Characterizing Quantum Advantage in Machine Learning

Hsin-Yuan (Robert) Huang, Richard Kueng, Michael Broughton, Masoud Mohseni, Ryan Babbush, Sergio Boixo, Hartmut Neven, Jarrod McClean and John Preskill

Google AI Quantum
Institute of Quantum Information and Matter (IQIM), Caltech
Johannes Kepler University Linz

[1] **Information-theoretic bounds on quantum advantage in machine learning**, *arXiv:2101.02464*.
[2] **Power of data in quantum machine learning**, *arXiv:2011.01938*.

JOHANNES KEPLER UNIVERSITY LINZ

Google AI Quantum

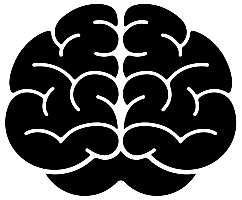
INSTITUTE FOR QUANTUM INFORMATION AND MATTER

Motivation

- Machine learning (ML) has received great attention in the quantum community these days.

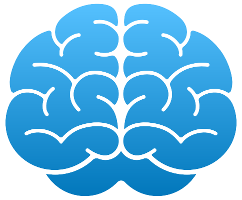
Classical ML
for quantum physics/chemistry

The goal ◎:
Solve challenging quantum many-body problems **better** than traditional classical algorithms



Enhancing ML
with quantum computers

The goal ◎:
Design quantum ML algorithms that yield significant **advantage** over any classical algorithm



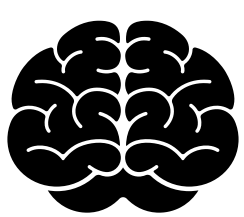
“Solving the quantum many-body problem with artificial neural networks.” *Science* 355.6325 (2017): 602-606.
"Learning phase transitions by confusion." *Nature Physics* 13.5 (2017): 435-439.
"Supervised learning with quantum-enhanced feature spaces." *Nature* 567.7747 (2019): 209-212.

Motivation

- Yet, many fundamental questions remain to be answered.

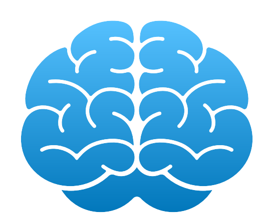
Classical ML
for quantum physics/chemistry

The question:
How can ML be more useful
than non-ML algorithms?



Enhancing ML
with quantum computers

The question:
What are the advantages of
quantum ML in general?



"Solving the quantum many-body problem with artificial neural networks." *Science* 355.6325 (2017): 602-606.
"Learning phase transitions by confusion." *Nature Physics* 13.5 (2017): 435-439.
"Supervised learning with quantum-enhanced feature spaces." *Nature* 567.7747 (2019): 209-212.

General Setting

- In this work, we focus on training an ML model to predict

$$x \mapsto f_{\mathscr{E}}(x) = \mathrm{Tr}(O\mathscr{E}(|x\rangle\langle x|)),$$

  where x is a classical input, $\mathscr{E}$ is an **unknown** CPTP map, and O is an observable.

- This is **very general**: includes any function computable by a quantum computer.

Example 1

Predicting outcomes of physical experiments

x : parameters describing the experiment

$\mathscr{E}$: the physical process in the experiment

O : what the scientist measure

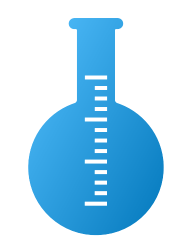
Example 2

Predicting ground state properties of a physical system

x : parameters describing a physical system

$\mathscr{E}$: a process for preparing ground state

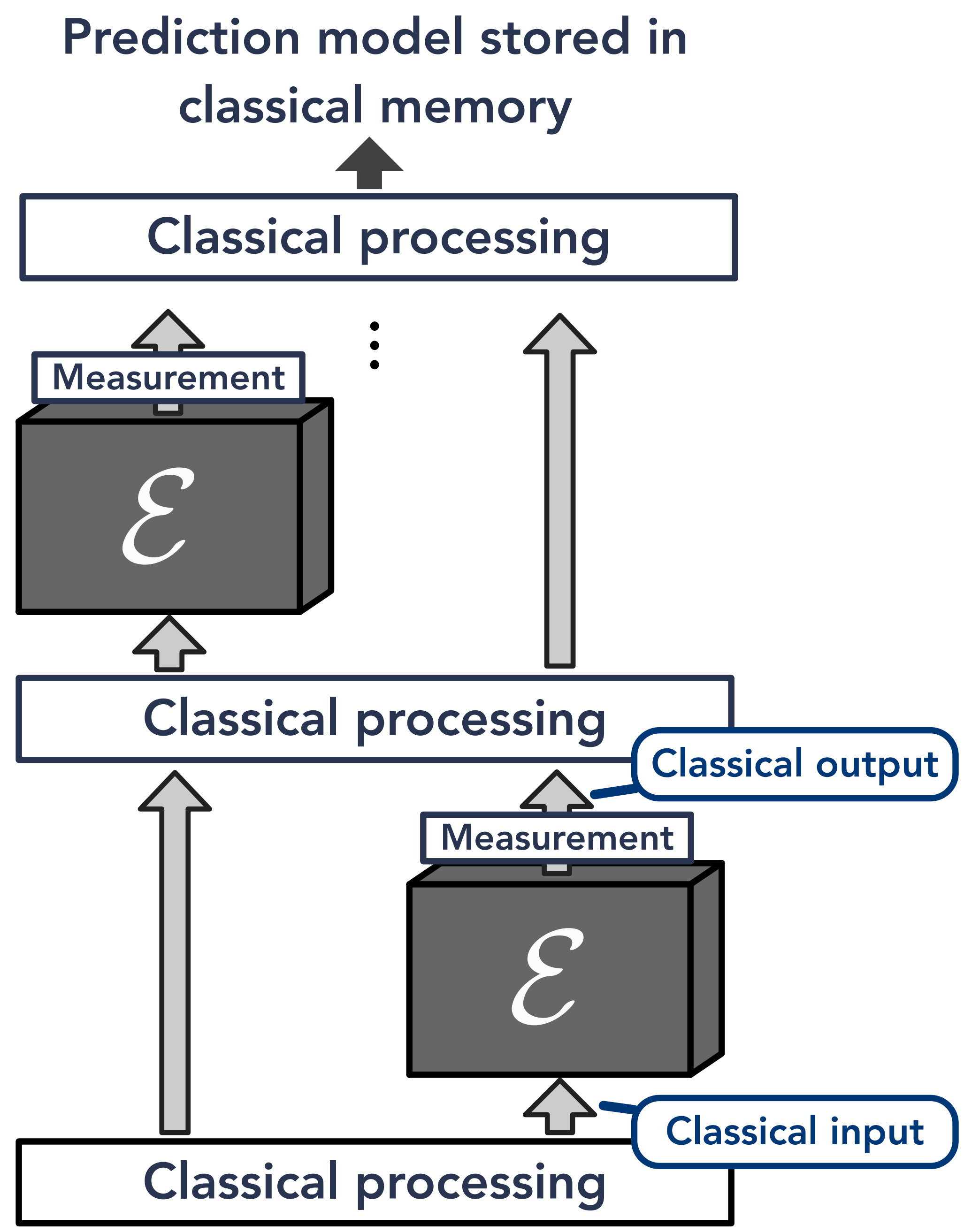
O : the property we want to predict



General Setting

Classical setting

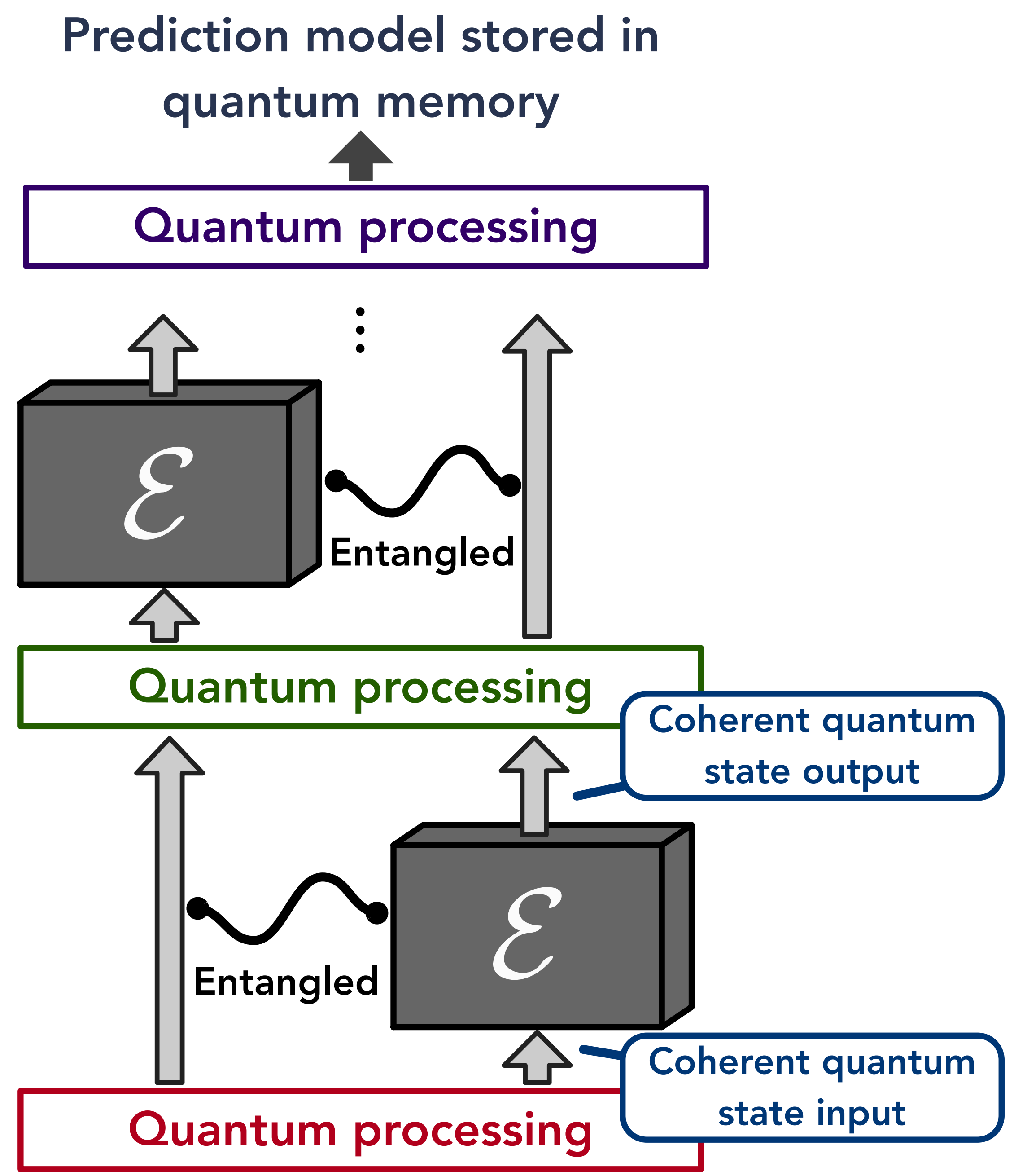
- Classical data from each experiment.
- Each query begins with a choice of classical input x and ends with an arbitrary POVM measurement.
- A prediction model $h(x) \approx f_{\mathcal{E}}(x)$ is created after learning.



