How to learn your quantum state (and how not to)

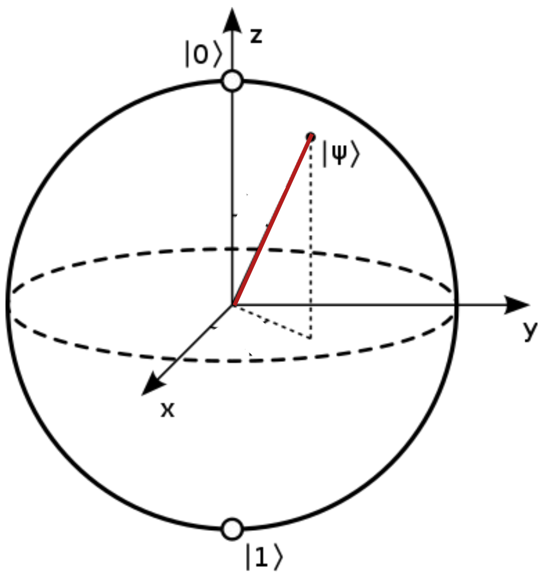
Yihui Quek

Stanford → Berlin

Freie Universität Berlin

quekpottheories

yihuiquek3.14@gmail.com

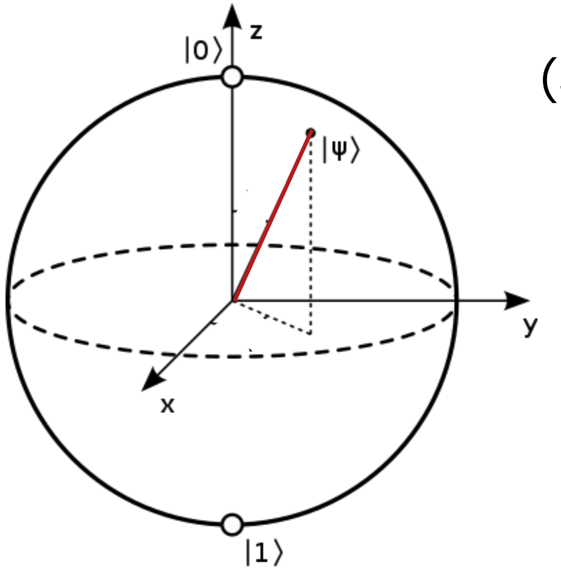


Quantum learning primer: quantum state tomography

Quantum State Tomography

a fundamental task in any quantum experiment

Task (pure-state version): Given physical system in unknown quantum state $|\psi\rangle$, estimate $|\psi\rangle$.



(Single-qubit intuition)

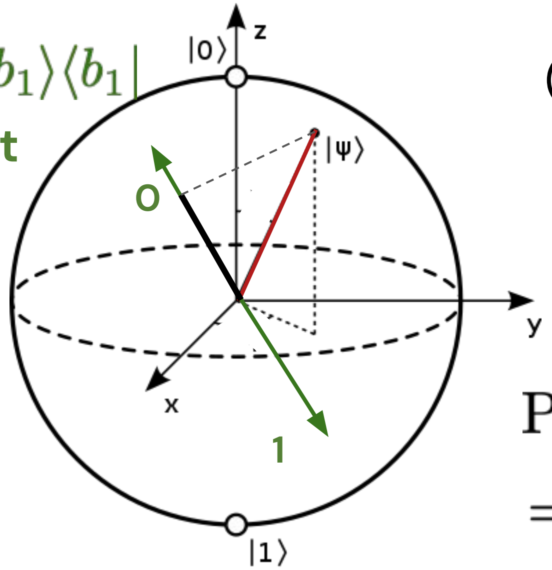
Quantum State Tomography

a fundamental task in any quantum experiment

Task (pure-state version): Given physical system in unknown quantum state $|\psi\rangle$, estimate $|\psi\rangle$ from results of repeated measurements.

$E_1 = |b_1\rangle\langle b_1|$

measurement operator



(Single-qubit intuition)

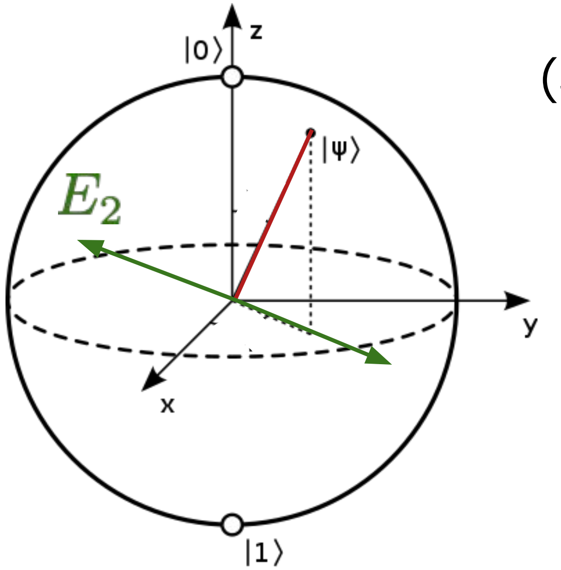
Measurement outcomes are probabilistic!

$$\Pr(\text{measurement} \to 0) = |\langle\psi|b_1\rangle|^2$$

Quantum State Tomography

a fundamental task in any quantum experiment

Task (pure-state version): Given physical system in unknown quantum state $|\psi\rangle$, estimate $|\psi\rangle$ from results of repeated measurements.



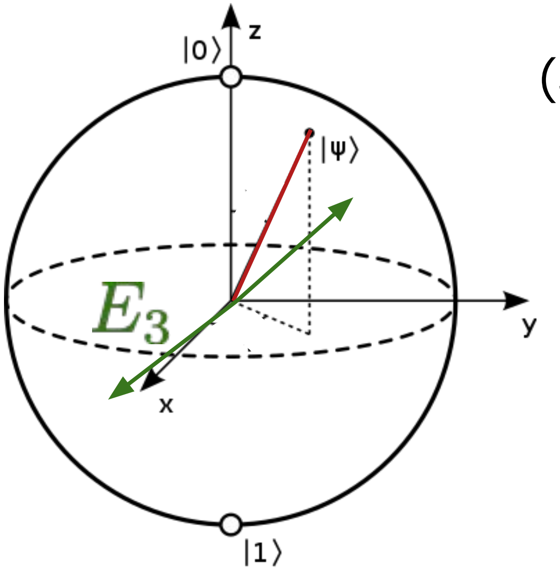
(Single-qubit intuition)

Measurement outcomes are probabilistic!

Quantum State Tomography

a fundamental task in any quantum experiment

Task (pure-state version): Given physical system in unknown quantum state $|\psi\rangle$, estimate $|\psi\rangle$ from results of repeated measurements.



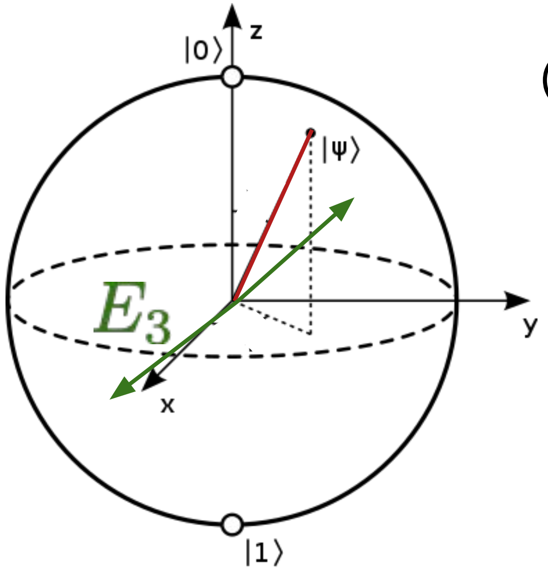
(Single-qubit intuition)

Measurement outcomes are probabilistic!

Quantum State Tomography

a fundamental task in any quantum experiment

Task (pure-state version): Given physical system in unknown quantum state $|\psi\rangle$, estimate $|\psi\rangle$ from results of repeated measurements.



(Single-qubit intuition)

Measurement outcomes are probabilistic!

Then process all collected measurement outcomes to get an estimate.

Quantum State Tomography

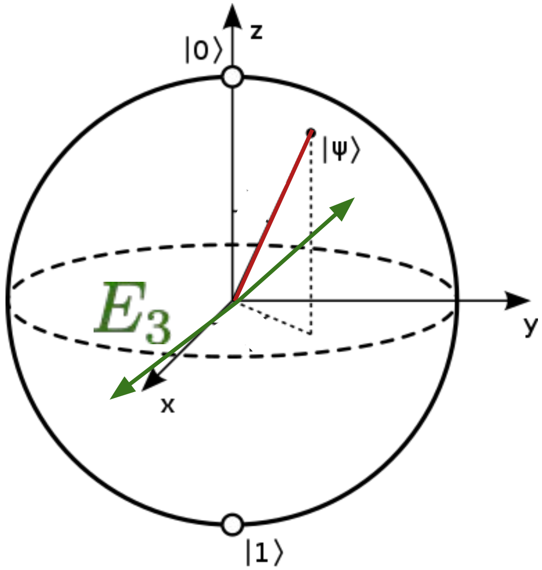
a fundamental task in any quantum experiment

Task (pure-state version): Given physical system in unknown quantum state $|\psi\rangle$, estimate $|\psi\rangle$ from results of repeated measurements.

(Single-qubit intuition)

For many qubits/mixed states, same story.

Problem: Need many measurements



Then process all collected measurement outcomes to get an estimate.

Full quantum state tomography is sample-inefficient.

Full quantum state tomography is sample-inefficient.

Task:
Given identical copies of an unknown n-qubit quantum state ρ, approximate the full density matrix of ρ (2^n x 2^n).

Full quantum state tomography is sample-inefficient.

Task:
Given identical copies of an unknown n-qubit quantum state ρ, approximate the full density matrix of ρ (2^n x 2^n).

copies of the state that need to be prepared and measured

Full quantum state tomography is sample-inefficient.

Task:
Given identical copies of an unknown n-qubit quantum state ρ, approximate the full density matrix of ρ (2^n x 2^n).

[HHJ+, OW STOC'16]

copies of the state that need to be prepared and measured

$2^{O(n)}$ necessary and sufficient.

Bad for experiments: **exponential in n** (number of qubits)!

Full quantum state tomography is sample-inefficient.

Are there efficient ways to learn quantum states?

Exponential in
n (number of qubits)

Full quantum state tomography is sample-inefficient.

Are there efficient ways to learn quantum states?

Exponential in
n (number of qubits)

Ideas:

1) Assume the quantum circuit that produces the state* is “simple“.
   a) Clifford circuits?
   b) Easily simulable → easily learnable?

*Actually, output distributions of circuits

Full quantum state tomography is sample-inefficient.

Are there efficient ways to learn quantum states?

Exponential in
n (number of qubits)

Ideas:

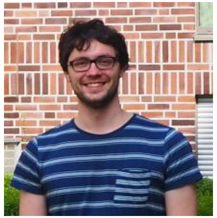
1) Assume the quantum circuit that produces the state* is "simple".
   a) Clifford circuits?
   b) Easily simulable → easily learnable?

2) Change the definition of `learning', so that one only needs to learn `some properties' of the state.

Aaronson, 2007. The learnability of quantum states *Proc. R. Soc. A.*4633089–3114

*Actually, output distributions of circuits

Part I





Learnability of the output distributions of local quantum circuits

Jonas Haferkamp*
Freie Universität Berlin
Helmholtz-Zentrum Berlin
jonas.haferkamp@t-online.de

Jens Eisert
Freie Universität Berlin
Helmholtz-Zentrum Berlin
Fraunhofer Heinrich Hertz Institute
jense@zedat.fu-berlin.de

Dominik Hangleiter
University of Maryland
NIST
dominik.hangleiter@gmail.com

Marios Ioannou
Freie Universität Berlin
marios_ioannou@outlook.com

Alexander Nietner
Freie Universität Berlin
a.nietner@fu-berlin.de

Yihui Quek
Freie Universität Berlin
Stanford University
yquek@stanford.edu

Marcel Hinsche
Freie Universität Berlin
hinsche.marcel@gmail.com

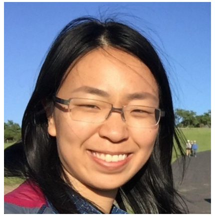
Jean-Pierre Seifert
Technical University Berlin
FhG SIT Darmstadt
jeanpierreseifert@gmail.com

Ryan Sweke
Freie Universität Berlin
rsweke@gmail.com

November 27, 2021

Abstract

There is currently a large interest in understanding the potential advantages quantum devices can offer for probabilistic modelling. To this end, in this work we investigate the probably approximately correct (PAC) learnability of the discrete distributions obtained by measuring, in the computational basis, the output states of local quantum circuits. More specifically, we study both (a) generator learning and (b) evaluator learning. For both problems, one is given some type of oracle access to the

arXiv: 2110.05517 + ongoing work

Quantum Born machines: the next big thing in QML?

Quantum Born machines: the next big thing in QML?

PHYSICAL REVIEW A

covering atomic, molecular, and optical physics and quantum information

Highlights Recent Accepted Collections Authors Referees Search Press About Staff

Access by Ecole Normale Superieure Lyon Go Mobile »

Differentiable learning of quantum circuit Born machines

Jin-Guo Liu and Lei Wang

Phys. Rev. A **98**, 062324 – Published 19 December 2018

10

More

Article

nature › npj quantum information › articles › a

Article | Open Access | Published: 27 May 2019

A generative modeling a benchmarking and traini quantum circuits

Marcello Benedetti, Delfina Garcia-Pintos, Oscar Perdomo, Vicente Leyton-Ortega, Yunseong Nam & Alejandro Perdomo-Ortiz

npj Quantum Information **5**, Article number: 45 (2019) | Cite this article

10k Accesses | **94** Citations | **19** Altmetric | Metrics

nature › npj quantum information › articles › article

Article | Open Access | Published: 08 July 2020

The Born supremacy: quantum advantage and training of an Ising Born machine

Brian Coyle, Daniel Mills, Vincent Danos & Elham Kashefi

npj Quantum Information **6**, Article number: 60 (2020) | Cite this article

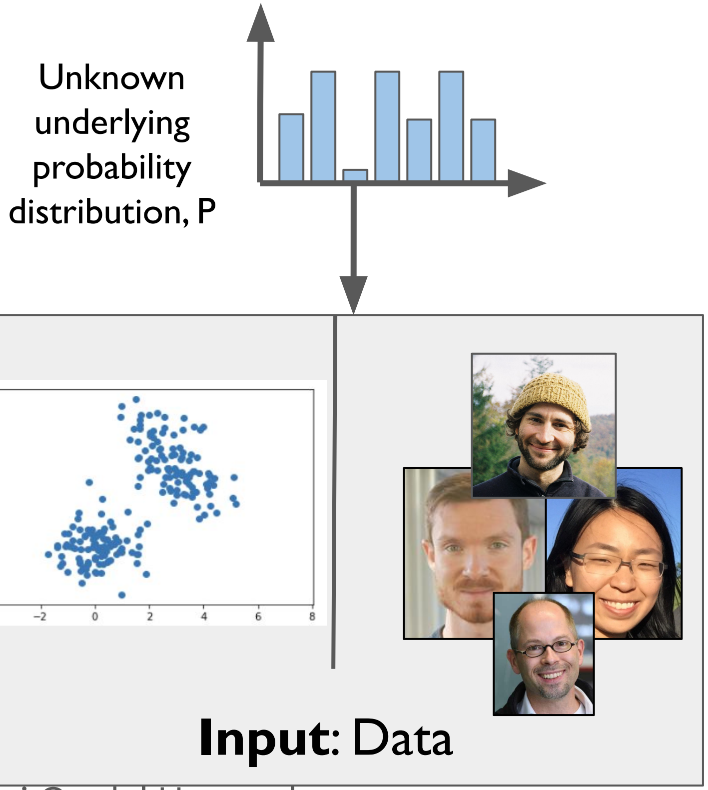
6288 Accesses | **31** Citations | **19** Altmetric | Metrics

Quantum Born machines are a type of generative model.

