2025.10.11 / Seoul, South Korea / Complexity Theory Study

The Complexity of NISQ

SungBin Lee

Dept. of Physics and Astronomy, Seoul National University

$$\mathsf{BPP} \subsetneq \mathsf{NISQ} \subsetneq \mathsf{BQP}$$

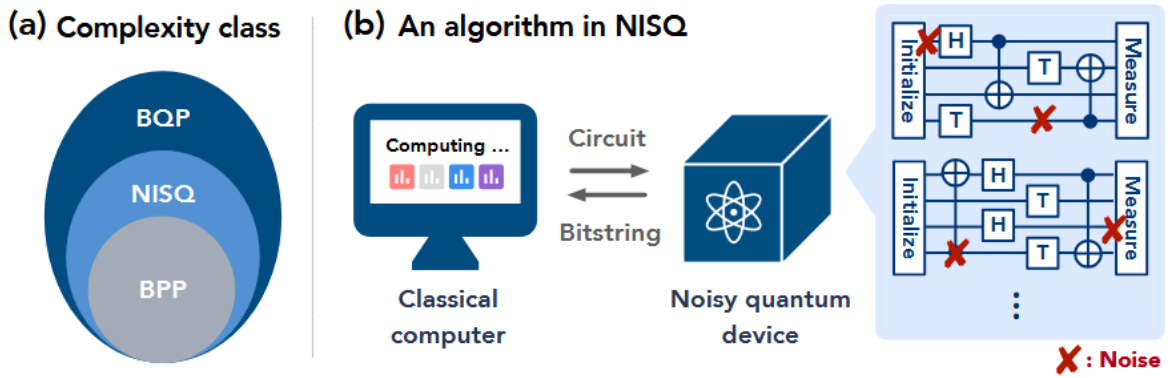
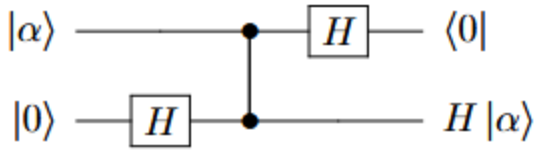


Figure 1: *Illustration of the* NISQ *complexity class:* (a) Complexity classes: NISQ contains problems that can be solved by classical computation (BPP), and some problems that can be solved by quantum computation (BQP). (b) NISQ algorithm: An algorithm in the complexity class NISQ is modeled by a hybrid quantum-classical algorithm, where a classical computer can specify the circuit to run on a noisy quantum device and the device would run a noisy version of the circuit and return a random classical bitstring obtained from noisy measurement.

What I already knew before this presentation

$$\mathrm{P} \subseteq \mathrm{BPP} \subseteq \mathrm{BQP} \subseteq \mathrm{PP} = \mathrm{PostBQP} = \mathrm{PostIQP} = \mathrm{PostQAOA}$$

pf) Any quantum circuit can be representated in H, T, CZ
If postselection is allowed, we can replace midcircuit H into



$$\mathrm{BPP} \subseteq \Sigma_2^{\mathrm{P}} \cap \Pi_2^{\mathrm{P}}, \text{ if } \mathrm{NP} \subseteq \mathrm{BPP} \rightarrow \mathrm{PH} \text{ collapses}$$

$$\Sigma_1^{\mathrm{P}} \subseteq \mathrm{PostBPP} \subseteq \mathrm{BPP}^{\mathrm{NP}} \subseteq \Sigma_3^{\mathrm{P}} \rightarrow \Sigma_2^{\mathrm{P}} \subseteq \mathrm{P}^{\mathrm{PostBPP}} \subseteq \Sigma_4^{\mathrm{P}} \subseteq \mathrm{PH} \subseteq \mathrm{P}^{\mathrm{PP}} = \mathrm{P}^{\#\mathrm{P}} = \mathrm{P}^{\mathrm{PostBQP}}$$

$$\mathrm{P}^{\mathrm{Apx_c}\#\mathrm{P}} \subseteq \Sigma_3^{\mathrm{P}} \subseteq \mathrm{PH} \subseteq \mathrm{P}^{\#\mathrm{P}} = \mathrm{P}^{\mathrm{Apx_c GapP}} = \mathrm{P}^{\mathrm{GapP}}$$

The Complexity of NISQ

The Complexity of NISQ

Sitan Chen* UC Berkeley
Jordan Cotler† Harvard University
Hsin-Yuan Huang‡ Caltech
Jerry Li§ Microsoft Research

October 14, 2022

Abstract

The recent proliferation of NISQ devices has made it imperative to understand their computational power. In this work, we define and study the complexity class NISQ, which is intended to encapsulate problems that can be efficiently solved by a classical computer with access to a NISQ device. To model existing devices, we assume the device can (1) noisily initialize all qubits, (2) apply many noisy quantum gates, and (3) perform a noisy measurement on all qubits. We first give evidence that BPP $\subsetneq$ NISQ $\subsetneq$ BQP, by demonstrating super-polynomial oracle separations among the three classes, based on modifications of Simon's problem. We then consider the power of NISQ for three well-studied problems. For unstructured search, we prove that NISQ cannot achieve a Grover-like quadratic speedup over BPP. For the Bernstein-Vazirani problem, we show that NISQ only needs a number of queries logarithmic in what is required for BPP. Finally, for a quantum state learning problem, we prove that NISQ is exponentially weaker than classical computation with access to noiseless constant-depth quantum circuits.

What is NISQ complexity class?

1. **Noisy quantum gates.** The device can execute noisy two-qubit logic gates. Using quantum logic gates (as opposed to, say, more general non-unitary CPTP maps) is standard in existing quantum devices, and it is well-understood that in real-world settings, they will be subject to noise. For concreteness, we consider the standard model of local depolarizing noise per qubit. However, our results extend to more general noise models; see Remarks A.4, C.19 and E.5.
2. **Noisy state preparation at the start.** The quantum devices have a fixed number of qubits and as such cannot bring in fresh qubits during the computation. This means that the device must prepare all qubits at the start. Notably, since we assume all quantum gates are subject to noise, this means all qubits will accrue entropy throughout the computation.
3. **Noisy measurement at the end.** The quantum devices are limited to perform noisy measurements only at the end of the computation, which means the measurement is performed on all qubits simultaneously. From a physical perspective, this constraint arises due to the difficulty of isolating subsets of qubits and measuring them without decohering the residual qubits.

Definition 2.1 (NISQ complexity class, informal)**.** NISQ *contains all problems that can be solved by a polynomial-time probabilistic classical algorithm with access to a noisy quantum device. To solve a problem of size n, the classical algorithm can access a noisy quantum device that can:*

1. *Prepare a noisy* poly(n)*-qubit all-zero state;*
2. *Apply arbitrarily many layers of noisy two-qubit gates;*
3. *Perform a noisy computational basis measurements on all the qubits simultaneously.*

All quantum operations are subject to a constant amount of depolarizing noise per qubit.

Inclusion vs Strict inclusion

$\mathrm{BPP} \subsetneq \mathrm{NISQ} \leftrightarrow$ NISQ have super-polynomial speedup over classical algorithms

$\mathrm{NISQ} \subsetneq \mathrm{BQP} \leftrightarrow$ NISQ not as powerful as FTQC

c.f. Simon's problem; for $f : \{0,1\}^n \to \{0,1\}^n, f(x)$ is 2-to-1 with $f(x) = f(x \oplus s)$ or 1-to-1

There is a classical oracle O_1 s.t. $\mathrm{BPP}^{O_1} \subsetneq \mathrm{NISQ}^{O_1}$

"Robustified" $\tilde{f} : \{0,1\}^{n'} \to \{0,1\}^n, n' \gg n, \forall x \in \{0,1\}^n, \exists A_x \subset \{0,1\}^{n'}$ s.t. $\tilde{f}(z) = f(x)\ \forall z \in A_x$

$\tilde{f}$ is robust to noise, allows NISQ to achieve super–polynomial speed up

There is a classical oracle O_2 s.t. $\mathrm{NISQ}^{O_2} \subsetneq \mathrm{BQP}^{O_2}$

"Lifted" $\tilde{f}$, gives exponentially little information to any noisy access

Requires exponentially more queries than FTQC. In fact, $\mathrm{NISQ}^{O_2} \subsetneq \mathrm{BPP}^{\mathrm{QNC}^0}$

$\mathrm{BPP}^{\mathrm{QNC}[f(n)]}$: depth-$f(n)$ FTQC, $\mathrm{BPP}^{\mathrm{QNC}^i}$ when $f(n) = O(\log^i(n)) \to \mathrm{BPP}^{\mathrm{QNC}} \subseteq \mathrm{BQP}$

Query complexity on more practical problems

Problem	BPP	BQP	NISQ_λ
Grover: $f(x) = 1$ when $x = w$(unknown), $f(x) = 0$ rest	$O(2^n)$	$O(2^{n/2})$	at least $O(\lambda 2^n)$
Bernstein-Vazrani: $f(x) = x \cdot s$ for unknown s	$O(n)$	$O(1)$	at most $O\left(\frac{\log n}{1-24\lambda}\right)$
Shadow tomography: learn $\lvert\mathrm{tr}(\rho P)\rvert$ up to const error	$O(4^n)$	at most $O(n)$	at least $O(e^{\lambda n})$

NISQ complexity class

A.1 Definition of the complexity class

We begin by recalling the single-qubit depolarizing channel D_λ.

Definition A.1 (Single-qubit depolarizing channel). *Given $\lambda \in [0,1]$. We define the* single-qubit depolarizing channel *to be $D_\lambda[\rho] \triangleq (1-\lambda)\rho + \lambda(I/2)$, where ρ is a single-qubit density matrix.*

Definition A.2 (Depth-1 unitary). *Given $n > 0$. An n-qubit unitary U is a* depth-1 unitary *if U can be written as a tensor product of two-qubit unitaries.*

We consider noisy quantum circuits with noise level λ to be defined as follows.

Definition A.3 (Output of a noisy quantum circuit). *Let $\lambda \in [0,1]$ and $n \in \mathbb{N}$. Given $T \in \mathbb{N}$ and a sequence of T depth-1 unitaries $U_1, \ldots, U_T$, the output of the corresponding* λ-noisy depth-T quantum circuit *is a random n-bit string $s \in \{0,1\}^n$ sampled from the distribution*

$$p(s) = \langle s | \, D_\lambda^{\otimes n}\left[U_T \ldots D_\lambda^{\otimes n}\left[U_2 D_\lambda^{\otimes n}\left[U_1 D_\lambda^{\otimes n}[|0^n\rangle\langle 0^n|]U_1^\dagger\right]U_2^\dagger\right]\ldots U_T^\dagger\right]|s\rangle \ , \tag{1}$$

where every quantum operation is followed by a layer of single-qubit depolarizing channel. When $\lambda = 0$, we say that this circuit is noiseless.

NISQ complexity class

Remark A.4. *We work with the single-qubit depolarizing channel as it is the most standard model for local noise. One could also consider stronger noise models, e.g. every qubit is randomly corrupted with probability λ by an adversary rather than randomly decohered. Tautologically, the lower bounds we prove in this work will translate to such stronger models. We also prove our upper bounds, namely Theorem 2.2 and 2.5, under this stronger model (see Remarks C.19 and E.5).*

Definition A.5 (Noisy quantum circuit oracle)**.** *We define NQC_λ to be an oracle that takes in an integer n and a sequence of depth-1 n-qubit unitary $\{U_k\}_{k=1,\dots,T}$ for any $T \in \mathbb{N}$ and outputs a random n-bit string s according to Eq. (1).*

We define the time to query NQC_λ with T depth-1 n-qubit unitaries to be $\Theta(nT)$, which is linear in the time to write down the input to the query.

