On the complexity of quantum random circuit sampling

Yunchao Liu (UC Berkeley)

With Dorit Aharonov (Hebrew U), Xun Gao (Harvard), Zeph Landau and Umesh Vazirani (UC Berkeley)

STOC 2023; arxiv: 2211.03999

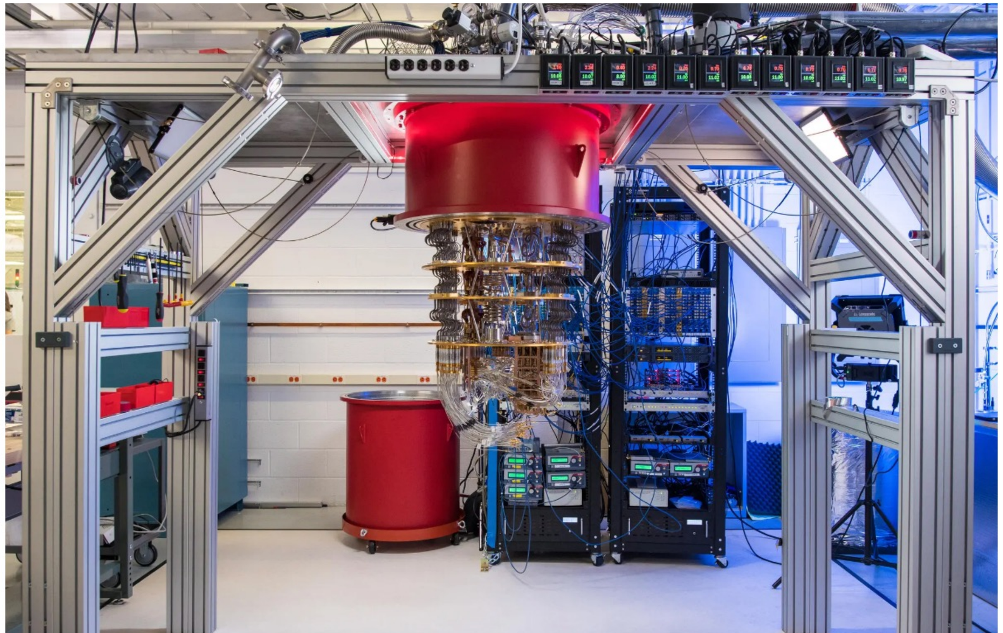
YITP long term workshop "Quantum Information, Quantum Matter and Quantum Gravity" Oct., 2023

Quantum supremacy experiments

The New York Times

Google Claims a Quantum Breakthrough That Could Change Computing

Give this article 602



Random circuit sampling (RCS):

- Use current noisy intermediate scale quantum (NISQ) devices to sample from a random quantum circuit
- Use a statistical test to evaluate how good the device is performing
- Claim that the same performance cannot be achieved classically

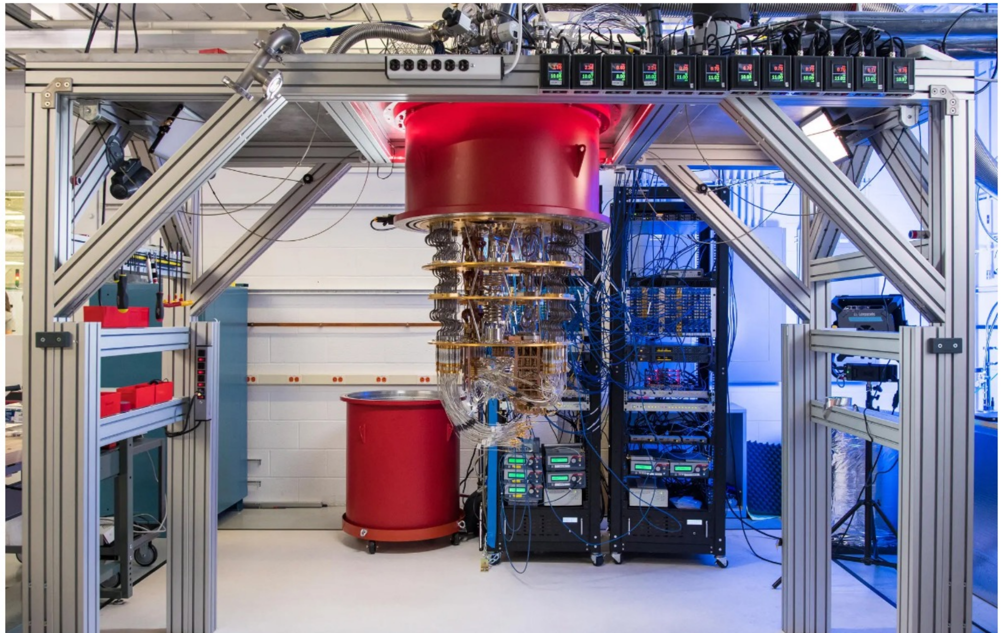
Google and USTC's 53-70 qubit experiments represent a great advance in physics experiments, and exploring the high complexity regime of quantum mechanics

