nature communications

Article

https://doi.org/10.1038/s41467-025-58545-4

Unconditional advantage of noisy qudit quantum circuits over biased threshold circuits in constant depth

Received: 15 November 2024

Accepted: 26 March 2025

Published online: 15 April 2025

Check for updates

Michael de Oliveira [1,2,3,4] ✉, Sathyawageeswar Subramanian [5] ✉, Leandro Mendes[1] & Min-Hsiu Hsieh[1] ✉

The rapid evolution of quantum devices fuels concerted efforts to experimentally establish quantum advantage over classical computing. Many demonstrations of quantum advantage, however, rely on computational assumptions and face verification challenges. Furthermore, steady advances in classical algorithms and machine learning make the issue of provable, practically demonstrable quantum advantage a moving target. In this work, we unconditionally demonstrate that parallel quantum computation can exhibit greater computational power than previously recognized. We prove that polynomial-size biased threshold circuits of constant depth—which model neural networks with tunable expressivity—fail to solve certain problems solvable by small constant-depth quantum circuits with local gates, for values of the bias that allow quantifiably large computational power. Additionally, we identify a family of problems that are solvable in constant depth by a universal quantum computer over prime-dimensional qudits with bounded connectivity, but remain hard for polynomial-size biased threshold circuits. We thereby bridge the foundational theory of non-local games in higher dimensions with computational advantage on emerging devices operating on a wide range of physical platforms. Finally, we show that these quantum advantages are robust to noise across all prime qudit dimensions with all-to-all connectivity, enhancing their practical appeal.

Quantum technologies, particularly quantum computing, have recently made significant progress. This includes a continuous increase in the number of qubits/qudits on quantum processors[1], notable reductions in error rates for native operations[2,3], and extended coherence times[4] across various hardware platforms. Additionally, these technical advances have culminated in several notable breakthroughs in error-correction experiments, marking significant progress toward early fault-tolerant quantum computation[5–7]. However, achieving the theoretical computational advantages promised by landmark quantum algorithms, such as integer factorization or search, remains limited by the substantial resource requirements, which current quantum hardware is still far from meeting[8–10].

This state of affairs raises the question of understanding whether the evolving small and faulty quantum devices could still support quantum advantages that are realizable with low resource requirements, over comparable classical devices (or computational models). In particular, a series of pioneering studies showed that under plausible complexity theoretic assumptions—e.g., that the polynomial hierarchy does not collapse to the third level—certain classes of quantum circuits are exponentially hard to simulate classically while

[1]Hon Hai (Foxconn) Quantum Computing Research Center, Taipei, Taiwan, ROC. [2]International Iberian Nanotechnology Laboratory, Braga, Portugal. [3]LIP6, Sorbonne Université, Paris, France. [4]INESC TEC, Porto, Portugal. [5]Department of Computer Science and Technology, University of Cambridge, Cambridge, UK. ✉e-mail: michaeldeoliveira848@gmail.com; sathynius2@gmail.com; minhsiuh@gmail.com

being more resource-efficient and hardware-friendly. Conditional hardness results of this kind include IQP circuits[11], Boson sampling experiments[12], and random circuit sampling[13]. On the other hand, the reliance on complexity-theoretic assumptions and the challenge of addressing noise in current quantum experiments reveal a significant gap between theoretical models and practical implementation, complicating the verification of correctness. Moreover, advancements in classical algorithms, simulation techniques, and artificial intelligence continue to escalate the race to achieve quantum advantage beyond classical capabilities.

Rather than comparing the capabilities of shallow-depth quantum circuits to general models of classical computation, recent research has shifted towards comparing them to their classical shallow-depth counterparts, highlighting the potential for unconditional quantum advantages on near-term hardware. The seminal work of Bravyi et al.[14] demonstrated a search problem that can be solved by constant-depth quantum circuits using only 2-qubit Clifford gates, but which no constant-depth classical circuit with bounded fan-in gates can solve, without relying on any complexity-theoretic assumptions. This first *unconditional* separation between shallow-depth quantum and classical circuits sparked renewed interest in the field. It underscored the potential for significant computational and practical advantages in *parallel* quantum computations. In both classical and quantum computing, circuit width is a good measure of parallelism (e.g., the number of processors) and depth is a good measure of runtime, with constant-depth circuits capturing the computations that can be performed by a polynomial number of processors running for a constant amount of time. Several extensions of Bravyi et al.'s result followed, enhancing the separation to average-case hardness[15], introducing noise resilience[16], demonstrating that the quantum advantage extends to larger constant-depth logical circuits with unbounded fan-in Boolean gates[17], and identifying problems of greater practical interest[18].

While this prior work has established that parallel quantum computations can show advantages over comparable parallel classical computations, these advantages have only been demonstrated for classical circuit classes with limited practical applicability. To broaden the scope of quantum advantage, it is thus essential to explore circuit classes beyond those previously considered. A prime candidate for this exploration is circuits that are allowed to use *threshold* gates, which output one if the Hamming weight of the input string meets or exceeds a threshold k, and zero otherwise, mirroring the behavior of a Heaviside step function. Constant-depth circuits comprising such threshold gates belong to the complexity class TC^0 [19]. A point of interest is that these circuits serve as a canonical theoretical model for vanilla neural networks[20–23], and have even been useful in obtaining mathematical results about the expressivity of transformer architectures that underlie large language models such as ChatGPT[24,25]. A pertinent question, which could provide foundational insights into parallel quantum computations and the computational potential of quantum machine learning models, is whether quantum circuits can outperform threshold circuits in constant depth.

Here, we extend the scope of provable quantum advantages in parallel computation to more advanced and potentially practical classical parallel models for all qudits of prime dimensions. In particular, our results focus on a classical circuit class that captures the computational power of *threshold* circuits in a well-defined and measurable way through a parameterized bias, allowing us to quantify the potential extent of quantum advantage over these models. By incorporating realistic noise models into shallow quantum circuits, we demonstrate that these quantum advantages are robust to noise, preserving their feasibility under practical conditions. Finally, we quantify the resource costs for demonstrating quantum advantage, presenting a hierarchy of quantum advantage experiments and their associated resource requirements.

Results

Background

In this paper, we consider classes of shallow-depth quantum circuits—circuits of constant depth independent of the input length, using a polynomial number of gates that have bounded fan-in (i.e., each gate has a fixed, constant number of input and output wires) and are drawn from a finite, qudit universal quantum gate set. Over qubits, this class is commonly referred to as QNC^0. Throughout this paper, we refer to qudit systems of prime dimension p as "qupits" (with the associated state space $\mathbb{C}^p$). Concurrently, we will compare shallow quantum circuits, as efficient and hardware-friendly as possible, with circuits in the $\mathsf{bPTC}^0(k)$ class described by constant depth classical circuits composed of unbounded fan-in gates that compute biased polynomial threshold functions (PTFs), which may be non-linear but are constrained by a bias parameter k[26]. A PTF with bias k is defined as follows,

$$f(x)=\begin{cases} P(x), & \sum_{i=1}^{n} x_i \le k \\ 1, & \sum_{i=1}^{n} x_i > k \end{cases}; , \tag{1}$$

with $P:\mathbb{F}_2^n \to \mathbb{F}_2$ being a polynomial over $\mathbb{F}_2=\{0,1\}$ and k restricting the maximum degree of P. We note that the common definition in the literature takes $P:\mathbb{R}_2^n \to \mathbb{R}$, and sets $f(x)=\frac{1}{2}(1+\mathrm{sgn}(P(x)))$. On the other hand here we are interested in polynomials of degree at most k over $\mathbb{F}_2^n$, with threshold behavior determined by the bias parameter k.

This parameter constrains their behavior to function as the Boolean OR when the input string has a Hamming weight of at least k. Conversely, the behavior can be inverted, in which case the bias parameter constrains it to function as the Boolean AND, applying to input strings with a Hamming weight of at most $n-k$. For a constant, non-zero k (i.e., $k=\mathcal{O}(1)$), this class corresponds to Boolean circuits with unbounded fan-in AND, OR, and NOT gates of constant depth, known as the circuit class AC^0[27]. Notably, AC^0 represents the largest classical circuit class for which unconditional quantum advantages have been established[17]. When $k=\Theta(n)$, it encompasses the strictly larger TC^0 circuit class.

For intermediate scalings, this class provides access to unbounded fan-in gates with biased yet non-trivial activation regions, see Fig. 1 and refer to SI section D2 for our decomposition algorithm, which translates k biased activation functions into $\mathsf{bPTC}^0(k)$ circuits. Furthermore, it includes majority gates with small fan-in, capturing some

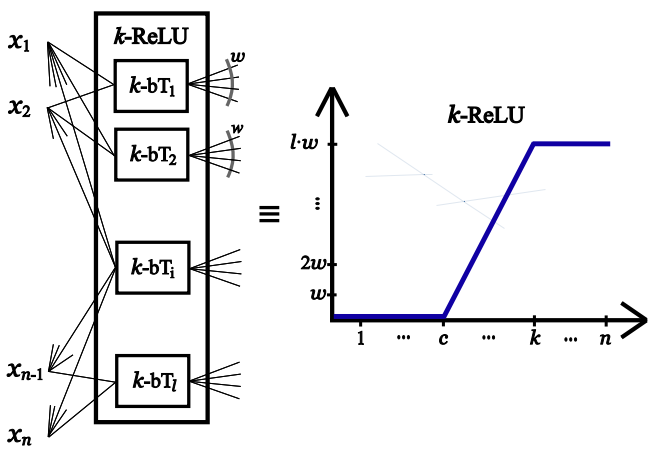

Fig. 1 | Representation of a k-ReLU gate within $\mathsf{bPTC}^0(k)$, constructed using multiple biased threshold gates (k-bT). This gate is equivalent to a ReLU gate, defined as $f(x)=\max\{0,x-c\}$ (where the center is shifted from 0 to c), up to an integer precision w for any input string with a Hamming weight bounded by k. Our scheme considers $\mathrm{ReLU}:\{0,1\}^n \mapsto \{0,n-c\}$, which takes n-bit strings as input and interprets their Hamming weight as the input x.