We now define NISQ algorithms, which are classical algorithms with access to the noisy quantum circuit oracle. This provides a formal definition for hybrid noisy quantum-classical computation.

Definition A.6 (NISQ algorithm)**.** *A NISQ_λ algorithm with access to λ-noisy quantum circuits is defined as a probabilistic Turing machine M that can query NQC_λ to obtain an output bitstring s for any number of times, and is denoted as $A_\lambda \triangleq M^{\mathrm{NQC}_\lambda}$. The runtime of A_λ is given by the classical runtime of M plus the sum of the times to query NQC_λ.*

NISQ complexity class

We now define NISQ algorithms, which are classical algorithms with access to the noisy quantum circuit oracle. This provides a formal definition for hybrid noisy quantum-classical computation.

Definition A.6 (NISQ algorithm)**.** *A* NISQ_λ *algorithm with access to* λ*-noisy quantum circuits is defined as a probabilistic Turing machine* M *that can query* NQC_λ *to obtain an output bitstring* s *for any number of times, and is denoted as* $A_\lambda \triangleq M^{\mathrm{NQC}_\lambda}$*. The runtime of* A_λ *is given by the classical runtime of* M *plus the sum of the times to query* NQC_λ*.*

The NISQ complexity class for decision problems is defined as follows. Observe that the following recovers the definition for BPP when M^{NQC_λ} in the definition of A_λ above is replaced by M.

Definition A.7 (NISQ complexity)**.** *A language* $L \subseteq \{0,1\}^*$ *is in* NISQ *if there exists a* NISQ_λ *algorithm* A_λ *for some constant* $\lambda > 0$ *that decides* L *in polynomial time, that is, such that*

- *for all* $x \in \{0,1\}^*$, A_λ *produces an output in time* $\mathrm{poly}(|x|)$*, where* $|x|$ *is the length of* x*;*
- *for all* $x \in L$, A_λ *outputs* 1 *with probability at least* $2/3$*;*
- *for all* $x \notin L$, A_λ *outputs* 0 *with probability at least* $2/3$*.*

Classical & Quantum oracle

Definition A.8 (Classical oracle O). *A* classical oracle O *is a function from* $\{0,1\}^n$ *to* $\{0,1\}^m$ *for some* $n, m \in \mathbb{N}$. *The* $(n+m)$*-qubit unitary* U_O *corresponding to the classical oracle* O *is given by* $U_O |x\rangle |y\rangle = |x\rangle |y \oplus O(x)\rangle$ *for all* $x \in \{0,1\}^n, y \in \{0,1\}^m$.

Definition A.9 (Classical algorithm with access to O). *A* classical algorithm M^O with access to O *is a probabilistic Turing machine* M *that can query* O *by choosing an* n*-bit input* x *and obtaining the* m*-bit output* $O(x)$.

Definition A.10 (Quantum algorithm with access to O). *A* quantum algorithm Q^O with access to O *is a uniform family of quantum circuits* $\{U_n\}_n$, *where* U_n *is an* n'*-qubit quantum circuit given by*

$$U_n \triangleq V_{n,k}(U_O \otimes I) \cdots (U_O \otimes I) V_{n,2}(U_O \otimes I) V_{n,1} \,,$$

for some integer $k \in \mathbb{N}$ *and* n'*-qubit unitaries* $V_{n,1}, \ldots, V_{n,k}$ *given as the product of many depth-*1 *unitaries. Here,* I *denotes the identity matrix over* $n' - n$ *qubits.*

Classical & Quantum oracle

Definition A.11 (Noisy quantum circuit oracle with access to O). *We define* NQC_λ^O *to be an oracle that takes in an integer* n' *and a sequence of* n'*-qubit unitaries* $\{U_k\}_{k=1,\dots,T}$ *for any* $T \in \mathbb{N}$*, where* U_k *can either be a depth-1 unitary or* $U_O \otimes I$*, to a random* n*-bit string* s *sampled according to the distribution*

$$p(s) = \langle s | \, D_\lambda^{\otimes n'} \left[U_T \dots D_\lambda^{\otimes n'} \left[U_2 D_\lambda^{\otimes n'} \left[U_1 D_\lambda^{\otimes n'} [|0^{n'}\rangle\langle 0^{n'}|] U_1^\dagger \right] U_2^\dagger \right] \dots U_T^\dagger \right] |s\rangle \, .$$

Definition A.12 (NISQ algorithm with access to O). *Let* $\lambda \in [0,1]$*. A* NISQ_λ *algorithm* $A_\lambda^O = (M^{\mathrm{NQC}_\lambda})^O$ *with access to* O *is a probabilistic Turing machine* M *that has the ability to classically query* O *by choosing the* n*-bit input* x *to obtain the* m*-bit output* $O(x)$*, as well as the ability to query* NQC_λ^O *by choosing* n' *and* $\{U_k\}_{k=1,\dots,T}$ *to obtain a random* n'*-bit string* s*. The runtime of* A_λ^O *is given by the sum of the classical runtime of* M*, the number of classical queries to* O*, and the sum of the times to query* NQC_λ^O*.*

With this definition in hand, we can extend the usual notions of relativized complexity to NISQ:

Definition A.13 (Relativized NISQ). *Given a sequence of oracles* $O : \{0,1\}^n \to \{0,1\}^{m(n)}$ *parametrized by* $n \in \mathbb{N}$*, a language* $L \subseteq \{0,1\}^*$ *is in* NISQ^O *if there exists a constant* $\lambda > 0$ *and a* NISQ_λ *algorithm* A_λ^O *with access to* O *that decides* L *in polynomial time.*

Noisy circuits and noiseless bounded-depth circuits

Definition A.14 (Noiseless hybrid quantum-classical computation of bounded depth)**.** *A* noiseless depth-T algorithm *is a* NISQ_0 *algorithm* A *that only queries* NQC_0 *on sequences of depth-1* n*-qubit unitaries* $\{U_k\}_{k=1,\ldots,T'}$ *for* $1 \leq T' \leq T$.

Definition A.15 ($\mathsf{BPP}^{\mathsf{QNC}}$)**.** *Let* $f : \mathbb{N} \to \mathbb{N}$ *be a nondecreasing function. A language* $L \subseteq \{0,1\}^*$ *is in* $\mathsf{BPP}^{\mathsf{QNC}[f(n)]}$ *if there is a noiseless depth-*$f(n)$ *algorithm* A *that decides* L *in polynomial time. When* $f(n) = O(\log^i(n))$, *we denote this class by* $\mathsf{BPP}^{\mathsf{QNC}^i}$. *We also define* $\mathsf{BPP}^{\mathsf{QNC}} \triangleq \cup_{i \geq 0} \mathsf{BPP}^{\mathsf{QNC}^i}$.

Note that $\mathsf{BPP}^{\mathsf{QNC}}$ is contained in the class BQP, as BQP can implement arbitrary *polynomial*-depth quantum computation.

We can also define noiseless depth-T algorithms with access to a classical oracle, as well as relativized versions of $\mathsf{BPP}^{\mathsf{QNC}^i}$ which we denote by $(\mathsf{BPP}^{\mathsf{QNC}^i})^O$, completely analogously to what is done in Section A.2.

Tree representation

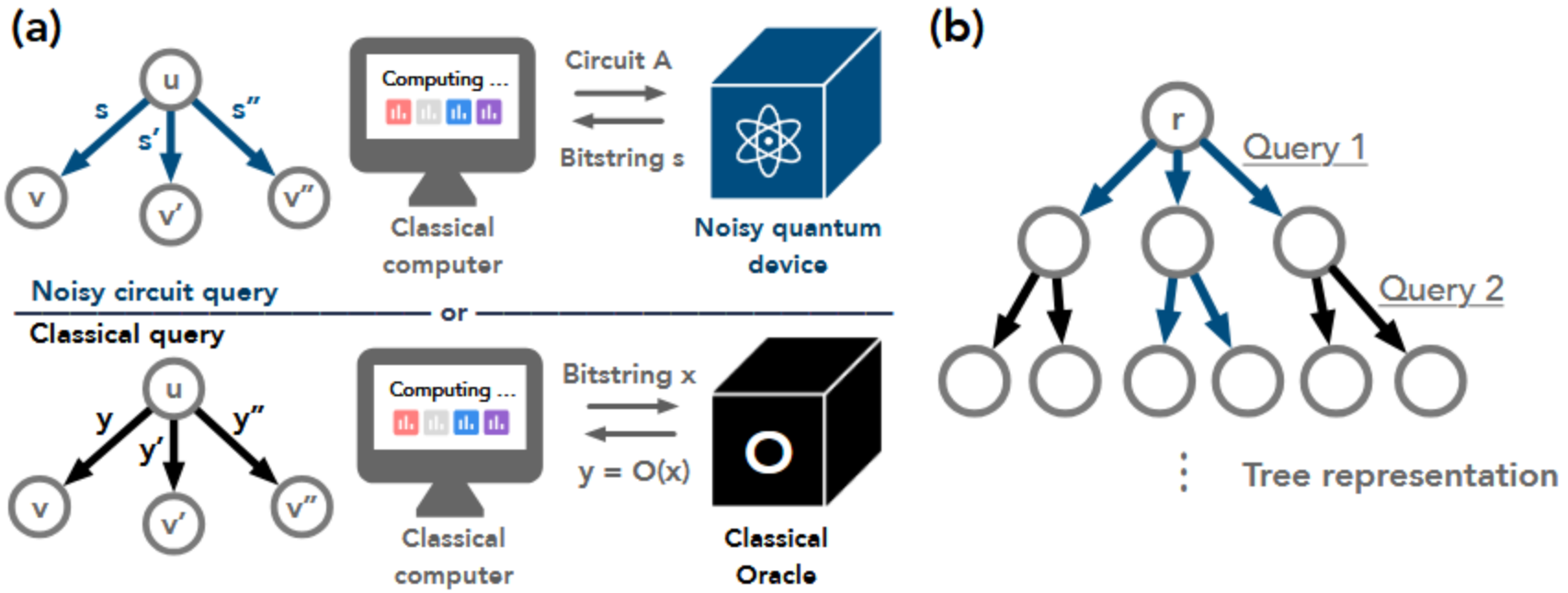


Figure 2: Illustration of the tree representation for NISQ algorithms. (a) At every memory state u of the classical computer/algorithm, it could either make a noisy circuit query or a classical query. (b) The tree representation with a mix of noisy circuit queries and classical queries.