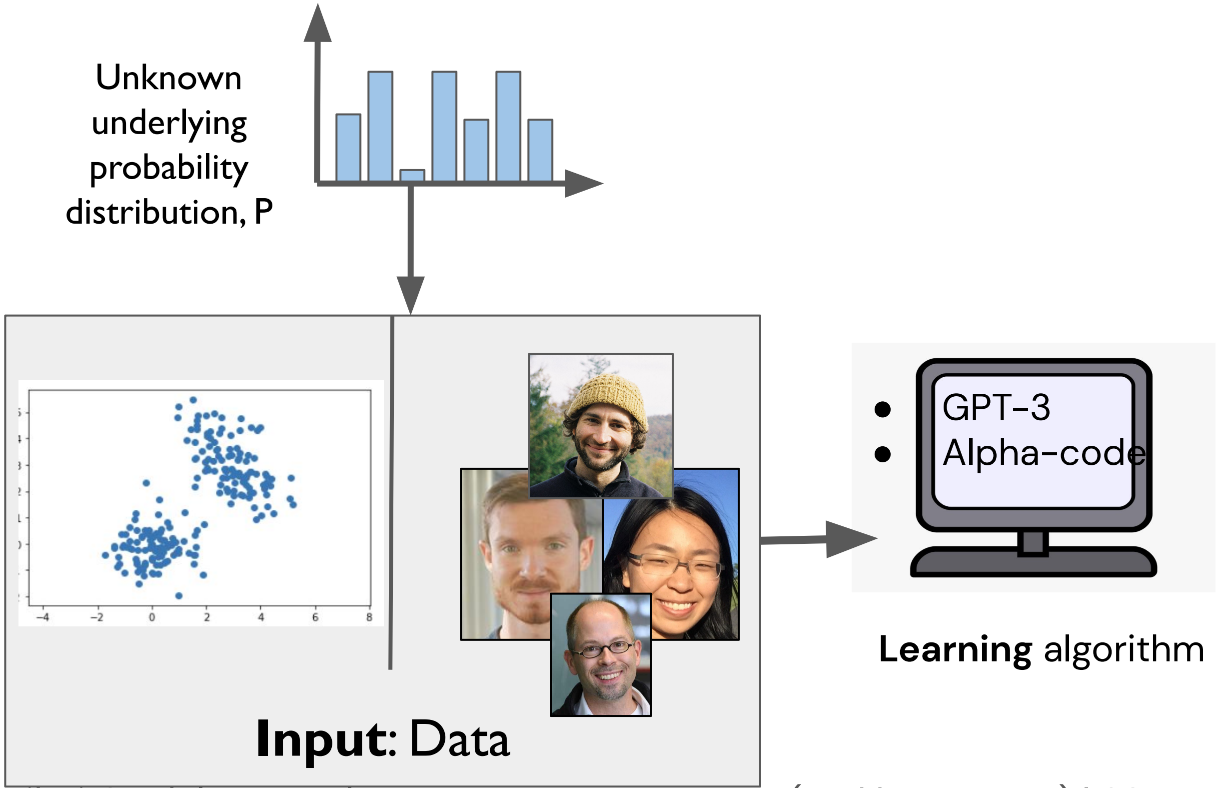
Generative modelling: learning to generate samples from an unknown distribution

More formally in: [KMR+94] Kearns et al. On the learnability of discrete distributions. 1994

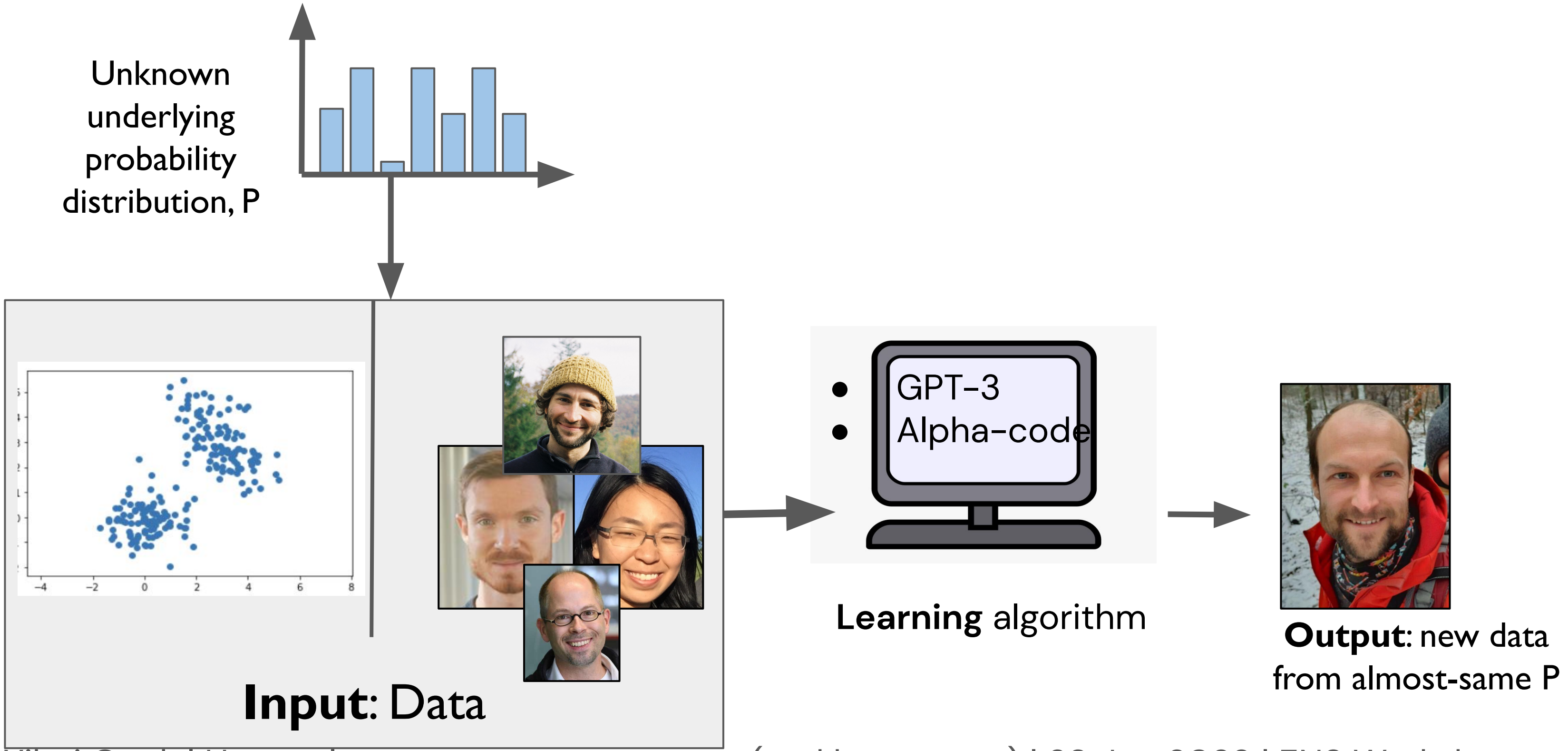
Generative modelling (classical) – real footage



Generative modelling (classical) – real footage

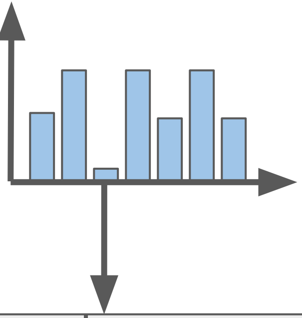


Generative modelling (classical) – real footage

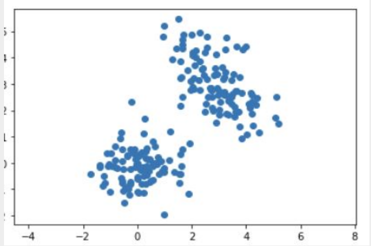


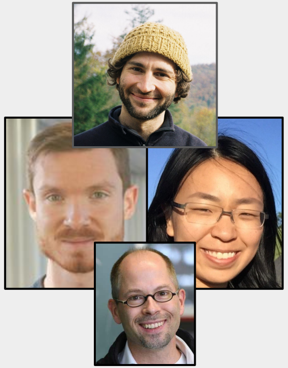
Generative modelling – quantum possibilities

Unknown underlying probability distribution, P

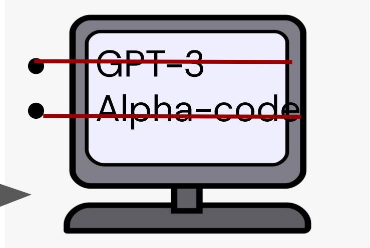


- **Quantum GAN**
- **Boltzmann machine**
- **Hidden markov model**





Input: Data

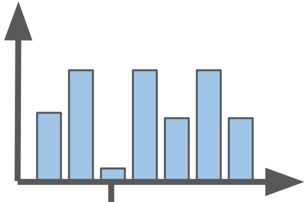


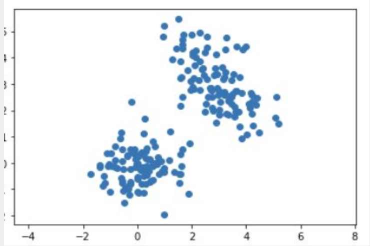
Quantum
Learning algorithm

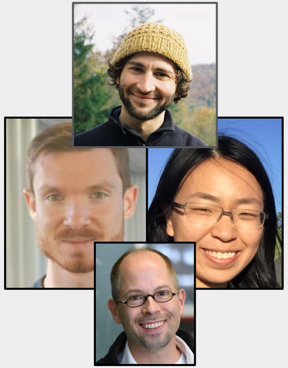
Output: new data from almost-same P

Generative modelling – quantum possibilities

Unknown underlying probability distribution, P



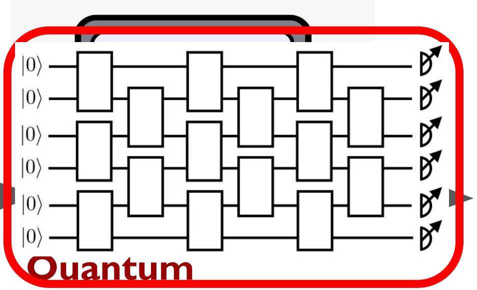




Input: Data

QML proposal:
- **Parametrized Quantum circuit**

= Quantum Circuit Born Machine



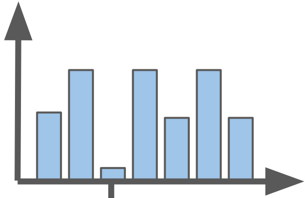
Learning algorithm

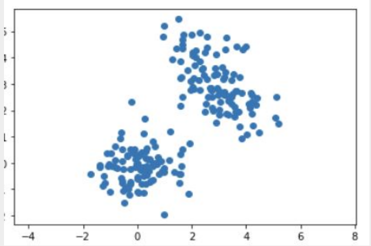
Output: new data from almost-same P

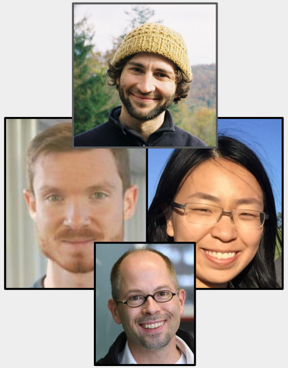
Generative modelling – quantum po ilities

QUANTUM ADVANTAGE?

Unknown underlying probability distribution, P



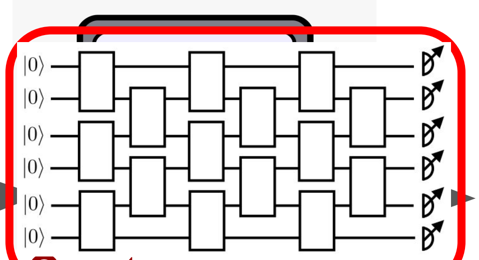




Input: Data

Q posal:

- Parametrized Quantum circuit



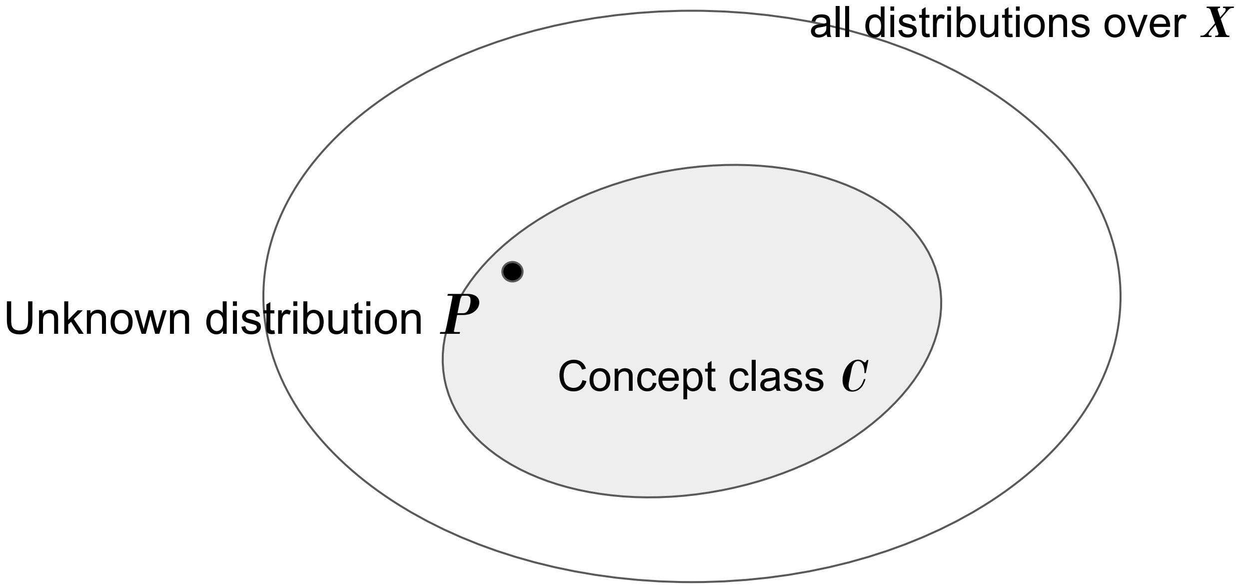
Quantum Learning algorithm

[CMDK20] Coyle et al., npj Quantum Information 2020.
[NDL+20] Niu, arXiv:2010.11983, 2020.

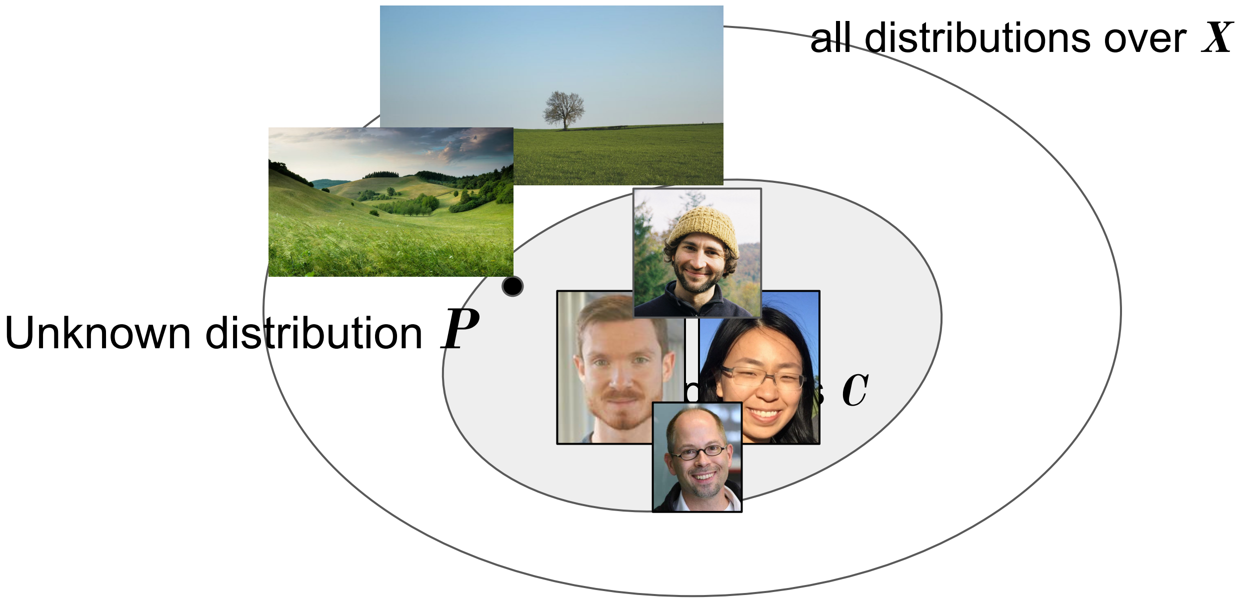
= Quantum Circuit Born Machine

Output: new data from almost-same P

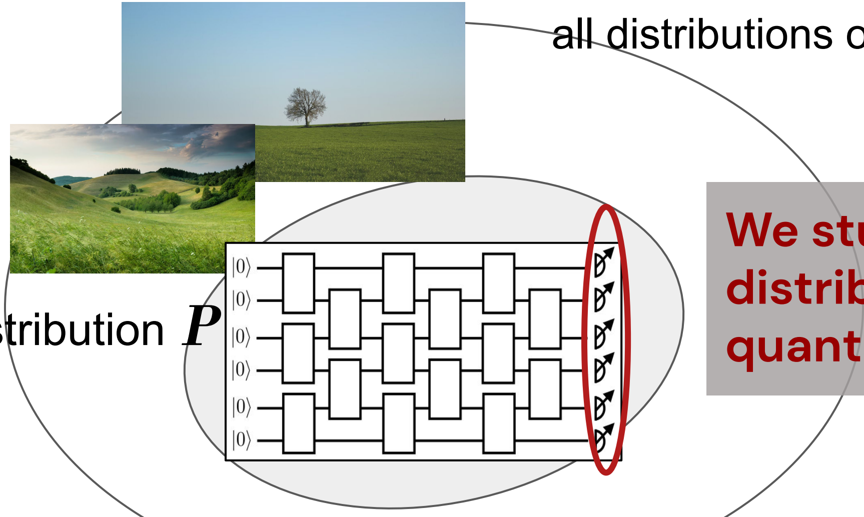
Observe that the class of distributions C is an important prior



Observe that the class of distributions C is an important prior



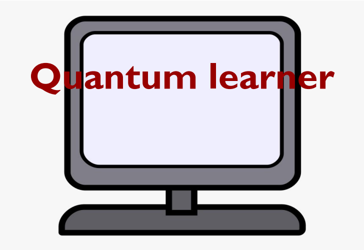
Observe that the class of distributions C is an important prior



We study C = output distributions of local quantum circuits.

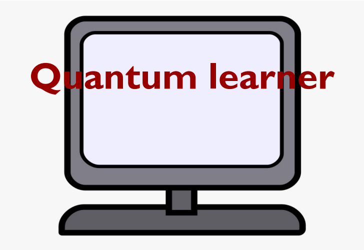
The idea: if quantum learner has an advantage in learning a ***quantum circuit***, can hope for a *generic* quantum advantage.

