Learning to predict ground state properties of gapped Hamiltonians

Provably efficient machine learning for quantum many-body problems

Richard Kueng

Department of Computer Science,
Johannes Kepler University Linz, Austria

March 6th, 2023

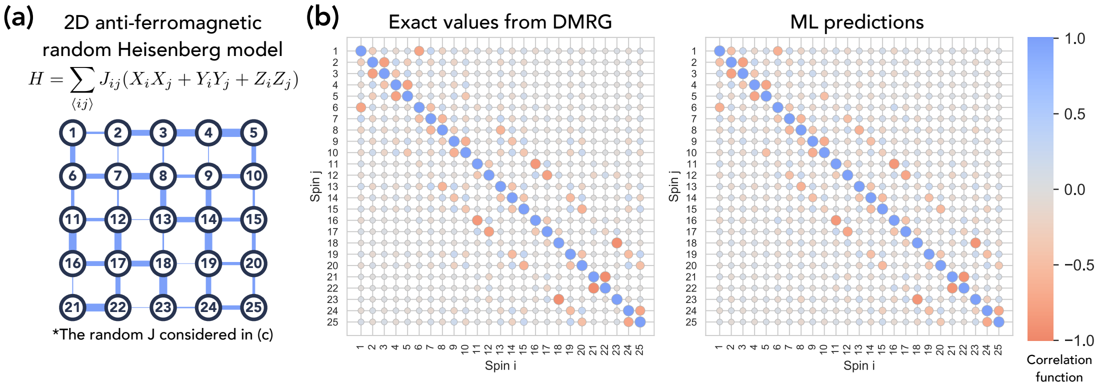
joint work with **H.Y. Huang, G. Torlai, V. Albert** and **J. Preskill**

outline

motivation

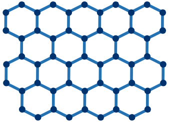
ground state problem in quantum many-body physics

Motivating numerics: 2D Heisenberg model $(n = 25, m = 40)$



High-level vision

$\mathbf{H}_{\mathrm{tot}} = \sum_j \boldsymbol{H}$



Parameters describing a physical Hamiltonian

$\Rightarrow$

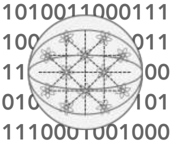
direct computation

$\rho(x) = \boldsymbol{v}_{\min}(x)\boldsymbol{v}_{\min}(x)^\dagger$

(expensive: $D = 2^n$)

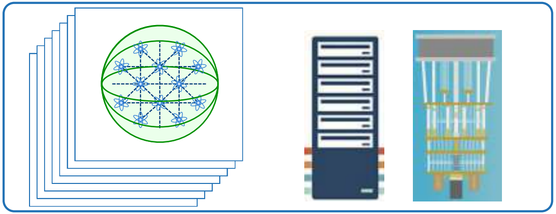
$\Rightarrow$

$\mathrm{tr}\left(\boldsymbol{O}\rho(x)\right)$



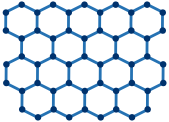
Classical representation of the ground state

High-level vision



$x \in [-1,1]^m$

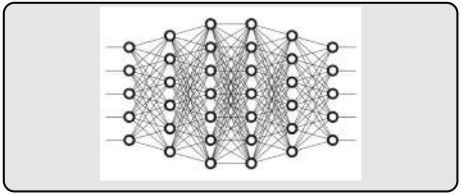
$\mathbf{H}_{\mathrm{tot}}(x) = \sum_j \boldsymbol{H}(x)$



Parameters describing a physical Hamiltonian

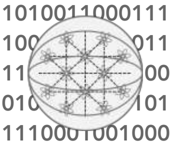
$\Rightarrow$

$(x_\ell, \boldsymbol{\rho}(x_\ell)) \Downarrow x_\ell \overset{unif}{\sim} [-1,1]^m$



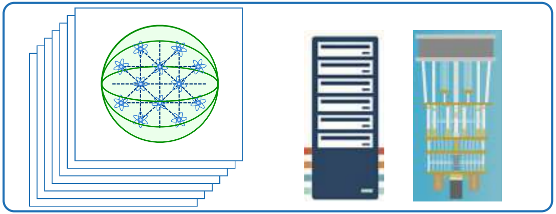
$\Rightarrow$

$\mathrm{tr}\left(\boldsymbol{O}\boldsymbol{\rho}_{\mathrm{train}}(x)\right)$



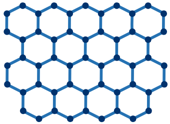
Classical representation of the ground state

High-level vision



$x \in [-1,1]^m$

$\mathbf{H}_{\mathrm{tot}}(x) = \sum_j \boldsymbol{H}(x)$



Parameters describing a physical Hamiltonian

$\Rightarrow$

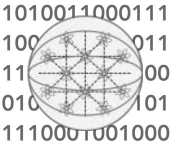
$(x_\ell, \boldsymbol{\rho}(x_\ell)) \Downarrow x_\ell \overset{unif}{\sim} [-1,1]^m$

$$\boldsymbol{\rho}_{\mathrm{train}}(x) = \sum_{\ell=1}^N \kappa(x, x_\ell)\, \boldsymbol{\rho}(x_\ell)$$

κ : neural tangent kernel

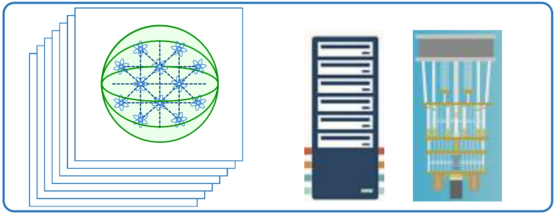
$\Rightarrow$

$\mathrm{tr}\left(\boldsymbol{O}\boldsymbol{\rho}_{\mathrm{train}}(x)\right)$



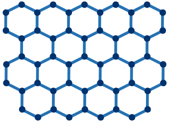
Classical representation of the ground state

High-level vision



$x \in [-1,1]^m$

$\mathbf{H}_{\text{tot}}(x) = \sum_j \boldsymbol{H}(x)$



Parameters describing a physical Hamiltonian

$\Rightarrow$

$(x_\ell, \boldsymbol{\rho}(x_\ell)) \Downarrow x_\ell \overset{unif}{\sim} [-1,1]^m$

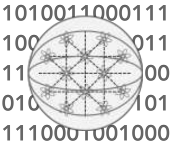
$$\boldsymbol{\rho}_{\text{train}}(x) = \sum_{\ell=1}^{N} \kappa\,(x, x_\ell)\,\boldsymbol{\rho}(x_\ell)$$

~~κ : neural tangent kernel~~

κ : ℓ_2-Dirichlet kernel

$\Rightarrow$

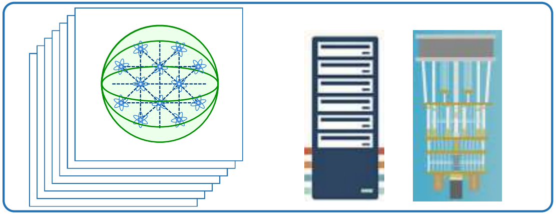
$\mathrm{tr}\,(\boldsymbol{O}\boldsymbol{\rho}_{\text{train}}(x))$



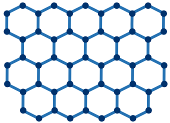
Classical representation of the ground state

High-level vision

spoiler: assumptions on $\boldsymbol{H}(x), \boldsymbol{O}$ ensure
$\mathbb{E}_{x \overset{unif}{\sim} [-1,1]^m} \left| \mathrm{tr}\left(\boldsymbol{O}\boldsymbol{\rho}_{\mathrm{train}}(x)\right) - \mathrm{tr}\left(\boldsymbol{O}\boldsymbol{\rho}(x)\right)\right|^2 \leq \epsilon$
$(\mathrm{MSE} \leq \epsilon)$ with $\mathrm{poly}(m) = \mathrm{poly}(n)$ scaling in
- training data size
- runtime + memory



$x \in [-1,1]^m$
$\mathbf{H}_{\mathrm{tot}}(x) = \sum_j \boldsymbol{H}(x)$



Parameters describing a physical Hamiltonian

$\Rightarrow$

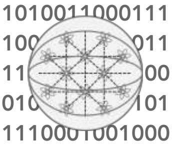
$(x_\ell, \boldsymbol{\rho}(x_\ell)) \Downarrow x_\ell \overset{unif}{\sim} [-1,1]^m$

$$\boldsymbol{\rho}_{\mathrm{train}}(x) = \sum_{\ell=1}^{N} \kappa\left(x, x_\ell\right) \boldsymbol{\rho}(x_\ell)$$

