Information-theoretic bounds on quantum advantage in machine learning

arXiv:2101.02464

Hsin-Yuan (Robert) Huang, Richard Kueng, and John Preskill

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

JOHANNES KEPLER UNIVERSITY LINZ

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

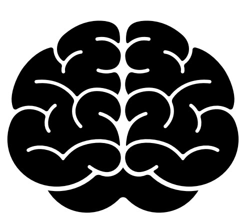
“Solving the quantum many-body problem with artificial neural networks.” *Science* 355.6325 (2017): 602-606.
"Learning phase transitions by confusion." *Nature Physics* 13.5 (2017): 435-439.
"Supervised learning with quantum-enhanced feature spaces." *Nature* 567.7747 (2019): 209-212.

Motivation

- Yet, many fundamental questions remain to be answered.

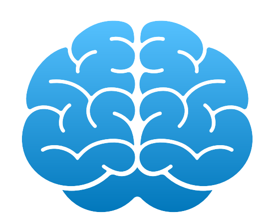
Classical ML
for quantum physics/chemistry

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

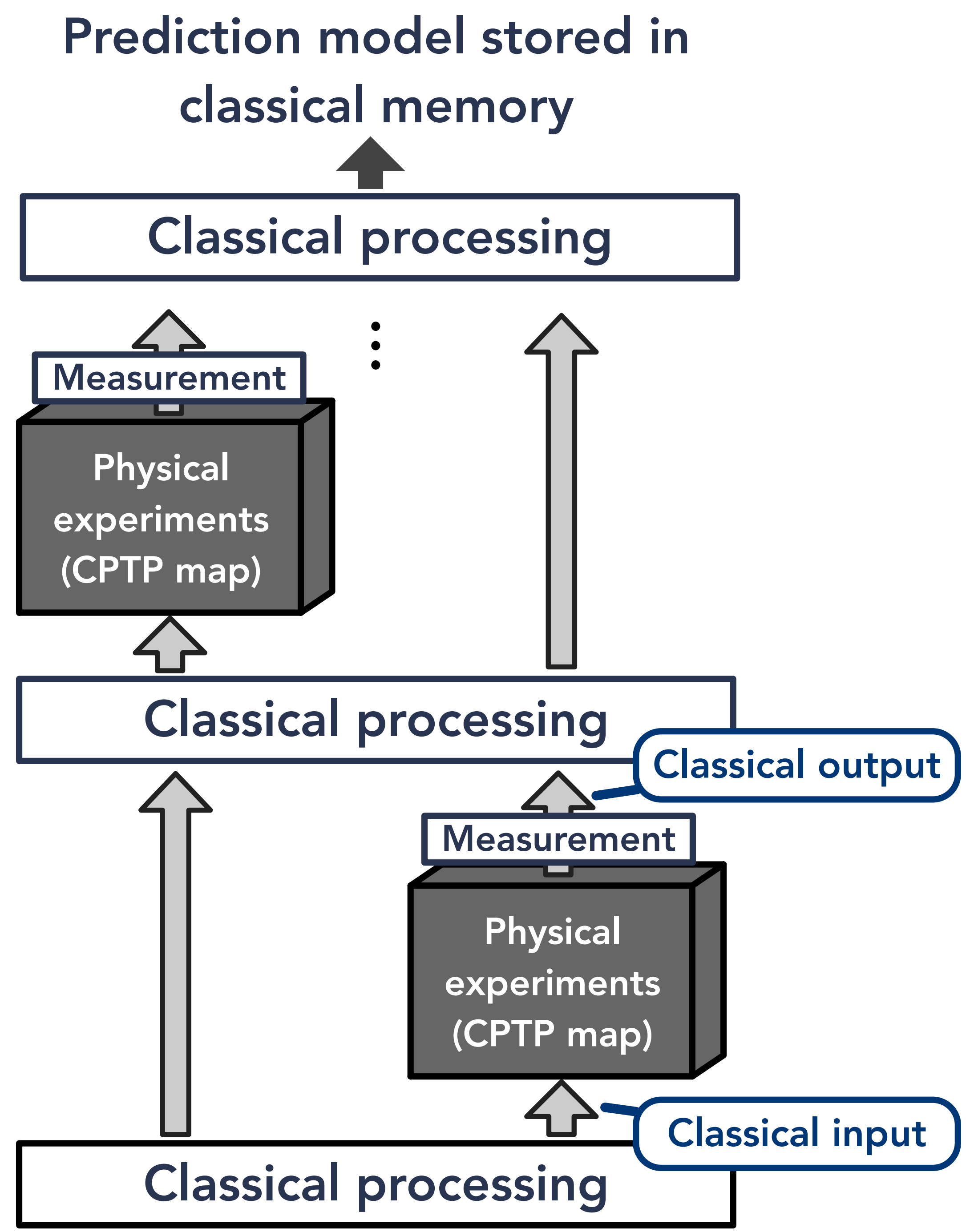
$\mathscr{E}$: a process for preparing ground state

O : the property we want to predict

General Setting

Classical machine learning

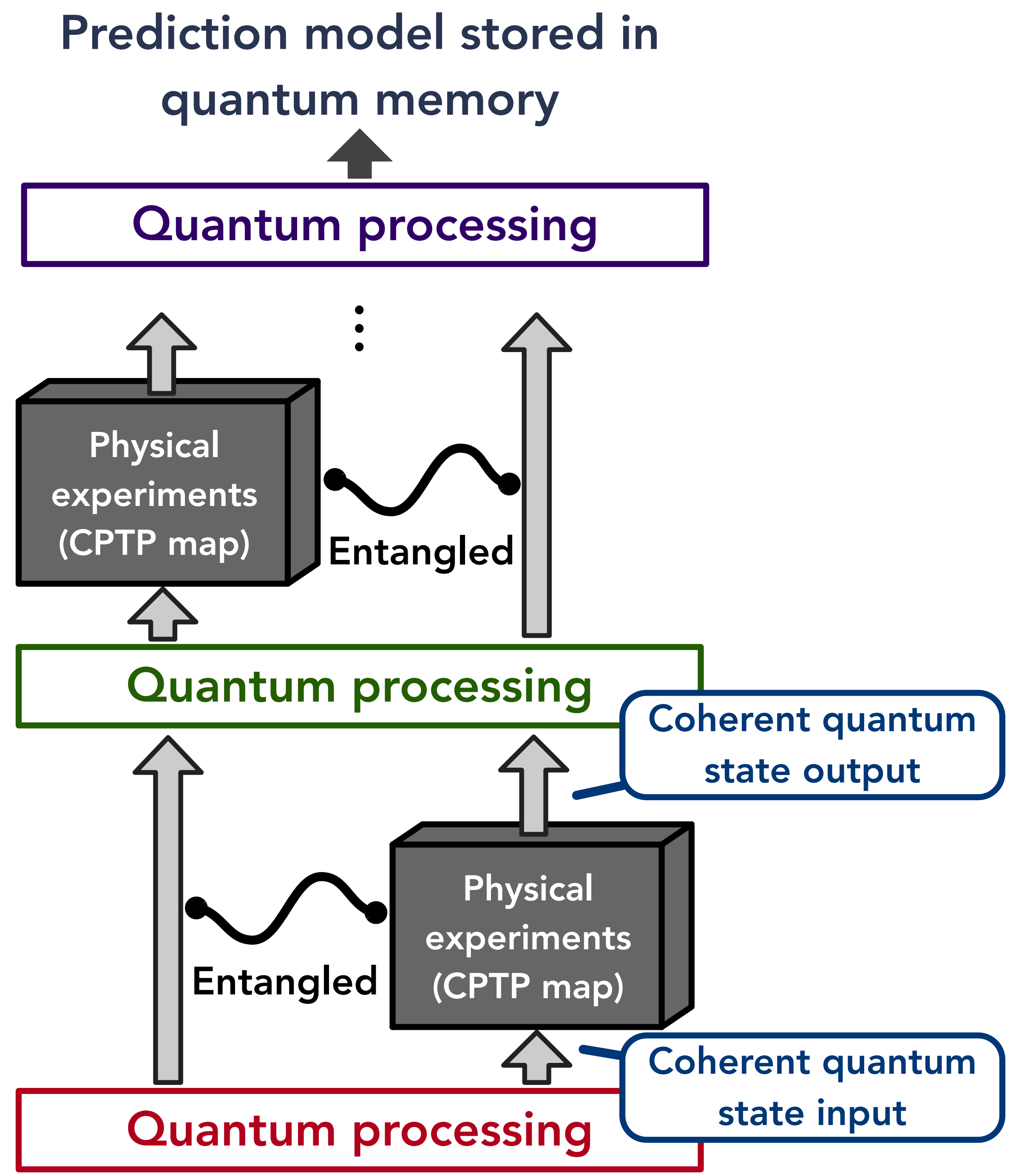
- Learning agents can actively perform experiments to learn a prediction model.
- Each query begins with a choice of classical input x and ends with an arbitrary POVM measurement.
- A prediction model $h(x) \approx f_{\mathscr{E}}(x)$ is created after learning.



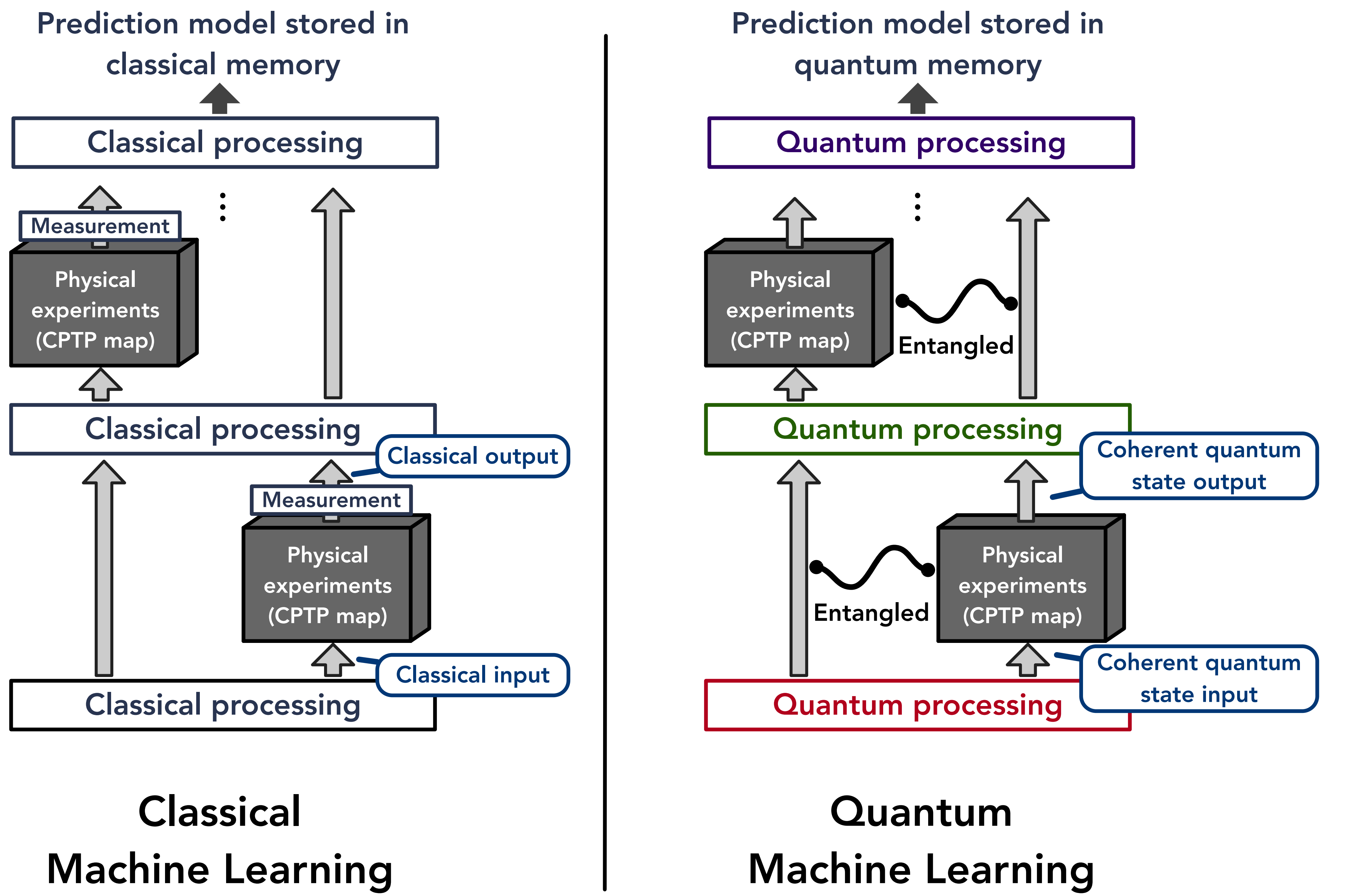
General Setting

Quantum machine learning

- Similar to classical ML setting.
- Each query consists of an arbitrary access to the CPTP map $\mathscr{E}$ (the input can be entangled, and no measurement at the end).
- A prediction model $h(x) \approx f_{\mathscr{E}}(x)$ is stored in a quantum memory instead of a classical memory.



