# Classical Verification of Quantum Learning

Matthias C. Caro

Freie Universität Berlin

October 17, 2023 – IPAM Workshop “Mathematical Aspects of Quantum Learning”

Based on arXiv:2306.04843 + ongoing work

# My collaborators

Marcel Hinsche

Marios Ioannou

Alexander Nietner

Ryan Sweke

# My collaborators


Marcel Hinsche


Marios Ioannou


Alexander Nietner


Ryan Sweke


Jens Eisert

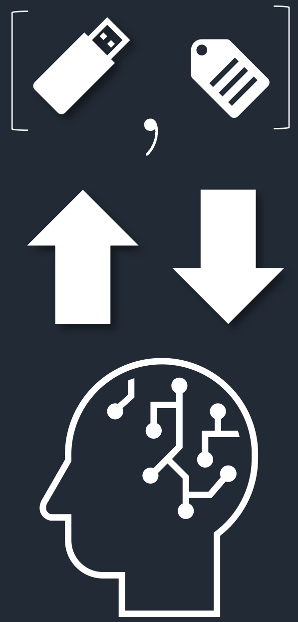
# Motivation

What this talk is about

# Verifying Classical Learning [1]

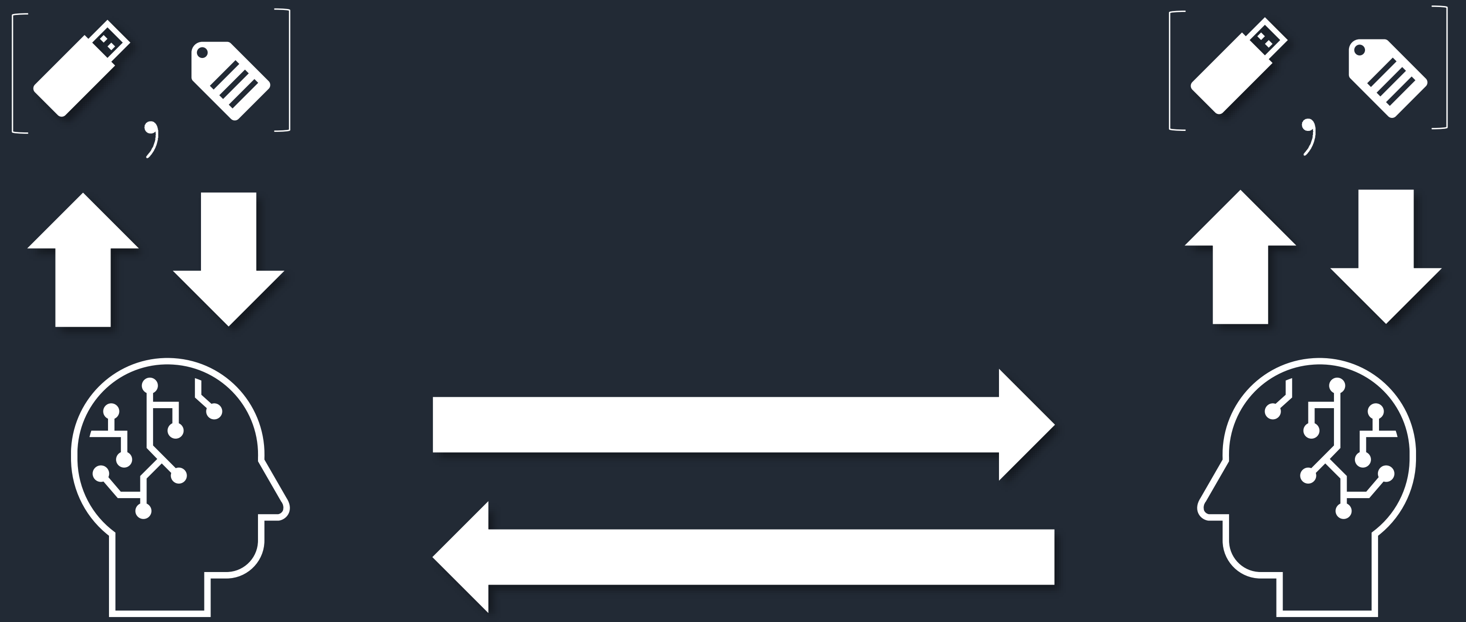
[1] S. Goldwasser et al.; *ITCS 2021*

# Verifying Classical Learning [1]



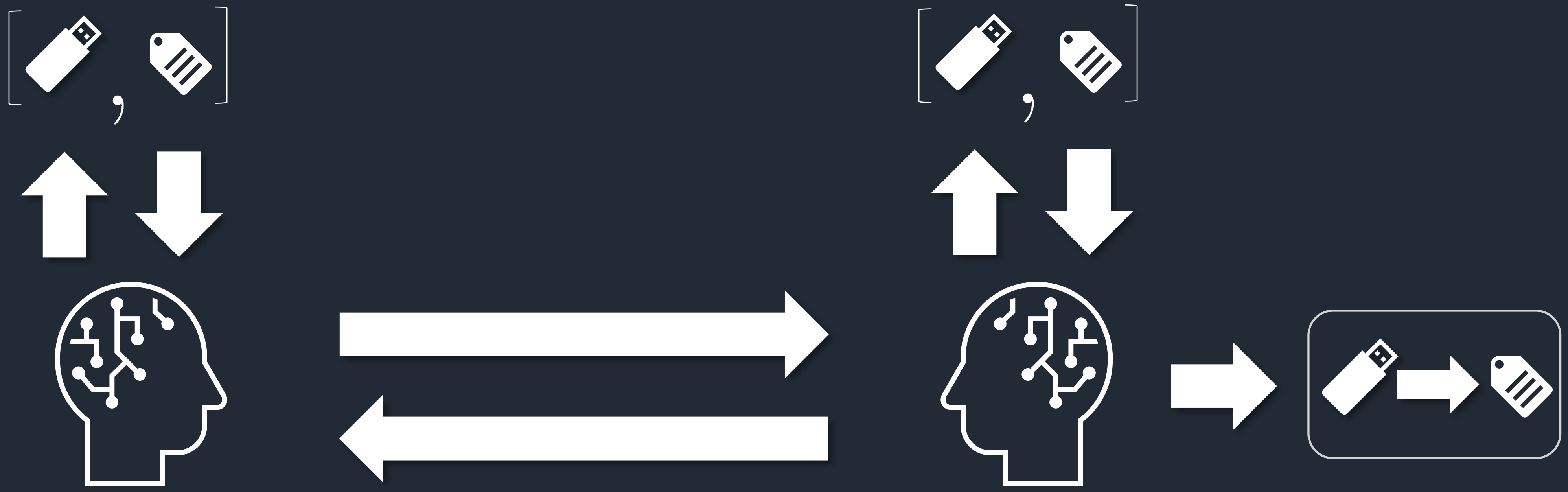
[1] S. Goldwasser et al.; *ITCS 2021*

# Verifying Classical Learning [1]



[1] S. Goldwasser et al.; *ITCS 2021*

# Verifying Classical Learning [1]

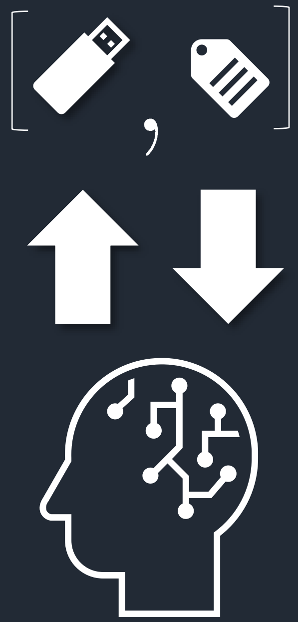



[1] S. Goldwasser et al.; *ITCS 2021*

# Verifying Quantum Learning [2]

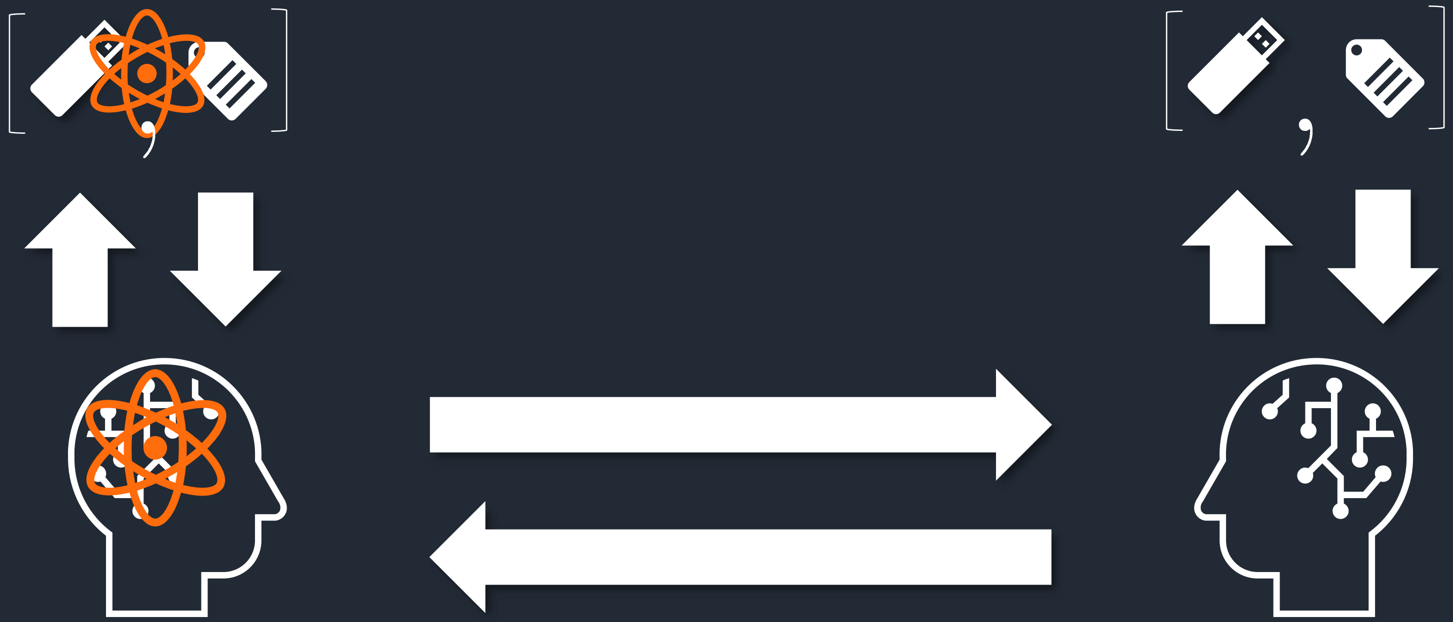
[2] **M.C.C.**, M. Hinsche, M. Ioannou, A. Nietner, and R. Sweke; *arXiv:2306.04843*

# Verifying Quantum Learning [2]



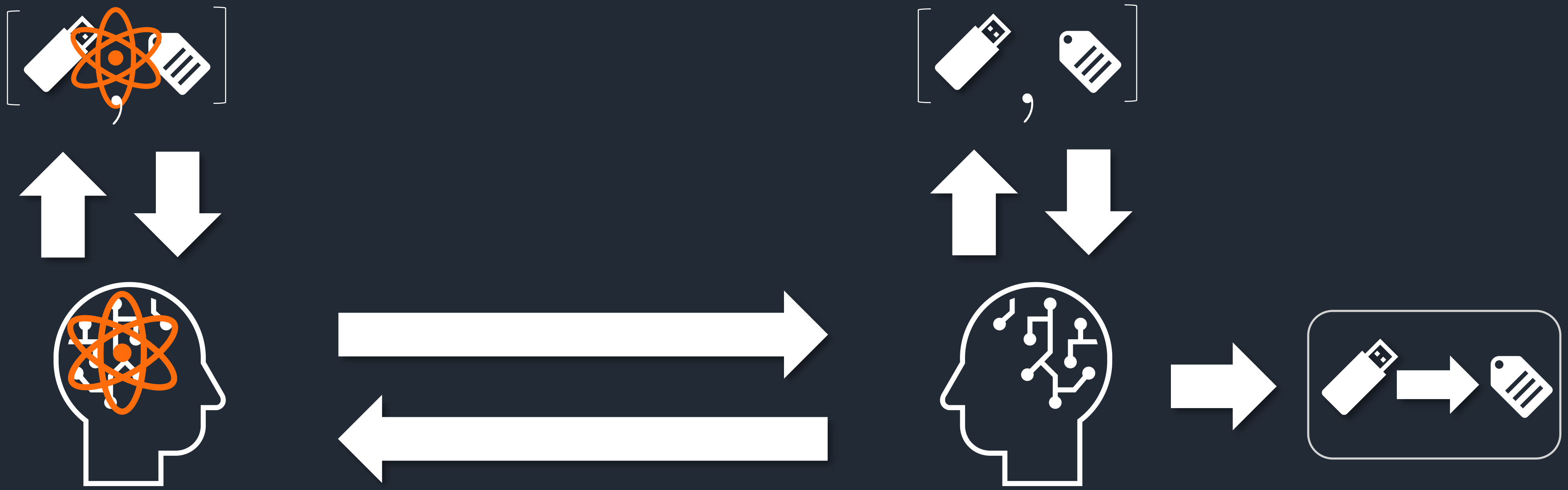
[2] **M.C.C.**, M. Hinsche, M. Ioannou, A. Nietner, and R. Sweke; *arXiv:2306.04843*

# Verifying Quantum Learning [2]



[2] **M.C.C.**, M. Hinsche, M. Ioannou, A. Nietner, and R. Sweke; *arXiv:2306.04843*

# Verifying Quantum Learning [2]



[2] **M.C.C.**, M. Hinsche, M. Ioannou, A. Nietner, and R. Sweke; *arXiv:2306.04843*

# Guiding Question

**Question:**

Is there an (agnostic) learning problem that simultaneously...

1. ...is intractable for classical learners but efficiently solvable by quantum learners using quantum data,
2. ...and can be efficiently verifiably delegated from a classical verifier to an untrusted quantum prover?

# Outline

# Quantum Computational Learning Theory – An Intro

How to learn classical hypotheses from quantum data

# Classical Learning Theory Crash Course

# Classical Learning Theory Crash Course

Ingredients for $\alpha$-agnostic learning:

# Classical Learning Theory Crash Course

Ingredients for $\alpha$-agnostic learning:

- Benchmark class $\mathcal{B}$

# Classical Learning Theory Crash Course

Ingredients for $\alpha$-agnostic learning:

- Benchmark class $\mathcal{B}$
- Model class $\mathcal{M}$

# Classical Learning Theory Crash Course

Ingredients for $\alpha$-agnostic learning:

- Benchmark class $\mathcal{B}$
- Model class $\mathcal{M}$
- Distribution class $\mathfrak{D}$ over $\mathcal{X}_n \times \{0,1\} = \{0,1\}^n \times \{0,1\}$

# Classical Learning Theory Crash Course

Ingredients for $\alpha$-agnostic learning:

- Benchmark class $\mathcal{B}$
- Model class $\mathcal{M}$
- Distribution class $\mathfrak{D}$ over $\mathcal{X}_n \times \{0,1\} = \{0,1\}^n \times \{0,1\}$
- Oracle $\mathsf{O}$

# Classical Learning Theory Crash Course

Ingredients for $\alpha$-agnostic learning:

- Benchmark class $\mathcal{B}$
- Model class $\mathcal{M}$
- Distribution class $\mathfrak{D}$ over $\mathcal{X}_n \times \{0,1\} = \{0,1\}^n \times \{0,1\}$
- Oracle $\mathsf{O}$

# Classical Learning Theory Crash Course

Ingredients for $\alpha$-agnostic learning:

- Benchmark class $\mathcal{B}$
- Model class $\mathcal{M}$
- Distribution class $\mathfrak{D}$ over $\mathcal{X}_n \times \{0,1\} = \{0,1\}^n \times \{0,1\}$
- Oracle $\mathsf{O}$

**Definition:**

# Classical Learning Theory Crash Course

Ingredients for $\alpha$-agnostic learning:

- Benchmark class $\mathcal{B}$
- Model class $\mathcal{M}$
- Distribution class $\mathfrak{D}$ over $\mathcal{X}_n \times \{0,1\} = \{0,1\}^n \times \{0,1\}$
- Oracle $\mathsf{O}$

**Definition:**

$\mathcal{A}$ is an $\alpha$-agnostic learner for $\mathcal{B}$ via $\mathcal{M}$ w.r.t. $\mathfrak{D}$ from $\mathsf{O}$-access if:

# Classical Learning Theory Crash Course

Ingredients for $\alpha$-agnostic learning:

- Benchmark class $\mathcal{B}$
- Model class $\mathcal{M}$
- Distribution class $\mathfrak{D}$ over $\mathcal{X}_n \times \{0,1\} = \{0,1\}^n \times \{0,1\}$
- Oracle $\mathsf{O}$

**Definition:**

$\mathcal{A}$ is an $\alpha$-agnostic learner for $\mathcal{B}$ via $\mathcal{M}$ w.r.t. $\mathfrak{D}$ from $\mathsf{O}$-access if:

For any $\mathcal{D} \in \mathfrak{D}$, given access to $\mathsf{O}(\mathcal{D})$ and $\varepsilon, \delta$: With prob. $\geq 1 - \delta$, $\mathcal{A}$ outputs $m \in \mathcal{M}$ s.t.

