Electronic Colloquium on Computational Complexity, Report No. 58 (2025)

Generalised Linial–Nisan Conjecture is False for DNFs

Yaroslav Alekseev *Technion* Mika Göös *EPFL* Ziyi Guan *EPFL* Gilbert Maystre *EPFL*
Artur Riazanov *EPFL* Dmitry Sokolov *EPFL* Weiqiang Yuan *EPFL*

May 5, 2025

Abstract

Aaronson (STOC 2010) conjectured that *almost k-wise independence* fools constant-depth circuits; he called this the *generalised Linial–Nisan conjecture*. Aaronson himself later found a counterexample for depth-3 circuits. We give here an improved counterexample for depth-2 circuits (DNFs). This shows, for instance, that Bazzi's celebrated result (k-wise independence fools DNFs) cannot be generalised in a natural way. We also propose a way to circumvent our counterexample: We define a new notion of pseudorandomness called *local couplings* and show that it fools DNFs and even decision lists.

1 Introduction

Linial and Nisan [LN90] conjectured that "k-wise independent" distributions fool constant-depth circuits (class AC^0). More specifically, a distribution $\mathcal{D}$ over $\{0,1\}^n$ is called *k-independent* if the marginal distribution on every k-sized subset of bits is uniform. We say that $\mathcal{D}$ *δ-fools* a circuit C if the circuit cannot distinguish $\mathcal{D}$ from the uniform distribution on $\{0,1\}^n$:

$$\left| \Pr_{\boldsymbol{x}\sim\mathcal{D}}[C(\boldsymbol{x})=1] - \Pr_{\boldsymbol{x}\sim\{0,1\}^n}[C(\boldsymbol{x})=1] \right| \;\le\; \delta.$$

The Linial–Nisan conjecture was first proved for depth-2 circuits (DNFs and CNFs) by Bazzi [Baz09] (with a simplification by Razborov [Raz09]) and then for every AC^0-circuit by Braverman [Bra11]. Indeed, Braverman showed that every size-s AC^0-circuit is $o(1)$-fooled by $\operatorname{poly}(\log s)$-independence.

Aaronson [Aar10] asked whether the Linial–Nisan conjecture could be strengthened to hold also for "almost k-wise independence," a seemingly modest generalisation. We say that a distribution $\mathcal{D}$ over $\{0,1\}^n$ is *(ε,k)-independent* if for every subset $I\subseteq[n]$, $|I|=k$, the marginal distribution on the bits in I is multiplicatively close to uniform in the sense that for every $\alpha\in\{0,1\}^I$,

$$(1-\varepsilon)2^{-k} \;\le\; \Pr_{\boldsymbol{x}\sim\mathcal{D}}[\boldsymbol{x}_I=\alpha] \;\le\; (1+\varepsilon)2^{-k}.$$

Generalised Linial–Nisan Conjecture (GLN). *Let $\mathcal{D}$ be a $(1/n^{\Omega(1)}, n^{\Omega(1)})$-independent distribution over $\{0,1\}^n$. Then $\mathcal{D}$ $o(1)$-fools every AC^0-circuit of size $2^{n^{o(1)}}$.*

ISSN 1433-8092

Aaronson's original motivation for this conjecture was to resolve a problem in quantum complexity theory. He showed that a positive resolution of GLN would imply the separation $\mathsf{BQP} \not\subseteq \mathsf{PH}$ relative to an oracle. (This separation was subsequently proved by Raz and Tal [RT19] by a different approach.) Later, Aaronson himself found a counterexample to GLN for depth-3 circuits [Aar11], but he still re-posed the conjecture (and thought it "plausible") for depth-2 circuits. Our main result here is to refute the GLN conjecture in this remaining case.

Theorem 1 (Main result). *There exists a $(1/n^{\Omega(1)}, n^{\Omega(1)})$-independent distribution $\mathcal{D}$ over $\{0,1\}^n$ and a $O(\log^3 n)$-width DNF formula F such that*

$$\Pr_{\boldsymbol{x}\sim\mathcal{D}}[F(\boldsymbol{x}) = 1] - \Pr_{\boldsymbol{x}\sim\{0,1\}^n}[F(\boldsymbol{x}) = 1] \;\geq\; \Omega(1).$$

Let us make two notes about the parameters here. First, our formula F will have quasi-polynomial size, whereas Aaronson's depth-3 counterexample has only polynomial size; hence his example achieves slightly better parameters (at the cost of larger depth). Second, our construction can be varied to produce the following tradeoff: by increasing the DNF width to any $w \leq n^{o(1)}$, we can make the distribution $(\exp(-w^{\Omega(1)}), n^{\Omega(1)})$-independent (see Section 2.5).

1.1 Implications and related work

One consequence of the failure of the GLN conjecture is to the construction of *pseudorandom generators* (PRGs) for DNFs. It is known that for $k \geq \Omega(\log n)$ there exist $(o(1), k)$-independent distributions with support size $2^{O(k)}$ [NN93, AGHP92], which is smaller than $n^{\Omega(k)}$ that is required for truly k-wise independent distributions [CGH+85]. Thus Theorem 1 rules out a natural approach ("output an almost k-wise independent distribution") to improving the seed length of PRGs. For the current state-of-the-art PRGs for DNFs, see [DETT10, Tal17, Lyu22]; see also the survey [HH24].

Another lesson from Theorem 1 is to the further development of circuit lower bound methods. We find it important to seek alternative proofs of central theorems such as Bazzi's [Baz09, Raz09] and its extensions [Bra11]. The existing proofs use the *polynomial method* to approximate a DNF with a low-degree polynomial. Is there a more "combinatorial" proof of Bazzi's theorem? One such more combinatorial approach is the *top-down* lower bound method [HJP93, GRSS23], which often uses entropy-based arguments to analyse circuits. We interpret the failure of GLN as a challenge to such top-down methods. While the method is in a formal sense *complete* (it can prove any lower bound that is true), the typical entropy counting arguments have a hard time distinguishing almost k-wise independent distributions from truly k-wise independent ones, suggesting that any top-down proof of Bazzi's theorem would require substantially new ideas.

Finally, we mention that—besides (almost) k-wise independence—several other notions of pseudorandomness have been considered in the literature [BIVW16, BDF+22, Hoz25].

1.2 Workaround: Local couplings

To complement our main result, we also propose a way to circumvent the failure of GLN by proposing a new notion of pseudorandomness called *local couplings*. This notion is useful for fooling depth-2 circuit models, but not depth-3 models; in particular, we show the following claims:

- *(C1)* Local couplings fool DNFs (query complexity analogue of NP).
- *(C2)* Local couplings fool decision lists (query complexity analogue of P^{NP}).

(C3) Local couplings do *not* fool depth-3 circuits (query complexity analogue of $\Sigma_2\mathsf{P}$).

Definition 2 (Local couplings). A pair of jointly distributed random variables $(\boldsymbol{x}, \boldsymbol{y}) \in (\{0,1\}^n)^2$ is an ε-*semi-coupling* if for every $y \in \mathrm{supp}(\boldsymbol{y})$ and $i \in [n]$,

$$\Pr[\boldsymbol{x}_i \neq \boldsymbol{y}_i \mid \boldsymbol{y} = y] \leq \varepsilon.$$

We say that $(\boldsymbol{x}, \boldsymbol{y})$ is an ε-*coupling* if both $(\boldsymbol{x}, \boldsymbol{y})$ and $(\boldsymbol{y}, \boldsymbol{x})$ are ε-semi-couplings.

The notion of a local coupling was somewhat implicit in Aaronson's analysis [Aar11] of his depth-3 counterexample. Local couplings are also a stronger variant of a notion proposed by Zhandry [Zha25] that he called "substitution distance."

Claims *(C1)–(C2)*. A width-k *decision list* is a sequence of pairs $\{(T_i, a_i)\}_{i\in[m]}$ where T_i are k-terms (conjunctions of at most k literals) and $a_i \in \{0,1\}$ are output values. A decision list defines $f \colon \{0,1\}^n \to \{0,1\}$ as follows: $f(x) = a_i$ where $i = \min\{i \in [m] \mid T_i(x) = 1\}$. The decision list width of a function is polynomially equivalent to the number of DNF queries necessary to compute the function [GKPW19, Appendix A]. In other words it is indeed a query complexity analogue of P^{NP}. The following theorem (Section 3.1) formalises *(C1)–(C2)* when $\boldsymbol{y}$ is uniformly distributed.

