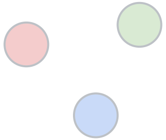
Quantum AI

The role of data, precomputation, and communication in a quantum learning landscape

Jarrod McClean
@JarrodMcClean
Senior Staff Research Scientist

Google

How we do science - a learning perspective

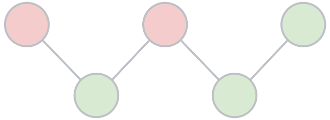


Reagents / Reactants
(Input chemicals)

$\times N \longrightarrow \infty$

How long does this typically take?
(reaction rates)

Does this ever happen?
(synthesizability)



Products
(Output chemicals)

F=ma

$$i\partial_t|\Psi\rangle = H|\Psi\rangle$$

(sometimes just 1 copy)

Wait

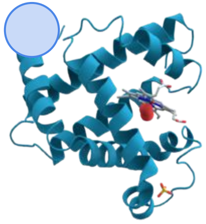
?

Quantum AI

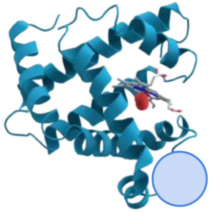
McClean et al. J. Chem. Phys. 155, 150901 (2021)

Quantum simulation and fast-forwarding

H



$$i\partial_t|\Psi\rangle = H|\Psi\rangle$$



Rapidly developing field - numerically exact evolutions in time sublinear in # basis functions*!

Exact classical competition is a hard exponential wall - entanglement truncation challenging

Even a quantum computer has limits - No Fast-Forwarding Theorem**

$$[\text{Simulation time}] > c \times [\text{Physical Time}]$$

$$c \gg 1$$

Chemical reaction times ~ hours?

* Babbush, Berry, McClean, Neven, *NPJ Quantum Information* Vol 5 No. 92 (2019)
** Berry, Ahokas, Cleve, Sanders, *Communications in Mathematical Physics* 270, 359 (2007)

Why do we focus on stationary states so often?

$$i\partial_t|\Psi\rangle = H|\Psi\rangle$$

Dynamics - BQP

$$\mathcal{H}\,|\psi\rangle = E\,|\psi\rangle$$

Stationary(often ground) States - QMA

Thermodynamics is predictive of longer time behavior

$t \longrightarrow \infty$ is predictive of $t >> 1$, I actually care about

Limited to physical time scales

+

Makes predictions on time scales much longer than physical

Thermo can’t be better for all systems (physical systems special? e.g. overlap assumptions in QPE)
Worst cases are like state enumeration / diagonalization (is this a useful perspective?)

The predictive power of (free) energies

Can all physically interesting questions be answered by some reduced model?

Recent*

- Does a system thermalize?
- Does a system have an electronic gap?
- Will molecule X ever form from constituents Y?

→ **Undecidable**

Physical undecidability** - as the system evolves in time, there are sudden, qualitative changes that cannot be predicted in any way except evolving forward in time and seeing if it happens, and no answer in finite time can indicate if it will never happen (for all systems).

Undecidability formally broken by **advice** in some cases

Data is a restricted form of **advice**

* Infinite systems specified by unit and translational invariance
Cubitt, Perez-Garcia, Wolf Nature Vol. 528, pg 207–211 (2015)
Shiraishi, Matsumoto arXiv:2012.13889 (2020)

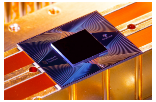
** Unpredictability and undecidability in dynamical systems
Moore, *Phys. Rev. Lett.* 64, 2354 (1990)

Quantum machine learning & data advantage

Computationally limited problems - Simple inputs, known computational procedure

x

Key to factor
Hamiltonian to simulate
...



Compute ~ n

Compute ~ exp(n)

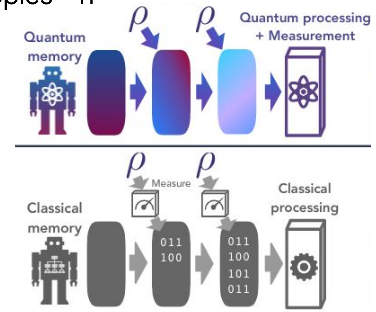
Data limited problems - Limited by availability of data, no computation possible to overcome lack of data

ρ (limited copies)

Transduced quantum state
Analog simulation state
Output of computation
...

Copies ~ n

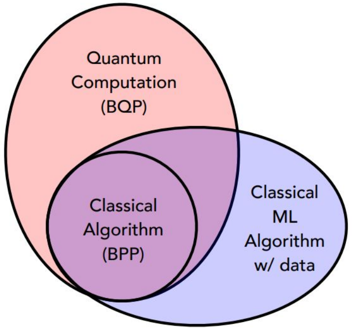
Compute ~ ?



Copies ~ exp(n)

Compute ~ ?

Data assisted problems - Known computational procedure, complexity can change with available data (~advice)



Quantum AI

A motivating example for data assisted problems

Given circuit U_{QNN} & some input $\vec{x}_i \rightarrow |x_i\rangle = \sum_{k=1}^{p} x_i^k |k\rangle$

Task - compute $y_i = f(x_i) = \langle x_i | U_{\text{QNN}}^\dagger O U_{\text{QNN}} | x_i \rangle$

Some data $\{(x_i, y_i)\}_i$

Arbitrary length quantum circuit

Hermitian operator

Direct simulation at least as hard as BQP,
must be a powerful function of x_i !

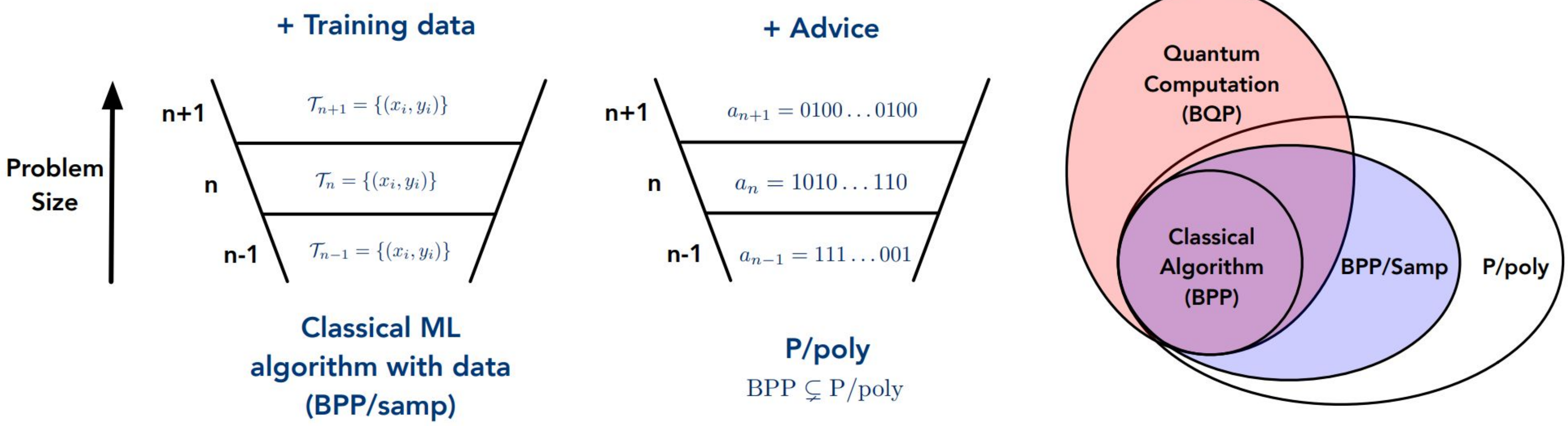
$$f(x_i) = \left(\sum_{k=1}^{p} x_i^{k*} \langle k|\right) U_{\text{QNN}}^\dagger O U_{\text{QNN}} \left(\sum_{l=1}^{p} x_i^l |l\rangle\right)$$
$$= \sum_{k=1}^{p} \sum_{l=1}^{p} B_{kl} x_i^{k*} x_i^l,$$