Our setting: quantum generative modelling



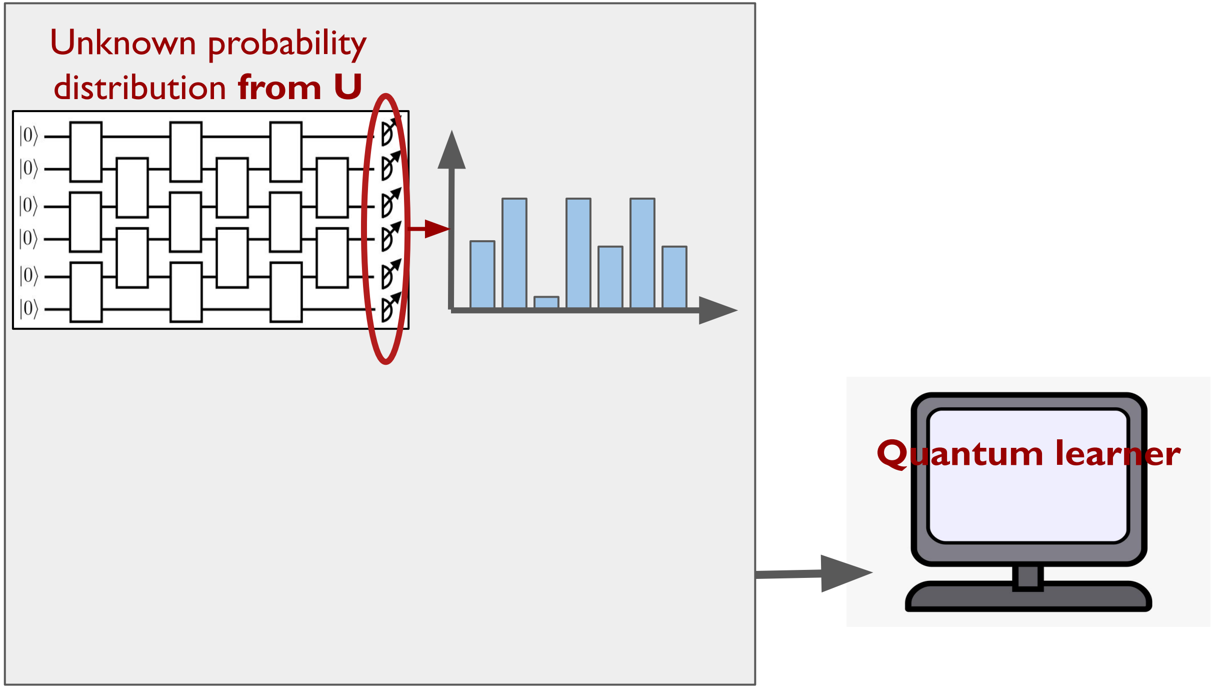
Our setting: quantum generative modelling

NEW input: Not just *any old* probability distribution, but…



Our setting: quantum generative modelling

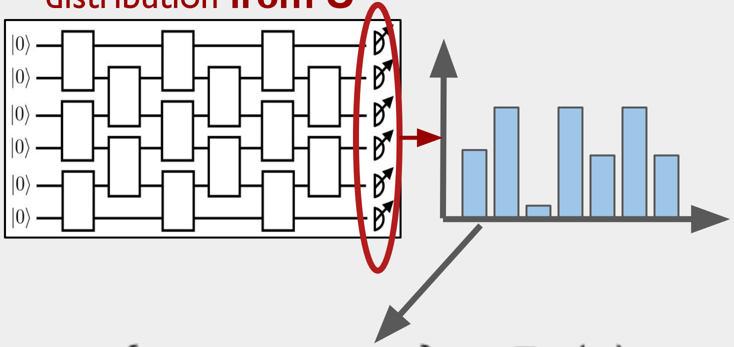
NEW input: Not just *any old* probability distribution, but…



Our setting: quantum generative modelling

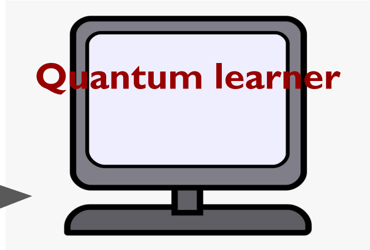
NEW input: Not just *any old* probability distribution, but…

Unknown probability distribution **from U**



$$\{x_1, x_2, \ldots, x_k\} \sim P_U(x)$$

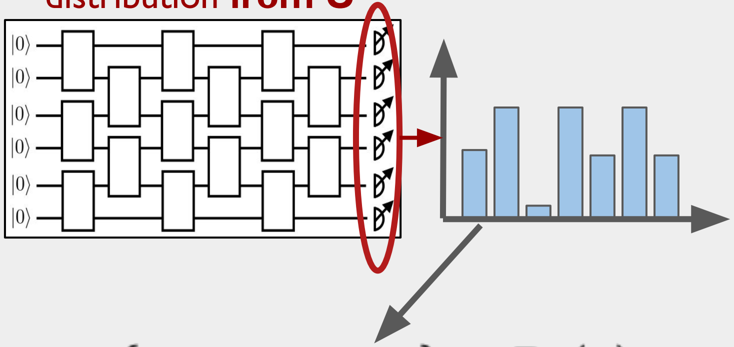
Input: Samples **from a quantum circuit**



Our setting: quantum generative modelling

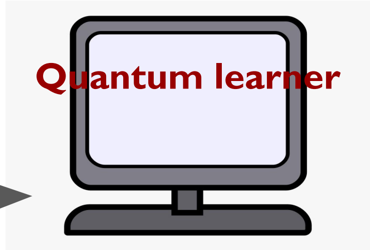
NEW input: Not just *any old* probability distribution, but…

Unknown probability distribution **from U**



$$\{x_1, x_2, \ldots, x_k\} \sim P_U(x)$$

Input: Samples **from a quantum circuit**

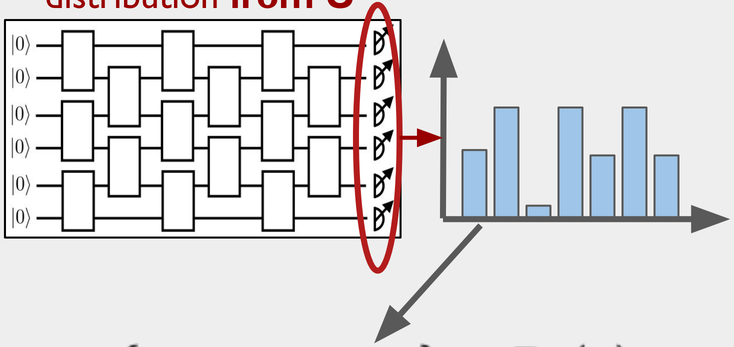


NEW output:

Our setting: quantum generative modelling

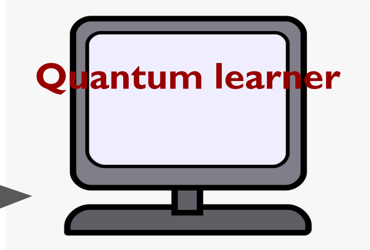
NEW input: Not just *any old* probability distribution, but…

Unknown probability distribution **from U**



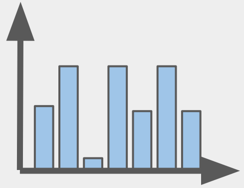
$$\{x_1, x_2, \ldots, x_k\} \sim P_U(x)$$

Input: Samples **from a quantum circuit**



NEW output:

Output: model for

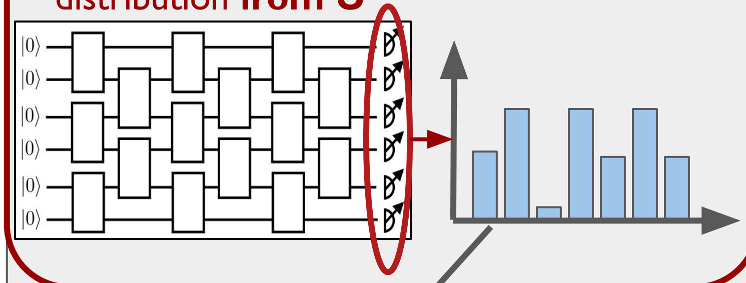


Our setting: quantum generative modelling

NEW input: Not just *any old* probability distribution, but

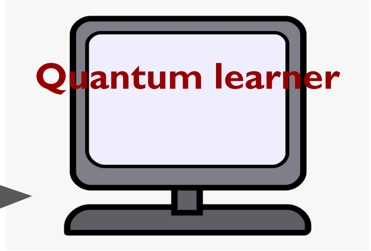
Why??

Unknown probability distribution **from U**



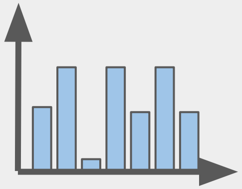
$$\{x_1, x_2, \ldots, x_k\} \sim P_U(x)$$

Input: Samples **from a quantum circuit**



NEW output:

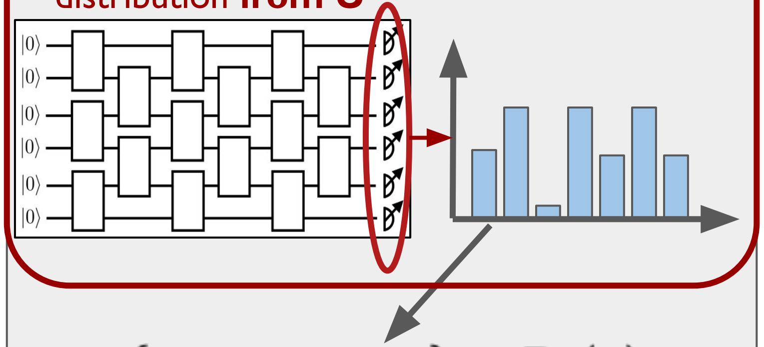
Output: model for



Our setting: quantum generative modelling

NEW input: Not just *any old* probability distribution, but

Unknown probability distribution **from U**



$$\{x_1, x_2, \ldots, x_k\} \sim P_U(x)$$

Input: Samples **from a quantum circuit**

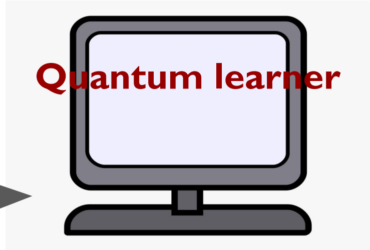
Why??

This model class is

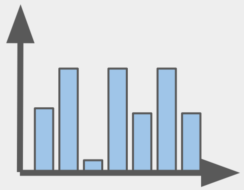
- quite **expressive** relative to other known models

[GSP+19] Glasser et al., Advances in Neural Information Processing Systems 2019

- **can only be run on a QC**



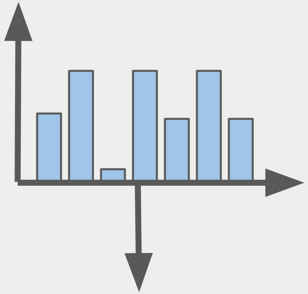
Output: model for

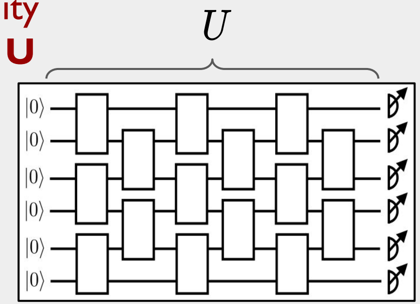


Our setting: quantum generative modelling

NEW input: Not just *any old* probability distribution, but…

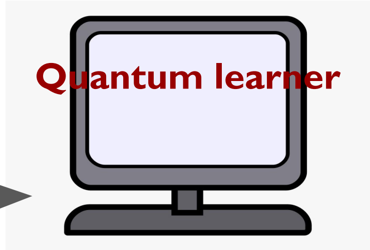
Unknown probability distribution **from U**





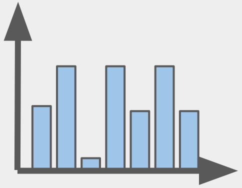
$$\{x_1, x_2, \ldots, x_k\} \sim P_U(x)$$

Input: Samples **from a quantum circuit**



Loosely: a way of **generating new samples from** P_U (cf. GANs)

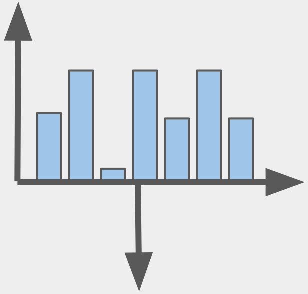
Output: model for

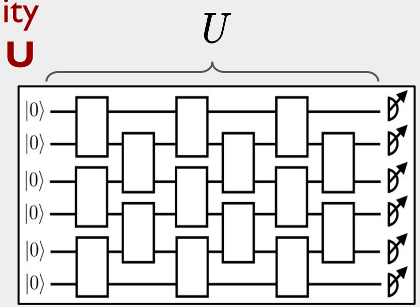


Our setting: quantum generative modelling

NEW input: Not just *any old* probability distribution, but…

Unknown probability distribution **from U**



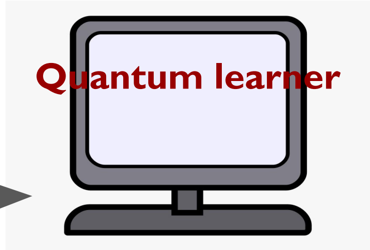


$$\{x_1, x_2, \ldots, x_k\} \sim P_U(x)$$

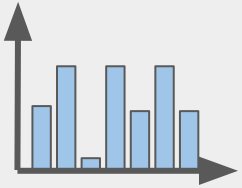
Input: Samples **from a quantum circuit**

Study **time complexity**

hard = superpoly(n)

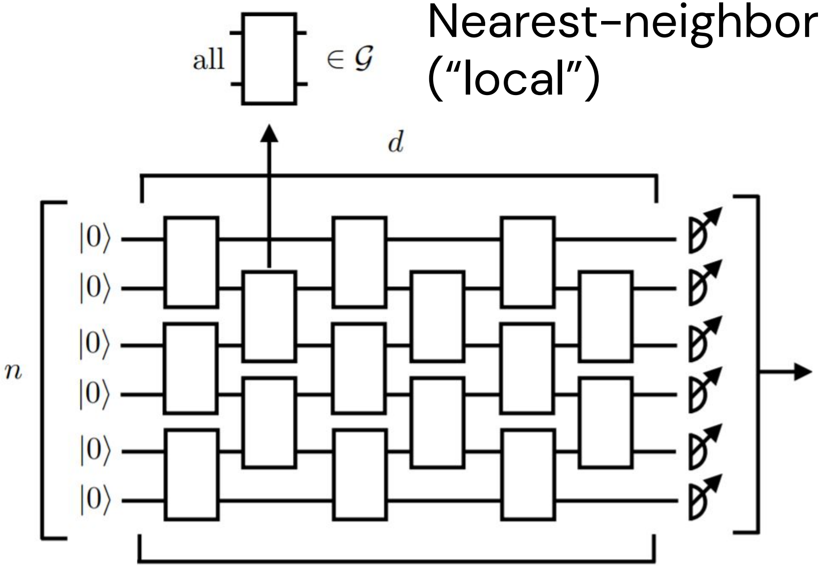


Output: model for



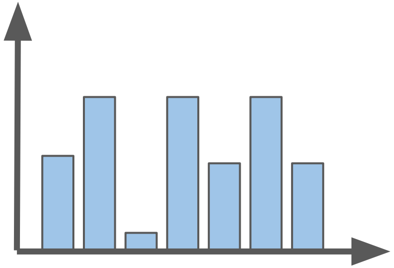
The distributions we wish to learn

Nearest-neighbor gates only
("local")



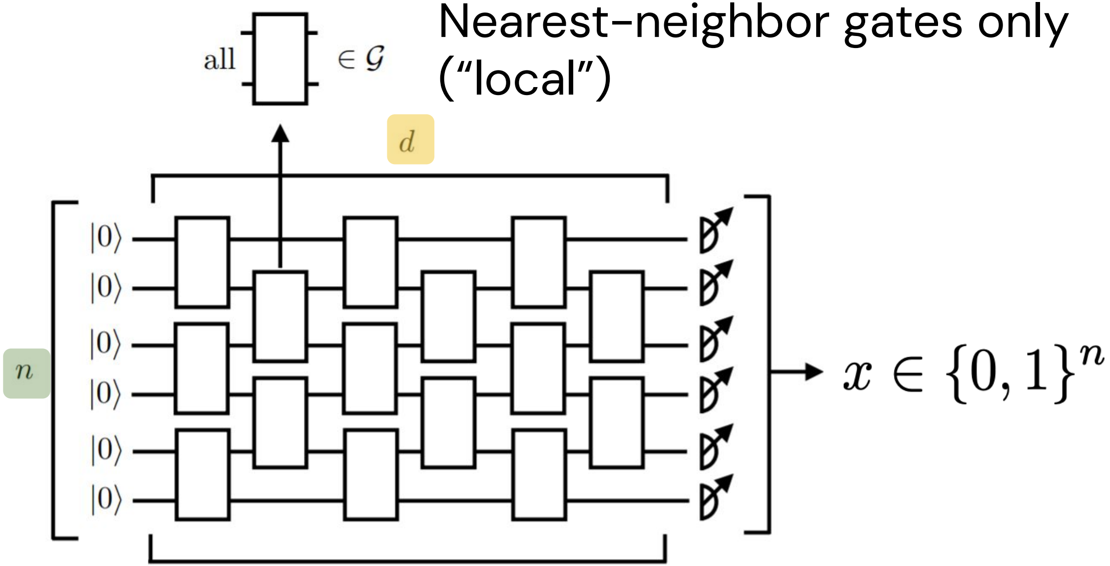
$x \in \{0,1\}^n$

$$P_U(x) = |\langle x|U|0\rangle|^2$$



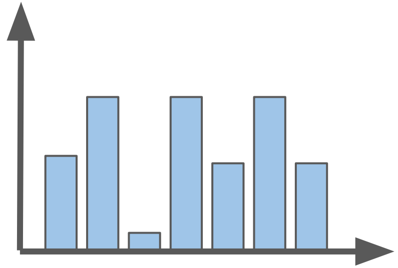
Goal: learn this distribution

The distributions we wish to learn



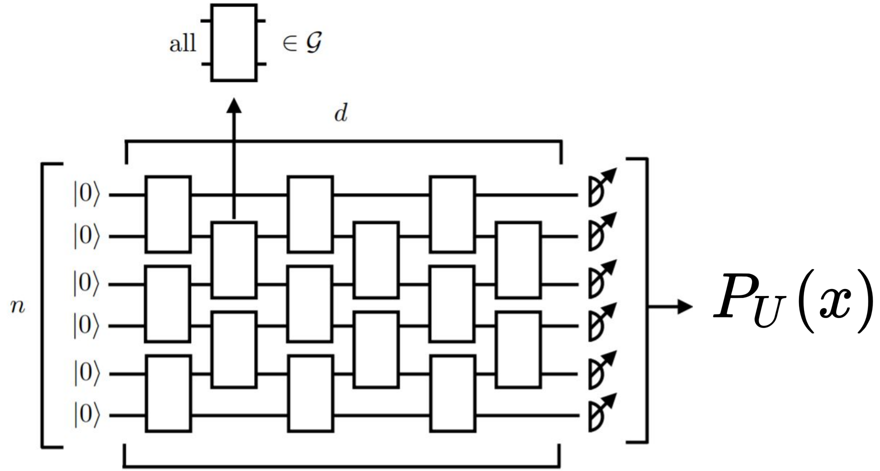
Nearest-neighbor gates only ("local")

$$P_U(x) = |\langle x|U|0\rangle|^2$$

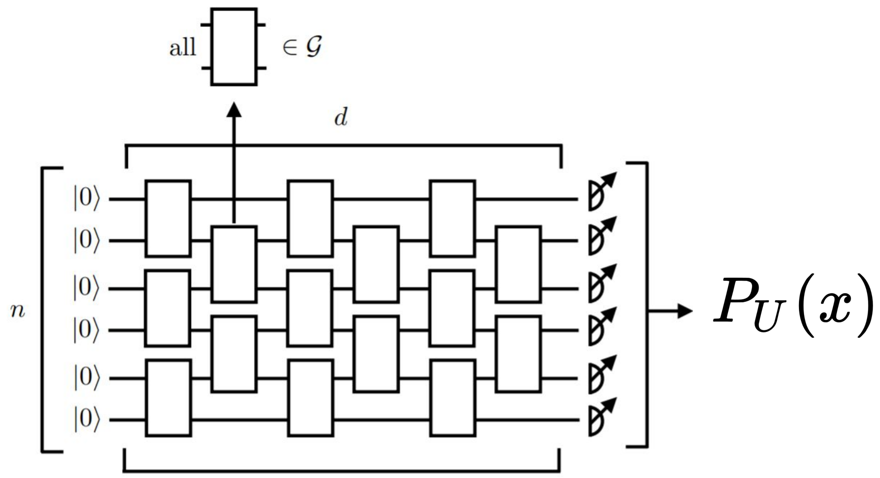


Goal: learn this distribution

What parameters should matter?



