Electronic Colloquium on Computational Complexity, Report No. 151 (2015)

New Extractors for Interleaved Sources

Eshan Chattopadhyay*
Department of Computer Science,
University of Texas at Austin
eshanc@cs.utexas.edu

David Zuckerman †
Department of Computer Science,
University of Texas at Austin
diz@cs.utexas.edu

September 14, 2015

Abstract

We study how to extract randomness from a C-interleaved source, that is, a source comprised of C independent sources whose bits or symbols are interleaved. We describe a simple approach for constructing such extractors that yields:

- For some $\delta > 0, c > 0$, explicit extractors for 2-interleaved sources on $\{0,1\}^{2n}$ when one source has min-entropy at least $(1-\delta)n$ and the other has min-entropy at least $c \log n$. The best previous construction, by Raz and Yehudayoff [RY11], worked only when both sources had entropy rate $1-\delta$.
- For some $c > 0$ and any large enough prime p, explicit extractors for 2-interleaved sources on $[p]^{2n}$ when one source has min-entropy rate at least .51 and the other source has min-entropy rate at least $(c \log n)/n$.

We use these to obtain the following applications:

- We introduce the class of any-order-small-space sources, generalizing the class of small-space sources studied by Kamp et al. [KRVZ11]. We construct extractors for such sources with min-entropy rate close to 1/2. Using the Raz-Yehudayoff construction would require entropy rate close to 1.
- For any large enough prime p, we exhibit an explicit function $f : [p]^{2n} \to \{0,1\}$ such that the randomized best-partition communication complexity of f with error $1/2 - 2^{-\Omega(n)}$ is at least $.24n \log p$. Previously this was known only for a tiny constant instead of .24, for $p = 2$ [RY11].
- We introduce non-malleable extractors in the interleaved model. For any large enough prime p, we give an explicit construction of a weak-seeded non-malleable extractor for sources over $[p]^n$ with min-entropy rate .51. Nothing was known previously, even for almost full min-entropy.

*Research supported in part by NSF Grant CCF-1218723.
†Research supported in part by NSF Grant CCF-1218723.

ISSN 1433-8092

1 Introduction

Extracting truly random bits from various naturally-arising weak random sources is a major area of study in computer science, and has applications in various areas such as cryptography, coding theory, communication complexity, and distributed computing. An extractor is defined to be a procedure that takes input from a weak random source and outputs a distribution that is close to uniform.

The first model of a weak random source was considered by Von Neumann [vN51], where he showed how to extract from a source with independent and biased bits. Various other models of weak random sources were considered [Blu86, SV86, CGH+85], but it was realized that devising such extractors was impossible for any general class of weak random sources lacking significant independence between different parts.

To get around this difficulty, the notion of a seeded extractor was introduced by Nisan and Zuckerman [NZ96], where the seeded extractor is given access to a small number of uniformly random bits to extract randomness from a weak source X. The min-entropy of a weak source X is a standard way of measuring of the amount of randomness in X, and is defined as $H_\infty(X) = \min_{s \in \text{support}(X)} \{1/\log(\Pr[X = s])\}$. The min-entropy rate of X is given by $H_\infty(X)/n$. By a long line of work ending with [LRVW03, DKSS09, GUV09], we now have explicit seeded extractors with almost optimal parameters.

In recent years, there has been renewed interest in the original problem of constructing seedless extractors for weak random sources. In particular, a line of work has focused on devising seedless extractors that takes as input C independent weak sources $X_1, \ldots, X_C$, and outputs a distribution close to uniform. This problem was originally considered by Chor and Goldreich [CG88], who showed how to extract from two independent sources (on $\{0,1\}^n$) each with min-entropy at least $(\frac{1}{2} + \delta)n$. Such extractors are called as two-source extractors. They also constructed a different two-source extractor which works when one source has min-entropy $(\frac{1}{2} + \delta)n$, and the other source has only polylogarithmic min-entropy.

However, there was no progress on this result for around 20 years until the work of Bourgain [Bou05], who achieved a small improvement over [CG88], and showed how to extract from two independent sources each with min-entropy $0.49n$, based on techniques from the area of additive combinatorics. Raz [Raz05] constructed a different two-source extractor when one source has min-entropy at least $(\frac{1}{2} + \delta)n$ and the other source has poly-logarithmic min-entropy at least $O(\log n)$, with a more flexibility in parameters than the construction of [CG88]. Finally, the authors recently constructed two-source extractors for polylogarithmic min-entropy with one bit output [CZ15]. Subsequently, Li [Li15] improved the output length to $\Omega(k)$ bits.

1.1 Interleaved Sources

Raz and Yehudayoff [RY11] introduced a natural generalization of the class of independent sources, which we call interleaved sources. We formally define this class of sources.

Notation Let $[n]$ denote the set $\{1, \ldots, n\}$. For any string $s \in [R]^n$ and $i \in [n]$, let s_i denote the symbol in the ith coordinate of s. For any permutation $t : [n] \to [n]$, define the string $w = (s)_t \in [R]^n$ such that $w_i = s_{t(i)}$ for $i = 1, \ldots, n$. For distributions D_1 and D_2, we use $|D_1 - D_2|$ to denote the statistical distance. See Section 3 for more preliminaries.

Definition 1.1 (Interleaved Sources)**.** *Let $X_1, \ldots, X_C$ be arbitrary independent sources on $[R]^n$ and let $t : [Cn] \to [Cn]$ be any permutation. Then $Z = (X_1 \circ \ldots \circ X_C)_t$ is a C-interleaved source.*

One way that such sources can arise naturally is when the independent sources are communicated remotely to an extractor and packets of bits from different sources arrive in a fixed but unknown order. We show that extractors for interleaved sources can be used to construct extractors for certain samplable sources, thus extending the line of work initiated by Trevisan and Vadhan [TV00]. We discuss this in Section 1.2. Further, Raz and Yehudayoff [RY11] showed that such extractors have applications in communication complexity (see Section 1.3) and proving lower bounds for arithmetic circuits.

Previous Results

The only known construction of an extractor for the class of interleaved sources is due to Raz and Yehudayoff [RY11]. They constructed extractors for 2-interleaved sources on $\{0,1\}^{2n}$ when both sources have min-entropy rate at least $1 - \beta$, with output length $\Omega(\beta n)$ and exponentially small error.

The constant β in the result of [RY11] is tiny and arises from a multilinear exponential sum estimate from [BGK06] (which is based on sum-product estimates on finite fields [BKT04, Kon03]). Thus, the only known construction required both the sources to have almost full min-entropy.

The analysis of the extractor in [RY11] requires estimating a non-trivial exponential sum, and is quite involved.

Our Results

We develop a simple technique that yields explicit extractors that work for lower min-entropy rates. In particular, our method yields explicit extractors for min-entropy rate 0.51 for two interleaved sources, when the sources are over a field of large enough (constant) characteristic.

We show how to convert any two-source extractor that is a function of the sum of its inputs into an extractor for a 2-interleaved source. Our method of converting a two-source extractor into an extractor for interleaved sources is based on explicit constructions of certain combinatorial sets, which we call (r, s)-spanning sets. These spanning sets are essentially subspace-evasive sets with different parameters than studied earlier (see Section 2.1 for more details). It turns out that the columns of parity check matrices of linear codes with good erasure list-decodability form spanning sets with good parameters. We discuss this in detail later.

Next, we observe that an existing two-source extractor from [CG88] is a function of the sum of the inputs. This leads to our construction of an extractor for 2-interleaved sources with one source having min-entropy at least $(1 - \alpha)n$ and the other source having min-entropy at least $\lambda \log n$ (for some $\alpha, \lambda > 0$). Next, we show that for any large enough constant prime p, if the 2-interleaved source is on $[p]^{2n}$, we can extract when one source has min-entropy rate at least 0.51 and the other source has min-entropy rate at least $c \log n/n$. We give various related constructions achieving different tradeoffs between min-entropy, error, and output length. This is summarized in Table 1.

We show that random sets are (r, s)-spanners with high probability (see Lemma 5.10). By our proof technique, any improved construction of an (r, s)-spanning set matching the probabilistic

p	k_1	k_2	Output Length	Error	Reference	Remarks
2	$\geq (1-\beta)n$	$\geq (1-\beta)n$	γn, $\gamma < \beta$	$2^{-\Omega(n)}$	[RY11]	Not strong
2	$\geq (1-\alpha)n$	$\geq 10\lambda \log n$	$\lambda \log n$	$n^{-\Omega(1)}$	This work, Theorem 6.4	Strong in X
2	$\geq (1-\alpha)n$	$\geq 10\lambda \log n$	Output in $\mathbb{Z}_M$, $M = n^{\lambda}$	$2^{-\Omega(k_2)}$	This work, Theorem 6.2	Strong in X
any $p > 2^{\frac{c}{\delta}}$	$\geq (\frac{1}{2}+\delta)n \log p$	$\geq c_1(\delta, \lambda, p) \log n$	$\lambda \log n$	$n^{-\Omega(1)}$	This work, Theorem 6.5	Strong in X
any $p > 2^{\frac{c}{\delta}}$	$\geq (\frac{1}{2}+\delta)n \log p$	$\geq (\frac{1}{2}+\delta)n \log p$	$\Omega(n)$	$n^{-\Omega(1)}$	This work, Theorem 6.6	Not strong
any $p > 2^{\frac{c}{\delta}}$	$\geq (\frac{1}{2}+\delta)n \log p$	$\geq c_2(\delta, \lambda, p) \log n$	1 bit	$2^{-\Omega(k_2)}$	This work, Theorem 6.7	Strong in X
any $p > 2^{\frac{c}{\delta}}$	$\geq (\frac{1}{2}+\delta)n \log p$	$\geq c_1(\delta, \lambda, p)\lambda \log n$	$\Omega(k_2)$	$2^{-\Omega(k_2)}$	This work, Theorem 6.9	Semi-explicit construction
2	$\geq \gamma n$, any constant γ	$\geq \gamma n$	$\lambda \log n$	$n^{-\Omega(1)}$	This work, Theorem 6.11	Assuming Generalized Paley Graph Conjecture

Table 1: Results on Extractors for 2-Interleaved Sources. The setting is as follows: $Z = (X \circ Y)_t$ is an arbitrary 2-interleaved source on $[p]^{2n}$, where X and Y are independent sources on $[p]^n$ (for some prime p) with min-entropy k_1 and k_2 respectively, and $t : [2n] \to [2n]$ is an arbitrary permutation. Let α be a small enough constant and c a large enough constant. Also let $\lambda > 1$ be any constant. (We also list the result of [RY11] in Table 1.

method will yield extractors for 2-interleaved sources on $\{0,1\}^{2n}$ that have essentially the same min-entropy requirement as standard (non-interleaved) setting.

1.2 Any-Order-Small-Space-Sources

Trevisan and Vadhan [TV00] introduced the problem of constructing seedless extractors for the class of samplable sources (the weak random source is generated by an efficient algorithm) and constructed explicit extractors based on some complexity-theoretic assumptions. Subsequently, Kamp et al. [KRVZ11] introduced a class of samplable sources called small-space sources, where the algorithm generating the source has bounded space. They constructed seedless extractors for such sources with linear min-entropy. Most sources considered previously (for seedless extraction) can be computed in small-space (see [KRVZ11] for more details). In particular, extractors for small-space sources also extract from bit-fixing sources and symbol-fixing sources, and thus have applications in cryptography [KZ07].

We introduce a natural generalization of small-space sources. For this, we recall the definition of small-space sources from [KRVZ11].

Definition 1.2 (Small-Space Sources [KRVZ11])**.** *A space s source X on $[r]^n$ is generated by a r-way branching program of length n and width 2^s in the following way: The r-way branching program is a layered graph with $n+1$ layers and a single start vertex. Each edge is labeled with a variable X_j, a probability value and a symbol in $[r]$. Further all edges between the ith and $(i+1)$th layer are labelled with same variable X_i. The output of the source is a random walk starting from the start vertex, assigning the symbol on the edge to the corresponding variable and finally outputting the generated string.*

Note that in the above definition, the variable assigned to an edge is known (for example, all edges between the ith and $(i+1)$th layers have the variable X_i assigned to it). We introduce the natural generalization where the branching program is oblivious but the variable assigned to an edge is unknown. In particular, for an unknown permutation $t : [n] \to [n]$, all edges between the ith and $(i+1)$th layers have the variable $X_{t(i)}$ assigned to it.