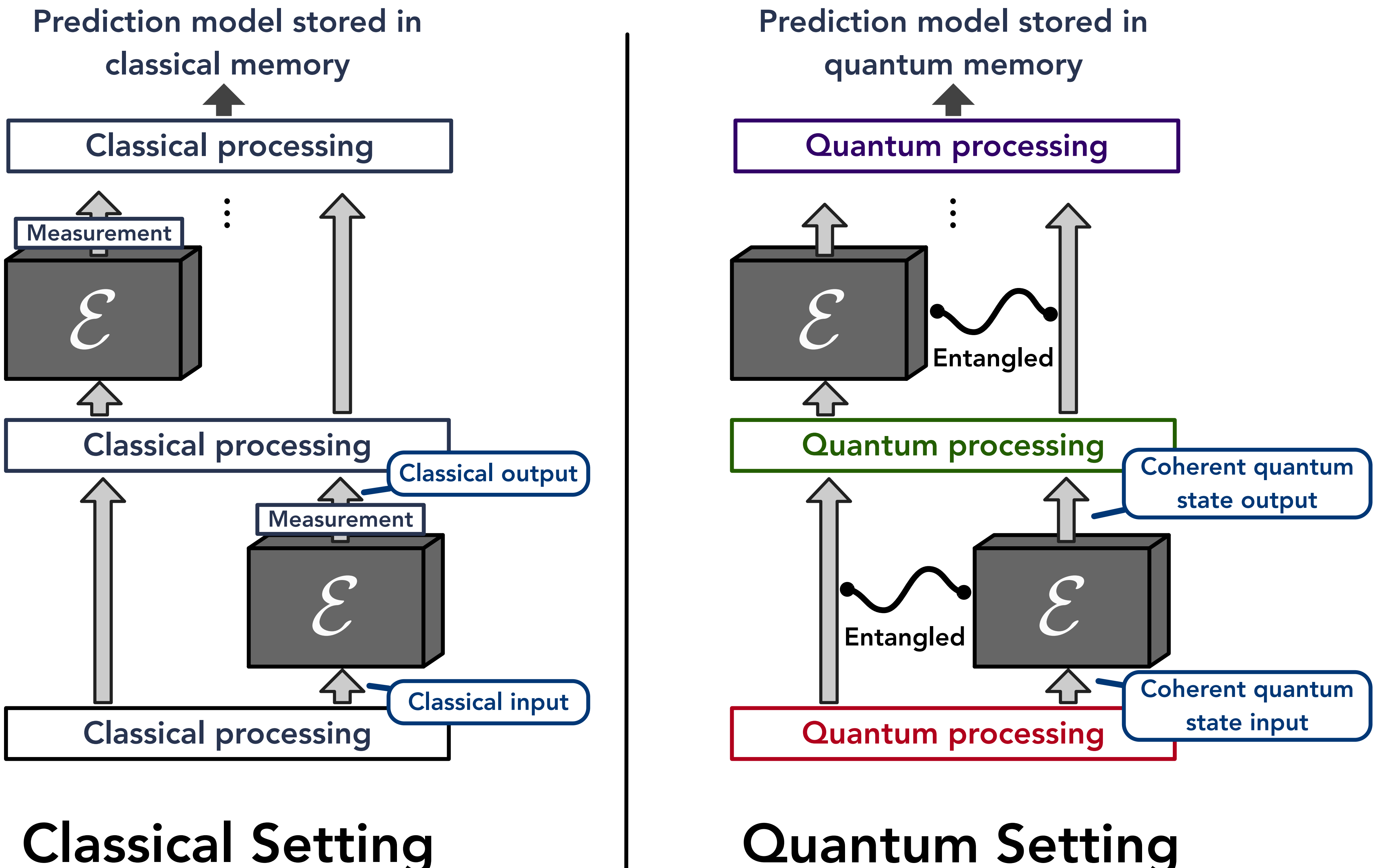
General Setting

Quantum setting

- Quantum data from each experiment.
- Each query consists of a quantum access to the CPTP map $\mathscr{E}$ (quantum input + quantum output).
- A prediction model $h(x) \approx f_{\mathscr{E}}(x)$ is created after learning.



General Setting



The setup is closely related to Quantum Algorithmic Measurements by Aharonov, Cotler, Qi

Main Questions

Information-theoretic aspect:

Do we need significantly more experiments in the classical setting compared to the quantum setting to learn $f_{\mathscr{E}}(x)$?

[1] **Information-theoretic bounds on quantum advantage in machine learning**, *arXiv:2101.02464*.

Computational aspect:

Could classical ML use data to efficiently compute $f_{\mathscr{E}}(x) = \mathrm{Tr}(O\mathscr{E}(|x\rangle\langle x|))$ even if $f_{\mathscr{E}}(x)$ is hard to compute with a classical computer?

[2] **Power of data in quantum machine learning**, *arXiv:2011.01938*.

Information-theoretic aspect

Theorem (Huang, Kueng, Preskill; 2021 [1])

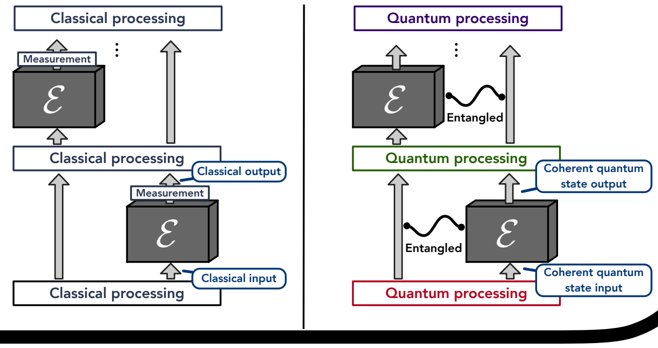
Consider any observable O, any family of CPTP maps $\mathscr{F} = \{\mathscr{E}\}$ with n-qubit input and m-qubit output, and any input distribution $\mathscr{D}$.

Suppose a quantum ML uses N_{Q} queries to the unknown CPTP map $\mathscr{E}$ to learn a prediction model $h_{\mathrm{Q}}(x)$ that achieves a prediction error of

$$\mathbb{E}_{x\sim\mathscr{D}} \left| h_{\mathrm{Q}}(x) - f_{\mathscr{E}}(x) \right|^2 \leq \epsilon,$$

then there is a classical ML using $N_{\mathrm{C}} \leq \mathscr{O}(mN_{\mathrm{Q}}/\epsilon)$ to learn a prediction model $h_{\mathrm{C}}(x)$ that achieves a prediction error of

$$\mathbb{E}_{x\sim\mathscr{D}} \left| h_{\mathrm{C}}(x) - f_{\mathscr{E}}(x) \right|^2 \leq \mathscr{O}(\epsilon)\,.$$



[1] **Information-theoretic bounds on quantum advantage in machine learning**, *arXiv:2101.02464*.

Information-theoretic aspect

Theorem (Huang, Kueng, Preskill; 2021 [1])

Concept/hypothesis class in statistical learning theory

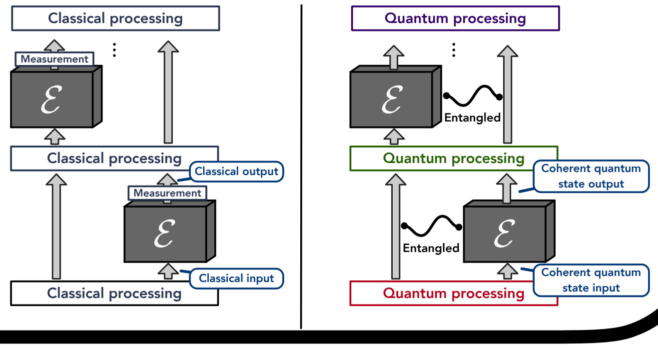
Consider any observable O, any family of CPTP maps $\mathscr{F} = \{\mathscr{E}\}$ with n-qubit input and m-qubit output, and any input distribution $\mathscr{D}$.

