Quantum Advantage from Sampling Shallow Circuits: Beyond Hardness of Marginals

Daniel Grier
UC San Diego, CA, USA

Daniel M. Kane
UC San Diego, CA, USA

Jackson Morris
UC San Diego, CA, USA

Anthony Ostuni
UC San Diego, CA, USA

Kewen Wu
Institute for Advanced Study, Princeton, NJ, USA

Abstract

We construct a family of distributions $\{\mathcal{D}_n\}_n$ with $\mathcal{D}_n$ over $\{0,1\}^n$ and a family of depth-7 quantum circuits $\{C_n\}_n$ such that $\mathcal{D}_n$ is produced exactly by C_n with the all zeros state as input, yet any constant-depth classical circuit with bounded fan-in gates evaluated on any binary product distribution has total variation distance $1 - e^{-\Omega(n)}$ from $\mathcal{D}_n$. Moreover, the quantum circuits we construct are geometrically local and use a relatively standard gate set: Hadamard, controlled-phase, CNOT, and Toffoli gates. All previous separations of this type suffer from some undesirable constraint on the classical circuit model or the quantum circuits witnessing the separation.

Our family of distributions is inspired by the Parity Halving Problem of Watts, Kothari, Schaeffer, and Tal (STOC, 2019), which built on the work of Bravyi, Gosset, and König (Science, 2018) to separate shallow quantum and classical circuits for relational problems.

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

Keywords and phrases Shallow circuits, sampling, quantum circuits

Digital Object Identifier 10.4230/LIPIcs.ITCS.2026.73

Related Version *Full Version*: `https://arxiv.org/abs/2510.07808` [17]

Funding *Daniel M. Kane*: Supported in part by NSF Medium Award CCF-2107079.
Kewen Wu: Supported by the National Science Foundation under Grant No. DMS-2424441, and by the IAS School of Mathematics.

Acknowledgements AO thanks Farzan Byramji for suggestions improving the presentation. KW thanks David Gosset for helpful discussions motivating the problem.

1 Introduction

One of, if not *the* primary direction in the study of quantum computing is to exhibit computational tasks that can be performed far more efficiently on a quantum computer than on a classical one. There are a number of promising candidates [28, 1, 7], but the quantum superiority of many such algorithms relies on unproven assumptions about computational hardness.

To obtain *unconditional* quantum-classical separations, one must consider classical models of computation against which there are known unconditional lower bounds. Bravyi, Gosset, and König gave the first result of this kind by constructing a search problem which could

17th Innovations in Theoretical Computer Science Conference (ITCS 2026).
Editor: Shubhangi Saraf; Article No. 73; pp. 73:1–73:14
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

be solved by constant-depth quantum circuits, but not constant-depth classical circuits [8]. Formally, $\mathsf{FQNC}^0 \not\subseteq \mathsf{FNC}^0$. Since then, there have been many improvements to this result that consider stronger classical circuit families, different error models, and/or different topologies [36, 18, 9, 19, 10, 15].

Nevertheless, these results still fundamentally use the search paradigm for separating the quantum and classical circuit models (or, in fact, sometimes generalizations of search [18, 16]). Intuitively, one can think of these search problems as follows: the input to the problem is both the number of qubits n *and* a specification of a constant-depth quantum circuit Q, and the goal is to output any bit string in the support of the distribution after measuring $Q\,|0^n\rangle$ in the computational basis. One might wonder if the specification of the quantum circuit is even necessary. That is, is there a single quantum circuit for every n that gives rise to a hard-to-sample distribution? In fact, Bravyi, Gosset, and König asked exactly this question in their original work [8, Section 5].

There are a few reasons why we might want such a separation. First, one goal for proofs of quantum advantage is to help distill the core aspects of quantum computers that make them more powerful than their classical counterparts. Clearly then, a separation from a single family of distributions is desirable in its simplicity. Moreover, such results give complexity-theoretic support for certain quantum advantage experiments in which changing the underlying circuit is extremely difficult [40, 13, 38].

Watts and Parham [37] were the first to answer the challenge of [8] by constructing a family of constant-depth quantum circuits with output distributions that cannot be sampled (even approximately) by constant-depth classical circuits with bounded fan-in gates. Unfortunately, their result has two significant caveats. First, it imposes a strict requirement on the number of input bits to the classical circuit. Second, the quantum circuits they construct contain more-or-less arbitrary single-qubit gates (at least outside the Clifford hierarchy).

These two properties combine to make the "quantum" contribution to the quantum-classical separation less clear. To see this, first notice that the usual method of converting between gate sets does not apply in the constant-depth regime, since the Solovay-Kitaev theorem incurs a polylogarithmic depth overhead [23]. This implies that the choice of which single-qubit gates to allow in the quantum circuit model could ultimately affect which kinds of separations are possible. This consideration has been put into sharp relief by recent work that gives a *product* distribution which cannot be sampled (even approximately) by constant-depth classical circuits with uniformly random input bits [35, 21].

In other words, it is possible to obtain a quantum vs. classical separation with a quantum circuit model that has no entangling gates (as only single-qubit rotation gates are needed to sample from a product distribution), undermining the claim that the separation is related to the powers of quantum mechanics. Indeed, if instead we allowed our classical circuit to have random inputs of arbitrary bias, then they, too, could easily produce the desired distribution. While the result of [37] does allow for classical circuits with biased input bits, the restriction on the number of input bits leaves open the possibility that larger classical circuits may still be able to sample from the distribution.

The main contribution of this work is the construction of a family of distributions that achieves the best properties from all prior works:

▶ **Theorem 1** (Informal Version of Theorem 2). *There is a uniform family of constant-depth quantum circuits $\{Q_n\}_n$ such that*

- Discrete gate set: *Q_n is constructed from Hadamard, controlled-phase,* CNOT*, and Toffoli gates. Furthermore, Q_n has a depth-7, geometrically local implementation.*

- Quantum advantage: *Let* $\{C_n \colon \{0,1\}^* \to \{0,1\}^n\}_n$ *be a family of constant-depth classical circuits (i.e.,* NC^0*), and consider the following two distributions:* C_n *applied to any product distribution; and measuring* $Q_n |0^n\rangle$ *in the computational basis. The total variation distance between these two distributions is* $1 - e^{-\Omega(n)}$.

Our distribution in Theorem 1 is based on a relational problem given by Watts, Kothari, Schaeffer, and Tal [36] to strengthen the previously mentioned separation between NC^0 and QNC^0 [8]. Despite this similar construction, our results are incomparable, as they lower bound the stronger class of AC^0, but our results are in the distributional, rather than relational, setting.

