Online Linear Extractors for Independent Sources

Yevgeniy Dodis
New York University
dodis@cs.nyu.edu

Siyao Guo
New York University Shanghai
siyao.guo@nyu.edu

Noah Stephens-Davidowitz
Cornell University
noahsd@gmail.com

Zhiye Xie
New York University Shanghai
zx572@nyu.edu

Abstract

In this work, we characterize *online linear extractors*. In other words, given a matrix $A \in \mathbb{F}_2^{n \times n}$, we study the convergence of the iterated process $\boldsymbol{S} \leftarrow A\boldsymbol{S} \oplus \boldsymbol{X}$, where $\boldsymbol{X} \sim D$ is repeatedly sampled independently from some fixed (but unknown) distribution D with (min)-entropy at least k. Here, we think of $\boldsymbol{S} \in \{0,1\}^n$ as the *state* of an online extractor, and $\boldsymbol{X} \in \{0,1\}^n$ as its input.

As our main result, we show that the state $\boldsymbol{S}$ converges to the uniform distribution for all input distributions D with entropy $k > 0$ if and only if the matrix A has no non-trivial invariant subspace (i.e., a non-zero subspace $V \subsetneq \mathbb{F}_2^n$ such that $AV \subseteq V$). In other words, a matrix A yields an online linear extractor if and only if A has no non-trivial invariant subspace. For example, the linear transformation corresponding to multiplication by a generator of the field $\mathbb{F}_{2^n}$ yields a good online linear extractor. Furthermore, for any such matrix convergence takes at most $\widetilde{O}(n^2(k+1)/k^2)$ steps.

We also study the more general notion of condensing—that is, we ask when this process converges to a distribution with entropy at least ℓ, when the input distribution has entropy greater than k. (Extractors corresponding to the special case when $\ell = n$.) We show that a matrix gives a good condenser if there are relatively few vectors $\boldsymbol{w} \in \mathbb{F}_2^n$ such that $\boldsymbol{w}, A^T\boldsymbol{w}, \ldots, (A^T)^{n-k-1}\boldsymbol{w}$ are linearly dependent. As an application, we show that the very simple cyclic rotation transformation $A(x_1, \ldots, x_n) = (x_n, x_1, \ldots, x_{n-1})$ condenses to $\ell = n - 1$ bits for any $k > 1$ if n is a prime satisfying a certain simple number-theoretic condition.

Our proofs are Fourier-analytic and rely on a novel lemma, which gives a tight bound on the product of certain Fourier coefficients of any entropic distribution.

1 Introduction

An *extractor* is a deterministic algorithm that takes input $\boldsymbol{X} \sim D$ sampled from some sufficiently nice distribution D and outputs nearly uniformly random $\boldsymbol{Y} \in \{0,1\}^n$. An *online extractor* is a deterministic algorithm with a state $\boldsymbol{S} \in \{0,1\}^n$ that takes inputs $\boldsymbol{X}_1 \sim D_1, \boldsymbol{X}_2 \sim D_2, \ldots, \boldsymbol{X}_m \sim D_m$ one at a time, updating its state after each input. We say that it *extracts* from $D_1, \ldots, D_m$ if the state $\boldsymbol{S}$ is statistically close to random at the end of this process. This naturally models the idea of "gradually accumulating entropy" from entropic sources.

We are interested in perhaps the simplest possible setting, when the $D_i = D$ are independent and identical but otherwise arbitrary entropic distributions over $\{0,1\}^n$, and when the extractor is

linear (over $\mathbb{F}_2$). In other words, on input $\boldsymbol{X} \in \{0,1\}^n$, the state $\boldsymbol{S} \in \{0,1\}^n$ is updated by the procedure

$$\boldsymbol{S} \leftarrow A\boldsymbol{S} \oplus \boldsymbol{X}$$

for some fixed linear transformation $A \in \mathbb{F}_2^{n \times n}$.

We then ask the natural question

Which matrices $A \in \mathbb{F}_2^{n \times n}$ are good extractors?

In other words, for which matrices A does the process $\boldsymbol{S} \leftarrow A\boldsymbol{S} \oplus \boldsymbol{X}$ always converge to uniform when $\boldsymbol{X}$ is sampled independently from *any* distribution with non-zero entropy?

We first notice that there is a natural obstruction that prevents some matrices $A \in \mathbb{F}_2^{n \times n}$ from extracting. As an illustrative example, suppose that A is the "rotation" map defined by $A(x_1, \ldots, x_n) = (x_n, x_1, x_2, \ldots, x_{n-1})$. Then, A clearly fails to extract from the uniform distribution over $\{0^n, 1^n\}$.

More generally, suppose that there exists a subspace $V \subset \mathbb{F}_2^n$ with dimension $0 < \dim(V) < n$ such that $AV \subseteq V$. Such a subspace is called a *non-trivial invariant subspace*. (The trivial invariant subspaces are $\{0^n\}$ and $\mathbb{F}_2^n$.) Then, if X is sampled from the uniform distribution over V, it is not hard to see that the distribution of the state $\boldsymbol{S}$ will itself remain uniform over V after each run of the extractor $\boldsymbol{S} \leftarrow A\boldsymbol{S} \oplus \boldsymbol{X}$. (Here and elsewhere, we assume without loss of generality that the starting state is 0^n.) So, A completely fails to extract from this distribution, even though it clearly has (min-)entropy.

Our main theorem is a proof that this is the only obstruction, i.e., that a matrix A extracts from all entropic distributions *if and only* if A has no non-trivial invariant subspaces. In fact, we show that this property implies that A extracts after relatively few samples, just $\widetilde{O}(n^2(k+1)/k^2)$ samples. (Notice that n/k samples is the best that one could possibly hope for.)

Theorem 1.1 (Informal, see Theorems 4.2 and 4.3)**.** *A matrix $A \in \mathbb{F}_2^{n \times n}$ extracts from arbitrary entropic distributions if and only if A has no non-trivial invariant subspace.*

Specifically, if A has no non-trivial invariant subspace and the input has min-entropy $k > 0$, then the distribution of the state will be 2^{-n}-close to uniform after $m \leq O(n^2(k+1)/k^2 \cdot \log(2n/k))$ steps.

We note that, while the property of having a non-trivial invariant subspace might seem rather opaque, it is efficiently checkable: A has no non-trivial invariant subspace (and thus is a good extractor) if and only if its characteristic polynomial is irreducible [Cla13]. Moreover, there are very sparse matrices A having this property. For example, if A is the linear transformation corresponding to multiplication by a generator of the finite field $\mathbb{F}_{2^n}$, then A is a good extractor which can be easily implemented in time $O(n)$.[1] Thus, we show very simple linear-time, online linear extractors that work for any (unknown) distribution with non-zero min-entropy.