Theorem 3. *Let f be computed by a width-k decision list. For any ε-coupling $(\boldsymbol{x}, \boldsymbol{y})$,*

$$\Pr[f(\boldsymbol{x}) \neq f(\boldsymbol{y})] \leq 2k\varepsilon.$$

Claim *(C3)*. Aaronson's [Aar11] original counterexample involved a distribution $\mathcal{D}$ related to a certain *surjectivity* function, which can be computed by a small depth-3 circuit. We observe (Section 3.2) that Aaronson's distribution $\mathcal{D}$ can indeed be locally coupled with the uniform distribution, which implies that local couplings do not fool depth-3 circuits (Claim *(C3)*). We can furthermore conclude (using Theorem 3) that $\mathcal{D}$ fools decision lists—this claim was already made earlier by Aaronson [Aar11, Theorem 3], but his proof contained a mistake,[1] which we can now fix with the notion of local couplings. Finally, we also show (Section 3.3) that an ε-semi-coupling is not enough by itself to fool DNFs—one truly needs the two-sided condition of an ε-coupling.

2 Counterexample

In this section, we prove Theorem 1 by constructing a DNF formula F and an associated almost k-independent distribution $\mathcal{D}$ that F can distinguish from uniform. We first (Section 2.1) construct a weak example that distinguishes $\mathcal{D}$ from uniform with advantage $1/\mathrm{poly}(\log n)$. Then (Section 2.4) we amplify this advantage to $\Omega(1)$ by using a standard majority trick.

[1] The mistake is acknowledged on the author's homepage. An implication of this result would have been to show the separation $\Pi_2\mathsf{P} \neq \mathsf{P}^{\mathsf{NP}}$ relative to a random oracle. That result however follows (using a function different from surjectivity) from the more recent result that PH is infinite in the random oracle model [HRST17].

2.1 Construction

Our starting point is the *address* function $\textsc{Addr}\colon \{0,1\}^m \times \{0,1\}^{2^m} \to \{0,1\}$ defined as $\textsc{Addr}(a,p) \coloneqq p_a$. Here we write p_a to mean $p_{\mathrm{int}(a)}$ where $\mathrm{int}(a) \in [2^m]$ is the integer corresponding naturally to the bitstring a. Let us first observe that $\textsc{Addr}$ together with the uniform distribution over $\textsc{Addr}^{-1}(1)$ "almost works" as the counterexample in Theorem 1. For $(\boldsymbol{a},\boldsymbol{p}) \sim \textsc{Addr}^{-1}(1)$ the distribution of $\boldsymbol{p}$ is already $(o(1), 2^{\Omega(m)})$-independent. The reason the whole $(\boldsymbol{a},\boldsymbol{p})$ does not have the same property is that for example, fixing all bits of $\boldsymbol{a}$ to some a forces $\boldsymbol{p}_a = 1$, so for some $I \subseteq [m+2^m]$ containing all bits describing $\boldsymbol{a}$ and the bit $\boldsymbol{p}_a$ the settings of $(\boldsymbol{a},\boldsymbol{p})_I$ with $\boldsymbol{p}_a = 0$ have probability zero.

To avoid this issue we hide the bits of the address using the usual *tribes* function:

$$\textsc{Tribes}(A) \coloneqq \bigvee_{j\in[r]} \bigwedge_{k\in[m]} A_{k,j}.$$

The input here is an $m \times r$ boolean matrix and the function returns 1 iff the matrix contains an all-1 column. It is well-known [O'D14, §4.2] that if we choose $r \coloneqq \lceil 2^m \ln 2 \rceil$, the function becomes *balanced*, meaning that $\Pr_{\boldsymbol{A}\sim\{0,1\}^{m\times r}}[\textsc{Tribes}(\boldsymbol{A})] = 1/2 + o(1)$.

A natural attempt to define a counterexample would be to consider the distinguishing function $\textsc{Addr}((\textsc{Tribes}(A^1),\ldots,\textsc{Tribes}(A^m)),p)$. This does not work since this function requires polynomial DNF width as the negation of $\textsc{Tribes}$ reduces to it. We fix this by replacing the $\textsc{Addr}$ function by its *monotone* version: $\textsc{mAddr}\colon \{0,1\}^m \times \{0,1\}^{2^m} \to \{0,1\}$ is defined as

$$\textsc{mAddr}(a,p) \coloneqq \begin{cases} 0 & \text{if } |a| < m/2 \\ p_a & \text{if } |a| = m/2 \\ 1 & \text{if } |a| > m/2 \end{cases}, \quad \text{where } |a| \text{ is the Hamming weight of } a.$$

We are now ready to define our function $f\colon (\{0,1\}^{m\times r})^m \times \{0,1\}^{2^m} \to \{0,1\}$ by (see also Figure 1)

$$f(A^1,\ldots,A^m,p) \coloneqq \textsc{mAddr}((\textsc{Tribes}(A^1),\ldots,\textsc{Tribes}(A^m)),p), \tag{1}$$

Note that input size of f is $n \coloneqq m^2 r + 2^m$. The following constructs a narrow DNF for f.

Claim 4. *There exists a* DNF F *of width* $O(\log^2 n)$ *that computes* f*.*

Proof. A DNF is commonly viewed as a collection of 1-*certificates*: f is computable by a k-DNF iff for each point $x \in f^{-1}(1)$ there exists a certificate comprised of a subset of input bits $I \subseteq [n]$ of size k and $\alpha \in \{0,1\}^I$ such that $x'_I = \alpha$ implies $f(x') = 1$. Hence it is enough to provide a certificate of width $O(m^2) = O(\log^2 n)$ for each 1-input of f. Consider a 1-input $x = (A^1,\ldots,A^m,p)$ and let $a \coloneqq (\textsc{Tribes}(A^1),\ldots,\textsc{Tribes}(A^m))$. If $|a| > m/2$, a 1-certificate is simply a set of matrices $H \subseteq [m]$ of size $|H| = m/2+1$ together with a 1^m-column in each of those matrices. Such certificates fix $(m/2+1)\cdot m$ variables. A similar idea can certify 1-inputs with $|a| = m/2$, at the cost of adding the corresponding bit p_a. □

We now define $\mathcal{D}$ as a distribution of the random variable $\boldsymbol{x}$ defined below.

Definition 5. Let $\boldsymbol{x} = (\boldsymbol{A}^1,\boldsymbol{A}^2,\ldots,\boldsymbol{A}^m,\boldsymbol{p})$ over $\{0,1\}^n$ be sampled as follows:

1. Sample $\boldsymbol{A}^i \sim \{0,1\}^{m\times r}$ uniformly and independently for each $i \in [m]$.
2. Sample $\boldsymbol{p} \sim \{0,1\}^{2^m}$ uniformly and independently.

