Characterizing Quantum Supremacy in Near-Term Devices

S. Boixo

S. Isakov, V. Smelyanskiy, R. Babbush, M. Smelyanskiy,
N. Ding, Z. Jiang, M. J. Bremner, J. Martinis, H. Neven

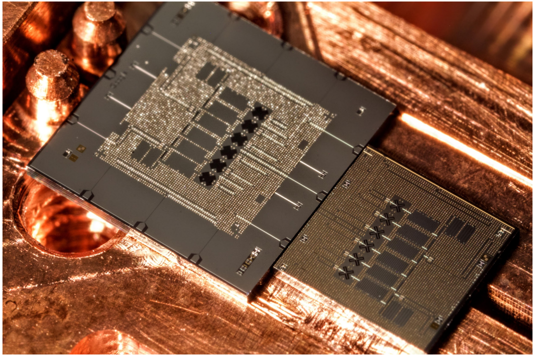
Google

January 19th

Beyond-classical Computing
AKA Quantum Supremacy, J. Preskill, 2012

With a quantum device

- perform a well-defined computational task
- beyond the capabilities of state-of-the-art classical supercomputers
- in the near-term
  - without error correction (shallow circuits with high fidelity gates).



Not necessarily solving a practical problem.

Beyond-classical computing in the near-term

- We want a computational task which requires direct simulation of quantum evolution.
  - Cost exponential in number of qubits.
  - Typical of chaotic systems (no shortcuts).
- Specific figure of merit for the computational task, related to fidelity.
- Relation to Computational Complexity.
  - Previous work in sampling problems, such as BosonSampling (Aaronson and Arkhipov) and Commuting Circuits (M. Bremner et. al.).
  - Recent conjecture by Aaronson and Chen: for a random circuit C of depth $\sim \sqrt{n}$ there is no polynomial-time classical algorithm that guesses if $|\langle 0^n| C |0^n\rangle|^2$ is greater than the mediam of $|\langle x| C |0^n\rangle|^2$ with success probability $1/2 + \Omega(1/2^n)$.
  - Nevertheless, formal Computational Complexity is asymptotic, requires error correction (Strong Church-Turing Thesis). We don't know how to satisfy the previous conjecture in the near term.

Random Universal Quantum Circuits

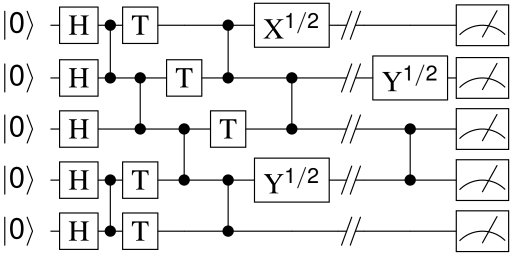


Figure: Vertical lines correspond to controlled-phase gates .

- Random quantum circuits are examples of quantum chaos.
- Classically sampling $p_U(x) = |\langle x| U |0\rangle|^2$ expected to require direct simulations. Cost in $2D$ exponential in $\propto \min(n, d\sqrt{n})$, with n qubits, d depth. (7×7 qubits requires $d \simeq 40$ with current constraints.)
- Good benchmark for quantum computers.
- New results in computational complexity.

Porter-Thomas distribution

- (Pseudo-)random circuit U (random gates from universal set)

$$|\Psi\rangle = U|0\rangle = \sum_{j=1}^{N} c_i |x_i\rangle \ .$$

- Sample the output distribution with probabilities

$$p_i = |c_i|^2 = |\langle x_i| U |\Psi\rangle|^2 \ .$$

- Real and imaginary parts of c_i are distributed (quasi) uniformly on a $2N$ dimensional sphere (Hilbert space) if the circuit (or Hamiltonian evolution) has sufficient depth (evolution time).
  - The distribution of c_i is, up to finite moments, Gaussian with mean 0 and variance $\propto 1/N$.
- Porter-Thomas distribution: $\mathrm{Pr}(Np) = e^{-Np}$.

Verification and uniformity test

- There is no polynomial witness for this sampling problem. This problem is much harder than NP.
  - This is required for near-term (few qubits) supremacy.
- The PT distribution is very flat: $p(x_j) \sim 1/N$.
- If we don't know anything about $p(x_j)$ (black-box setting) we need $\Theta(\sqrt{N})$ measurements to distinguish from uniform over bit-strings.
- The ℓ_1 distance between PT and uniform distribution is

$$\sum_j |p(x_j) - 1/N| = 2/e \,.$$

- If we calculate $p(x_j)$ given circuit U, we can distinguish these distributions with a constant number of measurements.
  - Hardware verification.

