How to learn a Hamiltonian

Talk in a slide

Hamiltonian learning is about what is possible to learn in a quantum world.

Spooky things happen that cannot happen in classical settings.

I'll explain what I like about this problem.

1. Motivating Hamiltonian learning
2. Defining the basic objects
3. An example: Hamiltonian learning from real-time evolution
4. The broader landscape

Motivation: experiments at scale

I'm building a quantum device to explore a system's behavior/
experimentally validate a prediction/
do something cool.

- How do I know that I succeeded?
- How do I benchmark my system?
- How do I diagnose issues?
- How do I know what's going on in general?

Examples from Wright, *How to learn a quantum state*

Motivation: experiments at scale

Quantum teleportation over 143 kilometres using active feed-forward

Xiao-Song Ma ✉, Thomas Herbst, Thomas Scheidl, Daqing Wang, Sebastian Kropatschek, William Naylor, Bernhard Wittmann, Alexandra Mech, Johannes Kofler, Elena Anisimova, Vadim Makarov, Thomas Jennewein, Rupert Ursin & Anton Zeilinger ✉

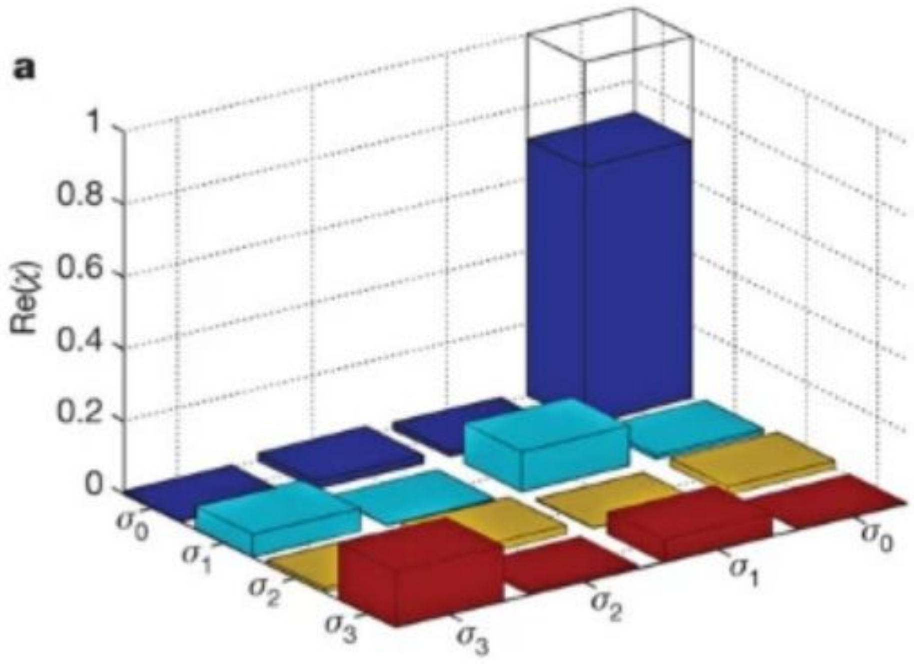
Nature **489**, 269–273 (2012) | Cite this article

Teleporting one qubit between from La Palma to Tenerife

605 *runs*

Examples from Wright, *How to learn a quantum state*

Figure 4: Quantum process tomography of quantum teleportation without feed-forward.



Motivation: experiments at scale

Scalable multiparticle entanglement of trapped ions

H. Häffner, W. Hänsel, C. F. Roos, J. Benhelm, D. Chek-al-kar, M. Chwalla, T. Körber, U. D. Rapol, M. Riebe, P. O. Schmidt, C. Becher, O. Gühne, W. Dür & R. Blatt