We formally define this class of sources.

Definition 1.3 (Any-Order-Small-Space-Sources)**.** *An any-order-space s source X on $[r]^n$ is generated by an r-way branching program of length n and width 2^s and a permutation $t : [n] \to [n]$ in the following way: The r-way branching program is a layered graph with $n+1$ layers and a single start vertex. Each edge is labeled with a variable X_j, a probability value and a symbol in $[r]$. Further all edges between the ith and $(i+1)$th layer are labelled with same variable $X_{t(i)}$. The output of the source is a random walk starting from the start vertex, assigning the symbol on the edge to the corresponding variable and finally outputting the generated string.*

Our Results

To construct extractors for the class of any-order-oblivious-small-space sources, we reduce it to the task of extracting from 2-interleaved sources by adapting the technique of [KRVZ11] to our situation.

Consider an arbitrary any-order-space $s = \delta n/2$ source X on $[p]^n$ (for some constant p) with min-entropy $k = (\frac{1}{2} + \delta) n \log p$. By conditioning on the state of the p-way branching program at the $\frac{n}{2}$th layer, it follows by Lemma 4.7 that X is $2^{-\Omega(n)}$-close to a source $Z = (Y_1 \circ Y_2)_t$, where Y_1 and Y_2 are independent sources on $[p]^{\frac{n}{2}}$ with $\min\{H_\infty(Y_1), H_\infty(Y_2)\} \geq \frac{\delta n \log p}{8}$ and $\max\{H_\infty(Y_1), H_\infty(Y_2)\} \geq (\frac{1}{2} + \frac{\delta}{8}) \frac{n \log p}{2}$, and $t : [n] \to [n]$ is a permutation.

It thus follows that all our extractor constructions for 2-interleaved sources also extract from any-order-small-space sources (by splitting the input string into two equal parts and applying the extractor).

Using this reduction, we obtain the first explicit construction of an extractor for any-order-oblivious-small-space sources with min-entropy rate close to $\frac{1}{2}$ (by using the extractor from Theorem 6.5).

Theorem 1.4. *There exists $c > 0$ such that for any $\delta \geq 2\delta_1 > 0$ and any prime $p > 2^{\frac{c}{\delta}}$, there exists an explicit function $\text{ext} : [p]^n \to \{0,1\}^m$, $m = O(\log n)$, such that if X is an any-order-oblivious space $s = \delta_1 n$ source on $[p]^n$ with min-entropy $(\frac{1}{2} + \delta) \log p$, then*

$$|\text{ext}(X) - U_m| = n^{-\Omega(1)}.$$

We note that using our reduction, the extractor from [RY11] can be used to extract from any-order-small-space sources with min-entropy rate very close to 1.

1.3 Applications to Communication Complexity

Communication complexity was introduced by Yao [Yao79]. There has been an extensive amount of research done in this area and various models of communication have been considered (see [KN97] for formal definitions and background). We recall the definition of the randomized best-partition communication complexity of an arbitrary function $f : [R]^{2n} \to \{0,1\}$, which generalizes the usual setting where the partition of inputs is known.

Let Alice and Bob be two players who want to collectively compute f following a protocol Π and having access to a common random string r. Fix an arbitrary partition of the set $[2n]$ into 2 subsets of equal size, say S and T. For arbitrary $x, y \in [R]^n$, Alice is given x and Bob receives y and the goal is to compute $f(z)$ with probability at least $1 - \epsilon$, where $z \in [R]^{2n}$ such that $z_S = x$ and $z_T = y$.

For any protocol Π, the randomized communication cost of f with respect to an equi-partition $S, T \subset [2n]$ denoted by $R^{\epsilon}_{\Pi,S,T}(f)$, is defined to be the maximum communication between Alice and Bob over all inputs x, y in the scenario described above. The best-partition communication complexity of f, denoted by $R^{best,\epsilon}(f)$ is defined as:

$$R^{best,\epsilon}(f) = \min_{\Pi}\{\min_{\substack{S,T:|S|=|T|=n,\\ S\cup T=[2n]}} R^{\epsilon}_{\Pi,S,T}(f)\}.$$

Lower bounds on the best-partition communication complexity of f implies lower bounds on branching programs computing f ([AM86]) and also has implications in time/space tradeoffs for VLSI circuits.

Raz and Yehudayoff [RY11] proved the following lower bound.

Theorem 1.5 ([RY11])**.** *For some $\beta > 0$, there exists an explicit function $f : \{0,1\}^{2n} \to \{0,1\}$ such that the randomized best-partition communication complexity of f with error $\epsilon = \frac{1}{2} - 2^{-\beta n}$ is at least βn.*

The constant β in the above theorem is, however, extremely small and arises from arguments in additive combinatorics. A similar bound also follows from their work for inputs on $[R]^{2n}$ (for any constant R) and it appears nontrivial to use their techniques to obtain better bounds for β.

Our Results

We obtain the following result.

Theorem 1.6. *There exists $c > 0$ such that for any $\delta, \gamma > 0$ and any prime $p > 2^{\frac{c}{\delta}}$, there exists an explicit function $f : [p]^{2n} \to \{0,1\}$ such that the randomized best-partition communication complexity of f with error $\epsilon = \frac{1}{2} - p^{-\gamma n}$ is at least $(\frac{1}{4} - \delta - \gamma)n \log p$.*

We prove this using a well known technique of lower bounding randomized communication complexity by discrepancy. Our explicit function is the 1-bit extractor constructed in Theorem 6.7.

However, we need to analyze the error of the extractor more carefully to obtain the above bound. We prove Theorem 1.6 in Section 8.

1.4 Interleaved-Non-Malleable Extractors

Non-malleable extractors were introduced by Dodis and Wichs [DW09], where it was shown that explicit constructions of non-malleable extractor with good parameters can be used to design almost optimal protocols for privacy amplification, which is a very well studied problem in cryptography. Recently, non-malleable extractors were also used in constructing explicit two-source extractors [CZ15]. We introduce the natural generalization of non-malleable extractors in the interleaved model.

We first recall the definition of a non-malleable extractor.

Definition 1.7 (Non-Malleable Extractor)**.** *A function* $\mathrm{nmExt} : [R]^{2n} \to \{0,1\}^m$ *is a non-malleable extractor for min-entropy* k *and error* ϵ *if the following holds: If* X *is a source (on* $[R]^n$*) with min-entropy* k*,* $f : [R]^n \to [R]^n$ *is any function with no fixed points, then*

$$|\mathrm{nmExt}(X \circ U_{[R]^n}) \circ \mathrm{nmExt}(X \circ f(U_{[R]^n})) \circ U_{[R]^n} - U_m \circ \mathrm{nmExt}(X \circ f(U_{[R]^n}) \circ U_{[R]^n}| < \epsilon.$$

The first explicit construction of a non-malleable extractors was given in [DLWZ14], with subsequent improvements of parameters achieved in [CRS12, Li12]. However these constructions require min-entropy $> 0.49n$. In a recent work [CGL15], the min-entropy required was improved to $O(\log^2 n)$.

We initiate the study of non-malleable extractors in the interleaved model, where the extractor is guaranteed to work even when symbols from the source X and tampered seed $U_{[R]^n}$ arrive to the non-malleable extractor in a fixed but unknown interleaved order.

We formally define interleaved-non-malleable extractors.

Definition 1.8 (Interleaved-Non-Malleable Extractor)**.** *A function* $\mathrm{nmExt} : [R]^{2n} \to \{0,1\}^m$ *is a non-malleable extractor in the any-order model for min-entropy* k *and error* ϵ *if the following holds: If* X *is a source (on* $[R]^n$*) with min-entropy* k*,* $f : [R]^n \to [R]^n$ *is any function with no fixed points and* $t : [2n] \to [2n]$ *is any permutation, then*

$$|\mathrm{nmExt}((X \circ U_{[R]^n})_t) \circ \mathrm{nmExt}((X \circ f(U_{[R]^n}))_t) \circ U_{[R]^n} - U_m \circ \mathrm{nmExt}((X \circ f(U_{[R]^n}))_t) \circ U_{[R]^n}| < \epsilon.$$

In the above definition, when the seed has some min-entropy instead of being uniform, we say that the interleaved-non-malleable extractor is weak-seeded.

Our Results

We give the first explicit construction of an interleaved-non-malleable extractor. Further our non-malleable extractor is weak-seeded.

Theorem 1.9. *There exists* $\lambda > 0$ *such that for any* $\delta > 0$*,* $c > c(\delta)$ *and any prime* $p > 2^{\frac{\lambda}{\delta}}$*, there exists an explicit function* $\mathrm{nmExt} : \mathbb{F}_p^{2n} \to \{0,1\}^m$*,* $m = O(\log n)$*, such that if* X*,* Y *are independent sources on* $\mathbb{F}_p^n$ *with min-entropy* k_1, k_2 *respectively, satisfying* $k_1 > (\frac{1}{2} + \delta)n \log p$ *and*

$k_2 > c\max\{m, \log n\}$, $t : [2n] \to [2n]$ *is any injective map and* $f : \mathbb{F}_p^n \to \mathbb{F}_p^n$ *is any function with no fixed points, then*

$$|\mathrm{nmExt}((X \circ Y)_t) \circ \mathrm{nmExt}((X \circ f(Y))_t) \circ Y - U_m \circ \mathrm{nmExt}((X \circ f(Y))_t) \circ Y| = n^{-\Omega(1)}.$$

As before, if we are allowed to run the non-malleable extractor in sub-exponential time, we can extract $\Omega(n)$ bits at error $2^{-\Omega(n)}$. See Theorem 7.4 for more details.

Organization

We outline our constructions in Section 2. We introduce preliminaries in Section 3 and recall some known explicit constructions and other tools in Section 4. In Section 6, we present our extractor constructions for 2-interleaved sources. In Section 7, we present our constructions of interleaved-non-malleable extractors. We present the proof of Theorem 1.6 in Section 8.

2 Outline of Constructions

2.1 Extractors for 2-Interleaved Sources

Our extractor for interleaved sources exploits the existence of good 2-source extractors which are functions of $X + Y$. To do this, we encode our source in a new way. Our encoding is based on explicit constructions of certain combinatorial sets, which we call spanning vectors.

Definition 2.1. *A set of vectors* $S \subseteq \mathbb{F}_p^{\bar{\ell}}$ *is* (r, s)*-spanning if the span of any* r *vectors of* S *has dimension at least* s*.*

Note that this is the same as a subspace-evasive set: Any $(s-1)$-dimensional subspace contains at most $(r-1)$ vectors in the set. However our parameters are quite different than studied previously [Gur11, DL12].

Our explicit constructions of spanning vectors are based on using the columns of a parity check matrix of a linear codes with good erasure list-decodability. Informally, a (e, L)-erasure list-decodable code $\mathcal{C}$ satisfies the property that at most L codewords agree on any particular subset of co-ordinates of size $n - e$. This property can then be used to lower bound the rank of any subset of e columns of the parity check matrix of $\mathcal{C}$. We refer the reader to Section 5 for more details.

We define the following encoding based on spanning vectors.

Definition 2.2. *For any* (r, s)*-spanning set* $S = \{v_1, \ldots, v_\ell\} \subseteq \mathbb{F}_p^{\bar{\ell}}$ *of size* ℓ*, the function* $\mathrm{enc} : \mathbb{F}_p^\ell \to \mathbb{F}_p^{\bar{\ell}}$ *defined as*

$$\mathrm{enc}(z) = \sum_{i=1}^{\ell} z_i v_i$$

is called an (r, s)*-encoding from* $\mathbb{F}_p^\ell$ *to* $\mathbb{F}_p^{\bar{\ell}}$*.*

Consider the following setting: Let $Z = (X \circ Y)_t$ be any 2-interleaved source on $\{0,1\}^{2n}$, where X and Y are arbitrary independent sources on $\{0,1\}^n$ with min-entropy k_1 and k_2 respectively, and $t : [2n] \to [2n]$ is any permutation.

Our first step is to use an (n, s)-encoding enc from $\mathbb{F}_2^{2n}$ to $\mathbb{F}_2^{\bar{n}}$ to encode Z. Thus,

$$\text{enc}(Z) = X' + Y',$$

where

$$X' = \sum_{i=1}^{n} X_i v_{t(i)}, \quad Y' = \sum_{i=j}^{n} Y_j v_{t(n+j)}.$$

where $S = \{v_1, \ldots, v_{2n}\}$ is a (n, s)-spanning set of vectors.