Quantum supremacy experiments

The New York Times

Google Claims a Quantum Breakthrough That Could Change Computing

Give this article 602



Random circuit sampling (RCS):

- Use current noisy intermediate scale quantum (NISQ) devices to sample from a random quantum circuit
- Use a statistical test to evaluate how good the device is performing
- Claim that the same performance cannot be achieved classically

This talk: recent progress on understanding the computational complexity of RCS

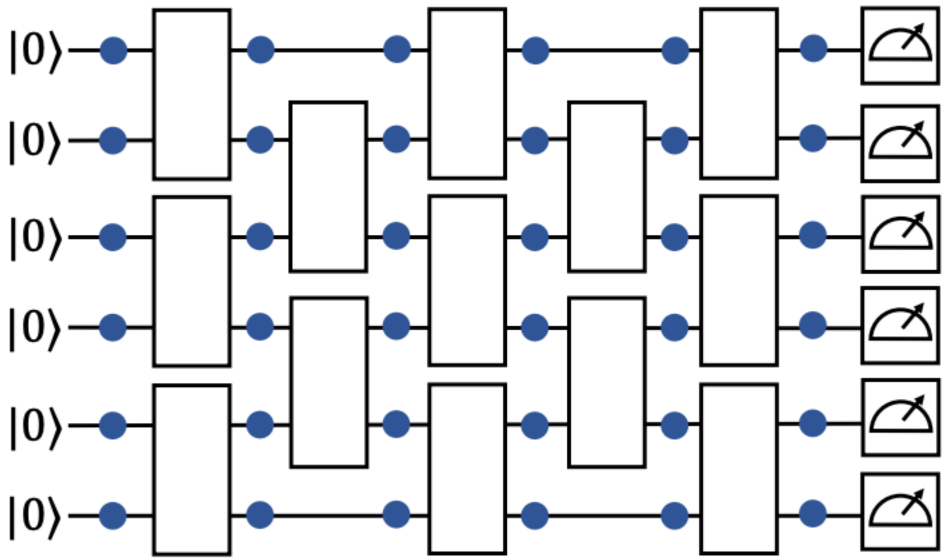
Outline

1. Overview of RCS and our main result
2. Prior work on the computational complexity of RCS
3. Proof sketch
4. Discussion & conclusions

Part I: Overview of RCS and our main result

RCS experiments

- Sample a random circuit C on n qubits with depth d
  - $d = \Omega(\log n)$ for anti-concentration
- Fix the circuit, obtain M samples from the noisy distribution $\tilde{p}(C, x)$, $x \in \{0,1\}^n$
- Compute a statistical measure $F(C, x_1, \ldots, x_M)$
  - Takes $\exp(n)$ time
- Repeat the procedure for a few circuits



(b) Noisy RCS

At each step, each qubit is subject to an arbitrarily small constant amount of noise

Google’s quantum supremacy experiment

- In 2019 Google implemented RCS on up to n=53 qubits with circuits of depth up to d=20; the device has ~1% noise per gate
  - Given M samples from the device $x_1, \ldots, x_M$, compute $XEB = 2^n \frac{1}{M} \sum_i p_{ideal}(x_i) - 1$
  - A score between 0 (uniform sampling) and 1 (ideal RCS); experiment XEB = 0.002
- [Pan, Chen, Zhang’21] developed a brute force tensor network algorithm to achieve the same task using 512 GPUs in 15 hours
  - However, this algorithm runs in exponential time and is unlikely to be practical with larger system size
- A race between classical and quantum: since 2019 there has been larger RCS experiments (Google 2023 70 qubits); will there be new spoofing algorithms for larger experiments?

Motivation: explore the high complexity regime

- Extended Church-Turing thesis: any "reasonable" model of computation can be *efficiently* simulated on a probabilistic Turing machine
- Quantum supremacy: experimental violation of ECT using NISQ devices; two aspects:
- Computational complexity: does the model of noisy RCS violate the ECT in the asymptotic sense?
- Finite size experiments: does current 53-70 qubit experiments take a lot of resources to simulate classically?

