Parity vs. $\mathbf{AC^0}$ with Simple Quantum Preprocessing

Joseph Slote

Department of Computing and Mathematical Sciences, California Institute of Technology, Pasadena, CA, USA

Abstract

A recent line of work [5, 27, 12, 6, 28] has shown the unconditional advantage of constant-depth quantum computation, or $\mathbf{QNC^0}$, over $\mathbf{NC^0}$, $\mathbf{AC^0}$, and related models of classical computation. Problems exhibiting this advantage include search and sampling tasks related to the parity function, and it is natural to ask whether $\mathbf{QNC^0}$ can be used to help compute parity itself. Namely, we study $\mathbf{AC^0} \circ \mathbf{QNC^0}$ – a hybrid circuit model where $\mathbf{AC^0}$ operates on measurement outcomes of a $\mathbf{QNC^0}$ circuit – and we ask whether $\mathbf{PAR} \in \mathbf{AC^0} \circ \mathbf{QNC^0}$.

We believe the answer is negative. In fact, we conjecture $\mathbf{AC^0} \circ \mathbf{QNC^0}$ cannot even achieve $\Omega(1)$ correlation with parity. As evidence for this conjecture, we prove:

- When the $\mathbf{QNC^0}$ circuit is ancilla-free, this model can achieve only negligible correlation with parity, even when $\mathbf{AC^0}$ is replaced with any function having LMN-like decay in its Fourier spectrum.
- For the general (non-ancilla-free) case, we show via a connection to nonlocal games that the conjecture holds for any class of postprocessing functions that has approximate degree $o(n)$ and is closed under restrictions. Moreover, this is true even when the $\mathbf{QNC^0}$ circuit is given arbitrary quantum advice. By known results [8], this confirms the conjecture for linear-size $\mathbf{AC^0}$ circuits.
- Another approach to proving the conjecture is to show a switching lemma for $\mathbf{AC^0} \circ \mathbf{QNC^0}$. Towards this goal, we study the effect of quantum preprocessing on the decision tree complexity of Boolean functions. We find that from the point of view of decision tree complexity, nonlocal channels are no better than randomness: a Boolean function $f$ precomposed with an $n$-party nonlocal channel is together equal to a randomized decision tree with worst-case depth at most $\text{DT}_{\text{depth}}[f]$.

Taken together, our results suggest that while $\mathbf{QNC^0}$ is surprisingly powerful for search and sampling tasks, that power is “locked away” in the global correlations of its output, inaccessible to simple classical computation for solving decision problems.

2012 ACM Subject Classification Theory of computation → Quantum computation theory; Theory of computation → Circuit complexity

Keywords and phrases QNC0, AC0, Nonlocal games, k-wise indistinguishability, approximate degree, switching lemma, Fourier concentration

Digital Object Identifier 10.4230/LIPIcs.ITCS.2024.92


Funding Joseph Slote: supported by Chris Umans’ Simons Foundation Investigator Grant.

Acknowledgements We are grateful to Chris Umans and Thomas Vidick for numerous valuable discussions and for the opportunity to share this work with Henry Yuen and the quantum group at Columbia University in the fall of 2022. We are also grateful to Atul Singh Arora, discussions with whom inspired this project. Finally we thank the anonymous ITCS 2024 reviewers for their generous and meticulous feedback on an earlier draft.

Introduction

In 2017, Bravyi, Gosset, and König [5] proved a breakthrough unconditional separation between constant-depth quantum circuits, or $\mathbf{QNC^0}$, and constant-depth bounded fan-in classical circuits, or $\mathbf{NC^0}$. The authors showed that for a certain relational problem solvable...
by \(QNC^0\) circuits, any randomized \(NC^0\) circuit solving the same problem with high probability must have logarithmic depth. The realization that unconditional proofs of quantum advantage were possible – albeit over weak models of classical computation – inspired an exciting series of results strengthening and generalizing the work of Bravyi, Gosset, and König. There are now separations against stronger classical circuit models such as constant depth circuits with unbounded fan-in, or \(AC^0\) \([27]\), average-case separations \([10]\), separations between more intricate interactive models \([12]\), separations that remain even for quantum circuits subject to noise (e.g., \([6]\)), and separations for sampling problems with no input \([28]\), among others.

Although these separations are for comparatively weak models of computation, they are concrete non-oracle, non-query separations, and are free from complexity-theoretic assumptions, making them important companions to the query complexity and conditional separations studied since the founding of quantum computer science. One notable feature of these \(QNC^0\) separations, however, is that they are all for search or sampling problems; decision separations appear to be absent from this list.

On the surface, there is a somewhat trivial reason for this: \(QNC^0\) cannot solve interesting decision problems alone. Indeed, any single output qubit in a constant-depth quantum circuit can only depend on constantly-many input qubits, so any \(QNC^0\) circuit with one output bit may be simulated by randomized \(NC^0\). However, this “lightcone barrier” may be removed by instead measuring all qubits in the quantum circuit and then applying a classical Boolean function \(f\) to the result. As long as \(f\) depends on all of its inputs, it might be possible for \(f\) to leverage \(QNC^0\)’s search and sampling prowess for decision-making ends. Given Bene Watts et al.’s search separation between \(QNC^0\) and \(AC^0\) \([27]\), a natural class of Boolean functions to choose for this postprocessing is \(AC^0\) itself. This gives rise to the following definition, which does not appear to have been studied before.

\[\text{Definition 1.}\] Let \(AC^0 \circ QNC^0\) denote the model of computation composed of a \(QNC^0\) circuit \(C\), followed by a computational basis measurement, and then an \(AC^0\) function \(f\) applied to the result. This process defines the randomized Boolean function \(f \circ C : \{0,1\}^n \rightarrow \mathcal{M}(\{-1,1\})\) from the hypercube to the set \(\mathcal{M}(\{-1,1\})\) of probability measures on \(\{-1,1\}\).

In this work we take a \(QNC^0\) circuit to be a polynomial-size constant-depth quantum circuit composed of arbitrary 2-qubit unitary gates. Ancilla qubits are allowed and are initialized in the state \(|0^m\rangle\) for \(m \in \text{poly}(n)\). No geometric locality or clean computation constraints are assumed. A formal definition appears later as Definition 12.

Certainly \(QNC^0 \subseteq AC^0 \circ QNC^0\), so the search separation between \(QNC^0\) and \(AC^0\) in Bene Watts et al. is also a search separation between \(AC^0 \circ QNC^0\) and \(AC^0\). Moreover, this modification obviates the lightcone barrier mentioned above and allows us to ask meaningful questions about decision separations between concrete models of quantum and classical computation.

Specifically, Bene Watts et al. \([27]\) show exponential advantage of \(QNC^0\) over \(AC^0\) for (a variant of) the “parity halving problem”:

\[\text{Parity halving. Given } x \in \{0,1\}^n \text{ with the promise } |x| \equiv 0 \mod 2, \text{ output any even string if } |x| \equiv 0 \mod 4 \text{ and any odd string otherwise.}\]

Given the form of this problem, it is natural to ask whether parity is itself computable by a hybrid model such as \(AC^0 \circ QNC^0\).

\[\text{Remark 2.}\] Before summarizing our progress on the parity question, we pause to note another reason to study \(AC^0 \circ QNC^0\) coming from the rich subject of quantum-classical interactive proofs. A central project in this area is the classical verification of quantum
computations [11]. In a landmark work, Mahadev gave a cryptographic protocol for this task [17]; however, whether or not this task may be accomplished without cryptographic hardness assumptions remains open despite many efforts [11]. It therefore makes sense to consider the question in simpler contexts, such as where the prover and verifier are replaced with $\text{QNC}^0$ and $\text{AC}^0$ respectively and interact for constantly-many rounds to establish the correctness of a $\text{QNC}^0$ computation. With this perspective we see that $\text{AC}^0 \circ \text{QNC}^0$ models the first round of interaction in such a proof system.

Parity vs. $\text{AC}^0 \circ \text{QNC}^0$: Overview and organization

We conjecture that $\text{AC}^0 \circ \text{QNC}^0$ cannot approximate parity ($\text{Par}$) on average, over both choice of uniformly random input $x \sim \mathcal{U}([0, 1]^n)$ and the randomness in $f \circ C$. It is convenient to take $\text{Par}$ and $f \circ C$ to be $(\pm 1)$-valued and phrase this in terms of the correlation

$$\mathbb{E}_x[(f \circ C)(x) \cdot \text{Par}(x)],$$

proportional to the advantage of $f \circ C$ over random guessing for computing parity.

▶ Conjecture 3. $\text{AC}^0 \circ \text{QNC}^0$ cannot achieve correlation $\Omega(1)$ with the parity function. That is, fix a polynomial size bound $p(n)$ and constant depth $d$. Then for all sequences $\{(f_n, C_n)\}_n$ of circuits such that $\text{size}(f_n), \text{size}(C_n) \leq p(n)$ and $\text{depth}(f_n), \text{depth}(C_n) \leq d$, we have

$$\mathbb{E}_x[(f_n \circ C_n)(x) \cdot \text{Par}_n(x)] \to 0 \quad \text{as} \quad n \to \infty.$$

Although proving correlation bounds against $\text{AC}^0$ is a well-understood topic with many techniques (among them Håstad’s switching lemma [13] and Razborov-Smolensky [22, 25]), when $\text{QNC}^0$ precomputation is added these approaches cannot be used directly. The pursuit of new techniques leads us to connections with many-player nonlocal games, approximate degree bounds, and new directions for generalizing Håstad’s switching lemma. Evidence for Conjecture 3 is laid out as follows.