~~κ : neural tangent kernel~~
κ : ℓ_2-Dirichlet kernel

$\Rightarrow$

$\mathrm{tr}\left(\boldsymbol{O}\boldsymbol{\rho}_{\mathrm{train}}(x)\right)$

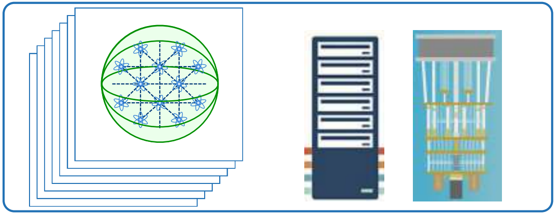


Classical representation of the ground state

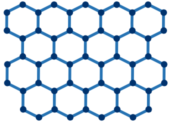
High-level vision

spoiler: assumptions on $\boldsymbol{H}(x), \boldsymbol{O}$ ensure
$$\mathbb{E}_{x \overset{unif}{\sim} [-1,1]^m} \left| \mathrm{tr}\left(\boldsymbol{O}\boldsymbol{\rho}_{\mathrm{train}}(x)\right) - \mathrm{tr}\left(\boldsymbol{O}\boldsymbol{\rho}(x)\right)\right|^2 \leq \epsilon$$
$(\mathrm{MSE} \leq \epsilon)$ with $\mathrm{poly}(m) = \mathrm{poly}(n)$ scaling in

- training data size (improvement to $\mathrm{polylog}(n)$)
- runtime + memory (Lewis *et al.* 2301.13169)



$x \in [-1,1]^m$
$\mathbf{H}_{\mathrm{tot}}(x) = \sum_j \boldsymbol{H}(x)$



Parameters describing a physical Hamiltonian

$\Rightarrow$

$(x_\ell, \boldsymbol{\rho}(x_\ell)) \Downarrow x_\ell \overset{unif}{\sim} [-1,1]^m$

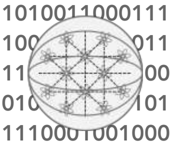
$$\boldsymbol{\rho}_{\mathrm{train}}(x) = \sum_{\ell=1}^{N} \kappa(x, x_\ell)\, \boldsymbol{\rho}(x_\ell)$$

~~κ : neural tangent kernel~~

κ : ℓ_2-Dirichlet kernel

$\Rightarrow$

$\mathrm{tr}\left(\boldsymbol{O}\boldsymbol{\rho}_{\mathrm{train}}(x)\right)$



Classical representation of the ground state

Data format: classical shadows

ML insight: compression s.t. $\hat{\boldsymbol{\sigma}}(\alpha \boldsymbol{\rho}_1) + \beta \hat{\boldsymbol{\sigma}}(\boldsymbol{\rho}_2) = \hat{\boldsymbol{\sigma}}(\alpha \boldsymbol{\rho}_1 + \beta \boldsymbol{\rho}_2)$

Data format: classical shadows

ML insight: compression s.t. $\hat{\boldsymbol{\sigma}}(\alpha\boldsymbol{\rho}_1) + \beta\hat{\boldsymbol{\sigma}}(\boldsymbol{\rho}_2) = \hat{\boldsymbol{\sigma}}(\alpha\boldsymbol{\rho}_1 + \beta\boldsymbol{\rho}_2)$

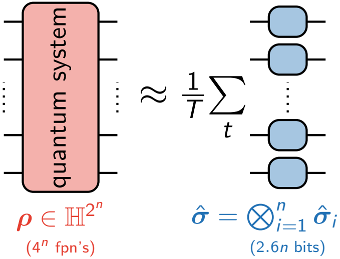
our solution : Monte Carlo sampling with quantum architectures

Data format: classical shadows

ML insight: compression s.t. $\hat{\boldsymbol{\sigma}}(\alpha \boldsymbol{\rho}_1) + \beta \hat{\boldsymbol{\sigma}}(\boldsymbol{\rho}_2) = \hat{\boldsymbol{\sigma}}(\alpha \boldsymbol{\rho}_1 + \beta \boldsymbol{\rho}_2)$

our solution : Monte Carlo sampling with quantum architectures

– Monte Carlo paradigm ensures tractable approximations



$$\mathrm{tr}\left(\boldsymbol{O}_1 \boldsymbol{\rho}\right) \approx \frac{1}{T} \sum_{t=1}^T \mathrm{tr}\left(\boldsymbol{O}_1 \boldsymbol{\sigma}_t\right)$$

$$\vdots$$

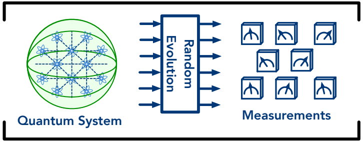
$$\mathrm{tr}\left(\boldsymbol{O}_L \boldsymbol{\rho}\right) \approx \frac{1}{T} \sum_{t=1}^T \mathrm{tr}\left(\boldsymbol{O}_L \boldsymbol{\sigma}_t\right)$$

Data format: classical shadows

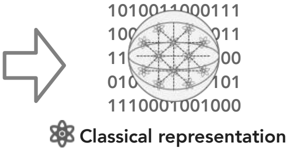
ML insight: compression s.t. $\hat{\boldsymbol{\sigma}}(\alpha\boldsymbol{\rho}_1)+\beta\hat{\boldsymbol{\sigma}}(\boldsymbol{\rho}_2)=\hat{\boldsymbol{\sigma}}(\alpha\boldsymbol{\rho}_1+\beta\boldsymbol{\rho}_2)$

our solution : Monte Carlo sampling with quantum architectures

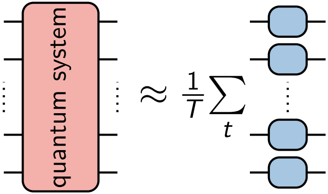
- Monte Carlo paradigm ensures tractable approximations
- sampling process outsourced to quantum simulator



Few rounds of randomized measurements



Classical representation of the quantum system



$$\approx \frac{1}{T}\sum_t$$

$\boldsymbol{\rho} \in \mathbb{H}^{2^n}$ (4^n fpn's)

$\hat{\boldsymbol{\sigma}} = \bigotimes_{i=1}^n \hat{\boldsymbol{\sigma}}_i$ ($2.6n$ bits)

$$\mathrm{tr}\left(\boldsymbol{O}_1\boldsymbol{\rho}\right) \approx \tfrac{1}{T}\sum_{t=1}^T \mathrm{tr}\left(\boldsymbol{O}_1\boldsymbol{\sigma}_t\right)$$

$$\vdots$$

$$\mathrm{tr}\left(\boldsymbol{O}_L\boldsymbol{\rho}\right) \approx \tfrac{1}{T}\sum_{t=1}^T \mathrm{tr}\left(\boldsymbol{O}_L\boldsymbol{\sigma}_t\right)$$

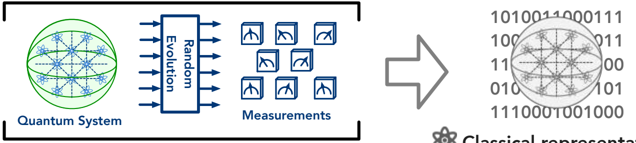
Data format: classical shadows

ML insight: compression s.t. $\hat{\boldsymbol{\sigma}}(\alpha\boldsymbol{\rho}_1) + \beta\hat{\boldsymbol{\sigma}}(\boldsymbol{\rho}_2) = \hat{\boldsymbol{\sigma}}(\alpha\boldsymbol{\rho}_1 + \beta\boldsymbol{\rho}_2)$

our solution : Monte Carlo sampling with quantum architectures

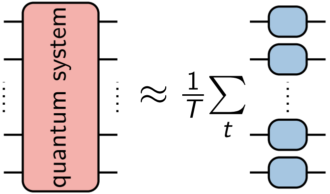
– Monte Carlo paradigm ensures tractable approximations

– sampling process outsourced to quantum simulator



Few rounds of randomized measurements

Classical representation of the quantum system



$\boldsymbol{\rho} \in \mathbb{H}^{2^n}$ (4^n fpn's)

$\hat{\boldsymbol{\sigma}} = \bigotimes_{i=1}^{n} \hat{\boldsymbol{\sigma}}_i$ ($2.6n$ bits)

$$\mathrm{tr}(\boldsymbol{O}_1\boldsymbol{\rho}) \approx \frac{1}{T}\sum_{t=1}^{T} \mathrm{tr}(\boldsymbol{O}_1\boldsymbol{\sigma}_t)$$

$$\vdots$$

$$\mathrm{tr}(\boldsymbol{O}_L\boldsymbol{\rho}) \approx \frac{1}{T}\sum_{t=1}^{T} \mathrm{tr}(\boldsymbol{O}_L\boldsymbol{\sigma}_t)$$

