Title: Neural-Shadow Quantum State Tomography

Speakers: Victor Wei

Series: Machine Learning Initiative

Date: November 10, 2023 - 2:30 PM

URL: https://pirsa.org/23110056

Abstract: Quantum state tomography (QST) is the art of reconstructing an unknown quantum state through measurements. It is a key primitive for developing quantum technologies. Neural network quantum state tomography (NNQST), which aims to reconstruct the quantum state via a neural network ansatz, is often implemented via a basis-dependent cross-entropy loss function. State-of-the-art implementations of NNQST are often restricted to characterizing a particular subclass of states, to avoid an exponential growth in the number of required measurement settings. In this talk, I will discuss an alternative neural-network-based QST protocol that uses shadow-estimated infidelity as the loss function, named "neural-shadow quantum state tomography" (NSQST). After introducing NNQST and the classical shadow formalism, I will present numerical results on the advantage of NSQST over NNQST at learning the relative phases, NSQST's noise robustness, and NSQST's advantage over direct shadow estimation. I will also briefly discuss the future prospects of the protocol with different variational ansatz and randomized measurements, as well as its experimental feasibility.

Zoom link https://pitp.zoom.us/j/94167105773?pwd=TXR3TUtwNjV4VFB4SEpvTkhqd29SUT09

Neural-Shadow Quantum State Tomography

Victor Wei, Bill Coish, Pooya Ronagh, and Christine Muschik

arXiv:2305.01078 (2023)

