Sparse extractor families for all the entropy

Andrej Bogdanov[*]   Siyao Guo[†]

Abstract

We consider the problem of extracting entropy by sparse transformations, namely functions with a small number of overall input-output dependencies. In contrast to previous works, we seek extractors for essentially all the entropy without any assumption on the underlying distribution beyond a min-entropy requirement. We give two simple constructions of *sparse extractor families*, which are collections of sparse functions such that for any distribution X on inputs of sufficiently high min-entropy, the output of most functions from the collection on a random input chosen from X is statistically close to uniform.

For strong extractor families (i.e., functions in the family do not take additional randomness) we give upper and lower bounds on the sparsity that are tight up to a constant factor for a wide range of min-entropies. We then prove that for some min-entropies weak extractor families can achieve better sparsity.

We show how this construction can be used towards more efficient parallel transformation of (non-uniform) one-way functions into pseudorandom generators. More generally, sparse extractor families can be used instead of pairwise independence in various randomized or nonuniform settings where preserving locality (i.e., parallelism) is of interest.

---

[*]`andrejb@cse.cuhk.edu.hk`. Department of Computer Science and Engineering and Institute for Theoretical Computer Science and Communications, Chinese University of Hong Kong. Work supported by grants RGC GRF CUHK410309 and CUHK410111.

[†]`syguo@cse.cuhk.edu.hk`. Department of Computer Science and Engineering, Chinese University of Hong Kong.

1 Introduction

Randomness extractors [NZ93, Sha04], have numerous applications in complexity theory and cryptography. For example, in computational complexity they are used to isolate satisfying assignments of boolean formulas [VV86] and in constructing pseudorandom generators for space-bounded computation [Nis92, RR99]. One notable use in cryptography is in the construction of pseudorandom generators from one-way functions [HILL99, HRV10, VZ11].

In many applications of extractors, including the above ones, it is important that the extractors recover essentially all the entropy of the input distribution. A popular choice in such scenarios is to instantiate the extractor by a pairwise-independent hash function family [BBR88, ILL89]. Pairwise-independent functions are appealing because they have a variety of implementations, ranging from very simple ones [CW79] to very efficient ones [IKOS08].

Mansour, Nisan, and Tiwari [MNT93] observed that pairwise-independent hash functions must be "dense" in the sense that a typical output in a typical function in the family must depend on a linear number of inputs. So despite their numerous nice properties, in terms of the number of input-output dependencies, pairwise-independent functions are quite complex. Motivated by an application to local cryptography, Bogdanov and Rosen [BR11] recently gave a way to bypass this barrier in the context of hardness amplification of "local" functions.

In this work we study sparse extractors for all the entropy, these are extractors with a small number of overall input-output dependencies. We consider the more general notion of sparse extractor families. An *extractor family* for distributions of min-entropy k over $\{0,1\}^n$ with error ε is a distribution H on functions $\{0,1\}^n \times \{0,1\}^s \to \{0,1\}^m$ where $m \leq s + k$ such that for every distribution X over $\{0,1\}^n$ of min-entropy k, the statistical distance between $(H, H(X, U_s))$ and (H, U_m) is at most ε (where U_s and U_m are uniformly random). The extractor family is *strong* if $s = 0$, i.e. H does not take any additional randomness beyond X.

Without the sparsity restriction extractors and extractor families are essentially the same object, as the randomness used to choose an extractor from the family can be included in the seed. Once we take sparsity into consideration, however, extractor families allow for more flexibility. This advantage is especially pronounced in the case of strong extractors: Any single strong extractor in which some output bit depends only on ℓ input bits cannot extract from a source that fixes all those ℓ bits. In contrast, we show strong extractor families can achieve much better sparsity.

In this work we prove three results regarding sparse extractor families. First we give a simple construction of sparse extractor families for all the entropy. Then we show that the sparsity of our construction is optimal up to constant factors for a wide range of the min-entropy parameter. Finally, we show that an equally simple construction of weak extractor families achieves better sparsity. Thus when sparsity is required, weak extractors can provably outperform strong ones.

We also show our weak extractor family gives a somewhat improved nonuniform construction of local pseudorandom generators from local one-way functions, based on recent work of Vadhan and Zheng [VZ11]. In general our results can be useful in randomized or nonuniform settings where hashing is used and obtaining or preserving small input-output dependencies (i.e., parallelism) is of interest.

1.1 Our results

Let $h: \{0,1\}^n \to \{0,1\}^m$ be a function. We say output j of h *depends* on input i if there exists assignments $x, x' \in \{0,1\}^n$ that differ only in the ith coordinate such that $h(x)_j \neq h(x')_j$. We say

h is *s-sparse* if the number of input-output pairs (i,j) such that output j depends on input i is at most s, and h is *ℓ-local* if every output j depends on at most ℓ inputs i.

Theorem 1. *Let K be a sufficiently large constant and n, k, m, δ be parameters such that $1 \le m \le k \le n$, and $0 < \delta < 1$. Let $H(x) = Mx$, where M is an $m \times n$ matrix over $GF(2)$ where each entry equals 1 independently with probability*

$$p = \min\left\{\frac{1}{m} \cdot \log \frac{m}{\delta} \ln \frac{Kn}{m}, \frac{1}{2}\right\}.$$

Then H is a strong extractor family for min-entropy k with error at most $\frac{1}{2}\sqrt{\delta + K \cdot 2^{-k+m}}$.

By a large deviation bound, all but an δ-fraction of H are $O(nmp)$-sparse. The best error we can hope for [RT97] is $\Omega(\sqrt{2^{-k+m}})$, which is achieved by a pairwise independent hash function family. Perhaps the simplest construction of such a family is to choose each entry of M independently at random with probability $p = 1/2$. When $p < 1/2$, Theorem 1 shows the sparsity can be reduced dramatically at the cost of increasing the error by a little. For example if we set $\delta = 2^{-k+m}$, we obtain a $O(n \log m \log(n/m))$ sparse strong extractor family whose error is within a constant factor of optimal.

Our main negative result shows that this sparsity is necessary for a large range of values of k and when δ is constant.

Theorem 2. *Suppose $n^{0.99} \le m \le n/6$. There exists a distribution $\mathcal{D}$ over distributions $\overline{X}$ on $\{0,1\}^n$ of min-entropy $1.5m$ each so that for every function $h\colon \{0,1\}^n \to \{0,1\}^m$ of sparsity $0.001 n \log m \log(2n/m)$, the expected statistical distance between $h(\overline{X})$ and the uniform distribution over $\{0,1\}^m$ is at least $1 - e^{-m^{\Omega(1)}}$.*

Applying Yao's minimax principle or just take the covex combination of distributions $\overline{X}$ as the bad distribution, we conclude that the sparsity in Theorem 1 is optimal up to constant factor for this range of parameters. The next results concerns weak extractor families.