Convergence to chaos (Porter-Thomas distribution)

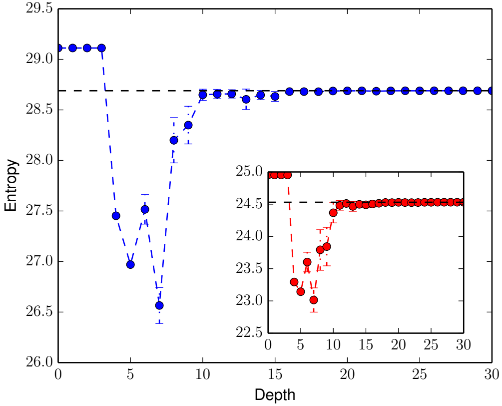


Figure: Depth required for PT distribution. Dashed line is known entropy for PT. 2D circuit 7 $\times$ 6 qubits. Inset 6 $\times$ 6 qubits.

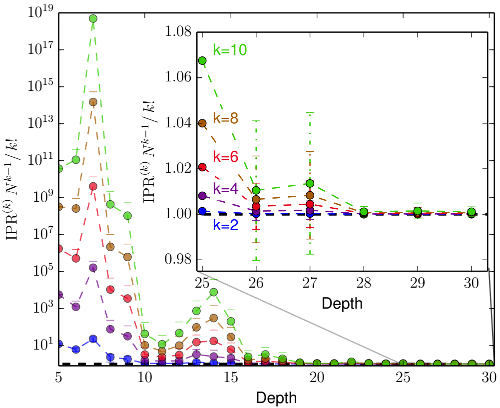


Figure: Participation ratios $\mathrm{PR}^{(k)} \simeq N\langle p^k \rangle$ with $k = 2, 4, 6, 8, 10$, normalized to 1 for PT distribution. Related to t-designs. 7 $\times$ 6 qubits.

For decoupling in random circuits (Brown & Fawzi), and anti-concentration on sparse-IQP (Bremner et. al.), depth scales like $\tilde{O}(\sqrt{n})$. We expect this for 2-designs, entropy (Nahum et. al.), out-of-time ordered correlator,...

Sampling from ideal circuit U

Sample $S = \{x_1, \ldots, x_m\}$ of bit-strings x_j from circuit U (measurements in the computational basis).

$$\log \mathrm{Pr}_U(S) = \sum_{x_j \in S} \log p_U(x_j) = -m\,\mathrm{H}(p_U) + O(m^{1/2}) \ ,$$

where $H(p_U)$ is the entropy of PT

$$\mathrm{H}(p_U) = -\int_0^\infty pN^2 e^{-Np} \log p \, dp = \log N - 1 + \gamma \ .$$

and $\gamma \simeq 0.577$.

Sampling with polynomial classical circuit $A_{\mathrm{pcl}}(U)$

A *polynomial* classical algorithm $A_{\mathrm{pcl}}(U)$ produces sample $S_{\mathrm{pcl}} = \{x_1^{\mathrm{pcl}}, \ldots, x_m^{\mathrm{pcl}}\}$. The probability $\mathrm{Pr}_U(S_{\mathrm{pcl}})$ that this sample S_{pcl} is observed from the output $|\psi\rangle$ of the circuit U is

$$\log \mathrm{Pr}_U(S_{\mathrm{pcl}}) = -m\,\mathrm{H}(p_{\mathrm{pcl}}, p_U) + O(m^{1/2}) \ ,$$

where

$$\mathrm{H}(p_{\mathrm{pcl}}, p_U) \equiv -\sum_{j=1}^{N} p_{\mathrm{pcl}}(x_j|U) \log p_U(x_j)$$

is the cross entropy.

Sampling with polynomial classical circuit $A_{\mathrm{pcl}}(U)$ (II)

We are interested in the average over $\{U\}$ of random circuits

$$\mathbb{E}_U\left[\mathrm{H}(p_{\mathrm{pcl}}, p_U)\right] = \mathbb{E}_U\left[\sum_{j=1}^{N} p_{\mathrm{pcl}}(x_j|U)\log\frac{1}{p_U(x_j)}\right] \ .$$

Because U is chaotic, Hilbert space has exponential dimension, and $A_{\mathrm{pcl}}(U)$ is polynomial, we conjecture that p_{pcl} and p_U are (almost) uncorrelated (see next). We can take averages independently.

