Extractors for circuit sources

Emanuele Viola*

December 20, 2011

Abstract

We obtain the first deterministic extractors for sources generated (or sampled) by small circuits of bounded depth. Our main results are:

(1) We extract $k(k/nd)^{O(1)}$ bits with exponentially small error from n-bit sources of min-entropy k that are generated by functions $f : \{0,1\}^\ell \to \{0,1\}^n$ where each output bit depends on $\leq d$ input bits. In particular, we extract from NC^0 sources, corresponding to $d = O(1)$.

(2) We extract $k(k/n^{1+\gamma})^{O(1)}$ bits with super-polynomially small error from n-bit sources of min-entropy k that are generated by $\mathrm{poly}(n)$-size AC^0 circuits, for any $\gamma > 0$.

As our starting point, we revisit the connection by Trevisan and Vadhan (FOCS 2000) between circuit lower bounds and extractors for sources generated by circuits. We note that such extractors (with very weak parameters) are equivalent to lower bounds for generating distributions (FOCS 2010; with Lovett, CCC 2011). Building on those bounds, we prove that the sources in (1) and (2) are (close to) a convex combination of high-entropy "bit-block" sources. Introduced here, such sources are a special case of affine ones. As extractors for (1) and (2) one can use the extractor for low-weight affine sources by Rao (CCC 2009).

Along the way, we exhibit an explicit boolean function $b : \{0,1\}^n \to \{0,1\}$ such that $\mathrm{poly}(n)$-size AC^0 circuits cannot generate the distribution $(Y, b(Y))$, solving a problem about the complexity of distributions.

Independently, De and Watson (RANDOM 2011) obtain a result similar to (1) in the special case $d = o(\lg n)$.

*Supported by NSF grant CCF-0845003. Email: viola@ccs.neu.edu

1 Introduction

Access to a sequence of uniform and independent bits (or numbers) is crucial to efficient computation, but available sources of randomness appear to exhibit biases and correlations. So a significant amount of work is put into "purifying" such sources, by applying a *deterministic* function, known as *extractor*, that given as input a weak, n-bit source of randomness outputs m bits that are close to uniform over $\{0,1\}^m$ (in statistical distance).

The theoretical investigation of this problem goes back to von Neumann [vN51]. Since then, many researchers have been analyzing increasingly complex sources, modeled as probability distributions D with high min-entropy k (i.e., $\Pr[D = a] \leq 2^{-k}$ for every a), see e.g. [Blu86, CG88, SV86].

In 2000, Trevisan and Vadhan [TV00] consider sources that can be generated, or sampled, *efficiently*. That is, the n-bit source is the output of a small circuit $C : \{0,1\}^\ell \to \{0,1\}^n$ on a uniform input. As they write, "one can argue that samplable distributions are a reasonable model for distributions actually arising in nature." They point out that even extracting 1 bit from from such sources of min-entropy $k = n - 1$ entails a circuit lower bound for related circuits. On the other hand, assuming the existence of a function computable in time $2^{O(n)}$ that requires Σ_5 circuits of size $2^{\Omega(n)}$, Trevisan and Vadhan obtain extractors (for min-entropy $k = (1 - \Omega(1))n$). The gap between their positive and negative result prevents one from obtaining unconditional results even for "restricted" classes of circuits for which we do have lower bounds, such as the class AC^0 of unbounded fan-in circuits of constant depth.

The word "restricted" is in quotes because seemingly crippled circuits such as AC^0 or DNF turn out to have surprising power when it comes to *sampling* as opposed to *computing* [Vio10]. In fact, until this work it was an open problem to exhibit *any* explicit distribution on n bits with min-entropy $n - 1$ that cannot be sampled in AC^0! The solution to this problem is obtained in this paper as a corollary to our main results: extractors for sources sampled by restricted classes of circuits, which we simply call *circuit sources*.

A main difficulty in obtaining such extractors is that circuit sources are not easily broken up in *independent* blocks, a property that is heavily exploited to obtain extractors for various sources, including independent sources (see e.g. [Li11a] and the references therein), bit-fixing sources [CGH$^+$85, KZ07, GRS06, Rao09], and small-space sources [KRVZ11]. One type of sources that somewhat escaped this "independence trend," and that is especially important for this work, is the affine one over the field with two elements, i.e., distributions that are uniform over an affine sub-space of $\{0,1\}^n$ of dimension k. Here a line of works exploits the *algebraic structure* to obtain extractors [BKS$^+$10, Bou07, Rao09, BSK09, Yeh10, Li11b, Sha11]. But again, algebraic structure does not seem present in circuit sources, at first sight.

1.1 Our results

We obtain the first extractors for sources generated by various types of circuits, such as AC^0. This is achieved by exhibiting new reductions that show that those sources are (close to) a convex combination of (a special case of) affine sources. Depending on which affine extractor is used, one extracts from circuit sources with various parameters. We state next

some extractors obtained using Rao's extractor for low-weight affine sources [Rao09].

The following theorem extracts from *local* sources, i.e., n-bit sources that are the output distribution of a function $f : \{0,1\}^\ell \to \{0,1\}^n$ where each bit f_i depends on $\le d$ input bits. We extract $m \ge k(k/n)^{O(1)}$ bits. The theorem and the discussion below give a more refined bound on m. The notation $\tilde{\Omega}$ hides logarithmic factors; all logarithms are in base 2.

Theorem 1.1 (Extractor for local sources)**.** *For some $\rho > 0$, any $d = d(n), k = k(n)$:*

There is an explicit function $\text{Ext} : \{0,1\}^n \to \{0,1\}^m$ *that extracts m bits with error $\epsilon \le 2^{-m^{\Omega(1)}}$ from any d-local source with min-entropy k, provided $2dn/k < m^\rho$, for:*

(1) $m = \Omega(k(k/n)^2 \lg(d)/\lg(4n/k)d^3) = \tilde{\Omega}(k(k/n)^2 d^3)$, or

(2) $m = \Omega(k(k/n)/d^2 2^d)$.

Note that Theorem 1.1.(1) extracts from some sublinear entropy $k = n^{1-\Omega(1)}$ and simultaneously polynomial locality $d = n^{\Omega(1)}$. Also, from NC^0 sources ($d = O(1)$) of min-entropy $k = \Omega(n)$, Theorem 1.1 (either setting) extracts $\Omega(n)$ bits with error $2^{-n^{\Omega(1)}}$. The error can be improved to $2^{-\Omega(n)}$ using Bourgain's extractor [Bou07] (cf. [Yeh10, Li11b]).

We also obtain extractors for AC^0 sources, with output length $m \ge k(k/n^{1+\gamma})^{O(1)}$.

Theorem 1.2 (Extractor for AC^0 sources)**.** *For some $\rho > 0$, any $\gamma > 0, d = O(1), k = k(n)$:*

There is an explicit extractor $\text{Ext} : \{0,1\}^n \to \{0,1\}^m$ *with output length $m = k(k/n^{1+\gamma})$ and error $1/n^{\omega(1)}$ for sources with min-entropy k that are generated by AC^0 circuits $C : \{0,1\}^{n^d} \to \{0,1\}^n$ of depth d and size n^d, provided $n^{1+\gamma}/k < m^\rho$.*

The unspecified constant ρ in the "provided" sentences in the above theorems arises from a corresponding unspecified constant in Rao's work [Rao09]. Later in §2.3 we sketch how this constant can be made $\rho = 1 - \alpha$ for any constant $\alpha > 0$. This makes Theorem 1.2 apply provided just $k > n^{2/3+\Omega(1)}$, while if $d = n^{o(1)}$ Theorem 1.1.(1) applies provided $k > n^{3/4+\Omega(1)}$, and if $d = o(\lg n)$ Theorem 1.1.(2) applies provided $k > n^{2/3+\Omega(1)}$.

Assuming a sufficiently good affine extractor, the "provided" sentences are dropped altogether. For example, in the case $d = O(1)$, Theorem 1.1.(2) always extracts $\Omega(k(k/n))$ bits. This is interesting for $k \ge c\sqrt{n}$, and we do not know how to handle smaller values of k even for $d = 2$.

Rao's extractor, and hence the extractor in Theorem 1.1 and 1.2, is a somewhat elaborate algorithm. It is natural to try to obtain simpler extractors. For affine sources, this is investigated in the recent works [BSK09, Li11b]. For local sources, in this paper we show that the majority function extracts one bit, albeit with worse parameters than the previous theorems. More bits can be obtained by truncating the hamming weight of the source, resulting in a simple, *symmetric* extractor.