The idea is to argue that the independent sources X' and Y' (on $\{0,1\}^{\bar{n}}$) have enough min-entropy. Since (by construction) the span of the set of vectors $\{v_{t(1)}, \ldots, v_{t(n)}\}$ has dimension at least s, Lemma 4.9 implies that $H_\infty(X') = k_1' \geq k_1 - (n-s)$. Similarly $H_\infty(Y') = k_2' \geq k_2 - (n-s)$.

We now associate $\mathbb{F}_2^{\bar{n}}$ with $\mathbb{F}_{2^{\bar{n}}}$. A character sum estimate of Karatsuba[1] [Kar71, Kar91] (we use a slightly more precise bound from [Shp13], see Theorem 4.1) implies that for any nonprincipal multiplicative character χ of $\mathbb{F}_{2^{\bar{n}}}^*$,

$$E_{X'}|\mathbb{E}_{Y'}[\chi(X' + Y')]| \leq 2^{-\delta k_2'}$$

whenever: $k_1 \geq (\frac{1}{2} + 3\delta)\bar{n} + (n - s)$ and $k_2 \geq 4 \log \bar{n} \log p + (n - s)$.

Suppose k_1 and k_2 satisfy these conditions.

We then follow a standard approach and define the function:

$$\text{ext}(Z) = \log_g(X' + Y') \pmod{M},$$

where $M = 2^{\delta k_2'/2}$ and g is a primitive element of $\mathbb{F}_{2^{\bar{n}}}$. Using a version of the Abelian XOR lemma (see Lemma 4.5), it follows that ext is an extractor with output length $\delta k_2'/2$ and error $2^{-\Omega(k_2')}$. Further the extractor is strong in the source X. However, the running time of this extractor is subexponential since it involves computing discrete logs over finite fields. This gives us a semi-explicit extractor construction.

To get a polynomial time extractor, we compute discrete log over a smaller multiplicative subgroup of $\mathbb{F}_{2^{\bar{n}}}^*$. Let $M | 2^{\bar{n}} - 1$ and $M = n^\lambda$ for any constant λ (we show in Theorem 6.2 that we can ensure that there is always such an M). Define the function:

$$\text{ext}_1(Z) = \text{enc}(Z)^{\frac{2^{\bar{n}}-1}{M}}.$$

Thus $\text{ext}_1(Z)$ is a distribution on the multiplicative subgroup $G = \{x^{\frac{2^{\bar{n}}-1}{M}} : x \in \mathbb{F}_{2^{\bar{n}}}^*\}$ (of $\mathbb{F}_{2^{\bar{n}}}^*$) of size M (in fact $\text{ext}_1(Z)$ is a distribution on $G \cup \{0\}$, but $\Pr[\text{ext}_1(Z) = 0] = 2^{-\Omega(n)}$ and hence we ignore this and add this to the error). Let g be a generator of G. It now follows by using the character sum estimate of Karatsuba [Kar71] that the function:

$$\text{ext}(Z) = \log_g(\text{ext}_1(Z))$$

is an extractor.

We need to find a generator g of G efficiently. For this, we use an efficient algorithm of Shoup [Sho90] for finding a small set of elements such that one of them is a primitive element of $\mathbb{F}_{2^{\bar{n}}}$. We

[1] this character sum was also used in [CG88] for constructing explicit two-source extractors.

use a straightforward method to find g from this set in polynomial time. We achieve output length of $\lambda \log n$ and error $n^{-\Omega(1)}$. The extractor is strong in the source X.

Reducing the Min-Entropy Rate For some c and any $\delta > 0$, let $p > 2^{\frac{c}{\delta}}$ be any prime. When the source $Z = (X \circ Y)_t$ is on $[p]^{2n}$, we can reduce the min-entropy rate requirement of the source X to $(\frac{1}{2} + \delta)$. The construction follows the same outline as above (using (n, s)-encodings from $\mathbb{F}_p^{2n}$ to $\mathbb{F}_p^{\bar{n}}$), and the improvement is achieved by using the fact that over alphabet $[p]$, we can construct (n, n)-spanning sets in $\mathbb{F}_p^{\bar{n}}$ with $\bar{n} = n(1 + \frac{\delta}{5})$ (using explicit codes from [GI02]). The output length of the extractor obtained is $\lambda \log n$ (for any constant λ) and achieves error $n^{-\Omega(1)}$. Further the extractor is strong in the source X.

Improving the Output Length We improve the output length of the above extractor to $\Omega(n)$ when both sources X and Y (on $[p]^n$) have min-entropy at least $(\frac{1}{2} + \delta)n \log p$. Our construction is as follows. Let SExt be an explicit strong seeded extractor for linear min-entropy with linear output length and polynomially small error with seed seed length $O(\log n)$, for example from the work of [GUV09]. Let $Z_{[n]}$ denote the projection of Z to the first n co-ordinates and let ext_p denote the extractor constructed in the previous paragraph (for 2-interleaved sources on $[p]^{2n}$). Our extractor is the following function:

$$\mathrm{ext}_{p,long}(Z) = \mathrm{SExt}(Z_{[n]}, \mathrm{ext}_p(Z)).$$

We sketch the proof of correctness. Without loss of generality, suppose that X has more symbols in $Z_{[n]}$ than the source Y. Let $S \subseteq [n]$ be the co-ordinates of X which are in $Z_{[n]}$ and let X_S denote the projection of X to the co-ordinates indexed by S. Let $T \subset [n]$ be the co-ordinates of Y which are in $Z_{[n]}$ and let Y_T denote the projection of Y to the co-ordinates indexed by T. Further, we use $X_S \circ Y_T$ to denote $Z_{[n]}$. Note that, by assumption $|S| \geq \frac{n}{2}$ and $|T| \leq \frac{n}{2}$. It follows by Lemma 4.7 that $Y|Y_T$ is close to a source with min-entropy $> \frac{\delta n \log p}{2}$ with probability $1 - 2^{-\Omega(n)}$. Also note that X_S a source with min-entropy $\geq \delta n \log p$.

Consider such a good fixing $Y_T = y_T$. Since X and $Y|Y_T = y_T$ have enough min-entropy, it follows that even under this fixing, $W = \mathrm{ext}_p(Z)$ is close to uniform. We now use the property that ext_p is strong with respect to the source X_S, i.e., $|(X_S, W) - (X_S, U_d)| \leq n^{-\Omega(1)}$. Using a probability lemma from [Sha06], it follows that for any $W = w$, $|X_S - (X_S|(W = w))| \leq n^{-\Omega(1)}$ (using that w is of length $O(\log n)$).

Hence, $\mathrm{SExt}(X_S \circ Y_T, W)|Y_T = y_T$ is $n^{-\Omega(1)}$-close to the convex combination: $\sum_w \Pr[(W|Y_T = y_T) = w]\mathrm{SExt}(X_S \circ Y_T, w)|Y_T = y_T$. Since as observed above, $W|Y_T = y_T$ is $n^{-\Omega(1)}$-close to U_d, it follows that $\mathrm{SExt}(X_S \circ Y_T, W)|Y_T = y_T$ is $n^{-\Omega(1)}$-close to $\mathrm{SExt}(X_S \circ y_T, U_d)$. The correctness now follows using the fact that SExt is a seeded extractor for linear min-entropy.

Probabilistic Method We show in Lemma 5.10, that a random set $S \subset \mathbb{F}_2^n$ of size $2n$ is an $(n, n - 2\sqrt{n})$-spanning set with high probability. Thus, using the proof technique described above, any explicit construction of such a set will yield explicit extractors for 2-interleaved sources on $\{01\}^{2n}$ when one source has min-entropy at least $0.51n$ and the other source has min-entropy at least $cn^{\frac{1}{2}}$. We leave it as an interesting open problem to explicitly construct such a set S.[2]

We give formal proofs of the above extractor constructions and other related constructions in Section 6.

[2]This is related to finding explicit constructions of binary erasure list-decodable codes with almost optimal parameters. See Section 5 for more details.

2.2 Interleaved-Non-Malleable Extractors

For some $c > 0$ and any $\delta > 0$, let $p > 2^{\frac{c}{\delta}}$ be any prime. Let X be a source on $[p]^n$ with min-entropy k_1 and Y be a weak-seed on $[p]^n$ with min-entropy k_2. Let $f : [p]^n \to [p]^n$ be any function with no fixed points. Thus the non-malleable extractor has access to $Z = (X \circ Y)_t$ for an artitrary permutation $t : [2n] \to [2n]$. Let Z_f denote the tampered source $(X \circ f(Y))_t$.

We show that the extractor ext_p constructed for 2-interleaved sources (described in the previous section) is also non-malleable. We prove it in the following way. Recall the construction of ext_p:

$$\text{enc}(Z) = \sum_{i=1}^{2n} Z_i v_i, \quad \text{ext}_1(Z) = \text{enc}(Z)^{\frac{p^{\bar{n}}-1}{M}}, \quad \text{ext}_p(Z) = \log_g(\text{ext}_1(Z)),$$

where $S = \{v_1, \ldots, v_{2n}\}$ is an (n, n)-spanning set in $\mathbb{F}_p^{\bar{n}}$, $M = \text{poly}(n)$, $\bar{n} = n(1 + \frac{\delta}{5})$ and g is a generator of the multiplicative subgroup $G = \{x^{\frac{2^{\bar{n}}-1}{M}} : x \in \mathbb{F}_{2^{\bar{n}}}^*\}$.

Since ext_p is a distribution on $\mathbb{Z}_M$, it follows by a version of the Abelian XOR lemma proved in [DLWZ14] that to prove non-malleability, it is enough to prove the bound:

$$|\mathbb{E}[\psi_a(\text{ext}_p(Z))\psi_b(\text{ext}_p(Z_f))]| \leq n^{-\Omega(1)},$$

for all additive characters ψ_a and ψ_b (of $\mathbb{Z}_M$) such that ψ_a is nontrivial. When ψ_b is the trivial character, the above quantity can be bounded by the fact that ext_p is an extractor for 2-interleaved sources. Thus, suppose both ψ_a and ψ_b are nontrivial.

It follows that

$$|\mathbb{E}[\psi_a(\text{ext}_p(Z))\psi_b(\text{ext}_p(Z_f))]| = |\mathbb{E}[\chi_a(\text{enc}(Z))\chi_b(\text{enc}(Z_f))]|$$

where χ_a and χ_b are nonprincipal multiplicative characters of $\mathbb{F}_{2^{\bar{n}}}^*$.

Further, $Z = \sum_{i=1}^n X_i v_{t(i)} + \sum_{j=1}^n Y_j v_{t(j)}$ and $Z_f = \sum_{i=1}^n X_i v_{t(i)} + \sum_{j=1}^n f(Y)_j v_{t(j)}$. Thus,

$$Z = X' + Y', \quad Z_f = X' + f'(Y'),$$

where $X' = \sum_{i=1}^n X_i v_{t(i)}$, $Y' = \sum_{i=j}^n Y_j v_{t(n+j)}$ and $f' = L \circ f \circ L^{-1}$, L being the one-one linear map $L(z) = \sum_{i=1}^n z_i v_{t(n+i)}$. Thus,

$$|\mathbb{E}[\psi_a(\text{ext}_p(Z))\psi_b(\text{ext}_p(Z_f))]| = |\mathbb{E}[\chi_a(X' + Y')\chi_b(X' + f'(Y'))]|.$$

Using the work of Dodis et al. [DLWZ14], we can prove the required upper bound on the quantity on the right hand side if f' does not have any fixed points. We indeed show that f' has no fixed points (by using the fact that L is one-one and f has no fixed points). This completes the proof sketch. The non-malleable extractor outputs $\lambda \log n$ bits (for any constant λ) and achieves error $n^{-\Omega(1)}$.

See Section 7 for more details.