General Setting



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

Main Question

Information-theoretic aspect:

Do classical ML need significantly more experiments (query complexity) than quantum ML to predict $f_{\mathscr{E}}(x) = \mathrm{Tr}(O\mathscr{E}(|x\rangle\langle x|))$?

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

Main Theorem

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.*

Main Theorem

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

Concept/hypothesis class in statistical learning theory

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

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

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

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

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

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

Main Theorem

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 $h_Q(x)$ that achieves a predict

Average prediction error

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

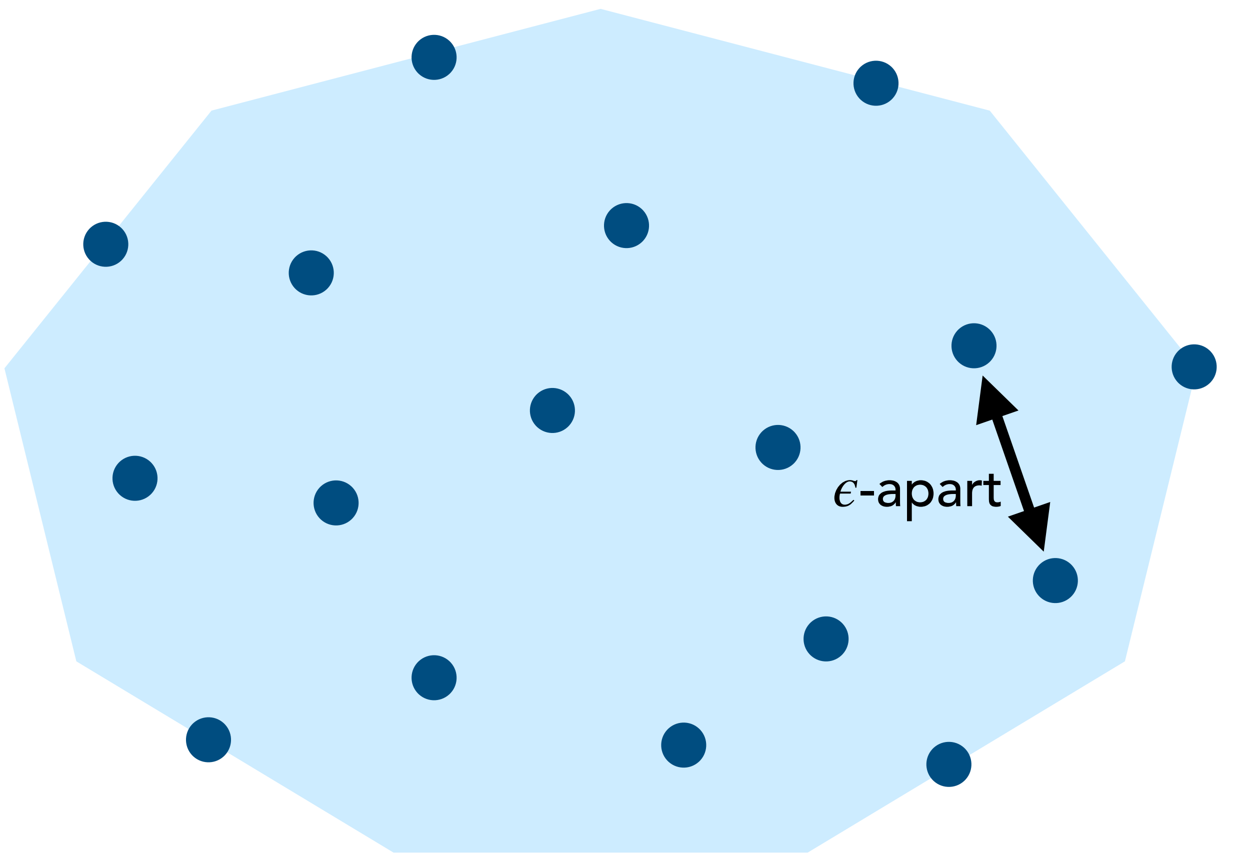
then there is a classical ML using $N_C \leq \mathscr{O}(mN_Q/\epsilon)$ to learn a prediction model $h_C(x)$ that achieves a prediction error of

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

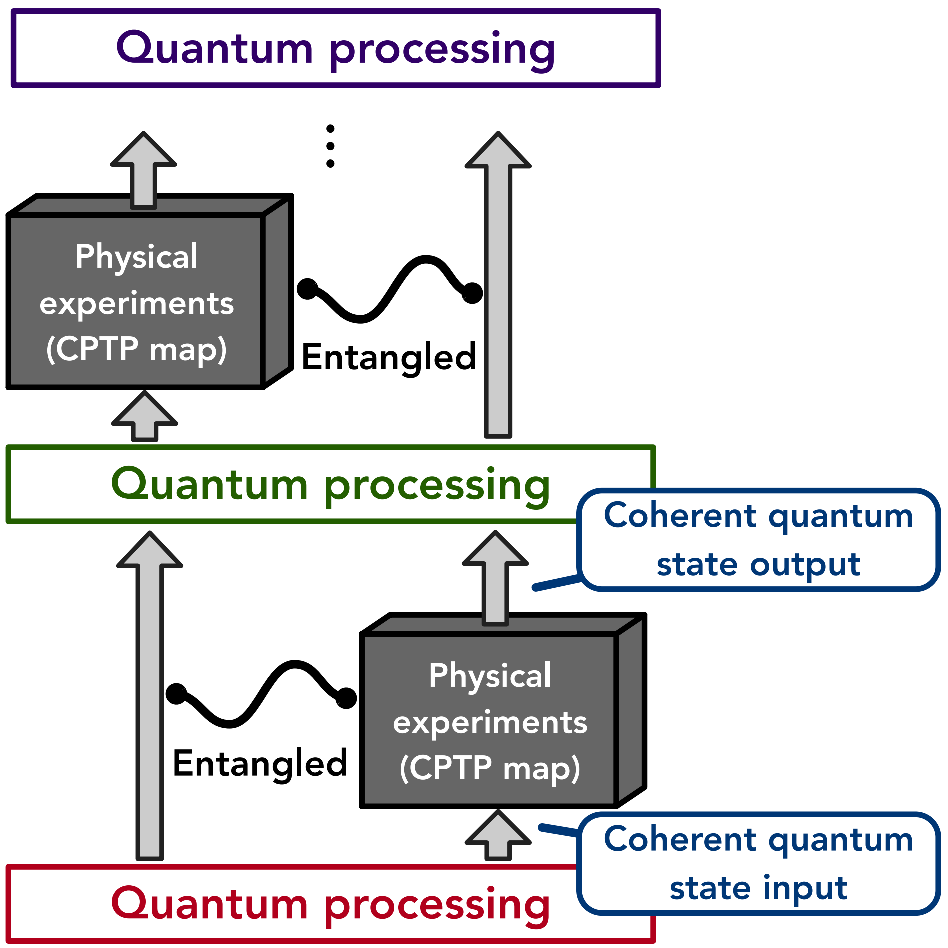
[1] **Information-theoretic bounds on quantum advantage in machine learning**, *arXiv:2101.02464*.