Theorem 1.3. *There is a symmetric, explicit, deterministic extractor* $\text{Ext} : \{0,1\}^n \to \{0,1\}^m$ *that extracts $m = \Omega(\lg\lg n - \lg d)$ bits with error $\epsilon = (d/\lg n)^{\Omega(1)}$ from any n-bit source with shannon entropy $k \ge n - n^{0.49}$ whose bits are each computable by a decision tree of depth d. To extract $m = 1$ bit, one can take* $\text{Ext} := \text{majority}$.

For example setting $d = \sqrt{\lg n}$ we extract $\Omega(\lg \lg n)$ bits with error $(1/\lg n)^{\Omega(1)}$.

While the parameters of Theorem 1.3 are much weaker than those of the previous extractors, we remark that any symmetric extractor for $(d = \omega(1))$-local sources needs min-entropy $k \geq n(1 - O(\lg d)/d) = n(1 - o(1))$, as can be seen by breaking the source in chunks of d bits and generating a balanced string in each. Also note that the extractor in Theorem 1.3 extracts from shannon entropy, as opposed to min-entropy. It can be shown that any extractor needs shannon entropy $k \geq n(1 - o(1))$ to extract even 1 bit with error $o(1)$.

Extractors vs. the complexity of distributions. As our starting point, we revisit the aforementioned connection between extractors and circuit lower bounds by Trevisan and Vadhan [TV00]: We observe that obtaining extractors (with very weak parameters) for circuit sources is equivalent to proving sampling lower bounds for the same circuits [Vio10, LV11]. For example we record the following no-overhead incarnation of this equivalence. Let $\text{Ext} : \{0,1\}^n \to \{0,1\}$ be any function, and assume for simplicity that Ext is balanced. Then Ext is an extractor with error $< 1/2$ (a.k.a. disperser) for sources of min-entropy $k = n - 1$ generated by circuits of size s if and only if for every $b \in \{0,1\}$ circuits of size s cannot generate the uniform distribution over $\text{Ext}^{-1}(b)$. For general k and possibly unbalanced Ext, we have that Ext is such an extractor if and only if for every $b \in \{0,1\}$ circuits of size s cannot generate a distribution of min-entropy k supported in $\text{Ext}^{-1}(b)$.

By the "if" direction, the sampling bounds in [Vio10] yield extractors with very weak parameters $(d < \lg n, k = n - 1, m = 1, \epsilon < 1/2)$.

The "only if" direction is a slight variant of [TV00, Prop. 3.2]. We use it next in combination with our extractors to address the challenge of exhibiting a boolean function b such that small AC^0 circuits cannot sample $(Y, b(Y))$, raised in [Vio10] (cf. [LV11]). (Actually we use another slight variant of [TV00, Prop. 3.2] which has some overhead but more easily gives a polynomial-time samplable distribution.)

Theorem 1.4. *There is an explicit map $b : \{0,1\}^* \to \{0,1\}$ such that for every $d = O(1)$:*

Let $C : \{0,1\}^{n^d} \to \{0,1\}^{n+1}$ be an AC^0 circuit of size n^d and depth d. The distribution $C(X)$ for uniform X has statistical distance $\geq 1/2^{n^{1-\Omega(1)}}$ from the distribution $(Y, b(Y))$ for uniform $Y \in \{0,1\}^n$.

For b one can take the first bit of the extractor in Theorem 1.2 for $k = n^{1-\Omega(1)}$.

This theorem is also interesting in light of the fact that small AC^0 circuits are able to generate the distribution $(x, \text{Ext}(x))$ where Ext is some affine extractor for min-entropy $\geq (1/2 + \Omega(1))n$. Specifically, for Ext one can choose the inner product function, which has been shown by several researchers to be an extractor, and whose corresponding distribution $(x, \text{Ext}(x))$ can be sampled by small AC^0 circuits [IN96] (cf. [Vio05]). Thus the above theorem is an explanation for the fact that affine extractors for sub-linear min-entropy are more complicated.

1.2 Techniques

A main technical contribution is to show that local sources are (close to) a convex combination of a special case of affine sources which we call "bit-block." A bit-block source is a source in which each output bit is either a constant or a literal (i.e., X_i or $1 - X_i$), and the number of occurrences of each literal is bounded by a parameter we call "block-size."

Definition 1.5 (Bit-block source)**.** *A random variable Y over $\{0,1\}^n$ is a* bit-block source *with block-size w and entropy k if there exist:*

(1) a partition of $[n]$ into $k+1$ sets $B_0, B_1, \ldots, B_k$ such that $|B_i| \leq w$ for every $i \geq 1$,

(2) a string $b_0 \in \{0,1\}^{|B_0|}$, and

(3) k non-constant functions $f_i : \{0,1\} \to \{0,1\}^{|B_i|}$ for any $i \geq 1$,

such that Y can be generated as follows: let $(X_1, \ldots, X_k)$ be uniform over $\{0,1\}^k$, set $Y_{B_0} = b_0$, and for $i \geq 1$ set $Y_{B_i} = f_i(X_i)$. (Where Y_S denotes the bits of Y indexed by S.)

The next theorem shows that local sources Y are a convex combination of bit-block sources, up to an error ϵ. That is, there is an algorithm to sample Y that, except for an error probability of ϵ, outputs a sample from a bit-block source.

Theorem 1.6 (Local is convex combo of bit-block)**.** *Let $f : \{0,1\}^\ell \to \{0,1\}^n$ be a d-local function such that for uniform $X \in \{0,1\}^\ell$, $f(X)$ has min-entropy $\geq k$.*

Then, letting $s := 10^{-5}k(k/n)^2 \lg(d)/\lg(4n/k)d^3 = \tilde{\Omega}(k^3/n^2d^3)$, we have that $f(X)$ is 2^{-s}-close to a convex combination of bit-block sources with entropy $k' = s$ and block-size $\leq 2dn/k$.

Bit-block sources with block-size w are a special case of affine sources of weight w. The latter sources, defined in [Rao09], are generated as $a_0 + \sum_{i=1}^k X_i b_i$, where $a_0, b_1, \ldots, b_k \in \{0,1\}^n$, $(X_1, \ldots, X_k)$ is uniform in $\{0,1\}^k$, and the b_i are independent vectors of hamming weight $\leq w$. To write a bit-block source as in Definition 1.5 as a low-weight affine one, define each vector b_i as 0 except for $b_i|_{B_i} := f_i(1) - f_i(0)$, and vector a_0 as $a_0|_{B_0} = b_0$, $a_0|_{B_i} = f_i(0)$.

Rao [Rao09] extracts $m = k(1 - o(1))$ bits from affine sources of weight $w < k^\rho$. So we obtain the extractor for local sources in Theorem 1.1.(1) by combining Theorem 1.6 with [Rao09]. To obtain Theorem 1.1.(2) we prove a corresponding variant of Theorem 1.6.

Intuition behind the proof of Theorem 1.6. We now explain the ideas behind the proof of Theorem 1.6.(1). Let $f : \{0,1\}^{nd} \to \{0,1\}^n$ be a d-local map, whose output distribution $Y = f(X)$ has min-entropy k. We describe an algorithm to sample Y such that with high probability the algorithm outputs a sample from a high-entropy bit-block source. For the description it is useful to consider the bipartite graph associated to f, where an output variable y_i is adjacent to the $\leq d$ input variables x_j it depends on.

The algorithm.

Note there at most $k/2$ input variables x_j with degree $\geq 2dn/k$. Fix those uniformly at random, and consider the random variable X where the other bits are chosen uniformly at random. Note the output min-entropy is still $\geq k - k/2 = \Omega(k)$.

Now the idea is to iteratively select high-influence input variables, and let their neighborhoods be a block in the bit-block source. (Recall the influence of a variable x on a function is the probability over the choice of the other variables that the output still depends on x.)

Iterate while $H_\infty(f(X)) \geq \Omega(k)$: Since $H_\infty(f(X)) \geq \Omega(k)$, there is an output random variable Y_i with shannon entropy $\geq \Omega(k/n)$. Otherwise, the overall shannon entropy of the output is $\leq o(k/n)n = o(k)$, and shannon entropy is larger than min-entropy.

Consequently, Y_i has high variance: $\min\{\Pr[Y_i = 0], \Pr[Y_i = 1]\} \geq \tilde{\Omega}(k/n)$.

Now, Y_i only depends on d input variables X_j. By the edge isoperimetric inequality [Har64, Har76], there is an input variable X_j that has influence $\geq \tilde{\Omega}(k/nd)$.