At most quadratic function on entries of x_i with p^2 coefficients!

No data - hard quantum circuit
With data - Almost trivial learning task!

(More generally, need ~ $\left(\frac{p^2}{\epsilon^2}\right)$ data pts)

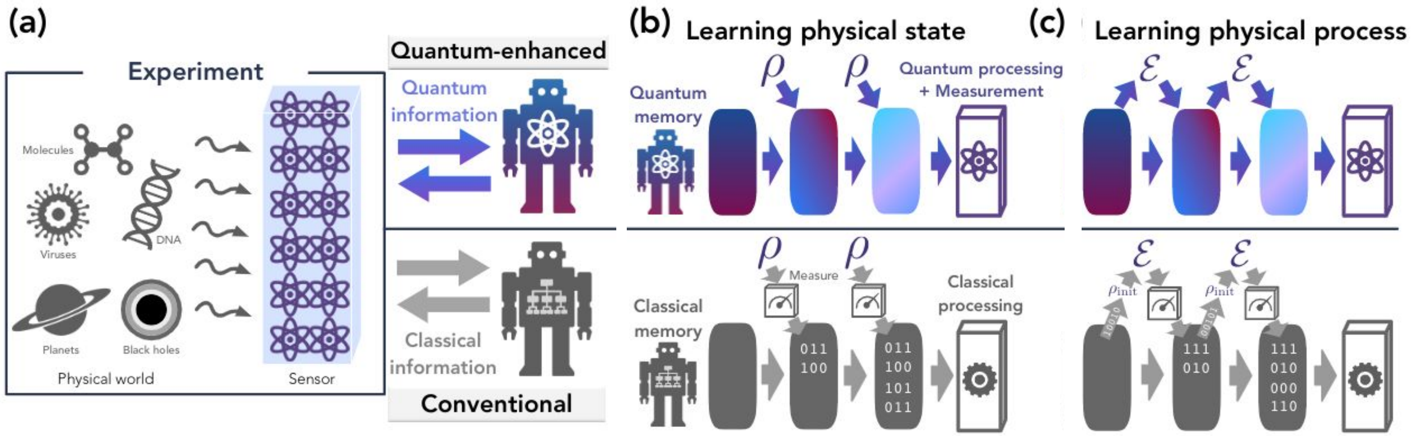
The power of data in quantum machine learning*



* Hsin-Yuan (Robert) Huang, Michael Broughton, Masoud Mohseni, Ryan Babbush, Sergio Boixo, Hartmut Neven, **Jarrod R. McClean** “Power of data in quantum machine learning” Nature Communications, Vol.12, No. 2631 (2021)

Quantum AI

Quantum memory and quantum-enhanced experiments



This work - Exponential advantage with exactly 2 copies on 2 different tasks and efficient classical compute (and additional proofs)

Quantum advantage in learning from experiments
Huang, Broughton, Cotler, Chen, Li, Mohseni, Neven, Babbush, Kueng, Preskill, McClean
Science 376, 6598 (2022)

What's the simplest task we can have an advantage on?

Setup -

N copies of $\rho \propto (I + \alpha P)$

$$\alpha \in (-1, 1)$$

P Is a general n-qubit pauli operator e.g.

$$Z \otimes X \otimes I \otimes Z \otimes \ldots = Z_1 X_2 Z_4 \ldots$$

Form of state is known

α, P unknown

Note -
State is un-entangled but **not** factorizable
Can be realized in depth 1 Clifford circuit

Task -

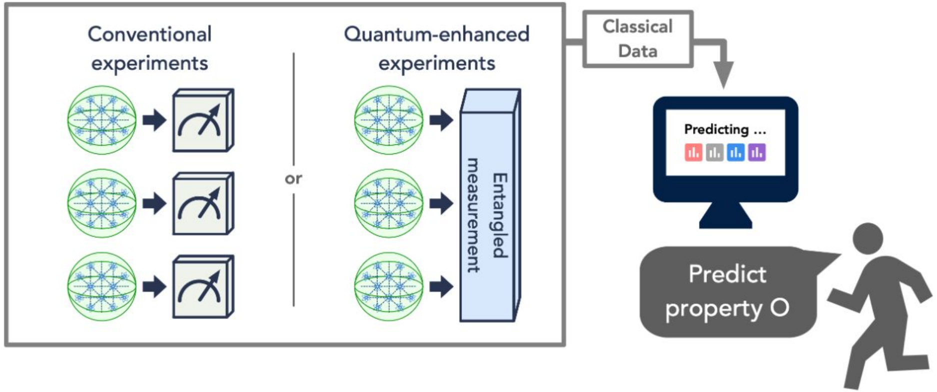
Take any measurements you want on the N copies

Collect classical data signature of state

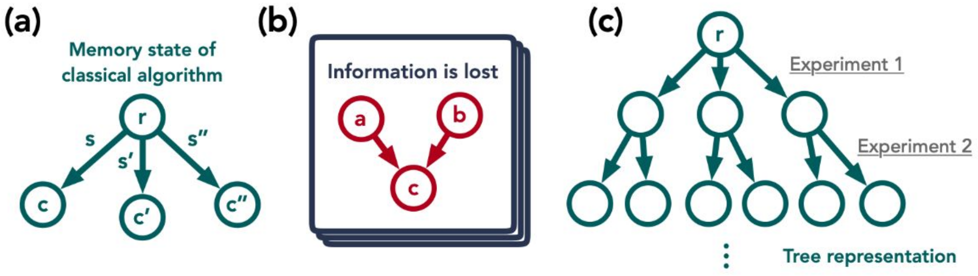
Given new Pauli operator O, predict $|\mathrm{Tr}[\rho O]|$

(**Alternatively**) Given 2 candidate Pauli operators Q_1, Q_2

Determine $|\mathrm{Tr}[Q_1 \rho]| > |\mathrm{Tr}[Q_2 \rho]|$?