The ancilla-free case

In Section 1 we prove Conjecture 3 when $\text{QNC}^0$ is restricted to be ancilla-free. A key feature of such $\text{QNC}^0$ circuits is that they correspond to unitary transformations, and we find in this case the correlation of $f \circ C$ with $\text{Par}$ is controlled by the Fourier tail of $f$. Recall the Fourier tail of a Boolean function $f$ is given by

$$W^2[f] := \sum_{|S| \geq 1} \hat{f}(S)^2.$$

Appealing to the Linial-Mansour-Nisan-type (LMN-type) estimates of the Fourier tail of $\text{AC}^0$ [15], we obtain the following strong correlation bound.

▶ Theorem 4 (Ancilla-free $\text{QNC}^0$, general $\text{AC}^0$ case). If $C$ is an ancilla-free $\text{QNC}^0$ circuit and $f$ is an $\text{AC}^0$ function then

$$\mathbb{E}_x[(f \circ C)(x) \cdot \text{Par}_n(x)] \leq 2^{-n/polylog(n)}.$$
Reducing to nonlocal games

To move beyond ancilla-free QNC$^0$, in Section 2 we reduce Conjecture 3 to a question about the value of a certain class of nonlocal games, which we call $n$-player parity games and which are parameterized by a postprocessing Boolean function $f$. Through a connection to the notion of $k$-wise indistinguishability introduced in [4], we show the quantum value of a parity game is controlled by the approximate degree of the associated $f$.

Recall for $\varepsilon > 0$ the $\varepsilon$-approximate degree of a $(0, 1)$-valued Boolean function $f$ is given by

$$\deg_{\varepsilon}[f] = \min \{ \deg(g) \mid g : \{0, 1\}^n \to \mathbb{R} \text{ a polynomial with } \| f - g \|_{\infty} \leq \varepsilon \}.$$ 

Of course, $\deg_{\varepsilon}[f] \leq n$ for any $n$-variate $f$ and $\varepsilon > 0$. By convention $\deg[f] := \deg_{1/3}[f]$. A function class $\mathcal{F} = (\mathcal{F}_n)_{n \geq 1}$ is a sequence of sets of $n$-variate Boolean functions, and we extend approximate degree to function classes via $\deg[\mathcal{F}](n) := \max_{f \in \mathcal{F}_n} \deg[f]$. With this notation, we have the following theorem.

**Theorem 5** (Corollary 20, Section 2). Suppose function class $\mathcal{F}$ is closed under inverse-polynomial-sized restrictions. Then if $\deg[\mathcal{F}] \in o(n)$, $\mathcal{F} \circ \text{QNC}^0$ cannot achieve $\Omega(1)$ correlation with $\text{Par}_n$, even if $\text{QNC}^0$ is given arbitrary quantum advice.

It follows from Theorem 5 that Conjecture 1 would be confirmed in full generality if $\deg[\text{AC}^0] \in o(n)$, a notorious open problem [9]. Such a bound is already known for large subclasses of AC$^0$, however: for example, for AC$^0$ circuits of size $O(n)$ (termed LC$^0$), we may appeal to the recent bounds of [8] to conclude:

**Theorem 6** (General AC$^0$, linear-size AC$^0$ case). Suppose $f \in \text{AC}^0$ has size $O(n)$. Then $f \circ \text{QNC}^0$ achieves correlation at most $1/n$ with $\text{Par}_n$. This holds even if $\text{QNC}^0$ is given arbitrary quantum advice. That is,

$$E[(\text{LC}^0 \circ \text{QNC}^0)/q\text{poly}] \cdot \text{Par}_n \in \text{negl}(n).$$

(This is proved as Corollary 21 in Section 2).

Is the difficulty of proving approximate degree bounds for AC$^0$ a barrier for resolving Conjecture 3? It seems unlikely: the reduction to approximate degree bounds is via a series of substantial relaxations and it would be surprising if all the required converses held. In fact, we conclude Section 2 with a self-contained approximation theory question (Question 1) concerning a notion of blockwise approximate degree which may be easier to solve than $\deg[\text{AC}^0]$ but would still imply Conjecture 3.

Towards an AC$^0 \circ$ QNC$^0$ switching lemma

In Section 3 we chart a different route to resolving Conjecture 3, aiming to prove a switching lemma for our hybrid AC$^0 \circ$ QNC$^0$ circuits. Recall that Håstad’s original switching lemma is used to argue that (very roughly) randomly fixing a large fraction of inputs to an AC$^0$ circuit with high probability yields a function that can be computed by a shallow decision tree. At the same time, PAR retains maximum decision tree complexity under the same restrictions, so this leads to AC$^0$ correlation bounds.

In comparison to Håstad’s switching lemma and its descendants, a challenge with AC$^0 \circ$ QNC$^0$ circuits is that QNC$^0$ can correlate, spread out, and bias random restrictions before they reach the bottom layer of DNFs or CNFs in the AC$^0$ circuit. If QNC$^0$ were

---

1 For $(\pm 1)$-valued $f$, we use the same definition after making the standard identification $+1 \leftrightarrow 0, -1 \leftrightarrow 1$. 
replaced with randomized NC<sup>0</sup> this problem could be readily addressed by considering each deterministic circuit in the distribution, applying standard arguments there, and computing the expected correlation with parity across circuits in the distribution. But unlike randomized computation, and as discussed e.g., in [1], a recurring theme in quantum complexity theory is the impossibility of “pulling out the quantumness” from a quantum circuit.

Contrary to this theme, however, we show that when QNC<sup>0</sup> is replaced by an n-party nonlocal channel N, it is possible to pull out the quantumness in a particular sense:

**Theorem (Theorem 25, restated).** Let f : {0, 1}^m → {0, 1} be any Boolean function and consider an n-party nonlocal channel N, where the i<sup>th</sup> party receives one bit and responds with m<sub>i</sub> ≥ 0 bits, such that \( \sum_{i=1}^{n} m_i = m \). Then the random function f ◦ N is equal to a randomized decision tree Γ such that \( \text{depth}(T) \leq \text{DT}_{\text{depth}}[f] \) for all \( T \in \text{Supp}(\Gamma) \).

(This theorem is proved in Section 3 as Theorem 25.) By an n-party nonlocal channel we mean the channel corresponding to a quantum strategy in an n-player nonlocal game: parties receiving one bit of input each may measure disjoint systems of a shared quantum state as part of their responses but are not allowed to communicate. A formal definition appears as Definition 13. In fact, Theorem 25 is true not only for nonlocal channels, but for any channel where parties obey the no-signaling property; that is, the output of any subset \( S \subset [n] \) of the parties is a function only of the inputs to those parties in \( S \). A formal definition of no-signaling channels appears as Definition 24.

The regime where Theorem 25 is truly interesting is when \( \text{DT}_{\text{depth}}[f] \geq \log(n) \). Then f may depend on all the input coordinates and (potentially) make great use of the processing power afforded by no-signaling channels. Theorem 25 says that to the contrary, precomposition of f by any no-signaling channel has no effect on the (randomized) decision tree complexity of f.

How does Theorem 25 connect to Conjecture 3? As we detail in Section 2, the replacement of QNC<sup>0</sup> by the channel N is essentially without loss of generality from the point of view of Conjecture 3. Unfortunately, however, AC<sup>0</sup> circuits can easily have maximum decision tree complexity, so Theorem 25 cannot be immediately applied. Instead, we believe Theorem 25 stands as a striking example of the inability of classical postprocessing to make use of the search and sampling power of quantum and super-quantum models of computation. Additionally, we hope that this theorem’s proof technique, which involves tracking the interplay between a decision tree for f and the no-signaling channel N, represents the style of argument that could eventually lead to a switching lemma for AC<sup>0</sup> ◦ QNC<sup>0</sup>.

**Outlook**

Taken together, these results suggest QNC<sup>0</sup> cannot render its power in a way AC<sup>0</sup> or other simple models of classical computation can access for the purpose of making decisions. Several questions for further research are posed in Section 4.

**Related work**

Unlike the quantum-classical separations surveyed in the introduction, which show quantum upper bounds and classical lower bounds, this paper aims to prove a lower bound against a concrete model of quantum computation. The pursuit of lower bounds against quantum circuits for computational problems is a nascent area and very little is known.

One quantum circuit model where lower bounds have received some concerted study is QAC<sup>0</sup> [18, 14, 21, 23, 19]. A superset of QNC<sup>0</sup> circuits, QAC<sup>0</sup> additionally allows for arbitrarily-large Toffoli gates,

\[
|x_1, \ldots, x_k, x_{k+1}\rangle \mapsto |x_1, \ldots, x_k, x_{k+1} \oplus (\bigwedge_{i=1}^{k} x_i)\rangle,
\]
which are quantum analogues of classical AND gates with unbounded fan-in. In this setting correlation with parity is also a central open question, and there is growing evidence that QAC⁰ cannot achieve Ω(1) correlation with parity either. Recent work has shown negligible correlation bounds between QAC⁰ and parity when a) the QAC⁰ circuit is restricted to depth 2 [23], and b) when the QAC⁰ circuit is of any depth d and is restricted to O(n^{1/d})-many ancillas [19]. In fact, the second result is a corollary to a Pauli-basis analogue of the LMN theorem for the same subclass of QAC⁰ [19].