Suppose a quantum ML uses N_Q queries to the unknown CPTP map $\mathscr{E}$ to learn a prediction model $h_\mathrm{Q}(x)$ that achieves a prediction error of

$$\mathbb{E}_{x\sim\mathscr{D}} \left| h_\mathrm{Q}(x) - f_\mathscr{E}(x) \right|^2 \leq \epsilon,$$

then there is a classical ML using $N_\mathrm{C} \leq \mathcal{O}(mN_\mathrm{Q}/\epsilon)$ to learn a prediction model $h_\mathrm{C}(x)$ that achieves a prediction error of

$$\mathbb{E}_{x\sim\mathscr{D}} \left| h_\mathrm{C}(x) - f_\mathscr{E}(x) \right|^2 \leq \mathcal{O}(\epsilon)\,.$$



[1] **Information-theoretic bounds on quantum advantage in machine learning**, *arXiv:2101.02464.*

Information-theoretic aspect

Theorem (Huang, Kueng, Preskill; 2021 [1])

Consider any observable O, any family of CPTP maps $\mathscr{F} = \{\mathscr{E}\}$ with n-qubit input and m-qubit output, and any input distribution $\mathscr{D}$.

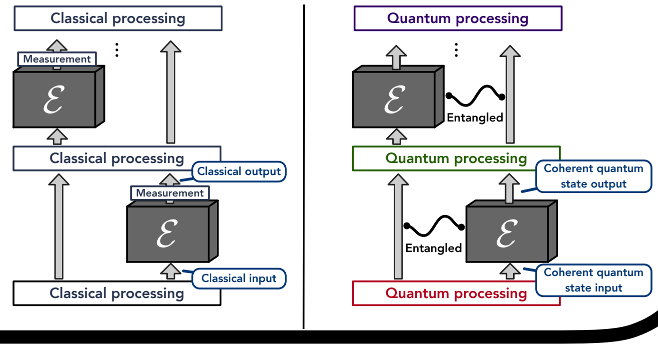
Suppose a quantum ML uses N_{Q} queries to the unknown CPTP map $\mathscr{E}$ to learn a prediction model $h_{\mathrm{Q}}(x)$ that achieves a predic[tion error of]

Average prediction error

$$\mathbb{E}_{x\sim\mathscr{D}} \left| h_{\mathrm{Q}}(x) - f_{\mathscr{E}}(x) \right|^2 \leq \epsilon,$$

then there is a classical ML using $N_{\mathrm{C}} \leq \mathcal{O}(mN_{\mathrm{Q}}/\epsilon)$ to learn a prediction model $h_{\mathrm{C}}(x)$ that achieves a prediction error of

$$\mathbb{E}_{x\sim\mathscr{D}} \left| h_{\mathrm{C}}(x) - f_{\mathscr{E}}(x) \right|^2 \leq \mathcal{O}(\epsilon)\,.$$



[1] **Information-theoretic bounds on quantum advantage in machine learning**, *arXiv:2101.02464*.

Implication of $N_\mathrm{C} \leq \mathcal{O}(m N_\mathrm{Q}/\epsilon)$

- Quantum ML setting may likely only be available far in the future. (need quantum memory to store data)
- Classical ML setting is readily available. (only need classical memory to store data)
- Learning from classical data can be as powerful as learning from coherent quantum data.

[1] **Information-theoretic bounds on quantum advantage in machine learning**, *arXiv:2101.02464.*

Non-Implication of $N_{\mathrm{C}} \leq \mathcal{O}(m N_{\mathrm{Q}}/\epsilon)$

- ML models trained on classical computers are computationally as powerful as those running on quantum computers?
- **No!** We only consider data efficiency, not computational complexity.
- We can consider quantum algorithms for the classical setting (learning only from classical data stored in classical memory).
- Quantum computers can potentially optimize/compute faster.

[1] **Information-theoretic bounds on quantum advantage in machine learning**, *arXiv:2101.02464.*

[2] **Power of data in quantum machine learning**, *arXiv:2011.01938.*

Implication of $N_{\mathrm{C}} \leq \mathcal{O}(m N_{\mathrm{Q}}/\epsilon)$

- Learning from classical data can be as powerful as learning from coherent quantum data.
- ML models running quantum computers can train/predict faster than classical computers.
- Boosts our hope for using near-term quantum devices + classical computers to address challenging quantum problems in physics/chemistry (more to come in my next paper).

[1] **Information-theoretic bounds on quantum advantage in machine learning**, *arXiv:2101.02464.*

Main Questions

Information-theoretic aspect:

Do we need significantly more experiments in the classical setting compared to the quantum setting to learn $f_{\mathscr{E}}(x)$?

[1] **Information-theoretic bounds on quantum advantage in machine learning**, *arXiv:2101.02464*.

Computational aspect:

Could classical ML use data to efficiently compute $f_{\mathscr{E}}(x) = \mathrm{Tr}(O\mathscr{E}(|x\rangle\langle x|))$ even if $f_{\mathscr{E}}(x)$ is hard to compute with a classical computer?

[2] **Power of data in quantum machine learning**, *arXiv:2011.01938*.

Main Questions

Information-theoretic aspect:

Do we need significantly more experiments in the classical setting compared to the quantum setting to learn $f_{\mathscr{E}}(x)$? **No!**

[1] Information-theoretic bounds on quantum advantage in machine learning, *arXiv:2101.02464*.

Computational aspect:

Could classical ML use data to efficiently compute $f_{\mathscr{E}}(x) = \mathrm{Tr}(O\mathscr{E}(|x\rangle\langle x|))$ even if $f_{\mathscr{E}}(x)$ is hard to compute with a classical computer?

[2] Power of data in quantum machine learning, *arXiv:2011.01938*.

Exponential advantage

- The theorem holds only for average-case prediction error.
- Other measures of prediction error (e.g., worst-case) admits provable exponential advantage.

$$\max_x \left| h(x) - f_{\mathscr{E}}(x) \right|^2 \text{ instead of } \mathbb{E}_{x \sim \mathscr{D}} \left| h(x) - f_{\mathscr{E}}(x) \right|^2$$

[1] **Information-theoretic bounds on quantum advantage in machine learning**, *arXiv:2101.02464*.

Main Questions

Information-theoretic aspect:

Do we need significantly more experiments in the classical setting compared to the quantum setting to learn $f_{\mathscr{E}}(x)$? **No!**

[1] **Information-theoretic bounds on quantum advantage in machine learning**, *arXiv:2101.02464*.

Computational aspect:

Could classical ML use data to efficiently compute $f_{\mathscr{E}}(x) = \mathrm{Tr}(O\mathscr{E}(|x\rangle\langle x|))$ even if $f_{\mathscr{E}}(x)$ is hard to compute with a classical computer?

[2] **Power of data in quantum machine learning**, *arXiv:2011.01938*.

Main Questions

Exponential separation
for worst-case error.

Information-theoretic aspect:

Do we need significantly more experiments in the classical setting compared to the quantum setting to learn $f_{\mathscr{E}}(x)$? **No!**

[1] **Information-theoretic bounds on quantum advantage in machine learning**, *arXiv:2101.02464*.

Computational aspect:

Could classical ML use data to efficiently compute $f_{\mathscr{E}}(x) = \mathrm{Tr}(O\mathscr{E}(|x\rangle\langle x|))$ even if $f_{\mathscr{E}}(x)$ is hard to compute with a classical computer?

[2] **Power of data in quantum machine learning**, *arXiv:2011.01938*.

Main Questions

> Exponential separation for worst-case error.

Information-theoretic aspect:

Do we need significantly more experiments in the classical setting compared to the quantum setting to learn $f_{\mathscr{E}}(x)$? **No!**

[1] **Information-theoretic bounds on quantum advantage in machine learning**, *arXiv:2101.02464*.

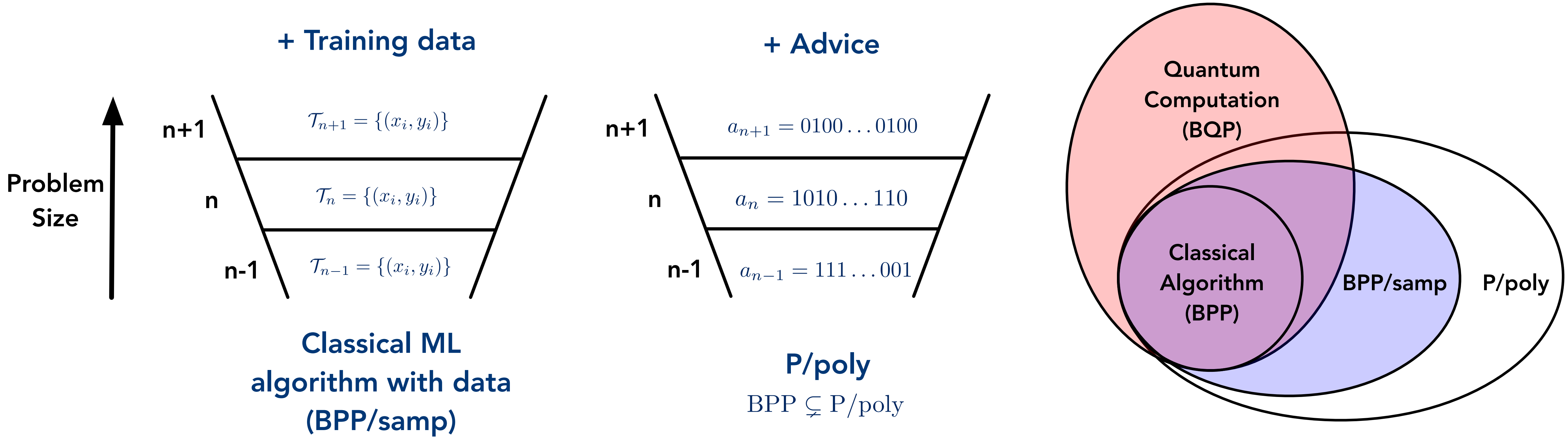
Computational aspect:

Could classical ML use data to efficiently compute $f_{\mathscr{E}}(x) = \mathrm{Tr}(O\mathscr{E}(|x\rangle\langle x|))$ even if $f_{\mathscr{E}}(x)$ is hard to compute with a classical computer?