Our proof of Theorem 1.1 is Fourier-analytic; the main technical tool is a novel lemma (Lemma 3.1) concerning certain products of Fourier coefficients of distributions with entropy k. Specifically, for linearly independent $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_r \in \mathbb{F}_2^n$, we give a tight bound on the product of the product of the associated Fourier coefficients. (The worst case is essentially a linear transformation of the uniform distribution over a Hamming ball.)

[1] Indeed, multiplication by the generator corresponds to one cyclic rotation and one conditional XOR with a fixed string corresponding to the coefficients of the irreducible polynomial generating the field.

Online linear condensers. We also consider a more general question. Recall that a *condenser* is a deterministic algorithm that takes as input $\boldsymbol{X} \sim D$ sampled from a sufficiently nice distribution and outputs $\boldsymbol{Y} \in \{0,1\}^n$ that has relatively large entropy (but is not necessarily close to uniform). In our setting, we are interested in the following question.

> *For which matrices A does the process $\boldsymbol{S} \leftarrow A\boldsymbol{S} \oplus \boldsymbol{X}$ converge to a distribution with at least ℓ bits of entropy, whenever X is sampled independently from some (unknown) distribution with more than k bits of entropy?*

Notice that our extractor question from above corresponds to the the special case when $k = 0$ and $\ell = n$.

Here, our result is necessarily a bit more complicated (though the proof is simple and uses the same Fourier-analytic tools). Specifically, we define the *A-rank* of a vector $\boldsymbol{w} \in \mathbb{F}_2^n$ as the dimension of the subspace spanned by $\boldsymbol{w}, A\boldsymbol{w}, \ldots, A^{n-1}\boldsymbol{w}$. Notice that a matrix A has a non-trivial invariant subspace if and only if there is a non-zero vector $\boldsymbol{w} \in \mathbb{F}_2^n$ with A-rank less than n—so that this notion of A-rank is naturally related to the idea of non-trivial invariant subspaces discussed above. And, notice that the obstruction that we ran into with rotation arose from the existence of the vector 1^n with rank equal to 1, which can cause our condenser to "get stuck at one bit of entropy." There is a similar obstruction caused by the uniform distribution over the subspace orthogonal to 1^n (i.e., the subspace of vectors with even Hamming weight) that can cause our condenser to "get stuck at $n-1$ bits of entropy."

More generally, a vector with A-rank r means that "we can get stuck on distributions with entropy r or entropy $n-r$." So, if we are going to condense from k bits to ℓ bits, we must have $k > \min\{n-r, r\}$ and $\ell \leq \max\{r, n-r\}$.

We prove that low-rank vectors are essentially the only possible obstruction to condensing. In particular, a matrix A is a good condenser if it has a small number of vectors with small A-rank. (Again, while this might seem rather opaque, it is easy to count the vectors with a given A-rank by computing the characteristic and minimal polynomials of A [Cla13].) In fact, for technical reasons, it is more natural to study vectors with low A^T-rank, rather than vectors with low A-rank. (Since A^T and A have the same characteristic and minimal polynomials, A-rank and A^T-rank are closely related.)

Theorem 1.2 (Informal, see Theorem 5.4)**.** *For any invertible $A \in \mathbb{F}_2^{n \times n}$, if there are at most N vectors in $\{0,1\}^n$ with A^T-rank less than r, then A condenses any distribution with $k > g := n - r$ bits of min-entropy to a distribution with at least $\ell = n - \log_2 N$ bits of min-entropy. In particular, the state will have entropy at least $\ell - 2^{-n}$ after $m = \widetilde{O}(n^2(k-g+1)/(k-g)^2)$ steps.*

As an application, we show that rotation does in fact condense from $k > 1$ bits of entropy to $n-1$ bits—and that it only requires $m = \widetilde{O}(n^2 k/(k-1)^2)$ steps to do so—when n is a prime satisfying a simple number-theoretic condition.

1.1 Related work

To the best of our knowledge, our question of linear extractors from independent, identically distributed (IID) sources was not explicitly considered by prior work, but several works considered somewhat related models.

The closest such model is our recent prior work [DGSX21], which was motivated by a very practical question of analyzing the bit-level complexity of fast entropy accumulation in real-world

random number generators (RNGs), such as the Fortuna RNG used by Windows 10 [Fer19]. That work also studied online linear extractors, but only for a specific class of natural distributions that arise in practice and only for hyper-efficient linear transformations A that simply permute the bits of the state. Indeed, in [DGSX21], we were primarily concerned with the practical question of optimizing the exact number of samples needed to extract from such distributions for fixed $n \in \{32, 64\}$ using these extremely fast linear transformations.[2] From a technical point of view, both works use Fourier-analytic techniques, but the details are quite different. The main Fourier-analytic tool in [DGSX21] is a bound on the Fourier coefficients of the class of natural distributions that we study there. Here, our main tool is Lemma 3.1, which applies to arbitrary entropic distributions.

Starting with Chor and Goldreich [CG88], many papers (see [BIW04, KRVZ11, CZ19] and references therein) studied the much harder question of randomness extraction from several independent (but *not* identical) arbitrary entropic sources. Unlike our work, these extractors *cannot* be linear, and, to the best of our knowledge, no online extractor is known to extract from this general class of courses. However, if one sufficiently restricts the distribution family to be more structured, online extraction is sometimes possible—even by extremely efficient functions. For example, the classical work of Santha and Vazirani [SV86] showed that simply applying bit-wise XOR is a good extractor for independent (but not necessarily identical) SV-sources. In fact, in some cases online extraction becomes possible even without assuming independence, as long as each new source comes from certain very structured family conditioned on the previous sources [BEG15, BBEG18].

The classical work of von Neumann [von51] studied the question of randomness extraction from IID coin flips with an a-priori unknown bias, and his extractor happened to be online. Elias [Eli72] improved the rate of von Neumann's extractor, but sacrificed the online property to do so.

The works of [CDKT19, DGH+04] explicitly considered online extractors in various idealized computational models (such as the random oracle model). These extractors are highly non-linear.

In the setting of so-called "seeded extractors", where an additional random seed is available for extraction, the power of simple, linear extractors goes back to the leftover hash lemma [HILL99], and the streaming analog of this question (corresponding to a very long source X) was studied by [BTRS02].