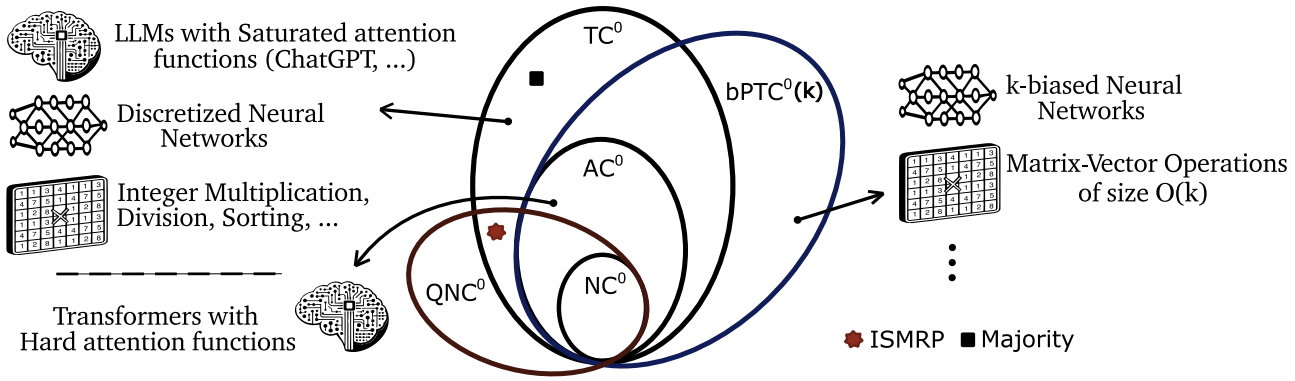

Fig. 2 | Relationships between the key classical and quantum circuit (complexity) classes considered. Notably, Majority is in TC^0 but not AC^0. The ISMR family, introduced here, separates constant-depth quantum circuits from biased polynomial threshold circuits (bias $k = \omega(\log n)$). The latter class contains AC^0, and can solve NC^1-complete problems for super-logarithmic biases and input lengths, suggesting non-trivial overlap with TC^0. We also note computational tasks and models of significant practical value, such as neural networks. For example, Large Language Models (LLMs) that use the transformer architecture can be simulated by AC^0 circuits when the attention mechanism is limited in certain natural ways[24,77]. Meanwhile, TC^0, the standard for modeling discretized neural networks[20,78], can simulate LLMs with realistic constraints on variable precision and autoregression[25,79,80], in the absence of more complicated elements such as feedback loops. The biased polynomial threshold circuits we study can simulate neural network variants and approximate activation functions controlled by the bias parameter k.

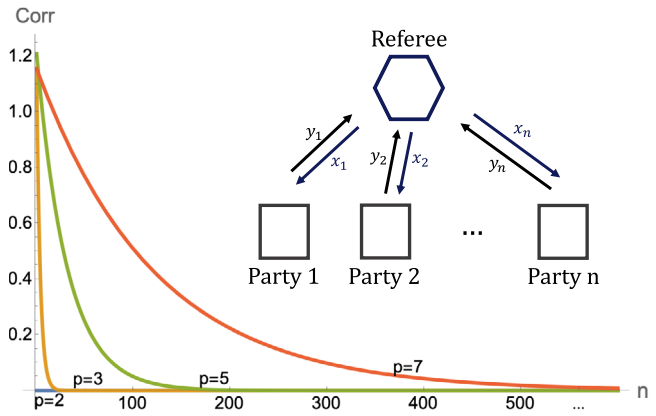

Fig. 3 | Representation of the n-party Modular XOR games $\mathcal{G}_p$. Each party P_i receives an input $x_i \in \mathbb{F}_p$ from a referee, forming a combined input string x with an ℓ_1 – norm $|x| = kp$, where $k \in \mathbb{N}$. Without further communication, each party responds with $y_i \in \mathbb{F}_p$, resulting in a collective output $y = (y_1, y_2, ..., y_n)$. The players win if the output y has ℓ_1-norm equal to the additive inverse of k modulo p, i.e., $|y| = -k \bmod p$. For context, $\mathcal{G}_2$ relates to the Mermin-Peres non-local game used in prior works[14,17]. The plot illustrates the upper bounds on the (winning) correlation of classical strategies for $\mathcal{G}_2$, $\mathcal{G}_3$, $\mathcal{G}_5$, and $\mathcal{G}_7$, whereas the optimal quantum strategy attains a maximum correlation of 1 for all $\mathcal{G}_p$.

of the computational power of TC^0. Notably, even a single biased threshold gate with bias $k = \omega(\log n)$ is known to require Boolean circuits of superpolynomial size (i.e., $\Omega(n^{\text{polylog}(n)})$) using unbounded fan-in AND, OR, and NOT gates to simulate it[26]. From this, and given that bounded fan-in constant-depth classical circuits are described by the NC^0 circuit class, we derive the following sequence of inclusions for classical parallel computational classes:

$$NC^0 \subsetneq AC^0 \subsetneq \text{bPTC}^0(k) \text{ for } k = \omega(\log n). \tag{2}$$

Thus, we highlight that biased polynomial threshold circuits of constant depth provide a valuable framework for investigating quantum computational power, especially by analyzing the impact of various bias parameters and their practical relevance (see Fig. 2).

Finally, we focus on relational or search problems, where the inputs x are n-bit strings, and the outputs y are m-bit strings. Formally, we have valid input-output pairs $(x, y) \in \mathcal{R}$ for some relation $\mathcal{R} \subset \{0,1\}^n \times \{0,1\}^m$. As in previous results, these relations will have non-local games embedded into them. In this paper, we introduce the family of qudit XOR non-local games, designated Modular XOR games, described in Fig. 3. Building on this class of multi-party non-local games, we introduce a corresponding family of relational problems, that we term Inverted Strict Modular Relation Problems (ISMRP). For any prime p, the ISMR problem $\mathcal{R}_p$ is defined as follows. Given an input $x \in \mathbb{F}_2^n$ such that $|x| \mod p = 0$, the goal is to output a string from the set

$$\mathcal{R}_p(x) := \left\{ y \,|\, y \in \mathbb{F}_2^m : |y| = -\left(\frac{|x|}{p}\right) \bmod p \right\}, \tag{3}$$

where $|z| = \sum_{i=1}^n z_i$ is the ℓ_1-norm for any string $z \in \mathbb{F}_p^m$, which is equal to the Hamming weight in the case of bitstrings. We typically choose the output length m to be slightly larger than n.

These problems possess some intriguing structural properties, as the set of valid output strings is determined entirely by the modular residue of the input's ℓ_1-norm $\bmod p$. To describe how well a circuit can solve this search problem, we employ a correlation measure suited to its modular nature:

$$\text{Corr}_{\mathcal{D}}(f, g) = \mathop{\mathbb{E}}_{x \sim \mathcal{D}} \left[\text{Re}\left(e^{i\frac{2\pi}{p}(|f(x)| - |g(x)|)} \right) \right]. \tag{4}$$

Here $\mathcal{D}$ is a distribution over input strings, and f, g are $\mathbb{F}_p^m$-valued functions. Our notion of correlation for ISMR problems extends standard correlation for Boolean functions to mappings from $\{0,1\}^n$ to cyclic groups $(\mathbb{F}_p, +)$, measuring the deviation between the ℓ_1-norm of the true output $|f(x)|$ and an estimate $|g(x)|$. Perfect alignment, with $|f(x)| - |g(x)| \equiv 0 \pmod p$, maximizes correlation, while other values decrease correlation, with penalties growing as deviations approach $\frac{p-1}{2}$. Thus, it accounts for "how wrong" an output is.

Higher dimensional qudits

Qudit-based quantum computation has generated significant interest in recent years, harnessing multidimensional quantum states that are more common in nature compared to two-level systems. They offer greater accuracy and efficiency in information storage and processing, along with improved noise resilience[28,29]. For example, complex entangled states such as multidimensional Greenberger-Horne-Zeilinger (GHZ) states and cluster states have demonstrated higher noise robustness compared to their qubit counterparts[30]. Additionally,

algorithmic adaptations for qudits have empirically been observed to offer advantages, particularly in quantum simulations of complex systems, where qudits serve as a natural simulation platform. These capabilities have driven the development of qudit-based computing models across various hardware platforms. Moreover, the use of qudits could facilitate quantum advantage experiments that, in turn, may also serve as critical benchmarks for evaluating future quantum devices.

This motivates our first contribution, demonstrating separations in computational power between quantum circuits operating over prime dimensional qudits and biased polynomial threshold circuits of constant depth.

Theorem 1. For every prime p and large enough n, the search problem $\mathcal{R}_p$ with n-bit inputs and $\mathcal{O}(n\log(n)^{p-1})$-bit outputs admits a constant-depth quantum circuit consisting of local gates over $\mathcal{O}(n)$ "qupits", consisting of $o(n^2)$ gates (i.e., of sub-quadratic size), that has constant correlation with $\mathcal{R}_p$. In contrast, any polynomial-size biased polynomial threshold circuit of depth d and bias $k = n^{1/(5d)}$, with access to random bits, has exponentially small correlation $\exp\left(-\Omega\left(n^{3/5-\mathcal{O}(1)}\right)\right)$ with $\mathcal{R}_p$.

As noted, biased polynomial threshold circuits serve as a useful theoretical model for neural networks, and the extent of expressivity and computational power modeled by such circuits is controlled by the bias parameter. Our results demonstrate that biased threshold circuits with bias parameter values satisfying $k = \mathcal{O}(n^{1/d})$, require superpolynomial size to solve a relational problem that small quantum shallow-depth circuits over qupits solve efficiently. This strengthens the provable quantum advantages over classical models with significantly more computational power than those in prior studies[14,17,31] (see also Eq. (2)). Crucially, we prove our bounds for the maximum bias parameter for the search problems considered—any larger bias k would enable these circuits to efficiently compute functions like parity and solve the problem with high probability using linear-sized circuits. Thus, our results push the possible quantum advantage against to this class of circuits, in terms of bias, to its theoretical limit.

Previous studies have suggested that qudit non-local games with significant gaps in quantum versus classical winning probabilities could imply computational separations between classical and quantum circuit classes[32], though such separations were not explicitly demonstrated. In Theorem 1, the family of relational problems $\mathcal{R}_p$ for each prime dimension corresponds to a family of Modular XOR non-local games that we define (Fig. 3). For these non-local games, we constructively prove that classical strategies achieve, at best, exponentially lower success probabilities than quantum strategies as the number of parties increases. The translation to quantum computational advantage is then achieved using shallow-depth quantum circuits in each prime qudit dimension over a corresponding standard, minimal *finite* gate set, with the precise instantiation of these circuits provided in the Supplementary Information (SI) Section B2. This approach broadens quantum advantages to an infinite family of problems and demonstrates the feasibility of implementing constant-depth realizations of such advantages on existing quantum hardware.