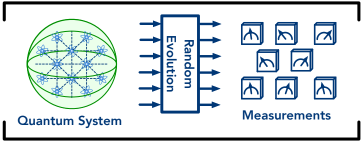
- combination of **quantum software** and **conventional software**
- simple **quantum software**, **conventional memory & runtime** is also cheap

Data format: classical shadows

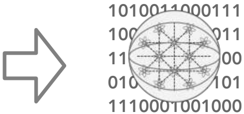
ML insight: compression s.t. $\hat{\boldsymbol{\sigma}}(\alpha\boldsymbol{\rho}_1) + \beta\hat{\boldsymbol{\sigma}}(\boldsymbol{\rho}_2) = \hat{\boldsymbol{\sigma}}(\alpha\boldsymbol{\rho}_1 + \beta\boldsymbol{\rho}_2)$

our solution : Monte Carlo sampling with quantum architectures

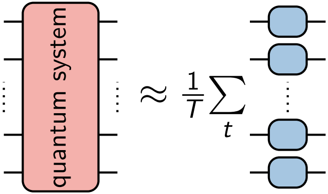
- Monte Carlo paradigm ensures tractable approximations
- sampling process outsourced to quantum simulator



Few rounds of randomized measurements



Classical representation of the quantum system



$\boldsymbol{\rho} \in \mathbb{H}^{2^n}$ (4^n fpn's)

$\hat{\boldsymbol{\sigma}} = \bigotimes_{i=1}^n \hat{\boldsymbol{\sigma}}_i$ ($2.6n$ bits)

$$\operatorname{tr}(\boldsymbol{O}_1\boldsymbol{\rho}) \approx \frac{1}{T}\sum_{t=1}^T \operatorname{tr}(\boldsymbol{O}_1\boldsymbol{\sigma}_t)$$
$$\vdots$$
$$\operatorname{tr}(\boldsymbol{O}_L\boldsymbol{\rho}) \approx \frac{1}{T}\sum_{t=1}^T \operatorname{tr}(\boldsymbol{O}_L\boldsymbol{\sigma}_t)$$

- combination of **quantum software** and **conventional software**
- simple **quantum software**, **conventional memory & runtime** is also cheap
- works for every state, but only efficient for local observables $\Rightarrow$ **locality assumption**

main result

Theorem 1 (Learning to predict ground state representations; informal). *For any smooth family of Hamiltonians $\{H(x) : x \in [-1,1]^m\}$ in a finite spatial dimension with a constant spectral gap, the classical machine learning algorithm can learn to predict a classical representation of the ground state $\rho(x)$ of $H(x)$ that approximates few-body reduced density matrices up to a constant error ϵ when averaged over x. The required training data size N and computation time are polynomial in m and linear in the system size n.*

H.Y. Huang, R. Kueng, G. Torlai, V.A. Albert, J. Preskill. *Provably efficient ML for many-body problems*. **Science 377**, eabk3333 (2022) (and https://arxiv.org/abs/2106.12627)

proof of the main result

three steps: (i) signal processing, (ii) bridge to ground state problem, (iii) classical shadows

Proof part 1: signal processing

Theorem

Consider a function $f : [-1,1]^m \to \mathbb{R}$ *(think:* $f(x) = \mathrm{tr}\left(\boldsymbol{O}\rho(x)\right)$*) that obeys*

(i) $\mathbb{E}_{x \overset{unif}{\sim}[-1,1]^m} \|\nabla_x f(x)\|_2^2 \leq C$ *(controlled average gradient size)*

(ii) $|f(x)| \leq B$ *almost surely (bounded magnitude).*

Use $N = B^2 m^{\mathcal{O}(C/\epsilon)}$ *uniform samples* $(x_\ell, f(x_\ell))$ *with* $x_\ell \overset{unif}{\sim} [-1,1]^m$ *to construct*

$$\tilde{f}(x) = \frac{1}{N}\sum_{\ell=1}^{N} \kappa_\Lambda(x, x_\ell) f(x_\ell) \quad \text{with} \quad \kappa_\Lambda(x, x_\ell) = \sum_{\substack{k \in \mathbb{Z}^m \\ \|k\|_2 \leq \Lambda}} \mathrm{e}^{\mathrm{i}\pi\langle k, x-y\rangle}, \ \Lambda = \mathcal{O}(C/\epsilon).$$

Then, $\mathbb{E}_{x \overset{unif}{\sim}[-1,1]^m} \left|\tilde{f}(x) - f(x)\right| \leq \epsilon$ *(*$\mathrm{MSE} \leq \epsilon$*) with high probability.*

Proof part 1: signal processing

$f(x) = \boldsymbol{F}^{-1}\boldsymbol{F}f(x)$ ← Fourier series plays nicely with MSE

Proof part 1: signal processing

$$\begin{aligned} f(x) &= \boldsymbol{F}^{-1}\boldsymbol{F}f(x) \\ &= \boldsymbol{F}^{-1}\boldsymbol{T}_\Lambda \boldsymbol{F}f(x) + \boldsymbol{F}^{-1}\left(\mathrm{Id} - \boldsymbol{T}_\Lambda\right)\boldsymbol{F}f(x) \\ &\approx \boldsymbol{F}^{-1}\boldsymbol{T}_\Lambda \boldsymbol{F}f(x) \end{aligned}$$

← Fourier series plays nicely with MSE

← truncation in frequency domain (use $\mathbb{E}_x \|\nabla_x f(x)\|_2^2 \leq C$)

Proof part 1: signal processing

$$
\begin{aligned}
f(x) &= \boldsymbol{F}^{-1}\boldsymbol{F}f(x) && \leftarrow \text{Fourier series plays nicely with MSE} \\
&= \boldsymbol{F}^{-1}\boldsymbol{T}_\Lambda \boldsymbol{F}f(x) + \boldsymbol{F}^{-1}\left(\mathrm{Id} - \boldsymbol{T}_\Lambda\right)\boldsymbol{F}f(x) && \\
&\approx \boldsymbol{F}^{-1}\boldsymbol{T}_\Lambda \boldsymbol{F}f(x) && \leftarrow \text{truncation in frequency domain} \\
&= \sum_{\|k\|_2 \leq \Lambda} \mathrm{e}^{\mathrm{i}\pi\langle k,x\rangle} \frac{1}{2^m}\int_{[-1,1]^m} \mathrm{d}^m y \mathrm{e}^{-\mathrm{i}\pi\langle k,y\rangle} f(y) && \quad \text{(use } \mathbb{E}_x \|\nabla_x f(x)\|_2^2 \leq C\text{)} \\
&= \frac{1}{2^m}\int_{[-1,1]^m}\left(\sum_{\|k\|_2\leq\Lambda} \mathrm{e}^{\mathrm{i}\pi\langle k,x-y\rangle}\right) f(y)\mathrm{d}^m y && \leftarrow \ell_2\text{-Dirichlet kernel emerges}
\end{aligned}
$$

Proof part 1: signal processing

$$
\begin{aligned}
f(x) &= \boldsymbol{F}^{-1}\boldsymbol{F}f(x) && \leftarrow \text{Fourier series plays nicely with MSE}\\
&= \boldsymbol{F}^{-1}\boldsymbol{T}_\Lambda\boldsymbol{F}f(x) + \boldsymbol{F}^{-1}\left(\mathrm{Id}-\boldsymbol{T}_\Lambda\right)\boldsymbol{F}f(x) &&\\
&\approx \boldsymbol{F}^{-1}\boldsymbol{T}_\Lambda\boldsymbol{F}f(x) && \leftarrow \text{truncation in frequency domain}\\
&= \sum_{\|k\|_2\leq\Lambda} \mathrm{e}^{\mathrm{i}\pi\langle k,x\rangle}\tfrac{1}{2^m}\int_{[-1,1]^m}\mathrm{d}^m y \mathrm{e}^{-\mathrm{i}\pi\langle k,y\rangle} f(y) && \quad (\text{use } \mathbb{E}_x\|\nabla_x f(x)\|_2^2\leq C)\\
&= \tfrac{1}{2^m}\int_{[-1,1]^m}\left(\textstyle\sum_{\|k\|_2\leq\Lambda}\mathrm{e}^{\mathrm{i}\pi\langle k,x-y\rangle}\right) f(y)\mathrm{d}^m y && \leftarrow \ell_2\text{-Dirichlet kernel emerges}\\
&= \tfrac{1}{2^m}\int_{[-1,1]^m}\kappa_\Lambda(x,y) f(y)\mathrm{d}^m y &&\\
&\approx \tfrac{1}{N}\textstyle\sum_{\ell=1}^N \kappa_\Lambda(x,x_\ell)f(x_\ell) \text{ with } x_\ell \overset{unif}{\sim} [-1,1]^m && \leftarrow \text{Monte Carlo approximation,}
\end{aligned}
$$

