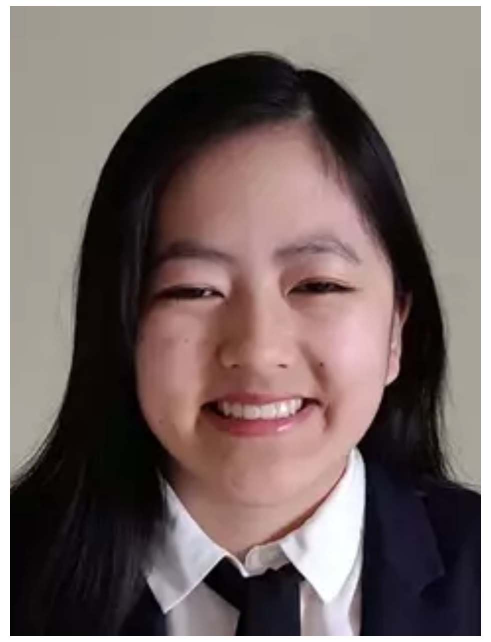
Learning quantum states and unitaries of bounded gate complexity

arXiv:2310.19882

Haimeng Zhao*
haimengzhao@icloud.com
Tsinghua => Caltech



Laura Lewis*

Ishaan Kannan*

Yihui Quek

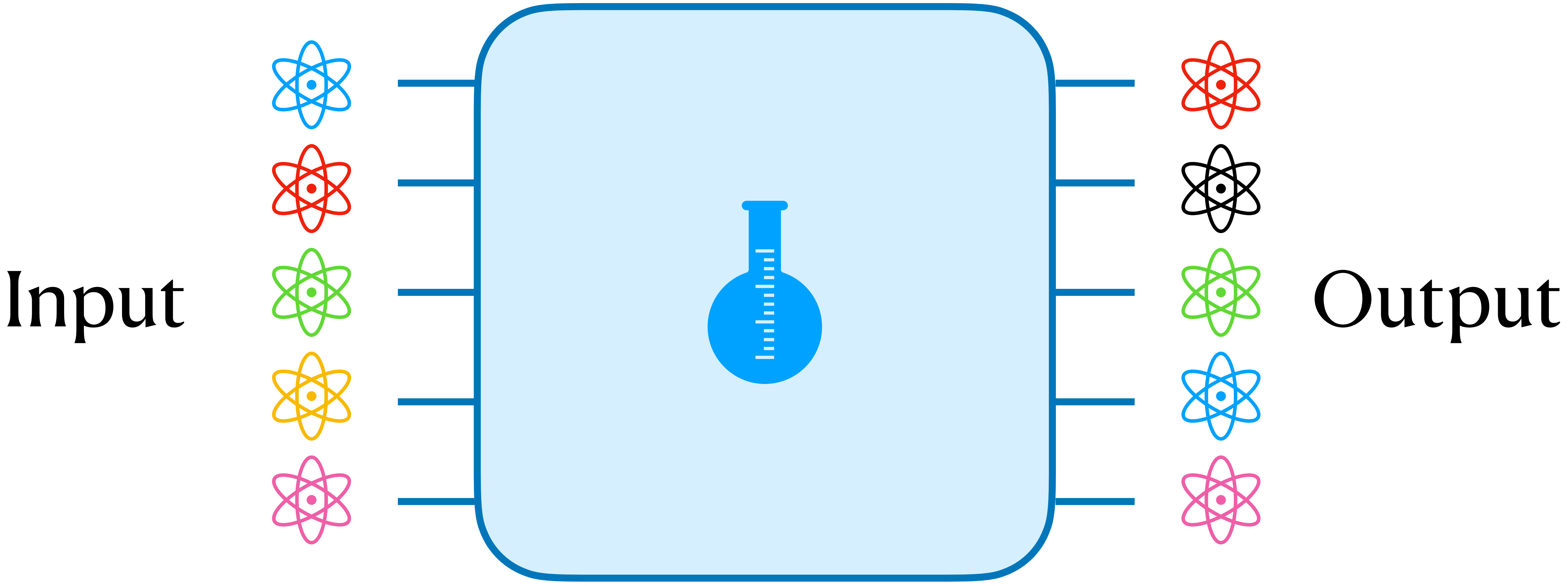
Hsin-Yuan Huang (Robert)

Matthias Caro

AQIS 2024, Sapporo, 30th Aug. 2024

Motivation
How to learn from Nature?

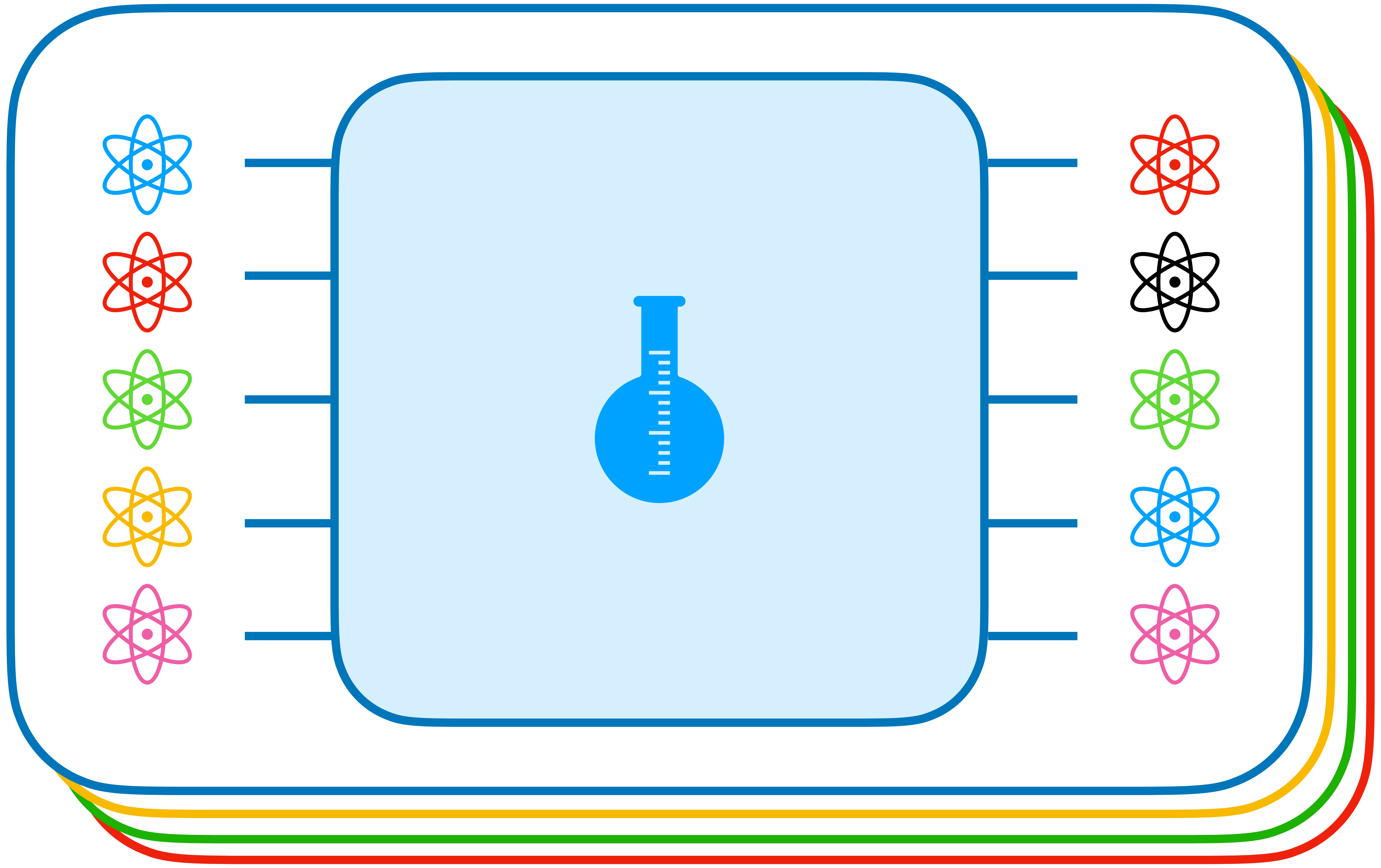
- Do experiments



unknown quantum process
in experiments

Motivation
How to learn from Nature?

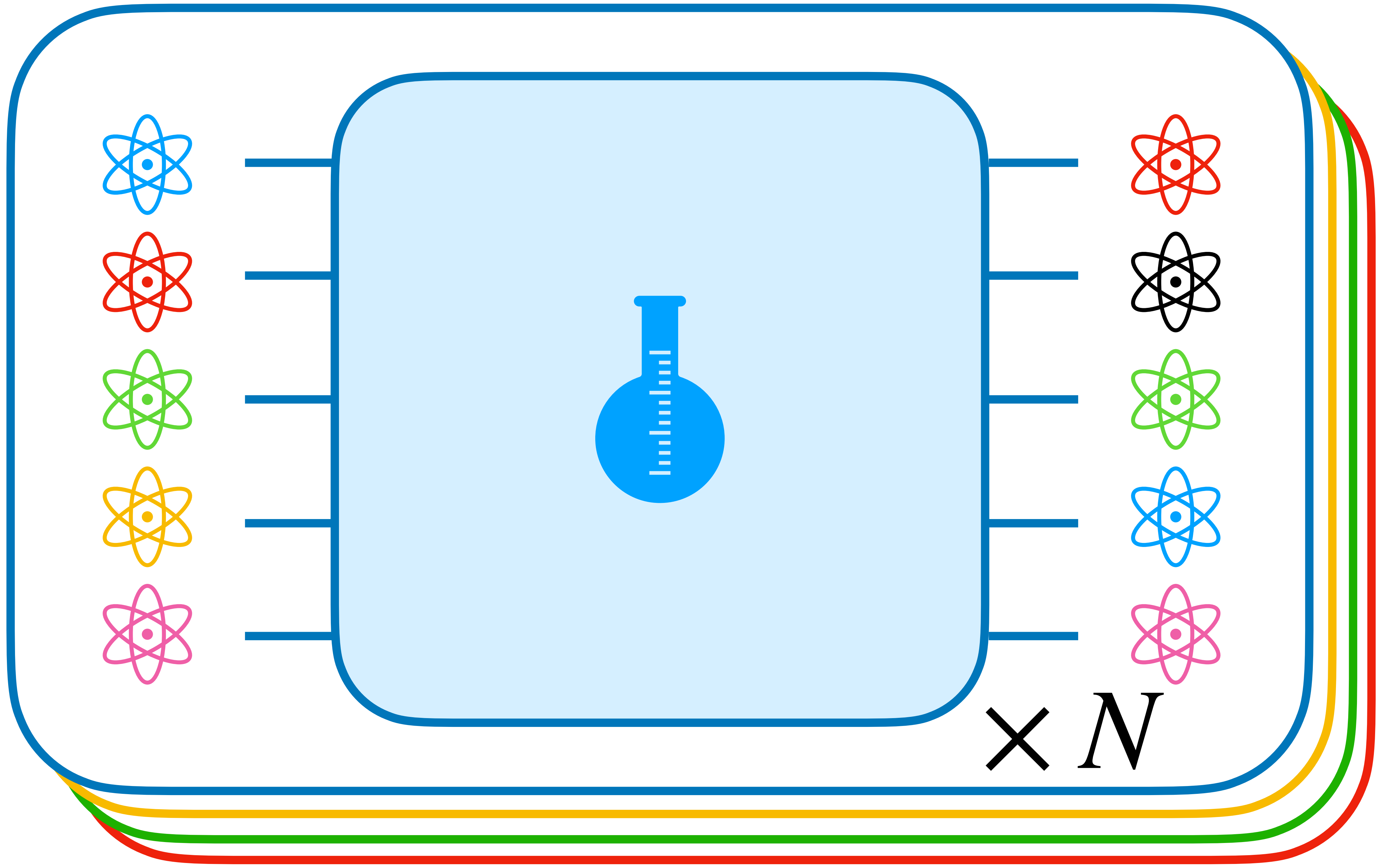
- Do experiments => Collect many samples