Tree representation

Definition B.1 (Tree representation for NISQ algorithms). *Given oracle $O : \{0,1\}^n \to \{0,1\}^m$, a NISQ_λ algorithm with access to O can be associated with a pair $(\mathcal{T}, \mathcal{A})$ as follows. The* learning tree *$\mathcal{T}$ is a rooted tree, where each node in the tree encodes the transcript of all classical query and noisy quantum circuit results the algorithm has seen so far. The tree satisfies the following properties:*

- *Each node u is associated with a value $p_O(u)$ corresponding to the probability that the transcript observed so far is given by the path from the root r to u. In this way, $\mathcal{T}$ naturally induces a distribution over its leaves. For the root r, $p_O(r) = 1$.*
- *At each non-leaf node u, we either classically query the oracle O at an input $x \in \{0,1\}^n$, or run a λ-noisy quantum circuit A with access to O.*

  (i) Classical query: u has a single child node v connected via an edge $(u, x, O(x))$, and we define

  $$p_O(v) = p_O(u)\,.$$

  (ii) Noisy circuit query: The children v of u are indexed by the possible $s \in \{0,1\}^{n'}$ that could be obtained as a result. We refer to the edge between u and v as (u, A, s). We denote by $|\phi_O(A)\rangle$ the output state of the circuit so that the probability of traversing (u, A, s) from node u to child v is given by $|\langle s|\phi_O(A)\rangle|^2$. We define

  $$p_O(v) = p_O(u) \cdot |\langle s|\phi_O(A)\rangle|^2\,.$$

- *If the total number of classical/quantum queries to O made along any root-to-leaf path is at most N, we say that the* query complexity *of the algorithm is at most N.*

Tree representation and upper bound of TVD

Lemma B.2. *Given learning tree $\mathcal{T}$ corresponding to a NISQ_λ algorithm with query complexity N, suppose $\mathcal{T}'$ is a learning tree obtained from $\mathcal{T}$ as follows. For every node u at which a noisy quantum circuit A is run, replace A by another circuit A' such that the new induced distribution over children of u is at most ε-far from the original distribution in total variation. Then the distributions over leaves of $\mathcal{T}$ and $\mathcal{T}'$ are at most εN-far in total variation.*

Proof. Consider the sequence of trees $\mathcal{T}^{(i)}$ where $\mathcal{T}^{(0)} = \mathcal{T}$ and $\mathcal{T}^{(i)}$ is given by taking all u in layer i of $\mathcal{T}^{(i-1)}$ that run some noisy quantum circuit A and replacing them with the corresponding circuit A' from $\mathcal{T}'$. By design, $\mathcal{T}^{(N)} = \mathcal{T}'$. Let $p^{(i)}$ denote the distribution over leaves of $\mathcal{T}^{(i)}$. It suffices to show that $d_{\mathrm{TV}}(p^{(i)}, p^{(i-1)}) \leq \varepsilon$.

Note that $p^{(i-1)}$ specifies some mixture over distributions p_v, where p_v is the distribution over leaves conditioned on reaching node v in the i-th layer. In particular, in this mixture, v is sampled by sampling parent node u by running the NISQ algorithm corresponding to $\mathcal{T}'$ for $i-1$ steps and then running the corresponding quantum circuit A from $\mathcal{T}$. In contrast, $p^{(i)}$ is a mixture over the same distributions p_v, but v is sampled by running the NISQ algorithm corresponding to $\mathcal{T}'$ for i steps and then running the corresponding quantum circuit A' from $\mathcal{T}'$. These two distributions over v are at most ε-far in total variation, so the two mixture distributions are also at most ε-far in total variation as claimed. □

Le Cam's theorem: complexity and TVD

Our lower bounds will be based on Le Cam's method– see Section 4.3 of [48] for an overview in the context of the tree formalism of Definition B.1. In every case we will reduce to some *distinguishing task* in which the algorithm must discern whether the oracle it has access to comes from one family of oracles or from another. For example, for unstructured search, the distinguishing task will be whether the oracle corresponds to some element in the search domain or whether the oracle is the identity channel.

Lemma B.3 (Le Cam's two-point method, see e.g. Lemma 4.14 from [48]). *Let* $\{O_i\}_{i\in S_0}$ *and* $\{O_i\}_{i\in S_1}$ *be two disjoint sets of oracles. Given a tree* $\mathcal{T}$ *as in Definition B.1 corresponding to a* NISQ *algorithm that makes* N *oracle queries, let* p_i *denote the induced distribution over leaves when the algorithm has access to* O_i*. If* $d_{TV}(\mathbb{E}_{i\sim D_0}[p_0], \mathbb{E}_{i\sim D_1}[p_1]) < 1/3$*, there is no algorithm* $\mathcal{A}$ *that maps transcripts* T *corresponding to leaves of* $\mathcal{T}$ *to* $\{0,1\}$ *which can distinguish between* S_0 *and* S_1 *with advantage* $1/3$.[1]

Quantum query complexity lower bounds via hybrid argument

B.2 Basic hybrid argument

Here we describe a standard template for showing quantum query complexity lower bounds via a hybrid argument.

Lemma B.4. *Let $\mathcal{E}_0, \mathcal{E}_1$ be quantum channels on n qubits such that for all pure states σ, we have $\|(\mathcal{E}_0 - \mathcal{E}_1)[\sigma]\|_{\mathsf{tr}} \leq \varepsilon$. Let A be any depth-T quantum circuit with access to one of the two channels, and let $s \in \{0,1\}^n$ be the random string output by the circuit. Let p_0, p_1 denote the distribution over s when A has access to $\mathcal{E}_0, \mathcal{E}_1$ respectively. Then $d_{TV}(p_0, p_1) \leq \varepsilon T$.*

Quantum query complexity lower bounds via hybrid argument

Proof. Let $\mathcal{E} = \mathcal{E}_s$ for $s \in \{0,1\}$, and define the channel $\mathcal{U}_i$ which acts by $\mathcal{U}_i(\sigma) = U_i \sigma U_i^\dagger$ where U_i is an associated unitary operator. We proceed via a hybrid argument. The output state of the circuit is given by

$$\sigma^s = \mathcal{U}_T \circ \mathcal{E} \circ \cdots \circ \mathcal{U}_2 \circ \mathcal{E} \circ \mathcal{U}_1[|0^n\rangle\langle 0^n|]$$

for some unitaries $U_1, \ldots, U_T$. For $s' = 1 - s$ and $1 \leq i \leq T$ define

$$\sigma^{(i)} \triangleq \mathcal{U}_T \circ \mathcal{E}_s \circ \cdots \circ \mathcal{U}_{i+1} \circ \mathcal{E}_s \circ \mathcal{U}_i \circ \mathcal{E} \circ \cdots \circ \mathcal{U}_2 \circ \mathcal{E} \circ \mathcal{U}_1[|0^n\rangle\langle 0^n|]\,.$$

Then

$$\begin{aligned}\|\sigma^s - \sigma^{s'}\|_{\mathsf{tr}} &= \left\|\sum_{i=1}^{T} \sigma^{(i)} - \sigma^{(i-1)}\right\|_{\mathsf{tr}} \leq \sum_{i=1}^{T} \|\sigma^{(i)} - \sigma^{(i-1)}\|_{\mathsf{tr}} \\ &\leq \sum_{i=1}^{T} \|(\mathcal{E} - \mathcal{E}_s) \circ \mathcal{U}_{i-1} \circ \mathcal{E} \circ \cdots \circ \mathcal{U}_2 \circ \mathcal{E} \circ \mathcal{U}_1[|0^n\rangle\langle 0^n|]\|_{\mathsf{tr}} \leq T \sup_{\sigma} \|(\mathcal{E} - \mathcal{E}_s)[\sigma]\|_{\mathsf{tr}},\end{aligned}$$

where the supremum is over all density matrices. By convexity of the trace norm, this bound still holds when the supremum is restricted to pure states σ. By assumption, the above quantity is εT. The total variation distance between p_1 and p_2 as defined in lemma statement is simply the L_1 distance between the diagonals of σ^s and $\sigma^{s'}$, which is upper bounded by $\|\sigma^s - \sigma^{s'}\|_{\mathsf{tr}} \leq \varepsilon T$. □

Super-polynomial oracle separation between BPP and NISQ

Theorem C.1 (Restatement of Theorem 2.2). $\mathsf{BPP}^{O_1} \subsetneq \mathsf{NISQ}^{O_1}$ *relative to a classical oracle* O_1.

Our basic strategy is to modify the Simon's oracle into a new classical oracle such that the new oracle is robust to noise. We note that a NISQ algorithm is unable to implement known fault-tolerant quantum computation schemes that can run for any arbitrary quantum circuit with a polynomial number of gates. However, we will still take inspiration from a fault-tolerant quantum computation scheme [53] to define a certain "robustified Simon's oracle" relative to which we obtain a super-polynomial separation between BPP and NISQ. As we will show, because the fault-tolerant scheme of [53] is robust not just to local depolarizing noise but to arbitrary local noise occurring with sufficiently small constant rate, the NISQ algorithm that we give will ultimately be robust under this stronger noise model as well (see Remark C.19).

“Robustified” Simon’s problem

C.1.2 Robustified Simon’s problem

Given a large enough integer n, we consider Simon’s problem over $n' = 2^{\Theta(\log(n)^c)}$ bits for a constant $0 < c < 1$. Here $1/c$ corresponds to the constant c_2 from Theorem 10 of [53]. We consider $r = \Theta(\log\log(n'))$ and encode each of the n' bits using m^r bits. Because $m = \mathcal{O}(1)$, we have $m^r n' = 2^{\Theta(\log(n)^c)} < n$ for large enough n.

Given a classical function $f_s : \{0,1\}^{n'} \to \{0,1\}^{n'}$ from Simon’s problem with secret string $s \in \{0,1\}^{n'}$, we define a classical function $\widetilde{f}_s : \{0,1\}^n \to \{0,1\}^{m^r n'}$ as follows. Let x be an n-bit string. We focus on the first $m^r n'$ bits of x and divide them into n' m^r-bit strings as $x_1, \ldots, x_{n'}$. We first define $\widetilde{f}_s^0 : \{0,1\}^n \to \{0,1\}^{n'}$ as follows,

$$\widetilde{f}_s^0(x) \triangleq \begin{cases} f_s(b_1 \ldots b_{n'}), & \text{if } \exists!\, b_1, \ldots, b_{n'} \in \{0,1\}, \text{ s.t. } x_i \in A_{b_i}^{(r)}, \forall i = 1, \ldots, n', \\ 0^{n'}, & \text{otherwise} \end{cases}. \tag{7}$$

We use $\exists!$ to denote “there exists a unique choice”. Because $A_0^{(r)}$ and $A_1^{(r)}$ are disjoint, there either exists a unique choice of $b_1, \ldots, b_{n'}$ or does not exist any choice of $b_1, \ldots, b_{n'}$ that satisfies $x_i \in A_{b_i}^{(r)}, \forall i = 1, \ldots, n'$. Letting $[\widetilde{f}_s^0(x)]_k$ denote the kth bit of $\widetilde{f}_s^0(x)$, we define the function $\widetilde{f}_s : \{0,1\}^n \to \{0,1\}^{m^r n'}$ by

“Robustified” Simon’s problem

We use $\exists!$ to denote “there exists a unique choice”. Because $A_0^{(r)}$ and $A_1^{(r)}$ are disjoint, there either exists a unique choice of $b_1, \ldots, b_{n'}$ or does not exist any choice of $b_1, \ldots, b_{n'}$ that satisfies $x_i \in A_{b_i}^{(r)}, \forall i = 1, \ldots, n'$. Letting $[\widetilde{f}_s^0(x)]_k$ denote the kth bit of $\widetilde{f}_s^0(x)$, we define the function $\widetilde{f}_s : \{0,1\}^n \to \{0,1\}^{m^r n'}$ by

$$\widetilde{f}_s(x) \triangleq \left(\underbrace{[\widetilde{f}_s^0(x)]_1, \ldots, [\widetilde{f}_s^0(x)]_1}_{m^r \text{ times}}, \ldots, \underbrace{[\widetilde{f}_s^0(x)]_{n'}, \ldots, [\widetilde{f}_s^0(x)]_{n'}}_{m^r \text{ times}} \right) \in \{0,1\}^{m^r n'}. \tag{8}$$

The function $\widetilde{f}_s$ can be considered as the robust version of f_s, where the output bitstring is stable over a large number of bitstrings.