Theorem 1 may also be of modest philosophical interest. Recall that randomness extractors convert weak sources of randomness into a near uniform distribution. In influential work, Trevisan and Vadhan provided extractors for distributions produced by polynomial size circuits, claiming that these "sampl[e]able distributions are a reasonable model for distributions actually arising in nature" [30]. A recent follow up by Ball, Goldin, Dachman-Soled, and Mutreja instead argues that a better choice for "natural sources" are those generated by quantum circuits, since the universe is governed by quantum phenomena [4]. Our main result shows that even in the extremely restricted circuit regime, these two beliefs dramatically differ.

Open Problems

The obvious next question in this line of inquiry, raised earlier in [37, Section 2], is whether a similar distributional separation exists between the classes of AC^0 and QNC^0. In fact, it appears even the weaker task of separating AC^0 from QAC^0 (for sampling distributions) is open. There are several known distributions based on pseudorandom objects that cannot be accurately sampled in AC^0 [24, 5, 33, 34]; it is unclear whether shallow quantum circuits can sample them. We remark that while our distribution is based on a problem from [36] which separates AC^0 and QNC^0 for relational problems, our distribution *can* be sampled by an AC^0 circuit (see [17, Remark 5.10]).

Another direction is to refine the quantum gate set. The quantum circuits in our main separation result only require Hadamard, controlled-phase, and Toffoli gates, as opposed to the rotation gates required to generate the (1/3)-biased product distribution used in previous separations [35, 21]. Still, one may wish to further limit the gate set, especially in light of the fact that Hadamard and Toffoli gates are quantum universal [2]. Unfortunately, the standard techniques to simulate the controlled-phase gates in this manner do not naively work in our setting (see [17, Subsection 4.1]), and we leave the minimal gate set required to separate NC^0 and QNC^0 for sampling distributions as an open question.

One final direction deserving of further investigation is hardness amplification for sampling. Our proof of Theorem 1 crucially uses a direct product theorem for sampling in NC^0 (see [17, Subsection 5.4]), which allows us to amplify a weak separation between NC^0 and QNC^0. A similar direct product theorem (or more generally, a hardness amplification result) for sampling in AC^0 would likely be useful in addressing open separations. Note that such a theorem was asked for by Chattopadhyay, Goodman, and Zuckerman [11] who gave an analogous result for read-once branching programs.

2 The Proof Outline

In this section, we provide an overview of the proof of Theorem 2 – a more precise and quantitative version of Theorem 1 parameterized by *locality*. The details can be found in the full version of the paper [17]. A function $f \colon \{0,1\}^* \to \{0,1\}^n$ is d-local if no output bit depends on more than d input bits. Note that any family of NC^0 circuits computes functions of constant locality. We will often refer directly to a distribution as d-local if it is the result of applying a d-local function to random inputs drawn from a product distribution.

▶ **Theorem 2.** *Let $d \geq 1$ be an integer. There exists a constant $c_d > 0$ depending only on d such that the following holds. There is a uniform family of distributions $\{\mathcal{D}_n\}_{n \geq c_d}$ with $\mathcal{D}_n$ over $\{0,1\}^n$ such that*

- *There exists a family of geometrically local depth-7 quantum circuits $\{C_n\}_{n \geq c_d}$ where $\mathcal{D}_n$ is produced exactly by C_n on input $\left|0^{2n}\right\rangle$. In addition, the quantum circuits only uses Hadamard, controlled-phase,* CNOT*, and Toffoli gates, and measurements in the computational basis. Moreover, Hadamard gates are only applied in the first and last layers, i.e., $\{C_n\}_n$ is in the second level of the Fourier Hierarchy [27].*
- *For all $n \geq c_d$, $\mathcal{D}_n$ has total variation distance $1 - e^{-n/c_d}$ from any d-local distribution with any binary product distribution as input.*

▶ Remark 3. Given Theorem 2, it is natural to wonder whether every distribution produced by NC^0 circuits can be sampled by QNC^0 circuits. The following example shows that this is not the case. Consider the distribution $\mathcal{P}$ over $\{0,1\}^n$ which takes value 0^n with probability 1/2 and 1^n otherwise. A classical circuit can easily produce $\mathcal{P}$ by having each output bit mirror the same input bit. QNC^0 circuits, however, cannot generate $\mathcal{P}$, as doing so is equivalent to preparing a nekomata (first defined in [25]), i.e., a state of the form

$$|\psi\rangle = \frac{|0^n\rangle |\psi_0\rangle + |1^n\rangle |\psi_1\rangle}{\sqrt{2}}.$$

A lightcone argument shows that $\Omega(\log n)$ depth is necessary to prepare such a state, as is shown in [36].

In Section 2.1, we will review the Parity Halving Problem of [36], and explain how to derive a distributional version of the problem that can be exactly sampled by a shallow quantum circuit, but seemingly cannot be accurately sampled by a function of low locality. In Section 2.2, we will sketch a proof that this distribution has constant distance from every d-local distribution. That is, there is a distribution which exhibits a constant distance separation between classical and quantum sampling with shallow circuits. To boost this separation to an optimal one, we highlight and apply a direct product theorem implicit in [21] in Section 2.3.

2.1 Quantum Sampling and a Classical Reduction

As mentioned, the authors of [36] define the *Parity Halving Problem* (PHP): a relation problem over bit strings which is solvable by a shallow quantum circuit, but any randomized AC^0 (and therefore NC^0) circuit can only succeed on a trivial fraction of inputs. It is defined as follows:

▶ **Definition 4** (Parity Halving Problem)**.** *Given $x \in \{0,1\}^n$ with $|x| \equiv 0 \pmod 2$, return $y \in \{0,1\}^n$ which satisfies $|y| \equiv \frac{|x|}{2} \pmod 2$.*

Their initial observation is that the PHP can be solved with certainty on all instances by a QNC^0 circuit with polynomial size quantum advice, i.e., PHP is in the class $\mathsf{QNC}^0/\mathsf{qpoly}$. This circuit is shown on the left in Figure 1.