Proof part 2: bridge to ground state problem

Theorem (streamlined insight from sampling theory)

Uniform sampling efficiently interpolates functions $f : [-1,1]^m \to \mathbb{R}$ *that obey (i)* $\mathbb{E}_{x \overset{iid}{\sim} [-1,1]^m} \|\nabla_x f(x)\|_2^2 \leq C$ *and (ii)* $|f(x)| \leq B$.

now, we set $f(x) = \mathrm{tr}\left(\boldsymbol{O}\boldsymbol{\rho}(x)\right)$ with $\boldsymbol{\rho}(x)$ ground state of $\boldsymbol{H}(x)$ and $\boldsymbol{O}$ sum of local terms

Proof part 2: bridge to ground state problem

Theorem (streamlined insight from sampling theory)

Uniform sampling efficiently interpolates functions $f : [-1,1]^m \to \mathbb{R}$ that obey (i) $\mathbb{E}_{x \overset{iid}{\sim} [-1,1]^m} \|\nabla_x f(x)\|_2^2 \leq C$ and (ii) $|f(x)| \leq B$.

now, we set $f(x) = \mathrm{tr}(\boldsymbol{O}\boldsymbol{\rho}(x))$ with $\boldsymbol{\rho}(x)$ ground state of $\boldsymbol{H}(x)$ and $\boldsymbol{O}$ sum of local terms

condition (ii): $|f(x)| = |\mathrm{tr}(\boldsymbol{O}\boldsymbol{\rho}(x))| \leq \|\boldsymbol{O}\|_\infty \|\boldsymbol{\rho}(x)\|_1 = \|\boldsymbol{O}|_\infty \leq B$

Proof part 2: bridge to ground state problem

Theorem (streamlined insight from sampling theory)

Uniform sampling efficiently interpolates functions $f : [-1,1]^m \to \mathbb{R}$ that obey (i) $\mathbb{E}_{x \overset{iid}{\sim} [-1,1]^m} \|\nabla_x f(x)\|_2^2 \leq C$ and (ii) $|f(x)| \leq B$.

now, we set $f(x) = \mathrm{tr}(\boldsymbol{O}\boldsymbol{\rho}(x))$ with $\boldsymbol{\rho}(x)$ ground state of $\boldsymbol{H}(x)$ and $\boldsymbol{O}$ sum of local terms

condition (ii): $|f(x)| = |\mathrm{tr}(\boldsymbol{O}\boldsymbol{\rho}(x))| \leq \|\boldsymbol{O}\|_\infty \|\boldsymbol{\rho}(x)\|_1 = \|\boldsymbol{O}|_\infty \leq B$

condition (i) follows ground state properties of 'nice' Hamiltonians and locality

quasi-adiabatic continuation and Lieb-Robinson bounds imply

Proof part 2: bridge to ground state problem

Theorem (streamlined insight from sampling theory)

Uniform sampling efficiently interpolates functions $f : [-1,1]^m \to \mathbb{R}$ *that obey* *(i)* $\mathbb{E}_{x \overset{iid}{\sim} [-1,1]^m} \|\nabla_x f(x)\|_2^2 \leq C$ *and (ii)* $|f(x)| \leq B$.

now, we set $f(x) = \mathrm{tr}\left(\boldsymbol{O}\rho(x)\right)$ with $\rho(x)$ ground state of $\boldsymbol{H}(x)$ and $\boldsymbol{O}$ sum of local terms

condition (ii): $|f(x)| = |\mathrm{tr}\left(\boldsymbol{O}\rho(x)\right)| \leq \|\boldsymbol{O}\|_\infty \|\rho(x)\|_1 = \|\boldsymbol{O}\|_\infty \leq B$

condition (i) follows ground state properties of 'nice' Hamiltonians and locality

quasi-adiabatic continuation and Lieb-Robinson bounds imply

Proposition: $\textsc{SpectralGap}(\boldsymbol{H}(x)) \geq \gamma = \Omega(1)$ for all $x \in [-1,1]^m$ ensures $\|\nabla_x \mathrm{tr}\left(\boldsymbol{O}\rho(x)\right)\|_2^2 \leq C_\gamma \left(\sum_l \|\boldsymbol{O}_l\|_\infty\right)^2 = C_\gamma B^2$ everywhere.

Proof part 2: bridge to ground state problem

Corollary (many-body restatement of main sampling theorem)

Let $\boldsymbol{H}(x) = \sum_j \boldsymbol{H}_j(x)$ *with* $x \in [-1,1]^m$ *be a parametrized family of 'geometrically local' n-qubit Hamiltonians with a constant spectral gap throughout and let* $\boldsymbol{O} = \sum_j \boldsymbol{O}_j$ *be a sum of local observables such that* $\sum_j \|\boldsymbol{O}_j\|_\infty \leq B$. *Then, a total of* $N = B^2 m^{\mathcal{O}(B^2/\epsilon)}$ *labeled ground states* $(x_\ell, \boldsymbol{\rho}(x_\ell))$ *with* $x_\ell \overset{unif}{\sim} [-1,1]^m$ *allows us to interpolate to new ground states:*

$$\boldsymbol{\rho}_{\mathrm{train}}(x) = \frac{1}{N}\sum_{\ell=1}^{N} \kappa_\Lambda(x, x_\ell)\boldsymbol{\rho}(x_\ell) \quad \text{with} \quad \kappa_\Lambda(x, x_\ell) = \sum_{k\in\mathbb{Z}^m:\ \|k\|_2\leq\Lambda} \mathrm{e}^{\mathrm{i}\pi\langle k, x-x_\ell\rangle}, \ \Lambda = \mathcal{O}(B^2/\epsilon).$$

With high probability, $\mathbb{E}_{x\overset{unif}{\sim}[-1,1]^m} \left|\mathrm{tr}\left(\boldsymbol{O}\tilde{\boldsymbol{\rho}}(x)\right) - \mathrm{tr}\left(\boldsymbol{O}\boldsymbol{\rho}(x)\right)\right|^2 \leq \epsilon$ *(*$\mathrm{MSE} \leq \epsilon$*).*

- for $B = \mathrm{const}$ and $\epsilon = \mathrm{const}$, $N = \mathrm{poly}(m) = \mathrm{polylog}(D)$ (efficient training size)
- constant spectral gap is *strong physical assumption* ('deep within a phase')
- procedure is not (yet) efficient: training data $\rho(x_\ell) \in \mathbb{H}_D$ is gigantic ($D = 2^n$)

Proof part 2: bridge to ground state problem

Corollary (many-body restatement of main sampling theorem)

Let $\boldsymbol{H}(x) = \sum_j \boldsymbol{H}_j(x)$ with $x \in [-1,1]^m$ be a parametrized family of 'geometrically local' n-qubit Hamiltonians with a constant spectral gap throughout and let $\boldsymbol{O} = \sum_j \boldsymbol{O}_j$ be a sum of local observables such that $\sum_j \|\boldsymbol{O}_j\|_\infty \leq B$. Then, a total of $N = B^2 m^{\mathcal{O}(B^2/\epsilon)}$ labeled ground states $(x_\ell, \boldsymbol{\rho}(x_\ell))$ with $x_\ell \overset{unif}{\sim} [-1,1]^m$ allows us to interpolate to new ground states:

$$\boldsymbol{\rho}_{\mathrm{train}}(x) = \frac{1}{N}\sum_{\ell=1}^{N} \kappa_\Lambda(x, x_\ell)\boldsymbol{\rho}(x_\ell) \quad \text{with} \quad \kappa_\Lambda(x,x_\ell) = \sum_{k \in \mathbb{Z}^m:\ \|k\|_2 \leq \Lambda} \mathrm{e}^{\mathrm{i}\pi\langle k, x - x_\ell\rangle}, \ \Lambda = \mathcal{O}(B^2/\epsilon).$$

With high probability, $\mathbb{E}_{x \overset{unif}{\sim}[-1,1]^m} \left|\mathrm{tr}\left(\boldsymbol{O}\tilde{\boldsymbol{\rho}}(x)\right) - \mathrm{tr}\left(\boldsymbol{O}\boldsymbol{\rho}(x)\right)\right|^2 \leq \epsilon$ *(*$\mathrm{MSE} \leq \epsilon$*).*

