A Counterexample to the Generalized Linial-Nisan Conjecture

Scott Aaronson*

Abstract

In earlier work [1], we gave an oracle separating the relational versions of BQP and the polynomial hierarchy, and showed that an oracle separating the decision versions would follow from what we called the *Generalized Linial-Nisan (GLN) Conjecture*: that "almost k-wise independent" distributions are indistinguishable from the uniform distribution by constant-depth circuits. The original Linial-Nisan Conjecture was recently proved by Braverman [7]; we offered a $200 prize for the generalized version. In this paper, we save ourselves $200 by showing that the GLN Conjecture is false, at least for circuits of depth 3 and higher.

As a byproduct, our counterexample also implies that $\Pi_2^p \not\subset \mathsf{P}^{\mathsf{NP}}$ relative to a random oracle with probability 1. It has been conjectured since the 1980s that PH is infinite relative to a random oracle, but the highest levels of PH previously proved separate were NP and coNP.

Finally, our counterexample implies that the famous results of Linial, Mansour, and Nisan [12], on the structure of AC^0 functions, cannot be improved in several interesting respects.

1 Introduction

Proving an oracle separation between BQP and PH is one of the central open problems of quantum complexity theory. In a recent paper [1], we reported the following progress on the problem:

(1) We constructed an oracle relative to which $\mathsf{FBQP} \not\subset \mathsf{FBPP}^{\mathsf{PH}}$, where FBQP and $\mathsf{FBPP}^{\mathsf{PH}}$ are the "relational" versions of BQP and PH respectively (that is, the versions where there are many valid outputs, and an algorithm's task is to output any one of them).

(2) We proposed a natural *decision* problem, called Fourier Checking, which is provably in BQP (as an oracle problem) and which we conjectured was not in PH.

(3) We showed that Fourier Checking has a property called *almost k-wise independence*, and that no $\mathsf{BPP}_{\mathsf{path}}$ or SZK problem shares that property. This allowed us to give oracles relative to which BQP was outside those classes, and to reprove all the known oracle separations between BQP and classical complexity classes in a unified way.

(4) We conjectured that no PH problem has the almost k-wise independence property, and called that the *Generalized Linial-Nisan (GLN) Conjecture*. Proving the GLN Conjecture would imply the existence of an oracle relative to which $\mathsf{BQP} \not\subset \mathsf{PH}$.

*MIT. Email: aaronson@csail.mit.edu. This material is based upon work supported by the National Science Foundation under Grant No. 0844626. Also supported by a DARPA YFA grant and the Keck Foundation.

This paper does nothing to modify points (1)-(3) above: the unconditional results in [1] are still true, and we still conjecture not only that there exists an oracle relative to which $\mathsf{BQP} \not\subset \mathsf{PH}$, but that Fourier Checking is such an oracle.

However, we will show that the hope of proving Fourier Checking$\not\in \mathsf{PH}$ by proving the GLN Conjecture was unfounded:

> *The GLN Conjecture is false, at least for Π_2^p and higher levels of the polynomial hierarchy.*

We prove this by giving an explicit counterexample: a family of depth-three AC^0 circuits that distinguish the uniform distribution over n-bit strings from an $\widetilde{O}(k/n)$-almost k-wise independent distribution, with constant bias.[1]

Our counterexample was inspired by a recent result of Beame and Machmouchi [3], giving a Boolean function with quantum query complexity $\Omega(n/\log n)$ that is computable by a depth-three AC^0 circuit. This disproved a conjecture, relayed to us earlier by Beame, stating that every AC^0 function has quantum query complexity $n^{1-\Omega(1)}$. Like the Beame-Machmouchi counterexample, ours involves inputs $X = x_1 \ldots x_N \in [M]^N$ that are lists of positive integers, with the x_i's encoded in binary to obtain a Boolean problem; as well as a function $f : [M]^N \rightarrow \{0,1\}$ that uses two alternating quantifiers to express a "global" property of X. In Beame and Machmouchi's case, the property in question was that the function $x(i) := x_i$ is 2-to-1; in our case, the property is that $x(i)$ is surjective.[2]

Our counterexample makes essential use of depth-*three* circuits, and we find it plausible that the GLN Conjecture still holds for depth-*two* circuits (i.e., for DNF formulas).[3] As shown in [1], proving the GLN Conjecture for depth-two circuits would yield an oracle relative to which $\mathsf{BQP} \not\subset \mathsf{AM}$, which is already a longstanding open problem.

Given that the GLN Conjecture resisted attacks for two years (and indirectly motivated the beautiful works of Razborov [16] and Braverman [7] on the original LN Conjecture), our counterexample cannot have been *quite* as obvious as it seems in retrospect! Perhaps Andy Drucker (personal communication) summarized the situation best: almost k-wise independent distributions seem to be much better at fooling *people* than at fooling circuits.

1.1 Further Implications

Besides falsifying the GLN Conjecture, our counterexample has several other interesting implications for PH and AC^0.

Firstly, we are able to use our counterexample to prove that $(\Pi_2^p)^A \not\subset \mathsf{P}^{\mathsf{NP}^A}$ with probability 1 relative to a random oracle A. Indeed, we conjecture that our counterexample can even be used to prove $(\Pi_2^p)^A \not\subset (\Sigma_2^p)^A$ with probability 1 for a random oracle A. The seminal work of Yao [18] showed PH infinite relative to *some* oracle, but it has been an open problem for almost thirty years to prove PH infinite relative to a *random* oracle (see the book of Håstad [17] for discussion). Motivation for this problem comes from a surprising result of Book [6], which says that if PH

[1] Note that depth-three AC^0 circuits correspond to the second level of PH, depth-four circuits correspond to the third level, and so on.

[2] Beame and Machmouchi [3] also mention the surjectivity property, in Corollary 6 of their paper.

[3] Indeed, we originally formulated the conjecture for depth-two circuits only, before (rashly) extending it to arbitrary depths.

collapses relative to a random oracle, then it also collapses in the unrelativized world. Our result, while simple, appears to represent the first "progress" toward separating PH by random oracles since the original result of Bennett and Gill [5] that $\mathsf{P} \neq \mathsf{NP} \neq \mathsf{coNP}$ relative to a random oracle with probability 1.[4]

Secondly, our counterexample shows that the celebrated results of Linial, Mansour, and Nisan [12], on the Fourier spectrum of AC^0 functions, *cannot* be improved in several important respects. In particular, Linial et al. showed that every Boolean function $f : \{0,1\}^n \rightarrow \{0,1\}$ in AC^0 has *average sensitivity* $O(\text{polylog}(n))$. However, we observe that this result fails completely if we consider a closely-related measure, the *average block-sensitivity*. Indeed, there exists a reasonably-balanced Boolean function $f \in \mathsf{AC}^0$ such that every 1-input can be modified in $\Omega(n/\log n)$ disjoint ways to produce a 0-input, and almost every 0-input can be modified in $\Omega(n/\log n)$ disjoint ways to produce a 1-input. What makes this behavior interesting is that one normally associates it with (say) PARITY, the canonical function *not* in AC^0!

Linial et al. [12] also showed that every Boolean function $f \in \mathsf{AC}^0$ has a *low-degree approximating polynomial*: that is, a real polynomial $p : \{0,1\}^n \rightarrow \mathbb{R}$, of degree $O(\text{polylog}(n))$, such that

$$\mathop{\mathrm{E}}_{X \in \{0,1\}^n} \left[\left(p(X) - f(X) \right)^2 \right] = o(1).$$

However, using our counterexample, we will show that such a polynomial p *cannot* generally be written as a linear combination of terms, $p = \sum_C \alpha_C C$, where the coefficients satisfy the following bound:

$$\sum_C |\alpha_C| \, 2^{-|C|} = n^{o(1)}.$$

In other words, such a polynomial cannot be "low-fat" in the sense defined by Aaronson [1], but must instead involve "massive cancellations" between positive and negative terms. This gives the first example of a Boolean function f that can be approximated in L_2-norm by a low-degree polynomial, but *not* by a low-degree low-fat polynomial—thereby answering another one of the open questions from [1].