2 Preliminaries

2.1 Entropy and statistical distance

For an integer $n \geq 1$, we write $[n] := \{0, \ldots, n-1\}$. For a distribution D over $\{0,1\}^n$ and $\boldsymbol{x} \in \{0,1\}^n$, we write $D(\boldsymbol{x}) := \Pr_{\boldsymbol{X} \sim D}[\boldsymbol{X} = \boldsymbol{x}]$ for the probability that D assigns to $\boldsymbol{x}$. The statistical distance between two distributions D_1 and D_2 over $\{0,1\}^n$ is

$$\mathsf{SD}(D_1, D_2) := \frac{1}{2} \cdot \sum_{\boldsymbol{x} \in \{0,1\}^n} |D_1(\boldsymbol{x}) - D_2(\boldsymbol{x})| \,.$$

[2] In contrast, we are interested in the more theoretical question of extracting from arbitrary entropic sources with arbitrary n. In exchange for this generality, we sacrifice the extreme efficiency achieved in [DGSX21] (which was the primary goal of that work). Indeed, in [DGSX21] we show that very efficient linear transformations A can extract from a natural class of sources in just a bit more than n/k steps, while it is easy to see that $n-k$ steps are necessary for an online linear extractor to extract from arbitrary entropic sources. Indeed, all of the different linear transformations that we considered in [DGSX21] are conjugates of rotation, and are therefore equivalent in our setting of arbitrary entropic sources, while in the model of [DGSX21] their convergence rates are quite different. (In [DGSX21], we were also happy to converge to at most, e.g., $n - \varepsilon$ bits of entropy, while here we are interested in asymptotic convergence.)

We say D_1 is ε-close to D_2 if $\mathsf{SD}(D_1, D_2) \le \varepsilon$. The *min-entropy* of D is

$$H_{\min}(D) := \min_{x \in \{0,1\}^n} \log_2(1/D(x)) \,.$$

2.2 Basic Fourier analysis

For a distribution D over $\{0,1\}^n$ and $\boldsymbol{w} \in \{0,1\}^n$, we define the Fourier coefficient of D at $\boldsymbol{w}$ as

$$\widehat{D}(\boldsymbol{w}) := \mathop{\mathbb{E}}_{\boldsymbol{X} \sim D}[(-1)^{\langle \boldsymbol{X}, \boldsymbol{w} \rangle}] = \Pr_{\boldsymbol{X} \sim D}[\langle \boldsymbol{X}, \boldsymbol{w} \rangle = 0 \bmod 2] - \Pr_{\boldsymbol{X} \sim D}[\langle \boldsymbol{X}, \boldsymbol{w} \rangle = 1 \bmod 2] \,.$$

Claim 2.1. *For any distribution D over $\{0,1\}^n$,*

$$H_{\min}(D) \ge n - \log_2 \Big(\sum_{\boldsymbol{w} \in \{0,1\}^n} |\widehat{D}(\boldsymbol{w})| \Big).$$

and

$$\mathsf{SD}(D, U) \le \frac{1}{2} \sum_{\boldsymbol{w} \in \{0,1\}^n, \boldsymbol{w} \neq \boldsymbol{0}} |\widehat{D}(\boldsymbol{w})| \,,$$

where U is the uniform distribution over $\{0,1\}^n$.

Proof. Recall that for any $\boldsymbol{x} \in \{0,1\}^n$,

$$D(\boldsymbol{x}) = \frac{1}{2^n} \sum_{\boldsymbol{w} \in \{0,1\}^n} \widehat{D}(\boldsymbol{w}) (-1)^{\langle \boldsymbol{x}, \boldsymbol{w} \rangle} \le \frac{1}{2^n} \sum_{\boldsymbol{w} \in \{0,1\}^n} |\widehat{D}(\boldsymbol{w})| \,.$$

Therefore,

$$H_{\min}(D) = \min_{x \in \{0,1\}^n} \log_2(1/D(\boldsymbol{x})) \ge n - \log_2 \Big(\sum_{\boldsymbol{w} \in \{0,1\}^n} |\widehat{D}(\boldsymbol{w})| \Big) \,.$$

Moreover, note that $\widehat{D}(\boldsymbol{0}) = 1$,

$$\big| D(\boldsymbol{x}) - \frac{1}{2^n} \big| = \big| \frac{1}{2^n} \sum_{\boldsymbol{w} \in \{0,1\}^n, \boldsymbol{w} \neq 0} \widehat{D}(\boldsymbol{w}) (-1)^{\langle \boldsymbol{x}, \boldsymbol{w} \rangle} \big| \le \frac{1}{2^n} \sum_{\boldsymbol{w} \in \{0,1\}^n, \boldsymbol{w} \neq \boldsymbol{0}} |\widehat{D}(\boldsymbol{w})| \,.$$

Therefore,

$$\mathsf{SD}(D, U) = \frac{1}{2} \cdot \sum_{\boldsymbol{x} \in \{0,1\}^n} |D(\boldsymbol{x}) - \frac{1}{2^n}| \le \frac{1}{2} \cdot 2^n \cdot \big(\frac{1}{2^n} \sum_{\boldsymbol{w} \in \{0,1\}^n, \boldsymbol{w} \neq \boldsymbol{0}} |\widehat{D}(\boldsymbol{w})| \big) = \frac{1}{2} \sum_{\boldsymbol{w} \in \{0,1\}^n, \boldsymbol{w} \neq \boldsymbol{0}} |\widehat{D}(\boldsymbol{w})| \,.$$

□

The Fourier coefficients arise naturally in our context because they interact nicely with both convolution and linear transformations, as this next well-known claim shows.

Claim 2.2. *For distributions $D_1, \ldots, D_m$ over $\{0,1\}^n$ and linear transformations $A_1, \ldots, A_m \in \mathbb{F}_2^{n \times n}$, let D be the distribution given by*

$$\Pr_{\boldsymbol{X} \sim D}[\boldsymbol{X} = \boldsymbol{x}] = \Pr_{\boldsymbol{X}_1 \sim D_1, \ldots, \boldsymbol{X}_m \sim D_m}[A_1 \boldsymbol{X}_1 \oplus \cdots \oplus A_m \boldsymbol{X}_m = \boldsymbol{x}] \,,$$

where the $\boldsymbol{X}_i$ are independent. Then,

$$\widehat{D}(\boldsymbol{w}) = \widehat{D_1}(A_1^T \boldsymbol{w}) \cdots \widehat{D_m}(A_m^T \boldsymbol{w}) \,.$$

for any $\boldsymbol{w} \in \{0,1\}^n$.