$\times N$

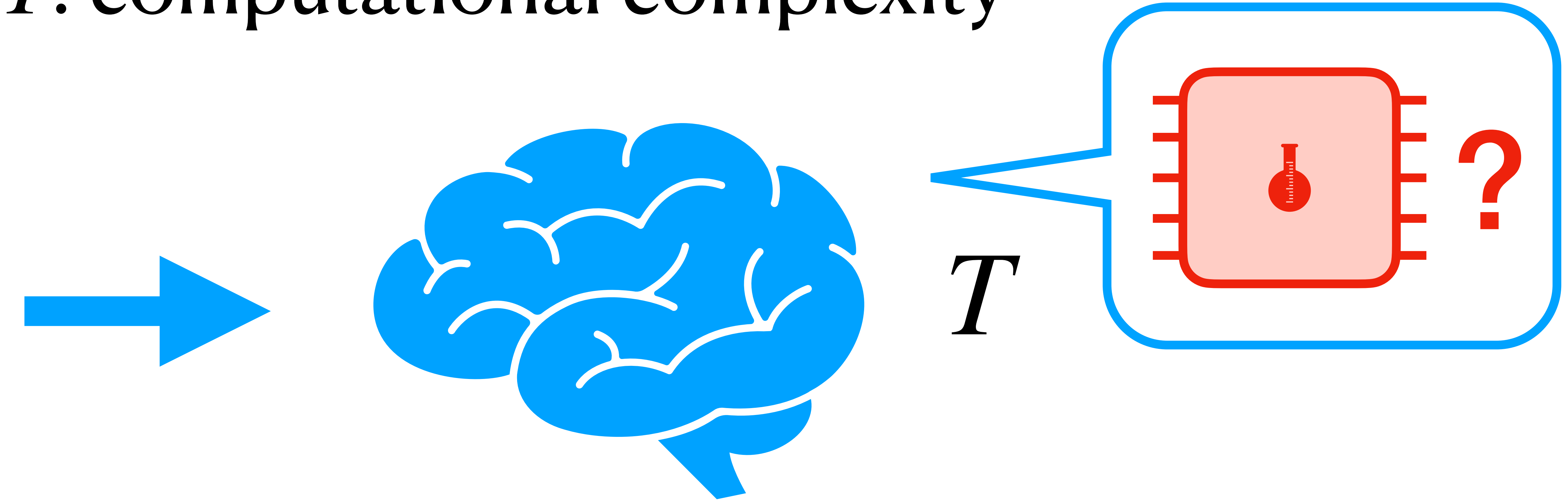
Motivation
How to learn from Nature?

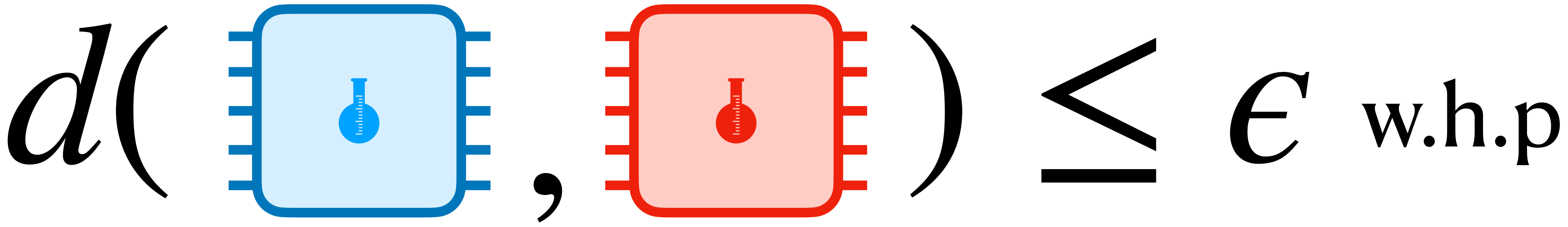
- Do experiments => Collect many samples => Try to learn the underlying mechanism



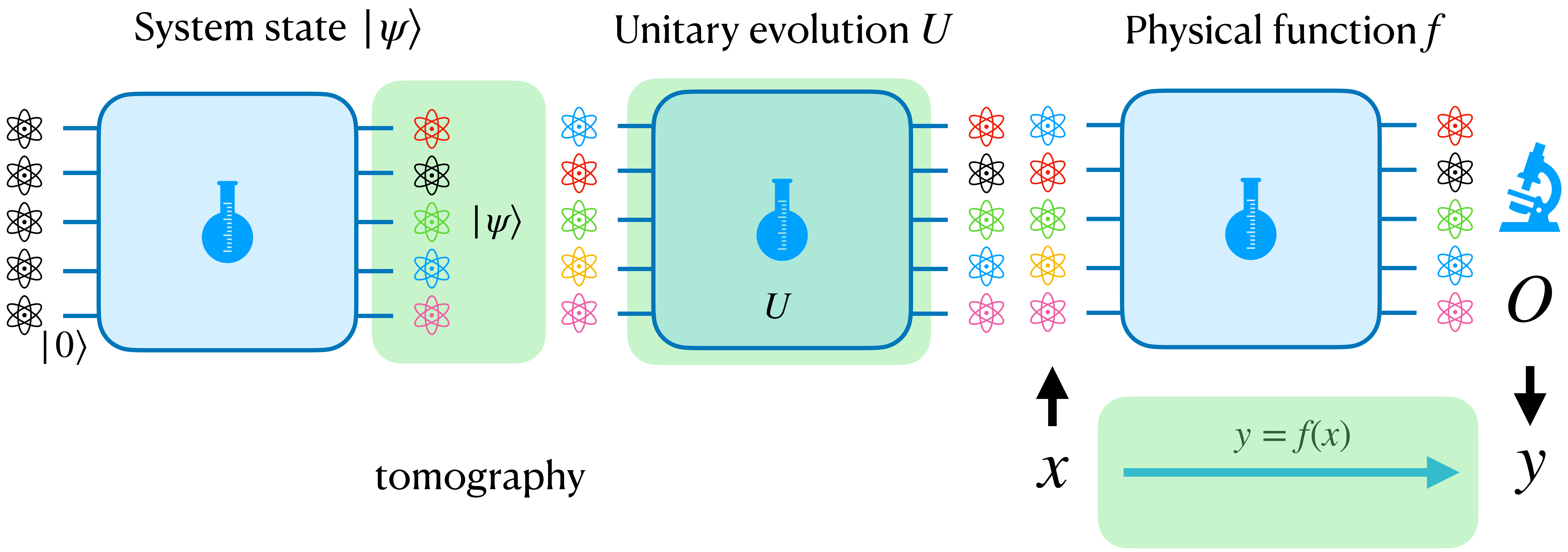
N: sample complexity

T: computational complexity



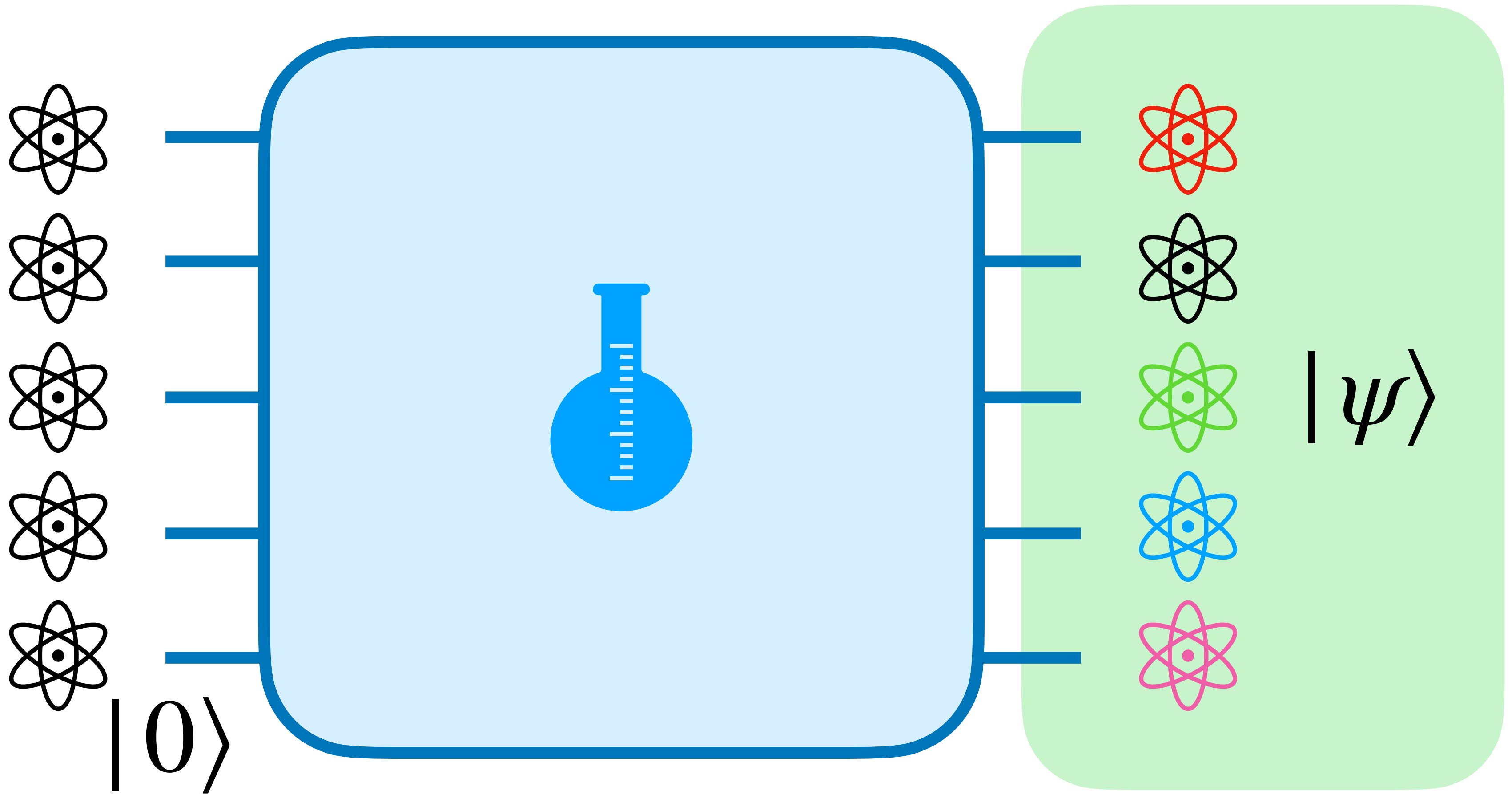


Motivation
What to learn?