Let $U_{\widetilde{f}_s}$ be the unitary from Eq. (9).

$$U_{\widetilde{f}_s} |x\rangle |y\rangle = |x\rangle |y \oplus \widetilde{f}_s(x)\rangle, \quad \forall x \in \{0,1\}^n, y \in \{0,1\}^{m^r n'}. \tag{9}$$

We denote by $O_{\widetilde{f}_s}$ the oracle which applies this unitary. Then we have the following theorem, which is the main result of this section and implies Theorem 2.2:

Oracle separation between BPP and NISQ

We denote by $O_{\tilde{f}_s}$ the oracle which applies this unitary. Then we have the following theorem, which is the main result of this section and implies Theorem 2.2:

Theorem C.7. *For λ sufficiently small, there is a NISQ_λ algorithm which, given oracle access to $O_{\tilde{f}_s}$, can determine whether f_s is 2-to-1 or 1-to-1 with constant advantage in time at most $\mathcal{O}(\mathrm{poly}(n))$. By contrast, any classical algorithm with access to $O_{\tilde{f}_s}$ requires at least $\Omega(\mathrm{superpoly}(n))$ time, to determine whether f_s is 2-to-1 or 1-to-1 with constant advantage. Thus, relative to oracles O of this form, $\mathsf{BPP}^O \subsetneq \mathsf{NISQ}^O$.*

CSS Code and Codeword

C.1.1 Recursively-defined concatenated code

We consider a Calderbank-Shor-Steane (CSS) code built from two classical linear codes C_1, C_2, where $C_1 \triangleq C$ is a punctured doubly-even self-dual code and $C_2 \triangleq C^\perp$ (we refer the reader to [53] for background on these notions). We consider C_1, C_2 to be over m classical bits. The corresponding CSS code encodes a single logical qubit into m physical qubits. Let $\mathbf{1}_m$ denote the all-ones vector of length m (when the subscript is clear from context, we will omit it). The two code words in the CSS code are given by

$$|S_0\rangle = \frac{1}{\sqrt{|C^\perp|}} \sum_{w \in C^\perp} |w\rangle\,, \quad |S_1\rangle = \frac{1}{\sqrt{|C^\perp|}} \sum_{w \in C^\perp} |w \oplus \mathbf{1}_m\rangle\,, \tag{2}$$

where $\oplus$ denotes addition over $\mathbb{Z}_2^m$ (i.e., it is the bit-wise XOR). Denote by d the number of errors that can be corrected by the CSS code. The two parameters m and d are both considered to be constant. We define

$$A_0 \triangleq \left\{ w \oplus x \,\middle|\, w \in C^\perp, x \in \{0,1\}^m, |x| \leq d \right\} \tag{3}$$

$$A_1 \triangleq \left\{ w \oplus x \,\middle|\, w \in C^\perp \oplus \mathbf{1}, x \in \{0,1\}^m, |x| \leq d \right\}, \tag{4}$$

where $C^\perp \oplus \mathbf{1}$ denotes the set $\{x \oplus \mathbf{1} \,|\, x \in C^\perp\}$ and $|x|$ is the number of 1's in x.

Lemma C.2 (Disjointness of A_0 and A_1). *With the above definitions, we have*

$$A_0 \cap A_1 = \emptyset. \tag{5}$$

CSS Code and Codeword

Definition C.3 (Basis of the concatenated code). *For* $r = 1$, $B_0^{(1)} \triangleq C^\perp$ *and* $B_1^{(1)} \triangleq C^\perp \oplus \mathbf{1}$. *For* $r > 1$, *we define* $B_0^{(r)}, B_1^{(r)}$ *recursively,*

$$B_0^{(r)} \triangleq \left\{ (v_1, \ldots, v_m) \in \{0,1\}^{m^r} \,\middle|\, w \in C^\perp, v_i \in B_{w_i}^{(r-1)}, \forall i = 1, \ldots, m \right\},$$
$$B_1^{(r)} \triangleq \left\{ (v_1, \ldots, v_m) \in \{0,1\}^{m^r} \,\middle|\, w \in C^\perp \oplus \mathbf{1}, v_i \in B_{w_i}^{(r-1)}, \forall i = 1, \ldots, m \right\}.$$

The two code words in the recursively-defined concatenated code are then given by

$$|R_b\rangle = \frac{1}{\sqrt{|B_b^{(r)}|}} \sum_{x \in B_b^{(r)}} |x\rangle, \quad b \in \{0,1\}. \tag{6}$$

For each r, we also define two sets $A_0^{(r)}, A_1^{(r)}$ over m^r-bit strings that correspond to the neighborhoods around $B_0^{(r)}, B_1^{(r)}$ induced by errors.

Definition C.4 (Neighborhood of $B_0^{(r)}, B_1^{(r)}$). *For* $r = 1$, $A_0^{(1)} \triangleq A_0$ *and* $A_1^{(1)} \triangleq A_1$. *By Eq.* (5), $A_0^{(r)} \cap A_1^{(r)} = \emptyset$. *For* $r > 1$, *we define* $A_0^{(r)}, A_1^{(r)}$ *recursively,*

$$A_0^{(r)} \triangleq \left\{ (v_1, \ldots, v_m) \in \{0,1\}^{m^r} \,\middle|\, w_0 \in C^\perp, x_0 \in \{0,1\}^m, |x_0| \le d, v_i \in A_{w_{0i}}^{(r-1)} \ \forall i \ s.t. \ x_{0i} = 0 \right\},$$
$$A_1^{(r)} \triangleq \left\{ (v_1, \ldots, v_m) \in \{0,1\}^{m^r} \,\middle|\, w_1 \in C^\perp \oplus \mathbf{1}, x_1 \in \{0,1\}^m, |x_1| \le d, v_i \in A_{w_{1i}}^{(r-1)} \ \forall i \ s.t. \ x_{1i} = 0 \right\}.$$

CSS Code and Codeword

Lemma C.5 (Structure of $A_0^{(r)}$ and $A_1^{(r)}$). *For all $r \geq 1$, we have*

$$A_0^{(r)} \oplus \mathbf{1} = A_1^{(r)}.$$

Proof. We consider a proof by induction on $r \geq 1$. By definition of A_0 and A_1, we have $A_0^{(1)} \oplus \mathbf{1} = A_1^{(1)}$, which establishes the base case of $r = 1$. For $r > 1$, we show that for any $(v_1, \ldots, v_m) \in A_0^{(r)}$, we have $(v_1, \ldots, v_m) \oplus \mathbf{1} \in A_1^{(r)}$. Consider w_0, x_0 corresponding to $(v_1, \ldots, v_m)$. Using $v_i \in A_{w_{0i}}^{(r-1)}$ for all i with $x_{0i} = 0$ and the inductive hypothesis that $A_0^{(r-1)} \oplus \mathbf{1} = A_1^{(r-1)}$, we have $v_i \oplus \mathbf{1} \in A_{w_{0i} \oplus 1}^{(r-1)}$ for all i with $x_{0i} = 0$. Hence, by considering $w_1 = w_0 \oplus \mathbf{1}$ and $x_1 = x_0$, we have $(v_1, \ldots, v_m) \oplus \mathbf{1} \in A_1^{(r)}$. Similarly, we can show that for any $(v_1, \ldots, v_m) \in A_1^{(r)}$, we have $(v_1, \ldots, v_m) \oplus \mathbf{1} \in A_0^{(r)}$. Therefore, we have shown that $A_0^{(r)} \oplus \mathbf{1} = A_1^{(r)}$. □

Lemma C.6 (Disjointness of $A_0^{(r)}$ and $A_1^{(r)}$). *For all $r \geq 1$, we have*

$$A_0^{(r)} \cap A_1^{(r)} = \emptyset.$$

Oracle separation between BPP and NISQ proof 1

C.1.5 Proof of super-polynomial separation between NISQ and BPP

We are now ready to complete the proof of the oracle separation between NISQ and BPP.

Proof of Theorem 2.2. Let us decompose our total Hilbert space $\mathcal{H}$ as $\mathcal{H} \simeq \mathcal{H}_{\text{main},1} \otimes \mathcal{H}_{\text{main},2} \otimes \mathcal{H}_{\text{anc},1} \otimes \mathcal{H}_{\text{anc},2}$ where

$$\mathcal{H}_{\text{main},1} \simeq (\mathbb{C}^2)^{\otimes(m^r n')}, \ \mathcal{H}_{\text{main},2} \simeq (\mathbb{C}^2)^{\otimes(n-m^r n')}, \ \mathcal{H}_{\text{anc},1} \simeq (\mathbb{C}^2)^{\otimes(m^r n')}, \ \mathcal{H}_{\text{anc},2} \simeq (\mathbb{C}^2)^{\otimes\mathcal{O}(\text{poly}(n))}.$$

We begin with a state on $\mathcal{H}$ initialized in the all-zero state. By Lemma C.15, we can prepare a state ρ on $\mathcal{H}_{\text{main},1}$, using the ancillas on $\mathcal{H}_{\text{anc},2}$, such that ρ is $(r, d/2)$-deviated from $\rho^0 = V^{\otimes n'} H^{\otimes n'} |0^{n'}\rangle\langle 0^{n'}| H^{\otimes n'} (V^{\otimes n'})^\dagger$. By Lemma C.16, we can prepare a state σ on $\mathcal{H}_{\text{anc},1}$, using the ancillas on $\mathcal{H}_{\text{anc},2}$, such that σ is $(r, d/2)$-deviated from $\sigma^0 = V^{\otimes n'} |0^{n'}\rangle\langle 0^{n'}| (V^{\otimes n'})^\dagger$.

At this point in the algorithm, our qubits on $\mathcal{H}_{\text{main},2}$ are no longer in the all-zero state due to the local noise. We do not care what the state is and suppose that the state is given by $\rho^{(2)}$. We proceed by applying our oracle unitary $U_{\tilde{f}_s}$ to $(\rho \otimes \rho^{(2)}) \otimes \sigma$ on $\mathcal{H}_{\text{main},1} \otimes \mathcal{H}_{\text{main},2} \otimes \mathcal{H}_{\text{anc},1}$. Since the oracle unitary acts as the identity on $\mathcal{H}_{\text{main},2}$ by construction, we can equivalently just apply $U_{\tilde{f}_s^*}$ to $\rho \otimes \sigma$ on $\mathcal{H}_{\text{main},1} \otimes \mathcal{H}_{\text{anc},1}$. Doing so and subsequently neglecting the $\mathcal{H}_{\text{anc},1}$ register (corresponding to tracing out the qubits), we obtain

Oracle separation between BPP and NISQ proof 2

$$\rho' = \mathrm{tr}_{\mathcal{H}_{\mathrm{anc},1}}\left\{U_{\tilde{f}_s^\star}(\rho\otimes\sigma)U_{\tilde{f}_s^\star}^\dagger\right\}.$$

But by Lemma C.18, this state is only $(r, d/2)$-deviated from

$$\rho^1 = \mathrm{tr}_{\mathcal{H}_{\mathrm{anc},1}}\left\{U_{\tilde{f}_s^\star}(\rho^0\otimes\sigma^0)U_{\tilde{f}_s^\star}^\dagger\right\}.$$

If f_s is a 1-to-1 function, then

$$\rho^1 = V^{\otimes n'}\left(\frac{1}{2^{n'}}\sum_{z\in\{0,1\}^{n'}}|z\rangle\langle z|\right)(V^{\otimes n'})^\dagger,$$

whereas if f_s is a 2-to-1 function we have

$$\rho^1 = V^{\otimes n'}\left(\frac{1}{2^{n'}}\sum_{z\in\{0,1\}^{n'}}\frac{1}{\sqrt{2}}\left(|z\rangle+|z\oplus s\rangle\right)\cdot\frac{1}{\sqrt{2}}\left(\langle z|+\langle z\oplus s|\right)\right)(V^{\otimes n'})^\dagger$$

Oracle separation between BPP and NISQ proof 3

where s is the hidden string. Applying Hadamards to the encoded qubits of ρ', measuring in the computational basis, and applying classical post-processing via recursive majority vote as per Lemma C.17, we will obtain an n' bit string z_0 which with probability $1 - \mathcal{O}(1/n')$ is sampled from the distribution $\mathcal{D}$ defined as follows. If f_s is 1-to-1 function then $\mathcal{D}$ will be the uniform distribution over n' bit strings, whereas if f_s is a 2-to-1 function then $\mathcal{D}$ will be the uniform distribution over n' bit strings subject to the constraint $z_0 \cdot s = 0 \pmod 2$.

