Information-Theoretic Lower Bounds for QML

How to prove limitations for QML using information theory

“Swiss Army Knife” of QIT Lower Bounding Strategies for QML

“Swiss Army Knife” of QIT Lower Bounding Strategies for QML

- Step 1: Learning requires extracting information

“Swiss Army Knife” of QIT Lower Bounding Strategies for QML

- Step 1: Learning requires extracting information

➢ $\mathtt{QMI}_m \geq \cdots$

"Swiss Army Knife" of QIT Lower Bounding Strategies for QML

- Step 1: Learning requires extracting information

➢ $\mathtt{QMI}_m \geq \cdots$

- Step 2: m examples carry at most m times as much information as a single example

"Swiss Army Knife" of QIT Lower Bounding Strategies for QML

- Step 1: Learning requires extracting information
  - $\mathrm{QMI}_m \geq \cdots$

- Step 2: m examples carry at most m times as much information as a single example
  - $\mathrm{QMI}_m \leq m \cdot \mathrm{QMI}_1$

“Swiss Army Knife” of QIT Lower Bounding Strategies for QML

- Step 1: Learning requires extracting information

➢ $\mathrm{QMI}_m \geq \cdots$

- Step 2: m examples carry at most m times as much information as a single example

➢ $\mathrm{QMI}_m \leq m \cdot \mathrm{QMI}_1$ ⬅ Warning: This step is in general sub-optimal [1]!

[1] S. Bab Hadiashar, A. Nayak, and P. Sinha; *arXiv:2301.02227*

“Swiss Army Knife” of QIT Lower Bounding Strategies for QML

- Step 1: Learning requires extracting information

➢ $\mathrm{QMI}_m \geq \cdots$

- Step 2: m examples carry at most m times as much information as a single example

➢ $\mathrm{QMI}_m \leq m \cdot \mathrm{QMI}_1$

"Swiss Army Knife" of QIT Lower Bounding Strategies for QML

- Step 1: Learning requires extracting information

➢ $\mathrm{QMI}_m \geq \cdots$

- Step 2: m examples carry at most m times as much information as a single example

➢ $\mathrm{QMI}_m \leq m \cdot \mathrm{QMI}_1$

- Step 3: In "hard" learning problems, a single example doesn't carry too much information.

"Swiss Army Knife" of QIT Lower Bounding Strategies for QML

- Step 1: Learning requires extracting information
  - ➢ $\mathrm{QMI}_m \geq \cdots$
- Step 2: m examples carry at most m times as much information as a single example
  - ➢ $\mathrm{QMI}_m \leq m \cdot \mathrm{QMI}_1$
- Step 3: In "hard" learning problems, a single example doesn't carry too much information.
  - ➢ $\mathrm{QMI}_1 \leq \cdots$

Quantum Shadow Tomography [2]

ρ

[2] S. Aaronson; *STOC 2018*

Quantum Shadow Tomography [2]

Unknown!

[2] S. Aaronson; *STOC 2018*

Quantum Shadow Tomography [2]

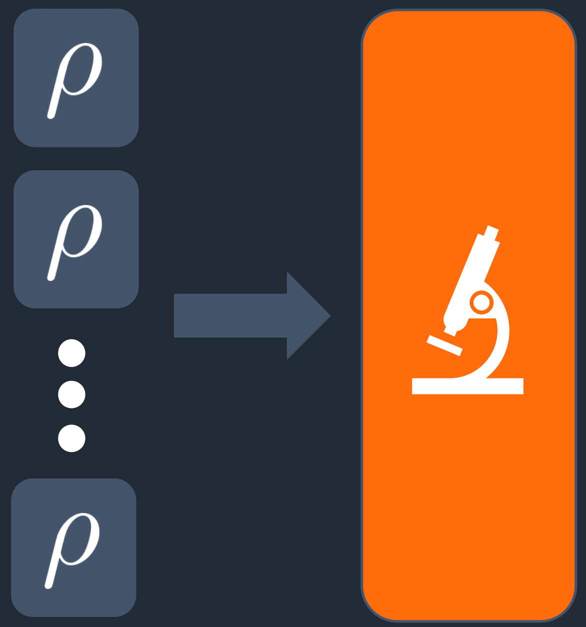
ρ

[2] S. Aaronson; *STOC 2018*

Quantum Shadow Tomography [2]