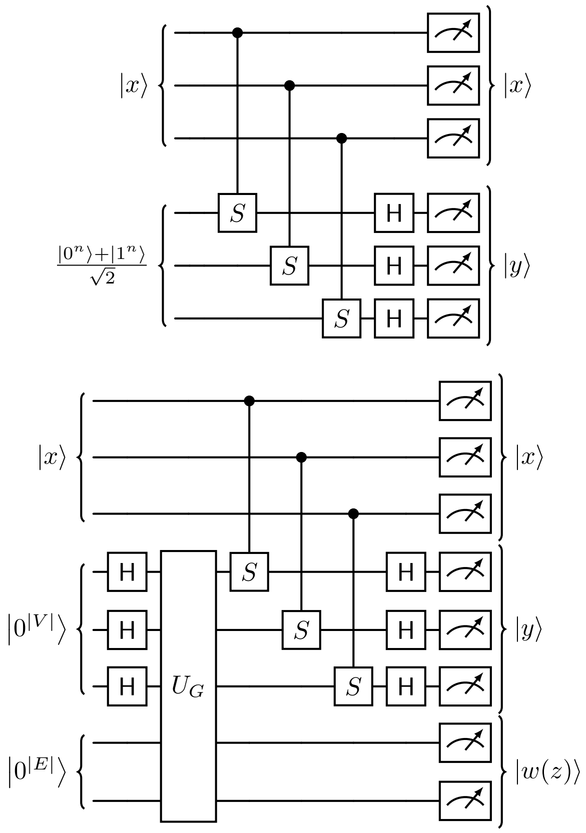

Figure 1 On the left is a $\mathsf{QNC}^0/\mathsf{qpoly}$ circuit which solves the Parity Halving Problem and on the right is a QNC^0 circuit which solves the Relaxed Parity Having Problem over a graph $G = (V, E)$. Here U_G is the $(|V| + |E|)$-qubit unitary which acts as $U_G |z\rangle |b\rangle = |z\rangle \bigotimes_{e=(u,v)\in E} |b_e \oplus z_u \oplus z_v\rangle$ for all $z \in \{0, 1\}^V$ and $b \in \{0, 1\}^E$.

To see that this circuit does indeed solve the PHP, note that after the CS gates are applied the resulting state is

$$|x\rangle \otimes \frac{|0^n\rangle + \mathrm{i}^{x_1+\cdots+x_n} |1^n\rangle}{\sqrt{2}} = |x\rangle \otimes \frac{|0^n\rangle + \mathrm{i}^{|x|} |1^n\rangle}{\sqrt{2}}$$
$$= \begin{cases} |x\rangle \otimes \frac{|0^n\rangle+|1^n\rangle}{\sqrt{2}} & \text{if } |x| \equiv 0 \pmod 4, \\ |x\rangle \otimes \frac{|0^n\rangle-|1^n\rangle}{\sqrt{2}} & \text{if } |x| \equiv 2 \pmod 4. \end{cases}$$

Finally, applying $\mathsf{H}^{\otimes n}$ to $\frac{|0^n\rangle+(-1)^b|1^n\rangle}{\sqrt{2}}$ yields a uniform superposition over bit strings of parity b.

In order to obtain a relational separation between NC^0 and QNC^0 without the need for quantum advice[1], the authors of [36] define a variant of the PHP as follows:

[1] Actually, the Relaxed Parity Halving Problem even serves to separate QNC^0 and AC^0, but we make mention of it here as it serves as motivation for our sampling separation.

▶ **Definition 5** (Relaxed Parity Halving Problem). *Fix a graph $G = (V, E)$. Given $x \in \{0,1\}^V$ with $|x| \equiv 0 \pmod 2$, return $y \in \{0,1\}^V$ and $w \in \{0,1\}^E$ for which there exists $z \in \{0,1\}^V$ such that*

$$z_u \oplus z_v = w_{(u,v)} \quad \forall (u,v) \in E \quad \textit{and} \quad |y| \equiv \langle z, x\rangle + \frac{|x|}{2} \pmod 2.$$

If G has a cycle then it may not be the case that for each $w \in \{0,1\}^E$ there exists a z such that w and z together satisfy the first condition above. However, when the underlying graph G is a tree then such a z exists for each w and preparing a "poor man's cat state" suffices to solve the Relaxed Parity Halving Problem over G. A poor man's cat state is a state proportional to $|z\rangle + |\overline{z}\rangle$ where $\overline{z}$ is the bitwise negation of z. The key observation is that there *is* a QNC^0 circuit which prepares a poor man's cat state, conditioned on the measurement outcome of another register. Indeed, the state $\frac{1}{\sqrt{2^{|V|-1}}} \sum_{z \in \{0,1\}^n, z_1 = 0} |w(z)\rangle \otimes \frac{|z\rangle + |\overline{z}\rangle}{\sqrt{2}}$ (where $w(z)$ and z satisfy the first constraint of Definition 5) can be prepared by a QNC^0 circuit so long as the maximum degree of G is constant. Finally, applying the PHP circuit, treating the Z register of the poor man's cat state as if it were the cat state, yields a uniformly random string of parity $\frac{|x|}{2} + \langle x, z\rangle$.

This circuit is shown on the right in Figure 1. In order to obtain lower bounds against NC^0 for the (R)PHP, standard locality arguments apply.

Recall our goal is to construct a *distributional* separation. A reasonable first attempt might be to consider the distribution $\mathcal{D}_{\mathrm{RPHP}}$ which is uniform over tuples (x, y, w) satisfying the relation. If generating this distribution is as hard as computing the RPHP relation, then classical hardness follows. Unfortunately, this is not the case. To gain some intuition, consider the classical PARITY function, which cannot be computed by shallow classical circuits [20, 29], and yet, a simple NC^0 circuit can sample from the distribution $(X, \mathsf{PARITY}(X))$ where X is a uniformly random bit string [3, 6]. Specifically, one can map the random bits

$$y_1, y_2, \ldots, y_n \to ((y_1 \oplus y_2) \circ (y_2 \oplus y_3) \circ \cdots \circ (y_{n-1} \oplus y_n), y_1 \oplus y_n),$$

where $\circ$ denotes concatenation [3, 6]. In fact, a similar construction classically samples from $\mathcal{D}_{\mathrm{RPHP}}$ (see [17, Subsection 5.2]).

We briefly digress to remark that this example is not an outlier. Indeed, the past decade or two has seen the study of sampling distributions blossom into a rich area, in many ways independent of computation, with exciting connections to fields such data structures [31, 24, 5, 35, 39, 21], extractors [30, 12, 32, 33, 4], pseudorandom generators [31, 24], and coding theory [26]. We refer the interested reader to the recent works [14, 35, 21, 39, 26, 22] and references within for more details.

To overcome the above barrier, consider the strings (x, y, w) subject to the constraint $|x| \equiv 1 \pmod 2$. The simple-but-key observation is that on input x with odd Hamming weight, the quantum circuit shown on the left in Figure 1 yields a uniformly random bit string y. (Note that w is always uniformly random.) Hence, if we replace $|x\rangle$ with some other state, we can now run our quantum circuit without necessarily invoking the promise on the Hamming weight of x, which gives us some added flexibility in our choice of distribution. In fact, we will simply replace each qubit with the single-qubit state $\sqrt{3/4}\,|0\rangle + \sqrt{1/4}\,|1\rangle$. That is, if we were to measure the qubits of the x-register, we would obtain the $(1/4)$-biased distribution for each bit. It is exactly this distribution of inputs for which we can show classical hardness. In the following subsection, we will highlight exactly *why* this distribution does not suffer from the same shortcoming as the PARITY example.