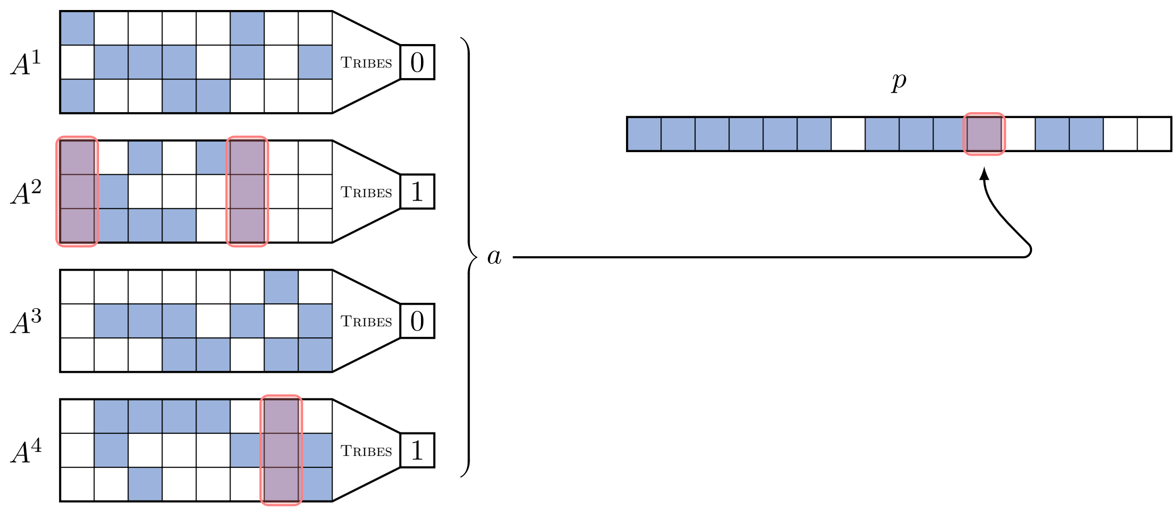

Figure 1: Illustration of $\text{mAddr}(\text{Tribes}(A^1), \ldots, \text{Tribes}(A^4), p)$. Blue cells correspond to 1-input bits, white cells correspond to 0-input bits. The address $a = (\text{Tribes}(A^1), \ldots, \text{Tribes}(A^4))$ is $(0, 1, 0, 1)$, so it satisfies $|a| = 4/2$. Hence the function outputs $p_{\text{int}(a)} = p_{11} = 1$.

3. Let $\boldsymbol{a} = \big(\text{Tribes}(\boldsymbol{A}^1), \ldots, \text{Tribes}(\boldsymbol{A}^m)\big)$; If $|\boldsymbol{a}| = m/2$, fix $\boldsymbol{p}_{\boldsymbol{a}} = 1$.

We show that $\mathcal{D}$ is $(n^{-1/5}, n^{1/5})$-independent, yet f distinguishes $\mathcal{D}$ from the uniform distribution, which together with Claim 4 implies the following weaker version of Theorem 1:

Lemma 6. *The distribution $\mathcal{D}$ as in Definition 5 is $(n^{-1/5}, n^{1/5})$-independent, but there is an $O(\log^2 n)$-DNF F such that* $\Pr_{\boldsymbol{x}\sim\mathcal{D}}[F(\boldsymbol{x}) = 1] - \Pr_{\boldsymbol{x}\sim\{0,1\}^n}[F(\boldsymbol{x}) = 1] = \Omega(\log^{-1/2} n)$.

We reduce the proof of Lemma 6 to the following two lemmas:

Lemma 7. $\Pr_{\boldsymbol{x}\sim\mathcal{D}}[f(\boldsymbol{x}) = 1] - \Pr_{\boldsymbol{x}\sim\{0,1\}^n}[f(\boldsymbol{x}) = 1] = \Omega\big(1/\sqrt{m}\big)$, *where m is as in Definition 5.*

Lemma 8. *The distribution $\mathcal{D}$ as in Definition 5 is (ε, k)-independent for $k \leq 2^m$, $\varepsilon \leq k2^{-m/2+1}$.*

Proof of Lemma 6. $\mathcal{D}$ is distributed over $\{0,1\}^n$ where $n \coloneqq m^2\lceil 2^m \ln 2\rceil + 2^m$. Apply Lemma 8 with $k = n^{1/5}$ and $\varepsilon = n^{1/5}2^{-m/2+1} \ll n^{-1/5}$. Then by Lemma 7 and Claim 4 there exists $O(m^2) = O(\log^2 n)$-DNF that $\Omega(1/\sqrt{m}) = \Omega(1/\sqrt{\log n})$-distinguishes $\mathcal{D}$ from $\mathcal{U}$, where the latter is uniformly distributed over $\{0,1\}^n$. □

2.2 Proof of Lemma 7

Let $\boldsymbol{x} \coloneqq (\boldsymbol{A}^1, \ldots, \boldsymbol{A}^m, \boldsymbol{p}) \sim \mathcal{D}$ and $\boldsymbol{y} \sim \{0,1\}^n$. Since the matrices $\boldsymbol{A}^i$ are uniformly generated, it is possible to couple $\boldsymbol{x}$ and $\boldsymbol{y}$ by defining $\boldsymbol{y} \coloneqq (\boldsymbol{A}^1, \ldots, \boldsymbol{A}^m, \boldsymbol{p}')$ where $\boldsymbol{p}' \sim \{0,1\}^{2^m}$. Note that the address part of each input coincides and in particular, they share the event $E \coloneqq$ "$|\boldsymbol{a}| = m/2$".

Observe that $\Pr[f(\boldsymbol{x}) = 1 \mid \neg E] = \Pr[f(\boldsymbol{y}) = 1 \mid \neg E]$ by the definition of $\mathcal{D}$: if $|\boldsymbol{a}| \neq m/2$ then Step (3) is not reached in the definition of $\mathcal{D}$ and $\boldsymbol{p}$ is uniform. On the other hand we have

$\Pr[F(\boldsymbol{x}) = 1 \mid E] = 1$. Indeed, if E holds, we have $F(\boldsymbol{x}) = \boldsymbol{p_a} = 1$ by the definition of f and $\mathcal{D}$.

$$\begin{aligned}\Pr[F(\boldsymbol{y}) = 1 \mid E] &= \Pr[\boldsymbol{p}'_{\boldsymbol{a}} = 1 \mid E] \\ &= \sum_{a\in\{0,1\}^m} \Pr[\boldsymbol{p}'_a = 1 \mid E \wedge \boldsymbol{a} = a] \Pr[\boldsymbol{a} = a \mid E] \\ &= \sum_{a\in\{0,1\}^m} \Pr[\boldsymbol{p}'_a = 1] \Pr[\boldsymbol{a} = a \mid E] = \frac{1}{2}.\end{aligned}$$

Thus, $\Pr[f(\boldsymbol{x}) = 1] - \Pr[f(\boldsymbol{y}) = 1] = \Pr[E]/2$ and so it remains to bound $\Pr[E]$. For that, we need the following simple fact:

Lemma 9. *Let $\boldsymbol{x}$ distributed over $\{0,1\}^n$ according to a product distribution such that $|\Pr[\boldsymbol{x}_i = 1] - 1/2| \le \varepsilon$. Then $\Delta(\boldsymbol{x}, \boldsymbol{u}) \coloneqq \max_{E\subseteq\{0,1\}^n} \left|\Pr[\boldsymbol{x} \in E] - \Pr[\boldsymbol{u} \in E]\right| \le 2n\varepsilon$, where $\boldsymbol{u} \sim \{0,1\}^n$.*

Proof. Let us couple $\boldsymbol{x}$ with $\boldsymbol{u}$ as follows: suppose $\Pr[\boldsymbol{x}_i = 1] = 1/2 + p$. We then set $\boldsymbol{u}_i \coloneqq \boldsymbol{x}_i$ with probability $1/(1 + 2|p|)$ and $\boldsymbol{u}_i \coloneqq [\![p > 0]\!] \coloneqq (1 \text{ if } p > 0 \text{ otherwise } 0)$ with probability $1-1/(1+2|p|) \le 2|p| \le 2\varepsilon$. Then $\Pr[\boldsymbol{U}_i = 1] = (1/2+p)/(1+2|p|) + [\![p > 0]\!](1-1/(1+2|p|)) = 1/2$, so $\boldsymbol{u}$ is indeed uniformly distributed. Then $\Pr[\boldsymbol{x} \neq \boldsymbol{u}] \le 2n\varepsilon$, so $\Delta(\boldsymbol{x}, \boldsymbol{u}) \le 2n\varepsilon$. □

Note that each bit $\boldsymbol{a}_i$ is close to being balanced:

$$\Pr[\boldsymbol{a}_i = 1] = 1 - (1 - 2^{-m})^r = 1 - (1/e + \Theta(2^{-m}))^{\ln 2} = 1/2 + \Theta(2^{-m}).$$

As all $\boldsymbol{a}_i$ are independent, we can use Lemma 9 to get sharp bounds on their sum being exactly $m/2$: $\Pr[E] \ge \Pr_{\boldsymbol{x}\sim\{0,1\}^m}[|\boldsymbol{x}| = m/2] - \Theta(m \cdot 2^{-m}) = \Omega(1/\sqrt{m})$.