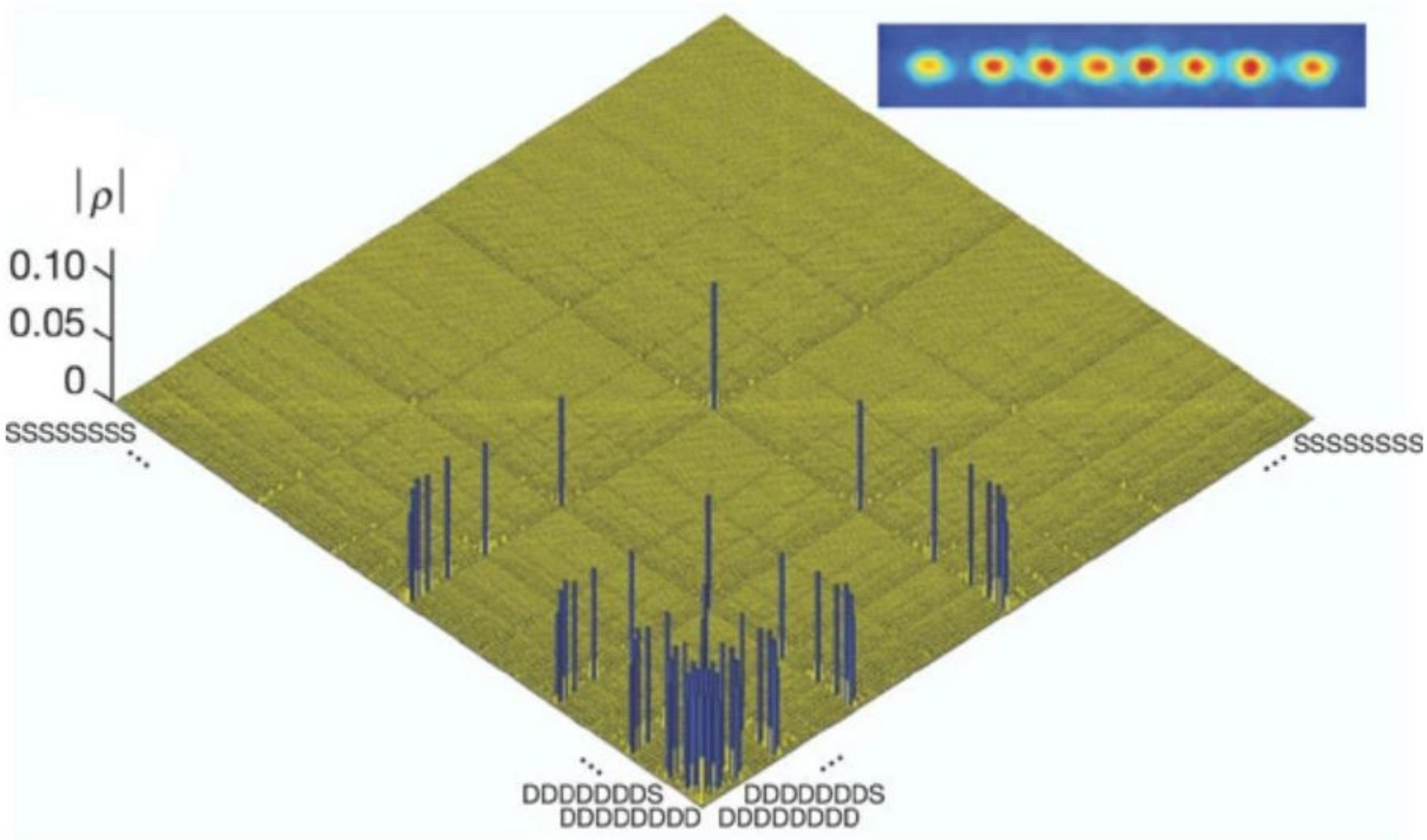
Nature **438**, 643–646 (2005) | Cite this article

Preparing an eight-qubit highly entangled state

656100 *runs*

Examples from Wright, *How to learn a quantum state*

Figure 1: Absolute values, $|\rho|$, of the reconstructed density matrix of a $|W_8\rangle$ state as obtained from quantum state tomography.



state

Motivation: experiments at scale

These papers use quantum state tomography, which is inherently not scalable:

runs is exponential in n = the number of qubits.

We want a protocol which scales as poly(n) for “physically reasonable” states.

Examples from Wright, *How to learn a quantum state*

Motivation: why learn Hamiltonians?