The complexity of noisy RCS

M samples	experiment	algorithm	Takes $\mathrm{poly}(M)$ time
	011111100101	100000010010	
	101000000011	111111010001	
	101100001010	100011011011	
	100100010100	001101100000	
	111110110111	011110000000	
	010001110010	111110010000	

No statistical test can tell the difference

A polynomial-time classical algorithm for noisy random circuit sampling
with Dorit Aharonov, Xun Gao, Zeph Landau, Umesh Vazirani; STOC 2023

The complexity of noisy RCS

- **Theorem.** [AGLLV'23] There is a classical algorithm that, on input a random circuit C on n qubits, outputs a sample from a distribution that is ε-close to the noisy output distribution $\tilde{p}(C)$ in total variation distance with success probability at least 0.99 over the choice of C in time $\text{poly}\left(n, \frac{1}{\varepsilon}\right) = (n/\varepsilon)^{O(1)}$

- The assumptions are anti-concentration ($\Omega(\log n)$ depth), and sufficient randomness in the gate set (see Discussion)
- Previously known $n^{O(\log 1/\varepsilon)}$ [Gao and Duan'18]
- Next: how to understand this result

Comparing classical simulation and quantum experiments

- **Theorem.** [AGLLV'23] There is a classical algorithm that, on input a random circuit C on n qubits, outputs a sample from a distribution that is ε-close to the noisy output distribution $\tilde{p}(C)$ in total variation distance with success probability at least 0.99 over the choice of C in time $\text{poly}\left(n, \frac{1}{\varepsilon}\right) = (n/\varepsilon)^{O(1)}$
- Statistical indistinguishability: two probability distributions cannot be distinguished by any statistical test on M samples (say with probability 0.51), if they are $0.01/M$ close in total variation distance
- By choosing $\varepsilon = 0.01/M$, we have running time $\text{poly}(n, M)$ to guarantee indistinguishability

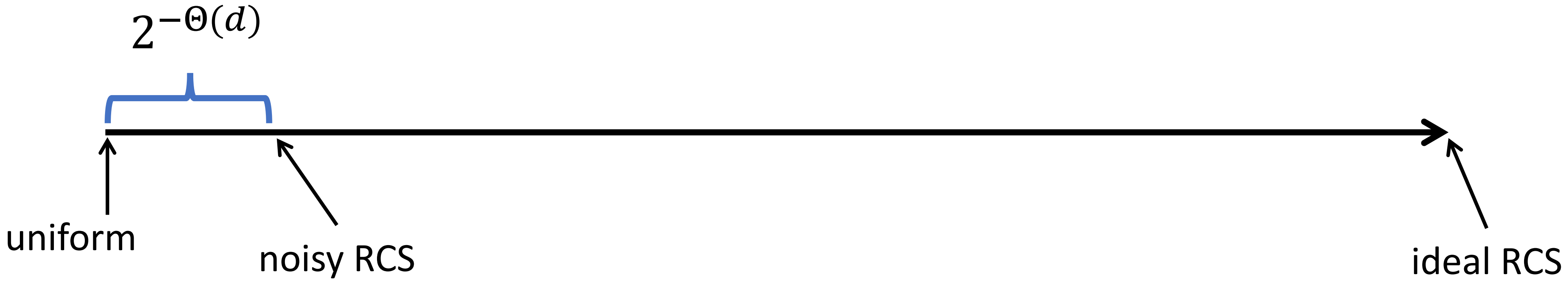
Comparing classical simulation and quantum experiments

- **Theorem.** [AGLLV'23] There is a classical algorithm that, on input a random circuit C on n qubits, outputs a sample from a distribution that is ε-close to the noisy output distribution $\tilde{p}(C)$ in total variation distance with success probability at least 0.99 over the choice of C in time $\mathrm{poly}\left(n, \frac{1}{\varepsilon}\right) = (n/\varepsilon)^{O(1)}$

- The running time of our algorithm is at most polynomial in the running time of the experiment ($\sim M$), in order to be indistinguishable from the experiment
- Currently the running time is not practical, $O(M^{1/\gamma})$ where γ is noise per gate