Proof idea:
Quantum lower bound

The set of CPTP maps $\mathscr{F} = \{\mathscr{E}\}$



Construct the maximum packing net M_ϵ^p

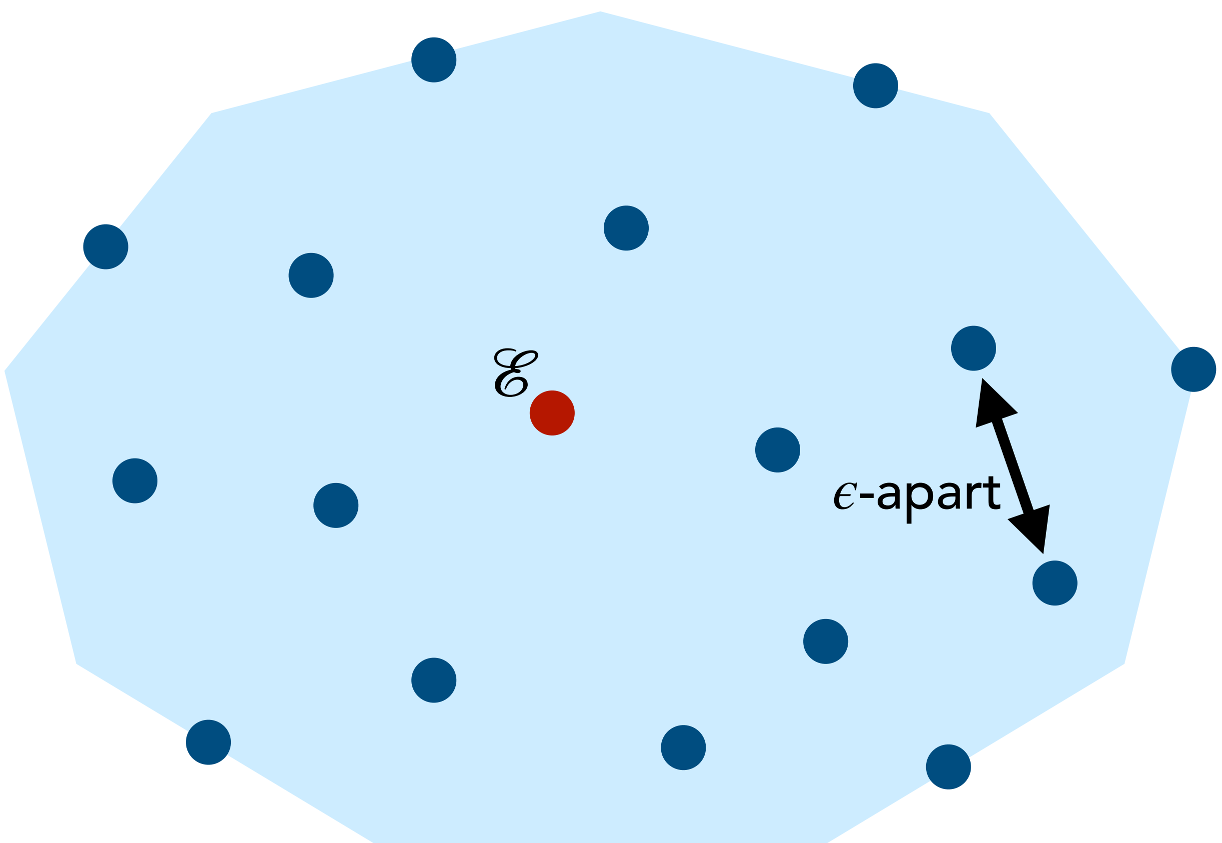


Quantum Machine Learning

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

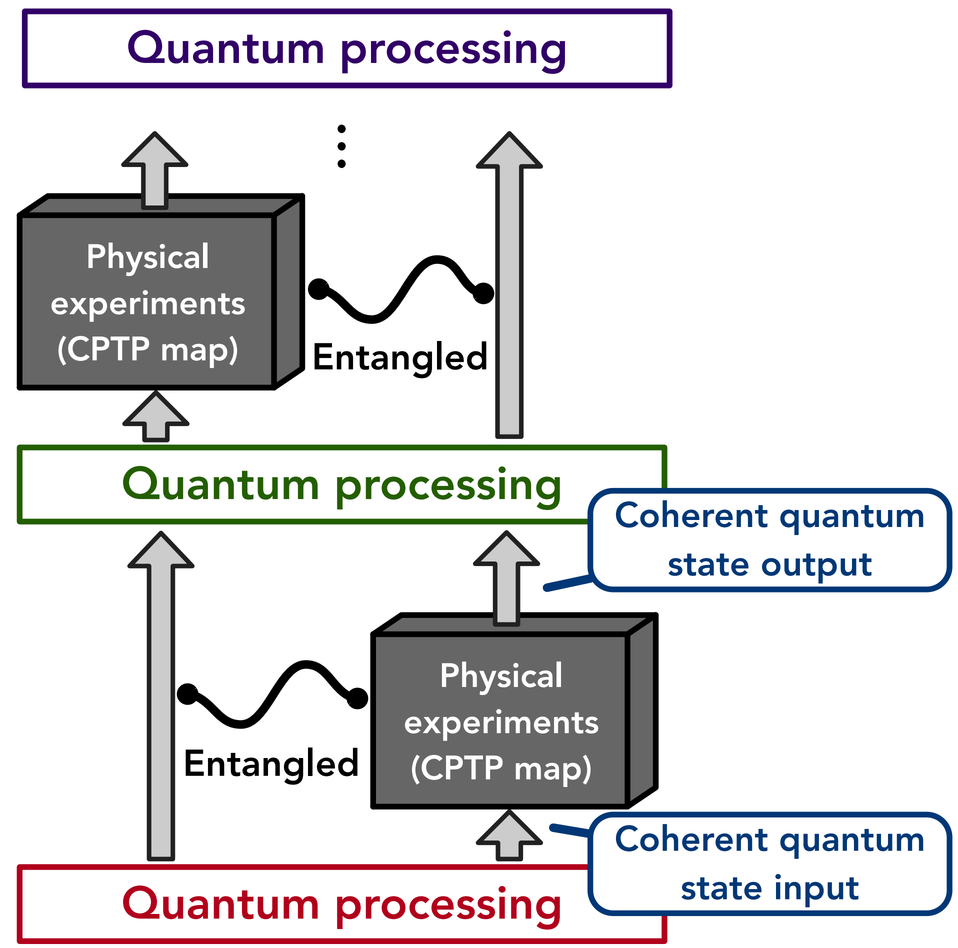
Proof idea:
Quantum lower bound

The set of CPTP maps $\mathscr{F} = \{\mathscr{E}\}$



Construct the maximum packing net M_ϵ^p

Alice chooses a CPTP map $\mathscr{E}$ among M_ϵ^p

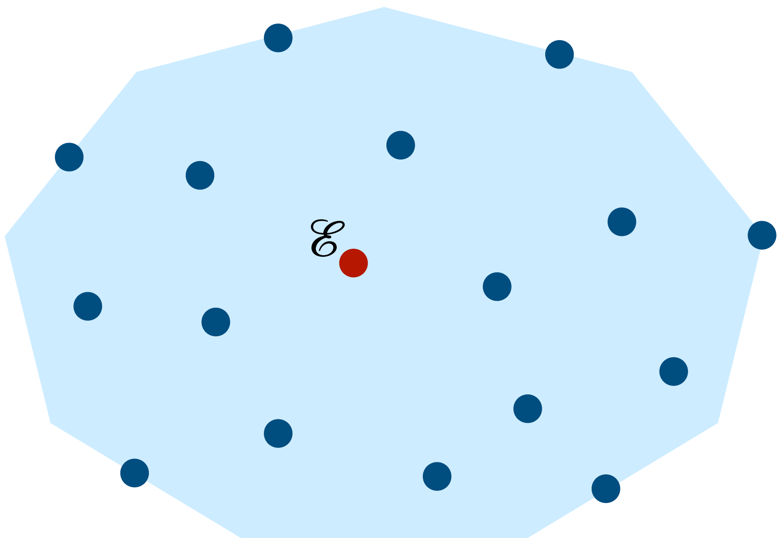


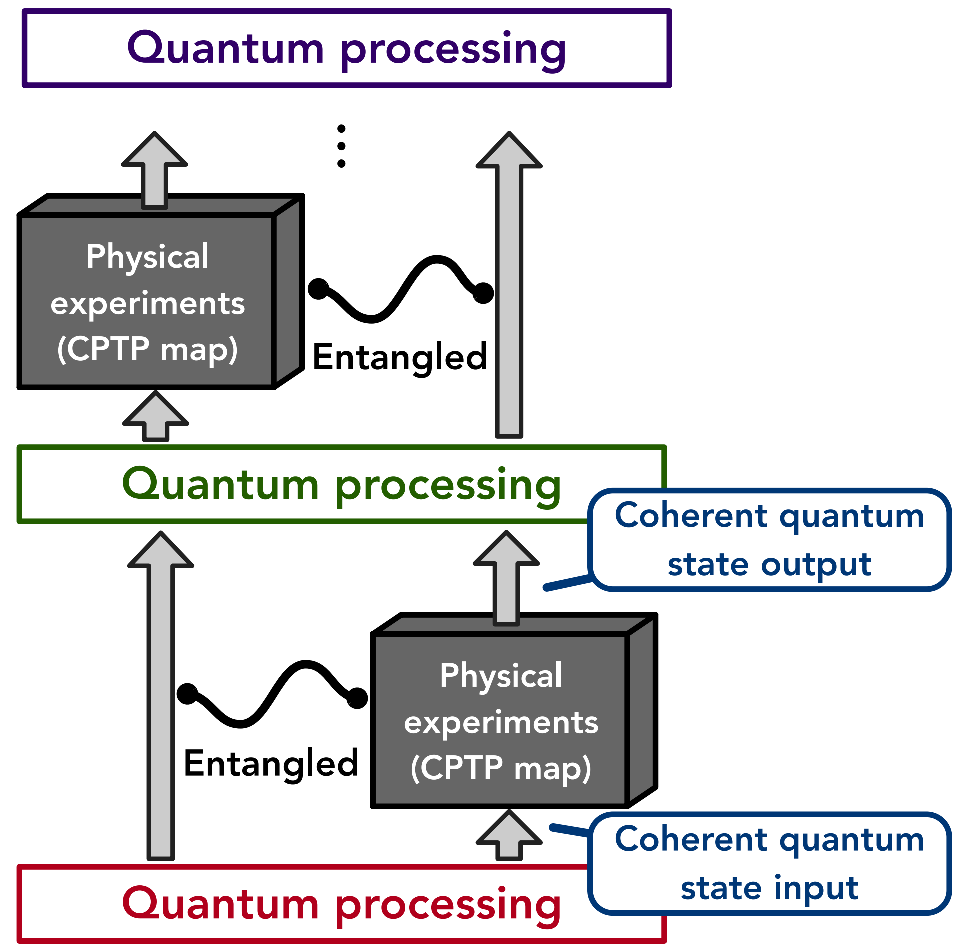
Quantum Machine Learning

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

Proof idea:
Quantum lower bound

1. Alice chooses a CPTP map $\mathscr{E}$ among packing net M_ϵ^p



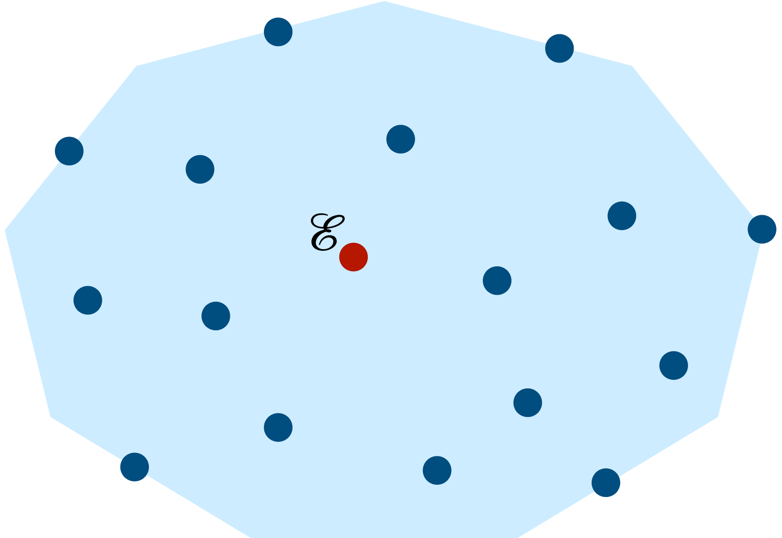


Quantum
Machine Learning

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

Proof idea:
Quantum lower bound

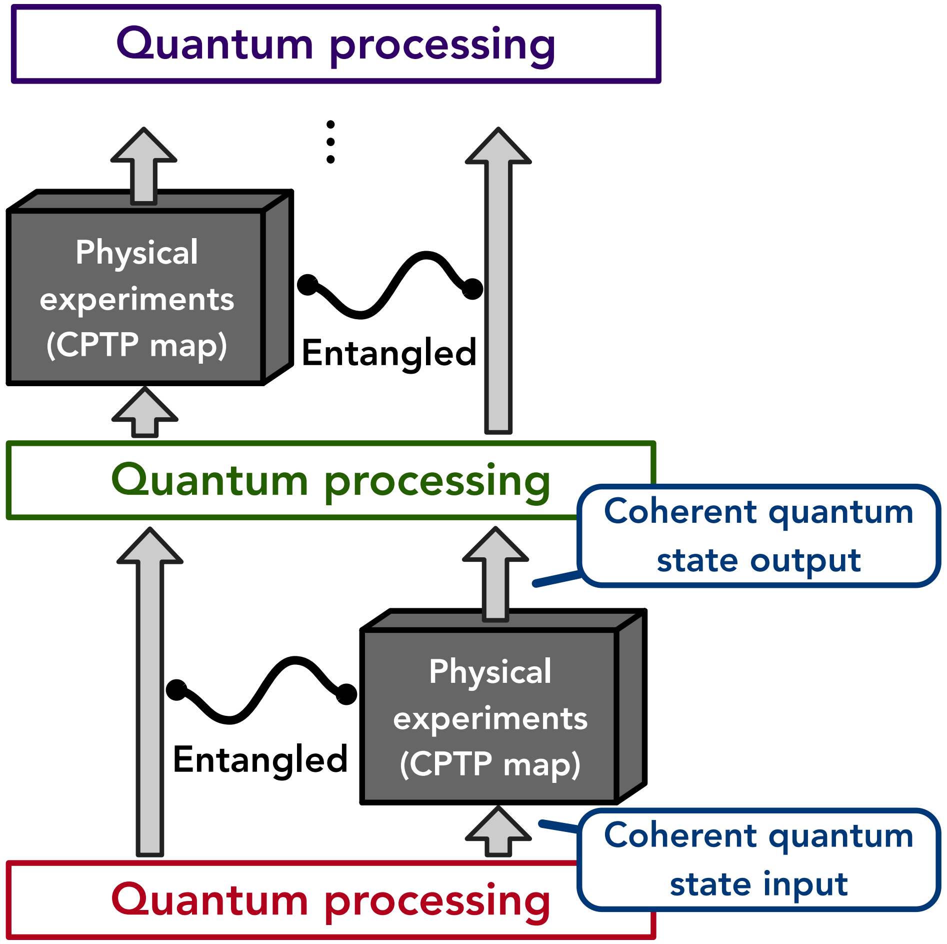
1. Alice chooses a CPTP map $\mathscr{E}$ among packing net M^p_ϵ



2. Bob uses the quantum machine learning algorithm to get

$$\rho_{N_Q,\mathscr{E}} = (\mathscr{E} \otimes I)\ldots C_2(\mathscr{E} \otimes I)C_1(\mathscr{E} \otimes I)(\rho_0),$$

where $C_1, C_2, \ldots$= quantum processing (CPTP maps),
And $\mathscr{E} \otimes I$ is the physical experiment to learn.