The best possible conventional experiments



Result summary

Best conventional strategy requires N ~ 2^n

to predict $|\mathrm{Tr}[\rho O]|$ to additive error < .25

with probability > .8

Sketch of proof

Reduction to discrimination task

Null hypothesis

$\rho = I/2^n$

O random

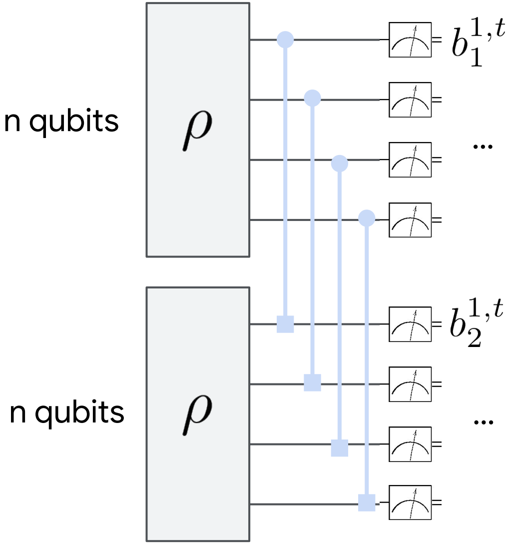
Alternate hypothesis

$\rho \propto (I + \alpha P)$

$\alpha \in \{+.9, -.9\}$ $O = P$

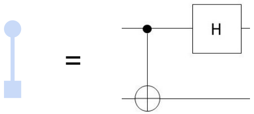
-Optimal discriminating POVM can be bounded using hypothesis structure
-Is independent of previous measurements
-Gives exp vanishing returns

The simplest quantum-enhanced experiment



Depth 1
clifford gates

Bell measurements



$$\rho \propto (I + \alpha P)$$

2n classical bits for N rounds creates Bell sketch of state $\{b_i^{k,t}\}$

$$O = \sigma_1 \otimes \sigma_2 \ldots \otimes \sigma_n$$

$$|s_k^{(t)}\rangle = \frac{1}{\sqrt{2}} I \otimes Z^{b_1^{k,t}} X^{b_2^{k,t}} (|00\rangle + |11\rangle) \qquad S_k^{(t)} = |s_k^{(t)}\rangle\langle s_k^{(t)}|$$

$$\hat{a}(O) = \frac{1}{N}\sum_{t=1}^{N}\prod_{k=1} \mathrm{Tr}\left[(\sigma_k \otimes \sigma_k) S_k^{(t)}\right] \longrightarrow$$ Estimate of $|\mathrm{Tr}[\rho O]|^2$

Samples ~ $\frac{1}{\epsilon^4}$ Computation time ~ nN

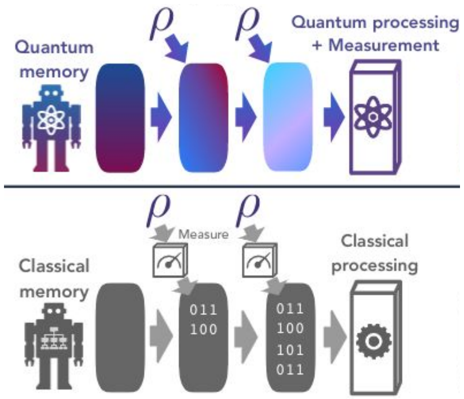
Summarizing the scale of the separation

N copies of $\rho \propto (I + \alpha P)$

Take any measurements you want on the N copies

Collect classical data signature of state

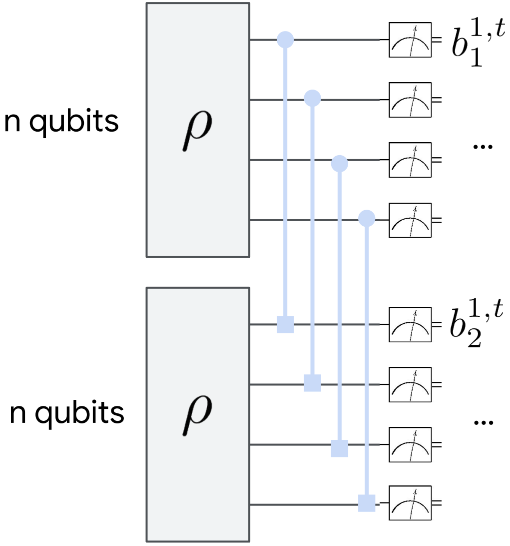
Given new Pauli operator O, predict $|\mathrm{Tr}[\rho O]|$ to error ϵ



Copies ~ $\dfrac{1}{\epsilon^4}$ Compute ~ n x Copies

Copies $> (2^n + 1)/.85$

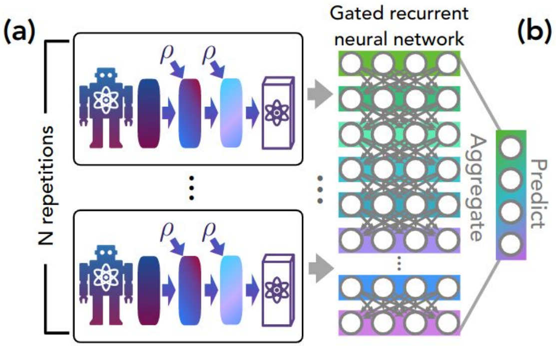
Bell measurements as a feature in learning



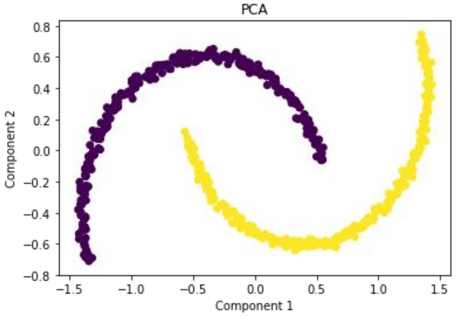
2n classical bits for N rounds creates Bell sketch of state $\{b_i^{k,t}\}$

Could we use this “**feature**” of a quantum state to learn this task?

Can state specific noise or features boost the performance? Or the performance of an adversary?

