Parity vs. 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 QNC^0, over NC^0, 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 QNC^0 can be used to help compute parity itself. Namely, we study $\mathsf{AC}^0 \circ \mathsf{QNC}^0$ – a hybrid circuit model where AC^0 operates on measurement outcomes of a QNC^0 circuit – and we ask whether $\textsc{Par} \in \mathsf{AC}^0 \circ \mathsf{QNC}^0$.

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

- When the QNC^0 circuit is ancilla-free, this model can achieve only negligible correlation with parity, even when 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 QNC^0 circuit is given arbitrary quantum advice. By known results [8], this confirms the conjecture for linear-size AC^0 circuits.
- Another approach to proving the conjecture is to show a switching lemma for $\mathsf{AC}^0 \circ \mathsf{QNC}^0$. 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 $\mathrm{DT}_{\mathrm{depth}}[f]$.

Taken together, our results suggest that while 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

Related Version *Full Version*: `https://arxiv.org/abs/2311.13679`

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 QNC^0, and constant-depth bounded fan-in classical circuits, or 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.

▶ **Definition 1.** *Let $\mathsf{AC}^0 \circ \mathsf{QNC}^0$ denote the model of computation composed of a QNC^0 circuit $\mathcal{C}$, followed by a computational basis measurement, and then an AC^0 function f applied to the result. This process defines the randomized Boolean function $f \circ \mathcal{C} : \{0,1\}^n \to \mathcal{M}(\{-1,1\})$ 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 $m \in \mathrm{poly}(n)$. No geometric locality or clean computation constraints are assumed. A formal definition appears later as Definition 12.

Certainly $\mathsf{QNC}^0 \subseteq \mathsf{AC}^0 \circ \mathsf{QNC}^0$, so the search separation between QNC^0 and AC^0 in Bene Watts et al. is also a search separation between $\mathsf{AC}^0 \circ \mathsf{QNC}^0$ 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":

> *Parity halving.* Given $x \in \{0,1\}^n$ with the promise $|x| \equiv 0 \mod 2$, output any even string if $|x| \equiv 0 \mod 4$ 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 $\mathsf{AC}^0 \circ \mathsf{QNC}^0$.

▶ Remark 2. Before summarizing our progress on the parity question, we pause to note another reason to study $\mathsf{AC}^0 \circ \mathsf{QNC}^0$ 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 QNC^0 and AC^0 respectively and interact for constantly-many rounds to establish the correctness of a QNC^0 computation. With this perspective we see that $\mathsf{AC}^0 \circ \mathsf{QNC}^0$ models the first round of interaction in such a proof system.

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

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

$$\mathop{\mathbb{E}}_{x}[(f \circ \mathcal{C})(x) \cdot \textsc{Par}(x)],$$

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

▶ **Conjecture 3.** $\mathsf{AC}^0 \circ \mathsf{QNC}^0$ *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* $\{(f_n, \mathcal{C}_n)\}_n$ *of circuits such that* $\mathrm{size}(f_n), \mathrm{size}(\mathcal{C}_n) \leq p(n)$ *and* $\mathrm{depth}(f_n), \mathrm{depth}(\mathcal{C}_n) \leq d$*, we have*

$$\mathop{\mathbb{E}}_{x}[(f_n \circ \mathcal{C}_n)(x) \cdot \textsc{Par}_n(x)] \to 0 \quad \textit{as} \quad n \to \infty\,.$$

Although proving correlation bounds against AC^0 is a well-understood topic with many techniques (among them Håstad's switching lemma [13] and Razborov-Smolensky [22, 25]), when 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 QNC^0 is restricted to be ancilla-free. A key feature of such QNC^0 circuits is that they correspond to unitary transformations, and we find in this case the correlation of $f \circ \mathcal{C}$ with PAR is controlled by the Fourier tail of f. Recall the *Fourier tail* of a Boolean function f is given by

$$\mathbf{W}^{\geq t}[f] := \textstyle\sum_{|S| \geq t} \widehat{f}(S)^2\,.$$

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

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

$$\mathop{\mathbb{E}}_{x}[(f \circ \mathcal{C})(x) \cdot \textsc{Par}_n(x)] \leq 2^{-n/\mathrm{polylog}(n)}\,.$$

This is proved as Corollary 8 in Section 1. The full statement holds for any Boolean function f with sufficient decay in the tail of the Fourier spectrum, including those outside of AC^0.

However, as we explain in the end of Section 1, the proof technique of Theorem 4 cannot extend to the case of general QNC^0 and we must find a different approach.

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 *ε-approximate degree* of a $(0,1)$-valued[1] Boolean function f is given by

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

Of course, $\widetilde{\deg}_\varepsilon[f] \leq n$ for any n-variate f and $\varepsilon > 0$. By convention $\widetilde{\deg}[f] := \widetilde{\deg}_{1/3}[f]$. A *function class* $\mathcal{F} = (\mathcal{F}_n)_{n \geq 1}$ is a sequence of sets $\mathcal{F}_n$ of n-variate Boolean functions, and we extend approximate degree to function classes via $\widetilde{\deg}[\mathcal{F}](n) := \max_{f \in \mathcal{F}_n} \widetilde{\deg}[f]$. 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 $\widetilde{\deg}[\mathcal{F}] \in o(n)$, $\mathcal{F} \circ \mathsf{QNC}^0$ cannot achieve $\Omega(1)$ correlation with $\textsc{Par}_n$, even if QNC^0 is given arbitrary quantum advice.*

It follows from Theorem 5 that Conjecture 1 would be confirmed in full generality if $\widetilde{\deg}[\mathsf{AC}^0] \in o(n)$, 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 $\mathcal{O}(n)$ (termed LC^0), we may appeal to the recent bounds of [8] to conclude:

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

$$\mathbb{E}[(\mathsf{LC}^0 \circ \mathsf{QNC}^0/\mathsf{qpoly}) \cdot \textsc{Par}_n] \in \operatorname{negl}(n).$$

(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 $\widetilde{\deg}[\mathsf{AC}^0]$ but would still imply Conjecture 3.