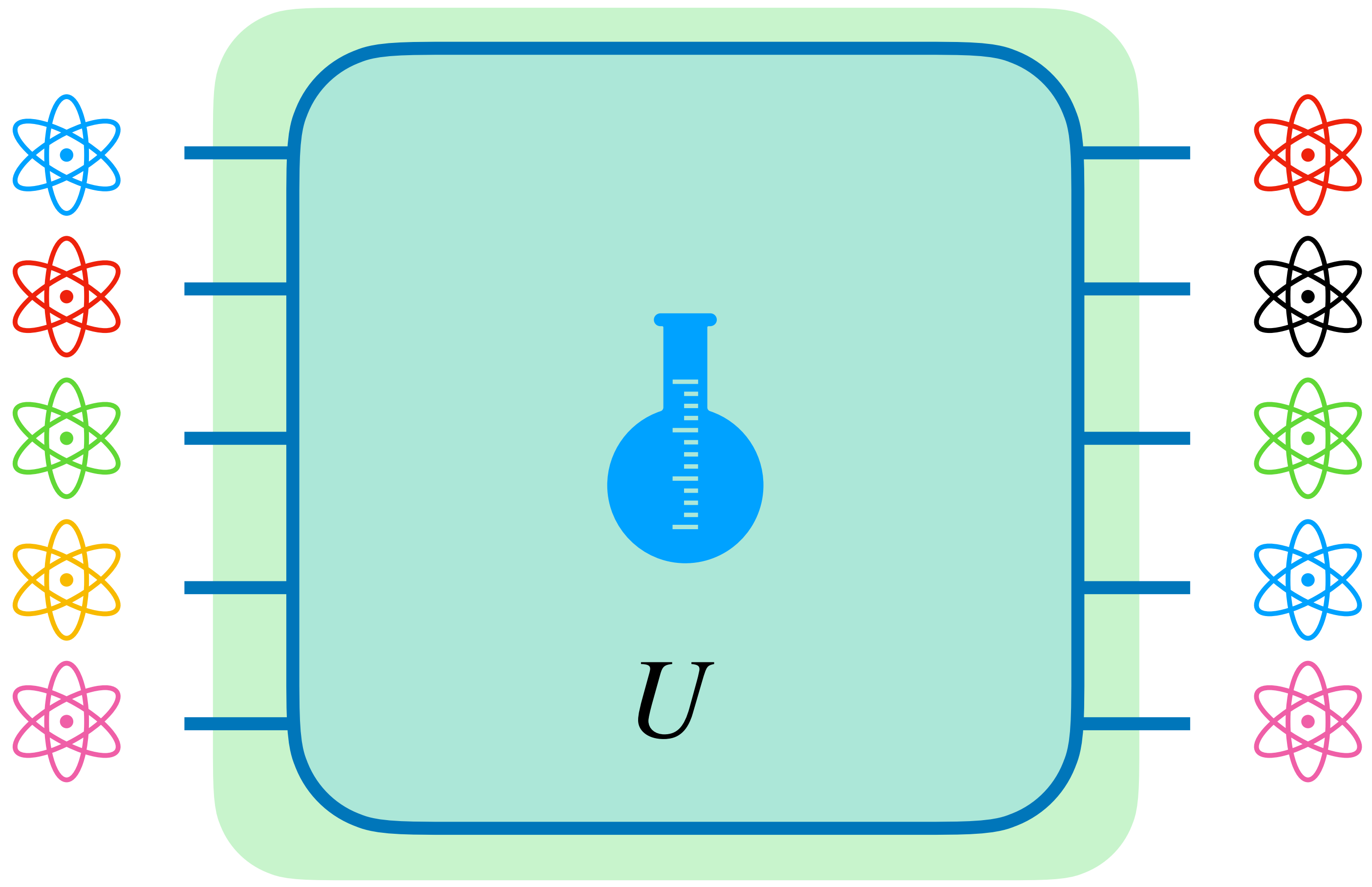
Motivation
Learning is hard in general!

System state $|\psi\rangle$



$$N = \Theta(2^n)$$

Unitary evolution U



$$N = \Theta(4^n)$$

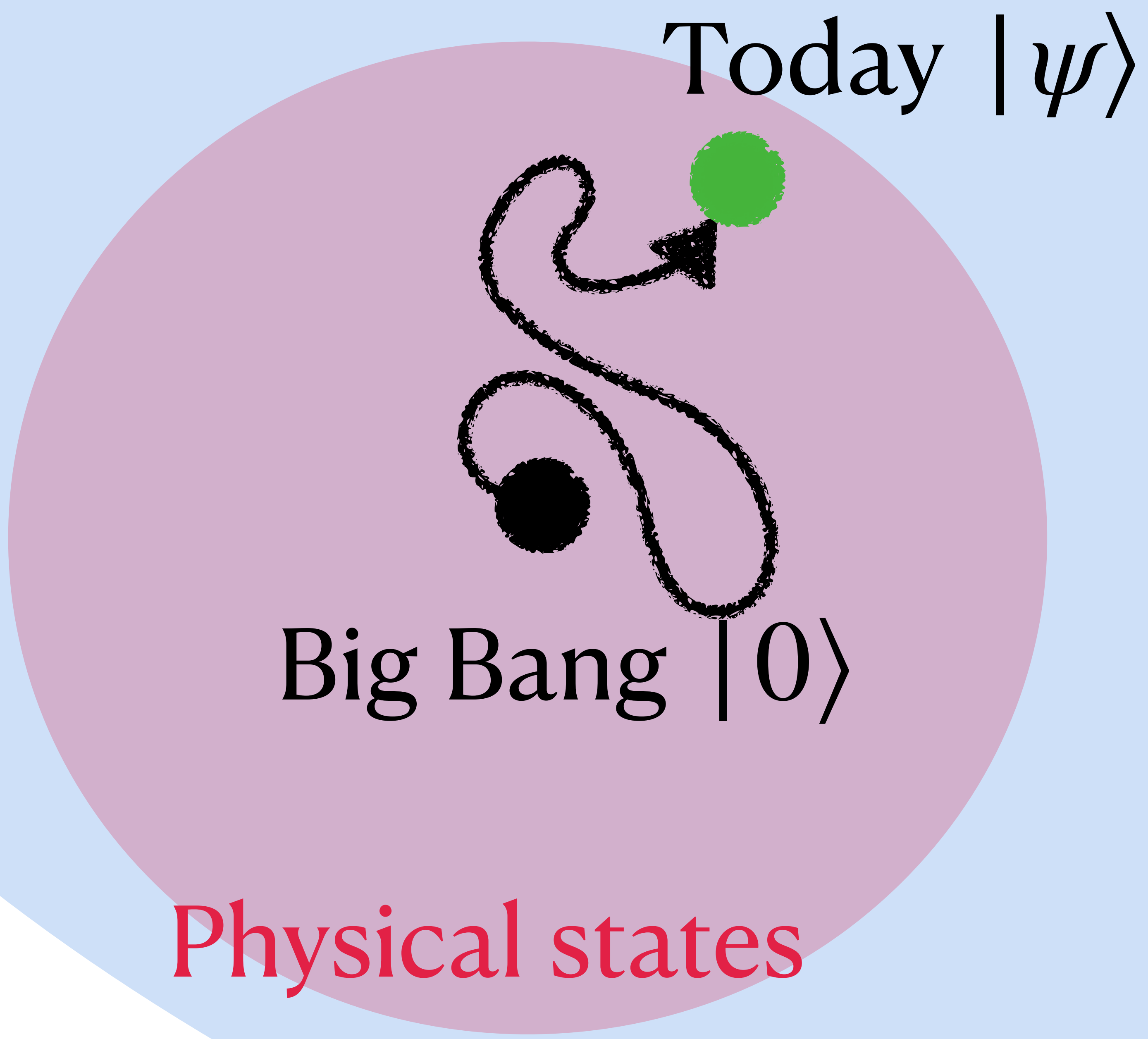
How is learning even possible?

$n \sim 10^{23}$!

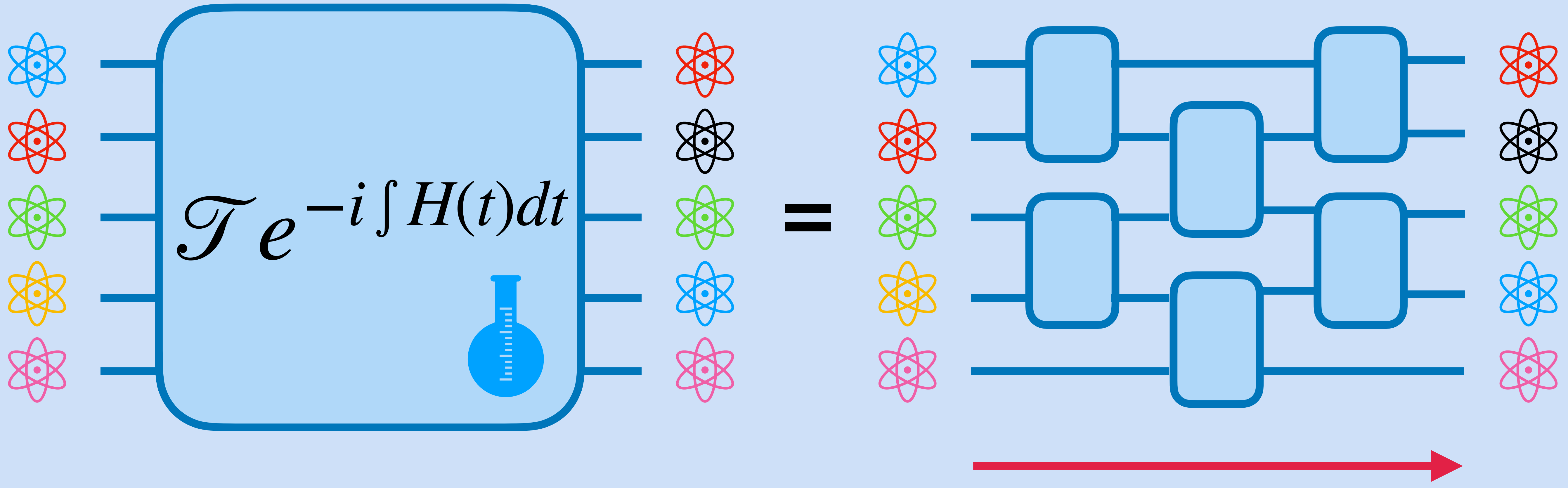
O'Donnell and Wright, STOC 2016
Haah, Kothari, O'Donnell, Tang, FOCS 2023