[2] **Power of data in quantum machine learning**, *arXiv:2011.01938*.

Computational aspect

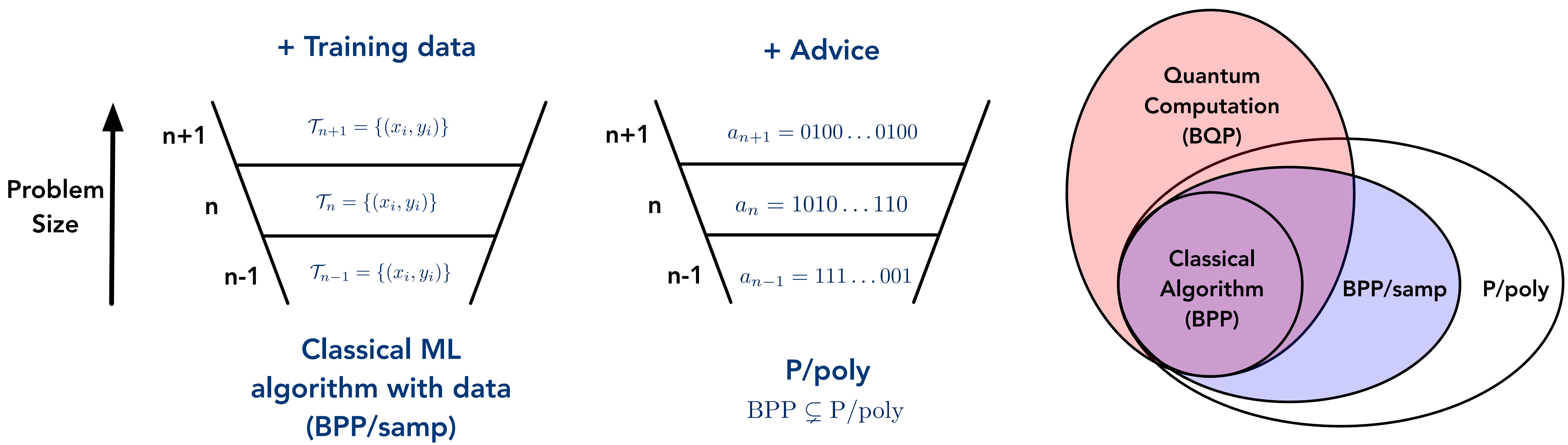
- The formal difference between classical ML and non-ML algorithm is that ML algorithm can learn from data.
- We define a complexity class for classical algorithm that could learn from sampled data (BPP/samp).
- BPP/samp is a restricted class of P/poly.



[2] **Power of data in quantum machine learning**, *arXiv:2011.01938.*

Computational aspect

- Classical algorithms learning from data could solve problems that can not be solved by non-ML algorithms.
- This is only true when data can not be computed in BPP. (such as data from quantum experiments)



[2] **Power of data in quantum machine learning**, *arXiv:2011.01938.*

Computational power of data

- For example, $|\psi_{\text{init}}\rangle$ = single-particle n-site Fermionic state, U = general interacting Hamiltonian evolution.
- Because U is a general $2^n \times 2^n$ unitary transformation, predicting property of $U|\psi_{\text{init}}\rangle$ is hard classically.

$$n = 3$$

$$|\psi_{\text{init}}\rangle = \alpha|100\rangle + \beta|010\rangle + \gamma|001\rangle$$

$$U|\psi_{\text{init}}\rangle = \alpha U|100\rangle + \beta U|010\rangle + \gamma U|001\rangle = \sum_{i=1}^{2^n} c_i|i\rangle$$

$$\langle\psi_{\text{init}}|U^\dagger O U|\psi_{\text{init}}\rangle = \sum_i \sum_j \overline{c_i} c_j \langle i|O|j\rangle$$

[1] **Power of data in quantum machine learning**, *arXiv:2011.01938.*

Computational power of data

- For example, $|\psi_{\rm init}\rangle$ = single-particle n-site Fermionic state, U = general interacting Hamiltonian evolution.
- However, given ~n^2 training data, predicting property of $U|\psi_{\rm init}\rangle$ can be done easily on a classical computer (equiv. to learning quadratic func.).

$$n = 3$$

$$|\psi_{\rm init}\rangle = \alpha|100\rangle + \beta|010\rangle + \gamma|001\rangle$$

$$U|\psi_{\rm init}\rangle = \alpha U|100\rangle + \beta U|010\rangle + \gamma U|001\rangle = \sum_{i=1}^{2^n} c_i|i\rangle$$

$$\langle\psi_{\rm init}|U^\dagger O U|\psi_{\rm init}\rangle = \sum_i \sum_j \overline{c_i} c_j \langle i|O|j\rangle = (\overline{\alpha}, \overline{\beta}, \overline{\gamma}) A (\alpha, \beta, \gamma)^T$$

[1] Power of data in quantum machine learning, *arXiv:2011.01938.*

Prediction error after training a kernel ML model

$$\mathbb{E}_x |g_K(x) - f_{\mathscr{E}}(x)| \leq \mathcal{O}\left(\sqrt{\frac{s_K}{N}}\right)$$

N: training data size

- $g_K(x)$ is the function learned by a kernel ML model.
- If s_K is **small**, then the kernel model can **accurately predict** $f(x) = \mathrm{Tr}(U^\dagger O U \rho(x))$. (irrespective of whether $f(x)$ is hard to compute without training data)
- **Quantum advantage** happens when s_C is large and s_Q is small.

$$s_K = \sum_{ij} (K^{-1})_{ij} f_{\mathscr{E}}(x_i) f_{\mathscr{E}}(x_j) \geq 0, \text{ where } K_{ij} = k(x_i, x_j).$$

[1] **Power of data in quantum machine learning**, *arXiv:2011.01938.*

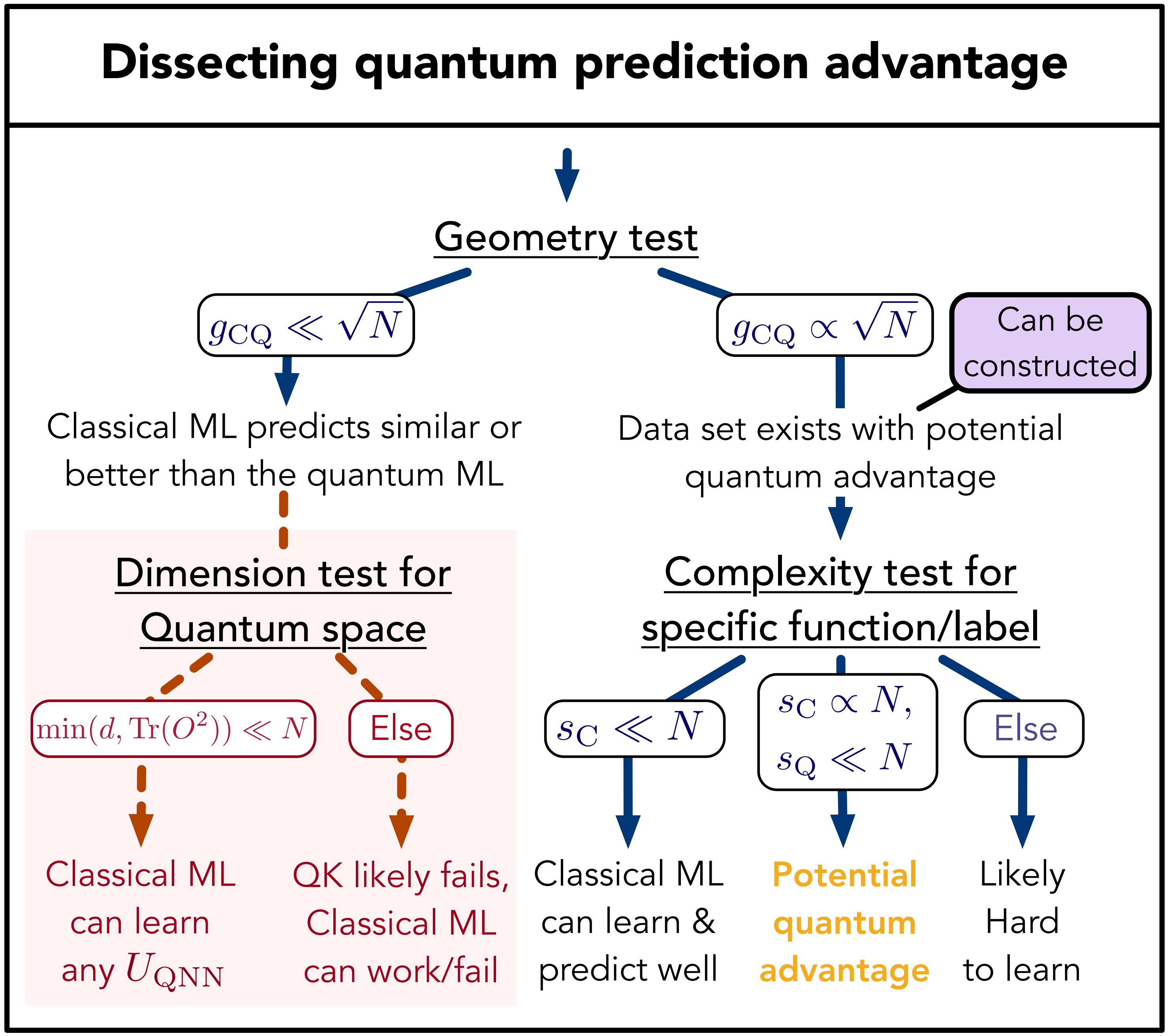
Geometric difference

- $s_C \leq g(K_C||K_Q)^2 s_Q$ where $g(K_C||K_Q) = \sqrt{\|\sqrt{K_Q}K_C^{-1}\sqrt{K_Q}\|_\infty} \geq 1$.
- If $g(K_C||K_Q)$ is small, **no function** f exists where the quantum ML outperforms classical ML.
- If $g(K_C||K_Q)$ is large, **a function** f exists where the quantum ML outperforms classical ML.