Towards an $\mathsf{AC}^0 \circ \mathsf{QNC}^0$ switching lemma

In Section 3 we chart a different route to resolving Conjecture 3, aiming to prove a switching lemma for our hybrid $\mathsf{AC}^0 \circ \mathsf{QNC}^0$ 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, $\textsc{Par}$ 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 $\mathsf{AC}^0 \circ \mathsf{QNC}^0$ 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 (± 1)-valued f, we use the same definition after making the standard identification $+1 \mapsto 0, -1 \mapsto 1$.

replaced with randomized NC^0 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^0 is replaced by an n-party nonlocal channel $\mathcal{N}$, it *is* possible to pull out the quantumness in a particular sense:

▶ **Theorem** (Theorem 25, restated). *Let $f : \{0,1\}^m \to \{0,1\}$ be any Boolean function and consider an n-party nonlocal channel $\mathcal{N}$, where the i^{th} party receives one bit and responds with $m_i \geq 0$ bits, such that $\sum_i m_i = m$. Then the random function $f \circ \mathcal{N}$ is equal to a randomized decision tree Γ such that $\mathrm{depth}(T) \leq \mathrm{DT}_{\mathrm{depth}}[f]$ for all $T \in \mathrm{Supp}(\Gamma)$.*

(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 $\mathrm{DT}_{\mathrm{depth}}[f] \geq \log(n)$. 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^0 by the channel $\mathcal{N}$ is essentially without loss of generality from the point of view of Conjecture 3. Unfortunately, however, AC^0 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 $\mathcal{N}$, represents the style of argument that could eventually lead to a switching lemma for $\mathsf{AC}^0 \circ \mathsf{QNC}^0$.

Outlook

Taken together, these results suggest QNC^0 cannot render its power in a way AC^0 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^0 [18, 14, 21, 23, 19]. A superset of QNC^0 circuits, QAC^0 additionally allows for arbitrarily-large Toffoli gates,

$$|x_1, \ldots, x_k, x_{k+1}\rangle \mapsto \left|x_1, \ldots, x_k,\ x_{k+1} \oplus (\wedge_{i=1}^{k} x_i)\right\rangle,$$

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 QAC^0 cannot achieve $\Omega(1)$ correlation with parity either. Recent work has shown negligible correlation bounds between QAC^0 and parity when a) the QAC^0 circuit is restricted to depth 2 [23], and b) when the QAC^0 circuit is of any depth d and is restricted to $\mathcal{O}(n^{1/d})$-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 QAC^0 [19].

The relationship between QAC^0 and $\mathsf{AC}^0 \circ \mathsf{QNC}^0$ is rather unclear, and they are likely incomparable as decision classes. In fact, as far as we know, it is even open whether $\mathsf{AC}^0 \subseteq \mathsf{QAC}^0$, let alone whether $\mathsf{AC}^0 \circ \mathsf{QNC}^0 \subseteq \mathsf{QAC}^0$ (noting the trivial containment $\mathsf{AC}^0 \subseteq \mathsf{AC}^0 \circ \mathsf{QNC}^0$).

The difficulty in comparing these models stems from a subtlety concerning the difference between unbounded fan-in and unbounded fan-out when implemented coherently. AC^0 circuits have no restriction on the *fan-out* of their gates, while the definition of QAC^0 appears to strongly limit outward propagation of information. If one augments QAC^0 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 QAC^0_f, and it is known QAC^0_f can compute parity exactly in depth 3 [18]. In view of existing lower bounds against QAC^0, it is expected that QAC^0 is strictly contained in QAC^0_f, and assuming this holds we immediately would have that the function version of $\mathsf{AC}^0 \circ \mathsf{QNC}^0$ is not in the function version of QAC^0. This follows, for example, from the fact that multi-output AC^0 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 QNC^0 – whether that interaction is coherent as in QAC^0 and QAC^0_f, or preceded by measurement as in $\mathsf{AC}^0 \circ \mathsf{QNC}^0$ – 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 $\mathsf{AC}^0 \circ \mathsf{QNC}^0$ model.

1 Lower bounds when QNC^0 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 QNC^0. By the celebrated work of Håstad [13] and Linial, Mansour, and Nisan [15], any $f \in \mathsf{AC}^0$ 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*

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

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

$$\mathop{\mathbb{E}}_x[f(x)g(x)] = \sum_{S \subseteq [n]} \widehat{f}(S)\widehat{g}(S) \, .$$

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

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

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

$$A^S := \bigotimes_i \begin{cases} A & \text{if } i \in S \\ \mathbb{1} & \text{otherwise.} \end{cases}$$

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

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

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 $\textsc{Par}$ is bounded as*

$$\mathop{\mathbb{E}}_x[\langle U^\dagger M_f U\rangle_x \cdot \textsc{Par}_n(x)] \;\leq\; \left(\mathbf{W}^{\geq 2^{-t}n}[f]\right)^{1/2} \, .$$

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

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

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 $\textsc{Par}_n \circ U^\dagger$, defined by applying $\textsc{Par}_n$ 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*

$$\begin{aligned} \mathop{\mathbb{E}}_x[\langle U^\dagger M_f U\rangle_x \cdot g(x)] &= \mathop{\mathbb{E}}_x[f(x) \cdot \langle U M_g U^\dagger\rangle_x] \\ &= 2^{-n} \operatorname{tr}[M_f U M_g U^\dagger] \, . \end{aligned}$$

Proof. Expanding the trace we have

$$\begin{aligned}\operatorname{tr}[M_f U M_g U^\dagger] &= \sum_z \langle z| \big(\textstyle\sum_y f(y)|y\rangle\langle y|\big) U \big(\textstyle\sum_x g(x)|x\rangle\langle x|\big) U^\dagger |z\rangle \\ &= \sum_{x,y,z} f(y)g(x) \langle z|y\rangle \langle y| U^\dagger |x\rangle \langle x| U |z\rangle \\ &= \sum_{x,y} f(y)g(x) \langle y| U |x\rangle \langle x| U^\dagger |y\rangle , \end{aligned} \tag{1}$$

while expanding the expectations we see

$$\mathop{\mathbb{E}}_x[\langle U^\dagger M_f U\rangle_x \cdot g(x)] = \frac{1}{2^n}\sum_{x,y} f(y)g(x) \langle x| U^\dagger |y\rangle \langle y| U |x\rangle = \mathop{\mathbb{E}}_y[f(y) \cdot \langle U M_g U^\dagger\rangle_y] .$$

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 $\langle A, B\rangle = \operatorname{tr}[A^\dagger B]$) 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*

$$\operatorname{tr}[Z_{[n]} U Z_S U^\dagger] = 0$$

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 $UZ_SU^\dagger$ is at most $|S|2^t$. Now if $|S|2^t < n$, then there is at least one coordinate j such that $UZ_SU^\dagger = V_{[n]\setminus j} \otimes \mathbb{1}_j$ for some $(n-1)$-qubit unitary $V_{[n]\setminus j}$, so

$$\operatorname{tr}[Z_{[n]} U Z_S U^\dagger] = \operatorname{tr}[Z_{[n]}(V_{[n]\setminus j} \otimes \mathbb{1}_j)] = \operatorname{tr}[Z_{[n]\setminus j} V_{[n]\setminus j}] \cdot \operatorname{tr}[Z] = 0$$

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.

