Most Juntas Saturate the Hardcore Lemma

Vinayak M. Kumar*

Abstract

Blanc, Hayderi, Koch, and Tan [FOCS, 2024] proved that Impagliazzo's Hardcore Lemma is quantitatively tight. We give a much simpler proof of this fact by showing that the lemma is tight for a random junta 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 such a hardcore set for circuits of size $s' \ll s$.

Theorem 1 ([Imp95, KS99, Hol05]). *Let $f : \{0,1\}^n \to \{0,1\}$ and $\delta, \gamma, s \leq \frac{2^n}{n}$. 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 at least 2δ 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 always be explained by a subset of "hard inputs" H on which the function looks random to small circuits. This phenomenon has found 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. If the following conjecture is true, the degradation is necessary.

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

[*] vmkumar@utexas.edu. Department of Computer Science, The University of Texas at Austin. Supported by a Jane Street Graduate Research Fellowship.

- *for every subset $H \subset \{0,1\}^n$ of density $\geq 2\delta$, there exists a circuit C of size $O\left(\frac{s\gamma^2}{\log(1/\delta)}\right)$ such that*

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

- *for all circuits C of size $\leq s$,*

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

Such a degradation was shown to be necessary in certain black-box models [LTW11], but an unconditional result seemed elusive, as giving an explicit f seems to require proving breakthrough circuit lower bounds. In a recent work, Blanc, Hayderi, Koch, and Tan [BHKT24] evaded this barrier by using a semi-explicit function.

Theorem 2. *Let $\delta, \gamma \in (0,1)$, integer n , $(1/\gamma^2) \leq s \leq 2^{\gamma^2 n}/\gamma^2 n$ be parameters with $s \geq \Omega(1/\gamma^2)$. There exists f such that*

- *for all $H \subset \{0,1\}^n$ of density $\geq 2\delta$, there exists a circuit C of size $O\left(s\gamma^2\right)$ such that*

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

- *For all circuits C of size $\leq s$,*

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

Hence, in the regime $\delta = \Theta(1)$ and $s \leq 2^{\gamma^2 n}/\gamma^2 n$, [BHKT24] show that the γ^2 decay in size is tight. They show with high probability, the majority of random functions on disjoint blocks of input bits witness this tightness. Intuitively, the "explicitness" of the function allows one to argue small circuits for any subdomain H, and the random portion of the construction allows a Shannon counting argument to deduce the lower bound over the whole domain. Structurally, their argument is very analytically involved; they first prove the result for junta complexity, and then lift it to circuit complexity using Fourier-analytic techniques and an analytic relaxation of junta complexity. These techniques are also used in [BHKT24] to tightly characterize the sample complexity overhead of smooth boosting.

In this note, we show that a random junta suffices to prove the tightness of the hardcore lemma in the regime $\delta = \Omega(1)$.

Theorem 3. *Let $\gamma, n, \delta \leq .49, s \leq 2^n/n$. If there exists constant $\varepsilon > 0$ such that $s \geq 1/\gamma^{2+\varepsilon}$, there exists a function f such that*

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

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

- *For all circuits C of size s,*

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

In fact, our example does not require an upper bound on s in terms of γ, we can give a small size approximating circuit for *any* distribution H over $\{0,1\}^n$, and the correlation over H does not degrade with δ. Furthermore, the simplicity of the construction lends itself to a simpler proof. Unfortunately, strengthening to an arbitrary distribution H comes at a cost: we only achieve optimal size decay when $s \geq (1/\gamma)^{2+\varepsilon}$, versus the optimal bound of $s \geq \Omega(1/\gamma^2)$. 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.

The main technical lemma is a 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 $f : \{0,1\}^n \to \{0,1\}$ 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 H is uniform over $\{0,1\}^n$. A short probabilistic argument proves the circuit size cannot be improved [ACR97, Theorem 4.1].

Assume $s = 2^k/k$ for some integer k. Intuitively, our proof of Theorem 3 will first use the classic Shannon argument to show that with high probability, a random k-junta cannot be $(1-\delta)$-approximated by circuits of size s. We then use Theorem 4 to show for any H, there exist circuits of size $O(\gamma^2 2^k / \log(\gamma^2 2^k) + k) = O(s\gamma^2)$ that $(1/2+\gamma)$-approximates f over H whenever $s \geq 1/\gamma^{1/2+\varepsilon}$. The combination of both of these claims imply Theorem 3.