# Classical Learning Theory Crash Course

Ingredients for $\alpha$-agnostic learning:

- Benchmark class $\mathcal{B}$
- Model class $\mathcal{M}$
- Distribution class $\mathfrak{D}$ over $\mathcal{X}_n \times \{0,1\} = \{0,1\}^n \times \{0,1\}$
- Oracle $\mathsf{O}$

**Definition:**

$\mathcal{A}$ is an $\alpha$-agnostic learner for $\mathcal{B}$ via $\mathcal{M}$ w.r.t. $\mathfrak{D}$ from $\mathsf{O}$-access if:

For any $\mathcal{D} \in \mathfrak{D}$, given access to $\mathsf{O}(\mathcal{D})$ and $\varepsilon, \delta$: With prob. $\geq 1-\delta$, $\mathcal{A}$ outputs $m \in \mathcal{M}$ s.t.

$$\mathbb{P}_{(x,y)\sim\mathcal{D}}[m(x) \neq y] \leq \alpha \cdot \min_{b\in\mathcal{B}} \mathbb{P}_{(x,y)\sim\mathcal{D}}[b(x) \neq y] + \varepsilon.$$

# Classical Learning Theory Crash Course

# Classical Learning Theory Crash Course

- If $\mathcal{A}$ uses $\mathcal{O}\left(\text{poly}\left(n, \frac{1}{\varepsilon}, \frac{1}{\delta}\right)\right)$ oracle calls and time, it is called sample- and computationally efficient, respectively.

# Classical Learning Theory Crash Course

- If $\mathcal{A}$ uses $\mathcal{O}\left(\text{poly}\left(n, \frac{1}{\varepsilon}, \frac{1}{\delta}\right)\right)$ oracle calls and time, it is called sample- and computationally efficient, respectively.
- For this talk:

# Classical Learning Theory Crash Course

- If $\mathcal{A}$ uses $\mathcal{O}\left(\mathsf{poly}\left(n, \frac{1}{\varepsilon}, \frac{1}{\delta}\right)\right)$ oracle calls and time, it is called sample- and computationally efficient, respectively.
- For this talk:
    - Focus on example oracles

# Classical Learning Theory Crash Course

- If $\mathcal{A}$ uses $\mathcal{O}\left(\mathbf{poly}\left(n, \frac{1}{\varepsilon}, \frac{1}{\delta}\right)\right)$ oracle calls and time, it is called sample- and computationally efficient, respectively.
- For this talk:
  - Focus on example oracles
  - Uniform input marginal: $\mathcal{D}_{\mathcal{X}_n} = \mathcal{U}_n$

# Classical Learning Theory Crash Course

- If $\mathcal{A}$ uses $\mathcal{O}\left(\mathbf{poly}\left(n, \frac{1}{\varepsilon}, \frac{1}{\delta}\right)\right)$ oracle calls and time, it is called sample- and computationally efficient, respectively.
- For this talk:
  - Focus on example oracles
  - Uniform input marginal: $\mathcal{D}_{\mathcal{X}_n} = \mathcal{U}_n$
  - Functional case vs Distributional case

# Classical Learning Theory Crash Course

- If $\mathcal{A}$ uses $\mathcal{O}\left(\mathbf{poly}\left(n, \frac{1}{\varepsilon}, \frac{1}{\delta}\right)\right)$ oracle calls and time, it is called sample- and computationally efficient, respectively.
- For this talk:
  - Focus on example oracles
  - Uniform input marginal: $\mathcal{D}_{\mathcal{X}_n} = \mathcal{U}_n$
  - Functional case vs Distributional case

$\forall \mathcal{D} \in \mathfrak{D}: \ \mathcal{D} = (\mathcal{D}_{\mathcal{X}_n}, f)$
for some Boolean $f : \mathcal{X}_n \to \{0, 1\}$

# Classical Learning Theory Crash Course

- If $\mathcal{A}$ uses $\mathcal{O}\left(\textbf{poly}\left(n, \frac{1}{\varepsilon}, \frac{1}{\delta}\right)\right)$ oracle calls and time, it is called sample- and computationally efficient, respectively.
- For this talk:
  - Focus on example oracles
  - Uniform input marginal: $\mathcal{D}_{\mathcal{X}_n} = \mathcal{U}_n$
  - Functional case vs Distributional case

$\forall \mathcal{D} \in \mathfrak{D}: \ \mathcal{D} = (\mathcal{D}_{\mathcal{X}_n}, f)$
for some Boolean $f : \mathcal{X}_n \to \{0, 1\}$

Realizable special case: $f$ always in $\mathcal{B}$

# Classical Learning Theory Crash Course

- If $\mathcal{A}$ uses $\mathcal{O}\left(\text{poly}\left(n, \frac{1}{\varepsilon}, \frac{1}{\delta}\right)\right)$ oracle calls and time, it is called sample- and computationally efficient, respectively.
- For this talk:
  - Focus on example oracles
  - Uniform input marginal: $\mathcal{D}_{\mathcal{X}_n} = \mathcal{U}_n$
  - Functional case vs Distributional case

$\forall \mathcal{D} \in \mathfrak{D} : \; \mathcal{D} = (\mathcal{D}_{\mathcal{X}_n}, f)$
for some Boolean $f : \mathcal{X}_n \to \{0, 1\}$

no extra assumption

Realizable special case: $f$ always in $\mathcal{B}$

# Boolean Fourier Analysis Crash Course

# Boolean Fourier Analysis Crash Course

# Boolean Fourier Analysis Crash Course

**Definition:**

The Fourier transform of $\phi: \mathcal{X}_n \to \mathbb{R}$ is defined as

# Boolean Fourier Analysis Crash Course

**Definition:**

The Fourier transform of $\phi: \mathcal{X}_n \to \mathbb{R}$ is defined as

$$\hat{\phi}: \mathcal{X}_n \to \mathbb{R},\ \hat{\phi}(s) = \mathbb{E}_{x \sim \mathcal{U}_n}[\phi(x)(-1)^{s \cdot x}].$$

# Boolean Fourier Analysis Crash Course

**Definition:**

The Fourier transform of $\phi: \mathcal{X}_n \to \mathbb{R}$ is defined as

$$\hat{\phi}: \mathcal{X}_n \to \mathbb{R},\ \hat{\phi}(s) = \mathbb{E}_{x\sim\mathcal{U}_n}[\phi(x)(-1)^{s\cdot x}].$$

Connecting learning and Fourier analysis:

# Boolean Fourier Analysis Crash Course

**Definition:**

The Fourier transform of $\phi: \mathcal{X}_n \to \mathbb{R}$ is defined as

$$\hat{\phi}: \mathcal{X}_n \to \mathbb{R},\ \hat{\phi}(s) = \mathbb{E}_{x \sim \mathcal{U}_n}[\phi(x)(-1)^{s \cdot x}].$$

Connecting learning and Fourier analysis:

For a distribution $\mathcal{D} = (\mathcal{U}_n, \varphi)$, with $\varphi(z) = \mathbb{E}_{(x,y)\sim\mathcal{D}}[y|x = z]$, and a function $f: \mathcal{X}_n \to \{0,1\}$, if we write $\mathrm{g} = (-1)^f = 1 - 2f$ and $\phi = 1 - 2\varphi$, then we get:

# Boolean Fourier Analysis Crash Course

**Definition:**

The Fourier transform of $\phi: \mathcal{X}_n \to \mathbb{R}$ is defined as

$$\hat{\phi}: \mathcal{X}_n \to \mathbb{R},\ \hat{\phi}(s) = \mathbb{E}_{x\sim\mathcal{U}_n}[\phi(x)(-1)^{s\cdot x}].$$

Connecting learning and Fourier analysis:

For a distribution $\mathcal{D} = (\mathcal{U}_n, \varphi)$, with $\varphi(z) = \mathbb{E}_{(x,y)\sim\mathcal{D}}[y|x = z]$, and a function $f: \mathcal{X}_n \to \{0,1\}$, if we write $\mathrm{g} = (-1)^f = 1 - 2f$ and $\phi = 1 - 2\varphi$, then we get:

$$\mathbb{P}_{(x,y)\sim\mathcal{D}}[f(x) \neq y] = \frac{1 - \sum_{s\in\mathcal{X}_n} \hat{g}(s)\, \hat{\phi}(s)}{2}.$$

# Boolean Fourier Analysis Crash Course

# Boolean Fourier Analysis Crash Course

Example applications of Fourier-based (classical) learning:

# Boolean Fourier Analysis Crash Course

Example applications of Fourier-based (classical) learning:

- Realizably learning functions with low-degree Fourier concentration [3,4]

[3] N. Linal, Y. Mansour, and N. Nisan, *J. ACM 40.3, 607-620 (1993)*
[4] A. Eskenazis and P. Ivanisvili, *STOC 2022*

# Boolean Fourier Analysis Crash Course

Example applications of Fourier-based (classical) learning:

- Realizably learning functions with low-degree Fourier concentration [3,4]
- Realizably learning Fourier-sparse functions such as DNFs or decision trees (from queries) [5,6]


[3] N. Linal, Y. Mansour, and N. Nisan, *J. ACM 40.3, 607-620 (1993)*
[4] A. Eskenazis and P. Ivanisvili, *STOC 2022*
[5] O. Goldreich and L. Levin, *STOC 1989*
[6] E. Kushilevitz and Y. Mansour, *SIAM J. Comput. 22.6 (1993)*

# Boolean Fourier Analysis Crash Course

Example applications of Fourier-based (classical) learning:

- Realizably learning functions with low-degree Fourier concentration [3,4]
- Realizably learning Fourier-sparse functions such as DNFs or decision trees (from queries) [5,6]
- Agnostic parity learning (folklore, from queries)


[3] N. Linal, Y. Mansour, and N. Nisan, *J. ACM 40.3, 607-620 (1993)*
[4] A. Eskenazis and P. Ivanisvili, *STOC 2022*
[5] O. Goldreich and L. Levin, *STOC 1989*
[6] E. Kushilevitz and Y. Mansour, *SIAM J. Comput. 22.6 (1993)*

# Boolean Fourier Analysis Crash Course

Example applications of Fourier-based (classical) learning:

- Realizably learning functions with low-degree Fourier concentration [3,4]
- Realizably learning Fourier-sparse functions such as DNFs or decision trees (from queries) [5,6]
- Agnostic parity learning (folklore, from queries)
- 2-agnostic Fourier-sparse learning (from queries) [2]

[3] N. Linal, Y. Mansour, and N. Nisan, *J. ACM 40.3, 607-620 (1993)*
[4] A. Eskenazis and P. Ivanisvili, *STOC 2022*
[5] O. Goldreich and L. Levin, *STOC 1989*
[6] E. Kushilevitz and Y. Mansour, *SIAM J. Comput. 22.6 (1993)*
[2] **M.C.C.**, M. Hinsche, M. Ioannou, A. Nietner, and R. Sweke; *arXiv:2306.04843*

# Learning From Superposition Examples

# Learning From Superposition Examples

$|\psi_{\mathcal{D}}\rangle$

$|\psi_{\mathcal{D}}\rangle$

$\vdots$

$|\psi_{\mathcal{D}}\rangle$

# Learning From Superposition Examples

$|\psi_{\mathcal{D}}\rangle$

$|\psi_{\mathcal{D}}\rangle$

⋮

$|\psi_{\mathcal{D}}\rangle$

$|\psi_{\mathcal{D}}\rangle = \sum_{(x,y)\in\{0,1\}^{n+1}} \sqrt{\mathcal{D}(x,y)}\,|x,y\rangle$ [7]

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

# Learning From Superposition Examples

Unknown!

$|\psi_{\mathcal{D}}\rangle$

$|\psi_{\mathcal{D}}\rangle$

⋮

$|\psi_{\mathcal{D}}\rangle$

$$|\psi_{\mathcal{D}}\rangle = \sum_{(x,y)\in\{0,1\}^{n+1}} \sqrt{\mathcal{D}(x,y)}\,|x,y\rangle \text{ [7]}$$