Illusive Hilbert space $\sim 2^n$

Motivation
Physical constraints



reachable in bounded time



local Hamiltonian
bounded time

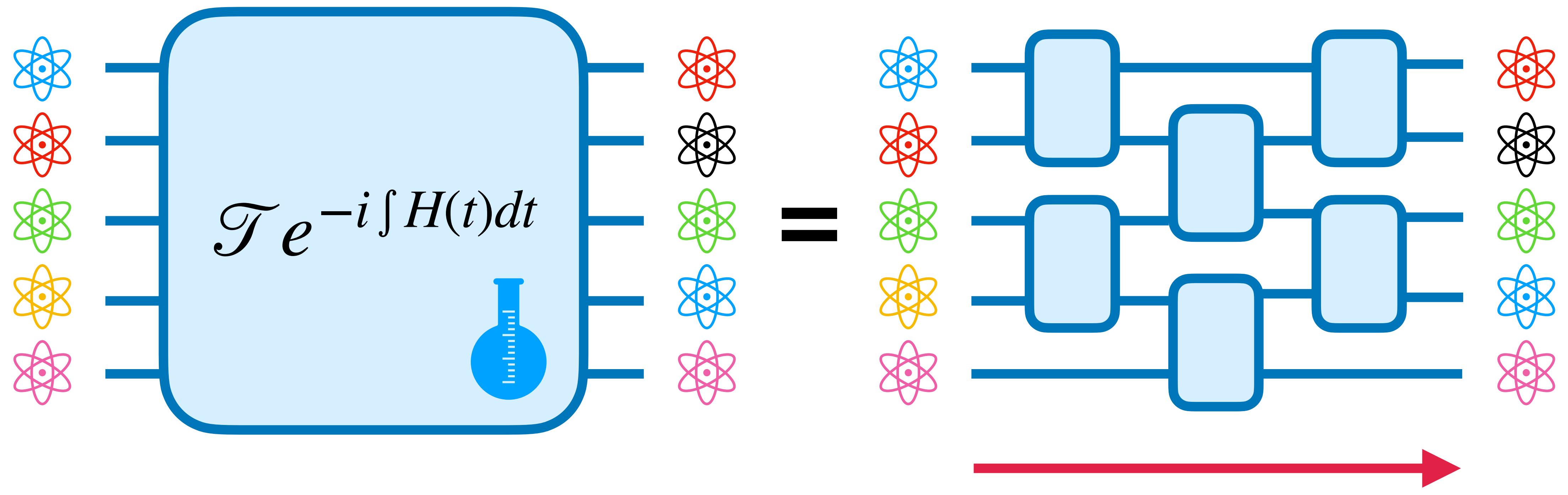
gates complexity $\leq G$

G two-qubit gates

In this work, we don't need "geometric" locality, nor discrete gate-set

Poulin, Qarry, Somma, Verstraete, PRL 2011

Main Question



local Hamiltonian
bounded time

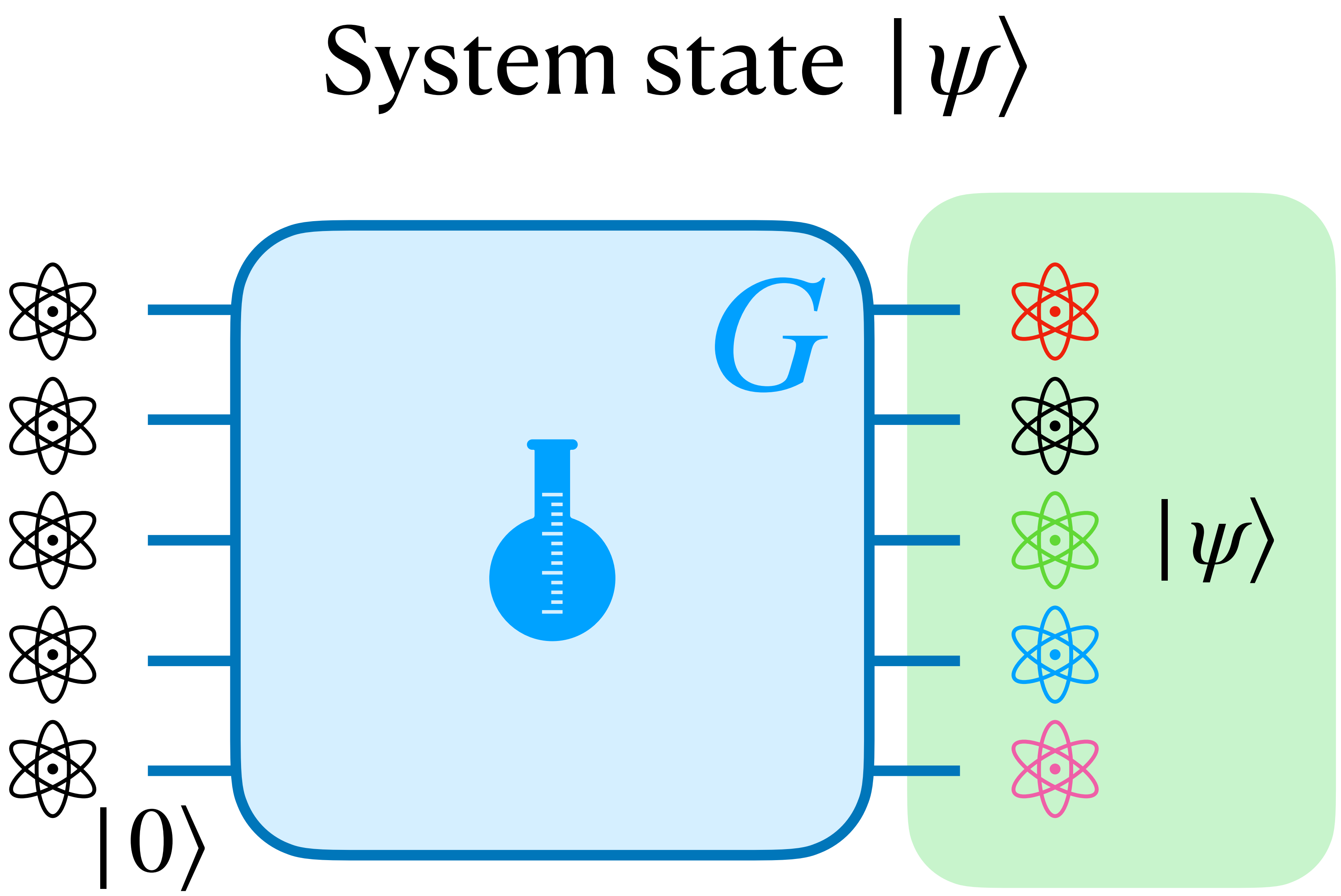
circuit complexity $\leq G$
G two-qubit gates

Can we efficiently learn states/unitaries of bounded gate complexity?

Applications on near-term quantum devices: limited G (both digital and analog)

Relating different notions of complexity: the complexity of learning and creating

Results
sample complexity



Learning $|\psi\rangle$ in ϵ trace distance requires $N = \tilde{\Theta}\left(\frac{G}{\epsilon^2}\right)$ samples.

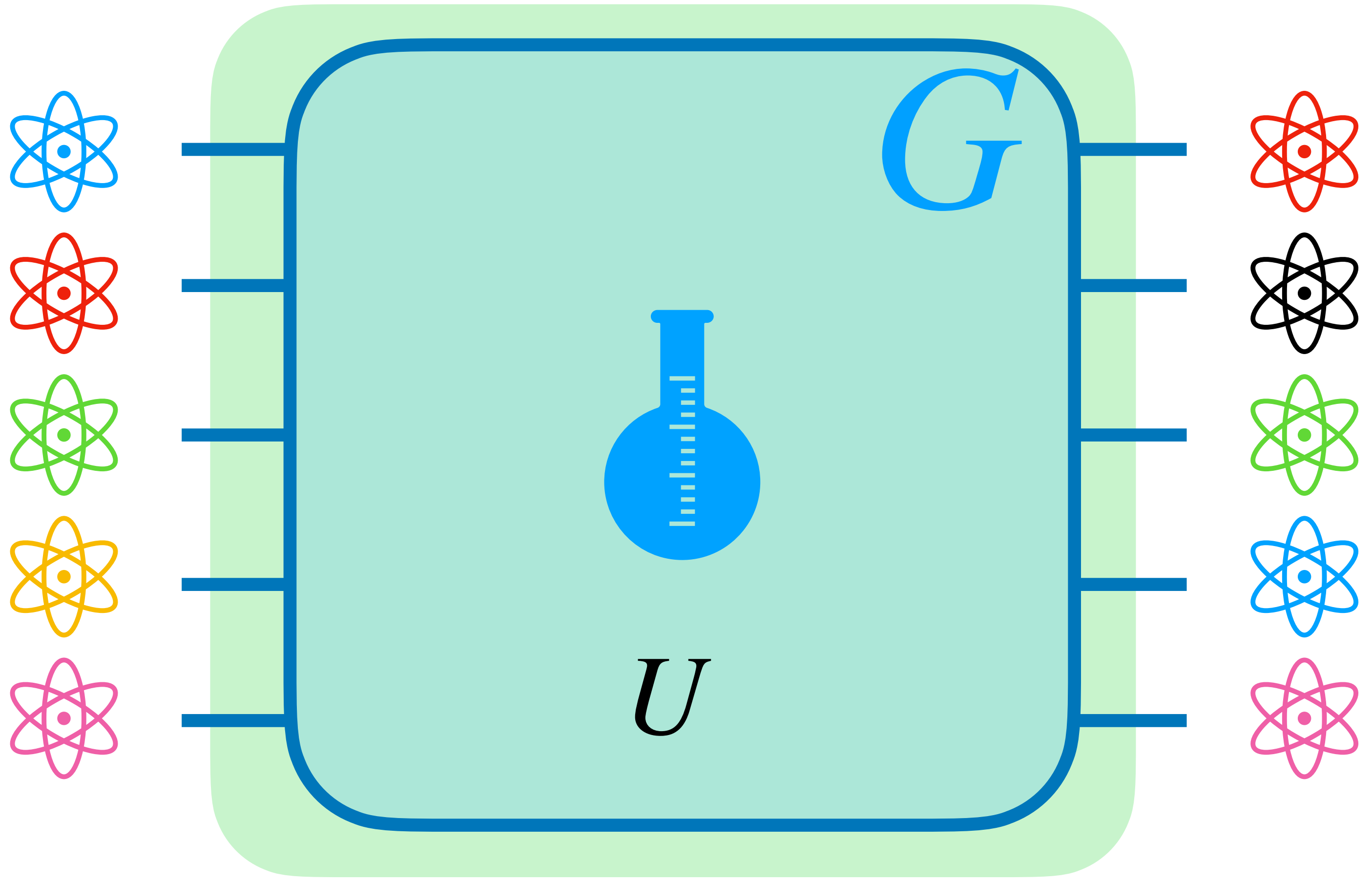
1. Complexity of learning = complexity of creating (information theoretically)
2. Completely independent of system size n. Can learn $n = 10000, G = 10$ states!
3. Non-adaptive/incoherent scheme is already optimal.