3 Preliminaries

3.1 Notation

We use capital letters to denote distributions and their support. We use corresponding small letters to denote a sample from the source.
We use $[l]$ to denote the set $\{1, 2, \ldots, l\}$ and $[a, b]$ to denote the set $\{a, a+1 \ldots, b\}$.
We use U_m to denote the uniform distribution over $\{0,1\}^m$.
For any set S, let U_S denote the uniform distribution on S. Also let $s \sim S$ denote a uniform draw from S.
For any string $s \in [R]^n$ and $i \in [n]$, let s_i denote the symbol at the ith coordinate of s. For any one-one map $t : [n] \to [n]$, define the string $w = (s)_t \in [R]^n$ such that $w_i = s_{t(i)}$ for $i = 1, \ldots, n$. Further for any $t \subset [n]$, let s_T denote the $|T|$ length string that is the projection of s onto the co-ordinates indexed by T.
For any $x \in [p]^{n_1}$, $y \in [p]^{n_2}$ and disjoint subsets $S, T \subset [n_1 + n_2]$ with $|S| = n_1$, $|T| = n_2$, we define $z = x_S \circ y_T$ such that $z_S = x$ and $z_T = y$.
For any integer $M > 0$, let $e_M(x) = e^{\frac{2\pi i x}{M}}$.

3.2 Min-Entropy and Flat Distributions

Definition 3.1. *The min-entropy of a source X is defined as:* $H_\infty(X) = \min_{s \in \text{support}(X)} \left\{ \frac{1}{\log(\Pr[X=s])} \right\}$.

Definition 3.2. *A distribution (source) D is flat if it is uniform over a set S.*

Definition 3.3. *A (n,k)-source is a distribution on $\{0,1\}^n$ with min-entropy k.*

Any (n,k)-source is a convex combination of flat sources supported on sets of size 2^k [Zuc97].

3.3 Statistical distance and Convex Combination of Distributions

Definition 3.4. *Let D_1 and D_2 be two distributions on a set S. The statistical distance between D_1 and D_2 is defined to be:* $|D_1 - D_2| = \frac{1}{2} \sum_{s \in S} |\Pr[D_1 = s] - \Pr[D_2 = s]|$.

A distribution D_1 is ϵ-close to another distribution D_2 if $|D_1 - D_2| \leq \epsilon$.

Definition 3.5. *For random variables X and Y, we use $X|Y$ to denote a random variable with distribution:* $\Pr[(X|Y) = x] = \sum_{y \in \text{support}(Y)} \Pr[Y = y] \cdot \Pr[X = x | Y = y]$.

4 Some Known Explicit Constructions and Other Tools

To construct our extractors, we use a variety of tools. We first set up these tools in this section and present our extractor constructions in the next section.

4.1 A 2-Source Extractor

The following double character sum estimate was obtained by Karatsuba [Kar71, Kar91]. We state a slightly more precise bound from [Shp13].

Theorem 4.1 ([Kar71, Kar91, Shp13])**.** *Let p be any prime. Let χ be a nonprincipal multiplicative character of $\mathbb{F}_{p^n}^*$. For any subsets $A, B \subseteq \mathbb{F}_{p^n}$, the following holds: For any integer $\lambda > 0$,*

$$\sum_{a \in A} \left| \sum_{b \in B} \chi(a+b) \right| \leq |A|^{\frac{2\lambda-1}{2\lambda}} (|B| p^{\frac{n}{4\lambda}} + |B|^{\frac{1}{2}} p^{\frac{n}{2\lambda}}).$$

The above theorem can be equivalently restated as a result on 2-source extractors.

Theorem 4.2. *Let p be any prime. Let χ be a nonprincipal multiplicative character of $\mathbb{F}_{p^n}^*$. For any $\delta > 0$ and independent sources X, Y on $\mathbb{F}_{p^n}$ with min-entropy k_1, k_2 respectively, satisfying $k_1 \geq \left(\frac{1}{2} + 3\delta\right) n \log p$ and $k_2 \geq 4 \log n \log p$, we have*

$$\mathbb{E}_{x \sim X} |\mathbb{E}_{y \sim Y} [\chi(x+y)]| \leq 2^{-\delta k_2}.$$

Proof. Let X, Y be flat sources on sets A and B respectively. Thus $|A| = 2^{k_1}$ and $|B| = 2^{k_2}$. Setting $\lambda = \frac{n \log p}{k_2}$ in Theorem 4.1 (so that $|B| = 2^{k_2} = p^{\frac{n}{\lambda}}$), we have

$$\begin{aligned}
\mathbb{E}_{x \sim X} |\mathbb{E}_{y \sim Y} [\chi(x+y)]| &\leq |A|^{-\frac{1}{2\lambda}} (p^{\frac{n}{4\lambda}} + |B|^{-\frac{1}{2}} p^{\frac{n}{2\lambda}}) \\
&\leq |A|^{-\frac{1}{2\lambda}} (p^{\frac{n}{2\lambda}} + 1) \\
&\leq p^{-\frac{3\delta n}{2\lambda}} + |A|^{-\frac{1}{2\lambda}} \\
&\leq 2p^{-\frac{3\delta n}{2\lambda}} \\
&= 2^{1 - \frac{3k_2 \delta n \log p}{2n \log p}} \leq 2^{-\delta k_2}.
\end{aligned}$$

□

4.2 A Seeded Extractor

We recall an explicit construction of a strong seeded extractor with optimal parameters.

Theorem 4.3 ([GUV09])**.** *There exists a constant $\alpha > 0$ such that for all $n, k \in \mathbb{N}$, there exists an explicit strong seeded extractor* $\mathrm{Ext} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$*, where $d = O\left(\frac{n}{\epsilon}\right)$ and $m = (1-\alpha)k$.*

4.3 Abelian XOR Lemmas

The following lemma is known as Vazirani's XOR Lemma.

Lemma 4.4. *Let D be a distribution over $\mathbb{Z}_M$ such that for every nontrivial additive character ψ of $\mathbb{Z}_M$, we have $|\mathbb{E}[\psi(D)]| \leq \epsilon$. Then, we have*

$$|D - U_M| \leq \epsilon \sqrt{M}.$$

Let $\sigma_M : \mathbb{Z}_N \to \mathbb{Z}_M$ be defined as $\sigma_M(x) = x \pmod M$. The following general version of the above XOR lemma was proved in [Rao07].

Lemma 4.5 ([Rao07]). *Let D be a distribution over $\mathbb{Z}_N$ such that for every non-trivial additive character ψ of $\mathbb{Z}_N$, we have $|\mathbb{E}[\psi(D)]| \leq \epsilon$. Then, for any $M < N$, we have*

$$|\sigma_M(D) - U_M| \leq O(\epsilon \log N \sqrt{M}) + O(M/N).$$

We also record a more generalized form of the XOR Lemma [DLWZ14].

Lemma 4.6 ([DLWZ14]). *Let D_1, D_2 be distributions over $\mathbb{Z}_N$ such that for arbitrary characters ψ, ϕ of $\mathbb{Z}_N$, we have $|\mathbb{E}[\psi(D_1)\phi(D_2)]| \leq \epsilon$, whenever ψ is nontrivial. Then, for any $M < N$, we have*

$$|(\sigma_M(D_1), \sigma_M(D_2)) - (U_M, \sigma_M(D_2))| = O(\epsilon(\log N)^2 M) + O(M/N).$$

4.4 Probability Lemmas

The following result follows from a lemma proved in [MW97].

Lemma 4.7 ([MW97]). *Let X, Y be random variables with supports $S, T \subseteq V$ such that (X, Y) is ϵ-close to a distribution with min-entropy k. Further suppose that the random variable Y can take at most l values. Then*

$$\Pr_{y \sim Y}\left[(X|Y = y) \text{ is } 2\epsilon^{1/2}\text{-close to a source with min-entropy } k - \log l - \log\left(\frac{1}{\epsilon}\right)\right] \geq 1 - 2\epsilon^{1/2}.$$

We also need the following lemma.

Lemma 4.8 ([Sha06]). *Let Y be a random variable taking values in $\{0,1\}^d$. Suppose $|(X, Y) - (X, U_d)| \leq \epsilon$. Then for any $y \in \text{support}(Y)$, $|X - (X|Y = y)| \leq 2^{d+1}\epsilon$.*

Lemma 4.9. *Let X be a source on $\mathbb{F}_p^n$ with min-entropy k. Let $V = \{v_1, \ldots, v_n\}$ be a collection of vectors such that $\dim(\text{span}\{V\}) \geq n - A$. Then $X_V = \sum_i x_i v_i : x \sim X$ is a source with min-entropy $\geq k - A \log p$.*

4.5 Finding Primitive Elements in Finite fields

There is no known deterministic polynomial time algorithm to find any primitive element of a finite field $\mathbb{F}_{p^n}$. However, there are efficient algorithms known for a weaker task, where the algorithm is only required to output a small set of elements with the guarantee that one of the elements is primitive. The following result is due to Shoup [Sho90].

Theorem 4.10 ([Sho90]). *Let $p > 0$ be any prime. For all $n > 0$, there exists a deterministic procedure which takes as input n, runs in time $\text{poly}(n)$, and outputs a set $S = \{a_1, \ldots, a_l\}$, $l = \text{poly}(n)$, such that S contains a primitive element of $\mathbb{F}_{p^n}$.*

5 Constructing Spanning Vectors

A key ingredient in our extractor construction are explicit constructions of spanning vectors. Recall that a set of vectors $S \subseteq \mathbb{F}_p^{\bar{\ell}}$ is (r, s)-spanning if the span of any r vectors of S has dimension at least s (see Definition 2.1). Our constructions of spanning vectors are simple and are based on explicit linear codes. Recall that a linear code of block length n, dimension k and distance d over any field $\mathbb{F}$ is a k dimensional subspace over $\mathbb{F}$ with the number of zero co-ordinates of any vector in this subspace being at most $n - d$. The relative rate of the code is k/n and the relative distance is d/n.

We show that the columns of the parity check matrix of any linear code with good erasure list-decoding radius (defined below) can be used as a spanning set.

Definition 5.1 (Erasure List-Decoding Radius [Gur03])**.** *We say that a linear code $[n, k, d]$ code $\mathcal{C}$ over a finite field $\mathbb{F}$ is (e, L)-erasure list-decodable if for every for every $r \in \mathbb{F}^{n-e}$ and $T \subseteq [n]$ of size $n - e$, $|\{c \in \mathcal{C} : c_T = r\}| \leq L$.*

We now establish a simple connection between erasure list-decodable codes and spanning sets.

Lemma 5.2. *Let $\mathcal{C}$ be a linear $[n, k, d]$ code over a finite field $\mathbb{F}$, which is (e, L)-earasure list-decodable. Let H be parity check matrix of $\mathcal{C}$, and let S be the set of columns of H. Then $S \subset \mathbb{F}^{n-k}$ is a (r, s)-spanning set of size n, with $r = e$ and $s = e - \log_{|\mathbb{F}|}(L)$.*

Proof. Since $\mathcal{C}$ is (e, L)-erasure list-decodable, it follows that the size of the null space of any e columns of the parity check matrix H is at most L. By the rank-nullity theorem, it follows that the rank of the sub-matrix of H restricted to these e columns is at least $e - \log_{|\mathbb{F}|}(L)$. Thus by definition, the set of columns of H form a $(e, e - \log_{|\mathbb{F}|}(L))$-spanning set. □

The following lemma relates the minimum distance of a code to its erasure list-decoding radius, and can be seen as an analogue of the Johnson bound for erasure list-decoding.

Lemma 5.3 ([Gur04])**.** *Let $\mathcal{C}$ be a code with block length n and relative distance δ over an alphabet of size q. Then for any $\epsilon > 0$, $\mathcal{C}$ is a (e, L)-erasure list-decodable code, where $e = \left(\frac{q}{q-1} - \epsilon\right)\delta n$ and $L = \frac{q}{(q-1)\epsilon}$.*

Combining the above results, the following lemma is immediate.

Lemma 5.4. *For any $\delta > 0$, let $\mathcal{C}$ be a binary linear code with relative distance $\frac{1}{4} + \delta$, and block length $2n$. Then the columns of the parity check matrix of H form a (r, s)-spanning set, with $r = n$ and $s = n - \log\left(\frac{1}{\delta}\right)$.*

Proof. Using Lemma 5.3, it follows that $\mathcal{C}$ is $(n, \frac{1}{\delta})$-erasure list-decodable. Now applying Lemma 5.2, the lemma follows directly. □

A similar result follows for the case of q-ary linear codes.

Lemma 5.5. *For any $\delta > 0$, let $\mathcal{C}$ be a linear code with relative distance $\frac{q-1}{2q} + \delta$ and block length $2n$ over a finite field of size q. Then the columns of the parity check matrix of H form a (r, s)-spanning set, with $r = n$ and $s = n - \log\left(\frac{q}{(q-1)\delta}\right)$.*

To instantiate the above results, we recall some explicit code constructions. Using standard code concatenation, there are known constructions of binary linear codes achieving the Zyablov bound.