2.3 Proof of Lemma 8

We need to show that for every $I \subseteq [n]$ of size k and for every $\alpha \in \{0,1\}^I$ we have $(1-\varepsilon)\cdot 2^{-k} \le \Pr[\boldsymbol{x}_I = \alpha] \le (1+\varepsilon)\cdot 2^{-k}$. We now classify the bits of I and α. Let $I_i \subseteq [m] \times [r]$ for $i \in [m]$ be the set of bits of $\boldsymbol{A}^i$ in I. Let $J \subseteq \{0,1\}^m$ be the set of bit indices of $\boldsymbol{p}$ that belong to I (we identify the indices with their bit representations). Let $\alpha^i \in \{0,1\}^{I_i}$ and $\beta \in \{0,1\}^J$ be the corresponding parts of α.

Since $\boldsymbol{A}^1, \dots, \boldsymbol{A}^m$ are uniformly distributed it suffices to show that

$$(1-\varepsilon)2^{-|J|} \le \Pr[\boldsymbol{p}_J = \beta \mid \forall i \in [m] : \boldsymbol{A}^i_{I_i} = \alpha^i] \le (1+\varepsilon)2^{-|J|}.$$

Let $J^{m/2} \coloneqq \{s \in J \mid |s| = m/2\}$. Intuitively the only non-uniformity in $\boldsymbol{x}_I$ is introduced when $\boldsymbol{a} \in J^{m/2}$ as this is the only case where $\boldsymbol{p}$ is changed from uniform. We make this intuition precise in the following claim.

Claim 10. *For any event E that is a function of $\boldsymbol{A}^1, \dots, \boldsymbol{A}^m$ we have*

$$(1 - \Pr[\boldsymbol{a} \in J^{m/2} \mid E])2^{-|J|} \le \Pr[\boldsymbol{p}_J = \beta \mid E] \le (1 + \Pr[\boldsymbol{a} \in J^{m/2} \mid E])2^{-|J|}.$$

Proof. Let $J_i \coloneqq \{j \in J^{m/2} \mid \beta_j = i\}$ for $i \in \{0,1\}$. By the total probability law we get

$$\begin{aligned}\Pr[\boldsymbol{p}_J = \beta \mid E] &= \Pr[\boldsymbol{p}_J = \beta \mid E \wedge \boldsymbol{a} \in J_0] \Pr[\boldsymbol{a} \in J_0 \mid E] \\ &\quad + \Pr[\boldsymbol{p}_J = \beta \mid E \wedge \boldsymbol{a} \in J_1] \Pr[\boldsymbol{a} \in J_1 \mid E] \\ &\quad + 2^{-|J|} \Pr[\boldsymbol{a} \notin J^{m/2} \mid E] \qquad (2) \\ &= 0 + 2^{-(|J|-1)} \Pr[\boldsymbol{a} \in J_1 \mid E] + 2^{-|J|} \Pr[\boldsymbol{a} \notin J^{m/2} \mid E] \qquad (3) \\ &= 2^{-|J|}(\Pr[\boldsymbol{a} \notin J^{m/2} \mid E] + 2\Pr[\boldsymbol{a} \in J_1 \mid E]). \qquad (4)\end{aligned}$$

In (2) and (3) we use that given $\boldsymbol{a}$ the event E is independent from $\boldsymbol{p}$. Since (4) is minimized when $J_1 = \emptyset$ and maximized when $J_1 = J^{m/2}$, we have the claim. $\square$

Now let E be the event "$\forall i \in [m]\colon \boldsymbol{A}^i_{I_i} = \alpha^i$". Let us compute $\Pr[\boldsymbol{a} = s \mid E]$ for $s \in \{0,1\}^m$. Since $s \in J^{m/2}$ we have $|s| = m/2$, wlog let $s = 0^{m/2}1^{m/2}$. Since the bits of $\boldsymbol{a}$ denoted by $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_m$ are independent and E is a conjunction of independent events we have

$$\begin{aligned}\Pr[\boldsymbol{a} = s \mid E] &= \prod_{\ell \in [m/2]} \Pr[\boldsymbol{a}_\ell = 0 \mid E] \cdot \prod_{\ell \in [m] \setminus [m/2]} \Pr[\boldsymbol{a}_\ell = 1 \mid E] \\ &\le \prod_{\ell \in [m/2]} \Pr[\boldsymbol{a}_\ell = 0 \mid \boldsymbol{A}^\ell_{I_\ell} = \alpha^\ell]\end{aligned}$$

Let us fix $\ell \in [m/2]$ and bound $\Pr[\boldsymbol{a}_\ell = 0 \mid \boldsymbol{A}^\ell_{I_\ell} = \alpha_\ell]$. By definition $\boldsymbol{a}_\ell = \textsc{Tribes}(\boldsymbol{A}^\ell)$, so it equals 0 iff no column of $\boldsymbol{A}^\ell$ is all-1, in particular all columns that do not contain bits of I_ℓ must not be all-1. For each of these columns the probability that it is not all-1 is $1 - 2^{-m}$. Since there are at least $\lceil 2^m \ln 2 \rceil - |I_\ell|$ such columns we get

$$\begin{aligned}\Pr[\boldsymbol{a} = s \mid E] &\le \prod_{\ell \in [m/2]} (1 - 2^{-m})^{\lceil 2^m \ln 2 \rceil - |I_\ell|} \\ &= (1 - 2^{-m})^{m/2 \cdot \lceil 2^m \ln 2 \rceil} (1 - 2^{-m})^{-\sum_{\ell \in [m/2]} |I_\ell|} \\ &\le 2^{-m/2} (1 - 2^{-m})^{-k} \\ &\le 2^{-m/2+1}\end{aligned}$$

Thus, $\Pr[\boldsymbol{a} \notin J^{m/2} \mid E] \le |J| 2^{-m/2+1} = k 2^{-m/2+1}$, so we conclude the proof by Claim 10.

2.4 Amplification

In this section we reduce Theorem 1 to Lemma 6. The construction is a simple variation of the majority vote of several instances of f. We prove that our construction indeed amplifies the distinguishing probability in the following lemma.

Lemma 11. *Suppose $\boldsymbol{x}$ is distributed over $\{0,1\}^n$ and there exists a function $g\colon \{0,1\}^n \to \{0,1\}$ such that*

$$\Pr[g(\boldsymbol{x}) = 1] - \Pr_{\boldsymbol{u} \sim \{0,1\}^n}[g(\boldsymbol{u}) = 1] \ge \delta,$$

for some δ depending on n. Let $\alpha = (\Pr[g(\boldsymbol{x}) = 1] + \Pr[g(\boldsymbol{u}) = 1])/2$. Then for $t = 2/\delta^2$ we have

$$\Pr\left[\sum_{i \in [t]} g(\boldsymbol{x}_i) \ge t \cdot \alpha\right] - \Pr\left[\sum_{i \in [t]} g(\boldsymbol{u}_i) \ge t \cdot \alpha\right] \ge \Omega(1),$$

where $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_t$ are independent samples of $\boldsymbol{x}$ and $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_t \sim \{0,1\}^n$.