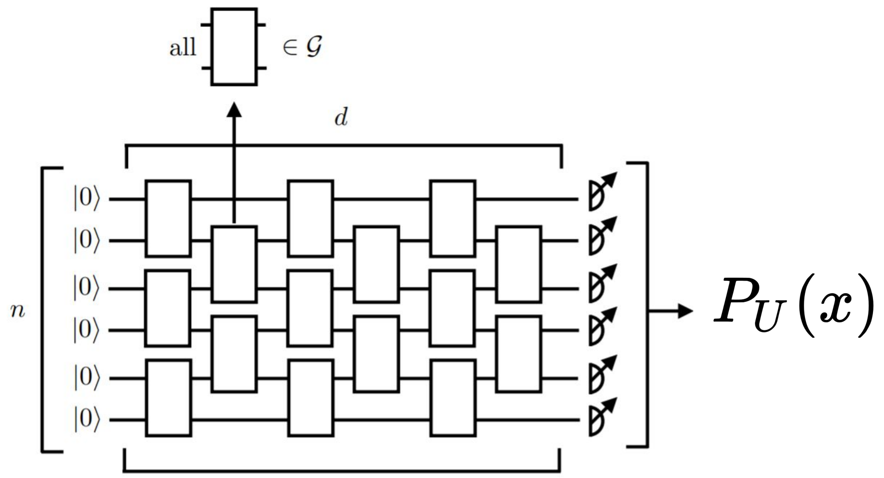
What parameters should matter?



Two extremes:

- **Depth *d* = 1 :**
  C = { product distributions over $\{0,1\}^n$ } **(easy)**
- **Depth *d* → ∞ (and universal gates):**
  C = { all distributions over $\{0,1\}^n$ } **(hard)**

What parameters should matter?

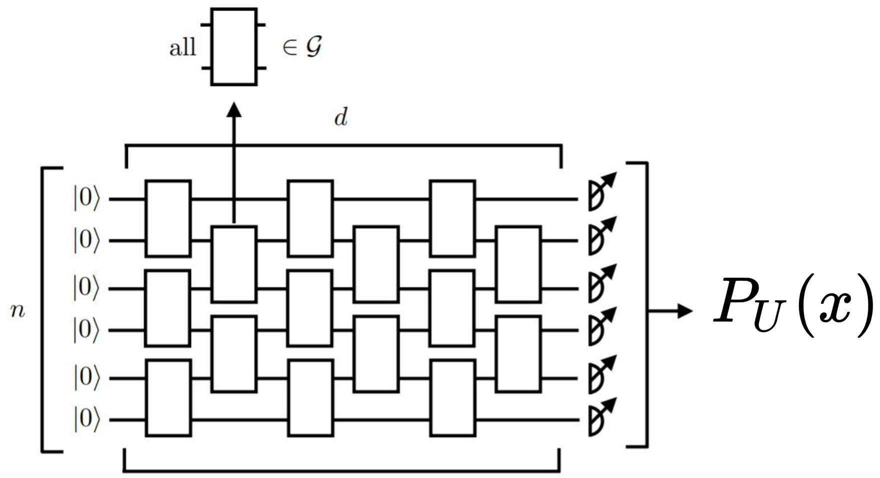


$$\mathcal{C} = \{P_U \mid \bullet\ \mathrm{depth}(U) \le d,\ \bullet \text{ gates from } \mathcal{G}\}$$

Two extremes:

- Depth ***d* = 1 :**
  C = { product distributions over $\{0,1\}^n$ } **(easy)**
- **Depth *d* → ∞ (and universal gates):**
  C = { all distributions over $\{0,1\}^n$ } **(hard)**

What parameters should matter?



$\mathcal{C} = \{P_U \mid$ • depth$(U) \leq d$,
• gates from $\mathcal{G}\}$

Two extremes:

- **Depth** ***d*** **= 1 :**
  C = { product distributions over $\{0,1\}^n$ } **(easy)**
- **Depth** ***d*** **→ ∞ (and universal gates):**
  C = { all distributions over $\{0,1\}^n$ } **(hard)**

(in terms of n = problem size)

How to lower-bound the complexity of your task

Your task,
$P_n(\Theta)$

(in terms of n = problem size)

How to lower-bound the complexity of your task

Your task,
$P_n(\Theta)$

Requires creativity

For some Θ's

A known hard problem Q

(in terms of n = problem size)

How to lower-bound the complexity of your task

Your task,
$P_n(\Theta)$

Requires creativity

For some Θ's

A known hard problem Q

This means: a solver for $P_n(\Theta)$ can be made into* a solver for Q.

*efficiently (≅with at most poly(n) amount of the resource you care about)

"Q reduces to P_n" (means P_n is *harder* than Q)

(in terms of n = problem size)

How to lower-bound the complexity of your task

Your task, $P_n(\Theta)$

Requires creativity

For some Θ's

A known hard problem Q

Requires calculation

Lower bound for complexity of Q

$= E(n)$ (E for enormous!)

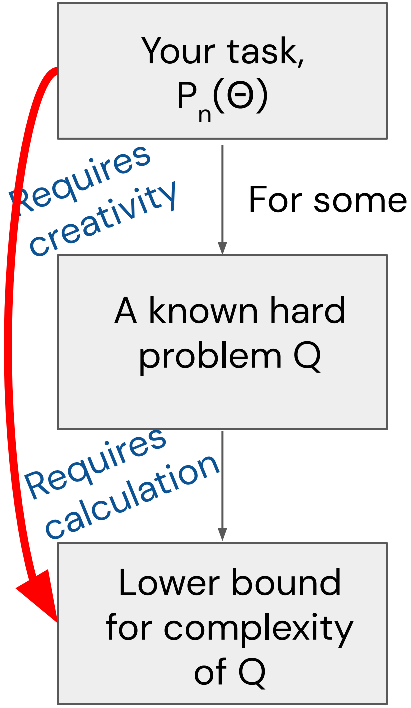
This means: a solver for $P_n(\Theta)$ can be made into* a solver for Q.

*efficiently (≅with at most poly(n) amount of the resource you care about)

"Q reduces to P_n" (means P_n is *harder* than Q)

(in terms of n = problem size)

How to lower-bound the complexity of your task

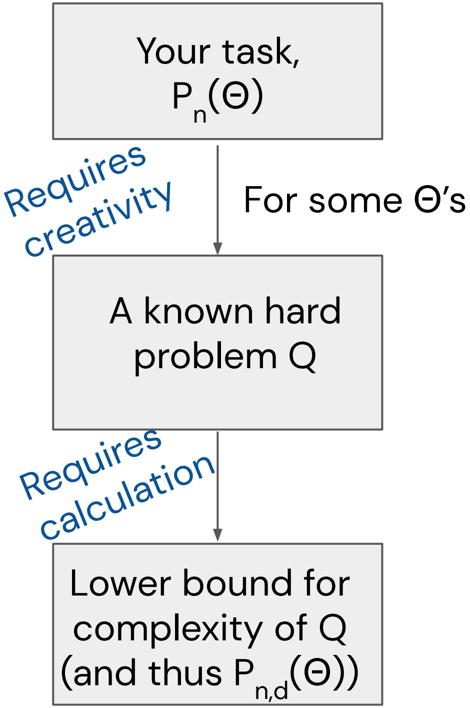


$= E(n)$

Conclusion:

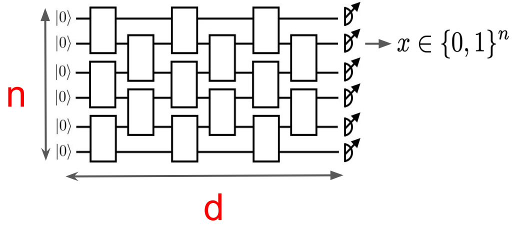
$P_n(\Theta)$ requires complexity $E(n)$ to solve in the worst-case.
→ INEFFICIENT (if E is enormous)

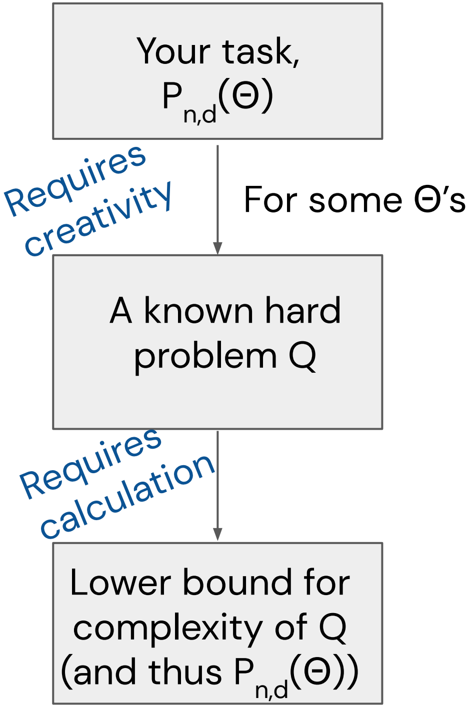
How we use this schema



Θ = {a quantum gateset, a circuit depth}
$P_n(\Theta)$ = learn the output distributions of depth-d quantum circuits on n qubits with only nearest-neighbor gates

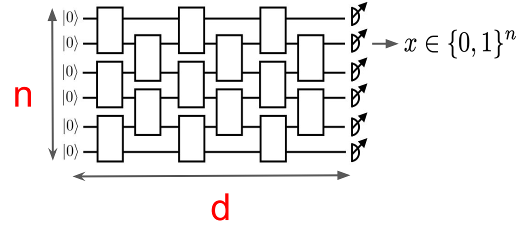
How we use this schema





Θ = {a quantum gateset, a circuit depth}

$P_{n,d}(\Theta)$ = generative modelling for local quantum circuits



How we use this schema

Your task, $P_{n,d}(\Theta)$

Requires creativity

For some Θ's

A known hard problem Q

Requires calculation

Lower bound for complexity of Q (and thus $P_{n,d}(\Theta)$)

Θ = {a quantum gateset, a circuit depth}

$P_{n,d}(\Theta)$ = generative modelling for local quantum circuits

Θ = Cliffords + 1T at depth $d = n^{\Omega(1)}$

Q = *Learning Parities with Noise*

Cliffords embed parities

Parities:

Concatenation of k bits + 1 bit

$$\mathbb{P}_{parities}(\vec{x}\|y) = \begin{cases} \frac{1}{2^k} & \text{if} \quad y = \vec{x}\cdot\vec{s} \\ 0 & else \end{cases}$$

where s is a secret k-bit string.

Cliffords embed parities

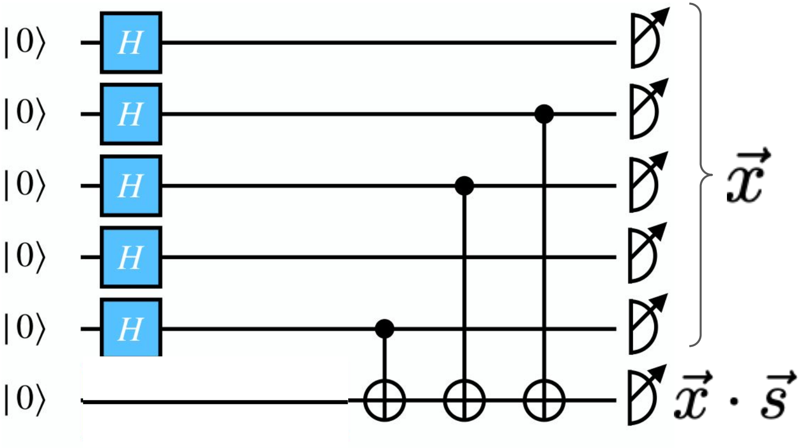
Parities:

Concatenation of k bits + 1 bit

$$\mathbb{P}_{parities}(\vec{x}\|y) = \begin{cases} \frac{1}{2^k} & \text{if} \quad y = \vec{x}\cdot\vec{s} \\ 0 & else \end{cases}$$

where s is a secret k-bit string.

Quantum circuit:



Cliffords + 1T embed noisy parities

Noisy Parities: flip the last bit with probability η

Concatenation of k bits + 1 bit

$$\mathbb{P}_{\text{noisy parities}}(\vec{x}\|y) = \begin{cases} \frac{1}{2^k} \cdot (1-\eta) & \text{if} \quad y = \vec{x}\cdot\vec{s} \\ \frac{1}{2^k} \cdot \eta & else \end{cases}$$

where s is a secret k-bit string.

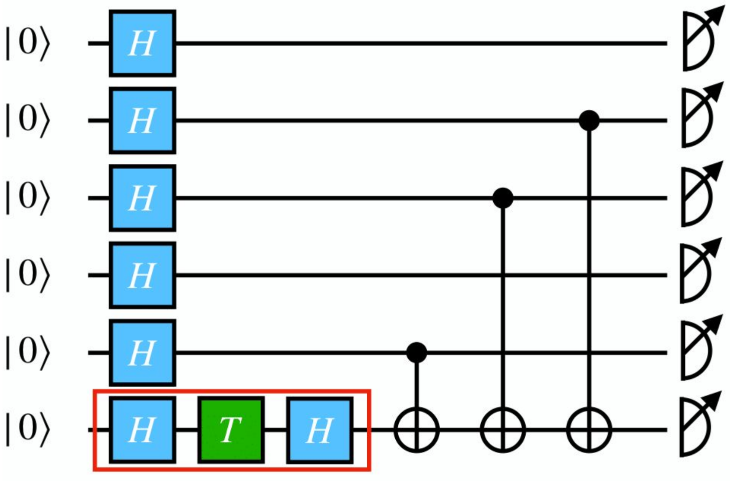
Cliffords + 1T embed noisy parities

Noisy Parities: flip the last bit with probability η

Concatenation of k bits + 1 bit

$$\mathbb{P}_{\text{noisy parities}}(\vec{x}\|y) = \begin{cases} \frac{1}{2^k} \cdot \boxed{(1-\eta)} & \text{if} \quad y = \vec{x} \cdot \vec{s} \\ \frac{1}{2^k} \cdot \boxed{\eta} & else \end{cases}$$

where s is a secret k-bit string.

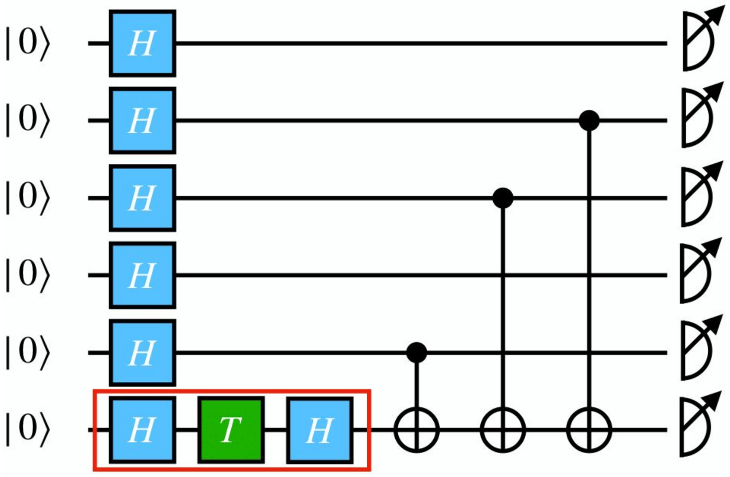


Cliffords + 1T embed noisy parities

Noisy Parities: flip the last bit with probability η

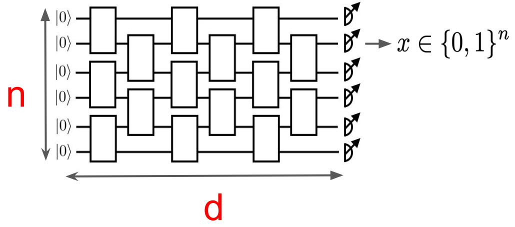
Concatenation of k bits + 1 bit

$$\mathbb{P}_{\text{noisy parities}}(\vec{x}\|y) = \begin{cases} \frac{1}{2^k} \cdot \boxed{(1-\eta)} & \text{if} \quad y = \vec{x} \cdot \vec{s} \\ \frac{1}{2^k} \cdot \boxed{\eta} & else \end{cases}$$



where s is a secret k-bit string.