Proof. We have

$$\begin{aligned}\mathbb{E}[(-1)^{\langle \boldsymbol{w},\boldsymbol{X}\rangle}] &= \mathbb{E}[(-1)^{\langle \boldsymbol{w},A_1\boldsymbol{X}_1\oplus\cdots\oplus A_m\boldsymbol{X}_m\rangle}] \\ &= \mathbb{E}[(-1)^{\langle \boldsymbol{w},A_1\boldsymbol{X}_1\rangle}]\cdots\mathbb{E}[(-1)^{\langle \boldsymbol{w},A_m\boldsymbol{X}_m\rangle}] \\ &= \mathbb{E}[(-1)^{\langle A_1^T\boldsymbol{w},\boldsymbol{X}_1\rangle}]\cdots\mathbb{E}[(-1)^{\langle A_m^T\boldsymbol{w},\boldsymbol{X}_m\rangle}] \\ &= \widehat{D_1}(A_1^T\boldsymbol{w})\cdots\widehat{D_m}(A_m^T\boldsymbol{w})\ .\end{aligned}$$

□

For a distribution D over $\{0,1\}^n$, integer $\ell \geq 1$, and linear transformation $A : \mathbb{F}_2^n \to \mathbb{F}_2^n$, we write $D_A^{(\ell)}$ for the distribution obtained by sampling $\boldsymbol{X}_1,\ldots,\boldsymbol{X}_\ell$ independently and returning $\boldsymbol{X}_1 \oplus A\boldsymbol{X}_2 \oplus \cdots \oplus A^{\ell-1}\boldsymbol{X}_\ell$.

2.3 Properties of (near)-uniform distribution over the Hamming ball

The (near)-uniform distribution over the Hamming ball with a given min-entropy plays an important role in our analysis.

Definition 2.3. *For $r,n \in \mathbb{N}, k \in \mathbb{R}$, suppose $1 \leq r \leq n$, and $n-r < k \leq n$, we define $D^*_{r,k}$ over $\{0,1\}^n$ as follows,*

$$D^*_{r,k}(\boldsymbol{x}) := \begin{cases} 2^{-k} & \sum_{i=1}^r x_i < d^* \\ p^* & \sum_{i=1}^r x_i = d^* \\ 0 & \textit{otherwise,}\end{cases}$$

where $d^* := \min\{0 \leq d \leq r \ :\ 2^{n-r}\cdot(\binom{r}{0}+\binom{r}{1}+\cdots+\binom{r}{d}) \geq 2^k\}$, and

$$p^* := \frac{1}{\binom{r}{d^*}}\cdot\left(2^{-(n-r)} - 2^{-k}\cdot\left(\binom{r}{0}+\binom{r}{1}+\cdots+\binom{r}{d^*-1}\right)\right)\ .$$

(I.e., d^* and p^* are chosen to make $D^*_{r,k}$ a probability distribution.)

Lemma 2.4. *Let $1 \leq r \leq n$ and $n-r < k \leq n$, and let $D^*_{r,k}$ be defined as above. Then, for $1 \leq i \leq r$,*

$$\widehat{D^*_{r,k}}(\boldsymbol{e}_i) \leq 1 - \frac{c\cdot d^*}{r} \leq \left(1 - \frac{c(r+k-n)}{6r\log(2r/(r+k-n))}\right)\ ,$$

where $c := 1-2^{-(r+k-n)} \geq \min(\frac{1}{2}, \frac{r+k-n}{2})$.

Proof. By symmetry, for $1 \leq i \leq r$, $j \in \mathbb{N}$,

$$r\cdot\widehat{D^*_{r,k}}(\boldsymbol{e}_i) = \sum_{i'=1}^r \widehat{D^*_{r,k}}(\boldsymbol{e}_{i'}) = \sum_{i'=1}^r (1 - 2\Pr_{\boldsymbol{x}\sim D^*_{r,k}}[x_{i'}=1]) = r - 2\mathop{\mathbb{E}}_{\boldsymbol{x}\sim D^*_{r,k}}[\sum_{i'=1}^r x_{i'}]\ .$$

Let $p_j := \Pr_{\boldsymbol{x}\sim D^*_{r,k}}[\sum_{i'=1}^r x_{i'} = j]$. We have that

$$p_j := \begin{cases} 2^{n-r-k}\binom{r}{j} & 0 \leq j \leq d^*-1 \\ 2^{n-r}\binom{r}{d^*}\cdot p^* & j = d^* \\ 0 & \text{otherwise.}\end{cases}$$

For $1 \le j \le d^* - 1$, it holds that

$$j \cdot p_j + (d^* - j) \cdot p_{d^*-j} \ge (p_j + p_{d^*-j}) \cdot (d^*/2)$$

because $p_j \le p_{d^*-j}$ if and only if $j \le d^* - j$ [3]. Hence,

$$\begin{aligned}
2 \mathop{\mathbb{E}}_{\boldsymbol{x} \sim D^*_{r,k}} \Big[\sum_{i'=1}^{r} x_{i'} \Big] &= \sum_{j=0}^{d^*} (j \cdot p_j + (d^* - j) \cdot p_{d^*-j}) \\
&\ge \sum_{j=1}^{d^*-1} (p_j + p_{d^*-j}) \cdot (d^*/2) + 2d^* \cdot p_{d^*} \\
&= d^* \cdot \sum_{i=0}^{d^*} p_i + d^* \cdot (p_{d^*} - p_0) \\
&= d^*(1 + p_{d^*} - p_0) \\
&\ge d^* \cdot c
\end{aligned}$$

where the last inequality is due to $p_{d^*} \ge 0$. Hence

$$\widehat{D^*_{r,k}}(\boldsymbol{e}_i) = 1 - \frac{2\,\mathbb{E}_{\boldsymbol{x} \sim D^*_{r,k}}[\sum_{i'=1}^{r} x_{i'}]}{r} \le 1 - \frac{c \cdot d^*}{r} \,.$$

The first inequality in the theorem statement follows.

To finish the proof, we prove that for $k \in \mathbb{R}, n - r < k \le n$,

$$d^* \ge \frac{r + k - n}{6 \log(2r/(r + k - n))} \,.$$

We rely on some basic facts about binary entropy function listed in Appendix A. For $p \in (0, 1)$, the binary entropy function is $H(p) := p \log_2(1/p) + (1 - p) \log_2(1/(1 - p))$. By Fact A.1, we have

$$2^{r+k-n} \le \sum_{i=0}^{d^*} \binom{r}{i} \le 2^{rH(d^*/r)} \,.$$

If $k \le n - 1$, then $d^* \le r/2$. The desired conclusion follows by instantiating $rH(d^*/r) \ge r + k - n$ in Claim A.2. If $n - 1 < k \le n$, then $d^* > r/2 > \frac{r+k-n}{6 \log(2r/(r+k-n))}$ because $\frac{r+k-n}{6 \log(2r/(r+k-n))} \le r/6$ for all $k \le n$. □

3 Our main lemma

Lemma 3.1. *For $r, n \in \mathbb{N}, k \in \mathbb{R}$, suppose $1 \le r \le n$, $n - r < k \le n$, $\mathbb{F}_2$-linearly independent vectors $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_r \in \{0,1\}^n$, and a distribution D over $\{0,1\}^n$ with at least min-entropy k, we have*

$$\prod_{i=1}^{r} |\widehat{D}(\boldsymbol{w}_i)| \le 2^{-c(r+k-n)/6 \log_2(2r/(r+k-n))} \,.$$

where $c = 1 - 2^{-(r+k-n)}$.