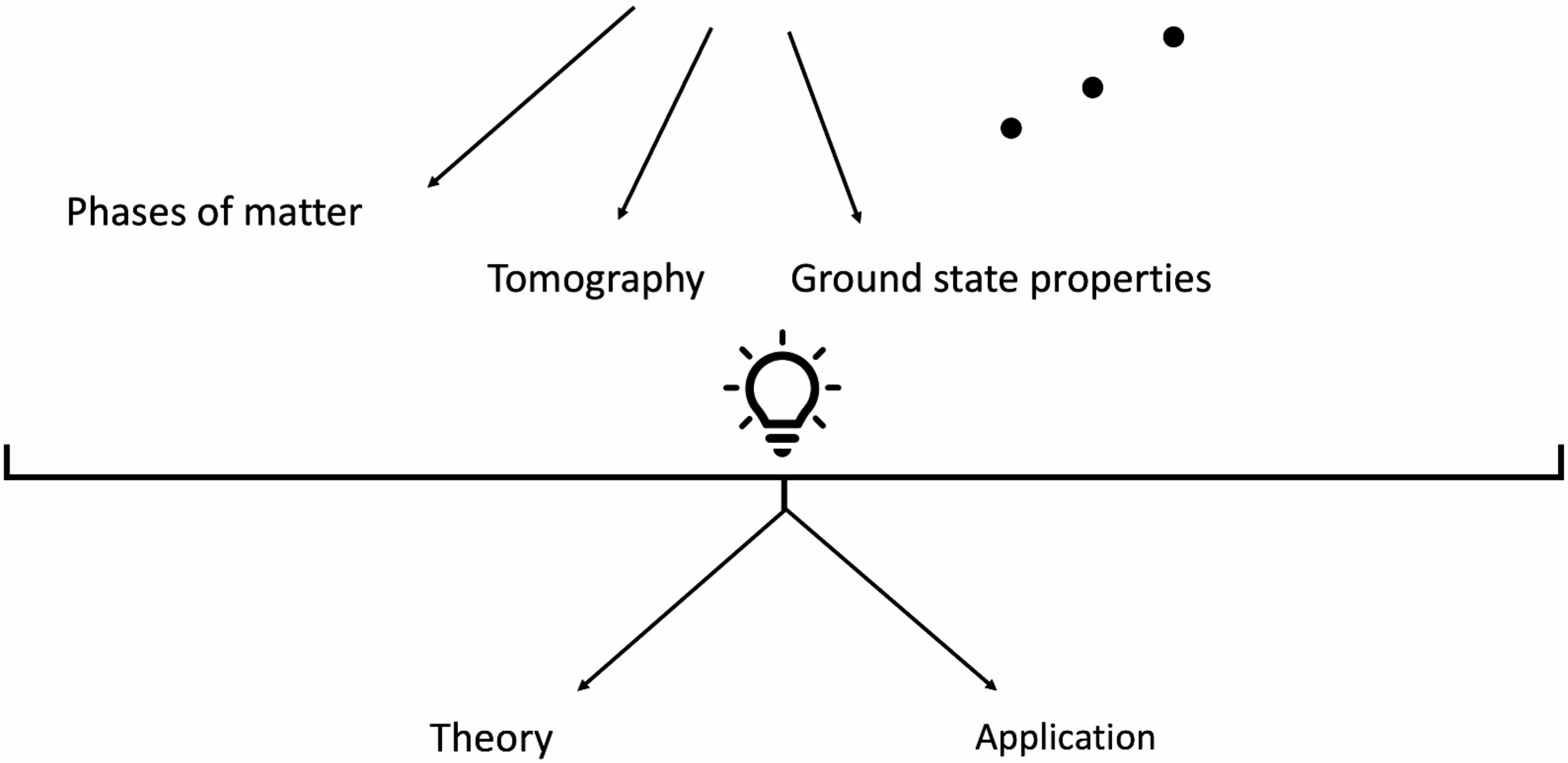
PIQuIL Seminar Talk, Nov. 10th, 2023

Quantum Interactions

McGill UNIVERSITY

IQC Institute for Quantum Computing

QUANTUM INTELLIGENCE LAB



Overview

- 1) Neural network quantum state tomography (NNQST)
- 2) Classical shadow formalism
- 3) Neural-shadow quantum state tomography (NSQST)
- 4) Advantages over NNQST for time-evolved states
- 5) Scalable advantages over direct shadow estimation (*new)
- 6) Noise robustness
- 7) Experimental prospects and future directions

Quantum state tomography

- Quantum state tomography is extremely important for verifying experimentally prepared quantum states.

- Full quantum state tomography requires one to specify $4^n - 1$ independent parameters, therefore exponentially many measurements.

- The exponential scaling makes full quantum state tomography unfeasible for larger quantum systems