The relationship between QAC⁰ and AC⁰ o QNC⁰ is rather unclear, and they are likely incomparable as decision classes. In fact, as far as we know, it is even open whether AC⁰ ⊆ QAC⁰, let alone whether AC⁰ o QNC⁰ ⊆ QAC⁰ (noting the trivial containment AC⁰ ⊆ AC⁰ o QNC⁰).

The difficulty in comparing these models stems from a subtlety concerning the difference between unbounded fan-in and unbounded fan-out when implemented coherently. AC⁰ circuits have no restriction on the fan-out of their gates, while the definition of QAC⁰ appears to strongly limit outward propagation of information. If one augments QAC⁰ with the so-called fan-out gate – which is a CNOT gate with any number of target qubits,

\[ |x_1, \ldots, x_k \rangle \mapsto |x_1, x_1 \oplus x_2, \ldots, x_1 \oplus x_k \rangle, \]

one obtains the circuit model QAC⁰ f, and it is known QAC⁰ f can compute parity exactly in depth 3 [18]. In view of existing lower bounds against QAC, it is expected that QAC⁰ is strictly contained in QAC⁰ f, and assuming this holds we immediately would have that the function version of AC⁰ o QNC⁰ is not in the function version of QAC⁰. This follows, for example, from the fact that multi-output AC⁰ circuits easily implement the classical reversible fan-out gate, \((x_1, \ldots, x_k) \mapsto (x_1, x_1 \oplus x_2, \ldots, x_1 \oplus x_k)\). It is safe to say the interaction of nonlocal gates with QNC⁰ – whether that interaction is coherent as in QAC⁰ and QAC⁰ f, or preceded by measurement as in AC⁰ o QNC⁰ – is only beginning to be understood.

A separate area where quantum circuit lower bounds have been very successfully developed is for state preparation problems. We do not attempt a survey here, but just mention they were crucial to the resolution of the NLTS conjecture [2] and make use of ideas from error correction, which partially originate in sampling lower bounds from classical complexity [16]. However, it is not clear how to transfer these methods to quantum circuit lower bounds for computational problems in the AC⁰ o QNC⁰ model.

1 Lower bounds when QNC⁰ is ancilla-free

Here we show any Boolean function \(f\) with small Fourier tail retains a small top-degree coefficient when composed with ancilla-free QNC⁰. By the celebrated work of Håstad [13] and Linial, Mansour, and Nisan [15], any \(f \in AC^0\) is an example – but this theorem addresses a broader set of functions. On the other hand, as we discuss at the end of the section, once ancillas are allowed, the theorem no longer holds for such a general class of functions.

Recall a function \(f : \{-1, 1\}^n \to \mathbb{R}\) admits a unique Fourier decomposition

\[ f = \sum_{S \subseteq [n]} \hat{f}(S) \chi_S, \]

where \(\chi_S(x) := \prod_{i \in S} x_i\) is the \(S\)th Fourier character (see e.g., [20] for more). We will later make use of the familiar Plancherel theorem, which states for any \(f, g : \{-1, 1\}^n \to \mathbb{R}\) that

\[ \mathbb{E}_x[f(x)g(x)] = \sum_{S \subseteq [n]} \hat{f}(S)\hat{g}(S). \]
Let us briefly connect this perspective to quantum observables. Given a Boolean function $f : \{\pm 1\}^n \to \mathbb{R}$ we define its Von Neumann observable as

$$M_f := \sum_x f(x) \langle x | x \rangle.$$

An identity we will use is

$$M_{\chi_S} = Z^S,$$

where the operator $Z$ here is the Pauli operator $\left( \begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix} \right)$, and generally for any 1-qubit operator $A$ we use the notation

$$A_S := \bigotimes_i \begin{cases} A & \text{if } i \in S \\ \mathbb{I} & \text{otherwise}. \end{cases}$$

Any Von Neumann observable $M$ (that is, any Hermitian operator) has expectation value on state $\rho$ given by

$$\langle M \rangle_{\rho} := \text{tr}[M \rho],$$

and when $M = M_f$ and $x \in \{0, 1\}^n$ we note the identity

$$\langle M_f \rangle_x := \langle M_f \rangle_{|x\rangle\langle x|} = f(x).$$

With this notation, we prove the following.

**Theorem 7** (Correlation bound for ancilla-free QNC$^0$). Let $f : \{\pm 1\}^n \to \mathbb{R}$ and $U$ an ancilla-free QNC$^0$ circuit of depth $t$. Then the correlation of $f \circ U$ and $\text{PAR}_n$ is bounded as

$$\mathbb{E}_x [\langle U^\dagger M_f U \rangle_x \cdot \text{PAR}_n(x)] \leq \left( W \geq 2^{-n} [f] \right)^{1/2}.$$

For example, when $f$ is an AC$^0$ circuit, we may use an LMN-type Fourier concentration bound, such as from [26], to obtain:

**Corollary 8.** If $U$ is an ancilla-free QNC$^0$ $n$-qubit circuit of depth $t$, and $f : \{\pm 1\}^n \to \{\pm 1\}$ is implemented by an AC$^0$ circuit of depth $d$ and size $s$, we have

$$\mathbb{E}_x [\langle U^\dagger M_f U \rangle_x \cdot \text{PAR}_n(x)] \leq \sqrt{2} \cdot \exp \left( \frac{-n}{2^t + O(\log s)^{d-1}} \right).$$

The proof of Theorem 7 relies on two brief lemmas. The first says that when measuring correlations, we could just as well have compared the correlation of $f$ alone to the random function $\text{PAR}_n \circ U^\dagger$, defined by applying $\text{PAR}_n$ to the output of $U^\dagger |x\rangle$.

**Lemma 9** (Symmetry of correlation). Let $f, g : \{\pm 1\}^n \to \{\pm 1\}$ and $U$ any $n$-qubit unitary. Then

$$\mathbb{E}_x [\langle U^\dagger M_f U \rangle_x \cdot g(x)] = \mathbb{E}_x [f(x) \cdot \langle U M_g U^\dagger \rangle_x]$$

$$= 2^{-n} \text{tr}[M_f U M_g U^\dagger].$$
Proof. Expanding the trace we have
\[
\text{tr}[M_f U M_g U] = \sum_z \langle z | \left( \sum_y f(y) |y\rangle y U \left( \sum_x g(x) |x\rangle x \right) U^\dagger |z \rangle
\]
\[
= \sum_{x,y,z} f(y) g(x) \langle z | y \rangle U^\dagger |x \rangle (x) U |z \rangle
\]
\[
= \sum_{x,y} f(y) g(x) \langle y | U |x \rangle (x) U^\dagger |y\rangle ,
\]
while expanding the expectations we see
\[
\mathbb{E}_x [(U^\dagger M_f U)_x \cdot g(x)] = \frac{1}{2^n} \sum_{x,y} f(y) g(x) \langle x | U^\dagger |y\rangle \langle y | U |x \rangle = \mathbb{E}_y [f(y) \cdot \langle U M_g U^\dagger y\rangle].
\]
Identifying the center expression with (a multiple of) (1) and changing variables completes the lemma.

The second lemma roughly says when Fourier characters \(Z_S\) and \(Z_T\) correspond to sets \(S, T\) of very different cardinality, they remain orthogonal (with respect to the inner product \langle A, B \rangle = \text{tr}[A^\dagger B]\) after an application of \(U\).

Lemma 10 (Lightcone lemma). Suppose \(U\) is a depth-\(t\) ancilla-free quantum circuit and \(|S| 2^t < n\). Then
\[
\text{tr}[Z_{[n]} U Z_S U^\dagger] = 0
\]

Proof. The number of qubits on which \(Z_S\) acts nontrivially at most doubles upon conjugation by each layer in \(U\). Therefore the number of non-identity coordinates in \(U Z_S U^\dagger\) is at most \(|S| 2^t\). Now if \(|S| 2^t < n\), then there is at least one coordinate \(j\) such that \(U Z_S U^\dagger = V_{[n]\setminus j} \otimes \mathbb{1}_j\) for some \((n-1)\)-qubit unitary \(V_{[n]\setminus j}\), so
\[
\text{tr}[Z_{[n]} U Z_S U^\dagger] = \text{tr}[Z_{[n]} (V_{[n]\setminus j} \otimes \mathbb{1}_j)] = \text{tr}[Z_{[n]\setminus j} V_{[n]\setminus j}] \cdot \text{tr}[Z] = 0
\]
because \(Z\) is traceless.

With these lemmas in hand, we can give the proof of Theorem 1 in a single display:

Proof of Theorem 7.
\[
\mathbb{E}_x [(U^\dagger M_f U)_x \cdot \chi_{[n]}(x)] = \mathbb{E}_x [f(x) \cdot \langle U Z_{[n]} U^\dagger \rangle_x] \quad \text{(Lemma 9)}
\]
\[
= \sum_{S \subseteq [n]} \hat{f}(S) \cdot \langle U Z_{[n]} U^\dagger \rangle_x (S) \quad \text{(Plancherel)}
\]
\[
= \sum_{S \subseteq [n]} \hat{f}(S) \mathbb{E}_x [(U Z_{[n]} U^\dagger)_x \cdot \chi_S(x)]
\]
\[
= 2^{-n} \text{tr}[Z_{[n]} U^\dagger Z_S U] \quad \text{(Lemma 9)}
\]
\[
= 0 \quad \text{if } |S| 2^t < n \quad \text{(Lemma 10)}
\]
\[
= \sum_{S \subseteq [n]} \hat{f}(S) \cdot \langle U^\dagger Z_{[n]} U \rangle (S)
\]
\[
\sum_{|S| \geq 2^{t-n}} \hat{f}(S) \cdot \langle U^\dagger Z_{[n]} U \rangle (S)
\]

\[\]
\[
\begin{align*}
&\leq \left( \sum_{|S| \geq 2^{-1} n} \hat{f}(S)^2 \right)^{1/2} \left( \sum_{|S| \geq 2^{-1} n} \langle U^\dagger Z_n U \rangle(S)^2 \right)^{1/2} \\
&\leq \left( W^{2^{-1} n} [f] \right)^{1/2}.
\end{align*}
\]

(Cauchy-Schwarz)

One may ask whether this proof approach extends to QNC⁰ circuits with ancillas. Although it might be possible to prove slight generalizations, we present an example demonstrating that any proof approach using an LMN-type theorem as a black box will fail for general QNC⁰ circuits. This is essentially because functions with Fourier decay are not closed under composition.