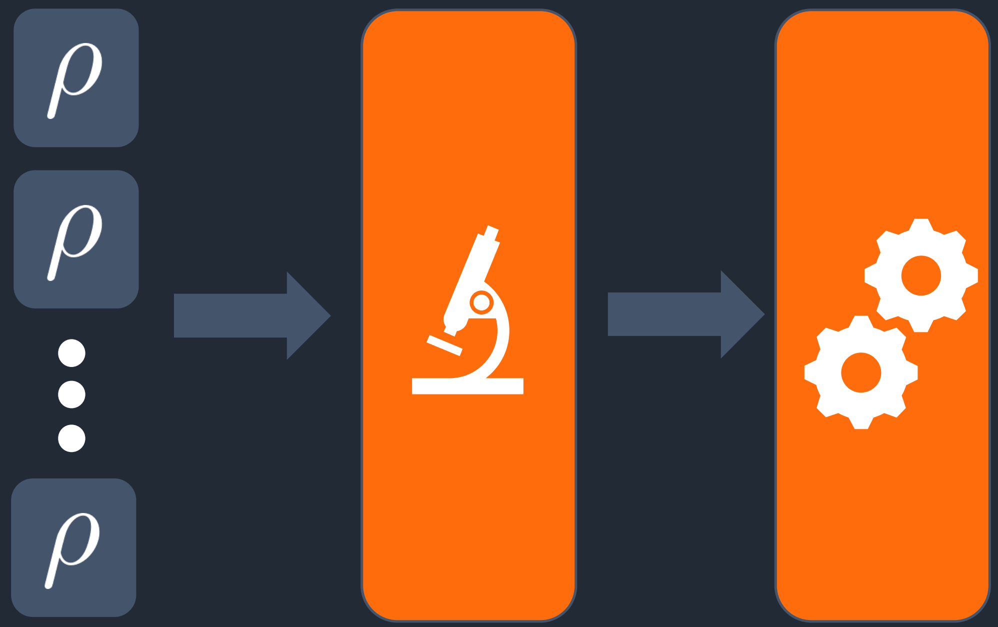
[2] S. Aaronson; *STOC 2018*

Quantum Shadow Tomography [2]



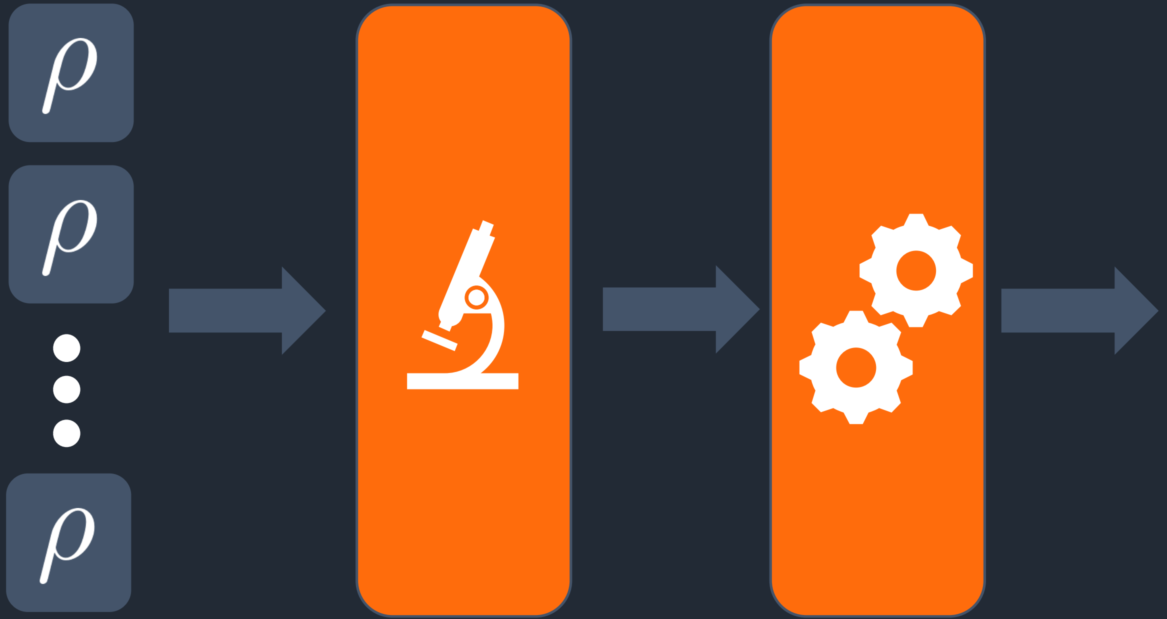
[2] S. Aaronson; *STOC 2018*

Quantum Shadow Tomography [2]



[2] S. Aaronson; *STOC 2018*

Quantum Shadow Tomography [2]

