Electronic Colloquium on Computational Complexity, Revision 1 of Report No. 163 (2025)

Most Juntas Saturate the Hardcore Lemma

Vinayak M. Kumar*

Abstract

Consider a function that is mildly hard for size-s circuits. For sufficiently large s, Impagliazzo's hardcore lemma guarantees a constant-density subset of inputs on which the same function is extremely hard for circuits of size $s' \ll s$. Blanc, Hayderi, Koch, and Tan [FOCS 2024] recently showed that the degradation from s to s' in this lemma is quantitatively tight in certain parameter regimes. We give a simpler and more general proof of this result in almost all parameter regimes of interest by showing that a random junta witnesses the tightness of the hardcore lemma with high probability.

1 Introduction

Let $f : \{0,1\}^n \to \{0,1\}$ be a Boolean function such that every circuit of size s errs on at least a δ-fraction of inputs. How can we amplify the hardness of this function? One approach is to restrict the domain: given a fixed size-s circuit, we select a subset of inputs of density at least 2δ in which half the points come from the error region and half are correct. Such a set forms a *hardcore set*, because on this region the circuit cannot do better than random guessing. Is it possible that there exists a *single* subset of density 2δ that is simultaneously hard for all size-s circuits? Impagliazzo's hardcore lemma establishes the existence of a $\Omega(\delta)$-density hardcore set for *all* circuits of size $s' \ll s$. The version of this lemma with the smallest size degradation from s to s' is the following.

Theorem 1 ([Imp95, KS99, BHK09])**.** *Let $f : \{0,1\}^n \to \{0,1\}$ and $\delta, \gamma, n \leq s \leq \frac{2^n}{n}$. Suppose that for all circuits C of size at most s,*

$$\Pr_{x \sim \{0,1\}^n}[C(x) = f(x)] \leq 1 - \delta.$$

Then there exists a subset $H \subset \{0,1\}^n$ of density $\Omega(\delta)$ such that for all circuits C of size $O\left(\frac{s\gamma^2}{\log(1/\delta)}\right)$, we have

$$\Pr_{x \sim H}[C(x) = f(x)] \leq \frac{1}{2} + \gamma.$$

Conceptually, the theorem says that circuit hardness can be explained by a subset of "hard inputs" H on which the function looks random to small circuits.[1] This phenomenon has found

*vmkumar@utexas.edu. Department of Computer Science, The University of Texas at Austin. Supported in part by NSF Grant CCF-2312573, a Simons Investigator Award (#409864, David Zuckerman), a Jane Street Graduate Research Fellowship, and a UT Austin Dean's Prestigious Fellowship Supplement.

[1] Holenstein [Hol05] gives a set H of optimal density 2δ, but suffers a larger degradation from s to $O\left(\frac{s\gamma^2}{n}\right)$.

ISSN 1433-8092

applications throughout computer science, including hardness amplification [O'D02, Tre05], pseudorandomness [STV99, CL21], cryptography [Hol05], algorithmic fairness [CDV24], combinatorics [RTTV08], and learning theory [BHK09, KS99].

While the hardcore lemma is a remarkable result, a natural question is whether the size degradation $s \to \frac{s\gamma^2}{\log(1/\delta)}$ is necessary. This is formalized by the following conjecture.

Conjecture 1. *For any $\delta \in (0,1), \gamma \in (0, \frac{1}{2})$, $n \in \mathbb{N}$ large enough, and $\Omega\left(\frac{\log(1/\delta)}{\gamma^2}\right) \leq s \leq O\left(\frac{2^n}{n}\right)$, there exists an f such that*

- *for all circuits C of size $\leq s$,*
$$\Pr_{x \sim \{0,1\}^n}[C(x) = f(x)] \leq 1 - \delta.$$
- *for every subset $H \subset \{0,1\}^n$ of density $\geq \Omega(\delta)$, there exists a circuit C of size $O\left(\frac{s\gamma^2}{\log(1/\delta)}\right)$ such that*
$$\Pr_{x \sim H}[C(x) = f(x)] \geq \frac{1}{2} + \gamma.$$

Such a degradation was shown to be necessary for a certain class of proofs [LTW11], but an unconditional result remained elusive, as proving this theorem appears to require constructing an explicit, mildly approximable function that demands large circuits to strongly approximate it. This felt tantamount to proving breakthrough circuit lower bounds. In a recent work, Blanc, Hayderi, Koch, and Tan [BHKT24] evaded this barrier by arguing about a nonexplicit function. In fact, the proof of Theorem 1 first shows the result for $H \sim \{0,1\}^n$ being the weaker notion of a δ-smooth distribution (i.e., no string has more than $\frac{1}{\delta 2^n}$ probability of being sampled), and then uses the distribution to extract a density-δ subset. [BHKT24] proves a tightness result over the more general δ-smooth distributions.

Theorem 2 ([BHKT24, Theorem 2])**.** *Let $\delta \in (0,1), \gamma \in \left[\Omega\left(\frac{1}{\sqrt{n}}\right), \frac{1}{2}\right)$, $n \in \mathbb{N}$ large enough. For $s \in [\Omega(\frac{1}{\gamma^2}), O(\frac{2^{\gamma^2 n}}{\gamma^4 n})]$, there exists $f : \{0,1\}^n \to \{0,1\}$ such that*

- *For all circuits C of size $\leq s$,*
$$\Pr_{x \sim \{0,1\}^n}[C(x) = f(x)] \leq 1 - \delta.$$
- *for all δ-smooth distributions $H \sim \{0,1\}^n$, there exists a circuit C of size $O\left(s\gamma^2\right)$ such that*
$$\Pr_{x \sim H}[C(x) = f(x)] \geq \frac{1}{2} + \delta\gamma.$$

The above is a reparametrized version of what appears in [BHKT24], which includes dependencies on δ from their proof (where δ was assumed to be constant).[2] Hence, in the regime $\delta = \Theta(1)$, $\gamma \geq \Omega(1/\sqrt{n})$ and $s = O(2^{\gamma^2 n}/(\gamma^4 n))$, [BHKT24] shows that the γ^2-factor decay in size is tight. Structurally, their argument is very analytically involved and is in multiple stages. In what follows, we say f γ-correlates with g over H if $\Pr_{x \sim H}[f(x) = g(x)] \geq \frac{1}{2} + \gamma$, and f γ-approximates g over H if $\Pr_{x \sim H}[f(x) = g(x)] \geq \gamma$.