Set uniformly at random $N(N(x_j)) \setminus \{x_j\}$, where $N(.)$ denotes "neighborhood," and put x_j aside. (x_j is candidate to contributing to the entropy of the bit-block source.)

Go to the next iteration.

Set uniformly at random all unfixed input variables that were not put aside.

Finally, set uniformly at random the variables that were put aside, and output $f(X)$.

We now analyze this algorithm. First note each iteration fixes $\leq |N(N(x_j))| \leq 2d^2n/k$ variables. Since we iterate as long as $H_\infty(f(X)) \geq \Omega(k)$, we do so $t = \Omega(k(k/nd^2))$ times.

Also, when we set $N(N(x_j)) \setminus \{x_j\}$, with probability at least the influence $\geq \tilde{\Omega}(k/nd)$ the value of x_j influences the output variables in $N(x_j)$. Those variables correspond to a block, which note has size $|N(x_j)| \leq 2dn/k$.

Fixing $N(N(x_j)) \setminus \{x_j\}$ ensures that future actions of the algorithm will not alter the distribution of $N(x_j)$ over the choice of x_j.

Hence out of t iterations we expect $t\Omega(k/nd)$ blocks to be non-constant, corresponding to the entropy of the bit-block source. By a chernoff bound we indeed have $t\Omega(k/nd)$ such blocks with high probability at the final step. This concludes the overview of the proof of Theorem 1.6.(1).

The above argument can be implemented in various ways depending what influence bound one uses. For example one obtains Theorem 1.6.(2) using the simple bound that any non-constant function on d bits has a variable with influence $\geq 1/2^d$.

Finally, we mention that a decomposition of local sources into bit-block ones also appears in [Vio10]. However that decomposition is tailored for a different type of results, and is incomparable with the present decomposition. In particular, it is still an open problem if the negative results in [Vio10] about generating the uniform distribution over n-bit strings with hamming weight αn can be improved to handle $d \geq \lg n$ or larger statistical distance.

Handling AC^0 sources. To handle AC^0 sources we use the previous extractor for local sources in combination with random restrictions [FSS84, Ajt83, Yao85, Hås87]. The switching lemma [Hås87] guarantees that after fixing all but a small fraction q of the input bits to constant, the AC^0 source collapses to a source with small locality.

The problem we face is that the restriction may have destroyed the min-entropy of the source. We show that, in fact, with high probability a random restriction that leaves free a

q fraction of the input variables decreases the entropy by a factor $\Omega(q)$ at most. This is the best possible up to constant factors as can be seen by considering the identity map.

Lemma 1.7 (Restrictions preserve min-entropy)**.** *Let $f : \{0,1\}^\ell \to \{0,1\}^n$ be a function such that $H_\infty(f(X)) = k$. Let ρ be a random restriction that independently sets variables to $\star$, 1, and 0 with probabilities q, $(1-q)/2$, and $(1-q)/2$. For every $\epsilon > 0$:*

$$\Pr_\rho \left[H_\infty(f_\rho(X)) \geq kq/4 - \lg(1/\epsilon)/2 \right] \geq 1 - \epsilon.$$

Note the only restriction this lemma puts on f is entropy.

The proof of this lemma builds on [LV11]. Specifically, we use an isoperimetric inequality for noise, that was highlighted in that work, to bound the collision probability of the restriction of f. The lemma then follows from the fact that the logarithm of the collision probability equals min-entropy up to constant factors.

Putting everything together, we arrive at the following result stating that any high-entropy AC^0 map is close to a convex combination of high-entropy bit-block sources.

Corollary 1.8 (AC^0 is convex combo of bit-block)**.** *For every $d = O(1), \gamma > 0$:*

Any n-bit distribution generated by an AC^0 circuit of depth d and size n^d is $1/n^{\omega(1)}$-close to a convex combination of affine sources with entropy $k(k/n^{1+\gamma})$. (In fact, bit-block sources with block-size $n^{1+\gamma}/k$.)

Intuition behind the proof of Theorem 1.3. We now explain how we prove that the majority function extracts one bit with error $o(1)$ from sources with locality $d = O(1)$ and min-entropy $k = n - o(\sqrt{n})$.

First, we use an information-theoretic argument from [Raz98, EIRS01, SV10] to argue that for all but $o(\sqrt{n})$ bits in the output, any w bits are close (error $o(1)$) to being uniform, for some $w = \omega(1)$.

Then we use the following key idea: since the distribution is local, if w bits are *close* to being uniform, then they are *exactly* uniform. This is because the granularity of the probability mass of those w bits is 2^{-wd}, which we can set to be larger than the error.

Hence, we have a distribution where all but $o(\sqrt{n})$ bits are w-wise independent. We show how to extract from such distributions using the bounded-independence central limit theorem [DGJ+10]. Indeed, this theorem guarantees that the sum of the w-wise independent bits behaves like a binomial distribution, up to some error. In particular, it has standard deviation $\Omega(\sqrt{n})$, and so the $o(\sqrt{n})$ "bad" bits over which we do not have control are unlikely to be able to influence the value of majority, which will be roughly unbiased.

Concurrent work. De and Watson [DW11] independently obtain extractors for sources with locality $d = o(\lg n)$. Their result is similar to our Theorem 1.1.(2). Their proof also uses Rao's extractor, but is otherwise different.

Organization. In §2 we prove Theorem 1.6 that local sources are a convex combination of bit-block sources, and then obtain extractors for local sources, proving Theorem 1.1. We also discuss various ways to optimize the parameters. In §3 we prove our Lemma 1.7 bounding the entropy loss when applying a restriction, and obtain our extractor for AC^0 sources, proving Theorem 1.2. We also prove Theorem 1.4, the negative result for generating $(Y, b(Y))$ in AC^0. In §4 we obtain the simpler extractor, proving Theorem 1.3. Finally, in §5 we conclude and discuss open problems.

2 From local to bit-block

In this section we prove Theorem 1.6, restated next, that local sources are a convex combination of bit-block sources. We then use this to obtain extractors for local sources, proving Theorem 1.1. We then discuss various ways to improve the parameters.

Theorem 1.6 (Local is convex combo of bit-block)**.** *Let $f : \{0,1\}^\ell \to \{0,1\}^n$ be a d-local function such that for uniform $X \in \{0,1\}^\ell$, $f(X)$ has min-entropy $\geq k$.*

Then, letting $s := 10^{-5}k(k/n)^2 \lg(d)/\lg(4n/k)d^3 = \tilde{\Omega}(k^3/n^2d^3)$, we have that $f(X)$ is 2^{-s}-close to a convex combination of bit-block sources with entropy $k' = s$ and block-size $\leq 2dn/k$.

We start with some preliminaries for the proof. First we need a few basic results regarding entropy, both the shannon entropy H and the min-entropy H_∞.

Claim 2.1. *For any $x, y \in (0, 1/2]$, if $H(x) \geq y$ then $x \geq 0.06y/\lg(1/y)$.*

Proof. Since $y \leq 1/2$ we have $y/\lg(1/y) \leq 1/2$. Hence if $x \geq 0.03$ then we are done. Otherwise, it can be verified that $2\lg(1/x) \leq 1/x^{2/3}$. It also holds for any $x \leq 1/2$ that $H(x) \leq 2x\lg(1/x)$, and so our assumption implies $2x\lg(1/x) \geq y$. Combining these two facts we get that $x^{1/3} \geq y$, and so $2\lg(1/x) \leq 6\lg(1/y)$. Using again that $x \geq y/2\lg(1/x)$, we get $x \geq y/6\lg(1/y) \geq 0.06y/\lg(1/y)$. □

Claim 2.2. *For any distribution X, $H(X) \geq H_\infty(X)$.*

Proof. Immediate from the definition. □

Claim 2.3. *Let $f : \{0,1\}^\ell \to \{0,1\}^n$ be a function, and let X be uniform over $\{0,1\}^\ell$. Let X' be a distribution over $\{0,1\}^\ell$ where s bits are constant and the other $\ell - s$ are uniform and independent. Then $H_\infty(f(X')) \geq H_\infty(f(X)) - s$.*

Proof. The claim holds because, for every a, $\Pr[f(X) = a] \geq 2^{-s} \cdot \Pr[f'(X) = a]$. □

Then we need the notion of *influence* of a variable x_i on a boolean function $f : \{0,1\}^d \to \{0,1\}$. Recall this is the probability over the choice of a uniform input $X \in \{0,1\}^d$ that flipping the value of the variable x_i changes the output of f. Kahn, Kalai, and Linial show that a near-balanced f has a variable with influence $\Omega(\lg(d)/d)$, improving on the lower bound $\Omega(1/d)$ which follows from the edge isoperimetric inequality over the hypercube [Har64, Har76]. This improvement is not essential to our results; it makes the final bound a bit better, and cleaner in the case $d = n^{\Omega(1)}$.