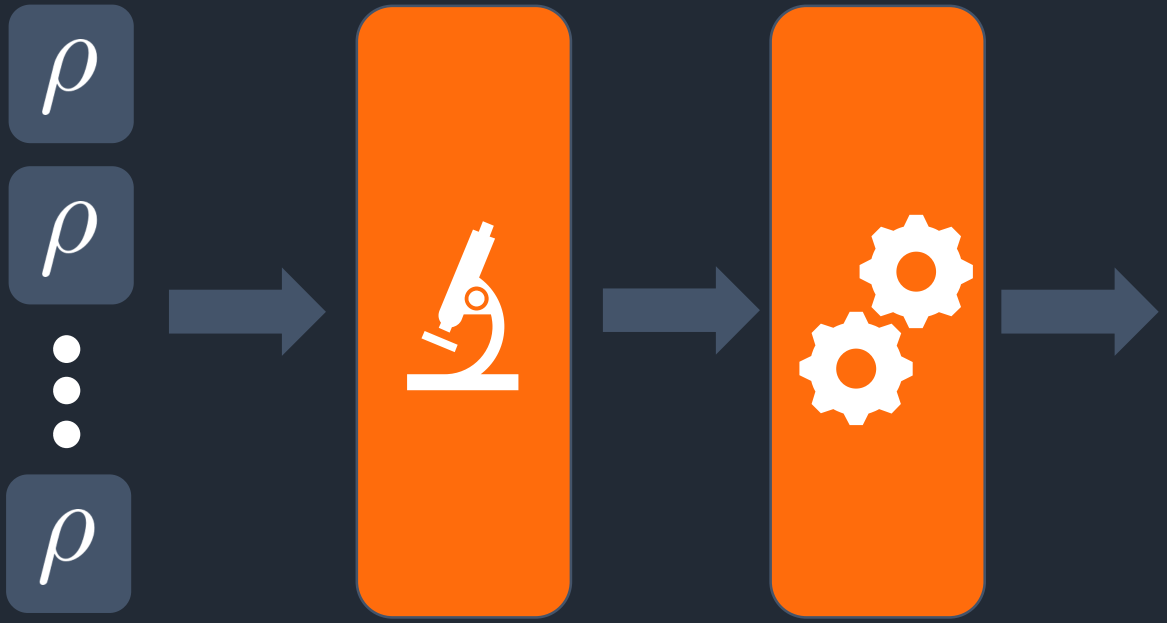


$$\hat{\rho}: \ |\mathrm{tr}[O_i\hat{\rho}] - \mathrm{tr}[O_i\rho]| \leq \varepsilon$$
$$\text{for } 1 \leq i \leq M, \text{ w.h.p.}$$

[2] S. Aaronson; *STOC 2018*

Quantum Shadow Tomography [2]

$\tilde{O}\left(\frac{\log^2(M)\cdot n}{\varepsilon^4}\right)$ copies [3]!



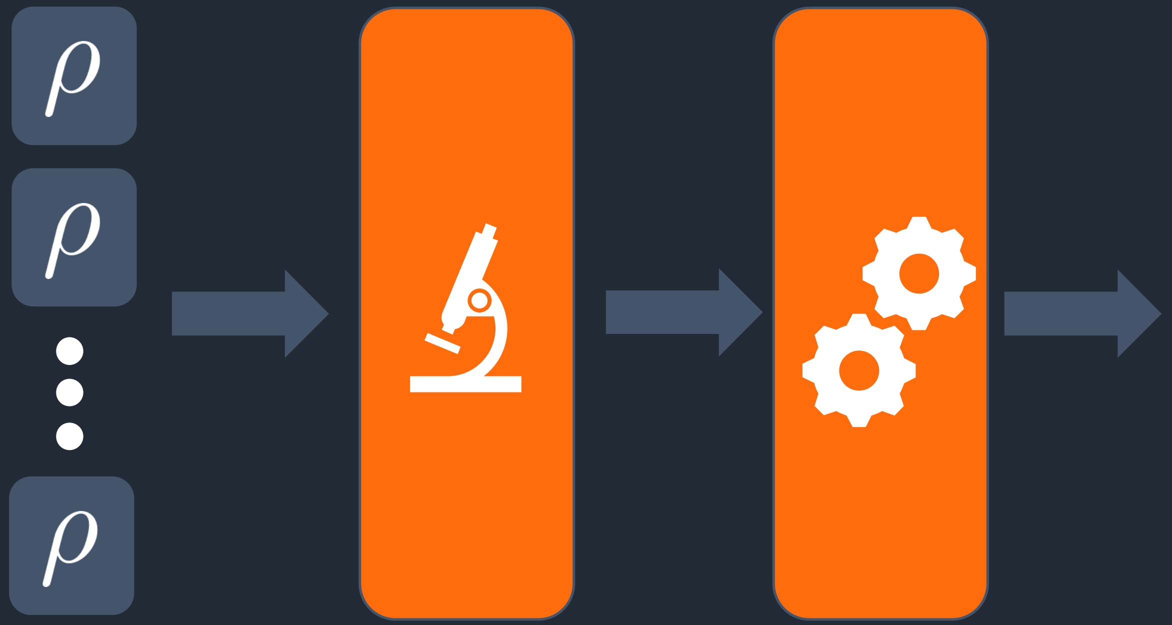
$\hat{\rho}$: $|\mathrm{tr}[O_i\hat{\rho}] - \mathrm{tr}[O_i\rho]| \leq \varepsilon$
for $1 \leq i \leq M$, w.h.p.