[2] [BHKT24, Theorem 2] gives, for any parameters s and γ, a tightness result for a function of input length $n(s, \gamma)$. We have reparametrized to the standard convention of fixing the input length to n, and examining which s, γ are possible (in terms of n).

- Let $k = 1/\gamma^2$. They first prove that for any δ-smooth distribution H, the majority on k bits is $\frac{\delta}{\sqrt{k}}$-correlated with a 1-junta over H, but no $0.01k$-junta can $\frac{1}{4}$-correlate with it over the uniform distribution.[3]

- They then bootstrap this result to show that, for any δ-smooth H, the majority of k random functions on k disjoint $\frac{n}{k}$-bit input blocks is $\frac{\delta}{\sqrt{k}}$-correlated with a size-$O\left(\frac{2^{n/k}}{n/k}\right)$ circuit over H, but requires size $\Omega\left(\frac{k2^{n/k}}{n/k}\right)$ to $\frac{1}{4}$-correlate with (over the uniform distribution). They accomplish this by introducing an analytic relaxation of junta complexity, using Fourier-analytic noise-stability arguments to equate this relaxation to the original measure (up to constant factors), and then using coupling arguments, Fourier-analytic calculations, and subgaussian concentration to analyze the relaxed junta complexity of the random-function ensemble.

We note this junta-to-circuit lifting theorem is interesting in its own right, and towards proving this, they establish a novel direct sum theorem. These techniques are also used in [BHKT24] to tightly characterize the sample-complexity overhead of smooth boosting.

In this note, we give a very short proof that a random junta saturates the hardcore lemma in the regime $\delta = \Omega(1)$ over *arbitrary* distributions.

Theorem 3. *Let $\delta \in (0, 0.49), \gamma \in (0, \frac{1}{2}), n \in \mathbb{N}$ large enough, and $n \le s \le O\left(\frac{2^n}{n}\right)$. If there exists a constant $\varepsilon > 0$ such that $s \ge 1/\gamma^{2+\varepsilon}$, there exists a function f such that*

- *For all circuits C of size s,*

$$\Pr_{x \sim \{0,1\}^n}[C(x) = f(x)] \le 1 - \delta.$$

- *for all distributions H over $\{0,1\}^n$, there exists a circuit C of size $O(s\gamma^2)$ such that*

$$\Pr_{x \sim H}[C(x) = f(x)] \ge \frac{1}{2} + \gamma.$$

Formally, Theorem 2 and Theorem 3 are incomparable: Theorem 2 holds only for $\gamma \ge \Omega(1/\sqrt{n})$ and $s \in [\Omega(1/\gamma^2), O(2^{\gamma^2 n}/(\gamma^4 n))]$, while Theorem 3 holds for any γ and any $s \in [\Omega(1/\gamma^{2+\varepsilon}), O(2^n/n)]$. While incomparable, we note the former region is extremely restrictive, and does not include the common setting of $\gamma = \frac{1}{n}$. Meanwhile, the latter region contain almost all possible (γ, s) pairs. The latter region is short of subsuming the former interval only by an arbitrarily small polynomial factor of $\gamma^{-\varepsilon}$. Removing this ε-slack remains open.

When $s \ge 1/\gamma^{2+\varepsilon}$, Theorem 3 improves on Theorem 2 in two ways:

- Theorem 2 constructs circuits with correlation $\delta\gamma$, while Theorem 3 has correlation γ that does not degrade with δ.

- Theorem 2 requires H to be δ-smooth, but Theorem 3 makes no assumption about H.

[3] A k-junta $f : \{0,1\}^n \to \{0,1\}$ is a Boolean function depending only on a subset of $k < n$ input variables. The junta complexity of a function is the smallest k such that the function is a k-junta.

These two points allow our theorem to remain meaningful even for $\delta = o(1)$.

In summary, our Theorem 3 is stronger in the regime $s \geq 1/\gamma^{2+\varepsilon}$ or $\gamma = O(1/\sqrt{n})$, but gives inferior bounds when $1/\gamma^2 \leq s \leq (1/\gamma)^{2+\varepsilon}$ and $\gamma = \Omega(1/\sqrt{n})$.[4] We also note that Conjecture 1 remains open. In particular, it would be interesting to pin down the optimal dependence of the circuit-size decay on δ.

Our main technical lemma is the strengthening of a beautiful result of Andreev, Clementi, and Rolim [ACR97], which shows that arbitrary Boolean functions can be approximated by small-size circuits.

Theorem 4. *For an arbitrary function $f : \{0,1\}^n \to \{0,1\}$, $\frac{27}{2^{n/2}} < \gamma < \frac{1}{2}$, and any distribution H over $\{0,1\}^n$, there exists a circuit C of size $O\left(\frac{\gamma^2 2^n}{\log(\gamma^2 2^n)} + n\right)$ such that*

$$\Pr_{x \sim H}[C(x) = f(x)] \geq \frac{1}{2} + \gamma.$$

This result was proven by Andreev, Clementi, and Rolim [ACR97] in the case where H is uniform over $\{0,1\}^n$. A short probabilistic argument proves that the circuit size cannot be improved [ACR97, Theorem 4.1].

1.1 Proof Overview

Assume $s = 2^k/k$ for some integer k, and let $H \sim \{0,1\}^n$ be an arbitrary distribution. Intuitively, our proof of Theorem 3 will first use the classic Shannon argument to show that, with high probability, a random function on the first k bits cannot be $(1-\delta)$-approximated by circuits of size s. Letting H' be the induced distribution of H on the first k bits, we can use Theorem 4 to see that there exist circuits of size $O\left(\frac{\gamma^2 2^k}{\log(\gamma^2 2^k)} + k\right) = O(s\gamma^2)$ that γ-correlate with f over H' whenever $s \geq 1/\gamma^{2+\varepsilon}$. The combination of both of these claims implies Theorem 3.

