ACM (JACM)*, 56(4):1–34, 2009.

[Kar71] A.A. Karatsuba. On a certain arithmetic sum. *Soviet Math Dokl., 12, 1172-1174*, 1971.

[Kar91] AA Karatsuba. The distribution of values of Dirichlet characters on additive sequences. In *Doklady Acad. Sci. USSR*, volume 319, pages 543–545, 1991.

[KKL89] Jeff Kahn, Gil Kalai, and Nathan Linial. *The influence of variables on Boolean functions.* Citeseer, 1989.

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

[Li11] Xin Li. Improved constructions of three source extractors. In *2011 IEEE 26th Annual Conference on Computational Complexity*, pages 126–136. IEEE, 2011.

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

[Li15a] Xin Li. Non-malleable condensers for arbitrary min-entropy, and almost optimal protocols for privacy amplification. In *Theory of Cryptography Conference*, pages 502–531. Springer, 2015.

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

[Li16] Xin Li. Improved two-source extractors, and affine extractors for polylogarithmic entropy. In *2016 IEEE 57th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 168–177. IEEE, 2016.

[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*, pages 1144–1156, 2017.

[Li19] Xin Li. Non-malleable extractors and non-malleable codes: Partially optimal constructions. In *34th Computational Complexity Conference, CCC 2019, July 18-20, 2019, New Brunswick, NJ, USA*, pages 28:1–28:49, 2019.

[LRVW03] Chi-Jen Lu, Omer Reingold, Salil Vadhan, and Avi Wigderson. Extractors: Optimal up to constant factors. In *Proceedings of the thirty-fifth annual ACM symposium on Theory of computing*, pages 602–611, 2003.

[Mau92] Ueli M Maurer. Protocols for secret key agreement by public discussion based on common information. In *Annual International Cryptology Conference*, pages 461–470. Springer, 1992.

[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.

[MR10] Rajeev Motwani and Prabhakar Raghavan. *Randomized algorithms*. Chapman & Hall/CRC, 2010.

[MU02] Elchanan Mossel and Christopher Umans. On the complexity of approximating the VC dimension. *Journal of Computer and System Sciences*, 65(4):660–671, 2002.

[MW97] Ueli Maurer and Stefan Wolf. Privacy amplification secure against active adversaries. In *Annual International Cryptology Conference*, pages 307–321. Springer, 1997.

[NZ96] Noam Nisan and David Zuckerman. Randomness is linear in space. *Journal of Computer and System Sciences*, 52(1):43–52, 1996.

[Ram30] Frank P. Ramsey. On a problem of formal logic. *Proceedings of the London Mathematical Society, Series 2*, 30:264–286, 1930.

[Rao09] Anup Rao. Extractors for low-weight affine sources. In *2009 24th Annual IEEE Conference on Computational Complexity*, pages 95–101. IEEE, 2009.

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

[Sha04] Ronen Shaltiel. Recent developments in explicit constructions of extractors. In *Current Trends in Theoretical Computer Science: The Challenge of the New Century, Vol. 1: Algorithms and Complexity*, pages 189–228. World Scientific, 2004.

[SV86] Miklos Santha and Umesh V Vazirani. Generating quasi-random sequences from semi-random sources. *Journal of computer and system sciences*, 33(1):75–87, 1986.

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

[Uma99] Christopher Umans. Hardness of approximating Σ_2^p minimization problems. In *40th Annual Symposium on Foundations of Computer Science (Cat. No. 99CB37039)*, pages 465–474. IEEE, 1999.

[Vad12] Salil P Vadhan. Pseudorandomness. *Foundations and Trends® in Theoretical Computer Science*, 7(1–3):1–336, 2012.

[Vio14] Emanuele Viola. Extractors for circuit sources. *SIAM Journal on Computing*, 43(2):655–672, 2014.

[vN51] J. von Neumann. Various techniques used in connection with random digits. *Applied Math Series*, 12:36–38, 1951. Notes by G.E. Forsythe, National Bureau of Standards. Reprinted in *Von Neumann's Collected Works*, 5:768-770, 1963.

[WZ93] Avi Wigderson and David Zuckerman. Expanders that beat the eigenvalue bound: Explicit construction and applications. In *Proceedings of the twenty-fifth annual ACM symposium on Theory of computing*, pages 245–251, 1993.

[Zuc90] David Zuckerman. General weak random sources. In *Proceedings [1990] 31st Annual Symposium on Foundations of Computer Science*, pages 534–543. IEEE, 1990.

[Zuc96] David Zuckerman. On unapproximable versions of NP-complete problems. *SIAM Journal on Computing*, 25(6):1293–1304, 1996.

[Zuc97] David Zuckerman. Randomness-optimal oblivious sampling. *Random Structures & Algorithms*, 11(4):345–367, 1997.

[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, 2006.