[2] S. Aaronson; *STOC 2018*
[3] C. Badescu and R. O'Donnell; *STOC 2021*

Quantum Shadow Tomography [2]

$\tilde{\mathcal{O}}\left(\frac{\log^2(M)\cdot n}{\varepsilon^4}\right)$ copies [3]!

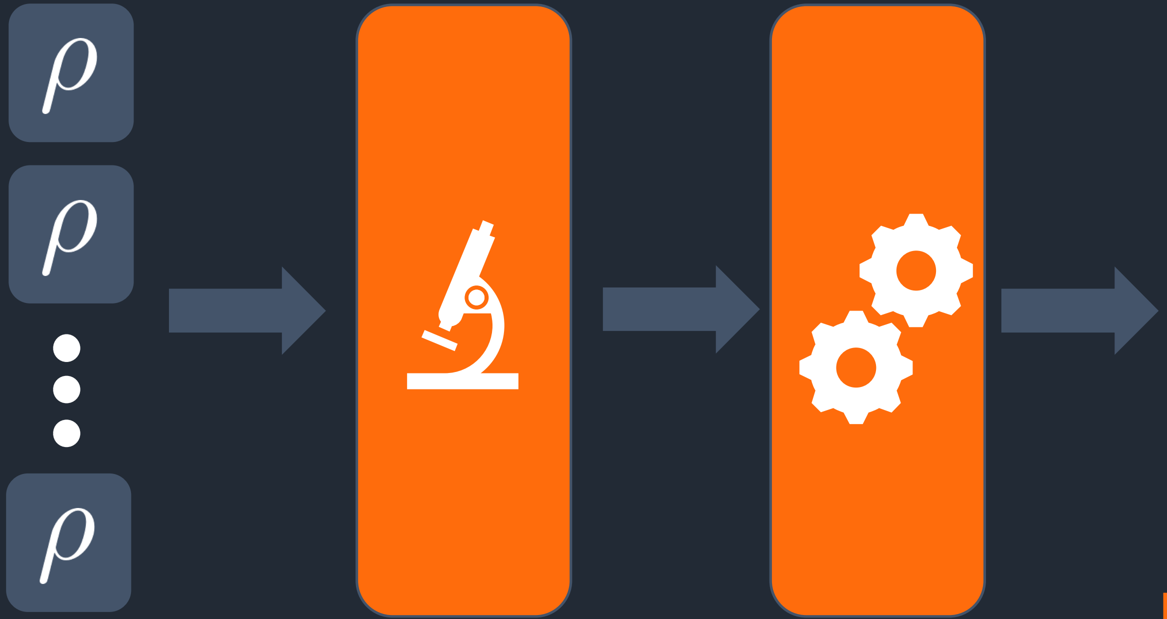
$\Omega\left(\frac{\min\{4^n, \log M\}}{\varepsilon^2}\right)$ copies [2]



$\hat{\rho}$: $|\mathrm{tr}[O_i\hat{\rho}] - \mathrm{tr}[O_i\rho]| \leq \varepsilon$
for $1 \leq i \leq M$, w.h.p.

[2] S. Aaronson; *STOC 2018*
[3] C. Badescu and R. O'Donnell; *STOC 2021*

Quantum Shadow Tomography [2]



$\tilde{\mathcal{O}}\left(\frac{\log^2(M)\cdot \mathrm{n}}{\varepsilon^4}\right)$ copies [3]!

$\Omega\left(\frac{\min\{4^n, \log M\}}{\varepsilon^2}\right)$ copies [2]

$\hat{\rho}$: $|\mathrm{tr}[O_i\hat{\rho}] - \mathrm{tr}[O_i\rho]| \leq \varepsilon$ for $1 \leq i \leq M$, w.h.p.