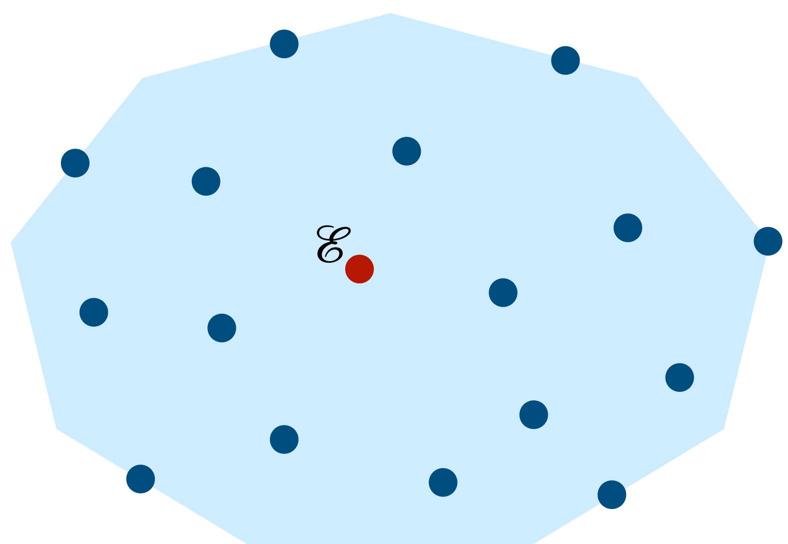


Quantum Machine Learning

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

Proof idea:
Quantum lower bound

1. Alice chooses a CPTP map $\mathscr{E}$ among packing net M_ϵ^p



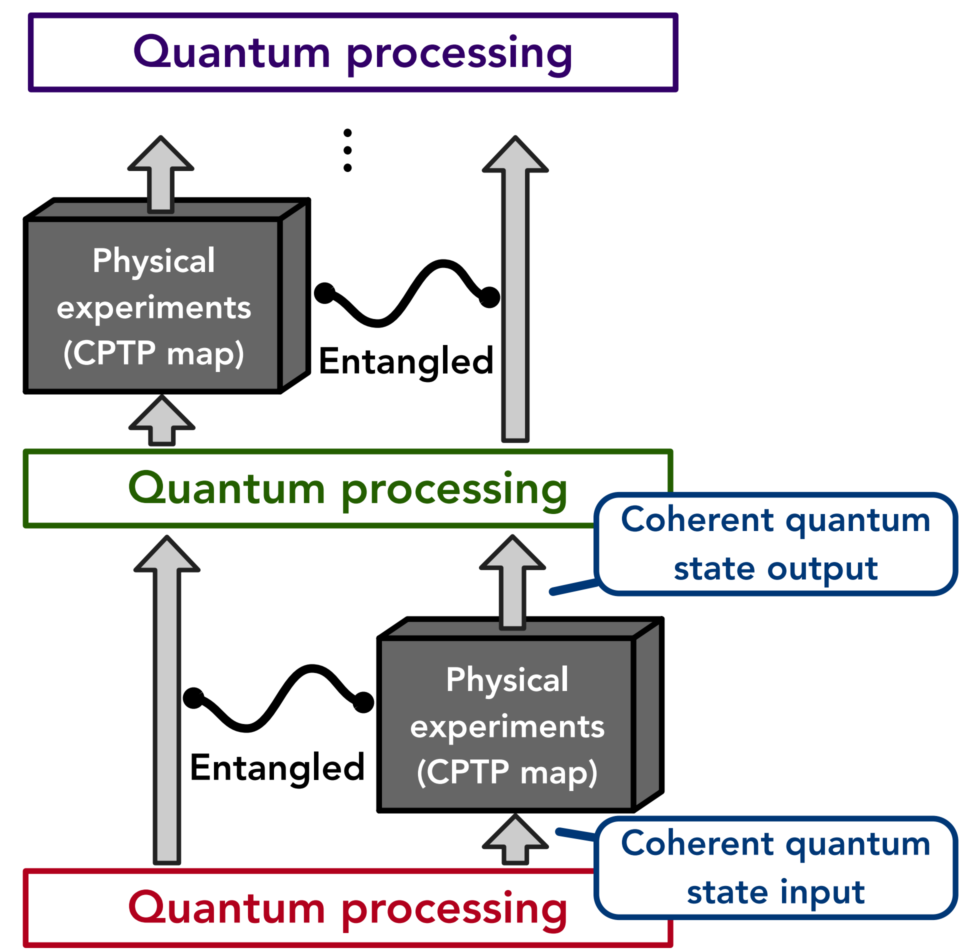
2. Bob uses the quantum machine learning algorithm to get

$$\rho_{N_Q,\mathscr{E}} = (\mathscr{E} \otimes I)\ldots C_2(\mathscr{E} \otimes I)C_1(\mathscr{E} \otimes I)(\rho_0),$$

where $C_1, C_2, \ldots$= quantum processing (CPTP maps),
And $\mathscr{E} \otimes I$ is the physical experiment to learn.

3. Bob can use $\rho_{N_Q,\mathscr{E}}$ to predict $f_{\mathscr{E}}(x) = \mathrm{Tr}(O\mathscr{E}(|x\rangle\langle x|))$

to ϵ-error, so Bob could determine $\mathscr{E}$ (bc. of packing net).



Quantum Machine Learning

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

Proof idea:
Quantum lower bound

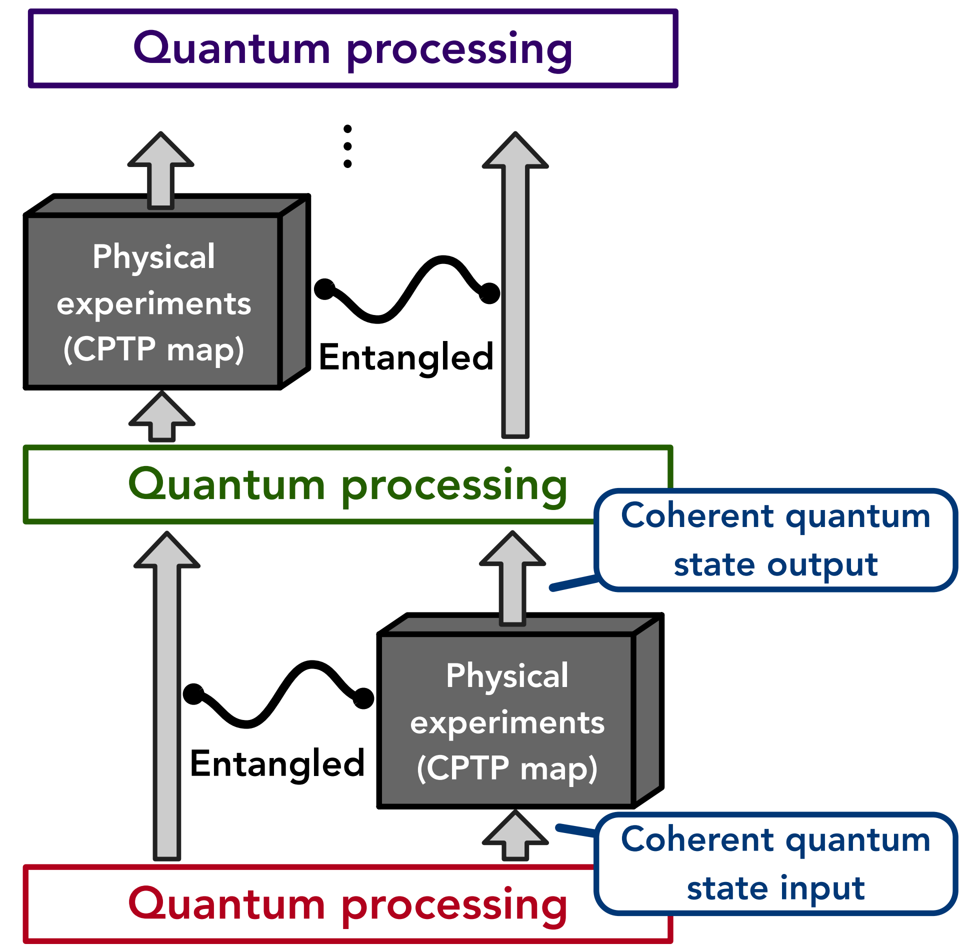
1. Alice chooses a CPTP map $\mathcal{E}$ among packing net M_ϵ^p

2. Bob uses the quantum machine learning algorithm to get
$$\rho_{N_Q,\mathcal{E}} = (\mathcal{E} \otimes I)\ldots C_2(\mathcal{E} \otimes I)C_1(\mathcal{E} \otimes I)(\rho_0).$$

3. Bob can use $\rho_{N_Q,\mathcal{E}}$ to determine $\mathcal{E}$ (bc. of packing net).

⇩

Mutual information between $\mathcal{E} \in M_\epsilon^p$ and $\rho_{N_Q,\mathcal{E}}$ is at least an order of $\log(|M_\epsilon^p|)$.



Quantum Machine Learning

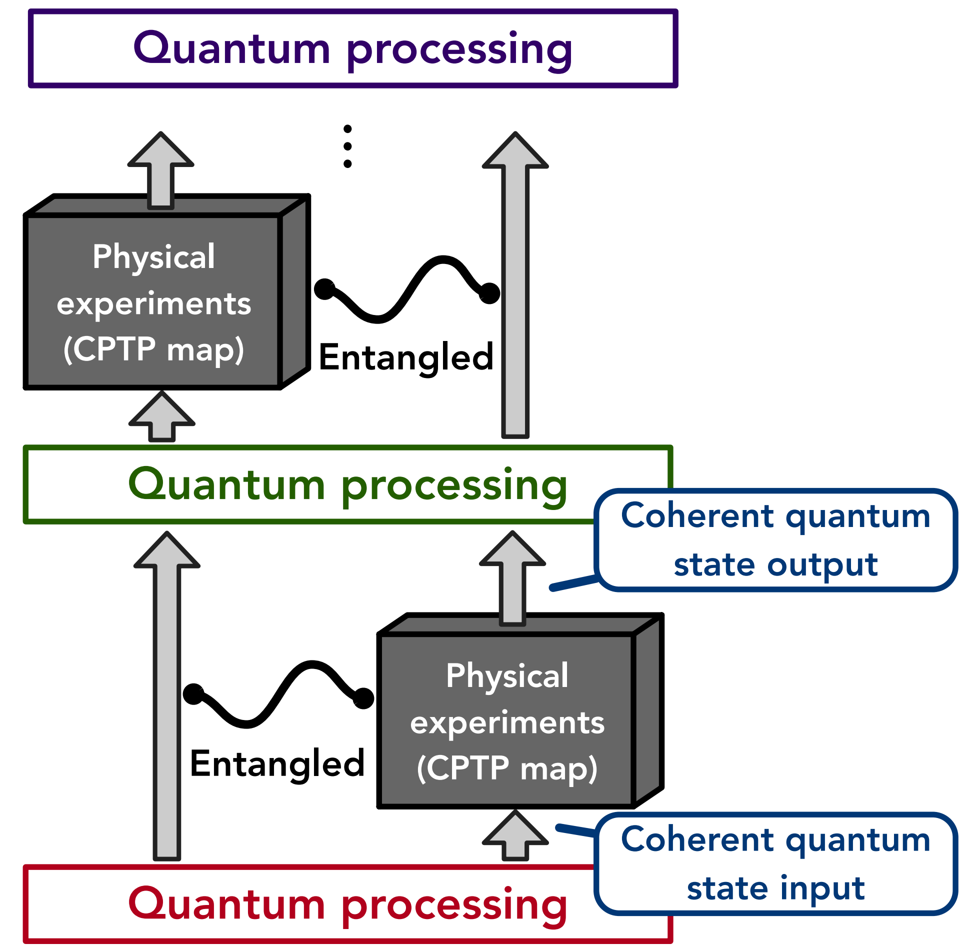
[1] **Information-theoretic bounds on quantum advantage in machine learning**, *arXiv:2101.02464*.

Proof idea:
Quantum lower bound

Mutual information between $\mathscr{E} \in M_\epsilon^p$ and $\rho_{N_Q,\mathscr{E}}$ is at least an order of $\log(|M_\epsilon^p|)$.

$$(\mathscr{E} \otimes I)\ldots\underline{C_2}(\mathscr{E} \otimes I)C_1(\mathscr{E} \otimes I)(\rho_0)$$

Does not increase the information

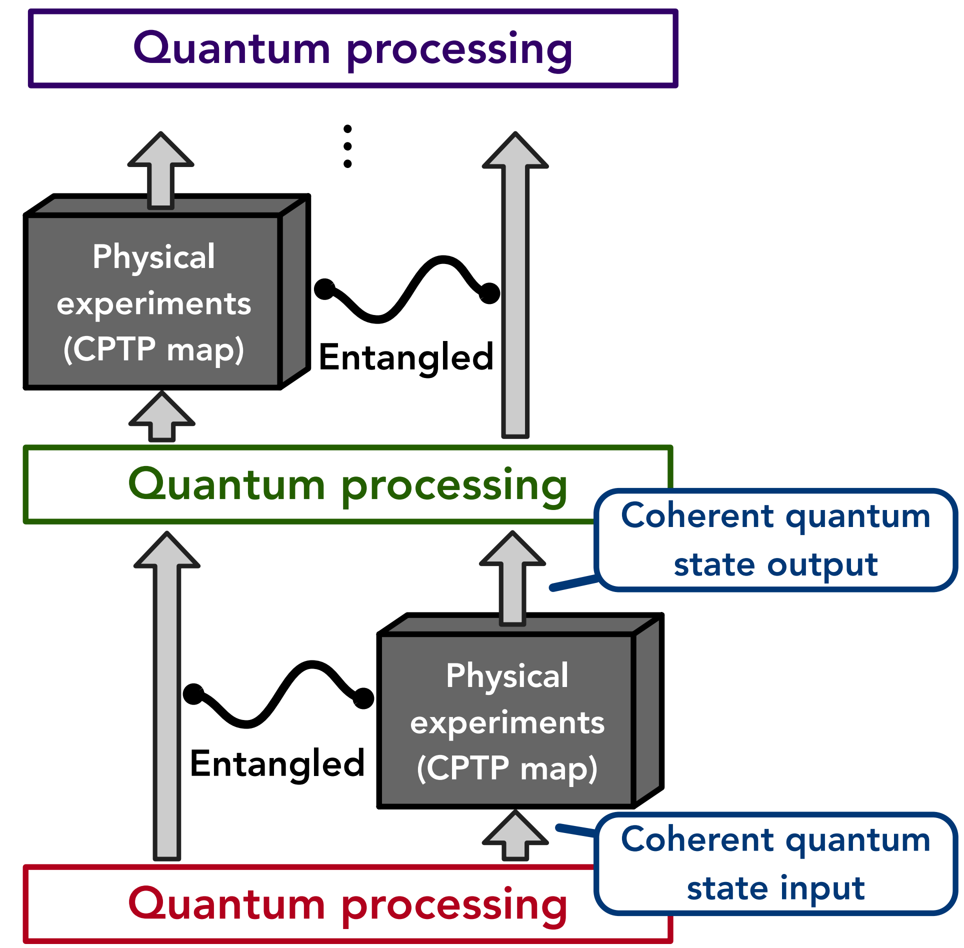


Quantum Machine Learning

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

Proof idea:
Quantum lower bound

Mutual information between $\mathscr{E} \in M_\epsilon^p$ and $\rho_{N_Q,\mathscr{E}}$ is at least an order of $\log(|M_\epsilon^p|)$.