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

# Learning From Superposition Examples

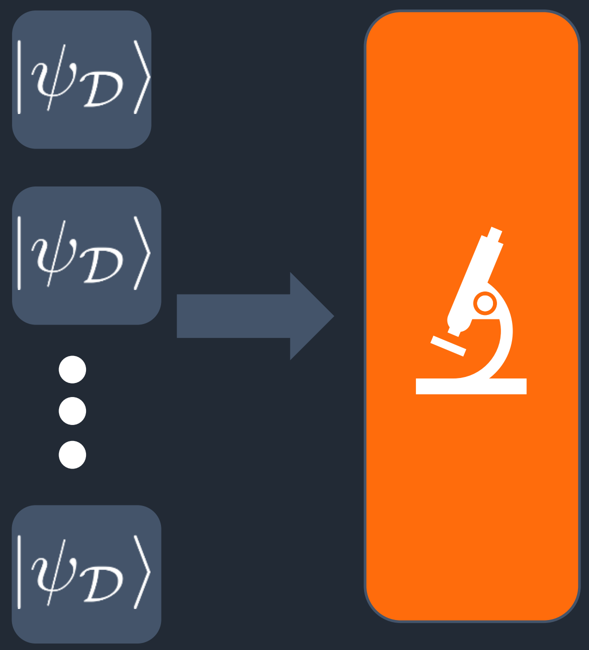
$|\psi_{\mathcal{D}}\rangle$

$|\psi_{\mathcal{D}}\rangle$

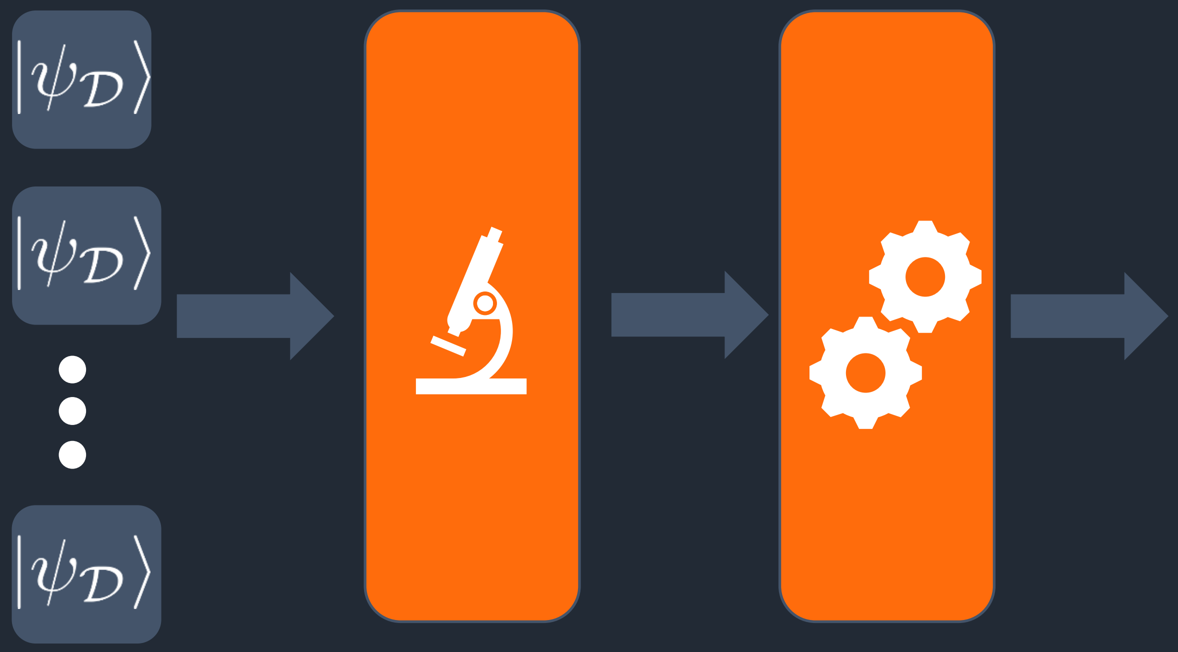
$\vdots$

$|\psi_{\mathcal{D}}\rangle$

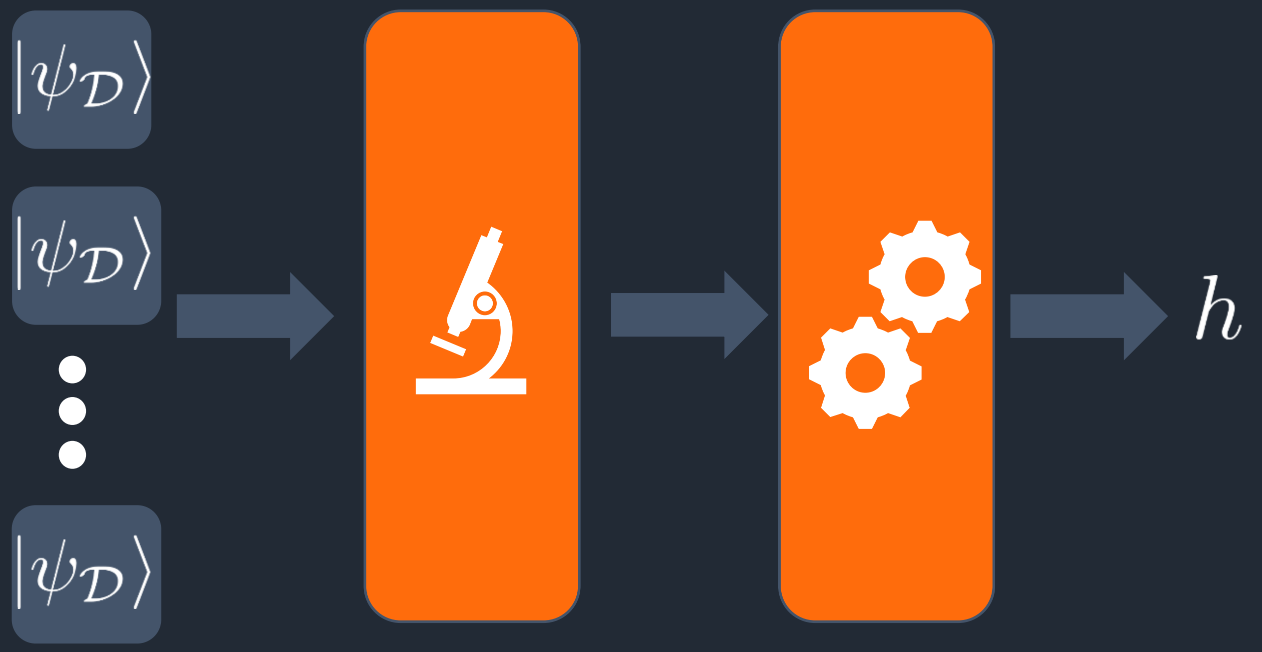
# Learning From Superposition Examples



# Learning From Superposition Examples



# Learning From Superposition Examples



# Learning From Superposition Examples

# Learning From Superposition Examples

Quantum Fourier sampling central to learning from $|\psi_{(\mathcal{U}_n,f)}\rangle$

# Learning From Superposition Examples

Quantum Fourier sampling central to learning from $|\psi_{(\mathcal{U}_n, f)}\rangle$

- Parity learning [8,9,10]

[8] E. Bernstein and U. Vazirani; *SIAM Journal on Computing 26.5, 1411-1473 (1997)*
[9] A.B. Grilo, I. Kerenidis, and T. Zijlstra; *Phys. Rev. A 99, 032314 (2019)*
[10] M.C.C.; *Quantum Information Processing 19, 172 (2020)*

# Learning From Superposition Examples

Quantum Fourier sampling central to learning from $|\psi_{(\mathcal{U}_n,f)}\rangle$

- Parity learning [8,9,10]
- DNF learning [7,11]

[8] E. Bernstein and U. Vazirani; *SIAM Journal on Computing 26.5, 1411-1473 (1997)*
[9] A.B. Grilo, I. Kerenidis, and T. Zijlstra; *Phys. Rev. A 99, 032314 (2019)*
[10] M.C.C.; *Quantum Information Processing 19, 172 (2020)*
[7] N.H. Bshouty and J.C. Jackson; *SIAM Journal on Computing 28.3, 1136-1153 (1998)*
[11] V. Kanade, A. Rocchetto, and S. Severini; *QIC 19.15&16, 1261–1278 (2019)*

# Learning From Superposition Examples

Quantum Fourier sampling central to learning from $|\psi_{(\mathcal{U}_n,f)}\rangle$

- Parity learning [8,9,10]
- DNF learning [7,11]
- Fourier-sparse learning [12]

[8] E. Bernstein and U. Vazirani; *SIAM Journal on Computing 26.5, 1411-1473 (1997)*
[9] A.B. Grilo, I. Kerenidis, and T. Zijlstra; *Phys. Rev. A 99, 032314 (2019)*
[10] M.C.C.; *Quantum Information Processing 19, 172 (2020)*
[7] N.H. Bshouty and J.C. Jackson; *SIAM Journal on Computing 28.3, 1136-1153 (1998)*
[11] V. Kanade, A. Rocchetto, and S. Severini; *QIC 19.15&16, 1261–1278 (2019)*
[12] S. Arunachalam et al.; *Quantum 5, 587 (2021)*

# Learning From Superposition Examples

Quantum Fourier sampling central to learning from $|\psi_{(\mathcal{U}_n, f)}\rangle$

- Parity learning [8,9,10]
- DNF learning [7,11]
- Fourier-sparse learning [12]

No quantum Fourier sampling known for learning from $|\psi_{\mathcal{D}}\rangle$ unless $\mathcal{D} = (\mathcal{U}_n, f)$ up to noise

?


[8] E. Bernstein and U. Vazirani; *SIAM Journal on Computing 26.5, 1411-1473 (1997)*

[9] A.B. Grilo, I. Kerenidis, and T. Zijlstra; *Phys. Rev. A 99, 032314 (2019)*

[10] M.C.C.; *Quantum Information Processing 19, 172 (2020)*

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

[11] V. Kanade, A. Rocchetto, and S. Severini; *QIC 19.15&16, 1261–1278 (2019)*

[12] S. Arunachalam et al.; *Quantum 5, 587 (2021)*

# Learning From Superposition Examples

Quantum Fourier sampling central to learning from $|\psi_{(\mathcal{U}_n,f)}\rangle$

- Parity learning [8,9,10]
- DNF learning [7,11]
- Fourier-sparse learning [12]

No quantum Fourier sampling known for learning from $|\psi_{\mathcal{D}}\rangle$ unless $\mathcal{D} = (\mathcal{U}_n, f)$ up to noise

?


[8] E. Bernstein and U. Vazirani; *SIAM Journal on Computing 26.5, 1411-1473 (1997)*
[9] A.B. Grilo, I. Kerenidis, and T. Zijlstra; *Phys. Rev. A 99, 032314 (2019)*
[10] M.C.C.; *Quantum Information Processing 19, 172 (2020)*
[7] N.H. Bshouty and J.C. Jackson; *SIAM Journal on Computing 28.3, 1136-1153 (1998)*
[11] V. Kanade, A. Rocchetto, and S. Severini; *QIC 19.15&16, 1261–1278 (2019)*
[12] S. Arunachalam et al.; *Quantum 5, 587 (2021)*

# A New Data Oracle for Agnostic Quantum Learning

How we redefine quantum data for agnostic learning

# Mixture-of-Superpositions (MoS) Examples

# Mixture-of-Superpositions (MoS) Examples

One quantum
example, please!

# Mixture-of-Superpositions (MoS) Examples

One quantum example, please!

MoS Oracle for $\mathcal{D}$

# Mixture-of-Superpositions (MoS) Examples

One quantum example, please!

MoS Oracle for $\mathcal{D}$

Random function $f \sim F_{\mathcal{D}}$

# Mixture-of-Superpositions (MoS) Examples

One quantum example, please!

MoS Oracle for $\mathcal{D}$

Random function $f \sim F_{\mathcal{D}}$

$|\psi_{(\mathcal{U}_n, f)}\rangle$

# Mixture-of-Superpositions (MoS) Examples

One quantum example, please!

MoS Oracle for $\mathcal{D}$

Random function $f \sim F_{\mathcal{D}}$

$|\psi_{(\mathcal{U}_n, f)}\rangle$

$|\psi_{(\mathcal{U}_n, f)}\rangle$

# Mixture-of-Superpositions (MoS) Examples

# Mixture-of-Superpositions (MoS) Examples

**Definition:**

$$\rho_{\mathcal{D}} = \mathbb{E}_{f \sim F_{\mathcal{D}}} \left[ \left| \psi_{(\mathcal{U}_n, f)} \right\rangle \left\langle \psi_{(\mathcal{U}_n, f)} \right| \right]$$

# Mixture-of-Superpositions (MoS) Examples

**Definition:**

$$\rho_{\mathcal{D}} = \mathbb{E}_{f \sim F_{\mathcal{D}}} \left[ \left| \psi_{(\mathcal{U}_n, f)} \right\rangle \left\langle \psi_{(\mathcal{U}_n, f)} \right| \right]$$

Here, $F_{\mathcal{D}}$ is the distribution that $\mathcal{D}$ induces over all functions:

# Mixture-of-Superpositions (MoS) Examples

**Definition:**

$$\rho_{\mathcal{D}} = \mathbb{E}_{f \sim F_{\mathcal{D}}} \left[ \left| \psi_{(\mathcal{U}_n, f)} \right\rangle \left\langle \psi_{(\mathcal{U}_n, f)} \right| \right]$$

Here, $F_{\mathcal{D}}$ is the distribution that $\mathcal{D}$ induces over all functions:

$$\mathbb{P}_{f \sim F_{\mathcal{D}}}[f = \tilde{f}] = \prod_{z \in \mathcal{X}_n} \mathbb{P}_{(x,y) \sim \mathcal{D}}[\tilde{f}(z) = y | x = z]$$

# Mixture-of-Superpositions (MoS) Examples

**Definition:**

$$\rho_{\mathcal{D}} = \mathbb{E}_{f\sim F_{\mathcal{D}}}\left[ \left| \psi_{(\mathcal{U}_n, f)} \right\rangle \left\langle \psi_{(\mathcal{U}_n, f)} \right| \right]$$

Here, $F_{\mathcal{D}}$ is the distribution that $\mathcal{D}$ induces over all functions:

$$\mathbb{P}_{f\sim F_{\mathcal{D}}}[f = \tilde{f}] = \prod_{z\in\mathcal{X}_n} \mathbb{P}_{(x,y)\sim\mathcal{D}}[\tilde{f}(z) = y | x = z]$$

Sanity checks:

# Mixture-of-Superpositions (MoS) Examples

**Definition:**

$$\rho_{\mathcal{D}} = \mathbb{E}_{f\sim F_{\mathcal{D}}}\left[\left|\psi_{(\mathcal{U}_n,f)}\right\rangle\left\langle\psi_{(\mathcal{U}_n,f)}\right|\right]$$

Here, $F_{\mathcal{D}}$ is the distribution that $\mathcal{D}$ induces over all functions:

$$\mathbb{P}_{f\sim F_{\mathcal{D}}}[f=\tilde{f}] = \prod_{z\in\mathcal{X}_n}\mathbb{P}_{(x,y)\sim\mathcal{D}}[\tilde{f}(z)=y|x=z]$$

Sanity checks:

1. If $\mathcal{D} = (\mathcal{U}_n, f)$, then $\rho_{\mathcal{D}} = |\psi_{(\mathcal{U}_n,f)}\rangle\langle\psi_{(\mathcal{U}_n,f)}|$.

# Mixture-of-Superpositions (MoS) Examples

**Definition:**

$$\rho_{\mathcal{D}} = \mathbb{E}_{f\sim F_{\mathcal{D}}}\left[\left|\psi_{(\mathcal{U}_n,f)}\right\rangle\left\langle\psi_{(\mathcal{U}_n,f)}\right|\right]$$

Here, $F_{\mathcal{D}}$ is the distribution that $\mathcal{D}$ induces over all functions:

$$\mathbb{P}_{f\sim F_{\mathcal{D}}}[f = \tilde{f}] = \prod_{z\in\mathcal{X}_n} \mathbb{P}_{(x,y)\sim\mathcal{D}}[\tilde{f}(z) = y | x = z]$$

Sanity checks:

1. If $\mathcal{D} = (\mathcal{U}_n, f)$, then $\rho_{\mathcal{D}} = |\psi_{(\mathcal{U}_n,f)}\rangle\langle\psi_{(\mathcal{U}_n,f)}|$.
2. Measuring $\rho_{\mathcal{D}}$ in the computational basis = Sampling from $\mathcal{D}$.

# Quantum Fourier Sampling from MoS

# Quantum Fourier Sampling from MoS

Quantum Fourier sampling central to learning from $|\psi_{(\mathcal{U}_n,f)}\rangle$

No quantum Fourier sampling known for learning from $|\psi_{\mathcal{D}}\rangle$ unless $\mathcal{D} = (\mathcal{U}_n, f)$ up to noise

# Quantum Fourier Sampling from MoS

Quantum Fourier sampling central to learning from $|\psi_{(\mathcal{U}_n, f)}\rangle$

No quantum Fourier sampling known for learning from $|\psi_{\mathcal{D}}\rangle$ unless $\mathcal{D} = (\mathcal{U}_n, f)$ up to noise

# Quantum Fourier Sampling from MoS

Quantum Fourier sampling central to learning from $|\psi_{(\mathcal{U}_n, f)}\rangle$

No quantum Fourier sampling known for learning from $|\psi_{\mathcal{D}}\rangle$ unless $\mathcal{D} = (\mathcal{U}_n, f)$ up to noise

**Theorem:**

# Quantum Fourier Sampling from MoS

Quantum Fourier sampling central to learning from $|\psi_{(\mathcal{U}_n,f)}\rangle$

No quantum Fourier sampling known for learning from $|\psi_{\mathcal{D}}\rangle$ unless $\mathcal{D} = (\mathcal{U}_n, f)$ up to noise

**Theorem:**

Given $\rho_{\mathcal{D}}$ for $\mathcal{D} = (\mathcal{U}_n, \varphi)$, perform Hadamard gates on all qubits, measure in the computational basis to obtain outcome $(t, b) \in \{0,1\}^{n+1}$.

# Quantum Fourier Sampling from MoS

Quantum Fourier sampling central to learning from $|\psi_{(\mathcal{U}_n,f)}\rangle$

No quantum Fourier sampling known for learning from $|\psi_{\mathcal{D}}\rangle$ unless $\mathcal{D} = (\mathcal{U}_n, f)$ up to noise

**Theorem:**

Given $\rho_{\mathcal{D}}$ for $\mathcal{D} = (\mathcal{U}_n, \varphi)$, perform Hadamard gates on all qubits, measure in the computational basis to obtain outcome $(t, b) \in \{0,1\}^{n+1}$.