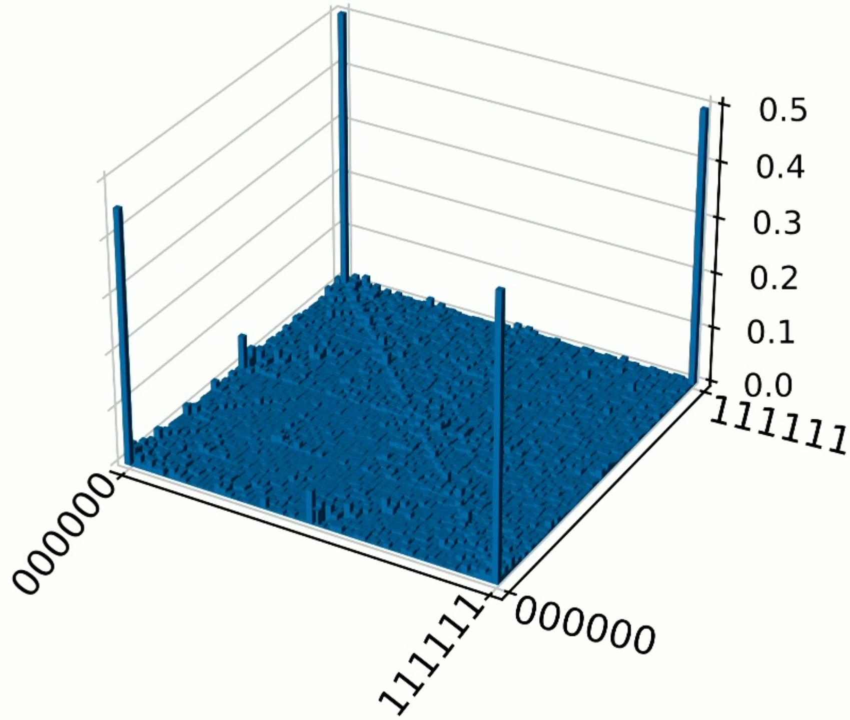


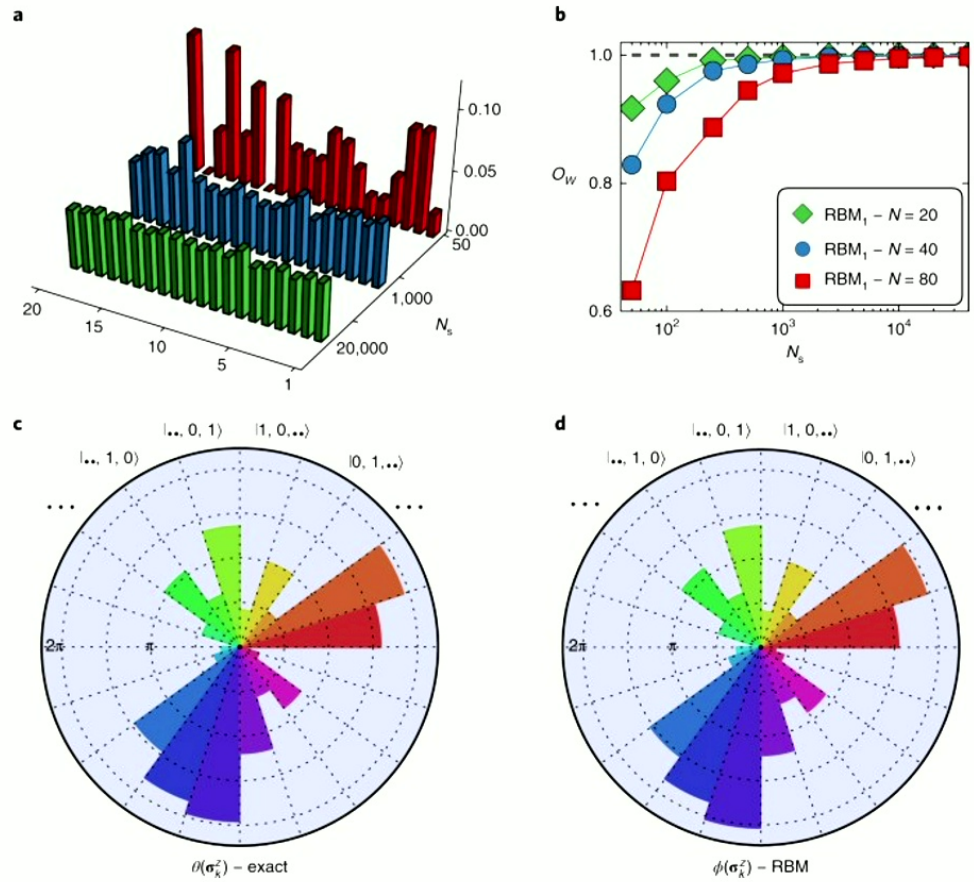
Image credit: Mach. Learn.: Sci. Technol. 3 01LT01

Neural network quantum state tomography (pure state)

Torlai *et al. Nature Phys* **14**, 447–450 (2018).

Neural-network quantum state tomography

Giacomo Torlai, Guglielmo Mazzola, Juan Carrasquilla, Matthias Troyer, Roger Melko & Giuseppe Carleo



Target state:

$$|\Psi_W\rangle = \frac{1}{\sqrt{N}}(|100\ldots\rangle + \ldots + |\ldots 001\rangle)$$

Variational ansatz:

$$\psi_{\lambda,\mu}(\mathbf{x}) = \sqrt{\frac{p_\lambda(\mathbf{x})}{Z_\lambda}}\underbrace{e^{i\phi_\mu(\mathbf{x})/2}}_{\text{relative phases}}$$

NNQST training: loss function

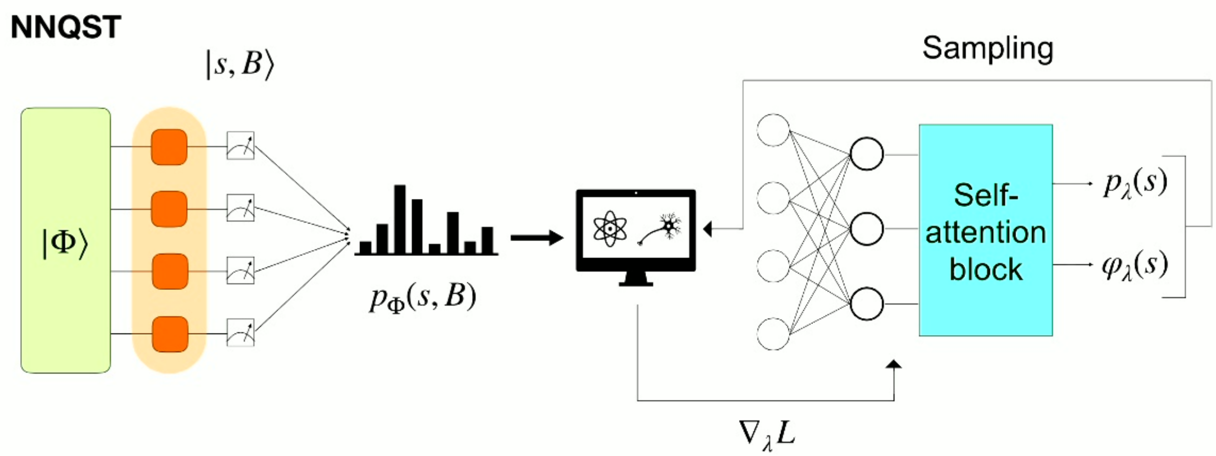
The probability distribution can be learnt from direct measurements.

But learning the relative phases requires measurements in rotated bases.

The training dataset consists of measurements over a set of Pauli bases {B}.

$$L_\lambda \approx -\frac{1}{|\mathcal{D}_T|} \sum_{|s,B\rangle \in \mathcal{D}_T} \ln p_{\psi_\lambda}(s,B)$$

$$p_{\psi_\lambda}(s,B) = \Big| \sum_{\substack{t \in \{0,1\}^n \\ \langle s,B|t\rangle \neq 0}} \langle s,B|t\rangle\langle t|\psi_\lambda\rangle \Big|^2$$



NNQST training: choosing bases

Two important points:

1. The chosen set of bases should be "informationally complete".

   Ex. If the prepared state is the unique ground state of a known Hamiltonian, choose the Hamiltonian's Pauli strings as bases.