$$-\mathbb{E}_U\left[\log p_U(x_j)\right] \approx -\int_0^\infty Ne^{-Np}\log p\, dp = \log N + \gamma \ .$$

$$\mathbb{E}_U\left[\mathrm{H}(p_{\mathrm{pcl}}, p_U)\right] = \log N + \gamma \equiv \mathrm{H}_0 \ .$$

Chaotic sensitivity to perturbations
Residual correlation after a single Pauli error

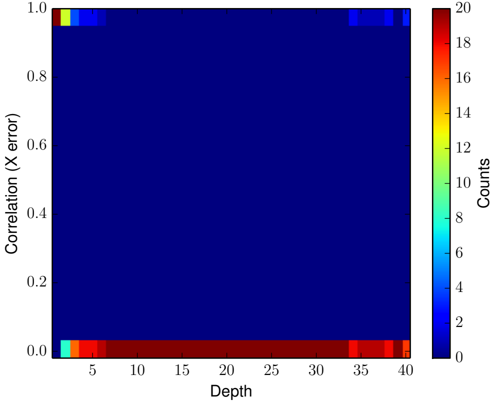


Figure: Correlation after a single X error at different depths. 5×4 qubits.

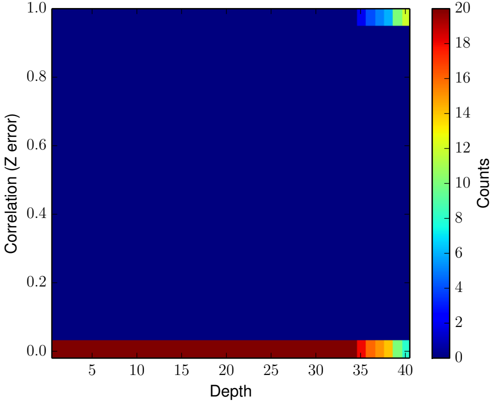


Figure: Correlation after a single Z error at different depths. 5×4 qubits.

Cross entropy and fidelity

- For algorithm A (quantum or classical of any cost) define the cross entropy difference

$$\alpha \equiv \Delta \mathrm{H}(p_A) \equiv \mathrm{H}_0 - \mathrm{H}(p_A, p_U) \ .$$

- The output of an evolution with fidelity $\tilde{\alpha}$ is

$$\rho = \tilde{\alpha} U \left|0\right\rangle\left\langle 0\right| U^\dagger + (1 - \tilde{\alpha})\sigma_U \ ,$$

with $p_{\mathrm{exp}}(x) = \left\langle x\right| \rho \left|x\right\rangle = \tilde{\alpha} p_U(x) + (1 - \tilde{\alpha}) \left\langle x\right| \sigma_U \left|x\right\rangle$.

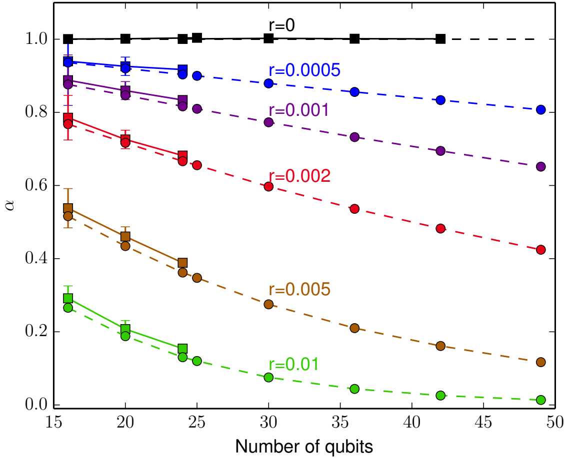
- We again conjecture that $\left\langle x\right| \sigma_U \left|x\right\rangle$ is uncorrelated with $p_U(x)$.

$$\begin{aligned}
\alpha &= \mathbb{E}_U[\Delta \mathrm{H}(p_{\mathrm{exp}})] \\
&= \mathrm{H}_0 + \sum_j \left(\tilde{\alpha} p_U(x_j) + (1 - \tilde{\alpha}) \left\langle x_j\right| \sigma_U \left|x_j\right\rangle\right) \log p_U(x_j) \\
&= \mathrm{H}_0 - \tilde{\alpha}\mathrm{H}(p_U) - (1 - \tilde{\alpha})\mathrm{H}_0 = \tilde{\alpha} \ .
\end{aligned}$$

- The cross entropy α approximates the fidelity $\tilde{\alpha}$.

Numerics and theory for realistic 2D circuits

Cross entropy difference □ and estimated fidelity ∘.



r is two-qubit gate error rate. $\alpha = 1$ for PT distribution. Depth 25.