- for $B = \mathrm{const}$ and $\epsilon = \mathrm{const}$, $N = \mathrm{poly}(m) = \mathrm{polylog}(D)$ (efficient training size)
- constant spectral gap is *strong physical assumption* ('deep within a phase')
- procedure is not (yet) efficient: training data $\boldsymbol{\rho}(x_\ell) \in \mathbb{H}_D$ is gigantic ($D = 2^n$)

Proof part 3: data compression with classical shadows

- take-home message from previous slide:
  $\boldsymbol{\rho}(x) \approx \boldsymbol{\rho}_{\mathrm{train}}(x) = \frac{1}{N}\sum_{\ell=1}^{N} \kappa_{\Lambda}(x, x_\ell)\boldsymbol{\rho}(x_\ell)$

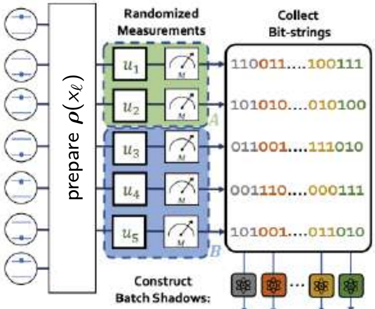
Proof part 3: data compression with classical shadows

- take-home message from previous slide:

  $$\boldsymbol{\rho}(x) \approx \boldsymbol{\rho}_{\mathrm{train}}(x) = \frac{1}{N}\sum_{\ell=1}^{N} \kappa_\Lambda(x, x_\ell)\boldsymbol{\rho}(x_\ell)$$

- linear data compression (sketching) ensures

  $$\hat{\boldsymbol{\sigma}}\left(\boldsymbol{\rho}(x)\right) \approx \hat{\boldsymbol{\sigma}}\left(\boldsymbol{\rho}_{\mathrm{train}}(x)\right) = \frac{1}{N}\sum_{\ell=1}^{N} \kappa_\Lambda(x, x_\ell)\hat{\boldsymbol{\sigma}}(\boldsymbol{\rho}(x_\ell))$$



$$\hat{\boldsymbol{\sigma}}(x_\ell) = \left\{\boldsymbol{S}^{(1)}, \ldots, \boldsymbol{S}^{(T)}\right\}$$

$$\boldsymbol{S}^{(t)} = \boldsymbol{S}_1^{(t)} \otimes \cdots \otimes \boldsymbol{S}_n^{(t)}$$

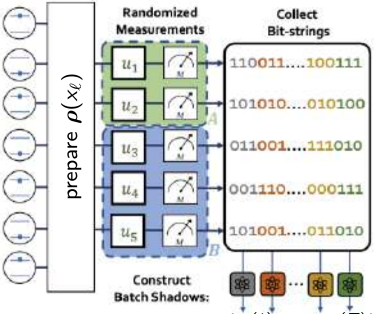
Proof part 3: data compression with classical shadows

- take-home message from previous slide:

  $\boldsymbol{\rho}(x) \approx \boldsymbol{\rho}_{\mathrm{train}}(x) = \frac{1}{N}\sum_{\ell=1}^{N} \kappa_\Lambda(x, x_\ell)\boldsymbol{\rho}(x_\ell)$
- linear data compression (sketching) ensures

  $\hat{\boldsymbol{\sigma}}\left(\boldsymbol{\rho}(x)\right) \approx \hat{\boldsymbol{\sigma}}\left(\boldsymbol{\rho}_{\mathrm{train}}(x)\right) = \frac{1}{N}\sum_{\ell=1}^{N} \kappa_\Lambda(x, x_\ell)\hat{\boldsymbol{\sigma}}(\boldsymbol{\rho}(x_\ell))$
- classical shadows (Monte Carlo) accumulate samples

  if $\boldsymbol{\sigma}(x_\ell)$ contains T shots, then $\hat{\sigma}_{\mathrm{train}}(x)$ contrains $NT \gg q$



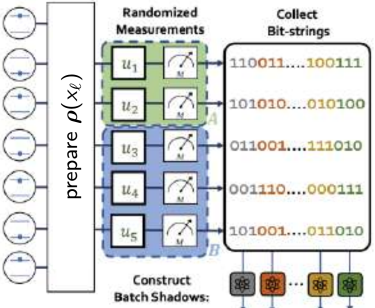
$\hat{\boldsymbol{\sigma}}(x_\ell) = \left\{\boldsymbol{S}^{(1)}, \ldots, \boldsymbol{S}^{(T)}\right\}$

$\boldsymbol{S}^{(t)} = \boldsymbol{S}_1^{(t)} \otimes \cdots \otimes \boldsymbol{S}_n^{(t)}$

Proof part 3: data compression with classical shadows

- take-home message from previous slide:
  $\rho(x) \approx \rho_{\mathrm{train}}(x) = \frac{1}{N}\sum_{\ell=1}^{N} \kappa_\Lambda(x, x_\ell)\rho(x_\ell)$
- linear data compression (sketching) ensures
  $\hat{\sigma}\left(\rho(x)\right) \approx \hat{\sigma}\left(\rho_{\mathrm{train}}(x)\right) = \frac{1}{N}\sum_{\ell=1}^{N} \kappa_\Lambda(x, x_\ell)\hat{\sigma}(\rho(x_\ell))$
- classical shadows (Monte Carlo) accumulate samples
  if $\sigma(x_\ell)$ contains T shots, then $\hat{\sigma}_{\mathrm{train}}(x)$ contrains $NT \gg q$

Proposition: minimal classical shadows with $T = 1$ already ensure MSE ϵ; moreover $\mathrm{STORAGE}(\hat{\sigma}(x_\ell)) = 2.6\mathrm{bits}$
$\Rightarrow$ cheap data acquisition, storage + processing



$\hat{\sigma}(x_\ell) = \left\{ S^{(1)}, \ldots, S^{(T)} \right\}$

$S^{(t)} = S_1^{(t)} \otimes \cdots \otimes S_n^{(t)}$

Theorem (Learning to predict ground state properties with classical shadows)

Let $\boldsymbol{H}(x) = \sum_j \boldsymbol{H}_j(x)$ with $x \in [-1,1]^m$ be a parametrized family of 'geometrically local' n-qubit Hamiltonians with a constant spectral gap throughout and let $\boldsymbol{O} = \sum_j \boldsymbol{O}_j$ be a sum of local observables such that $\sum_j \|O_j\|_\infty \leq B$. Then, a total of $N = B^2 m^{\mathcal{O}(B^2/\epsilon)}$ labeled ground state sketches $(x_\ell, \boldsymbol{\sigma}(x_\ell))$ (minimal classical shadows) with $x_\ell \overset{unif}{\sim} [-1,1]^m$ allows us to interpolate to new ground state sketches:

$$\tilde{\boldsymbol{\sigma}}(x) = \frac{1}{N}\sum_{\ell=1}^{N} \kappa_\Lambda(x, x_\ell)\boldsymbol{\sigma}(\rho(x_\ell)) \quad \textit{with} \quad \kappa_\Lambda(x, x_\ell) = \sum_{k \in \mathbb{Z}^m:\ \|k\|_2 \leq \Lambda} \mathrm{e}^{\mathrm{i}\pi\langle k, x - x_\ell\rangle},\ \Lambda = \mathcal{O}(B^2/\epsilon).$$

With high probability, this interpolation obeys $\mathbb{E}_{x \overset{unif}{\sim}[-1,1]^m} \left|\mathrm{tr}\left(O\tilde{\boldsymbol{\sigma}}(x)\right) - \mathrm{tr}\left(\boldsymbol{O}\boldsymbol{\rho}(x)\right)\right|^2 \leq \epsilon$ (MSE $\leq \epsilon$). Moreover, all computational resources (data compression, storage, training, prediction) are bounded by $\mathcal{O}\left(nB^2 m^{\mathcal{O}(B^2/\epsilon)}\right)$.