Results
sample complexity

= trace distance between Choi states

$$\mathrm{d_{avg}}(U, V) = \sqrt{\mathbb{E}_{|\psi\rangle}[\mathrm{d_{tr}}(U|\psi\rangle, V|\psi\rangle)^2]}$$

Unitary evolution U



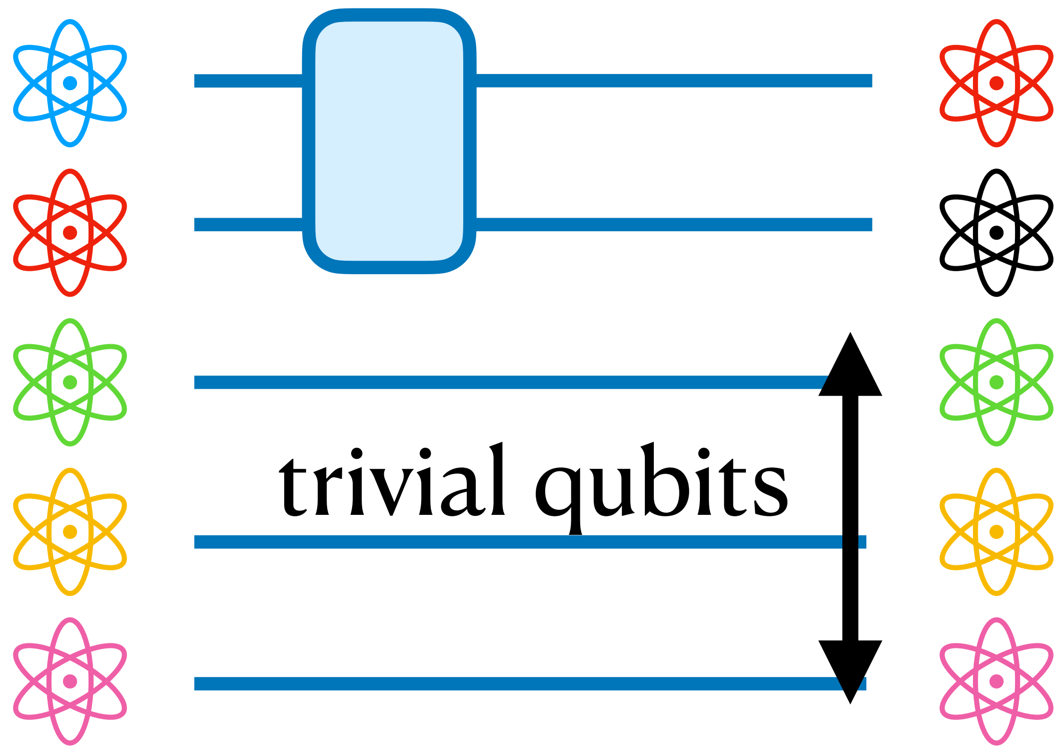
Learning U in average-case distance

requires $N = \tilde{\Theta}(G)$ queries.

1. Non-adaptive/incoherent query is already optimal (in G).
2. Learning in worst-case distance (diamond norm) requires $N = \exp(\Omega(\min\{G, n\}))/\epsilon$
3. ϵ-dependence: $\tilde{O}(\min\{1/\epsilon^2, \sqrt{2^n}/\epsilon\})$, $\Omega(1/\epsilon)$, Heisenberg scaling open

Grover

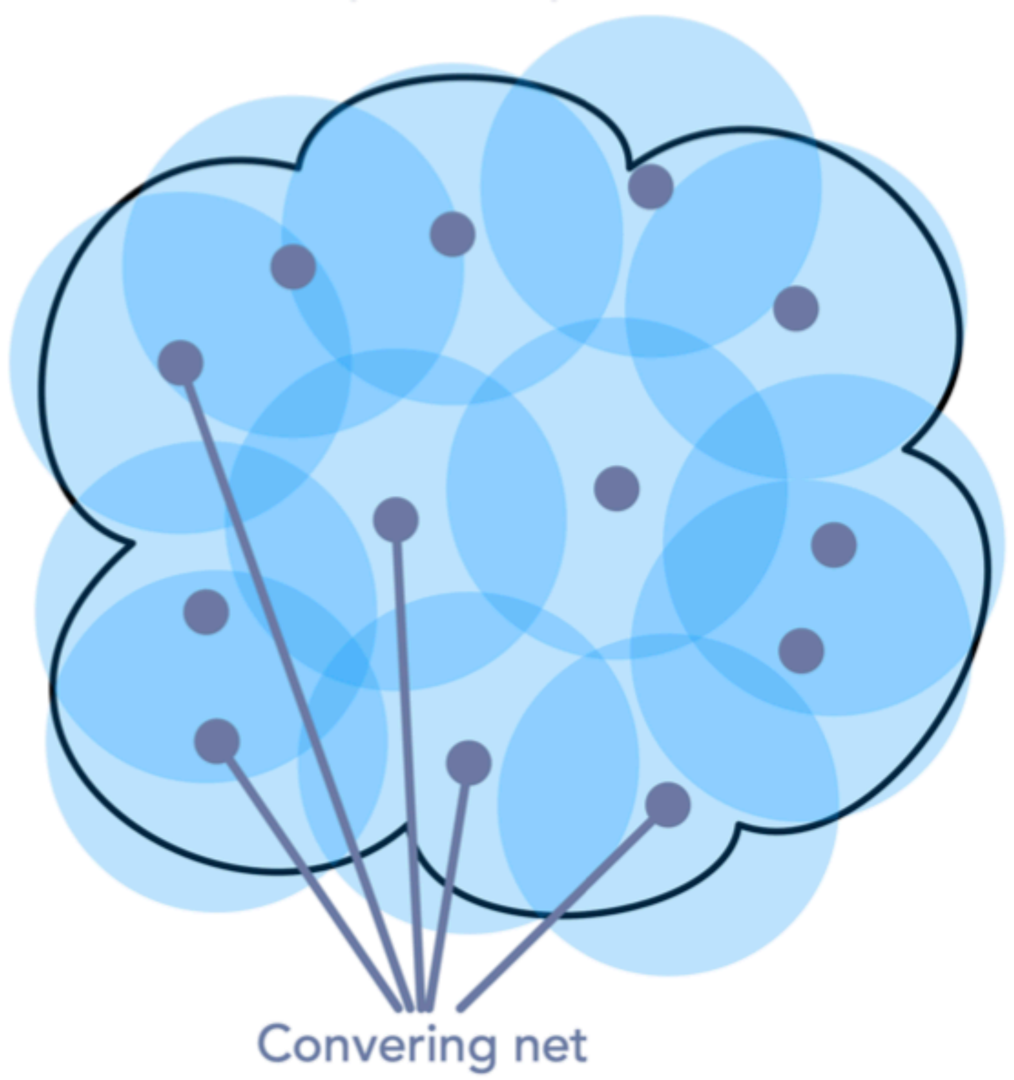
Proof sketch
Upper bound: the learning algorithm



1. **Junta learning:** measure to identify non-trivial qubits, remove n-dependence

2. **Hypothesis selection:** construct a covering net $\mathcal{N}$ that covers the set of G-gate states with ϵ-balls
$\log|\mathcal{N}| = \tilde{\Theta}(G)$

3. Find the best candidate by estimating all distances with classical shadow $N = O\left(\frac{\log|\mathcal{N}|}{\epsilon^2}\right) \leq \tilde{O}\left(\frac{G}{\epsilon^2}\right)$