We also emphasize that this separation is achieved in the setting of average-case hardness with uniform input distributions over binary inputs. To do this, our method elegantly manages *non-uniform input distributions* for qupit non-local games, handling the necessary dit-to-bit mappings to ensure that we use uniform (rather than *biased*) random restrictions for the biased polynomial threshold circuits. This approach allows us to obtain average-case $\mathrm{bPTC}^0(k)$ hardness with respect to the uniform distribution on binary inputs for the search problems that demonstrate the separation. Additionally, it avoids the need to compute explicit success probability bounds for each game. In particular, we show that for our family of games parameterized by a prime p, the correlation of any classical strategy with a winning answer decreases exponentially with the number of parties in the game (see Section IV A). This correlation decay alone is sufficient to achieve comparable separations. We thus sidestep the obstacle of explicitly computing bounds on winning *probabilities*: this has been emphasized in prior research, which predominantly relied on established quantum-classical distinctions in the winning probabilities for non-local games[14,17]. Few developments beyond the conventional qubit setting were considered or achieved previously[33]. Finally, our search problems $\mathcal{R}_p$ remain binary, aligning naturally with Boolean circuits. They do not impose hardness by making clever use of arithmetic operations over other prime fields, making traditional Razborov-Smolensky type lower bounds inapplicable to this classical Boolean gate setting.

For the quantum solutions, we have generalized a method that translates optimal quantum strategies for these non-local games into small constant-depth quantum circuits that address the corresponding qupit search problems. We specifically leverage the use of generalized GHZ states in optimal strategies and the ability to generate multipartite entanglement between non-neighboring qupits in constant depth. We have derived local unitary (LU)-equivalent generalized GHZ states and introduced a technique to efficiently describe the support of standard measurement outcomes of these states, factoring in phase dependencies. This enables us to use them in combination with an additional correction function that is computable in constant depth, approximating the outcome of the optimal quantum strategy. Moreover, it facilitates the determination of the success probability, highlighting the quantum circuits' superiority over classical approaches.

To understand the potential advantages of parallel quantum computation, we first need a candidate problem with an efficient quantum solution. A significant part of the challenge subsequently lies in proving lower bounds on the resources (for example, the circuit size or depth) required by the corresponding classical parallel models to solve this problem. Our approach introduces a novel multi-output multi-switching lemma that serves as a reduction tool, breaking down multi-output biased polynomial threshold circuits into simpler classical computational models, such as decision trees, to which locality and light-cone arguments can be applied (see Section IV A). In Theorem 1, this new lemma plays a pivotal role by simplifying these circuits into forms that reveal their locality, directly linking their performance to the optimal classical strategies for the corresponding non-local games. More broadly, as these tools address not only binary decision problems but also search-type problems, this technique may hold independent significance, as neural networks are fundamentally sequence-to-sequence models. Therefore, this approach can be applied to relational problems beyond those examined here. Furthermore, parameterization by the bias parameter k provides a continuum of classical computational power levels, allowing for broader applicability. In particular, we use this framework to clarify the relationship between hardware requirements and the potential for quantum advantage.

Qubits

Among all prime qudit dimensions, only two—$p = 2$ and $p = 3$—allow for an intuitive translation from quantum lower and classical upper bounds on correlations to corresponding bounds on success probabilities. Specifically, our results show that polynomial-size biased polynomial threshold circuits (of small bias) cannot solve the $\mathcal{R}_2$ and $\mathcal{R}_3$ problems with success probabilities appreciably larger than 1/2 and 1/3, both of which correspond to random guessing. In contrast, quantum circuits can solve the qubit case exactly, while in the qutrit case, the problem is solvable with a probability larger than and bounded away from 1/2 by a constant.

In the qubit case, we can in fact demonstrate something stronger: that if classical circuits are required to solve $\mathcal{R}_2$ exactly—that is, to produce a valid output string with certainty for all inputs—quantum advantages or separations of even greater magnitude are attained.

Theorem 2. For large enough n, the search problem $\mathcal{R}_2$ with n-bit inputs and $\mathcal{O}(n \log n)$-bit outputs can be solved by a constant-depth qubit quantum circuit with $o(n^2)$ gates, which on any valid input x outputs y such that $(x, y) \in \mathcal{R}_2$ with certainty. In contrast, the size s of any depth-d k-biased polynomial threshold circuit, with access to random bits, that computes any valid y is lower bounded as in Table 1.

We remark that this exact-case analysis is well-motivated by the fact that the quantum circuit in the qubit case is actually a deterministic solution to $\mathcal{R}_2$, and therefore it is fair to compare it with classical deterministic circuits—to draw a complexity theoretic analogy, this is similar to comparing EQP and P, as opposed to the more common comparison of BQP and BPP. We also note that for $k = \mathcal{O}(1)$, constant-depth k-biased polynomial threshold circuits are equivalent to constant-depth logical circuits with unbounded fan-in Boolean gates (i.e., AC^0), and so our bounds in the exact-case augment and improve on prior work[17].

In proving Theorem 2, we have developed a technique to capture the classical-quantum circuit separation using the algebraic normal form (ANF), a standard representation for Boolean functions (see Section IV B). Importantly, this approach avoids the need for non-local or contextual games[32], which have traditionally been essential in prior work, while also showing that previous quantum advantages might benefit from new techniques to improve resource requirements and bring them closer to practical quantum demonstrations.

We thus derive two lower bounds on the classical circuit size in the qubit setting. The most robust demonstration of quantum advantage arises in the setting of average-case hardness, where deviations of the classical success probability away from random guessing can be directly bounded. However, our exact-case hardness analysis reveals a quantum advantage of a greater magnitude, establishing higher resource requirements for classical circuits to match the performance of the quantum circuits they must compete with. This suggests that quantum advantage experiments against shallow-depth classical circuits could be achieved with fewer resources.

Noise-robustness

Finally, quantum systems are unavoidably affected by noise, and error correction is a much more complex process in the quantum realm, bringing into question the *robustness* of a computational advantage under noise. This question is of significance to both theory and practice, especially as we navigate the NISQ era. Even for very powerful quantum computational models, the introduction of noise often dramatically diminishes computational advantages that they may offer over their classical counterparts: for instance, recent work shows that even small constant error rates result in a collapse of the power of multi-prover interactive proofs where the provers share entanglement (MIP^*) from RE to multi-prover interactive proofs without shared entanglement (MIP)[34]. Our third main result is to prove that all our separations are robust to noise: even if all steps of the quantum computation are affected by local stochastic noise, there is a family of modified relation problems that these noisy quantum circuits, when provided with logical magic states, can solve, but is hard for noiseless $bPTC^0(k)$ circuits.

Theorem 3. For every prime p and large enough n, there exists a search problem $\mathfrak{R}_p$ with n-bit inputs and $\mathcal{O}(n \cdot \mathrm{poly}(\log n))$-bit outputs, such that for local stochastic noise with physical error rate below a constant threshold, there is a noise-resilient constant-depth quantum circuit over qupits with local gates and all-to-all connectivity, equipped with logical $|T^{1/p}\rangle$-magic states, that has constant correlation with $\mathfrak{R}_p$. In contrast, any (even noiseless) depth-d polynomial-size biased polynomial threshold circuit with bias $k = n^{1/(5d)}$ has exponentially small correlation $\exp\left(-\Omega\left(n^{3/5-\mathcal{O}(1)}\right)\right)$ with $\mathfrak{R}_p$.

The search problem $\mathfrak{R}_p$ is defined using the ISMR problem $\mathcal{R}_p$, accounting for quantum error correction (see SI Section C).

Table 1 | Lower bounds on the size of $bPTC^0(k)$ circuits solving the $\mathcal{R}_2$ problem, for different regimes of the value of k

$bPTC^0(k)$/rpoly	Exact hardness
$k = \mathcal{O}(1)$	$s = \Omega\left(\exp\left(\left(\frac{\sqrt{n}}{(\log n)^{3/2+\mathcal{O}(1)}}\right)^{\frac{1}{d-1}}\right)\right)$
$k = n^{1/(5d)}$	$s = \Omega\left(\exp\left(\left(\frac{n^{3/10}}{(\log n)^{3/2+\mathcal{O}(1)}}\right)^{\frac{1}{d-1}}\right)\right)$

Local stochastic noise is a standard model used in quantum error correction, favored due to its ability to account for gate-level noise, noisy input state preparation, and noisy measurements while allowing for (weakly) non-local errors. Locality means that the probability of any given Pauli error decays exponentially with the number of sites it affects non-trivially. Additionally, this model effectively captures fabrication faults and aligns with standard physical descriptions of noise, wherein errors become exponentially less probable as their weight increases[35]. Theorem 3 proves that our computational separations are robust to the presence of qupit generalized local stochastic noise in the quantum circuits.

This result improves upon prior work in three main ways. First, it extends unconditional separations between noisy, shallow-depth quantum circuits and classical circuit classes. In particular, this is achieved for all prime qudit dimensions and against the largest classical circuit class to date. Notably, while most qudit dimensions require logical T-type magic states as advice in the general formulation of Theorem 3, we have also demonstrated a noise-resilient quantum advantage using Clifford circuits over qubits that do not rely on such advice. Additionally, as a corollary, this establishes separations against NC^0 for each prime qudit dimension, potentially enabling near-term quantum advantage experiments due to the reduced resource requirements for this class and the favorable error-resilience properties of qudits.

Second, our approach extends the framework for noise-robust quantum advantages beyond the qubit Clifford model introduced in ref. 16, by addressing qudit non-Clifford gates. Specifically, we demonstrate that for a particular CSS-type error correction code, this extension is achievable through the use of logical T-type magic states, which can themselves be affected by local stochastic noise, along with qudit magic-state injection protocols (see Section IV C). The need for these more complex quantum circuits arises from the inability to violate Bell inequalities with stabilizer states for any qupit dimension beyond qubits, as shown in refs. 36–38, which also suggests that the same limitation extends to quantum-classical separations in circuit complexity. Additionally, as part of Theorem 3, we design new quantum circuits in the form of non-adaptive Clifford circuits with input-independent advice states to solve ISMR problems fault-tolerantly. These quantum circuits essentially give rise to our definition of the $\mathfrak{R}_p$ search problems.