Theorem 5.6. *For any $\epsilon, \gamma > 0$, there exists an explicit construction of a binary linear code with relative distance $\delta = \frac{1}{4} + \epsilon$ and relative rate $R \geq \max_{0<r<1-H(\delta+\epsilon)} r\left(1 - \frac{\delta}{H^{-1}(1-r)-\epsilon}\right)$.*

Over larger alphabets, the following explicit codes were constructed in the work of Guruswami and Indyk [GI02].

Theorem 5.7 ([GI02])**.** *There exists $c > 0$ such that for every $\gamma > 0$ and any prime $p > 2^{\frac{c}{\gamma}}$ there is an efficient construction of a linear code $C \subset \mathbb{F}_p^n$ with relative distance $\delta = \frac{1}{2} - \frac{1}{4p}$ and rate $R = \frac{1}{2} - \gamma$.*

Using the above codes, we now have explicit constructions of spanning sets.

Lemma 5.8. *There exist constants $\gamma > 0$ and c such that for any n, there exists an explicit $(n, n-c)$-spanning set $S \subset \mathbb{F}_{2^{\bar{n}}}$ of size $2n$, where $\bar{n} = 2n(1-\gamma)$.*

Proof. Let H be the parity check matrix of the explicit linear code $C \subset \mathbb{F}_2^{2n}$ from Theorem 5.6 for relative distance $\frac{1}{4} + \delta$, for some small constant δ. Let $S = \{v_1, \ldots, v_{2n}\}$ be the set of columns of H. Thus $S \subset \mathbb{F}_2^{\bar{n}}$, $\bar{n} = 2n(1-\gamma)$, γ being the relative rate of the code. Applying Lemma 5.4, the result is now immediate. □

Lemma 5.9. *There exists $c > 0$ such that for any $\gamma > 0$ and any prime $p > 2^{\frac{c}{\gamma}}$, there is an efficient construction of an explicit set $(n, n-C)$-spanning set $S \subset \mathbb{F}_{2^{\bar{n}}}$ of size $2n$, where $\bar{n} = n(1+2\gamma)$ and $C = \frac{2c}{\gamma}$.*

Proof. Let H be the parity check matrix of the explicit linear code $C \subset \mathbb{F}_p^{2n}$ from Theorem 5.7 with relative distance $\frac{1}{2} - \frac{1}{4p}$ and rate $\frac{1}{2} - \gamma$. Let $S = \{v_1, \ldots, v_{2n}\}$ be the set of columns of H. The result now follows by Lemma 5.5. □

We show that random sets are (r, s)-spanning sets with overwhelmingly high probability. Guruswami's existence proof of subspace evasive [Gur11] targets different parameters and does not apply here. This lemma is more related to the existence of good erasure list-decodable codes.

Lemma 5.10. *Let S be a random subset of $\mathbb{F}_2^n$ of size $2n$. Then,*

$$\Pr[S \text{ is not a } (n, n-2\sqrt{n})\text{-spanning set }] \leq 2^{-n}.$$

Proof. Let $t > 0$. Consider any subset $R \subset S$, $|R| = n$. By standard arguments, it follows that

$$\Pr[\dim(\mathrm{span}(R)) \leq n - t] \leq \binom{n}{t}(2^{-t})^t \leq \left(\frac{n}{2^t}\right)^t.$$

Thus,

$$\Pr[\exists\ R \subset S,\ |R| = n \text{ with } \dim(\mathrm{span}(R)) \leq n - t] \leq \binom{2n}{n}\left(\frac{n}{2^t}\right)^t \leq 2^{2n - t^2 + t\log n}$$

The lemma follows by setting $t = 2\sqrt{n} + 1$. □

6 Extractors for 2-Interleaved Sources

6.1 Extractors for 2-Interleaved Sources on $\{0,1\}^{2n}$

Our extractor constructions are based on encoding the interleaved-sources using spanning vectors. Recall that any (r,s)-encoding from $\mathbb{F}_p^{\ell} \to \mathbb{F}_p^{\bar{\ell}}$ is defined in the following way: For any (r,s)-spanning set $S = \{v_1, \ldots, v_\ell\} \subseteq \mathbb{F}_p^{\bar{n}}$, the function $\mathrm{enc} : \mathbb{F}_p^{\ell} \to \mathbb{F}_p^{\bar{\ell}}$ defined as

$$\mathrm{enc}(z) = \sum_{i=1}^{n} z_i v_i$$

is an (r,s)-encoding from $\mathbb{F}_p^{\ell} \to \mathbb{F}_p^{\bar{\ell}}$.

The following is a key lemma in our extractor constructions.

Lemma 6.1 (Main Lemma)**.** *Fix any $\delta > 0$. Let p be any prime and let $Z = (X \circ Y)_t$ be any 2-interleaved source on $\mathbb{F}_p^{2n}$, where X and Y are independent sources on $\mathbb{F}_p^{n}$ with min-entropy k_1 and k_2 respectively, and $t : [2n] \to [2n]$ is any permutation. Also suppose χ is any nonprincipal multiplicative character of $\mathbb{F}_{p^{\bar{n}}}^{*}$ and* enc *is an arbitrary (n,s)-encoding from $\mathbb{F}_p^{2n}$ to $\mathbb{F}_p^{\bar{n}}$. Then,*

$$\mathbb{E}_X |\mathbb{E}_Y[\chi(\mathrm{enc}(Z))]| \leq 2^{-\delta(k_2 - (n-s)\log p)},$$

whenever

- $k_1 \geq (\frac{1}{2} + 3\delta)\bar{n}\log p + (n-s)\log p$*, and*
- $k_2 \geq 4\log\bar{n}\log p + (n-s)\log p$*.*

Proof. For any $z \in \mathbb{F}_p^{2n}$, let

$$\mathrm{enc}(z) = \sum_{i=1}^{2n} z_i v_i$$

where $S = \{v_1, \ldots, v_{2n}\} \subset \mathbb{F}_p^{\bar{n}}$ is (n,s)-spanning.

We have,

$$\chi(\mathrm{enc}(Z)) = \chi\left(\sum_{i=1}^{2n} Z_i v_i\right) = \chi\left(\sum_{i=1}^{n} X_i v_{t(i)} + \sum_{j=1}^{n} Y_j v_{t(n+j)}\right)$$

Define the following independent sources:

$$X' = \sum_{i=1}^{n} x_i v_{t(i)} : x \sim X, \qquad Y' = \sum_{j=1}^{n} y_j v_{t(n+j)} : y \sim Y.$$

Using Lemma 4.9, it follows that: $k_1' = H_\infty(X') \geq k_1 - (n-s)\log p$ and $k_2' = H_\infty(Y') \geq k_2 - (n-s)\log p$.

Thus, we have

$$\begin{aligned}\mathbb{E}_X|\mathbb{E}_Y[\chi(\mathrm{enc}(Z))]| &= \mathbb{E}_{x\sim X}\left|\mathbb{E}_{y\sim Y}\left[\chi\left(\sum_{i=1}^{n} x_i v_{t(i)} + \sum_{j=1}^{n} y_j v_{t(n+j)}\right)\right]\right| \\ &= \mathbb{E}_{X'}\left|\mathbb{E}_{Y'}\left[\chi\left(X'+Y'\right)\right]\right| \\ &= 2^{-\delta k_2'}\end{aligned}$$

where the last inequality follows using Theorem 4.2. □

Using the above main lemma, we construct extractors for 2-interleaved sources on $\mathbb{F}_2^{2n}$.

Theorem 6.2. *For some $\delta > 0$ and any $\lambda > 0$, there exists an explicit function* ext : $\{0,1\}^{2n} \to [M]$, *$M = n^\lambda$, such that if X and Y are independent sources on $\mathbb{F}_2^n$ with min-entropy k_1, k_2 respectively satisfying $k_1 > (1-\delta)n$ and $k_2 > 10\max\{\log n, \log M\}$, $t : [2n] \to [2n]$ is any permutation, then*

$$|\mathrm{ext}((X \circ Y)_t) \circ X - U_M \circ X| = 2^{-\Omega(k_2)}.$$

Proof. Let H be the parity check matrix of a code $C \subset \mathbb{F}_2^{2n}$ with relative distance $= \frac{1}{4} + \delta_1$ (for some small constant δ_1) and constant rate R, where we fix R as follows. Let R_Z be the rate of the code from Theorem 5.6. Let $\epsilon_1 << R_Z$ be a small constant. We choose R in the interval $[R_Z - \epsilon_1, R_Z]$ such that $\bar{n} = 2n(1-R)$ is divisible by integer m, $m = \lambda \log n$. Since $2R_Z\epsilon_1 n >> m$, we can indeed find such an R. Fix $M = 2^m - 1$. We note that $M|2^{\bar{n}} - 1$. Set $\delta = \frac{R}{6}$.

Let $S = \{v_1, \ldots, v_{2n}\}$ be the set columns of H. By Lemma 5.8, S is $(n, n-C)$-spanning, for some constant C. We interpret each v_i as being an element in the field $\mathbb{F}_{2^{\bar{n}}}$. Consider the multiplicative subgroup:

$$G = \{x^{\frac{2^{\bar{n}}-1}{M}} : x \in \mathbb{F}_{2^{\bar{n}}}^*\}.$$

A generator g of G can be found efficiently in the following way: Using Theorem 4.10, we can efficiently construct a set $S = \{a_1, \ldots, a_l\}$, $l = \mathrm{poly}(n)$, such that one of the a_i's, say a_j, is a primitive element of $\mathbb{F}_{2^{\bar{n}}}$. Let $S' = \{a_1^{\frac{2^{\bar{n}}-1}{M}}, \ldots, a_l^{\frac{2^{\bar{n}}-1}{M}}\}$. We note that $a_j^{\frac{2^{\bar{n}}-1}{M}} \in S'$ is an element of order M. Thus, it is enough to enumerate over the elements in S' and compute the order of each element. Since the order of any element in S' is bounded by $M = \mathrm{poly}(n)$, the search procedure can be implemented efficiently.

Let $Z = (X \circ Y)_t$. For any $z \in \mathbb{F}_2^{2n}$, define the functions:

$$\mathrm{enc}(z) = \sum_{i=1}^{2n} z_i v_i, \qquad \mathrm{ext}_1(z) = (\mathrm{enc}(z))^{\frac{2^{\bar{n}}-1}{M}}, \qquad \mathrm{ext}(z) = \log_g(\mathrm{ext}_1(z)).$$

We note that ext_1 and ext are efficiently computable functions. Further note that enc is an $(n, n-C)$-encoding from $\mathbb{F}_2^{2n}$ to $\mathbb{F}_2^{\bar{n}}$.

Using the above lemma, we prove the following claim.

Claim 6.3. *Let $\psi(x) = e_M(\beta x)$, $\beta \neq 0 \pmod M$, be any nontrivial character of the additive group $\mathbb{Z}_M$.*

Then,

$$\mathbb{E}_X \left| \mathbb{E}_Y [\psi(\mathrm{ext}_2((X \circ Y)_t))] \right| \leq 2^{-\delta k_2}.$$

We note that Theorem 6.2 follows directly from Claim 6.3 by using Lemma 4.4. Thus it is enough to prove Claim 6.3.

Proof of Claim 6.3. We have,

$$\begin{aligned} \psi(\mathrm{ext}(z)) &= e_M(\beta \log_g(\mathrm{ext}_1(z))) \\ &= \chi\left(\mathrm{enc}(z)\right), \end{aligned}$$

where $\chi(x) = e_M(\beta \log_g(x))$ is a nonprincipal multiplicative character of $\mathbb{F}^*_{2^{\bar{n}}}$ of order $\frac{M}{\gcd(M,\beta)}$. Thus, we have

$$\begin{aligned} \mathbb{E}_X \left| \mathbb{E}_Y [\psi(\mathrm{ext}_2((X \circ Y)_t))] \right| &= \mathbb{E}_{x \sim X} \left| \mathbb{E}_{y \sim Y} \left[\chi\left(\mathrm{enc}(Z)\right)\right] \right| \\ &\leq 2^{-\delta k_2}, \end{aligned}$$

where the inequality follows from Lemma 6.1. □

□

It is direct from the above theorem, that if we insist that the output of the above extractor is a bit string, we have the following result.