Formally, the distribution witnessing the separation is defined as follows:

▶ **Definition 6** (The $\mathcal{D}_{\mathsf{host}}(\mathcal{T})$ Distribution). *Let $\mathcal{T} = (V, E)$ be a tree with undirected edges. A sample $(X, Y, W) \sim \mathcal{D}_{\mathsf{host}}(\mathcal{T})$ is drawn as follows: first sample $X \sim \mathcal{U}^V_{1/4}$ and $Z \sim \mathcal{U}^V_{1/2}$. Define $W \in \{0,1\}^E$ by setting $W_e = Z_u \oplus Z_v$ for each $e = \{u, v\} \in E$. If X has odd Hamming weight, then sample $Y \in \{0,1\}^V$ uniformly at random; otherwise, sample Y as a uniform $|V|$-bit string of parity $\langle Z, X\rangle + |X|/2 \pmod 2$.*

In [17, Section 4], we show how to sample $\mathcal{D}_{\mathsf{host}}(\mathcal{T})$ exactly using QNC^0 circuits with the help of ancilla. The circuit is obtained by slightly modifying the construction given in [36] for the RPHP:

▶ **Proposition 7.** *Let $\mathcal{T} = (V, E)$ be a tree and let $\Delta \geq 2$ be its maximum vertex degree. Then there exists a geometrically local quantum circuit C such that the following holds.*

- *C has depth $2\Delta + 1$ and only uses Hadamard, controlled-phase, CNOT, and Toffoli gates. Moreover, Hadamard gates are only applied in the first and last layers.*
- *Let $\mathcal{P}$ be the distribution obtained by measuring $C\left|0^{5|V|-1}\right\rangle$ in the computational basis. Then the marginal distribution of the first $3|V| - 1$ coordinates of $\mathcal{P}$ is exactly $\mathcal{D}_{\mathsf{host}}(\mathcal{T})$.*

We refer to the target distribution as $\mathcal{D}_{\mathsf{host}}$ because it essentially "hosts" the following distribution, $\mathcal{D}_{\mathsf{hard}}(n, m)$, defined below:

▶ **Definition 8** (The $\mathcal{D}_{\mathsf{hard}}(n, m)$ Distribution). *A sample $(x, y) \sim \mathcal{D}_{\mathsf{hard}}(n, m)$ is drawn as follows: first sample $x \sim \mathcal{U}^n_{1/4}$ according to the $(1/4)$-biased product distribution. If x has odd Hamming weight, then sample $y \in \{0,1\}^m$ uniformly at random; otherwise x has even Hamming weight, and sample y as a uniform m-bit string of parity $|x|/2 \pmod 2$.*

$\mathcal{D}_{\mathsf{hard}}$ is the distribution which we are able to prove classical hardness for in a more straightforward way. Observe that the relationship between $\mathcal{D}_{\mathsf{hard}}$ and $\mathcal{D}_{\mathsf{host}}$ is analogous to that between the PHP and RPHP. The reduction from $\mathcal{D}_{\mathsf{host}}$ to $\mathcal{D}_{\mathsf{hard}}$ is given in [17, Subsection 5.3].

▶ **Lemma 9.** *Let $\mathcal{T} = (V, E)$ be a tree and let $v^* \in V$ be arbitrary. Define $K = \sum_{v \in V} |P_v|$, where P_v is the set of edges on the unique path between v^* and v. Then there exists a 5-local function $\mathsf{red}\colon \{0,1\}^{3|V|-1} \times \{0,1\}^* \to \{0,1\}^{2|V|+K}$ such that*

$$\mathsf{red}\left(\mathcal{D}_{\mathsf{host}}(\mathcal{T}), \mathcal{U}^*_{1/2}\right) = \mathcal{D}_{\mathsf{hard}}(|V|, |V| + K).$$

2.2 Classical Hardness

The classical lower bound of Theorem 2 is largely derived from the following hardness result. Let $\mathrm{tow}(x)$ denote the tower of 2's of height x (e.g., $\mathrm{tow}(3) = 2^{2^2}$).

▶ **Theorem 10.** *Let $d \geq 1$ be an integer. Assume $n \geq \mathrm{tow}(30d)$ and $m \leq n^2/\mathrm{tow}(30d)$. Then any d-local distribution has total variation distance at least 0.24 from $\mathcal{D}_{\mathsf{hard}}(n, m)$.*

Combining Theorem 10 with Lemma 9 easily gives the following corollary.

▶ **Corollary 11.** *Let $\mathcal{T} = (V, E)$ be a tree and let $v^* \in V$ be arbitrary. Define $K = \sum_{v \in V} |P_v|$, where P_v is the set of edges on the unique path between v^* and v. Additionally, let $d \geq 1$ be an integer, and assume $|V| \geq \mathrm{tow}(30(d + 5))$ and $|V| + K \leq |V|^2/\mathrm{tow}(30(d + 5))$. Then any d-local distribution has total variation distance at least 0.24 from $\mathcal{D}_{\mathsf{host}}(\mathcal{T})$.*

Proof. Assume by contradiction there exists a d-local function $f\colon \{0,1\}^* \to \{0,1\}^{3|V|-1}$ and a product distribution Π over $\{0,1\}^*$ such that the distribution of f applied to samples drawn from Π, denoted $f(\Pi)$, is δ-close to $\mathcal{D}_{\mathsf{host}}(\mathcal{T})$ for some $\delta < 0.24$. Define a new function $g\colon \{0,1\}^* \to \{0,1\}^{2|V|+K}$ by

$$g(\Pi) = \mathsf{red}(f(\Pi), \{0,1\}^*),$$

where red is defined as in Lemma 9. Then g is $(d+5)$-local by Lemma 9 and δ-close to $\mathcal{D}_{\mathsf{hard}}(|V|, |V|+K)$ by the data processing inequality. This contradicts Theorem 10. ◀

Before sketching the main ideas behind the proof of Theorem 10, a few remarks are in order. First, a tighter analysis can yield distance $\frac{1}{4} - \varepsilon$, assuming $m \leq O_\varepsilon(n^2/\mathrm{tow}(30d))$; this is near optimal, as the 2-local[2] distribution $\mathcal{U}_{1/4}^n \times \mathcal{U}_{1/2}^m$ achieves distance $\frac{1}{4} - o(1)$. Second, the quadratic upper bound on m in Theorem 10 is necessary; we show $\mathcal{D}_{\mathsf{hard}}(n,m)$ is $O(1)$-local when $m \geq \Omega(n^2)$ in [17, Proposition 5.8]. Finally, it is necessary that $x \sim \mathcal{U}_{1/4}^n$ and not $x \sim \mathcal{U}_{1/2}^n$, as the latter *can* be exactly sampled (see [17, Proposition 5.9]), though any bias other than $0, 1$, or $1/2$ will be hard.