Theorem 3. *Let K be a sufficiently large constant and n, k, c, m be parameters such that $1 \le k \le n$, $1 \le s < m$ and $c > 1$. Let $H\colon \{0,1\}^n \times \{0,1\}^s \to \{0,1\}^m$ be given by $H(x,r) = Mx + Br$, where M is an $m \times n$ random matrix in which each entry equals 1 independently with probability*

$$p = \min\left\{\frac{K}{m} \cdot \ln \frac{n}{\ln c}, \frac{1}{2}\right\},$$

and B is an $m \times s$ ($m > s$) matrix of full rank where every set of at most $m/2K$ rows is linearly independent. Then H is an extractor family for min-entropy k with error $\frac{1}{2}\sqrt{c \cdot 2^{-k-s+m}}$.

The construction of a matrix B with the desired properties and $O(m)$ sparsity is a well studied problem in the theory of low density parity check codes [Gal62, SS96]. Capalbo et al. [CRVW02] give an explicit construction with $s = \alpha m$ for some constant $\alpha < 1$ and every m, which is optimal up to the choice of the constant α. Instantiating Theorem 3 with this matrix, and setting $c = 2$, we obtain a family of $O(n \log n)$ sparse extractors with error $O(\sqrt{2^{-k-s+m}})$, which is optimal up to constant factor. (If $m = k + s - O(1)$, the output contains almost all the entropy from the source plus all the entropy invested by the seed and the error is an arbitrarily small constant.) By using a larger value of c, we can reduce the sparsity at the cost of increasing the error.

We observe that using the randomized encoding of Applebaum et al. [AIK04a], these extractors can be made to have constant locality at the cost of increasing the seed length r to $O(n \log n)$ bits.

For certain parameters the weak extractor family from Theorem 3 bypasses the limitation on strong extractor families from Theorem 2. For example, for constant statistical distance and $m = n^{0.99}$ Theorem 2 implies that for a strong extractor to produce even a constant fraction of the entropy a sparsity of $\Omega(n(\log n)^2)$ is necessary, while Theorem 3 says that an $O(n \log n)$-sparse weak extractor family can extract all but $O(1)$ bits of entropy.

1.2   An application

Pseudorandom generators from one-way functions   The construction of pseudorandom generators from one-way functions of Håstad et al. [HILL99] does not in general preserve locality. Haitner, Reingold, and Vadhan [HRV10] gave a construction that is more efficient and can be implemented in NC^1. Recently Vadhan and Zheng [VZ11] gave an even simpler variant of this construction. In combination with the "compiler" of Applebaum, Ishai, and Kushilevitz [AIK04b], one obtains a generic locality-preserving transformation of one-way functions into pseudorandom generators.

Applying the transformation of Applebaum et al. may have an adverse effect on seed length, as it may grow quadratically. However, the construction of Vadhan and Zheng is extremely simple; it is obtained by applying extractor to a sequence of "blocks", each of which inherits the locality of the one-way function f. Instantiating the extractor by the construction from Theorem 3, we obtain a transformation of *nonuniform* one-way functions into nonuniform pseudorandom generators that preserves output locality logarithmic in the size of the adversary with the same seed length as the one obtained by Vadhan and Zheng. Using an additional idea of Applebaum et al., the transformation can be made to preserve constant output locality at the expense of increasing the seed length. We describe this application in Section 5 (see Proposition 9).

1.3   Related work

Sparse extractors for restricted sources   Motivated by certain applications, Zhou and Bruck [ZB11] show that low density random matrices can efficiently extract random bits from some restricted noisy sources, such as bit fixing sources and Markov sources. Our Theorem 1 shows that essentially the same construction extracts from arbitrary sources of given min-entropy.

Extractors in NC^0   Applebaum, Ishai and Kushilevitz [AIK06] give a weak extractor in NC^0 (thus sparsity $O(n)$) works for min-entropy $k = (1 - O(1))n$, but suffers $\Omega(n)$ entropy loss. Our extractor family from Theorem 3 matches these parameters. The construction from [AIK06] does not appear to extend to distributions of smaller min-entropy or allow for smaller entropy loss, while ours does. However, they provide a single extractor that works for all distributions, while we only give an extractor family.

Locally computable extractors   A *locally computable extractor* [Lu04, Vad03] is an extractor in which after the seed is fixed, the output as a whole depends on a small number of input bits. Such extractors are used to implement private-key encryption in the bounded storage model [Mau92]. We observe that the notions of locally computable extractors and sparse extractor families are fundamentally different. This is best illustrated in the regime in which we extract all the entropy,

3

which is of main interest in this work. A lower bound of Vadhan [Vad03] shows that when the output length m is linear in the min-entropy k, then even after the seed is fixed the output of the extractor as a whole must depend on at least a linear fraction of the input. Thus $o(n)$-locally computable extractors are not possible when $m = \Omega(k)$. Although this is inevitable, our results show that it is possible to make a small number of input-output dependencies.

We observe that the locally computable extractors of Lu [Lu04] and De and Trevisan [DT09] are also sparse, but they extract only a fixed root of the min-entropy k. This is sufficient for bounded storage cryptography, but not for the application we describe in Section 5.

1.4 Our proofs

Bogdanov and Rosen [BR11] proved a quantitatively weaker version of Theorem 1 that achieves sparsity $O(n(\log n)^3)$ instead of the optimal $O(n \log(m) \log(n/m))$. (They did not attempt to determine the dependence on m and they can achieve sparsity $O(n \log(m) \log(n/m)^2)$) In their proof, H is viewed as a collection of boolean functions $(h_1, \ldots, h_m), h_i \colon \{0,1\}^n \to \{0,1\}$. They show that for most choices of h_1, conditioning on $h_1(x)$ reduces the min-entropy k of x by at most $1 + 1/\text{poly}(k)$ bits (unless k is very small), and so this bit extraction can be applied iteratively for m steps.

One drawback of this argument is that as i gets larger and the min-entropy of x conditioned on $h_1(x), \ldots, h_{i-1}(x)$ becomes smaller, the density of the functions h_i must keep increasing (as required by our lower bound). To achieve our optimal (up to constant) sparsity, we must analyze the effect of all the functions $h_1, \ldots, h_m$ simultaneously.