There are many choices for what reasonable hypothesis classes should be:

- Quantum circuits describe systems from quantum computers;
- Tensor networks describe systems from classical simulation;
- Hamiltonians describe systems from models of physics;

We believe these classes are interchangeable in some senses, but this is not rigorous.

Motivation: a basic epistemic question

1. The Hamiltonian is a description of the interactions in a system at the particle scale;
2. We expect the particle-scale features to be “feelable”;
3. So, we should be able to run experiments to find them efficiently.

If it’s not possible, then the Hamiltonian has an undetectable degree of freedom.

Hamiltonian learning models the task of a physicist:

> Can I determine the underlying interactions of a system from measuring it?

Background: Pauli matrices

One qubit is a 2×2 Hermitian matrix.

A useful basis is the basis of Pauli matrices.

- $\operatorname{tr}(PQ) = 0$ unless $P = Q$;
- reflections;
- nice product structure.

$$I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

$$Y = \begin{pmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix}, \qquad Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

$\times$	X	Y	Z
X	I	$\mathrm{i}Z$	$-\mathrm{i}Y$
Y	$-\mathrm{i}Z$	I	$\mathrm{i}X$
Z	$\mathrm{i}Y$	$-\mathrm{i}X$	I

Background: Pauli matrices

We will work in a system of n qubits, i.e. complex matrices of dimension $2^n \times 2^n$.

The analogous basis is that of tensor products of Pauli matrices:

$$P = P^{(1)} \otimes P^{(2)} \otimes \ldots \otimes P^{(n)}$$

$\mathrm{tr}(PQ) = 0$ unless $P = Q$; then, $\mathrm{tr}(PQ) = \mathrm{tr}(I) = 2^n$;

We use the notation e.g.

$$Z_2 = I \otimes Z \otimes I \otimes \ldots \otimes I.$$

$$I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

$$Y = \begin{pmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix}, \qquad Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

$\times$	X	Y	Z
X	I	$\mathrm{i}Z$	$-\mathrm{i}Y$
Y	$-\mathrm{i}Z$	I	$\mathrm{i}X$
Z	$\mathrm{i}Y$	$-\mathrm{i}X$	I

Background: Pauli matrices

$$I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

$$Y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Example:

$$(X_1 Y_2 Z_3)\,(Z_2 Z_3 Z_4) = X_1\,(Y_2 Z_2)\,(Z_3 Z_3)\,Z_4 = i\,X_1 X_2 Z_4$$

The **support** of a Pauli: $\mathrm{supp}(X_1 X_2 Z_4) = \{1, 2, 4\}$.

The support of an operator A is the set of qubits which A is not the identity on.

The **commutator** $[A, B] = AB - BA$.

$[X, Y] = 2iZ$

$[P_i, Q_j] = 0$ when $i \neq j$.

×	X	Y	Z
X	I	iZ	$-iY$
Y	$-iZ$	I	iX
Z	iY	$-iX$	I

Background: local Hamiltonian

A *local Hamiltonian* on *n* sites is

$$H = \lambda_1 E_1 + \ldots + \lambda_m E_m$$

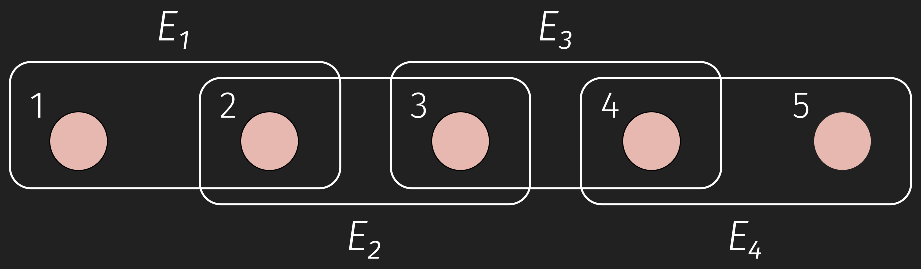
where every E_a is a Pauli with *O(1)* support and $-1 \le \lambda_a \le 1$.

The degree of *H* is the degree of the interaction graph, *G*, with vertices *[n]* and hyperedges $\{supp(E_a)\}_a$.