Example 11. Consider the following “Trojan horse” function on an even number of bits \( n = 2m \):

\[
h : \{\pm 1\}^{2m} \rightarrow \{\pm 1\}
\]

\[
x \mapsto \begin{cases} 
\chi_{[m]}(x) & \text{if } x_{[m+1,2m]} = 11 \cdots 1 \\
1 & \text{otherwise}.
\end{cases}
\]

By direct computation one finds the Fourier coefficients of \( h \) are given by

\[
\hat{h}(S) = \begin{cases} 
1 - 2^{-m} & S = \emptyset, \\
-2^{-m} & S \subseteq [m], S \neq \emptyset \\
2^{-m} & [m+1,2m] \subseteq S \\
0 & \text{otherwise}.
\end{cases}
\]

This means for any \( t \geq 1 \), the \( t \)th Fourier tail of \( h \) is \( W^{2^{-1} n} [h] \in O(2^{-n/2}) \). Thus by Theorem 7, for any ancilla-free QNC⁰ circuit \( C \), \( h \circ C \) has negligible correlation with parity.

On the other hand, consider the (deterministic) function \( C : \{\pm 1\}^{m} \rightarrow \{\pm 1\}^{2m} \) given by \( x \mapsto x11 \cdots 1 \). Certainly \( C \) can be implemented in QNC⁰, and we have \( h \circ C = \chi_{[m]} = \text{PAR}_m \).

This example shows that exponential Fourier decay of \( f \) is not sufficient to entail Conjecture 3 for general AC⁰ ∘ QNC⁰ circuits. We must take a different approach that exploits finer structural properties of AC⁰ and QNC⁰.

2 Lower bounds against AC⁰ ∘ QNC⁰ via nonlocal games

Here we pass from QNC⁰ to nonlocal games to make an argument that works for general QNC⁰. First let us fix ideas about QNC⁰.

Definition 12 (QNC⁰). An \( n \)-input, depth-\( d \) QNC⁰ circuit \( C \) is a quantum circuit composed of \( d \) layers of arbitrary 2-qubit gates, acting on an input register of \( n \) qubits and an ancilla register of \( m \in \text{poly}(n) \) qubits initialized to \( |0^m\rangle \). Via measurement of the entire output of \( C \) in the computational basis, the circuit \( C \) effects a randomized mapping from \( n \) bits of input to \( n + m \in \text{poly}(n) \) bits of output. A QNC⁰ circuit with \( v \) qubits of quantum advice, has \( v \) out of \( m \) ancilla qubits initialized to a \( v \)-qubit state, not necessarily a product state. For general \( v \in \text{poly}(n) \), this is denoted by the class QNC⁰/qpoly.

We will show a reduction from QNC⁰ circuits to nonlocal channels.
The definition of the nonlocal channel is completed by a choice of shared state \( \psi \in (\bigotimes_{i=1}^n \mathcal{H}_i) \otimes \mathcal{H}_{\text{ref}} \) and works as follows. Upon receipt of an input string \( x \in \{0,1\}^n \), the \( n \) players and one referee perform the joint measurement \( (M_{(1,x_1)}, \ldots, M_{(n,x_n)}, M_{\text{ref}}) \) on \( |\psi\rangle \), resulting in the outcomes \( y_1, \ldots, y_n \), and \( y_{\text{ref}} \). The output of the channel is the \((nk+m)\)-long string \( y = y_1 \cdots y_n \| y_{\text{ref}} \).

**Definition 14** (No-signaling channel). An \((n,k,m)\) no-signaling channel is defined analogously, except the correlations among parties may be general no-signaling correlations. (A very detailed definition of such channels is given in Definition 24)

**Definition 15** (Parity games). Let \( n, k, m \) be fixed and consider \( f : \{0,1\}^{kn+m} \rightarrow \{0,1\} \). The \((n,k,f)\) parity game is played by \( n \) entangled and non-communicating players, with the \( i^{th} \) player receiving input bit \( x_i \) from \( x \) drawn uniformly from \( \{0,1\}^n \). A \( (\text{quantum}) \) parity game strategy is an \((n,k,m)\) nonlocal channel with output string \( y \). Players win when \( f(y) = \text{Par}(x) \). We say a parity game strategy has advantage \( \varepsilon \) if its winning probability is at least \( 1/2 + \varepsilon \).

As a final piece of notation, for Boolean \( f \) let \( \neg f \) denote its negation. We are prepared to give our reduction to parity games.

**Lemma 16.** Fix \( n \geq 1, m \in \text{poly}(n) \), let \( C \) be a \( n \)-qubit, depth-\( d \) \( \text{QNC}^0 \) circuit with \( m \) ancilla and arbitrary quantum advice, and let \( f : \{0,1\}^{n+m} \rightarrow \{0,1\} \) be any Boolean function. Suppose \( f \circ C \) has correlation \( \varepsilon \) with \( \text{PAR}_n \). Then for some \( n' \geq n/(2^d+1) \) there is a quantum strategy for the \((n',2^d,f)\) or \((n',2^d,\neg f)\) parity game with advantage \( \varepsilon/2 \).

**Proof.** Suppose \( f \circ C \) has correlation \( \varepsilon \) with \( \text{PAR} \). For each input qubit \( j \) denote by \( L_j \) the set of output qubits in the forward lightcone of \( j \). Consider the graph with vertices the input qubits \( [n] \) and edges drawn between qubits \( j \) and \( k \) when \( L_j \) and \( L_k \) have nonempty intersection. Then \( G \) has degree at most \( 2^d \), so there exists an independent set \( S \subseteq [n] \) of size at least \( n/(2^d+1) \).

For each \( y \in \{0,1\}^S \), define the circuit \( C_y \) to be \( C \) but where for \( j \in S^c \), the \( j^{th} \) input is hardcoded to \( y_j \). Then \( C_y \) is a circuit on at least \( n/(2^d+1) \) variables such that the forward lightcones of input qubits are pairwise disjoint. Such a circuit defines an \((n',2^d,f)\) nonlocal channel for some \( n' \geq 2^{-d} + 1 \) and \( m' = n + n' \cdot 2^d \). (Note this \( m' \) is without loss of generality because we may freely assign a player some output bits of the referee if their lightcone is smaller than \( 2^d \).)

As a result, this restriction represents a strategy for the \((n',2^d,f)\) parity game. Moreover, we have
\[
\mathbb{E}(f \circ C)(x) \cdot \text{Par}(x) = \mathbb{E}_{y \sim \{0,1\}^n} \mathbb{E}_{z \sim \{0,1\}^z}[f \circ C_y(z) \cdot \text{Par}(y) \cdot \text{Par}(z)]
\]

Therefore since \( f \circ C \) has \( \varepsilon \) correlation with parity on \( n \) bits, for at least one \( y \), \( f \circ C_y \) or \( \neg f \circ C_y \) must have at least \( \varepsilon \) correlation (in magnitude) with parity on \( n/d \) bits. This is exactly half the advantage of the strategy defined by \( C_y \).

Lemma 16 shows that bounds on the value of parity games translate into correlation bounds for \( \text{AC}^0 \circ \text{QNC}^0 \) with \( \text{Par} \). How might we analyze parity games? They are in some sense “flipped” versions of XOR games, where parity is computed on the inputs to the players, rather than the outputs. However, it is not clear whether the rich collection of techniques developed to analyze XOR games is applicable here. Instead, we bound the no-signaling value of the game by taking the perspective of distinguishability.

For any \((n, k, 0)\) no-signaling channel \( N \), begin by rewriting the correlation as

\[
\mathbb{E}(f \circ N)(x) \cdot \text{Par}(x) = \mathbb{E}(f \circ N)(x \mid x \text{ even}) - \mathbb{E}(f \circ N)(x \mid x \text{ odd})/2.
\]

Let \( \mathcal{U}_{\text{even}} \) and \( \mathcal{U}_{\text{odd}} \) denote the uniform distribution on even and odd bitstrings of length \( n \) respectively, and consider the pushforwards of \( \mathcal{U}_{\text{even}} \) and \( \mathcal{U}_{\text{odd}} \) through \( N \):

\[
\mu := N(\mathcal{U}_{\text{even}}) \quad \text{and} \quad \nu := N(\mathcal{U}_{\text{odd}}).
\]

So \( \mu \) and \( \nu \) are distributions on strings of length \( N := nk \), and

\[
\mathbb{E}(f \circ N)(x) \cdot \text{Par}(x) = \frac{\mathbb{E}[f(\mu)] - \mathbb{E}[f(\nu)]}{2} = \Pr[f(\mu) = 1] - \Pr[f(\nu) = 1].
\]

Therefore the correlation of \( f \circ N \) with parity can be phrased in terms of \( f \)’s ability to distinguish the distributions \( \mu \) and \( \nu \).