Let us now provide an overview of Theorem 10's proof. Fix an arbitrary d-local function $f\colon \{0,1\}^* \to \{0,1\}^{n+m}$ and an arbitrary product distribution Π over $\{0,1\}^*$ as input. Our goal is to show that the distribution $f(\Pi)$ is 0.24-far from $\mathcal{D}_{\mathsf{hard}}(n,m)$. One immediate challenge in working with d-local functions is that the locality constraint is "one-sided." Even though no output bit is influenced by many input bits, there may exist an input bit that affects every single output bit. The resulting output distribution, then, can have complicated correlations, which muddle the analysis.

The Structured Case: A First Attempt

To warm-up, we first consider the idealized setting where there are r "non-connected" output bits, by which we mean no two such output bits depend on a common input bit. In particular, the r marginal distributions of $f(\Pi)$ projected onto the individual coordinates are independent. Here, one should view r as large, say $\Omega_d(n)$. We proceed via a concentration vs. anticoncentration dichotomy, present in various forms in the works [31, 14, 35, 21, 22]. Specifically, we classify each of the r output bits according to how their corresponding marginal distribution compares to the marginal distribution of the target distribution.

At a high level, we would like to argue that either many of these output bits have marginal distributions which are far from those of the target distribution, in which case we can combine the marginal errors, or many of these output bits are close to the "correct" marginal distribution, in which case a more complicated anticoncentration argument shows that a specific potential function highlights a noticeable discrepancy between the two distributions. To this end, we call an output bit b Type-1 if the marginal distribution $f(\Pi)|_b$ is δ-far in total variation distance from the marginal distribution $\mathcal{D}_{\mathsf{hard}}(n,m)|_b$, and call it Type-2 otherwise. Here, $\delta = O_d(1)$ is some small threshold parameter.

---

[2] The distribution is 2-local if the input bits are unbiased coins. When we allow input bits with mixed bias of $1/4$ and $1/2$, the distribution is 1-local.

Suppose at least $r/2$ of the non-connected output bits are $\mathsf{Type\text{-}1}$. Note that since total variation distance is closed under projection, a single Type-1 neighborhood already gives distance δ. To strengthen the bound, one can take advantage of independence and apply standard concentration inequalities, as in the proof of [21, Lemma 4.2], to conclude $f(\Pi)$ has distance roughly $1 - e^{-\delta^2 \cdot r}$ from $\mathcal{D}_{\mathsf{hard}}(n, m)$.[3] For $r \gg 1/\delta^2$, this is at least 0.24, as desired.

The more involved case is when at least $r/2$ output bits are $\mathsf{Type\text{-}2}$. Here, rather than directly comparing $f(\Pi)$ to $\mathcal{D}_{\mathsf{hard}}(n, m)$, we compare the expectation of a complex-valued potential function $h(x, y) = \mathrm{i}^{|x|+2|y|}$ over samples (x, y) drawn from the two distributions. Direct calculations show that $\mathbb{E}_{x,y}[h(x, y)] \approx 1/2$ for $(x, y) \sim \mathcal{D}_{\mathsf{hard}}(n, m)$ (see [17, Claim 5.3]) and that $|\mathbb{E}[\mathrm{i}^A]|$ is bounded away from 1 for any integral random variable A suitably far from constant modulo 4 (see [17, Claim 5.5]). It is tempting to argue that by the independence of the non-connected output bits, we can fix the value of the input bits not affecting any $\mathsf{Type\text{-}2}$ output bits to view $\mathbb{E}_{x,y}[h(x, y)]$ as a product of many independent random variables with magnitudes bounded away from 1. Then we could conclude that for each of these input conditionings, $|\mathbb{E}_{x,y}[h(x, y)]| \ll 0.01$ for $(x, y) \sim f(\Pi)$, which would give the desired distance of 0.24 (using [17, Lemma 3.4]).

The problem, however, is that the contributions of the remaining output bits can compensate for those of the non-connected output bits. For example, consider the string $z_1, 1 - z_1, z_2, 1 - z_2, \ldots, z_k, 1 - z_k$, where the z_k's are independent random bits. There are k independent bits, yet the string's Hamming weight is fixed at k. Thus, we cannot reason about $\mathbb{E}_{x,y}[h(x, y)]$ solely from the non-connected output bits. Instead, we need to consider the *neighborhood* of each output bit b, i.e., the set of output bits that are also influenced by the input bits determining b.

The Structured Case: Refining the Output Structure

To fix our analysis, let us change our assumption from there being r non-connected output bits to there being r non-connected neighborhoods. Here, we refer to two neighborhoods N_1, N_2 as non-connected if for every pair of output bits $b_1 \in N_1$ and $b_2 \in N_2$, the input bits that determine b_1 are disjoint from those that determine b_2. We can similarly classify each neighborhood as $\mathsf{Type\text{-}1}$ or $\mathsf{Type\text{-}2}$ depending on the distance of its marginal distribution to that of the target distribution. The analysis in the case of many $\mathsf{Type\text{-}1}$ neighborhoods is performed almost identically to the previous scenario, but now we are able to reason more carefully when there are many $\mathsf{Type\text{-}2}$ neighborhoods.

Indeed, consider a $\mathsf{Type\text{-}2}$ neighborhood $N = (x', y')$, where x' is the output bits contained in the first n indices (corresponding to x) and y' is the output bits contained in the latter m indices (corresponding to y). If we can show that $|x'| + 2|y'| \pmod 4$ is not too close to being a constant, then the potential function argument sketched above can actually be carried out. To this end, let b be the output bit that defines the neighborhood $N = N(b)$, and consider the effect of conditioning on $b = 0$ vs. on $b = 1$.

First suppose b is in the x part. In this case, we can write $x' = (b, x'')$ and express $|x'| + 2|y'| \pmod 4$ as $b + |x''| + 2|y'| \pmod 4$. Recall that N is a $\mathsf{Type\text{-}2}$ neighborhood, so it should resemble a product distribution. In particular, the distribution of $|x''| + 2|y'| \pmod 4$ conditioned on $b = 0$ should be roughly the same as when conditioned on $b = 1$. Observe then, that $1 + |x''| + 2|y'| \pmod 4$ and $|x''| + 2|y'| \pmod 4$ should have noticeably different

[3] There is a small subtlety here, in that if the set of $\mathsf{Type\text{-}1}$ output bits fully contains the last m output bits, then those output bits are not a product distribution in $\mathcal{D}_{\mathsf{hard}}(n, m)$. For simplicity, we will assume that this does not occur, although the full statement of [21, Lemma 4.2] is robust enough to still apply in that scenario.

distributions, as we are essentially comparing a binomial distribution with its shift. Since b should be close to (1/4)-biased, it takes both values with constant probability, so $|x'| + 2|y'| \pmod 4$ cannot be too close to any fixed value. A similar analysis shows that if b is in the y part, then we are comparing the density of $|x'|+2|y'| \pmod 4$ and $|x'|+2(|y'|+1) \pmod 4$.