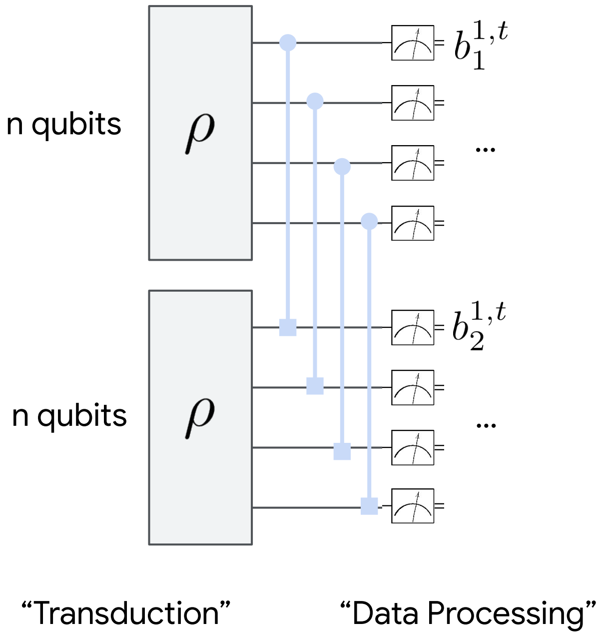


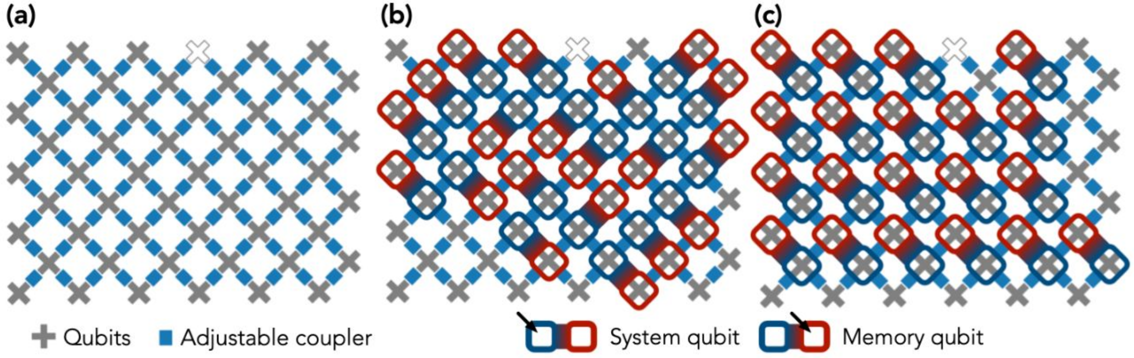
Supervised



Unsupervised

Using our chip to understand performance on real data



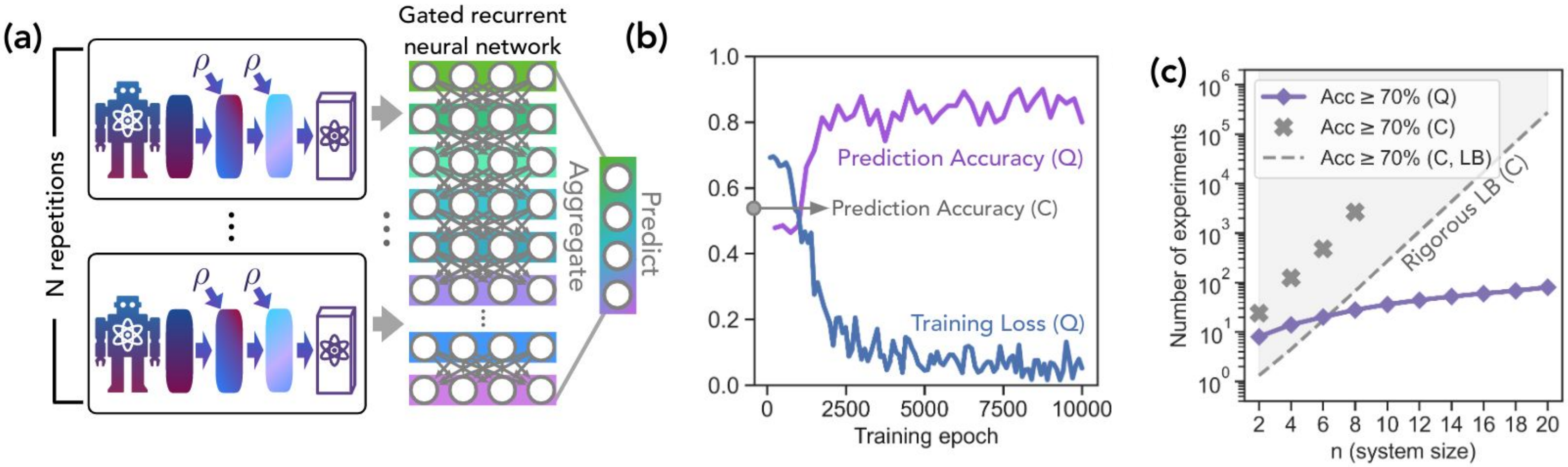


Experimental demonstration of advantage

N copies of $\rho \propto (I + \alpha P)$

Given 2 candidate Pauli operators Q_1, Q_2

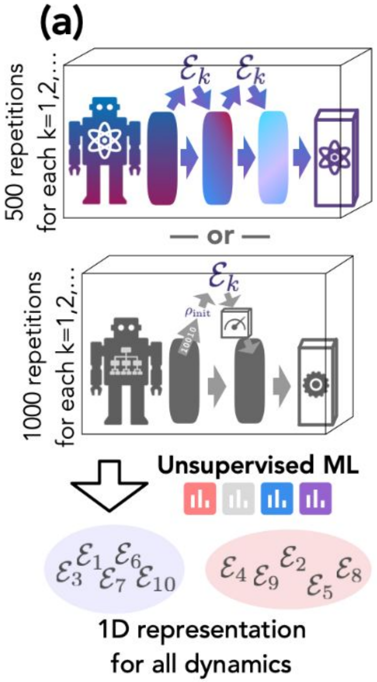
Determine $|\mathrm{Tr}[Q_1\rho]| > |\mathrm{Tr}[Q_2\rho]|?$