- analysis extends to infinite-width neural networks (neural tangent kernel)
- for $\epsilon, B = \text{const}$, $\mathcal{O}(nB^2 m^{\mathcal{O}(B^2/\epsilon)}) = \text{poly}(n) = \text{polylog}(D)$ (efficient cost throughout)
- remaining question: where does training data $(x_\ell, \boldsymbol{\sigma}(x_\ell))$ come from?
- (i) numerical simulations, (ii) quantum computers (quantum-to-classical conversion)

Theorem (Learning to predict ground state properties with classical shadows)

Let $\boldsymbol{H}(x) = \sum_j \boldsymbol{H}_j(x)$ with $x \in [-1,1]^m$ be a parametrized family of 'geometrically local' n-qubit Hamiltonians with a constant spectral gap throughout and let $\boldsymbol{O} = \sum_j \boldsymbol{O}_j$ be a sum of local observables such that $\sum_j \|O_j\|_\infty \leq B$. Then, a total of $N = B^2 m^{\mathcal{O}(B^2/\epsilon)}$ labeled ground state sketches $(x_\ell, \boldsymbol{\sigma}(x_\ell))$ (minimal classical shadows) with $x_\ell \overset{unif}{\sim} [-1,1]^m$ allows us to interpolate to new ground state sketches:

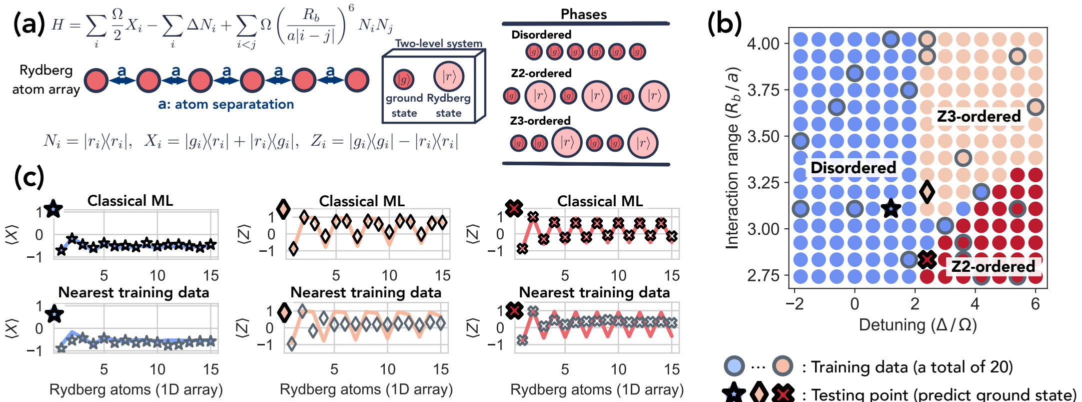
$$\tilde{\boldsymbol{\sigma}}(x) = \frac{1}{N}\sum_{\ell=1}^{N} \kappa_\Lambda(x, x_\ell)\boldsymbol{\sigma}(\rho(x_\ell)) \quad \textit{with} \quad \kappa_\Lambda(x, x_\ell) = \sum_{k \in \mathbb{Z}^m:\ \|k\|_2 \leq \Lambda} \mathrm{e}^{\mathrm{i}\pi\langle k, x - x_\ell\rangle},\ \Lambda = \mathcal{O}(B^2/\epsilon).$$

With high probability, this interpolation obeys $\mathbb{E}_{x \overset{unif}{\sim} [-1,1]^m} \left|\mathrm{tr}\left(O\tilde{\boldsymbol{\sigma}}(x)\right) - \mathrm{tr}\left(\boldsymbol{O}\boldsymbol{\rho}(x)\right)\right|^2 \leq \epsilon$ (MSE $\leq \epsilon$). Moreover, all computational resources (data compression, storage, training, prediction) are bounded by $\mathcal{O}\left(nB^2 m^{\mathcal{O}(B^2/\epsilon)}\right)$.

- analysis extends to infinite-width neural networks (neural tangent kernel)
- for $\epsilon, B = \mathrm{const}$, $\mathcal{O}(nB^2 m^{\mathcal{O}(B^2/\epsilon)}) = \mathrm{poly}(n) = \mathrm{polylog}(D)$ (efficient cost throughout)
- remaining question: where does training data $(x_\ell, \boldsymbol{\sigma}(x_\ell))$ come from?
- (i) numerical simulations, (ii) quantum computers (quantum-to-classical conversion)

numerical experiments

Numerics: 1D chain of $n = 51$ Rydberg atoms

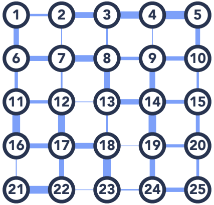


*Solid lines in the six line plots indicate exact values from DMRG

Numerics: 2D Heisenberg model with $n = 25$ spins

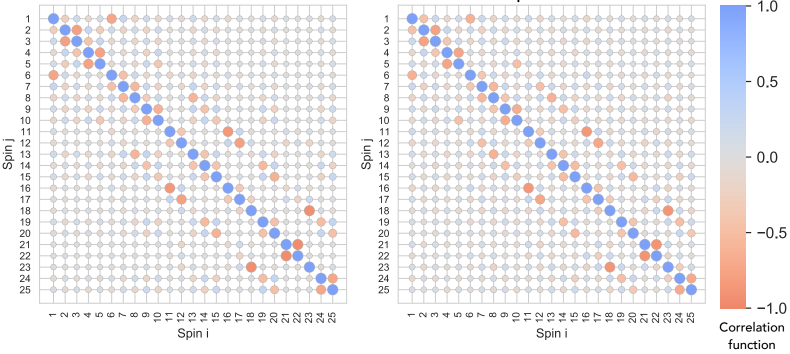
(a) 2D anti-ferromagnetic random Heisenberg model

$$H = \sum_{\langle ij \rangle} J_{ij}(X_i X_j + Y_i Y_j + Z_i Z_j)$$

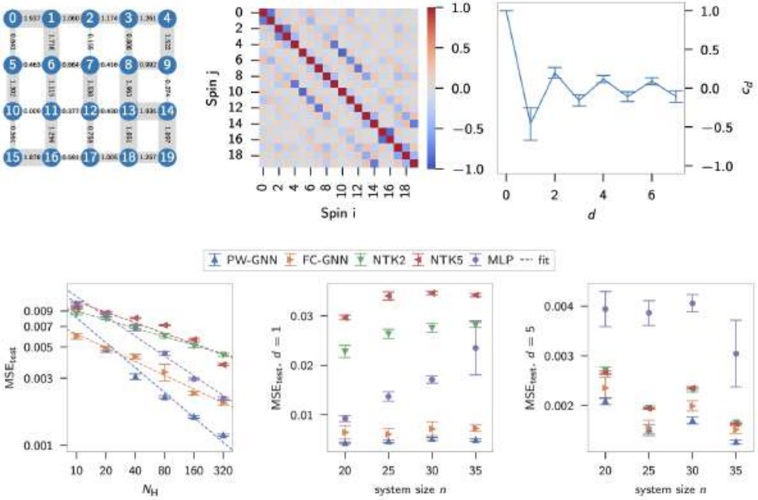


*The random J considered in (c)

(b)



Numerics: 2D Heisenberg model with different ML models



comparison between

- ○ NKT (green,red)
- ○ MLP (purple)
- ○ GNN (blue, orange)

collaboration with Caltech and Hochreiter group (JKU)

[Tran *et al*, NeurIPS workshop 22]

synopsis

Synopsis

- **ambition:** provide rigorous theory support for efficient fruitful between quantum computing/experiments (data generation) and classical ML (generalization)

Synopsis

- **ambition:** provide rigorous theory support for efficient fruitful between quantum computing/experiments (data generation) and classical ML (generalization)
- **concrete objective:** learn to predict properties of parametrized many-body ground states

Synopsis

- **ambition:** provide rigorous theory support for efficient fruitful between quantum computing/experiments (data generation) and classical ML (generalization)
- **concrete objective:** learn to predict properties of parametrized many-body ground states
- **result:** strong assumptions (constant spectral gap & locality) $\Rightarrow$ quasi-polynomial guarantees

Synopsis

- **ambition:** provide rigorous theory support for efficient fruitful between quantum computing/experiments (data generation) and classical ML (generalization)
- **concrete objective:** learn to predict properties of parametrized many-body ground states
- **result:** strong assumptions (constant spectral gap & locality) $\Rightarrow$ quasi-polynomial guarantees

- our **constructive proof** combines three ingredients (each of which is easy for experts)

Synopsis

- **ambition:** provide rigorous theory support for efficient fruitful between quantum computing/experiments (data generation) and classical ML (generalization)
- **concrete objective:** learn to predict properties of parametrized many-body ground states
- **result:** strong assumptions (constant spectral gap & locality) $\Rightarrow$ quasi-polynomial guarantees

- our **constructive proof** combines three ingredients (each of which is easy for experts)
  (i) *harmonic analysis:* approximate slow-varying and bounded function by uniform sampling
  (ii) *many-body physics:* locality & constant spectral gap ensure slow variation + boundedness
  (iii) *classical shadows* allow efficient + linear data compression ('sketching')