$$\begin{aligned}
\mathop{\mathbb{E}}_x[\langle U^\dagger M_f U\rangle_x \cdot \chi_{[n]}(x)] &= \mathop{\mathbb{E}}_x[f(x) \cdot \langle U Z_{[n]} U^\dagger\rangle_x] && \text{(Lemma 9)} \\
&= \sum_{S\subseteq[n]} \widehat{f}(S) \cdot \widehat{\langle U Z_{[n]} U^\dagger\rangle}(S) && \text{(Plancherel)} \\
&= \sum_{S\subseteq[n]} \widehat{f}(S) \underbrace{\mathop{\mathbb{E}}_x[\langle U Z_{[n]} U^\dagger\rangle_x \cdot \chi_S(x)]}_{} \\
&\qquad = 2^{-n} \operatorname{tr}[Z_{[n]} U^\dagger Z_S U] && \text{(Lemma 9)} \\
&\qquad = 0 \quad \text{if } |S|2^t < n && \text{(Lemma 10)} \\
&= \sum_{\substack{S\subseteq[n] \\ |S|\geq 2^{-t}n}} \widehat{f}(S) \cdot \widehat{\langle U^\dagger Z_{[n]} U\rangle}(S)
\end{aligned}$$

$$\leq \left(\sum_{|S|\geq 2^{-t}n} \widehat{f}(S)^2\right)^{1/2} \left(\sum_{|S|\geq 2^{-t}n} \widehat{\langle U^\dagger Z_{[n]} U\rangle}(S)^2\right)^{1/2} \quad \text{(Cauchy-Schwarz)}$$

$$\leq \left(\mathbf{W}^{\geq 2^{-t}n}[f]\right)^{1/2}. \qquad \blacktriangleleft$$

One may ask whether this proof approach extends to QNC^0 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 QNC^0 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 $n = 2m$:

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

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

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

This means for any $t \geq 1$, the t^{th} Fourier tail of h is $\mathbf{W}^{\geq t}[h] \in \mathcal{O}(2^{-n/2})$. Thus by Theorem 7, for any ancilla-free QNC^0 circuit $\mathcal{C}$, $h \circ \mathcal{C}$ has negligible correlation with parity.

On the other hand, consider the (deterministic) function $C : \{\pm 1\}^m \to \{\pm 1\}^{2m}$ given by $x \mapsto x11\cdots 1$. Certainly C can be implemented in QNC^0, and we have $h \circ C = \chi_{[m]} = \textsc{Par}_m$.

This example shows that exponential Fourier decay of f is not sufficient to entail Conjecture 3 for general $\mathsf{AC}^0 \circ \mathsf{QNC}^0$ circuits. We must take a different approach that exploits finer structural properties of AC^0 and QNC^0.

2 Lower bounds against $\mathsf{AC}^0 \circ \mathsf{QNC}^0$ via nonlocal games

Here we pass from QNC^0 to nonlocal games to make an argument that works for general QNC^0. First let us fix ideas about QNC^0.

▶ **Definition 12** (QNC^0). *An n-input, depth-d QNC^0 circuit $\mathcal{C}$ 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 $\mathcal{C}$ in the computational basis, the circuit $\mathcal{C}$ effects a randomized mapping from n bits of input to $n + m \in \text{poly}(n)$ bits of output. A* QNC^0 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* $\mathsf{QNC}^0/\mathsf{qpoly}$.

We will show a reduction from QNC^0 circuits to nonlocal channels.

▶ **Definition 13** (Nonlocal channel). *Let $n, k \geq 1$ and $m \geq 0$. An (n, k, m) nonlocal channel is the randomized mapping defined by a quantum strategy in a nonlocal game where n parties receive one bit of input each and respond with k bits each, along with a referee response of m bits.*

Concretely, each party $i \in [n]$ is assigned a local Hilbert space $\mathcal{H}_i$ and for each $b \in \{0,1\}^n$, a POVM

$$M_{(i,b)} = \left\{M^y_{(i,b)} : y \in \{0,1\}^k\right\}$$

on $\mathcal{H}_i$. There is also a referee Hilbert space $\mathcal{H}_{\text{ref}}$ with a fixed POVM

$$M_{\text{ref}} = \left\{M^y_{\text{ref}} : y \in \{0,1\}^m\right\}.$$

The definition of the nonlocal channel is completed by a choice of shared state $|\psi\rangle \in \left(\bigotimes_{i=1}^n \mathcal{H}_i\right) \otimes \mathcal{H}_{\text{ref}}$ 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 $\left(M_{(1,x_1)}, \ldots, M_{(n,x_n)}, M_{\text{ref}}\right)$ on $|\psi\rangle$, resulting in the outcomes $y_1, \ldots, y_n$, and y_{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 $f : \{0,1\}^{kn+m} \to \{0,1\}$. 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 (quantum) parity game* strategy *is an (n, k, m) nonlocal channel with output string y. Players win when $f(y) = \textsc{Par}(x)$. We say a parity game strategy has* advantage *ε 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 $\mathcal{C}$ be a n-qubit, depth-d QNC^0 circuit with m ancilla and arbitrary quantum advice, and let $f : \{0,1\}^{n+m} \to \{0,1\}$ be any Boolean function. Suppose $f \circ \mathcal{C}$ has correlation ε with $\textsc{Par}_n$. 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 $f \circ \mathcal{C}$ has correlation ε with $\textsc{Par}$. 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 $y \in \{0,1\}^{S^c}$, define the circuit $\mathcal{C}_y$ to be $\mathcal{C}$ but where for $j \in S^c$, the j^{th} input is hardcoded to y_j. Then $\mathcal{C}_y$ 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 $(n', 2^d, m')$ nonlocal channel for some $n' \geq 2^{-d} + 1$ and $m' = n + m - n'2^d$. (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

$$\begin{aligned}\mathop{\mathbb{E}}_{x}[(f\circ\mathcal{C})(x)\cdot \textsc{Par}(x)] &= \mathop{\mathbb{E}}_{y\sim\{0,1\}^{S^c}}\mathop{\mathbb{E}}_{z\sim\{0,1\}^{S}}[f\circ\mathcal{C}_y(z)\cdot\textsc{Par}(y||z)]\\ &= \mathop{\mathbb{E}}_{y\sim\{0,1\}^{S^c}}\textsc{Par}(y)\mathop{\mathbb{E}}_{z\sim\{0,1\}^{S}}[f\circ\mathcal{C}_y(z)\cdot\textsc{Par}(z)]\,.\end{aligned}$$

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

Lemma 16 shows that bounds on the value of parity games translate into correlation bounds for $\mathsf{AC}^0\circ\mathsf{QNC}^0$ with $\textsc{Par}$. 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 $\mathcal{N}$, begin by rewriting the correlation as

$$\mathbb{E}[(f\circ\mathcal{N})(x)\cdot\textsc{Par}(x)] = \frac{\mathbb{E}[(f\circ\mathcal{N})(x)\mid x \text{ even}] - \mathbb{E}[(f\circ\mathcal{N})(x)\mid x\text{ odd}]}{2}\,.$$

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 $\mathcal{N}$:

$$\mu := \mathcal{N}(\mathcal{U}_{\text{even}}) \qquad \text{and} \qquad \nu := \mathcal{N}(\mathcal{U}_{\text{odd}})\,.$$

So μ and ν are distributions on strings of length $N := nk$, and

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

Therefore the correlation of $f\circ\mathcal{N}$ with parity can be phrased in terms of f's ability to distinguish the distributions μ and ν.