In the main body of the paper, we will actually show a slightly weaker version of Theorem 4 with circuit size of $O\left(\frac{\gamma^2 2^n}{\log(\gamma^2 2^n)} + n^2\right)$. 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 (this can be seen as the "approximate version" of Lupanov's theorem), and its 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. Due to this, we hope to bring attention to this result and give a self-contained and simpler 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].

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 extra 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 starting point is to observe 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 either the constant 0 or 1 function will give a $\left(\frac{1}{2}+\Omega(2^{-n/2})\right)$-approximation of f. For a size-approximation tradeoff, 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 only depending 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 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 $1/2+\Omega(\sqrt{2^{-(n-k)}})$ with f. Setting $k=\log(\gamma^2 2^n)$ gives the result for a random function.[1]

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 $1/2+\gamma$ 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 $C \oplus g$ 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. It turns out that a 4-wise uniform string has square-root anticoncentration with constant probability. Implementing the usual 4-wise uniform generator construction naively into 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. Hence, we can argue that the average number of subcubes of $f \oplus \mathcal{C}$ with squareroot anticoncentration is at least a constant proportion. Fixing $C \in \mathcal{C}$ that achieves this average, it follows there is a function g on the first k bits that approximates $f \oplus C$ well. It follows $C \oplus g$ approximates f well.

2 Preliminaries

All logarithms are in base 2. 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.

We now define k-wise uniformity.

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

A crucial property about 4-wise uniform strings are that they enjoy squareroot anticoncentration with constant probability just like a fully random string.

[1] From our introduction, one might think proving Theorem 4 for a random function suffices to show Theorem 3. This argument shows for a fixed distribution H, a random function will have a small-size approximator, but will not say that a random function will have small-size approximators for *all* H.

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*

$$\Pr_{x \sim X}\left[\left|\sum v_i x_i\right| \geq \frac{||v||_2}{\sqrt{3}}\right] \geq \frac{2}{11}.$$

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 $G : \{0,1\}^m \to \{0,1\}^{2^n}$ such that for each $x \in \{0,1\}^m$, there exists a circuit C_x of size $O(n^2)$ such that $C_x(i) = G(x)_i$.*

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

$$G(v) := \left(\iota\left(\sum_{i<4} v_i x^i\right)\right)_{x \in \mathbb{F}_{2^n}}.$$

This a 4-wise uniform generator (see [HH24, Theorem 2.1.3]). Notice that as a function of i, $G(x)_i$ 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] was to give a 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 $G : \{0,1\}^m \to \{0,1\}^{2^n}$ such that for each $x \in \{0,1\}^m$, there exists a circuit C_x of size $O(n)$ such that $C_x(i) = G(x)_i$.*

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

3 Tightness of Impagliazzo's Hardcore Lemma

We will now prove a lemma which 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, and let H be any distribution over $\{0,1\}^n$, and γ be a parameter. 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_x of size r such that $C_x(i) = G(x)_i$. 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 $\ell := \log(1/\gamma^2)$. Let H' be the distribution over $\{0,1\}^{n-\ell}$ defined by the probability mass function

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

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

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

Let G be the 4-wise uniform generator guaranteed by the hypothesis. By Theorem 5, we have that for each c and random s,

$$\Pr_{s\sim\{0,1\}^m}\left[\left|\sum_{y\sim\{0,1\}^\ell} H_c(y)(-1)^{f(c\circ y)+G(s)_{c\circ y}}\right| \geq \sqrt{\frac{\sum H(x)^2}{3}}\right] \geq \frac{2}{11}.$$

Letting $I_c(s)$ denote the indicator of the above event, it follows that

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

and therefore there exists a choice of s such that $\Pr_{c\sim H'}[I_c(s) = 1] \geq 2/11$. Fix such s. 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 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

$$\begin{aligned}
\mathbb{E}_{x\sim H}[(-1)^{f(x)+g(x)}] &= \mathbb{E}_{c\sim H'}\left[\sum_{y\in\{0,1\}^\ell} H_c(y)(-1)^{f(c\circ y)+g(c\circ y)} =\right] \\
&= \mathbb{E}_{c\sim H'}\left[(-1)^{h(c)} \sum_{y\in\{0,1\}^{\log(1/\gamma^2)}} H_c(y)(-1)^{f(c\circ y)+G(s)_{c\circ y}}\right] \\
&= \mathbb{E}_{c\sim H'}\left[\left|\sum_{y\in\{0,1\}^\ell} H_c(y)(-1)^{f(c\circ y)+G(s)_y}\right|\right] \\
&\geq \frac{2}{11}\cdot\sqrt{\frac{\sum_{y\in\{0,1\}^\ell} H_c(y)^2}{3}} \\
&\geq \frac{2}{11}\sqrt{\frac{(1/2^\ell)\sum H_c(y)}{3}} \\
&\geq \frac{\gamma}{9},
\end{aligned}$$

where the penultimate inequality is an application of Cauchy-Schwartz.