$g(K_C||K_Q)$ measures the difference between how quantum ML vs classical ML sees the relation between data.

[1] **Power of data in quantum machine learning**, *arXiv:2011.01938.*

A flowchart for understanding quantum advantage



[2] **Power of data in quantum machine learning**, *arXiv:2011.01938.*

Limitation of Quantum kernel methods

- When the quantum states $\rho(x^i)$ for the training set span a large dimension quantum Hilbert space, all inputs are too far apart, so

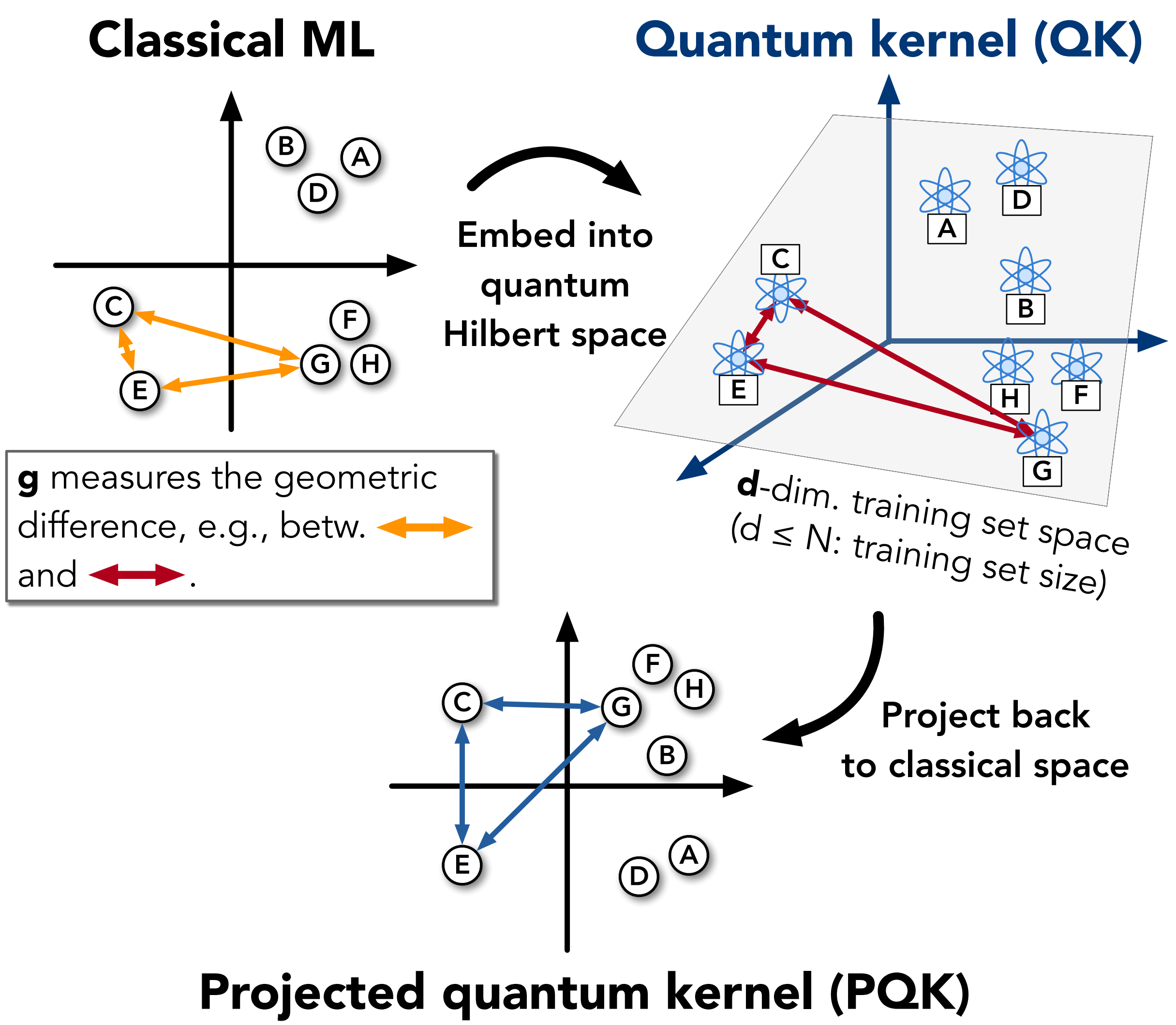
$$K^{\mathrm{Q}} \approx I \quad \text{and} \quad g_{\mathrm{CQ}} = \sqrt{\|\sqrt{K_{\mathrm{Q}}} K_{\mathrm{C}}^{-1} \sqrt{K_{\mathrm{Q}}}\|_\infty} \approx 1.$$

- This means classical ML can often compete or outperform quantum kernel methods in learning any quantum models.
- One could rigorously show that for simple quantum models, quantum kernel need exponential number of data, while classical ML only need linear.
- We see classical ML outperforming quantum kernel throughout numerics.

Prediction error bound for QK: $\mathbb{E}_x |g(x) - \mathrm{Tr}(O^U \rho(x))| \leq \mathcal{O}\left(\sqrt{\frac{\min(d, \mathrm{Tr}(O^2))}{N}} + \sqrt{\frac{\log(1/\delta)}{N}}\right)$

One solution

- Large quantum Hilbert space dimension makes quantum ML suffers more than classical ML.
- Projects quantum states back to classical space, e.g. using reduced observable or classical shadow [1].
- Define kernel in the classical space.
- We call this the projected quantum kernel (PQK).



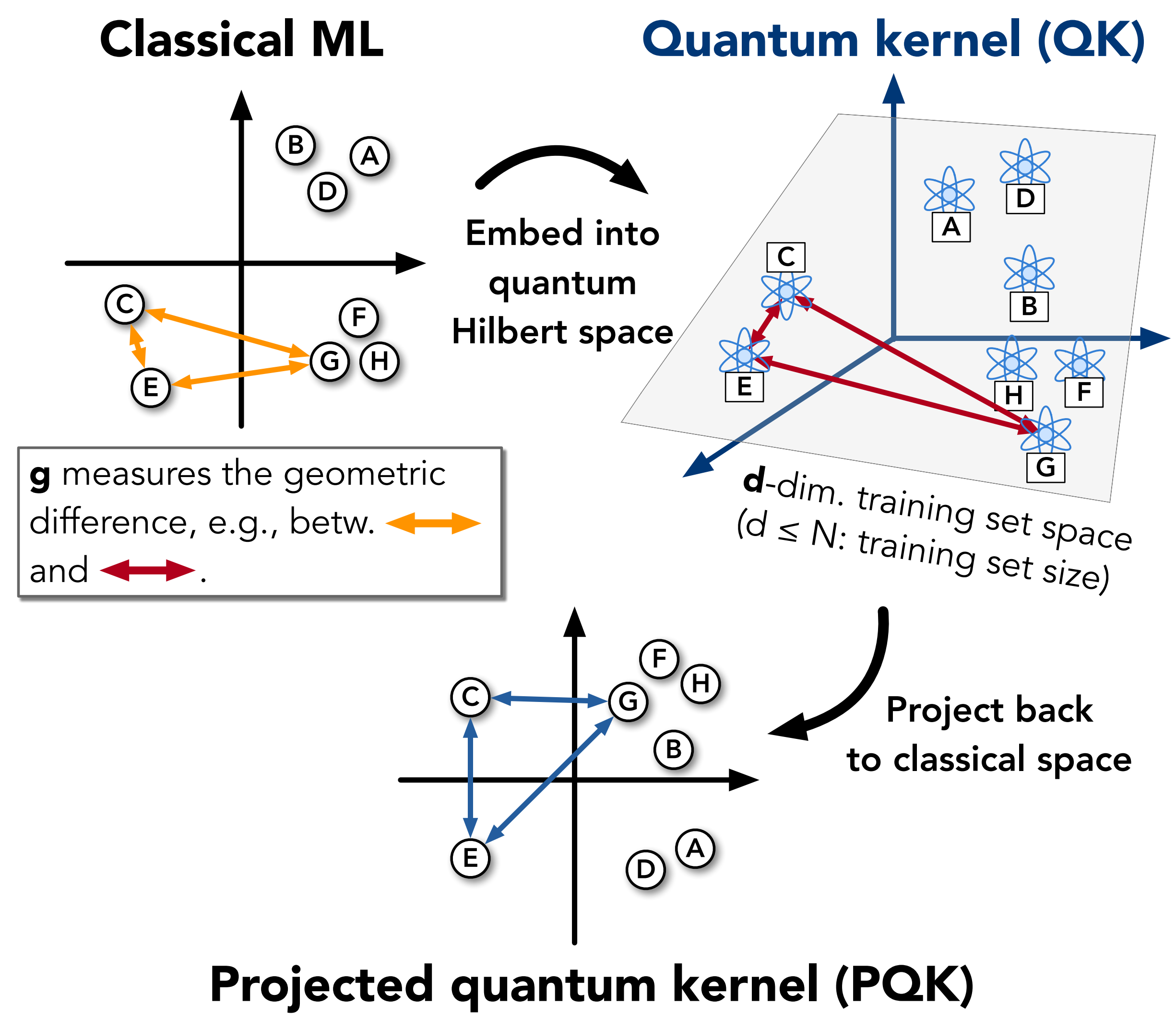
[1] Predicting many properties of a quantum system from very few measurements

Projected quantum kernel

- PQK requires quantum computer to compute (by going through QK).
- PQK results in much higher geometric difference. (because QK has $g \approx 1$)
- Simple-to-prove rigorous advantage in a learning problem based on discrete logarithm [1].