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

$$\mu_S = \nu_S. \tag{2}$$

Here the notation μ_S denotes the marginal distribution of μ 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 $\mathcal{N}$, the marginal μ_S (resp. ν_S) 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 $[n]$, $(U_{\text{even}})_T = (\mathcal{U}_{\text{odd}})_T = \mathcal{U}(\{0,1\}^{|T|})$, so we must have $\mu_S=\nu_S$.

So all small marginals of μ and ν 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 μ and ν on $\{\pm1\}^N$ are k-wise indistinguishable if for all $S\subset[N]$ with $|S|\le k$, $\mu_S=\nu_S$.*

Additionally, for $f:\{0,1\}^n\to\{0,1\}$, we say f is ε-fooled by k-wise indistinguishability if for any pair μ,ν of k-wise indistinguishable distributions,

$$|\Pr[f(\mu)=1]-\Pr[f(\nu)=1]|\le\varepsilon\,.$$

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.

▶ **Theorem 18** ([4, Theorem 1.2]). *Let $f : \{0,1\}^n \to \{0,1\}$ and $\varepsilon > 0$. Then f is ε-fooled by k-wise indistinguishability if and only if $\widetilde{\deg}_{\varepsilon/2}[f] \leq k$.*

With this fact, Lemma 16, and the above discussion, we are ready prove the main theorem in this section.

We say a class of Boolean functions $\mathcal{F} = (\mathcal{F}_n)_{n\geq 1}$ is *closed under inverse-polynomial restrictions* if for all $f \in \mathcal{F}_n$ and all $S \subseteq [n]$ with $n \in \mathrm{poly}(|S|)$, fixing the bits in S^c yields a function still in $\mathcal{F}$:

$$f\restriction_{S^c \leftarrow x} \in \mathcal{F}_{|S|} \quad \forall x \in \{0,1\}^{|S^c|}.$$

Note that AC^0 is closed under inverse-polynomial restrictions.

▶ **Theorem 19.** *Suppose $\mathcal{F}$ is a class of Boolean functions closed under negations and inverse-polynomial restrictions. Let m be fixed and suppose there is an $f \in \mathcal{F}$ on $N = \mathrm{poly}(m)$ 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 ε with $\textsc{Par}_m$. Then there is a $g \in \mathcal{F}$ on $n \geq m/2$ variables with $\widetilde{\deg}_{\varepsilon/2}[g] \geq n/2^d - 1$.*

Proof. By Lemma 16, there is an $m' \geq m/(2^d+1)$ and an $(m', 2^d, N - 2^d m')$ nonlocal channel $\mathcal{N}$ such that $f \circ \mathcal{N}$ or $\neg f \circ \mathcal{N}$ achieves correlation ε with $\textsc{Par}_{m'}$.

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 $(m', 2^d, 0)$ nonlocal channel $\mathcal{N}^{R\leftarrow r}$. 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 $\mathcal{N}^{R\leftarrow r}$ still yields correlation ε with $\textsc{Par}$. Define $g := f\restriction_{R\leftarrow r}$ or $g := \neg f\restriction_{R\leftarrow r}$ as appropriate and put $\mathcal{E} := \mathcal{N}^{R\leftarrow r}$. Then $g \in \mathcal{F}$ is a function on $n := 2^d m'$ bits and

$$\mathop{\mathbb{E}}_x[(g \circ \mathcal{E})(x) \cdot \textsc{Par}(x)] \geq \varepsilon.$$

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

$$\widetilde{\deg}_{\varepsilon/2}[g] \geq \frac{n}{2^d} - 1. \qquad \blacktriangleleft$$

▶ **Corollary 20.** *Suppose function class $\mathcal{F}$ is closed under inverse-polynomial-sized restrictions. Then if $\widetilde{\deg}[\mathcal{F}] \in o(n)$, $\mathcal{F} \circ \mathsf{QNC}^0$ cannot achieve $\Omega(1)$ correlation with $\textsc{Par}$.*

The burning question, then, is whether $\widetilde{\deg}[\mathsf{AC}^0] \in o(n)$. In fact, the approximate degree of 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 AC^0.

▶ **Theorem** ([8, Theorem 5]). *Let $p(n) \in \mathrm{poly}(n)$. Then the class of AC^0 circuits of linear size, denoted by LC^0, has*

$$\widetilde{\deg}_{1/p(n)}[\mathsf{LC}^0] \in o(n).$$

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

▶ **Theorem.** *For any $\delta > 0$, there is a function $f \in \mathsf{AC}^0$ with $\widetilde{\deg}[f] \in \Omega(n^{1-\delta})$.*

The lower bound of $\Omega(n^{1-\delta})$-for-any-δ 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 $\widetilde{\deg}[\mathsf{AC}^0] \in \Theta(n/\log n)$ either. Several questions – including now Conjecture 3 – could be settled if the gap between $\Omega(n^{1-\delta})$-for-any-δ and n for $\widetilde{\deg}[\mathsf{AC}^0]$ were closed.

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

▶ **Corollary 21.** *Let $\mathcal{C}$ 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 $\mathcal{O}(n)$. Then $f \circ \mathcal{C}$ achieves negligible correlation with $\textsc{Par}_n$.*

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 $\widetilde{\deg}[\mathsf{AC}^0]$.