1.2 The Future of BQP and PH

While this paper rules out the GLN approach, at least three plausible avenues remain for proving an oracle separation between BQP and PH.

(1) Our original idea for proving FOURIER CHECKING$\notin$ PH was to use a direct random restriction argument—and while we were unable to make such an argument work, we have also found nothing to rule it out.

(2) Besides almost k-wise independence, the other "obvious" property of FOURIER CHECKING that might be useful for lower bounds is its close connection with the MAJORITY function. Indeed, given as input the truth table of a Boolean function $f : \{0,1\}^n \rightarrow \{-1,1\}$, estimating a *single* Fourier coefficient $\widehat{f}(s) := \frac{1}{2^{n/2}} \sum_x (-1)^{x \cdot s} f(x)$ is easily seen to be equivalent to solving

[4] Though "working from the opposite direction," Cai [9] proved the beautiful result that PH $\neq$ PSPACE relative to a random oracle with probability 1. Note that any relativized world where PH is infinite must also satisfy PH $\neq$ PSPACE. Cai [8] also proved that BH is infinite with probability 1, where BH represents the Boolean hierarchy over NP, a subclass of $\mathsf{P}_{||}^{\mathsf{NP}}$.

MAJORITY, which is known to be hard for AC^0. Thus, in proving FOURIER CHECKING$\notin \mathsf{PH}$, the difficulty is "merely" to show that checking the answers to 2^n overlapping MAJORITY instances is not significantly easier for an AC^0 circuit than checking the answer to one instance. While the usual hybrid argument fails in this case, one could hope for some other reduction—possibly a non-black-box reduction—showing that if FOURIER CHECKING is in AC^0, then MAJORITY is as well.

(3) Recently, Fefferman and Umans [10] proposed a beautiful alternative approach to the relativized BQP versus PH question. Like approach (2) above, their approach is based on a hoped-for reduction from MAJORITY. However, they replace FOURIER CHECKING by a different candidate problem, which involves Nisan-Wigderson combinatorial designs [15] rather than the Fourier transform. They show that their candidate problem is in BQP, and also show that it is *not* in PH, assuming (roughly speaking) that the analysis of the NW generator can be improved in a direction that people have wanted to improve it in for independent reasons. Fefferman and Umans' conjecture follows from the GLN Conjecture,[5] but is much more tailored to a specific pseudorandom generator, and is completely unaffected by our counterexample.

1.3 Organization

The rest of the paper is organized as follows. Section 2 provides background on AC^0, (almost) k-wise independence, and the (Generalized) Linial-Nisan Conjecture; then Section 3 presents our counterexample. Section 4 uses the counterexample to prove that $\Pi_2^p \not\subset \mathsf{P}^{\mathsf{NP}}$ relative to a random oracle, and Section 5 gives implications of the counterexample for the noise sensitivity and approximate degree of AC^0 functions. Section 6 concludes with some discussion and open problems.

2 Background

We refer the reader to [1] for details on the original and generalized Linial-Nisan Conjectures, as well as their relationship to BQP and PH. In this section, we give a brief recap of the definitions, conjectures, and results that are relevant to our counterexample.

By AC^0, we mean the class of Boolean function families $\{f_n\}_{n\geq 1}$ such that each $f_n : \{0,1\}^n \rightarrow \{0,1\}$ is computable by a circuit of AND, OR, and NOT gates with constant depth, unbounded fanin, and size $n^{O(1)}$. Here *depth* means the number of alternating layers of AND and OR gates; NOT gates are not counted. Abusing notation, we will often use phrases like "AC^0 circuit of size $2^{n^{o(1)}}$," which means the size is now superpolynomial but the depth is still $O(1)$. We will also generally drop the subscript of n.

Throughout the paper we abbreviate probability expressions such as $\Pr_{X\sim\mathcal{D}}[f(X)]$ by $\Pr_{\mathcal{D}}[f]$. Let $\mathcal{U}$ be the uniform distribution over n-bit strings, so that $\Pr_{\mathcal{U}}[X] = 1/2^n$ for all $X \in \{0,1\}^n$. A distribution $\mathcal{D}$ over $\{0,1\}^n$ is called *k-wise independent* (for $k \leq n$) if $\mathcal{D}$ is uniform on every subset of at most k bits. A central question in pseudorandomness and cryptography is what computational resources are needed to distinguish such a "pretend-uniform" distribution from the "truly-uniform" one. In 1990, Linial and Nisan [13] famously conjectured that *n^{ε}-wise independence fools AC^0 circuits:*

[5] As, indeed, *anything* follows from the GLN Conjecture.

Conjecture 1 (Linial-Nisan or LN Conjecture) *Let $\mathcal{D}$ be any $n^{\Omega(1)}$-wise independent distribution over $\{0,1\}^n$, and let $f : \{0,1\}^n \rightarrow \{0,1\}$ be computed by an AC^0 circuit of size $2^{n^{o(1)}}$. Then*

$$\left| \Pr_{\mathcal{D}}[f] - \Pr_{\mathcal{U}}[f] \right| = o(1).$$

(The actual parameters in the LN Conjecture are considerably stronger than the above, but also more complicated to state. We chose weaker parameters that suffice for our discussion.)

After seventeen years of almost no progress, Bazzi [2] finally proved Conjecture 1 for the special case of depth-2 circuits. Shortly afterward, Razborov [16] gave a dramatically simpler proof of Bazzi's theorem, and shortly after *that*, Braverman [7] proved the full Conjecture 1:

Theorem 2 (Braverman's Theorem [7]) *Let $f : \{0,1\}^n \rightarrow \{0,1\}$ be computed by an AC^0 circuit of size S and depth d, and let $\mathcal{D}$ be a $\left(\log \frac{S}{\varepsilon}\right)^{7d^2}$-wise independent distribution over $\{0,1\}^n$. Then for all sufficiently large S,*

$$\left| \Pr_{\mathcal{D}}[f] - \Pr_{\mathcal{U}}[f] \right| \leq \varepsilon.$$

Even before the work of Razborov [16] and Braverman [7], we had proposed a deceptively modest-seeming generalization of Conjecture 1, motivated by the application to the BQP versus PH question mentioned previously. To state the generalization, we need some more terminology. Let $X = x_1 \ldots x_n \in \{0,1\}^n$ be a string. Then a *literal* is an expression of the form x_i or $1 - x_i$, and a *k-term* is a product of k literals (each involving a different x_i), which is 1 if the literals all take on prescribed values and 0 otherwise.

Definition 3 (almost k-wise independence) *Given a distribution $\mathcal{D}$ over $\{0,1\}^n$ and a k-term C, we say that C is ε-fooled by $\mathcal{D}$ if*

$$1 - \varepsilon \leq \frac{\Pr_{\mathcal{D}}[C]}{\Pr_{\mathcal{U}}[C]} \leq 1 + \varepsilon.$$

(Note that $\Pr_{\mathcal{U}}[C]$ is just 2^{-k}.) Then $\mathcal{D}$ is ε-almost k-wise independent if every k-term C is ε-fooled by $\mathcal{D}$.

In other words, there should be no assignment to any k bits, such that conditioning on that assignment gives us much information about whether X was drawn from $\mathcal{D}$ or from $\mathcal{U}$. We can now state the conjecture that we falsify.