$$(\mathscr{E} \otimes I)\ldots C_2(\mathscr{E} \otimes I)C_1(\mathscr{E} \otimes I)(\rho_0)$$

Each query increases information by at most order m

Quantum Machine Learning

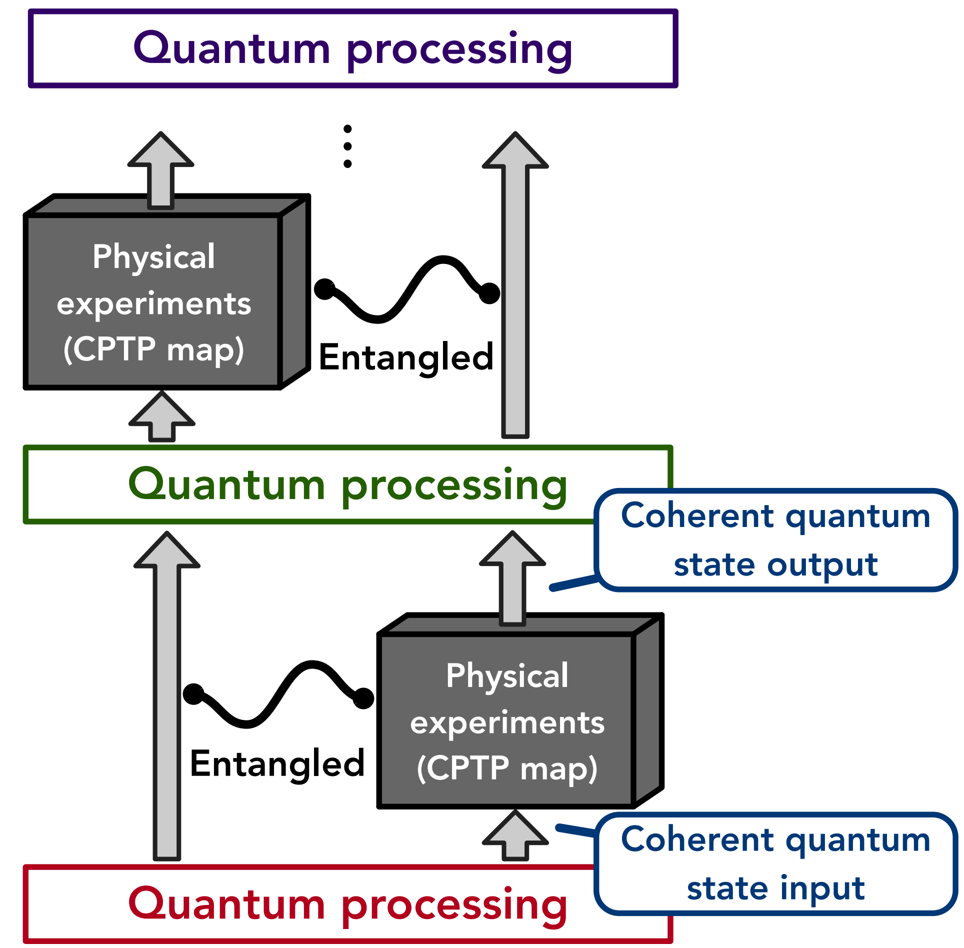
[1] **Information-theoretic bounds on quantum advantage in machine learning**, *arXiv:2101.02464*.

Proof idea: Quantum lower bound

Mutual information between $\mathscr{E} \in M_\epsilon^p$ and $\rho_{N_Q,\mathscr{E}}$ is at least an order of $\log(|M_\epsilon^p|)$.

$$(\mathscr{E} \otimes I)\ldots C_2(\mathscr{E} \otimes I)C_1(\mathscr{E} \otimes I)(\rho_0)$$

Mutual information is upper bounded by order mN_Q



Quantum Machine Learning

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

Proof idea:
Quantum lower bound

Mutual information between $\mathscr{E} \in M_\epsilon^p$ and $\rho_{N_Q,\mathscr{E}}$ is at least an order of $\log(|M_\epsilon^p|)$.

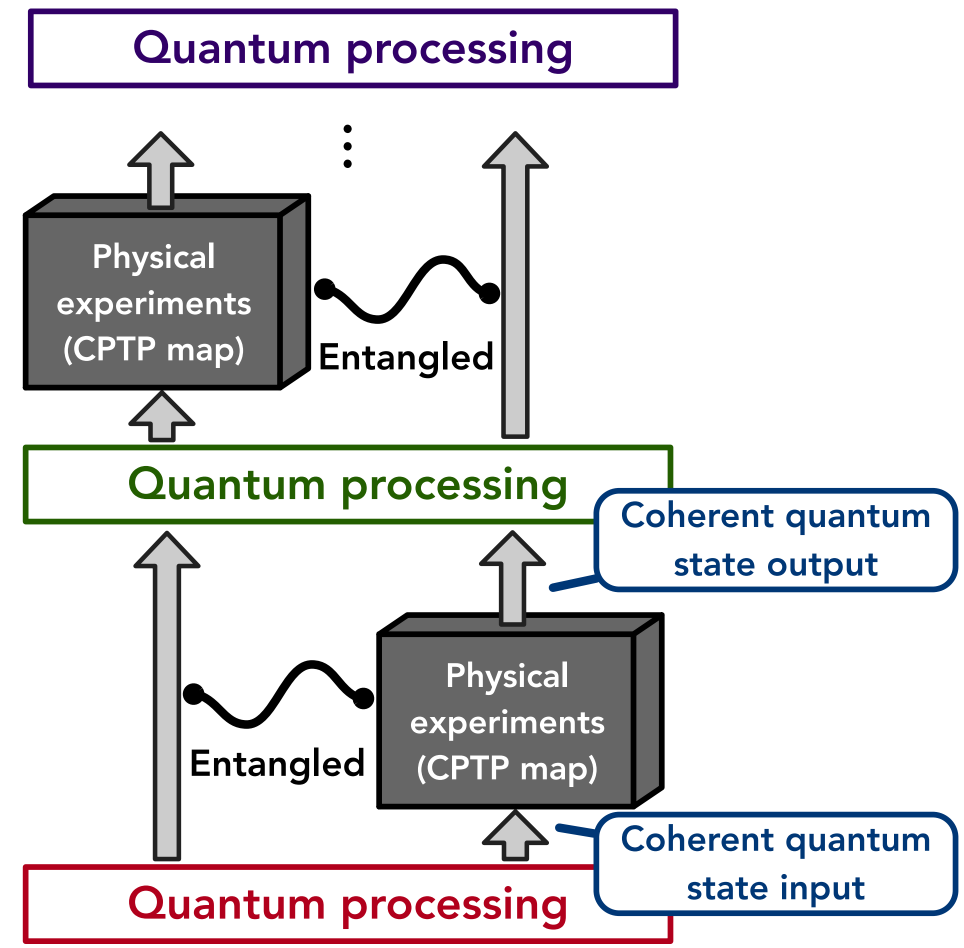
+

$$(\mathscr{E} \otimes I)\ldots C_2(\mathscr{E} \otimes I)C_1(\mathscr{E} \otimes I)(\rho_0)$$