The role of circuit depth

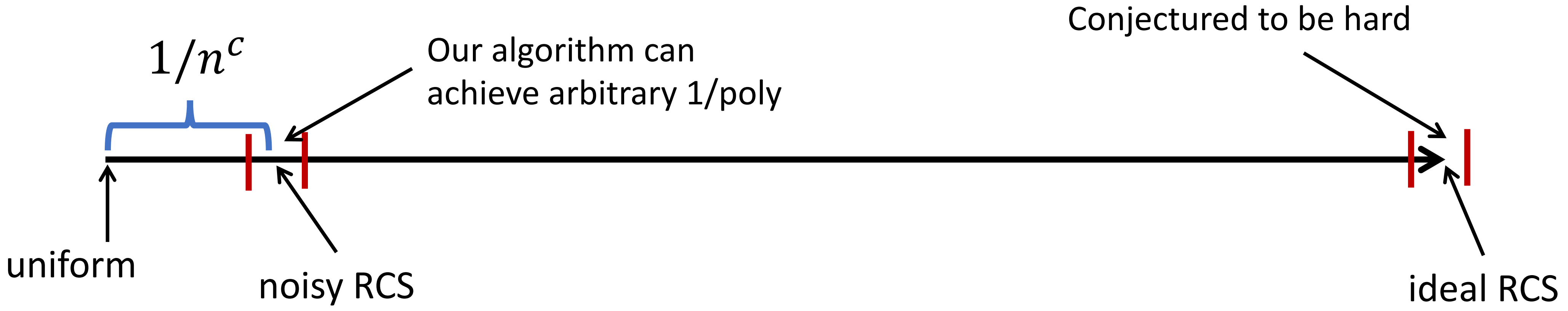
- Experimentally, need enough samples to detect a non-trivial quantum signal



- Due to noise, the output distribution of noisy RCS is $2^{-\Theta(d)}$ close to uniform
- Experimentally needs at least $M = 2^{\Omega(d)}$ samples
- In general, both the experiment and our algorithm have running time exponential in d

The role of circuit depth, $d = \Theta(\log n)$

- Anti-concentration is a central assumption for both the experiment and our algorithm, needs $d = \Omega(\log n)$
- Want the experiment to have polynomial sample complexity, needs $d = O(\log n)$
- Therefore, $d = \Theta(\log n)$ is the sweet spot for scalable quantum supremacy [Deshpande et al'21]



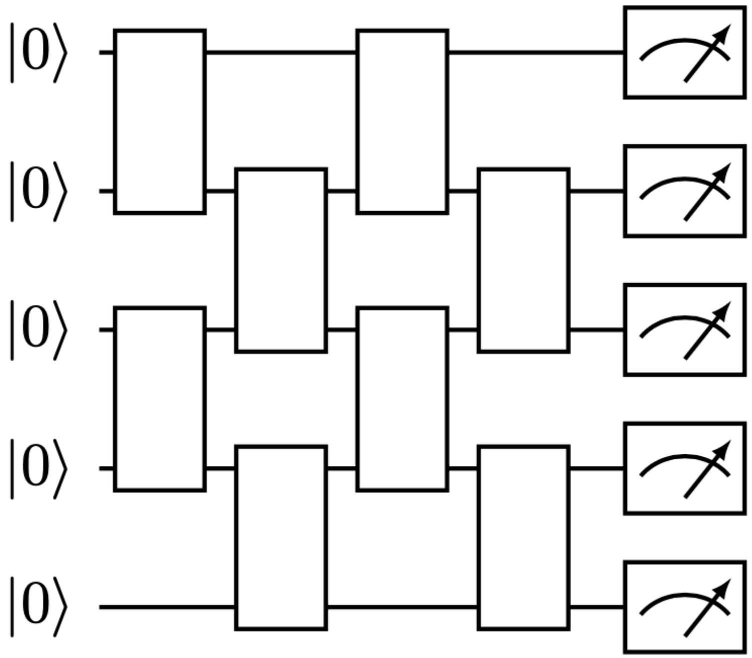
Summary

- The running time of our algorithm is at most polynomial in the running time of the experiment, in order to be indistinguishable from the experiment
  - Assuming anti-concentration ($\Omega(\log n)$ depth), and sufficient randomness in the gate set
  - In particular, at $d = \Theta(\log n)$, both the experiment and our algorithm have $\mathrm{poly}(n)$ running time

- Therefore, noisy RCS cannot be the basis of a scalable experimental violation of the extended Church-Turing thesis

Part II: Prior work on the computational complexity of RCS

The first genre: ideal RCS



1101100000101111100010011011011110011110100101110101,
1111000010101010111011101110000000010001111011101001,
00010001011010100010110010000101000000110100001010010...