If we repeat the entire procedure n' times, then with probability $(1 - \mathcal{O}(1/n'))^{n'} = \Omega(1)$ we obtain n' such bit strings $z_0, z_1, ..., z_{n'-1}$. If this event, call it $\mathcal{E}$, happens, then by solving the n' linear equations $z_i \cdot s = 0 \pmod 2$ for $i = 0, 1, ..., n' - 1$, we can determine whether s is the all-zero string meaning f_s is 1-to-1, or some non-trivial string in which case f_s is 2-to-1. In general, if $\mathcal{E}$ does not happen and we have obtained some arbitrary string s, we can check that this situation is the case by querying the classical oracle at $f_s(0)$ and $f_s(s)$. So by repeating the entire procedure $\mathcal{O}(\log(1/\delta))$ times, with probability at least $1 - \delta$ the event $\mathcal{E}$ will happen at least once, and we will be able to determine if f_s is 1-to-1 or 2-to-1. □

(r,k)-sparse set and (r,k)-deviated

Definition C.8 ((r,k)-sparse set). *An (r,k)-sparse set of qubits over many blocks of the m^r qubits is defined recursively as follows. A set A of qubits over many blocks of m qubits is $(1,k)$-sparse if and only if every block has at most k qubits that are in A. A set A of qubits over many blocks of m^r qubits is (r,k)-sparse if and only if for every block, by treating the m^r qubits as m sub-blocks of m^{r-1} qubits, there are at most k sub-blocks that are not $(r-1,k)$-sparse.*

Definition C.9 ((r,k)-deviate). *A state ρ is said to be (r,k)-deviated from ρ' if k is the minimum integer such that there exists an (r,k)-sparse set of qubits A, such that $\rho_{A^c} = \rho'_{A^c}$. Here, we denote ρ_{A^c} to be the reduced density matrix of ρ on the qubits not in set A.*

V is the encoding map of recursive CSS code

(noiseless) quantum circuit Q with $2n'$ input qubits, depth t, and v locations. Let C be a quantum computation code with gates $\mathcal{G}$ that corrects d errors. Let

$$V : \mathbb{C}^2 \to (\mathbb{C}^2)^{\otimes m^r}$$

be the encoding map for the code given by recursively concatenating C a total of $r = \mathcal{O}(\log\log(v/\delta))$ times. Using key lemmas for establishing the threshold theorem from [53] (Theorem 10 therein), we can extract the following two results:

NQC is (r,d)-deviated from the following

Theorem C.12 (Lemma 8 and 10 from [53]). *There is an absolute constant $\lambda_c \in [0,1]$ such that for any $\lambda < \lambda_c$, there exists a λ-noisy quantum circuit Q' which can initialize ancillary qubits at any time (these ancillary qubits are also subject to qubit-wise noise of λ) during the computation and satisfies the following. Q' operates on $m^r n'$ qubits and has depth $\mathcal{O}(t \operatorname{polylog}(v/\delta))$, and the output state ρ of Q' is (r,d)-deviated from*

$$V^{\otimes n'} Q |0^{n'}\rangle\langle 0^{n'}| Q^\dagger (V^{\otimes n'})^\dagger$$

with probability $1-\delta$ over the local noise.

Theorem C.13 (Lemma 8, 9, and 10 from [53]). *There is an absolute constant $\lambda_c \in [0,1]$ such that for any $\lambda < \lambda_c$, there exists a λ-noisy quantum circuit Q' which can initialize ancillary qubits at any time (these ancillary qubits are also subject to qubit-wise noise of λ) during the computation, and a classical postprocessing algorithm $\mathcal{A}$ based on recursive majority vote, that satisfies the following. Q' operates on $m^r n'$ qubits and has depth $\mathcal{O}(t \operatorname{polylog}(v/\delta))$. Let σ be any n'-qubit state. Let $\mathcal{D}$ be the n'-bit string distribution generated by measuring $Q\sigma Q^\dagger$ in the computational basis. For any state ρ that is (r,d)-deviated from*

$$V^{\otimes n'} \sigma (V^{\otimes n'})^\dagger,$$

applying a λ-noisy computational basis measurements on the output state of Q' given input state ρ, followed by the classical algorithm $\mathcal{A}$, produces a distribution $\mathcal{D}'$ equal to $\mathcal{D}$ with probability $1-\delta$ over the local noise.

NQC is (r,d)-deviated from the following

Lemma C.15. *Suppose $n' \leq \exp(\log^c n)$ for $0 < c < 1$ a sufficiently small constant. There exists an absolute constant $\lambda_c \in [0,1]$ such that for any non-negative $\lambda < \lambda_c$, there exists a λ-noisy quantum circuit which operates on n' input qubits and $\mathrm{poly}(n)$ ancillary qubits and has $\mathrm{polylog}(n)$ layers, such that with probability at least $1 - \mathcal{O}(1/n')$ over the local noise, the output state is $(r, d/2)$-deviated from the state*

$$V^{\otimes n'} H^{\otimes n'} |0^{n'}\rangle\langle 0^{n'}| H^{\otimes n'} (V^{\otimes n'})^\dagger \tag{10}$$

for $r = \log\log(n')$.

Lemma C.16. *Suppose $n' \leq \exp(\log^c n)$ for $0 < c < 1$ a sufficiently small constant. There exists an absolute constant $\lambda_c \in [0,1]$ such that for any non-negative $\lambda < \lambda_c$, there exists a λ-noisy quantum circuit which operates on n' input qubits and $\mathrm{poly}(n)$ ancillary qubits and has $\mathrm{polylog}(n)$ layers, such that with probability at least $1 - \mathcal{O}(1/n')$ over the local noise, the output state is $(r, d/2)$-deviated from the state*

$$V^{\otimes n'} |0^{n'}\rangle\langle 0^{n'}| (V^{\otimes n'})^\dagger$$

for $r = \log\log(n')$.

Lemma C.17. *Suppose $n' \leq \exp(\log^c n)$ for $0 < c < 1$ a sufficiently small constant and for $r = \mathrm{polylog}(n')$. There exists an absolute constant $\lambda_c \in [0,1]$ such that for any non-negative $\lambda < \lambda_c$, there exists a λ-noisy quantum circuit Q' which operates on $m^r n'$ input qubits and $\mathrm{poly}(n)$ ancillary qubits and has $\mathrm{polylog}(n')$ layers, such that the following holds. Let $\mathcal{A}$ be the classical post-processing procedure based on recursive majority vote from Theorem C.13.*

For k satisfying $k \leq d$, let input state ρ be (r,k)-deviated from the state

$$V^{\otimes n'} \sigma (V^{\otimes n'})^\dagger.$$

Stability of the "robustified" classical oracle

First, recall the unitary $U_{\widetilde{f}_s}$ from Eq. (9). Note that by construction, $\widetilde{f}_s$ only depends non-trivially on the first $m^r n' < n$ bits of its input. By defining $\widetilde{f}_s^* : \{0,1\}^{m^r n'} \to \{0,1\}^{m^r n'}$ to be the function $\widetilde{f}_s$ restricted to the first $m^r n'$ bits, we can rewrite (9) as

$$U_{\widetilde{f}_s} |x_1\rangle |x_2\rangle |y\rangle = |x_1\rangle |x_2\rangle |y \oplus \widetilde{f}_s^*(x_1)\rangle , \quad \forall x_1 \in \{0,1\}^{m^r n'}, x_2 \in \{0,1\}^{n-m^r n'}, y \in \{0,1\}^{m^r n'}.$$

We see from the above equation that $U_{\widetilde{f}_s}$ acts trivially on the $|x_2\rangle$ part of the input state. So let us define $U_{\widetilde{f}_s^*}$ as the restriction of $U_{\widetilde{f}_s}$ to its $|x_1\rangle$ and $|y\rangle$, subsystems, namely

$$U_{\widetilde{f}_s^*} |x_1\rangle |y\rangle = |x_1\rangle |y \oplus \widetilde{f}_s^*(x_1)\rangle , \quad \forall x_1 \in \{0,1\}^{m^r n'}, y \in \{0,1\}^{m^r n'}.$$

With the above notations for the oracle, we can now prove the following lemma showing the classical function $\widetilde{f}_s^*$ preserves the deviation metric. We note that, on the other hand, the ordinary Simon's function f_s does not have the same property.

Lemma C.18 (Stability of the robustified classical oracle). *Consider a Hilbert space $\mathcal{H}$ which decomposes into subsystems as $\mathcal{H} \simeq \mathcal{H}_{\text{main},1} \otimes \mathcal{H}_{\text{anc},1}$ where $\mathcal{H}_{\text{main},1} \simeq (\mathbb{C}^2)^{\otimes(m^r n')}$ and $\mathcal{H}_{\text{anc},1} \simeq (\mathbb{C}^2)^{\otimes(m^r n')}$. Further let $\rho^0 = V^{\otimes n'} H^{\otimes n'} |0^{n'}\rangle\langle 0^{n'}| H^{\otimes n'} (V^{\otimes n'})^\dagger$ and $\sigma^0 = V^{\otimes n'} |0^{n'}\rangle\langle 0^{n'}| (V^{\otimes n'})^\dagger$.*

Given any $k \leq d$, if $\rho \otimes \sigma$ is (r,k)-deviated from $\rho^0 \otimes \sigma^0$, then $\mathrm{tr}_{\mathcal{H}_{\text{anc},1}}\left\{U_{\widetilde{f}_s^*}(\rho \otimes \sigma)U_{\widetilde{f}_s^*}^\dagger\right\}$ is (r,k)-deviated from $\mathrm{tr}_{\mathcal{H}_{\text{anc},1}}\left\{U_{\widetilde{f}_s^*}(\rho^0 \otimes \sigma^0)U_{\widetilde{f}_s^*}^\dagger\right\}$.

"Lifted" Simon's problem for NISQ and BQP

C.2 NISQ vs. BQP

In this section we show an oracle separation between NISQ and BQP via a simple "lifting" of Simon's problem. In fact, we will actually be able to separate NISQ and $\mathsf{BPP}^{\mathsf{QNC}^0}$ relative to this oracle.