1. $\mathbb{P}[\mathrm{b} = 0] = \frac{1}{2} = \mathbb{P}[\mathrm{b} = 1]$

# Quantum Fourier Sampling from MoS

Quantum Fourier sampling central to learning from $|\psi_{(\mathcal{U}_n, f)}\rangle$

No quantum Fourier sampling known for learning from $|\psi_{\mathcal{D}}\rangle$ unless $\mathcal{D} = (\mathcal{U}_n, f)$ up to noise

**Theorem:**

Given $\rho_{\mathcal{D}}$ for $\mathcal{D} = (\mathcal{U}_n, \varphi)$, perform Hadamard gates on all qubits, measure in the computational basis to obtain outcome $(t, b) \in \{0,1\}^{n+1}$.

1. $\mathbb{P}[\mathrm{b} = 0] = \frac{1}{2} = \mathbb{P}[\mathrm{b} = 1]$
2. $\left|\mathbb{P}[\mathrm{t} = \mathrm{s}|\mathrm{b} = 1] - \left(\hat{\phi}(s)\right)^2\right| \leq \frac{1}{2^n}$

# Fourier Approximation from MoS

# Fourier Approximation from MoS

What is quantum Fourier sampling from MoS good for?

# Fourier Approximation from MoS

What is quantum Fourier sampling from MoS good for?

# Fourier Approximation from MoS

What is quantum Fourier sampling from MoS good for?

**Theorem:**

# Fourier Approximation from MoS

What is quantum Fourier sampling from MoS good for?

**Theorem:**

Let $\varepsilon > 2^{-\left(\frac{n}{2}-2\right)}$. With $\tilde{\mathcal{O}}(1/\varepsilon^4)$ copies of $\rho_{\mathcal{D}}$ for $\mathcal{D} = (\mathcal{U}_n, \varphi)$, we can efficiently produce a $\tilde{\phi}: \{0,1\}^n \to [-1,1]$ such that

# Fourier Approximation from MoS

What is quantum Fourier sampling from MoS good for?

**Theorem:**

Let $\varepsilon > 2^{-\left(\frac{n}{2}-2\right)}$. With $\tilde{\mathcal{O}}(1/\varepsilon^4)$ copies of $\rho_{\mathcal{D}}$ for $\mathcal{D} = (\mathcal{U}_n, \varphi)$, we can efficiently produce a $\tilde{\phi}: \{0,1\}^n \to [-1,1]$ such that

*1.* $\|\tilde{\phi} - \hat{\phi}\|_\infty \leq \varepsilon$, and

# Fourier Approximation from MoS

What is quantum Fourier sampling from MoS good for?

**Theorem:**

Let $\varepsilon > 2^{-\left(\frac{n}{2}-2\right)}$. With $\tilde{\mathcal{O}}(1/\varepsilon^4)$ copies of $\rho_{\mathcal{D}}$ for $\mathcal{D} = (\mathcal{U}_n, \varphi)$, we can efficiently produce a $\tilde{\phi}: \{0,1\}^n \to [-1,1]$ such that

1. $\|\tilde{\phi} - \hat{\phi}\|_\infty \leq \varepsilon$, and
2. $\|\tilde{\phi}\|_0 \leq \mathcal{O}\left(\frac{1}{\varepsilon^2}\right)$,

# Fourier Approximation from MoS

What is quantum Fourier sampling from MoS good for?

**Theorem:**

Let $\varepsilon > 2^{-\left(\frac{n}{2}-2\right)}$. With $\tilde{\mathcal{O}}(1/\varepsilon^4)$ copies of $\rho_{\mathcal{D}}$ for $\mathcal{D} = (\mathcal{U}_n, \varphi)$, we can efficiently produce a $\tilde{\phi}: \{0,1\}^n \to [-1,1]$ such that

1. $\|\tilde{\phi} - \hat{\phi}\|_\infty \leq \varepsilon$, and
2. $\|\tilde{\phi}\|_0 \leq \mathcal{O}\left(\frac{1}{\varepsilon^2}\right)$,

with high success probability.

# Fourier Approximation from MoS

What is quantum Fourier sampling from MoS good for?

**Theorem:**

Let $\varepsilon > 2^{-\left(\frac{n}{2}-2\right)}$. With $\tilde{\mathcal{O}}(1/\varepsilon^4)$ copies of $\rho_{\mathcal{D}}$ for $\mathcal{D} = (\mathcal{U}_n, \varphi)$, we can efficiently produce a $\tilde{\phi}: \{0,1\}^n \to [-1,1]$ such that

1. $\|\tilde{\phi} - \hat{\phi}\|_\infty \le \varepsilon$, and
2. $\|\tilde{\phi}\|_0 \le \mathcal{O}\left(\frac{1}{\varepsilon^2}\right)$,

with high success probability.

For the next results, all $\varepsilon$ are $\varepsilon > 2^{-\left(\frac{n}{2}-2\right)}$.

# Fourier Approximation from MoS

**Proof Idea:**

# Fourier Approximation from MoS

**Proof Idea:**

- We can sample from a distribution that is close to $\left\{\left(\hat{\phi}(s)\right)^2\right\}_s$ up to error $\frac{1}{2^n} < \frac{\varepsilon^2}{16}$.

# Fourier Approximation from MoS

**Proof Idea:**

- We can sample from a distribution that is close to $\left\{\left(\hat{\phi}(s)\right)^2\right\}_s$ up to error $\frac{1}{2^n} < \frac{\varepsilon^2}{16}$.
- The DKW Theorem [13] (discrete version [14]) implies:
  With high probability, $\mathcal{O}\left(\frac{1}{\tau^2}\right)$ samples suffice to approximate the probability distribution to accuracy $\tau$.

[13] A. Dvoretzky, J. Kiefer, and J. Wolfowitz, *Ann. Math. Stat. 27.3, 642-669 (1956)*
[14] M.R. Kosorok, *Introduction to empirical processes and semiparametric inference (2008)*

# Fourier Approximation from MoS

**Proof Idea:**

- We can sample from a distribution that is close to $\left\{\left(\hat{\phi}(s)\right)^2\right\}_s$ up to error $\frac{1}{2^n} < \frac{\varepsilon^2}{16}$.
- The DKW Theorem [13] (discrete version [14]) implies:
  With high probability, $\mathcal{O}\left(\frac{1}{\tau^2}\right)$ samples suffice to approximate the probability distribution to accuracy $\tau$.
- Use DKW with $\tau = \frac{\varepsilon^2}{8}$ relying on the approximate sampling.

[13] A. Dvoretzky, J. Kiefer, and J. Wolfowitz, *Ann. Math. Stat. 27.3, 642-669 (1956)*
[14] M.R. Kosorok, *Introduction to empirical processes and semiparametric inference (2008)*

# Agnostic Parity and Fourier-Sparse Quantum Learning from MoS

# Agnostic Parity and Fourier-Sparse Quantum Learning from MoS

What is quantum Fourier approximation from MoS good for?

# Agnostic Parity and Fourier-Sparse Quantum Learning from MoS

What is quantum Fourier approximation from MoS good for?

# Agnostic Parity and Fourier-Sparse Quantum Learning from MoS

What is quantum Fourier approximation from MoS good for?

**Corollary (1-Agnostic Proper Parity Learning from MoS):**

# Agnostic Parity and Fourier-Sparse Quantum Learning from MoS

What is quantum Fourier approximation from MoS good for?

**Corollary (1-Agnostic Proper Parity Learning from MoS):**
Using $\tilde{\mathcal{O}}\left(\frac{1}{\varepsilon^4}\right)$ copies of $\rho_{\mathcal{D}}$, we can efficiently produce a bit string $s \in \{0,1\}^n$ such that

# Agnostic Parity and Fourier-Sparse Quantum Learning from MoS

What is quantum Fourier approximation from MoS good for?

**Corollary (1-Agnostic Proper Parity Learning from MoS):**
Using $\tilde{\mathcal{O}}\left(\frac{1}{\varepsilon^4}\right)$ copies of $\rho_{\mathcal{D}}$, we can efficiently produce a bit string $s \in \{0,1\}^n$ such that

$$\mathbb{P}_{(x,b)\sim\mathcal{D}}[b \neq s \cdot x] \leq \min_{t\in\{0,1\}^n} \mathbb{P}_{(x,b)\sim\mathcal{D}}[b \neq t \cdot x] + \varepsilon\,.$$

# Agnostic Parity and Fourier-Sparse Quantum Learning from MoS

What is quantum Fourier approximation from MoS good for?

**Corollary (1-Agnostic Proper Parity Learning from MoS):**
Using $\tilde{\mathcal{O}}\left(\frac{1}{\varepsilon^4}\right)$ copies of $\rho_{\mathcal{D}}$, we can efficiently produce a bit string $s \in \{0,1\}^n$ such that

$$\mathbb{P}_{(x,b)\sim\mathcal{D}}[b \neq s \cdot x] \leq \min_{t\in\{0,1\}^n} \mathbb{P}_{(x,b)\sim\mathcal{D}}[b \neq t \cdot x] + \varepsilon\,.$$

Note: This is (at least) LPN-hard from classical examples!

# Agnostic Parity and Fourier-Sparse Quantum Learning from MoS

What is quantum Fourier approximation from MoS good for?

# Agnostic Parity and Fourier-Sparse Quantum Learning from MoS

What is quantum Fourier approximation from MoS good for?

# Agnostic Parity and Fourier-Sparse Quantum Learning from MoS

What is quantum Fourier approximation from MoS good for?

**Corollary (2-Agnostic Fourier-Sparse Learning from MoS):**

# Agnostic Parity and Fourier-Sparse Quantum Learning from MoS

What is quantum Fourier approximation from MoS good for?

**Corollary (2-Agnostic Fourier-Sparse Learning from MoS):**
Using $\tilde{\mathcal{O}}\left(\frac{k^4}{\varepsilon^4}\right)$ copies of $\rho_{\mathcal{D}}$, we can efficiently produce a randomized hypothesis $h: \{0,1\}^n \to \{0,1\}$ such that

# Agnostic Parity and Fourier-Sparse Quantum Learning from MoS

What is quantum Fourier approximation from MoS good for?

**Corollary (2-Agnostic Fourier-Sparse Learning from MoS):**
Using $\tilde{\mathcal{O}}\left(\frac{k^4}{\varepsilon^4}\right)$ copies of $\rho_{\mathcal{D}}$, we can efficiently produce a randomized hypothesis $h: \{0,1\}^n \to \{0,1\}$ such that

$$\mathbb{P}_{(x,b)\sim\mathcal{D}}\left[b \neq h(x)\right] \leq 2 \min_{\substack{\tilde{f}:\mathcal{X}_n \to \{0,1\} \\ \text{Fourier}-k-\text{sparse}}} \mathbb{P}_{(x,b)\sim\mathcal{D}}[b \neq \tilde{f}(x)] + \varepsilon\,.$$

# Guiding Question

**Question:**

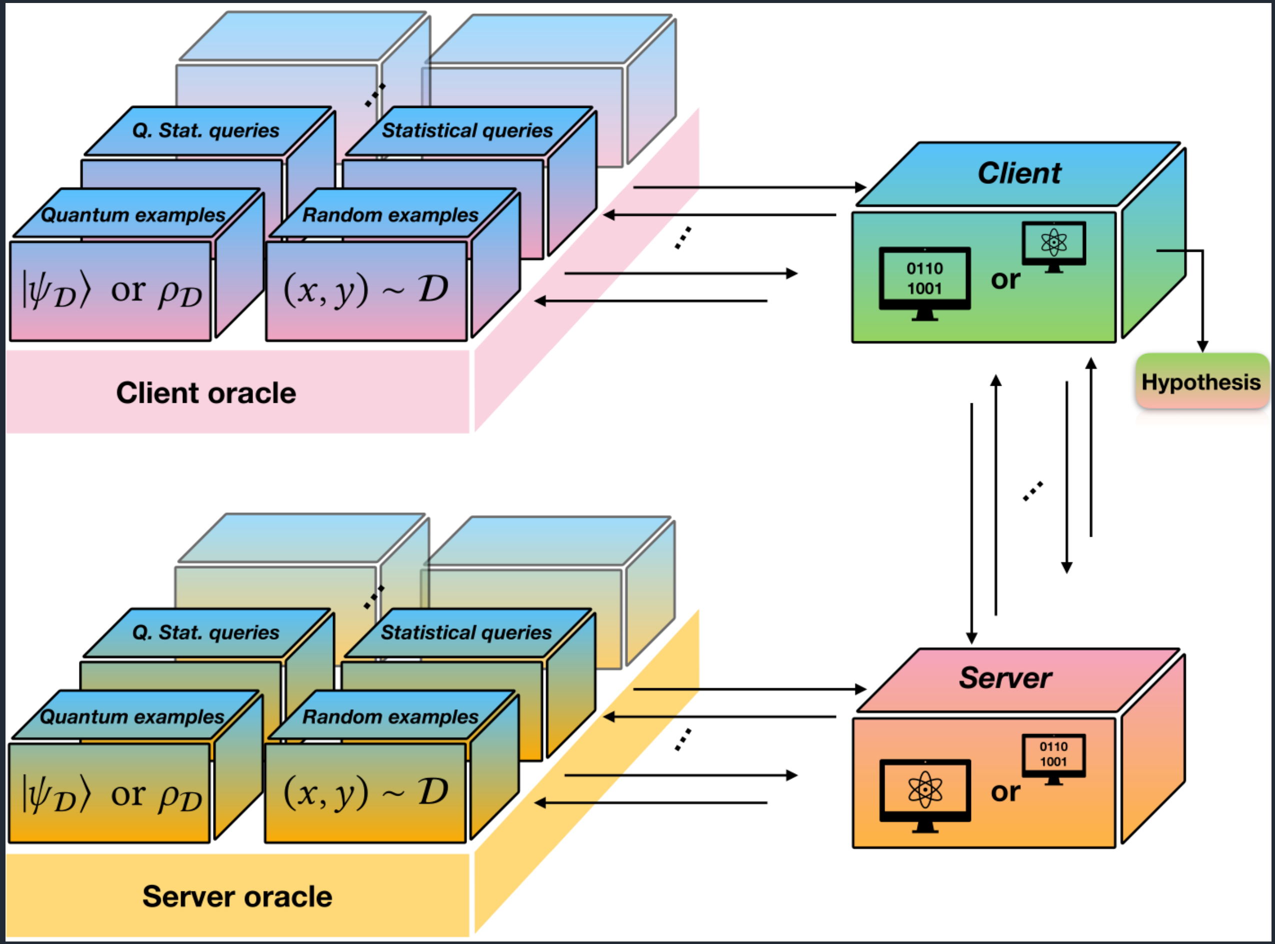
Is there an (agnostic) learning problem that simultaneously...

1. ...is intractable for classical learners but efficiently solvable by quantum learners using quantum data,
2. ...and can be efficiently verifiably delegated from a classical verifier to an untrusted quantum prover?

# Interactive Classical Verification of Quantum Learning

How classical clients can verifiably delegate learning to quantum servers

# Verifying Quantum Learners



# Two Assumptions

# Two Assumptions

1. $\mathcal{D} = (\mathcal{U}_n, \varphi)$ has no small non-zero Fourier coefficients:

# Two Assumptions

1. $\mathcal{D} = (\mathcal{U}_n, \varphi)$ has no small non-zero Fourier coefficients:

$$\mathcal{D} \in \mathfrak{D}_{\mathcal{U}_n;\geq\vartheta} := \left\{(\mathcal{U}_n, \varphi) \mid \hat{\phi} \neq 0 \ \Rightarrow |\hat{\phi}| \geq \vartheta\right\}$$

# Two Assumptions

1. $\mathcal{D} = (\mathcal{U}_n, \varphi)$ has no small non-zero Fourier coefficients:

$$\mathcal{D} \in \mathfrak{D}_{\mathcal{U}_n;\geq\vartheta} := \left\{(\mathcal{U}_n, \varphi) \mid \hat{\phi} \neq 0 \ \Rightarrow |\hat{\phi}| \geq \vartheta\right\}$$

2. $\mathcal{D} = (\mathcal{U}_n, \varphi)$ has $L_2$-bounded bias:

# Two Assumptions

1. $\mathcal{D} = (\mathcal{U}_n, \varphi)$ has no small non-zero Fourier coefficients:

$$\mathcal{D} \in \mathfrak{D}_{\mathcal{U}_n;\geq\vartheta} := \left\{(\mathcal{U}_n, \varphi) \mid \hat{\phi} \neq 0 \ \Rightarrow |\hat{\phi}| \geq \vartheta\right\}$$

2. $\mathcal{D} = (\mathcal{U}_n, \varphi)$ has $L_2$-bounded bias:

$$\mathcal{D} \in \mathfrak{D}_{\mathcal{U}_n;[a^2,b^2]} := \left\{(\mathcal{U}_n, \varphi) \mid \mathbb{E}_{x\sim\mathcal{U}_n}[(\phi(x))^2] \in [a^2, b^2]\right\}$$

# Two Assumptions

1. $\mathcal{D} = (\mathcal{U}_n, \varphi)$ has no small non-zero Fourier coefficients:

$$\mathcal{D} \in \mathfrak{D}_{\mathcal{U}_n;\geq\vartheta} := \left\{ (\mathcal{U}_n, \varphi) \mid \hat{\phi} \neq 0 \ \Rightarrow |\hat{\phi}| \geq \vartheta \right\}$$

2. $\mathcal{D} = (\mathcal{U}_n, \varphi)$ has $L_2$-bounded bias:

$$\mathcal{D} \in \mathfrak{D}_{\mathcal{U}_n;[a^2,b^2]} := \left\{ (\mathcal{U}_n, \varphi) \mid \mathbb{E}_{x\sim\mathcal{U}_n}[(\phi(x))^2] \in [a^2, b^2] \right\}$$

Note: Both are satisfied (for suitable $\vartheta$, $a$, and $b$) if $\mathcal{D}$ is known to correspond to a noisy parity or Fourier-sparse function.