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

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

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

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

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

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

$$\widetilde{\mathrm{bdeg}}_k[\mathsf{AC}^0] \stackrel{?}{\leq} n/k - 1\,.$$

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

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

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 $f \in \mathsf{AC}^0$, the left-hand side holding for arbitrary constant k is equivalent to $\widetilde{\deg}[\mathsf{AC}^0] \in o(n)$, 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 $\mathcal{E}$ from the proof of Theorem 19, with n/k players responding with k bits each. Defining μ and ν as the pushforwards of uniform distributions over even and odd bitstrings as before, it is true that μ and ν 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, μ and ν 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 $\widetilde{\deg}[\mathsf{AC}^0] \stackrel{?}{\in} o(n)$. We can compare the two questions head-to-head as follows. Let $\mathcal{P}_k$ be all the relabelings of P:

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

For any $P \in \mathcal{P}_k$, let $\mathrm{bdeg}_P[f]$ 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:

$$\widetilde{\deg}[\mathsf{AC}^0] < n/k \iff \forall f \in \mathsf{AC}^0, \exists g, \forall P \in \mathcal{P}_k, \mathrm{bdeg}_P[g] \le n/k \text{ and } \|f-g\|_\infty \le 1/3$$
$$\widetilde{\mathrm{bdeg}}_k[\mathsf{AC}^0] < n/k \iff \forall f \in \mathsf{AC}^0, \forall P \in \mathcal{P}_k, \exists g, \mathrm{bdeg}_P[g] \le n/k \text{ and } \|f-g\|_\infty \le 1/3\,.$$

3 Towards a switching lemma for $\mathsf{AC}^0 \circ \mathsf{QNC}^0$

Recall that our approach in Section 1 fails because circuits with LMN-style Fourier decay are not suitably closed under precomposition by QNC^0. In fact this is true even under precomposition by 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 $0 \le \delta \le 1$,*

$$\Pr_{\rho\sim\mathbf{R}_\delta}[\mathrm{DT}_{\mathrm{depth}}(f\restriction_\rho) > t] \le (C\delta w)^t\,,$$

where C is a universal constant.

Here $\mathbf{R}_\delta$ is the distribution of random restrictions with star probability δ (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 $\mathsf{AC}^0 \circ \mathsf{QNC}^0$ would mean to track the passage of random restrictions through QNC^0 – a tall order given that 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 $m \ge 0$ and $k, w, n \ge 1$ and suppose $f : \{0,1\}^{kn+m} \to \{0,1\}$ is a DNF of width w and $\mathcal{N}$ is an (n,k,m) nonlocal channel. Then for each restriction ρ there exists a distribution Γ_ρ over decision trees such that $(f\circ\mathcal{N})\restriction_\rho = \{T\}_{T\sim\Gamma_\rho}$ and
>
> $$\Pr_{\rho\sim\mathbf{R}_\delta}\Pr_{T\sim\Gamma_\rho}[\mathrm{depth}(T) > t] \le (C\delta w)^t\,.$$

By Lemma 16 such a switching lemma would be sufficient to show correlations bounds between $f \circ \mathsf{QNC}^0$ 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 τ is equal to a probability distribution Γ of decision trees with* $\mathrm{depth}(\tau') \leq \mathrm{depth}(\tau)$ *for all* $\tau' \in \mathrm{Supp}(\Gamma)$.

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 ν a probability measure on $\mathcal{M}(X)$ we may define the expected distribution

$$\mathop{\mathbb{E}}_{\mu\sim\nu}[\mu] := \left\{x \text{ w.p. } \Pr_{\mu\sim\nu}\Pr_{z\sim\mu}[z = x]\right\}_{x\in X} \tag{3}$$

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 μ on the set of channels from n to N bits, we use $\mathbb{E}_{\mathcal{N}\sim\mu}\mathcal{N}$ to denote the channel defined pointwise as:

$$\left(\mathop{\mathbb{E}}_{\mathcal{N}\sim\mu}\mathcal{N}\right)(x) := \mathop{\mathbb{E}}_{\mathcal{N}\sim\mu}[\mathcal{N}(x)]. \tag{4}$$

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*

$$\mathcal{N}^T(x) := \left\{y \quad \text{w.p.} \sum_{\substack{z\in\{\pm1\}^N \\ z_T = y}} \Pr[\mathcal{N}(x) = z]\right\}_{y\in\{\pm1\}^{|T|}}.$$

▶ **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 $x_{B(T)\setminus\{\perp\}}$ 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

$$f \circ \mathcal{N}(x) = \left\{b \quad \text{w.p.} \Pr_{y\sim\mathcal{N}(x)}[f(y) = b]\right\}_{b\in\{\pm1\}}.$$

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 Γ over decision trees such that*

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

(ii) *For all $\tau \in \mathrm{Supp}(\Gamma)$, $\mathrm{DT}_{depth}(\tau) \leq \mathrm{DT}_{depth}(f)$.*

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

$$\mathbb{E}[f \circ \mathcal{N}] = \sum_{S\subset[n]} a_S \chi_S \qquad \text{with} \qquad a_S := \mathop{\mathbb{E}}_{x}\left[\mathbb{E}[f \circ \mathcal{N}] \cdot \chi_S(x)\right]$$

We pause to note the related fact that in terms of the expected output $\mathbb{E}[f \circ \mathcal{N}]$, the degree of any function f does not increase under composition with an NSC: $\deg(f) \geq \deg(\mathbb{E}[f \circ \mathcal{N}])$.

This claim has a very simple direct proof[2] and we emphasize that it is not equivalent to Theorem 25. For example, there are Boolean functions g with $\deg(g) = n^{2/3}$ but $\mathrm{DT}_{\mathrm{depth}}(g) = n$ (see Example 3 in [7]). One could imagine a Boolean function h with $\deg(h) \approx \mathrm{DT}_{\mathrm{depth}}(h) \in o(n)$ 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 requiries some bookkeeping. The idea is to begin with τ's root vertex variable y_i and locally decompose the univariate channel $\mathcal{N}^{\{i\}}(x) \rightsquigarrow y_i$ into a distribution of deterministic functions $\{y_{i,\omega}(x_i)\}_{\omega\sim\mu}$. This decomposition of the root vertex induces a probabilistic decomposition $\{\tau'_\omega \circ \mathcal{N}'_\omega\}_{\omega\sim\mu}$ of the entire hybrid computation where the root variable y_i in τ' has been replaced with an $x_{B(i)}$ and the left and right subtrees of τ 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 $\mathcal{N}'$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 $Y \in \mathrm{Supp}(\mathcal{N}^J(x))$ define the *conditional channel* as

$$\mathcal{N}(x \mid y_J = Y) := \big\{y \quad \text{w.p. } \Pr[\mathcal{N}(x) = y \mid y_J = Y]\big\}_{y \in \{\pm 1\}^N},$$

and for $T \subseteq [N]$ the *reduced conditional channel*

$$\mathcal{N}^T(x \mid y_J = Y) := \Big\{y \quad \text{w.p.} \sum_{\substack{z \in \{\pm 1\}^N \\ z_T = y}} \Pr[\mathcal{N}(x) = z \mid z_J = Y]\Big\}_{y \in \{\pm 1\}^{|T|}}.$$