[3] Note that $p_j = 2^{n-r-k} \cdot \binom{r}{j}$, $p_{d^*-j} = 2^{n-r-k} \cdot \binom{r}{d^*-j}$ for $1 \le j \le d^* - 1$. If $p_j \le p_{d^*-j}$, it implies $\binom{r}{j} \le \binom{r}{d^*-j}$. Since $(j + d^* - j)/2 = d^*/2 \le r/2$, it implies $j \le d^* - j$. Conversely, if $j \le d^* - j$, by the same reason it implies $\binom{r}{j} \le \binom{r}{d^*-j}$ and thus $p_j \le p_{d^*-j}$.

Proof of Lemma 3.1. Let $D^*_{r,k}$ be defined as in Definition 2.3. We show that products of Fourier coefficients at independent vectors is maximized by the products of Fourier coefficients at basis vectors for $D^*_{r,k}$.

Claim 3.2. *For $\mathbb{F}_2$-linearly independent vectors $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_r \in \{0,1\}^n$ and any distribution D over $\{0,1\}^n$ with min-entropy $k \le n$. we have*

$$\prod_{i=1}^{r} |\widehat{D}(\boldsymbol{w}_i)| \le \prod_{i=1}^{r} \widehat{D^*_{r,k}}(\boldsymbol{e}_i) \ ,$$

where $\boldsymbol{e}_i \in \{0,1\}^n$ is the ith standard basis vector.

Combining with Lemma 2.4, we have

$$\prod_{i=1}^{r} |\widehat{D}(\boldsymbol{w}_i)| \le \prod_{i=1}^{r} \widehat{D^*_{r,k}}(\boldsymbol{e}_i) \le \Big(1 - \frac{c(r+k-n)}{6r\log(2r/(r+k-n))}\Big)^r \ .$$

The desired conclusion follows (notice $(1-x)^r \le 2^{-rx}$ for $x \ge 0$). Now we prove Claim 3.2.

Proof of Claim 3.2. Let $A \in \mathbb{F}_2^{n\times n}$ be an invertible linear transformation such that $A^T \boldsymbol{w}_i = \boldsymbol{e}_i$ for all i. Then, $H_{\min}(AD) = H_{\min}(D)$ and $\widehat{AD}(\boldsymbol{e}_i) = \widehat{D}(\boldsymbol{w}_i)$. So, by applying the linear transformation A, we may assume without loss of generality that $\boldsymbol{w}_i = \boldsymbol{e}_i$. By possibly flipping some bits, we may also assume that $\widehat{D}(\boldsymbol{e}_i) \ge 0$, so that it suffices to prove that

$$\prod_{i=1}^{r} \widehat{D}(\boldsymbol{e}_i) \le \prod_{i=1}^{r} \widehat{D^*_{r,k}}(\boldsymbol{e}_i) \ ,$$

For $1 \le i < j \le r$, let $\pi : \{0,1\}^n \to \{0,1\}^n$ be the map that swaps the ith and jth coordinates and leaves all other coordinates untouched. Let D' be the distribution given by $D'(\boldsymbol{x}) = (D(\boldsymbol{x}) + D(\pi(\boldsymbol{x})))/2$. Notice that $H_{\min}(D') \ge H_{\min}(D)$. Furthermore,

$$\prod_{k=1}^{r} \widehat{D'}(\boldsymbol{e}_k) = \frac{(\widehat{D}(\boldsymbol{e}_i) + \widehat{D}(\boldsymbol{e}_j))^2}{4} \cdot \prod_{k\notin\{i,j\}} \widehat{D}(\boldsymbol{e}_k) \ge \prod_{k=1}^{r} \widehat{D}(\boldsymbol{e}_k) \ ,$$

where the last inequality follows from the fact that $(a+b)/2 \ge \sqrt{ab}$ for $a, b \ge 0$. Therefore, we may assume without loss of generality that $D(\boldsymbol{x}) = D(\pi(\boldsymbol{x}))$. By a similar argument, we may assume that $D(\boldsymbol{x}) = D(\boldsymbol{x}')$ for any $\boldsymbol{x}, \boldsymbol{x}' \in \{0,1\}^n$ with $\sum_{i=1}^r x_i = \sum_{i=1}^r x'_i$.

Now, suppose that there exists a vector $\boldsymbol{x} \in \{0,1\}^n$ and an index $1 \le i \le r$ such that $x_i = 1$, $D(\boldsymbol{x}) > 0$ and $D(\boldsymbol{x} \oplus \boldsymbol{e}_i) < 2^{-k}$. Then, let D' be the distribution that is identical to D except that $D'(\boldsymbol{x}) = D(\boldsymbol{x}) - p$ and $D'(\boldsymbol{x} \oplus \boldsymbol{e}_i) = D(\boldsymbol{x} \oplus \boldsymbol{e}_i) + p$, where $0 < p \le \min\{D(\boldsymbol{x}),\ 2^{-k} - D(\boldsymbol{x} \oplus \boldsymbol{e}_i)\}$. Clearly, $H_{\min}(D') \ge k$ and $\prod_{i=1}^r \widehat{D'}(\boldsymbol{e}_i) > \prod_{i=1}^r \widehat{D}(\boldsymbol{e}_i)$.

So, by replacing D with D', we may assume without loss of generality that no such $\boldsymbol{x}$ and i exist. Together with the above assumption that $D(\boldsymbol{x}) = D(\boldsymbol{x}')$ whenever $\sum_{i=1}^r x_i = \sum_{i=1}^r x'_i$, this uniquely characterizes the distribution D. I.e., $D = D^*_{r,k}$. The result follows. □

□

4 Extractability

In this section, we characterize the matrices A that yield online extractors.

Definition 4.1. *We say that a subspace $S \subseteq \mathbb{F}_2^n$ is an invariant subspace of $A \in \mathbb{F}_2^{n\times n}$ or A-invariant if for every $\boldsymbol{w} \in S$, $A\boldsymbol{w} \in S$. We say S is non-trivial if $S \neq \{\mathbf{0}\}$ and $S \neq \mathbb{F}_2^n$.*

There is a rich theory of invariant subspaces that is beyond the scope of this work. (See, e.g., [Cla13].) For our purposes, it suffices to note simply that the invariant subspaces can be computed efficiently. In particular, the invariant subspaces correspond to factors of the characteristic and minimal polynomials of A, and A has no non-trivial invariant subspace if and only if the characteristic polynomial of A is irreducible.