# Two Assumptions

1. $\mathcal{D} = (\mathcal{U}_n, \varphi)$ has no small non-zero Fourier coefficients:

$$\mathcal{D} \in \mathfrak{D}_{\mathcal{U}_n; \geq \vartheta} := \left\{ (\mathcal{U}_n, \varphi) \mid \hat{\phi} \neq 0 \;\Rightarrow |\hat{\phi}| \geq \vartheta \right\}$$

2. $\mathcal{D} = (\mathcal{U}_n, \varphi)$ has $L_2$-bounded bias:

$$\mathcal{D} \in \mathfrak{D}_{\mathcal{U}_n; [a^2, b^2]} := \left\{ (\mathcal{U}_n, \varphi) \mid \mathbb{E}_{x \sim \mathcal{U}_n}[(\phi(x))^2] \in [a^2, b^2] \right\}$$

Note: Both are satisfied (for suitable $\vartheta$, $a$, and $b$) if $\mathcal{D}$ is known to correspond to a noisy parity or Fourier-sparse function.

➢Classical agnostic learning is hard in these settings!

# Interactive Verification of Fourier Spectrum Approximation

# Interactive Verification of Fourier Spectrum Approximation

Central ingredient in our quantum learning procedures:

# Interactive Verification of Fourier Spectrum Approximation

Central ingredient in our quantum learning procedures:

Quantum Fourier sampling + Fourier spectrum approximation

# Interactive Verification of Fourier Spectrum Approximation

Central ingredient in our quantum learning procedures:

Quantum Fourier sampling + Fourier spectrum approximation

➢ Can we classically verify this quantum subroutine?

# Interactive Verification of Fourier Spectrum Approximation

# Interactive Verification of Fourier Spectrum Approximation

# Interactive Verification of Fourier Spectrum Approximation

**Theorem:**

Let $\vartheta > 2^{-\left(\frac{n}{2}-3\right)}$ and $\varepsilon > 2\sqrt{b^2 - a^2}$. There is a classical-quantum pair $(V, P)$ that, for any $\mathcal{D} \in \mathfrak{D}_{\mathcal{U}_n;\geq\vartheta} \cap \mathfrak{D}_{\mathcal{U}_n;[a^2,b^2]}$, achieves:

# Interactive Verification of Fourier Spectrum Approximation

**Theorem:**

Let $\vartheta > 2^{-\left(\frac{n}{2}-3\right)}$ and $\varepsilon > 2\sqrt{b^2 - a^2}$. There is a classical-quantum pair $(V, P)$ that, for any $\mathcal{D} \in \mathfrak{D}_{\mathcal{U}_{n};\geq\vartheta} \cap \mathfrak{D}_{\mathcal{U}_{n};[a^2,b^2]}$, achieves:

*1.* $P$ uses $\tilde{\mathcal{O}}\left(\frac{1}{\vartheta^4}\right)$ copies of $\rho_{\mathcal{D}}$ and time $\tilde{\mathcal{O}}\left(\frac{n}{\vartheta^4}\right)$,

# Interactive Verification of Fourier Spectrum Approximation

**Theorem:**

Let $\vartheta > 2^{-\left(\frac{n}{2}-3\right)}$ and $\varepsilon > 2\sqrt{b^2 - a^2}$. There is a classical-quantum pair $(V, P)$ that, for any $\mathcal{D} \in \mathfrak{D}_{\mathcal{U}_n;\geq\vartheta} \cap \mathfrak{D}_{\mathcal{U}_n;[a^2,b^2]}$, achieves:

1. $P$ uses $\tilde{\mathcal{O}}\left(\frac{1}{\vartheta^4}\right)$ copies of $\rho_{\mathcal{D}}$ and time $\tilde{\mathcal{O}}\left(\frac{n}{\vartheta^4}\right)$,
2. $V$ uses $\tilde{\mathcal{O}}\left(\frac{b^4}{\varepsilon^4\vartheta^4}\right)$ classical examples and time $\tilde{\mathcal{O}}\left(\frac{nb^4}{\varepsilon^4\vartheta^4}\right)$,

# Interactive Verification of Fourier Spectrum Approximation

**Theorem:**

Let $\vartheta > 2^{-\left(\frac{n}{2}-3\right)}$ and $\varepsilon > 2\sqrt{b^2 - a^2}$. There is a classical-quantum pair $(V, P)$ that, for any $\mathcal{D} \in \mathfrak{D}_{\mathcal{U}_{n;\geq\vartheta}} \cap \mathfrak{D}_{\mathcal{U}_{n;[a^2,b^2]}}$, achieves:

1. $P$ uses $\tilde{\mathcal{O}}\left(\frac{1}{\vartheta^4}\right)$ copies of $\rho_{\mathcal{D}}$ and time $\tilde{\mathcal{O}}\left(\frac{n}{\vartheta^4}\right)$,
2. $V$ uses $\tilde{\mathcal{O}}\left(\frac{b^4}{\varepsilon^4\vartheta^4}\right)$ classical examples and time $\tilde{\mathcal{O}}\left(\frac{nb^4}{\varepsilon^4\vartheta^4}\right)$,
3. If $V$ interacts with $P$, then $V$ outputs $\tilde{\phi}$ s.t. $\left\|\tilde{\phi} - \hat{\phi}\right\|_1 \leq \varepsilon$ and $\left\|\tilde{\phi}\right\|_0 \leq \mathcal{O}\left(\frac{1}{\vartheta^2}\right)$ with high probability.

# Interactive Verification of Fourier Spectrum Approximation

**Theorem:**

Let $\vartheta > 2^{-\left(\frac{n}{2}-3\right)}$ and $\varepsilon > 2\sqrt{b^2 - a^2}$. There is a classical-quantum pair $(V, P)$ that, for any $\mathcal{D} \in \mathfrak{D}_{\mathcal{U}_n;\geq\vartheta} \cap \mathfrak{D}_{\mathcal{U}_n;[a^2,b^2]}$, achieves:

1. $P$ uses $\tilde{\mathcal{O}}\left(\frac{1}{\vartheta^4}\right)$ copies of $\rho_{\mathcal{D}}$ and time $\tilde{\mathcal{O}}\left(\frac{n}{\vartheta^4}\right)$,
2. $V$ uses $\tilde{\mathcal{O}}\left(\frac{b^4}{\varepsilon^4\vartheta^4}\right)$ classical examples and time $\tilde{\mathcal{O}}\left(\frac{nb^4}{\varepsilon^4\vartheta^4}\right)$,
3. If $V$ interacts with $P$, then $V$ outputs $\tilde{\phi}$ s.t. $\left\|\tilde{\phi} - \hat{\phi}\right\|_1 \leq \varepsilon$ and $\left\|\tilde{\phi}\right\|_0 \leq \mathcal{O}\left(\frac{1}{\vartheta^2}\right)$ with high probability.
4. If $V$ interacts with any $P'$, then $V$ accepts and outputs $\left\|\tilde{\phi} - \hat{\phi}\right\|_1 > \varepsilon$ only with small probability.

# Interactive Verification of Fourier Spectrum Approximation

**Proof idea:**

# Interactive Verification of Fourier Spectrum Approximation

**Proof idea:**

*1.* $V$ asks $P$ to send a list of all non-zero Fourier coefficients.

# Interactive Verification of Fourier Spectrum Approximation

**Proof idea:**

1. $V$ asks $P$ to send a list of all non-zero Fourier coefficients.
2. $P$ uses Fourier approximation from MoS to create such a list and sends it to $V$.

# Interactive Verification of Fourier Spectrum Approximation

**Proof idea:**

1. $V$ asks $P$ to send a list of all non-zero Fourier coefficients.
2. $P$ uses Fourier approximation from MoS to create such a list and sends it to $V$.
3. $V$ uses classical examples to produce their own Fourier coefficient estimates for the coefficients that $P$ sent.

# Interactive Verification of Fourier Spectrum Approximation

**Proof idea:**

1. $V$ asks $P$ to send a list of all non-zero Fourier coefficients.
2. $P$ uses Fourier approximation from MoS to create such a list and sends it to $V$.
3. $V$ uses classical examples to produce their own Fourier coefficient estimates for the coefficients that $P$ sent.
4. If the total accumulated Fourier weight of the estimates is large enough, $V$ is happy. Otherwise, $V$ rejects.

# Interactive Verification of Quantum Agnostic Learning

# Interactive Verification of Quantum Agnostic Learning

# Interactive Verification of Quantum Agnostic Learning

**Theorem:**

# Interactive Verification of Quantum Agnostic Learning

**Theorem:**

Let $\vartheta > 2^{-\left(\frac{n}{2}-3\right)}$ and $\varepsilon > 2\sqrt{b^2 - a^2}$. There is a classical-quantum pair $(V, P)$ that, for any $\mathcal{D} \in \mathfrak{D}_{\mathcal{U}_n;\geq\vartheta} \cap \mathfrak{D}_{\mathcal{U}_n;[a^2,b^2]}$, achieves:

# Interactive Verification of Quantum Agnostic Learning

**Theorem:**

Let $\vartheta > 2^{-\left(\frac{n}{2}-3\right)}$ and $\varepsilon > 2\sqrt{b^2 - a^2}$. There is a classical-quantum pair $(V, P)$ that, for any $\mathcal{D} \in \mathfrak{D}_{\mathcal{U}_n;\geq\vartheta} \cap \mathfrak{D}_{\mathcal{U}_n;[a^2,b^2]}$, achieves:

1. $P$ uses $\tilde{\mathcal{O}}\left(\frac{1}{\vartheta^4}\right)$ copies of $\rho_{\mathcal{D}}$ and time $\tilde{\mathcal{O}}\left(\frac{n}{\vartheta^4}\right)$,

# Interactive Verification of Quantum Agnostic Learning

**Theorem:**

Let $\vartheta > 2^{-\left(\frac{n}{2}-3\right)}$ and $\varepsilon > 2\sqrt{b^2 - a^2}$. There is a classical-quantum pair $(V, P)$ that, for any $\mathcal{D} \in \mathfrak{D}_{\mathcal{U}_n;\geq\vartheta} \cap \mathfrak{D}_{\mathcal{U}_n;[a^2,b^2]}$, achieves:

*1.* $P$ uses $\tilde{\mathcal{O}}\left(\frac{1}{\vartheta^4}\right)$ copies of $\rho_{\mathcal{D}}$ and time $\tilde{\mathcal{O}}\left(\frac{n}{\vartheta^4}\right)$,

*2.* $V$ uses $\tilde{\mathcal{O}}\left(\frac{b^4}{\varepsilon^4\vartheta^4}\right)$ classical examples and time $\tilde{\mathcal{O}}\left(\frac{nb^4}{\varepsilon^4\vartheta^4}\right)$,

# Interactive Verification of Quantum Agnostic Learning

**Theorem:**

Let $\vartheta > 2^{-\left(\frac{n}{2}-3\right)}$ and $\varepsilon > 2\sqrt{b^2 - a^2}$. There is a classical-quantum pair $(V, P)$ that, for any $\mathcal{D} \in \mathfrak{D}_{\mathcal{U}_n;\geq\vartheta} \cap \mathfrak{D}_{\mathcal{U}_n;[a^2,b^2]}$, achieves:

1. $P$ uses $\tilde{\mathcal{O}}\left(\frac{1}{\vartheta^4}\right)$ copies of $\rho_{\mathcal{D}}$ and time $\tilde{\mathcal{O}}\left(\frac{n}{\vartheta^4}\right)$,
2. $V$ uses $\tilde{\mathcal{O}}\left(\frac{b^4}{\varepsilon^4\vartheta^4}\right)$ classical examples and time $\tilde{\mathcal{O}}\left(\frac{nb^4}{\varepsilon^4\vartheta^4}\right)$,
3. If $V$ interacts with $P$, then $V$ outputs a bit string $s \in \{0,1\}^n$ s.t. $\mathbb{P}_{(x,y)\sim\mathcal{D}}[s \cdot x \neq y] \leq \min_{t\in\mathcal{X}_n} \mathbb{P}_{(x,y)\sim\mathcal{D}}[t \cdot x \neq y] + \varepsilon$ with high prob.

# Interactive Verification of Quantum Agnostic Learning

**Theorem:**

Let $\vartheta > 2^{-\left(\frac{n}{2}-3\right)}$ and $\varepsilon > 2\sqrt{b^2 - a^2}$. There is a classical-quantum pair $(V, P)$ that, for any $\mathcal{D} \in \mathfrak{D}_{\mathcal{U}_n;\geq\vartheta} \cap \mathfrak{D}_{\mathcal{U}_n;[a^2,b^2]}$, achieves:

1. $P$ uses $\tilde{\mathcal{O}}\left(\frac{1}{\vartheta^4}\right)$ copies of $\rho_{\mathcal{D}}$ and time $\tilde{\mathcal{O}}\left(\frac{n}{\vartheta^4}\right)$,
2. $V$ uses $\tilde{\mathcal{O}}\left(\frac{b^4}{\varepsilon^4\vartheta^4}\right)$ classical examples and time $\tilde{\mathcal{O}}\left(\frac{nb^4}{\varepsilon^4\vartheta^4}\right)$,
3. If $V$ interacts with $P$, then $V$ outputs a bit string $s \in \{0,1\}^n$ s.t. $\mathbb{P}_{(x,y)\sim\mathcal{D}}[s \cdot x \neq y] \leq \min_{t\in\mathcal{X}_n} \mathbb{P}_{(x,y)\sim\mathcal{D}}[t \cdot x \neq y] + \varepsilon$ with high prob.
4. If $V$ interacts with any $P'$, then $V$ accepts and outputs a bit string $s \in \{0,1\}^n$ s.t. $\mathbb{P}_{(x,y)\sim\mathcal{D}}[s \cdot x \neq y] > \min_{t\in\mathcal{X}_n} \mathbb{P}_{(x,y)\sim\mathcal{D}}[t \cdot x \neq y] + \varepsilon$ with low prob.