2. The chosen bases should be "nearly diagonal".

   It means B should have few X or Y elements.

   This is necessary for efficient classical post-processing, for evaluating $p_{\psi_\lambda}(s,B) = \Big| \sum_{\substack{t\in\{0,1\}^n \\ \langle s,B|t\rangle\neq 0}} \langle s,B|t\rangle\langle t|\psi_\lambda\rangle \Big|^2$

   Ex. For the W state, Torlai *et al.* chose the following bases:

$$\{X,X,Z,Z,\dots\}, \{Z,X,X,Z,\dots\}, \{Z,Z,X,X,\dots\}$$

$$\{X,Y,Z,Z,\dots\}, \{Z,X,Y,Z,\dots\}, \{Z,Z,X,Y,\dots\}$$

NNQST prediction: Pauli observables and fidelity

1) For Pauli observables, we use the local estimator and collect samples from the neural network ansatz

$$\langle \mathcal{O}^{ND} \rangle = \sum_{\boldsymbol{\sigma}} |\psi_{\boldsymbol{\lambda}}(\boldsymbol{\sigma})|^2 \mathcal{O}_L(\boldsymbol{\sigma}) \simeq \frac{1}{n} \sum_{k=1}^{n} \mathcal{O}_L(\boldsymbol{\sigma}_k) \qquad \mathcal{O}_L(\boldsymbol{\sigma}) = \sum_{\boldsymbol{\sigma}'} \sqrt{\frac{p_{\boldsymbol{\lambda}}(\boldsymbol{\sigma}')}{p_{\boldsymbol{\lambda}}(\boldsymbol{\sigma})}} \mathcal{O}_{\boldsymbol{\sigma}\boldsymbol{\sigma}'}$$

The variance of this local estimator is a constant, independent of the system size or the observable weight ("weight" counts non-identity Pauli matrices in O).

2) For fidelity to another state, we can also define an estimator as fraction of amplitudes, with constant variance.

Detailed proofs can be found in Havlicek, *Quantum* 7, 938 (2023).

NNQST prediction: Rényi-2 entropy

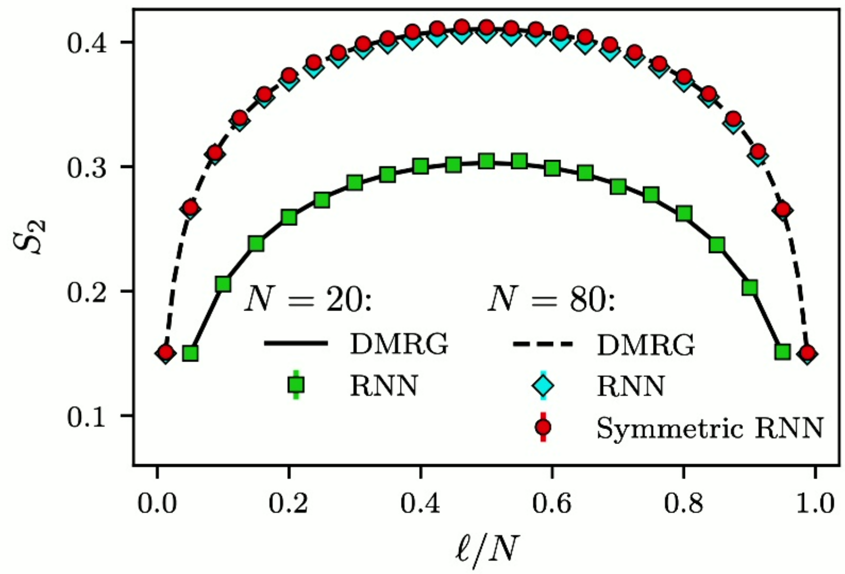
An important non-linear function of interest is Rényi-2 entropy.

$$S_\alpha(A) = \frac{1}{1-\alpha} \log \left(\mathrm{Tr} \rho_A^\alpha \right)$$

$$\langle \mathrm{Swap}_A \rangle = \sum_{\boldsymbol{\sigma}, \tilde{\boldsymbol{\sigma}}} \psi_\lambda^*(\boldsymbol{\sigma}_A \boldsymbol{\sigma}_B) \psi_\lambda^*(\tilde{\boldsymbol{\sigma}}_A \tilde{\boldsymbol{\sigma}}_B) \psi_\lambda(\tilde{\boldsymbol{\sigma}}_A \boldsymbol{\sigma}_B) \psi_\lambda(\boldsymbol{\sigma}_A \tilde{\boldsymbol{\sigma}}_B)$$
$$= \mathrm{Tr} \rho_A^2 = \exp[-S_2(A)].$$

Estimating Swap operator also has a bounded variance, independent of the subsystem size.