Invariant subspaces arise naturally in this context. Indeed, if $S \subset \mathbb{F}_2^n$ is a non-trivial invariant subspace of A, then A will completely fail to extract from the uniform distribution over S. We make this observation formal in Theorem 4.2.

Theorem 4.2. *For $A \in \mathbb{F}_2^{n\times n}$, if there exists a non-trivial A-invariant subspace with dimension r, then there exists a distribution D over $\{0,1\}^n$ with min-entropy r such that $D_A^{(m)} = D$ for all m.*

Proof. Let S be an A-invariant subspace with dimension r. Let D be the uniform distribution over S with min-entropy r. Recall that D_A^m is the distribution obtained by sampling $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_\ell$ independently from D and returning $\boldsymbol{X}_1 \oplus A\boldsymbol{X}_2 \oplus \cdots \oplus A^{m-1}\boldsymbol{X}_m$. Because S is A-invariant, it holds that $\boldsymbol{y} := A\boldsymbol{X}_2 \oplus \cdots \oplus A^{m-1}\boldsymbol{X}_m$ is in the subspace S, and $\boldsymbol{X}_1 \oplus y$ is uniformly distributed over S for an independent $y \in S$. Therefore for all m, $D_A^{(m)}$ is the uniform distribution over S. □

Perhaps more surprisingly, the next theorem shows that this is the only restriction. In particular, if A has no non-trivial invariant subspace, then A extracts from any source with min-entropy k after $\widetilde{O}(n^2(k+1)/k^2)$ steps.

Theorem 4.3. *For $A \in \mathbb{F}_2^{n\times n}$, if A has no non-trivial invariant subspace, then for $k > 0$, and any distribution D over $\{0,1\}^n$ with min-entropy at least k,*

$$\mathsf{SD}(D_A^{(m)}, U) \leq 2^{n-1-\lfloor m/n \rfloor \cdot \frac{ck}{6\log_2(2n/k)}}$$

where $c = 1 - 2^{-k}$.

Proof. Because the orthogonal subspace of an A-invariant subspace is A^T-invariant, A^T also has no non-trivial invariant subspace. For any non-zero $\boldsymbol{w}$, it must therefore be the case that $\boldsymbol{w}_1 := \boldsymbol{w}, \boldsymbol{w}_2 := \ldots, \boldsymbol{w}_n := (A^T)^{n-1}\boldsymbol{w}$, are linearly independent. Otherwise the span of $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_n$ would be a non-trivial A^T-invariant subspace. By applying Lemma 3.1 with $r = n$, we obtain

$$\prod_{i=0}^{n-1} |\widehat{D}((A^T)^i\boldsymbol{w})| = \prod_{i=1}^{n} |\widehat{D}(\boldsymbol{w}_i)| \leq 2^{-ck/6\log_2(2n/k)}\,, \tag{1}$$

where $c = 1 - 2^{-k}$. Therefore, for any non-zero $\boldsymbol{w}$,

$$|\widehat{D_A^{(m)}}(\boldsymbol{w})| = \prod_{i=0}^{m-1} |\widehat{D}((A^T)^i\boldsymbol{w})| = \prod_{j=0}^{\lfloor m/n \rfloor - 1} \prod_{i=0}^{n-1} |\widehat{D}((A^T)^{jn+i}\boldsymbol{w})| \leq 2^{-\lfloor m/n \rfloor \cdot \frac{ck}{6\log_2(2n/k)}}$$

where the last inequality is due to $(A^T)^{jn}\boldsymbol{w} \neq 0$ and (1). By applying Claim 2.1,

$$\mathsf{SD}(D_A^{(m)}, U) \leq \frac{1}{2} \cdot \sum_{\boldsymbol{w}\in\{0,1\}^n, \boldsymbol{w}\neq \mathbf{0}} |\widehat{D_A^{(m)}}(\boldsymbol{w})| \leq 2^{n-1-\lfloor m/n \rfloor \cdot \frac{ck}{6\log_2(2n/k)}} ,$$

as desired. □

We note in passing that the matrix A corresponding to multiplication by a generator of a finite field is a particularly nice example satisfying the condition of Theorem 4.3. That is, if we interpret $\boldsymbol{y} = (y_1, \ldots, y_n) \in \{0,1\}^n$ as the polynomial $y_1 + y_2 t + \cdots + y_n t^{n-1} \in \mathbb{F}_2[t]/p(t)$ for some irreducible polynomial $p(t) \in \mathbb{F}_2[t]$ of degree n. Then, the matrix A corresponding to multiplication by t has no non-trivial invariant subspace[4] and thus yields a good extractor. This matrix has the convenient property that it is quite sparse—with all columns except the last having a single non-zero entry.

5 Condensibility

We now turn our attention to online linear condensers. Our results will be in terms of the concept of the A-rank of a vector $\boldsymbol{w} \in \mathbb{F}_2^n$, defined below.

Definition 5.1. *For any $A \in \mathbb{F}_2^{n\times n}$, the A-orbit of a vector $\boldsymbol{w} \in \{0,1\}^n$ is the set $\{A^k\boldsymbol{w}\}_{k=0}^{\infty}$. The linear orbit $[\boldsymbol{w}]$ of $\boldsymbol{w}$ is the subspace spanned by A-orbit of $\boldsymbol{w}$.*

Definition 5.2. *For any $A \in \mathbb{F}_2^{n\times n}$, the A-rank of a vector $\boldsymbol{w} \in \{0,1\}^n$ is the maximal integer r such that the set of vectors $\{\boldsymbol{w}, A\boldsymbol{w}, \ldots, A^{r-1}\boldsymbol{w}\}$ is linearly independent. We use $\mathrm{rank}_A(w)$[5] to denote A-rank of $\boldsymbol{w}$.*

One can efficiently compute the number of vectors with a given A-rank by computing the minimal polynomial of A [Cla13].

Proposition 5.3. *For $A \in \mathbb{F}_2^{n\times n}$, $\boldsymbol{w} \in \{0,1\}^n$ with the A-rank r, the linear orbit $[\boldsymbol{w}]$ is an invariant subspace of dimension r. Moreover,*

$$[\boldsymbol{w}] = \mathrm{span}(\boldsymbol{w}, A\boldsymbol{w}, \ldots, A^{r-1}\boldsymbol{w}) .$$

The above proposition shows that the A-rank of $\boldsymbol{w}$ characterizes the minimal invariant subspace V containing $\boldsymbol{w}$: if the A-rank of $\boldsymbol{w}$ is r, then the first r vectors in the A-orbit are linear independent and thus generate V. In particular, if A has no non-trivial invariant subspace, then every $\boldsymbol{w} \in \mathbb{F}_2^n \setminus \{0^n\}$ has A-rank n.