# Interactive Verification of Quantum Agnostic Learning

# Interactive Verification of Quantum Agnostic Learning

# Interactive Verification of Quantum Agnostic Learning

**Theorem:**

# Interactive Verification of Quantum Agnostic Learning

**Theorem:**

Let $\vartheta > 2^{-\left(\frac{n}{2}-3\right)}$ and $\varepsilon > 4k\sqrt{b^2 - a^2}$. There is a classical-quantum pair $(V, P)$ that, for any $\mathcal{D} \in \mathfrak{D}_{\mathcal{U}_n;\geq\vartheta} \cap \mathfrak{D}_{\mathcal{U}_n;[a^2,b^2]}$, achieves:

# Interactive Verification of Quantum Agnostic Learning

**Theorem:**

Let $\vartheta > 2^{-\left(\frac{n}{2}-3\right)}$ and $\varepsilon > 4k\sqrt{b^2 - a^2}$. There is a classical-quantum pair $(V, P)$ that, for any $\mathcal{D} \in \mathfrak{D}_{\mathcal{U}_n;\geq\vartheta} \cap \mathfrak{D}_{\mathcal{U}_n;[a^2,b^2]}$, achieves:

1. $P$ uses $\tilde{\mathcal{O}}\left(\frac{1}{\vartheta^4}\right)$ copies of $\rho_{\mathcal{D}}$ and time $\tilde{\mathcal{O}}\left(\frac{n}{\vartheta^4}\right)$,

# Interactive Verification of Quantum Agnostic Learning

**Theorem:**

Let $\vartheta > 2^{-\left(\frac{n}{2}-3\right)}$ and $\varepsilon > 4k\sqrt{b^2 - a^2}$. There is a classical-quantum pair $(V, P)$ that, for any $\mathcal{D} \in \mathfrak{D}_{\mathcal{U}_n;\geq\vartheta} \cap \mathfrak{D}_{\mathcal{U}_n;[a^2,b^2]}$, achieves:

1. $P$ uses $\tilde{\mathcal{O}}\left(\frac{1}{\vartheta^4}\right)$ copies of $\rho_{\mathcal{D}}$ and time $\tilde{\mathcal{O}}\left(\frac{n}{\vartheta^4}\right)$,
2. $V$ uses $\tilde{\mathcal{O}}\left(\frac{b^4k^4}{\varepsilon^4\vartheta^4}\right)$ classical examples and time $\tilde{\mathcal{O}}\left(\frac{nb^4k^4}{\varepsilon^4\vartheta^4}\right)$,

# Interactive Verification of Quantum Agnostic Learning

**Theorem:**

Let $\vartheta > 2^{-\left(\frac{n}{2}-3\right)}$ and $\varepsilon > 4k\sqrt{b^2 - a^2}$. There is a classical-quantum pair $(V, P)$ that, for any $\mathcal{D} \in \mathfrak{D}_{\mathcal{U}_n;\geq\vartheta} \cap \mathfrak{D}_{\mathcal{U}_n;[a^2,b^2]}$, achieves:

1. $P$ uses $\tilde{\mathcal{O}}\left(\frac{1}{\vartheta^4}\right)$ copies of $\rho_{\mathcal{D}}$ and time $\tilde{\mathcal{O}}\left(\frac{n}{\vartheta^4}\right)$,
2. $V$ uses $\tilde{\mathcal{O}}\left(\frac{b^4k^4}{\varepsilon^4\vartheta^4}\right)$ classical examples and time $\tilde{\mathcal{O}}\left(\frac{nb^4k^4}{\varepsilon^4\vartheta^4}\right)$,
3. If $V$ interacts with $P$, then $V$ outputs a randomized hypothesis $h$ s.t. $\mathbb{P}_{(x,y)\sim\mathcal{D}}[h(x) \neq y] \leq \min_{\tilde{f} \text{ Fourier-k-sparse}} \mathbb{P}_{(x,y)\sim\mathcal{D}}\left[\tilde{f}(\boldsymbol{x}) \neq y\right] + \varepsilon$ with high prob.

# Interactive Verification of Quantum Agnostic Learning

**Theorem:**

Let $\vartheta > 2^{-\left(\frac{n}{2}-3\right)}$ and $\varepsilon > 4k\sqrt{b^2 - a^2}$. There is a classical-quantum pair $(V, P)$ that, for any $\mathcal{D} \in \mathfrak{D}_{\mathcal{U}_n;\geq\vartheta} \cap \mathfrak{D}_{\mathcal{U}_n;[a^2,b^2]}$, achieves:

1. $P$ uses $\tilde{\mathcal{O}}\left(\frac{1}{\vartheta^4}\right)$ copies of $\rho_{\mathcal{D}}$ and time $\tilde{\mathcal{O}}\left(\frac{n}{\vartheta^4}\right)$,
2. $V$ uses $\tilde{\mathcal{O}}\left(\frac{b^4k^4}{\varepsilon^4\vartheta^4}\right)$ classical examples and time $\tilde{\mathcal{O}}\left(\frac{nb^4k^4}{\varepsilon^4\vartheta^4}\right)$,
3. If $V$ interacts with $P$, then $V$ outputs a randomized hypothesis $h$ s.t. $\mathbb{P}_{(x,y)\sim\mathcal{D}}[h(x) \neq y] \leq \min_{\tilde{f}\text{ Fourier}-\text{k}-\text{sparse}} \mathbb{P}_{(x,y)\sim\mathcal{D}}\left[\tilde{f}(\boldsymbol{x}) \neq y\right] + \varepsilon$ with high prob.
4. If $V$ interacts with any $P'$, then $V$ accepts and outputs a hypothesis $h$ s.t. $\mathbb{P}_{(x,y)\sim\mathcal{D}}[h(x) \neq y] > \min_{\tilde{f}\text{ Fourier}-\text{k}-\text{sparse}} \mathbb{P}_{(x,y)\sim\mathcal{D}}\left[\tilde{f}(\boldsymbol{x}) \neq y\right] + \varepsilon$ with low prob.

# Close-to-Optimality of our Verification

# Close-to-Optimality of our Verification

Do you find the "$\varepsilon > 2\sqrt{b^2 - a^2}$" and "$\varepsilon > 4k\sqrt{b^2 - a^2}$" a bit weird?

# Close-to-Optimality of our Verification

Do you find the "$\varepsilon > 2\sqrt{b^2 - a^2}$" and "$\varepsilon > 4k\sqrt{b^2 - a^2}$" a bit weird?

We did, so we looked into it for the parity case:

# Close-to-Optimality of our Verification

Do you find the "$\varepsilon > 2\sqrt{b^2 - a^2}$" and "$\varepsilon > 4k\sqrt{b^2 - a^2}$" a bit weird?

We did, so we looked into it for the parity case:

# Close-to-Optimality of our Verification

Do you find the "$\varepsilon > 2\sqrt{b^2 - a^2}$" and "$\varepsilon > 4k\sqrt{b^2 - a^2}$" a bit weird?

We did, so we looked into it for the parity case:

**Theorem:**

The interactive verification guarantees cannot at the same time

# Close-to-Optimality of our Verification

Do you find the "$\varepsilon > 2\sqrt{b^2 - a^2}$" and "$\varepsilon > 4k\sqrt{b^2 - a^2}$" a bit weird?

We did, so we looked into it for the parity case:

**Theorem:**

The interactive verification guarantees cannot at the same time

- work for all $\varepsilon \geq \frac{1}{3}\sqrt{b^2 - a^2}$, and

# Close-to-Optimality of our Verification

Do you find the "$\varepsilon > 2\sqrt{b^2 - a^2}$" and "$\varepsilon > 4k\sqrt{b^2 - a^2}$" a bit weird?

We did, so we looked into it for the parity case:

**Theorem:**

The interactive verification guarantees cannot at the same time

- work for all $\varepsilon \geq \frac{1}{3}\sqrt{b^2 - a^2}$, and
- have a verifier $V$ with $n$-independent sample complexity.

# Close-to-Optimality of our Verification

**Proof idea (sketchy; inspired by [15]):**

[15] S. Mutreja and J. Shafer, *COLT 2023*

# Close-to-Optimality of our Verification

**Proof idea (sketchy; inspired by [15]):**

Classical-quantum pair $(V, P)$

[15] S. Mutreja and J. Shafer, *COLT 2023*

# Close-to-Optimality of our Verification

**Proof idea (sketchy; inspired by [15]):**

Classical-quantum pair $(V, P)$

➢Classical distinguisher between random noisy parities and fully uniform distribution

[15] S. Mutreja and J. Shafer, *COLT 2023*

# Close-to-Optimality of our Verification

**Proof idea (sketchy; inspired by [15]):**

Classical-quantum pair $(V, P)$

- ➢Classical distinguisher between random noisy parities and fully uniform distribution
  - ➢Violating a known $\Omega(n)$ sample complexity lower bound ⚡

[15] S. Mutreja and J. Shafer, *COLT 2023*

# Close-to-Optimality of our Verification

**Proof idea:**

# Close-to-Optimality of our Verification

**Proof idea:**

samples
$(x_i, y_i) \sim \mathcal{D}$

# Close-to-Optimality of our Verification

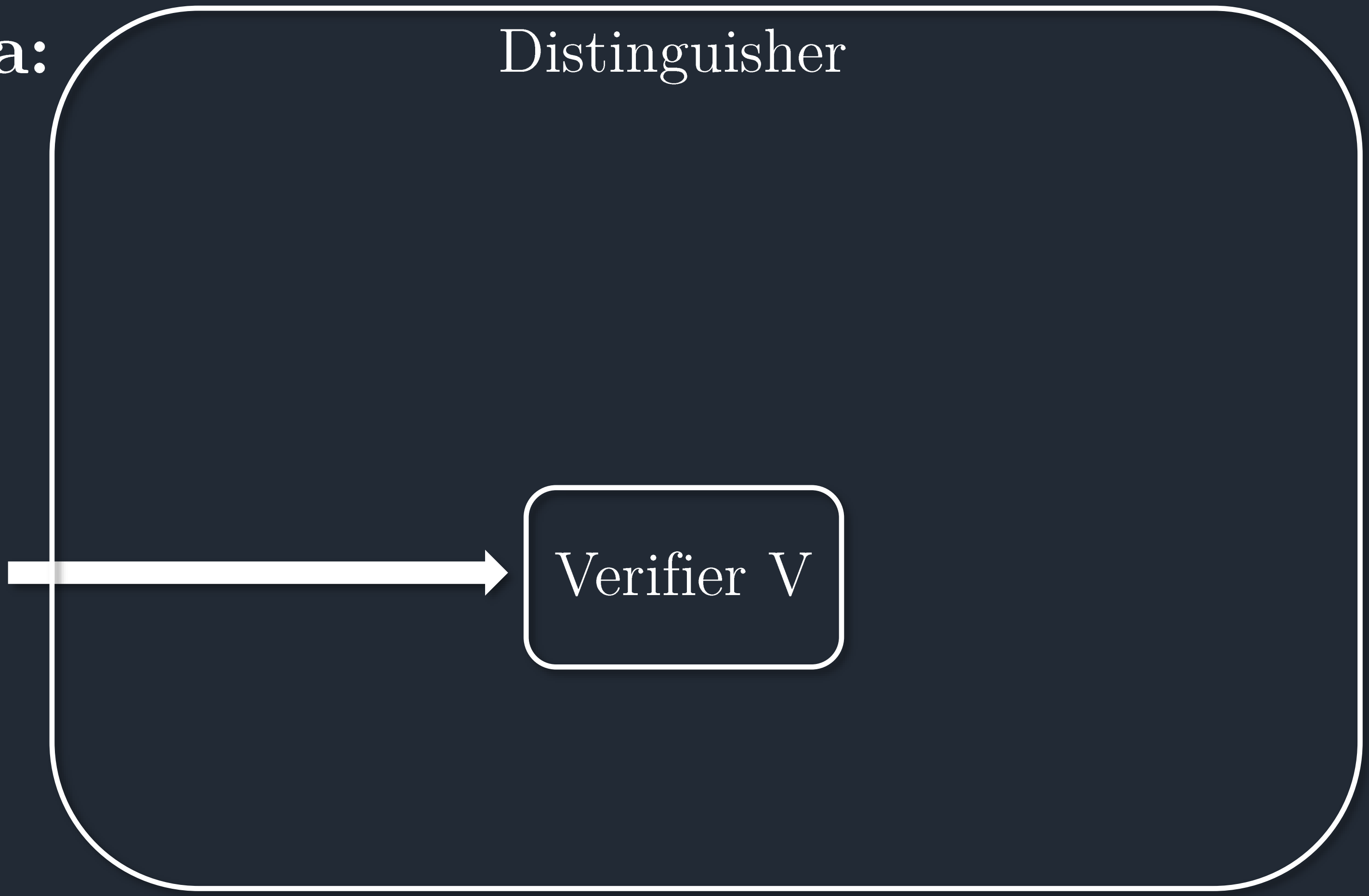
**Proof idea:**

Distinguisher

samples
$(x_i, y_i) \sim \mathcal{D}$

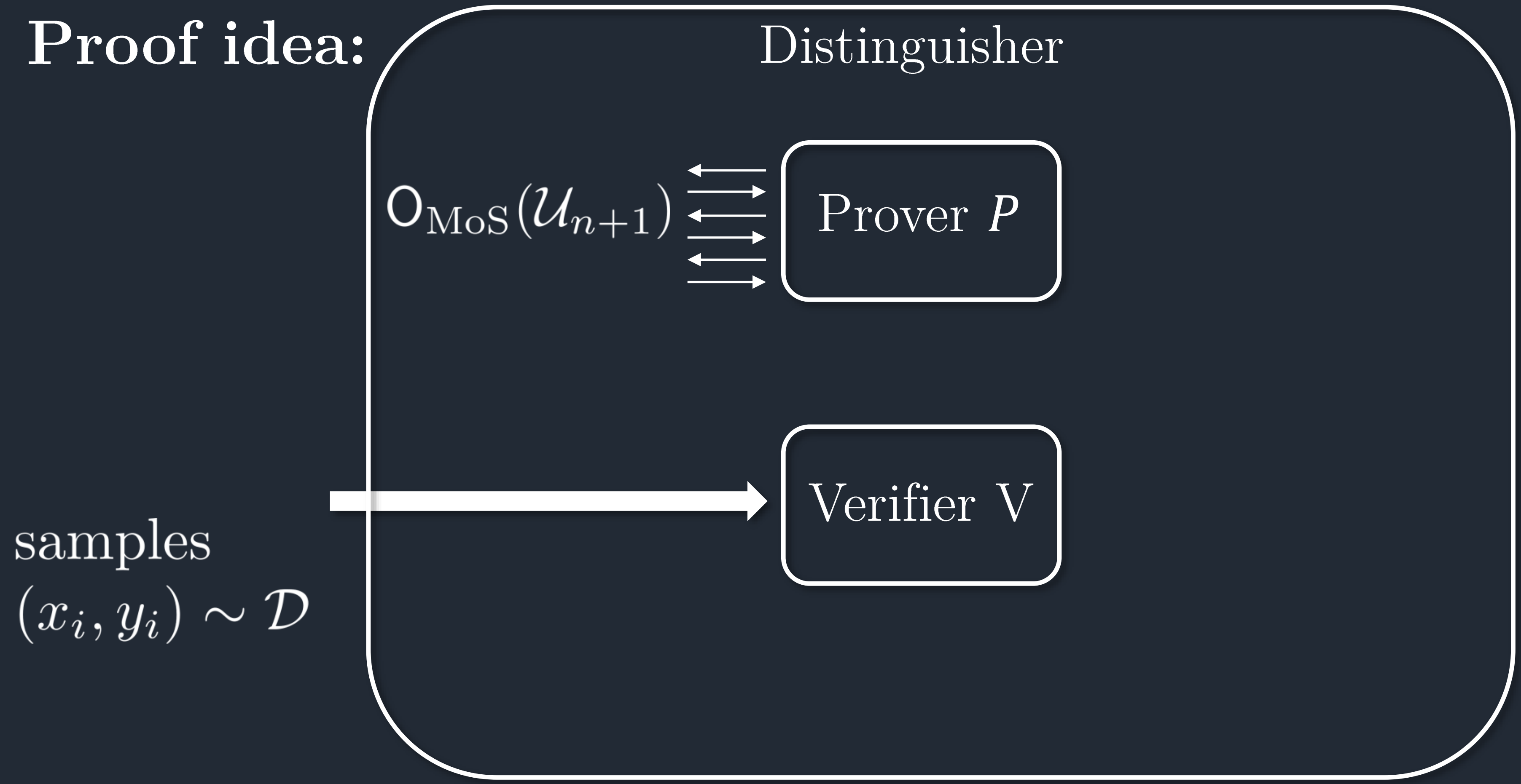
# Close-to-Optimality of our Verification