1D Ising model:

$$H = Z_1 Z_2 + Z_2 Z_3 + Z_3 Z_4 + \ldots$$

Local Hamiltonians model *spin systems with few-body interactions.*



Background: local Hamiltonian

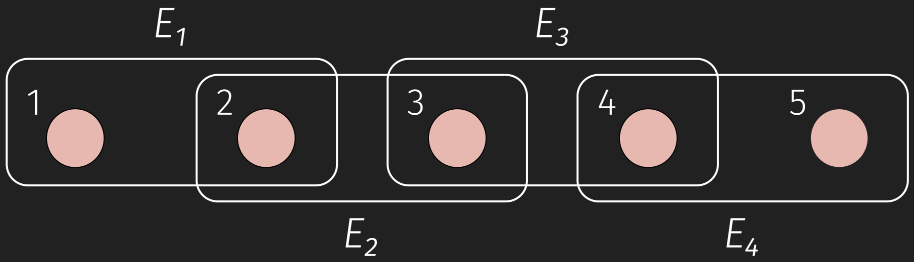
A *local Hamiltonian* on *n* sites is

$$H = \lambda_1 E_1 + \ldots + \lambda_m E_m$$

where every E_a is a Pauli with $O(1)$ support and $-1 \leq \lambda_a \leq 1$.

The degree of H is the degree of the interaction graph, G, with vertices $[n]$ and hyperedges $\{supp(E_a)\}_a$.

Local Hamiltonians model *spin systems with few-body interactions.*



XXX Heisenberg model (aka "quantum max-cut"):

$$H = -\sum_{(u,v)\in G} (X_u X_v + Y_u Y_v + Z_u Z_v) + h \sum_{i\in[n]} Z_i$$

What are the mathematical consequences of locality?

The evolution of an operator A is $e^{-iHt}Ae^{iHt}$.

To control this, we can try to truncate it to low degree. Consider the series expansion

$$e^{-\mathrm{i}Ht}Ae^{\mathrm{i}Ht} = \Big(I - \mathrm{i}Ht + \frac{1}{2}(\mathrm{i}Ht)^2 - \dots\Big)A\Big(I + \mathrm{i}Ht + \frac{1}{2}(\mathrm{i}Ht)^2 + \dots\Big)$$

Generally, $\| -iHt \| \sim mt$, which means we can only truncate at degree $\sim mt$.

But we can use the locality structure to get a better bound.

$$e^{-\mathrm{i}Ht}Ae^{\mathrm{i}Ht} = \sum_{k=0}^{\infty} \frac{1}{k!}[-\mathrm{i}Ht, A]_k$$

$$= A + [-\mathrm{i}Ht, A] + \frac{1}{2!}[-\mathrm{i}Ht, [-\mathrm{i}Ht, A]] + \dots$$

Locality implies that some series converge quickly

$$e^{-\mathrm{i}Ht} A e^{\mathrm{i}Ht} = \sum_{k=0}^{\infty} \frac{1}{k!}[-\mathrm{i}Ht, A]_k = \sum_{k=0}^{\infty} \frac{(-\mathrm{i}t)^k}{k!}[H, A]_k$$

Lemma. If $|supp(A)| = O(1)$, $\| [H, A]_k \| < k!\, C^k$ for a constant C.

So, we can truncate at order t.

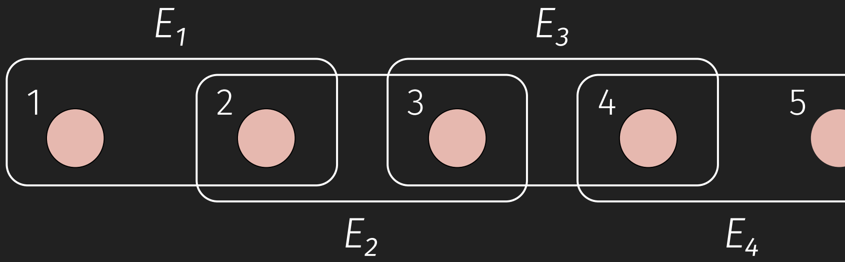
In physics language, "if A is local then $e^{-iHt}Ae^{iHt}$ is quasilocal".