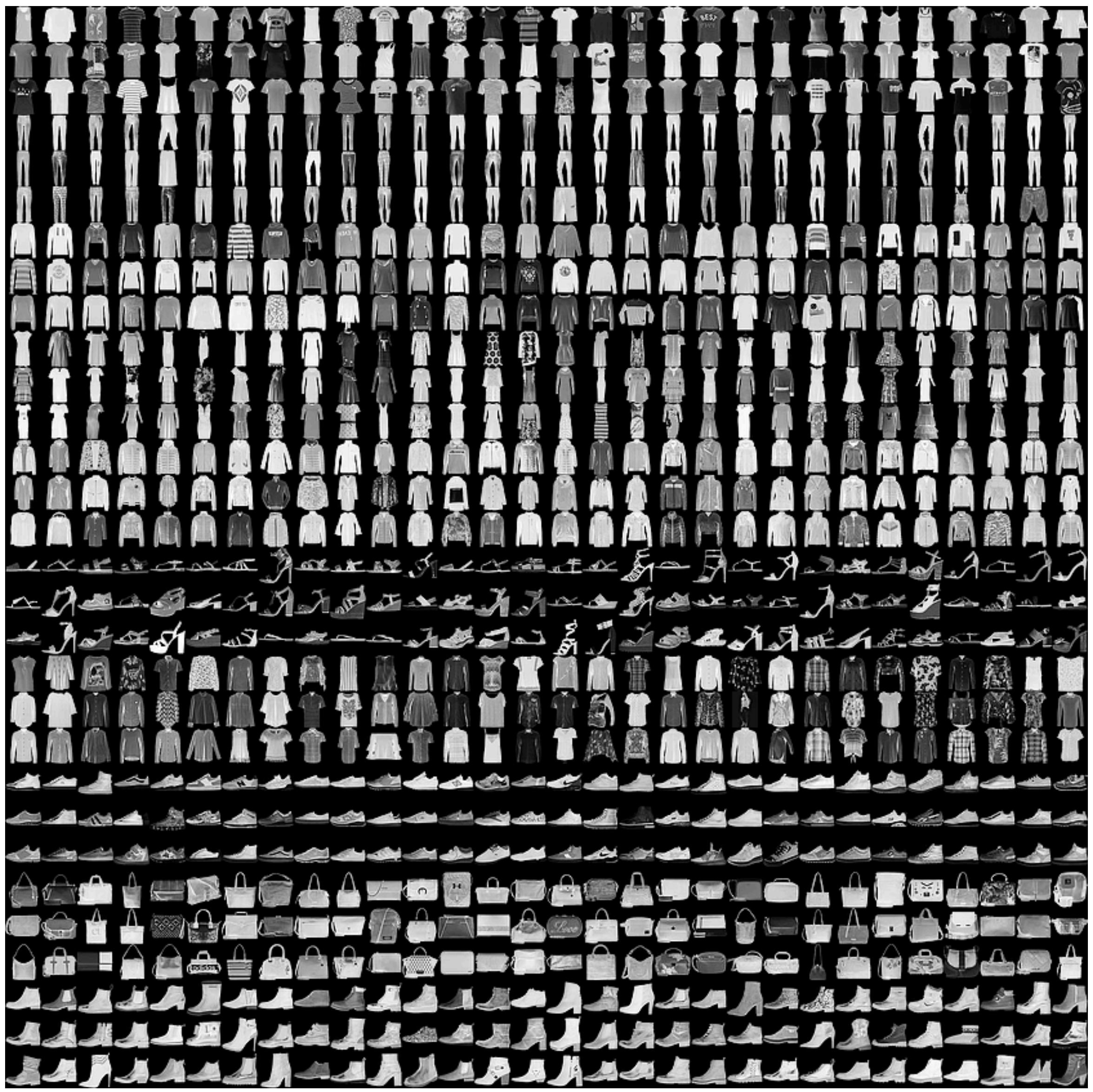
$$y(x) = \begin{cases} +1, & \log_g(x) \in [s, s + \frac{p-3}{2}], \\ -1, & \log_g(x) \notin [s, s + \frac{p-3}{2}], \end{cases}$$

- The proof that QK can learn the above problem is much more complicated [1].



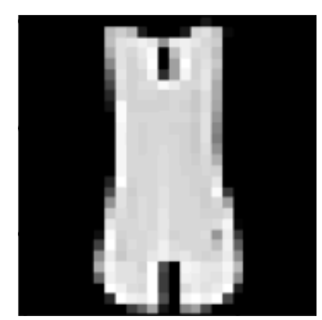
[1] A rigorous and robust quantum speed-up in supervised machine learning

Experiments



Fashion-MNIST

- MNIST is too easy (can predict well with one pixel) and overused.
- Fashion-MNIST is a harder alternative with the same format.
- We focus on binary classification (dresses versus shirts)



How well it works in practice

Data source: Fashion-MNIST → PCA → n components → length n vector → x_i

E1

$$|x_i\rangle = \bigotimes_{j=1}^{n} e^{-iX_j x_{ij}} |0^n\rangle$$

E2

$$|x_i\rangle = U_Z(x_i) H^{\otimes n} U_Z(x_i) H^{\otimes n} |0^n\rangle$$

$$U_Z(x_i) = \exp\left(\sum_{j=1}^{n} x_{ij} Z_j + \sum_{j=1}^{n} \sum_{j'=1}^{n} x_{ij} x_{ij'} Z_j Z_{j'}\right)$$

E3

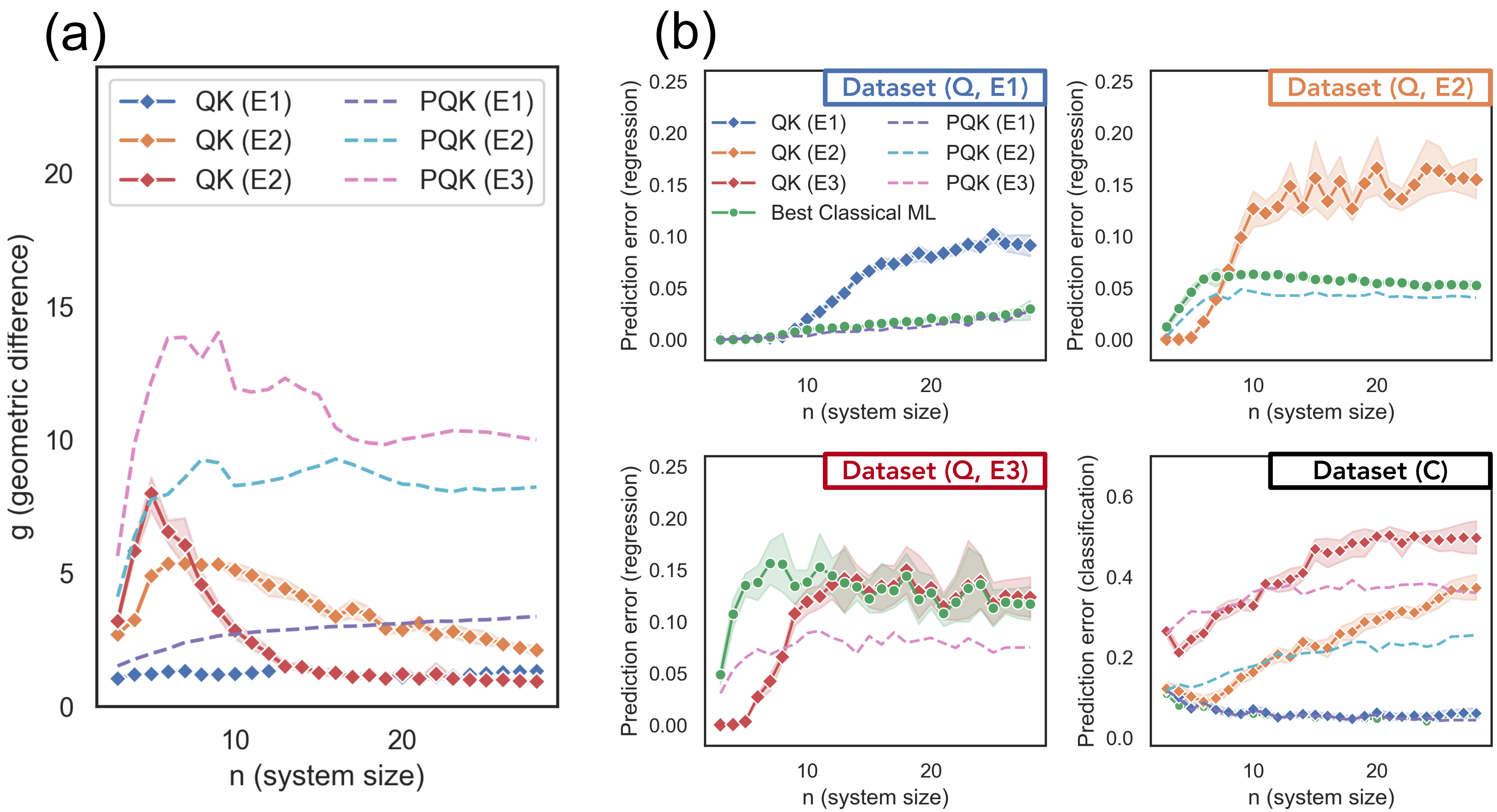
$$|x_i\rangle = \left(\prod_{j=1}^{n} \exp\left(-i\frac{t}{T} x_{ij} \left(X_j X_{j+1} + Y_j Y_{j+1} + Z_j Z_{j+1}\right)\right)\right)^T \bigotimes_{j=1}^{n+1} |\psi_j\rangle$$