Restricted channel version [4]

[2] S. Aaronson; *STOC 2018*

[3] C. Badescu and R. O'Donnell; *STOC 2021*

[4] **M.C.C.**; *arXiv:2212.04471*

Quantum Shadow Tomography Lower Bound – Proof

Quantum Shadow Tomography Lower Bound – Proof

Setup:

Quantum Shadow Tomography Lower Bound – Proof

Setup:

- $N = \min\{2^n, \sqrt{\log M}\}$ and $K \sim c^{N^2}$

Quantum Shadow Tomography Lower Bound – Proof

Setup:

- $N = \min\{2^n, \sqrt{\log M}\}$ and $K \sim c^{N^2}$
- By probabilistic argument, $\exists\ (N/2)$-dim. subspace projectors $\Pi_1, \ldots, \Pi_K$ s.t. the states $\sigma_i = (1 - 6\varepsilon)\frac{\mathbb{I}}{N} + 6\varepsilon \cdot \frac{2\Pi_i}{N}$ satisfy $\left|\mathrm{tr}[\Pi_i \sigma_i] - \frac{1}{2}\right| = 3\varepsilon$ and $\left|\mathrm{tr}\left[\Pi_i \sigma_j\right] - \frac{1}{2}\right| \leq \frac{\varepsilon}{2}$ for all $i \neq j$.

Quantum Shadow Tomography Lower Bound – Proof

Setup:

- $N = \min\{2^n, \sqrt{\log M}\}$ and $K \sim c^{N^2}$
- By probabilistic argument, $\exists$ $(N/2)$-dim. subspace projectors $\Pi_1, \ldots, \Pi_K$ s.t. the states $\sigma_i = (1-6\varepsilon)\frac{\mathbb{I}}{N} + 6\varepsilon \cdot \frac{2\Pi_i}{N}$ satisfy $\left|\mathrm{tr}[\Pi_i \sigma_i] - \frac{1}{2}\right| = 3\varepsilon$ and $\left|\mathrm{tr}\left[\Pi_i \sigma_j\right] - \frac{1}{2}\right| \leq \frac{\varepsilon}{2}$ for all $i \neq j$.

➢Shadow tomography with accuracy ε allows to distinguish the σ_i, thus extracts $\geq \log K \sim N^2$ bits of information.

Quantum Shadow Tomography Lower Bound – Proof

- Shadow tomography with accuracy ε allows to distinguish the σ_i, thus extracts $\geq \log K \sim N^2 = \min\{4^n, \log M\}$bits of information.

Quantum Shadow Tomography Lower Bound – Proof

- Shadow tomography with accuracy ε allows to distinguish the σ_i, thus extracts $\geq \log K \sim N^2 = \min\{4^n, \log M\}$bits of information.

“Swiss army knife” in action:

Quantum Shadow Tomography Lower Bound – Proof

➢ Shadow tomography with accuracy ε allows to distinguish the σ_i, thus extracts $\geq \log K \sim N^2 = \min\{4^n, \log M\}$ bits of information.

"Swiss army knife" in action:

- Step 1: $\rho_{CQ} = \mathbb{E}_{i\sim[K]}\left[|i\rangle\langle i| \otimes \sigma_i^{\otimes m}\right]$ has QMI $I(C{:}\,Q)_\rho \gtrsim N^2$.

Quantum Shadow Tomography Lower Bound – Proof

- Shadow tomography with accuracy ε allows to distinguish the σ_i, thus extracts $\geq \log K \sim N^2 = \min\{4^n, \log M\}$ bits of information.

"Swiss army knife" in action:

- Step 1: $\rho_{CQ} = \mathbb{E}_{i\sim[K]}\left[|i\rangle\langle i| \otimes \sigma_i^{\otimes m}\right]$ has QMI $I(C{:}Q)_\rho \gtrsim N^2$.
- Step 2: As ρ_{CQ} is a CQ state, $I(C{:}Q)_\rho \leq \cdots \leq m \cdot \left(\log N - S(\sigma_i)\right)$.

Quantum Shadow Tomography Lower Bound – Proof

- Shadow tomography with accuracy ε allows to distinguish the σ_i, thus extracts $\geq \log K \sim N^2 = \min\{4^n, \log M\}$bits of information.

“Swiss army knife” in action:

- Step 1: $\rho_{CQ} = \mathbb{E}_{i\sim[K]}\left[|i\rangle\langle i| \otimes \sigma_i^{\otimes m}\right]$ has QMI $I(C:Q)_\rho \gtrsim N^2$.
- Step 2: As ρ_{CQ} is a CQ state, $I(C:Q)_\rho \leq \cdots \leq m \cdot \left(\log N - S(\sigma_i)\right)$.
- Step 3: By looking at the spectrum, $S(\sigma_i) \geq \cdots \geq \log N - \mathcal{O}(\varepsilon^2)$.

Quantum Shadow Tomography Lower Bound – Proof

➢Shadow tomography with accuracy ε allows to distinguish the σ_i, thus extracts $\geq \log K \sim N^2 = \min\{4^n, \log M\}$bits of information.