Locality implies that some series converge quickly

Lemma. If $|supp(A)| = O(1)$, $\| [H, A]_k \| < k!\, C^k$ for a constant C.

We consider a specific example:

$$H = X_1 Y_2 + X_2 Y_3 + X_3 Y_4 + \ldots$$
$$A = Z_1$$



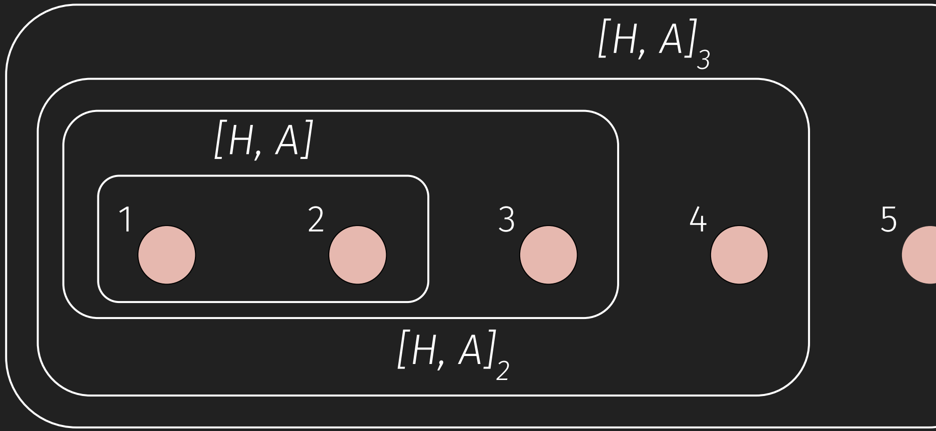
$$[H, A] = \left[\sum_{i=1}^{n-1} X_i Y_{i+1}, Z_1\right] = \sum_{i=1}^{n-1} [X_i Y_{i+1}, Z_1] = [X_1 Y_2, Z_1] = -2\mathrm{i} Y_1 Y_2$$

$$[H, A]_2 = 4 Z_1 + 4 Y_1 Z_2 Y_3$$

Locality implies that some series converge quickly

Lemma. If $|supp(A)| = O(1)$, $\| [H, A]_k \| < k!\, C^k$ for a constant C.

We consider a specific example:



$$H = X_1Y_2 + X_2Y_3 + X_3Y_4 + \dots$$
$$A = Z_1$$

$$[H, A] = \left[\sum_{i=1}^{n-1} X_iY_{i+1}, Z_1\right] = \sum_{i=1}^{n-1}[X_iY_{i+1}, Z_1] = [X_1Y_2, Z_1] = -2\mathrm{i}Y_1Y_2$$

$$[H, A]_2 = 4Z_1 + 4Y_1Z_2Y_3$$

Background: the learning task

Input: description of the terms $E_1, \ldots, E_m$;

some kind of access to H.

Output: estimates of the coefficients $\lambda_1, \ldots, \lambda_m$.

NB: this is *parameter learning*. More common classically is structure learning.

Parameter learning in the quantum setting is already non-trivial.

structure learning of (classical) MRFs: [Klivans, Meka '17]

An example: learning from real-time evolutions

Input: description of the terms $E_1, \ldots, E_m$;

ability to apply e^{-iHt} for every $t > 0$.

Output: estimates of the coefficients $\lambda_1, \ldots, \lambda_m$.

Track *evolution time*: applying e^{-iHt} costs t, sum over the entire algorithm

Simplification: "statistical query model".

We can estimate every $\mathrm{tr}(P e^{-iHt} Q e^{iHt})/2^n$ for $|supp(P)|, |supp(Q)| = O(1)$ to ε error.

evolution time: $O(t \log(n)/\varepsilon^2)$

(Classical intuition: estimating every low-weight Fourier coefficient of f from $O(\log(n)/\varepsilon^2)$ random queries.)

[Huang, Kueng, Preskill '20]; [Haah, Kothari, T '24]

A simple Hamiltonian learning algorithm

Input: description of the terms $E_1, \ldots, E_m$;

α-estimates of every $\mathrm{tr}(Pe^{-iHt}Qe^{iHt})/2^n$
for $|supp(P)|, |supp(Q)| = O(1)$

Output: estimates of the coefficients $\lambda_1, \ldots, \lambda_m$.

total evolution time: $O(t \log(n)/\alpha^2)$

Consider a term $E_1 = X_1 Y_2$.