In the main body of the paper, we will actually show a slightly weaker version of Theorem 4 with circuit size $O\left(\frac{\gamma^2 2^n}{\log(\gamma^2 2^n)} + n^2\right)$. The proof of this claim is drastically simpler than that of Theorem 4, but still recovers Theorem 3.

Although the weaker version of Theorem 4 suffices, it is our impression that the result of [ACR97] is not as well known to the community as it should be. It is a complete resolution to a very natural question in circuit complexity (it is the "approximate version" of Lupanov's theorem [Lup71]), and the ideas behind the construction are quite useful (e.g., they appear to have been rediscovered in the construction of covering codes of Rabani and Shpilka [RS10]). This is potentially due to the paper being quite technical and terse, as well as evading search engines. For this reason, we hope to bring attention to this result by giving an exposition and simpler, self-contained proof of Theorem 4 in the appendix, assuming a basic consequence of the fourth-moment method [Ber97] and the existence of asymptotically good codes encodable by linear-size circuits [Spi96]. We now provide a proof overview below.

1.1.1 Overview of Theorem 4

For ease of exposition, we give an overview of Theorem 4 when H is fixed to be uniform over $\{0,1\}^n$. Extending the given arguments to arbitrary distributions H just requires a couple of extra

[4] For general $s = \Omega(1/\gamma^2)$, we get approximating circuits of size $O\left(s\gamma^2 \cdot \frac{\log s}{\log(s\gamma^2)}\right)$. See Remark 1.

modifications. This overview is morally the same as one provided by Trevisan [Tre09], but perhaps with a slightly different point of view in the latter half.

The initial observation is that a random function can actually be efficiently approximated. By standard anticoncentration results, the bias of a random function will be at least $\Omega(2^{-n/2})$ with constant probability, in which case either the constant 0 or 1 function will give a $\left(\frac{1}{2} + \Omega(2^{-n/2})\right)$-approximation of f (we will henceforth refer to this as *square-root anticoncentration*). For a size–approximation trade-off, we can split the truth table of f into 2^k subcubes according to the first k bits of the input, and then approximate each subcube by its majority bit. This is a function depending only on the first k bits of input and can therefore be implemented by a size-$O(2^k/k)$ circuit. Since each subcube will have 2^{n-k} bits, it follows that for a random function, the majority bit will give a $\left(\frac{1}{2} + \Omega(\sqrt{2^{-(n-k)}})\right)$-approximation of the subcube with constant probability. Hence, with high probability, a constant fraction of subcubes will be $\left(\frac{1}{2} + \Omega(\sqrt{2^{-(n-k)}})\right)$-approximated (say, by Chernoff), and thus this circuit will have overall approximation $\frac{1}{2} + \Omega\left(\sqrt{2^{-(n-k)}}\right)$ with f. Setting $k = \log(\gamma^2 2^n)$ gives the result for a random function.

Does this argument work for an arbitrary (rather than a random) f? Clearly not: one can pick any f that is unbiased on each of these 2^k subcubes (e.g., the parity function), and the constructed circuit will have no correlation with f.

What if we could "reduce to the random case" by artificially adding random noise to the truth table of f, and then approximating this noisy function with a circuit on the first k bits? To be more precise, say we had a distribution $\mathcal{C}$ over size-s circuits such that the truth table of $C \sim \mathcal{C}$ was a random string. Then we know $f \oplus \mathcal{C}$ will be a random function, and consequently can be $\left(\frac{1}{2} + \gamma\right)$-approximated by a function g on the first k bits with high probability. We can then fix such a $C \in \mathcal{C}$, and deduce that $C \oplus g$ is a good approximator for f with size $O(2^k/k + s)$.

Unfortunately, a fully random truth table can only be generated by maximally sized circuits. However, we could hope to use a *pseudorandom* string instead. The only property used about the randomness of f was that its truth table had square-root anticoncentration on subcubes. It turns out 4-wise uniform strings have square-root anticoncentration with constant probability [Ber97], motivating us to look at this primitive. Implementing the usual 4-wise uniform-generator construction naively in a circuit immediately gives a distribution $\mathcal{C}$ over circuits of size $O(n^2)$ such that the truth table of $C \sim \mathcal{C}$ is a 4-wise independent string. With more effort, one can get a distribution over $O(n)$-size circuits, which is optimal. By the fourth-moment method [Ber97], we can argue that the average number of subcubes of $f \oplus \mathcal{C}$ with square-root anticoncentration is at least a constant proportion. Fixing $C \in \mathcal{C}$ that achieves this average, it follows that there is a function g on the first k bits that approximates $f \oplus C$ well by our previous analysis. Consequently, $C \oplus g$ is a circuit of size $O(2^k/k + n)$ that approximates f well, as desired.

2 Preliminaries

All logarithms are in base 2. $[n] \coloneqq \{1, \ldots, n\}$. $\mathbb{F}_{2^n}$ denotes the finite field of 2^n elements, and each element will be identified by either a string in $\{0,1\}^{2^n}$ or integer in $[2^n]$ in the natural way. For a distribution D, $d \sim D$ is an element sampled from D. If S is a set, we denote $s \sim S$ to be a uniformly random element from S. $\circ$ denotes string concatenation. We consider circuits with arbitrary gates of fan-in 2 and arbitrary depth. A Boolean function $f : \{0,1\}^n \to \{0,1\}$ is a k-junta if $f(x) = g(x_S)$ for some subset $S \subset [n]$ of size k. For $x, y \in \{0,1\}^n$, we denote the distance between

x and y to be the quantity $|\{i \in [n] : x_i \neq y_i\}|$.