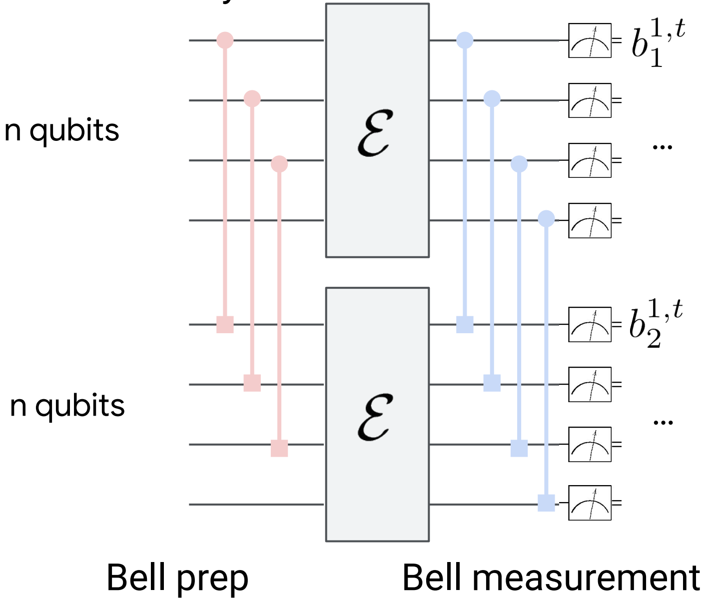
Trained on noiseless data n < 8

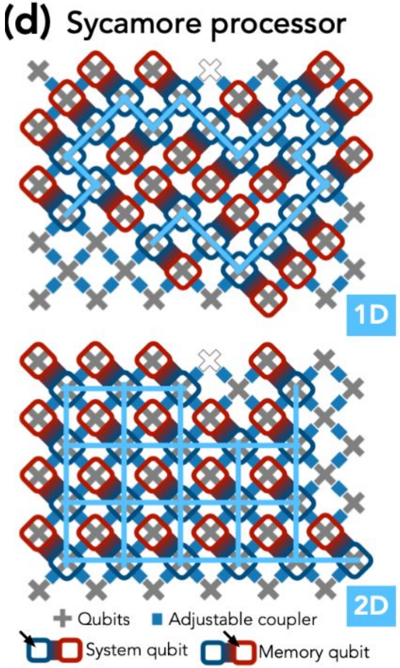
Test on n up to 20 (= 40 physical qubits)

We’ve learned about states… how about processes?



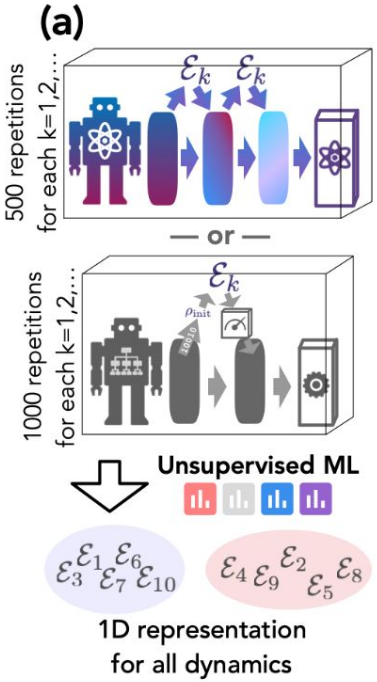
Given access to a process $\mathcal{E}$ N times, determine if
(a) random time-reversal symmetric evolution
(b) random unitary





	Number of qubits	Number of gates	Circuit depth
1D dynamics	40	842	40
2D dynamics	40	1388	54

Unsupervised discovery



Trials

$b_i^{1,t}$

...

Avg and Variance

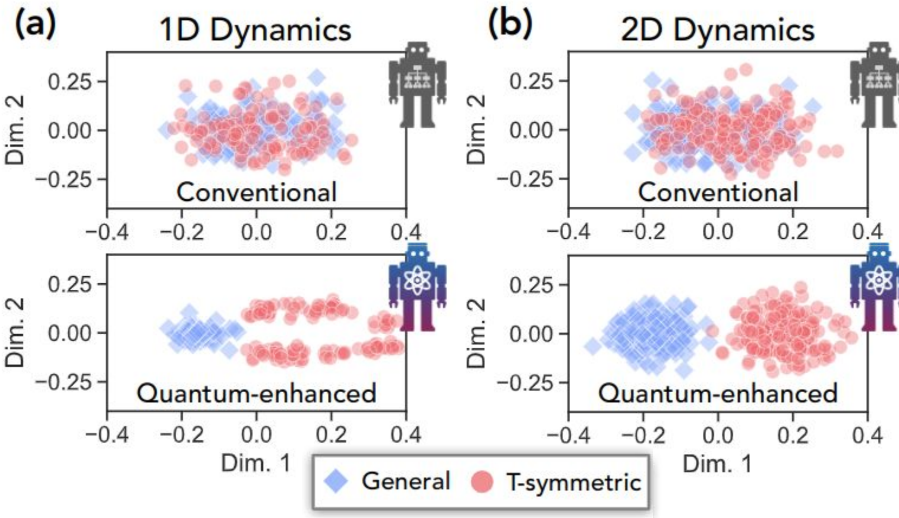
$(\mathbb{E}[b_i^1], \mathrm{Var}[b_i^1])$

...

Feature vector for each process $\vec{b}_i$

Squared exp kernel

$$K(\vec{b}_i, \vec{b}_j) = \exp(-\gamma||\vec{b}_i - \vec{b}_j||^2)$$



Kernel PCA

Quantum AI

SWAPs and virtual distillation to the quantum PCA

Corrupted quantum data → $$\rho = f\rho_0 + \sum_i p_i \rho_i$$

Uncorrupted data ($f\rho_0$) Orthogonal errors ($\sum_i p_i \rho_i$)

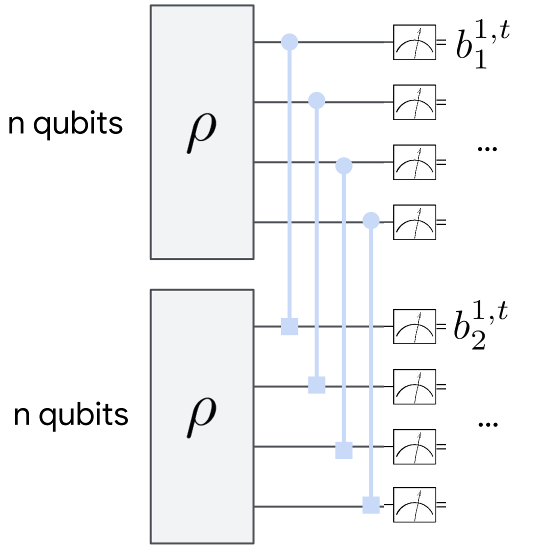
1 SWAP → Virtual distillation

SWAP as a generator → $\exp\left(-i\delta t \mathrm{SWAP}\right)$ → Quantum PCA

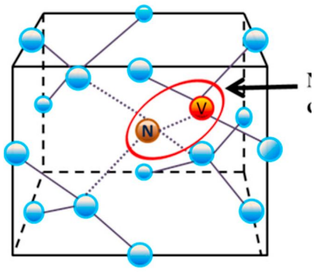
This work: Proof in a conventional scenario that exponential number of copies are required to learn about principal component vs constant in quantum enhanced setting.