Third, our work extends shallow-depth computational separations and error-correction mechanisms across arbitrary prime qudit dimensions, demonstrating their robustness. Previous research used the minimum weight perfect matching decoder, which performs poorly in higher dimensions. By using a different decoder, we show that we can still perform corrections and recover the desired states with exponentially high confidence. Specifically, we illustrate how the qupit surface code, when combined with the hard decision renormalization decoder, supports fault-tolerant implementation of the necessary quantum circuits. This advancement includes the development of single-shot logical state preparation for qupits. We have also extended the 3D block construction from[39] to higher dimensions, showing that a particular measurement pattern yields a reduced state corresponding to a logical GHZ_2 state, up to local Clifford corrections.

Table 2 | Estimates of the input sizes *n* required to demonstrate quantum advantage using constant-depth quantum circuits in both noise-free and noise-affected settings, based on the quantum upper and classical lower bounds determined in Section II B, Section II C, and Section II D

	NC^0 regime				$bPTC^0(k)$ regime			
Quantum circuit type	**Min F^***	**$F=2$**	⋯	**$F=8$**	**$k=1$**	**$k=2$**	⋯	**$k=n^{1/(5d)}$**
Qubits	2540	9364	⋯	1.6×10^{8}	4.3×10^{13}	5.5×10^{16}	⋯	5.0×10^{26}
Noisy qubits	-	7.0×10^{10}	⋯	1.5×10^{23}	1.8×10^{38}	8.0×10^{44}	⋯	1.1×10^{75}
Qutrits	125162	1952660	⋯	1.3×10^{10}	5.5×10^{14}	1.1×10^{18}	⋯	2.3×10^{30}
Noisy qutrits	-	3.5×10^{11}	⋯	9.0×10^{30}	3.5×10^{52}	9.0×10^{60}	⋯	3.4×10^{101}
Ququints	1.0×10^{8}	1.5×10^{9}	⋯	8.0×10^{12}	9.5×10^{14}	1.7×10^{18}	⋯	5.0×10^{30}
Noisy ququints	-	3.0×10^{14}	⋯	7.0×10^{33}	5.5×10^{52}	1.4×10^{61}	⋯	7.0×0^{101}

For the noise-free case, we consider the depth of four quantum circuits solving the $\mathcal{R}_p$ problems. These circuits, featuring $2n$ gates and all-to-all connectivity, generate the shortest possible solution strings, creating harder instances for classical circuits to replicate. The classical lower bounds for each problem are obtained from and depend on the deviation between the optimal classical and quantum winning strategies for the XOR non-local games $\mathcal{G}_p$. We also examine the minimal fan-in (F^*) scenario by comparing quantum circuits with classical circuits of equal locality—that is, having the same fan-in as the quantum circuits in each layer. This yields the lowest resource estimates for direct comparisons. Additionally, we analyze exact-case hardness bounds for the qubit setting to establish lower resource estimates. For the remaining qudit dimensions, we rely on average-case hardness to derive comparable estimates.
In the noisy setting, we analyze depth-9 error-corrected quantum circuits for qubits and depth-11 circuits for qudits. The latter qudit circuits include additional logical operations implemented in a noise-resilient manner, requiring noise-tolerant versions based on the qudit surface code. We assume a code distance of order $\log(n)$, as no specific error threshold is defined for the local stochastic noise. This threshold is left as a parameter for further investigation and potential alignment with quantum hardware advancements.

Resource estimates

When testing computational advantages with physical implementations, it is essential to pinpoint the circuit depth d and input size n (i.e., the number of input qubits) where a transition in circuit size occurs. Specifically, at what depths and input sizes does quantum advantage emerge? To address this, we estimate these values by solving for the points where our new asymptotic lower bounds on the size of the best classical circuits match our corresponding quantum upper bounds. In Table 2, we present the input size n for a given depth d, corresponding to the shallowest quantum circuits (with all-to-all connectivity) in each qudit dimension that outperform their classical counterparts in solving the respective ISMR problems.

For context, the transition point for Shor's factoring algorithm is estimated to be ~1700 qubits, 10^{36} Toffoli gates, and a circuit depth of 10^{25} [40], while for the HHL quantum matrix inversion algorithm it is roughly 10^{8} qubits and a depth of 10^{29} [10].

In comparison, recent advances in quantum hardware have prioritized scaling up the number of qubits over extending coherence times, leading to a greater emphasis on shallower quantum circuits[5,41]. Notably, the quantum advantages over constant-depth classical circuits with bounded fan-in gates, as studied in this work, require only thousands of qudits across various dimensions. These setups can demonstrate classical intractability in tasks such as Bell violations[42,43]. Classical circuits for these problems must have a depth of at least $d=\Omega(\log n)$, showing a clear quantum advantage when a quantum circuit solves the same problem at a strictly smaller depth. In practical scenarios with noise, the quantum circuit depth may increase by a constant factor, while the minimal classical circuit depth remains at least $d=\Omega(\frac{\log n}{\log\log n})$. These noise-driven increases in the input size and other parameters required to observe quantum advantage is still significantly smaller, in terms of total resource counts, than what is required in other previous quantum advantage demonstrations.

Progressing up the hierarchy of computational power toward demonstrations of unconditional quantum advantage, the challenge lies in outperforming larger classical constant-depth circuit classes, such as biased polynomial threshold circuits. Achieving such quantum advantages would require significantly greater quantum resources, yet could still compare favorably to other *conditional* quantum advantage demonstrations in noise-free settings. These quantum circuits maintain the advantage of shallow depth, potentially making them more practical for near-term quantum hardware. However, comparisons between noise-affected quantum circuits and noise-free classical circuits often demand unrealistic resources, exposing a fundamental imbalance in such analyses. Classical computing benefits from decades of refinement, and classical error-correcting codes may also be necessary to achieve exponentially high efficiency rates akin to those expected of error-corrected quantum circuits[44]. Although noise levels in quantum and classical systems are unlikely to converge[6,45,46], resource estimates that account for noise on both sides—combined with advances in quantum hardware—are anticipated to bring these comparisons closer to what is expected from noise-free quantum models.

Beyond the estimates in Table 2, our bounds extend to all prime qudit dimensions and explore quantum circuits with varying hardware connectivities. In noise-resilient scenarios, all-to-all connectivity is required, while in noise-free comparisons, architectures can range from $p+1$-dimensional configurations for qudit dimension p to full all-to-all connectivities. Among these, our estimates represent the lowest obtained for equivalent unconditional quantum separations, with the qubit case achieving the smallest resource requirements due to our analysis of exact-case hardness bounds. As mentioned before, while this setting reflects a less robust form of quantum advantage it nevertheless significantly reduces resource requirements, bringing theoretical predictions closer to the capabilities of current quantum devices. These estimates could be improved for all qudit dimensions by requiring classical circuits to better match the performance of the quantum circuits, and by adding connectivity restrictions to classical circuits. While these constraints are not formally part of the definition of shallow-depth circuit classes such as NC^0, they reflect the limitations of realistic classical hardware and could lower the input size needed to demonstrate quantum advantage[47]. More generally, tighter lower-bound techniques and the discovery of computational problems with greater quantum advantages could further refine these estimates.

Finally, our estimates indicate that these quantum advantage experiments could serve as powerful quantum benchmarks, providing a systematic framework for evaluating and comparing computational capabilities. To explore larger quantum computational advantages, one could test classical circuits with larger fan-in gates, establishing new benchmarks and creating a structured "ladder of quantum advantages" to assess increasingly stronger computational separations. This hierarchy can be expanded by examining the ability of shallow quantum circuits to outperform more advanced classical circuit classes, such as the biased polynomial threshold circuits analyzed in this work. Adjusting the bias parameter within these circuits further allows one to tune or amplify potential quantum advantages offered by specific quantum circuits and architectures. Notably, the levels within

both hierarchies are separated by small multiplicative factors, making them an effective tool for benchmarking hardware improvements and guiding steady progress in quantum computing.

Discussion

In this paper, we advance the growing body of work demonstrating unconditional separations between the computational power of classical and quantum shallow-depth circuits, focusing on extending such results to hold against the largest classical circuit classes studied to date in this context. Specifically, we show that small local shallow-depth quantum circuits can efficiently solve search problems that polynomial-size circuits of k-biased threshold gates fail to solve with significant probability, even on average, and even for appreciably large bias parameter k.

We have developed a family of non-local games for qupits of each prime dimension and utilized the difference in winning probabilities between classical and quantum strategies to demonstrate that the computational separation established for constant-depth qubit circuits extends to constant-depth quantum circuits over higher-dimensional quantum systems. This, combined with the fact that these circuits are defined over a standard, minimal finite qupit gate set, allows for constant-depth realizations using the elementary operations available on standard quantum computing devices and fault-tolerant implementations. These explicit quantum advantages clarify the theoretical landscape and have practical relevance, as many quantum computing platforms naturally operate in higher dimensions[29,48]. Furthermore, we hope our estimates inspire further advances in proof techniques and parameter optimization, narrowing the gap between theoretical predictions and the capabilities of near-term quantum devices toward achieving experimental quantum advantages.

Considering an infinite gate set, such as {all single-qubit *gates*, CNOT} as assumed in refs. 18,49,50 would certainly enable realizing all our qudit circuits using qubit circuits in a noise-free setting[51,52]. However, these realizations are not feasible in real hardware in constant depth[53] or in a fault-tolerant manner. Thus, in a more realistic context, any finite minimal gate set defined over a specific qudit dimension would require decomposition into its native gates if used to solve one of the ISMR problems from another prime dimension. In this regard, employing Solovay-Kitaev-type decompositions, these gates would likely necessitate log-depth decompositions to achieve suitable approximations. Therefore, under this hardware-realistic definition of constant-depth qudit quantum circuits, we conjecture that for each prime p, there exists a relational problem—in particular, our ISMR problem $\mathcal{R}_p$—that cannot be solved by a constant-depth quantum circuit using a minimal universal gate set for qudits of any dimension $q \neq p$, but can be solved by such a circuit using a gate set for dimension p. We propose that the ISMR problems could thus play a similar role for qudit constant-depth quantum circuit classes as Razborov-Smolensky-type modular problems do for the classical $\mathrm{AC}^0[p]$ circuit classes. If validated, this conjecture could reveal important aspects of parallel quantum computations that depend on specific system dimensions and might also guide hardware manufacturers in considering quantum system dimensions beyond qubits.