Theorem 6.4. *For some $\delta > 0$ and any $\lambda > 0$, there exists an explicit function $\mathrm{ext} : \{0,1\}^{2n} \to \{0,1\}^m$, $m = \lambda \log n$, such that if X, Y are independent sources on $\mathbb{F}_2^n$ with min-entropy k_1, k_2 respectively satisfying $k_1 > (1-\delta)n$ and $k_2 > 10 \max\{\log n, m\}$, $t : [2n] \to [2n]$ is any permutation, then*

$$\left| \mathrm{ext}((X \circ Y)_t) \circ X - U_m \circ X \right| = n^{-\Omega(1)}.$$

6.2 Extracting from 2-Interleaved Sources on $\mathbb{F}_p^{2n}$

If the sources X and Y are on $\mathbb{F}_p^n$ (for some large enough prime p), we can reduce the min-entropy rate requirement of the source X to about $\frac{1}{2}$.

Theorem 6.5. *There exists $c > 0$ such that for any $\delta, \lambda > 0$ and any prime $p > 2^{\frac{c}{\delta}}$, there exists an explicit function $\mathrm{ext}_p : \mathbb{F}_p^{2n} \to \{0,1\}^m$, $m = \lambda \log n$, such that if X and Y are independent sources on $\mathbb{F}_p^n$ with min-entropy k_1, k_2 respectively, satisfying $k_1 > (\frac{1}{2} + \delta) n \log p$ and $k_2 > \max\{5 \log n \log p, \frac{3m}{\delta}\}$, $t : [2n] \to [2n]$ is any injective map, then*

$$\left| \mathrm{ext}_p((X \circ Y)_t) \circ X - U_m \circ X \right| = n^{-\Omega(1)}.$$

Proof. Let $S = \{v_1, \ldots, v_{2n}\}$ be an explicit (n,n-C)-spanning set in $\mathbb{F}_p^{\bar{n}}$ from Lemma 5.9. Further, as in the proof of Theorem 6.2, we choose the rate of the code in Lemma 5.9 such that $m|\bar{n}$ and $m = \lambda \log_p n$. Thus we can ensure that $\bar{n} \leq n(1 + \frac{\delta}{5})$.

Let $M = n^\lambda$. For any $z \in \mathbb{F}_p^{2n}$, define the functions:

$$\text{enc}(z) = \sum_{i=1}^{2n} z_i v_i, \qquad \text{ext}_1(z) = (\text{enc}(z))^{\frac{p^{\bar{n}}-1}{M}}, \qquad \text{ext}(z) = \log_g(\text{ext}_1(z))$$

where g is a generator of $G = \{x^{\frac{p^{\bar{n}}-1}{M}} : x \in \mathbb{F}_{p^{\bar{n}}}^*\}$. The proof now follows using Lemma 6.1 and Lemma 4.4. □

6.3 Improving the Output Length

The output length of the extractor in Theorem 6.5 is $\Omega(\log n)$. We improve the output length to $\Omega(n)$ bits when the min-entropy rate of both the sources (on $\mathbb{F}_p^n$) are slightly more than $\frac{1}{2}$.

A general technique to improve the output length extractors was introduced by Shaltiel [Sha06]. In particular, Shaltiel showed that the function:

$$\text{SExt}(X, 2\text{ext}(X, Y)) \circ \text{SExt}(Y, 2\text{ext}(X, Y))$$

is 2-source extractor with longer output length, where 2ext is a 2-source extractor with short output length and SExt is a seeded extractor set to appropriate parameters.

However this does not work in our case since it requires access to the individual sources X and Y. Surprisingly, we show that the construction: $\text{SExt}(((X \circ Y)_t)_{[n]}, 2\text{ext}_p((X \circ Y)_t))$ can be proved to be an extractor.

Theorem 6.6. *There exists $c > 0$ such that for any $\delta > 0$ and any prime $p > 2^{\frac{c}{\delta}}$, there exists an explicit function $\text{ext}_{p,long} : \mathbb{F}_p^{2n} \to \{0,1\}^m$, $m = \Omega(n)$, such that if X and Y are independent sources on $\mathbb{F}_p^n$ with min-entropy k_1, k_2 respectively satisfying $k_1 > (\frac{1}{2} + \delta)n \log p$ and $k_2 > (\frac{1}{2} + \delta)n \log p$, $t : [2n] \to [2n]$ is any injective map, then*

$$|\text{ext}_{p,long}((X \circ Y)_t) - U_m| = n^{-\Omega(1)}.$$

Proof. Let SExt be the seeded-extractor from Theorem 4.3 with parameters $\beta = \delta$, $\alpha = \delta/2$ and $\epsilon = n^{-\Omega(1)}$. Let the seed length of SExt with this setting of the parameters be $d = \lambda \log n$. Let $Z = (X \circ Y)_t$. Define

$$\text{ext}_{p,long}(Z) = \text{SExt}(Z_{[n]}, \text{ext}_p(Z)),$$

where ext_p is the extractor from Theorem 6.5 designed to extract from 2-interleaved sources with one source at min-entropy $k_1 \geq (\frac{1}{2} + \delta)n \log p$ and the other source with min-entropy $k_2 \geq \frac{\delta n \log p}{2}$ with error $\epsilon_p = n^{-2\lambda}$ and output length $m_p = \lambda \log n$.

Let $S = \{i \in [n] : Z_i = X_i\}$ and $T = \{j \in [n] : Z_j = Y_j\}$. Also let $\bar{S} = [n] \setminus S$ and $\bar{T} = [n] \setminus T$. Without loss of generality, we can assume that $|S| \geq \frac{n}{2}$. It follows from Lemma 4.7 that there exists a set $Good_y$ such that for any $y_T \in Good_y$, $Y_{\bar{T}}|Y_T = y_T$ is $2^{-\Omega(n)}$-close to a source with entropy more than $\frac{\delta n \log p}{2}$, and $\Pr[Y_t \in Good_y] > 1 - 2^{-\Omega(n)}$.

Let $y_T \in Good_y$. It follows by the setting of ext_p that

$$|(\mathrm{ext}_p(Z|Y_T = y_T) \circ X_S - U_m \circ X_S| \leq n^{-2\lambda}.$$

Using Lemma 4.8, it follows that

$$|X_S - (X_S|(\mathrm{ext}_p(Z|Y_T = y_T) = e))| \leq n^{-\lambda+1}. \tag{1}$$

Let $p_{y_T} = \Pr[Y_T = y_T]$ and let $p_{e|y_T} = \Pr[\mathrm{ext}_p(Z|Y_T = y_T) = e]$.

Using the above estimates, we have

$$\begin{aligned}
|\mathrm{ext}_{p,long}(Z) - U_m| &\leq \sum_{y_T} p_{y_T} |\mathrm{SExt}(X_S \circ y_T, \mathrm{ext}_p(Z|Y_T = y_T)) - U_m| \\
&\leq \left(\sum_{y_T \in Good_y} p_{y_T} |\mathrm{SExt}(X_S \circ y_T, \mathrm{ext}_p(Z|Y_T = y_T)) - U_m| \right) + 2^{-\Omega(n)} \\
&\leq \sum_{y_T \in Good_y} p_{y_T} \left(\sum_e p_{e|y_T} |\mathrm{SExt}(X_S \circ y_T, e) - U_m| + n^{-\lambda+1} \right) + 2^{-\Omega(n)} \\
&\leq \left(\sum_{y_T \in Good_y} p_{y_T} |\mathrm{SExt}(X_S \circ y_T, U_d) - U_m| \right) + n^{-\Omega(1)} \\
&= n^{-\Omega(1)}.
\end{aligned}$$

where the last line follows from the fact that X_S has min-entropy at least $\delta n \log p$. □

6.4 One Bit Extractors for 2-Interleaved Sources on $\mathbb{F}_p^{2n}$ with Exponentially Small Error

Note that all our extractor constructions so far have polynomially small error if we insist that the output of the extractor is a bit string. Here we show how to achieve exponentially small error for 2-interleaved sources on $\mathbb{F}_p$, for any large enough prime. However we can output only 1 bit.

Theorem 6.7. *There exists $c > 0$ such that for any $\delta > 0$ and any prime $p > 2^{\frac{c}{\delta}}$, there exists an explicit function $\mathrm{ext}_{1bit} : \mathbb{F}_p^{2n} \to \{0,1\}$, such that if X and Y are independent sources on $\mathbb{F}_p^n$ with min-entropy k_1, k_2 respectively, satisfying $k_1 > (\frac{1}{2} + \delta) n \log p$ and $k_2 > 5 \log n \log p$, $t : [2n] \to [2n]$ is any injective map, then*

$$|\mathrm{ext}_{1bit}((X \circ Y)_t) \circ X - U_1 \circ X| = 2^{-\Omega(k_2)}.$$

Proof. Let $S = \{v_1, \ldots, v_{2n}\}$ be an explicit (n,n-C)-spanning set in $\mathbb{F}_p^{\bar{n}}$ from Lemma 5.9. Define the functions:

$$\mathrm{enc}(z) = \sum_{i=1}^{2n} z_i v_i, \qquad \mathrm{ext}(z) = \mathrm{QR}\left(\mathrm{enc}(z)\right),$$

where QR is the quadratic character of $\mathbb{F}_{p^{\bar{n}}}^*$. The proof now follows using Lemma 6.1. □

6.5 Semi-Explicit Extractors for 2-Interleaved Sources with Linear Output Length and Exponentially Small Error

We note that the extractors constructed so far have either achieved linear output length or exponentially small error, but not both simultaneously. We show that if we allow the extractors to run in sub-exponential time, then we can indeed construct such extractors. (Note that the trivial algorithm to find such an extractor runs in doubly exponential time.)

Theorem 6.8. *For some $\delta > 0$, there exists a semi-explicit function* $\text{ext} : \{0,1\}^{2n} \to \{0,1\}^m$, *such that if X and Y are independent sources on $\mathbb{F}_2^n$ with min-entropy k_1, k_2 respectively satisfying $k_1 > (1-\delta)n$ and $k_2 > 10\max\{\log n, m\}$, $t : [2n] \to [2n]$ is any permutation, then*

$$|\text{ext}((X \circ Y)_t) \circ X - U_m \circ X| = 2^{-\Omega(k_2)}.$$

Proof. Let $S = \{v_1, \ldots, v_{2n}\}$ be an explicit $(n, n-C)$-spanning set in $\mathbb{F}_2^{\bar{n}}$ constructed using Lemma 5.8. Let $m = \frac{\delta k_2}{2}$. For any $z \in \mathbb{F}_p^{2n}$, define the functions:

$$\text{enc}(z) = \sum_{i=1}^{2n} z_i v_i, \qquad \text{ext}_1(z) = \log_g(\text{enc}(z)), \qquad \text{ext}(z) = \text{ext}_1(z) \pmod{2^m}$$

where g is a generator of $\mathbb{F}_{2^{\bar{n}}}^*$. The proof now follows using Lemma 6.1 and Lemma 4.5. □

Using the $(n, n-C)$-spanning sets from Lemma 5.9 to encode the sources, we obtain the following theorem using Lemma 6.1.

Theorem 6.9. *There exists $c > 0$ such that for any $\delta > 0$ and any prime $p > 2^{\frac{c}{\delta}}$, there exists a semi-explicit function* $\text{ext} : \mathbb{F}_p^{2n} \to \{0,1\}^m$, *such that if X, Y are independent sources on $\mathbb{F}_p^n$ with min-entropy k_1, k_2 respectively satisfying $k_1 > (\frac{1}{2} + \delta)n \log p$ and $k_2 > \max\{5 \log n \log p, \frac{3m}{\delta}\}$, $t : [2n] \to [2n]$ is any permutation, then*

$$|\text{ext}((X \circ Y)_t) \circ X - U_m \circ X| = 2^{-\Omega(k_2)}.$$

6.6 Extractors for 2-Interleaved Sources with Linear Min-Entropy Under the Generalized Paley Graph Conjecture

In this section, we show how to construct extractors for sources with linear min-entropy under the widely believed Generalized Paley Graph Conjecture.

Generalized Paley Graph Conjecture. *Let χ be any non-principal multiplicative character of $\mathbb{F}_{p^n}^*$. For any constant $\delta > 0$, and arbitrary subsets $A, B \subseteq \mathbb{F}_{p^n}$ satisfying $|A|, |B| > p^{\delta n}$, we have*

$$\left| \sum_{a \in A, b \in B} \chi(a+b) \right| \leq p^{-\gamma(\delta)n} |A||B|.$$

Assuming the above conjecture, we obtain the following improved version of Lemma 6.1.