Conjecture 4 (Generalized Linial-Nisan or GLN Conjecture) *Let $\mathcal{D}$ be a $1/n^{\Omega(1)}$-almost $n^{\Omega(1)}$-wise independent distribution over $\{0,1\}^n$, and let $f : \{0,1\}^n \rightarrow \{0,1\}$ be computed by an AC^0 circuit of size $2^{n^{o(1)}}$. Then*

$$\left| \Pr_{\mathcal{D}}[f] - \Pr_{\mathcal{U}}[f] \right| = o(1).$$

Note that, for Conjecture 4 not to be ruled out *immediately*, it is essential that our definition of ε-fooling was *multiplicative* rather than additive. For suppose we had merely required that, on every subset of indices $S \subseteq [n]$ with $|S| \leq k$, the marginal distribution $\mathcal{D}_S$ was ε-close in variation distance to the uniform distribution. Then it would be easy to construct almost k-wise distributions

$\mathcal{D}$ that were distinguishable from the uniform distribution even by DNF formulas. For example, the uniform distribution over all sequences $X = x_1 \dots x_N \in [N]^N$ that are *permutations* (with the x_i's appropriately coded in binary) is one such $\mathcal{D}$.

This paper shows that, even with the more careful multiplicative definition of ε-fooling, there is *still* a counterexample to Conjecture 4—although we have to work harder and use higher-depth circuits to construct it. The failure of Conjecture 4 means that Braverman's Theorem is "essentially optimal," in the sense that one cannot relax the k-wise independence condition to almost k-wise independence. This demonstrates a striking contrast between k-wise independence and almost k-wise independence in terms of their implications for pseudorandomness.

3 The Counterexample

Fix a positive integer m, and let $M := 2^m$. Then it will be useful think of the input $X = x_1 \dots x_N$ as belonging to the set $[M]^N$, where $N := \lceil Mm \ln 2 \rceil$. However, to make contact with the original statement of the GLN Conjecture, we can easily encode such an X as an n-bit string where $n := Nm$, by writing out each x_i in binary. Abusing notation, we will speak interchangeably about X as an element of $\{0,1\}^n$ or of $[M]^N$.

Let the *image* of X, or $\mathrm{Im}_X := \{x_1, \dots, x_N\}$, be the set of integers that appear in X. Then define the *surjectivity function*, $f_{\mathrm{Surj}} : \{0,1\}^n \rightarrow \{0,1\}$ by $f_{\mathrm{Surj}}(X) = 1$ if $\mathrm{Im}_X = [M]$ and $f_{\mathrm{Surj}}(X) = 0$ otherwise. A first easy observation is that $f_{\mathrm{Surj}} \in \mathsf{AC}^0$.

Lemma 5 *f_{Surj} is computable by an AC^0 circuit of depth 3 and size $O(NMm)$.*

Proof. For all $i \in [N]$ and $y \in [M]$, let $\Delta(x_i, y)$ denote the m-term that evaluates to 1 if $x_i = y$ and to 0 otherwise. Then

$$f_{\mathrm{Surj}}(X) = \bigwedge_{y \in [M]} \bigvee_{i \in [N]} \Delta(x_i, y).$$

■

Now let $\mathcal{U}$ be the uniform distribution over $[M]^N$, so that $\Pr_{\mathcal{U}}[X] = 1/M^N$ for all $X \in [M]^N$. Also, given an input $X \in [M]^N$, we define a distribution $\mathcal{D}(X)$ over "perturbed" versions of X via the following procedure:

(1) Choose y uniformly at random from $[M]$.

(2) For each $i \in [N]$ such that $x_i = y$, change x_i to a uniform, independent sample from $[M] \setminus \{y\}$.

Then we let $\mathcal{D} := \mathcal{D}(\mathcal{U})$ be the distribution over inputs Z obtained by first drawing an X from $\mathcal{U}$, and then sampling Z from $\mathcal{D}(X)$. Notice that $\mathrm{Im}_Z \neq [M]$ and hence $f(Z) = 0$ for all Z in the support of $\mathcal{D}$.

Here is an observation that will be helpful later. Given a sample $Z = z_1 \dots z_N$ from $\mathcal{D}$, we can define a distribution $\mathcal{D}^{\mathrm{inv}}(Z)$ over perturbed versions of Z via the following "inverse" procedure:

(1) Choose y uniformly at random from $[M] \setminus \mathrm{Im}_Z$.

(2) For each $i \in [N]$, change z_i to y with independent probability $1/M$.

We claim that $\mathcal{D}^{\mathrm{inv}}$ is indeed the inverse of $\mathcal{D}$.

Claim 6 $\mathcal{D}^{\text{inv}}(\mathcal{D}(\mathcal{U})) = \mathcal{U}$.

Proof. Let $\mathcal{D}_y(X)$ be the variant of $\mathcal{D}(X)$ where we fix the element $y \in [M]$ in step (1), so that $\mathcal{D}(X) = \mathrm{E}_{y \in [M]} \mathcal{D}_y(X)$. Similarly, let $\mathcal{D}_y^{\text{inv}}(Z)$ be the variant of $\mathcal{D}^{\text{inv}}(Z)$ where we fix the element $y \in [M] \setminus \mathrm{Im}_Z$. Then it is easy to see that, for every fixed $y \in [M]$, we have $\mathcal{D}_y^{\text{inv}}(\mathcal{D}_y(\mathcal{U})) = \mathcal{U}$. For choosing each x_i uniformly at random, then changing it randomly if equals y, then changing it *back* to y with probability $1/M$, is just a more complicated way of choosing x_i uniformly at random.

Now let $\mathrm{Hist}(X)$ be the *histogram* of X: that is, the multiset $\{h_1, \ldots, h_M\}$ where $h_y := |\{i : x_i = y\}|$. Then we can conclude from the above that, for every $y \in [M]$,

$$\begin{aligned}\mathrm{Hist}\left(\mathcal{D}^{\text{inv}}(\mathcal{D}(\mathcal{U}))\right) &= \mathrm{Hist}\left(\mathcal{D}^{\text{inv}}(\mathcal{D}_y(\mathcal{U}))\right) \\ &= \mathrm{Hist}\left(\mathcal{D}_y^{\text{inv}}(\mathcal{D}_y(\mathcal{U}))\right) \\ &= \mathrm{Hist}(\mathcal{U}).\end{aligned}$$

Call a distribution $\mathcal{A}$ over $[M]^N$ *symmetric* if $\Pr_{\mathcal{A}}[X]$ depends only on $\mathrm{Hist}(X)$. Notice that $\mathcal{U}$ is symmetric, and that if $\mathcal{A}$ is symmetric, then $\mathcal{D}(\mathcal{A})$ and $\mathcal{D}^{\text{inv}}(\mathcal{A})$ are both symmetric also. This means that from $\mathrm{Hist}\left(\mathcal{D}^{\text{inv}}(\mathcal{D}(\mathcal{U}))\right) = \mathrm{Hist}(\mathcal{U})$, we can conclude that $\mathcal{D}^{\text{inv}}(\mathcal{D}(\mathcal{U})) = \mathcal{U}$ as well. ■

We now show that the function f_{Surj} distinguishes $\mathcal{D}$ from $\mathcal{U}$ with constant bias.

Lemma 7 $\mathrm{E}_{\mathcal{U}}[f_{\text{Surj}}] - \mathrm{E}_{\mathcal{D}}[f_{\text{Surj}}] \geq 1/e - o(1)$.

Proof. By construction, we have $\mathrm{E}_{\mathcal{D}}[f_{\text{Surj}}] = 0$. On the other hand,

$$\mathop{\mathrm{E}}_{\mathcal{U}}[f_{\text{Surj}}] = \Pr_{\mathcal{U}}[|\mathrm{Im}_X| = M].$$

Think of $N = M \ln M + O(1)$ balls, which are thrown uniformly and independently into M bins. Then $|\mathrm{Im}_X|$ is just the number of bins that receive at least one ball. Using the Poisson approximation, we have

$$\lim_{M \to \infty} \Pr_{\mathcal{U}}[|\mathrm{Im}_X| = M] = \frac{1}{e},$$

and therefore $\mathrm{E}_{\mathcal{U}}[f_{\text{Surj}}] \geq 1/e - o(1)$. ■

To show that the distribution $\mathcal{D}$ is almost k-wise independent, we first need a technical claim, to the effect that almost k-wise independence behaves well with respect to restrictions. Given a k-term C, let $V(C)$ be the set of variables that occur in C. Also, given a set S of variables that contains $V(C)$, let $U_S(C)$ be the set of all $2^{|S|-k}$ terms B such that $V(B) = S$ and $B \Longrightarrow C$.