Mutual information is upper bounded by order mN_Q

=

$$N_Q \geq \Omega(\log(|M_\epsilon^p|)/m)$$

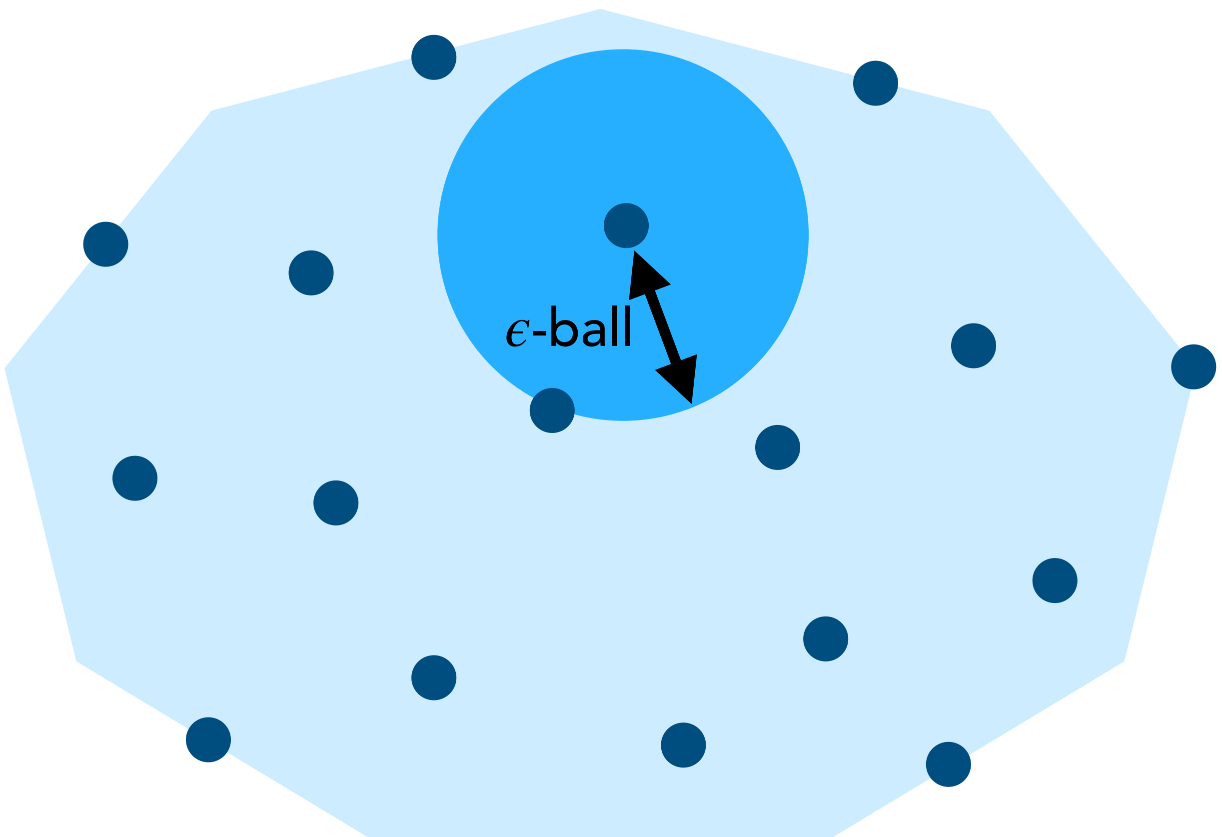


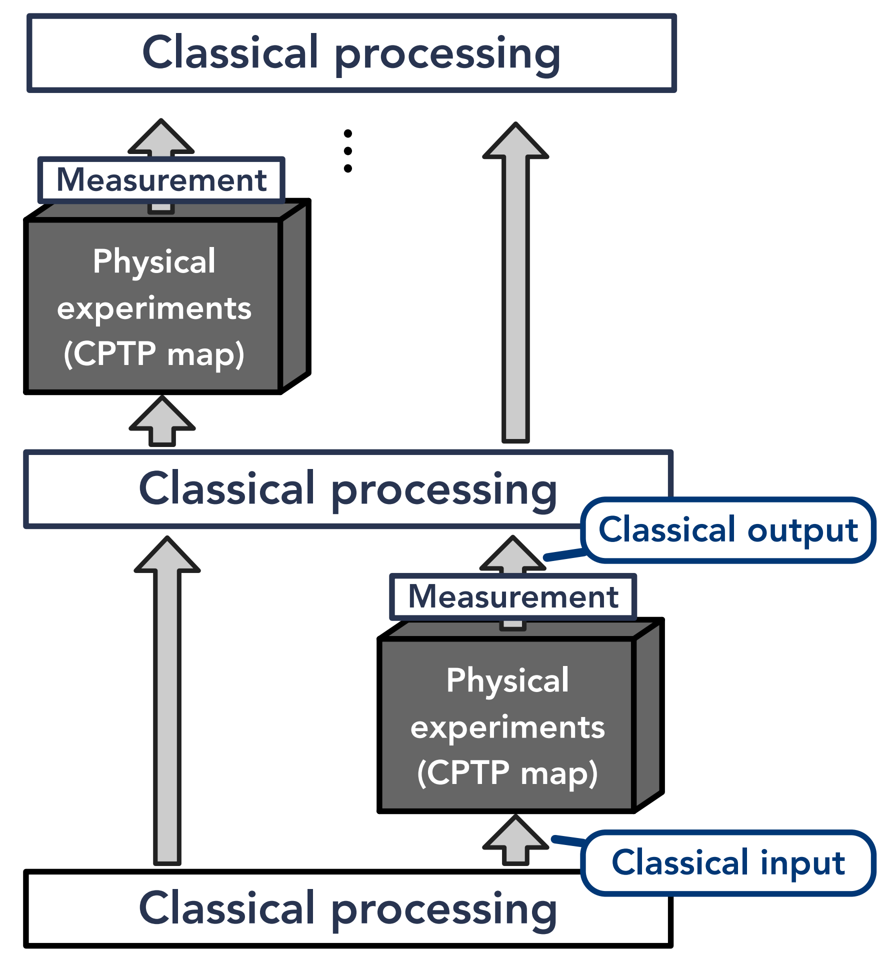
Quantum Machine Learning

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

Proof idea:
Classical upper bound

The set of CPTP maps $\mathscr{F} = \{\mathscr{E}\}$





Construct the maximum packing net M_ϵ^p

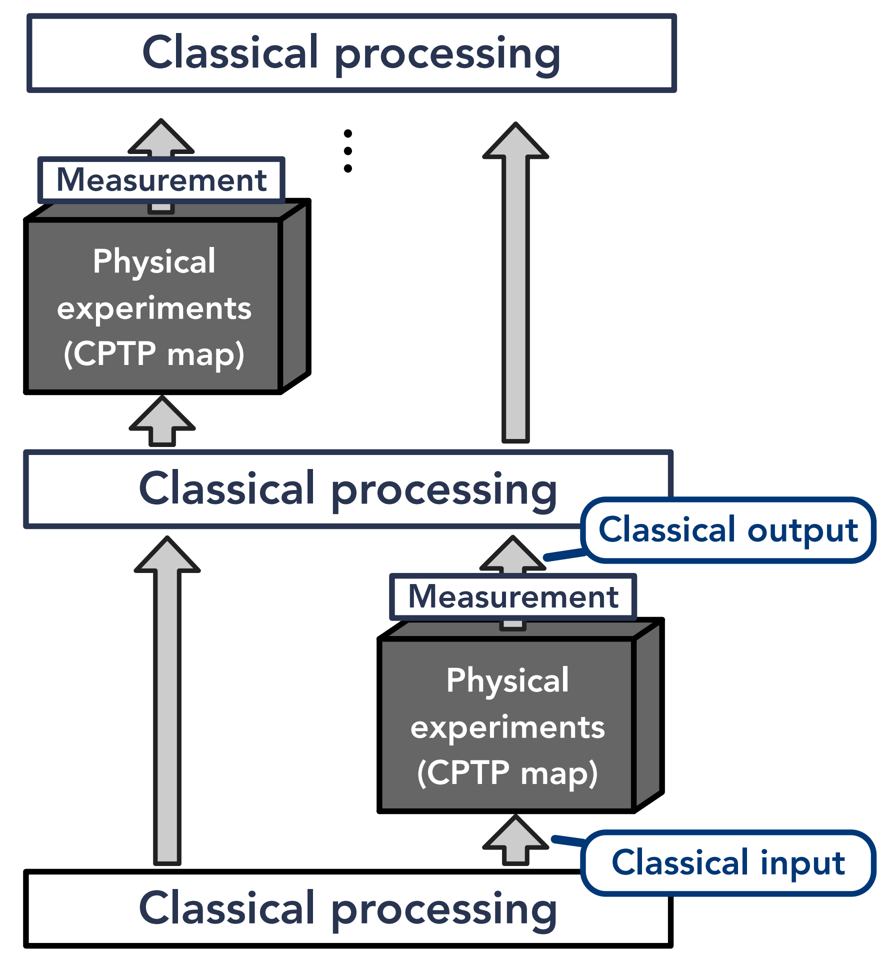
M_ϵ^p covers the entire set $\mathscr{F}$ with ϵ-ball.

Classical Machine Learning

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

Proof idea:
Classical upper bound

1. Randomly select inputs $x_1, \ldots, x_{N_C}$ from distribution $\mathscr{D}$



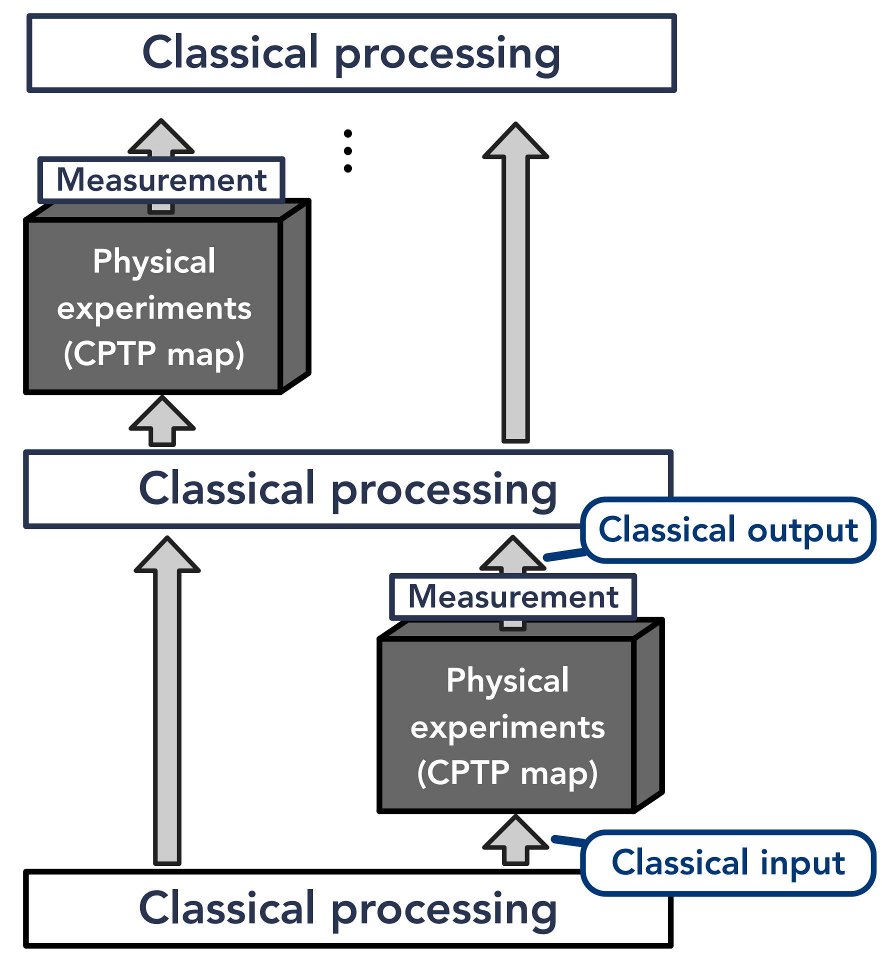
Classical Machine Learning

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

Proof idea:
Classical upper bound

1. Randomly select inputs $x_1, \ldots, x_{N_C}$ from distribution $\mathscr{D}$

2. Measure observable O on the output state of the CPTP map that takes in input x_i to obtain outcome o_i.



Classical Machine Learning

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

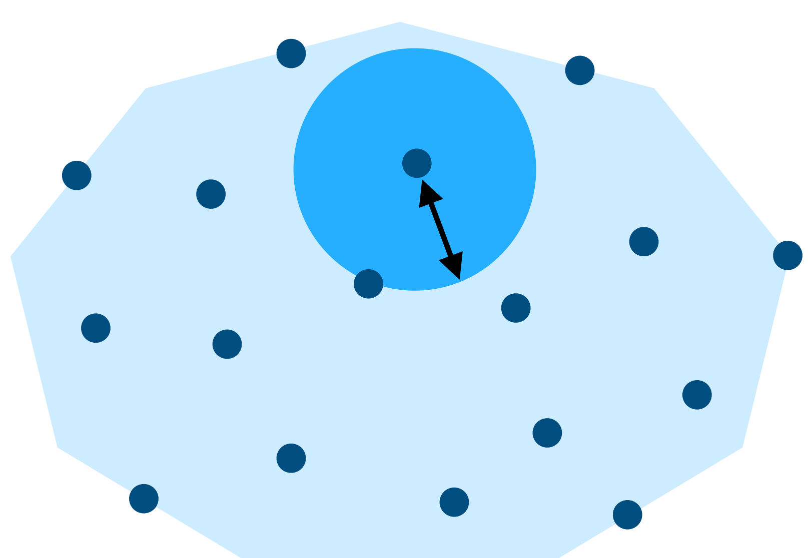
Proof idea:
Classical upper bound

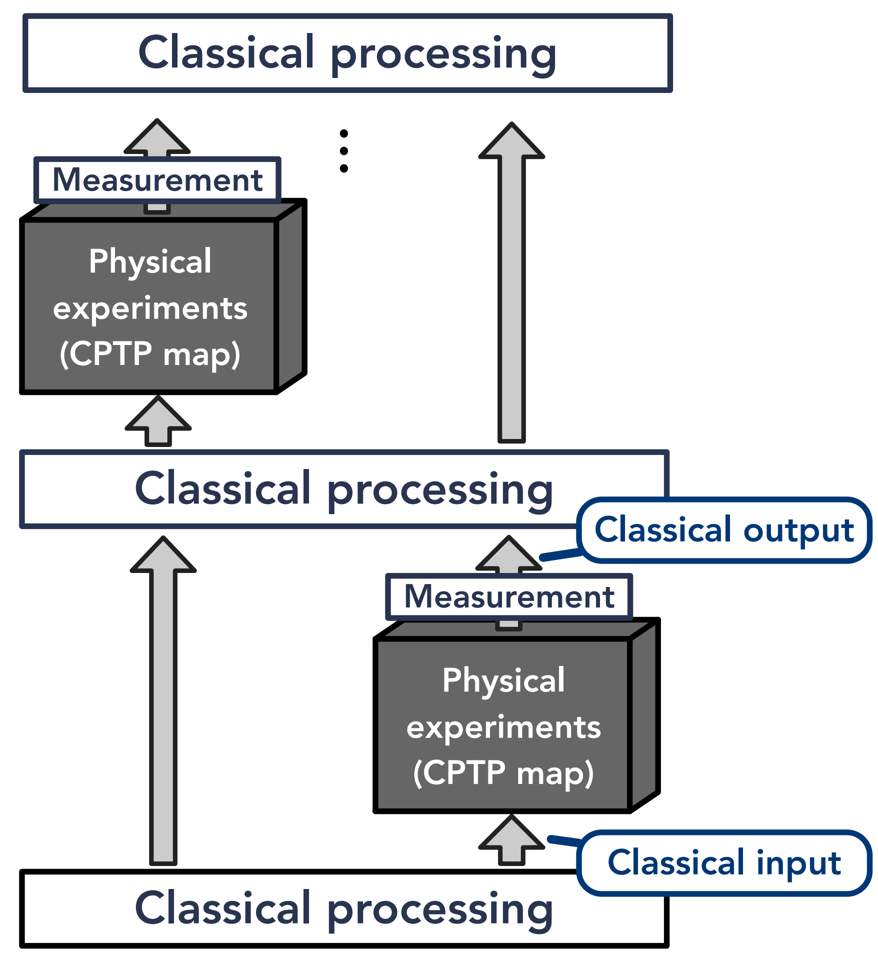
1. Randomly select inputs $x_1, \ldots, x_{N_C}$ from distribution $\mathscr{D}$