$$T = 20 \qquad t = \frac{n}{3}$$

Label: (C) - Original Fashion-MNIST labels,

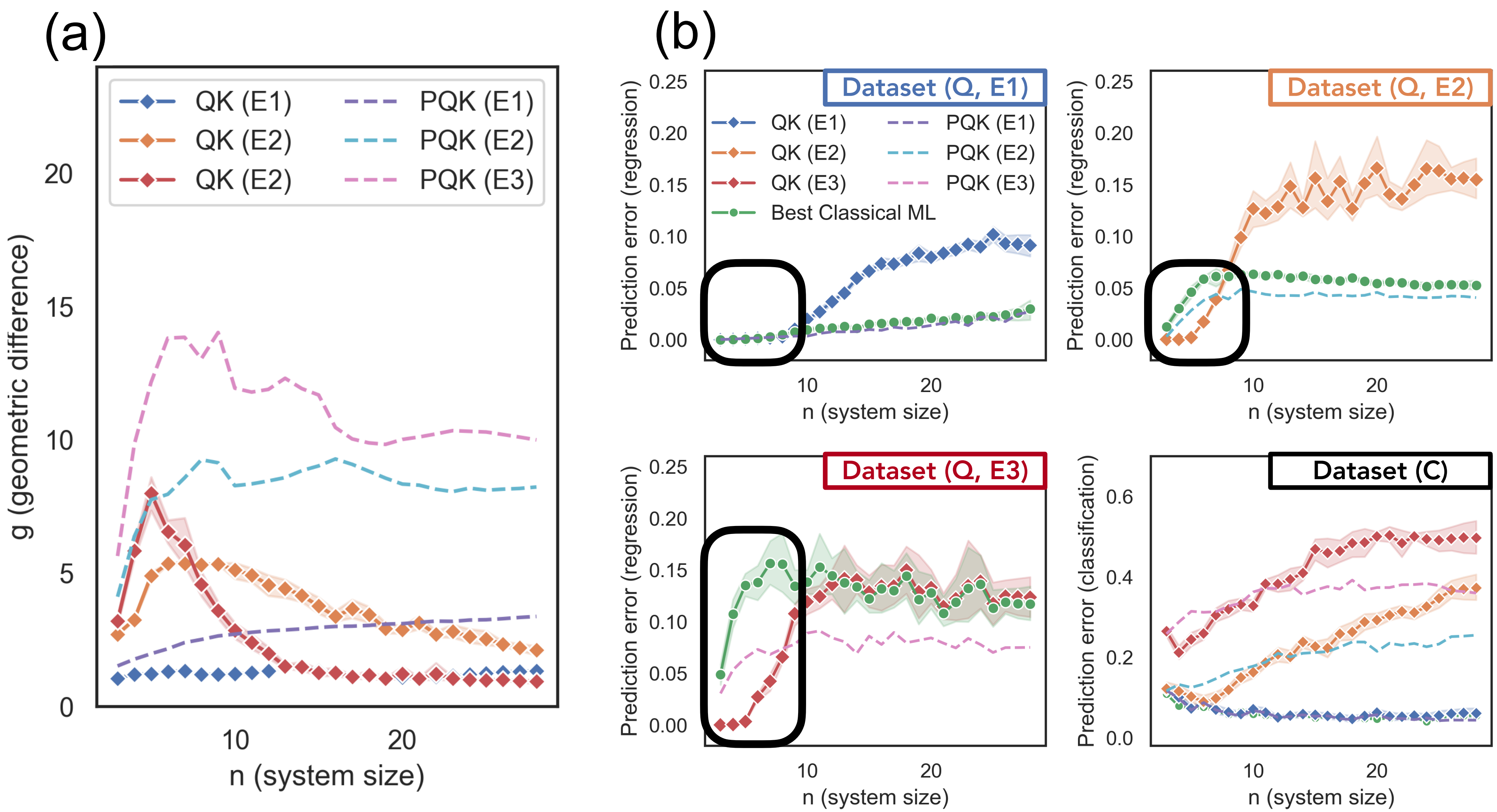
(Q) - Local magnetization after random Heisenberg evolution

Experiments



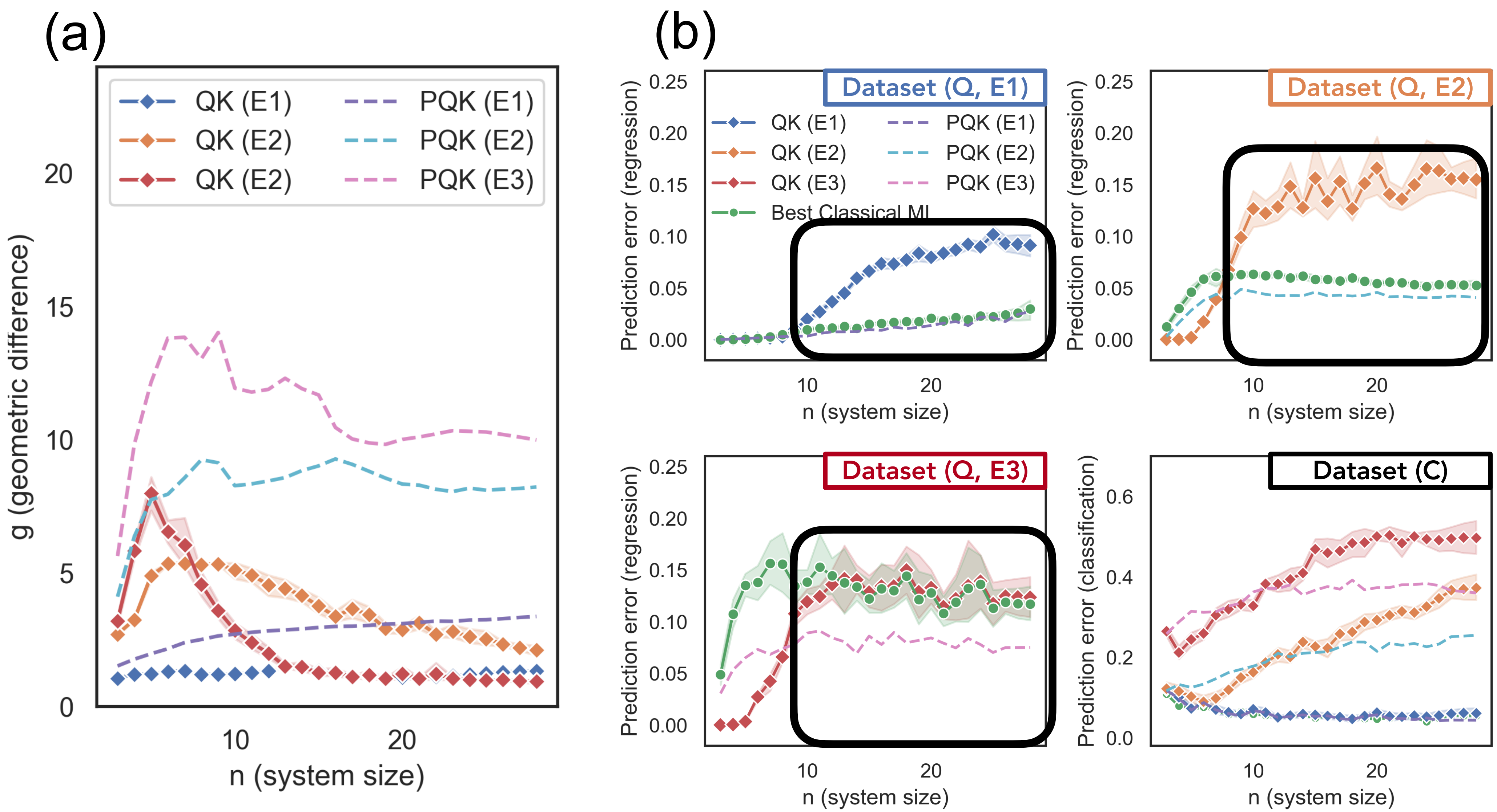
1. Green line is classical ML. Other lines are quantum ML.
2. QK is quantum kernel ML model proposed in [Havlicek, Nature, 2019].
3. PQK is our proposed modified QML to increase geometric difference.

Experiments

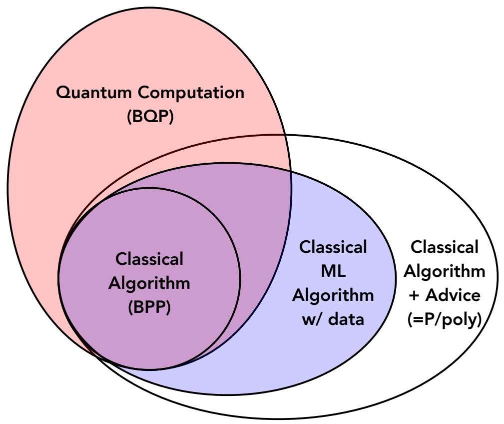


1. Green line is classical ML. Other lines are quantum ML.
2. QK is quantum kernel ML model proposed in [Havlicek, Nature, 2019].
3. PQK is our proposed modified QML to increase geometric difference.