Our next theorem gives a partial characterization of matrices A that yield good online linear condensers in terms of A^T-rank and the number of vectors with small A^T-rank. This yields a natural generalization of Theorem 4.3.

[4] To see this, suppose for contradiction that there exists a non-trivial t-invariant subspace $V \subset \mathbb{F}_2[t]/p(t)$. Then, for any $x \in V$, we must have that $x, tx, \ldots, t^{n-1}x$ are linearly dependent (since otherwise V is either not invariant or $V = \mathbb{F}_2^n$ is non-trivial). Since $\mathbb{F}_2[t]/p(t)$ is a field, if $V \neq \{0\}$, we must also have that $1, t, \ldots, t^{n-1}$ are linearly independent. This means that t is a root of a polynomial with degree at most $n-1$, contradicting the assumption that p is irreducible.

[5] In linear algebra, our notation $\mathrm{rank}_A(w)$ is the same as the maximal dimension of a Krylov subspace generated by A and $\boldsymbol{w}$.

Theorem 5.4. *For any invertible $A \in \mathbb{F}_2^{n\times n}$, if there are at most N vectors in $\{0,1\}^n$ with A^T-rank less than r, then for any real number $g := n - r < k \le n$ and any distribution D over $\{0,1\}^n$ with min-entropy at least k,*

$$H_{\min}(D_A^{(m)}) \ge n - \log_2(N + 2^{n - \lfloor m/n \rfloor \cdot \frac{c(k-g)}{6\log_2(2r/(k-g))}})$$

where $c = 1 - 2^{-(k-g)}$.

Proof. For any $\boldsymbol{w} \in \{0,1\}^n$ with A^T-rank at least r, then there are at least r-linear independent vectors among $\boldsymbol{w}, \ldots, (A^T)^{n-1}\boldsymbol{w}$, denoted as $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_r$. By Lemma 3.1, it implies

$$\prod_{i=0}^{n-1} |\widehat{D}((A^T)^i \boldsymbol{w})| \le \prod_{i=1}^{r} |\widehat{D}(\boldsymbol{w}_i)| \le 2^{-\frac{c(k-g)}{6\log_2(2r/(k-g))}} ,$$

where $c = 1 - 2^{-(k-g)}$. Moreover, because A is invertible, $(A^T)^n \boldsymbol{w}$ has the same A^T-rank as $\boldsymbol{w}$. We have that,

$$|\widehat{D_A^{(m)}}(\boldsymbol{w})| = \prod_{i=0}^{m-1} |\widehat{D}((A^T)^i \boldsymbol{w})| \le \prod_{j=0}^{\lfloor m/n \rfloor - 1} \prod_{i=0}^{n-1} |\widehat{D}((A^T)^{jn+i} \boldsymbol{w})| \le 2^{-\lfloor m/n \rfloor \cdot \frac{c(k-g)}{6\log_2(2r/(k-g))}}$$

Because there are at most N vectors with A^T-rank less than r, it holds that

$$\sum_{\boldsymbol{w}\in\{0,1\}^n} |\widehat{D_A^{(m)}}(\boldsymbol{w})| = \sum_{\boldsymbol{w}:\mathrm{rank}_{A^T}(\boldsymbol{w})<r} |\widehat{D_A^{(m)}}(\boldsymbol{w})| + \sum_{\boldsymbol{w}:\mathrm{rank}_{A^T}(\boldsymbol{w})\ge r} |\widehat{D_A^{(m)}}(\boldsymbol{w})| \le N \cdot 1 + 2^n \cdot 2^{-\lfloor m/n \rfloor \cdot \frac{c(k-g)}{6\log_2(2r/(k-g))}} .$$

By applying Claim 2.1,

$$H_{\min}(D) = n - \log_2 \Big(\sum_{\boldsymbol{w}\in\{0,1\}^n} |\widehat{D}(\boldsymbol{w})| \Big) \ge n - \log_2(N + 2^{n - \lfloor m/n \rfloor \cdot \frac{c(k-g)}{6\log_2(2r/(k-g))}}) ,$$

as desired. □

Theorem 5.4 implies that *any* distribution with $> n - r$ bits of min-entropy can be condensed into at least $n - \log_2 N$ bits. Notice that Theorem 5.4 is non-vacuous if $N < 2^r$. Moreover, the constraint $k > n - r$ is tight. If there exists a vector with A^T-rank r, then there is an A^T-invariant subspace V of dimension r, which in particular contains 2^r vectors of A^T-rank at most r. Then, by Theorem 4.2 the distribution D that is uniform over the subspace orthogonal to V has min-entropy $n - r$ but $D_A^{(m)} = D$ for all m.

Rotation. Finally, as an application of this result, we show that rotation yields a good condenser for some n. (Moreover, if we assume an additional minor condition on the distribution D, we actually get an extractor.)

We write rot_n for the linear transformation over $\{0,1\}^n$ which rotates the coordinates of a vector $\boldsymbol{x}$ by 1. In other words,

$$\mathsf{rot}_n((x_1, \ldots, x_n)) := (x_n, x_1, x_2, \ldots, x_{n-1}) .$$

Our first observation is that $\{\boldsymbol{x} : x_i = x_{i+d}, \forall 1 \leq i \leq n-d\}$ is an invariant subspace of any rotation when $d < n$ is a divisor of n. By Theorem 4.2, rot_n therefore cannot extract from sources with min-entropy d for $d < n$ a divisor of n. Moreover, rotations in general cannot condense from a single bit of randomness because of the invariant subspace $\{0^n, 1^n\}$ and cannot condense beyond $n-1$ bits of randomness because of the invariant subspace $\{\boldsymbol{x} : x_1 \oplus \cdots \oplus x_n = 0\}$. Therefore, the best we can hope for is to condense from $k > 1$ bits of entropy to $n-1$ bits of entropy for n prime.

We show that rot_n does in fact achieve this as long as n is a prime satisfying a natural number-theoretic condition. Indeed, this follows from Theorem 5.4 together with the following lemma due to Vazirani [Vaz87].

Lemma 5.5 ([Vaz87])**.** *If n is a prime such that 2 generates $\mathbb{Z}_n^*$ (e.g., 5, 29, 37), then all $\boldsymbol{w} \in \{0,1\}^n \setminus \{1^n, 0^n\}$ have rot_n-rank at least $n-1$.*

Plugging into Theorem 5.4 yields the following. In particular, for such primes, rot_n condenses from $k > 1$ bits to $n-1$ bits in at most $m = \widetilde{O}(n^2 k/(k-1)^2)$ steps.