Claim 8 *Given a k-term C and a set S containing $V(C)$, suppose every term $B \in U_S(C)$ is ε-fooled by $\mathcal{D}$. Then C is ε-fooled by $\mathcal{D}$.*

Proof. It suffices to check the claim in the case $|S| = k+1$, since we can then use induction on k. Let $S = V(C) \cup \{x\}$ for some variable $x \notin V(C)$. Then $U_S(C)$ contains two terms: $C_0 := C \wedge \overline{x}$ and $C_1 := C \wedge x$. By the law of total probability, we have $\Pr_{\mathcal{D}}[C] = \Pr_{\mathcal{D}}[C_0] + \Pr_{\mathcal{D}}[C_1]$ and $\Pr_{\mathcal{U}}[C] = \Pr_{\mathcal{U}}[C_0] + \Pr_{\mathcal{U}}[C_1]$. Hence

$$\min\left\{\frac{\Pr_{\mathcal{D}}[C_0]}{\Pr_{\mathcal{U}}[C_0]}, \frac{\Pr_{\mathcal{D}}[C_1]}{\Pr_{\mathcal{U}}[C_1]}\right\} \leq \frac{\Pr_{\mathcal{D}}[C]}{\Pr_{\mathcal{U}}[C]} \leq \max\left\{\frac{\Pr_{\mathcal{D}}[C_0]}{\Pr_{\mathcal{U}}[C_0]}, \frac{\Pr_{\mathcal{D}}[C_1]}{\Pr_{\mathcal{U}}[C_1]}\right\}.$$

So if C_0 and C_1 are both ε-fooled by $\mathcal{D}$, then C is ε-fooled as well. ■

Given an input $X = x_1 \dots x_N$, recall that $\Delta(x_i, y)$ denotes a term that evaluates to 1 if $x_i = y$, and to 0 if $x_i \neq y$. Then let a *proper k-term* C be a product of the form $\Delta(x_{i_1}, y_1) \cdot \dots \cdot \Delta(x_{i_k}, y_k)$, where $1 \leq i_1 < \dots < i_k \leq N$ and $y_1, \dots, y_k \in [M]$.

We now prove the central fact, that $\mathcal{D}$ is almost k-wise independent.

Lemma 9 *$\mathcal{D}$ is $2k/M$-almost k-wise independent for all $k \leq M/2$.*

Proof. Notice that a Boolean k-term can involve bits from at most k different x_i's. So by Claim 8, to show that any Boolean k-term is ε-fooled by $\mathcal{D}$, it suffices to show that any *proper* k-term

$$C = \Delta(x_{i_1}, y_1) \cdot \dots \cdot \Delta(x_{i_k}, y_k)$$

is ε-fooled by $\mathcal{D}$.

We first upper-bound $\Pr_{\mathcal{D}}[C]$. Recall that to sample an input Z from the distribution $\mathcal{D}$, we first sample an X from $\mathcal{U}$, and then sample Z from $\mathcal{D}(X)$. Suppose $C(X) = 1$. Then the only way we can get $C(Z) = 0$ is if, when we perturb the input X to obtain $\mathcal{D}(X)$, some $\Delta(x_{i_j}, y_j)$ changes from TRUE to FALSE. But for each $j \in [k]$, this can happen only if $y = y_j$, which occurs with probability $1/M$. So by the union bound,

$$\Pr_{\mathcal{D}}[C] \geq \Pr_{\mathcal{U}}[C] \cdot \left(1 - \frac{k}{M}\right).$$

We can similarly upper-bound $\Pr_{\mathcal{U}}[C]$. By Claim 6, to sample an input X from $\mathcal{U}$, we can first sample a Z from $\mathcal{D}$, and then sample X from $\mathcal{D}^{\text{inv}}(Z)$. Suppose $C(Z) = 1$. Then we can only get $C(X) = 0$ if, when we perturb Z to $\mathcal{D}^{\text{inv}}(Z)$, some $\Delta(z_{i_j}, y_j)$ changes from TRUE to FALSE. But each z_i changes with probability at most $1/M$. So by the union bound,

$$\Pr_{\mathcal{U}}[C] \geq \Pr_{\mathcal{D}}[C] \cdot \left(1 - \frac{k}{M}\right).$$

Combining the upper and lower bounds, and using the fact that $k \leq M/2$, we have

$$1 - \frac{k}{M} \leq \frac{\Pr_{\mathcal{D}}[C]}{\Pr_{\mathcal{U}}[C]} \leq 1 + \frac{2k}{M}.$$

■

Combining Lemmas 5, 7, and 9, and recalling that $n = Nm$, we obtain the following.

Theorem 10 *Conjecture 4 (the GLN Conjecture) is false. Indeed, there exists a family of Boolean functions $f_{\text{Surj}} : \{0,1\}^n \rightarrow \{0,1\}$, computable by AC^0 circuits of size $O(n^2)$, depth 3, and bottom fanin $O(\log n)$, as well as an $O((k \log^2 n)/n)$-almost k-wise independent distribution $\mathcal{D}$ over $\{0,1\}^n$, such that $\mathrm{E}_{\mathcal{D}}[f_{\text{Surj}}] - \mathrm{E}_{\mathcal{U}}[f_{\text{Surj}}] = \Omega(1)$.*

4 Random Oracle Separations

In this section, we reuse the function f_{Surj} and distribution $\mathcal{D}$ from Section 3 to show that $(\Pi_2^p)^A \not\subset \mathsf{P}^{\mathsf{NP}^A}$ with probability 1 relative to a random oracle A. The central observation here is simply that $\mathcal{D}$ has support on a *constant* fraction of $[M]^N$—and that therefore, any algorithm that computes $f_{\text{Surj}}(X)$ on a $1 - \varepsilon$ fraction of inputs $X \in [M]^N$ must also distinguish $\mathcal{D}$ from $\mathcal{U}$ with constant bias. The following lemma makes this implication precise.

Lemma 11 *Let B be a random variable such that $\Pr_{\mathcal{U}}[B = f_{\text{Surj}}] \geq 0.92$. Then $\Pr_{\mathcal{U}}[B] - \Pr_{\mathcal{D}}[B] \geq 0.022 - o(1)$.*

Proof. For convenience, let us adopt the convention that all probabilities are implicitly the limiting probabilities as $m \to \infty$; this introduces at most an $o(1)$ additive error. Then $\Pr_{\mathcal{U}}[f_{\text{Surj}}] = 1/e$, so

$$\Pr_{\mathcal{U}}[B] \geq \Pr_{\mathcal{U}}[f_{\text{Surj}}] - \Pr_{\mathcal{U}}[B \neq f_{\text{Surj}}] \geq \frac{1}{e} - 0.08 > 0.287.$$