For $f, g : \{0,1\}^n \to \{0,1\}$, we say f γ-correlates with g over H if $\Pr_{x\sim H}[f(x) = g(x)] \geq \frac{1}{2} + \gamma$, and f γ-approximates g over H if $\Pr_{x\sim H}[f(x) = g(x)] \geq \gamma$. If no H is specified, it is assumed to be the uniform distribution over $\{0,1\}^n$.

We now define k-wise uniformity.

Definition 1 (k-wise uniformity [HH24])**.** *A distribution D over $\{0,1\}^n$ is a* k-wise uniform distribution *if, for all subsets $T \subset [n]$ of size k, the marginal distribution $(x_T)_{x\sim D}$ is uniform over $\{0,1\}^k$. A function $G : S \to \{0,1\}^n$ is a* k-wise uniform generator *if $(G(s))_{s\sim S}$ is a k-wise uniform distribution.*

A crucial property of 4-wise uniform strings is that they enjoy square-root anticoncentration with constant probability, just like a fully random string.

Theorem 5 ([Ber97, Corollary 3.1])**.** *Let X be a 4-wise uniform distribution over $\{-1,1\}^n$, and let $v \in \mathbb{R}^n$. We have*

$$\Pr_{x\sim X}\left[\left|\sum_{i=1}^n v_i x_i\right| \geq \sqrt{\frac{\sum_{i=1}^n v_i^2}{3}}\right] \geq \frac{2}{11}.$$

We will want 4-wise generators such that, for a fixed seed, each output bit can be locally computed in small circuit size. The standard construction of 4-wise uniform generators serves this purpose for us.

Theorem 6. *There exists a 4-wise uniform generator $G : S \to \{0,1\}^{2^n}$ such that, for each $s \in S$ and $x \in [2^n]$, there exists a circuit C_s of size $O(n^2)$ with $C_s(x) = G(s)_x$.*

Proof. Define $\iota : \mathbb{F}_{2^n} \to \{0,1\}$ to map $x \in \mathbb{F}_{2^n}$ to the first bit of the binary encoding of x. Let $G : \mathbb{F}_{2^n}^4 \to \mathbb{F}_2^{2^n}$ be defined by the evaluation map

$$G(s) \coloneqq \left(\iota\left(\sum_{i=1}^4 s_i x^{i-1}\right)\right)_{x\in\mathbb{F}_{2^n}}.$$

This is a 4-wise uniform generator (see [HH24, Theorem 2.2]). Notice that, as a function of x, $G(s)_x$ is an evaluation of a degree-3 polynomial, which can be done in a circuit of size $O(n^2)$ by grade-school multiplication (better multiplication algorithms are known, but this suffices). □

This theorem suffices to prove Theorem 3. A technical contribution of [ACR97], and a key ingredient behind Theorem 4, is a 4-wise uniform generator that can be locally computed in linear circuit size, which is the best one could hope for.

Theorem 7 ([ACR97])**.** *There exists a 4-wise uniform generator $G : S \to \{0,1\}^{2^n}$ such that, for each $s \in S$ and $x \in [2^n]$, there exists a circuit C_s of size $O(n)$ with $C_s(x) = G(s)_x$.*

We give a self-contained proof of this in the appendix.

3 Tightness of Impagliazzo's Hardcore Lemma

We will now prove a lemma that shows how to construct small circuit approximators for arbitrary functions using 4-wise uniform generators.

Lemma 1. *Let $f : \{0,1\}^n$ be an arbitrary Boolean function, let H be any distribution over $\{0,1\}^n$, and let $\gamma \in (\frac{27}{2^{n/2}}, \frac{1}{2})$. Let $G : \{0,1\}^m \to \{0,1\}^{2^n}$ be a 4-wise uniform generator such that, for each $s \in \{0,1\}^m$, there exists a circuit C_s of size r with $C_s(x) = G(s)_x$. Then there exists a circuit C of size $O\left(\frac{\gamma^2 2^n}{\log(\gamma^2 2^n)} + r\right)$ such that*

$$\Pr_{x\sim H}[C(x) = f(x)] \geq \frac{1}{2} + \gamma.$$

Proof. Denote $\ell \coloneqq \lfloor \log(1/363\gamma^2) \rfloor$, and note $n - \ell \geq 1$. Let H' be the distribution over $\{0,1\}^{n-\ell}$ defined by the probability mass function

$$H'(c) \coloneqq \Pr_{x\sim H}[x \in c \times \{0,1\}^\ell].$$

This is the induced distribution of H on the subcubes defined by the first $n - \ell$ bits of the input. For each c, define the conditional distribution over the subcube $c \times \{0,1\}^\ell$ by the function

$$H_c(y) \coloneqq \Pr_{x\sim H}[x = c \circ y \mid x \in c \times \{0,1\}^\ell] = \frac{\Pr_{x\sim H}[x = c \circ y]}{H'(c)}.$$

Let G be the 4-wise uniform generator guaranteed by the hypothesis. For a subcube c, denote the indicator variable

$$I_c(s) \coloneqq \mathbf{1}\left(\left|\sum_{y\in\{0,1\}^\ell} H_c(y)(-1)^{f(c\circ y)+G(s)_{c\circ y}}\right| \geq \sqrt{\frac{\sum_{y\in\{0,1\}^\ell} H_c(y)^2}{3}}\right).$$

By Theorem 5, we have that for each c and random s, $\Pr_{s\sim\{0,1\}^m}[I_c(s) = 1] \geq \frac{2}{11}$. Hence,

$$\mathbb{E}_{s\sim\{0,1\}^m}\left[\mathbb{E}_{c\sim H'}[I_c(s)]\right] = \mathbb{E}_s\left[\sum_c H'(c) I_c(s)\right] \geq \frac{2}{11}\sum_c H'(c) = \frac{2}{11},$$

and by an averaging argument there exists a choice of s such that $\Pr_{c\sim H'}[I_c(s) = 1] \geq 2/11$. Fix such an s. Now for any c with $I_c(s) = 1$, we can use Cauchy-Schwarz to lower bound

$$\left|\sum_{y\in\{0,1\}^\ell} H_c(y)(-1)^{f(c\circ y)+G(s)_{c\circ y}}\right| \geq \sqrt{\frac{\sum_{y\in\{0,1\}^\ell} H_c(y)^2}{3}} \geq \sqrt{\frac{(1/2^\ell)\sum_{y\in\{0,1\}^\ell} H_c(y)}{3}} = \sqrt{\frac{1}{2^\ell \cdot 3}}$$

Now define $h : \{0,1\}^{n-\ell} \to \{0,1\}$ by the map

$$h(c) = \mathbf{1}\left(\sum_{y\sim\{0,1\}^\ell} H_c(y)(-1)^{f(x)+G(s)_{c\circ y}} < 0\right),$$

which encodes whether the subcube is positively or negatively correlated with the 4-wise uniform string. We now set our approximator $g : \{0,1\}^n \to \{0,1\}$ to be $g(c \circ y) \coloneqq h(c) \oplus G(s)_{c \circ y}$.