Recall virtual distillation $\rho \to \rho^2$

Uses destructive SWAP

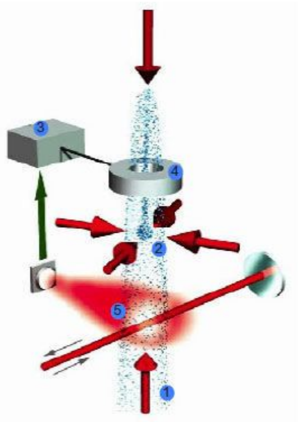


Truly multi-qubit quantum sensors? Arrays?



(a)

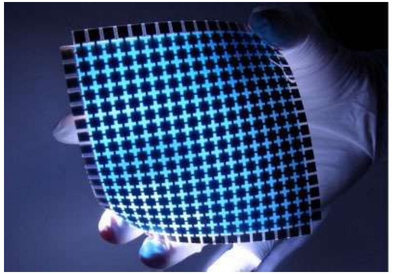
Single qubit



Ensemble of single qubits

$$|GHZ\rangle_{general} = \frac{|0\rangle^{\otimes n} + |1\rangle^{\otimes n}}{\sqrt{2}}$$

Stretched single qubits



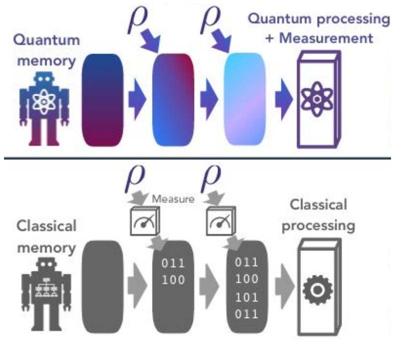
Requirements -

Multiple coherent qubits w/ independent information
Ask more interesting questions than 1 parameter
Some limited, but essential, “quantumness” in the data

Quantum AI

Quantum data

Punchline - IF we could find a suitable data source, our cloud quantum devices **today** allow us to learn things that are otherwise inaccessible.



(Recall computational vs data advantage)

This work

- Proofs of advantage in state learning, process learning, and quantum PCA
- Experimental demonstration of state and process learning using up to 40 physical qubits & 1300 gates

Outlook

- Inspire work on quantum data sources & sensors (beyond quadratic)
- Deeper connection to physics? Interferometry?
- Other tasks with 2-copy + Clifford advantage?
- Beyond Bell features?
- Can these proof techniques tell us something about existing learning tasks or quantum techniques?
- Moving away from kernel methods?

Making a gamble with classical data

Quantum data is interesting for future discovery of the universe (recall the impact of CCD cameras on telescopes - see "The Perfect Theory"), but most data we work with today, even from quantum systems, seems **classical.**

There are a few pieces of evidence that QC might help for **classical data** (sampling hard distributions, learning problems based on discrete log, linear algebra routines, ...) but a lot of pieces of evidence that it will be hard to achieve in practice

Immediate path for everyone opening Nielsen and Chuang

1. Stick N features of classical data into Log N qubits
2. Read about Holevo's bound limiting you to Log N bits of information out, get sad

$$\vec{x} \rightarrow |x\rangle = \sum_{k}^{N} x^k |k\rangle$$

Naive amplitude encoding + expected values limits you to quadratic functions on data - pretty weak models

Rotation based encoding and calling data multiple times (Data re-uploading) can get trig functions and higher degree polynomials - but can be hard to design in some cases [Schuld et al 2020]

In fault tolerance, about as easy to encode data in a way that allows non-linear functions over compact intervals for each feature by applying a unitary to the state

$$\vec{x} \rightarrow |x\rangle = \sum_{j}^{M} \sum_{k}^{N} e^{-2\pi i j x^k / M} |j\rangle |k\rangle$$

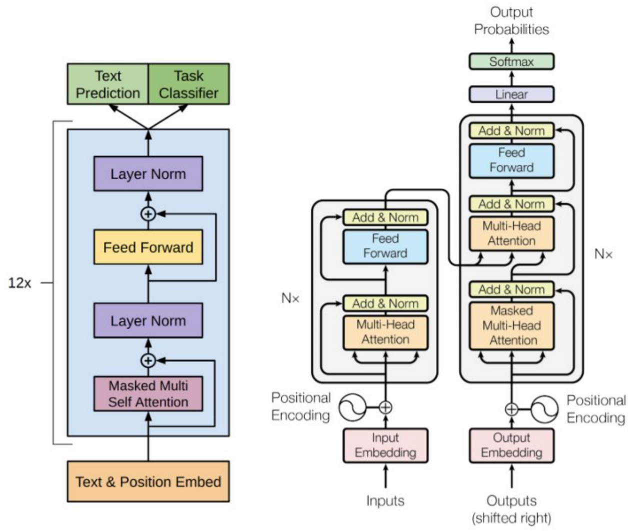
Even when amplitude encoding or feature-by-feature is enough - loading data always scales with ~N, many successful classical algorithms scale like N or < N – advantage lost?

Current premier models

Bard - 137 billion parameters ($\sim 10^{11}$)

GPT 3 - 175 billion parameters ($\sim 10^{11}$)

GPT 4 - 1.76 Trillion parameters ($\sim 10^{12}$) (Speculated)



When # of key parameters like weights or # of data points ~ 10^{12} any scalings worse than linear can be catastrophic and determines architecture / algorithm success or failure.

The same may be true in the quantum case independent of the large Hilbert space dimension

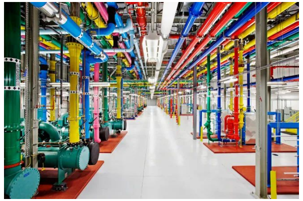
Quantum AI

Quantum pre-computation

Loading data always scales with ~N, most successful classical algorithms scale like N or < N – advantage lost?

Is it ever reasonable to discount the cost of data loading? If we do, how do our wins change?

Quantum data center

$\vec{x} \rightarrow$  (Mass production theorems) [Kretschmer 2022] $\rightarrow |x\rangle$, $\rightarrow |x\rangle$, $\rightarrow |x\rangle$ … $\longrightarrow$ User

Classically we use caches, database indexes, lookup tables, not to change asymptotic complexity, but practical time to solution.

$$\mathcal{A} : x, y, \rho \to z, \sigma.$$

$$\mathcal{P} : \bar{x}(\mathcal{A}, x), |\Gamma(\mathcal{A}, x)\rangle, y, \rho \to z, \sigma,$$

Let ρ define a program, known how to consume copies of ρ to implement $\exp(-i\rho t)$

$$R = \mathbb{I} - 2|b\rangle\langle b| = e^{-i\pi|b\rangle\langle b|}.$$

$$|x\rangle \propto A^{-1}|b\rangle$$

(can also post-select)

Question - What is the full class of uniquely quantum pre-computations we can do to accelerate time to solution?