2. Measure observable O on the output state of the CPTP map that takes in input x_i to obtain outcome o_i.

3. Output the function h_C from M_ϵ^p that minimizes

$$\frac{1}{N_C}\sum_{i=1}^{N_C} |h_C(x_i) - o_i|^2.$$





Classical Machine Learning

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

Proof idea:
Classical upper bound

1. Randomly select inputs $x_1, \ldots, x_{N_C}$ from distribution $\mathcal{D}$

2. Measure observable O on the output state of the CPTP map that takes in input x_i to obtain outcome o_i.

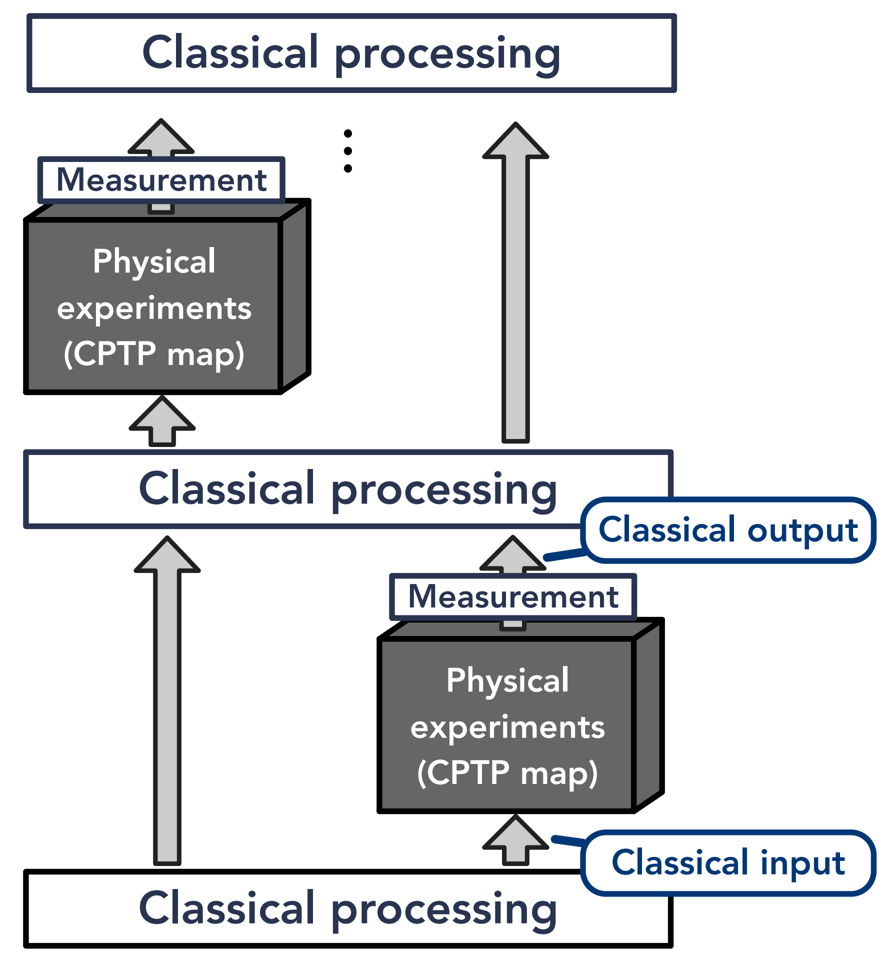
3. Output the function h_C from M_ϵ^p that minimizes

$$\frac{1}{N_C} \sum_{i=1}^{N_C} |h_C(x_i) - o_i|^2.$$

Prediction error $\mathbb{E}_{x \sim \mathcal{D}} \left| h_C(x) - f_{\mathcal{E}}(x) \right|^2 \leq \mathcal{O}(\epsilon)$

using $N_C = \mathcal{O}(\log(|M_\epsilon^p|)/\epsilon)$.

A proper/complicated statistical analysis gives this.



Classical Machine Learning

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

Proof idea:
Combining the two bounds

Quantum lower bound

$$N_Q \geq \Omega(\log(|M_\epsilon^p|)/m).$$

Classical upper bound

$$N_C \leq \mathcal{O}(\log(|M_\epsilon^p|)/\epsilon).$$

$$N_C \leq \mathcal{O}(mN_Q/\epsilon).$$

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

Main Theorem

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.*

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

- Quantum ML can perform better than classical ML when ϵ is small or when m is large.
- This can still be useful in practice!
- But the advantage in query complexity is limited as above in any quantum problem.

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

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

- The Quantum ML setting requires coherent accesses to $\mathscr{E}$ + large quantum memory.

- The Classical ML setting only use fixed measurement after each $\mathscr{E}$ + large classical memory.

- Quantum ML setting may likely only be available far in the future.

- Classical ML setting is just as powerful after getting moderately more data. And is readily available.

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

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

- ML models trained on classical computers are computationally as powerful as those running on quantum computers?
- **No!** We only consider query complexity, not computational complexity.
- We can consider quantum computers running in the classical ML setting (learning only from measurement data stored in classical memory).
- Quantum computers can optimize/compute faster than classical computers! E.g., see [2] for discussion on computational complexity.

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

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

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 query complexity, not computational complexity.

- We can consider quantum computers runnin
(learning only from measurement data store

Complexity class of classical ML algorithms trained on data is strictly bigger than BPP

- Quantum computers can optimize/compute faste an classical computers! E.g., see [2] for discussion on computational complexity.

[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 $N_{\mathrm{C}} \leq \mathcal{O}(mN_{\mathrm{Q}}/\epsilon)$

- Classical ML setting is just as powerful as quantum ML setting after getting moderately more data.
- Quantum computers can optimize/compute ML models faster than classical computers.
- => Near-term quantum devices + classical computers may be able to address challenging quantum problems in physics/chemistry.

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

Exponential advantage

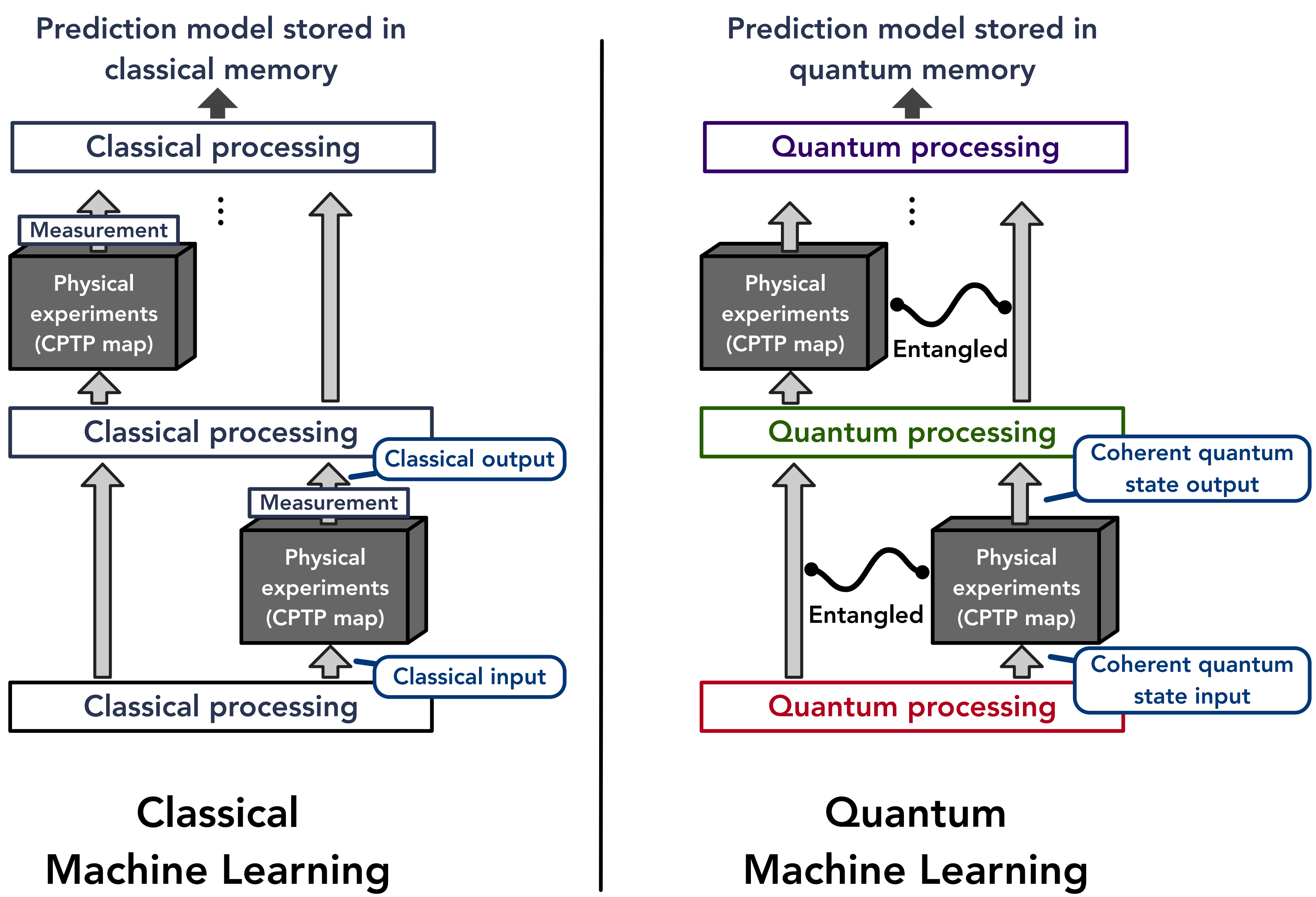
- The theorem holds only for average-case prediction error.
- Other measures of prediction error (e.g., worst-case) admits provable exponential advantage.

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

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

Exponential advantage

- We give an example where the CPTP map takes no input.



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

Exponential advantage

- The physical experiment prepares an unknown quantum system and we want to predict expectation values of Pauli observables on the unknown quantum system.
- The input x describes which Pauli observable we would like to predict.
- The output $f_{\mathscr{E}}(x)$ is the expectation of the Pauli observable on the unknown quantum system.
- Goal: Learn a model $h(x)$ such that $h(x) \approx f_{\mathscr{E}}(x)$.

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

Exponential advantage

What we know so far:

We can always achieve an average prediction error

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

with a classical ML that uses a number of experiments similar to the optimal quantum ML.

But what about:

Can we achieve a worst-case prediction error

$$\max_x \left| h_{\mathrm{C}}(x) - f_{\mathscr{E}}(x) \right|^2 \leq \mathcal{O}(\epsilon)$$

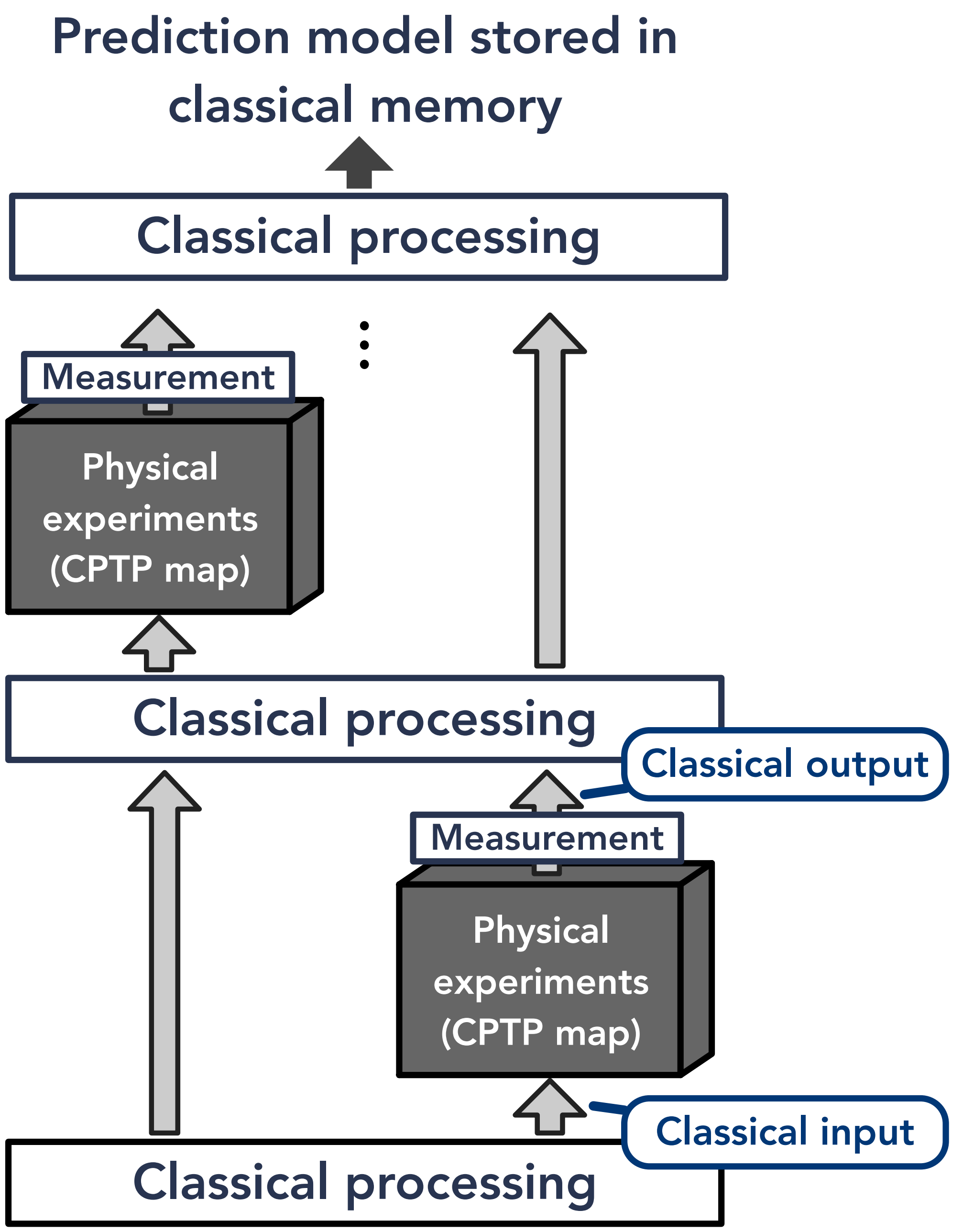
with a classical ML that uses a number of experiments similar to the optimal quantum ML?