We can now write the correlation between f and g (with respect to H) as

$$\begin{aligned}
\mathbb{E}_{x\sim H}[(-1)^{f(x)+g(x)}] &= \mathbb{E}_{c\sim H'}\left[\sum_{y\in\{0,1\}^\ell} H_c(y)(-1)^{f(c\circ y)+g(c\circ y)}\right] \\
&= \mathbb{E}_{c\sim H'}\left[(-1)^{h(c)}\sum_{y\in\{0,1\}^\ell} H_c(y)(-1)^{f(c\circ y)+G(s)_{c\circ y}}\right] \\
&= \mathbb{E}_{c\sim H'}\left[\left|\sum_{y\in\{0,1\}^\ell} H_c(y)(-1)^{f(c\circ y)+G(s)_y}\right|\right] \\
&\geq \Pr_{c\sim H'}[I_c(s)] \cdot \mathbb{E}_{c\sim H'}\left[\left|\sum_{y\in\{0,1\}^\ell} H_c(y)(-1)^{f(c\circ y)+G(s)_y}\right| : I_c(s) = 1\right] \\
&\geq \frac{2}{11}\cdot\sqrt{\frac{1}{2^\ell\cdot 3}} \geq 2\gamma.
\end{aligned}$$

Hence,

$$\Pr_{x\sim H}[f(x) = g(x)] = \frac{\mathbb{E}_{x\sim H}[(-1)^{f(x)+g(x)}]+1}{2} \geq \frac{1}{2}+\gamma.$$

Since h depends only on the first $n-\ell$ bits, it can be constructed in circuit size $O\left(\frac{2^{n-\ell}}{n-\ell}\right) = O\left(\frac{\gamma^2 2^n}{\log(\gamma^2 2^n)}\right)$. By construction, $G(s)_y$ can be computed by some size-r circuit C_s. Therefore, g γ-correlates with f and can be computed by a circuit of size $O\left(\frac{\gamma^2 2^n}{\log(\gamma^2 2^n)} + r\right)$, as desired. □

We are now ready to prove Theorem 3.

Proof of Theorem 3. We will first prove the first item. Pick a function $g : \{0,1\}^k \to \{0,1\}$ uniformly at random, and define $f(x) \coloneqq g(x_{\leq k})$ to be g on the first k bits of the input. By a Chernoff bound, we know that for a fixed circuit $C : \{0,1\}^k \to \{0,1\}$ of size $\leq s$ and $\delta < 0.49$,

$$\Pr_f\left[\Pr_x[g(x) = C(x)] \geq 1-\delta\right] \leq 2^{-\Omega(2^k)}.$$

Set $2^k = \Omega(s \log s)$. Taking a union bound over all circuits of size s (of which there are $s^{O(s)}$) we note that, with probability at most $s^{O(s)}2^{-\Omega(2^k)} < 1$, g can be $(1-\delta)$-approximated by a size-s circuit. Fix a g that cannot. We now show f cannot be approximated by any circuit $C : \{0,1\}^n \to \{0,1\}$ of size s. Restricting the last $n-k$ bits of C will result in a circuit on the first k bits of size at most s. Hence, we can deduce

$$\Pr_x[f(x) = C(x)] = \mathbb{E}_{x_{>k}}\left[\Pr_{x_{\leq k}}[g(x_{\leq k}) = C(x_{\leq k}, x_{>k})]\right] < 1-\delta$$

as desired.

We will now prove this f satisfies the second item. Consider an arbitrary distribution H over $\{0,1\}^n$. The marginal distribution of H on the first k bits will be some distribution H' over $\{0,1\}^k$. As $\gamma > s^{-1/2} > 27 \cdot 2^{-k/2}$ for large enough n, Lemma 1 and Theorem 6 give a circuit $C : \{0,1\}^k \to \{0,1\}$ of size $O\left(\frac{\gamma^2 2^k}{\log(\gamma^2 2^k)} + k^2\right)$ such that $\Pr_{x\sim H'}[g(x) = C(x)] \geq \frac{1}{2} + \gamma$. Letting $C' : \{0,1\}^n \to \{0,1\}$ be the circuit that applies C to the first k bits, we note that

$$\Pr_{x\sim H}[f(x) = C'(x)] = \Pr_{x\sim H'}[g(x) = C(x)] \geq \frac{1}{2} + \gamma.$$

As C' has the same size as C, and $\frac{\gamma^2 2^k}{\log(\gamma^2 2^k)} + k^2 = O\left(s\gamma^2 \cdot \left(\frac{\log s}{\log(\gamma^2 s)}\right) + \log^2 s\right) = O_\varepsilon(s\gamma^2)$ when $s \geq 1/\gamma^{2+\varepsilon}$, C' is a size-$O_\varepsilon(s\gamma^2)$ circuit that γ-correlates with f over H. Consequently, f is a k-junta that cannot be $(1-\delta)$-approximated by a size-s circuit, but, over any H, γ-correlates with a size-$O_\varepsilon(s\gamma^2)$ circuit, as desired. □