Lemma 2.4 ([KKL88]). *Let $f : \{0,1\}^d \to \{0,1\}$ be a function that equals one (or zero) with probability $p \leq 1/2$. Then there is a variable with influence at least $0.2p \lg(d)/d$.*

2.1 Proof of Theorem 1.6

Consider the bipartite graph with input side $\{x_1, \ldots, x_\ell\}$ and output side $\{y_1, \ldots, y_n\}$, where each variable y_i is connected to input variables it depends on. We call the x_i and y_i both nodes and variables. By assumption each variable y_i has degree $\leq d$. For a node v we denote by $N(v)$ the neighborhood of v; for a set V we denote by $N(V)$ the union of the neighborhoods of the nodes in V. In particular $N(N(v))$ is the two-step neighborhood of v.

Let

$$r := k/n \in [0,1].$$

Note that there are $\leq k/2$ input nodes with degree $> 2dn/k = 2d/r$, for else we would have $> dn$ edges.

Now we devise an algorithm to generate $f(X)$, and then we analyze it to prove the theorem. The algorithm works in stages. At stage i we work with a function $f_i : \{0,1\}^{L_i} \times \{0,1\}^{W_i} \to \{0,1\}^n$. These functions are obtained from f by fixing more and more input variables. L_i and W_i are disjoint sets of input variables, and we use the notation $\{0,1\}^{L_i}$ instead of $\{0,1\}^{|L_i|}$ to maintain the variable names throughout the algorithm. The sets W_i and L_i will satisfy the invariant $N(L_i) \bigcap N(W_i) = \emptyset$.

Algorithm

1. Set $L_0 := \emptyset, W_0 := \{x_1, \ldots, x_\ell\}$.
2. Let $f_0 : \{0,1\}^{L_0} \times \{0,1\}^{W_0} \to \{0,1\}^n$ be the function obtained from f by setting uniformly at random the $\leq k/2$ input variables with degree $> 2dn/k$. Remove those variables from W_0.
3. For stage $i = 0$ to $t-1$, where $t := \frac{k}{4}\frac{r}{2d^2}$:
   (a) Pick a variable $x \in W_i$ with maximal influence on f_i, that is, the probability over uniform input to f_i that flipping the value of x changes the output;
   (b) Let $L_{i+1} := L_i \bigcup \{x\}$; $W_{i+1} := W \setminus N(N(x))$; let $f_{i+1} : \{0,1\}^{L_{i+1}} \times \{0,1\}^{W_{i+1}} \to \{0,1\}^n$ be the function obtained from f_i by setting uniformly at random the input variables in $N(N(x)) \setminus \{x\}$.
4. Set uniformly at random the variables in W_t. Let $f' : \{0,1\}^{L_t} \to \{0,1\}^n$ be the resulting function
5. Set uniformly at random the variables in L_t and output the value of f'.

First, note that the algorithm generates the same distribution as $f(X)$ for uniform $X \in \{0,1\}^\ell$, because each bit is set uniformly at random independently of the others.

Note that throughout the execution of the algorithm, the invariant $N(L_i) \bigcap N(W_i) = \emptyset$ is maintained. This is because when we move a variable x into L_i at Step (3b) we remove $N(N(x))$ from W_i.

Claim 2.5. *At every stage $i < t$, the variable x picked at Step* (3a) *has influence $q \geq 0.003r \lg(d)/\lg(4/r)d$.*

Proof. First note $H_\infty(f_0) \geq k/2$ (cf. Claim 2.3). Let $i < t$ be an arbitrary stage.

Write $f_i(X)|_A$ for the restriction of $f_i(X)$ to the output variables in set A.

At every stage we set $< 2d(d-1)/r$ variables. So

$$H_\infty(f_i(X)) = H_\infty(f_i(X)|_{N(L_i)}) + H_\infty(f_i(X)|_{N(W_i)}) \geq k/2 - t2d(d-1)/r.$$

Since $H_\infty(f_i(X)|_{N(L_i)}) \leq |L_i| \leq t$, we obtain

$$H_\infty(f_i(X)|_{N(W_i)}) \geq k/2 - t2d(d-1)/r - t \geq k/2 - t2d^2/r \geq k/4,$$

by our choice of $t = \frac{k}{4}\frac{r}{2d^2}$.

Let $p \in [0, 1/2]$ be the number such that the maximum Shannon entropy of an output variable $y \in N(W_i)$ is $H(p)$. Bounding min-entropy from above by Shannon entropy, and using the sub-additivity of Shannon entropy, we see

$$k/4 \leq H_\infty(f_i(X)|_{N(W_i)}) \leq H(f_i(X)|_{N(W_i)}) \leq |N(W_i)|H(p) \leq nH(p). \tag{1}$$

Hence,

$$H(p) \geq \frac{k}{4n}.$$

By Claim 2.1, we get

$$p \geq 0.06\frac{k}{4n\lg(4n/k)}.$$

Now let $y \in N(W_i)$ be a variable such that $\Pr[y = b] = p$ for some $b \in \{0,1\}$.

By Lemma 2.4 there is an input variable x that has influence at least $0.2p\lg(d)/d \geq 0.003r\lg(d)/\lg(4/r)d$ on y. Note $x \in W_i$, since $y \in N(W_i)$. This concludes the proof of the claim. □

Claim 2.6. *The source $f'(X)$ at Step 5 is a bit-block source with block-size $\leq 2d/r$. Its blocks are $B_0 = \overline{N(L_t)}$, and for $L_t = \{x_{j_1}, \ldots, x_{j_t}\}$ and $h \geq 1$, $B_h = N(x_{j_h})$.*

Proof. By the invariant that $N(L_i) \bigcap N(W_i) = \emptyset$, and since L_{i+1} is obtained by moving a variable from W_i to L_i, we have that for any $x, x' \in L_t$, $N(x) \bigcap N(x') = \emptyset$. Hence the neighborhoods of variables in L_t form a partition of $N(L_t)$. Each set in the partition has size at most $\leq 2d/r$ by the bound on the degree of each input variable. □

It remains to bound the entropy of f'. Say that stage i is *good* if, letting x be the variable picked at Step 5, after the choice for the variables in $N(N(x)) \setminus \{x\}$, the output variables in $N(x)$ take two distinct values over the choice of x. Note that if the latter is the case then it is also the case for f' in Step 5, because after we set the variables in $N(N(x)) \setminus \{x\}$, the output variables in $N(x)$ depend only on x, and x is not set until Step 5.

Hence, the entropy of the bit-block source f' is the number of good stages. By Claim 2.5 each stage is good with probability $q \geq 0.003r \lg(d)/\lg(4/r)d$. Note that although the stages are not independent, the claim guarantees that each stage is good with probability $\geq q$ regardless of the outcomes of the previous stages. This is sufficient to apply the standard chernoff bound.

For example, one can use a bound by Panconesi and Srinivasan [PS97], with a compact proof by Impagliazzo and Kabanets [IK10]. Letting Z_i be the indicator variable of a stage being bad, the claim guarantees that for any $S \subseteq [t]$, $\Pr[\wedge_{i \in S} Z_i = 1] \leq (1-q)^{|S|}$. Theorem 1.1 in [IK10] implies that the probability of having more than $t(1-q/2)$ bad stages is at most

$$2^{-tD(1-q/2||1-q)} = 2^{-tD(q/2||q)} \leq 2^{-tq/5},$$

where D denotes relative entropy with logarithms in base 2, and the inequality can be verified numerically.

Hence we have $\geq tq/2$ good stages, except with probability $2^{-tq/5}$. Noting that

$$tq = 0.000375 \cdot kr^2 \lg(d)/\lg(4/r)d^3$$

concludes the proof. □

We now discuss how to improve the parameters in Theorem 1.6 in special cases.

Small locality. When the locality d is small, it is beneficial to use the following simple bound on influence: any non-constant function $f : \{0,1\}^d \to \{0,1\}$ has a variable with influence $\geq 2/2^d$. Using this, the bound in Claim 2.5 can be replaced with $q := 2/2^d$. (In the proof of Claim 2.5, after we guarantee $H_\infty(f_i(X)|_{N(W_i)}) \geq k/4 > 0$ we know there is a non-constant output variable.) Following the proof of Theorem 1.6, this guarantees $tq/2 \geq \Omega(k(k/n)/d^2 2^d)$ good stages except with probability $2^{-tq/5}$.