Unfortunately, there is a problem with this latter case. Suppose the neighborhood N does not contain any bits in the x part. Then we are comparing the density of $2|y'| \pmod 4$ and $2(|y'| + 1) \pmod 4$, or equivalently, $|y'| \pmod 2$ and $|y'| + 1 \pmod 2$. The y part is (1/2)-biased, so $|y'| \pmod 2$ *can* have the same distribution as $|y'| + 1 \pmod 2$! Note that it is this fact which allows for the previously described sampling algorithm for (X, PARITY). Hence, we must make one further refinement to our assumption.

The Structured Case: A Final Adjustment

Now instead of simply assuming there are r non-connected neighborhoods, we insist that all r neighborhoods are generated by output bits in the x part. Moreover, we will only require the non-connectedness property on bits in the x part of the neighborhoods. This second condition actually makes the analysis more challenging, but we will later see it is necessary for this model case to be obtainable. We once more redefine Type-1 and Type-2 neighborhoods; this time we classify neighborhoods based only on their marginals on the first n output bits. The case of many Type-1 neighborhoods essentially works as before (see [17, Lemma 5.2]), so it remains to address the case where at least $r/2$ of the neighborhoods are Type-2.

To obtain some structure in the y part, we exploit our assumption on the size of m. Since we have $m \leq n^2/\mathrm{tow}(30d)$, most pairs of neighborhoods do not intersect in the last m output bits. Quantitatively, we can find $C \approx r^2/(md^2) \gg 1$ non-connected Type-2 neighborhoods that do not intersect in the y part. Without loss of generality, assume they are $N(1), N(2), \ldots, N(C)$. By fixing the value of all the input bits that do not affect $1, 2, \ldots, C$, the contributions to h from these neighborhoods are now independent. In particular, the expectation of h becomes a product of expectations over the output of the neighborhood. As noted above, we can conclude the expectation over the neighborhood $N = (x', y')$ is bounded away from 1 if $|x'| + 2|y'| \pmod 4$ is not too close to any fixed value.

This ends up being a bit difficult to show directly, since while the C neighborhoods are disjoint in the y part, they may be connected. Fortunately, the variance of $|x'|+2|y'| \pmod 4$ over a random such fixing of the input bits follows from that of $|x'| \pmod 2$, where we do have non-connectedness in the x part. By the previous argument of considering $|x'|$ conditioned on the output bit b being 0 vs. being 1, we are able to prove $|x'| \pmod 2$ is typically not too close to constant (see [17, Claim 5.7]). This concludes the analysis of many Type-2 neighborhoods (see [17, Lemma 5.6]), as well as the proof of Theorem 2 under certain ideal assumptions.

Reduction to the Structured Case

Previously, we assumed a rather strong structure: $r = \Omega_d(n)$ many output bits generating neighborhoods that are non-connected in $[n]$. This, of course, is not a structure readily present in an arbitrary d-local function f. To reduce to this case, a standard approach (appearing in, e.g., [31, 8, 34, 14, 35, 21, 22]) is to strategically condition on bits to express an arbitrary d-local function as a convex combination of functions with the desired structure. In other words, if we can find some set S of input bits whose removal induces many non-connected neighborhoods of the form we want, then we can express $f(\Pi)$ as

$$f(\Pi) = \mathop{\mathbb{E}}_{\rho \in \{0,1\}^S} [f_\rho(\Pi)],$$

where $f_\rho\colon \{0,1\}^* \to \{0,1\}^{n+m}$ is defined as f with the input bits in S fixed to their values in ρ. Observe that each f_ρ has the structured form we already know how to analyze, regardless of the actual values the bits in S are set to. More formally, we have:

1. If most of the non-connected neighborhoods are Type-1, then $f_\rho(\Pi)$ is $\approx (1 - e^{-\Omega_d(r)})$-far from $\mathcal{D}_{\mathsf{hard}}(n,m)$, and
2. Otherwise, $\mathbb{E}_{(x,y)\sim\mathcal{D}_{\mathsf{hard}}(n,m)}[h(x,y)] - \mathbb{E}_{(x,y)\sim f_\rho(\Pi)}[h(x,y)] \geq 0.49$.

By a union bound argument (see [17, Lemma 3.4]), these results on the conditioned functions can be combined to obtain $\|f(\Pi) - \mathcal{D}_{\mathsf{hard}}(n,m)\|_{\mathsf{TV}} \gtrsim 0.245 - 2^{|S|} \cdot e^{-\Omega_d(r)}$. Then as long as $r \gg |S|$, we obtain the desired distance bound of 0.24.

At this point, the remaining task is combinatorial. We construct a bipartite graph whose left side is the first n output bits, and whose right side is the input bits. Note that each left vertex has maximum degree d. We want to remove s right vertices to obtain r non-connected neighborhoods of the prescribed form, where $r \gg s$. Ideally, we would like r to be as large as possible to maximize the total variation distance. Fortunately, this task has already been done for us. By [21, Corollary 4.11], we can take $s \ll r$ and $r = \Omega_d(n)$, as desired.

For the sake of completeness, we briefly highlight the main idea behind the proof of [21, Corollary 4.11]. The key observation is that locality, while only explicitly constraining the left vertices, also constrains the right ones, since it upper bounds the number of edges by dn. Thus while we cannot forbid high-degree right vertices, there cannot be many of them. This implies that we can "affordably" remove all right vertices above a particular degree threshold, and greedily find non-connected vertices on the left side. A more involved analysis (see [21, Corollary 4.8]) provides better parameters than one could obtain via this naive approach, and an even more involved analysis guarantees non-connected left neighborhoods, rather than just vertices. Still, the proofs morally operate in a similar way to the strategy described. This completes the sketch of the proof of Theorem 2; the full details can be found in [17, Section 5].

We conclude by remarking that the above analysis is fairly robust, and it allows one to rule out the sampleability of a number of simple distributions by shallow circuits. Thus, the specific distributions we have chosen to consider are primarily a function of what can be produced by shallow quantum circuits, rather than what can be forbidden for shallow classical ones.