Remark 1. *If Lemma 1 and Theorem 7 is used, rather than Lemma 1 and Theorem 6, we instead get γ-correlating circuits of size $O\left(s\gamma^2 \cdot \left(\frac{\log s}{\log(\gamma^2 s)}\right) + \log s\right) = O\left(s\gamma^2 \cdot \frac{\log s}{\log(s\gamma^2)}\right)$ for all $s = \Omega(1/\gamma^2)$.*

Remark 2. *Upon seeing the above proof, one might notice that it suffices to prove that a* random *function has small approximating circuits over arbitrary H, rather than an* arbitrary *function. Section 1.1.1 gives a very simple argument to efficiently approximate a random function, so one might wonder why* arbitrary *functions are considered. The arbitrariness of H seems to force us to consider arbitrary f (but this is not a rigorous claim). For example, can construct H that renders the approximating circuit for the random function discussed in Section 1.1.1 useless.*

Acknowledgements

We thank Guy Blanc, Geoffrey Mon and David Zuckerman for illuminating discussions. We also thank Jeffrey Champion, Sabee Grewal, Rocco Servedio, and anonymous reviewers for comments on a draft of this work.

References

[ACR97] Alexander E. Andreev, Andrea E.F. Clementi, and José D.P. Rolim. Optimal bounds for the approximation of boolean functions and some applications. *Theoretical Computer Science*, 180, 1997. doi:10.1016/S0304-3975(96)00217-4. [pp. 4, 6, 11]

[Ber97] Bonnie Berger. The fourth moment method. *SIAM Journal on Computing*, 1997. doi:10.1137/S0097539792240005. [pp. 4, 5, 6]

[BHK09] Boaz Barak, Moritz Hardt, and Satyen Kale. The uniform hardcore lemma via approximate bregman projections. In *SODA*, 2009. doi:10.1137/1.9781611973068.129. [pp. 1, 2]

[BHKT24] Guy Blanc, Alexandre Hayderi, Caleb Koch, and Li-Yang Tan. The sample complexity of smooth boosting and the tightness of the hardcore theorem. In *FOCS*, 2024. doi:10.1109/FOCS61266.2024.00092. [pp. 2, 3]

[CDV24] Sílvia Casacuberta, Cynthia Dwork, and Salil Vadhan. Complexity-theoretic implications of multicalibration. In *STOC*, 2024. doi:10.1145/3618260.3649748. [p. 2]

[CL21] Lijie Chen and Xin Lyu. Inverse-exponential correlation bounds and extremely rigid matrices from a new derandomized xor lemma. In *STOC*, 2021. doi:10.1145/3406325.3451132. [p. 2]

[GDP73] S. I. Gelfand, R. L. Dobrushin, and M. S. Pinsker. On the complexity of coding. In *Second International Symposium on Information Theory*, pages 177–184, 1973. [pp. 11, 12]

[HH24] Pooya Hatami and William M. Hoza. Paradigms for unconditional pseudorandom generators. *Foundations and Trends® in Theoretical Computer Science*, 16, 2024. doi:10.1561/0400000109. [p. 6]

[Hol05] Thomas Holenstein. Key agreement from weak bit agreement. In *STOC*, 2005. doi:10.1145/1060590.1060688. [pp. 1, 2]

[Imp95] Russell Impagliazzo. Hard-core distributions for somewhat hard problems. In *FOCS*, 1995. doi:10.1109/SFCS.1995.492584. [p. 1]

[KS99] Adam R. Klivans and Rocco A. Servedio. Boosting and hard-core sets. In *FOCS*, 1999. doi:10.1109/SFFCS.1999.814638. [pp. 1, 2]

[LTW11] Chi-Jen Lu, Shi-Chun Tsai, and Hsin-Lung Wu. Complexity of hard-core set proofs. *Computational Complexity*, 2011. doi:10.1007/s00037-011-0003-7. [p. 2]

[Lup71] O. B Lupanov. Complexity of formula realization of functions of logical algebra. *Journal of Symbolic Logic*, 36(3):547–548, 1971. doi:10.2307/2270028. [pp. 4, 11]

[O'D02] Ryan O'Donnell. Hardness amplification within NP. In *STOC*, 2002. doi:10.1145/509907.510015. [p. 2]

[RS10] Yuval Rabani and Amir Shpilka. Explicit construction of a small ϵ-net for linear threshold functions. *SIAM Journal on Computing*, 2010. doi:10.1137/090764190. [p. 4]

[RTTV08] Omer Reingold, Luca Trevisan, Madhur Tulsiani, and Salil Vadhan. Dense subsets of pseudorandom sets. In *FOCS*, 2008. doi:10.1109/FOCS.2008.83. [p. 2]

[Spi96] Daniel A. Spielman. Linear-time encodable and decodable error-correcting codes. *IEEE Transactions on Information Theory*, 1996. doi:10.1109/18.556668. [pp. 4, 11, 12]

[STV99] Madhu Sudan, Luca Trevisan, and Salil Vadhan. Pseudorandom generators without the xor lemma (extended abstract). In *STOC*, 1999. doi:10.1145/301250.301397. [p. 2]

[Tre05] Luca Trevisan. On uniform amplification of hardness in np. In *STOC*, 2005. doi:10.1145/1060590.1060595. [p. 2]

[Tre09] Luca Trevisan. Approximating a boolean function via small circuits. https://lucatrevisan.wordpress.com/2009/11/06/approximating-a-boolean-function-via-small-circuits/, 2009. Accessed: 2025-08-08. [pp. 5, 11]

A Optimal Approximating Circuits: Proofs of Theorem 7 and Theorem 4

In this section, we give motivation and a proof of the fundamental result of Andreev, Clementi, and Rolim [ACR97]. The exposition here has nontrivial overlap with that of Trevisan [Tre09].

A.1 Motivation

A classical probabilistic argument of Shannon says that there exists functions mapping n bits to one bit that have circuit complexity $\Omega(2^n/n)$. What is remarkable is that this argument is *tight*; Lupanov's theorem states that *any* function $f : \{0,1\}^n \to \{0,1\}$ can be computed in circuit size $O(2^n/n)$ [Lup71].