To do this, we upper bound the probability that two samples x, x' collide under h, that is the probability that $h(x + x') = 0$. For a fixed pair (x, x'), each entry $h_i(x + x')$ of $h(x + x')$ is biased towards zero. We can think of $h_i(x + x')$ as a random variable that takes value zero with some probability $p(x, x')$, and is unbiased otherwise. Intuitively, our analysis shows that the unbiased components of this distribution dominate in collisions. Several technical complications arise in the formal argument. One useful tool that allows us to analyze the case when most of the components of $h(x + x')$ are unbiased is Hölder's inequality.

To give an idea of our proof of Theorem 2, let's make the simplifying assumption that h is ℓ-local, where $\ell = \gamma(n/m) \log m \log(n/m)$. We give a heuristic argument why we expect the output of h to be far from uniform when h is linear. Let X be the p-biased distribution over $\{0,1\}^n$ (each bit takes value 1 independently with probability p) and p is chosen so that $H(p) = m/n$, where $H(p)$ is the binary entropy of p. Then the distribution X has Shannon entropy m. However, every output bit of $h(X)$ is $(1-2p)^\ell$-biased, and we chose the parameters so that $(1-2p)^\ell = m^{-\Omega(\gamma)}$. By choosing γ small enough, we can ensure that every output bit of h has, say, $m^{-1/2}$ bits of *entropy deficiency*, so by the sub-additivity of Shannon entropy $h(X)$ has $m^{1/2}$ fewer bits of entropy than a uniformly random variable over $\{0,1\}^m$. So $h(X)$ does not "look" random in terms of Shannon entropy.

To turn this heuristic argument into a proof we need to handle several issues, the most interesting of which is replacing entropy deficiency by statistical distance from the uniform distribution. One advantage of measuring entropy deficiency is that entropy is subadditive, which allows us to ignore the dependencies between the various outputs of $h(X)$ in the above argument. In contrast, to obtain a good lower bound on statistical distance we must take into account these dependencies. Here we apply tail bound for read t family [GLSS12]. To extend the analysis from linear functions to general ones we apply an elegant idea of Viola [Vio05] of shifting X by a random offset.

We establish Theorem 3 by a relatively straightforward probabilistic calculation.

1.5 Open problems

In terms of seed length our sparse extractors are quite poor. For example the size of the family H in Theorem 1 is exponential in n. By a standard probabilistic argument it can be shown that a random sample of H of size $O(n/\varepsilon^2)$ is as effective as the whole family while incurring an additional penalty of only ε in statistical distance. Consequently there is no existential obstacle to sparse extractor families with short seed. It remains to see if such families can be found efficiently.

For weak sparse extractors, our work leaves open two possible improvements. First, we do not know if the sparsity of our weak extractors is the best possible. Second, we do not know what is the minimal size of a sparse weak extractor family. It could be that even a family of size 1, i.e. a single sparse weak extractor, is sufficient. Such an extractor could be used to obtain a uniform construction of local pseudorandom generators from local one-way functions. Could there be a single sparse weak extractor of sparsity linear in n that extracts $k - k^{0.99}$ bits of min-entropy for every source over $\{0,1\}^n$ of min-entropy k?

2 Proof of Theorem 1

To prove Theorem 1 it is sufficient to show that for every set S of size 2^k, the statistical distance between $(H, H(X))$ and (H, U) is at most $\frac{1}{2}\sqrt{\delta + O(2^{-k+m})}$, where X is chosen at random from S and U is uniformly random. In fact we will show for every $x_0 \in S$,

$$\Pr_{H,X}[H(X) = H(x_0)] \leq \frac{1 + \delta + O(2^{-k+m})}{2^m}$$

from where

$$\Pr_{H,X,X'}[H(X) = H(X')] \leq \max_{x_0 \in S} \Pr_{H,X}[H(X) = H(x_0)] \leq \frac{1 + \delta + O(2^{-k+m})}{2^m}$$

where X and X' are independent samples from S. This is sufficient to establish Theorem 1 using the relation between collision probability and statistical distance from Claim 10 in Appendix A.

Proof. When $p = 1/2$ the analysis is standard, so we will assume that $p = \frac{1}{m} \cdot \log(m/\delta) \ln(15n/m) < 1/2$. Since entries of M are chosen independently from each other, for any $y \in \{0,1\}^n$, we have

$$\Pr_H[H(y) = 0] = \Pr_a[\langle a, y \rangle = 0]^m = \left(\frac{1 + (1-2p)^{|y|}}{2}\right)^m = \frac{1}{2^m}\sum_{i=0}^{m}\binom{m}{i}(1-2p)^{i|y|}$$

Here $a \sim \{0,1\}^m$ is chosen from the p-biased distribution. Let S_0 be the set $\{x_0 + x \colon x \in S\}$. Then

$$\Pr_{H,X}[H(X) = H(x_0)] = \Pr_{H,y \sim S_0}[H(y) = 0]$$
$$= E_{y \sim S_0}\left[\frac{1}{2^m}\sum_{i=0}^{m}\binom{m}{i}(1-2p)^{i|y|}\right] = \frac{1}{2^m}\sum_{i=0}^{m}\binom{m}{i}E_{y \sim S_0}[(1-2p)^{i|y|}]$$

Let $a_i = E_{y \sim S_0}[(1-2p)^{i|y|}]$. We now upper bound the sum $\sum_{i=0}^{m}\binom{m}{i}a_i$ by $1 + \delta + O(2^{-k+m})$. We will consider two cases: When i is small – specifically, $i \leq k/(2\log(m/\delta))$, we show that a_i

decreases at a rate faster than $(m/\delta)^{-i}$, so the sum is dominated by the term $i = 0$. When i is large, we want to bound both a_i and $\binom{m}{i}$ by its largest possible value. To achieve this, we have to further split the large i's into "mildly large" and "very large" ones and apply the argument to each summation separately. The resulting contribution is $O(2^{-k+m})$. Notice that, in the case $m \leq k/(2\log(m/\delta))$, we need not consider the contribution of large i's thus we can improve p to be $\frac{1}{k} \cdot \log(m/\delta)\ln(15n/k)$ by using the same analysis for small i's.