Experimental proposal

1. Implement a random universal circuit U (chaotic evolution).
2. Take large sample $S_{\mathrm{exp}} = \{x_1^{\mathrm{exp}}, \ldots, x_m^{\mathrm{exp}}\}$ of bit-strings x in the computational basis ($m \sim 10^3 - 10^6$).
3. Compute quantities $\log p_U(x_j^{\mathrm{exp}})$ with supercomputer.

Cross entropy difference (figure of merit)

$$\alpha = \frac{1}{m}\sum_{j=1}^{m} \log p_U(x_j^{\mathrm{exp}}) + \log 2^n + \gamma \pm \frac{\kappa}{\sqrt{m}}\;, \qquad \kappa \simeq 1,\; \gamma = 0.577$$

Measure and extrapolate α (size, depth, T gates).
Fit to theory: α approx. circuit fidelity, chaotic state very sensitive to errors.

$$\alpha \approx \exp(-r_1 g_1 - r_2 g_2 - r_{\mathrm{init}} n - r_{\mathrm{mes}} n)\;,$$

$r_1, r_2 \ll 1$ one and two-qubit gates Pauli error rates, $g_1, g_2 \gg 1$ number of one and two-qubit gates, $r_{\mathrm{init}}, r_{\mathrm{mes}} \ll 1$ initialization and measurement error rates.

Complex Ising models from universal circuits

- As in a path integral, the output amplitude of U is

$$\langle x| U |0\rangle = \sum_{\{s^t\}} \prod_{t=0}^{d} \langle s^t| U^{(t)} |s^{t-1}\rangle , \quad |s^d\rangle = |x\rangle \ .$$

  where $|s^t\rangle = \otimes_{j=1}^{n} |s_j^t\rangle$ is the computational basis, $s_j^t = \pm 1$, and $U^{(t)}$ are gates at clock cycle t.

- Gates give Ising couplings between spins s_j^k, like in path integral QMC. For instance, for $\mathrm{X}^{1/2}$ gates

$$\frac{i\pi}{4} H_s^{\mathrm{X}^{1/2}} = \frac{i\pi}{2} \sum_{j=1}^{n} \sum_{k=0}^{d(j)} \alpha_j^k \frac{1 + s_j^{k-1} s_j^k}{2} \ .$$

  where $\alpha_j^k = 1$ if a $\mathrm{X}^{1/2}$ gate was applied at qubit j in (clock cycle) k.

Computational complexity

- For universal circuits, $p_U(x) = \lambda|Z|^2$ is proportional to the partition function $Z = \sum_s e^{i\theta H_x(s)}$ of an Ising model $H_x(s) = h_x \cdot s + s \cdot \hat{J} \cdot s$ with complex temperature $i\theta (= i\pi/8)$ and no structure.
- Z has a strong sign problem: $Z = \sum_j M_j e^{i\theta E_j}$, number of paths M_j for phase θE_j exponentially larger than $|Z|$.
- Worst-case complexity: Z can not be probabilistically approximated asymptotically with an NP-oracle (is #P-hard). (Fujii and Morimae 2013, Goldberg and Guo 2014).
- Computational complexity conjecture: average case = worst case complexity. There is no structure. (Bremner et. al. 2015).
- Theorem: if $p_U(x)$ can be classically sampled, then Z can be approximated with an NP-oracle (Bremner et. al. 2015). Contradiction.
- Connection to complex Ising model gives interesting perspective.
    - Complex temperature corresponding to Clifford gates independently known to be easy.

Simulation time

% of comm	# of sockets	# of fused	Avg. time per gate (sec)	Time per Depth-25 (sec)
5×4 *circuit*: 20 qubits, 10.3 gates per level, **17 MB of memory**				
0.0%	1	0.00	0.00015	**0.039**
6×4 *circuit*: 24 qubits, 12.5 gates per level, **268 MB of memory**				
0.0%	1	7.01	0.0041	**1.294**
6×5 *circuit*: 30 qubits, 16.2 gates per level, **17 GB of memory**				
0.0%	1	5.64	0.349	**141.3**
6×6 *circuit*: 36 qubits, 19.5 gates per level, **1 TB of memory**				
6.2%	64	5.40	0.76	**369.0**
7×6 *circuit*: 42 qubits, 23.0 gates per level, **70 TB of memory**				
11.2%	4,096	5.54	1.72	**989.0**

On Edison, a Cray XC30 with 5,576 nodes. Each node is dual-socket Intel®Xeon E5 2695-V2 with 12 cores per socket, 2.4GHz. 64GB per node (32GB per socket). Nodes connected via Cray Aries with Dragonfly topology. (Mikhail Smelyanskiy).