Large locality but small-depth decision tree. If the locality is large, but we have the additional guarantee that each output bit of the source is a decision tree of depth d' (e.g., $d' = \lg d$), then we can use the fact that every decision tree has an influential variable [OSSS05] (cf. [Lee10]). This replaces a factor $(\lg d)/d$ with $\Omega(1/d')$, guaranteeing $tq/2 = \Omega(k(k/n)/\lg(4n/k)d^2 d')$.

This improvement using [OSSS05] actually gives hope for a more dramatic improvement on Theorem 1.6; see §5.

2.2 Extractor for local sources

In this section we complete the proof of the extractor for local sources, restated next.

Theorem 1.1 (Extractor for local sources)**.** *For some $\rho > 0$, any $d = d(n), k = k(n)$:*

There is an explicit function $\text{EXT} : \{0,1\}^n \to \{0,1\}^m$ *that extracts m bits with error $\epsilon \leq 2^{-m^{\Omega(1)}}$ from any d-local source with min-entropy k, provided $2dn/k < m^\rho$, for:*

(1) $m = \Omega(k(k/n)^2 \lg(d)/\lg(4n/k)d^3) = \tilde{\Omega}(k(k/n)^2 d^3)$, or

(2) $m = \Omega(k(k/n)/d^2 2^d)$.

As we mentioned, we obtain it using Rao's extractor for low-weight affine sources. We now define these sources, observe that bit-block sources are a special case of them, and finally state Rao's extractor and prove Theorem 1.1.

Definition 2.7 ([Rao09])**.** *A distribution (or source) Y over $\{0,1\}^n$ is n-bit affine with entropy k and weight w if there are k linearly independent vectors $b_1, b_2, \ldots, b_k$ each of hamming weight $\leq w$, and a vector a_0, such that Y can be generated by choosing uniform $X \in \{0,1\}^k$ and outputting $a_0 + \sum_i X_i b_i \in \{0,1\}^n$.*

Remark 2.8. *A bit-block source with min-entropy k and block-size w (cf. Definition 1.5) is affine with entropy k and weight w (with the additional restrictions that the vectors b_i in Definition 2.7 have disjoint support). Indeed, one can define each vector b_i as 0 except for $b_i|_{B_i} := f_i(1) - f_i(0)$, and vector a_0 as $a_0|_{B_0} = b_0$, $a_0|_{B_i} = f_i(0)$.*

Theorem 2.9 ([Rao09], Theorem 1.3)**.** *There exist constants c, ρ such that for every $k(n) > \lg^c n$ there is an explicit extractor* $\text{EXT} : \{0,1\}^n \to \{0,1\}^{k(1-o(1))}$ *with error $1/2^{k^{\Omega(1)}}$ for affine sources with weight $w < k^\rho$ and min-entropy k.*

By Remark 2.8, Theorem 2.9 applies as stated to bit-block sources of entropy k and weight $w < k^\rho$.

Proof of Theorem 1.1. Theorem 1.6 guarantees that any d-local source with min-entropy k is 2^{-s} close to a convex combination of bit-block sources with entropy s and block-size $\leq 2dn/k$, where $s = \Omega(k(k/n)^2 \lg(d)/\lg(4n/k)d^3)$.

For a sufficiently small $\rho > 0$, Rao's extractor (Theorem 2.9) extracts $m := s/2$ bits with error $1/2^{s^{\Omega(1)}}$ from any such bit-block source, as long as $2dn/k \leq s^\rho \Leftarrow 2dn/k \leq m^\rho$. Thus the overall error is $\leq 1/2^{s^{\Omega(1)}} + 2^{-s} = 1/2^{m^{\Omega(1)}}$.

This proves Theorem 1.1.(1).

To prove Theorem 1.1.(2) we reason similarly, using the improvement on Theorem 1.6 for small locality discussed in the paragraph "Small locality" above. □

2.3 Optimizing Rao's extractor

In this section we sketch how to optimize Rao's extractor (Theorem 2.9) to obtain $\rho = 1 - \epsilon$ for any fixed $\epsilon > 0$. This improvement can be obtained using the same steps as in Rao's proof, but optimizing a few results used there. We are grateful to Rao for his help with the material in this section.

First, Rao uses a parity check function $P : \{0,1\}^n \to \{0,1\}^t$ for a code of distance wk^α, with output length $t = O(w^2 k^{2\alpha} \lg^2 n)$. The parameter w corresponds to the weight (or

block-size) of the source, and the squaring turns out to be problematic to the optimization. However using better codes (e.g., [ABN+92]) one can make $t = O(wk^\alpha \lg n)$.

Second, Rao uses the strong, linear, seeded extractor obtained in [RRV02] building on Trevisan's extractor [Tre01]. The dependence on n in the seed length of this extractors if $O(\lg n)$, and for the current improvement it is important to reduce to one the constant hidden in $O(.)$. This for example can be achieved using an extractor in [GUV09] to condense entropy before applying [RRV02].

Finally, one needs to observe that Theorem 3.1 in [Rao09] although being stated for fixed constants 0.7 and 0.9, can actually be obtained for constants arbitrarily close to 1.

3 From AC^0 to local

In this section we obtain extractors for AC^0 sources, proving Theorem 1.2, restated next. Then we prove the negative result for generating $(Y, b(Y))$ in AC^0, Theorem 1.4.

Theorem 1.2 (Extractor for AC^0 sources)**.** *For some $\rho > 0$, any $\gamma > 0, d = O(1), k = k(n)$:*

There is an explicit extractor $\textsc{Ext} : \{0,1\}^n \to \{0,1\}^m$ with output length $m = k(k/n^{1+\gamma})$ and error $1/n^{\omega(1)}$ for sources with min-entropy k that are generated by AC^0 circuits $C : \{0,1\}^{n^d} \to \{0,1\}^n$ of depth d and size n^d, provided $n^{1+\gamma}/k < m^\rho$.

To prove this theorem we bound the entropy loss associated to random restrictions, and then recall the switching lemma.

The effect of restrictions on min-entropy. Recall that a restriction ρ on n variables is a map $\rho : [n] \to \{0, 1, \star\}$. We denote by f_ρ the function obtained from f by applying the restriction.

We now state and prove a lemma that bounds the entropy loss incurred when applying a random restriction to a function.

Lemma 1.7 (Restrictions preserve min-entropy)**.** *Let $f : \{0,1\}^\ell \to \{0,1\}^n$ be a function such that $H_\infty(f(X)) = k$. Let ρ be a random restriction that independently sets variables to $\star$, 1, and 0 with probabilities q, $(1-q)/2$, and $(1-q)/2$. For every $\epsilon > 0$:*

$$\Pr_\rho \left[H_\infty(f_\rho(X)) \geq kq/4 - \lg(1/\epsilon)/2 \right] \geq 1 - \epsilon.$$

The proof of this lemma relies on the following isoperimetric inequality for noise, see [LV11] for a proof.

Lemma 3.1. *Let $A \subseteq \{0,1\}^\ell$ and $\alpha := |A|/2^\ell$. For any $0 \leq p \leq 1/2$, let E be a noise vector of i.i.d. bits with $\Pr[1] = p$; let X be uniform in $\{0,1\}^\ell$:*

$$\alpha^2 \leq \Pr_{X,E}[X \in A \wedge X + E \in A] \leq \alpha^{1/(1-p)} \leq \alpha^{1+p}.$$

Proof of Lemma 1.7. The idea is to bound $H_\infty(f_\rho(X))$ using the collision probability $\Pr_{X,Y}[f_\rho(X) = f_\rho(Y)]$ of f_ρ, which in turn can be analyzed via Lemma 3.1.

Specifically, note that the joint distribution $(f_\rho(X), f_\rho(Y))$ where ρ is a random restriction with parameter q as in the statement of the lemma, and X and Y are uniform and independent, is the same as the joint distribution $(f(X), f(X+E))$ where X is uniform and E is noise vector where each bit is set to 1 independently with probability $p := q/2 \leq 1/2$.

For any $a \in \{0,1\}^n$, let $A_a := f^{-1}(a)$; also denote by X_ρ the result of applying the restriction ρ to X. By Lemma 3.1:

$$\begin{aligned}\Pr_{\rho,X,Y}[f_\rho(X) = f_\rho(Y)] &= \sum_a \Pr_{\rho,X,Y}[X_\rho \in A_a \wedge Y_\rho \in A_a] \leq \sum_a (|A_a|/2^\ell)^{1+p} \\ &\leq \max_a (|A_a|/2^\ell)^p \leq 2^{-pk},\end{aligned}$$

where the last inequality is the assumption that $H_\infty(f(X)) \geq k$.