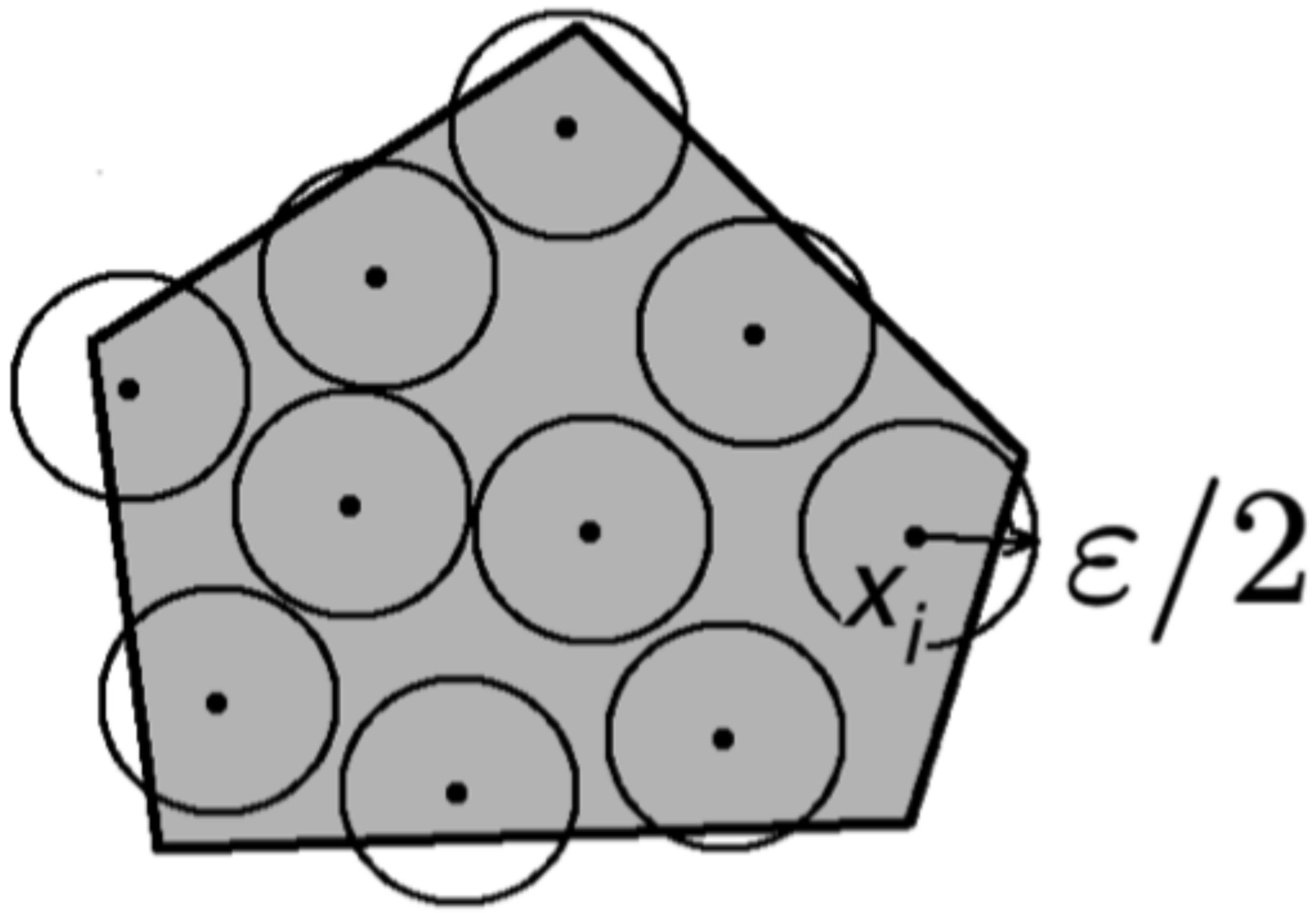


Unitary: $O(G\min\{1/\epsilon^2, \sqrt{2^n}/\epsilon\})$

via Choi states + quantum phase estimation

Chen, Nadimpalli, Yuen, SODA 2023
Bădescu and O'Donnell, STOC 2021
Huang, Kueng, Preskill, Nat Phys 2020

Proof sketch
Lower bound: information theory



1. **Learning** => **distinguish** elements of a packing net $\mathscr{P}$

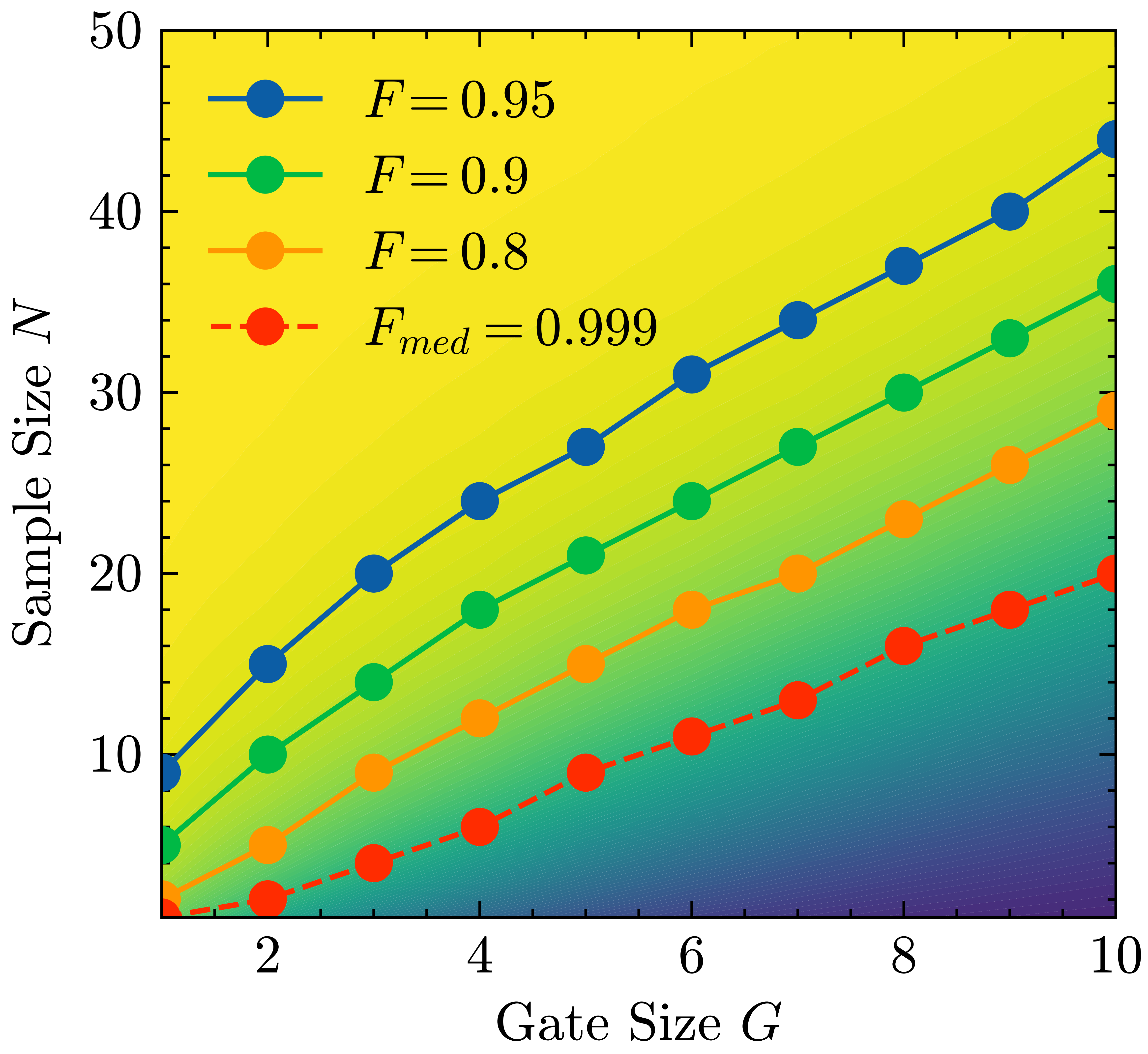
   $\log|\mathscr{P}| = \tilde{\Theta}(G)$

2. **Fano's inequality:** distinguishing requires $\Omega(\log|\mathscr{P}|)$ bits of info

3. **Holevo's theorem:** each sample gives $\tilde{O}(\epsilon^2)$ bits of info => $N \geq \Omega\left(\frac{\log|\mathscr{P}|}{\epsilon^2}\right) = \tilde{\Omega}\left(\frac{G}{\epsilon^2}\right)$

$\log|\mathscr{P}| \approx \log|\mathscr{N}|$: a general way to prove matching sample complexity bound

Numerical experiments



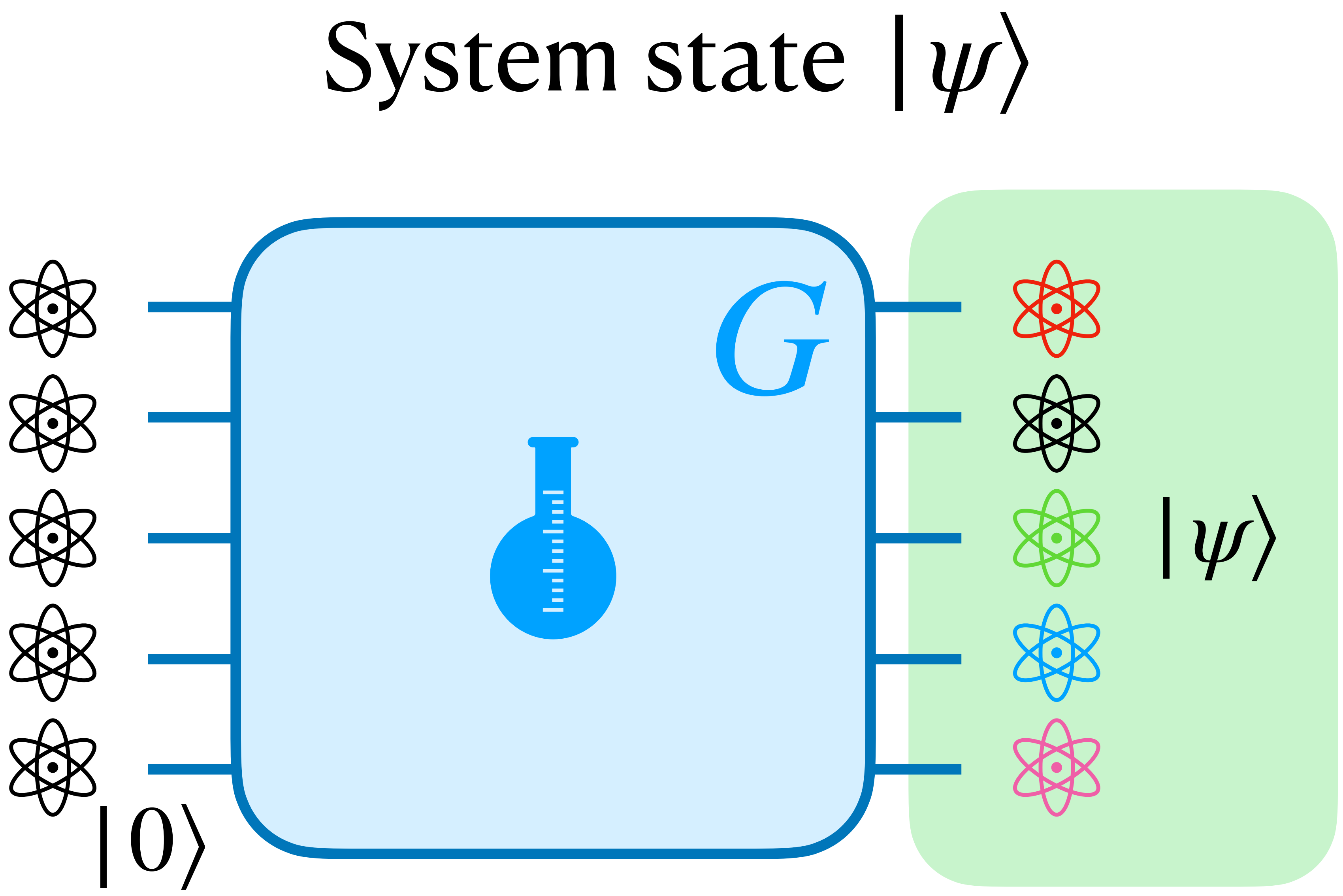
Learning random G-gate states on $n = 10000$ qubits