Proof. Let $p_x = \mathbb{E}[g(\boldsymbol{x})]$. Since $\mathbb{E}\left[\sum_{i \in [t]} g(\boldsymbol{x}_i)\right] = t \cdot p_x$, we have by Hoeffding inequality,

$$\Pr\left[\sum_{i \in [t]} g(\boldsymbol{x}_i) \ge \alpha t\right] = 1 - \Pr\left[\sum_{i \in [t]} g(\boldsymbol{x}_i) < \alpha t\right] \ge 1 - e^{-2t^2(p_x - \alpha)^2/t} \ge 1 - e^{-(t\delta)^2/2t}.$$

Similarly, we can conclude that $\Pr\left[\sum_{i\in[t]} g(\boldsymbol{u}_i) \geq \alpha t\right] \leq e^{-(t\delta)^2/2t}$, hence,

$$\Pr\left[\sum_{i\in[t]} g(\boldsymbol{x}_i) \geq \alpha t\right] - \Pr\left[\sum_{i\in[t]} g(\boldsymbol{u}_i) \geq \alpha t\right] \geq 1 - 2e^{-t\delta^2/2}$$

With $t = 2/\delta^2$, we conclude the proof. □

We now need to show that a narrow DNF can check whether $\sum_{i\in[t]} f(x_i) \geq \alpha t$. In fact, this is true for any monotone function composed with a narrow DNF:

Lemma 12. *Let $f\colon \{0,1\}^n \to \{0,1\}$ be a function that can be computed by a ℓ-DNF D. Let $g\colon \{0,1\}^t \to \{0,1\}$ be a monotone function. Then $g \circ f^t(x_1,\ldots,x_t) \coloneqq g(f(x_1),\ldots,f(x_t))$ can be computed by a $t\ell$-DNF.*

Proof. Since f can be computed by a ℓ-DNF, a 1-certificate of f is a satisfying assignment for one term of D, which has size at most ℓ. Since g is monotone we can certify that $g \circ f^t(x_1,\ldots,x_t) = 1$ by giving a 1-certificate that $D(x_i) = 1$ for every $i \in [t]$ where that is the case. Such certificate has size at most $t\ell$, which implies that $g \circ f^t$ can be computed by a $t\ell$-DNF. □

Finally, we need to show that independent copies of an (ε, k)-independent distribution comprise an $(O(\varepsilon t), k)$-independent distribution:

Lemma 13. *If $\mathcal{D}$ is (ε, k)-independent, then the product distribution $\mathcal{D}^t$ is $(O(\varepsilon t), k)$-independent.*

Proof. Suppose $\boldsymbol{x} \sim \mathcal{D}$. Let $\boldsymbol{x}^t \sim \mathcal{D}^t$ be t independent copies of $\boldsymbol{x}$. Fix $I \in \binom{[n\cdot t]}{k}$ and $\alpha \in \{0,1\}^I$. For every $i \in [t]$, we define I_i and α_i to be the positions of I and α respectively in $\boldsymbol{x}_i$. Then,

$$\Pr[\boldsymbol{x}^t_I = \alpha] = \prod_{i\in[t]} \Pr[(\boldsymbol{x}_i)_{I_i} = \alpha_i] = \prod_{i\in[t]} \Pr[\boldsymbol{x}_{I_i} = \alpha_i].$$

Since $\boldsymbol{x}$ is (ε, k)-independent, for every $i \in [t]$, $(1-\varepsilon)\cdot 2^{-|I_i|} \leq \Pr[\boldsymbol{x}_{I_i} = \alpha_i] \leq (1+\varepsilon)\cdot 2^{-|I_i|}$. Hence, for small enough ε:

$$(1-2t\varepsilon)\cdot 2^{-k} \leq 2^{-\sum_{i\in[t]}|I_i|}\cdot(1-\varepsilon)^t \leq \Pr[\boldsymbol{x}^t_I = \alpha] \leq 2^{-\sum_{i\in[t]}|I_i|}\cdot(1+\varepsilon)^t \leq 2^{-k}\cdot(1+2t\varepsilon).$$ □

Proof of Theorem 1. Let s be a natural number to be fixed later. Let $\mathcal{D}$ be the $(s^{-1/5}, s^{1/5})$-independent distribution in Lemma 6. Let D be the $O(\log^2 s)$-DNF such that

$$\Pr_{\boldsymbol{x}\sim\mathcal{D}}[D(\boldsymbol{x}) = 1] - \Pr_{\boldsymbol{u}\sim\{0,1\}^s}[D(\boldsymbol{u}) = 1] = \Omega(1/\sqrt{\log m}).$$

From Lemma 13, for every t, $\mathcal{D}^t$ is $(O(t\cdot s^{-1/5}), s^{1/5})$-independent. By Lemma 11 for $\varphi(x_1,\ldots,x_t) \coloneqq [\![\sum_{i=1}^t D(x_i) \geq \alpha t]\!] \coloneqq (1$ if $\sum_{i=1}^t D(x_i) \geq \alpha t$, otherwise 0), when $t = O(\log s)$,

$$\Pr_{\boldsymbol{x}^t\sim\mathcal{D}^t}\left[\varphi(\boldsymbol{x}^t)\right] - \Pr_{\boldsymbol{u}^t\sim\{0,1\}^{st}}\left[\varphi(\boldsymbol{u}^t)\right] = \Omega(1).$$

Moreover, φ can be computed by a $O(t\cdot\log^2 s)$-DNF from Lemma 12. Choosing $t = O(\log s)$ and $t\cdot s = n$ we get that there exists a $(O(\log n\cdot n^{-1/5}), \Omega(n/\log n)^{1/5})$-independent distribution $\mathcal{D}^t$ over $\{0,1\}^n$ that can be $\Omega(1)$-distinguished from the uniform by a $O(\log^3 n)$-DNF, which implies the claim. □

2.5 Variation: Tradeoff between width and error

We finally sketch an extension of our construction that gives a tradeoff between DNF width and ε.

Theorem 14. *For any $w \geq \Omega(\log n)$ there exists a function $f_w\colon \{0,1\}^n \to \{0,1\}$ computable by a $w^{O(1)}$-DNF and an $(n^{-\Omega(w)}, n^{\Omega(1)})$-independent distribution $\mathcal{D}$ over $\{0,1\}^n$ such that*

$$\Pr_{\boldsymbol{x}\sim\mathcal{D}}[f_w(\boldsymbol{x})] - \Pr_{\boldsymbol{x}\sim\{0,1\}^n}[f_w(\boldsymbol{x})] \geq \Omega(1).$$

Proof sketch. We define a "monotone xor" of the functions $\textsc{Addr}$ as follows: $g\colon (\{0,1\}^m)^w \times (\{0,1\}^{2^m})^w \to \{0,1\}$ where $g(a^1,\ldots,a^w,p^1,\ldots,p^w) \coloneqq p^1_{a^1} \oplus \cdots \oplus p^w_{a^w}$ if $|a| = wm/2$, if $|a| \neq wm/2$ the value of g is 1 iff $|a| > wm/2$. The distinguisher f_w is then defined by hiding the bits of a in $\textsc{Tribes}$ instances:

$$f_w(A^1,\ldots,A^{mw},p^1,\ldots,p^w) \coloneqq g(\textsc{Tribes}(A^1),\ldots,\textsc{Tribes}(A^{mw}),p^1,\ldots,p^w).$$

We sample $\boldsymbol{x}$ from the distribution $\mathcal{D}$ in two steps: (1) Sample $\boldsymbol{x} = (\boldsymbol{A}^1,\ldots,\boldsymbol{A}^{mw},\boldsymbol{p}^1,\ldots,\boldsymbol{p}^w)$ uniformly at random. (2) If for $\boldsymbol{a} = \textsc{Tribes}^{mw}(\boldsymbol{A})$ it happens that $|\boldsymbol{a}| = wm/2$ and $g(\boldsymbol{a},\boldsymbol{p}) = 0$, we flip a random bit among $\boldsymbol{p}^1_{\boldsymbol{a}^1},\ldots,\boldsymbol{p}^w_{\boldsymbol{a}^w}$.

The $\Omega(1/\sqrt{mw})$-distinguishability of $\mathcal{D}$ from the uniform distribution by f_w is shown analogously to Lemma 7. Then according to Section 2.4 we increase the width of the DNF by the factor $O(mw)$ to get a $\Omega(1)$-distinguisher. The result then follows by choosing the appropriate constants in Ω and big-O.