**Proof idea:**



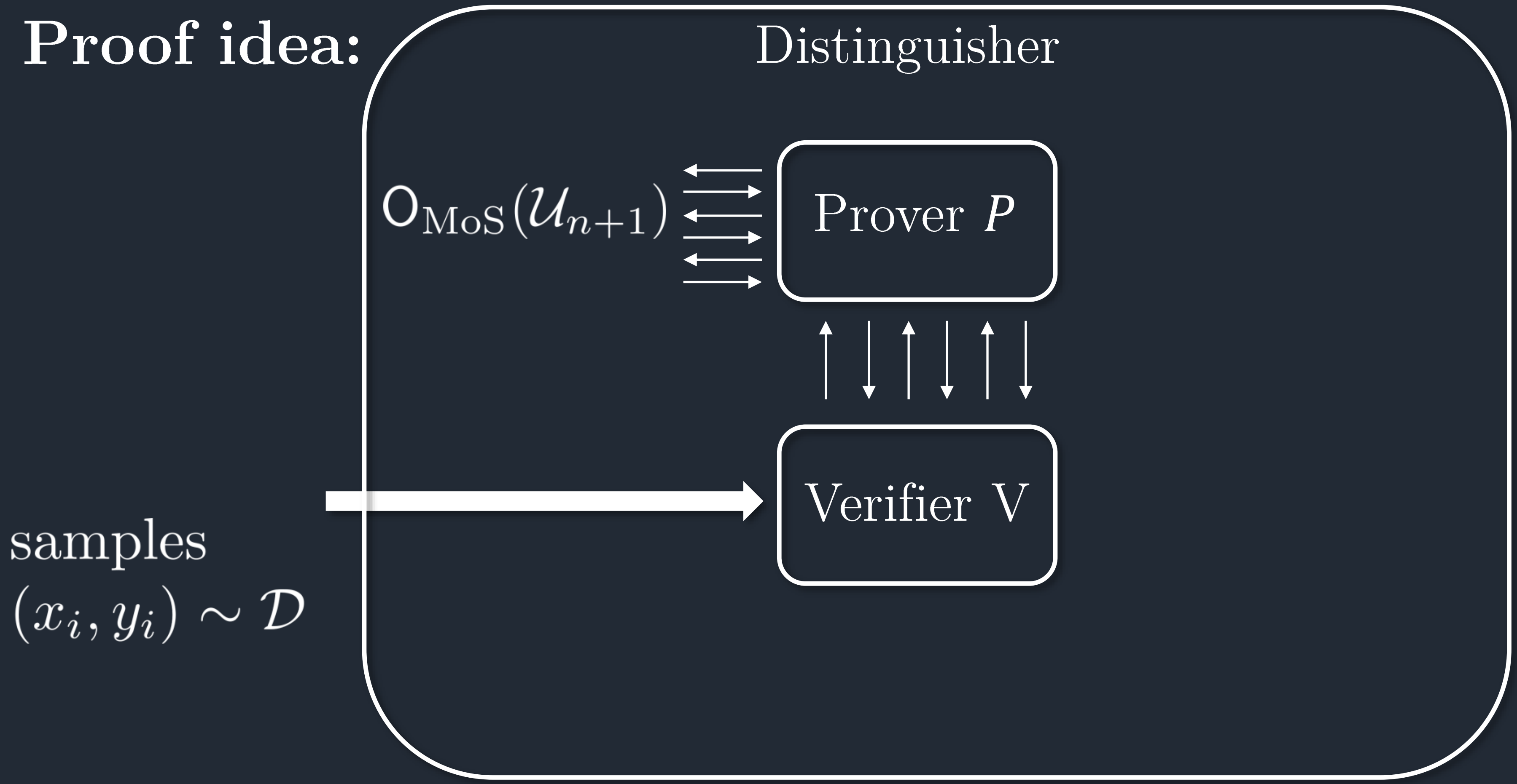
# Close-to-Optimality of our Verification

**Proof idea:**



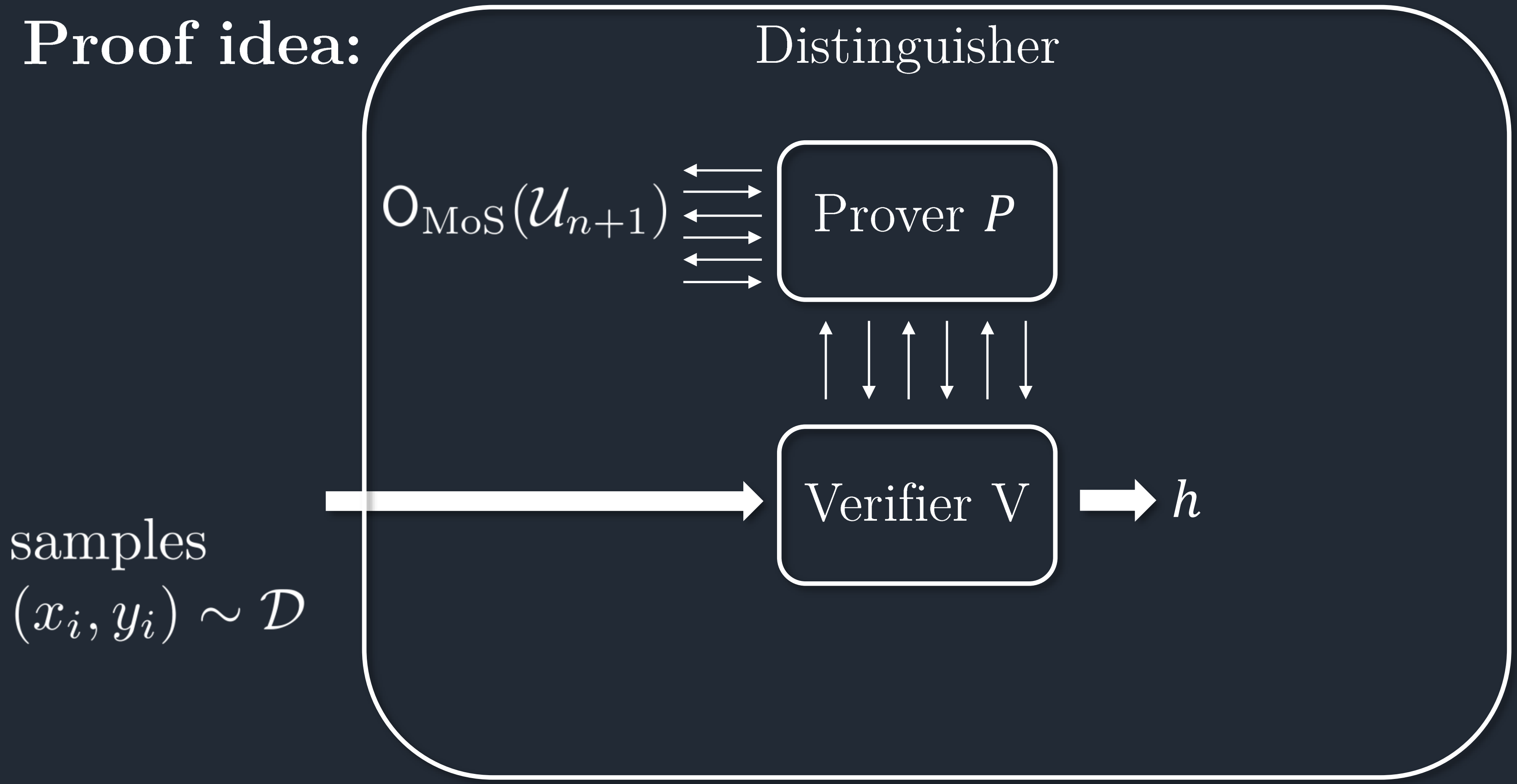
# Close-to-Optimality of our Verification

**Proof idea:**



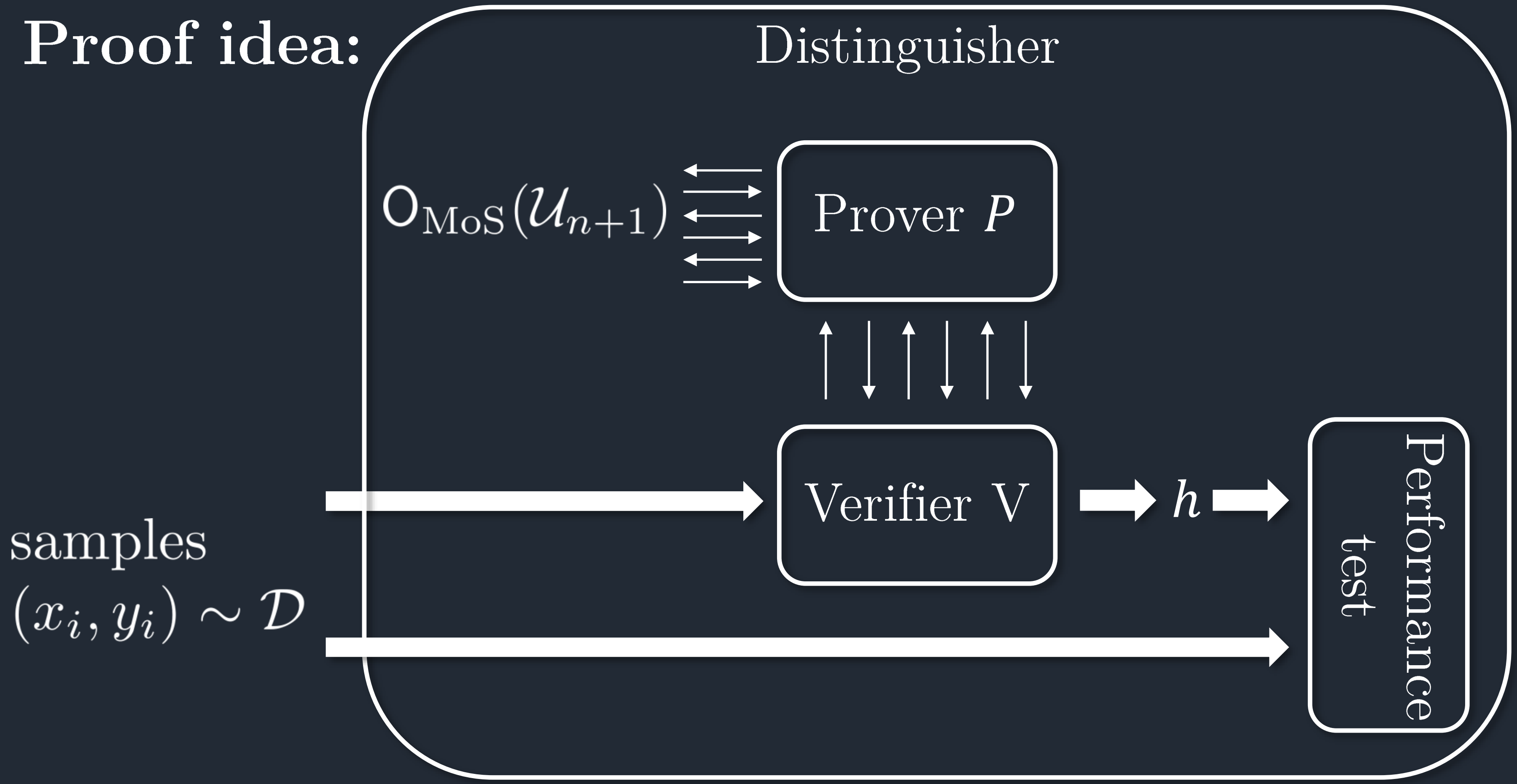
# Close-to-Optimality of our Verification

**Proof idea:**



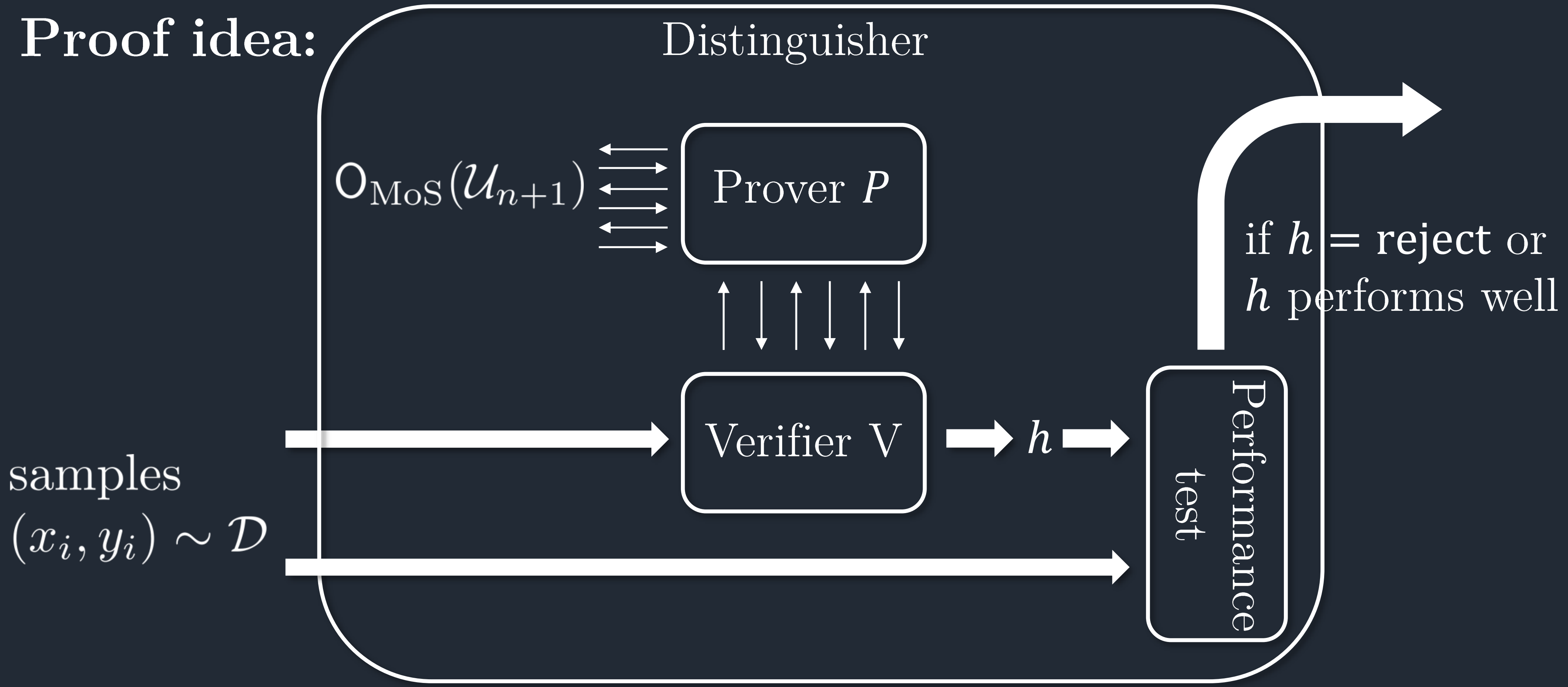
# Close-to-Optimality of our Verification