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

$$\mathcal{G}^{\{i\}}(x \mid y_{[n]\setminus i} = 00\cdots 0) = \Big\{\prod_j x_j \quad \text{w.p. } 1\Big\},$$

the parity of all n bits of x. All the same, some structure remains after conditioning:

▶ **Proposition 26.** *For $T, J \subseteq [N]$, let $x, x' \in \{\pm 1\}^n$ be such that $x_{B(J\cup T)} = x'_{B(J\cup T)}$. Then for all $Y \in \mathrm{Supp}(\mathcal{N}^J(x))$,*

$$\mathcal{N}^T(x \mid y_J = Y) \;=\; \mathcal{N}^T(x' \mid y_J = Y).$$

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

[2] Consider the Fourier expansion $f = \sum_{S\subseteq[N]} \widehat{f}(S)\chi_S$. Then $\mathbb{E}[(f \circ \mathcal{N})(x)] = \mathbb{E}\left[\sum_{S\subseteq[N]} \widehat{f}(S)\chi_S \circ \mathcal{N}(x)\right] = \sum_S \widehat{f}(S)\,\mathbb{E}[\chi_S \circ \mathcal{N}(x)] = \sum_S \widehat{f}(S)\,\mathbb{E}[\chi_S \circ \mathcal{N}^S(x)]$, a linear combination of functions of at most $|S|$ variables each for $|S| \le \deg(f)$.

This proposition says $\mathcal{N}^T(x \mid y_J = Y)$ is a function of $x_{B(J \cup T)}$ only. Thus if we fix variables $x_{B(J)}$ we recover a smaller NSC:

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

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

and otherwise let $\mathcal{N}'$ be the $(n-1)$-to-$(N-1)$ NSC

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

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 $(\mathcal{N}', B')$ 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 $\mathcal{T}$ *on n variables with ℓ leaves consists of the data $(\tau, \mathcal{G}_1, \dots, \mathcal{G}_\ell)$, where*

- **(i)** *The first argument τ is a rooted binary tree with ℓ 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 ι of τ is associated with an n-to-1 channel $\mathcal{G}_\iota : \{\pm 1\}^n \to \mathcal{M}\{\pm 1\}$.*

A hybrid tree defines a channel $\mathcal{T}_\tau(\mathcal{G}_1, \dots, \mathcal{G}_\ell) : \{\pm 1\}^n \to \mathcal{M}\{\pm 1\}$ as follows. Computation on input $x \in \{\pm 1\}^n$ proceeds just as with standard decision trees until a leaf ι is reached, at which point the distribution $\mathcal{G}_\iota(x)$ 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 $\mathcal{T}$,

$$\mathcal{T}\big(\mathcal{G}_1, \dots, \mathcal{G}_{\iota-1}, \mathop{\mathbb{E}}_{\omega\sim\mu}[\mathcal{G}_\omega], G_{\iota+1}, \dots, \mathcal{G}_\ell\big) = \mathop{\mathbb{E}}_{\omega\sim\mu}\big[\mathcal{T}(\mathcal{G}_1, \dots, \mathcal{G}_{\iota-1}, \mathcal{G}_\omega, \mathcal{G}_{\iota+1}, \dots, \mathcal{G}_\ell)\big]$$

▷ Claim 30. For any hybrid decision trees $\mathcal{T}_\tau(\mathcal{G}_1, \dots, \mathcal{G}_\ell)$ and $\mathcal{T}_{\tau'}(\mathcal{G}_{\iota 1}, \dots, \mathcal{G}_{\iota\ell'})$,

$$\begin{aligned}\mathcal{T}_\tau\big(\mathcal{G}_1, \dots, \mathcal{G}_{\iota-1}, \mathcal{T}_{\tau'}(\mathcal{G}_{\iota 1}, \dots, \mathcal{G}_{\iota\ell'}), G_{\iota+1}, \dots, \mathcal{G}_\ell\big)& \\ = \mathcal{T}_{\tau \circ_\iota \tau'}(\mathcal{G}_1, \dots, \mathcal{G}_{\iota-1}, \mathcal{G}_{\iota 1}, \dots, \mathcal{G}_{\iota\ell'}, \mathcal{G}_{\iota+1}, \dots, \mathcal{G}_\ell),&\end{aligned}$$

where $\tau \circ_\iota \tau'$ is τ with the ι^{th} leaf replaced with τ'.

▷ Claim 31. Suppose τ is a decision tree and $(\mathcal{N}, B)$ is an NSC. Then either:

- **(i)** $\tau \circ \mathcal{N} = \mathbb{E}_{\omega\sim\mu}[\tau_\omega \circ \mathcal{N}_\omega]$ where $\mathrm{depth}(\tau_\omega) \le \mathrm{depth}(\tau) - 1$, $|\mathrm{Supp}(\mu)| \le 2$, and each $\mathcal{N}_\omega$ is an NSC, or
- **(ii)** $\tau \circ \mathcal{N} = \mathbb{E}_{\omega\sim\mu}\Big[\mathcal{T}_{\tau^*}\big(\tau_{\omega_L} \circ \mathcal{N}_{\omega_L}, \tau_{\omega_R} \circ \mathcal{N}_{\omega_R}\big)\Big]$, where $|\mathrm{Supp}(\mu)| \le 3$, τ^* has one internal node, $\mathrm{depth}(\tau_{\omega_L}), \mathrm{depth}(\tau_{\omega_R}) \le \mathrm{depth}(\tau) - 1$, and each $\mathcal{N}_{\omega_L}, \mathcal{N}_{\omega_R}$ is an NSC; or
- **(iii)** (Base case) $\tau \circ \mathcal{N}(x) = \{b \text{ w.p. } 1\}$ for all x, for some fixed $b \in \{\pm 1\}$.

Proof. If τ 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 τ. There are two cases depending on the value of $B(i)$.