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

□

We note here that Lemma 1 and Theorem 7 together imply Theorem 4. However, we will prove the main theorem below without depending on Theorem 7.

Proof of Theorem 3. Pick a function $g : \{0,1\}^k \to \{0,1\}$ uniformly at random, where k will be chosen soon. We then define our function $f(x) \coloneqq g(x_{\leq k})$ to be g on the first k bits of the input. By a Chernoff bound, we know that for a fixed circuit C and $\delta < .49$,

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

Union bounding all circuits of size s (of which there are $s^{O(s)}$), and taking $k = \Theta(\log s/\log\log s)$, we notice with probability at most $s^{O(s)}2^{-\Omega(2^k)} < 1$, f can be $(1-\delta)$-approximated by a size s circuit. Fix an f that cannot.

Now consider arbitrary distribution H over $\{0,1\}^n$. The marginal distribution of H on the first k bits will be some distribution H' over $\{0,1\}^k$. By Lemma 1 and Theorem 6, there exists 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

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

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 $s \geq 1/\gamma^{2+\varepsilon}$, the desired result follows. □

Acknowledgements

We thank Guy Blanc and David Zuckerman for references and illuminating discussions.

References

[ACR97] Alexander E. Andreev, Andrea E.F. Clementi, and JoséD.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. 3, 5, 9, 10]

[Ber97] Bonnie Berger. The fourth moment method. *SIAM Journal on Computing*, 1997. doi:10.1137/S0097539792240005. [pp. 3, 5]

[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. [p. 1]

[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. [p. 2]

[CDV24] Sílvia Casacuberta, Cynthia Dwork, and Salil Vadhan. Complexity-theoretic implications of multicalibration. In *STOC*, 2024. doi:10.1145/3618260.3649748. [p. 1]

[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. 1]

[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. [pp. 4, 5]

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

[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. [p. 1]

[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]

[O'D02] Ryan O'Donnell. Hardness amplification within NP. In *STOC*, 2002. doi:10.1145/509907.510015. [p. 1]

[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. 3]

[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. 1]

[Spi96] Daniel A. Spielman. Linear-time encodable and decodable error-correcting codes. *IEEE Transactions on Information Theory*, 1996. doi:10.1109/18.556668. [pp. 3, 10]

[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. 1]

[Tre05] Luca Trevisan. On uniform amplification of hardness in np. In *STOC*, 2005. doi:10.1145/1060590.1060595. [p. 1]

[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. [p. 3]

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

In this section, we will provide an alternate exposition of Theorem 7, the 4-wise uniform generator constructed in [ACR97]. The core idea is the same as that of [ACR97], but we will use modern tools to help modularize and simplify the construction and its analysis.

The starting point of the construction is to consider the simpler task of a 2-wise uniform generator. There is a classic pairwise uniform generator mapping a nonzero seed of length $\log n$ to a string of size $n-1$, which is to simply output all nonempty $\mathbb{F}_2$-linear combinations of the seed. That is, $G(s) := (\langle s, r\rangle)_{r \in \mathbb{F}_2^{\log n} \setminus \{0\}}$. Notice for a fixed s, the output of the r'th bit as a function of r is simply some parity of a subset of bits in r, which is trivially a circuit of size $O(\log n)$ as desired.

Now 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

$$G(s)_x + G(s)_y = \langle s, x\rangle + \langle s, y\rangle = \langle s, x+y\rangle = G(s)_{x+y}.$$

Hence, we do not even have 3-wise uniformity, as the bits in indices $x, y, x+y$ are correlated. This motivates the following idea: what if we only focused on a subset of indices $X \subset \{0,1\}^n$ such that all distinct $x_1, x_2, x_3, x_4 \in S$ are linearly independent. Can we show that $G(s)_{x \in X}$ is a 4-wise independent string? Yes.