Computational complexity theory assumption:
Noisy parities is HARD (superpolynomial) to learn from samples



How we use this schema

Your task, $P_{n,d}(\Theta)$

Requires creativity

For some Θ's

A known hard problem Q

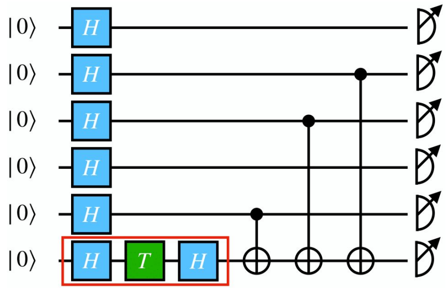
Requires calculation

Lower bound for complexity of Q (and thus $P_{n,d}(\Theta)$)

Θ = {a quantum gateset, a circuit depth}

$P_{n,d}(\Theta)$ = generative modelling for local quantum circuits

Θ = Cliffords + 1T at depth $d = n^{\Omega(1)}$

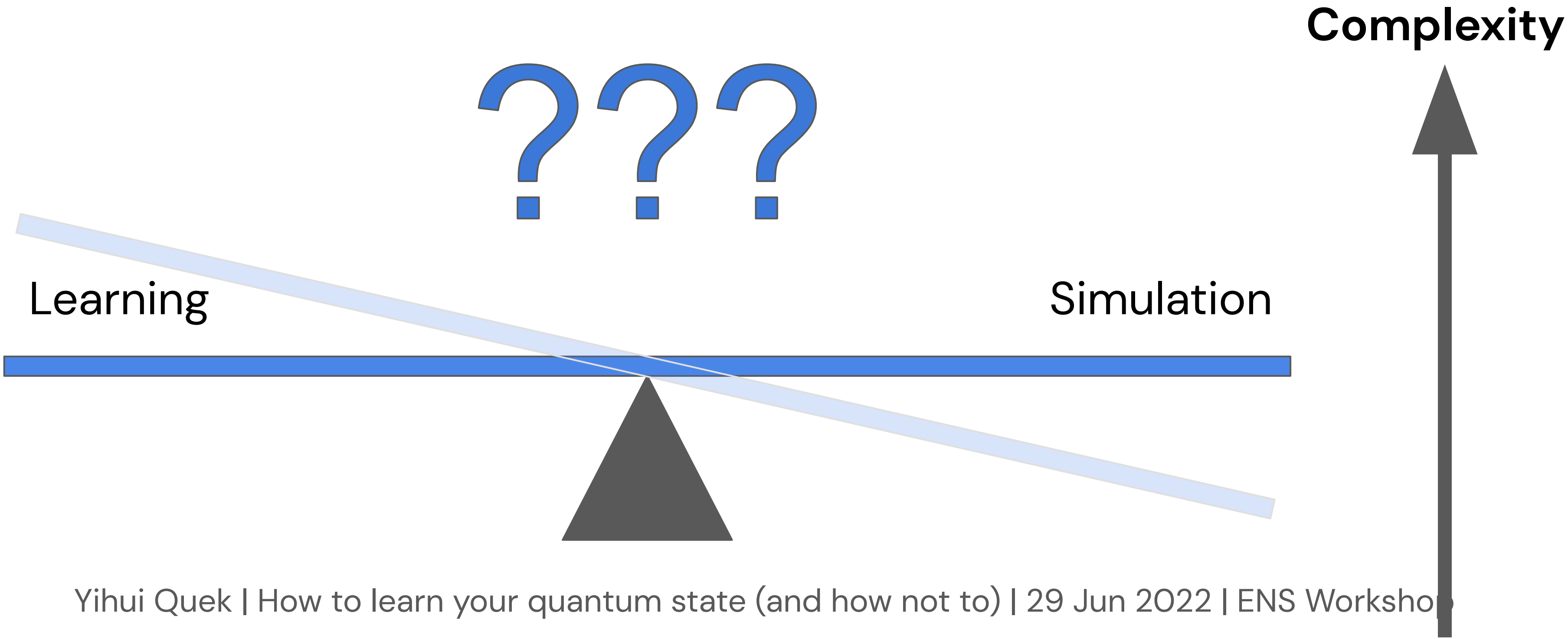


Q = *Learning Parities with Noise*

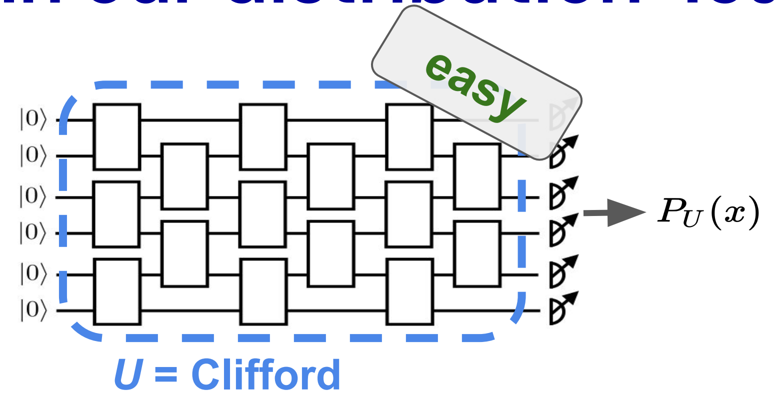
Conclusion:

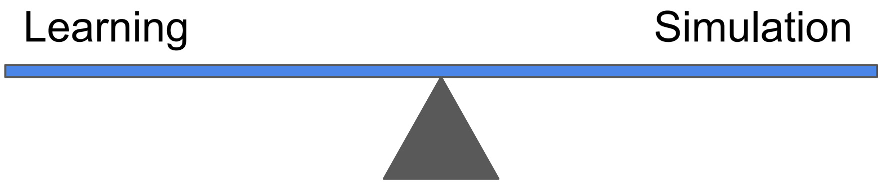
Learning local quantum circuits of depth $n^{\Omega(1)}$ is hard in the worst-case.

Learnability vs simulatability: the case of Cliffords

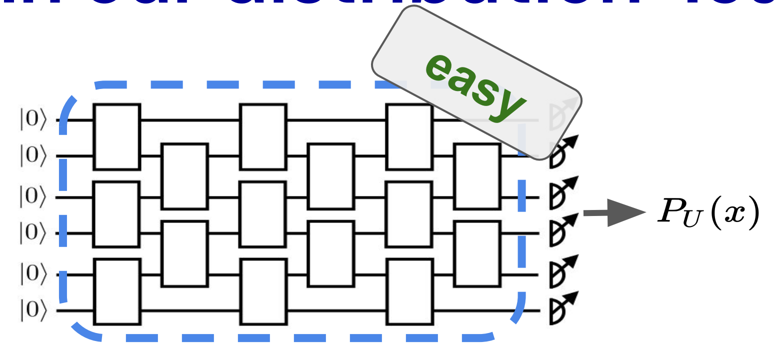


In our distribution-learning setting

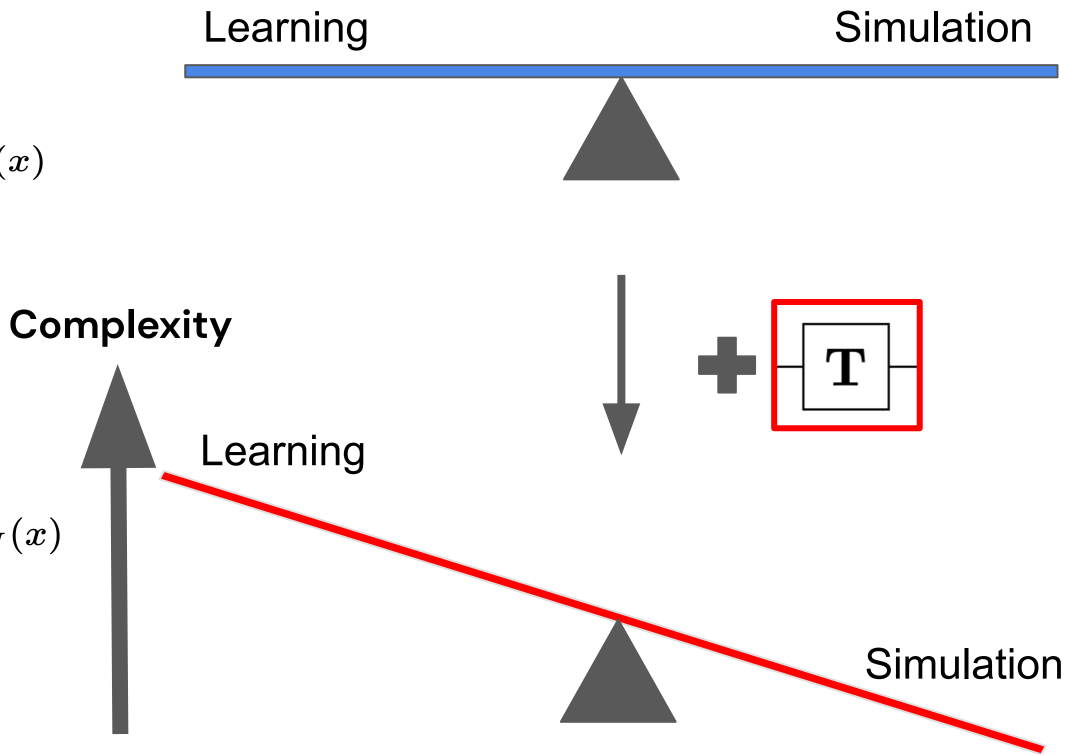


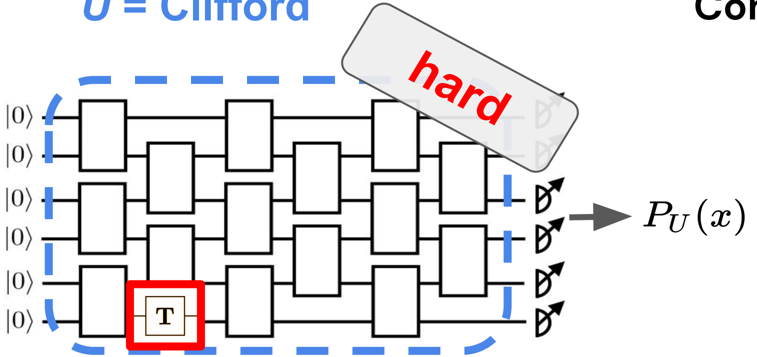


In our distribution-learning setting



U = Clifford

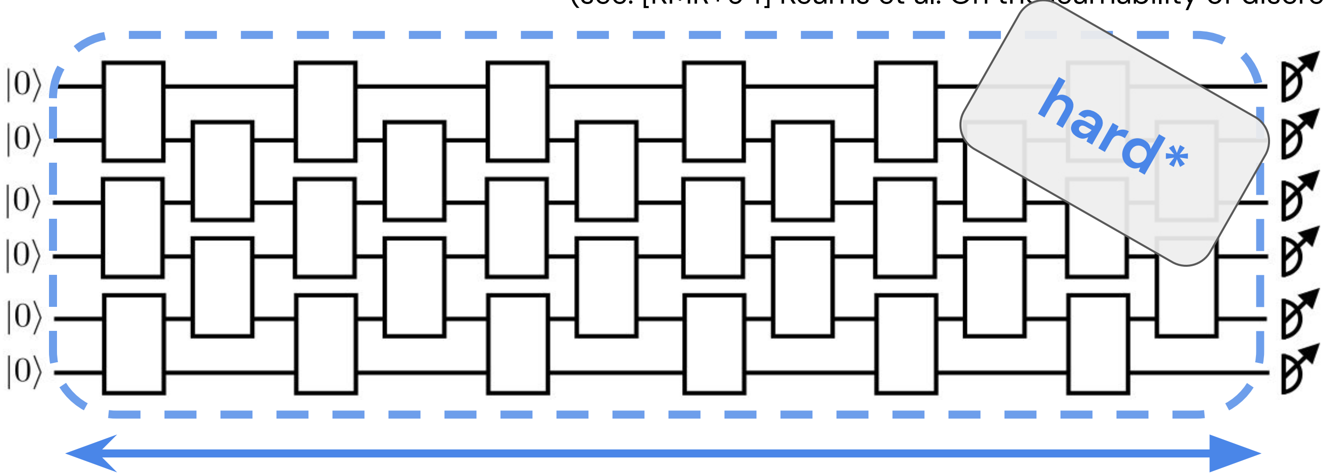




***U*= Clifford + one T gate**

Our contribution: Hardness vs depth

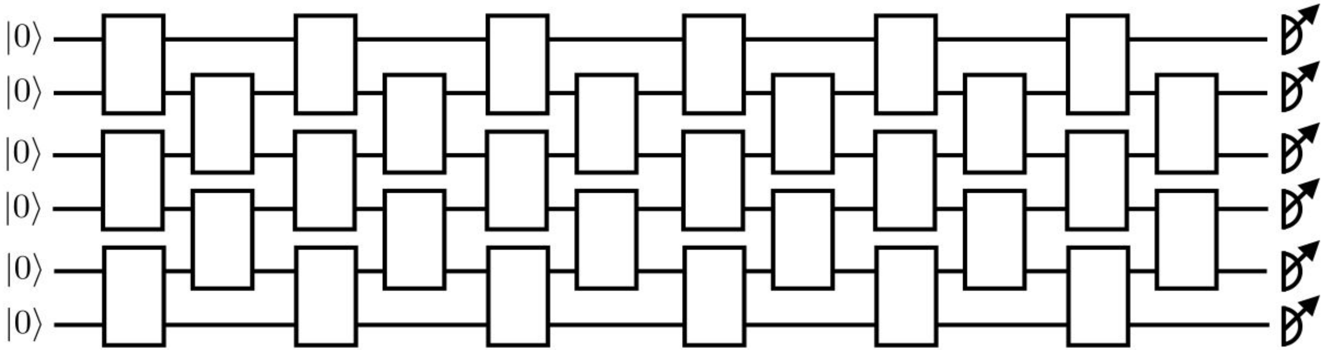
Proof: Embed a *pseudo-random function* into the output distribution
(see: [KMR+94] Kearns et al. On the learnability of discrete distributions. 1994)



$d = n^{\Omega(1)}$

Our contribution: Hardness vs depth

We also considered statistical queries (SQ) – a *weaker* form of sampling.

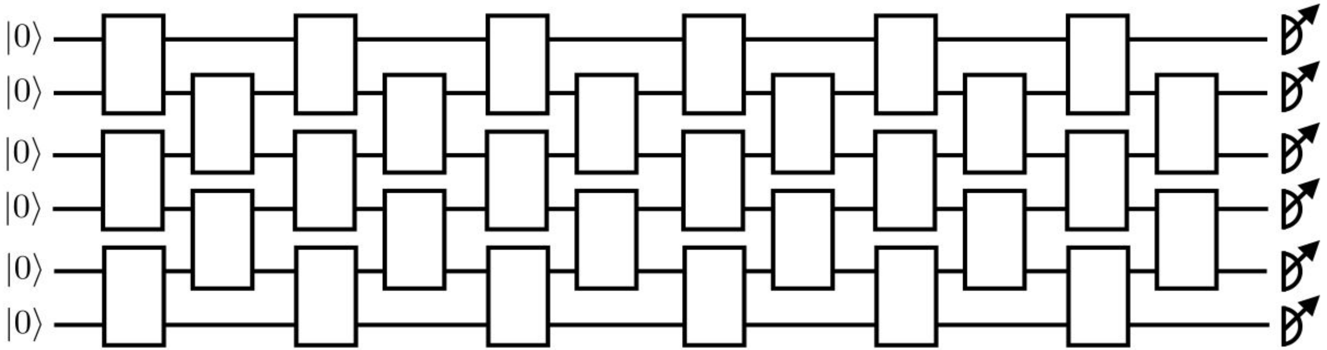


SQ = learner doesn't treat samples individually but only sees statistical averages $\mathbb{E}_{x\sim P_U}[\phi(x)]$ – e.g. ϕ is gradient)

This is what many algorithms actually do!

Our contribution: Hardness vs depth

We also considered statistical queries (SQ) – a *weaker* form of sampling. Expect hardness at lower depths.

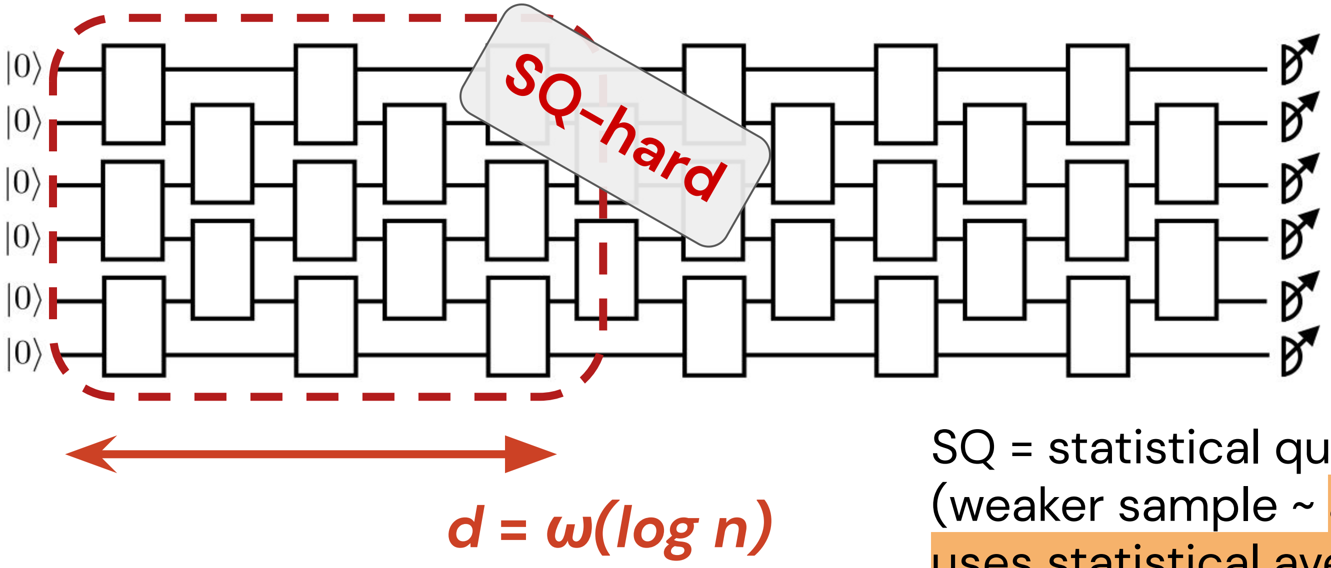


SQ = learner doesn't treat samples individually but only sees statistical averages $\mathbb{E}_{x\sim P_U}[\phi(x)]$ – e.g. ϕ is gradient)

This is what many algorithms actually do!