2.3 Boosting the Separation

Combining our results thus far produces a separation with constant total variation distance. In order to prove the stronger separation in Theorem 2, we consider the distribution $\mathcal{D}_{\mathsf{host}}(\mathcal{T})^k = \mathcal{D}_{\mathsf{host}}(\mathcal{T}) \times \cdots \times \mathcal{D}_{\mathsf{host}}(\mathcal{T})$. Certainly, if our quantum circuit can generate $\mathcal{D}_{\mathsf{host}}(\mathcal{T})$, then it can also generate $\mathcal{D}_{\mathsf{host}}(\mathcal{T})^k$. Moreover, we can apply the following direct product theorem implicit in [21] (and formalized in [17, Subsection 5.4]) to show the overlap of the target distribution with that produced by classical circuits decays exponentially quickly.

▶ **Theorem 12** (Direct Product Theorem). *Let $d, \ell \geq 1$ be integers, and let $\mathcal{D}$ be a distribution over $\{0,1\}^\ell$. Suppose that for any d-local function $f\colon \{0,1\}^* \to \{0,1\}^\ell$ and binary product distribution Π on $\{0,1\}^*$, we have*

$$\|f(\Pi) - \mathcal{D}\|_{\mathsf{TV}} \geq \delta.$$

Then for any integer $k \geq 1$, d-local function $g\colon \{0,1\}^* \to \{0,1\}^{\ell k}$, and binary product distribution Ξ on $\{0,1\}^*$, we have

$$\left\|g(\Xi) - \mathcal{D}^k\right\|_{\mathsf{TV}} \geq 1 - 4\exp\left\{-\left(\frac{\delta^2}{16d\ell}\right)^{4d\ell} \cdot k\right\}.$$

The proof of Theorem 12, much like the proof of Theorem 2, uses a graph elimination result derived in [21]. In this context, such a result allows one to find many independent groups of output bits corresponding to instances of $\mathcal{D}$. Since the marginal distributions of $\mathcal{D}$ and $f(\Pi)$ disagree on each group, we can use a standard concentration inequality to derive a strong error bound.

Proof of Theorem 2. Let $n = \mathrm{tow}(40d)$, and let $\mathcal{T} = (V, E)$ be the spanning tree of the $\sqrt{n} \times \sqrt{n}$ square grid obtained by including all the edges in the first row, as well as all the edges in each column. Observe that the diameter of $\mathcal{T}$ is $3\sqrt{n}$, so K (as defined in Corollary 11) is at most $3n^{3/2}$. In particular, our choice of n guarantees $|V| + K \leq |V|^2/\mathrm{tow}(30(d+5))$. Thus, Corollary 11 implies any d-local distribution at least 0.24-far from $\mathcal{D}_{\mathsf{host}}(\mathcal{T})$. Applying Theorem 12, we can boost this error to conclude that any d-local distribution has distance from $\mathcal{D}_{\mathsf{host}}(\mathcal{T})^k$ at least

$$1 - 4\exp\left\{-\left(\frac{0.24^2}{16d(3n-1)}\right)^{4d(3n-1)} \cdot k\right\}.$$

Let $c_d > 0$ be a sufficiently large constant depending only on d. For any integer $N \geq c_d$, express $N = k\cdot(3n-1)+r$ with $0 \leq r < 3n-1$, and define the distribution $\mathcal{D}_N = \mathcal{D}_{\mathsf{host}}(\mathcal{T})^k \times 0^r$. Since total variation distance is closed under projection, we find that any d-local distribution has distance from $\mathcal{D}_N$ at least

$$1 - 4\exp\left\{-\left(\frac{0.24^2}{16d(3n-1)}\right)^{4d(3n-1)} \cdot \frac{N-r}{3n-1}\right\} \geq 1 - e^{-N/c_d}$$

for large enough c_d.

We conclude by noting that Proposition 7 gives a depth-7 quantum circuit that exactly samples $\mathcal{D}_{\mathsf{host}}(\mathcal{T})^k$ on input $\left|0^{k(5n-1)}\right\rangle$ by considering the marginal distribution on $k(3n-1)$ specific coordinates. By padding with r extra zeros, a similar circuit on $k(5n-1) + r \leq 2N$ inputs can also sample $\mathcal{D}_N$. ◀

▶ Remark 13. Our setting of $\mathcal{T}$ is motivated by common topological choices in implementations. One could alternatively set $\mathcal{T}$ to be a balanced binary tree to minimize K, but this would not affect the final bound.

References

1. Scott Aaronson and Alex Arkhipov. The computational complexity of linear optics. *Theory Of Computing*, 9(4):143–252, 2013. doi:10.4086/TOC.2013.V009A004.
2. Dorit Aharonov. A simple proof that Toffoli and Hadamard are quantum universal. *arXiv preprint quant-ph/0301040*, 2003.
3. Lászió Babai. Random oracles separate PSPACE from the polynomial-time hierarchy. *Information Processing Letters*, 26(1):51–53, 1987. doi:10.1016/0020-0190(87)90036-6.
4. Marshall Ball, Eli Goldin, Dana Dachman-Soled, and Saachi Mutreja. Extracting randomness from samplable distributions, revisited. In *2023 IEEE 64th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1505–1514. IEEE, 2023. doi:10.1109/FOCS57990.2023.00092.
5. Chris Beck, Russell Impagliazzo, and Shachar Lovett. Large deviation bounds for decision trees and sampling lower bounds for AC0-circuits. In *2012 IEEE 53rd Annual Symposium on Foundations of Computer Science*, pages 101–110. IEEE, 2012. doi:10.1109/FOCS.2012.82.

6 Ravi B Boppana and Jeffrey C Lagarias. One-way functions and circuit complexity. *Information and Computation*, 74(3):226–240, 1987. `doi:10.1016/0890-5401(87)90022-8`.

7 Adam Bouland, Bill Fefferman, Chinmay Nirkhe, and Umesh Vazirani. On the complexity and verification of quantum random circuit sampling. *Nature Physics*, 15(2):159–163, 2019.

8 Sergey Bravyi, David Gosset, and Robert König. Quantum advantage with shallow circuits. *Science*, 362(6412):308–311, 2018.

9 Sergey Bravyi, David Gosset, Robert König, and Marco Tomamichel. Quantum advantage with noisy shallow circuits. *Nature Physics*, 16(10):1040–1045, 2020.

10 Libor Caha, Xavier Coiteux-Roy, and Robert König. A colossal advantage: 3D-local noisy shallow quantum circuits defeat unbounded fan-in classical circuits. *arXiv preprint arXiv:2312.09209*, 2023.

11 Eshan Chattopadhyay, Jesse Goodman, and David Zuckerman. The space complexity of sampling. In *13th Innovations in Theoretical Computer Science Conference,(ITCS 2022)*, 2022.