Some open questions

- Practical computations with near-term small low-depth high-fidelity quantum circuits. Details matter.
  - Quantum chemistry.
  - Approximate optimization.
- Experimental proof of error-correction.
- Solve the control problem (see, i.e., D-Wave).
- Improve fidelity (coherent and incoherent errors).
- Complexity theory without full error correction. (Bremner et. al. arXiv:1610.01808.)
- Improve bounds in $2D$ random circuits: anti-concentration, t-designs, complexity bounds.
- Optimal classical-simulation algorithms. Details matter.

Conclusions

- We expect to be able to approximately sample the output distribution of shallow random circuits of 7×7 qubits with significant fidelity in the near term.
- We don't know how to approximately sample the output distribution of shallow random quantum circuits of ≈ 48 qubits with state-of-the-art supercomputers ($d \sim 40$).
- Beyond-classical computing.
- New method to benchmark complex quantum circuits efficiently.
- Relation to computational complexity.
- The cross entropy method applies to other sampling problems: chaotic Hamiltonians, commuting quantum circuits.

Numerical distribution with digital noise

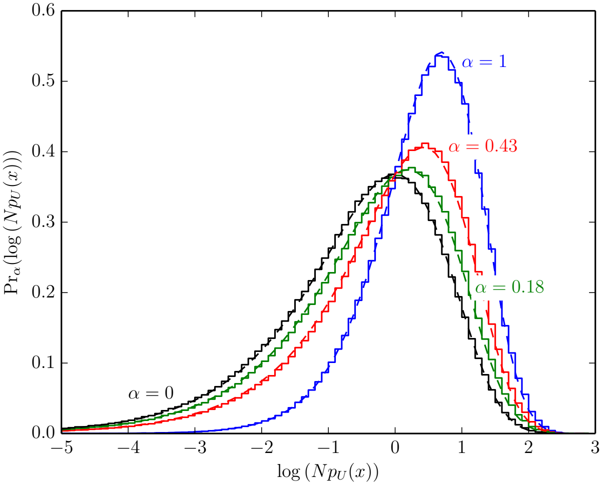


Figure: Numerical distribution with digital noise fits well a mixture of ideal + completely mixed distribution.

Porter-Thomas distribution

Histogram of the output distribution for different values of the two-qubit gate error rate *r*.

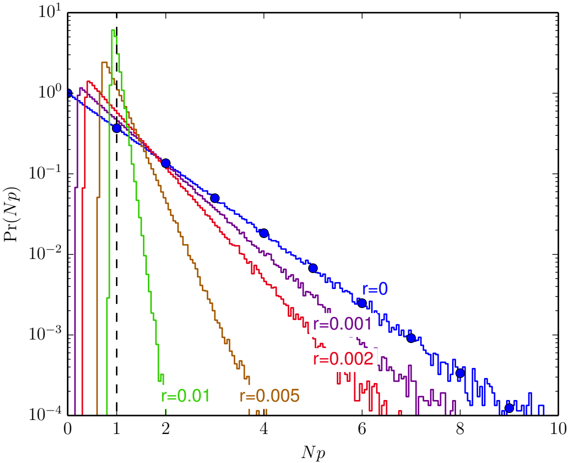


Figure: Circuit with 5 $\times$ 4 qubits (2D lattice) and depth 40.

Convergence to chaos (II)

For decoupling in random circuits (Brown & Fawzi), and anti-concentration on sparse-IQP (Bremner et. al.), depth scales like $\tilde{O}(\sqrt{n})$. We expect this for 2-designs, entropy (Nahum et. al.), out-of-time ordered correlator,...

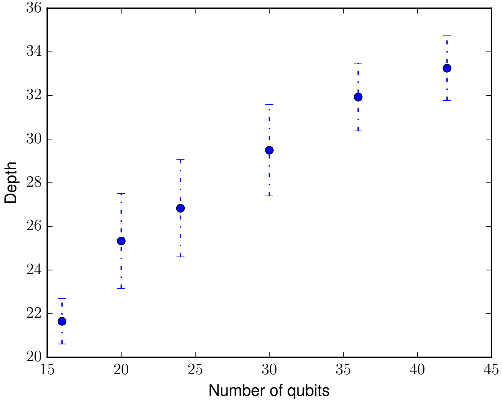


Figure: First cycle such that the entropy remains within $2^{-n/2}$ of PT entropy.