Our contribution: Hardness vs depth

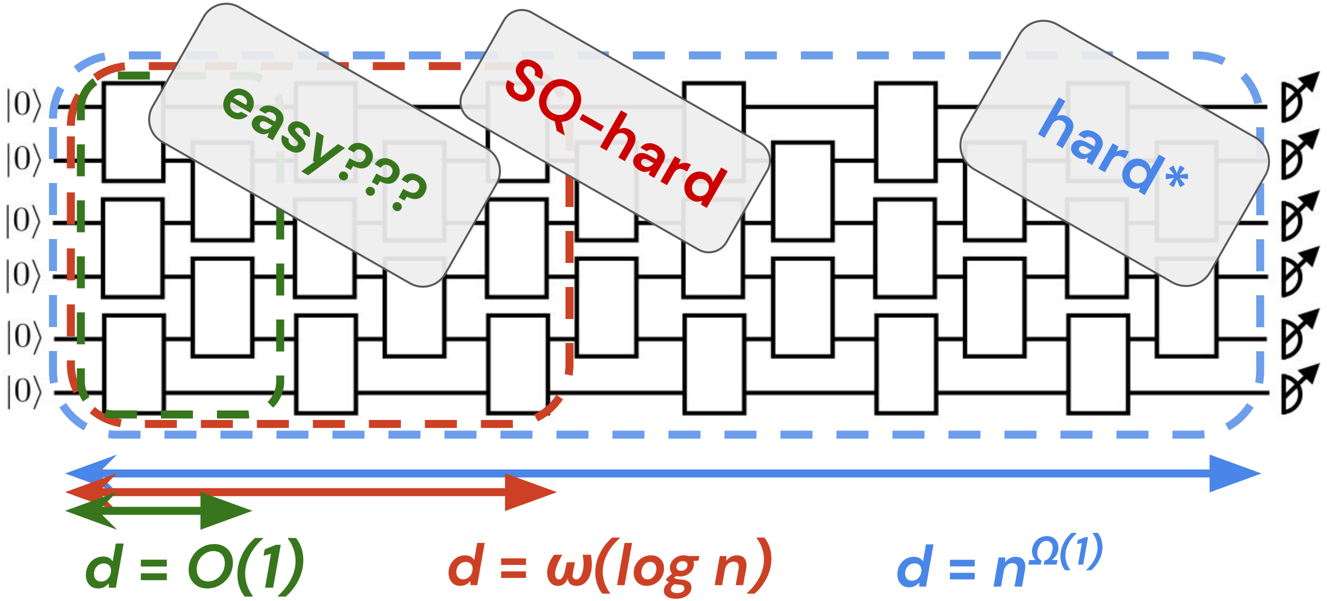
Proof: embed *Learning Parities* (easy from samples – *exponentially hard* from statistical queries).



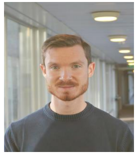
SQ = statistical query
(weaker sample ~ algorithm that uses statistical averages only)

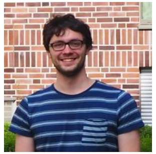
What we DON'T know so far

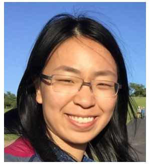
Learning **shallow *d=O(1)*** circuit output distributions??



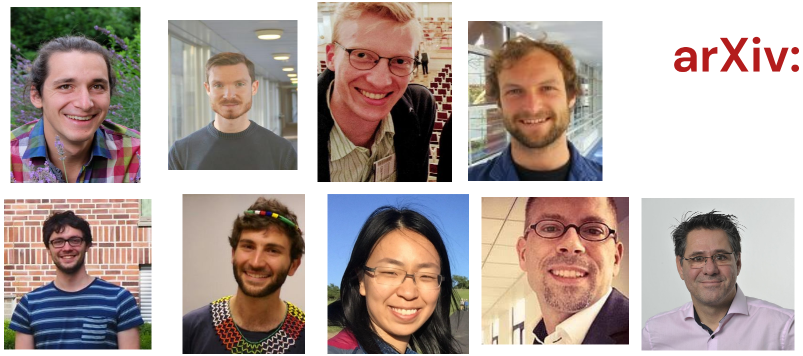
Part I takeaways



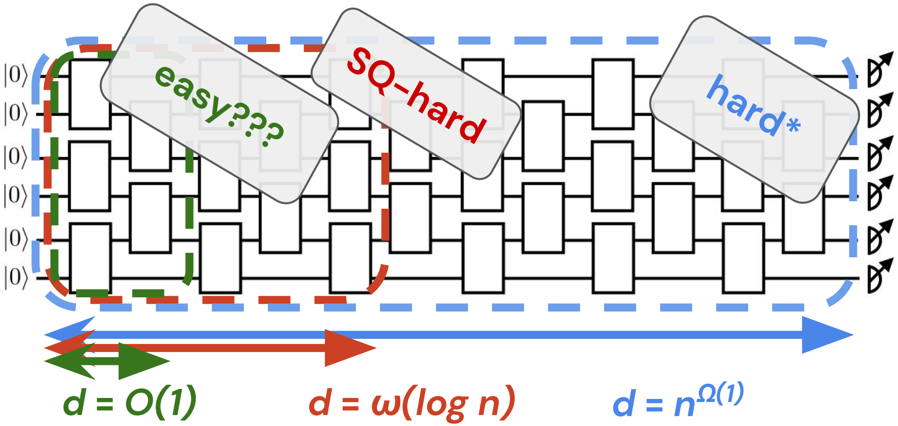
arXiv: 2110.05517

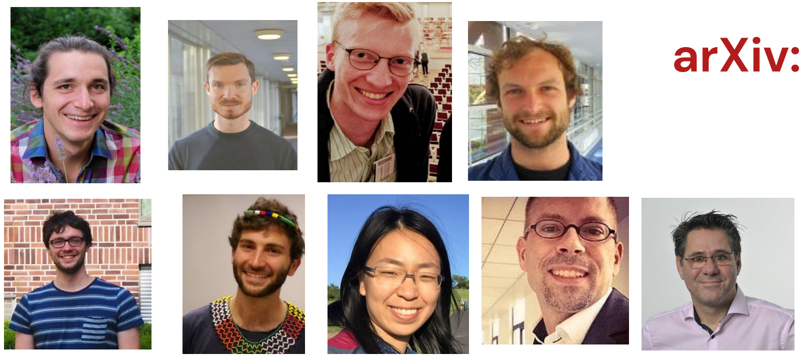


arXiv: 2110.05517

Part I takeaways

- First rigorous insights into learnability of quantum circuit output distributions.

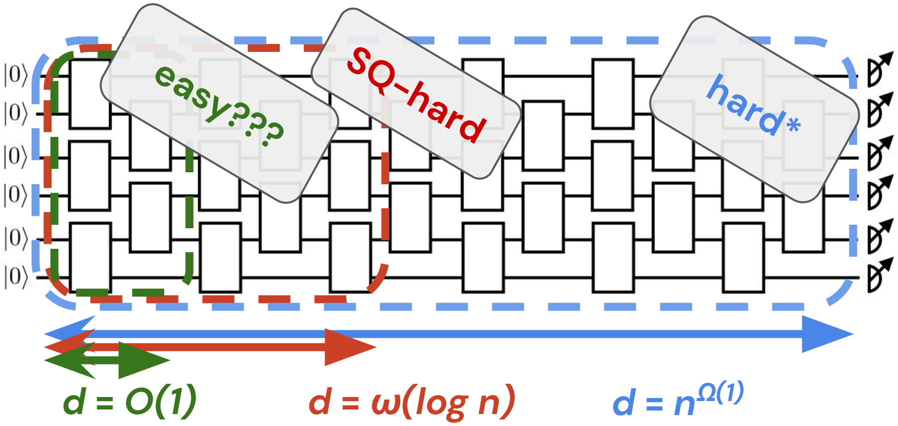


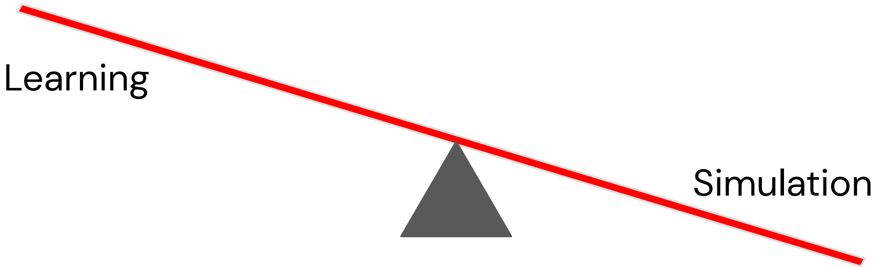


arXiv: 2110.05517

Part I takeaways

- First rigorous insights into learnability of quantum circuit output distributions.
- A single T-gate makes distribution learning of Cliffords hard.





Back to the big picture.

Are there efficient ways to learn quantum states?

Ideas:

1) Assume the quantum circuit that produces the state* is “simple“.
   a) Clifford circuits?
   b) Easily simulable → easily learnable?
2) Change the definition of `learning’, so that one only needs to learn `some properties’ of the state.

Aaronson, 2007. The learnability of quantum states *Proc. R. Soc. A.*4633089–3114

*Actually, output distributions of circuits

Back to the big picture.

Are there efficient ways to learn quantum states?

Ideas:

1) Assume the quantum circuit that produces the state* is “simple“.
   a) Clifford circuits? ✅ Easy to learn indeed, but...
   b) Easily simulable → easily learnable?
2) Change the definition of `learning’, so that one only needs to learn `some properties’ of the state.

Aaronson, 2007. The learnability of quantum states *Proc. R. Soc. A*.4633089–3114

*Actually, output distributions of circuits

Back to the big picture.

Are there efficient ways to learn quantum states?

Ideas:

1) Assume the quantum circuit that produces the state* is "simple".
   a) Clifford circuits? ✅ Easy to learn indeed, but…
   b) Easily simulable → easily learnable? ❌ …not because they're easily simulable.
2) Change the definition of `learning', so that one only needs to learn `some properties' of the state.

Aaronson, 2007. The learnability of quantum states *Proc. R. Soc. A*.4633089–3114

*Actually, output distributions of circuits

Back to the big picture.

Are there efficient ways to learn quantum states?

Ideas:

1) Assume the quantum circuit that produces the state* is “simple“.
   a) Clifford circuits? ✅ Easy to learn indeed, but...
   b) Easily simulable → easily learnable? ❌ ...not because they’re easily simulable.

2) Change the definition of `learning’, so that one only needs to learn `some properties’ of the state.

Aaronson, 2007. The learnability of quantum states *Proc. R. Soc. A*.4633089–3114

*Actually, output distributions of circuits

Back to the big picture.

Are there efficient ways to learn quantum states?

Ideas:

1) Assume the quantum circuit that produces the state* is “simple“.
   a) Clifford circuits? ✅ Easy to learn indeed, but…
   b) Easily simulable → easily learnable? ❌ …not because they’re easily simulable.

2) Change the definition of `learning’, so that one only needs to learn `some properties’ of the state.

Aaronson, 2007. The learnability of quantum states *Proc. R. Soc. A.*4633089–3114

Learnable in model A
→ Learnable in model B?

*Actually, output distributions of circuits

Part II

[AQS'21] arXiv: 2102.07171, NeurIPS 2021 (Spotlight)

Private learning implies quantum stability

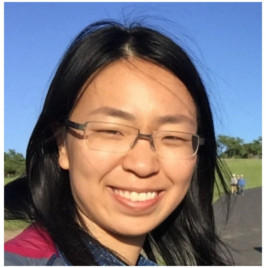
Srinivasan Arunachalam*[1], Yihui Quek†[2], and John Smolin‡[1]

[1]IBM Quantum, IBM T.J. Watson Research Center, Yorktown Heights, USA
[2] Information Systems Laboratory, Stanford University, USA

February 16, 2021

Abstract

Learning an unknown n-qubit quantum state ρ is a fundamental challenge in quantum computing. Information-theoretically, it is well-known that tomography requires exponential in n many copies of an unknown state ρ in order to estimate it up to small trace distance. Motivated by computational learning theory, Aaronson and others introduced several (weaker) learning models: the PAC model of learning quantum states (Proc. of Royal Society A'07), shadow tomography (STOC'18) for learning "shadows" of a quantum state, a learning model that additionally requires learners to be differentially private (STOC'19), and the online model of learning quantum states (NeurIPS'18). In these models it was shown that an unknown quantum state can be learned "approximately well" using *linear* in n many copies of ρ. But is there any relationship between these learning models? In this paper we prove a sequence of (information-theoretic) implications from differentially-private PAC learning to online learning and then to quantum stability.



Redefining quantum learning

Critical observation: every ρ defines a function f_ρ :

$$f_\rho(E) = \mathrm{Tr}(E\rho)$$

Redefining quantum learning

Critical observation: every ρ defines a function f_ρ :

$$f_\rho(E) = \mathrm{Tr}(E\rho)$$

E = a two-outcome measurement

Redefining quantum learning

Critical observation: every ρ defines a function f_ρ mapping to [0,1]:

$$f_\rho(E) = \mathrm{Tr}(E\rho)$$

= probability of E accepting on ρ

E = a two-outcome measurement

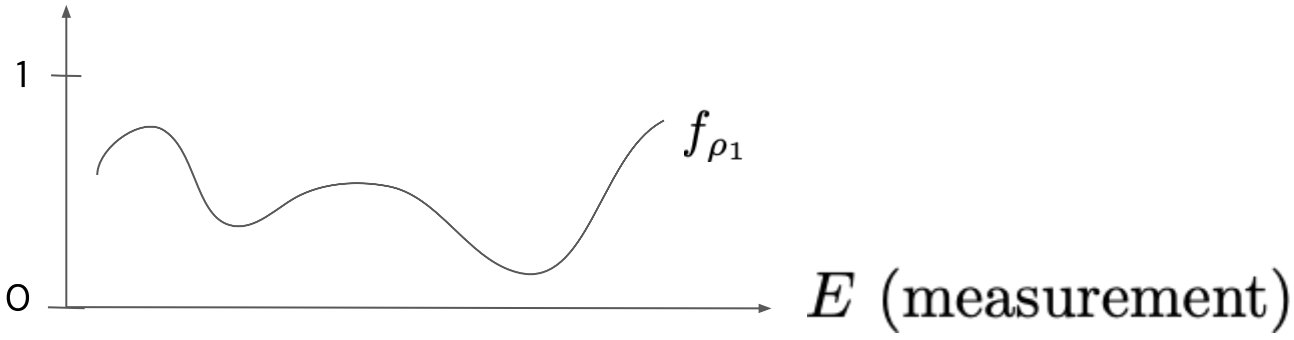
Redefining quantum learning

Critical observation: every ρ defines a function f_ρ mapping to [0,1]:

$$f_\rho(E) = \mathrm{Tr}(E\rho)$$

= probability of E accepting on ρ

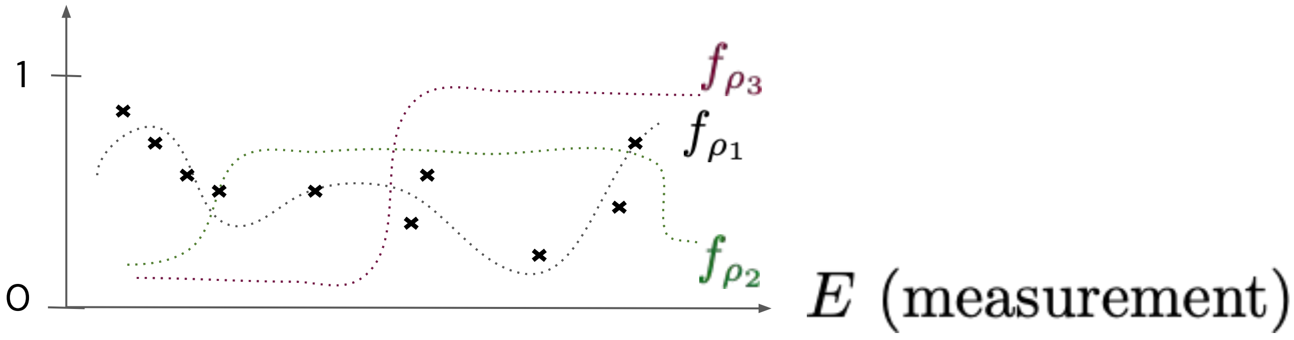
E = a two-outcome measurement



Redefining quantum learning

Critical observation: every ρ defines a function f_ρ mapping to [0,1]:

$$f_\rho(E) = \mathrm{Tr}(E\rho)$$

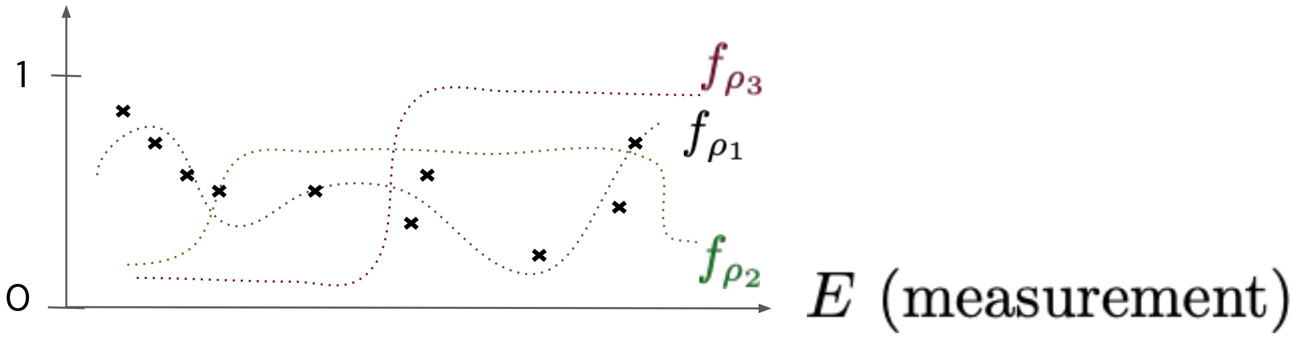


Redefining quantum learning

Critical observation: every ρ defines a function f_ρ mapping to [0,1]:

$$f_\rho(E) = \mathrm{Tr}(E\rho)$$

Data = the pairs $\{E_i, f_\rho(E_i)\}_i$



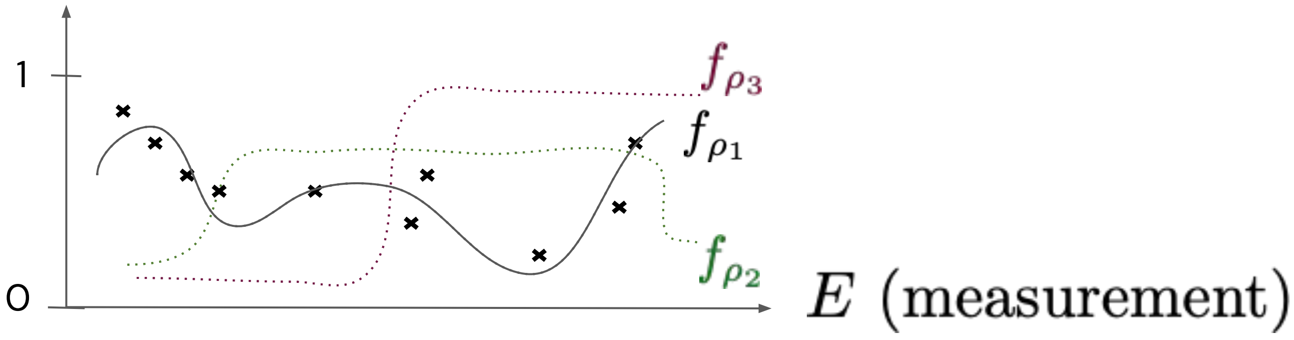
Redefining quantum learning

Critical observation: every ρ defines a function f_ρ mapping to [0,1]:

$$f_\rho(E) = \mathrm{Tr}(E\rho)$$

Data = the pairs
$\{E_i, f_\rho(E_i)\}_i$

Learning task:
Learn f_ρ from data.