The small i's. We show that if $i \leq k/(2\log(m/\delta))$, then $a_i \leq (m/\delta)^{-1.8i}$ and therefore

$$\sum_{i=0}^{k/(2\log(m/\delta))} \binom{m}{i} a_i \leq (1 + (m/\delta)^{-1.8})^m \leq e^{\delta^{1.8}/m^{0.8}} \leq 1 + \delta.$$

To bound a_i we apply Hölder's inequality, which says that for every $B \geq 1$:

$$a_i = \mathrm{E}_{y \sim S_0}[(1-2p)^{i|y|}] = \frac{1}{|S|} \sum_{y \in \{0,1\}^n} 1_{y \in S_0} \cdot (1-2p)^{i|y|} \leq \frac{1}{|S|} |S|^{1-1/B} \sum_{y \in \{0,1\}^n} \left((1-2p)^{Bi|y|}\right)^{1/B}$$

The last expression can be simplified to give

$$a_i \leq \left(\frac{(1 + (1-2p)^{Bi})^n}{|S|}\right)^{1/B} \leq \left(\frac{(1 + e^{-2piB})^n}{|S|}\right)^{1/B}.$$

We choose $B = k/(2i\log(m/\delta))$, which is at least one because $i \leq k/(2\log(m/\delta))$. By our choice of p, it follows that $2piB \geq \ln(15n/m)$ and so

$$a_i \leq \left(\frac{(1 + m/15n)^n}{|S|}\right)^{1/B} \leq \left(\frac{e^{m/15}}{2^k}\right)^{1/B} \leq \left(\frac{e^{k/15}}{2^k}\right)^{1/B} \leq (2^{-0.9})^{k/B} = (m/\delta)^{-1.8i}.$$

The large i's. We have that

$$a_i = \mathrm{E}_{y \sim S_0}[(1-2p)^{i|y|}] \leq \frac{1}{|S|} \sum_{y \in \{0,1\}^n} (1-2p)^{i|y|} = \frac{(1 + (1-2p)^i)^n}{|S|} \leq \frac{(1 + e^{-2pi})^n}{|S|}. \qquad (1)$$

When $i \geq m/4$, the last expression is at most $(1 + e^{-pm/2})^n/|S|$. By our choice of p, $pm/2 \geq \frac{1}{2}\log m \ln(15n/m)$. Optimizing for $\log m$, it can be calculated that this expression is at least $\ln n$ when n is sufficiently large. It then follows that

$$a_i \leq \frac{(1 + 1/n)^n}{|S|} \leq e 2^{-k} \qquad \text{and so} \qquad \sum_{i=m/4}^{m} \binom{m}{i} a_i \leq 2^m \cdot e 2^{-k} = e 2^{-k+m}.$$

Finally, we handle the i's in the range $k/(2\log(m/\delta)) < i < m/4$. Using (1) and the lower bound on i, we have that

$$a_i \leq \frac{(1 + m/15n)^n}{|S|} \leq \frac{e^{m/15}}{|S|} \leq 2^{0.1m-k}$$

and so

$$\sum_{k/(2\log(m/\delta))<i<m/4} \binom{m}{i} a_i \leq 2^{0.1m-k} \sum_{i=0}^{m/4} \binom{m}{i} \leq 2^{0.1m-k} \cdot 2^{H(1/4)m+O(1)} \leq 2^{-k+m+O(1)},$$

where H is the binary entropy function, and $H(1/4) \leq 0.9$. $\qquad \square$

3 Proof of Theorem 2

Let distribution $\overline{X}$ be a truncated variant of the p-biased distribution X where p is chosen so that $H(p) = 2m/n$ and $p \leq 1/2$. The distribution $\mathcal{D}$ on distributions is defined as follows. Choose y uniformly from $\{0,1\}^n$ and output $y + \overline{X}$. To prove Theorem 2, we will show for most choices of y, there exists a statistical test T_y that distinguishes $h(y + \overline{X})$ from the uniform distribution U and then argue the expected statistical distance over choice of y between $h(y + \overline{X})$ and U is large. We will define T_y shortly for $y \in \{0,1\}^n$ and first argue that for most choice of y, T_y distinguishes $h(y + X)$ from U. Then we will show how to define $\overline{X}$ of min-entropy $1.5m$ in a way that X and $\overline{X}$ are statistically close. Finally, we conclude that the expected statistical distance over choice of y between $h(y + \overline{X})$ and U is at least $1 - e^{-m^{\Omega(1)}}$.

The following bounds on p are obtained by plugging in $H(p) = 2m/n$ in Lemma 11 in Appendix B:

$$\frac{m}{3n \log_2(n/m)} \leq p \leq \frac{2m}{n \log_2(n/2m)}. \tag{2}$$

Now suppose h has sparsity $(m/2p)\beta \log m$, where β is a sufficiently small constant, say $\beta = 0.08$. Notice this β also satisfies $n \leq m^{1+2\beta}$ (since by assumption $m \geq n^{0.99}$). Partition the inputs of h' into two sets H and L, where H contains those inputs that participate in at least $m^{2-6\beta}/pn$ outputs of h', and L contains the rest. By Markov's inequality (using the assumption $n \leq m^{1+2\beta}$), H has size at most $m^{8\beta}\beta \log k$. For $x \in \{0,1\}^n$, let x_0 and x_1 denote its projections onto H and L, respectively. For every $y \in \{0,1\}^n$, we define the statistical test

$$T_y = \{z \in \{0,1\}^m : \Delta(h(x_0, y_1), z) \leq 1/2 - m^{-\beta}/4 \text{ for some } x_0, \}$$

where $\Delta(a, b)$ is relative Hamming distance between the strings a and b, i.e. the fraction of positions in which they differ.

Claim 4. *For sufficiently large k, $\Pr_X[h(X + y) \in T_y] \geq 1 - e^{-m^{3\beta}/2}$ for at a least $1 - 1 - e^{-m^{3\beta}/2}$ values of $y \in \{0,1\}^n$.*

In the proof we will need the following fact about Boolean functions $f \colon \{0,1\}^d \to \{0,1\}$

$$\Pr_{X,Y}[f(X + Y) \neq f(Y)] \leq \tfrac{1}{2} - \tfrac{1}{2}(1 - 2p)^d \tag{3}$$

where Y is uniformly distributed in $\{0,1\}^n$, and X is chosen independently from the p-biased distribution on $\{0,1\}^n$. This fact follows easily by Fourier analysis [O'D02] and was also used by Viola [Vio05] in a context related to ours.