Case i), $B(i) = \perp$**.** Observe that $\mathcal{N}^{\{i\}}(x)$ is the same distribution μ over $\{\pm 1\}$, independent of x. For $\omega \in \{\pm 1\}$ let τ_ω be the subtree of τ attached to the ω-valued edge of y_i. Put $\mathcal{N}_\omega = \mathcal{N}^{T\setminus\{i\}}(x \mid y_i = \omega)$. Then we have for $z \in \{\pm 1\}$,

$$\begin{aligned}\Pr[\tau \circ \mathcal{N}(x) = z] &= \textstyle\sum_{\omega\in\{\pm1\}} \Pr[\tau \circ \mathcal{N}(x) = z \mid D^i(x) = \omega] \Pr[\mathcal{N}^i(x) = \omega] \\ &= \textstyle\sum_{\omega\in\{\pm1\}} \Pr[\tau_\omega \circ \mathcal{N}(x \mid y_i = \omega) = z] \Pr[\mathcal{N}^i(x) = \omega] \\ &= \textstyle\sum_{\omega\in\{\pm1\}} \Pr[\tau_\omega \circ \mathcal{N}_\omega(x) = z] \Pr[\mathcal{N}^i(x) = \omega] \\ &= \Pr\big[\mathop{\mathbb{E}}_{\omega\sim\mu}[\tau_\omega \circ \mathcal{N}_\omega](x) = z\big]\end{aligned}$$

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

Case ii), $B(i) \neq \perp$**.** Let τ^* be the one-vertex tree consisting of the root vertex of τ relabeled with $x_{B(i)}$ and let τ_1, τ_{-1} be the left and right subtrees of τ respectively. Observe that $\mathcal{N}^{\{i\}}(x) = \mathcal{N}^{\{i\}}(x_{B(i)})$ is a univariate channel. Hence it can be decomposed as a convex combination

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

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

$$\tau \circ \mathcal{N} = \mathop{\mathbb{E}}_{(L,R)\sim\mu}\big[\mathcal{T}_{\tau^*}\big(\tau_L \circ \mathcal{N}_L^{(1)}, \tau_R \circ \mathcal{N}_R^{(-1)}\big)\big] \tag{5}$$

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

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

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":

$$\begin{aligned}\Pr[\tau \circ \mathcal{N}(x) = z] &= \sum_{L\in\{\pm1\}} \Pr[\tau \circ \mathcal{N}(x) = z \mid \mathcal{N}_i(x) = L] \Pr[\mathcal{N}_i(x) = L] \\ &= \sum_{L\in\{\pm1\}} \Pr[\tau \circ \mathcal{N}(x \mid y_i = L) = z](a_{(L,1)} + a_{(L,-1)}) \\ &= \sum_{L\in\{\pm1\}} \Pr[\tau_L \circ \mathcal{N}(x \mid x_{B(i)} = 1, y_i = L) = z]\big(\textstyle\sum_{R\in\{\pm1\}} a_{(L,R)}\big) \\ &= \sum_{L,R\in\{\pm1\}} a_{(L,R)} \Pr[\tau_L \circ \mathcal{N}_L^{(1)}(x) = z] \\ &= \sum_{L,R\in\{\pm1\}} a_{(L,R)} \Pr[\mathcal{T}_{\tau^*}(\tau_L \circ \mathcal{N}_L^{(1)}, \tau_R \circ \mathcal{N}_R^{(-1)})(x) = z] \\ &= \Pr\Big[\mathop{\mathbb{E}}_{(L,R)\sim\mu}[\mathcal{T}_{\tau^*}(\tau_L \circ \mathcal{N}_L^{(1)}, \tau_R \circ \mathcal{N}_R^{(-1)})](x) = z\Big],\end{aligned}$$

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

Proof of Theorem 25. Let τ 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 $\mathcal{T}_\tau(\mathcal{G}_1, \ldots, \mathcal{G}_\ell)$ is some hybrid tree and $\mathcal{G}_\iota = \tau' \circ \mathcal{N}$ for some nontrivial DT τ' and (potentially conditioned) NSC $\mathcal{N}$. Then depending on the case in Claim 3 we either have

$$\begin{aligned}
\mathcal{T}_\tau(\ldots, \underbrace{\tau \circ \mathcal{N}}_{\text{index } \iota}, \ldots) &= \mathcal{T}_\tau(\ldots, \mathbb{E}_{(L,R)\sim\mu}[\mathcal{T}_{\tau^*}(\tau_{\omega_L} \circ \mathcal{N}_{\omega_L}, \tau_{\omega_R} \circ \mathcal{N}_{\omega_R})], \ldots) && \text{(Claim 31.}i\text{)} \\
&= \mathop{\mathbb{E}}_{(\omega_L,\omega_R)\sim\mu} \left[\mathcal{T}_\tau(\ldots, \mathcal{T}_{\tau^*}(\tau_{\omega_L} \circ \mathcal{N}_{\omega_L}, \tau_{\omega_R} \circ \mathcal{N}_{\omega_R}), \ldots)\right] && \text{(Claim 29)} \\
&= \mathop{\mathbb{E}}_{(\omega_L,\omega_R)\sim\mu} \left[\mathcal{T}_{\tau \circ_\iota \tau^*}(\ldots, \tau_{\omega_L} \circ \mathcal{N}_{\omega_L}, \tau_{\omega_R} \circ \mathcal{N}_{\omega_R}, \ldots)\right], && \text{(Claim 30)}
\end{aligned}$$

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

$$\begin{aligned}
\mathcal{T}_\tau(\ldots, \underbrace{\tau \circ \mathcal{N}}_{\text{index } \iota}, \ldots) &= \mathcal{T}_\tau(\ldots, \mathbb{E}_{\omega\sim\mu}[\tau_\omega \circ \mathcal{N}_\omega], \ldots) && \text{(Claim 31.}ii\text{)} \\
&= \mathop{\mathbb{E}}_{\omega\sim\mu} \left[\mathcal{T}_\tau(\ldots, \tau_\omega \circ \mathcal{N}_\omega, \ldots)\right], && \text{(Claim 29)}
\end{aligned}$$

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

If we repeatedly make these transformations on the elements of Γ, we will eventually be left with a distribution over hybrid decision trees $(\tau, \mathcal{G}, \ldots)$ where each channel $\mathcal{G} = \tau' \circ \mathcal{N}$ 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 τ. ◀

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 $\mathsf{AC}^0 \circ \mathsf{QNC}^0$ non-advantage might use the language of Fourier decay.

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

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:

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

$$f \overset{\mathcal{C}}{\longmapsto} f \circ \mathcal{C}\,.$$

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

1 Scott Aaronson, DeVon Ingram, and William Kretschmer. The acrobatics of BQP. In *Proceedings of the 37th Computational Complexity Conference*, CCC '22, pages 1–17, Dagstuhl, DEU, September 2022. Schloss Dagstuhl – Leibniz-Zentrum fuer Informatik. doi:10.4230/LIPIcs.CCC.2022.20.

2 Anurag Anshu, Nikolas P. Breuckmann, and Chinmay Nirkhe. NLTS Hamiltonians from good quantum codes. In *Proceedings of the 55th Annual ACM Symposium on Theory of Computing*, pages 1090–1096, June 2023. arXiv:2206.13228 [cond-mat, physics:quant-ph]. doi:10.1145/3564246.3585114.

3 Robert Beals, Harry Buhrman, Richard Cleve, Michele Mosca, and Ronald de Wolf. Quantum lower bounds by polynomials. *Journal of the ACM*, 48(4):778–797, July 2001. doi:10.1145/502090.502097.