12 Anindya De and Thomas Watson. Extractors and lower bounds for locally samplable sources. *ACM Transactions on Computation Theory (TOCT)*, 4(1):1–21, 2012. `doi:10.1145/2141938.2141941`.

13 Yu-Hao Deng, Yi-Chao Gu, Hua-Liang Liu, Si-Qiu Gong, Hao Su, Zhi-Jiong Zhang, Hao-Yang Tang, Meng-Hao Jia, Jia-Min Xu, Ming-Cheng Chen, et al. Gaussian boson sampling with pseudo-photon-number-resolving detectors and quantum computational advantage. *Physical review letters*, 131(15):150601, 2023.

14 Yuval Filmus, Itai Leigh, Artur Riazanov, and Dmitry Sokolov. Sampling and certifying symmetric functions. In *Approximation, Randomization, and Combinatorial Optimization. (APPROX/RANDOM)*, 2023.

15 Sabee Grewal and Vinayak M Kumar. Improved circuit lower bounds and quantum-classical separations. *arXiv preprint arXiv:2408.16406*, 2024.

16 Daniel Grier, Nathan Ju, and Luke Schaeffer. Interactive quantum advantage with noisy, shallow Clifford circuits. *arXiv preprint arXiv:2102.06833*, 2021. `arXiv:2102.06833`.

17 Daniel Grier, Daniel M. Kane, Jackson Morris, Anthony Ostuni, and Kewen Wu. Quantum advantage from sampling shallow circuits: Beyond hardness of marginals. *arXiv preprint arXiv:2510.07808*, 2025. `doi:10.48550/arXiv.2510.07808`.

18 Daniel Grier and Luke Schaeffer. Interactive shallow Clifford circuits: quantum advantage against NC^1 and beyond. In *Proceedings of the 52nd Annual ACM Symposium on Theory of Computing*, pages 875–888, 2020. `doi:10.1145/3357713.3384332`.

19 Atsuya Hasegawa and François Le Gall. Quantum Advantage with Shallow Circuits Under Arbitrary Corruption. In *32nd International Symposium on Algorithms and Computation (ISAAC 2021)*, volume 212 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 74:1–74:16. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2021. `doi:10.4230/LIPIcs.ISAAC.2021.74`.

20 Johan Håstad. *Computational limitations for small depth circuits*. PhD thesis, Massachusetts Institute of Technology, 1986.

21 Daniel M Kane, Anthony Ostuni, and Kewen Wu. Locality bounds for sampling Hamming slices. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing*, pages 1279–1286, 2024. `doi:10.1145/3618260.3649670`.

22 Daniel M Kane, Anthony Ostuni, and Kewen Wu. Locally sampleable uniform symmetric distributions. In *Proceedings of the 57th Annual ACM Symposium on Theory of Computing*, pages 1807–1816, 2025. `doi:10.1145/3717823.3718243`.

23 Greg Kuperberg. Breaking the cubic barrier in the Solovay-Kitaev algorithm. *arXiv preprint arXiv:2306.13158*, 2023. `doi:10.48550/arXiv.2306.13158`.

24 Shachar Lovett and Emanuele Viola. Bounded-depth circuits cannot sample good codes. In *2011 IEEE 26th Annual Conference on Computational Complexity*, pages 243–251. IEEE, 2011. `doi:10.1109/CCC.2011.11`.

25 Gregory Rosenthal. Bounds on the QAC^0 complexity of approximating parity. In *12th Innovations in Theoretical Computer Science Conference (ITCS 2021)*, Leibniz International Proceedings in Informatics (LIPIcs), pages 32:1–32:20. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2021. doi:10.4230/LIPIcs.ITCS.2021.32.

26 Ronen Shaltiel and Jad Silbak. Explicit codes for poly-size circuits and functions that are hard to sample on low entropy distributions. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing*, pages 2028–2038, 2024. doi:10.1145/3618260.3649735.

27 Yaoyun Shi. Quantum and classical tradeoffs. *Theoretical computer science*, 344(2-3):335–345, 2005. doi:10.1016/J.TCS.2005.03.053.

28 Peter W Shor. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM review*, 41(2):303–332, 1999. doi:10.1137/S0036144598347011.

29 Roman 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*, pages 77–82, 1987. doi:10.1145/28395.28404.

30 Luca Trevisan and Salil Vadhan. Extracting randomness from samplable distributions. In *Proceedings 41st Annual Symposium on Foundations of Computer Science*, pages 32–42. IEEE, 2000. doi:10.1109/SFCS.2000.892063.

31 Emanuele Viola. The complexity of distributions. *SIAM Journal on Computing*, 41(1):191–218, 2012. doi:10.1137/100814998.

32 Emanuele Viola. Extractors for Turing-machine sources. In *International Workshop on Approximation Algorithms for Combinatorial Optimization*, pages 663–671. Springer, 2012. doi:10.1007/978-3-642-32512-0_56.

33 Emanuele Viola. Extractors for circuit sources. *SIAM Journal on Computing*, 43(2):655–672, 2014. doi:10.1137/11085983X.

34 Emanuele Viola. Sampling lower bounds: boolean average-case and permutations. *SIAM Journal on Computing*, 49(1):119–137, 2020. doi:10.1137/18M1198405.

35 Emanuele Viola. New sampling lower bounds via the separator. In *38th Computational Complexity Conference (CCC 2023)*, Leibniz International Proceedings in Informatics (LIPIcs), pages 26:1–26:23. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2023. doi:10.4230/LIPIcs.CCC.2023.26.

36 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 Symposium on Theory of Computing*, pages 515–526, 2019. doi:10.1145/3313276.3316404.

37 Adam Bene Watts and Natalie Parham. Unconditional quantum advantage for sampling with shallow circuits. *arXiv preprint arXiv:2301.00995*, 2023. doi:10.48550/arXiv.2301.00995.

38 Aaron W Young, Shawn Geller, William J Eckner, Nathan Schine, Scott Glancy, Emanuel Knill, and Adam M Kaufman. An atomic boson sampler. *Nature*, 629(8011):311–316, 2024.

39 Huacheng Yu and Wei Zhan. Sampling, flowers and communication. In *15th Innovations in Theoretical Computer Science Conference (ITCS 2024)*, Leibniz International Proceedings in Informatics (LIPIcs), pages 100:1–100:11. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2024. doi:10.4230/LIPIcs.ITCS.2024.100.

40 Han-Sen Zhong, Hui Wang, Yu-Hao Deng, Ming-Cheng Chen, Li-Chao Peng, Yi-Han Luo, Jian Qin, Dian Wu, Xing Ding, Yi Hu, et al. Quantum computational advantage using photons. *Science*, 370(6523):1460–1463, 2020.