Having initiated the consideration of error-corrected qudit circuits for unconditional quantum separations, we speculate whether prior work on high-dimensional error-correcting codes with improved parameters[54] could facilitate experimental demonstrations. We have reiterated the interest of magic states in the constant-depth fault-tolerant regimes[55,56], and remark that we believe that magic state factories are unlikely to be parallelizable to the extent of being realizable by constant-depth quantum circuits. It is hence of great interest to understand what the simplest classical circuit class capable of simulating these processes is, as well as introduce the capacity for adaptive operations during error-corrected circuit execution. This offers potential avenues to extend unconditional quantum-classical separations beyond what is currently known and lift larger conditional separations to the simplest fault-tolerant quantum circuit classes[57]. In parallel to such complexity theoretic questions, it would be important to consider more complex and alternative noise models[58] to better align with specific practical quantum computing architectures. In the same vein, it is also of interest to understand if there is a complexity phase transition for some values of the bias parameter k and the noise strength.

Finally, our computational separations could highlight potential advantages of quantum over classical machine learning models. Quantum machine learning has often been shown to outperform classical approaches by enabling the encoding of classically hard problems or leveraging quantum phenomena such as non-locality and contextuality. Some of these advantages can also be demonstrated using metrics that are standard for these learning models, such as the Kullback-Leibler divergence[59]. Our results complement this body of work. Additionally, biased polynomial threshold circuits naturally model a broad range of neural networks, including certain transformer architectures central to modern LLMs, leading us to conjecture that they support attention mechanisms beyond AC^0 and likely beyond TC^0 (for appropriate values of the bias). Consequently, we are optimistic that further work extending our line of investigation can reveal novel quantum advantages over classical machine learning models in these settings.

Moreover, a key challenge beyond proving quantum advantage is the difficulty of learning effective quantum solutions to various problems. In quantum machine learning, it has been demonstrated that models that are easy to train can often be classically simulated efficiently[60], thus limiting their quantum advantage. Conversely, more complex models that could potentially offer quantum benefits are theoretically more challenging to train. Therefore, identifying models that are well-suited for practical quantum advantage is crucial. Our research supports the notion that even simple and potentially easy-to-learn quantum circuits can outperform their classical counterparts significantly, reinforcing the idea that efficiently learnable quantum circuits can nevertheless provide practical advantages in information processing tasks[59,61–67].

Methods

We now provide a concise overview of the proofs our main results. We aim to strike a balance between rigor and accessibility, and give an intuitive discussion of both the classical and quantum techniques employed. We will also highlight the key technical improvements we have made over existing work in this area.

Proof of theorem 1: separations for higher dimensions (qupits)

We introduce an infinite family of ISMR problems that generalize the parity halving problem introduced by ref. 17, and the hidden linear function problem studied by ref. 14 reduces to the latter. The ISMRPs enable us to extend prior work on unconditional quantum advantage in two complementary directions: from qubits to all prime dimensional qudits, and beyond classical unbounded fan-in AND-OR circuits to the strictly more powerful class of biased polynomial threshold circuits.

The proof has two parts. First, we show that multi-output biased polynomial threshold circuits of polynomial size have poor correlation with the ISMR outputs as the input size grows. Second, we demonstrate that quantum circuits using qupits can solve all instances of the ISMR problems with a constant positive correlation (depending only on p), regardless of input size.

The key challenge is to establish upper bounds for the correlation with which biased polynomial threshold circuits of size s and depth d can solve the ISMR problems on typical inputs of length n. This is the most technical part of the proof, as no previous lower-bound techniques existed for multi-output biased polynomial threshold circuits of constant-depth. To address this, and constrain the computational

power of multi-output biased polynomial threshold circuits, we introduce a new multi-output multi-switching lemma that can reduce these circuits to decision forests. In this context, a decision forest consists of a single global decision tree with a set of decision trees at each one of its leaves, each computing a single output bit of the original circuit.

Lemma 1. Let f be a k-biased polynomial threshold circuit with m output bits and n input bits, of size s and depth d. Then, there is a random restriction whose probability p depends on s, d, k, t, and q, that reduces f to a decision forest with global decision tree depth $2t-2$, and maximum depth q of the decision trees at its leaves with probability at least $1-s\cdot 2^{-t+k}$.

By applying this lemma and selectively fixing variables in the global decision tree, we reduce the biased polynomial threshold circuits to m independent decision trees, each computing a single outcome bit. This reduction limits the complexity of the circuit, while the problem retains its structure and hardness over the variables left unfixed by these restrictions. Further exploiting this asymmetry using locality arguments such as lightcone techniques on the final decision trees allows us to relate the efficiency of the initial circuits in solving the $\mathcal{R}_p$ problem to that of classical strategies in solving a non-local game embedded within it over the remaining variables. For example, in the qubit case, Mermin's multi-player game[68] is implicitly integrated into $\mathcal{R}_2$ over the variables preserved by the random restriction. Thus, for the qubit case, our new multi-output multi-switching lemma immediately shows that, for large n, any biased polynomial threshold circuit of depth $d\geq 4$, size $s\leq \exp(n^{1/(2d-2)})$ and $k\leq n^{1/(5d)}$ solves $\mathcal{R}_2$ with a success probability bounded by

$$\frac{1}{2}+\exp\left(-\Omega\left(\frac{n^{2-o(1)}}{m^{1+o(1)}(k\cdot\log(s))^{2d}}\right)\right). \tag{5}$$

However, in the higher-dimensional setting, many essential technical tools were previously undeveloped. To the best of the authors' knowledge, this is the first time the family of modular XOR non-local games related to the respective ISMR problems has been defined. Consequently, we need to determine upper bounds on the efficiencies of any classical strategies, whereas previous work relied on established non-local games with known bounds on optimal winning strategies.

Lemma 2. Let w be any local probabilistic classical strategy that wins the Modular XOR non-local game $\mathcal{G}_p$ with n parties exchanging messages that are values in $\mathbb{F}_p$. For inputs drawn according to the uniform distribution over strings of length $n(p-1)$ with ℓ_1-norm satisfying $(\sum_{i=1}^n x_i) \bmod p = 0$ and a fixed binary to base-p encoding, the maximal correlation is bounded by $\mathrm{Corr}\left(w,\mathcal{G}_p\right)\leq (c_p)^{\frac{n}{p-1}}$, for a constant $c_p\in(0,1)$.

To establish these bounds for the non-local games $\mathcal{G}_p$—crucial for applying uniform random restrictions in the proof and deriving average-case hardness results—we consider encodings and decodings between uniform distributions over $\mathbb{F}_2^n$ and non-uniform distributions over $\mathbb{F}_p^n$. This approach is necessary because these non-local games are naturally defined as mappings of the form $\mathbb{F}_p^n\mapsto\mathbb{F}_p^m$. Additionally, these bounds rely on selecting an encoding that introduces a linear bias in the non-local games, as achieved by our chosen encoding, which can then be incorporated into our proof technique to establish tight upper bounds on the success probabilities of classical winning strategies (see SI Section B2).

Finally, by employing the same sequence of techniques—starting with the multi-output multi-switching lemma, followed by lightcone arguments, and considering blocks of bits that represent the dits of the non-local game, while leveraging the fact that these dits do not exhibit any specific structure assumed in previous works (e.g.,[31])—we relate the efficiency of the biased polynomial threshold circuit, as well as the intermediate NC^0 circuits, to the upper bounds on optimal classical strategies. Specifically, we show that for sufficiently large n and $q\in\mathbb{N}_{>0}$, any biased polynomial threshold circuit C of depth $d\geq 4$, size $s\leq\exp(n^{1/(2d-2)})$, bias parameter $k\leq n^{1/(5d)}$, and access to random strings $\mathrm{rpoly}\in\mathbb{F}_2^{\mathrm{poly}(n)}$ solves $\mathcal{R}_p$ with correlation bounded by

$$\mathrm{Corr}\left(C,\mathcal{R}_p\right)=\exp\left(-\Omega\left(\frac{n^{2-o(1)}}{m^{1+2/q}\log(s)^{2d-2}k^{2d}}\right)\right), \tag{6}$$

for a uniform input distribution over strings $x\in\mathbb{F}_2^n$ that satisfy $|x| \bmod p = 0$.

To establish the quantum lower bound on correlation for the ISMR problems, we construct circuits that solve these problems with constant positive correlation by translating optimal quantum strategies for modular XOR non-local games into constant-depth quantum circuits that produce equivalent output strings. This approach first generates a qudit generalized "poor man's cat state", which is LU-equivalent state to qudit generalized GHZ states, essential for optimal quantum strategies. These states serve as resources, and we show that a random instance from this class of states can be generated using constant-depth qupit circuits, along with a string $z\in\mathbb{F}_p^n$ that suffices to define the operator mapping the state to the respective p-dimensional GHZ state.

Using these resource states, we apply the rotations and measurements defined by the optimal quantum winning strategies to produce an output string. However, higher-dimensional poor man's cat states introduce undesired phase factors from multiple inner products between the state-defining string z and the input x, altering the expected measurement outcomes in subtle ways. To address this, we first present a concise representation of the output string's support based on the input and random strings defining the states. We then show that computing at least one of these inner products and incorporating it into the output string enhances efficiency in solving the original problems, while also demonstrating that a constant-depth classical circuit cannot compute at least one of these terms. Thus, we conclude that the quantum circuits solve these problems with constant, input-independent correlation. More formally, we prove that a constant-depth quantum circuit C_Q solves the ISMR problem $\mathcal{R}_p$ on a uniformly random input from $\mathbb{F}_2^n$ that satisfies the condition $\sum_{i=1}^n x_i \bmod p = 0$ with a high correlation, namely

$$\mathrm{Corr}\left(C_Q,\mathcal{R}_p\right)=\frac{p-1}{p^2}. \tag{7}$$

We remark that also for the qutrit case, the quantum advantage can be expressed through the success probability for the $\mathcal{R}_3$ problem, as the correlation function relates directly to success probabilities in both qubit and qutrit scenarios.

In summary, quantum advantages are achieved with qupit constant-depth circuits across various finite-dimensional connectivities. Each problem $\mathcal{R}_p$ can be addressed with qupit circuits of dimension p and p-dimensional connectivity. However, to optimize these quantum advantages, careful consideration of parameters is essential. Specifically, the interplay between the bias parameter k, circuit size, and the dimensionality of constant-depth quantum circuits must be optimized. For instance, in the Measurement-Based Quantum Computation (MBQC) paradigm, all-to-all connectivity allows for shallower circuits and improved resource estimates, as shown in Section II E.