We begin by describing the modification of Simon's problem we will consider. For $n \in \mathbb{N}$, given a function $f : \{0,1\}^n \to \{0,1\}^n$, we define the *lift* of f to be the function $\widetilde{f} : \{0,1\}^{2n} \to \{0,1\}^n$ given by

$$\widetilde{f}(x) \triangleq \begin{cases} f(x_1, \ldots, x_n) & \text{if } x_{n+1}, \ldots, x_{2n} = 0 \\ 0 & \text{otherwise} \end{cases}.$$

Given lifted function $\widetilde{f}$, we will abuse notation and let $O_{\widetilde{f}}$ denote both the classical oracle given by evaluating $\widetilde{f}$ as well as the quantum oracle

$$O_{\widetilde{f}} : |x\rangle\,|y\rangle \mapsto |x\rangle\,|y \oplus \widetilde{f}(x)\rangle\,.$$

It is not hard to see that in the absence of depolarizing noise, a minor modification of Simon's algorithm, which can be implemented in $\mathsf{BPP}^{\mathsf{QNC}^0}$, still works under this lifting. In contrast, for NISQ algorithms, we show the following:

Theorem C.20. *Let $\lambda \in [0,1]$. Any NISQ_λ algorithm which, given oracle access to $O_{\widetilde{f}}$ for any lift of a function $f : \{0,1\}^n \to \{0,1\}^n$ which is either 2-to-1 or 1-to-1, can determine whether f is 2-to-1 or 1-to-1 with constant advantage must have query complexity at least $\exp(\Omega(\lambda n))$. Thus, relative to oracles O of this form, $\mathsf{NISQ}^O \subsetneq \mathsf{BQP}^O$.*

Oracle separation between NISQ and BQP proof 1

Proof of Theorem C.20. Let $\mathcal{T}$ be the learning tree corresponding to a NISQ_λ algorithm that makes at most N classical or quantum oracle queries to $O_{\tilde{f}}$, as in Definition B.1. By Lemma B.2 and Lemma C.22, if we replace every noisy quantum circuit A in the tree with a noisy quantum circuit A' that makes queries to the identity oracle instead of to $O_{\tilde{f}}$, then the new distribution over the leaves of $\mathcal{T}$ is at most $N^2 \exp(-\Omega(\lambda n))$-far in total variation from the original distribution $p_{O_{\tilde{f}}}$; for $N = \exp(o(\lambda n))$, this quantity is $o(1)$. For convenience, denote this new distribution by p'_f.

To apply Lemma B.3, we wish to bound $d_{\mathrm{TV}}(\mathbb{E}_{f \text{ 1-to-1}}[p'_f], \mathbb{E}_{f \text{ 2-to-1}}[p'_f])$. But note that because the quantum circuits A' in the new learning tree are independent of the underlying function f, the learning tree is simply implementing a randomized classical query algorithm. We can thus think of p'_f as a mixture over distributions p'^r_f each corresponding to some fixing of the internal randomness r of the algorithm (here the coefficients of the mixture are independent of f). It thus suffices to bound $\sup_r d_{\mathrm{TV}}(\mathbb{E}_{f \text{ 1-to-1}}[p'^r_f], \mathbb{E}_{f \text{ 2-to-1}}[p'^r_f])$.

Oracle separation between NISQ and BQP proof 2

Henceforth fix any r. The rest of the argument follows the standard proof of the classical lower bound for Simons' algorithm. The algorithm queries the classical oracle at some deterministic sequence of inputs $x_1, \ldots, x_a$, which we may assume without loss of generality are distinct and lie in Ω. For any $y_1, \ldots, y_a$ which are all distinct, $(x_1, y_1), \ldots, (x_a, y_a)$ and r determine some leaf node ℓ of the tree. The probability of this leaf node under $\mathbb{E}_{f \text{ 1-to-1}}[p'^r_f]$ is $\frac{(2^n-a)!}{(2^n)!}$ and under $\mathbb{E}_{f \text{ 2-to-1}}[p'^r_f]$) is $\frac{M}{2^n-1}\frac{(2^n-a)!}{(2^n)!}$ where $M \triangleq 2^n - 1 - |\{x_i \oplus x_j \mid 1 \le i < j \le a\}|$. For any $y_1, \ldots, y_a$ for which there is a collision, the probability of the corresponding leaf node under $\mathbb{E}_{f \text{ 1-to-1}}[p'^r_f]$ is clearly 0. We conclude that the total variation between these two mixtures is upper bounded by the probability that there is a collision among $f(x_1), \ldots, f(x_a)$ for a random 2-to-1 function f. The latter is at most

$$\sum_{i=0}^{a-1} \frac{i}{2^n - 1 - \binom{i}{2}} \le \frac{a^2}{2^{n+1} - 2 - a^2},$$

so for $a \ll 2^{n/2}$, this quantity is $o(1)$. As $\min(\exp(\Omega(\lambda n)), 2^{n/2}) = \exp(\Omega(\lambda n))$, the theorem thus follows by Lemma B.3. □

Remark C.23. *The reader may observe that apart from the classical lower bound for Simon's problem, our proof of the lower bound in Theorem C.20 makes very little use of the fact that f is either a 2-to-1 or 1-to-1 function. In fact, the above argument shows more generally that for any search problem over a family of Boolean functions, the query complexity of any* NISQ *algorithm is essentially given by the classical query complexity for that problem.*

Trace distance under projection

Lemma C.21. *Given $n' \in \mathbb{N}$, let Ω denote some subset of $\{0,1\}^{n'}$, and let Π denote the projection to the span of $\{|x\rangle\}_{x\in\Omega}$. Then for any $\lambda \in [0,1]$ and any n'-qubit state $|\psi\rangle$,*

$$\mathrm{tr}(\Pi D_\lambda^{\otimes n'}[|\psi\rangle\langle\psi|]) \leq \sup_D \Pr_{a\sim D,\tilde{a}}[\tilde{a} \in \Omega], \tag{14}$$

where the supremum is over probability distributions over $\{0,1\}^{n'}$, and $\tilde{a}$ is the random string obtained by flipping each of the bits of a independently with probability $\lambda/2$.

Note the probability on the right-hand side of (14) is exponentially small when $\Omega \subset \{0,1\}^{2n}$ is the set of strings x for which $x_{n+1},\ldots,x_{2n} = 0$. We will now use this to show that the distribution over measurement outcomes from running a noisy quantum circuit that has query access to either $O_{\tilde{f}}$ or the identity oracle Id gives very little information about which oracle the circuit has access to.

TVD upper bound between non-trivial and identity oracle

Lemma C.22. *Let A be any λ-noisy quantum circuit which makes N oracle queries. If $p_{\tilde{f}}$ (respectively p_{id}) is the distribution over the random string s output by the circuit when the oracle is $O_{\tilde{f}}$ (respectively the identity oracle* Id*), then* $d_{TV}(p_{\tilde{f}}, p_{\mathrm{id}}) \leq N \exp(-\Omega(\lambda n))$.

Proof. Let n' denote the number of qubits on which A operates. For convenience, we denote by $\widehat{O}_{\tilde{f}}$ the channel given by pre-composing $O_{\tilde{f}}$ with $D_\lambda^{\otimes 3n}$. We will show that for all n'-qubit pure states σ, $\|(\widehat{O}_{\tilde{f}} \otimes D_\lambda^{\otimes n'-3n} - D_\lambda^{\otimes n'})[\sigma]\|_{\mathrm{tr}}$ is small so that we can apply Lemma B.4.

When $\Omega \subset \{0,1\}^{2n}$ is given by all strings whose last n bits are 0, then for any $a \in \Omega$, if $\tilde{a}$ is obtained by flipping each of the bits of a independently with probability $\lambda/2$, then $\Pr[\tilde{a} \in \Omega] \leq (1-\lambda/2)^n \leq \exp(-\lambda n/2)$. So by Lemma C.21, if $D_\lambda^{\otimes n'}[\sigma] = \sum_i \lambda_i |v_i\rangle \langle v_i|$, then $\sum_i \lambda_i \|\Pi' v_i\|^2 \leq \exp(-\lambda n/2)$, where Π' is the projection to the span of $\{|x\rangle |y\rangle |w\rangle\}_{x\in\Omega, y\in\{0,1\}^n, w\in\{0,1\}^{n'-3n}}$. If we write every v_i as $\sum_{x\in\{0,1\}^{2n}, y\in\{0,1\}^n, w\in\{0,1\}^{n'-3n}} v_{i,x,y,w} |x\rangle |y\rangle |w\rangle$, then

$$\sum_i \lambda_i \sum_{x\in\Omega, y, w} v_{i,x,y,w}^2 \leq \exp(-\lambda n/2).$$

TVD upper bound between non-trivial and identity oracle

In particular,

$$\begin{aligned}
&\|(O_{\tilde{f}} \otimes \mathrm{Id} - \mathrm{Id})[|v_i\rangle\langle v_i|]\|_{\mathsf{tr}} \\
&\quad \le \sqrt{2}\|(O_{\tilde{f}} \otimes \mathrm{Id} - \mathrm{Id})[|v_i\rangle\langle v_i|]\|_F \\
&\quad \le 2\sqrt{2}\Big(\sum_{x\in\Omega \text{ or } x'\in\Omega, y, y', w, w'} v_{i,x,y,w}^2 v_{i,x',y',w'}^2\Big)^{1/2} \\
&\quad \le 2\sqrt{2}\Big(1 - \Big(1 - \sum_{x\in\Omega,y,w} v_{i,x,y,w}^2\Big)^2\Big)^{1/2} \le 4\sqrt{\sum_{x\in\Omega,y,w} v_{i,x,y,w}^2}.
\end{aligned}$$

By Jensen's inequality, we can bound $\|(\widehat{O}_{\tilde{f}} \otimes D_\lambda^{\otimes n'-3n} - D_\lambda^{\otimes n'})[\sigma]\|_{\mathsf{tr}}$ by

$$4\sum_i \lambda_i \sqrt{\sum_{x\in\Omega,y,w} v_{i,x,y,w}^2} \le 4\sqrt{\sum_i \lambda_i \sum_{x\in\Omega,y,w} v_{i,x,y,w}^2} \le 4\exp(-\lambda n/4).$$

By taking the channels $\mathcal{E}_1$ and $\mathcal{E}_2$ in Lemma B.4 to be $\widehat{O}_{\tilde{f}} \otimes D_\lambda^{\otimes n'-3n}$ and $D_\lambda^{\otimes n'}$, we obtain the desired bound on $d_{\mathrm{TV}}(p_{\tilde{f}}, p_{\mathrm{id}})$. □

Example 1: Grover (unstructured search)

D Unstructured Search

In this section we show that there is no quadratic speedup for unstructured search in NISQ. Given $d \in \mathbb{N}$ and $i \in [d]$, we abuse notation and let O_i denote both the classical oracle $O_i : [d] \to \{0,1\}$ given by $O_i(x) = \mathbb{1}[x = i]$ as well as the quantum oracle

$$O_i : |x\rangle |w\rangle \mapsto (-1)^{\mathbb{1}[x=i]} |x\rangle |w\rangle \quad \forall\, x \in [d].$$

Theorem D.1. *Let $\lambda \in [0,1]$. Any NISQ_λ algorithm which, given oracle access to O_i for any $i \in [d]$, can determine i with probability $2/3$ must have query complexity at least $\widetilde{\Omega}(d\lambda)$.*

Example 1: Grover (unstructured search) proof 1

Proof of Theorem D.1. Let $\mathcal{T}$ be the learning tree corresponding to a NISQ algorithm which has access to O_i for some $0 \leq i \leq d$ and has query complexity N, as in Definition B.1. Let $\overline{T}$ be some choice of depth that we will tune later. We will convert $\mathcal{T}$ to a learning tree $\widehat{\mathcal{T}}$ corresponding to a bounded-depth noiseless NISQ algorithm, as in Definition D.2.