It remains to upper-bound $\Pr_{\mathcal{D}}[B]$. Using the Poisson approximation, for every fixed integer $k \geq 0$ we have

$$\Pr_{\mathcal{U}}[|\text{Im}_X| = M - k] = \frac{1}{e \cdot k!}.$$

By comparison, for every fixed $k \geq 1$ we have

$$\Pr_{\mathcal{D}}[|\text{Im}_X| = M - k] = \frac{1}{e \cdot (k-1)!}.$$

Now, once we condition on the value of $|\text{Im}_X|$, it is not hard to see that the distributions $\mathcal{D}$ and $\mathcal{U}$ are identical. Thus, since

$$\frac{\Pr_{\mathcal{D}}[|\text{Im}_X| = M - k]}{\Pr_{\mathcal{U}}[|\text{Im}_X| = M - k]} = \frac{e \cdot k!}{e \cdot (k-1)!} = k$$

increases with k, the way to maximize $\Pr_{\mathcal{D}}[B]$ is to set $B = 1$ for those inputs X such that k is as large as possible (in other words, such that $|\text{Im}_X|$ is as small as possible). Notice that

$$\begin{aligned}
\Pr_{\mathcal{U}}[(|\text{Im}_X| < M) \wedge B] &\leq \Pr_{\mathcal{U}}[B \neq f_{\text{Surj}}] \\
&\leq 0.08 \\
&< 1 - \frac{5}{2e} \\
&= \sum_{k=3}^{\infty} \frac{1}{e \cdot k!}.
\end{aligned}$$

It follows that

$$\begin{aligned}
\Pr_{\mathcal{D}}[B] &\leq \sum_{k=3}^{\infty} \Pr_{\mathcal{D}}[|\text{Im}_X| = M - k] \\
&= \sum_{k=3}^{\infty} \frac{1}{e \cdot (k-1)!} \\
&= 1 - \frac{2}{e} \\
&< 0.265.
\end{aligned}$$

Combining,

$$\Pr_{\mathcal{U}}[B] - \Pr_{\mathcal{D}}[B] > 0.287 - 0.265 = 0.022.$$

■

Recall that Lemma 9 showed the distribution $\mathcal{D}$ to be $2k/M$-almost k-wise independent. Examining the proof of Lemma 9, we can actually strengthen the conclusion to the following.

Lemma 12 *Let F be a k-DNF formula, with $k \leq M/2$. Then*

$$1 - \frac{k}{M} \leq \frac{\Pr_{\mathcal{D}}[F]}{\Pr_{\mathcal{U}}[F]} \leq 1 + \frac{2k}{M}.$$

Proof. Let $F = C_1 \vee \cdots \vee C_\ell$. Fix an input $X \in [M]^N$, and suppose $F(X) = 1$. Then there must be an $i \in [\ell]$ such that $C_i(X) = 1$. In the proof of Lemma 9, we actually showed that

$$\Pr_{\mathcal{D}(X)}[C_i] \geq 1 - \frac{k}{M}.$$

It follows that

$$\Pr_{\mathcal{D}(X)}[F] \geq 1 - \frac{k}{M},$$

and hence

$$\Pr_{\mathcal{D}}[F] \geq \Pr_{\mathcal{U}}[F] \cdot \left(1 - \frac{k}{M}\right).$$

Similarly,

$$\Pr_{\mathcal{U}}[F] \geq \Pr_{\mathcal{D}}[F] \cdot \left(1 - \frac{k}{M}\right).$$

The lemma now follows, using the assumption $k \leq M/2$. ■

By combining Lemma 12 with the standard diagonalization tricks of Bennett and Gill [5], we can now prove a random oracle separation between Π_2^p and P^{NP}.

Theorem 13 $(\Pi_2^p)^A \not\subset \mathsf{P}^{\mathsf{NP}^A}$ *with probability* 1 *relative to a random oracle A.*

Proof. We will treat the random oracle A as encoding, for each positive integer m, a random sequence of integers $X_m \in [M]^N$, where $M := 2^m$ and $N := \lceil Mm \ln 2 \rceil$. Let $f_{\text{Surj}} : [M]^N \to \{0,1\}$ be our usual surjectivity function; i.e. $f_{\text{Surj}}(X_m) = 1$ if and only if $\text{Im}_{X_m} = [M]$. Then let L be a unary language that contains 0^m if and only if $f_{\text{Surj}}(X_m) = 1$. Clearly $L \in (\Pi_2^p)^A$. It remains to show that $L \notin \mathsf{P}^{\mathsf{NP}^A}$ with probability 1 over A. Fix a $\mathsf{P}^{\mathsf{NP}^A}$ machine B^A, which runs in time $p(m)$ for some fixed polynomial p. Also, let $m_1, m_2, \ldots$ be a sequence of input lengths that are exponentially far apart, so that we do not need to worry about $B^A(0^{m_i})$ querying X_{m_j} for any $j > i$. We will treat X_{m_j} as fixed for all $j < i$, so that only $X := X_m := X_{m_i}$ itself is a random variable. Then $B^A(0^m)$ makes a sequence of at most $p(m)$ adaptive $\mathsf{NTIME}(p(m))$ queries to X, call them $Q_1, \ldots, Q_{p(m)}$. For each $t \in [p(m)]$, we can write a $p(m)$-DNF formula $F_t(X)$ which evaluates to TRUE if and only if $Q_t(X)$ accepts. Then by Lemma 12, we have

$$1 - \frac{p(m)}{M} \leq \frac{\Pr_{\mathcal{D}}[F_t]}{\Pr_{\mathcal{U}}[F_t]} \leq 1 + \frac{2p(m)}{M}.$$

This implies that

$$\left|\Pr_{\mathcal{D}}[F_t] - \Pr_{\mathcal{U}}[F_t]\right| \leq \frac{2p(m)}{M}.$$

So by the union bound, we have

$$\left|\Pr_{\mathcal{D}}\left[B^A(0^m)\right] - \Pr_{\mathcal{U}}\left[B^A(0^m)\right]\right| \leq \frac{2p(m)^2}{M},$$

even after we take into account the possible adaptivity of the queries. Clearly $2p(m)^2/M < 0.022$ for all sufficiently large m. So taking the contrapositive of Lemma 11,

$$\Pr_A\left[B^A(0^m) = f(X)\right] < 0.92$$

for all sufficiently large M. So as in the standard random oracle argument of Bennett and Gill [5], we have

$$\Pr_A\left[B^A \text{ decides } L\right] \leq \prod_{i=1}^{\infty} \Pr_A\left[B^A(0^{m_i}) = f(X_{m_i})\right] = 0.$$

Then taking the union bound over all $\mathsf{P}^{\mathsf{NP}^A}$ machines B^A,

$$\Pr_A\left[L \in \mathsf{P}^{\mathsf{NP}^A}\right] = 0$$

as well. ■

It is well-known that $\mathsf{P}^{\mathsf{NP}^A} = \mathsf{BPP}^{\mathsf{NP}^A}$ with probability 1 relative to a random oracle A. Thus, Theorem 13 immediately implies that $(\Pi_2^p)^A \not\subset \mathsf{BPP}^{\mathsf{NP}^A}$ relative to a random oracle A as well. Since the class $\mathsf{BPP}_{\mathsf{path}}$ is contained in $\mathsf{BPP}^{\mathsf{NP}}$ (as shown by Han, Hemaspaandra, and Thierauf [11]), we also obtain the new result that $(\Pi_2^p)^A \not\subset \mathsf{BPP}_{\mathsf{path}}^A$ relative to a random oracle A.

5 Implications for AC^0

In this section, we discuss two implications of our counterexample for AC^0 functions.

(1) Linial, Mansour, and Nisan [12] famously showed that every AC^0 function has average sensitivity $O(\text{polylog}\, n)$. By contrast, we show in Section 5.1 that there are reasonably-balanced AC^0 functions with average *block-sensitivity* almost linear in n (on both 0-inputs and 1-inputs). In other words, there exist AC^0 functions that counterintuitively behave almost like the PARITY function in terms of block-sensitivity!

(2) Linial et al. [12] also showed that every AC^0 function can be approximated in L_2-norm by a low-degree polynomial. By contrast, we show in Section 5.2 that there does not generally exist such a polynomial that *also* satisfies a reasonable sparseness condition on the coefficients (what Aaronson [1] called the "low-fat" condition).

5.1 The Average Block-Sensitivity of AC^0

Let us first recall the definition of average sensitivity.