Take P to not commute with E_1:
Let $Q = i\, P\, E_1$:

$P = Y_1$
$Q = i\, (Y_1)\, (X_1 Y_2) = Z_1 Y_2$

Then $\mathrm{tr}(Pe^{-iHt}Qe^{iHt})/2^n = 2\lambda_a t + O(t\^{}2)$ for $t << 1$.

Take time $t = \varepsilon$ and estimate to error $\alpha = \varepsilon^2$.

total evolution time: $O(\log(n)/\varepsilon^3)$

A simple Hamiltonian learning algorithm

$$\mathrm{tr}(Pe^{-\mathrm{i}Ht}Qe^{\mathrm{i}Ht})/2^n$$
$$= \mathrm{tr}(e^{\mathrm{i}Ht}Pe^{-\mathrm{i}Ht}Q/2^n)$$
$$= \mathrm{tr}\Big(\Big(P + [\mathrm{i}Ht, P] + \underbrace{F}_{\|F\|=O(t^2)}\Big)\frac{Q}{2^n}\Big)$$
$$= \mathrm{tr}((P + \mathrm{i}t[H, P])Q/2^n) + O(t^2)$$
$$= \mathrm{tr}\Big(\Big(P + \sum_{a=1}^{m} \mathrm{i}t\lambda_a[E_a, P]\Big)Q/2^n\Big) + O(t^2)$$
$$= 2\lambda_a t + O(t^2).$$

Consider a term $E_1 = X_1Y_2$.

Take P to not commute with E_1:
Let $Q = i\ P\ E_1$:

$P = Y_1$
$Q = i\ (Y_1)\ (X_1Y_2) = Z_1Y_2$

Then $\mathrm{tr}(Pe^{-iHt}Qe^{iHt})/2^n = 2\lambda_a t + O(t\text{^}2)$ for $t << 1$.

Take time $t = \varepsilon$ and estimate to error $\alpha = \varepsilon^2$.

total evolution time: $O(log(n)/\varepsilon^3)$

Improving by considering more of the series

It actually suffices to take

time t = $\Theta(1)$ and error $\alpha = \varepsilon$.

$$\{\lambda_a\}_a \mapsto \left\{\mathrm{tr}(P e^{-\mathrm{i}Ht} Q e^{\mathrm{i}Ht})/2^n\right\}_{P,Q}$$

The observables are a polynomial system in the parameters with a strong decay with degree, so we can solve for the coefficients.

total evolution time: $O(\log(n)/\varepsilon^2)$

Are we done?

No, log(n)/ε is possible! [Huang, Tong, Fang, Su '24]

[Haah, Kothari, T '24]

Learning in *1/ε* evolution time (the "Heisenberg limit")

Quantum-mechanical noise in an interferometer

Carlton M. Caves
Phys. Rev. D **23**, 1693 – Published 15 April 1981

Letter | Published: 11 September 2011

A gravitational wave observatory operating beyond the quantum shot-noise limit

The LIGO Scientific Collaboration

Nature Physics **7**, 962–965 (2011) | Cite this article

Letter | Published: 11 January 2016

Measurement noise 100 times lower than the quantum-projection limit using entangled atoms

Onur Hosten, Nils J. Engelsen, Rajiv Krishnakumar & Mark A. Kasevich

Nature **529**, 505–508 (2016) | Cite this article

Learning in 1/ε evolution time (the "Heisenberg limit")

Consider the single-qubit example:

$$H = \lambda Z; \qquad e^{-\mathrm{i}Ht} = \begin{pmatrix} 1 & \\ & e^{-\mathrm{i}\lambda t} \end{pmatrix}$$