And so

$$\Pr_\rho \left[\Pr_{X,Y}[f_\rho(X) = f_\rho(Y)] \leq 2^{-pk}/\epsilon \right] \geq 1 - \epsilon.$$

To conclude, note that for any ρ

$$\max_a \left(\Pr_X[f_\rho(X) = a] \right)^2 \leq \Pr_{X,Y}[f_\rho(X) = f_\rho(Y)],$$

and so with probability $\geq 1 - \epsilon$ over ρ we have

$$\max_a \left(\Pr_X[f_\rho(X) = a] \right)^2 \leq 2^{-pk}/\epsilon \Rightarrow H_\infty(f_\rho(X)) \geq pk/2 - \lg(1/\epsilon)/2.$$

□

The switching lemma. We also need to collapse an AC^0 source to a local source, which can be accomplished via the following standard corollary to the switching lemma [Hås87].

Lemma 3.2. *Let $f : \{0,1\}^\ell \to \{0,1\}$ be a function computable by a depth-d AC^0 circuit with s gates. Let ρ be a random restriction with $\Pr[\star] = q < 1/9^d$. The probability over ρ that f_ρ cannot be written as a decision tree of depth t is $\leq s(9q^{1/d}t)^t$.*

This lemma can be proved using [Tha09, Lemma 1] (cf. [Bea94]). The restriction is seen as the successive application of d restrictions with $\Pr[\star] = q^{1/d}$.

We can now prove Theorem 1.2.

Proof of Theorem 1.2. Let t be a slowly growing function, such as $t = \lg o(\lg n)$. Let Ext be the extractor in Theorem 1.1 for locality t and min-entropy $0.1k/n^\gamma$.

By Lemma 3.2 a random restriction with $\Pr[\star] = 1/n^\gamma$ will collapse all n circuits (computing the n output bits) to decision trees of depth t – in particular, 2^t-local functions – except for an error $1/n^{\omega(1)}$. (Here we use that $t = \omega(1)$.)

By Lemma 1.7, except for an error $1/n^{\lg n}$, the restricted source has min-entropy

$$k' \geq 0.25k/n^{\gamma} - \lg^2 n \geq 0.1k/n^{\gamma}.$$

The theorem now follows from Theorem 1.1.(2). The theorem extracts $m = \Omega(k'(k'/n)/2^{O(2^t)}) \geq \Omega(k(k/n^{1+3\gamma}))$ bits (since $2^{O(2^t)} \leq n^{\gamma}$), provided $m^{\rho} > 2n2^t/k'$ which is implied by $m^{\rho} > n^{1+2\gamma}/k$. The error is dominated by the error incurred by the restriction step, which is $1/n^{\omega(1)}$. □

Appealing to Theorem 1.1.(1) instead allows to improve the error from $1/n^{\omega(1)}$ to $1/n^{\Omega(\lg n)}$, at the price of requiring larger k.

Finally, we mention that Corollary 1.8, claiming that any high-entropy AC^0 distribution is close to a convex combination of high-entropy bit-block sources, can be proved along the same lines. Namely we can generate the distribution by first selecting a random restriction, and then the rest, and invoke Lemmas 3.2 and Lemma 1.7 and Theorem 1.6 (the "Small locality" version, see §2).

3.1 Negative result for generating $(Y, b(Y))$

We now prove Theorem 1.4.

Theorem 1.4. *There is an explicit map $b : \{0,1\}^* \to \{0,1\}$ such that for every $d = O(1)$:*

Let $C : \{0,1\}^{n^d} \to \{0,1\}^{n+1}$ be an AC^0 circuit of size n^d and depth d. The distribution $C(X)$ for uniform X has statistical distance $\geq 1/2^{n^{1-\Omega(1)}}$ from the distribution $(Y, b(Y))$ for uniform $Y \in \{0,1\}^n$.

For b one can take the first bit of the extractor in Theorem 1.2 for $k = n^{1-\Omega(1)}$.

Proof of Theorem 1.4. Define b to be the first output bit of the extractor in Theorem 1.2 for n-bit distributions of some min-entropy $k = n^{1-\Omega(1)}$ generated by circuits of size n^{d+a} and depth $d + a$ for a universal constant a to be set later.

Assume towards a contradiction that there is a circuit $C(X) = (Y, Z) \in \{0,1\}^n \times \{0,1\}$ as in the theorem such that the relevant statistical distance is $\leq 1/2^{n^{\delta}}$. Then for every a, $\Pr[C(X) = a] \leq 1/2^n + 1/2^{n^{\delta}} \leq 2/2^{n^{\delta}}$. So $H_{\infty}(C(X)) \geq n^{\delta} - 1$. Note that on uniform input U, $b(U) = 1$ with probability $p = 1/2 \pm o(1)$, and so $Z = 1$ also with probability $p' = 1/2 \pm o(1)$.

Consider the circuit C' that runs $C(X)$ to generate (Y, Z), and then if $Z = 1$ it outputs Y, otherwise outputs a uniform n-bit string. For a suitable choice of a, C' is implementable in size n^{d+a} and depth $d + a$.

Note that the min-entropy of $C'(X)$ is $\geq n^{\delta} - O(1)$, and that $b(C'(X)) = 1$ with probability $p' + (1 - p')p = 1/2 + \Omega(1)$. For a large enough $\delta < 1$, this contradicts Theorem 1.2. □

To get a lower bound of ϵ on the statistical distance, the above proof needs an extractor for min-entropy $\lg(1/\epsilon) - O(1)$. This prevents us from obtaining bounds such as $\epsilon = 1/2 - o(1)$. Obtaining such bounds for AC^0 seems an interesting direction.

4 A worse, simpler extractor

In this section we prove Theorem 1.3, restated next.

Theorem 1.3. *There is a symmetric, explicit, deterministic extractor* $\textsc{Ext} : \{0,1\}^n \to \{0,1\}^m$ *that extracts* $m = \Omega(\lg\lg n - \lg d)$ *bits with error* $\epsilon = (d/\lg n)^{\Omega(1)}$ *from any* n*-bit source with shannon entropy* $k \geq n - n^{0.49}$ *whose bits are each computable by a decision tree of depth* d*. To extract* $m = 1$ *bit, one can take* $\textsc{Ext} :=$ majority*.*

The proof combines several lemmas discussed next.

Lemma 4.1. *[Raz98, EIRS01, SV10] Let* $V = (V_1, \ldots, V_n)$ *be a random variable over* $\{0,1\}^n$ *such that* $H(V) \geq n - a$*. Then for any* $\epsilon > 0$ *and integer* q *there exists a set* $G \subseteq [n]$ *such that* $|G| \geq n - 16 \cdot q \cdot a/\epsilon^2$*, and for any distinct* $i_1, \ldots, i_q \in G$ *the distribution* $(V_{i_1}, \ldots, V_{i_q})$ *is* ϵ*-close to uniform.*

Proof sketch. By the chain rule, $H(V_1) + H(V_2|V_1) + \cdots + H(V_n|V_1 \ldots V_{n-1}) = H(V) \geq n - a$. Picking i uniformly in $[n]$, we see $E[1 - H(V_i|V_1 \ldots V_{i-1})] \leq a/n$. Let $b := \epsilon^2 n/(16q \cdot a)$. By Markov's inequality, $\Pr_i[1 - H(V_i|V_1 \ldots V_{i-1}) \geq b \cdot a/n] \leq 1/b = 16q \cdot a/(\epsilon^2 n)$.

That means there are at most $16q \cdot a/\epsilon^2$ "bad" values for i for which $H(V_i|V_1 \ldots V_{i-1}) \leq 1 - b \cdot a/n = \epsilon^2/16q$. Hence for any q "good" values for i, the entropy of the joint distribution of the corresponding variables is at least $q(1 - \epsilon^2/16q) = q - \epsilon^2/16$, which implies that the joint distribution is ϵ-close to uniform. □

Lemma 4.2 (Bounded independence central limit theorem [DGJ+10])**.** *There is* $C > 0$ *such that the following holds for every* n, ϵ*, and* $q \geq C \lg^2(1/\epsilon)/\epsilon^2$*:*

Let $U = (U_1, \ldots, U_n)$ *be the uniform distribution over* $\{0,1\}^n$*, and let* $X = (X_1, \ldots, X_n)$ *be any* q*-wise independent distribution over* $\{0,1\}^n$*. Then for any* $t \geq 0$*:*

$$\left| \Pr\left[\sum_i U_i \geq t\right] - \Pr\left[\sum_i X_i \geq t\right] \right| \leq \epsilon.$$

In particular, the classical central limit theorem for the sum of independent Bernoulli trials holds for q-wise independent trials up to error ϵ.