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

Exponential advantage

Classical ML setting

- It can perform arbitrary POVM measurement on the physical system (adaptively).
- Then analyze classical measurement data.

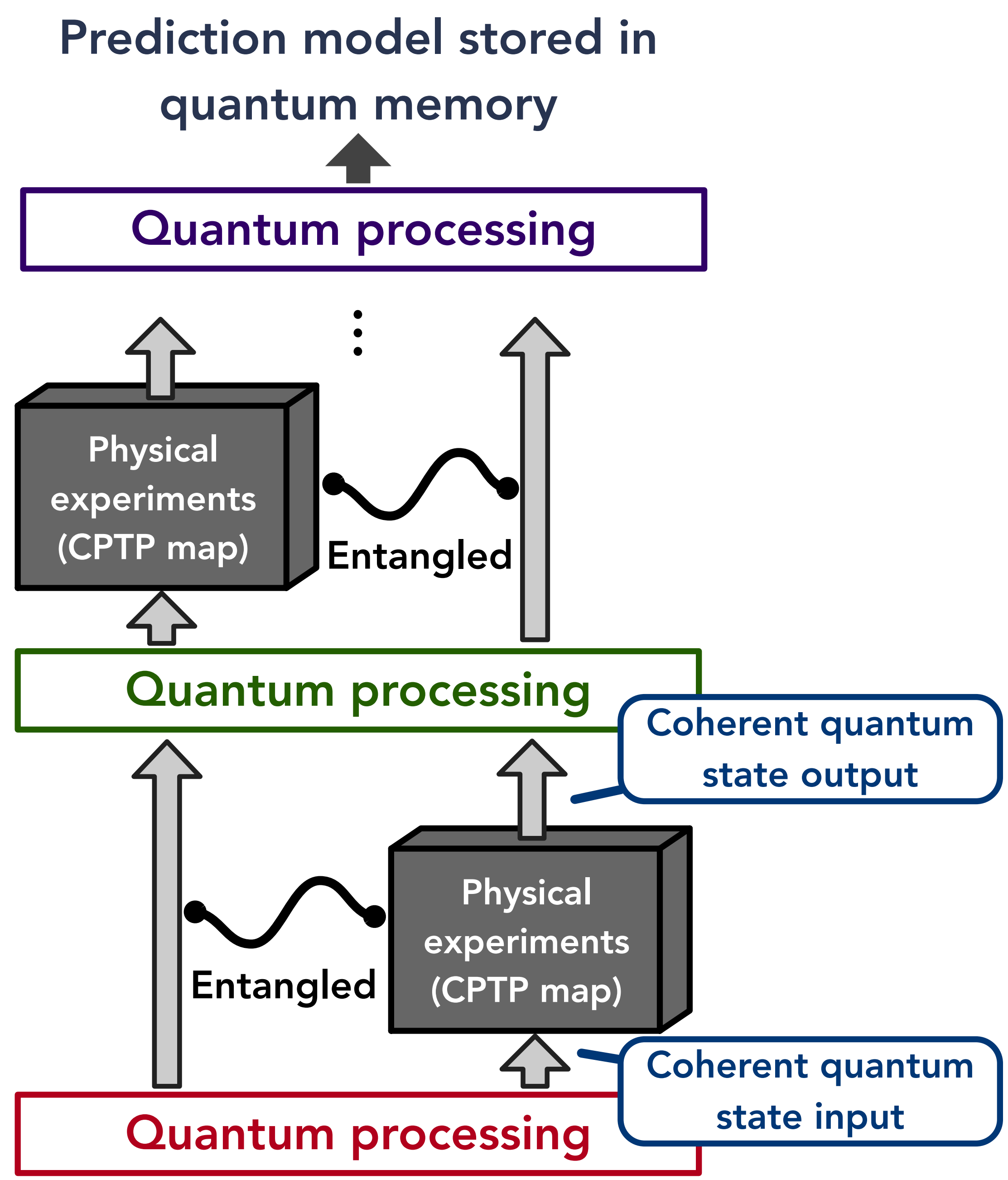


[1] **Information-theoretic bounds on quantum advantage in machine learning**, *arXiv:2101.02464*.

Exponential advantage

Quantum ML setting

- It can store quantum information from each physical experiment coherently in quantum memory.
- Then perform quantum data analysis on the quantum data.



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

Exponential advantage

Where could quantum advantage come from?

Classical ML suffers from uncertainty principle, especially when many observables are highly incompatible.

Quantum ML can store data in quantum memory and access higher-order function of the physical world, e.g., $\rho^{\otimes k}$.

Quantum memory enables the ability to reduce the effect of uncertainty principle [*, **].

[*] Shadow tomography of quantum states.

[**] The uncertainty principle in the presence of quantum memory.

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

Exponential advantage

- The input x describes which Pauli observable we would like to predict.
- The output $f_{\mathscr{E}}(x)$ is the expectation of the Pauli observable on the unknown quantum system.
- **Lower bound**: $\Omega(2^n)$ is necessary to predict all Pauli observables for classical ML (or any conventional experiments).
- **Upper bound**: $\mathcal{O}(n)$ is sufficient to predict all Pauli observables for quantum ML (based on a simple quantum algorithm).

Classical lower bound

- **Lower bound**: $\Omega\left(2^{n/3}\right)$ is necessary to predict all Pauli observables for the classical ML setting (i.e., adaptive single-copy measurement protocols).

- Consider a subset of states of the form $(I+P)/2^n$, where P is a tensor product of Pauli-X/Y/Z observable.

- If we can predict all Pauli observables, then we can discriminate completely mixed state vs one of the above states.

- The informationally maximal POVM is $\{w_i|\psi_i\rangle\langle\psi_i|\}$. A complicated information-theoretic proof shows that because
$$\frac{1}{4^n}\sum_P \langle\psi|P|\psi\rangle^2 = \frac{1}{2^n},$$
(a signature of high incompatibility)
we need at least $\Omega\left(2^{n/3}\right)$ measurements.

Classical lower bound

- **Lower bound**: $\Omega\left(2^{n/3}\right)$ is necessary to predict all Pauli observables for the classical ML setting (i.e., adaptive single-copy measurement protocols).

- The informationally maximal POVM is $\{w_i|\psi_i\rangle\langle\psi_i|\}$. A complicated information-theoretic proof shows that because
$$\frac{1}{4^n}\sum_P \langle\psi|P|\psi\rangle^2 = \frac{1}{2^n}, \text{ (a signature of high incompatibility)}$$
we need at least $\Omega\left(2^{n/3}\right)$ measurements.

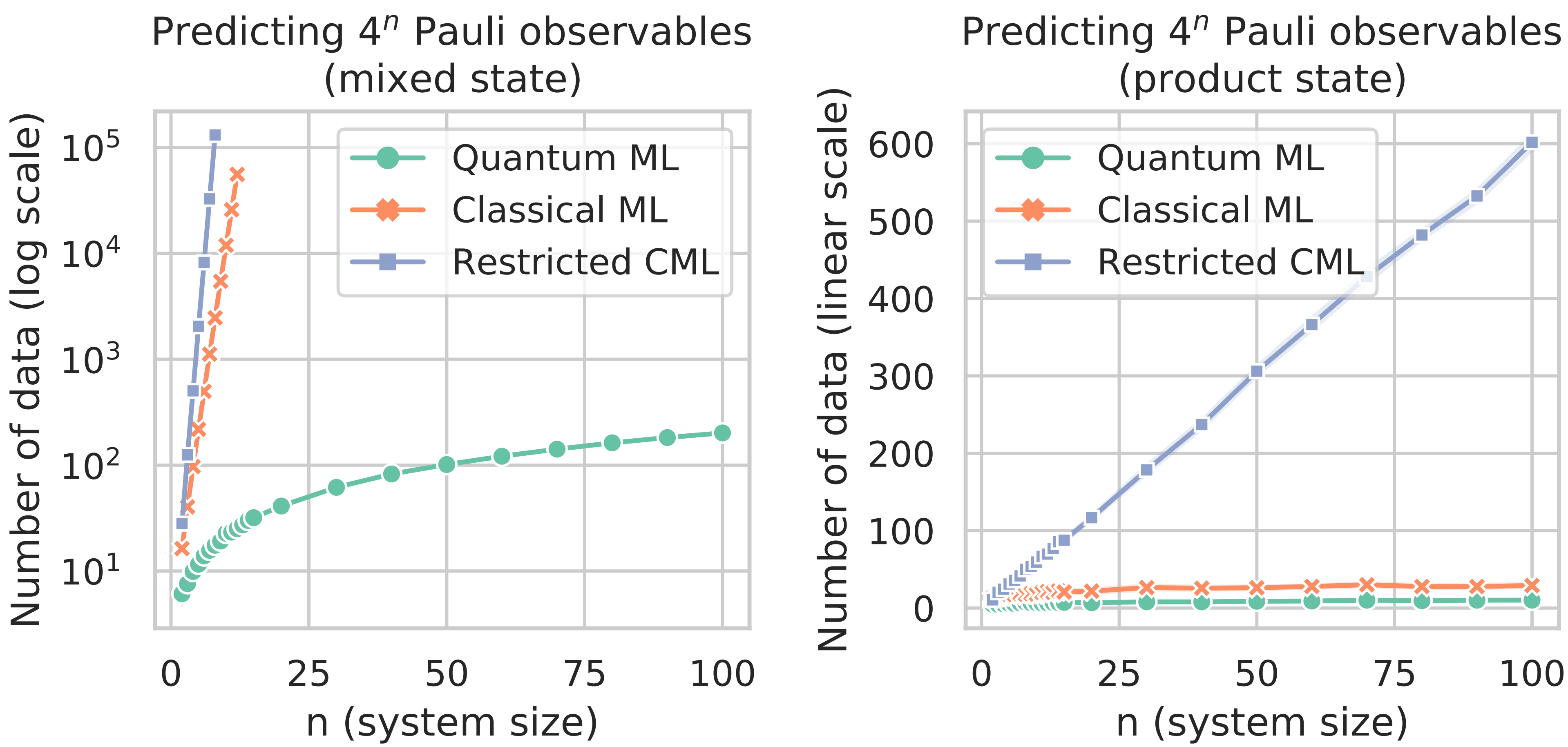
- The same proof can be extended to any set of traceless observables,
$$\frac{1}{N_O}\sum_O \langle\psi|O|\psi\rangle^2 = \delta, \text{ (a signature of high incompatibility)}$$
implies at least $\Omega(\delta^{-1/3})$ measurements to predict the set of observables.

Quantum upper bound

- **Upper bound**: $\mathcal{O}(n)$ is sufficient to predict all Pauli observables for quantum ML (based on a simple quantum algorithm).
- Two level protocol:
- Estimate $|\mathrm{Tr}(P\rho)|^2$ — $P \otimes P$ commutes for all pair of P, so we can simultaneously measure $(P \otimes P)$ on $\rho \otimes \rho$. And note that $\mathrm{Tr}((P \otimes P)(\rho \otimes \rho)) = \mathrm{Tr}(P\rho)^2$.
- Estimate $\mathrm{sign}(\mathrm{Tr}(P\rho))$ — Only consider P with $|\mathrm{Tr}(P\rho)| > \epsilon/2$. Perform coherent majority vote on n copies of ρ. This will not disturb the state much because the outcome happens with very high probability.

Worst-case prediction error

- Numerical experiments that achieve exponential advantage.



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

Conclusion

- A fundamental limit on quantum advantage in data efficiency for achieving average-case prediction error.
- An exponential separation between classical and quantum ML setting for achieving worst-case prediction error.