We will also make use of the following inequality of Gavinsky et al. [GLSS12]. A collection of indicator random variables $Z_1, \ldots, Z_m$ is called a *read t family* of functions if there exist independent random variables $X_1, \ldots, X_n$ such that each X_i

Then we will apply tail bound for *read t family* of functions [GLSS12] to show for most choice of (x, y) outcome concentrate on expectation. Indicator random variables $Z_1, \ldots, Z_m$ is a *read t family* if they can written as a function of independent random variables $X_1, \ldots, X_n$ where each X_i affects at most t of the Z_i's. Then for every $\varepsilon > 0$,

$$\Pr[Z \geq \mathrm{E}[Z] + \varepsilon m] \leq e^{-2\varepsilon^2 m/t}. \tag{4}$$

where $Z = Z_1 + \cdots + Z_m$.

Proof. We will show that for every choice of x_0, y_0, with probability $1 - e^{-\Omega(m^{1-\beta})}$ over the choice of x_1, y_1, $h(x+y) = h(x_0 + y_0, x_1 + y_1)$ is in T_y. Fix x_0, y_0 and consider the function $h^{x_0 + y_0}(x_1) = h(x_0 + y_0, x_1)$. Let $Z = Z_1 + \cdots + Z_m$, where

$$Z_i = \begin{cases} 1, & \text{if } h_i^{x_0 + y_0}(x_1 + y_1) \neq h_i^{x_0 + y_0}(y_1), \\ 0, & \text{otherwise.} \end{cases}$$

Suppose $h_i^{x_0 + y_0}$ depends on d_i inputs for $1 \leq i \leq m$, by (3) we have

$$\mathrm{E}[Z_i] \leq 1/2(1 - (1 - 2p)^{d_i}).$$

By linearity of expectation and $\sum_{i=1}^m d_i \leq (m/2p)\beta \log m$, we get

$$\mathrm{E}[Z] \leq m/2(1 - (1 - 2p)^{\beta \log m/2p}) \leq m(1/2 - m^{-\beta}/2).$$

Now we apply tail bound (4) to $Z_1, \ldots, Z_m$ with $t = m^{2-6\beta}/pn$ and $\varepsilon = m^{-\beta}/4$ to obtain

$$\Pr[Z \geq m(1/2 - m^{-\beta}/4)] \leq e^{-2\varepsilon^2 m/(m^{2-6\beta}/pn)} \leq e^{-m^{3\beta}}$$

where we used the estimate (2) to lower bound pn. In other words,

$$\Pr_{x_1, y_1}[\Delta(h_{x_0 + y_0}(y_1), h_{x_0 + y_0}(x_1 + y_1)) \leq 1/2 - m^{-\beta}/4] \geq 1 - e^{-m^{3\beta}}.$$

It follows that

$$\Pr_{x,y}[h(x+y) \in T_y] \geq \mathrm{E}_{x_0, y_0}\left[\Pr_{x_1, y_1}[\Delta(h_{x_0 + y_0}(y_1), h_{x_0 + y_0}(x_1 + y_1)) \leq 1/2 - m^{-\beta}/4]\right]$$
$$\geq 1 - e^{-m^{3\beta}}.$$

Applying Markov's inequality, we conclude that for at least $1 - e^{-m^{3\beta}/2}$ choices of y,

$$\mathrm{E}_X[h(X + y) \in T_y] \geq 1 - e^{-m^{3\beta}/2}. \qquad \square$$

Claim 5. *For any fixed $y \in \{0,1\}^n$, with probability $1 - 2^{-\Omega(m^{1-2\beta})}$ over the choice of a uniform $U \sim \{0,1\}^m$, U is not in T_y.*

Proof. Since H has size at most $m^{8\beta}\beta \log k$, the range of $h'(x_0, y_1)$ has at most $2^{m^{8\beta}\beta \log m}$ elements. For every such element $h(x_0, y_1)$, the probability that U is within distance $m/2 - m^{1-\beta}/4$ to $h(x_0, y_1)$ can be computed by Chernoff bounds to be at most $2^{-\Omega(m^{1-2\beta})}$. Taking a union bound over all such $h(x_0, y_1)$, we obtain

$$\Pr[U \in T_y] \leq 2^{m^{8\beta}\beta \log m} 2^{-\Omega(m^{1-2\beta})} = 2^{-\Omega(m^{1-2\beta})}$$

as long as $\beta < 1/10$ and m is sufficiently large. $\qquad \square$

From these two claims, it follows that for a $1 - e^{-m^{3\beta}/2}$ choices of y,

$$\Pr_X[h(X + y) \in T_y] - \Pr_U[U \in T_y] \geq 1 - 1 - e^{-m^{3\beta}/2} - 2^{-\Omega(m^{1-2\beta})}. \tag{5}$$

To finish the proof, we show how to replace X with another variable $\overline{X}$ of min-entropy at least $1.5m$ that is statistically close to it. We define $\overline{X}$ as follows: First, choose X from the p-biased distribution. If the Hamming weight of X is at least $0.9pn$, set $\overline{X} = X$. Otherwise, let $\overline{X}$ be uniformly random in $\{0,1\}^n$. We prove the following claim in Appendix C:

8

Claim 6. $\overline{X}$ *has min-entropy at least* $1.5m$.

Clearly the same conclusion holds for the distribution $\overline{X} + y$. The statistical distance between X and $\overline{X}$ is upper bounded by the probability that X has Hamming weight less than $0.9pn$. By Chernoff bounds, the probability of this is at most $\exp(-\Omega(pn))$, which using the lower bound (2) is at least $\exp(-\Omega(m/\log(n/m))) = \exp(-m^{\Omega(1)})$ (since $m \geq n^{-0.99}$). Applying the triangle inequality, for all y satisfying (5) we have

$$\Pr_{\overline{X}}[h(\overline{X} + y) \in T_y] - \Pr_U[U \in T_y] \geq 1 - e^{-m^{\Omega(1)}}.$$

We conclude that the expected statistical distance between $h(\overline{X} + y)$ and U_m for a random choice of y is at least $(1 - e^{-m^{\Omega(1)}})(1 - e^{-m^{\Omega(1)}}) = 1 - e^{-m^{\Omega(1)}}$.

4 Proof of Theorem 3

As in the proof of Theorem 1, it is sufficient to show for every set S of size 2^k and every x_0 in S and r_0 in $\{0,1\}^s$,

$$\Pr_{M,X,R}[MX + BR = Mx_0 + Br_0] \leq \frac{1 + (1/\delta) \cdot 2^{-k-s+m}}{2^m}$$

where the probability is taken over the random matrix M, X chosen uniformly from S and R chosen uniformly from $\{0,1\}^s$.