What can be said about \( \mu \) and \( \nu \)? We claim that on every set \( S \subset [N] \) of size at most \( N/k - 1 = n - 1 \), we must have

\[
\mu_S = \nu_S.
\]

(2)

Here the notation \( \mu_S \) denotes the marginal distribution of \( \mu \) on the coordinates in \( S \). To see (2), let \( T \subset [n] \) be the set of players whose outputs overlap \( S \). Then by the no-signaling property of \( N \), the marginal \( \mu_T \) (resp. \( \nu_T \)) is entirely determined by the marginal input distribution on \( T \); that is, \( (\mathcal{U}_{\text{even}})_T \) (resp. \( (\mathcal{U}_{\text{odd}})_T \)). And for any \( T \) a strict subset of \([n]\), \( (\mathcal{U}_{\text{even}})_T = (\mathcal{U}_{\text{odd}})_T = \mathcal{U}([0,1]^{|T|}) \), so we must have \( \mu_T = \nu_T \).

So all small marginals of \( \mu \) and \( \nu \) are information-theoretically indistinguishable. This is exactly \( k \)-wise indistinguishability, a generalization of \( k \)-wise independence introduced by Bogdanov et al. [4] and first used in the context of secret sharing.

Definition 17 (\( k \)-wise indistinguishability [4]). Two distributions \( \mu \) and \( \nu \) on \( \{\pm 1\}^N \) are \( k \)-wise indistinguishable if for all \( S \subset [N] \) with \( |S| \leq k \), \( \mu_S = \nu_S \).

Additionally, for \( f : \{0,1\}^n \to \{0,1\} \), we say \( f \) is \( \varepsilon \)-fooled by \( k \)-wise indistinguishability if for any pair \( \mu, \nu \) of \( k \)-wise indistinguishable distributions,

\[
|\Pr[f(\mu) = 1] - \Pr[f(\nu) = 1]| \leq \varepsilon.
\]

It turns out \( k \)-wise indistinguishability over the hypercube is intimately connected to approximate degree. By a linear programming duality argument, Bogdanov et al. proved the following.
In this section, we describe a sublinear upper bound of a large subclass of \( \text{AC}_0 \) circuits. Let \( m \) be fixed and suppose there is an \( f \in \mathcal{F} \) on \( N = \text{poly}(m) \) variables and an \( m \)-input QNC\(^0\) circuit \( \mathcal{C} \) of depth \( d \), with \( N - m \) ancilla qubits, and receiving arbitrary quantum advice, such that \( f \circ \mathcal{C} \) achieves correlation \( \varepsilon \) with \( \text{PAR}_m \). Then there is a \( g \in \mathcal{F} \) on \( n \geq m/2 \) variables with \( \tilde{\text{deg}}_{\varepsilon/2}[g] \geq n/2^d - 1 \).

**Proof.** By Lemma 16, there is an \( m' \geq m/(2^d + 1) \) and an \( \langle m', 2^d, N - 2^d m' \rangle \) nonlocal channel \( \mathcal{N} \) such that \( f \circ \mathcal{N} \) or \( -f \circ \mathcal{N} \) achieves correlation \( \varepsilon \) with \( \text{PAR}_{m'} \).

Suppose the referee measures their system and obtains outcome string \( r \). This event leads to an updated state shared among the parties in \( \mathcal{N} \) and thereby defines an \( \langle m', 2^d, 0 \rangle \) nonlocal channel \( \mathcal{N}^{R+r} \). By a similar averaging argument to the one used in Lemma 16, there is at least one outcome \( r \) of the referee register such that \( \mathcal{N}^{R+r} \) still yields correlation \( \varepsilon \) with \( \text{PAR} \). Define \( g := f|_{R=r} \) or \( g := -f|_{R=r} \) as appropriate and put \( \mathcal{E} := \mathcal{N}^{R+r} \). Then \( g \in \mathcal{F} \) is a function on \( n := 2^d m' \) bits and

\[
\mathbb{E}_x[(g \cdot \mathcal{E})(x) \cdot \text{PAR}(x)] \geq \varepsilon.
\]

Therefore, by the discussion above, we see \( g \) can \( \varepsilon \)-distinguish \( (n/2^d - 1) \)-wise indistinguishable distributions. Applying Theorem 18 we conclude that

\[
\tilde{\text{deg}}_{\varepsilon/2}[g] \geq \frac{n}{2^d} - 1.
\]

**Corollary 20.** Suppose function class \( \mathcal{F} \) is closed under inverse-polynomial-sized restrictions. Then if \( \text{deg}[\mathcal{F}] \in o(n) \), \( \mathcal{F} \circ \text{QNC}_0 \) cannot achieve \( \Omega(1) \) correlation with \( \text{PAR} \).

The burning question, then, is whether \( \text{deg}[\text{AC}_0] \in o(n) \). In fact, the approximate degree of \( \text{AC}_0 \) is a longstanding open problem and its resolution would lead to several consequences in complexity theory [9]. To get a sense of the difficulty of this question, consider that on one hand, a sublinear upper bound is known for a large subclass of \( \text{AC}_0 \).

**Theorem ([8, Theorem 5]).** Let \( p(n) \in \text{poly}(n) \). Then the class of \( \text{AC}_0 \) circuits of linear size, denoted by \( \text{LC}_0 \), has

\[
\tilde{\text{deg}}_{1/p(n)}[\text{LC}_0] \in o(n).
\]

Yet on the other hand, a series of works, most recently [24], show the following:

**Theorem.** For any \( \delta > 0 \), there is a function \( f \in \text{AC}_0 \) with \( \tilde{\text{deg}}[f] \in \Omega(n^{1-\delta}) \).
The lower bound of $\Omega(n^{1-\delta})$-for-any-$\delta$ is tantalizingly close to the trivial upper bound of $n$ for the approximate degree of any Boolean function, but as it stands it is not unreasonable to guess that $\deg[AC^0] \in O(n/\log n)$ either. Several questions – including now Conjecture 3 – could be settled if the gap between $\Omega(n^{1-\delta})$-for-any-$\delta$ and $n$ for $\deg[AC^0]$ were closed.

We may combine the sublinear lower bound on $LC^0$ from [8] with Theorem 19 to obtain:

**Corollary 21.** Let $C$ be an $n$-input, $m$-ancilla QNC$^0$ circuit with arbitrary advice. Suppose $f : \{0,1\}^{n+m} \to \{-1,1\}$ is defined by an $AC^0$ circuit of size $O(n)$. Then $f \circ C$ achieves negligible correlation with $\text{PAR}_n$.

### 2.1 Blockwise approximate degree

We conclude this section by laying out a self-contained question concerning the approximate degree of $AC^0$ with respect to a modified, “blockwise” notion of approximate degree. This question is sufficient to imply Conjecture 3 in full generality and may be easier to resolve than $\deg[AC^0]$.

Fix $k \geq 1$ (assuming $k$ divides $n$ for simplicity) and let $P$ be the partition of $[n]$ into “blocks” of size $k$:

$$P := \{ \{1,\ldots,k\}, \{k+1,\ldots,2k\}, \ldots, \{n-k+1,n\} \} .$$

For a monomial $\chi_S = \prod_{i \in S} x_i$ define the $(k)$-block degree $\text{bdeg}_k[\chi_S]$ to be the number of distinct blocks $B \in P$ having nonempty intersection with $S$. This definition extends naturally to the $k$-block degree $\text{bdeg}_k[f]$ of a Boolean function $f : \{0,1\}^n \to \{-1,1\}$ and to the approximate $k$-block degree $\widetilde{\text{bdeg}}_k[f]$ of $f$:

$$\widetilde{\text{bdeg}}_k[f] = \min \{ \text{bdeg}_k[g] \mid g : \{0,1\}^n \to \mathbb{R} \text{ a polynomial with } \| f - g \|_\infty \leq 1/3 \} .$$

Of course $\widetilde{\text{bdeg}}_k[f] \leq n/k$ for any function.

**Question 1.** For all constants $k$, does the following hold?

$$\widetilde{\text{bdeg}}_k[AC^0] \leq n/k - 1 .$$

As we explain below, this would be enough to prove Conjecture 3. Note the following, which are immediate and hold for all $f$:

$$\text{deg}[f] < \frac{n}{k} \implies \text{bdeg}_k[f] < \frac{n}{k} \implies \text{deg}[f] < n - k .$$

Moreover, these implications are sharp in that each one cannot generically imply anything stronger, as witnessed by a parity function on an appropriate subset of $[n]$. Regarding $f \in AC^0$, the left-hand side holding for arbitrary constant $k$ is equivalent to $\text{deg}[AC^0] \in o(n)$, while the far right-hand side follows directly from LMN-type Fourier tail bounds for $AC^0$.

**Proposition 22.** If the resolution to Question 1 is “yes”, then Conjecture 3 is true.

**Proof sketch.** Consider the referee-free nonlocal channel $E$ from the proof of Theorem 19, with $n/k$ players responding with $k$ bits each. Defining $\mu$ and $\nu$ as the pushforwards of uniform distributions over even and odd bitstrings as before, it is true that $\mu$ and $\nu$ are $(n/k-1)$-wise indistinguishable when viewed as distributions on $\{0,1\}^n$. However, they may also be viewed as distributions on the hypergrid $[2^k]^m$ for $m = n/k$.