4 Andrej Bogdanov, Yuval Ishai, Emanuele Viola, and Christopher Williamson. Bounded Indistinguishability and the Complexity of Recovering Secrets. In *Proceedings, Part III, of the 36th Annual International Cryptology Conference on Advances in Cryptology — CRYPTO 2016 - Volume 9816*, pages 593–618, Berlin, Heidelberg, August 2016. Springer-Verlag. doi:10.1007/978-3-662-53015-3_21.

5 Sergey Bravyi, David Gosset, and Robert König. Quantum advantage with shallow circuits. *Science*, 362(6412):308–311, October 2018. Publisher: American Association for the Advancement of Science. doi:10.1126/science.aar3106.

6 Sergey Bravyi, David Gosset, Robert König, and Marco Tomamichel. Quantum advantage with noisy shallow circuits. *Nature Physics*, 16(10):1040–1045, October 2020. Number: 10 Publisher: Nature Publishing Group. doi:10.1038/s41567-020-0948-z.

7 Harry Buhrman and Ronald de Wolf. Complexity measures and decision tree complexity: a survey. *Theoretical Computer Science*, 288(1):21–43, October 2002. doi:10.1016/S0304-3975(01)00144-X.

8 Mark Bun, Robin Kothari, and Justin Thaler. Quantum algorithms and approximating polynomials for composed functions with shared inputs. In *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '19, pages 662–678, USA, January 2019. Society for Industrial and Applied Mathematics.

9 Mark Bun and Justin Thaler. Approximate Degree in Classical and Quantum Computing. *Foundations and Trends® in Theoretical Computer Science*, 15(3-4):229–423, December 2022. Publisher: Now Publishers, Inc. doi:10.1561/0400000107.

10 François Le Gall. Average-case quantum advantage with shallow circuits. In *Proceedings of the 34th Computational Complexity Conference*, CCC '19, pages 1–20, Dagstuhl, DEU, September 2020. Schloss Dagstuhl – Leibniz-Zentrum fuer Informatik. doi:10.4230/LIPIcs.CCC.2019.21.

11 Alexandru Gheorghiu, Theodoros Kapourniotis, and Elham Kashefi. Verification of quantum computation: An overview of existing approaches. *Theory of Computing Systems*, 63(4):715–808, July 2018. doi:10.1007/s00224-018-9872-3.

12 Daniel Grier and Luke Schaeffer. Interactive shallow Clifford circuits: Quantum advantage against NC1 and beyond. In *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2020, pages 875–888, New York, NY, USA, June 2020. Association for Computing Machinery. doi:10.1145/3357713.3384332.

13 J Håstad. Almost optimal lower bounds for small depth circuits. In *Proceedings of the eighteenth annual ACM symposium on Theory of computing*, STOC '86, pages 6–20, New York, NY, USA, November 1986. Association for Computing Machinery. doi:10.1145/12130.12132.

14 Peter Høyer and Robert Špalek. Quantum Circuits with Unbounded Fan-out. In Helmut Alt and Michel Habib, editors, *STACS 2003*, Lecture Notes in Computer Science, pages 234–246, Berlin, Heidelberg, 2003. Springer. doi:10.1007/3-540-36494-3_22.

15 Nathan Linial, Yishay Mansour, and Noam Nisan. Constant depth circuits, Fourier transform, and learnability. *Journal of the ACM*, 40(3):607–620, July 1993. doi:10.1145/174130.174138.

16 Shachar Lovett and Emanuele Viola. Bounded-Depth Circuits Cannot Sample Good Codes. In *2011 IEEE 26th Annual Conference on Computational Complexity*, pages 243–251, June 2011. ISSN: 1093-0159. doi:10.1109/CCC.2011.11.

17 Urmila Mahadev. Classical verification of quantum computations. In *2018 IEEE 59th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 259–267, 2018. doi:10.1109/FOCS.2018.00033.

18 Cristopher Moore. Quantum Circuits: Fanout, Parity, and Counting. Technical Report TR99-032, Electronic Colloquium on Computational Complexity (ECCC), September 1999. ISSN: 1433-8092. URL: https://eccc.weizmann.ac.il//eccc-reports/1999/TR99-032/.

19 Shivam Nadimpalli, Natalie Parham, Francisca Vasconcelos, and Henry Yuen. On the pauli spectrum of qac0, 2023. doi:10.48550/ARXIV.2311.09631.

20 Ryan O'Donnell. *Analysis of Boolean functions*. Cambridge University Press, Cambridge, England, June 2014.

21 Daniel Padé, Stephen Fenner, Daniel Grier, and Thomas Thierauf. Depth-2 QAC circuits cannot simulate quantum parity, May 2020. arXiv:2005.12169 [quant-ph]. doi:10.48550/arXiv.2005.12169.

22 A. A. Razborov. Lower bounds on the size of bounded depth circuits over a complete basis with logical addition. *Mathematical notes of the Academy of Sciences of the USSR*, 41(4):333–338, April 1987. doi:10.1007/BF01137685.

23 Gregory Rosenthal. Bounds on the QAC^0 Complexity of Approximating Parity. In James R. Lee, editor, *12th Innovations in Theoretical Computer Science Conference (ITCS 2021)*, volume 185 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 32:1–32:20, Dagstuhl, Germany, 2021. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. ISSN: 1868-8969. doi:10.4230/LIPIcs.ITCS.2021.32.

24 Alexander A. Sherstov. The approximate degree of DNF and CNF formulas. In *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2022, pages 1194–1207, New York, NY, USA, June 2022. Association for Computing Machinery. doi:10.1145/3519935.3520000.

25 R. Smolensky. Algebraic methods in the theory of lower bounds for Boolean circuit complexity. In *Proceedings of the nineteenth annual ACM symposium on Theory of computing*, STOC '87, pages 77–82, New York, NY, USA, January 1987. Association for Computing Machinery. doi:10.1145/28395.28404.

26 Avishay Tal. Tight bounds on the Fourier spectrum of AC0. In *Proceedings of the 32nd Computational Complexity Conference*, CCC '17, pages 1–31, Dagstuhl, DEU, July 2017. Schloss Dagstuhl – Leibniz-Zentrum fuer Informatik.

27 Adam Bene Watts, Robin Kothari, Luke Schaeffer, and Avishay Tal. Exponential separation between shallow quantum circuits and unbounded fan-in shallow classical circuits. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2019, pages 515–526, New York, NY, USA, June 2019. Association for Computing Machinery. doi:10.1145/3313276.3316404.

28 Adam Bene Watts and Natalie Parham. Unconditional Quantum Advantage for Sampling with Shallow Circuits, January 2023. arXiv:2301.00995 [quant-ph]. doi:10.48550/arXiv.2301.00995.