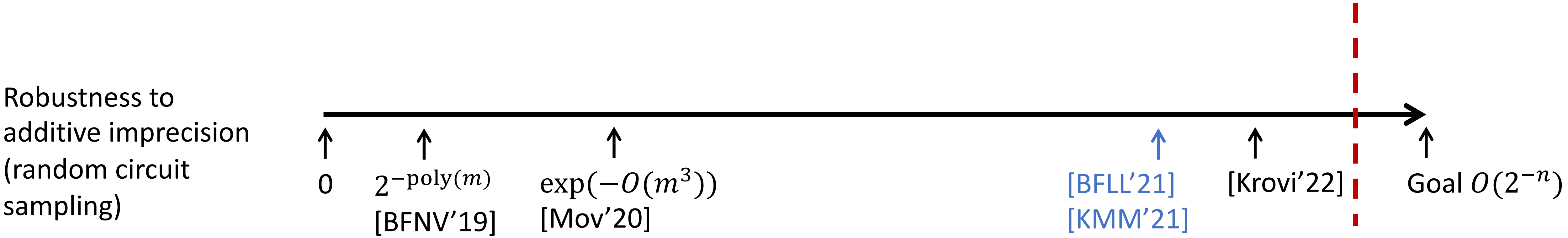
Hardness of ideal RCS:

Goal: Prove it is classically hard to sample from a distribution that is ε-close to the ideal distribution in total variation distance

By known reductions [Stockmeyer'85, AA'11], assuming anti-concentration, suffices to show #P hardness to compute $|\langle 0^n|C|0^n\rangle|^2$ within additive error $\varepsilon/2^n$ for a random circuit C

Improved robustness in the ideal regime

Task	Early result	Result of [BFLL'21] and [KMM'21]	Result of [Krovi'22]	Goal
Random circuit sampling (n qubits, m gates)	$2^{-\text{poly}(m)}$ [BFNV'19] $\exp(-O(m^3))$ [Mov'20]	$\exp\big(-O(m\log m)\big)$	$\exp\big(-O(m)\big)$	$O(2^{-n})$



The second genre: high noise regime

- Instead of being close to the output distribution of ideal RCS in TVD, actual experiments only achieve a tiny correlation with the ideal distribution due to noise; want to show this is still hard classically
- Linear cross entropy [Google'19] (n=53):
- Given M samples from the device $x_1, \ldots, x_M$, calculate the output probabilities of the ideal circuit, compute $2^n \frac{1}{M} \sum_i p_{ideal}(x_i) - 1$
  - In expectation, this equals 0 if the samples are uniform
  - If the samples are from p_{ideal}, this is related to the 2nd moment of p_{ideal}
- Intuition: if the experimental distribution is more correlated with p_{ideal}, then this quantity tends to be larger

The second genre: high noise regime

- Instead of being close to the output distribution of ideal RCS in TVD, actual experiments only achieve a tiny correlation with the ideal distribution due to noise; want to show this is still hard classically
- $XEB = 2^n \mathbb{E}_{C,x\sim p_{exp}} p_{ideal}(x) - 1 = 2^n \mathbb{E}_C \sum_x p_{exp}(x) p_{ideal}(x) - 1$
  - When exp = uniform, XEB = 0; when exp = ideal, XEB ≈ 1 (anti-concentration)
- Hard to estimate as p_{ideal} takes 2^n time to compute, but there are ways to compute at small sizes and heuristically extrapolate to large size
  - The heuristic extrapolation works well above $\log(n)$ depth
  - Google's experiment on n=53 qubits and d=20 achieves XEB=0.002, only achieves a tiny correlation with ideal RCS

Evidence of high complexity in noisy regime

- Focus on noisy regime: want to show even the tiny XEB (0.002 in Google's experiment) in experiments is hard to achieve classically
- [Aaronson and Gunn'19] formulated the XQUATH conjecture, which says that even a tiny correlation (order 2^{-n}) with the ideal RCS distribution is hard to achieve classically
  - Similar to the QUATH conjecture of [Aaronson and Chen'16]
  - The strong parameter (order 2^{-n}) was necessary to support the hardness of tiny XEB
- This provided a way to heuristically argue that even the very small XEB achieved in actual 53-70 qubit experiments was a classically difficult computational task