With this perspective, $\mu$ and $\nu$ are $m - 1$ indistinguishable. Repeating the proof of [4, Theorem 1.2] over this larger alphabet, we recover exactly the notion of blockwise degree. The rest of the argument is as before.■
Remark 23. It is unclear to us whether Question 1 is any easier than \( \tilde{\text{deg}}[\text{AC}^0] \in o(n) \). We can compare the two questions head-to-head as follows. Let \( \mathcal{P}_k \) be all the relabelings of \( P \):

\[
\mathcal{P}_k := \left\{ \{\pi(1), \ldots, \pi(k)\}, \{\pi(k+1), \ldots, \pi(2k)\}, \ldots, \{\pi(n-k+1), \ldots, \pi(n)\} \right\}_{\pi \in S_n}.
\]

For any \( P \in \mathcal{P}_k \), let \( \text{bdeg}_P[f] \) be the maximum number of blocks in \( P \) overlapped by some monomial in \( f \). Then we have the following characterization, where \( g \) ranges over real-valued multilinear polynomials on the hypercube as usual:

\[
\begin{align*}
\tilde{\text{deg}}[\text{AC}^0] < n/k & \iff \forall f \in \text{AC}^0, \exists g, \forall P \in \mathcal{P}_k, \text{bdeg}_P[g] \leq n/k \text{ and } \|f-g\| \leq 1/3 \\
\tilde{\text{bdeg}}_P[\text{AC}^0] < n/k & \iff \forall f \in \text{AC}^0, \exists P \in \mathcal{P}_k, \exists g, \text{bdeg}_P[g] \leq n/k \text{ and } \|f-g\| \leq 1/3.
\end{align*}
\]

### 3 Towards a switching lemma for \( \text{AC}^0 \circ \text{QNC}^0 \)

Recall that our approach in Section 1 fails because circuits with LMN-style Fourier decay are not suitably closed under precomposition by \( \text{QNC}^0 \). In fact this is true even under precomposition by \( \text{NC}^0 \), and the proof of the LMN theorem elegantly avoids an induction assumption phrased in terms of Fourier decay. Instead, the proof relies on a structural theorem about the effect of random restrictions on DNFs and CNFs – Håstad’s celebrated switching lemma:

**Theorem (Håstad [13]).** Suppose \( f \) is a width-\( w \) DNF. Then for any \( 0 \leq \delta \leq 1 \),

\[
\Pr_{\rho \sim \mathbf{R}_\delta} [\text{DT}_{\text{depth}}(f | \rho) > t] \leq (C\delta w)^t,
\]

where \( C \) is a universal constant.

Here \( \mathbf{R}_\delta \) is the distribution of random restrictions with star probability \( \delta \) (see e.g., [20, §4.3] for more). This theorem has received several proofs over time, but each rely on the well-controlled structure of random restrictions. To naively repeat the switching lemma argument directly on \( \text{AC}^0 \circ \text{QNC}^0 \) would mean to track the passage of random restrictions through \( \text{QNC}^0 \) – a tall order given that \( \text{QNC}^0 \) can destroy the independence and unbiasedness of random restrictions that switching arguments tend to rely on.

The situation may be slightly improved by instead considering a switching lemma for the model studied in Section 2. Recalling that \( f \circ \mathcal{N} \) is a randomized function, we may hope for a switching lemma of the following form:

**An imagined switching lemma for nonlocal channels.** Let \( m \geq 0 \) and \( k, w, n \geq 1 \) and suppose \( f : \{0,1\}^{kn+m} \rightarrow \{0,1\} \) is a DNF of width \( w \) and \( \mathcal{N} \) is an \( (n, k, m) \) nonlocal channel. Then for each restriction \( \rho \) there exists a distribution \( \Gamma_\rho \) over decision trees such that \( (f \circ \mathcal{N})|_\rho = \{T\}_{T \sim \Gamma_\rho} \) and

\[
\Pr_{\rho \sim \mathbf{R}_\delta} [\text{depth}(T) > t] \leq (C\delta w)^t.
\]

By Lemma 16 such a switching lemma would be sufficient to show correlations bounds between \( f \circ \text{QNC}^0 \) and parity for any DNF (or CNF) \( f \), which in turn are direct prerequisites to proving Conjecture 3. While this imagined switching lemma is currently out of reach, we contend it presents a useful challenge to existing switching lemma proof techniques. As a first step in this direction, we devote this section to a proof of a simpler but related structural result.
**Theorem (Informal).** Any no-signaling channel $\mathcal{N}$ composed with a decision tree $\tau$ is equal to a probability distribution $\Gamma$ of decision trees with $\text{depth}(\tau') \leq \text{depth}(\tau)$ for all $\tau' \in \text{Supp}(\Gamma)$.

Let us fix some notation. For a finite set $X$ let $\mathcal{M}(X)$ denote the set of probability measures on $X$. The set $\mathcal{M}(X)$ is convex, so for $\nu$ a probability measure on $\mathcal{M}(X)$ we may define the expected distribution

$$E[\nu] := \left\{ x \text{ w.p. } \Pr_{\mu \sim \nu} [z = x] \right\}_{x \in X}$$

(3)

Here we study Boolean channels, or functions of the form

$$\mathcal{N} : \{\pm 1\}^n \to \mathcal{M}(\{\pm 1\}^N).$$

For a probability measure $\mu$ on the set of channels from $n$ to $N$ bits, we use $E_{\mathcal{N} \sim \mu} \mathcal{N}$ to denote the channel defined pointwise as:

$$\left( E_{\mathcal{N} \sim \mu} \mathcal{N} \right)(x) := E_{\mathcal{N} \sim \mu} [\mathcal{N}(x)].$$

(4)

To be clear, $\mathcal{N}(x)$ is a probability measure on $\{\pm 1\}^N$, so in the right-hand side of (4) we are computing the expected distribution according to (3). Also, for $T \subseteq [N]$ define the reduced channel

$$\mathcal{N}^T(x) := \left\{ y \text{ w.p. } \sum_{z \in \{\pm 1\}^N \atop z_\tau = y} \Pr[\mathcal{N}(x) = z] \right\}_{y \in \{\pm 1\}^{\tau^\prime}}$$

**Definition 24 (No-signaling channel).** Consider a map $\mathcal{N} : \{\pm 1\}^n \to \mathcal{M}(\{\pm 1\}^N)$ and a “backwards lightcone” function $B : [N] \to [n] \cup \{\perp\}$. The pair $(\mathcal{N}, B)$ is a no-signaling channel (NSC) if for all $S \subseteq [n]$, for all $x, x' \in \{\pm 1\}^n$ with $x_S = x'_S$, we have $\mathcal{N}^{B^{-1}(S \cup \perp)}(x) = \mathcal{N}^{B^{-1}(S \cup \perp)}(x')$.

That is, a channel is an NSC if for any collection of output indices $T$, $\mathcal{N}^T(x)$ is a function of $x_{B(T) \cup \{\perp\}}$ only. Note also $\mathcal{N}^{B^{-1}(\perp)}$ is oblivious to the value of $x$ entirely – the outputs $B^{-1}(\perp)$ could be called the referee outputs.

Recall that for a Boolean function $f : \{\pm 1\}^N \to \{\pm 1\}$, $f \circ \mathcal{N}$ denotes the channel

$$f \circ \mathcal{N}(x) = \left\{ b \text{ w.p. } \Pr_{y \sim \mathcal{N}(x)} [f(y) = b] \right\}_{b \in \{\pm 1\}}.$$

The restriction structure on NSCs interacts nicely with decision trees:

**Theorem 25.** Given $f : \{\pm 1\}^N \to \{\pm 1\}$ and $\mathcal{N} : \{\pm 1\}^n \to \mathcal{M}(\{\pm 1\}^N)$ an NSC, there exists a distribution $\Gamma$ over decision trees such that

(i) For all $x$ the composition $f \circ \mathcal{N}(x) = \{\tau(x)\}_{\tau \sim \Gamma}$, so $E[f \circ \mathcal{N}(x)] = E_{\tau \sim \Gamma} [\tau(x)];$ and

(ii) For all $\tau \in \text{Supp}(\Gamma)$, $D_{\text{depth}}(\tau) \leq D_{\text{depth}}(f)$.

Recall that $f \circ \mathcal{N}$ is an $\mathcal{M}(\{\pm 1\})$-valued function on the hypercube, so $x \mapsto E[f \circ \mathcal{N}(x)]$ is a $[-1, 1]$-valued function on the hypercube, and accordingly has a multilinear Fourier expansion

$$E[f \circ \mathcal{N}] = \sum_{S \subseteq [n]} a_S X_S$$

with $a_S := E_x [E[f \circ \mathcal{N}] \cdot X_S(x)]$

We pause to note the related fact that in terms of the expected output $E[f \circ \mathcal{N}]$, the degree of any function $f$ does not increase under composition with an NSC: $\deg(f) \geq \deg(E[f \circ \mathcal{N}])$. 
Consider the Fourier expansion
\[ f = \sum_{S \subseteq [N]} \hat{f}(S) \chi_S. \]
Then \( \mathbb{E}[(f \circ \mathcal{N})(x)] = \mathbb{E} \left[ \sum_{S \subseteq [N]} \hat{f}(S) \chi_S \circ \mathcal{N}(x) \right] = \sum_{S} \hat{f}(S) \mathbb{E}[\chi_S \circ \mathcal{N}(x)]. \)
A linear combination of functions of at most \(|S|\) variables each for \(|S| \leq \deg(f)\).