Synopsis

- **ambition:** provide rigorous theory support for efficient fruitful between quantum computing/experiments (data generation) and classical ML (generalization)
- **concrete objective:** learn to predict properties of parametrized many-body ground states
- **result:** strong assumptions (constant spectral gap & locality) $\Rightarrow$ quasi-polynomial guarantees

- our **constructive proof** combines three ingredients (each of which is easy for experts)
  (i) *harmonic analysis:* approximate slow-varying and bounded function by uniform sampling
  (ii) *many-body physics:* locality & constant spectral gap ensure slow variation + boundedness
  (iii) *classical shadows* allow efficient + linear data compression ('sketching')

- **refined proof technique** improves training data bound (almost) exponentially [Lewis *et al*, 2023]

Synopsis

- **ambition:** provide rigorous theory support for efficient fruitful between quantum computing/experiments (data generation) and classical ML (generalization)
- **concrete objective:** learn to predict properties of parametrized many-body ground states
- **result:** strong assumptions (constant spectral gap & locality) $\Rightarrow$ quasi-polynomial guarantees

- our **constructive proof** combines three ingredients (each of which is easy for experts)
  - (i) *harmonic analysis:* approximate slow-varying and bounded function by uniform sampling
  - (ii) *many-body physics:* locality & constant spectral gap ensure slow variation + boundedness
  - (iii) *classical shadows* allow efficient + linear data compression ('sketching')

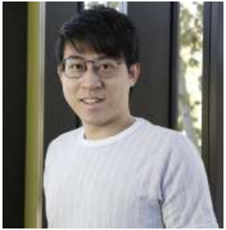
- **refined proof technique** improves training data bound (almost) exponentially [Lewis *et al*, 2023]
- **numerical experiments** highlight that practical performance is even better [Tran *et al*, 2023]

Synopsis

- **ambition:** provide rigorous theory support for efficient fruitful between quantum computing/experiments (data generation) and classical ML (generalization)
- **concrete objective:** learn to predict properties of parametrized many-body ground states
- **result:** strong assumptions (constant spectral gap & locality) $\Rightarrow$ quasi-polynomial guarantees

- our **constructive proof** combines three ingredients (each of which is easy for experts)
  - (i) *harmonic analysis:* approximate slow-varying and bounded function by uniform sampling
  - (ii) *many-body physics:* locality & constant spectral gap ensure slow variation + boundedness
  - (iii) *classical shadows* allow efficient + linear data compression ('sketching')

- **refined proof technique** improves training data bound (almost) exponentially [Lewis *et al*, 2023]
- **numerical experiments** highlight that practical performance is even better [Tran *et al*, 2023]
- **complexity-theoretic bottlenecks** for approaches that don't use training data
  reduction from rectilinear 3-SAT and integer FACTORIZATION


H-Y. Huang


R. Kueng


G. Torlai

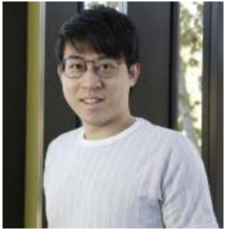

V. Albert


J. Preskill

Provably efficient machine learning for quantum many-body problems.

Science 377, eabk3333 (2022) (and https://arxiv.org/abs/2106.12627)

Thank you!


H-Y. Huang


R. Kueng


G. Torlai


V. Albert


J. Preskill

Provably efficient machine learning for quantum many-body problems.

Science 377, eabk3333 (2022) (and `https://arxiv.org/abs/2106.12627`)

Thank you!

Also, I am building up a team in Linz – PhD and postdoc positions are available.

Please help me spread the word.

Backup slide 1: spectral & locality control avg. gradient size

- Hamiltonian $\boldsymbol{H}_{\mathrm{tot}}(x)$ has spectral gap $\left|\lambda_1^{\uparrow}(\boldsymbol{H}_{\mathrm{tot}}(x)) - \lambda_2^{\uparrow}(\boldsymbol{H}_{\mathrm{tot}}(x))\right| \geq \gamma = \Omega(1)\ \forall x \in [-1,1]^m$
- observable $\boldsymbol{O}_{\mathrm{tot}} = \sum_j \boldsymbol{O}_j$ is sum of 'local' terms, e.g. $\boldsymbol{O}_1 = \tilde{\boldsymbol{O}}_1 \otimes \mathbb{I}^{\otimes(n-o)}$
- **quasi-adiabatic continuation:** $W_\gamma(t)$ is fast-decaying weight function around 0 and

$$(\partial/\partial\hat{u})\,\rho(x) = [\boldsymbol{D}_{\hat{u}}, \rho(x)] \text{ with } \boldsymbol{D}_{\hat{u}} = \int_{\mathbb{R}} W_\gamma(t) \mathrm{e}^{\mathrm{i}t\boldsymbol{H}_{\mathrm{tot}}(x)} \left((\partial/\partial\hat{u})\,\boldsymbol{H}_{\mathrm{tot}}(x)\right) \mathrm{e}^{-\mathrm{i}t\boldsymbol{H}_{\mathrm{tot}}(x)\mathrm{d}t}$$

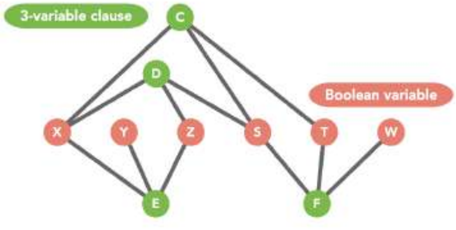
$$\Rightarrow \|\nabla_x \mathrm{tr}\,(\boldsymbol{O}\rho(x))\|_2 = \|\mathrm{tr}\,([\boldsymbol{O}, \boldsymbol{D}_{\hat{u}}]\,\rho(x))| \leq \|[\boldsymbol{O}, \boldsymbol{D}_{\hat{u}}]\|_\infty \|\rho(x)\|_1$$

$$\leq \sum_{j_1,j_2} \int_{\mathbb{R}} W_\gamma(t) \left\| \left[\boldsymbol{O}_{j_1}, \mathrm{e}^{\mathrm{i}t\boldsymbol{H}_{\mathrm{tot}}} \left((\partial/\partial\hat{u})\,\boldsymbol{H}_{j_2}\right) \mathrm{e}^{-\mathrm{i}t\boldsymbol{H}_{\mathrm{tot}}}\right] \right\|_\infty$$

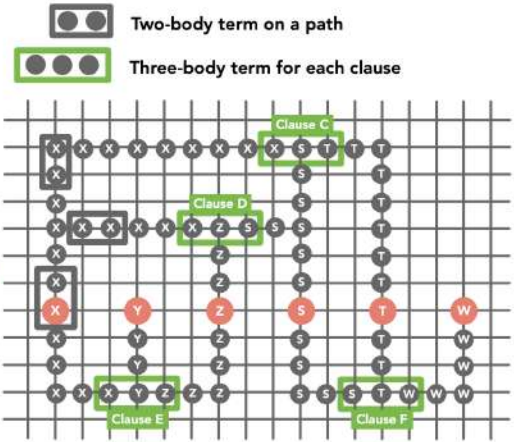
- **locality** implies that most $\boldsymbol{O}_{j_1}$ and $\boldsymbol{H}_{j_2}$ act nontrivially on very distant regions (tensor factors)
- **Lieb-Robinson bounds** ensure that the matrix exponentials don't change this too much

$\Rightarrow$ almost all matrices in (1) commute approximately $\Rightarrow$ small gradient size $\forall x \in [-1,1]^m$

Backup slide 2: reduction from rectilinear 3-SAT



Planar rectilinear 3SAT Problem



Qubit Hamiltonian on 2D grid

Backup slide 3: truncation in Fourier domain

Lemma (truncation)

Let $f : [-1,1]^m \to \mathbb{R}$ *and* $f_\Lambda(x) = \boldsymbol{F}^{-1}\boldsymbol{T}_\Lambda \boldsymbol{F} f(x) = \sum_{k \in \mathbb{Z}^m, \|k\|_2 \leq \Lambda} \mathrm{e}^{\mathrm{i}\pi\langle k,x\rangle} \hat{f}(k)$. *Then,*

$$\mathbb{E}_{x \overset{unif}{\sim} [-1,1]^m} |f(x) - f_\Lambda(x)| \leq \frac{1}{\pi^2\Lambda^2} \mathbb{E}_{x \overset{unif}{\sim} [-1,1]^m} \|\nabla_x f(x)\|_2^2 .$$