Leveraging the sampling advantage of autoregressive models, a figure from Hibat-Allah et al. *Phys. Rev. Research* **2**, 023358 (2020).



From NNQST (mixed state) to classical shadows

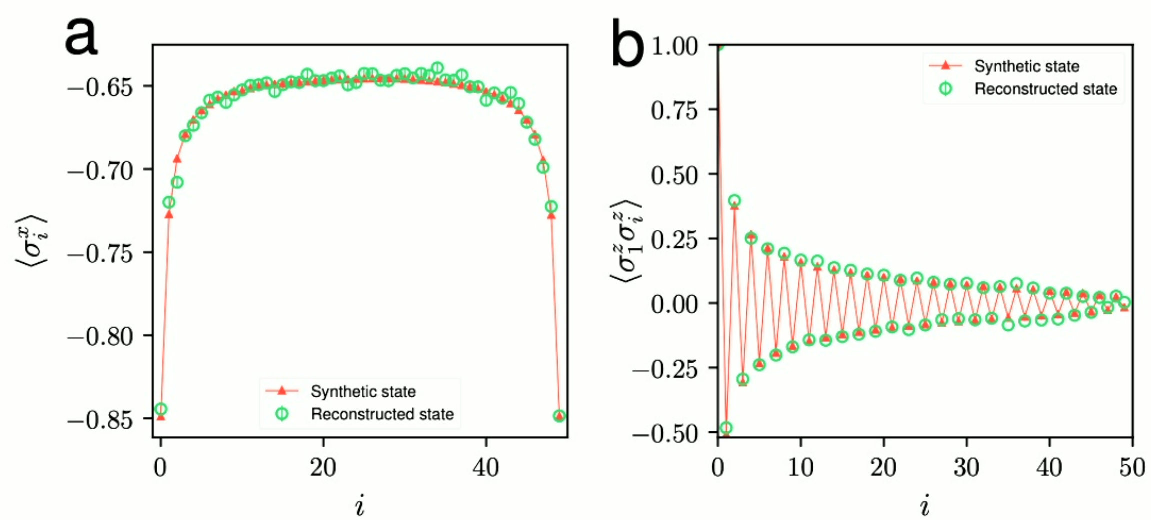
- Motivation: If it is hard to choose a set of informationally-complete bases, can we avoid basis selection by "pushing everything to a single probability distribution"?

Carrasquilla *et al. Nature Phys* **14**, 447–450 (2018).

Reconstructing quantum states with generative models

Juan Carrasquilla, Giacomo Torlai, Roger G. Melko & Leandro Aolita

Answer is yes! For an informationally-complete POVM with invertible overlap matrix, one can push everything to a typical generative model, and predict observables using the known overlap matrix.



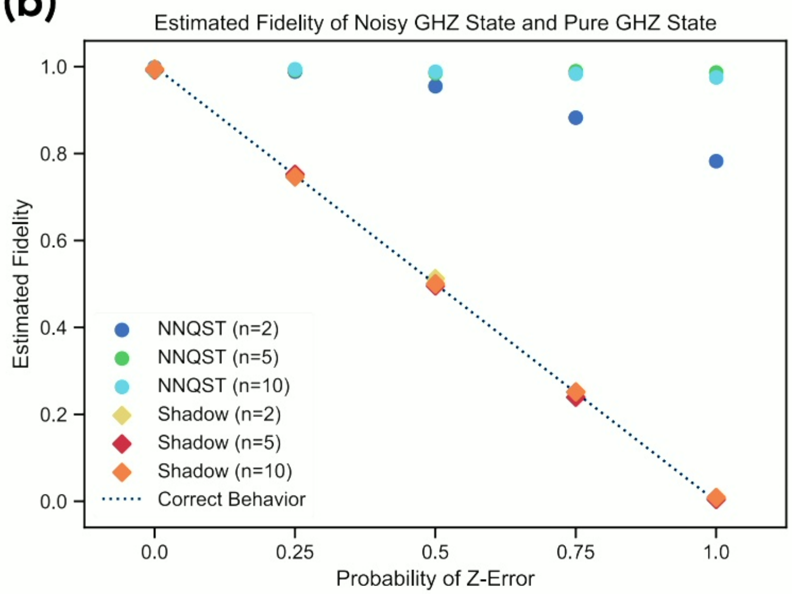
But…

- Unlike pure state ansatz, mixed state NNQST based on informationally-complete POVM has a hard time predicting non-local observables at large system size, such as fidelity to another state.

- Moreover, NNQST's classical fidelity does not reflect phase errors well even at small system sizes.

- To further explore the boundaries of tomography from a theoretical perspective, including the sample efficiency for different observables, classical shadow formalism was introduced.