Definition 14 (average sensitivity) *Given a string $X \in \{0,1\}^n$ and coordinate $i \in [n]$, let X^i denote X with the i^{th} bit flipped. Then given a Boolean function $f : \{0,1\}^n \to \{0,1\}$, the sensitivity of f at X, or $\mathrm{s}_X(f)$, is the number of i's such that $f\left(X^i\right) \neq f(X)$. Then the average sensitivity of f is*

$$\overline{\mathrm{s}}(f) := \mathop{\mathrm{E}}_{X \in \{0,1\}^n}\left[\mathrm{s}_X(f)\right].$$

Assuming f is non-constant, we can also define the average 0-sensitivity $\overline{\mathrm{s}}_0(f)$ and average 1-sensitivity $\overline{\mathrm{s}}_1(f)$ respectively, by

$$\overline{\mathrm{s}}_b(f) := \mathop{\mathrm{E}}_{X \in \{0,1\}^n :\ f(X)=b}\left[\mathrm{s}_X(f)\right].$$

Then Linial, Mansour, and Nisan [12] showed that *every* AC^0 *function has low average sensitivity:*

Theorem 15 ([12]) *Every Boolean function* $f : \{0,1\}^n \rightarrow \{0,1\}$ *computed by an* AC^0 *circuit of depth* d *satisfies* $\overline{\mathrm{s}}(f) = O\left(\log^d n\right)$.

We now recall the definition of *block-sensitivity*, a natural generalization of sensitivity introduced by Nisan [14].

Definition 16 (average block-sensitivity) *Given a string* $X \in \{0,1\}^n$ *and a subset of indices* $B \subseteq [n]$ *(called a "block"), let* X^B *denote* X *with the bits in* B *flipped. Then given a Boolean function* $f : \{0,1\}^n \rightarrow \{0,1\}$*, the block-sensitivity of* f *at* X*, or* $\mathrm{bs}_X(f)$*, is the largest* k *for which there exist* k *pairwise-disjoint blocks,* $B_1, \ldots, B_k$*, such that* $f\left(X^{B_i}\right) \neq f(X)$ *for all* $i \in [k]$*. Then the average block-sensitivity of* f *is*

$$\overline{\mathrm{bs}}(f) := \mathop{\mathrm{E}}_{X \in \{0,1\}^n} \left[\mathrm{bs}_X(f)\right].$$

Assuming f *is non-constant, we can also define the average* 0*-block-sensitivity* $\overline{\mathrm{bs}}_0(f)$ *and average* 1*-block-sensitivity* $\overline{\mathrm{bs}}_1(f)$ *respectively, by*

$$\overline{\mathrm{bs}}_b(f) := \mathop{\mathrm{E}}_{X \in \{0,1\}^n \,:\, f(X)=b} \left[\mathrm{bs}_X(f)\right].$$

We consider the following question: *does any analogue of Theorem 15 still hold if we replace sensitivity by block-sensitivity?*

We start with some simple observations. Call a Boolean function $f : \{0,1\}^n \rightarrow \{0,1\}$ *reasonably-balanced* if there exist constants $a, b \in (0,1)$ such that $a \leq \mathrm{E}_{\{0,1\}^n}[f] \leq b$ for every n. Then if we do not require f to be reasonably-balanced, it is easy to find an $f \in \mathsf{AC}^0$ such that $\overline{\mathrm{bs}}_0(f)$ and $\overline{\mathrm{bs}}_1(f)$ are both large. For example, the two-level AND-OR tree satisfies $\overline{\mathrm{bs}}_0(f) = \Theta(\sqrt{n})$ and $\overline{\mathrm{bs}}_1(f) = \Theta(\sqrt{n})$.

So let us require f to be reasonably-balanced. Even then, it is easy to find an $f \in \mathsf{AC}^0$ such that $\overline{\mathrm{bs}}(f) = \Omega(n/\log n)$. Given an input $X = x_1 \ldots x_N \in [N]^N$, define the *Tribes function* by $f_{\mathrm{Tribes}}(X) = 1$ if there exists an $i \in [N]$ such that $x_i = 1$, and $f_{\mathrm{Tribes}}(X) = 0$ otherwise. Then not only is f_{Tribes} in AC^0, it has an AC^0 circuit of depth 2 (i.e., a DNF formula). On the other hand, let X be any 0-input of f_{Tribes}; then we can change X to a 1-input by setting $x_i := 1$ for any i. So

$$\mathrm{bs}_X(f_{\mathrm{Tribes}}) \geq N = \Omega\left(\frac{n}{\log n}\right),$$

where $n := N \log_2 N$ is the bit-length of X. Hence $\overline{\mathrm{bs}}_0(f_{\mathrm{Tribes}}) = \Omega(n/\log n)$. Indeed $\overline{\mathrm{bs}}(f_{\mathrm{Tribes}}) = \Omega(n/\log n)$ as well, since

$$\lim_{N \rightarrow \infty} \Pr_X \left[f_{\mathrm{Tribes}}(X) = 0\right] = \frac{1}{e}.$$

By contrast, one can check that $\overline{\mathrm{bs}}_1(f_{\mathrm{Tribes}})$ is only $\Theta(\log n)$. Indeed, *any* Boolean function f that can be represented by a k-DNF formula satisfies $\overline{\mathrm{bs}}_1(f) \leq k$, since if a particular k-term C is satisfied, then there are at most k disjoint ways to make it unsatisfied.

The above observations might lead one to ask the following question: *does every reasonably-balanced* AC^0 *function* f *satisfy either* $\overline{\mathrm{bs}}_0(f) = O(\mathrm{polylog}\, n)$ *or* $\overline{\mathrm{bs}}_1(f) = O(\mathrm{polylog}\, n)$*?* We now show, alas, that the answer is still no.

Theorem 17 *There exists a reasonably-balanced Boolean function $f : \{0,1\}^n \rightarrow \{0,1\}$, computable by a depth-three AC^0 circuit, such that $\overline{\mathrm{bs}}_0(f) = \Omega(n/\log n)$ and $\overline{\mathrm{bs}}_1(f) = \Omega(n/\log n)$.*

Proof. Let f be the function f_{Surj} from our counterexample. As usual, we can think of an input X to f_{Surj} as belonging to either or $\{0,1\}^n$ or $[M]^N$, where $M = 2^m$, $N = \lceil Mm \ln 2 \rceil$, and $n = Nm$. As in Lemma 7, we have

$$\lim_{M \rightarrow \infty} \mathop{\mathrm{E}}_{[M]^N} [f_{\mathrm{Surj}}] = \frac{1}{e},$$

so f_{Surj} is reasonably-balanced.

To lower-bound $\overline{\mathrm{bs}}_1(f_{\mathrm{Surj}})$, consider an input $X = x_1 \ldots x_N \in [M]^N$ such that $f_{\mathrm{Surj}}(X) = 1$ or equivalently $\mathrm{Im}_X = [M]$. Given $y \in [M]$, let $C_y(X)$ be the set of all $i \in [N]$ such that $x_i = y$. Then we can change $f_{\mathrm{Surj}}(X)$ from 1 to 0, by changing x_i to an arbitrary element of $[M] \setminus \{y\}$ for each $i \in C_y(X)$. This implies that $\mathrm{bs}_X(f_{\mathrm{Surj}}) \geq M$. Indeed, we can improve the bound to $\mathrm{bs}_X(f_{\mathrm{Surj}}) \geq Mm$, by noticing that it suffices to change a single *bit* of x_i for each $i \in C_y(X)$. Hence

$$\overline{\mathrm{bs}}_1(f_{\mathrm{Surj}}) \geq Mm = \Omega\left(\frac{n}{\log n}\right).$$