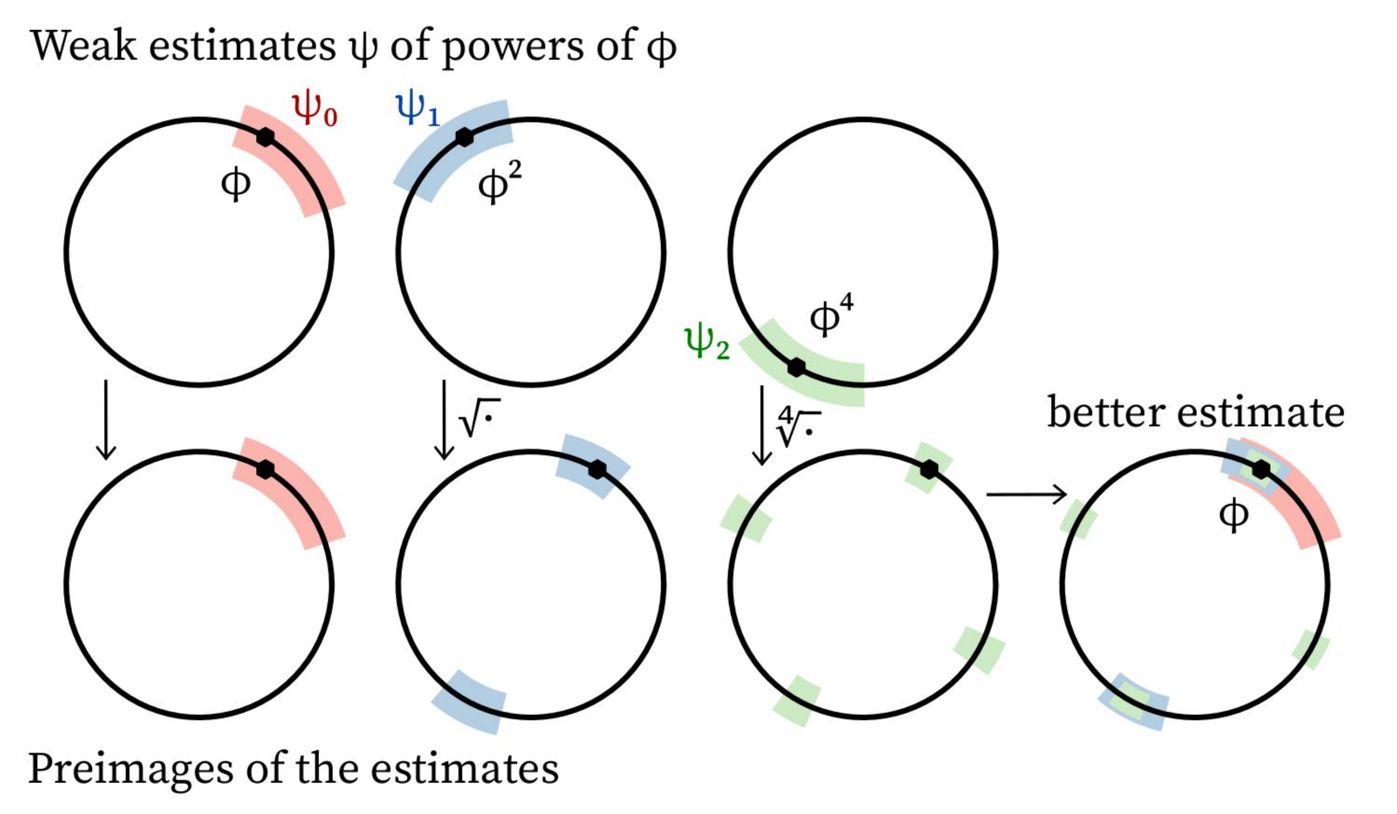
It suffices to estimate $\phi = e^{-i\lambda}$ to ε error.

Take $t = 1, 2, 4, 8, \ldots, 1/\varepsilon$ but error $\alpha = \Theta(1)$.

This gives constant-error estimates for $\phi^1, \phi^2, \phi^4, \ldots$

with evolution time $1/\varepsilon$.

Learning in 1/ε evolution time (the "Heisenberg limit")



Hamiltonian learning more broadly

Learning from dynamics: we are given the ability to evolve by e^{-iHt}.