Now we show the $(n^{-\Omega(w)}, n^{\Omega(1)})$-independence for f_w: analogously to Claim 10 one can show that to establish that $\mathcal{D}$ is $(O(\varepsilon),k)$-independent it suffices to bound $\Pr[\boldsymbol{a}^1 \in J_1 \wedge \cdots \wedge \boldsymbol{a}^w \in J_w \mid \boldsymbol{A}_I = \alpha]$ as $O(\varepsilon)$ for $J_1,\ldots,J_w \subseteq [2^m]$ and $I \subseteq ([m] \times [\lceil 2^m \ln 2\rceil])^{mw}$ such that $|J_1| + \cdots + |J_w| + |I| \leq k$. Now for every $j = (j_1,\ldots,j_w) \in J_1 \times \cdots \times J_w$ such that $|j| = mw/2$ we have analogously to Lemma 8 $\Pr[\boldsymbol{a} = j \mid \boldsymbol{A}_I = \alpha] \leq 2^{-mw/2+w}$ as long as $|I| \leq \lceil 2^m \ln 2\rceil$. Assuming that $|J| \leq k \leq 2^{m/4} = n^{\Omega(1)}$ we get that $\prod_{i\in[w]} |J_i| \leq 2^{mw/4}$ and therefore $\varepsilon \leq 2^{-mw/4+w} = n^{-\Omega(w)}$. □

3 Local couplings

3.1 Couplings fool decision lists: Proof of Theorem 3

Let $T_1,\ldots,T_M$ be the k-terms in the decision list defining f. It is sufficient to show that for $L(x) \coloneqq \min\{i \in [M] \mid T_i(x) = 1\}$ we have $\Pr[L(\boldsymbol{x}) \neq L(\boldsymbol{y})] \leq 2k\varepsilon$. We show that $\Pr[L(\boldsymbol{x}) \leq L(\boldsymbol{y})]$ and $\Pr[L(\boldsymbol{y}) \leq L(\boldsymbol{x})]$ are both high and conclude the statement from that. Let us show $\Pr[L(\boldsymbol{x}) \leq L(\boldsymbol{y})] \geq 1 - k\varepsilon$ using that $(\boldsymbol{x},\boldsymbol{y})$ is an ε-semi-coupling. Denoting $\mathrm{supp}(T_i) \subseteq [n]$ the set of input bits mentioned in the term T_i we write

$$\begin{aligned}
\Pr[L(\boldsymbol{x}) \leq L(\boldsymbol{y})] &= \sum_{i\in[N]} \Pr[L(\boldsymbol{x}) \leq i \mid L(\boldsymbol{y}) = i] \Pr[L(\boldsymbol{y}) = i] \\
&\geq \sum_{i\in[N]} \Pr[T_i(\boldsymbol{x}) = 1 \mid L(\boldsymbol{y}) = i] \Pr[L(\boldsymbol{y}) = i] \\
&\geq \sum_{i\in[N]} \Pr[\boldsymbol{x}_{\mathrm{supp}(T_i)} = \boldsymbol{y}_{\mathrm{supp}(T_i)} \mid L(\boldsymbol{y}) = i] \Pr[L(\boldsymbol{y}) = i] \\
&\geq \sum_{i\in[N]} \Pr[L(\boldsymbol{y}) = i] \Bigg(1 - \sum_{j\in\mathrm{supp}(T_i)} \Pr[\boldsymbol{x}_j \neq \boldsymbol{y}_j \mid L(\boldsymbol{y}) = i]\Bigg)
\end{aligned}$$

In order to conclude that $\Pr[L(\boldsymbol{x}) \le L(\boldsymbol{y})] \ge 1 - k\varepsilon$ it suffices to show that $\Pr[\boldsymbol{x}_j \ne \boldsymbol{y}_j \mid L(\boldsymbol{y}) = i] \le \varepsilon$. This follows from the total probability law:

$$\Pr[\boldsymbol{x}_j \ne \boldsymbol{y}_j \mid L(\boldsymbol{y}) = i] = \sum_{y\colon L(y)=i} \Pr[\boldsymbol{y} = y] \Pr[\boldsymbol{x}_j \ne \boldsymbol{y}_j \mid \boldsymbol{y} = y] \le \varepsilon.$$

Now the same argument shows that since $(\boldsymbol{y}, \boldsymbol{x})$ is an ε-semi-coupling we have $\Pr[L(\boldsymbol{x}) \ge L(\boldsymbol{y})] \ge 1 - k\varepsilon$. We conclude Theorem 3 by the union bound.

3.2 Surjectivity fools decision lists

Aaronson [Aar11] refuted the GLN conjecture by considering the following distribution:

Definition 15. For every $n = m^2 2^m$, let $N = m2^m$. Define $\mathcal{D}_n$ (or simply $\mathcal{D}$ when n is clear from the context) as the distribution of $\boldsymbol{x} = (\boldsymbol{x}_1, \dots, \boldsymbol{x}_N) \in (\{0,1\}^m)^N$ generated as follows:

1. Sample $\boldsymbol{x}' = (\boldsymbol{x}'_1, \dots, \boldsymbol{x}'_N) \sim (\{0,1\}^m)^N$.
2. Sample $\boldsymbol{y} \sim \{0,1\}^m$.
3. For each $i \in [N]$, let $\boldsymbol{x}_i := \boldsymbol{x}'_i$ if $\boldsymbol{x}'_i \ne \boldsymbol{y}$, otherwise $\boldsymbol{x}_i$ is sampled uniformly from $\{0,1\}^m \setminus \{\boldsymbol{y}\}$.

Aaronson proved the following.

Theorem 16 ([Aar11])**.** *For every $n = m^2 2^m$, $\mathcal{D}$ is $(k \cdot 2^{-m+1}, k)$-wise independent for all $k \le 2^{m-1}$. Moreover, there is a depth-3 AC^0 circuit $C\colon \{0,1\}^n \to \{0,1\}$ of size $O(n^2)$ such that*

$$\left| \Pr_{\boldsymbol{u} \sim \{0,1\}^n}[C(\boldsymbol{u}) = 1] - \Pr_{\boldsymbol{x} \sim \mathcal{D}}[C(\boldsymbol{x}) = 1] \right| \ge \Omega(1).$$

We prove that Aaronson's counterexample, however, cannot refute GLN conjecture for more restricted models, even decision lists.

Lemma 17. *For every $n = m^2 2^m$ and decision list $L\colon \{0,1\}^n \to \{0,1\}$ of width k,*

$$\left| \Pr_{\boldsymbol{u} \sim \{0,1\}^n}[L(\boldsymbol{u}) = 1] - \Pr_{\boldsymbol{x} \sim \mathcal{D}}[L(\boldsymbol{x}) = 1] \right| \le 2k \log^2 n / n.$$

Proof. Let $\boldsymbol{x}, \boldsymbol{x}'$ be as in Definition 15. Note that $\boldsymbol{x} \sim \mathcal{D}, \boldsymbol{x}' \sim \{0,1\}^n$. By Theorem 3, it suffices to show $\boldsymbol{x}$ is $\log^2 n / n = 2^{-m}$-coupled with $\boldsymbol{x}'$.

By definition, we need to show $(\boldsymbol{x}, \boldsymbol{x}')$ and $(\boldsymbol{x}', \boldsymbol{x})$ are 2^{-m}-semi-couplings. The former directly follows from Definition 15: for every $x' \in \{0,1\}^n$ and $i \in [N]$,

$$\Pr[\boldsymbol{x}_i \ne \boldsymbol{x}'_i \mid \boldsymbol{x}' = x'] = \Pr[\boldsymbol{x}'_i = \boldsymbol{y} \mid \boldsymbol{x}' = x'] = 2^{-m}.$$

Regarding the latter, fix any $x \in \operatorname{supp}(\mathcal{D})$, $i \in [N]$. For each $y \in \{0,1\}^m \setminus \operatorname{Im}(x)$ we have

$$\begin{aligned} \Pr[\boldsymbol{x}'_i \ne \boldsymbol{x}_i \mid \boldsymbol{x} = x \wedge \boldsymbol{y} = y] &= \Pr[\boldsymbol{x}'_i = y \mid \boldsymbol{x} = x \wedge \boldsymbol{y} = y] \\ &= \Pr[\boldsymbol{x}'_i = y \mid \boldsymbol{x}_i = x_i \wedge \boldsymbol{y} = y] \\ &= \frac{\Pr[\boldsymbol{x}'_i = y \wedge \boldsymbol{x}_i = x_i \mid \boldsymbol{y} = y]}{\Pr[\boldsymbol{x}_i = x_i \mid \boldsymbol{y} = y]} \\ &= \frac{(2^m - 1)^{-1} 2^{-m}}{(2^m - 1)^{-1}} = 2^{-m}. \end{aligned} \tag{5}$$