Sample linear in G, runtime e^G

GitHub

haimengzhao/bounded-gate-tomography

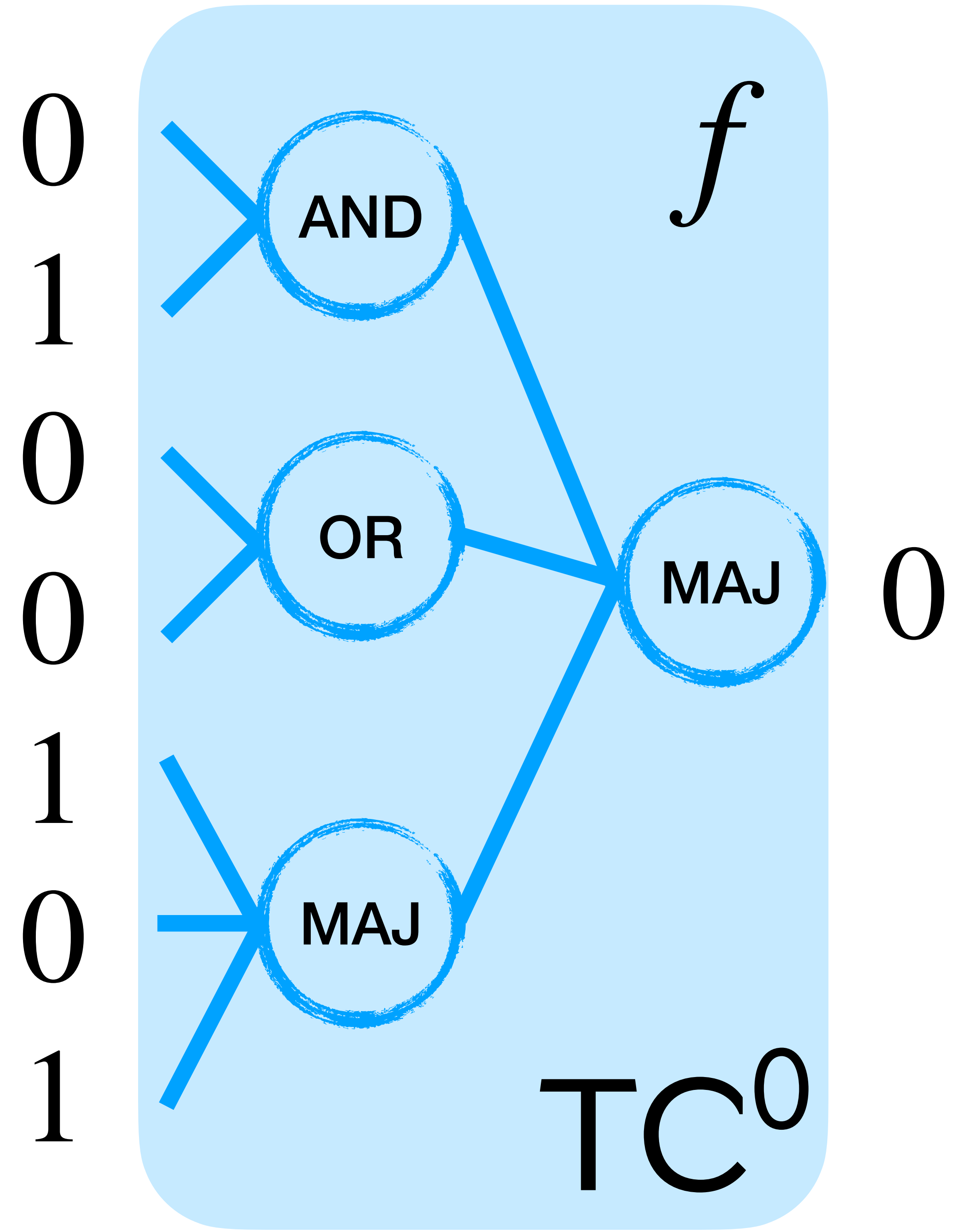
Results
computational complexity



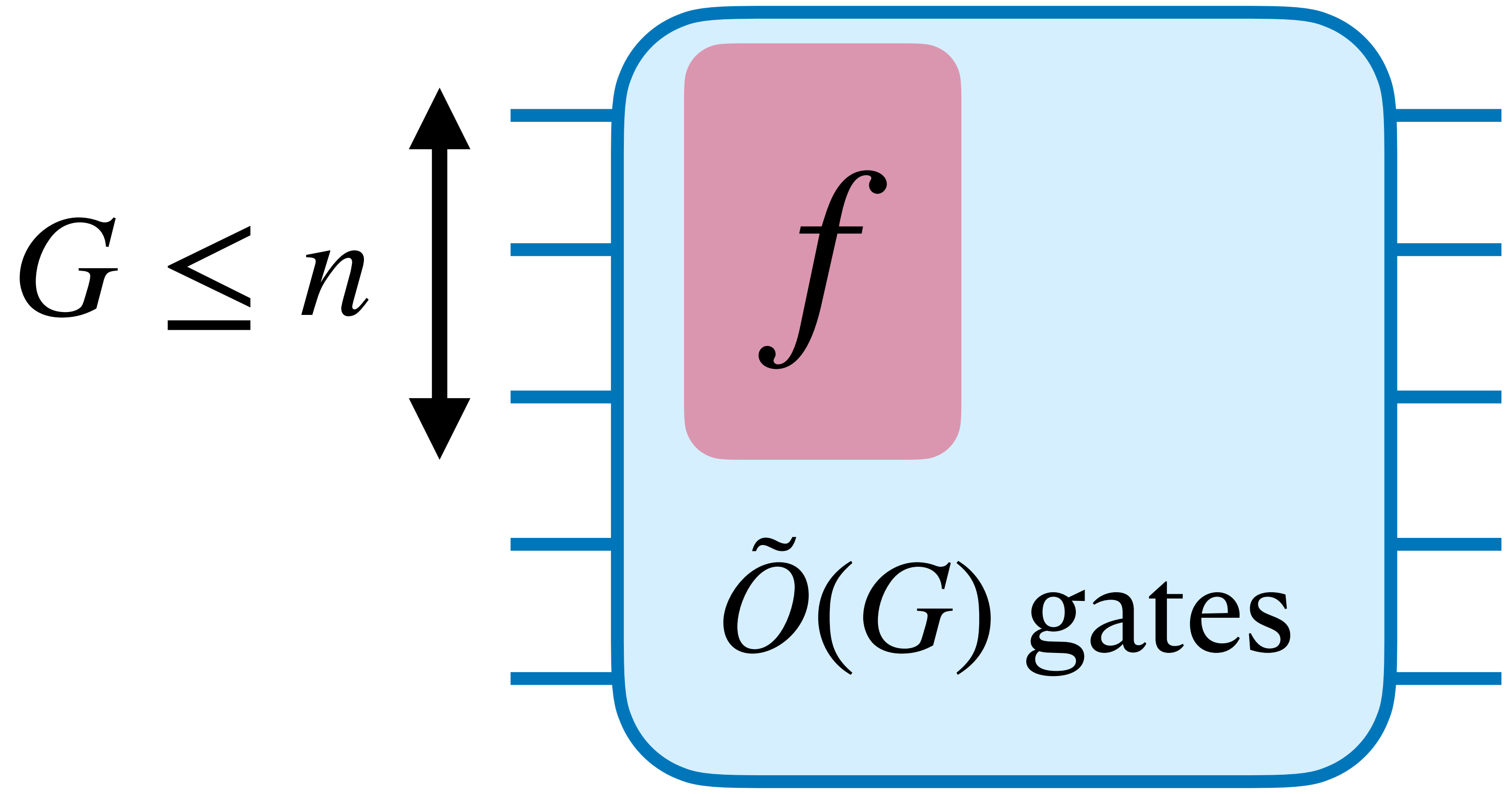
Learning $|\psi\rangle$ in ϵ trace distance requires $T = \exp(\Omega(\min\{G, n\}))$ time, if RingLWE is sub-exponential hard.

1. Complexity of learning = $e^{\text{complexity of creating}}$ (computationally), efficient $\log n$ gates
2. Even for quantum learners: RingLWE is expected to be hard for quantum computers.
3. Worst-case statement: efficiency possible with additional assumptions.
4. Same for average-case unitary learning

Proof sketch
computational complexity



pseudorandom functions



embed into states/unitaries

Learning breaks PRF/PRS

requires $T = e^{\Omega(G)}$ time

$G \leq n$

Ji, Liu, Song, CRYPTO 2018

Arunachalam, Grilo, Sundaram, SICOMP 2021

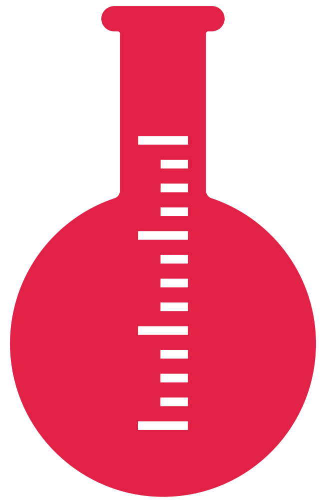
Message

sample size — gate complexity

$$N \approx \log T \approx G \approx t$$

compute time — evolution time

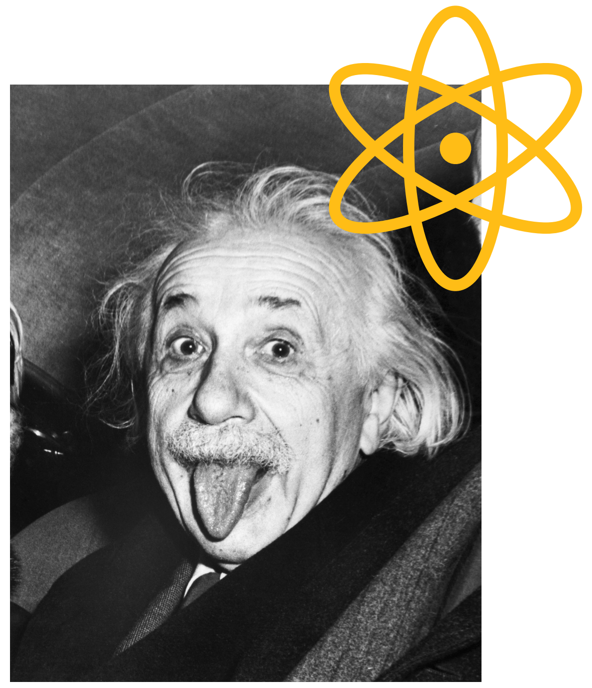
(Brown-Susskind conjecture)



Learning physical states/unitaries is information-theoretically easy,

but computationally hard!