Assume that $p \leq 1/2$. Let S_0 be the set $\{x + x_0 : x \in S\}$. Then

$$\Pr_{M,X,R}[MX + BR = Mx_0 + Br_0] = \Pr_{M,X,R}[M(X + x_0) = B(R + r_0)] = \Pr_{M,Y,R}[MY = BR]$$

where Y is a random element from S_0. Let M_i, B_i denote the ith row of M and B. Then

$$\Pr[MY = BR] = \mathrm{E}_{M,Y,R}\Big[\prod_{i=1}^m \frac{1 + (-1)^{M_iY + B_iR}}{2}\Big]$$

$$= \frac{1}{2^m} \sum_{T \subseteq [m]} \mathrm{E}_{M,Y,R}\big[(-1)^{\sum_{i \in T} M_iY + B_iR}\big] = \frac{1}{2^m} \sum_{T \subseteq [m]} \mathrm{E}_{M,Y}\big[(-1)^{\sum_{i \in T} M_iY}\big] \mathrm{E}_R\big[(-1)^{\sum_{i \in T} B_iR}\big].$$

Since any $t = m/2K$ rows of B are linearly independent, for every nonempty T of size at most t, $\sum_{i \in T} B_i \neq 0$ and so $\mathrm{E}[(-1)^{\sum_{i \in T} B_iR}] = 0$. On the other hand for every T of size at least t we have

$$\mathrm{E}_{M,Y}\big[(-1)^{\sum_{i \in T} M_iY}\big] = \frac{1}{2^k} \sum_{y \in S_0} \mathrm{E}_M\big[(-1)^{\sum_{i \in T} M_iy}\big] = \frac{1}{2^k} \sum_{y \in S_0} (1 - 2p)^{|y| \cdot |T|}$$

$$\leq \frac{1}{2^k} \sum_{y \in \{0,1\}^n} (1 - 2p)^{t|y|} \leq \frac{1}{2^k}(1 + (1 - 2p)^t)^n \leq \frac{e^{ne^{-2pt}}}{2^k}.$$

9

Since B has full rank, the condition $\sum_{i\in T} B_i = 0$ is satisfied for at most 2^{m-s} sets T. Hence,

$$\sum_{T\subseteq[m]} \mathrm{E}_{M,Y}[(-1)^{\sum_{i\in T} M_i Y}]\,\mathrm{E}_R[(-1)^{\sum_{i\in T} B_i R}] = 1 + \sum_{T:|T|>t} \mathrm{E}_{M,Y}[(-1)^{\sum_{i\in T} M_i y}]\,\mathrm{E}_R[(-1)^{\sum_{i\in T} B_i s}]$$

$$\leq 1 + \sum_{T:|T|>t} \frac{e^{ne^{-2pt}}}{2^k} |\mathrm{E}_R[(-1)^{\sum_{i\in T} B_i R}]|$$

$$\leq 1 + 2^{m-s} \cdot \frac{e^{ne^{-2pt}}}{2^k}$$

$$= 1 + \frac{e^{ne^{-2pt}}}{2^{s+k-m}}.$$

Plugging in $t = \frac{m}{2K}$ and $p = \frac{K}{m}\cdot\ln\frac{n}{\ln c}$ we get the desired bound.

5   Local pseudorandom generators from local one-way functions

A sequence of correlated random variables $X_1,\dots,X_m$ taking values in $\{0,1\}^n$ has (s,ε) *conditional pseudo-min-entropy* r if for every $1\leq i\leq m$, there exists a random variable Y_i jointly distributed with $X_1,\dots,X_{i-1}$ such that the min-entropy of Y conditioned on any choice of $X_1,\dots,X_{i-1}$ is at least r and for every circuit D of size s,

$$\big|\Pr[D(X_i)\mid X_1,\dots,X_{i-1}] - \Pr[D(Y_i)\mid X_1,\dots,X_{i-1}]\big| \leq \varepsilon.$$

Vadhan and Zheng [VZ11] give the following construction of conditional pseudo-min-entropy sequences from a one-way function $f\colon\{0,1\}^n\to\{0,1\}^n$. Let $z_i, 1\leq i\leq t$ be the random strings

$$z_i = {}_{o_i}\lfloor f(x_{i1})\circ x_{i1}\circ\cdots\circ f(x_{ik})\circ x_{ik}\rfloor_{n-o_i}$$

where $1\leq o_i\leq n$ is an offset, $x_{i1},\dots,x_{ik}$ are random strings, and ${}_f\lfloor y\rfloor_\ell$ denotes truncating the first f and last ℓ bits of y respectively. Let $X_j = z_{1j}z_{2j}\dots z_{tj}$, where z_{ij} denotes the jth bit of z_i. Vadhan and Zheng prove the following theorem (we state it in the nonuniform setting).

Theorem 7 (Vadhan and Zheng). *Suppose $f\colon\{0,1\}^n\to\{0,1\}^n$ is computable by a circuit of size $\mathrm{poly}(n)$ and is hard to invert on a $1/s$ fraction of inputs by circuits of size s. There exists offsets $o_1,\dots,o_t$ such that for every ε, $X_1,\dots,X_m$ has $(s^{\Omega(1)}/\mathrm{poly}(n\varepsilon),\varepsilon)$ conditional pseudo-min-entropy at least $t(1/2+\Omega((\log s)/n))$ where $k = O(n/\log s)$, $t = O((n/\log s)^2\log^2 n\log(1/\varepsilon))$ and $m = 2(k-1)n$.*

The following claim was proved in the uniform setting by Haitner, Reingold, and Vadhan. We need a nonuniform version of it, whose proof is analogous. We include it at the end of this section for completeness.