Crucially (5) holds since given $\boldsymbol{y} = y$ random variables $\{(\boldsymbol{x}_j, \boldsymbol{x}'_j)\}_{j \in [N]}$ are independent from each other. We conclude by the total probability law:

$$\Pr[\boldsymbol{x}'_i \ne \boldsymbol{x}_i \mid \boldsymbol{x} = x] = \sum_{y \in \{0,1\}^m \setminus \operatorname{Im}(x)} \Pr[\boldsymbol{y} = y \mid \boldsymbol{x} = x] \cdot \Pr[\boldsymbol{x}'_i \ne \boldsymbol{x}_i \mid \boldsymbol{x} = x, \boldsymbol{y} = y] = 2^{-m}. \qquad \square$$

3.3 Semi-couplings do not fool DNFs

In this section we give an example of a semi-coupling $(\boldsymbol{x}, \boldsymbol{u})$ where $\boldsymbol{u} \sim \{0,1\}^n$ such that $\boldsymbol{x}$ can be distinguished from $\boldsymbol{u}$ by a polylogarithmic-width DNF. First, observe that we can interpret the definition of $\boldsymbol{x}$ in Definition 5 as a coupling with the uniform distribution: we sample $\boldsymbol{A}^1, \ldots, \boldsymbol{A}^m, \boldsymbol{p}$ uniformly and then modify $\boldsymbol{p}$ in the location $\boldsymbol{a} = \textsc{Tribes}(\boldsymbol{A}^1), \ldots, \textsc{Tribes}(\boldsymbol{A}^m)$. With $\boldsymbol{p}'$ being the state of $\boldsymbol{p}$ before the change, that defines some coupling between $\boldsymbol{x}$ and the uniformly distributed $\boldsymbol{A}^1, \ldots, \boldsymbol{A}^m, \boldsymbol{p}'$. This, however, is not a semi-coupling, since if we fix $\boldsymbol{A}^1, \ldots, \boldsymbol{A}^m$ to some value such that $|\boldsymbol{a}| = m/2$ and fix $\boldsymbol{p}'$ such that $\boldsymbol{p}'_{\boldsymbol{a}} = 0$, then $0 = \boldsymbol{p}'_{\boldsymbol{a}} \neq \boldsymbol{p}_{\boldsymbol{a}} = 1$ with probability 1.

We modify the distribution from Definition 5 by replacing each bit of $\mathbf{p}$ with an instance of $\textsc{Tribes}$.

Lemma 18. *There exists a $n^{-0.6}$-semi-coupling $(\boldsymbol{x}, \boldsymbol{u})$ with $\boldsymbol{u} \sim \{0,1\}^n$ and an $O(\log^2 n)$-DNF that $\Omega(\log^{-1/2} n)$-distinguishes $\boldsymbol{x}$ from $\boldsymbol{u}$.*

Proof. Consider the smallest m such that $m^2\lceil 2^m \ln 2\rceil + 2^m \lceil 2^{2m} \ln 2\rceil \geq n$. We define the coupling as follows:

1. Sample $\boldsymbol{A} = \boldsymbol{A}^1, \ldots, \boldsymbol{A}^m \sim (\{0,1\}^{m \times \lceil 2^m \ln 2\rceil})^m$ uniformly.
2. Sample $\boldsymbol{P} = \boldsymbol{P}^1, \ldots, \boldsymbol{P}^{2^m} \sim (\{0,1\}^{2m \times \lceil 2^{2m} \ln 2\rceil})^{2^m}$ uniformly.
3. Take $\boldsymbol{Q} = \boldsymbol{P}$.
4. Define $\boldsymbol{a} \in \{0,1\}^m$ by $\boldsymbol{a}_i = \textsc{Tribes}(\boldsymbol{A}^i)$ for each $i \in [m]$.
5. If $|\boldsymbol{a}| = m/2$, choose $\boldsymbol{j} \sim [\lceil 2^{2m} \ln 2\rceil]$ and force $\boldsymbol{Q}^{\boldsymbol{a}}_{\ell,\boldsymbol{j}} := 1$ for each $\ell \in [2m]$.

Local coupling. We claim that $\boldsymbol{x} := (\boldsymbol{A}, \boldsymbol{Q})$ is 2^{-2m}-semi-coupled with $\boldsymbol{u} := (\boldsymbol{A}, \boldsymbol{P})$. Fix some $A \in \operatorname{supp}(\boldsymbol{A})$ and $P \in \operatorname{supp}(\boldsymbol{P})$. Then for bits of $\boldsymbol{x}$ that correspond to $\boldsymbol{A}$ the coupling condition is trivially satisfied as these bits are shared with $\boldsymbol{u}$. The remaining bits are indexed by $a \in \{0,1\}^m$, $i \in [2m]$, $j \in [\lceil 2^{2m} \ln 2\rceil]$, we need to bound the probability:

$$\Pr[\boldsymbol{P}^a_{i,j} \neq \boldsymbol{Q}^a_{i,j} \mid \boldsymbol{A} = A \wedge \boldsymbol{P} = P] = \Pr[\boldsymbol{Q}^a_{i,j} \neq P^a_{i,j} \mid \boldsymbol{A} = A \wedge \boldsymbol{P} = P]$$

If $|a| \neq m/2$ or $a \neq (\textsc{Tribes}(A^1), \ldots, \textsc{Tribes}(A^m))$, then this probability is 0 since (5) is not invoked and $\boldsymbol{P} = \boldsymbol{Q}$. If $|a| = m/2$ and $a = (\textsc{Tribes}(A^1), \ldots, \textsc{Tribes}(A^m))$ we have

$$\Pr[\boldsymbol{Q}^a_{i,j} \neq P^a_{i,j} \mid \boldsymbol{A} = A \wedge \boldsymbol{P} = P] \leq \Pr[\boldsymbol{j} = j] = 1/\lceil 2^{2m} \ln 2\rceil \leq 2^{-2m} \ll n^{-0.6}.$$

Distinguishability. We take the distinguishing function F from Lemma 7 and define the new distinguisher $F' \colon \operatorname{supp}(\boldsymbol{A}) \times \operatorname{supp}(\boldsymbol{P}) \to \{0,1\}$ as

$$F'(A^1, \ldots, A^m, P^1, \ldots, P^{2^m}) := F(A^1, \ldots, A^m, \textsc{Tribes}(P^1), \ldots, \textsc{Tribes}(P^{2^m})).$$

Let E be the event "$|\boldsymbol{a}| = m/2$". As in Lemma 7 we observe that $\Pr[\boldsymbol{P} = \boldsymbol{Q} \mid \neg E] = 1$, so $\Pr[F'(\boldsymbol{A}, \boldsymbol{P}) = 1 \mid \neg E] = \Pr[F'(\boldsymbol{A}, \boldsymbol{Q}) = 1 \mid \neg E]$. By the construction of $\boldsymbol{Q}$ and F' we have $\Pr[F'(\boldsymbol{A}, \boldsymbol{Q}) = 1 \mid E] = 1$. On the other hand

$$\begin{aligned}
\Pr[F'(\boldsymbol{A}, \boldsymbol{P}) = 1 \mid E] &= \Pr[F(\boldsymbol{A}, (\textsc{Tribes}(\boldsymbol{P}^1), \ldots, \textsc{Tribes}(\boldsymbol{P}^{2^m}))) = 1 \mid E] \\
\text{(by Lemma 9)} \quad &\leq \Pr_{\boldsymbol{x} \sim \{0,1\}^{2^m}}[F(\boldsymbol{A}, \boldsymbol{x}) = 1 \mid E] + O(2^{-2m} \cdot 2^m) \\
\text{(analogous to Lemma 7)} \quad &\leq 1/2 + O(2^{-m}) \leq 2/3.
\end{aligned}$$

Formally, to show the last inequality, we will do the following:

$$\Pr_{\boldsymbol{x}\sim\{0,1\}^{2^m}}[F(\boldsymbol{A},\boldsymbol{x})=1 \mid E] = \Pr_{\boldsymbol{x}\sim\{0,1\}^{2^m}}[\boldsymbol{x}_{\boldsymbol{a}}=1 \mid E]$$
$$= \sum_{a\in\{0,1\}^m} \Pr[\boldsymbol{x}_a = 1 \mid E \wedge \boldsymbol{a}=a]\Pr[\boldsymbol{a}=a \mid E]$$
$$= \sum_{a\in\{0,1\}^m} \Pr[\boldsymbol{x}_a = 1]\Pr[\boldsymbol{a}=a \mid E] = \frac{1}{2}.$$

Then as shown in Lemma 7 $\Pr[E] = \Omega(1/\sqrt{m})$. All together this gives us that F' $\Omega(1/\sqrt{m})$-distinguishes $\boldsymbol{x}$ and $\boldsymbol{u}$.