Proof of theorem 2: qubit exact-case hardness

As mentioned before, since the quantum circuits actually solve the search problem with certainty, it is fair to ask what the hardness of an analogous classical exact solution is, exploring the boundary between deterministic and probabilistic circuits. We use the term *exact-case* to

describe the ability of the circuit to solve the problem with certainty for all inputs. This differs from vanilla worst-case hardness, wherein the circuit is only required to succeed with a high enough (constant) success probability. In particular, from the average-case correlation bound in ref. 17, one can determine a lower bound on the size of an AC^0 circuit that solves this problem in the exact setting. A similar result follows for $\mathsf{bPTC}^0(k)$ from our average-case hardness result referenced in Equation (5). However, these bounds are not tight, as we have demonstrated with our exact-case hardness result, which implies that even larger AC^0 and $\mathsf{bPTC}^0(k)$ circuits are required to solve the problem exactly. Focusing on deterministic classical circuits in this way has the benefit of revealing quantum-classical advantages at input sizes that are orders of magnitude smaller.

To achieve this, we develop a deeper combinatorial understanding of the $\mathcal{R}_2$ problem. We observe first that the XOR of all the output bits is always equal to a fixed Boolean function, namely $\mathrm{LSB}:\{0,1\}^n\mapsto\{0,1\}$ which outputs the second least significant bit of the binary representation of the Hamming weight of the input $|x|$. In addition, each output bit produced by a $\mathsf{bPTC}^0(k)$ circuit can be viewed as a distinct Boolean function f_i: $\{0,1\}^n\mapsto\{0,1\}$. By examining these functions in their ANF—which represents each function as a polynomial over the field $\mathbb{F}_2$—we find that the XOR of the ANF of all these Boolean functions must match the ANF of the function $\mathrm{LSB}(x)$.

When the input distribution is supported over all the bit strings in $\mathbb{F}_2^n$ we would be able to use the ANF of the LSB function directly. However, $\mathcal{R}_2^m$ is a *promise problem*, in that only strings of even parity are considered valid inputs. Thus, we modify the discussion above to work for *partial* (or "partially defined") Boolean functions. To do this, we first prove a property about the ANF for all the (exponentially many) Boolean functions that equal the LSB function on our domain of interest. We do this by showing that all these functions require, in their ANF representation, at least $\Omega(n^2)$ degree-two terms. Secondly, we examine the capacity of decision trees, to which $\mathsf{bPTC}^0(k)$ circuits can be reduced under random restrictions, in generating terms of degree two. As an illustrative example of how the tree depth relates to the ANF of Boolean functions, consider the parity function: any decision tree computing the parity of n bits must have a depth of at least n. This ensures that the tree can produce all the degree one terms of the parity function's ANF. See SI Section B1 for more details.

Finally, these two components, in conjunction with our switching lemma (Lemma 1) that reduces $\mathsf{bPTC}^0(k)$ circuits to decision trees, allow us to determine a minimal depth of the decision trees that directly translate to the minimal size for this class of circuits: for sufficiently large n, any $\mathsf{bPTC}^0(k)/\mathrm{rpoly}$ circuit depth $d\geq 4$ and $k\leq n^{1/(5d)}$ that solves $\mathcal{R}_2^m$ must have size at least

$$s\geq \exp\left(\widetilde{\mathcal{O}}\left(\left(nk^{-d}m^{-1/2}\right)^{1/(d-1)}\right)\right). \tag{8}$$

Together with our quantum circuits for $\mathcal{R}_2^m$, this lower bound completes our proof of Theorem 2.

Proof of theorem 3: noise-resilient quantum advantage

At a high level, we define a set of problems related to noise-tolerant, constant-depth quantum circuits for solving ISMR problems. These problems are designed to demonstrate the noise robustness of our quantum advantages. Specifically, we show that the ISMR problem can be reduced to a problem in our new family when an AC^0 circuit can decode the output of these noise-tolerant constant-depth quantum circuits. This reduction extends the hardness results from the noiseless case to the noisy, achieving noise-robust separations with correlation bounds similar to those in Theorem 1.

We tackle two main technical challenges that go beyond previous noise-robust separations explored in refs. 16,69. First, for qudits of dimension $p\geq 3$, we must handle *non-Clifford* circuits, requiring us to show that these circuits can also be implemented fault-tolerantly—a new approach beyond prior work on unconditional separations between quantum and classical circuits. Second, we generalize quantum error-correction techniques used in ref. 16, including single-shot state preparation, decoding, and transversal constant-depth gate execution, to prime qudit dimensions while integrating methods for fault-tolerant non-Clifford circuits.

We address the first challenge by introducing logical advice states over qupits, enabling fault-tolerant non-Clifford operations through a qupit magic state injection protocol. We demonstrate that using a CSS-type error correction code that meets specific conditions (see SI Section A4), we can implement non-adaptive qupit Clifford circuits fault-tolerantly within the constant depth, for arbitrary prime dimensions, utilizing an advice state. Moreover, we show that for a code distance $l=\mathcal{O}(\mathrm{poly}\log n)$, if the circuit and advice are affected by local stochastic noise, respectively, $\mathcal{E}\sim\mathcal{N}(\mathrm{Q})$ and $\mathcal{E}_A\sim\mathcal{N}(\rho)$, then whenever both the physical error rates Q and ρ are below a threshold value scaling as $p_{th}=2^{-2^{\mathcal{O}(d)}}$, for any input $x\in\mathbb{F}^n$ we can show that $\Pr[\mathrm{DEC}^*(\mathcal{EC}(x))=C(x)]>0.99$, with DEC^* being the combined correction and decoding operation needed to retrieve the logical outcome from the encoded state generated by the fault-tolerant implementation $\mathcal{EC}$.

Our proof consists of two parts. First, we demonstrate that the qupit surface code meets all the necessary conditions for fault-tolerant implementation, specifically ensuring that it supports single-shot state preparation, transversal gate implementation in constant depth, and single-shot information retrieval using the selected decoder. Making these steps precise in higher-dimensional constant-depth quantum circuits requires new insights that extend beyond prior work. For the second part of our proof, we design new constant-depth non-adaptive Clifford circuits operating over qupits with input-independent advice states, capable of solving the ISMR problems.

Noise-resilient qupit Clifford circuits with quantum advice. In higher dimensions, errors do not simply manifest as defects at the endpoints of the qupit lattice, as they do in the case of qubits. Instead, for every error, defects are likely to be scattered throughout the lattice. This distribution of defects motivates our use of the hard-decision renormalization group (HDRG) decoder, which has been shown to have good error-correction properties beyond the qubit case, overcoming issues associated with the minimum weight perfect matching decoder[70,71]. In addition, for our setting, the information retrieval property necessitates that the code and decoder accurately perform logical $\overline{Z}$ measurements even under noisy conditions. To address this, we extend a result from ref. 72 to qupit generalized local stochastic noise. We demonstrate that the HDRG decoder's probability of failure decreases exponentially as the qupit lattice size increases. Consequently, the HDRG decoder also reliably yields the correct outcome for logical measurements, provided the physical error rate does not exceed a threshold. Specifically, we establish that $\Pr[\mathrm{Success}]\geq 1-\exp(-\Omega(m^\eta))$, where m is the surface code distance and η is a constant.

Although the transversal implementation of qudit Clifford operations in constant-depth, follows from the work of ref. 73, achieving single-shot state preparation in the qupit surface code is more complex. We show that single-shot state preparation can be achieved using the qudit surface code and the HDRG decoder for qupits of dimension $p\geq 2$. More precisely, we show that the 3D block construction from ref. 39 preparing logical Bell pairs can be adapted for single-shot state preparation of GHZ_2 qupit states, a capability not previously demonstrated. This involves defining functions necessary for the single-shot state preparation process, typically categorized as recovery and repair. The recovery function entails applying operations to retrieve the correct state from the randomness inherent in the noise-free process, while the repair function corrects the states based on the effects of noise that may occur

during the described noise-free state preparation circuit and the recovery function.

The recovery function is derived directly from our generalization of the state preparation process. Regarding the repair function, we establish its feasibility up to the single-shot properties, using the alternative lifting properties proposed in ref. 16 and considering a repair function based on the HDRG decoder. This shows that these logical states can be prepared for errors below a certain threshold with exponentially high confidence for increasing the lattice size. Finally, a detailed description of the code conditions and the full derivation of the solution described above can be found in SI Section C1.

Non-adaptive Clifford circuits with magic state injection. To redesign the quantum circuits from Theorem 1 for a noise-resilient architecture, we retain all Clifford operations, as they pose no challenges. However, addressing the non-Clifford rotations necessitates a new approach. The simplest solution is to use a gate teleportation device that incorporates the non-Clifford gate into the specified advice state. While these devices are adaptive, we must rely on non-adaptive gadgets to ensure noise resilience. Despite this constraint, we can implement these gadgets without adaptive correction. We account for the additional phases introduced by this absence and demonstrate that we can solve the ISMR problems through a more complex reduction based on the outcomes of these qupit circuits, leading to a significantly more intricate NC^0 reduction. Thus, we obtain constant-depth non-adaptive Clifford circuits that use the advice state $|T^{1/p}\rangle^{\otimes n}$, consisting of magic states $|T^{1/p}\rangle := \frac{1}{\sqrt{p}}\sum_{j=0}^{p-1} e^{\frac{i2\pi j}{p^2}}|j\rangle$, that solve the ISMR problem $\mathcal{R}_p$ with output strings of size $m = \mathcal{O}(n \cdot g^{p^3})$, on uniformly random input strings $x \in \mathbb{F}_2^n$ that satisfy the condition $\sum_{i=1}^n x_i \bmod p = 0$ with the exact correlation as in Eq. (7).

Furthermore, we demonstrate that a minimal universal gate set, which requires only the set of Clifford gates and a single magic gate, suffices. We prove that this magic gate is related to T-type qubit gates (see SI Section C2). Thus, we recover the standard noise-resilient circuit architecture, incorporating magic state injection gadgets for all qupit dimensions, thereby extending the scope of these results.

As a corollary, we also obtain the same separation for qubits. However, since these are based on Clifford gates, we obtain a direct separation between noisy QNC^0 circuits and $\mathsf{bPTC}^0(k)$ circuits without any advice states.

New switching lemma for biased PTF circuits

As mentioned, we prove a new multi-switching lemma for $\mathsf{bPTC}^0(k)$ circuits, obtaining tighter concentration bounds compared to prior work. We then use it to establish depth reduction for $\mathsf{bPTC}^0(k)$ circuits with multiple output bits. Our approach parallels the strategies employed for multi-output AC^0 circuits[17,74], suitably modified for $\mathsf{bPTC}^0(k)$ circuits. Using this tool, we analyze how $\mathsf{bPTC}^0(k)$ circuits reduce under random restrictions of the input variables, showing that we obtain *decision forests* of low complexity with high probability. These more tractable decision forests can then be compared with quantum circuits via lightcone arguments.