Claim 8. *Suppose $X_1,\dots,X_m$ (where X_i takes values in $\{0,1\}^t$) has (T,ε_1) conditional pseudo-min-entropy α. Let H be an extractor family for min-entropy α with error ε_2 so that every function in H is computable in size T_0. Then with probability at least $1/2$ over the choice of H the distribution $(H(X_1,R_1),\dots,H(X_m,R_m))$ is $(T-mT_0, m\varepsilon_1+2m^2\varepsilon_2)$-pseudorandom where $R_1,\dots,R_m\sim\{0,1\}^r$.*

Instantiating Claim 8 with the function family from Theorem 3 where we set the output length of function to be $t(1/2 + \Omega((\log s)/n)) + r$ and let d be the entropy loss. We obtain the following consequence for the function

$$G(x_{11}, \ldots, x_{tk}, r_1, \ldots, r_m) = (H(X_1, r_1), \ldots, H(X_m, r_m)).$$

Proposition 9. *Suppose* $f \colon \{0,1\}^n \to \{0,1\}^n$ *is an ℓ-local function computable by a circuit of size* $\mathrm{poly}(n)$ *and is hard to invert on a $1/s$ fraction of inputs by circuits of size s. With probability at least $1/2$ over the choice of H, $G \colon \{0,1\}^{nkt+mr} \to \{0,1\}^{nkt(1+\Omega((\log s)/n))+mr}$ is an $O(\ell \cdot (nkt \ln(t/\ln c) + mr))$-sparse, $(s^{\Omega(1)}/\mathrm{poly}(n\varepsilon), \mathrm{poly}(n)(\varepsilon + \sqrt{c}2^{-d/2}))$ pseudorandom generator where $k = O(n/\log s)$, $t = O((n/\log s)^2 \log^2 n \log(1/\varepsilon))$ and $m = 2(k-1)n$.*

We can improve the locality of G at the expense of increasing its input and output length by the factor of $O(\ln(t/\ln c))$ via the following transformation of Applebaum, Ishai, and Kushilevitz. For every output of G, which is obtained by applying a sparse linear transformation to some X_j and therefore has the form

$$X_{jk_1} + \cdots + X_{jk_t}$$

introduce auxiliary new inputs $r_{j3}, r_{j4}, \ldots, r_{j(t-1)}$ for G and replace its corresponding output by the tuple

$$(X_{jk_1} + X_{jk_2} + r_{j3}, r_{j3} + X_{jk3} + r_{j4}, \ldots, r_{jt-1} + X_{j(t-1)} + X_{jt}).$$

Call this new function G'. Applebaum et al. show that if G is $(s^{\Omega(1)}, s^{-\Omega(1)})$-pseudorandom, so is G'. Since every bit of X_j comes either from some input x_i or from some output $f(x_i)$, it follows that if f has locality ℓ, then G' has locality 3ℓ.

Proof of Claim 8. Let Y_i be the conditional min-entropy model for X_i. We consider the hybrid distributions

$$X^{(i)} = (H(X_1, R_1), \ldots, H(X_{i-1}, R_{i-1}), H(X_i, R_i), U_{i+1}, \ldots, U_m) \quad \text{and}$$
$$Y^{(i)} = (H(X_1, R_1), \ldots, H(X_{i-1}, R_{i-1}), H(Y_i, R_i), U_{i+1}, \ldots, U_m)$$

where $U_1, \ldots, U_m$ are uniformly random and independent. By the definition of conditional pseudo-min-entropy, for every i the distributions $X^{(i)}$ and $Y^{(i)}$ are $(T - mT_0, \varepsilon_1)$-indistinguishable. Because $\mathcal{H}$ is an extractor family, the distributions $(H, H(Y_i, R_i) \mid X_1, \ldots, X_{i-1})$ and (H, U_i) are within statistical distance ε_2 for any choice of $X_1, \ldots, X_{i-1}$. It follows that $(H, Y^{(i)})$ and $(H, X^{(i-1)})$ are within statistical distance at most ε_2, so by Markov's inequality $Y^{(i)}$ and $X^{(i-1)}$ are within statistical distance $2m\varepsilon_2$ with probability at least $1 - 1/2m$ over the choice of H. By a union bound, with probability at least $1/2$ over the choice of H, $Y^{(i)}$ and $X^{(i-1)}$ are $2m\varepsilon_2$-statistically close for all i. For such a choice of H, by the triangle inequality $X^{(m)}$ is $(T - mT_0, m\varepsilon_1 + 2m^2\varepsilon_2)$ indistinguishable from $X^{(0)}$. Since $X^{(m)} = (H(X_1R_1), \ldots, H(X_m, R_m))$ and $X^{(0)}$ is the uniform distribution, we obtain the desired conclusion. $\square$

References

[AIK04a]   B. Applebaum, Y. Ishai, and E. Kushilevitz. Cryptography in nc0. In *Foundations of Computer Science, 2004. Proceedings. 45th Annual IEEE Symposium on*, pages 166–175. IEEE, 2004.

[AIK04b]   B. Applebaum, Y. Ishai, and E. Kushilevitz. Cryptography in nc0. In *Foundations of Computer Science, 2004. Proceedings. 45th Annual IEEE Symposium on*, pages 166–175. IEEE, 2004.

[AIK06]   B. Applebaum, Y. Ishai, and E. Kushilevitz. On pseudorandom generators with linear stretch in nc 0. *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, pages 260–271, 2006.

[BBR88]   C.H. Bennett, G. Brassard, and J.M. Robert. Privacy amplification by public discussion. *SIAM journal on Computing*, 17:210, 1988.

[BR11]   A. Bogdanov and A. Rosen. Input locality and hardness amplification. *Theory of Cryptography*, pages 1–18, 2011.

[CRVW02]   Michael R. Capalbo, Omer Reingold, Salil P. Vadhan, and Avi Wigderson. Randomness conductors and constant-degree lossless expanders. In *Proceedings of the 34th Annual ACM Symposium on Theory of Computing*, pages 659–668, 2002.

[CW79]   J.L. Carter and M.N. Wegman. Universal classes of hash functions. *Journal of computer and system sciences*, 18(2):143–154, 1979.

[DT09]   A. De and L. Trevisan. Extractors using hardness amplification. *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, pages 462–475, 2009.

[Gal62]   R. Gallager. Low-density parity-check codes. *IRE Transactions on Information Theory*, 8(1):21–28, 1962.

[GLSS12]   D. Gavinsky, S. Lovett, M. Saks, and S. Srinivasan. A tail bound for read-k families of functions. *Arxiv preprint arXiv:1205.1478*, 2012.

[HILL99]   J. Håstad, R. Impagliazzo, L.A. Levin, and M. Luby. A pseudorandom generator from any one-way function. *SIAM J. Comput.*, 28(4):1364–1396, 1999.

[HRV10]   I. Haitner, O. Reingold, and S. Vadhan. Efficiency improvements in constructing pseudorandom generators from one-way functions. In *Proceedings of the 42nd ACM symposium on Theory of computing*, pages 437–446. ACM, 2010.

[IKOS08]   Y. Ishai, E. Kushilevitz, R. Ostrovsky, and A. Sahai. Cryptography with constant computational overhead. In *Proceedings of the 40th annual ACM symposium on Theory of computing*, pages 433–442. ACM, 2008.