Next consider an input $X = x_1 \ldots x_N \in [M]^N$ such that $|\mathrm{Im}_X| = M-1$. Then clearly $f_{\mathrm{Surj}}(X) = 0$. Let $A(X)$ be the set of indices $i \in [N]$ for which there exists at least one $j \neq i$ such that $x_i = x_j$. Then we have $|A(X)| \geq N - M$ by the pigeonhole principle. Also, for any $i \in A(X)$, let X^i be identical to X, except that we change x_i to the unique element of $[M] \setminus \mathrm{Im}_X$. Then clearly $\mathrm{Im}_{X^i} = [M]$ and $f_{\mathrm{Surj}}(X^i) = 1$. Therefore $\mathrm{bs}_X(f_{\mathrm{Surj}}) \geq |A(X)| \geq N - M$. Furthermore, as in Lemma 11, we have

$$\lim_{M \rightarrow \infty} \Pr_{[M]^N} [|\mathrm{Im}_X| = M - 1] = \frac{1}{e}$$

by the Poisson approximation. It follows that

$$\lim_{M \rightarrow \infty} \overline{\mathrm{bs}}_0(f_{\mathrm{Surj}}) \geq \frac{1/e}{1 - 1/e}(N - M) = \Omega\left(\frac{n}{\log n}\right).$$

■

5.2 The Inapproximability of AC^0 by Low-Fat Polynomials

Let us recall another basic result of Linial, Mansour, and Nisan [12].

Theorem 18 ([12]) *Let $f : \{0,1\}^n \rightarrow \{0,1\}$ be computed by an AC^0 circuit of depth d. Then for all $\varepsilon > 0$, there exists a multilinear polynomial $p : \{0,1\}^n \rightarrow \mathbb{R}$ of degree $O\left(\log^d(n/\varepsilon)\right)$ such that $\mathrm{E}_{\mathcal{U}}\left[(p-f)^2\right] \leq \varepsilon$.*

In this section, we ask whether one can extend Theorem 18 to get an approximating polynomial p that is not merely low-degree, but also representable using coefficients that are bounded in absolute value. The specific property that we want was called the "low-fat" property by Aaronson [1]:

Definition 19 (low-fat polynomials) *Given a multilinear polynomial $p : \{0,1\}^n \rightarrow \mathbb{R}$, define the fat content of p, or $\mathrm{fat}(p)$, to be the minimum of $\sum_C |\alpha_C| \, 2^{-|C|}$ over all representations $p = \sum_C \alpha_C C$ of p as a linear combination of terms (that is, products of x_i's and $(1 - x_i)$'s). Then we call p low-fat if $\mathrm{fat}(p) = n^{o(1)}$.*

One motivation for Definition 19 comes from [1], where it was pointed out that the Generalized Linial-Nisan Conjecture is *equivalent* (via linear programming duality) to the following conjecture:

Conjecture 20 (Low-Fat Sandwich Conjecture) *Let $f : \{0,1\}^n \rightarrow \{0,1\}$ be computed by an AC^0 circuit of size $2^{n^{o(1)}}$. Then there exist low-fat multilinear polynomials $p_\ell, p_u : \{0,1\}^n \rightarrow \mathbb{R}$, of degree $n^{o(1)}$, that "sandwich" f in the following sense:*

(i) $p_\ell(X) \leq f(X) \leq p_u(X)$ *for all* $X \in \{0,1\}^n$ *and*

(ii) $\mathrm{E}_{\mathcal{U}}[p_u - p_\ell] = o(1)$.

Without the adjective "low-fat," Conjecture 20 would be equivalent to the *original* Linial-Nisan Conjecture, as shown by Bazzi [2]. And indeed, Braverman [7] heavily exploited this equivalence in his proof of the original LN Conjecture.[6]

Of course, from the fact that the GLN Conjecture is false, we can immediately deduce that Conjecture 20 is false as well.

On the other hand, the notion of low-fat polynomials seems interesting even apart from Conjecture 20—for the low-fat condition is a kind of "sparseness" condition, which might be useful (for example) in learning theory. Furthermore, the falsehood of Conjecture 20 does not directly rule out the possibility of low-fat approximating polynomials for every AC^0 function, since Conjecture 20 talks only about *sandwiching* polynomials. However, with a bit more work, we now show the existence of an AC^0 function that has no low-fat, low-degree approximating polynomial of any kind.

Theorem 21 *There exists a Boolean function $f : \{0,1\}^n \rightarrow \{0,1\}$, computable by a depth-three AC^0 circuit, for which any multilinear polynomial $p : \{0,1\}^n \rightarrow \mathbb{R}$ that satisfies $\mathrm{E}_{\mathcal{U}}\left[(p-f)^2\right] = o(1)$ also satisfies $\deg(p)\,\mathrm{fat}(p) = \Omega\left(n/\log^2 n\right)$.*

Proof. Once again we let $f = f_{\mathrm{Surj}}$. Let p be a multilinear polynomial such that $\mathrm{E}_{\mathcal{U}}\left[(p-f)^2\right] = \varepsilon$. By definition, we can write p as a linear combination of terms, $p = \sum_C \alpha_C C$, such that $\sum_C |\alpha_C| \mathrm{E}_{\mathcal{U}}[C] = \mathrm{fat}(p)$. Hence

$$\begin{aligned}
\mathop{\mathrm{E}}_{\mathcal{U}}[p] - \mathop{\mathrm{E}}_{\mathcal{D}}[p] &= \sum_C \alpha_C \left(\mathop{\mathrm{E}}_{\mathcal{U}}[C] - \mathop{\mathrm{E}}_{\mathcal{D}}[C] \right) \\
&\leq \sum_C |\alpha_C| \left(\frac{2|C|}{M} \mathop{\mathrm{E}}_{\mathcal{U}}[C] \right) \\
&\leq \frac{2\,\mathrm{fat}(p)\deg(p)}{M},
\end{aligned}$$

[6] Technically, Braverman constructed polynomials that satisfied slightly different properties than (i) and (ii) from Conecture 20. However, we know from Bazzi's equivalence theorem [2] that it must be possible to satisfy those properties as well.

where the second line follows from Lemma 9. Also, let $\Delta := p - f_{\text{Surj}}$. Then as in the proof of Lemma 11, we have

$$\begin{aligned}
\varepsilon &= \mathop{\mathrm{E}}_{\mathcal{U}}\left[\Delta^2\right] \\
&= \sum_{k=0}^{M} \Pr_{\mathcal{U}}\left[|\mathrm{Im}_X| = M-k\right] \cdot \mathop{\mathrm{E}}_{\mathcal{U}}\left[\Delta^2 \mid |\mathrm{Im}_X| = M-k\right] \\
&\geq \sum_{k=0}^{M} \frac{\mathrm{E}_{\mathcal{U}}\left[\Delta^2 \mid |\mathrm{Im}_X| = M-k\right]}{e \cdot k!} - o\left(1\right),
\end{aligned}$$

whereas

$$\begin{aligned}
\mathop{\mathrm{E}}_{\mathcal{D}}\left[\Delta^2\right] &= \sum_{k=0}^{M} \Pr_{\mathcal{D}}\left[|\mathrm{Im}_X| = M-k\right] \cdot \mathop{\mathrm{E}}_{\mathcal{D}}\left[\Delta^2 \mid |\mathrm{Im}_X| = M-k\right] \\
&\leq \sum_{k=0}^{M} \frac{\mathrm{E}_{\mathcal{U}}\left[\Delta^2 \mid |\mathrm{Im}_X| = M-k\right]}{e \cdot (k-1)!} + o\left(1\right).
\end{aligned}$$

Combining, we find that

$$\mathop{\mathrm{E}}_{\mathcal{D}}\left[\Delta^2\right] = O\left(\varepsilon \log \frac{1}{\varepsilon}\right) + o\left(1\right).$$