"Swiss army knife" in action:

- Step 1: $\rho_{CQ} = \mathbb{E}_{i\sim[K]}\left[|i\rangle\langle i| \otimes \sigma_i^{\otimes m}\right]$ has QMI $I(C:Q)_\rho \gtrsim N^2$.
- Step 2: As ρ_{CQ} is a CQ state, $I(C:Q)_\rho \leq \cdots \leq m \cdot \left(\log N - S(\sigma_i)\right)$.
- Step 3: By looking at the spectrum, $S(\sigma_i) \geq \cdots \geq \log N - \mathcal{O}(\varepsilon^2)$.
- Combining everything gives $m \gtrsim \Omega\left(\frac{N^2}{\varepsilon^2}\right) = \Omega\left(\frac{\min\{4^n, \log M\}}{\varepsilon^2}\right)$. □

Quantum Computational Learning Theory

Quantum Computational Learning Theory

$|\psi_f\rangle$

$|\psi_f\rangle$

$\vdots$

$|\psi_f\rangle$

Quantum Computational Learning Theory

$|\psi_f\rangle$

$|\psi_f\rangle$

⋮

$|\psi_f\rangle$

$$|\psi_f\rangle = \sum_{x\in\{0,1\}^n} \sqrt{P(x)}\,|x, f(x)\rangle$$ [5]

[5] N.H. Bshouty and J.C. Jackson; *SIAM Journal on Computing 28.3, 1136-1153 (1998)*

Quantum Computational Learning Theory

Unknown!

$|\psi_f\rangle$

$|\psi_f\rangle$

⋮

$|\psi_f\rangle$

$$|\psi_f\rangle = \sum_{x\in\{0,1\}^n} \sqrt{P(x)}\,|x, f(x)\rangle \ [5]$$

[5] N.H. Bshouty and J.C. Jackson; *SIAM Journal on Computing 28.3, 1136-1153 (1998)*