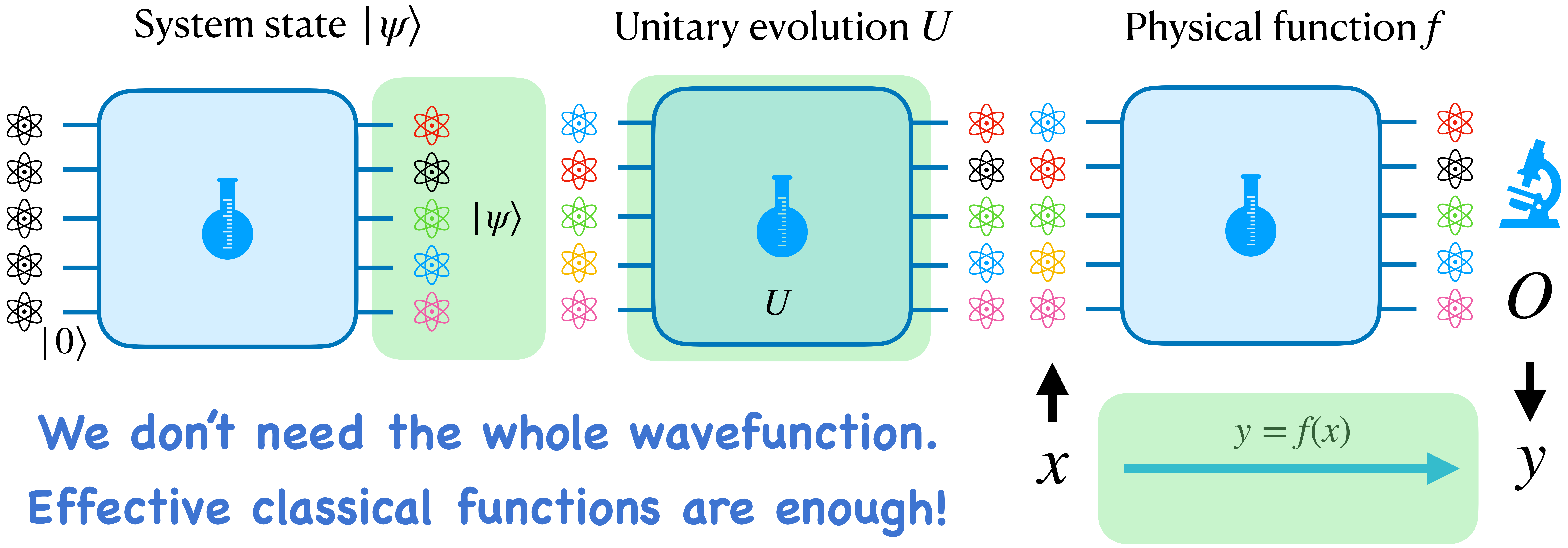
Reminder:
Doing math proofs needs
no data but is NP-hard!



In scientific discovery, a few samples might already be enough,

but coming up with a theory requires some real genius!

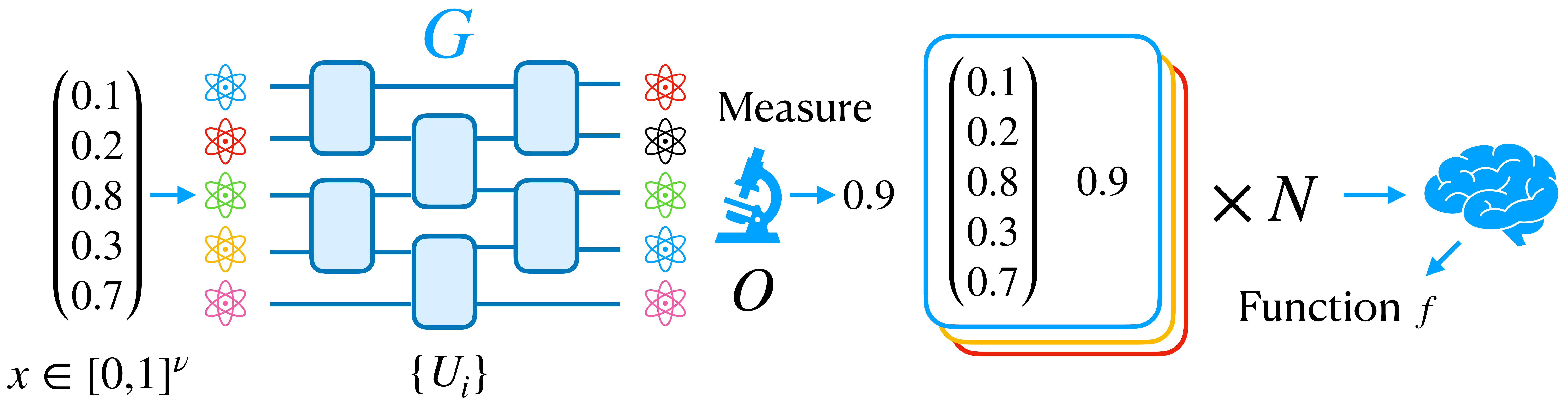
Motivation
What to learn?



We don't need the whole wavefunction.

Effective classical functions are enough!

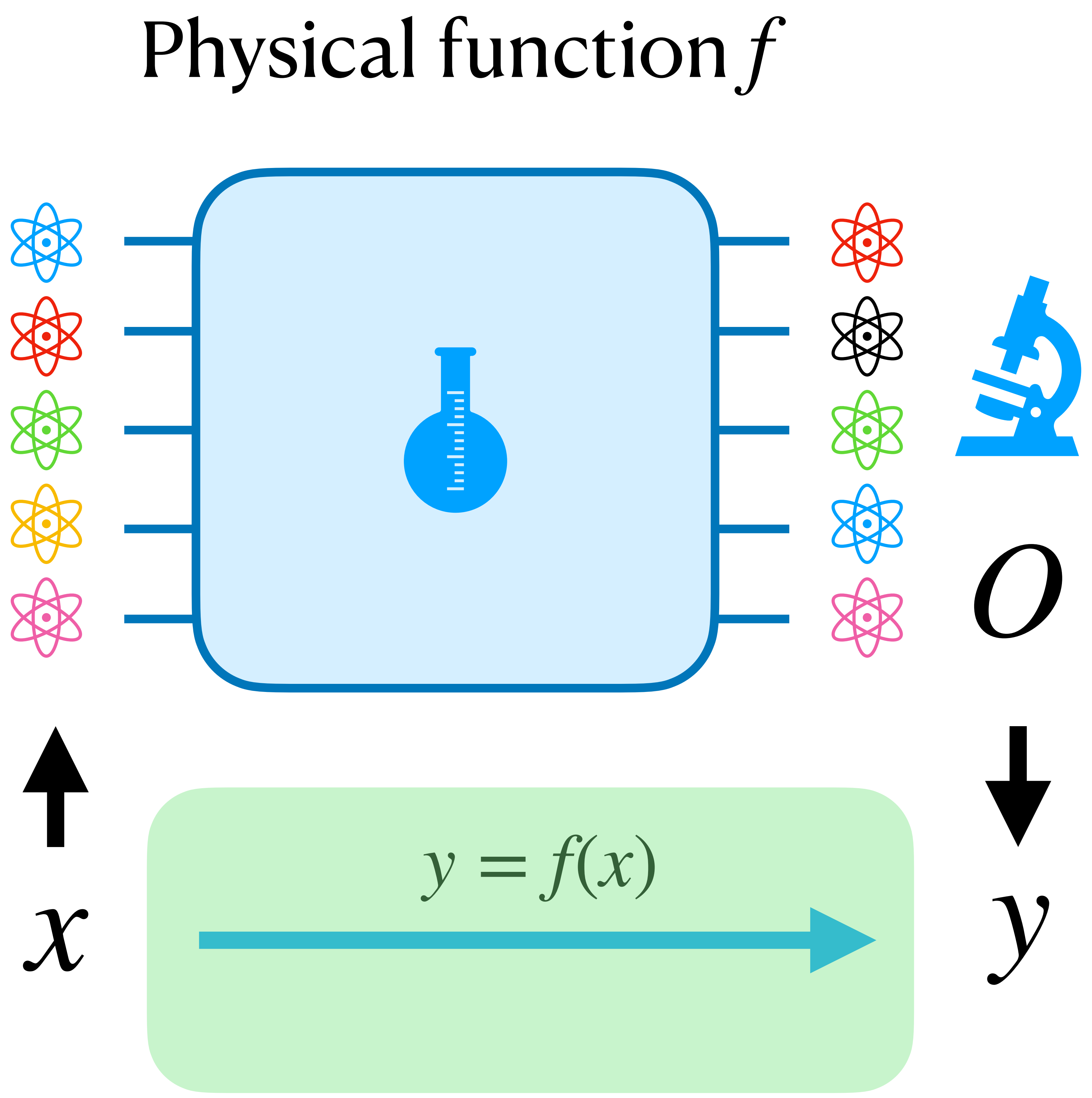
Physical functions



$$f(x) = \langle 0^n | U^\dagger(x, \{U_i\}) O U(x, \{U_i\}) | 0^n \rangle$$

What kind of functions are physically implementable?

Results



To implement/learn 1-Lip 1-bounded functions in ϵ infinite norm,
$G, N = \tilde{\Omega}(1/\epsilon^{\nu})$ gates/samples are needed

1. Certain well-behaved function class is not physical! $G = \Omega(\exp \nu)$
2. For these functions, no quantum advantage! $\tilde{\Theta}(1/\epsilon^{\nu})$ with classical ReLU NN.
3. More restricted function classes: possible physicality/advantage — Fourier integrable $O(1/\epsilon^2)$ **Gonon and Jacquier 2023**
4. Complement universal approximation theorems of QML — **Pérez-Salinas et al., PRA 2021; Schuld, Sweke, Meyer, PRA 2021; Manzano, Dechant, Tura, Dunjko 2023**

Proof Sketch

$$U_i|\psi\rangle = \begin{pmatrix} U_i^{11} & U_i^{12} \\ U_i^{21} & U_i^{22} \end{pmatrix} \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix} = \begin{pmatrix} U_i^{11}\psi_1 + U_i^{12}\psi_2 \\ U_i^{21}\psi_1 + U_i^{22}\psi_2 \end{pmatrix}$$

$$f(x) = \langle 0^n | U^\dagger(x, \{U_i\}) O U(x, \{U_i\}) | 0^n \rangle$$

$f(x)$ is a polynomial of $\{U_i\}$ with degree at most $2G$

=> fat-shattering/pseudo dimension $\leq \tilde{O}(G)$

Approximating 1-Lip 1-bounded functions require $\Omega(1/\epsilon^\nu)$

=> $G, N \geq \tilde{\Omega}(1/\epsilon^\nu)$

Caro and Datta, QMI 2020

Summary and Outlook

- Learning complexity = circuit complexity (information theoretically)
- Learning complexity = $e^{\text{circuit complexity}}$ (computationally)
- Possible efficient learning for more restricted states/unitaries?

  Clifford+T, MPS, Shallow Circuits, Fermion, Boson, etc.
- What physical properties does learning complexity relate to? Geometry? Phase? Thermalization? Entanglement?
- Worst-case unitary learning requires exponential samples.

  Du, Hsieh, Tao, arXiv last week
- Initiate the study of physical functions. Learning complexity? Q signal processing?
- Information-theoretic quantum no-free-lunch theorem (not covered)
- Technical: matching bounds, Heisenberg scaling, mixed states/channels, etc.