Claim 4.3. *Let* $f : \{0,1\}^n \to \{0,1\}^q$ *be a function such that each output bit is computed by a depth-*d *decision tree. Then for any event* $A \subseteq \{0,1\}^q$*, the probability that* $f(X) \in A$ *for a uniform* X *in* $\{0,1\}^n$ *equals* $a/2^{qd}$ *for some integer* a*.*

Proof. We can compute the whole q-bit output of the function by a decision tree of depth $q \cdot d$ whose leaves are labeled with q-bit strings. The decision tree simply simulates in turn the q decision trees of the q output bits of f. Since the events over X of outputting different leaves are disjoint, and each event has probability 2^{-qd}, the result follows. □

A denominator larger than 2^{-qd} would reflect in improved parameters for the extractor. But the following claim shows that no improvement is possible, even for the case in which each output bit is d-local.

Claim 4.4. *For any $d \geq q$ and $n \geq d(q-1)$ there is a d-local function $f : \{0,1\}^n \to \{0,1\}^q$ whose output is not uniform but is $1/2^{d(q-3)}$ close to uniform.*

Proof. Let $D := 2^d$. For simplicity we think of the output of f as $\{-1,1\}^q$ instead of $\{0,1\}^q$. Consider the function $g : \{0,1\}^d \to \{-1,1\}$ that is 1 on $D/4+1$ of the $D/2$ inputs where the last bit is 1 and on $D/4-1$ of the $D/2$ inputs where the last bit is 0. Notice $\Pr_Y[g(Y) = 1] = 1/2$. If Y_d is the last bit of Y, notice $\Pr_Y[Y_d = 1 | g(Y) = \pm 1] = (D/4 \pm 1)/(D/2) = 1/2 \pm 2/D$.

Now let $f_1, \ldots, f_{q-1}$ be the function g applied to disjoint d-bit inputs $Y^1, \ldots, Y^{q-1}$ (which is possible since $n \geq d(q-1)$); and let f_q equal to the product (XOR over $\{0,1\}$) of the last bits of $Y^1, \ldots, Y^{q-1}$ (which is possible since $d \geq q-1$).

Now we show that for any $a \in \{-1,1\}^q$, $|\Pr[f(X) = a] - 1/2^q| \leq 2^{q-1}/2^{d(q-1)}$. From this the claim about statistical distance follows by a union bound over all $\{0,1\}^q$ values a, which yields a bound $2^{2q-1}/2^{d(q-1)} \leq 2^{2d-d(q-1)}$ since $d \geq q$.

Indeed, the probability that the first $q-1$ bits of f agree with a is equal to $1/2^{q-1}$ since these output bits are uniform and independent. Condition on this happening. By the previous observation, the last bits of $Y^1, \ldots, Y^{q-1}$ are independent bits with $\Pr[Y^i = 1] = 1/2 \pm 2/D$. Hence, $E[Y^i] = \pm 4/D$. By independence, $E[f_q(X)] = \pm(4/D)^{q-1}$, which means $\Pr[f_q(X) = 1] = 1/2 \pm (4/D)^{q-1}$. Therefore, $\Pr[f(X) = a] = (1/2^{q-1})(1/2 \pm (4/D)^{q-1}) = 1/2^q + (2/D)^{q-1}$. □

4.1 Proof of Theorem 1.3

Let the entropy be $n - n^{0.5-\gamma}$. Set

$$q := \alpha \left(\frac{\lg n}{d} \right),$$

for a sufficiently small α depending on γ. We are going to extract $\Omega(\lg q)$ bits.

Apply Lemma 4.1 with $\epsilon := 0.5/2^{dq}$. For a small enough α, this gives that, except for at most

$$O(n^{0.5-\gamma} q 2^{dq}) = n^{0.5-\gamma} 2^{O(\alpha \lg n)} = n^{0.5-\gamma/2}$$

"bad" variables, any q "good" variables have a joint distribution that is $\leq 0.5/2^{dq}$ close to uniform.

By Claim 4.3, the joint distribution of those q variables is exactly uniform.

To summarize, the output distribution is q-wise independent, except for $t := n^{0.5-\gamma/2}$ bits that, we are going to think, are arbitrarily correlated with the output.

We are going to show how to extract from such sources.

Let $X \in [0, n-t]$ be the hamming weight of the q-wise independent part, and $Y \in [0,t]$ be the hamming weight of the rest. Let B be the sum of $n-t$ i.i.d. coin tosses (the binomial distribution).

By Lemma 4.2, there is an absolute constant η such that for any interval $[i,j]$, $|\Pr[X \in [i,j]] - \Pr[B \in [i,j]]| \leq \beta := (1/q)^\eta$.

Now, for a $\delta \in (0,1)$ to be determined later, partition $[0, n-t]$ into $s = q^\delta$ intervals whose measure w.r.t. B is $1/s \pm O(1/\sqrt{n-t})$, which is possible because B takes any fixed

value with probability at most $O(1/\sqrt{n-t})$ and because $s \leq \lg n = (n-t)^{o(1)}$ (so the greedy approach of collecting intervals won't stop before collecting s).

Now we bound the probability that $X+Y$, the hamming weight of the source, lands in any fixed interval $[i,j]$:

$$\begin{aligned}
\Pr[X+Y \in [i,j]] &\geq \Pr[X \in [i,j]] - \Pr[X \in [j-t,j]] \qquad \text{(Since } Y \geq 0\text{)} \\
&\geq \Pr[B \in [i,j]] - \Pr[B \in [j-t,j]] - 2\beta \\
&\geq 1/s - O(1/\sqrt{n-t}) - O(t/\sqrt{n-t}) - 2\beta \\
&\geq 1/s - O(\beta).
\end{aligned}$$

Repeating the argument for the upper bound, we get $|\Pr[X+Y \in [i,j]] - 1/s| = O(\beta) = O(1/q^{\eta})$.

Since we took $s = q^{\delta}$ intervals, for a sufficiently small δ we get that the statistical distance between $X+Y$ and the uniform distribution over intervals is $1/q^{\Omega(1)}$.

Assuming w.l.o.g. that s is a power of 2, we have extracted $\lg s = \Omega(\lg q)$ bits at distance $1/q^{\Omega(1)}$ from uniform.

To show that majority extracts one bit, one uses the same approach but instead of dividing into buckets one more simply argues that $\Pr[X+Y > n/2] = 1/2 \pm 1/q^{\Omega(1)}$.

5 Conclusion and open problems

Can one obtain a result like Theorem 1.1.(1) for depth-d decision trees? This would also allow to extract from AC^0 sources with error smaller than $1/n^{\lg n}$, a barrier for current techniques. The fact that every decision tree has an influential variable [OSSS05, Lee10] seems promising, but at the moment we are unable to carry through the proof in this case. On the other hand, the fact that depth $\lg n$ is sufficient for a decision tree to select a random variable from the input may also be used in a counterexample.

Can we extract from lower min-entropy in Theorems 1.1? Note one always needs $k > d$, since any distribution with min-entropy k can be obtained in a $d = k$ local fashion. So if d is polynomial then k must be polynomial as well. However for say $d = O(1)$ one may be able to handle $k = n^{o(1)}$.

Another question is whether we can extract with better parameters from an n-bit source where $n-t$ bits are k-wise independent. Say we want to extract one bit. We handled $t \approx \sqrt{n}$ in the proof of Theorem 1.3 using majority. If the $n-t$ bits were uniform, we could allow for greater entropy deficiency t by using Ben-Or and Linial's recursive-majority-of-3 function [BL90]. Can a similar improvement be obtained for bounded independence?

As a more general direction, we note that there are many other computational models besides AC^0 for which it will be important to derive extractors and the corresponding sampling lower bounds. As a starting point, one should derive such results for every model for which we currently have (classical) lower bounds, e.g. branching programs, turing machines, and polynomials. In fact, we view sampling lower bounds as a *third* type of lower bounds. The first type is the classical, worst-case one; the second is the average-case

one. Just like the second type gave substantial new information, in particular yielding new lower bounds of the first type (e.g. [HMP+93, HM04]) and new pseudorandom generators (e.g. [Nis91, INW94, Vio07]), we expect the third type to affirm itself as a central paradigm.