It remains to observe that the 1-certificate complexity of F' is at most $O(m^2)$: to the certificate of F in Claim 4 we add the certificate that $\textsc{Tribes}(P^j)=1$ where $j=(\textsc{Tribes}(A^1),\ldots,\textsc{Tribes}(A^m))$. Thus there exists a DNF of width $O(m^2)$ that computes F. □

In order to get the $\Omega(1)$-distinguishability we follow the amplification in Section 2.4:

Theorem 19. *There exists a $1/\sqrt{n}$-semi-coupling $(\boldsymbol{x},\boldsymbol{u})$ where $\boldsymbol{u}\sim\{0,1\}^n$ and a $O(\log^3 n)$-width* DNF *that $\Omega(1)$-distinguishes $\boldsymbol{x}$ from $\boldsymbol{u}$.*

Proof. The proof is identical to the one of Theorem 1. Take $\boldsymbol{x}'$ over $\{0,1\}^s$ that is $s^{-0.6}$-semi-coupled with $\boldsymbol{u}'\sim\{0,1\}^s$, then the random variable $\boldsymbol{x}$ comprised of $t=O(\log s)$ independent copies of $\boldsymbol{x}'$, $\boldsymbol{x}=\boldsymbol{x}'_1,\ldots,\boldsymbol{x}'_t$ is $s^{-0.6}$-semi-coupled with t independent copies of $\boldsymbol{u}'$, $\boldsymbol{u}=\boldsymbol{u}'_1,\ldots,\boldsymbol{u}'_t$. On the other hand by Lemma 12 and Lemma 11 there exists an $O(t\log^2 s)=O(\log^3 n)$-DNF that $\Omega(1)$-distinguishes $\boldsymbol{x}$ and $\boldsymbol{u}$. Since $s^{-0.6}\ll n^{-1/2}$ we get the claim. □

References

[Aar10] Scott Aaronson. BQP and the Polynomial Hierarchy. In *42nd ACM Symposium on Theory of Computing, STOC*, pages 141–150. ACM, 2010. doi:10.1145/1806689.1806711.

[Aar11] Scott Aaronson. A Counterexample to the Generalized Linial-Nisan Conjecture. *CoRR*, abs/1110.6126, 2011. arXiv:1110.6126.

[AGHP92] Noga Alon, Oded Goldreich, Johan Håstad, and René Peralta. Simple Constructions of Almost k-wise Independent Random Variables. *Random Structures and Algorithms*, 3(3):289–304, 1992. doi:10.1002/rsa.3240030308.

[Baz09] Louay Bazzi. Polylogarithmic independence can fool DNF formulas. *SIAM Journal on Computing*, 38(6):2220–2272, 2009. doi:10.1137/070691954.

[BDF+22] Andrej Bogdanov, Krishnamoorthy Dinesh, Yuval Filmus, Yuval Ishai, Avi Kaplan, and Akshayaram Srinivasan. Bounded Indistinguishability for Simple Sources. In *13th Innovations in Theoretical Computer Science Conference, ITCS*, volume 215 of *LIPIcs*, pages 26:1–26:18. Schloss Dagstuhl, 2022. doi:10.4230/LIPIcs.ITCS.2022.26.

[BIVW16] Andrej Bogdanov, Yuval Ishai, Emanuele Viola, and Christopher Williamson. Bounded Indistinguishability and the Complexity of Recovering Secrets. In *36th Advances in Cryptology, CRYPTO*, pages 593–618. Springer, 2016. doi:10.1007/978-3-662-53015-3_21.

[Bra11] Mark Braverman. Poly-logarithmic Independence Fools Bounded-Depth Boolean Circuits. *Communications of the ACM*, 54(4):108–115, 2011. doi:10.1145/1924421.1924446.

[CGH+85] Benny Chor, Oded Goldreich, Johan Hasted, Joel Freidmann, Steven Rudich, and Roman Smolensky. The Bit Extraction Problem or t-resilient Functions. In *26th Annual Symposium on Foundations of Computer Science, FOCS*. IEEE, 1985. doi:10.1109/sfcs.1985.55.

[DETT10] Anindya De, Omid Etesami, Luca Trevisan, and Madhur Tulsiani. Improved Pseudorandom Generators for Depth 2 Circuits. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, pages 504–517. Springer Berlin Heidelberg, 2010. doi:10.1007/978-3-642-15369-3_38.

[GKPW19] Mika Göös, Pritish Kamath, Toniann Pitassi, and Thomas Watson. Query-to-Communication Lifting for P^{NP}. *Computational Complexity*, 28(1):113–144, 2019.

[GRSS23] Mika Göös, Artur Riazanov, Anastasia Sofronova, and Dmitry Sokolov. Top-Down Lower Bounds for Depth-Four Circuits. In *64th Annual Symposium on Foundations of Computer Science, FOCS*, pages 1048–1055. IEEE, 2023. doi:10.1109/FOCS57990.2023.00063.

[HH24] Pooya Hatami and William Hoza. Paradigms for Unconditional Pseudorandom Generators. *Foundations and Trends in Theoretical Computer Science*, 16(1–2):1–210, 2024. doi:10.1561/0400000109.

[HJP93] Johan Håstad, Stasys Jukna, and Pavel Pudlák. Top-Down Lower Bounds for Depth 3 Circuits. In *34th Annual Symposium on Foundations of Computer Science, FOCS*, pages 124–129. IEEE Computer Society, 1993. doi:10.1109/SFCS.1993.366875.

[Hoz25] William Hoza. Fooling Near-Maximal Decision Trees. Technical report, ECCC, 2025. URL: https://eccc.weizmann.ac.il/report/2025/003/.

[HRST17] Johan Håstad, Benjamin Rossman, Rocco Servedio, and Li-Yang Tan. An Average-Case Depth Hierarchy Theorem for Boolean Circuits. *Journal of the ACM*, 64(5), 2017. doi:10.1145/3095799.

[LN90] Nathan Linial and Noam Nisan. Approximate inclusion-exclusion. *Combinatorica*, 10(4):349–365, 1990. doi:10.1007/BF02128670.

[Lyu22] Xin Lyu. Improved Pseudorandom Generators for AC^0 Circuits. In *37th Computational Complexity Conference, CCC*, volume 234 of *LIPIcs*, pages 34:1–34:25. Schloss Dagstuhl, 2022. doi:10.4230/LIPIcs.CCC.2022.34.

[NN93] Joseph Naor and Moni Naor. Small-Bias Probability Spaces: Efficient Constructions and Applications. *SIAM Journal on Computing*, 22(4):838–856, 1993. doi:10.1137/0222053.

[O'D14] Ryan O'Donnell. *Analysis of Boolean Functions*. Cambridge University Press, 2014.

[Raz09] Alexander Razborov. A Simple Proof of Bazzi's Theorem. *ACM Transactions Computational Theory*, 1(1):3:1–3:5, 2009. doi:10.1145/1490270.1490273.

[RT19] Ran Raz and Avishay Tal. Oracle separation of BQP and PH. In *51st ACM Symposium on Theory of Computing, STOC*, pages 13–23. ACM, 2019. doi:10.1145/3313276.3316315.

[Tal17] Avishay Tal. Tight Bounds on the Fourier Spectrum of AC^0. In *32nd Computational Complexity Conference, CCC*, pages 15:1–15:31. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2017. doi:10.4230/LIPICS.CCC.2017.15.

[Zha25] Mark Zhandry. Toward Separating QMA from QCMA with a Classical Oracle. In *16th Innovations in Theoretical Computer Science Conference, ITCS*, volume 325 of *LIPIcs*, pages 95:1–95:19. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2025. doi:10.4230/LIPICS.ITCS.2025.95.

ECCC ISSN 1433-8092

https://eccc.weizmann.ac.il