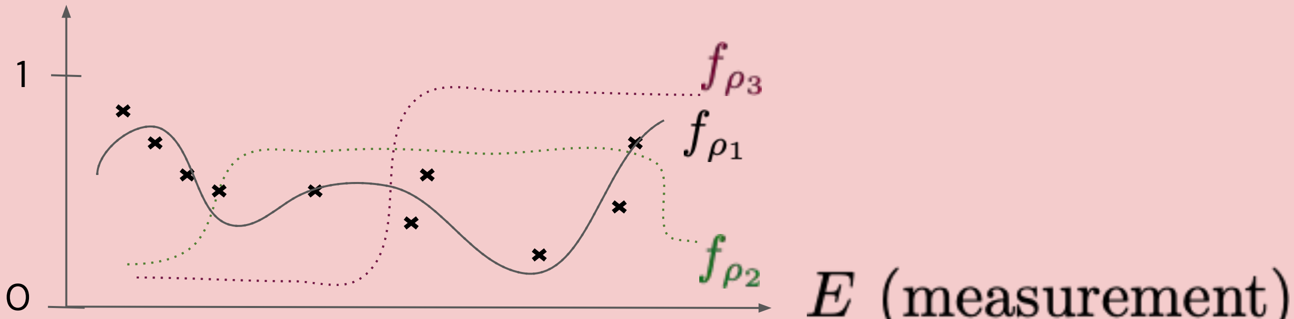


Redefining quantum learning

Critical observation: every ρ defines a function f_ρ mapping to [0,1]:

$$f_\rho(E) = \mathrm{Tr}(E\rho)$$

Learning q. state = estimate full $\rho\,(2^n \times 2^n \text{ matrix})$??

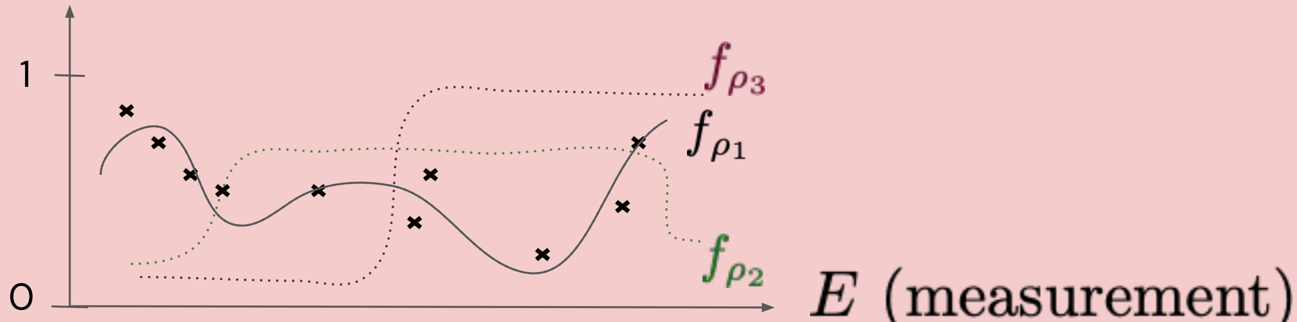


Redefining quantum learning

Critical observation: every ρ defines a function f_ρ mapping to [0,1]:

$$f_\rho(E) = \mathrm{Tr}(E\rho)$$

Learning q. state = ~~estimate full ρ $(2^n \times 2^n$ matrix)~~??
= Predict f_ρ

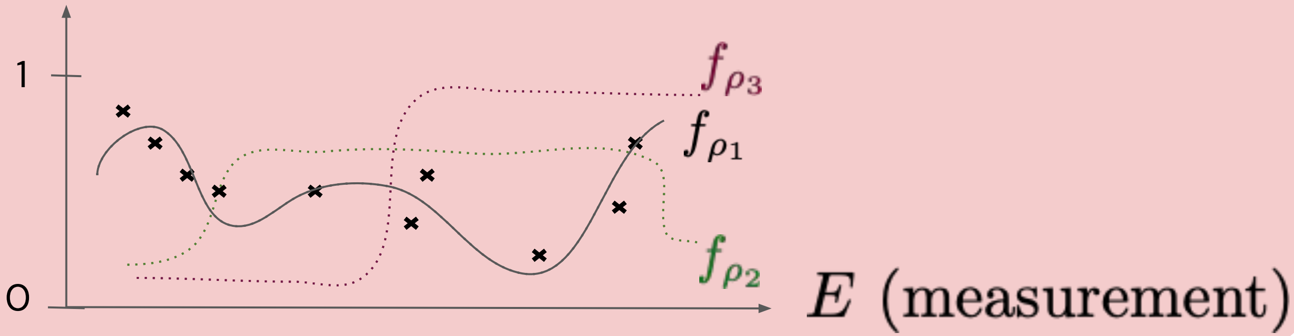


Redefining quantum learning

Critical observation: every ρ defines a function f_ρ mapping to [0,1]:

$$f_\rho(E) = \mathrm{Tr}(E\rho)$$ (***pretty-good*** tomography)

Learning q. state = ~~estimate full ρ ($2^n \times 2^n$ matrix)??~~
= Predict f_ρ



Example of q. learning model: Shadow tomography

[Shadow tomography of quantum states, Aaronson 2018]

Given:

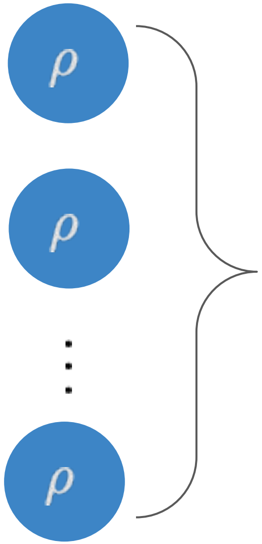
m possible measurements

Example of q. learning model: Shadow tomography

[Shadow tomography of quantum states, Aaronson 2018]

Given:

m possible measurements



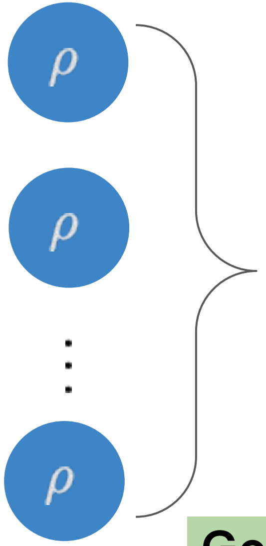
r copies of unknown
n-qubit ρ

Example of q. learning model: Shadow tomography

[Shadow tomography of quantum states, Aaronson 2018]

Given:

m possible measurements



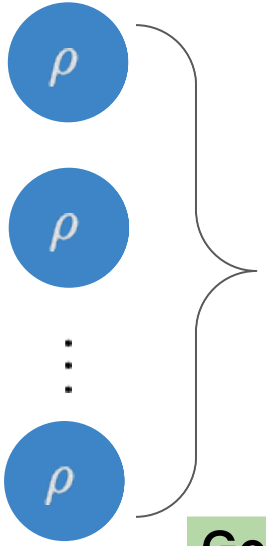
r copies of unknown
n-qubit ρ

Goal: What is smallest r to estimate $\mathrm{Tr}(\rho E_i)$ for all $i \in [m]$?

Example of q. learning model: Shadow tomography

Given:

m possible measurements



Observe: easier than estimating whole ρ

r copies of unknown
n-qubit ρ

Goal: What is smallest r to estimate $\mathrm{Tr}(\rho E_i)$ for all $i \in [m]$?

Another example: Online learning

Unknown n-qubit ρ. Repeat the following **rounds of interaction**:

Adversary

E_i

Maintains hypothesis σ

Online Learner

Online learning of quantum states, Aaronson, Chen, Nayak, Hazan, Kale, Nayak, NeurIPS 2018

Another example: Online learning

Unknown n-qubit ρ. Repeat the following **rounds of interaction**:

Adversary

E_i

I predict Tr(E_i σ)

Maintains hypothesis σ

Online Learner

Online learning of quantum states, Aaronson, Chen, Nayak, Hazan, Kale, Nayak, NeurIPS 2018

Another example: Online learning

Unknown n-qubit ρ. Repeat the following **rounds of interaction**:

Adversary

E_i

Maintains hypothesis σ

Online Learner

I predict Tr($E_i\ \sigma$)

You are **mistaken**/ you are correct

Actual value: Tr($E_i\ \rho$)

Optionally update σ

+ repeat T times!

Online learning of quantum states, Aaronson, Chen, Nayak, Hazan, Kale, Nayak, NeurIPS 2018

Another example: Online learning

Unknown n-qubit ρ. Repeat the following **rounds of interaction**:

Adversary

E_i

Maintains hypothesis σ

Online Learner

I predict $\mathrm{Tr}(E_i\, \sigma)$

You are **mistaken**/ you are correct

Mistake means:

$|\mathrm{Tr}(E_i\, \sigma) - \mathrm{Tr}(E_i\, \rho)| > \varepsilon$

Actual value: $\mathrm{Tr}(E_i\, \rho)$

Want to minimize: <u>worst-case number of mistakes made</u> in T rounds

"Pretty-good" tomography (and friends)

"Pretty-good" : learning f_ρ is often "easier" than learning whole ρ.

Linear in n

!!

Tomography: Exponential in n (number of qubits)

"Pretty-good" tomography (and friends)

Over limited set of measurements:
Shadow tomography

"Pretty-good" : learning f_ρ is often "easier" than learning whole ρ.

Linear in n

!!

Tomography: Exponential in n (number of qubits)

“Pretty-good” tomography (and friends)

Privacy preserving: **Private PAC learning**

w.h.p. over a distribution over measurements: **PAC learning**

Over limited set of measurements: **Shadow tomography**

“Pretty-good” : learning f_ρ is often “easier” than learning whole ρ.

Linear in n

!!

Tomography: Exponential in n (number of qubits)

“Pretty-good” tomography (and friends)

Privacy preserving: **Private PAC learning**

w.h.p. over a distribution over measurements: **PAC learning**

Sequential guesses with feedback: **Online learning**

Over limited set of measurements: **Shadow tomography**

“Pretty-good” : learning f_ρ is often “easier” than learning whole ρ.

Linear in n

!!

Tomography: Exponential in n (number of qubits)

"Pretty-good" tomography (and friends)

Privacy preserving: **Private PAC learning**

w.h.p. over a distribution over measurements: **PAC learning**

Over limited set of measurements: **Shadow tomography**

Sequential guesses with feedback: **Online learning**

By sending bits between Alice and Bob **Learning by communicating**

"Pretty-good" : learning f_ρ is often "easier" than learning whole ρ.

Linear in n

!!

Tomography: Exponential in n (number of qubits)

"Pretty-good" tomography (and friends)

How are they related?

Private PAC learning

PAC learning

Shadow tomography

Online learning

Learning by communicating

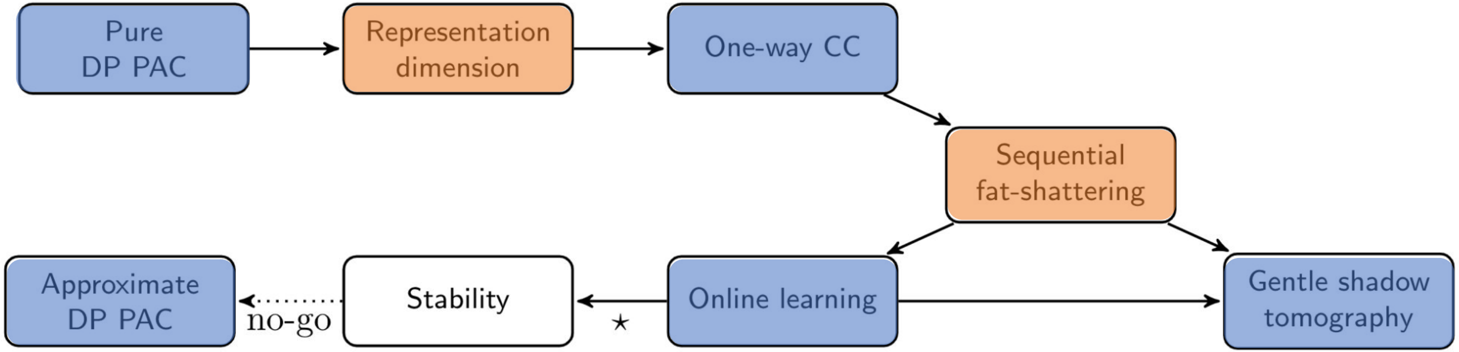
"Pretty-good" : learning f_ρ is often "easier" than learning whole ρ.

Linear in n

Tomography: Exponential in n (number of qubits)

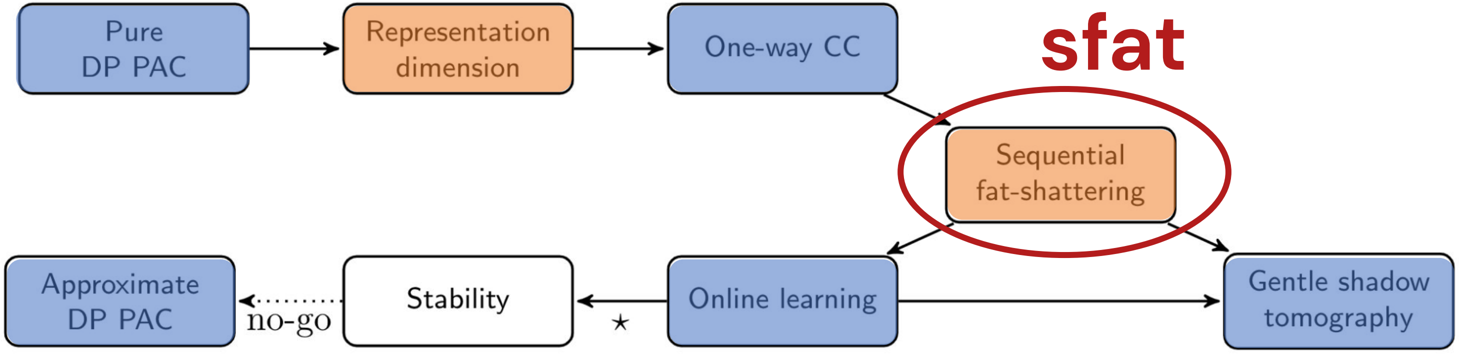
Our contribution: these models imply each other.

A web of implications between quantum learning models and combinatorial parameters



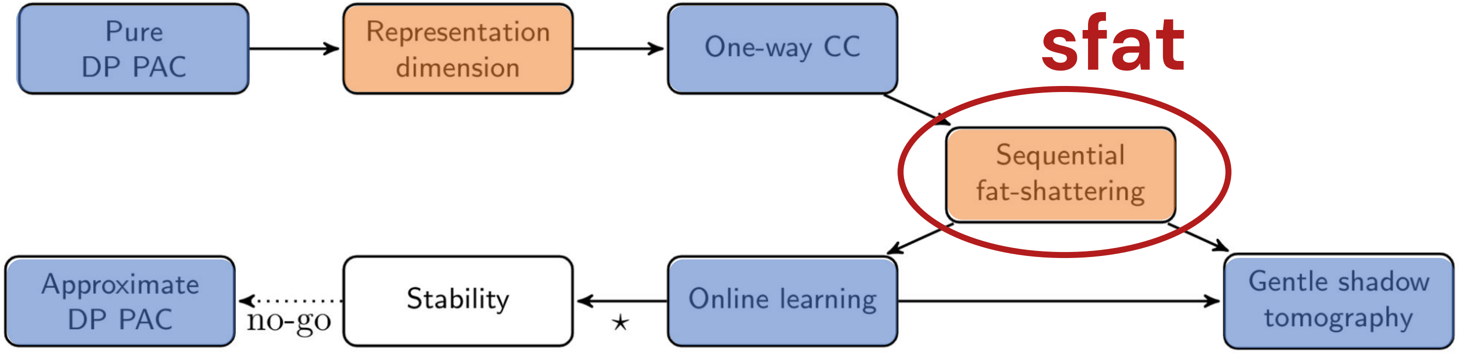
Our contribution: these models imply each other.

A web of implications between quantum learning models and combinatorial parameters = measures complexity of learning states *with special properties*



Our contribution: these models imply each other.

A web of implications between quantum learning models and combinatorial parameters = measures complexity of learning states *with special properties*



Warmup: Fat shattering dimension [Kearns, Schapire94]

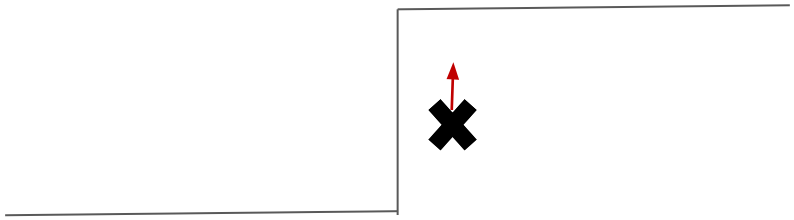
Given a function class C

Fat(C) = size of the largest set of points that is 'shattered' by C.

Shattered means:

For every pattern $(\pm\epsilon, \ldots \pm \epsilon)$, I can find a function in C that `fulfils that pattern of separation' from the points.

E.g. C is the class of threshold functions



Warmup: Fat shattering dimension [Kearns, Schapire94]

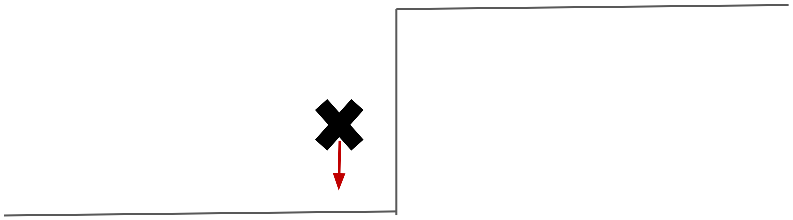
Given a function class C

Fat(C) = size of the largest set of points that is 'shattered' by C.

Shattered means:

For every pattern $(\pm\epsilon, \ldots \pm \epsilon)$, I can find a function in C that `fulfils that pattern of separation' from the points.