For example, there are Boolean functions \( g \) with \( \deg(g) = n^{2/3} \) but \( DT_{\text{depth}}(g) = n \) (see Example 3 in [7]). One could imagine a Boolean function \( h \) with \( \deg(h) \approx DT_{\text{depth}}(h) \in o(n) \) but where \( \mathbb{E}[h \circ \mathcal{N}] \) is “\( g \)-like”; any decision tree decomposition of \( \mathbb{E}[h \circ \mathcal{N}] \) contains a tree of depth \( n \) despite having \( \deg(\mathbb{E}[h \circ \mathcal{N}]) \in o(n) \). Theorem 25 says such an \( h, \mathcal{N} \) pair does not exist; precomposition by an NSC cannot increase the decision tree complexity of a function.

The proof of Theorem 25 requires some bookkeeping. The idea is to begin with \( \tau \)'s root vertex variable \( y_i \) and locally decompose the univariate channel \( \mathcal{N}^{(i)}(x) \sim y_i \) into a distribution of deterministic functions \( \{y_i, \omega(x_i)\} \omega_{1,\mu} \). This decomposition of the root vertex induces a probabilistic decomposition \( \{\tau' \circ \mathcal{N}'_j\} \omega_{1,\mu} \) of the entire hybrid computation where the root variable \( y_i \) in \( \tau' \) has been replaced with an \( x_{B(i)} \) and the left and right subtrees of \( \tau \) become compositions not with \( \mathcal{N} \), but with conditional versions of \( \mathcal{N} \) where \( x_{B(i)} \) and \( y_i \) have been fixed to certain values. This conditioning preserves the NSC-ness of the new \( \mathcal{N}'_j \)s, and the decomposition recurses down the tree.

We now introduce a notion of conditioning. For any \( n \)-to-\( N \) bit Boolean channel \( \mathcal{N} \), \( x \in \{\pm 1\}^n \), \( J \subseteq [N] \) and \( Y \in \text{Supp}(\mathcal{N}^T(x)) \) define the conditional channel as
\[ \mathcal{N}(x \mid y_J = Y) := \{ y \ \text{w.p.} \ \Pr[\mathcal{N}(x) = y \mid y_J = Y] \} \}_{y \in \{\pm 1\}^N}, \]
and for \( T \subseteq [N] \) the reduced conditional channel
\[ \mathcal{N}^T(x \mid y_J = Y) := \{ y \ \text{w.p.} \ \sum_{z \in \{\pm 1\}^N} \Pr[\mathcal{N}(x) = z \mid y_J = Y] \} \}_{y \in \{\pm 1\}^{|T|}}. \]

Note that \( T \)-reduced conditional no-signaling channels can depend on inputs outside \( B(T) \). Consider for example the \( n \)-to-\( n \)-bit NSC
\[ \mathcal{G}(x) = \begin{cases} \mathcal{U}\{\text{even strings}\} & x \text{ even} \\ \mathcal{U}\{\text{odd strings}\} & x \text{ odd}. \end{cases} \]
Now \( \mathcal{G}^{(i)}(x) \) is identically a Rademacher random variable (oblivious to \( x \) entirely), but
\[ \mathcal{G}^{(i)}(x \mid y_{|\{i\}|} = 00 \cdots 0) = \left\{ \prod_j x_j \ \text{w.p.} \ 1 \right\}, \]
the parity of all \( n \) bits of \( x \). All the same, some structure remains after conditioning:

\[ \begin{array}{c} \text{Proposition 26. For } T, J \subseteq [N], \text{ let } x, x' \in \{\pm 1\}^n \text{ be such that } x_{B(J \cup T)} = x'_{B(J \cup T)}. \text{ Then} \\ \text{for all } Y \in \text{Supp}(\mathcal{N}^T(x)), \\ \mathcal{N}^T(x \mid y_J = Y) = \mathcal{N}^T(x' \mid y_J = Y). \end{array} \]

\[ \text{ Proof. Let } x, x' \text{ be as in the proposition statement. We have from the definition of NSCs that } \mathcal{N}^{J \cup T}(x) = \mathcal{N}^{J \cup T}(x'). \text{ Certainly then } \mathcal{N}^{J \cup T}(x \mid y_J = Y) = \mathcal{N}^{J \cup T}(x' \mid y_J = Y) \text{ (we have taken the marginal of two equal distributions). The conclusion then follows from noticing that for any } U \subseteq V, \mathcal{N}^U = (\mathcal{N}^V)^U. \]

Consider the Fourier expansion \( f = \sum_{S \subseteq [N]} \hat{f}(S) \chi_S \). Then \( \mathbb{E}[(f \circ \mathcal{N})(x)] = \mathbb{E} \left[ \sum_{S \subseteq [N]} \hat{f}(S) \chi_S \circ \mathcal{N}(x) \right] = \sum_{S} \hat{f}(S) \mathbb{E}[\chi_S \circ \mathcal{N}(x)]. \)
This proposition says $N^T(x \mid y_1 = Y)$ is a function of $x_{B(I,T)}$ only. Thus if we fix variables $x_{B(I)}$ we recover a smaller NSC:

**Corollary 27.** Consider an $n$-to-$N$ NSC $(N, B)$, an $i \in [N]$, and $X, Y \in \{\pm 1\}$. If $B(i) = \perp$ let $N'$ be the $n$-to-$(N-1)$ NSC

\[
N' = N^{(N\setminus\{i\})}(x \mid y_1 = Y)
\]

and otherwise let $N'$ be the $(n-1)$-to-$(N-1)$ NSC

\[
N' = N^{(N\setminus\{i\})}(x_{B(i)} \mid x_{B(i)} = X, y_1 = Y).
\]

Define a new lightcone function $B'$ from $B$ as follows. Put $B(j) = \perp$ for all $j \in B^{-1}(B(i))$ and then remove $i$ from the domain of $B$. Then $(N', B')$ is an NSC.

Finally we introduce an object used internally in the proof of Theorem 25.

**Definition 28 (Hybrid Decision Tree).** A hybrid decision tree $T$ on $n$ variables with $\ell$ leaves consists of the data $(\tau, G_1, \ldots, G_\ell)$, where

(i) The first argument $\tau$ is a rooted binary tree with $\ell$ leaves labeled as follows. Each internal node is assigned $x_i$ for some $i \in [n]$, the edge to its left child is labeled 1, and the edge to its right child is labeled $-1$.

(ii) Each leaf $i$ of $\tau$ is associated with an $n$-to-1 channel $G_i : \{\pm 1\}^n \rightarrow M\{\pm 1\}$.

A hybrid tree defines a channel $T_{\tau}(G_1, \ldots, G_\ell) : \{\pm 1\}^n \rightarrow M\{\pm 1\}$ as follows. Computation on input $x \in \{\pm 1\}^n$ proceeds just as with standard decision trees until a leaf $i$ is reached, at which point the distribution $G_i(x)$ is returned.

Theorem 25 follows from these three claims. Proofs of the first two are immediate from the definitions.

**Claim 29.** For any hybrid decision tree $T$,

\[
T(G_1, \ldots, G_{\ell-1}, \mathbb{E}_{\omega \sim \mu}[G_\omega], G_{\ell+1}, \ldots, G_\ell) = \mathbb{E}_{\omega \sim \mu}[T(G_1, \ldots, G_{\ell-1}, G_\omega, G_{\ell+1}, \ldots, G_\ell)]
\]

**Claim 30.** For any hybrid decision trees $T_{\tau}(G_1, \ldots, G_\ell)$ and $T'_{\tau'}(G_1, \ldots, G_\ell)$,

\[
T_{\tau_0 \circ \tau'}(G_1, \ldots, G_{\ell-1}, T_{\tau}(G_{\ell+1}, \ldots, G_\ell), G_{\ell+1}, \ldots, G_\ell) = T_{\tau_0 \circ \tau'}(G_1, \ldots, G_{\ell-1}, G_{\ell+1}, \ldots, G_\ell, G_{\ell+1}, \ldots, G_\ell),
\]

where $\tau_0 \circ \tau'$ is $\tau$ with the $i$th leaf replaced with $\tau'$.

**Claim 31.** Suppose $\tau$ is a decision tree and $(N, B)$ is an NSC. Then either:

(i) $\tau \circ N = \mathbb{E}_{\omega \sim \mu}[\tau_{\omega} \circ N_\omega]$ where depth$(\tau_{\omega}) \leq$ depth($\tau$) $- 1$, $|\text{Supp}(\mu)| \leq 2$, and each $N_\omega$ is an NSC, or

(ii) $\tau \circ N = \mathbb{E}_{\omega \sim \mu}[T_{\tau}(\tau_{\omega_1} \circ N_{\omega_1}, \tau_{\omega_2} \circ N_{\omega_2})],$ where $|\text{Supp}(\mu)| \leq 3$, $\tau^*$ has one internal node, depth$(\tau_{\omega_1}),$ depth$(\tau_{\omega_2}) \leq$ depth($\tau$) $- 1$, and each $N_{\omega_1}, N_{\omega_2}$ is an NSC; or

(iii) (Base case) $\tau \circ N(x) = \{b \text{ w.p. 1}\}$ for all $x$, for some fixed $b \in \{\pm 1\}$.