Upon studying Shannon's argument, it is not hard to see that this lower-bound argument extends to circuits that only *approximate* rather than *compute*. Let $f : \{0,1\}^n \to \{0,1\}$ be a uniformly random function, and let $\gamma \in (0, \frac{1}{2})$ be such that $\gamma^2 2^n \geq 2$. For a fixed circuit C, we have by the Chernoff bound that

$$\Pr_f \left[\Pr_x [C(x) = f(x)] \geq \frac{1}{2} + \gamma \right] \leq 2^{-\Omega(\gamma^2 2^n)}.$$

Setting $s = \Theta\left(\frac{\gamma^2 2^n}{\log(\gamma^2 2^n)}\right)$ and taking a union bound over all $s^{O(s)}$ circuits C of size s tells us that the probability f is γ-correlated with a size-s circuit is at most $s^{O(s)} 2^{-\Omega(\gamma^2 2^n)} < 0.1$.

It is now natural to ask whether this is tight: for any Boolean function, is there a size-$O\left(\frac{\gamma^2 2^n}{\log(\gamma^2 2^n)}\right)$ circuit that γ-approximates f? This is not true by considering the parity function. Any function that does not depend on all n bits will be correct on the parity function on exactly half the inputs. Hence, an approximating circuit for parity must depend on all input bits and thus have circuit size at least n. We can then update our hypothesis and ask whether, for any Boolean function, there is a size-$O\left(\frac{\gamma^2 2^n}{\log(\gamma^2 2^n)} + n\right)$ circuit that γ-correlates with it.

A natural first approach is to try to use Lupanov's Theorem. In particular, we can construct a circuit that exactly computes f inside a subcube of volume $2\gamma 2^n$, and then otherwise outputs a fixed bit, whichever matches f better. This is guaranteed to exactly match f within the subcube and match on at least half the inputs outside the subcube, giving a γ-correlating circuit. However, this will be a circuit of size $O\left(\frac{\gamma 2^n}{\log(\gamma 2^n)}\right)$.

Andreev, Clementi, and Rolim give a circuit construction matching the probabilistic bound, effectively establishing the approximation analog of Lupanov's theorem. We believe this result to be fundamental, but unfortunately, it appears to be relatively unknown to the community. Hence, we give a modern and simplified presentation of a slightly stronger result here (Theorem 4).

Notice that Lemma 1 instantiated with Theorem 7 gives Theorem 4. Lemma 1 was proven in Section 3, so we now focus on the proof of Theorem 7. This was implicitly proven in [ACR97], but we give a shorter proof using asymptotically good codes encodable in linear time [GDP73, Spi96].

A.2 Overview of Theorem 7

The starting point of the construction is to consider the simpler task of a 2-wise uniform generator. There is a classic 2-wise uniform generator mapping a nonzero seed of length n to a string of

size $2^n - 1$, which is to simply output all nonempty $\mathbb{F}_2$-linear combinations of the seed; that is, $G(s) \coloneqq (\langle s, r\rangle)_{r \in \mathbb{F}_2^n \setminus \{0\}}$. Notice that, for a fixed s, the output of the rth bit as a function of r is simply some parity of a subset of bits in r, which is trivially a circuit of size $O(n)$, as desired.

Why is this generator not a 4-wise uniform generator? It is because of linear dependence. In particular, for nonzero vectors x, y, we have $G(s)_x + G(s)_y = G(s)_{x+y}$. Hence, we do not even have 3-wise uniformity, as the bits in indices x, y, and $x + y$ are correlated. This motivates the following idea: what if we only focused on a subset of indices $Y \subset \{0,1\}^n$ such that all distinct $x_1, x_2, x_3, x_4 \in Y$ are linearly independent? Can we show that $(G(s)_y)_{y \in Y}$ is a 4-wise independent string? Yes.

Lemma 2. *Let $Y \subset \mathbb{F}_2^n$ be a subset such that, for all subsets $X \subset Y$ of size 4, X is linearly independent. Then $G : \{0,1\}^n \to \{0,1\}^Y$ defined by $G(s)_y = \langle y, s\rangle$ is a 4-wise uniform generator.*

Proof. Consider arbitrary $X \subset Y$ of size 4. We will show that over a uniform s, the string $(\langle s, x\rangle)_{x \in X}$ is uniform. Notice this string is simply $M \cdot s$, where M is an $\mathbb{F}_2^{4 \times n}$ matrix whose rows are the elements of X. Hence, every preimage of this map has the same size, namely that of the kernel of M. It remains to show that the image of M is $\mathbb{F}_2^4$. But this is clear, as X consists of linearly independent vectors, implying that the rank of M is 4. □

In light of this, we will try to construct a linear-size circuit $h : \{0,1\}^n \to \{0,1\}^{16n}$ such that, for all distinct $x_1, \ldots, x_4$, the vectors $h(x_1), h(x_2), h(x_3), h(x_4)$ are linearly independent. Then our 4-wise generator $G : \{0,1\}^{16n} \to \{0,1\}^{2^n}$ would be $G(s)_x = \langle h(x), s\rangle$, which will be a linear-size circuit. How do we construct such an h?

Perhaps a natural approach is to make h randomly scatter the x's randomly among $\{0,1\}^{16n}$. This actually works, because for a fixed $a \leq 4$ and $x_1, \ldots, x_a \in \mathbb{F}_2^n$, the probability $h(x_1) + \cdots + h(x_a) = 0$ is 2^{-16n}. Taking a union bound over all tuples of size at most 4 gets the desired result. Of course, the issue is that this is not a linear-size circuit: a random h will have maximal circuit complexity. What if we let each bit of h be a random function on only constantly many bits of x?