E.g. C is the class of threshold functions



Warmup: Fat shattering dimension [Kearns, Schapire94]

Given a function class C

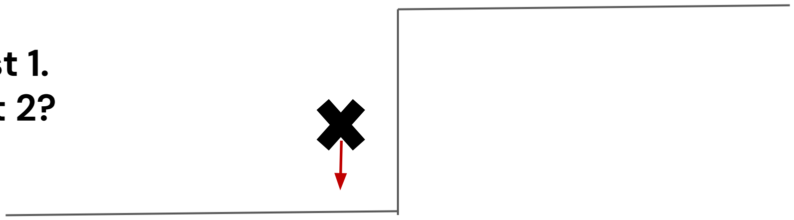
Fat(C) = size of the largest set of points that is 'shattered' by C.

Shattered means:

For every pattern $(\pm\epsilon, \ldots \pm \epsilon)$, I can find a function in C that `fulfils that pattern of separation' from the points.

E.g. C is the class of threshold functions

→ fat$_\epsilon$ (C) is at least 1.
Could it be at least 2?



Warmup: Fat shattering dimension [Kearns, Schapire94]

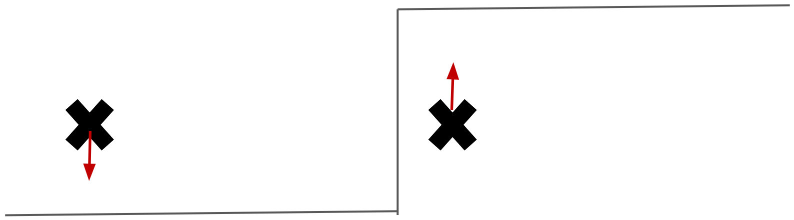
Given a function class C

Fat(C) = size of the largest set of points that is 'shattered' by C.

Shattered means:

For every pattern $(\pm\epsilon, \ldots \pm \epsilon)$, I can find a function in C that `fulfils that pattern of separation' from the points.

E.g. C is the class of threshold functions



Warmup: Fat shattering dimension [Kearns, Schapire94]

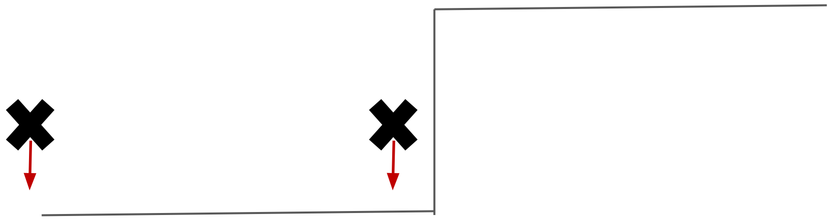
Given a function class C

Fat(C) = size of the largest set of points that is 'shattered' by C.

Shattered means:

For every pattern $(\pm\epsilon, \ldots \pm \epsilon)$, I can find a function in C that `fulfils that pattern of separation' from the points.

E.g. C is the class of threshold functions



Warmup: Fat shattering dimension [Kearns, Schapire94]

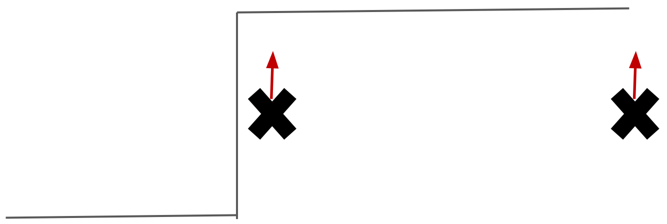
Given a function class C

Fat(C) = size of the largest set of points that is 'shattered' by C.

Shattered means:

For every pattern $(\pm\epsilon, \ldots \pm \epsilon)$, I can find a function in C that `fulfils that pattern of separation' from the points.

E.g. C is the class of threshold functions



Warmup: Fat shattering dimension [Kearns, Schapire94]

Given a function class C

Fat(C) = size of the largest set of points that is 'shattered' by C.

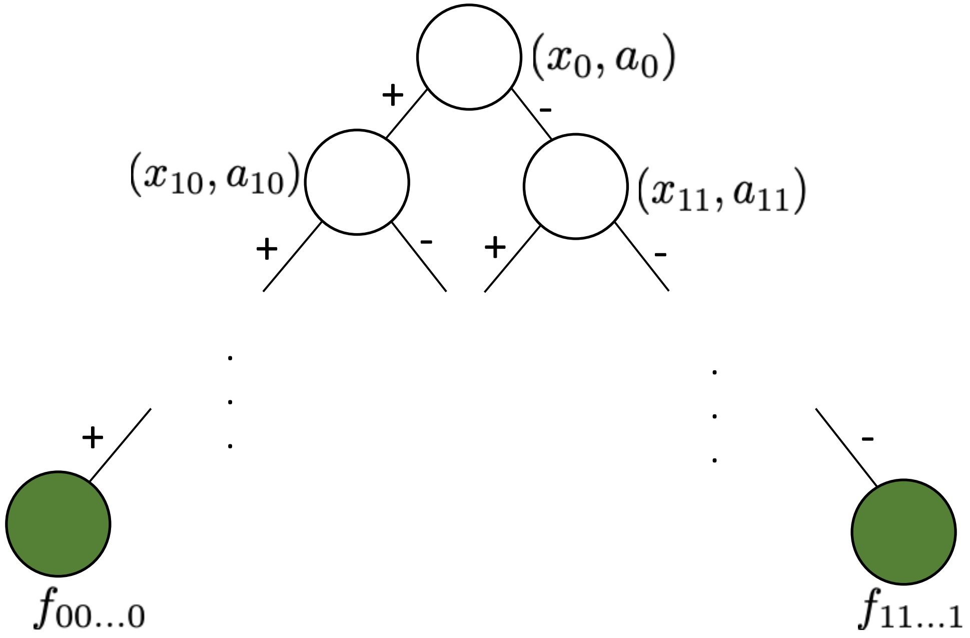
Shattered means:

For every pattern $(\pm\epsilon, \ldots \pm \epsilon)$, I can find a function in C that `fulfils that pattern of separation' from the points.

E.g. C is the class of threshold functions

→ fat_ϵ (C) is at least 1. Could it be at least 2?
Hmm... nope!

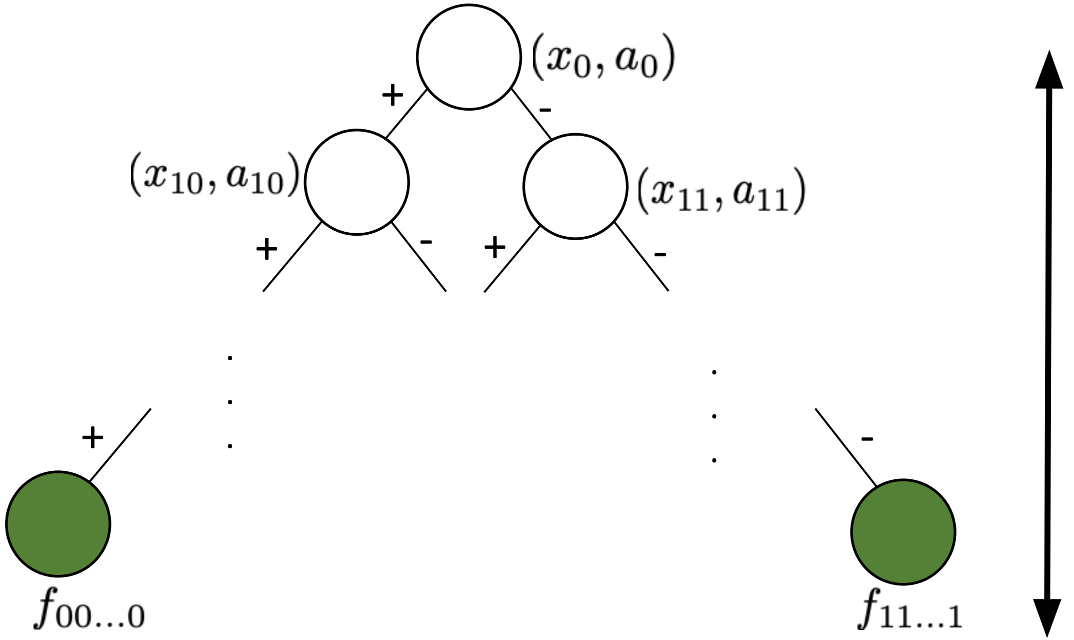
Sequential fat-shattering dimension [Rakhlin, Sridharan, Tewari 2010]



Internal nodes: points in x,y space

Leaf nodes: functions in function class

Sequential fat-shattering dimension [Rakhlin, Sridharan, Tewari 2010]



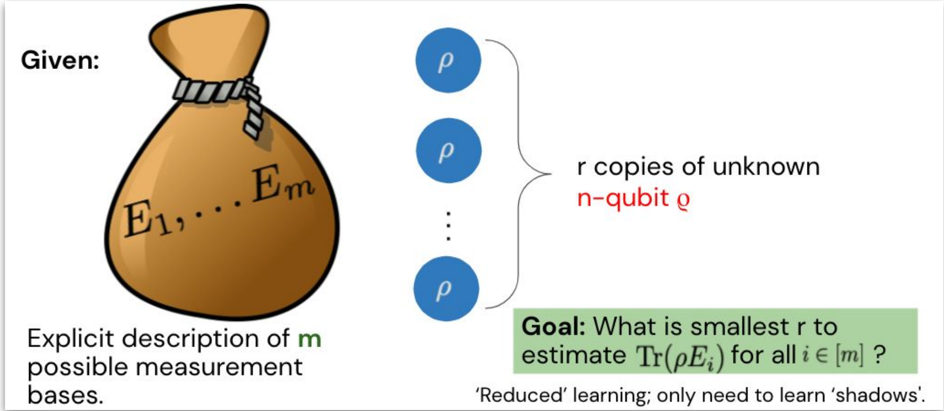
sfat = the depth of the deepest such tree

(*sfat* > *fat*)

For every path from root to leaf, there is a function that 'fulfils that pattern of separation'.
Difference from *fat*: points depend on earlier points.

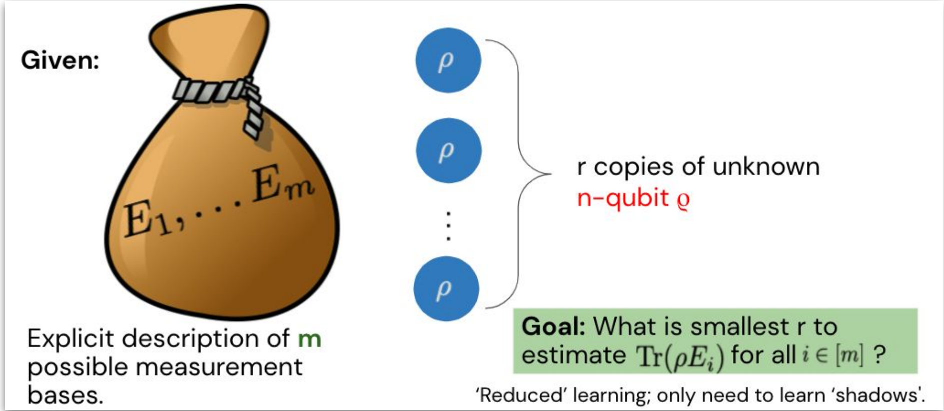
Application: faster shadow tomography

What's the complexity (=min r) of shadow tomography?



Application: faster shadow tomography

What's the complexity (=min r) of shadow tomography?



Trivial bound:

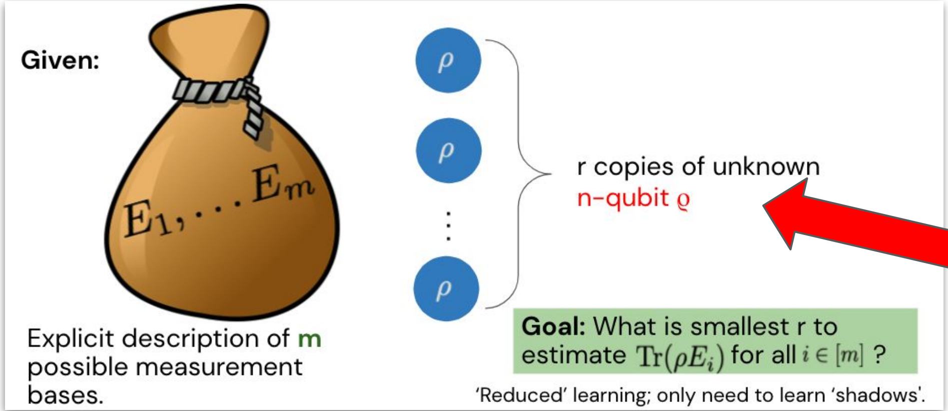
$$O(\mathrm{poly}(2^n, m))$$

Tomography to figure out ρ completely

Create "cheatsheet": Perform all m measurements on ρ, record outcome

Application: faster shadow tomography

What's the complexity (=min r) of shadow tomography?



Trivial bound:

$$O(\mathrm{poly}(2^n, m))$$

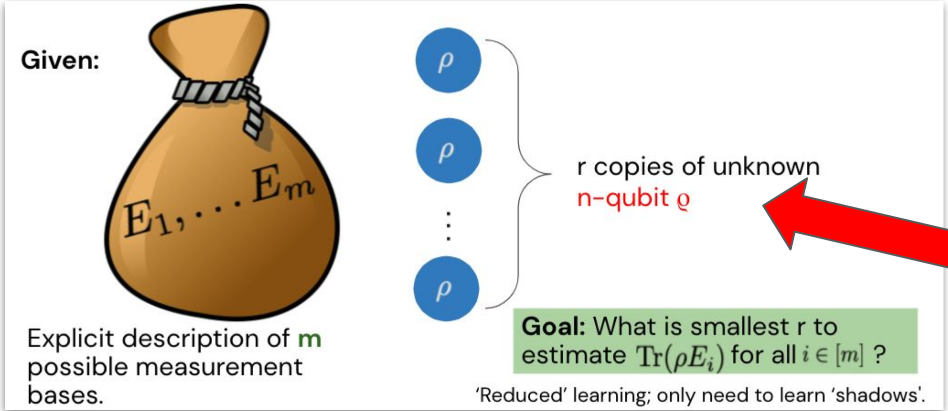
Our theorem (informal):

If ρ comes from some class $\mathcal{C}$ then

the sample complexity of shadow tomography is $O(\mathrm{poly}(\mathrm{sfat}(\mathcal{C}), \log m))$

Application: faster shadow tomography

What's the complexity (=min r) of shadow tomography for interesting classes of states?



Trivial bound:

$$O(\mathrm{poly}(2^n, m))$$

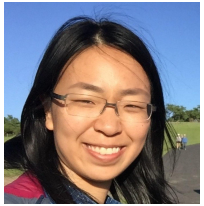
Our theorem (informal):
If ρ comes from some class $\mathcal{C}$ then

the sample complexity of shadow tomography is $O(\mathrm{poly}(\underbrace{\mathsf{sfat}(\mathcal{C})}^{< n}, \log m))$

e.g. for $\mathcal{C}$= low-rank states, bosonic states, noisy states

arXiv: 2102.07171/NeurIPS 2021

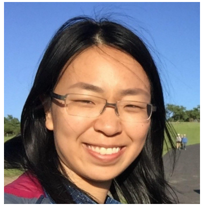
Part II takeaways

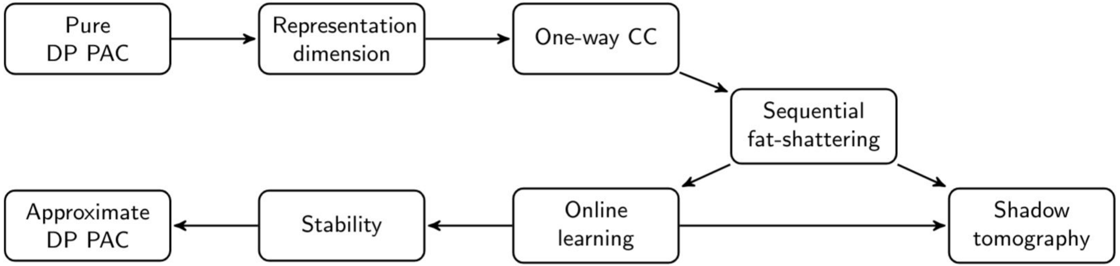

- There exist simpler/"reduced" versions of tomography – less resource-intensive but still capture "useful" information about state.

arXiv: 2102.07171/NeurIPS 2021

Part II takeaways

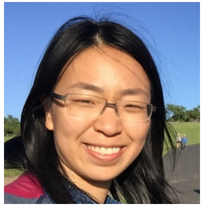




- There exist simpler/"reduced" versions of tomography – less resource-intensive but still capture "useful" information about state.
- Equivalences between **reduced models of quantum learning** allow us to port algorithms from one model to another.

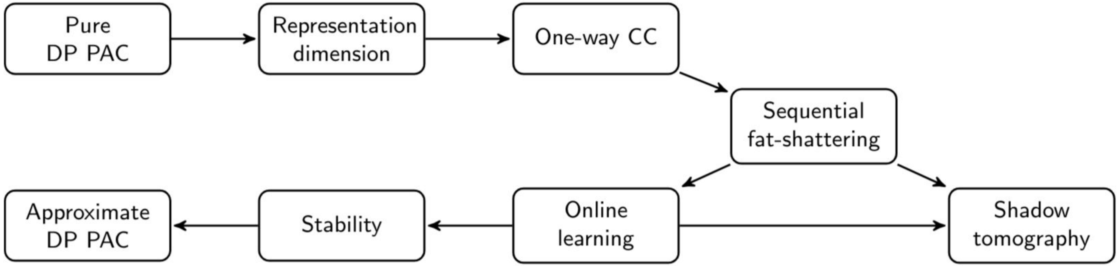


arXiv: 2102.07171/NeurIPS 2021

Part II takeaways


- There exist simpler/"reduced" versions of tomography – less resource-intensive but still capture "useful" information about state.
- Equivalences between **reduced models of quantum learning** allow us to port algorithms from one model to another.



- Application: speedups in **shadow tomography** and other learning models.

Open questions

- [Sampling problems on NISQ] Can we come up with a combinatorial dimension for learning *distributions*?

- [Learning vs Classical Simulation] We saw that simulability doesn't imply learnability. Does learnability imply simulability?

- What other notions of learning could avoid the curse of exponentiality?