Evidence of high complexity in noisy regime

- However, the work of [Gao et al'21] cast doubt on these arguments; specifically, it shows $2^{-O(d)}$ correlation can be achieved classically
  - However, even if the original strong conjectures are false, there could be a weaker conjecture that still supports the hardness of noisy experiments
  - The result only specifically targets the XEB test; the other statistical tests could still be hard to achieve classically
- This reopens the question: is there high complexity in noisy RCS experiments?
- We show that no statistical test can distinguish between the experiment with M samples and our $\text{poly}(M)$ time algorithm

Summary

- Ideal RCS is believed to be classically hard; currently we can prove weaker hardness results
  - Barriers to improving existing hardness results
- Noisy RCS is classically simulable in polynomial time
  - Assuming anti-concentration and randomness in the gate set
  - Not yet a practical algorithm

Interlude: progress on practical simulation

- $XEB = 2^n \mathbb{E}_{C,x\sim p_{alg}} p_{ideal}(x) - 1 = 2^n \mathbb{E}_C \sum_x p_{alg}(x) p_{ideal}(x) - 1$
  - When alg = uniform, XEB = 0; when alg = ideal, XEB = 1
- [Google'19] achieves 0.2% XEB, claims 10000 years classical running time on the largest supercomputer using the best algorithm then
- Since then, much progress has been made with practical tensor network algorithms
- [Pan, Chen and Zhang'21] used brute-force tensor network simulation to achieve the same XEB using 512 GPUs in 15 hours

Interlude: progress on practical simulation

- Problem: these brute force algorithms are inherently exponential time, therefore become impractical if the system size increases by a few qubits
- Currently, the largest RCS experiment on 70 qubits [Google'23] has not been challenged
- [Gao et al'21] algorithm is scalable with system size, but currently achieves 10% of the XEB of Google's 2019 experiment
- An interesting future direction is to develop practical implementations of our algorithm

Part III: Proof sketch

Prior argument: Feynman path integral

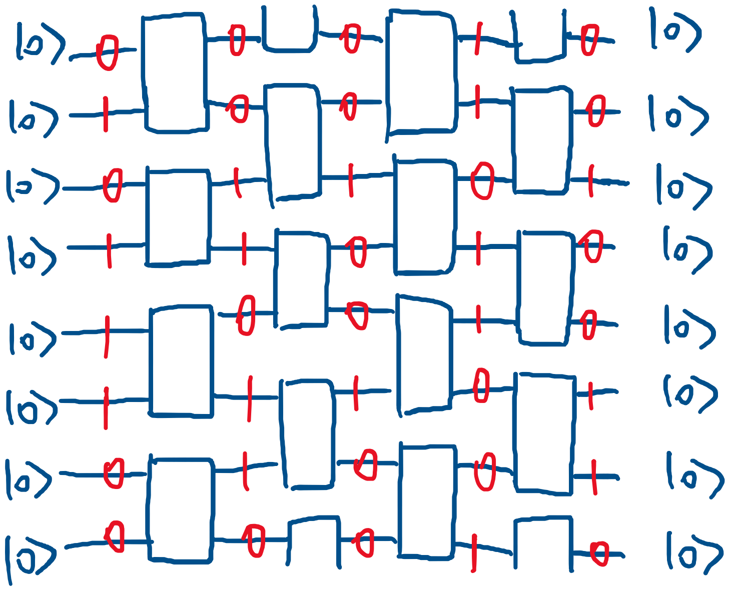
- Let $C = U_d \ldots U_2 U_1$ be a random circuit, Feynman path integral:

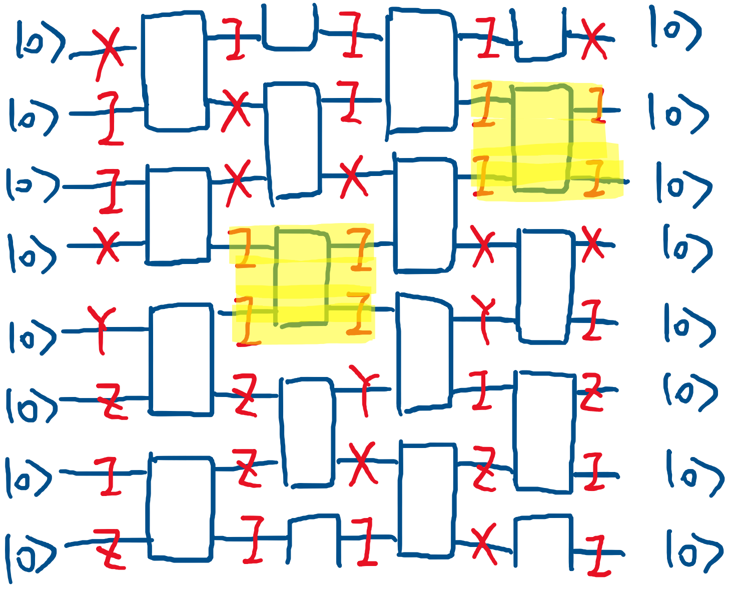
$$\langle 0^n|C|0^n\rangle = \sum_{x_1,\ldots,x_{d-1}\in\{0,1\}^n} \langle 0^n|U_d|x_{d-1}\rangle \; \langle x_{d-1}|U_{d-1}|x_{d-2}\rangle \cdots \langle x_1|U_1|0^n\rangle$$