Hence

$$\begin{aligned}
\mathop{\mathrm{E}}_{\mathcal{U}}\left[f_{\text{Surj}}\right] - \mathop{\mathrm{E}}_{\mathcal{D}}\left[f_{\text{Surj}}\right] &= \left(\mathop{\mathrm{E}}_{\mathcal{U}}\left[p\right] - \mathop{\mathrm{E}}_{\mathcal{U}}\left[\Delta\right]\right) - \left(\mathop{\mathrm{E}}_{\mathcal{D}}\left[p\right] - \mathop{\mathrm{E}}_{\mathcal{D}}\left[\Delta\right]\right) \\
&\leq \left(\mathop{\mathrm{E}}_{\mathcal{U}}\left[p\right] - \mathop{\mathrm{E}}_{\mathcal{D}}\left[p\right]\right) + \mathop{\mathrm{E}}_{\mathcal{U}}\left[\Delta\right] + \mathop{\mathrm{E}}_{\mathcal{D}}\left[\Delta\right] \\
&\leq \frac{2\,\mathrm{fat}\left(p\right)\deg\left(p\right)}{M} + \sqrt{\mathop{\mathrm{E}}_{\mathcal{U}}\left[\Delta^2\right]} + \sqrt{\mathop{\mathrm{E}}_{\mathcal{D}}\left[\Delta^2\right]} \\
&\leq \frac{2\,\mathrm{fat}\left(p\right)\deg\left(p\right)}{M} + O\left(\sqrt{\varepsilon \log \frac{1}{\varepsilon}}\right) + o\left(1\right),
\end{aligned}$$

where the third line follows from Cauchy-Schwarz. On the other hand, we know from Lemma 7 that

$$\mathop{\mathrm{E}}_{\mathcal{U}}\left[f_{\text{Surj}}\right] - \mathop{\mathrm{E}}_{\mathcal{D}}\left[f_{\text{Surj}}\right] \geq \frac{1}{e} - o\left(1\right).$$

So combining, if $\varepsilon = o\left(1\right)$, then

$$\mathrm{fat}\left(p\right)\deg\left(p\right) = \Omega\left(\frac{M}{e}\right) = \Omega\left(\frac{n}{\log^2 n}\right).$$

■

Since f_{Surj} has a depth-three AC^0 circuit, it follows from Theorem 18 that there exists a polynomial p of degree $O\left(\log^3 n\right)$ such that $\mathrm{E}_{\mathcal{U}}\left[\left(p - f_{\text{Surj}}\right)^2\right] = o\left(1\right)$. Thus, one corollary of Theorem 21 is a *separation* between low-degree approximation and low-degree low-fat approximation. In other words, there exists a Boolean function f (namely f_{Surj}) that can be well-approximated in L_2-norm by a polynomial of degree $O\left(\mathrm{polylog}\, n\right)$, but *not* by a low-fat polynomial of degree $O\left(\mathrm{polylog}\, n\right)$. This answers one of the open problems from [1].

6 Discussion

As we said before, we remain sanguine about the prospects for proving an oracle separation between BQP and PH. In our view, the lesson of our counterexample is simply that almost k-wise independence is too blunt of an instrument for this problem. Looking at the specific function f_{Surj} in the counterexample, we find two arguments in support of this position. Firstly, f_{Surj} is extremely different in character from Fourier Checking, or any of the other candidates for problems in $\mathsf{BQP} \setminus \mathsf{PH}$ (such as the ones studied by Fefferman and Umans [10]). Indeed, f_{Surj} is not even in BQP, as can be seen from the BBBV lower bound [4] for example.[7] Secondly, f_{Surj} is trivially in PH by construction—and for that reason, our counterexample does not really say anything unexpected about "the power of PH." To us, the unexpected part is simply the inability of *approximate local statistics* to "certify" a problem as outside PH, where *exact* local statistics succeed in doing so (as shown by Braverman [7]). But this is a surprise about proof techniques, not about complexity classes.

The obvious open problems are

(1) to solve the relativized BQP versus PH problem by whatever means, and

(2) to solve the relativized BQP versus AM problem, *possibly* by proving the depth-two GLN Conjecture.

We reiterate our offer of a $200 prize for problem (1) and a $100 prize for problem (2).

A third interesting problem is to show that our function $f_{\text{Surj}}(X)$ cannot be computed in Σ_2^p, on a $1-\varepsilon$ fraction of inputs $X \in [M]^N$. This would imply that $(\Pi_2^p)^A \not\subset (\Sigma_2^p)^A$ with probability 1 relative to a random oracle A. A fourth problem is whether one can say *anything* nontrivial about the block-sensitivity of AC^0 functions: for example, that every $f \in \mathsf{AC}^0$ has average block-sensitivity $O(n/\log n)$.

7 Acknowledgments

I thank Paul Beame for sharing a draft of the manuscript [3] with me, and Lane Hemaspaandra for pointing me to the paper [8] of Cai.

References

[1] S. Aaronson. BQP and the polynomial hierarchy. In *Proc. ACM STOC*, 2010. arXiv:0910.4698.

[2] L. Bazzi. Polylogarithmic independence can fool DNF formulas. In *Proc. IEEE FOCS*, pages 63–73, 2007.

[3] P. Beame and W. Machmouchi. The quantum query complexity of AC^0. Manuscript, 2010.

[4] C. Bennett, E. Bernstein, G. Brassard, and U. Vazirani. Strengths and weaknesses of quantum computing. *SIAM J. Comput.*, 26(5):1510–1523, 1997. quant-ph/9701001.

[7] Using the BBBV lower bound, one can show further that no BQP machine can distinguish the distributions $\mathcal{D}$ and $\mathcal{U}$ with constant bias.

[5] C. H. Bennett and J. Gill. Relative to a random oracle A, $P^A \neq NP^A \neq coNP^A$ with probability 1. *SIAM J. Comput.*, 10(1):96–113, 1981.

[6] R. V. Book. On collapsing the polynomial-time hierarchy. *Inform. Proc. Lett.*, 52(5):235–237, 1994.

[7] M. Braverman. Poly-logarithmic independence fools AC^0 circuits. In *Proc. IEEE Conference on Computational Complexity*, pages 3–8, 2009. ECCC TR09-011.

[8] J.-Y. Cai. Probability one separation of the Boolean Hierarchy. In *Proc. Intl. Symp. on Theoretical Aspects of Computer Science (STACS)*, pages 148–158, 1987.

[9] J.-Y. Cai. With probability one, a random oracle separates PSPACE from the polynomial-time hierarchy. 38(1):68–85, 1989.

[10] B. Fefferman and C. Umans. Pseudorandom generators and the BQP vs. PH problem. http://www.cs.caltech.edu/~umans/papers/FU10.pdf, 2010.

[11] Y. Han, L. Hemaspaandra, and T. Thierauf. Threshold computation and cryptographic security. *SIAM J. Comput.*, 26(1):59–78, 1997.

[12] N. Linial, Y. Mansour, and N. Nisan. Constant depth circuits, Fourier transform, and learnability. *J. ACM*, 40(3):607–620, 1993.

[13] N. Linial and N. Nisan. Approximate inclusion-exclusion. *Combinatorica*, 10(4):349–365, 1990. Earlier version in STOC'90.

[14] N. Nisan. CREW PRAMs and decision trees. *SIAM J. Comput.*, 20(6):999–1007, 1991.

[15] N. Nisan and A. Wigderson. Hardness vs. randomness. *J. Comput. Sys. Sci.*, 49(2):149–167, 1994.

[16] A. A. Razborov. A simple proof of Bazzi's theorem. *ACM Trans. on Computation Theory*, 1(1), 2009. ECCC TR08-081.

[17] J. Håstad. *Computational Limitations for Small Depth Circuits*. MIT Press, 1987.

[18] A. C-C. Yao. Separating the polynomial-time hierarchy by oracles (preliminary version). In *Proc. IEEE FOCS*, pages 1–10, 1985.