Corollary 5.6. *If n is a prime with 2 is a primitive root for $\mathbb{Z}_n^*$, then for any real number $1 < k \leq n$, and distribution D over $\{0,1\}^n$ with at least min-entropy k,*

$$H_{\min}(D^{(m)}_{\mathsf{rot}_n}) \geq n - \log_2(2 + 2^{n - \lfloor m/n \rfloor \cdot \frac{c(k-1)}{6\log_2(2(n-1)/(k-1))}}).$$

where $c = 1 - 2^{-(k-1)}$.

Finally, we note that our proof of Theorem 5.4 actually yields a statement about extraction as well, which we present here in the special case of rotation. Specifically, in the proof of Theorem 5.4, we used the trivial bound of $|\widetilde{D}(\boldsymbol{w})| \leq 1$ for low-rank $\boldsymbol{w}$. If we instead happen to know a better bound on the Fourier coefficient explicitly for the single non-zero low-rank vector for rotation, 1^n, we see that we can actually extract.

Theorem 5.7. *For primes n such that 2 generates $\mathbb{Z}_n^*$, and for $1 < k \leq n$, a distribution D over $\{0,1\}^n$ with at least min-entropy k,*

$$\mathsf{SD}(D^{(m)}_{\mathsf{rot}_n}, U) \leq \frac{1}{2} \cdot \left(|\widehat{D}(1^n)|^m + 2^{n - \lfloor m/n \rfloor \cdot \frac{c(k-1)}{6\log_2(2(n-1)/(k-1))}} \right)$$

where $c = 1 - 2^{-(k-1)}$.

Theorem 5.7 implies that for such primes n, rotation yields a good online linear extractor for distributions D with small $|\widehat{D}(1^n)|$ and min-entropy strictly larger than one. Notice that the two counterexamples that we discussed in the definition—the uniform distribution over $\{0^n, 1^n\}$, and the uniform distribution over all strings with even Hamming weight—show that one of these conditions alone is not enough.

References

[BBEG18] Salman Beigi, Andrej Bogdanov, Omid Etesami, and Siyao Guo. Optimal deterministic extractors for generalized Santha-Vazirani sources. In *RANDOM*, 2018.

[BEG15] Salman Beigi, Omid Etesami, and Amin Gohari. Deterministic randomness extraction from generalized and distributed Santha-Vazirani sources. In *ICALP*, 2015.

[BG13] Andrej Bogdanov and Siyao Guo. Sparse extractor families for all the entropy. In *ITCS*, 2013.

[BIW04] B. Barak, R. Impagliazzo, and A. Wigderson. Extracting randomness using few independent sources. In *FOCS*, 2004.

[BTRS02] Ziv Bar-Yossef, Luca Trevisan, Omer Reingold, and Ronen Shaltiel. Streaming computation of combinatorial objects. In *CCC*, 2002.

[CDKT19] Sandro Coretti, Yevgeniy Dodis, Harish Karthikeyan, and Stefano Tessaro. Seedless fruit is the sweetest: Random number generation, revisited. In *CRYPTO*, 2019.

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

[Cla13] Pete L. Clark. Linear algebra: Invariant subspaces. http://alpha.math.uga.edu/~pete/invariant_subspaces.pdf, 2013.

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

[DGH+04] Yevgeniy Dodis, Rosario Gennaro, Johan Håstad, Hugo Krawczyk, and Tal Rabin. Randomness extraction and key derivation using the CBC, Cascade and HMAC modes. In *CRYPTO*, 2004.

[DGSX21] Yevgeniy Dodis, Siyao Guo, Noah Stephens-Davidowitz, and Zhiye Xie. No time to hash: On superefficient entropy accumulation. In *CRYPTO*, 2021.

[Eli72] Peter Elias. The efficient construction of an unbiased random sequence. *Ann. Math. Statist.*, 43(3):865–870, 1972.

[Fer19] Niels Ferguson. The Windows 10 random number generation infrastructure. https://www.microsoft.com/security/blog/2019/11/25/going-in-depth-on-the-windows-10-random-number-generation-infrastructure/, 2019. [Online; posted October 2019].

[HILL99] J. Håstad, R. Impagliazzo, L.A. Levin, and M. Luby. A pseudorandom generator from any one-way function. *SIAM Journal on Computing*, 1999.

[KRVZ11] Jesse Kamp, Anup Rao, Salil P. Vadhan, and David Zuckerman. Deterministic extractors for small-space sources. *J. Comput. Syst. Sci.*, 77(1):191–220, 2011.

[SV86] Miklos Santha and Umesh V. Vazirani. Generating quasi-random sequences from semi-random sources. *J. Comput. Syst. Sci.*, 33(1):75–87, 1986.

[Vaz87] Umesh V. Vazirani. Efficiency considerations in using semi-random sources. In *STOC*, pages 160–168, 1987.

[von51] John von Neumann. Various techniques used in connection with random digits. In *Monte Carlo Method*, pages 36–38. National Bureau of Standards Applied Mathematics Series, 12, 1951.

A Facts about binary entropy function

Fact A.1. *For $0 \le d \le \ell/2$,*

$$\sum_{i=0}^{d} \binom{\ell}{i} \le 2^{\ell H(d/\ell)}.$$

Claim A.2. *[BG13] For every $p \in (0, 1/2]$,*

$$\frac{H(p)}{6\log_2(2/H(p))} \le p \le \frac{H(p)}{\log_2(1/H(p))}.$$

We include the proof from [BG13] for completeness.

Proof. The upper bound on p follows from the inequality $H(p) \ge p \log_2 1/p$. Applying twice we obtain

$$\frac{1}{p} \ge \frac{1}{H(p)} \log_2 \frac{1}{p} \ge \frac{1}{H(p)} \log_2 (\frac{1}{H(p)} \log_2 \frac{1}{p}) \ge \frac{1}{H(p) \log_2 \frac{1}{H(p)}}$$

because $1/p \ge 2$. For the lower bound, we apply $H(p) \le 2p \log_2 1/p$ twice to obtain

$$\frac{1}{p} \le \frac{2}{H(p)} \log_2 \frac{1}{p} \le \frac{2}{H(p)} \log (\frac{2}{H(p)} \log_2 \frac{1}{p}).$$

Now $2/H(p) \ge (1/p) \log_2(1/p) \ge \sqrt{\log_2(1/p)}$, which is true for every $p \in (0, 1]$. Therefore,

$$\frac{1}{p} \le \frac{2}{H(p)} \log_2 (\frac{8}{H(p)^3}) = \frac{6}{H(p)} \log_2 (\frac{2}{H(p)}).$$

□