Define $\widehat{\mathcal{T}}$ as follows. For every non-leaf node u, if the algorithm makes a single classical query at input j, then replace the edge $(u, x, O_i(x))$ to its child v by an edge (u, A, s) where A is a depth-1 quantum algorithm simulating the classical query. On the other hand, suppose that at u, the algorithm runs some λ-noisy quantum circuit A on $\text{poly}(d)$ qubits. If A makes fewer than $\overline{T}$ oracle queries in total, then consider the noiseless quantum circuit A' which simulates A by applying depolarizing noise at each layer. If A makes more than $\overline{T}$ queries, then replace A with the quantum circuit A' that simply measures the maximally mixed state in the computational basis, rather than the output state $|\phi_i(A)\rangle$. By Lemma D.16 and Pinsker's inequality, the total variation distance between the induced conditional distributions on children when $|\phi_i(A)\rangle$ gets measured versus when the maximally mixed state gets measured is at most $\sqrt{\frac{1}{2}\mathcal{I}(|\phi_i(A)\rangle)}$, and by Lemma D.15 this is at most $(1-\lambda)^{\overline{T}/2} \cdot \mathcal{O}(\sqrt{\log d})$.

Example 1: Grover (unstructured search) proof 2

Let p_i (respectively $\widehat{p}_i$) denote the distribution over leaves when running the NISQ algorithm given by $\mathcal{T}$ (respectively the noiseless algorithm given by $\widehat{\mathcal{T}}$). As the number of oracle queries is at most N, the depth of both trees is at most N, so we conclude that the total variation distance between p_i and $\widehat{p}_i$ is at most $N(1-\lambda)^{\overline{T}/2} \cdot \mathcal{O}(\sqrt{\log d})$ by Lemma B.2. We will take $\overline{T} = C\lambda^{-1}(\log\log d + \log N)$ for constant $C > 0$ so that this quantity is an arbitrarily small constant.

We conclude by Theorem D.4 that if $N \leq cd/\overline{T} = \Theta(d\lambda/(\log\log d + \log N))$, then the NISQ algorithm given by $\mathcal{T}$ cannot solve unstructured search with probability $2/3$. This concludes the proof of our $\widetilde{\Omega}(d\lambda)$ lower bound. □

Upper bound on the information in NQC

D.3 From bounded-depth to noisy computation

We now show how to extract from Theorem D.4 a lower bound against NISQ. We begin with the following basic lemma, a proof of which we include in Appendix H.1 for completeness, that quantifies the amount of information that is lost from running many layers of noisy computation:

Lemma D.15 (Lemma 8 from [87]). *Let A be a λ-noisy depth-T quantum circuit on n qubits with output state ρ. Then $\mathcal{I}(\rho) \triangleq n - S(\rho) \leq (1-\lambda)^T \cdot n$, where $S(\cdot)$ denotes von Neumann entropy.*

We will also use the following standard operational characterization of $I(\rho)$:

Lemma D.16 (See e.g. Lemma 2 from [87]). *Given any n-qubit state ρ and any POVM, the distributions p, q induced by respectively measuring ρ and $I/2^n$ with the POVM satisfy $KL(p\|q) \leq \mathcal{I}(\rho)$.*

Upper bound on the information in NQC

Given $S \subseteq [n]$ and mixed n-qubit state σ, let $\sigma|_S$ denote the restriction of σ to the subsystem indexed by S. That is,

$$\sigma_{!S} = \left(\mathrm{tr}(\sigma|_{S^c}) \cdot I/2^{|S^c|}\right) \otimes \sigma|_S. \tag{22}$$

Proof. Von Neumann entropy is invariant under unitary transformation, so it suffices to show that for any mixed state σ, $I(D_\lambda[\sigma]) \leq (1-\lambda) \cdot I(\sigma)$. Because $D_\lambda[\sigma] = \mathbb{E}_{S\sim\mu}[\sigma_{!S}]$ for μ the distribution over $S \subseteq [n]$ which includes each element of $[n]$ independently with probability λ. So by concavity of entropy, additivity of entropy for tensor products, and (22),

$$I(D_\lambda[\sigma]) \leq \sum_{k=0}^{n} \lambda^{n-k}(1-\lambda)^k \sum_{S\subset[n]:|S|=k} I(\sigma|_S) \leq \sum_{k=0}^{n} \binom{n}{k} \frac{k}{n} \lambda^{n-k}(1-\lambda)^k = (1-\lambda) I(\sigma),$$

where in the last step we used Lemma H.1 below. □

The proof above uses the following fact:

Lemma H.1 (Lemma 7 from [87]). *For any density matrix σ on n qubits and any $0 \leq k < n$,*

$$\binom{n}{k}^{-1} \sum_{S\subset[n]:|S|=k} I(\sigma|_S) \leq \frac{k}{n} I(\sigma).$$

Lower bound on the bounded-depth computation

Lemma D.3 (Eq. (7) in [91]). *For any quantum circuit A for unstructured search that makes T oracle queries, if $|\phi_i(A)\rangle$ denotes the output state when the underlying oracle is O_i, then*

$$\sum_{i=1}^{d} \| |\phi_i(A)\rangle - |\phi_0(A)\rangle \|^2 \leq 4T^2.$$

D.2 Lower bound against bounded-depth computation

We now use these tools to prove the following query complexity lower bound. This will be the main component in our proof of Theorem D.1.

Theorem D.4. *There is an absolute constant $c > 0$ for which the following holds. Let $d, T \in \mathbb{N}$ with $T \leq d$. Then no noiseless quantum algorithm of depth T with query complexity at most cd/T can, given oracle access to O_i for any $i \in [d]$, output i with probability $2/3$.*

Example 2: Bernstein-Vazirani

E Bernstein-Vazirani Problem

In this section we show that a NISQ algorithm can solve the Bernstein-Vazirani problem [92] with $\mathcal{O}(\log n)$ queries, whereas it is known that any classical algorithm requires $\Theta(n)$ queries. As with the upper bound in Section C.1, we will show that our algorithm is robust not just to local depolarizing noise, but also to arbitrary local noise that occurs with sufficiently small constant rate (see Remark E.5).

We begin by recalling the Bernstein-Vazirani problem on n bits. There is an unknown function $f : \{0,1\}^n \to \{0,1\}$ of the form $f(x) = s \cdot x \pmod 2$, where $s \in \{0,1\}^n$ is often called the *hidden string*. The goal is to determine the hidden string. In the quantum context, the classical oracle is rendered into a unitary O_f which acts as

$$O_f : |x\rangle \otimes |y\rangle \mapsto |x\rangle \otimes |y \oplus f(x)\rangle$$

for $|x\rangle$ a state on n qubits and $|y\rangle$ a state on one qubit. In the noiseless quantum setting, the best quantum algorithm can find the hidden string s in $\Theta(1)$ queries [92].

We prove the following result on the NISQ complexity of the Bernstein-Vazirani problem:

Theorem E.1. *Let $0 \leq \lambda < 1/24$. Then there is a NISQ_λ algorithm which can solve the Bernstein-Vazirani problem with probability at least $1 - \delta$ using $\mathcal{O}(\frac{1}{1-24\lambda} \log(n/\delta))$ queries.*

Example 2: Bernstein-Vazirani proof 1

Proof of Theorem E.1. By Lemma E.4, we see that for each bit i of the n output bits, the probability of being $|s_i\rangle$ is at least $(1-\lambda)^6$, which happens if the ancilla qubit is never corrupted in either of the two layers of noise prior to the application of the oracle (with probability $(1-\lambda)^2$), and if additionally the former of the two possible events in Lemma E.4 happens (with probability $(1-\lambda)^4$). Let $f(\lambda) \triangleq (1-\lambda)^6$ and note that for $\lambda \leq 1/10$, $f(\lambda) > 1/2$.

Let X_i be a random variable which equals zero if s_i is obtained correctly with our procedure, and equal to one otherwise. Letting Y_i be the average of M i.i.d. copies of X_i, then the Chernoff-Hoeffding bound tells us that

$$\text{Prob}\left(Y_i^{(j)} \geq \frac{1}{2}\right) \leq \exp\left(-2M\left(\frac{1}{2} - f(\lambda)\right)^2\right) .$$

This is an upper bound on the probability that if we repeat the Bernstein-Vazirani algorithm M times and employ the majority votes strategy on the ith site to determine s_i, then we will fail. The probability that we fail for at least one of the n sites is upper bounded by

$$\text{Prob}\left(\max_i Y_i \geq \frac{1}{2}\right) \leq \sum_{i=1}^{n} \exp\left(-2M\left(\frac{1}{2} - f(\lambda)\right)^2\right) \leq n \exp\left(-2M\left(\frac{1}{2} - f(\lambda)\right)^2\right) .$$

Example 2: Bernstein-Vazirani proof 2

So if we want the right-hand side to be at most δ, then we can pick some M such that

$$M = O\left(\frac{1}{(1-2f(\lambda))^2}\log(n/\delta)\right). \tag{21}$$

Since $(1-2f(\lambda))^2$ is an alternating series in λ, we can lower bound it by its first two terms as

$$(1-2f(\lambda))^2 \geq 1-24\lambda\,.$$

Then (21) can be written in a slightly simplified form as

$$M = \left(\frac{1}{1-24\lambda}\log(n/\delta)\right)$$

for $\lambda < 1/24$, as claimed.[3] □

[3] We have not attempted to optimize this threshold for λ for general local noise channels or particular choices of local noise channels. However, we note that with more careful bookkeeping one can show, e.g. when the local noise is depolarizing noise, that for any λ bounded away from 1 the above algorithm can solve the Bernstein-Vazirani problem with $\mathcal{O}(\log(n/\delta))$ queries.

Why $(1-\lambda)^6$??

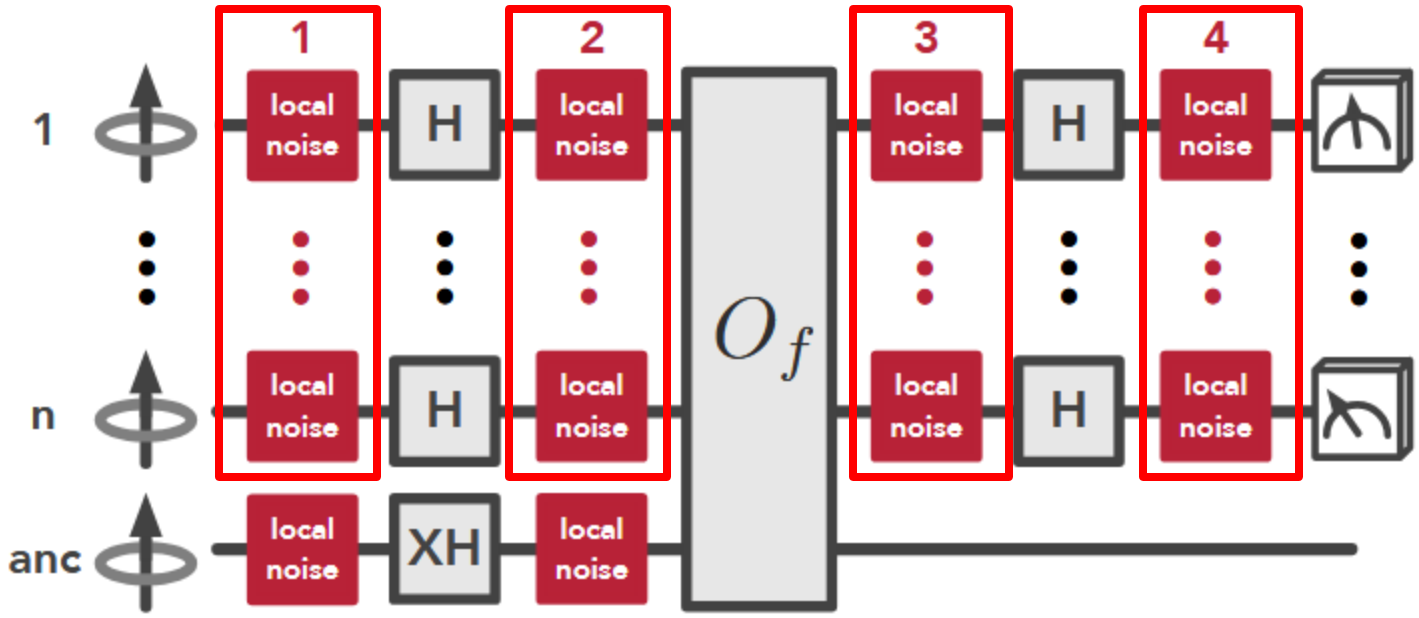


Figure 3: Bernstein-Vazirani algorithm in the presence of arbitrary noise (each box labeled by "local noise" denotes that with probability λ, an arbitrary, adversarially chosen single-qubit operation is applied). We have labeled the layers of noise for ease of reference in the proof.