Tighter multi-switching lemma for $\mathsf{bPTC}^0(k)$. One of our technical contributions is a new multi-switching lemma, which shows that a finite set of depth-2 $\mathsf{bPTC}^0(k)\circ\mathsf{AND}_w$ circuits reduces to a decision forest with high probability converging to unity as $1 - O(2^{-t})$.

Our proof employs an inductive approach similar to the one used by ref. 75 for AC^0 circuits. We conduct induction over the number of variables fixed by random restrictions and the circuits f_i corresponding to each output bit, aiming to lower bound the probability with which the latter reduce to depth-l decision trees. Simultaneously, when we encounter a circuit that does *not* reduce to a decision tree of depth-l with a set of variables fixed by random restrictions, we query the variables that remain "alive" to forcefully simplify this circuit. These variables then become part of the global decision tree, sequentially growing a decision forest. For this approach to work, we address the issue of fixing variables and clauses that describe these circuits by combining our induction with canonical decision trees from the witness method (also used in ref. 26).

The above induction technique necessitates the use of *downward-closed* random restrictions, which guarantees the monotonicity of decision tree size under random restrictions. The algorithm that constructs our canonical depth-l decision trees corresponding to the leaves of the decision forest has two properties that ensure this: all random restrictions reducing the circuit to a fixed depth-l decision tree overlap in the variables that they fix; consequently, for an arbitrary random restriction, fixing more variables does not increase the depth of the decision tree to which the initial circuit reduces.

Depth reduction for $\mathsf{bPTC}^0(k)$. Adopting the proof technique of ref. 74 for AC^0 circuits, we use our new multi-switching lemma to prove a new depth reduction lemma for $\mathsf{bPTC}^0(k)$ circuits. The multi-switching lemma first reduces the depth by 1, from d to $d-1$, for a computational object DF that is a decision forest feeding the inputs to a biased polynomial threshold circuit of depth d with layers of the circuit having $s_1, \ldots, s_d$ gates each, as demonstrated in the following lemma.

Lemma 3. A random restriction reduces DF to a decision forest, with a global decision tree of equal size as DF and decision trees of depth $l \geq \log s_1 + k + 2$ at the leaves feeding inputs to a k-biased polynomial threshold circuit of depth $d-1$ with layers of the circuit having $s_2 \ldots, s_d$ gates. In particular, for restriction probability p, this reduction in complexity happens with probability at least $1 - s_1 \cdot 2^k (400wp)^{t/2}$. Here t is the depth of the global decision tree, and w the size of the decision trees at the leaves of the initial decision forest.

This key result enables us to reduce the depth of the circuit by one each time, replacing the layer that disappears with a decision forest. By iteratively applying this insight, we can completely reduce the circuit to a decision forest.

Lemma 4. Iterative application of Lemma 3 completely reduces any n-input and m-output polynomial threshold circuit of depth d and bias k to a decision forest with global decision tree depth $2t-2$, and maximum depth q of the decision trees at its leaves. If the restriction probability in step i of the iteration is p_i, then the probability that the depth reduction succeeds is at least

$$1 - \left(\sum_{i=2}^{d-1} s_i \cdot 2^k \mathcal{O}(p_i l_i)^{t/2} + 2^k m^{1/q} \mathcal{O}(p_d \cdot l_d)^t \right). \tag{9}$$

Finally, we choose the probabilities $p_1, \ldots, p_d$ for the sequence of random restrictions, and the depths $l_1, \ldots, l_d$ of the local decision trees, such that for global decision tree depth t, local decision tree depths q, initial circuit size s, and circuit type parameterized by k, we succeed in reducing the circuit to a decision forest with the probability defined informally in Lemma 1. For the full derivation of the multi-output multi-switching lemma, refer to SI Section E; concurrent work of ref. 76 gives an independent derivation of a similar multi-switching lemma with different parameters.

Data availability

No additional datasets were generated or analyzed in this study. All relevant numerical values, derived from the analytical formulas described in the text, are presented in the manuscript and available from the authors upon request.

References

1. Kim, Y. et al. Evidence for the utility of quantum computing before fault tolerance. *Nature* **618**, 500 (2023).
2. Morgado, M. & Whitlock, S. Quantum simulation and computing with Rydberg-interacting qubits. *AVS Quantum Sci.* **3**, 023501 (2021).
3. Maldonado, T. J., Flick, J., Krastanov, S. & Galda, A. Error rate reduction of single-qubit gates via noise-aware decomposition into native gates. *Sci. Rep.* **12**, 6379 (2022).
4. Wang, Y. et al. Single-qubit quantum memory exceeding ten-minute coherence time. *Nat. Photonics* **11**, 646 (2017).
5. Bluvstein, D. et al. Logical quantum processor based on reconfigurable atom arrays. *Nature* **626**, 58–65 (2023).
6. Acharya, R. et al. Quantum error correction below the surface code threshold. https://arxiv.org/abs/2408.13687 (2024).
7. Ryan-Anderson, C. et al. High-fidelity teleportation of a logical qubit using transversal gates and lattice surgery. *Science* **385**, 1327 (2024).
8. Beverland, M. E. et al. Assessing requirements to scale to practical quantum advantage https://arxiv.org/abs/2211.07629 (2022).
9. Gidney, C. & Ekerå, M. How to factor 2048 bit RSA integers in 8 hours using 20 million noisy qubits. *Quantum* **5**, 433 (2021).
10. Scherer, A. et al. Concrete resource analysis of the quantum linear-system algorithm used to compute the electromagnetic scattering cross section of a 2D target. *Quantum Inf. Process.* **16,** https://doi.org/10.1007/s11128-016-1495-5 (2017).
11. Bremner, M. J., Jozsa, R. & Shepherd, D. J. Classical simulation of commuting quantum computations implies collapse of the polynomial hierarchy. *Proc. R. Soc. A Math. Phys. Eng. Sci.* **467**, 459 (2011).
12. Aaronson, S. & Arkhipov, A. The computational complexity of linear optics. In *Proc. Forty-Third Annual ACM Symposium on Theory of Computing, STOC'11* 333–342 https://doi.org/10.1145/1993636.1993682 (Association for Computing Machinery, 2011).
13. Bouland, A., Fefferman, B., Nirkhe, C. & Vazirani, U. On the complexity and verification of quantum random circuit sampling. *Nat. Phys.* **15**, 159 (2019).
14. Bravyi, S., Gosset, D. & König, R. Quantum advantage with shallow circuits. *Science* **362**, 308 (2018).
15. Le Gall, F. Average-case quantum advantage with shallow circuits. In *Proc. 34th Computational Complexity Conference (CCC 2019), Leibniz International Proceedings in Informatics* (LIPIcs) (ed. Shpilka, A.) Vol. 137, 21:1–21:20 https://doi.org/10.4230/LIPIcs.CCC.2019.21 (Schloss Dagstuhl—Leibniz-Zentrum für Informatik, 2019).
16. Bravyi, S., Gosset, D., Koenig, R. & Tomamichel, M. Quantum advantage with noisy shallow circuits. *Nat. Phys.* **16**, 1040 (2020).
17. Watts, A. B., Kothari, R., Schaeffer, L. & Tal, A. Exponential separation between shallow quantum circuits and unbounded fan-in shallow classical circuits. In *Proc. 51st Annual ACM SIGACT Symposium on Theory of Computing* 515–526 https://doi.org/10.1145/3313276.3316404 (Association for Computing Machinery, 2019).
18. Briët, J., Buhrman, H., Castro-Silva, D. & Neumann, N. M. P. Noisy decoding by shallow circuits with parities: classical and quantum (extended abstract). In *15th Innovations in Theoretical Computer Science Conference (ITCS 2024), Leibniz International Proceedings in Informatics (LIPIcs)*, Vol. 287, 21:1–21:11 (ed. Guruswami, V.) https://doi.org/10.4230/LIPIcs.ITCS.2024.21 (Schloss Dagstuhl—Leibniz-Zentrum für Informatik, 2024).
19. Parberry, I. & Schnitger, G. Parallel computation with threshold functions. *J. Comput. Syst. Sci.* **36**, 278–302 (1988).
20. Siu, K.-Y. & Bruck, J. On the power of threshold circuits with small weights. *SIAM J. Discret. Math.* **4**, 423 (1991).
21. Minsky, M. & Papert, S. *Perceptrons: An introduction to computational geometry*, **479** (The MIT Press, 1969).
22. Muroga, S. *Threshold Logic and its Applications* (John Wiley & Sons, 1971).
23. Baldi, P. & Vershynin, R. Polynomial threshold functions, hyperplane arrangements, and random tensors. *SIAM J. Math. Data Sci.* **1**, 699–729 (2019).
24. Merrill, W., Sabharwal, A. & Smith, N. A. Saturated transformers are constant-depth threshold circuits. *Trans. Assoc. Comput. Linguist.* **10**, 843–856 (2022).
25. Merrill, W. & Sabharwal, A. The parallelism tradeoff: Limitations of log-precision transformers. *Trans. Assoc. Comput. Linguist.* **11**, 531–545 (2023).
26. Kumar, V. M. Tight correlation bounds for circuits between AC^0 and TC^0. In *Proc. 38th Computational Complexity Conference, CCC'23* https://doi.org/10.4230/LIPIcs.CCC.2023.18 (Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2023).
27. Arora, S. & Barak, B. *Computational Complexity: A Modern Approach*, 1st ed. (Cambridge University Press, 2009).
28. Chi, Y. et al. A programmable qudit-based quantum processor. *Nat. Commun.* **13**, 1166 (2022).
29. Ringbauer, M. et al. A universal qudit quantum processor with trapped ions. *Nat. Phys.* **18**, 1053 (2022).
30. Reimer, C. et al. High-dimensional one-way quantum processing implemented on d-level cluster states. *Nat. Phys.* **15**, 148 (2019).
31. Caha, L., Coiteux-Roy, X. and Koenig, R. A colossal advantage: 3d-local noisy shallow quantum circuits defeat unbounded fan-in classical circuits, arXiv:2312.09209 (2023).
32. Aasnæss, S. *Comparing Two Cohomological Obstructions for Contextuality, and a Generalised Construction of Quantum Advantage with Shallow Circuits*, Ph.D. thesis (University of Oxford, 2021).
33. Lawrence, J. Mermin inequalities for perfect correlations in many-qutrit systems. *Phys. Rev. A* **95**, 042123 (2017).
34. Dong, Y. et al. The computational advantage of MIP* vanishes in the presence of noise. In *Proc. 39th Computational Complexity Conference (CCC 2024), Leibniz International Proceedings in Informatics (LIPIcs)* (ed. Santhanam, R.) Vol. 300, 30:1–30:71 https://doi.org/10.4230/LIPIcs.CCC.2024.30 (Schloss Dagstuhl—Leibniz-Zentrum für Informatik, 2024).
35. Bombín, H. Resilience to time-correlated noise in quantum computation. *Phys. Rev. X* **6**, 041034 (2016).
36. Gross, D. Hudson's theorem for finite-dimensional quantum systems. *J. Math. Phys.* **47**, 12 (2006)
37. Howard, M., Brennan, E. & Vala, J. Quantum contextuality with stabilizer states. *Entropy* **15**, 2340 (2013).
38. Meyer, U. I., Šupić, I., Markham, D. & Grosshans, F., Bell nonlocality from wigner negativity in qudit systems. https://arxiv.org/abs/2405.14367 (2024).
39. Raussendorf, R., Bravyi, S. & Harrington, J. Long-range quantum entanglement in noisy cluster states. *Phys. Rev. A* **71**, 062313 (2005).
40. Chevignard, C., Fouque, P.-A. & Schrottenloher, A. Reducing the number of qubits in quantum factoring. *Cryptology ePrint Archive* (2024).
41. Lubinski, T. et al. Application-oriented performance benchmarks for quantum computing, *IEEE Trans Quantum Eng.* (2023).
42. Shalm, L. K. et al. Strong loophole-free test of local realism. *Phys. Rev. Lett.* **115**, 250402 (2015).
43. Rauch, D. et al. Cosmic bell test using random measurement settings from high-redshift quasars. *Phys. Rev. Lett.* **121**, 080403 (2018).
44. Gál, A., Hansen, K. A., Koucký, M., Pudlák, P. & Viola, E. Tight bounds on computing error-correcting codes by bounded-depth circuits with arbitrary gates. In *Proc. Forty-Fourth Annual ACM Symposium on Theory of Computing* 479–494 (Association for Computing Machinery, 2012).