Lemma 6.10. *Assume the Generalized Paley graph Conjecture. Fix any $\delta > 0$ and any prime p. Let $Z = (X \circ Y)_t$ be any 2-interleaved source on $\mathbb{F}_p^{2n}$, where X and Y are independent sources on $\mathbb{F}_p^n$ with min-entropy k_1 and k_2 respectively, and $t : [2n] \to [2n]$ is any permutation. Also suppose χ is any nonprincipal multiplicative character of $\mathbb{F}_{p^{\bar{n}}}^*$ and* enc *is an arbitrary (n, s)-encoding from $\mathbb{F}_p^{2n}$ to $\mathbb{F}_p^{\bar{n}}$. Then, there exists $\gamma = \gamma(\delta)$ such that*

$$\mathbb{E}_X |\mathbb{E}_Y[\chi(\text{enc}(Z))]| \leq p^{-\gamma n},$$

whenever

- $k_1 \geq \delta \bar{n} \log p + (n - s) \log p$, *and*
- $k_2 \geq \delta \bar{n} \log p + (n - s) \log p$.

Proof. For any $z \in \mathbb{F}_p^{2n}$, let

$$\text{enc}(z) = \sum_{i=1}^{2n} z_i v_i$$

where $S = \{v_1, \ldots, v_{2n}\} \subset \mathbb{F}_p^{\bar{n}}$ is (n, s)-spanning.

We have,

$$\chi(\text{enc}(Z)) = \chi\left(\sum_{i=1}^{2n} Z_i v_i\right) = \chi\left(\sum_{i=1}^{n} X_i v_{t(i)} + \sum_{j=1}^{n} Y_j v_{t(n+j)}\right)$$

Define the following independent sources:

$$X' = \sum_{i=1}^{n} x_i v_{t(i)} : x \sim X, \qquad Y' = \sum_{j=1}^{n} y_j v_{t(n+j)} : y \sim Y.$$

Using Lemma 4.9, it follows that: $H_\infty(X') \geq k_1 - (n - s) \log p$ and $H_\infty(Y') \geq k_2 - (n - s) \log p$. Thus, we have

$$\begin{aligned}
\mathbb{E}_X |\mathbb{E}_Y[\chi(\text{enc}(Z))]| &= \mathbb{E}_{x \sim X} \left| \mathbb{E}_{y \sim Y} \left[\chi\left(\sum_{i=1}^{n} x_i v_{t(i)} + \sum_{j=1}^{n} y_j v_{t(n+j)} \right) \right] \right| \\
&= \mathbb{E}_{X'} \left| \mathbb{E}_{Y'} \left[\chi\left(X' + Y'\right) \right] \right| \\
&\leq p^{-\gamma n}
\end{aligned}$$

where the last inequality follows using the Generalized Paley Graph Conjecture. □

Using the above lemma, we have the following theorem.

Theorem 6.11. *Assume the Generalized Paley Graph Conjecture. For any $\delta, \lambda > 0$, there exists an explicit function $\text{ext}_{conjecture} : \{0,1\}^{2n} \to \{0,1\}^m$, $m = \lambda \log n$, such that if X and Y are independent sources with min-entropy δn each, and $t : [2n] \to [2n]$ is any permutation, then*

$$|\text{ext}_{conjecture}((X \circ Y)_t) - U_m| = n^{-\Omega(1)}.$$

Proof. Let $S = \{v_1, \ldots, v_{2n}\}$ be an explicit $(n, n-C)$-spanning set in $\mathbb{F}_p^{\bar{n}}$ constructed using Lemma 5.8. Further, as in the proof of Theorem 6.2, we choose the rate of the code in Lemma 5.9 such that $m|\bar{n}$ and $m = \lambda \log n$. Let $M = n^{\lambda}$. For any $z \in \mathbb{F}_2^{2n}$, define the functions:

$$\mathrm{enc}(z) = \sum_{i=1}^{2n} z_i v_i, \quad \mathrm{ext}_1(z) = (\mathrm{enc}(z))^{\frac{p^{\bar{n}}-1}{M}}, \quad \mathrm{ext}(z) = \log_g(\mathrm{ext}_1(z))$$

where g is a generator of $G = \{x^{\frac{2^{\bar{n}}-1}{M}} : x \in \mathbb{F}_{2^{\bar{n}}}^*\}$. The proof now follows using Lemma 6.10 and Lemma 4.4. □

We note that assuming the above conjecture, the output length of the above extractor can be improved to $\Omega(n)$ if both X and Y have min-entropy rate more than $\frac{1}{4}$ by using the proof method of Theorem 6.6.

7 Interleaved-Non-Malleable Extractors

In this section, we show that the proof technique developed in constructing extractors for 2-interleaved sources can be used to construct non-malleable extractors in the interleaved model.

Theorem 7.1. *There exists $\lambda_1 > 0$ such that for any $\delta, \lambda_2 > 0$, $c > c(\delta)$ and any prime $p > 2^{\frac{\lambda_1}{\delta}}$, there exists an explicit function $\mathrm{nmExt} : \mathbb{F}_p^{2n} \to \{0,1\}^m$, $m = \lambda_2 \log n$, such that if X, Y are independent sources on $\mathbb{F}_p^n$ with min-entropy k_1, k_2 respectively, satisfying $k_1 > (\frac{1}{2} + \delta) n \log p$ and $k_2 > c \max\{m, \log n\}$, $t : [2n] \to [2n]$ is any injective map and $f : \mathbb{F}_p^n \to \mathbb{F}_p^n$ is any function with no fixed points, then*

$$|\mathrm{nmExt}((X \circ Y)_t) \circ \mathrm{nmExt}((X \circ f(Y))_t) \circ Y - U_m \circ \mathrm{nmExt}((X \circ f(Y))_t) \circ Y| = n^{-\Omega(1)}.$$

To prove the above theorem, we recall a character sum estimate of Dodis et al. [DLWZ14].

Theorem 7.2. *For any $\delta > 0$ and $\eta < \frac{1}{2}$, suppose S and T are non-empty subsets of $\mathbb{F}_q$ satisfying $|S| > q^{\frac{1}{2}+\delta}$ and $|T| > \max\{(\frac{1}{\eta})^{\frac{7}{\delta}}, (\log q)^8\}$. Let $f : \mathbb{F}_q \to \mathbb{F}_q$ be any arbitrary function with no fixed points. For arbitrary multiplicative characters χ_a and χ_b, such that χ_a is nonprincipal, we have*

$$\sum_{y \in T} \left| \sum_{x \in S} \chi_a(x+y)\chi_b(x+f(y)) \right| < \eta |S||T|.$$

Proof of Theorem 7.1. We use encoding based on spanning vectors. In particular, let $S = \{v_1, \ldots, v_{2n}\}$ be an explicit $(n, n-C)$-spanning set in $\mathbb{F}_p^{\bar{n}}$ constructed using Lemma 5.9. Further, as in the proof of Theorem 6.2, we choose the rate of the code in Lemma 5.9 such that $m|\bar{n}$ and $m = \lambda_2 \log_p n$. Let $M = n^{\lambda_2}$. For any $z \in \mathbb{F}_p^{2n}$, define the functions:

$$\mathrm{enc}(z) = \sum_{i=1}^{2n} z_i v_i, \quad \mathrm{ext}_1(z) = (\mathrm{enc}(z))^{\frac{p^{\bar{n}}-1}{M}}, \quad \mathrm{ext}(z) = \log_g(\mathrm{ext}_1(z))$$

where g is a generator of $G = \{x^{\frac{p^{\bar{n}}-1}{M}} : x \in \mathbb{F}_{p^{\bar{n}}}^*\}$. We prove the following claim.

Claim 7.3. *Let ψ_a and ψ_b be arbitrary characters of the additive group $\mathbb{Z}_M$ such that ψ_a is non-trivial. Then,*

$$\mathbb{E}_{y\sim Y}\left|\mathbb{E}_{x\sim X}[\psi_a(\mathrm{nmExt}((X\circ Y)_t))\psi_b(\mathrm{nmExt}((X\circ f(Y))_t))]\right| = n^{-\Omega(1)}.$$

Before proving this claim, we note that Theorem 7.1 follows directly from Claim 7.3 by using Lemma 4.6.

Proof of Claim 7.3. Let $t([n]) = T_1$ and $t([n+1,2n]) = T_2$. Since S is (n,n)-spanning, it follows that the set $\{v_i : i \in T_1\}$ consists of linearly independent vectors. Similarly $\{v_j : j \in T_2\}$ is a set of linearly independent vectors.

Let $\psi_a(x) = e_M(ax)$, where $a \neq 0 \pmod M$. Also let $\psi_b(x) = e_M(bx)$. If $b = 0 \pmod M$, the claim follows from Lemma 6.1. Thus suppose $b \neq 0 \pmod M$.

We have,

$$\begin{aligned}\psi_a(\mathrm{nmExt}((X\circ Y)_t) &= e_M(a\log_g(\mathrm{ext}_1((X\circ Y)_t)))\\ &= \chi_a\left(\sum_{i=1}^{n} X_i v_{t(i)} + \sum_{j=1}^{n} Y_j v_{t(n+j)}\right)\\ &= \chi_a\left(X' + Y'\right)\end{aligned}$$

where $\chi_a(x) = e_M(a\log_g(x))$ is a nonprincipal multiplicative character of $\mathbb{F}^*_{p^{\bar{n}}}$ of order $\frac{M}{\gcd(M,a)}$, $X' = \sum_{i=1}^{n} x_i v_{t(i)} : x \sim X$ and $Y' = L(Y)$, $L : \mathbb{F}_p^n \to \mathbb{F}_p^{\bar{n}}$ being the injective linear map:

$$L(y) = \sum_{j=1}^{n} y_j v_{t(n+j)}.$$

Further,

$$\begin{aligned}\psi_b(\mathrm{nmExt}((X\circ f(Y))_t) &= e_M(b\log_g(\mathrm{ext}_1((X\circ Y)_t)))\\ &= \chi_b\left(\sum_{i=1}^{n} X_i v_{t(i)} + \sum_{j=1}^{n} f(Y)_j Y_{t(n+j)}\right)\\ &= \chi_b\left(X' + f'(Y')\right)\end{aligned}$$

where $f' = L\circ f\circ L^{-1}$ and $\chi_b(x) = e_M(b\log_g(x))$ is a nonprincipal multiplicative character of $\mathbb{F}^*_{p^{\bar{n}}}$ of order $\frac{M}{\gcd(M,b)}$.

We claim that f' has no fixed points. This can be proved in the following way. Suppose $f'(x) = x$ for some x. This implies that $f(L^{-1}(x)) = L^{-1}(x)$ and hence $f(w) = w$ for $w = L^{-1}(x)$. This contradicts our assumption on f. Thus f' has no fixed points.

It now follows from Theorem 7.2 that

$$\mathbb{E}_{x'\sim X'}|\mathbb{E}_{y'\sim Y'}[\chi_a(x'+y')\chi_b(x'+f'(y'))]| = n^{-\Omega(1)}.$$

□

□

If we allow the non-malleable extractor to run in sub-exponential time, then using the proof method of the above theorem, it can be shown that the extractor from Theorem 6.9 is non-malleable. Thus, we have the following result.

Theorem 7.4. *There exists $\lambda > 0$ such that for any $\delta > 0$, $c > c(\delta)$ and any prime $p > 2^{\frac{\lambda}{\delta}}$, there exists a semi-explicit function* $\mathrm{nmExt} : \mathbb{F}_p^{2n} \to \{0,1\}^m$, *$m = \Omega(n)$, such that if X, Y are independent sources on $\mathbb{F}_p^n$ with min-entropy k_1, k_2 respectively, satisfying $k_1 > (\frac{1}{2} + \delta)n \log p$ and $k_2 > c \max\{m, \log n\}$, $t : [2n] \to [2n]$ is any permutation and $f : \mathbb{F}_p^n \to \mathbb{F}_p^n$ is any function with no fixed points, then*

$$|\mathrm{nmExt}((X \circ Y)_t) \circ \mathrm{nmExt}((X \circ f(Y))_t) \circ Y - U_m \circ \mathrm{nmExt}((X \circ f(Y))_t) \circ Y| = 2^{-\Omega(k_2)}.$$

We note that under the Generalized Paley Graph Conjecture, we can reduce the min-entropy requirement of the source X in Theorem 7.1 to βn, for any constant $\beta > 0$.