[ILL89]   R. Impagliazzo, L.A. Levin, and M. Luby. Pseudo-random generation from one-way functions. In *Proceedings of the twenty-first annual ACM symposium on Theory of computing*, pages 12–24. ACM, 1989.

[Lu04]   C.J. Lu. Encryption against storage-bounded adversaries from on-line strong extractors. *Journal of Cryptology*, 17(1):27–42, 2004.

[Mau92]   U.M. Maurer. Conditionally-perfect secrecy and a provably-secure randomized cipher. *Journal of Cryptology*, 5(1):53–66, 1992.

[MNT93]   Y. Mansour, N. Nisan, and P. Tiwari. The computational complexity of universal hashing. *Theoretical Computer Science*, 107(1):121–133, 1993.

[Nis92]   N. Nisan. Pseudorandom generators for space-bounded computation. *Combinatorica*, 12(4):449–461, 1992.

[NZ93]   N. Nisan and D. Zuckerman. Randomness is linear in space. In *Symposium on Theory of Computing*, pages 235–244. Citeseer, 1993.

[O'D02]   Ryan O'Donnell. Hardness amplification within np. In *Proceedings of the thiry-fourth annual ACM symposium on Theory of computing*, pages 751–760. ACM, 2002.

[RR99]   R. Raz and O. Reingold. On recycling the randomness of states in space bounded computation. In *Proceedings of the thirty-first annual ACM symposium on Theory of computing*, pages 159–168. ACM, 1999.

[RT97]   J. Radhakrishnan and A. Ta-Shma. Tight bounds for depth-two superconcentrators. In *Proceedings of the 38th Annual Symposium on Foundations of Computer Science*, pages 585–594. IEEE, 1997.

[Sha04]   R. Shaltiel. Recent developments in explicit constructions of extractors. *Current Trends in Theoretical Computer Science: Algorithms and complexity*, 1:189, 2004.

[SS96]   M. Sipser and D.A. Spielman. Expander codes. *Information Theory, IEEE Transactions on*, 42(6):1710–1722, 1996.

[Vad03]   S. Vadhan. On constructing locally computable extractors and cryptosystems in the bounded storage model. *Advances in Cryptology-CRYPTO 2003*, pages 61–77, 2003.

[Vio05]   E. Viola. The complexity of constructing pseudorandom generators from hard functions. *Computational Complexity*, 13(3):147–188, 2005.

[VV86]   L.G. Valiant and V.V. Vazirani. Np is as easy as detecting unique solutions. *Theoretical Computer Science*, 47:85–93, 1986.

[VZ11]   S. Vadhan and J. Zheng. Characterizing pseudoentropy and simplifying pseudorandom generator constructions. Technical Report TR11-141, Electronic Colloquium on Computational Complexity, 2011.

[ZB11]   H. Zhou and J. Bruck. Linear extractors for extracting randomness from noisy sources. In *Information Theory Proceedings (ISIT), 2011 IEEE International Symposium on*, pages 1738–1742. IEEE, 2011.

A Statistical distance versus collision probability

Here we give a standard bound of the statistical distance of the distributions $(H, H(X))$ and (H, U) in terms of the collision probability of H. The proof is very well known when the distribution over H is uniform. In our application, however, it is not, so we include the proof for completeness.

Claim 10. *Let H be a distribution on functions from $\{0,1\}^n$ to $\{0,1\}^m$, X be a distribution over $\{0,1\}^n$ and U be the uniform distribution over $\{0,1\}^m$. The statistical distance between $(H, H(X))$ and (H, U) is at most*

$$\frac{1}{2}\sqrt{2^k \Pr_{H,X,X'}[H(X) = H(X')] - 1}.$$

where X and X' are independent samples from the same distribution.

Proof. We upper bound the ℓ_1 distance, which is twice the statistical distance:

$$
\begin{aligned}
\ell_1((H, h(X)), (H, U)) &= \sum_{h,b}\left|\Pr_{H,X}[H = h \wedge H(X) = b] - \Pr_H[H = h]2^{-m}\right| \\
&= \sum_{h,b}\sqrt{\Pr_H[H = h]} \cdot \sqrt{\Pr_H[H = h]}\left|\Pr_X[h(X) = b] - 2^{-m}\right| \\
&\leq \sqrt{\sum_{h,b}\Pr_H[H = h]} \cdot \sqrt{\sum_{h,b}\Pr_h[H = h]\left(\Pr_{H,X}[H(X) = b] - 2^{-m}\right)^2} \\
&= \sqrt{2^m} \cdot \sqrt{\Pr_{H,X,X'}[H(X) = H(X')] - 2^{-m}} \\
&= \sqrt{2^m \Pr_{H,X,X'}[H(X) = H(X')] - 1}. \qquad \square
\end{aligned}
$$

B Bounds on the inverse of entropy

Lemma 11. *For every $p \in (0, 1/2]$,*

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

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

$$\frac{1}{p} \geq \frac{1}{H(p)}\log_2\frac{1}{p} \geq \frac{1}{H(p)}\log_2\left(\frac{1}{H(p)}\log_2\frac{1}{p}\right) \geq \frac{1}{H(p)}\log_2\frac{1}{H(p)}$$

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

$$\frac{1}{p} \leq \frac{2}{H(p)}\log_2\frac{1}{p} \leq \frac{2}{H(p)}\log_2\left(\frac{2}{H(p)}\log_2\frac{1}{p}\right).$$

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

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

C   Proof of Claim 6

The maximum probability in $\overline{X}$ is attained by those strings in $\{0,1\}^n$ that have Hamming weight exactly $0.9pn$. Let a be such a string. Then

$$
\begin{aligned}
\Pr[\overline{X} = a] &\leq \Pr[X = a] + \Pr[U = a] \\
&= p^{0.9pn}(1-p)^{1.1(1-p)n} + 2^{-n} \\
&\leq 2^{-nH(p)} \cdot p^{-0.1pn} + 2^{-n} \\
&= 2^{-nH(p)} \cdot 2^{0.1p\log_2(1/p)n} + 2^{-n} \\
&\leq 2^{-nH(p)} \cdot 2^{0.1nH(p)} + 2^{-n} \\
&= 2^{-0.9nH(p)} + 2^{-n} \\
&= 2^{-1.8m} + 2^{-n} \leq 2^{-1.5m}.
\end{aligned}
$$

15