More concretely, say we pick subsets $S_1, \ldots, S_{16n} \subset [n]$ of constant size uniformly at random, and then let $g_i : \{0,1\}^{S_i} \to \{0,1\}$ be a random function. Define $h : \{0,1\}^n \to \{0,1\}^{16n}$ by $h(x) \coloneqq (g_1(x_{S_1}), \ldots, g_{16n}(x_{S_{16n}}))$. This is of linear circuit size, and we are hoping the randomness of the g_i keeps vectors linearly independent. Fix $x_1, \ldots, x_a$ for $a \leq 4$. We want the probability that $g_i((x_1)_{S_i}) + \cdots + g_i((x_a)_{S_i}) = 0$ for all i to be at most $1/\binom{2^n}{\leq 4}$. Unfortunately, this need not be true. Say $x_1, \ldots, x_4$ are within distance 2 of each other. Then the probability that a random constant-sized S_i satisfies $(x_1)_{S_i} = \cdots = (x_4)_{S_i}$ will be high. In this case, $g_i((x_j)_{S_i})$ will be guaranteed to all be the same for every $j \leq 4$, and so $g_i((x_1)_{S_i}) + \cdots + g_i((x_4)_{S_i}) = 0$. The key issue is that a randomly picked local view, S_i, might interpret $x_1, \ldots, x_4$ as the same. This motivates the final trick of first encoding x using an asymptotically good error-correcting code before picking our sets S_i. This will force different x's to have very different encodings, and then a random set S_i will indeed detect a difference. This will allow the randomness of g_i to prevent linear dependencies from happening.

But are there asymptotically good codes encodable in linear circuit size? Indeed, non-explicit constructions of such codes were known to exist since 1974, thanks to Gelfand, Dobrushin, and Pinsker [GDP73]. Non-explicit constructions suffice for our application, but we mention that Spielman codes are explicit constructions of such a code [Spi96].

Theorem 8 ([GDP73, Spi96])**.** *For any n there exists a small enough constant $\delta, m \leq 4n$, and an $O(n)$-sized circuit $C : \{0,1\}^n \to \{0,1\}^m$ such that for $x \neq y$, $C(x)$ and $C(y)$ have distance $\geq \delta m$.*

With this primitive, we can construct our 4-wise independence generator.

A.3 Proofs of Theorem 7 and Theorem 4

Proof of Theorem 7. Let $\mathsf{Enc} : \{0,1\}^n \to \{0,1\}^m$ be an $O(n)$-sized circuit implied by Theorem 8. We will show that there exist sets $S_1, \ldots, S_{16n} \subset [m]$, each of size $\leq \lceil 10/\delta \rceil$, and functions $g_1, \ldots, g_{16n}$, with $g_i : \{0,1\}^{S_i} \to \{0,1\}$, such that

$$G(s)_x \coloneqq \langle (g_1(\mathsf{Enc}(x)_{S_1}), \ldots, g_{16n}(\mathsf{Enc}(x)_{S_{16n}})), s \rangle$$

is a 4-wise uniform generator. Once we have this, the desired result follows, as for a fixed s, $\mathsf{Enc}(x)$ can be computed in linear circuit size, each g_i can be computed in constant circuit size, and the parity of any subset of the $16n$ bits can be done in linear circuit size.

By Lemma 2, it suffices to show the existence of $\{S_i\}, \{g_i\}$ such that for any $X \subset \mathbb{F}_2^n \setminus \{0\}$ of size ≤ 4, there exists $i \in [16n]$ such that $\sum_{x \in X} g_i(\mathsf{Enc}(x)_{S_i}) \neq 0$. This will be done by the probabilistic method. In particular, we will pick each S_i by selecting $\lceil 10/\delta \rceil$ elements uniformly and independently from $[m]$, and pick each $g_i : \{0,1\}^{S_i} \to \{0,1\}$ uniformly at random. Consider arbitrary $X \subset \{0,1\}^n \setminus \{0^n\}$ of cardinality at most 4.

First assume $2 \leq |X| \leq 4$. By construction of $\mathsf{Enc}(\cdot)$, the strings $\{\mathsf{Enc}(x)\}_{x \in X}$ will have pairwise distance $\geq \delta m$. Therefore, for a fixed $i \in [16n]$ and $x \neq x' \in X$, the probability that a random S_i satisfies $\mathsf{Enc}(x)_{S_i} = \mathsf{Enc}(x')_{S_i}$ is at most $(1-\delta)^{|S_i|}$. Hence, the probability that $\mathsf{Enc}(x)_{S_i} \neq \mathsf{Enc}(x')_{S_i}$ for some $x \neq x' \in X$ is at least

$$1 - \binom{4}{2}(1-\delta)^{|S_i|} = 1 - 6e^{-10} \geq \frac{99}{100}$$

by a union bound. Conditioned on this event, $\sum_{x \in X} g_i(\mathsf{Enc}(x)_{S_i})$ is a uniform bit for a random g_i. Thus, for a fixed i, $\sum_{x \in X} g_i(\mathsf{Enc}(x)_{S_i}) \neq 0$ with probability at least $\frac{99}{100} \cdot \frac{1}{2} \geq \frac{1}{3}$. Since each coordinate is independent, the probability that $\sum_{x \in X} g_i(\mathsf{Enc}(x)_{S_i}) = 0$ for all i is at most $(2/3)^{16n}$.

Now assume $X = \{x\}$. For any fixing of $\{S_i\}$ and for random $\{g_i\}$, $(g_i(\mathsf{Enc}(x)_{S_i}))_{i \in [16n]}$ is a uniformly random string in $\{0,1\}^{16n}$, and is consequently 0 with probability $2^{-16n} < (2/3)^{16n}$.

Thus, by a union bound, all subsets X of size at most 4 satisfy $\sum_{x \in X} g_i(\mathsf{Enc}(x)_{S_i}) \neq 0$ for some $i \in [16n]$ with probability at least

$$1 - \binom{2^n}{\leq 4}(2/3)^{16n} \geq 1 - 2^{4n}(2/3)^{16n} > 0,$$

implying the existence of such $\{S_i\}$ and $\{g_i\}$, and thereby yielding the result. □

With the help of Lemma 1, Theorem 4 is now immediate.

Proof of Theorem 4. Simply use the 4-wise uniform generator of Theorem 7 in Lemma 1. □

ECCC ISSN 1433-8092
https://eccc.weizmann.ac.il