- Intuition [Aaronson and Gunn'19]: each path contributes equally, there are exponentially (2^{nd}) many paths in total, if we sum over $\mathrm{poly}(n)$ random paths, only gets exponentially small correlation
- Therefore, conjecture that no classical algorithm can achieve better than $1/2^n$ correlation

Our algorithm: Pauli path integral

The contribution is uniform





Due to noise, the contribution decays exponentially with #non-I

Main idea: (1) in the Pauli basis the paths are nonuniform; order the paths by importance, only consider the most important paths
(2) Design an efficient algorithm to calculate those important paths; the algorithm uses the unitarity constraint

Idea (1): non-uniformity of Pauli paths

- Idea: consider Feynman path integral in Pauli basis, then the contribution from a low-weight path is much higher than a high-weight path due to noise
- Step 1: switch from vector basis to operator basis (think about density matrix)
- Step 2: the density matrix at each layer is a linear combination of Pauli operators; think about evolving Pauli operators
  - Vector basis: transition amplitude from i to j is $\langle j|U|i\rangle$
  - Pauli basis: "transition amplitude" from s_i to s_j is $\mathrm{Tr}(s_j U s_i U^\dagger)$; s_i, s_j are Paulis

$$|\langle 0^n|C|0^n\rangle|^2 = \sum_{s_0,\ldots,s_d \in \mathsf{P}_n} \mathrm{Tr}(|0^n\rangle\langle 0^n|\, s_d)\, \mathrm{Tr}\left(s_d U_d s_{d-1} U_d^\dagger\right) \cdots \mathrm{Tr}\left(s_1 U_1 s_0 U_1^\dagger\right) \mathrm{Tr}(s_0\, |0^n\rangle\langle 0^n|) = \sum_s f(C, s)$$

Idea (1): non-uniformity of Pauli paths

- Idea: consider Feynman path integral in Pauli basis, then the contribution from a low-weight path is much higher than a high-weight path due to noise
- Depolarizing noise: $I \to I; \quad X, Y, Z \to (1-\gamma) X, Y, Z$
- Pauli path integral:
  - $p(C, 0^n) = \sum_s f(C, s)$
  - $\tilde{p}(C, 0^n) = \sum_s (1-\gamma)^{|s|} f(C, s)$
- The contribution of a Pauli path in a noisy circuit decays exponentially with its Hamming weight
- Algorithm: compute $\sum_{s:|s| \leq \ell} (1-\gamma)^{|s|} f(C, s)$, choose $\ell = O(\log 1/\varepsilon)$