Proof. If $\tau$ is the trivial decision tree with no internal nodes, clearly we satisfy case iii. Otherwise, let $y_i$ be the variable at the root of $\tau$. There are two cases depending on the value of $B(i)$.
We will recursively break apart leaves of Case ii), Proof of Theorem 25. Case i), combined with the parent tree to become distributions over hybrid trees of greater depth. 

Let $\tau_{\omega}$ be the subtree of $\tau$ attached to the $\omega$-valued edge of $y_i$. Put $\mathcal{N}_{\omega} = \mathcal{N}^{(1)}(x \mid y_i = \omega)$. Then we have for $z \in \{\pm 1\}$,

$$
\Pr[\tau \circ \mathcal{N}(x) = z] = \sum_{\omega \in \{\pm 1\}} \Pr[\tau \circ \mathcal{N}(x) = z \mid D^i(x) = \omega] \Pr[\mathcal{N}^i(x) = \omega] = \sum_{\omega \in \{\pm 1\}} \Pr[\tau_{\omega} \circ \mathcal{N}(x \mid y_i = \omega) = z] \Pr[\mathcal{N}^i(x) = \omega] = \sum_{\omega \in \{\pm 1\}} \Pr[\tau_{\omega} \circ \mathcal{N}_{\omega}(x) = z] \Pr[\mathcal{N}^i(x) = \omega] = \Pr \left[ \sum_{\omega \in \{\pm 1\}} \Pr[\tau_{\omega} \circ \mathcal{N}_{\omega}](x) = z \right]
$$

as desired. Clearly $\tau_{\omega}$ is strictly shorter than $\tau$, and $\mathcal{N}_{\omega}$ is an NSC by Corollary 27.

Case ii), $B(i) \neq \perp$. Let $\tau^*$ be the one-vertex tree consisting of the root vertex of $\tau$ relabeled with $x_{B(i)}$ and let $\tau_L, \tau_R$ be the left and right subtrees of $\tau$ respectively. Observe that $\mathcal{N}^{(1)}(x) = \mathcal{N}^{(1)}(x_{B(i)})$ is a univariate channel. Hence it can be decomposed as a convex combination

$$
\mathcal{N}^{(1)}(x_{B(i)}) = a_{(1,1)} \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} + a_{(-1,-1)} \begin{bmatrix} 0 & 0 \\ 1 & 1 \end{bmatrix} + a_{(1,-1)} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + a_{(-1,1)} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.
$$

where only three of $a_{(L,R)}$ are nonzero. Let $\mu = \{(L,R) \text{ w.p. } a_{(L,R)}\}$. Then we claim

$$
\tau \circ \mathcal{N} = \sum_{(L,R) \sim \mu} \mathbb{E}_{(L,R) \sim \mu} \left[ \mathcal{T}_{\tau^*} (\tau_L \circ \mathcal{N}^{(1)}_{L}, \tau_R \circ \mathcal{N}^{(1)}_{R}) \right]
$$

where for $b, c \in \{\pm 1\}^2$,

$$
\mathcal{N}^{(b)}(x) = \mathcal{N}(x \mid x_{B(i)} = b, y_i = c).
$$

We check Eq. (5) pointwise. First consider an $x$ with $x_{B(i)} = 1$. We condition on the value of $y_i$, rearrange, and then “complete the tree”:

$$
\Pr[\tau \circ \mathcal{N}(x) = z] = \sum_{L \in \{\pm 1\}} \Pr[\tau \circ \mathcal{N}(x) = z \mid \mathcal{N}(x) = L] \Pr[\mathcal{N}(x) = L] = \sum_{L \in \{\pm 1\}} \Pr[\tau \circ \mathcal{N}(x \mid y_i = L) = z](a_{(L,1)} + a_{(L,-1)}) = \sum_{L \in \{\pm 1\}} \Pr[\tau_L \circ \mathcal{N}(x \mid x_{B(i)} = 1, y_i = L) = z] \left( \sum_{R \in \{\pm 1\}} a_{(R,L)} \right) = \sum_{L,R \in \{\pm 1\}} a_{(L,R)} \Pr[\tau_L \circ \mathcal{N}^{(1)}_L(x) = z] = \sum_{L,R \in \{\pm 1\}} a_{(L,R)} \Pr[\mathcal{T}_{\tau^*} (\tau_L \circ \mathcal{N}^{(1)}_{L}, \tau_R \circ \mathcal{N}^{(1)}_{R})(x) = z] = \Pr \left[ \sum_{(L,R) \sim \mu} \mathbb{E}_{(L,R) \sim \mu} \left[ \mathcal{T}_{\tau^*} (\tau_L \circ \mathcal{N}^{(1)}_{L}, \tau_R \circ \mathcal{N}^{(1)}_{R})(x) = z \right] \right],
$$

as desired. A similar argument goes through for $x_{B(i)} = -1$ by expanding over $R$ instead of $L$. \(\triangleleft\)

**Proof of Theorem 25.** Let $\tau$ be a depth-optimal decision tree for $f$. Construct the trivial hybrid tree $\mathcal{T}$ with no internal nodes and a single leaf with label $\tau \circ \mathcal{N}$. Put $\Gamma = \{\mathcal{T} \text{ w.p. } 1\}$. We will recursively break apart leaves of $\mathcal{T}$ into distributions of hybrid trees, which are then combined with the parent tree to become distributions over hybrid trees of greater depth.
This is done by repeated application of the following sequence of steps. Suppose $T_{\tau}(G_1, \ldots, G_\ell)$ is some hybrid tree and $G_\iota = \tau' \circ N$ for some nontrivial DT $\tau'$ and (potentially conditioned) NSC $N$. Then depending on the case in Claim 3 we either have

$$T_{\tau}(\ldots, \tau \circ N_{\text{index}} \tau', \ldots) = T_{\tau}(\ldots, E_{\omega_L \sim \mu}[\tau_{\omega_L} \circ N_{\omega_L}], \ldots)$$  (Claim 31.i)

$$= \mathbb{E}_{\omega_L, \omega_R \sim \mu}[T_{\tau}(\ldots, T_{\tau'}(\tau_{\omega_L} \circ N_{\omega_L}, \tau_{\omega_R} \circ N_{\omega_R}), \ldots)]$$  (Claim 29)

$$= \mathbb{E}_{\omega_L, \omega_R \sim \mu}[T_{\tau_{\omega_L, \omega_R}} \circ N(\tau_{\omega_L} \circ N_{\omega_L}, \tau_{\omega_R} \circ N_{\omega_R}), \ldots)]$$  (Claim 30)

where $\tau'$ has depth 1 and $\text{depth}(\tau_{\omega_L}), \text{depth}(\tau_{\omega_R}) \leq \text{depth}(\tau') - 1$, or we have

$$T_{\tau}(\ldots, \tau \circ N_{\text{index}} \tau', \ldots) = T_{\tau}(\ldots, \mathbb{E}_{\omega \sim \mu}[\tau_{\omega} \circ N_{\omega}], \ldots)$$  (Claim 31.ii)

$$= \mathbb{E}_{\omega \sim \mu}[T_{\tau}(\ldots, \tau_{\omega} \circ N_{\omega}, \ldots)]$$  (Claim 29)

where $\text{depth}(\tau_{\omega}) \leq \text{depth}(\tau') - 1$.

If we repeatedly make these transformations on the elements of $\Gamma$, we will eventually be left with a distribution over hybrid decision trees $(\tau, G_1, \ldots)$ where each channel $G = \tau' \circ N$ is in the base case of Claim 31. Such a hybrid tree is equal to a deterministic channel. Hence we are left with a distribution over deterministic channels that is trivially equivalent to a distribution of standard, deterministic decision trees.

Further, it’s easy to see that once done, the longest path in any tree of $\text{Supp}(\Gamma)$ is bounded by the longest path in the original tree $\tau$.

4 Discussion

We have seen several pieces of evidence for Conjecture 3, as well as highlighted new connections between quantum complexity theory, nonlocal games, and approximate degree.

If Conjecture 1 is ultimately proved true, we may wish to reach for a stronger no-advantage theorem closer to that of Beals et al. [3] from query complexity. A natural expression of $\text{AC}^0 \circ \text{QNC}^0$ non-advantage might use the language of Fourier decay.

\textbf{Question 2.} Does $\text{AC}^0 \circ \text{QNC}^0$ exhibit LMN-like Fourier decay? To make this precise for the randomized function $f \circ C$, consider the expectation over the randomness in $C$ to get an function $F : \{0, 1\}^n \to [-1, 1]$. Then we ask, is $W^{\geq t}[F] \in \mathcal{O}(\exp(-t))$?

As mentioned in the introduction, a similar result is known depth-$d$ QAC$^0$ circuits with at most $\mathcal{O}(n^{1/d})$ ancillas [19].

Finally, one may consider any number of variations on the theme of precomposing a Boolean function with QNC$^0$. It is natural to ask:

\textbf{Question 3.} View a QNC$^0$ circuit $C$ as a map from (randomized) Boolean functions to randomized Boolean functions:

$$f \xrightarrow{C} f \circ C.$$

By how much can this map increase influence, sensitivity, or other complexity measures of $f$?

Theorem 25 gives the answer “not at all” to a variant Question 3 where QNC$^0$ is replaced by nonlocal channels, and the complexity measure is randomized decision tree complexity.
References


<table>
<thead>
<tr>
<th></th>
<th>Reference</th>
<th>Details</th>
</tr>
</thead>
</table>