$\Rightarrow$ ϵ-error requires cutoff $\Lambda = \mathcal{O}(C/\epsilon)$ with $C \geq \mathbb{E}_x \|\nabla_x f(x)\|_2^2$

$\circ$ in words: average gradient size controls cutoff

Backup slide 3: truncation in Fourier domain

Lemma (truncation)

Let $f : [-1,1]^m \to \mathbb{R}$ *and* $f_\Lambda(x) = \boldsymbol{F}^{-1}\boldsymbol{T}_\Lambda \boldsymbol{F} f(x) = \sum_{k \in \mathbb{Z}^m, \|k\|_2 \leq \Lambda} \mathrm{e}^{\mathrm{i}\pi\langle k,x\rangle} \hat{f}(k)$. *Then,*

$$\mathbb{E}_{x \overset{unif}{\sim} [-1,1]^m} |f(x) - f_\Lambda(x)| \leq \frac{1}{\pi^2\Lambda^2} \mathbb{E}_{x \overset{unif}{\sim} [-1,1]^m} \|\nabla_x f(x)\|_2^2 .$$

$\Rightarrow$ ϵ-error requires cutoff $\Lambda = \mathcal{O}(C/\epsilon)$ with $C \geq \mathbb{E}_x \|\nabla_x f(x)\|_2^2$

$\circ$ in words: average gradient size controls cutoff

proof sketch: (standard Fourier argument with Parseval's identity)

$$\mathbb{E}_{x \overset{unif}{\sim} [-1,1]^m} \|\nabla_x f(x)\|_2^2 = \frac{1}{2^m} \int_{[-1,1]^m} \mathrm{d}^m x \left\| \sum_{k \in \mathbb{Z}^m} \mathrm{i}\pi k \mathrm{e}^{\mathrm{i}\pi\langle k,x\rangle} \hat{f}(k) \right\|_2^2 = \pi^2 \sum_{k \in \mathbb{Z}^m} \|k\|_2^2 \left| \hat{f}(k) \right|^2$$

$$\mathbb{E}_{x \overset{unif}{\sim} [-1,1]^m} |f(x) - f_\Lambda(x)| = \int_{[-1,1]^m} \mathrm{d}^m x \left| \sum_{\|k\|_2 > \Lambda} \mathrm{e}^{\mathrm{i}\pi\langle k,x\rangle} \hat{f}(k) \right| = \sum_{\|k\|_2 > \Lambda} \left| \hat{f}(k) \right|^2$$

Backup slide 4: sampling approximation

Lemma (sampling rate for MSE approximation)

Let $f_\Lambda(x)$ be a band-limited function with $|f_\Lambda(x)| \leq B$ a.s. Then, with high (constant) prob: $\tilde{f}_\Lambda(x) = \frac{1}{N}\sum_{\ell=1}^{N} \kappa_\Lambda(x, x_\ell) f(x_\ell)$ *with* $x_\ell \overset{unif}{\sim} [-1,1]^m$ *and* $k_\Lambda(x, x_\ell) = \sum_{k \in \mathbb{Z}^m:\ \|k\|_2 \leq \Lambda} \mathrm{e}^{\mathrm{i}\pi\langle k, x - x_\ell\rangle}$ *obeys:*

$$a = \mathbb{E}_{x \overset{unif}{\sim} [-1,1]^m} \left| \tilde{f}_\Lambda(x) - f_\Lambda(x) \right|^2 \leq B^2 m^{\mathcal{O}(\Lambda^2)}/N.$$

in words: bounded function value implies bounded (uniform) sampling rate N

Backup slide 4: sampling approximation

Lemma (sampling rate for MSE approximation)

Let $f_\Lambda(x)$ be a band-limited function with $|f_\Lambda(x)| \leq B$ a.s. Then, with high (constant) prob: $\tilde{f}_\Lambda(x) = \frac{1}{N}\sum_{\ell=1}^N \kappa_\Lambda(x, x_\ell) f(x_\ell)$ with $x_\ell \overset{unif}{\sim} [-1,1]^m$ and $k_\Lambda(x, x_\ell) = \sum_{k \in \mathbb{Z}^m:\ \|k\|_2 \leq \Lambda} \mathrm{e}^{\mathrm{i}\pi \langle k, x - x_\ell \rangle}$ obeys:

$$a = \mathbb{E}_{x \overset{unif}{\sim} [-1,1]^m} \left| \tilde{f}_\Lambda(x) - f_\Lambda(x) \right|^2 \leq B^2 m^{\mathcal{O}(\Lambda^2)}/N.$$

in words: bounded function value implies bounded (uniform) sampling rate N

proof sketch: (polynomial bound on lattice size + Hoeffding)

$$a = \frac{1}{2^m}\int_{[-1,1]^m} \mathrm{d}^m x \left| \sum_{\|k\|_2 \leq \Lambda} \mathrm{e}^{\mathrm{i}\pi\langle k, x\rangle} \left(\frac{1}{N}\sum_{\ell=1}^N \left(\mathrm{e}^{-\mathrm{i}\pi\langle k, x_\ell\rangle} f(x_\ell) - \mathbb{E}_{x_\ell}\left[\mathrm{e}^{-\mathrm{i}\pi\langle k, x_\ell\rangle} f(x_\ell)\right]\right)\right)\right|^2$$

$$= \sum_{k \in \mathbb{Z}^m : \|k\|_2 \leq \Lambda} \left| \frac{1}{N}\sum_{\ell=1}^N \left(\mathrm{e}^{-\mathrm{i}\pi\langle k, x_\ell\rangle} f(x_\ell) - \mathbb{E}_{x_\ell}\left[\mathrm{e}^{-\mathrm{i}\pi\langle k, x_\ell\rangle} f(x_\ell)\right]\right)\right|^2$$

Backup slide 4: sampling approximation

Lemma (sampling rate for MSE approximation)

Let $f_\Lambda(x)$ be a band-limited function with $|f_\Lambda(x)| \leq B$ a.s. Then, with high (constant) prob: $\tilde{f}_\Lambda(x) = \frac{1}{N}\sum_{\ell=1}^{N} \kappa_\Lambda(x,x_\ell) f(x_\ell)$ *with* $x_\ell \overset{unif}{\sim} [-1,1]^m$ *and* $k_\Lambda(x,x_\ell) = \sum_{k\in\mathbb{Z}^m:\ \|k\|_2\leq\Lambda} \mathrm{e}^{\mathrm{i}\pi\langle k, x-x_\ell\rangle}$ *obeys:*

$$a = \mathbb{E}_{x\overset{unif}{\sim}[-1,1]^m} \left|\tilde{f}_\Lambda(x) - f_\Lambda(x)\right|^2 \leq B^2 m^{\mathcal{O}(\Lambda^2)}/N.$$

in words: bounded function value implies bounded (uniform) sampling rate N

proof sketch: (polynomial bound on lattice size + Hoeffding)

$$a = \frac{1}{2^m}\int_{[-1,1]^m} \mathrm{d}^m x \left|\sum_{\|k\|_2\leq\Lambda} \mathrm{e}^{\mathrm{i}\pi\langle k,x\rangle}\left(\frac{1}{N}\sum_{\ell=1}^{N}\left(\mathrm{e}^{-\mathrm{i}\pi\langle k,x_\ell\rangle} f(x_\ell) - \mathbb{E}_{x_\ell}\left[\mathrm{e}^{-\mathrm{i}\pi\langle k,x_\ell\rangle} f(x_\ell)\right]\right)\right)\right|^2$$

$$= \sum_{k\in\mathbb{Z}^m:\|k\|_2\leq\Lambda}\left|\frac{1}{N}\sum_{\ell=1}^{N}\left(\mathrm{e}^{-\mathrm{i}\pi\langle k,x_\ell\rangle} f(x_\ell) - \mathbb{E}_{x_\ell}\left[\mathrm{e}^{-\mathrm{i}\pi\langle k,x_\ell\rangle} f(x_\ell)\right]\right)\right|^2$$

$$\leq \left|\{k\in\mathbb{Z}^m:\ \|k\|_2\leq\Lambda\}\right| \max_{\|k\|_2\leq\Lambda}\left|\frac{1}{N}\sum_{\ell=1}^{N}\left(\mathrm{e}^{-\mathrm{i}\pi\langle k,x_\ell\rangle} f(x_\ell) - \mathbb{E}_{x_\ell}\left[\mathrm{e}^{-\mathrm{i}\pi\langle k,x_\ell\rangle} f(x_\ell)\right]\right)\right|^2$$

$(2m+1)^{\Lambda^2} = m^{\mathcal{O}(\Lambda^2)}$ tail bound (Hoeffding)