(b)



Predicting many properties of a quantum system from very few measurements

Hsin-Yuan Huang, Richard Kueng & John Preskill

Huang *et al. Nature Phys* **16**, 1050–1057 (2020)

Classical shadow formalism

- Unlike **model-based** mixed state NNQST where the **inverse overlap matrix** is only used after a physical probability distribution has been learnt, classical shadow formalism inverts every **randomized** measurement snapshot using an analytically derived **inverted quantum channel**.

Classical shadow expression:

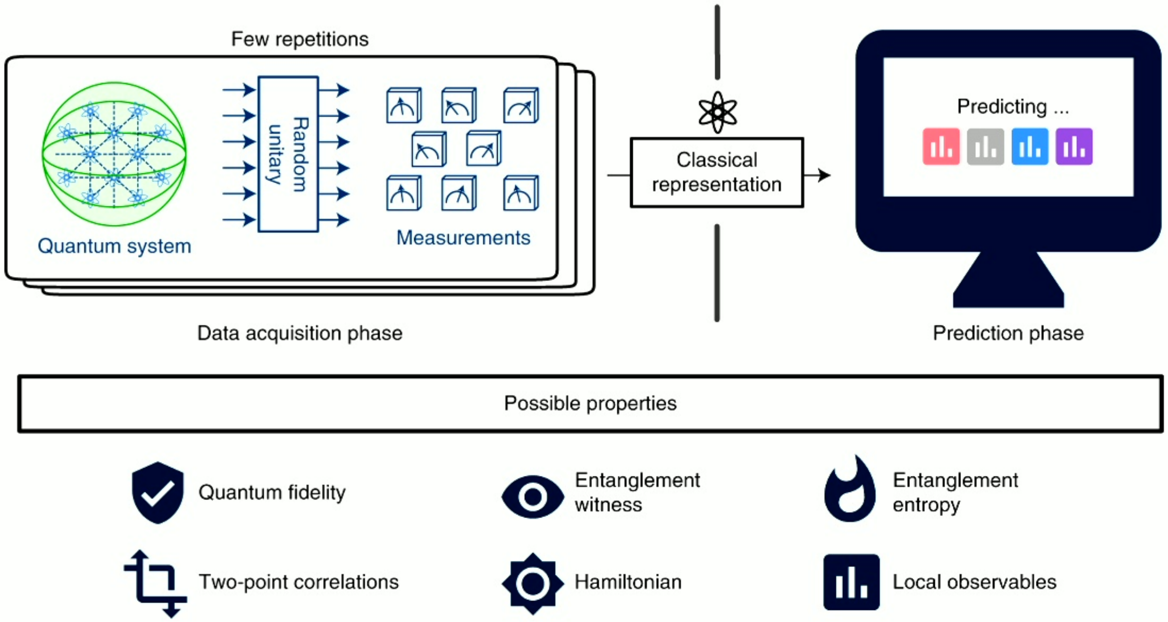
$$\hat{\rho} = \mathcal{M}^{-1}\left(U^\dagger |\hat{b}\rangle\langle\hat{b}| U\right)$$

Prediction:

$$\hat{o}_i = \mathrm{tr}\left(O_i \hat{\rho}\right) \quad \text{obeys} \quad \mathbb{E}\left[\hat{o}\right] = \mathrm{tr}\left(O_i \rho\right)$$

Variance bound (sample efficiency):

$$\mathrm{Var}\left[\hat{o}\right] = \mathbb{E}\left[\left(\hat{o} - \mathbb{E}\left[\hat{o}\right]\right)^2\right] \leq \left\| O - \frac{\mathrm{tr}(O)}{2^n}\mathbb{I} \right\|^2_{\text{shadow}}$$



Huang *et al. Nature Phys* **16**, 1050–1057 (2020)

Classical shadow: Pauli and Clifford

- Pauli shadow:
  1) measure in randomized Pauli bases with few single qubit gates
  2) ideal for local Pauli observables, with variance bound $4^{\text{locality}(O)}\|O\|_\infty^2$
  3) efficient classical post-processing with Pauli observable

- Clifford shadow:
  1) measure in randomized Clifford bases with O(n^2) entangling gates
  2) ideal for low-rank observables such as fidelity, with variance bound $3\text{tr}(O^2)$
  3) efficient classical post-processing with **only stabilizer states**

Non-linear functions:

Pauli shadows can be used efficiently predict Rényi-2 entropy with **small subsystem size**

Classical shadow: Pauli and Clifford

- Pauli shadow:

  1) measure in randomized Pauli bases with few single qubit gates

  2) ideal for local Pauli observables, with variance bound $4^{\mathrm{locality}(O)}\|O\|_\infty^2$

  3) efficient classical post-processing with Pauli observable

- Clifford shadow:

  1) measure in randomized Clifford bases with O(n^2) entangling gates

  2) ideal for low-rank observables such as fidelity, with variance bound $3\mathrm{tr}(O^2)$