Complexity measure: total evolution time

1. $poly(m, 1/\varepsilon)$
2. $\log(m)/\varepsilon^2$
3. $\log(m)/\varepsilon$

using locality-based series expansions, “Pauli analysis”, error amplification

1, [Cramer, Plenio, Flammia, Somma, Gross, Bartlett, Landon-Cardinal, Poulin, Liu '11];
2, [Haah, Kothari, T ‘24];
3, [Huang, Tong, Fang, Su ’24]

Hamiltonian learning more broadly

Learning from dynamics: we are given the ability to evolve by e^{-iHt}.

Complexity measure: total evolution time

Current algorithms use:

1. series expansions exploiting locality
2. "Pauli analysis",
3. error amplification

Learning from static states: we are given a state at equilibrium with respect to e^{-iHt}.

Gibbs state: $\rho_\beta \propto e^{-\beta H} / tr(e^{-\beta H})$

Complexity measure: sample complexity

1. polynomial samples
2. polynomial time (sum-of-squares hierarchy)
3. better β dependence (fine-tuned polynomial approx of e^x)

1, [Anshu, Arunachalam, Kuwahara, Soleimanifar '21];
2, [Bakshi, Liu, Moitra, T '24];
3, [Narayanan '24]

Hamiltonian learning more broadly

Learning from dynamics: we are given the ability to evolve by e^{-iHt}.

Complexity measure: total evolution time

Current algorithms use:

1. series expansions exploiting locality
2. “Pauli analysis”,
3. error amplification

Related problems:

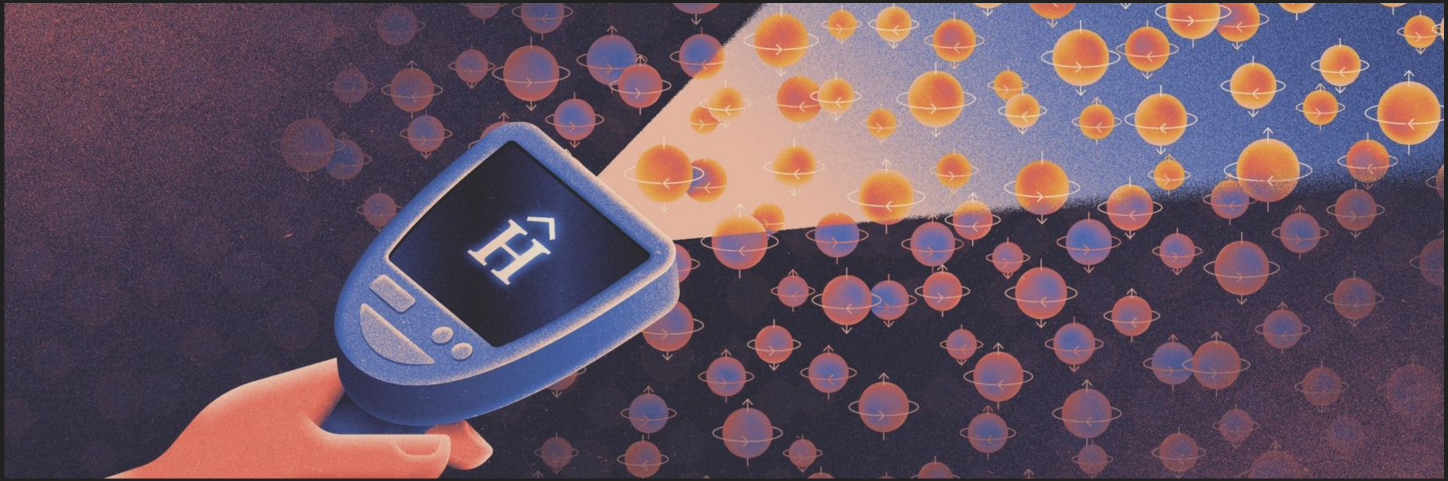
> Structure learning [Bakshi, Liu, Moitra, T ’24] refined analysis of series, “Pauli” Goldreich–Levin

> Testing [Gutiérrez ‘24] more Pauli analysis

> Agnostic learning [Grewal, Iyer, Kretschmer, Liang ‘24]

Learning from ground states: given ρ_∞

Thank you!



credit: Kristina Armitage