Quantum Computational Learning Theory

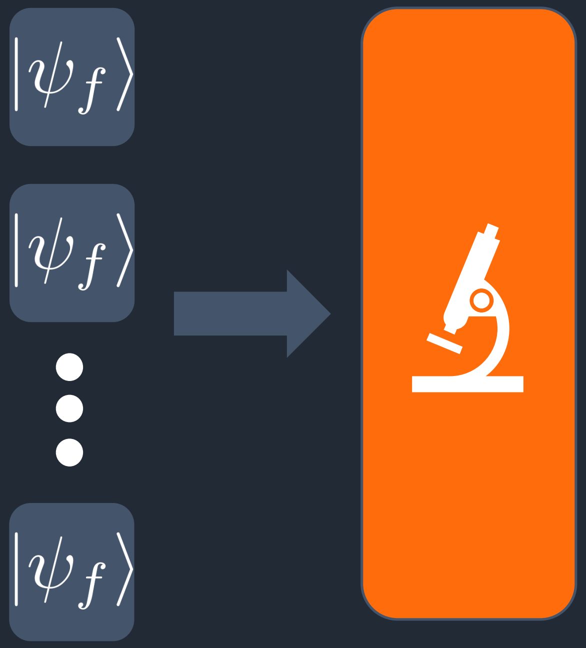
$|\psi_f\rangle$

$|\psi_f\rangle$

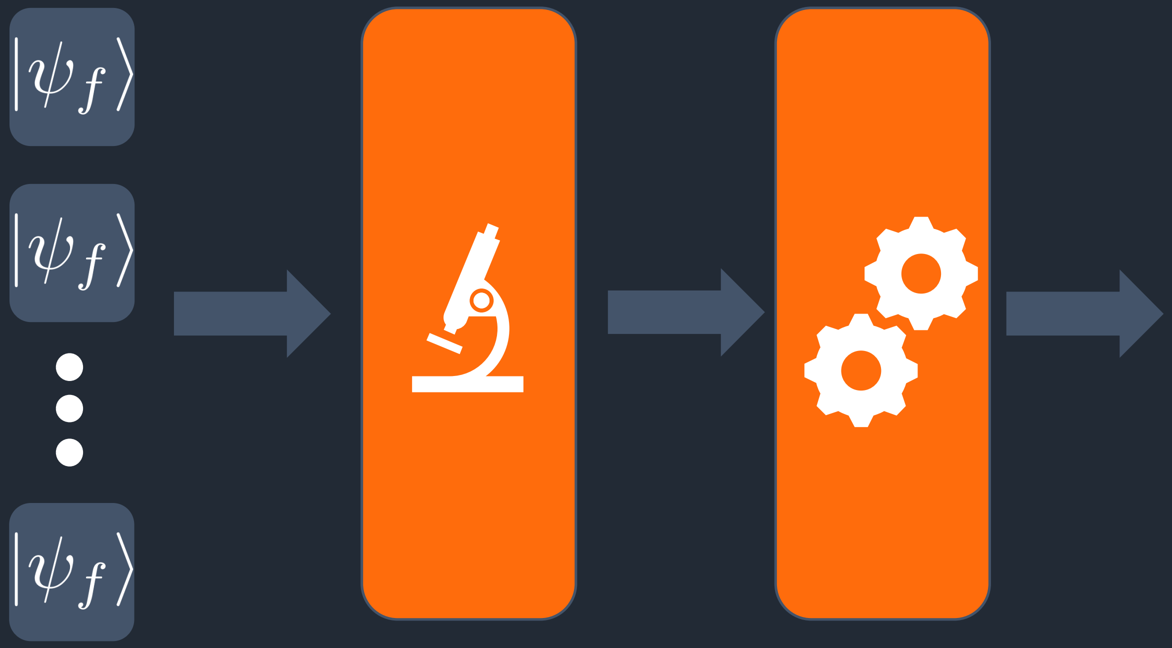
$\vdots$

$|\psi_f\rangle$

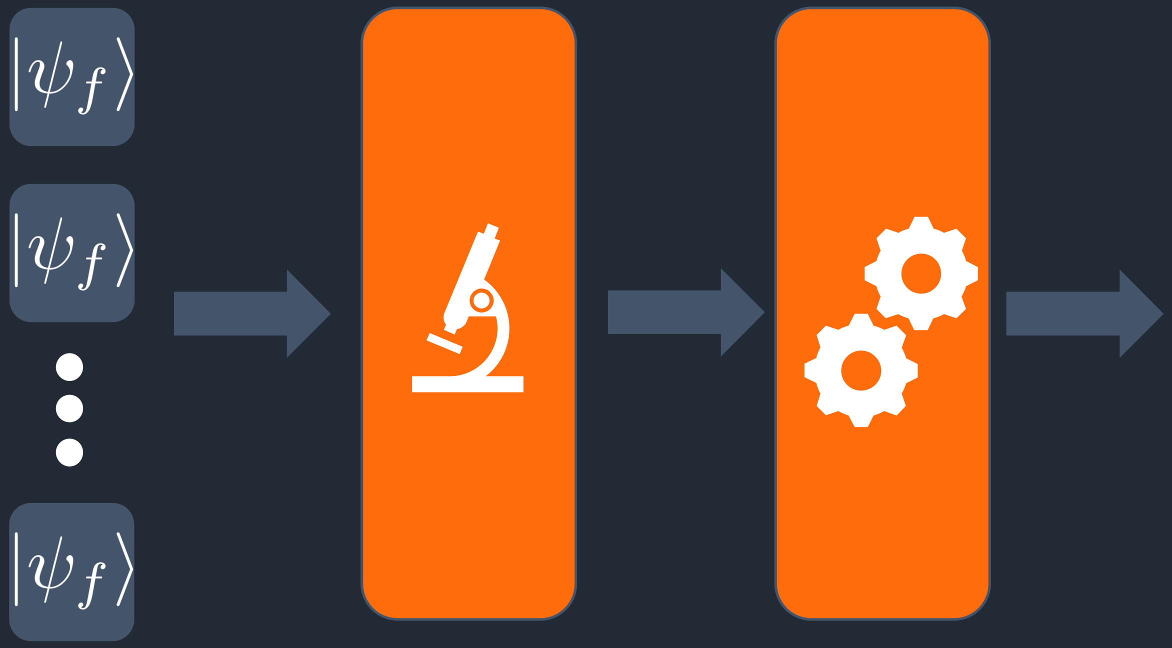
Quantum Computational Learning Theory



Quantum Computational Learning Theory



Quantum Computational Learning Theory



$\hat{f}\colon\ \mathbb{P}_{x\sim P}[\hat{f}(x) \neq f(x)] \leq \varepsilon$
w.h.p.

Quantum Computational Learning Theory

Quantum Computational Learning Theory

- [6] used the "swiss army knife" to show a quantum sample complexity lower bound for distribution-independent PAC learning that matches classical upper bounds up to log factors.

[6] S. Arunachalam and R. de Wolf; *ACM Sigact News 48.2 (2017): 41-67*

Quantum Computational Learning Theory

- [6] used the "swiss army knife" to show a quantum sample complexity lower bound for distribution-independent PAC learning that matches classical upper bounds up to log factors.
- [1] improved the "swiss army knife" to get rid of the log factors.

[6] S. Arunachalam and R. de Wolf; *ACM Sigact News 48.2 (2017): 41-67*
[1] S. Bab Hadiashar, A. Nayak, and P. Sinha; *arXiv:2301.02227*

Quantum Computational Learning Theory

- [6] used the "swiss army knife" to show a quantum sample complexity lower bound for distribution-independent PAC learning that matches classical upper bounds up to log factors.
- [1] improved the "swiss army knife" to get rid of the log factors.
  - [6] also got rid of the log factors, albeit with a completely different proof.