Why $(1-\lambda)^6$??

Lemma E.4. *If the ancilla qubit is not corrupted prior to application of the oracle in the Bernstein-Vazirani algorithm, then for every $i \in [n]$, with probability $(1-\lambda)^4$ the ith output bit is given by s_i and otherwise is given by a possibly incorrect bit.*

Proof. When we reach the step of the Bernstein-Vazirani algorithm where we apply the oracle, suppose that k qubits out of the first n have been corrupted in the first two layers of noise (see Figure 3). Further suppose that the qubits are located at $a_1, ..., a_k$, where $\{a_1, ..., a_k\} \subset \{1, ..., n\}$. Picking some permutation $\pi \in S_n$ such that $\pi(i) = a_i$ for $i = 1, ..., k$, we can write the state of the system as

$$\frac{1}{2^n} \sum_{\substack{z,z' \in \{0,1\}^k \\ x,x' \in \{0,1\}^{n-k}}} \beta_{z,z'}\, \pi(|z\rangle\langle z'| \otimes |x\rangle\langle x'|)\pi^\dagger \otimes |-\rangle\langle -|$$

for some coefficients $\{\beta_{z,z'}\}$ satisfying $\sum_z \beta_{z,z} = 1$. The rest of the protocol proceeds as follows. We apply O_f to get

$$\frac{1}{2^n} \sum_{\substack{z,z' \in \{0,1\}^k \\ x,x' \in \{0,1\}^{n-k}}} \beta_{z,z'}\, \pi(|z\rangle\langle z'| \otimes |x\rangle\langle x'|)\pi^\dagger \otimes |-\rangle\langle -|\, (-1)^{f_\pi(zx)+f_\pi(z'x')}\,.$$

Next we trace out the ancilla to find

$$\frac{1}{2^n} \sum_{\substack{z,z' \in \{0,1\}^k \\ x,x' \in \{0,1\}^{n-k}}} \beta_{z,z'}(-1)^{f_\pi(zx)+f_\pi(z'x')}\, \pi(|z\rangle\langle z'| \otimes |x\rangle\langle x'|)\pi^\dagger\,. \tag{20}$$

Why $(1-\lambda)^6$??

Following this we apply a third layer of local noise, apply $H^{\otimes n}$, and then apply a fourth layer of local noise again. Suppose that this procedure corrupts any number of the already corrupted qubits at positions $a_1, ..., a_k$, as well as ℓ qubits at positions different from $a_1, ..., a_k$. Suppose that the positions of these ℓ qubits are $a_{k+1}, ..., a_{k+\ell}$ where $\{a_1, ..., a_k, a_{k+1}, ..., a_{k+\ell}\} \subset \{1, ..., n\}$. Since the local noise and $H^{\otimes n}$ act qubit-wise, if we only want to track the uncorrupted qubits we can do as follows: at the outset we trace out the $k+\ell$ qubits which are to be corrupted, and then we apply $H^{\otimes(n-k-\ell)}$ to the residual qubits.

We implement this procedure presently. Defining $\sigma \in S_n$ by $\sigma(i) = a_i$ for $i = 1, ..., k+\ell$, we can rewrite (20) as

$$\frac{1}{2^n} \sum_{\substack{z,z' \in \{0,1\}^k \\ w,w' \in \{0,1\}^\ell \\ y,y' \in \{0,1\}^{n-k-\ell}}} \beta_{z,z',w,w'} (-1)^{f_\sigma(zwy)+f_\sigma(z'w'y')} \, \sigma(|z\rangle\langle z'| \otimes |w\rangle\langle w'| \otimes |y\rangle\langle y'|)\sigma^\dagger$$

for some coefficients $\{\beta_{z,z',w,w'}\}$ satisfying $\sum_{z,w} \beta_{z,z,w,w} = 1$. Letting $\{b_1, ..., b_{n-k-\ell}\} = \{1, ..., n\} \setminus \{a_1, ..., a_{k+\ell}\}$ be the uncorrupted registers, where we choose $b_1 < b_2 < \cdots < b_{n-k-\ell}$, tracing out everything but the $b_1, ..., b_{n-k-\ell}$ registers we find the residual pure state

$$\frac{1}{\sqrt{2^{n-k-\ell}}} \sum_{y \in \{0,1\}^{n-k-\ell}} (-1)^{y \cdot [s]_{b_1,...,b_{n-k-\ell}}} |y\rangle$$

Why $(1-\lambda)^6$??

where $[s]_{b_1,\ldots,b_{n-k-\ell}}$ denotes the $b_1,\ldots,b_{n-k-\ell}$ bits of the hidden string s. Applying $H^{\otimes(n-k-\ell)}$ we find

$$\frac{1}{2^{n-k-\ell}} \sum_{y,y' \in \{0,1\}^{n-k-\ell}} (-1)^{y\cdot([s]_{b_1,\ldots,b_{n-k-\ell}}+y')} |y'\rangle .$$

Finally, measuring in the computational basis, the probability of measuring $|[s]_{b_1,\ldots,b_{n-k-\ell}}\rangle$ is equal to one.

All in all, we have seen that if we perform the usual Bernstein-Vazirani algorithm, we obtain the hidden bit string s but with a fraction of its bits corrupted, corresponding precisely to the qubits that were corrupted by one of the four layers of local noise. □

Example 3: Shadow tomography

F Shadow Tomography

In this section we show that relative to a natural *quantum oracle*, there is also an exponential separation even between NISQ and $\mathsf{BPP}^{\mathsf{QNC}^0}$. The task we consider that witnesses this separation has been studied previously in the context of separations between algorithms with and without quantum memory [37, 48, 50] and is based on *shadow tomography*, i.e. predicting properties of an unknown state to which one has access via a state oracle. Informally, for an unknown state ρ, one would like to predict $|\mathrm{tr}(Q\rho)|$ for all Pauli operators $Q \in \{I, X, Y, Z\}^{\otimes n}$ given copies of ρ.

We will show that NISQ_λ algorithms with access to such an oracle require $1/(1-\lambda)^{\Omega(n)}$ copies of ρ to estimate all of these observables to within constant error. On the other hand, existing upper bounds [37] imply that $\mathsf{BPP}^{\mathsf{QNC}^0}$ algorithms only require $\mathcal{O}(n)$ copies of ρ.

Oracle O_ρ

Definition F.1. *Let ρ be an n-qubit state. We consider an oracle O_ρ given as a CPTP map that traces out n qubits in the* state *register and prepares the state ρ in the* state *register:*

$$O_\rho(\sigma) \triangleq \rho \otimes \mathrm{tr}_{\mathsf{state}}(\sigma),$$

for any integer $n' \geq n$ and any n'-qubit state σ.

For the purposes of showing a lower bound in this oracle model, we will assume that in between oracle queries, the algorithm can perform arbitrary noiseless quantum computation, and at the end it can perform a noiseless measurement in the computational basis. The only noise that gets applied is local depolarizing noise after any call to O_ρ. Note that this is a stronger model of computation than a λ-noisy quantum algorithm or a NISQ_λ algorithm, which merely makes the lower bound we show even stronger. Furthermore, as there is no notion of a classical oracle in this setting, it is not necessary to work with the tree formalism of Definition B.1.

F.2 Exponential lower bound

We are now ready to state the main oracle separation of this section:

Theorem F.2. *Let $\rho = \frac{1}{2^n}(I + s \cdot P)$ for some $s \in \{0,1\}$ and n-qubit Pauli operator P. Given access to the oracle O_ρ in Definition F.1, no NISQ_λ algorithm can determine either s or P with constant advantage unless it makes $\Omega((1-\lambda)^{-n})$ oracle queries.*

On the other hand, there is an algorithm in $\mathsf{BPP}^{\mathsf{QNC}^0}$ that, given $\mathcal{O}(n)$ oracle queries, can determine both s and P with high probability. In fact, even if ρ is an arbitrary state, there is an algorithm in $\mathsf{BPP}^{\mathsf{QNC}^0}$ that, given $\mathcal{O}(n)$ oracle queries, can estimate $|\mathrm{tr}(P\rho)|$ for all n-qubit Pauli operators P to within small constant error with high probability.

Shadow tomography proof

Lemma F.3. *For any $\lambda > 0$. Let $P \in \{I, X, Y, Z\}^{\otimes n}$. Any NISQ_λ algorithm that has oracle access to the state oracle O_ρ for either $\rho = \frac{1}{2^n}(I + P)$ or $\rho = \frac{I}{2^n}$ and can distinguish which oracle it has access to with at least $2/3$ probability must make $\Omega((1-\lambda)^{-|P|})$ queries, where $|P|$ denotes the number of non-identity components in P.*

Proof. Suppose the circuit operates on n' qubits. For convenience, for $s \in \{0, 1\}$ denote $O_{\frac{1}{2^n}(I+s\cdot P)}$ by O_s. We would like to show that for all n'-qubit states σ, $\|D_\lambda^{\otimes n'}[(O_1 - O_0)[\sigma]]\|_{\mathsf{tr}}$ is small so that we can apply Lemma B.4. But note that for any $Q \in \{X, Y, Z\}$, $D_\lambda[Q] = (1-\lambda)Q$, whereas $D_\lambda[I] = I$. We conclude that

$$\|D_\lambda^{\otimes n'}[(O_1 - O_0)[\sigma]]\|_{\mathsf{tr}} = \left\| D_\lambda^{\otimes n'}\left[\frac{P}{2^n} \otimes \mathrm{tr}_{\mathsf{state}}(\sigma)\right]\right\|_{\mathsf{tr}} \leq \left\| D_\lambda^{\otimes n}\left[\frac{P}{2^n}\right]\right\|_{\mathsf{tr}} = (1-\lambda)^{|P|}$$

By taking the channels $\mathcal{E}$ in Lemma B.4 to be $D_\lambda^{\otimes n'} \circ O_1$ and $D_\lambda^{\otimes n'} \circ O_0$, we conclude that no algorithm given by alternately querying the oracle followed by depolarizing noise, and running arbitrary noiseless quantum computation, and finally measuring in the computational basis can distinguish whether the underlying oracle is O_+ or O_- with at least $2/3$ probability unless it makes $\Omega((1-\lambda)^{-|P|})$ queries.

As this model is a stronger model of computation than NISQ_λ (note that there is no notion of a classical oracle in this setting), this implies the claimed lower bound for NISQ_λ. □