Quantum AI

“Accelerating Quantum Algorithms with Precomputation” Huggins, McClean arXiv:2305.09638 (2023)

Quantum communication complexity

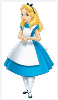
x   y

Alice has x, Bob has y, want to compute f(x,y), and the cost is only counted in terms of bits or qubits exchanged

$$\vec{x} \rightarrow |x\rangle = \sum_{k}^{N} x^k |k\rangle$$

In spite of Holevo's bound stating n qubits contain n bits of information, exponential quantum communication advantages are known. Sending log n qubits in place of n bits

Raz problem



Inputs	$\vec{x}$	$M = UZ_0U \rightarrow \{M_0, M_1\}$	V
Size	N	N x N	N x N

Quantum - Log N qubits
Classical - Ω(√N) bits

Output 0(1) if Vx is close to $M_0(M_1)$

Quantum AI

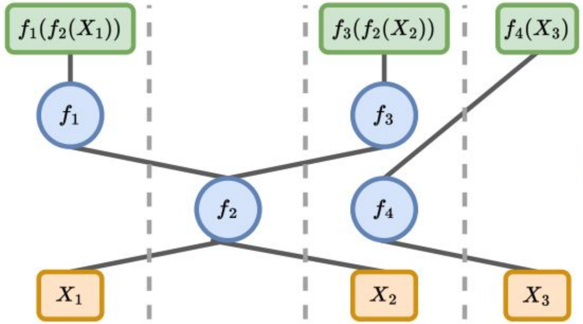
Some existing work in ML-like problems

Exponential communication advantaged shown for linear regression problem
Montanaro A, Shao C. Quantum communication complexity of linear regression arXiv:2210.01601 (2022)

Distributed quantum networks & models

Conjecture / opinion (controversial?) **-**

- The ability to make good quantum networks will be roughly coincident with really good quantum memories.
- Requirements beyond quantum memory - heralded transduction fidelity, rate, entanglement distillation - relatively modest.
- What we lack is **compelling, end-to-end costed out applications** to help motivate their development.

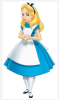


For an expressive class of functions, we find an exponential quantum communication advantage in the problem of inference and gradient determination.

Gilboa , McClean "Exponential Quantum Communication Advantage in Distributed Learning"arXiv:2310.07136 (2023)

Small communication advantage with ‘simple’ circuits

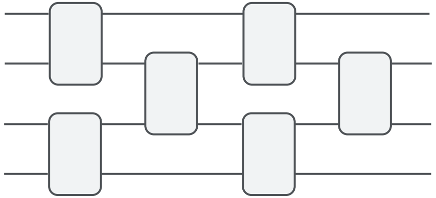
Raz problem



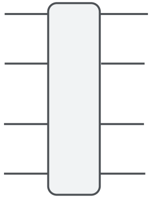
Inputs	$\vec{x}$	$M = UZ_0U \to \{M_0, M_1\}$	V
Size	N	N x N	N x N

Quantum - Log N qubits
Classical - Ω(√N) bits

Output 0(1) if Vx is close to M_0(M_1)



If M or V generated from polylog(N) ~ poly(n) and white box, it suffices to send circuit description.

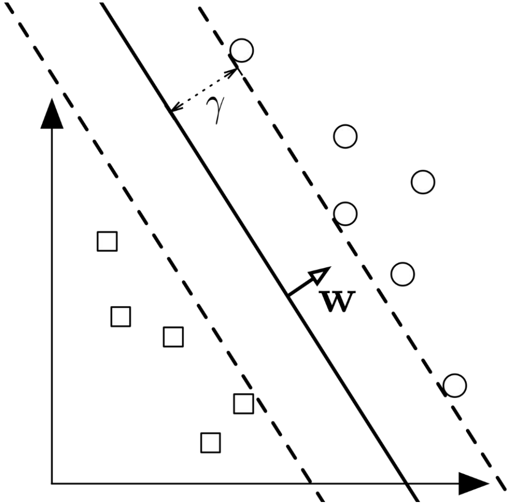


If M or V generated from polylog(N) ~ poly(n) and blackbox, it suffices to send Clifford classical shadows.

Lack of communication advantage, does not preclude computational advantage though.

- Suppose M or U was pseudo-random.
- Using black box + classical communication approach requires exponential classical computation under cryptographic assumptions

Problem structure can degrade advantage

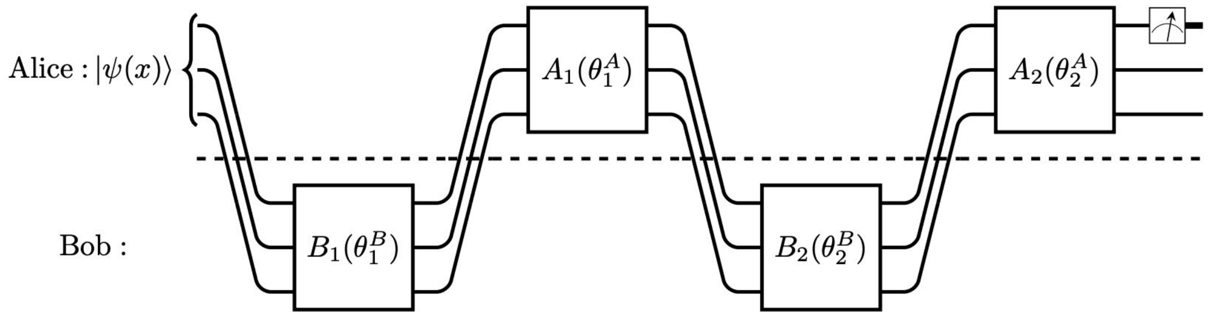


Problem 5 (Distributed Linear Classification). *Alice and Bob are given* $x, y \in S^N$*, with the promise that* $|x \cdot y| \geq \gamma$ *for some* $0 \leq \gamma \leq 1$*. Their goal is to determine the sign of* $x \cdot y$*.*

Lemma 7. *The quantum communication complexity of Problem 5 is* $\Omega\left(\sqrt{N/\max(1, \lceil \gamma N \rceil)}\right)$*. The randomized classical communication complexity of Problem 5 is* $O(\min(N, 1/\gamma^2))$*.*

If the classifier is really good (ɣ ~ 1), classical = quantum = O(1)

Exponential advantage in more general PQC models



$$|\varphi(\Theta, x)\rangle \equiv \left(\prod_{\ell=L}^{1} A_\ell(\theta_\ell^A, x) B_\ell(\theta_\ell^B, x)\right) |\psi(x)\rangle ,$$

$$\mathcal{L}(\Theta, x) \equiv \langle\varphi(\Theta, x)| \mathcal{P}_0 |\varphi(\Theta, x)\rangle ,$$

Problem 1 (Distributed Inference). *Alice and Bob each compute an estimate of $\langle\varphi| \mathcal{P}_0 |\varphi\rangle$ up to additive error ε.*