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

Proof. Consider arbitrary $Y \subset X$ of size 4. We will show over a uniform s, the string $(\langle s, y\rangle)_{y \in Y}$ is uniform. Notice this string is simply $M \cdot s$, where M is a $\mathbb{F}_2^{4 \times n}$ matrix whose rows are the elements of Y. 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 Y 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 $h : \{0,1\}^n \to \{0,1\}^{Cn}$ such that for all distinct $x_1, \dots, x_4$, the vectors $h(x_1), h(x_2), h(x_3), h(x_4)$ are linearly independent (i.e. we want a code of constant rate whose dual has distance ≥ 5 and is encodable by a linear-size circuit). 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 just try to scatter the x's randomly among the space $\{0,1\}^{4n}$. This actually works, because the probability for a fixed $a \leq 4$ and $x_1, \dots, x_a \in \mathbb{F}_2^n$, the probability $h(x_1) + \cdots + h(x_a) = 0$ is 2^{-16n}. Union bounding 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 constant many bits of x?

More concretely, say we pick subsets $S_1, \dots, S_{16n} \subset [n]$ of constant size uniformly at random, and then let $g_i : \{0,1\}^{S_i} \to \{0,1\}$ be a random function. Set $h(x) = (g_1(x), \dots, g_{16n}(x))$. This is of linear circuit size, and we are hoping the randomness of the g_i keeps vectors linearly independent. Fix $x_1, \dots, x_a$ for $a \leq 4$. We want the probability that $g_i(x_1) + \cdots + g_i(x_a) = 0$ for all i to be at most $1/\binom{2^n}{\leq 4}$. Unfortunately, this need not be true. Say $x_1, \dots, x_4$ 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, $g_i(x_j)$ will be guaranteed to all be the same, and so $g_i(x_1) + \cdots g_i(x_4) = 0$. In other words, the obstacle is that a randomly picked 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 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 will indeed detect a difference. This will allow the randomness of g_i to prevent dependencies from happening.

But are there asymptotically good codes that are encodable in linear circuit size? Indeed there are: Spielman codes.

Theorem 8 ([Spi96])**.** *For any n there exists a small enough constant δ_0 such that there exists a linear error correcting code of block length n, rate $k \geq n/4$, and distance $\delta_0 > 0$. Furthermore, there exists a linear size circuit $C : \{0,1\}^k \to \{0,1\}^n$ such that the image of C is the code.*

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

Theorem 9 ([ACR97])**.** *There exists a 4-wise uniform generator $G : \{0,1\}^{4n} \setminus \{0^{4n}\} \to \{0,1\}^{2^n}$ such that for each fixing of seed $x \in \{0,1\}^m \setminus \{0^{4n}\}$, there is a linear size circuit $C_s : \{0,1\}^n \to \{0,1\}$ such that $C_x(i) = G(x)_i$.*

Proof. Let $\mathsf{Enc} : \{0,1\}^n \to \{0,1\}^{4n}$ be a linear circuit that encodes an asymptotically good code with constant distance δ, as implied by Theorem 8. We will show that there exists sets $S_1, \ldots S_{16n} \subset [4n]$, each of size $10/\delta$, and functions $g_i : \{0,1\}^{S_i} \to \{0,1\}$ such that $G(s) := \langle (g_1(\mathsf{Enc}(x)), \ldots, g_{16n}(\mathsf{Enc}(x))), s \rangle$ is a 4-wise uniform generator.

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, some i has $\sum_{x \in X} g_i(x) \neq 0$. 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.

We do this by the probabilistic method. In particular, we will pick each S_i by selecting $10/\delta$ elements uniformly and independently from $[4n]$, 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.

If $2 \leq |X| \leq 4$, the strings $\{\mathsf{Enc}(x)\}_{x \in X}$ will have pairwise relative distance at least δ. Therefore, for a fixed $i \in [4n]$ and $x \neq x' \in X$, the probability a random S_i satisfies $(x)_{S_i} = (x')_{S_i}$ is at most $(1-\delta)^{|S_i|}$. Hence, the probability $(x)_{S_i} \neq (x')_{S_i}$ for all $x \neq x' \in X$ is at least

$$1 - \binom{5}{2}(1-\delta)^{|S_i|} = 1 - 10e^{-10} \geq \frac{99}{100}.$$

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

If $X = \{x\}$, then we note for any fixing of $\{S_i\}_{i \in [4n]}$, and for random $\{g_i\}$, $(g_i(x_{S_i}))_{i \in [4n]}$ is a uniformly random vector, and is consequently 0 with probability $\leq 2^{-16n} < (2/3)^{16n}$.

Thus by a union bound, all X are linearly independent with probability

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

implying the existence of such S_i, 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 on Lemma 1. □