  3) efficient classical post-processing with **only stabilizer states**

Non-linear functions:

Pauli shadows can be used efficiently predict Rényi-2 entropy with **small subsystem size**

*** Neither Pauli or Clifford shadows are guaranteed to efficiently predict high-weight Pauli observables or Rényi-2 entropy of large subsystem size.**

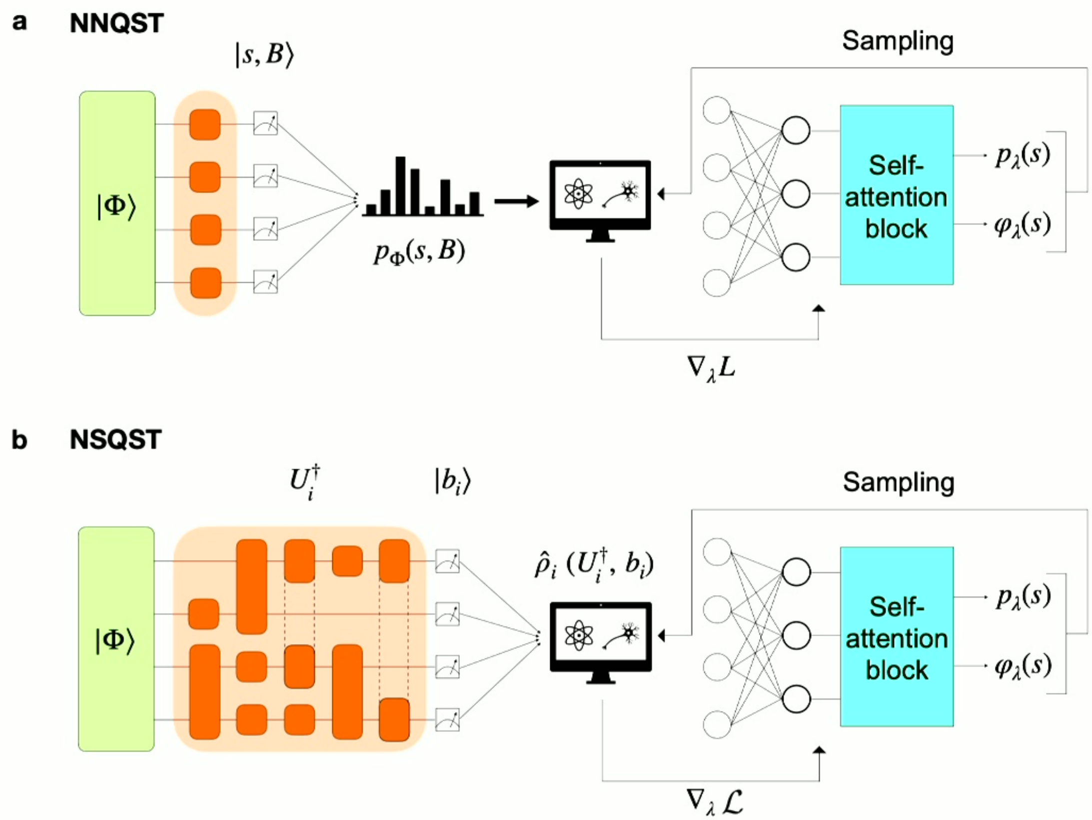
Combine NNQST and classical shadows?

Machine-learning Augmented Shadow Tomography (Part I)

Show affiliations

Cha, Peter ; Skaras, Tim ; Huang, Robert ; Carrasquilla, Juan ; McMahon, Peter ; Kim, Eun-Ah

- I decided to approach it differently, train neural network quantum state with classical shadows.
- In particular, I used Clifford shadows as training data and an autoregressive model from Bennewitz et al. *Nat Mach Intell* 4, 618–624 (2022).



NSQST loss function

- We use an infidelity-based loss function, where the infidelity is approximated by a set of Clifford shadows.

$$\begin{aligned}\mathcal{L}_\lambda(\mathcal{E}) &:= 1 - |\langle\psi_\lambda|\Phi\rangle|^2 \\ &\approx 1 - \frac{1}{N}\sum_{i=1}^{N}\mathrm{Tr}(O_\lambda\hat{\rho}_i) \\ &= 1 - \frac{1}{N}\sum_{i=1}^{N}\langle\psi_\lambda|\,\hat{\rho}_i(\mathcal{E},U_i,b_i)\,|\psi_\lambda\rangle\end{aligned}$$

- In the noiseless case, the Clifford shadow takes the expression $\hat{\rho}_i(U_i,b_i) := \mathcal{M}^{-1}(|\phi_i\rangle\langle\phi_i|) = (2^n+1)|\phi_i\rangle\langle\phi_i| - \mathbb{I}$

$$\mathcal{L}_\lambda(\mathbb{I}) \approx 2 - \frac{2^n+1}{N}\sum_{i=1}^{N}|\,\langle\phi_i|\psi_\lambda\rangle\,|^2$$