**Proof idea:**



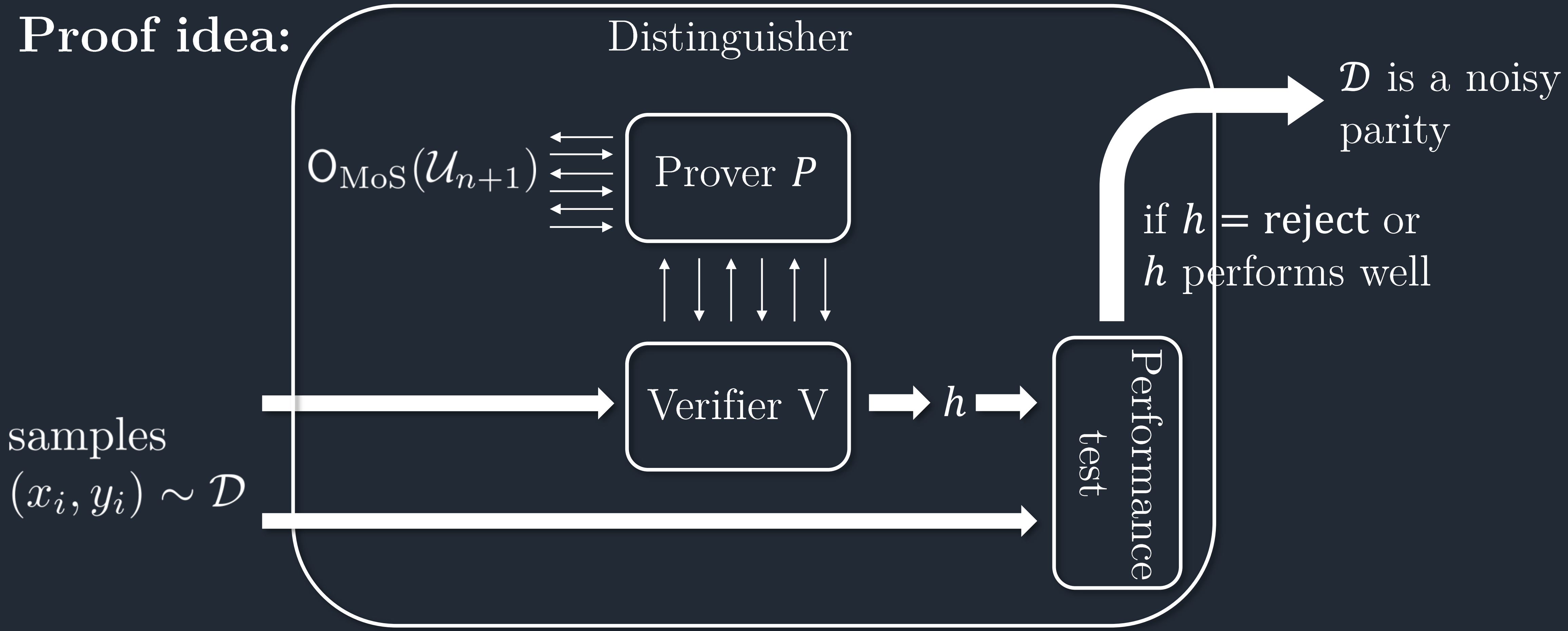
# Close-to-Optimality of our Verification

**Proof idea:**



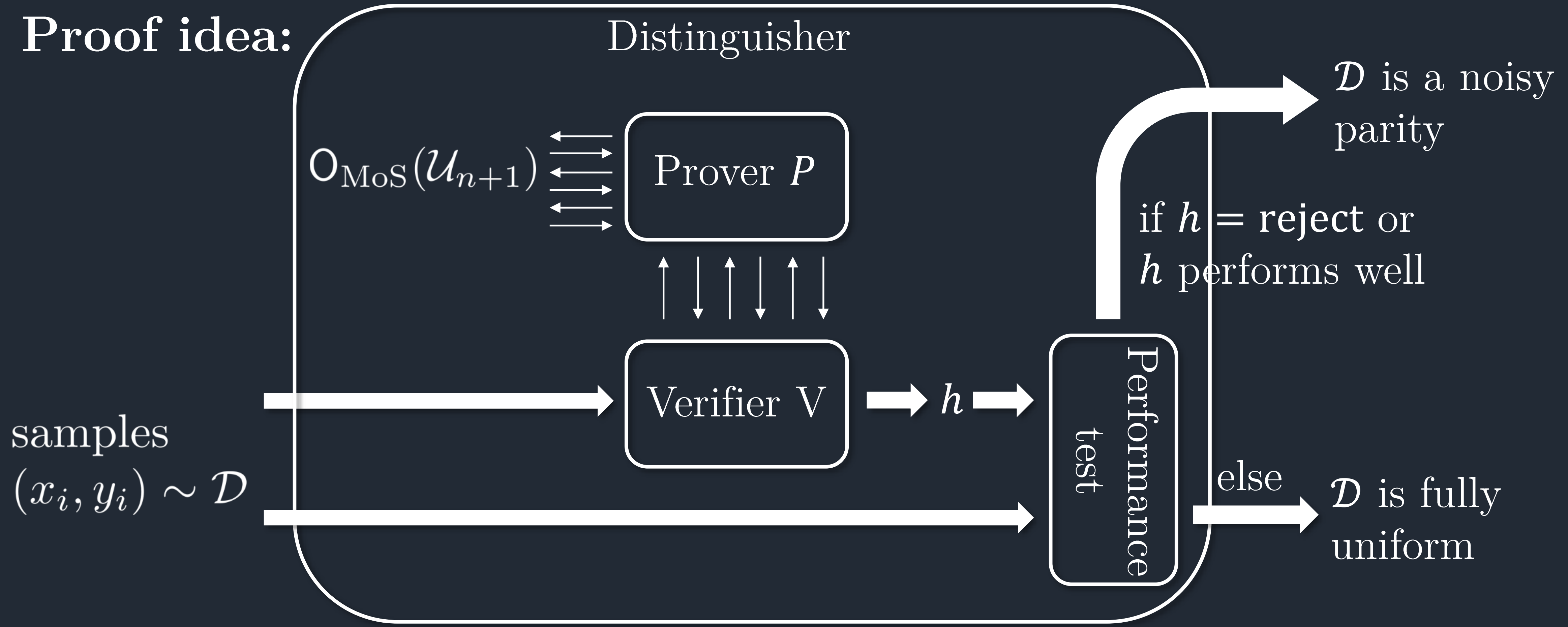
# Close-to-Optimality of our Verification

**Proof idea:**



# Close-to-Optimality of our Verification

**Proof idea:**



# Limitations on the Power of Quantum Data

Why quantum data is not all-powerful

# “Usefulness” of MoS for Distribution-Independent Agnostic Learning

# “Usefulness” of MoS for Distribution-Independent Agnostic Learning

So far: What can MoS examples do?

# "Usefulness" of MoS for Distribution-Independent Agnostic Learning

So far: What can MoS examples do?

Now: What can MoS examples **not** do?

# "Usefulness" of MoS for Distribution-Independent Agnostic Learning

So far: What can MoS examples do?

Now: What can MoS examples **not** do?

**Theorem:**

The quantum sample complexity of distribution-independent agnostic learning from copies of $\rho_{\mathcal{D}}$ is $\widetilde{\Theta}\left(\frac{\mathrm{VCdim} + \log(1/\delta)}{\varepsilon^2}\right)$.

# "Usefulness" of MoS for Distribution-Independent Agnostic Learning

So far: What can MoS examples do?

Now: What can MoS examples **not** do?

**Theorem:**

The quantum sample complexity of distribution-independent agnostic learning from copies of $\rho_{\mathcal{D}}$ is $\widetilde{\Theta}\left(\frac{\text{VCdim} + \log(1/\delta)}{\varepsilon^2}\right)$.

- This equals the classical sample complexity up to log-factors.

# "Usefulness" of MoS for Distribution-Independent Agnostic Learning

So far: What can MoS examples do?

Now: What can MoS examples **not** do?

**Theorem:**

The quantum sample complexity of distribution-independent agnostic learning from copies of $\rho_{\mathcal{D}}$ is $\widetilde{\Theta}\left(\frac{\mathrm{VCdim} + \log(1/\delta)}{\varepsilon^2}\right)$.

- This equals the classical sample complexity up to log-factors.
- This is (almost) the same limitation as for superposition examples [16].

[16] S. Arunachalam and R. de Wolf; *JMLR 19.1, 2879–2878 (2018)*

# “Usefulness” of MoS for Distribution-Independent Agnostic Learning

# “Usefulness” of MoS for Distribution-Independent Agnostic Learning

**Proof Idea:** We write $d = \texttt{VCdim}$, let $m$ be the number of copies.

# "Usefulness" of MoS for Distribution-Independent Agnostic Learning

**Proof Idea:** We write $d = \mathsf{VCdim}$, let $m$ be the number of copies.

1. Start from a shattered set $S = \{x_1, \ldots, x_d\}$, construct a set of $2^d$ distributions $\mathcal{D}_a, a \in \{0,1\}^d$, over $S \times \{0,1\}$ such that:

# "Usefulness" of MoS for Distribution-Independent Agnostic Learning

**Proof Idea:** We write $d = \mathsf{VCdim}$, let $m$ be the number of copies.

1. Start from a shattered set $S = \{x_1, \ldots, x_d\}$, construct a set of $2^d$ distributions $\mathcal{D}_a, a \in \{0,1\}^d$, over $S \times \{0,1\}$ such that:
   a) All $\mathcal{D}_a$ are "close-to-uniform" over $S \times \{0,1\}$.

# "Usefulness" of MoS for Distribution-Independent Agnostic Learning

**Proof Idea:** We write $d = \mathsf{VCdim}$, let $m$ be the number of copies.

1. Start from a shattered set $S = \{x_1, \ldots, x_d\}$, construct a set of $2^d$ distributions $\mathcal{D}_a, a \in \{0,1\}^d$, over $S \times \{0,1\}$ such that:
   a) All $\mathcal{D}_a$ are "close-to-uniform" over $S \times \{0,1\}$.
   b) A successful agnostic learner in particular has to be able to approximate $a$ in Hamming distance using $m$ copies of $\rho_{\mathcal{D}_a}$.

# "Usefulness" of MoS for Distribution-Independent Agnostic Learning

**Proof Idea:** We write $d = \mathsf{VCdim}$, let $m$ be the number of copies.

1. Start from a shattered set $S = \{x_1, \ldots, x_d\}$, construct a set of $2^d$ distributions $\mathcal{D}_a, a \in \{0,1\}^d$, over $S \times \{0,1\}$ such that:
   a) All $\mathcal{D}_a$ are "close-to-uniform" over $S \times \{0,1\}$.
   b) A successful agnostic learner in particular has to be able to approximate $a$ in Hamming distance using $m$ copies of $\rho_{\mathcal{D}_a}$.

➢ $\mathsf{QMI}(A; \rho_{\mathcal{D}_A}^{\otimes m}) \geq \Omega\big(\log\big(2^d\big)\big) = \Omega(d)$

# “Usefulness” of MoS for Distribution-Independent Agnostic Learning

**Proof Idea:** We write $d = \mathtt{VCdim}$, let $m$ be the number of copies.

*1.* $\mathrm{QMI}(A; \rho_{\mathcal{D}_A}^{\otimes m}) \geq \Omega(\log(2^d)) = \Omega(d)$

# "Usefulness" of MoS for Distribution-Independent Agnostic Learning

**Proof Idea:** We write $d = \mathtt{VCdim}$, let $m$ be the number of copies.

1. $\mathrm{QMI}(A; \rho_{\mathcal{D}_A}^{\otimes m}) \geq \Omega\big(\log\big(2^d\big)\big) = \Omega(d)$
2. $\mathrm{QMI}(A; \rho_{\mathcal{D}_A}^{\otimes m}) \leq m \cdot \mathrm{QMI}(A; \rho_{\mathcal{D}_A})$

# "Usefulness" of MoS for Distribution-Independent Agnostic Learning

**Proof Idea:** We write $d = \mathtt{VCdim}$, let $m$ be the number of copies.

1. $\mathrm{QMI}(A; \rho_{\mathcal{D}_A}^{\otimes m}) \geq \Omega\left(\log\left(2^d\right)\right) = \Omega(d)$
2. $\mathrm{QMI}(A; \rho_{\mathcal{D}_A}^{\otimes m}) \leq m \cdot \mathrm{QMI}(A; \rho_{\mathcal{D}_A})$
3. $\mathrm{QMI}(A; \rho_{\mathcal{D}_A}) \leq \mathcal{O}(\varepsilon^2 \log(d))$

# "Usefulness" of MoS for Distribution-Independent Agnostic Learning

**Proof Idea:** We write $d = \mathtt{VCdim}$, let $m$ be the number of copies.

1. $\mathrm{QMI}(A; \rho_{\mathcal{D}_A}^{\otimes m}) \geq \Omega\big(\log\big(2^d\big)\big) = \Omega(d)$
2. $\mathrm{QMI}(A; \rho_{\mathcal{D}_A}^{\otimes m}) \leq m \cdot \mathrm{QMI}(A; \rho_{\mathcal{D}_A})$
3. $\mathrm{QMI}(A; \rho_{\mathcal{D}_A}) \leq \mathcal{O}(\varepsilon^2 \log(d))$
   - This step uses the specific distributions $\mathcal{D}_a$ and involves a detailed eigenvalue analysis for the resulting MoS states.

# “Uselessness” of MoS for Verifying Distribution-Independent Agnostic Learning

So, MoS don’t help with distribution-independent agnostic learning...

# "Uselessness" of MoS for Verifying Distribution-Independent Agnostic Learning

So, MoS don't help with distribution-independent agnostic learning...

How much can they help a verifier interacting with an MoS prover?

# “Uselessness” of MoS for Verifying Distribution-Independent Agnostic Learning

So, MoS don’t help with distribution-independent agnostic learning...

How much can they help a verifier interacting with an MoS prover?

**Theorem (Consequence of [15]):**
The sample complexity of a verifier interacting with an MoS prover for distribution-independent agnostic learning is at most quadratically better than what the verifier can achieve alone.

[15] S. Mutreja and J. Shafer, *COLT 2023*

# "Uselessness" of MoS for Verifying Distribution-Independent Agnostic Learning

So, MoS don't help with distribution-independent agnostic learning...

How much can they help a verifier interacting with an MoS prover?

**Theorem (Consequence of [15]):**
The sample complexity of a verifier interacting with an MoS prover for distribution-independent agnostic learning is at most quadratically better than what the verifier can achieve alone.

[15] showed that interacting with a classical prover can also lead to an at most quadratic sample complexity improvement.

[15] S. Mutreja and J. Shafer, *COLT 2023*

# Conclusion and Outlook

What we talked about and where to go from here

# Guiding Question

**Question:**

Is there an (agnostic) learning problem that simultaneously...

1. ...is intractable for classical learners but efficiently solvable by quantum learners using quantum data,
2. ...and can be efficiently verifiably delegated from a classical verifier to an untrusted quantum prover?

# Summary

# Summary

- We proposed the *mixture-of-superpositions (MoS)* quantum oracle and demonstrated its power and limitations in agnostic learning.

# Summary

- We proposed the *mixture-of-superpositions (MoS)* quantum oracle and demonstrated its power and limitations in agnostic learning.

| | Oracle type | Problem type | | | |
|---|---|---|---|---|---|
| | | Fourier sampling | Heavy Fourier coefficient estimation | 1-agnostic parity learning | 2-agnostic Fourier-sparse learning |
| Functional | superposition examples = mixture-of-superpositions | ✓ | ✓ | ✓ | ✓ |
| Functional | superposition QSQ = mixture-of-superpositions QSQ | Probably ✗ | ✓ | ✓ | ✓ |
| Distributional | superposition examples | ? | ? | ? | ? |
| Distributional | superposition QSQ | Probably ✗ | ? | ? | ? |
| Distributional | mixture-of-superpositions | ✓ | ✓ | ✓ | ✓ |
| Distributional | mixture-of-superpositions QSQ | Probably ✗ | ✓ | ✓ | ✓ |

# Summary

- We proposed the *mixture-of-superpositions (MoS)* quantum oracle and demonstrated its power and limitations in agnostic learning.

- We identified learning problems that

# Summary

- We proposed the *mixture-of-superpositions (MoS)* quantum oracle and demonstrated its power and limitations in agnostic learning.

- We identified learning problems that
  - are believed to be classically hard,

# Summary

- We proposed the *mixture-of-superpositions (MoS)* quantum oracle and demonstrated its power and limitations in agnostic learning.

- We identified learning problems that
  - are believed to be classically hard,
  - but that become efficiently solvable for a classical verifier interacting with an untrusted quantum prover.

# Open Questions

# Open Questions

- Provable separation between superposition examples and mixture-of-superpositions examples for agnostic learning?

# Open Questions

- Provable separation between superposition examples and mixture-of-superpositions examples for agnostic learning?

- Verification in other quantum learning scenarios, e.g., distribution learning or learning with vs. w/o quantum memory (-> ongoing work)?

# Open Questions

- Provable separation between superposition examples and mixture-of-superpositions examples for agnostic learning?

- Verification in other quantum learning scenarios, e.g., distribution learning or learning with vs. w/o quantum memory (-> ongoing work)?

- NISQ-friendly variants of our framework, e.g., for variational quantum machine learning or for NISQ verifiers?

# Your Questions