[6] S. Arunachalam and R. de Wolf; *ACM Sigact News 48.2 (2017): 41-67*
[1] S. Bab Hadiashar, A. Nayak, and P. Sinha; *arXiv:2301.02227*

More Variations of “Swiss Army Knife”

More Variations of "Swiss Army Knife"

- Quantum state tomography [7,8]

[7] J. Haah et al., *STOC 2016*
[8] R. O'Donnell and J. Wright; *STOC 2016*

More Variations of "Swiss Army Knife"

- Quantum state tomography [7,8]
  - Extra ingredients: Communication protocol
    Fano's inequality [9]
    Packing nets
    Holevo's bound [10]

[7] J. Haah et al., *STOC 2016*
[8] R. O'Donnell and J. Wright; *STOC 2016*
[9] R.M. Fano; *MIT Press (1961)*
[10] A.S. Holevo; *Problemy Peredachi Informatsii 9.3 (1973): 3-11*

More Variations of “Swiss Army Knife”

- Quantum state tomography [7,8]
  - Extra ingredients: Communication protocol
    Fano‘s inequality [9]
    Packing nets
    Holevo‘s bound [10]
- Sample complexity lower bounds for classical shadows [11]

[7] J. Haah et al., *STOC 2016*
[8] R. O‘Donnell and J. Wright; *STOC 2016*
[11] H.-Y. Huang et al.; *Nat. Phys. 16, 1050–1057 (2020)*
[9] R.M. Fano; *MIT Press (1961)*
[10] A.S. Holevo; *Problemy Peredachi Informatsii 9.3 (1973): 3-11*

More Variations of "Swiss Army Knife"

- Quantum state tomography [7,8]
  - Extra ingredients: Communication protocol
    Fano's inequality [9]
    Packing nets
    Holevo's bound [10]
- Sample complexity lower bounds for classical shadows [11]
  - Extra ingredients: Communication protocol with adversary
    Fano's inequality [9]
    Holevo's bound [10]

[7] J. Haah et al., *STOC 2016*
[8] R. O'Donnell and J. Wright; *STOC 2016*
[11] H.-Y. Huang et al.; *Nat. Phys. 16, 1050–1057 (2020)*
[9] R.M. Fano; *MIT Press (1961)*
[10] A.S. Holevo; *Problemy Peredachi Informatsii 9.3 (1973): 3-11*

More Variations of “Swiss Army Knife”

- Quantum state tomography [7,8]
  - Extra ingredients: Communication protocol
    Fano‘s inequality [9]
    Packing nets
    Holevo‘s bound [10]
- Sample complexity lower bounds for classical shadows [11]
  - Extra ingredients: Communication protocol with adversary
    Fano‘s inequality [9]
    Holevo‘s bound [10]
- ...

[7] J. Haah et al., *STOC 2016*
[8] R. O‘Donnell and J. Wright; *STOC 2016*
[11] H.-Y. Huang et al.; *Nat. Phys. 16, 1050–1057 (2020)*
[9] R.M. Fano; *MIT Press (1961)*
[10] A.S. Holevo; *Problemy Peredachi Informatsii 9.3 (1973): 3-11*

Outlook: Learning With Restricted Quantum Capabilities

Outlook: Learning With Restricted Quantum Capabilities

- Shadow tomography lower bound $\Omega\left(\frac{\min\{4^n, \log M\}}{\varepsilon^2}\right)$ applies to arbitrary learning procedures (with entangled measurements).

Outlook: Learning With Restricted Quantum Capabilities

- Shadow tomography lower bound $\Omega\left(\frac{\min\{4^n, \log M\}}{\varepsilon^2}\right)$ applies to arbitrary learning procedures (with entangled measurements).

➢ What about more restricted procedures?

Outlook: Learning With Restricted Quantum Capabilities

- Shadow tomography lower bound $\Omega\left(\frac{\min\{4^n, \log M\}}{\varepsilon^2}\right)$ applies to arbitrary learning procedures (with entangled measurements).

➢ What about more restricted procedures?

- Significant recent progress on information-theoretic lower bounds for learners with restricted quantum capabilities [12,13,14,15,...]

[12] H.-Y. Huang, R. Kueng, and J. Preskill; *Phys. Rev. Lett. 126, 190505 (2021)*
[13] D. Aharonov, J. Cotler, and X.-L. Qi; *Nat Commun 13, 887 (2022)*
[14] S. Chen, J. Cotler, H.-Y. Huang, and J. Li; *FOCS 2021*
[15] **M.C.C.**; *arXiv:2212.04471*