45. Wang, Y. et al. Fault-tolerant one-bit addition with the smallest interesting color code. *Sci. Adv.* **10**, eado9024 (2024).
46. Schroeder, B., Pinheiro, E. & Weber, W.-D. Dram errors in the wild: a large-scale field study. In *Proc. Eleventh International Joint Conference on Measurement and Modeling of Computer Systems*, *SIGMETRICS'09 193–204* https://doi.org/10.1145/1555349.1555372 (Association for Computing Machinery, 2009).
47. Bharti, K. & Jain, R. On the power of geometrically-local classical and quantum circuits. arXiv:2310.01540 (2023).
48. González-Cuadra, D., Zache, T. V., Carrasco, J., Kraus, B. & Zoller, P. Hardware efficient quantum simulation of non-abelian gauge theories with qudits on rydberg platforms. *Phys. Rev. Lett.* **129**, 160501 (2022).
49. Watts, A. B. & Parham, N. Unconditional quantum advantage for sampling with shallow circuits. arXiv:2301.00995 (2023).
50. Takahashi, Y. & Tani, S. Collapse of the hierarchy of constant-depth exact quantum circuits. *Comput. Complex.* **25**, 849 (2016).
51. Barenco, A. et al. Elementary gates for quantum computation. *Phys. Rev. A* **52**, 3457 (1995).
52. Reck, M., Zeilinger, A., Bernstein, H. J. & Bertani, P. Experimental realization of any discrete unitary operator. *Phys. Rev. Lett.* **73**, 58 (1994).
53. Hu, Y., Melnyk, D., Wang, Y. & Wattenhofer, R. Space complexity of streaming algorithms on universal quantum computers. In *Proc. Theory and Applications of Models of Computation: 16th International Conference, TAMC 2020,* 275–286 (Springer, 2020).
54. Anwar, H., Brown, B. J., Campbell, E. T. & Browne, D. E. Fast decoders for qudit topological codes. *N. J. Phys.* **16**, 063038 (2014).
55. Mezher, R., Ghalbouni, J., Dgheim, J. & Markham, D. Fault-tolerant quantum speedup from constant depth quantum circuits. *Phys. Rev. Res.* **2**, 033444 (2020).
56. Paletta, L., Leverrier, A., Sarlette, A., Mirrahimi, M. & Vuillot, C. Robust sparse IQP sampling in constant depth. *Quantum* **8**, 1337 (2024).
57. Yoganathan, M., Jozsa, R. & Strelchuk, S. Quantum advantage of unitary clifford circuits with magic state inputs. *Proc. R. Soc. A* **475**, 20180427 (2019).
58. Hasegawa, A. & Le Gall, F. Quantum Advantage with shallow circuits under arbitrary corruption. In *Proc. 32nd International Symposium on Algorithms and Computation (ISAAC 2021), Leibniz International Proceedings in Informatics (LIPIcs)* Vol. 212, 74:1–7416 (eds Ahn, H.-K. & Sadakane, K. https://doi.org/10.4230/LIPIcs.ISAAC.2021.74 (Schloss Dagstuhl —Leibniz-Zentrum für Informatik, 2021).
59. Zhang, Z., Gong, W., Li, W. & Deng, D.-L. Quantum-classical separations in shallow-circuit-based learning with and without noises. *Commun Phys* **7**, 290 (2024).
60. Cerezo, M.et al. Does provable absence of barren plateaus imply classical simulability? or, why we need to rethink variational quantum computing, arXiv preprint arXiv:2312.09121 (2023).
61. Anschuetz, E. R. & Gao, X. Arbitrary polynomial separations in trainable quantum machine learning, arXiv:2402.08606 (2024).
62. Huang, H.-Y.et al. Learning shallow quantum circuits. STOC 2024: *Proceedings of the 56th Annual ACM Symposium on Theory of Computing*. 1343–1351 (2024).
63. Gao, X. & Duan, L.-M. Efficient representation of quantum many-body states with deep neural networks. *Nat. Commun.* **8**, 662 (2017).
64. Anschuetz, E. R., Hu, H.-Y., Huang, J.-L. & Gao, X. Interpretable quantum advantage in neural sequence learning. *PRX Quantum* **4**, 020338 (2023).
65. Bowles, J., Wright, V. J., Farkas, M., Killoran, N. & Schuld, M. Contextuality and inductive bias in quantum machine learning, arXiv:2302.01365 (2023).
66. Abbas, A. et al. The power of quantum neural networks. *Nat. Comput. Sci.* **1**, 403 (2021).
67. Du, Y., Hsieh, M.-H., Liu, T., You, S. & Tao, D. Learnability of quantum neural networks. *PRX Quantum* **2**, 040337 (2021).
68. Mermin, N. D. Extreme quantum entanglement in a superposition of macroscopically distinct states. *Phys. Rev. Lett.* **65**, 1838 (1990).
69. Grier, D., Ju, N. & Schaeffer, L. Interactive quantum advantage with noisy, shallow clifford circuits, arXiv:2102.06833 (2021).
70. Watson, F. H. E., Anwar, H. & Browne, D. E. Fast fault-tolerant decoder for qubit and qudit surface codes. *Phys. Rev. A* **92**, 032309 (2015).
71. Anwar, H. *Towards Fault-tolerant Quantum Computation with Higher-dimensional Systems*, Ph.D. thesis (UCL University College London, 2014).
72. Bravyi, S. & Haah, J. Quantum self-correction in the 3D cubic code model. *Phys. Rev. Lett.* **111**, https://doi.org/10.1103/physrevlett.111.200501 (2013).
73. Moussa, J. E. Transversal clifford gates on folded surface codes, *Phys. Rev. A* **94**, https://doi.org/10.1103/physreva.94.042316 (2016).
74. Rossman, B. An entropy proof of the switching lemma and tight bounds on the decision-tree size of AC^0 https://users.cs.duke.edu/~br148/logsize.pdf (2017).
75. Hastad, J. An average-case depth hierarchy theorem for higher depth. In *Proc. 2016 IEEE 57th Annual Symposium on Foundations of Computer Science (FOCS)* https://doi.org/10.1109/focs.2016.18 (IEEE, 2016).
76. Grewal, S. & Kumar, V. M. Improved circuit lower bounds and quantum-classical separations https://arxiv.org/abs/2408.16406 (2024).
77. Strobl, L., Merrill, W., Weiss, G., Chiang, D. & Angluin, D. What formal languages can transformers express? A survey. *Trans. Assoc. Comput. Linguist.* **12**, 543 (2024).
78. Bertoni, A. & Palano, B. Structural complexity and neural networks. In *Neural Nets: 13th Italian Workshop on Neural Nets, WIRN VIETRI 2002 Vietri sul Mare, Italy, May 30–June 1, 2002 Revised Papers 13* 190–216 (Springer, 2002).
79. Šíma, J. & Orponen, P. General-purpose computation with neural networks: a survey of complexity theoretic results. *Neural Comput.* **15**, 2727 (2003).
80. Parekh, O., Phillips, C. A., James, C. D. & Aimone, J. B. Constant-depth and subcubic-size threshold circuits for matrix multiplication. In *Proc. 30th on Symposium on Parallelism in Algorithms and Architectures* 67–76 (Association for Computing Machinery, 2018).

Acknowledgements

M.d.O. is supported by National Funds through the FCT—Fundação para a Ciência e a Tecnologia, I.P. (Portuguese Foundation for Science and Technology) within the project IBEX, with reference PTDC/CCI-COM/4280/2021, and via CEECINST/00062/2018 (EFG). S.S. is supported by a Royal Commission for the Exhibition of 1851 Research Fellowship. S.S. would like to thank Bruno Cavalar, Zhenjian Lu, and Ninad Rajgopal for insightful discussions.

Author contributions

The project was conceived by M.d.O., S.S., and M.-H.H. Theoretical results were proved by M.d.O. in discussion with S.S., with inputs from L.M. and M.-H.H. in parts of the proof of noise-robustness. Numerical implementations were performed by M.d.O. The first version of the manuscript was written by M.d.O. and subsequently improved by M.d.O. and S.S. with inputs from M.-H.H.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-025-58545-4.

Correspondence and requests for materials should be addressed to Michael de Oliveira, Sathyawageeswar Subramanian or Min-Hsiu Hsieh.

Peer review information *Nature Communications* thanks the anonymous reviewer(s) for their contribution to the peer review of this work. A peer review file is available.