8 Proof of Theorem 1.6

We briefly recall some definitions from communication complexity. We refer the reader to [KN97] for more background. For convenience, we define boolean functions with range $\{-1, 1\}$ (instead of $\{0, 1\}$).

Definition 8.1. *Let $f : [p]^{2n} \to \{-1, 1\}$ be any function. Fix any equi-partition of $[2n]$ into subsets S, T. For any rectangle R and probability distribution μ on $[p]^{2n}$, denote*

$$\mathrm{Disc}_{S,T}^{\mu,R}(f) = |\Pr_{\mu}[f(x_S, y_T) = 1 \text{ and } (x, y) \in R] - \Pr_{\mu}[f(x_S, y_T) = -1 \text{ and } (x, y) \in R]|.$$

Definition 8.2. *The discrepancy of $f : [p]^{2n} \to \{-1, 1\}$ with respect to an equi-partition of $[2n]$ into S, T and distribution μ on $[p]^{2n}$ is defined as:*

$$\mathrm{Disc}_{S,T}^{\mu}(f) = \left\{ \max_R \left(\mathrm{Disc}_{S,T}^{\mu,R}(f) \right) \right\}.$$

Definition 8.3. *The maximal-equipartition discrepancy of $f : [p]^{2n} \to \{0, 1\}$ with respect to a distribution μ on $[p]^{2n}$ is defined as:*

$$\mathrm{Disc}_{best}^{\mu}(f) = \max_{\substack{S,T:|S|=|T|=n, \\ S \cup T = [2n]}} \left\{ \mathrm{Disc}_{S,T}^{\mu}(f) \right\}.$$

The following theorem provides a method to lower bound randomized best-paritition communication complexity of f using its maximal-equi-partition discrepancy. A proof can be found in [KN97].

Theorem 8.4. *For every function $f : [p]^{2n} \to \{-1, 1\}$, every probability distribution μ on $[p]^{2n}$ and every $\epsilon \geq 0$,*

$$R^{best, \frac{1}{2} - \epsilon}(f) \geq \log \left(\frac{2\epsilon}{\mathrm{Disc}_{best}^{\mu}(f)} \right).$$

We now prove Theorem 1.6.

Proof of Theorem 1.6. We show that the explicit extractor from Theorem 6.7 is the required function. Recall the construction of the extractor.

Let $S = \{v_1, \ldots, v_{2n}\}$ be an explicit (n,n-C)-spanning set in $\mathbb{F}_p^{\bar{n}}$ constructed using Lemma 5.9, $\bar{n} = n(1 + 4\delta)$.

Define the functions:

$$\mathrm{enc}(z) = \sum_{i=1}^{2n} z_i v_i, \qquad \mathrm{ext}(z) = \mathrm{QR}\left(\mathrm{enc}(z)\right),$$

where QR is the quadratic character of $\mathbb{F}_{p^{\bar{n}}}^*$.

We claim that the randomized best partition discrepancy of ext with error $\frac{1}{2} - p^{-\gamma n}$ is at least $(\frac{1}{4} - \delta - \gamma) n \log p$.

Let μ be the uniform distribution on $[p]^{2n}$.

Claim 8.5. *For any equi-partition of $[2n]$ into disjoint subsets S and T,*

$$\log\left(\frac{1}{\mathrm{Disc}_{S,T}^{\mu}(\mathrm{ext})}\right) \geq \left(\frac{1}{4} - \delta\right) n \log p.$$

We note that the proof of Theorem 1.6 is direct from Claim 8.5 by using Theorem 8.4.

Proof of Claim 8.5. Fix any rectangle $R = X \times Y$, for arbitrary subsets $X, Y \subseteq [p]^n$. We have,

$$\mathrm{Disc}_{S,T}^{\mu,R}(ext) = \frac{|X||Y|}{p^{2n}} \left|\mathbb{E}_{x \in X, y \in Y}[QR\left(\mathrm{enc}(x_S \circ y_T)\right)]\right|$$

We note that if $|X| \leq p^{\frac{3n}{4}}$ or $|Y| \leq p^{\frac{3n}{4}}$, the claim follows easily.

Thus suppose $|X|, |Y| > p^{\frac{3n}{4}}$. We abuse notation and also use X, Y to denote the flat distributions supported on the sets X and Y respectively. Define the distribution $Z = (X \circ Y)_\pi$, where $\pi : [2n] \to [2n]$ is a permutation defined in the following way: Let $S = \{s_1, \ldots, s_n\}$ and $T = \{t_1, \ldots, t_n\}$ such that $s_1 \leq \ldots \leq s_n$ and $t_1 \leq \ldots \leq t_n$. For any $i \in [n]$, define $\pi(i) = s_i$ and for any $j \in [n+1, 2n]$, define $\pi(j) = t_j$ (thus, $\pi([n]) = S$ and $\pi([n+1, 2n]) = T$).

We note that enc is an (n, n)-encoding from $\mathbb{F}_p^{2n} \to \mathbb{F}_p^{\bar{n}}$. Thus,

$$\mathrm{enc}(Z) = X' + Y',$$

where X' and Y' are independent sources on $\mathbb{F}_p^{\bar{n}}$ with $H_\infty(X') = \log(|X|)$ and $H_\infty(Y') = \log(|Y|)$.

Using Theorem 4.1, with $\lambda = 1$, we have

$$\left|\mathbb{E}[QR\left(X' + Y'\right))]\right| \leq \left(\frac{p^{\bar{n}}}{|X||Y|}\right)^{\frac{1}{2}} + \left(\frac{p^{\frac{\bar{n}}{2}}}{|X|}\right)^{\frac{1}{2}}$$

Thus,

$$\begin{aligned}\mathrm{Disc}_{S,T}^{\mu,R}(ext) &\leq \left(\frac{|X||Y|}{p^{2n}}\right)\left(\left(\frac{p^{\bar{n}}}{|X||Y|}\right)^{\frac{1}{2}}+\left(\frac{p^{\frac{\bar{n}}{2}}}{|X|}\right)^{\frac{1}{2}}\right)\\ &\leq \frac{|X|^{\frac{1}{2}}|Y|^{\frac{1}{2}}}{p^{2n-\frac{\bar{n}}{2}}}+\frac{|X|^{\frac{1}{2}}}{p^{n-\frac{\bar{n}}{4}}}\\ &\leq p^{-(n-\frac{\bar{n}}{2})}+p^{-\frac{n}{2}+\frac{\bar{n}}{4}}\end{aligned}$$

Since the above estimate holds for any arbitrary rectangle R, we have

$$\log\left(\frac{1}{\mathrm{Disc}_{S,T}^{\mu}(\mathrm{ext})}\right) \geq \left(\frac{1}{4}-\delta\right) n \log p.$$

□

□

Acknowledgements

We thank anonymous referees for helpful comments.

References

[AM86] Noga Alon and Wolfgang Maass. Meanders, Ramsey Theory and Lower Bounds for Branching Programs. In *IEEE Symposium on Foundations of Computer Science*, pages 410–417, 1986.

[BGK06] J. Bourgain, A. A. Glibichuk, and S. V. Konyagin. Estimates for the number of sums and products and for exponential sums in fields of prime order. *Journal of the London Mathematical Society*, 73:380–398, 4 2006.

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

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

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

[CG88] Benny Chor and Oded Goldreich. Unbiased Bits from Sources of Weak Randomness and Probabilistic Communication Complexity. *Siam Journal on Computing*, 17:230–261, 1988.

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

[CGL15] Eshan Chattopadhyay, Vipul Goyal, and Xin Li. Non-malleable extractors and codes, with their many tampered extensions. *CoRR*, abs/1505.00107, 2015.

[CRS12] Gil Cohen, Ran Raz, and Gil Segev. Non-malleable extractors with short seeds and applications to privacy amplification. In *IEEE Conference on Computational Complexity*, pages 298–308, 2012.

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

[DKSS09] Zeev Dvir, Swastik Kopparty, Shubhangi Saraf, and Madhu Sudan. Extensions to the method of multiplicities, with applications to Kakeya sets and mergers. In *FOCS*, pages 181–190, 2009.

[DL12] Zeev Dvir and Shachar Lovett. Subspace evasive sets. In *Proceedings of the forty-fourth annual ACM symposium on Theory of computing*, pages 351–358. ACM, 2012.

[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.

[DW09] Yevgeniy Dodis and Daniel Wichs. Non-malleable extractors and symmetric key cryptography from weak secrets. In *STOC*, pages 601–610, 2009.

[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*, STOC '02, pages 812–821, New York, NY, USA, 2002. ACM.

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

[Gur04] Venkatesan Guruswami. *List Decoding of Error-Correcting Codes (Winning Thesis of the 2002 ACM Doctoral Dissertation Competition)*, volume 3282 of *Lecture Notes in Computer Science*. Springer, 2004.

[Gur11] Venkatesan Guruswami. Linear-algebraic list decoding of folded Reed-Solomon codes. In *Computational Complexity (CCC), 2011 IEEE 26th Annual Conference on*, pages 77–85. IEEE, 2011.

[GUV09] Venkatesan Guruswami, Christopher Umans, and Salil P. Vadhan. Unbalanced expanders and randomness extractors from Parvaresh–Vardy codes. *J. ACM*, 56(4), 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.

[KN97] Eyal Kushilevitz and Noam Nisan. *Communication complexity*. Cambridge University Press, 1997.

[Kon03] Sergei Konyagin. A sum-product estimate in fields of prime order. arXiv:math/0304217, 2003.

[KRVZ11] Jesse Kamp, Anup Rao, Salil P. Vadhan, and David Zuckerman. Deterministic extractors for small-space sources. *Journal of Computer and System Sciences*, 77:191–220, 2011.

[KZ07] Jesse Kamp and David Zuckerman. Deterministic Extractors for Bit-Fixing Sources and Exposure-Resilient Cryptography. *Siam Journal on Computing*, 36:1231–1247, 2007.

[Li12] Xin Li. Non-malleable extractors, two-source extractors and privacy amplification. In *FOCS*, pages 688–697, 2012.

[Li15] Xin Li. Improved constructions of two-source extractors. *Electronic Colloquium on Computational Complexity (ECCC)*, 2015.

[LRVW03] Chi-Jen Lu, Omer Reingold, Salil P. Vadhan, and Avi Wigderson. Extractors: optimal up to constant factors. In *STOC*, pages 602–611, 2003.

[MW97] Ueli M. Maurer and Stefan Wolf. Privacy amplification secure against active adversaries. In *CRYPTO*, pages 307–321, 1997.

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

[Rao07] Anup Rao. An exposition of Bourgain's 2-source extractor. *Electronic Colloquium on Computational Complexity (ECCC)*, 14(034), 2007.

[Raz05] Ran Raz. Extractors with weak random seeds. In *ACM Symposium on Theory of Computing*, pages 11–20, 2005.

[RY11] Ran Raz and Amir Yehudayoff. Multilinear formulas, maximal-partition discrepancy and mixed-sources extractors. *Journal of Computer and System Sciences*, 77:167–190, 2011.

[Sha06] Ronen Shaltiel. How to get more mileage from randomness extractors. In *21st Annual IEEE Conference on Computational Complexity (CCC 2006), 16-20 July 2006, Prague, Czech Republic*, pages 46–60, 2006.

[Sho90] Victor Shoup. Searching for primitive roots in finite fields. In *Proceedings of the 22nd Annual ACM Symposium on Theory of Computing, May 13-17, 1990, Baltimore, Maryland, USA*, pages 546–554, 1990.

[Shp13] Igor E Shparlinski. Additive decompositions of subgroups of finite fields. *SIAM Journal on Discrete Mathematics*, 27(4):1870–1879, 2013.

[SV86] Miklos Santha and Umesh V. Vazirani. Generating quasi-random sequences from semi-random sources. *Journal of Computer and System Sciences*, 33:75–87, 1986.

[TV00] Luca Trevisan and Salil P. Vadhan. Extracting Randomness from Samplable Distributions. In *IEEE Symposium on Foundations of Computer Science*, pages 32–42, 2000.

[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.

[Yao79] Andrew Chi-Chih Yao. Some complexity questions related to distributive computing. In *ACM Symposium on Theory of Computing*, pages 209–213, 1979.

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

ECCC ISSN 1433-8092

http://eccc.hpi-web.de