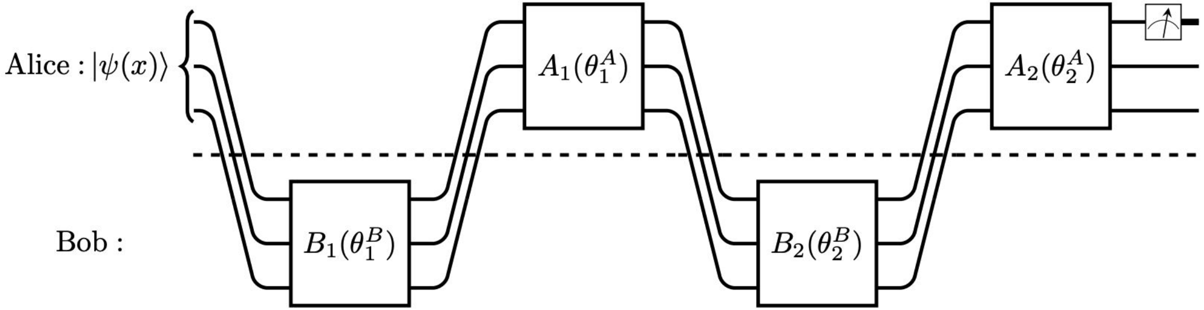
Problem 2 (Distributed Gradient Estimation). *Alice computes an estimate of $\nabla_A \langle\varphi| \mathcal{P}_0 |\varphi\rangle$, while Bob computes an estimate of $\nabla_B \langle\varphi| Z_0 |\varphi\rangle$, up to additive error ε in L^∞.*

Lemma 3. *i) The classical communication complexity of Problem 1 and Problem 2 is $\Omega(\max(\sqrt{N}, L))$.*

Lemma 1. *Problem 1 can be solved by communicating $O(\log N)$ qubits over $O(L/\varepsilon^2)$ rounds.*

Lemma 2. *Problem 2 can be solved with probability greater than $1-\delta$ by communicating $\tilde{O}(\log N (\log P)^2 \log(L/\delta)/\varepsilon^4)$ qubits over $O(L^2)$ rounds. The time and space complexity of the algorithm is $\sqrt{P}\, L\, \mathrm{poly}(N, \log P, \varepsilon^{-1}, \log(1/\delta))$.*

Model expressivity and privacy



$$|\varphi(\Theta, x)\rangle \equiv \left(\prod_{\ell=L}^{1} A_\ell(\theta_\ell^A, x) B_\ell(\theta_\ell^B, x) \right) |\psi(x)\rangle \,,$$

$$\mathcal{L}(\Theta, x) \equiv \langle \varphi(\Theta, x) | \mathcal{P}_0 | \varphi(\Theta, x) \rangle \,,$$

Lemma 9. *Let f be a p-times continuously differentiable function with period 1, and denote by $\hat{f}_{:M}$ the vector of the first M Fourier components of f. If $\left\| \hat{f}_{:M} \right\|_1 = 1$ then there exists a circuit of the form eq. (3.1) over $O(\log M)$ qubits such that*

$$\|\mathcal{L} - f\|_\infty \leq \frac{C}{M^{p-1/2}} \tag{F.4}$$

Corollary 1. *If Alice and Bob are implementing the quantum algorithm for gradient estimation described in Lemma 2, and all the communication between Alice and Bob is intercepted by an attacker, the attacker cannot extract more than $\tilde{O}(L^2 (\log N)^2 (\log P)^2 \log(L/\delta)/\varepsilon^4)$ bits of classical information about the inputs to the players.*

Expressivity - a double edged sword

Lemma 9. *Let f be a p-times continuously differentiable function with period 1, and denote by $\hat{f}_{:M}$ the vector of the first M Fourier components of f. If $\left\|\hat{f}_{:M}\right\|_1 = 1$ then there exists a circuit of the form eq. (3.1) over $O(\log M)$ qubits such that*

$$\|\mathcal{L} - f\|_\infty \leq \frac{C}{M^{p-1/2}} \tag{F.4}$$

Recall, if we want this approximation converge for each feature dimension separately, we can use the relatively easy to prepare state |x> using Log(N) Log(M) qubits

$$\vec{x} \rightarrow |x\rangle = \sum_j^M \sum_k^N e^{-2\pi i j x^k / M} |j\rangle |k\rangle$$

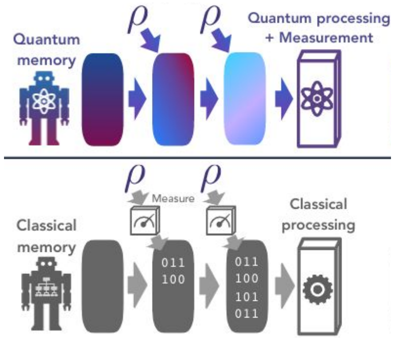
Only contains quadratic cross-feature terms, not complete across all N dimensions

Counting arguments tell you for this you need the multi-dimensional fourier state using N Log (M) qubits, which is the same complexity to just send the full state x classically → **no communication advantage**

Summary and outlook

Quantum data

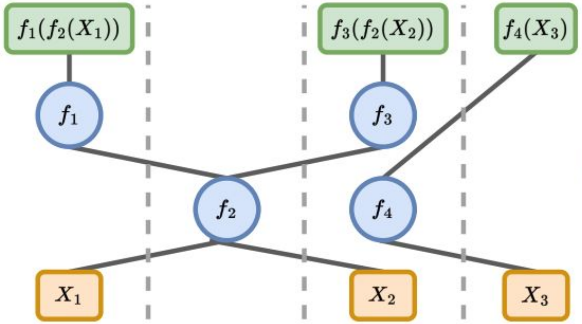
Punchline - IF we could find a suitable data source, our cloud quantum devices **today** allow us to learn things that are otherwise inaccessible.



(Recall computational vs data advantage)

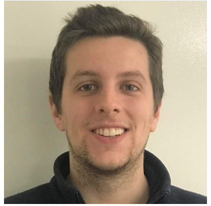
Classical data

Punchline(s) - Data changes the landscape of quantum advantage. If we can accept a future where stored quantum data states are stable and computers networked, we may find significant communication and privacy advantages in taking advantage of quantum encodings.



Acknowledgements

Hsin-Yuan (Robert) Huang



Michael Broughton

Brooks Foxen

Bill Huggins

Dar Gilboa

Exponential Quantum Communication Advantage in Distributed Learning
Gilboa , McClean
arXiv:2310.07136 (2023)

Accelerating Quantum Algorithms with Precomputation
Huggins, McClean
arXiv:2305.09638 (2023)

Quantum advantage in learning from experiments
Huang, Broughton, Cotler, Chen, Li, Mohseni, Neven, Babbush, Kueng, Preskill, McClean
Science 376, 6598 (2022)