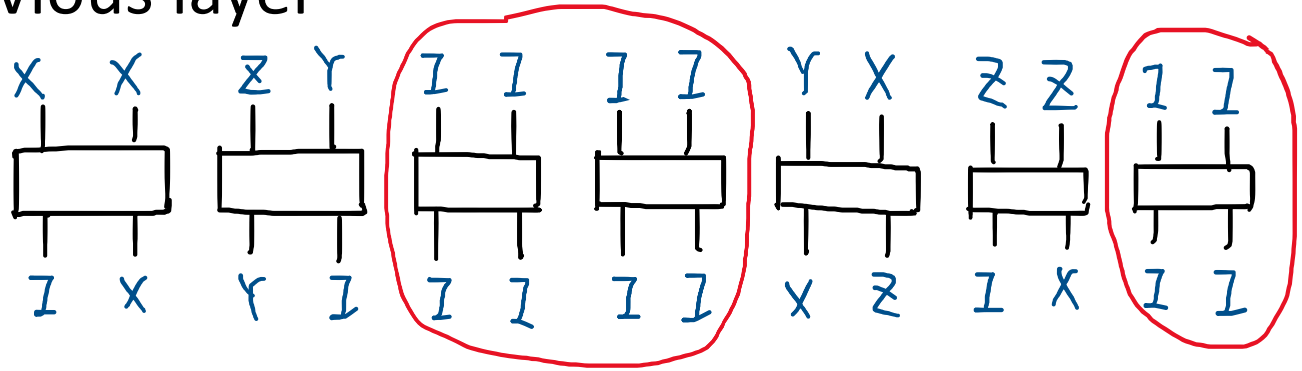
Bounding the truncation error

- Algorithm: compute $\sum_{s:|s|\leq \ell}(1-\gamma)^{|s|} f(C,s)$, choose $\ell = O(\log 1/\varepsilon)$ to achieve total variation distance ε
  - The bound is nontrivial as each $f(C,s)$ can be both positive and negative

- The proof uses two properties of random circuits:
- **Orthogonality:** $\mathbb{E}_C[f(C,s)f(C,s')] = 0$ when $s \neq s'$
- **Anti-concentration:** $\mathbb{E}_C \sum_{x\in\{0,1\}^n} p(C,x)^2 = O(1)\cdot 2^{-n}$
- Proof: use Cauchy-Schwarz to convert to L2; orthogonality kills all cross terms and gives a sum-of-square; that can be bounded using AC

Idea (2): efficient enumeration of Pauli paths

- Unitarity: identity only goes to identity; nonidentity only goes to nonidentity
  - $\langle j|U|i\rangle$ can be non-zero for any i, j
  - $\mathrm{Tr}(s_j U s_i U^\dagger)$ is only non-zero when both s_i, s_j are identity, or both are non-identity
  - A non-zero Pauli path must satisfy this constraint everywhere
- Continuity: the configuration of a layer cannot deviate too much from the previous layer



Idea (2): efficient enumeration of Pauli paths

- Unitarity: identity only goes to identity; nonidentity only goes to nonidentity
  - $\langle j|U|i\rangle$ can be non-zero for any i, j
  - $\mathrm{Tr}(s_j U s_i U^\dagger)$ is only non-zero when both s_i, s_j are identity, or both are non-identity
  - A non-zero Pauli path must satisfy this constraint everywhere
- Continuity: the configuration of a layer cannot deviate too much from the previous layer
  - Using this we design an enumeration algorithm that calculates all non-zero paths below weight ℓ in time $2^{O(\ell)} = \mathrm{poly}(1/\varepsilon)$

Part IV: Discussion & conclusions

Assumptions in our main result

- **Anti-concentration:** we assume anti-concentration $\mathbb{E}_C \sum_x p(C,x)^2 = O(1) \cdot 2^{-n}$, which is proven for certain architectures and is believed to hold above log depth for general architectures [Dalzell, Hunter-Jones, Brandão'20]

- What about sub logarithmic depth random circuits?
  - Theoretically, it is even unclear if ideal RCS is hard (hardness arguments assume anti-concentration); for example, [Napp et al'19] gave evidence that ideal RCS in 2D with very small depth is classically simulable
  - Existing RCS experiments rely on Porter-Thomas for statistical benchmarking, which is stronger than anti-concentration

Assumptions in our main result

- **Randomness in the gate set:** we assume the gate set is closed under random Pauli gates; this implies **orthogonality**
  - e.g., holds for Haar random 2-qubit gates, or fixed 2-qubit gate + Haar random single qubit gates

- What about less random gate sets?
  - Need at least *some* randomness for e.g. producing Porter-Thomas behavior
  - While we do not know if the result provably works for Google and USTC's gate sets, it works for a closely related gate set
    - Inserting random Z rotations

Conclusion

- RCS is an exciting experiment with multiple aspects:
  - Benchmarking quantum devices
  - Current back-and-forth with classical spoofing algorithms inspires the continued improvement of quantum devices
- Issues with scaling up:
  - Theoretically, we give strong negative evidence for RCS as a scalable experimental violation of the extended Church-Turing thesis
  - Practically, harder to perform verification as the system gets bigger
- It's an exciting time to start developing new proposals for near-term quantum computational advantage, with better complexity foundation

Future: the next challenge problem

- RCS and quantum supremacy experiments provided a clear target, which motivated a giant leap in the development of larger and better quantum devices

- The accumulated experimental advances and theoretical understanding in complexity theory provides the foundation for the next challenge problem for the next generation of NISQ devices