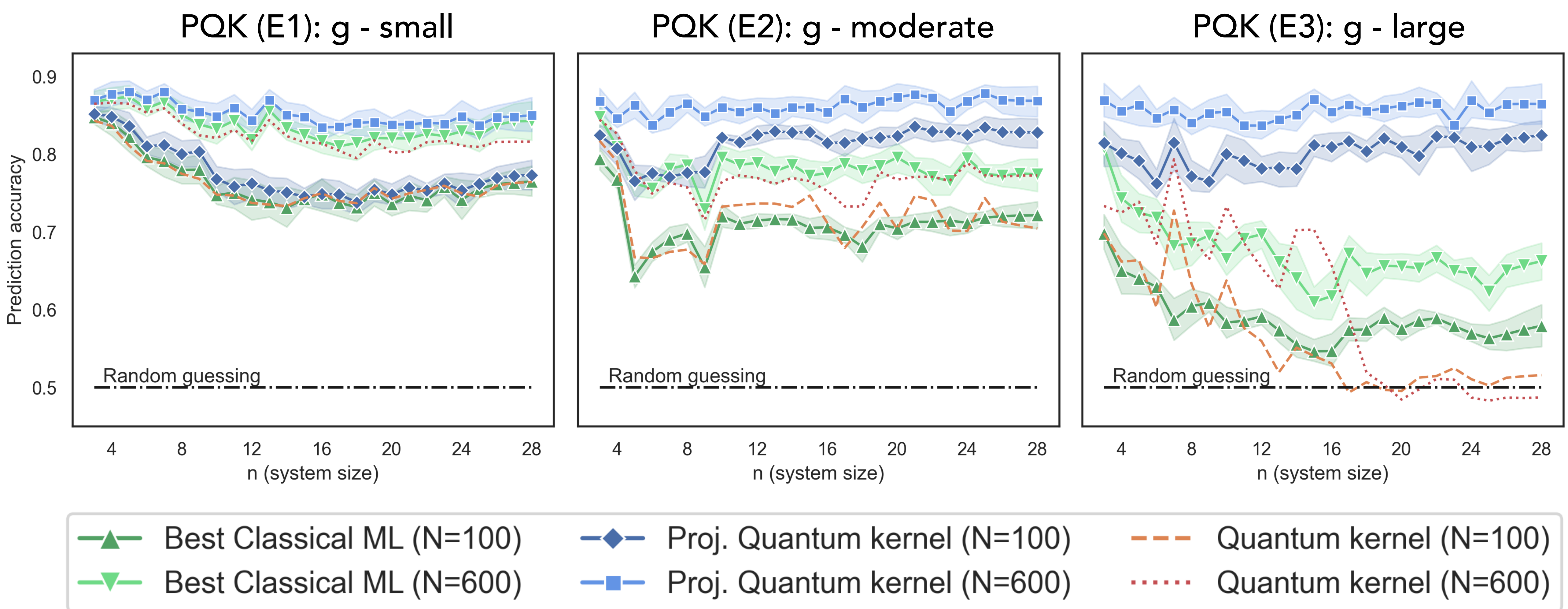
Experiments



1. Green line is classical ML. Other lines are quantum ML.
2. QK is quantum kernel ML model proposed in [Havlicek, Nature, 2019].
3. PQK is our proposed modified QML to increase geometric difference.



Experiments

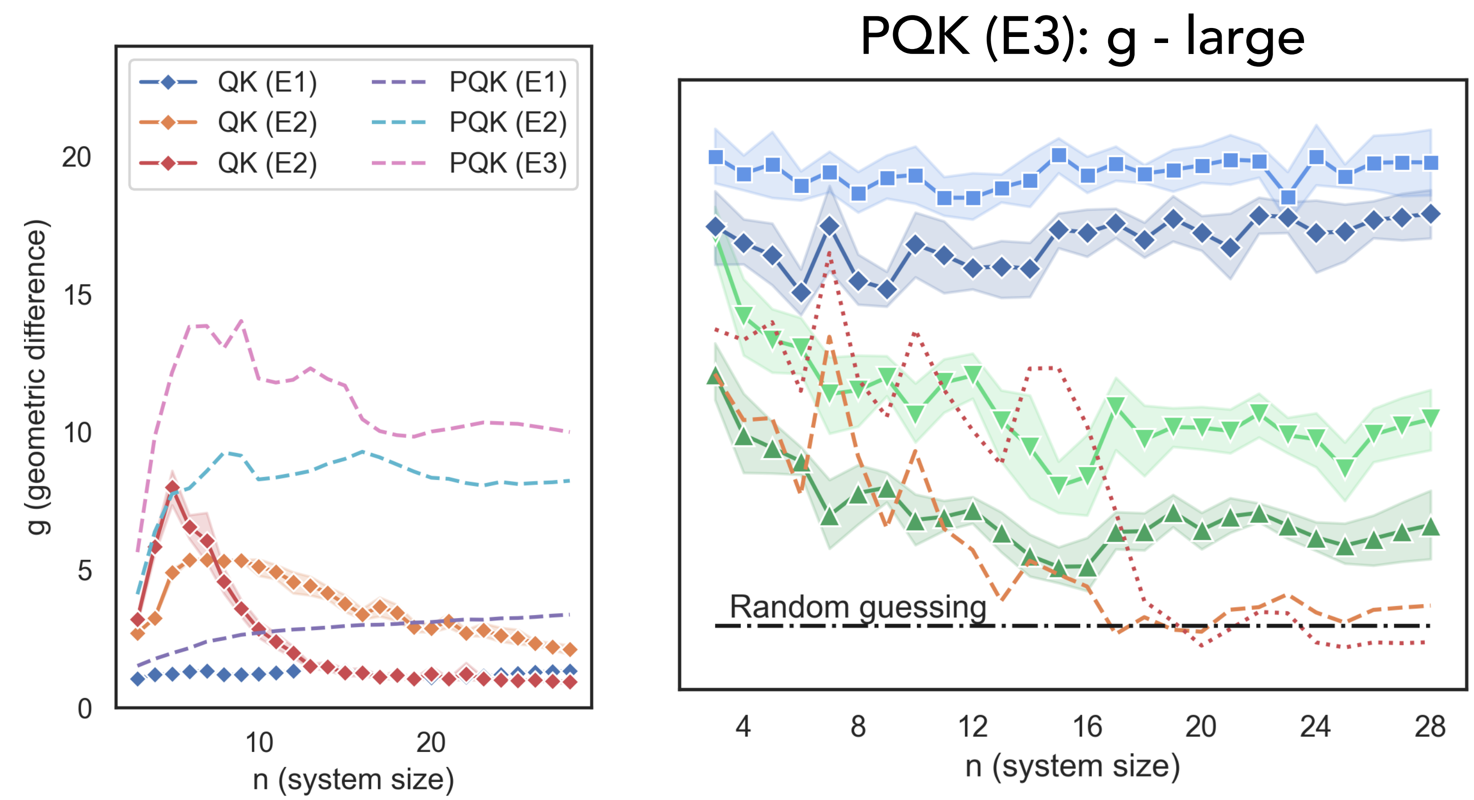


1. When geometric difference is large, data sets exist with large prediction advantage.
2. One can see significant advantage using quantum ML for these data sets.

Making sure things scale to large system size

TensorFlow Quantum

https://www.tensorflow.org/quantum



~ 1 petaflop/s peak, ~1 **exaflop** total

TF-Quantum Tutorial Implementation - https://www.tensorflow.org/quantum/tutorials/quantum_data
Blog Post - https://blog.tensorflow.org/2020/11/characterizing-quantum-advantage-in.html

Credit - Michael Broughton

Main Questions

Exponential separation
for worst-case error.

Information-theoretic aspect:

Do we need significantly more experiments in the classical setting compared to the quantum setting to learn $f_{\mathscr{E}}(x)$? **No!**

[1] **Information-theoretic bounds on quantum advantage in machine learning**, *arXiv:2101.02464*.

Computational aspect:

Could classical ML use data to efficiently compute $f_{\mathscr{E}}(x) = \mathrm{Tr}(O\mathscr{E}(|x\rangle\langle x|))$ even if $f_{\mathscr{E}}(x)$ is hard to compute with a classical computer?

[2] **Power of data in quantum machine learning**, *arXiv:2011.01938*.

Main Questions

Exponential separation
for worst-case error.

Information-theoretic aspect:

Do we need significantly more experiments in the classical setting compared to the quantum setting to learn $f_{\mathscr{E}}(x)$? **No!**

[1] **Information-theoretic bounds on quantum advantage in machine learning**, *arXiv:2101.02464*.

Computational aspect:

Could classical ML use data to efficiently compute $f_{\mathscr{E}}(x) = \mathrm{Tr}(O\mathscr{E}(|x\rangle\langle x|))$ even if $f_{\mathscr{E}}(x)$ is hard to compute with a classical computer? **Yes!**

[2] **Power of data in quantum machine learning**, *arXiv:2011.01938*.

Main Questions

Information-theoretic aspect:

Exponential separation for worst-case error.

Do we need significantly more experiments in the classical setting compared to the quantum setting to learn $f_{\mathscr{E}}(x)$? **No!**

[1] Information-theoretic bounds on quantum advantage in machine learning, *arXiv:2101.02464*.

Computational aspect:

Quantum ML is still computationally more powerful.

Could classical ML use data to efficiently compute $f_{\mathscr{E}}(x) = \mathrm{Tr}(O\mathscr{E}(|x\rangle\langle x|))$ even if $f_{\mathscr{E}}(x)$ is hard to compute with a classical computer? **Yes!**

[2] Power of data in quantum machine learning, *arXiv:2011.01938*.

Conclusion

- Learning from classical data is powerful for achieving small average-case prediction error.
- Data provide computational power that enables classical ML algorithms to become stronger than one expects.
- Data challenges quantum advantage in ML problems.
- But **quantum advantage** in prediction accuracy is still possible — more investigations are needed to fully claim quantum advantage.