Acknowledgments. We are very grateful to Amir Shpilka for extensive discussions. We also thank Anup Rao for a discussion on [Rao09] which resulted in §2.3, the organizers of the 2011 Dagstuhl seminar on complexity theory for the opportunity to present these results in March 2011, and the anonymous referees for their feedback.

References

[ABN+92] Noga Alon, Jehoshua Bruck, Joseph Naor, Moni Naor, and Ron M. Roth. Construction of asymptotically good low-rate error-correcting codes through pseudo-random graphs. *IEEE Transactions on Information Theory*, 38(2):509–516, 1992.

[Ajt83] Miklós Ajtai. Σ^1 [1]-formulae on finite structures. *Ann. Pure Appl. Logic*, 24(1):1–48, 1983.

[Bea94] Paul Beame. A switching lemma primer. Technical Report UW-CSE-95-07-01, Department of Computer Science and Engineering, University of Washington, November 1994. Available from http://www.cs.washington.edu/homes/beame/.

[BKS+10] Boaz Barak, Guy Kindler, Ronen Shaltiel, Benny Sudakov, and Avi Wigderson. Simulating independence: New constructions of condensers, ramsey graphs, dispersers, and extractors. *J. ACM*, 57(4), 2010.

[BL90] Michael Ben-Or and Nathan Linial. Collective coin-flipping. In Silvio Micali, editor, *Randomness and Computation*, pages 91–115. Academic Press, New York, 1990.

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

[Bou07] Jean Bourgain. On the construction of affine extractors. *Geometric And Functional Analysis*, 17:33–57, 2007.

[BSK09] Eli Ben-Sasson and Swastik Kopparty. Affine dispersers from subspace polynomials. In *Symposium on the Theory of Computing (STOC)*, pages 65–74, 2009.

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

[CGH+85] Benny Chor, Oded Goldreich, Johan Håstad, Joel Friedman, Steven Rudich, and Roman Smolensky. The bit extraction problem and t-resilient functions. In *26th IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 396–407, 1985.

[DGJ+10] Ilias Diakonikolas, Parikshit Gopalan, Ragesh Jaiswal, Rocco A. Servedio, and Emanuele Viola. Bounded independence fools halfspaces. *SIAM J. on Computing*, 39(8):3441–3462, 2010.

[DW11] Anindya De and Thomas Watson. Extractors and lower bounds for locally samplable sources. In *Workshop on Randomization and Computation (RANDOM)*, 2011.

[EIRS01] Jeff Edmonds, Russell Impagliazzo, Steven Rudich, and Jiri Sgall. Communication complexity towards lower bounds on circuit depth. *Computational Complexity*, 10(3):210–246, 2001.

[FSS84] Merrick L. Furst, James B. Saxe, and Michael Sipser. Parity, circuits, and the polynomial-time hierarchy. *Mathematical Systems Theory*, 17(1):13–27, 1984.

[GRS06] Ariel Gabizon, Ran Raz, and Ronen Shaltiel. Deterministic extractors for bit-fixing sources by obtaining an independent seed. *SIAM J. Comput.*, 36(4):1072–1094, 2006.

[GUV09] Venkatesan Guruswami, Christopher Umans, and Salil P. Vadhan. Unbalanced expanders and randomness extractors from parvaresh–vardy codes. *J. ACM*, 56(4), 2009.

[Har64] L. H. Harper. Optimal assignments of numbers to vertices. *SIAM Journal on Applied Mathematics*, 12(1):131–135, 1964.

[Har76] Sergiu Hart. A note on the edges of the n-cube. *Discrete Mathematics*, 14(2):157–163, 1976.

[Hås87] Johan Håstad. *Computational limitations of small-depth circuits*. MIT Press, 1987.

[HM04] Kristoffer Arnsfelt Hansen and Peter Bro Miltersen. Some meet-in-the-middle circuit lower bounds. In *29th Symposium on Mathematical Foundations of Computer Science (MFCS)*, Lecture Notes in Computer Science, Volume 3153, pages 334 – 345, 2004.

[HMP+93] András Hajnal, Wolfgang Maass, Pavel Pudlák, Mário Szegedy, and György Turán. Threshold circuits of bounded depth. *J. Comput. System Sci.*, 46(2):129–154, 1993.

[IK10] Russell Impagliazzo and Valentine Kabanets. Constructive proofs of concentration bounds. In *Workshop on Randomization and Computation (RANDOM)*, pages 617–631. Springer, 2010.

[IN96] Russell Impagliazzo and Moni Naor. Efficient cryptographic schemes provably as secure as subset sum. *Journal of Cryptology*, 9(4):199–216, 1996.

[INW94] Russell Impagliazzo, Noam Nisan, and Avi Wigderson. Pseudorandomness for network algorithms. In *26th ACM Symposium on the Theory of Computing (STOC)*, pages 356–364, 1994.

[KKL88] Jeff Kahn, Gil Kalai, and Nathan Linial. The influence of variables on Boolean functions. In *29th Symposium on Foundations of Computer Science (FOCS)*, pages 68–80, 1988.

[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.

[KZ07] Jesse Kamp and David Zuckerman. Deterministic extractors for bit-fixing sources and exposure-resilient cryptography. *SIAM J. Comput.*, 36(5):1231–1247, 2007.

[Lee10] Homin K. Lee. Decision trees and influence: an inductive proof of the osss inequality. *Theory of Computing*, 6(1):81–84, 2010.

[Li11a] Xin Li. Improved constructions of three source extractors. In *Conference on Computational Complexity (CCC)*, 2011.

[Li11b] Xin Li. A new approach to affine extractors and dispersers. In *Conference on Computational Complexity (CCC)*, 2011.

[LV11] Shachar Lovett and Emanuele Viola. Bounded-depth circuits cannot sample good codes. In *Conference on Computational Complexity (CCC)*, 2011. Invited and submitted to special issue of Computational Complexity.

[Nis91] Noam Nisan. Pseudorandom bits for constant depth circuits. *Combinatorica*, 11(1):63–70, 1991.

[OSSS05] Ryan O'Donnell, Michael E. Saks, Oded Schramm, and Rocco A. Servedio. Every decision tree has an influential variable. In *Symposium on Foundations of Computer Science (FOCS)*, pages 31–39. IEEE, 2005.

[PS97] Alessandro Panconesi and Aravind Srinivasan. Randomized distributed edge coloring

via an extension of the chernoff-hoeffding bounds. *SIAM J. Comput.*, 26(2):350–368, 1997.

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

[Raz98] Ran Raz. A parallel repetition theorem. *SIAM J. Comput.*, 27(3):763–803, 1998.

[RRV02] Ran Raz, Omer Reingold, and Salil P. Vadhan. Extracting all the randomness and reducing the error in Trevisan's extractors. *J. Comput. Syst. Sci.*, 65(1):97–128, 2002.

[Sha11] Ronen Shaltiel. Dispersers for affine sources with sub-polynomial entropy. In *Symposium on Foundations of Computer Science (FOCS)*. IEEE, 2011.

[SV86] Miklos Santha and Umesh V. Vazirani. Generating quasi-random sequences from semi-random sources. *J. of Computer and System Sciences*, 33(1):75–87, August 1986.

[SV10] Ronen Shaltiel and Emanuele Viola. Hardness amplification proofs require majority. *SIAM J. on Computing*, 39(7):3122–3154, 2010.

[Tha09] Neil Thapen. Notes on switching lemmas. http://www.math.cas.cz/∼thapen/, 2009.

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

[TV00] Luca Trevisan and Salil Vadhan. Extracting randomness from samplable distributions. In *Symposium on Foundations of Computer Science (FOCS)*, pages 32–42, 2000.

[Vio05] Emanuele Viola. On constructing parallel pseudorandom generators from one-way functions. In *20th Conference on Computational Complexity (CCC)*, pages 183–197. IEEE, 2005.

[Vio07] Emanuele Viola. Pseudorandom bits for constant-depth circuits with few arbitrary symmetric gates. *SIAM J. on Computing*, 36(5):1387–1403, 2007.

[Vio10] Emanuele Viola. The complexity of distributions. In *51th Symposium on Foundations of Computer Science (FOCS)*, pages 202–211. IEEE, 2010. To appear in SIAM J. on Computing.

[vN51] John von Neumann. Various techniques used in connection with random digits. *National Bureau of Standards, Applied Mathematics Series*, 12:36–38, 1951.

[Yao85] Andrew Yao. Separating the polynomial-time hierarchy by oracles. In *26th Symposium on Foundations of Computer Science (FOCS)*, pages 1–10. IEEE, 1985.

[Yeh10] Amir Yehudayoff. Affine extractors over prime fields. Unpublished manuscript, 2010.