NSQST loss function

- We use an infidelity-based loss function, where the infidelity is approximated by a set of Clifford shadows.

$$\begin{aligned}\mathcal{L}_\lambda(\mathcal{E}) &:= 1 - |\langle\psi_\lambda|\Phi\rangle|^2 \\ &\approx 1 - \frac{1}{N}\sum_{i=1}^{N}\mathrm{Tr}(O_\lambda\hat{\rho}_i) \\ &= 1 - \frac{1}{N}\sum_{i=1}^{N}\langle\psi_\lambda|\,\hat{\rho}_i(\mathcal{E}, U_i, b_i)\,|\psi_\lambda\rangle\end{aligned}$$

- In the noiseless case, the Clifford shadow takes the expression $\hat{\rho}_i(U_i, b_i) := \mathcal{M}^{-1}(|\phi_i\rangle\langle\phi_i|) = (2^n+1)|\phi_i\rangle\langle\phi_i| - \mathbb{I}$

$$\mathcal{L}_\lambda(\mathbb{I}) \approx 2 - \frac{2^n+1}{N}\sum_{i=1}^{N}|\langle\phi_i|\psi_\lambda\rangle|^2 \longrightarrow$$

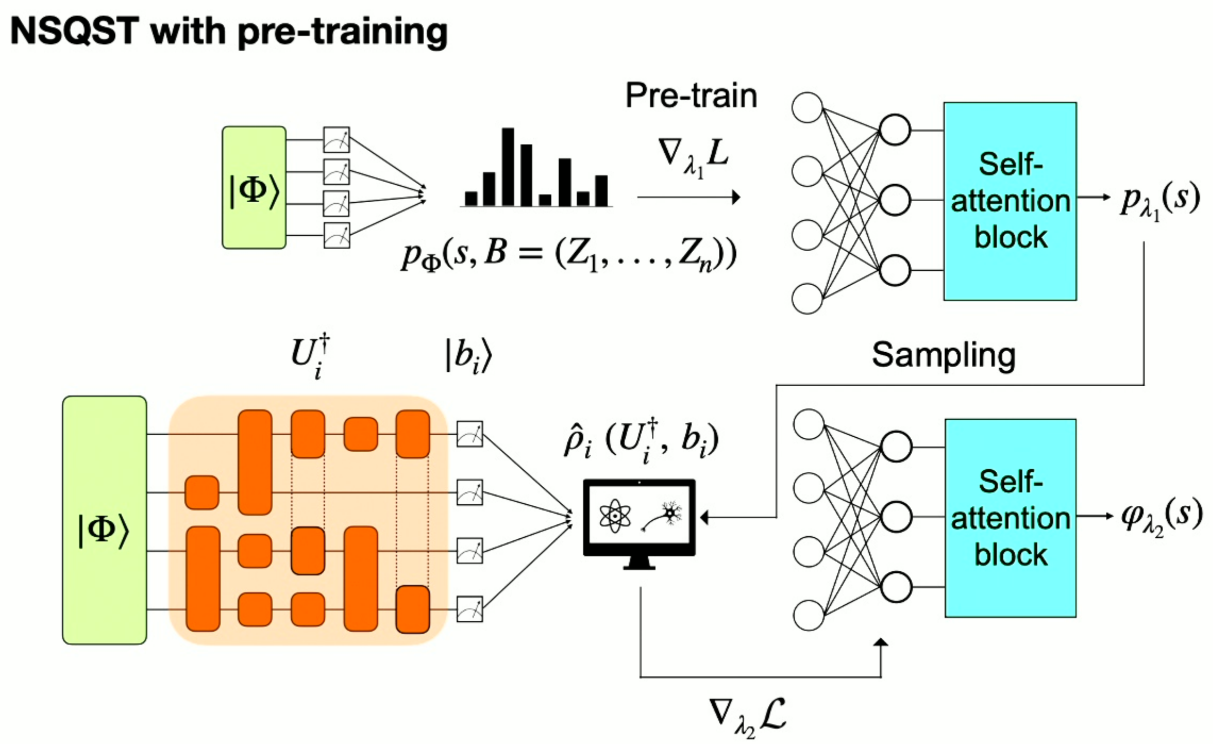
Cannot be computed exactly when system size is large, one of Clifford shadow's classical post-processing issue…

- Let's first approximate it using Monte Carlo method and try it on small systems. Just for now, a new set of Clifford shadows are measured in every iteration to avoid overfitting.

NSQST with pre-training

- Pre-train the probability distribution with direct measurements, then learn the relative phases with shadows.

- The pre-training step is about learning a physical probability distribution, beyond the information contained in the unphysical classical shadows.



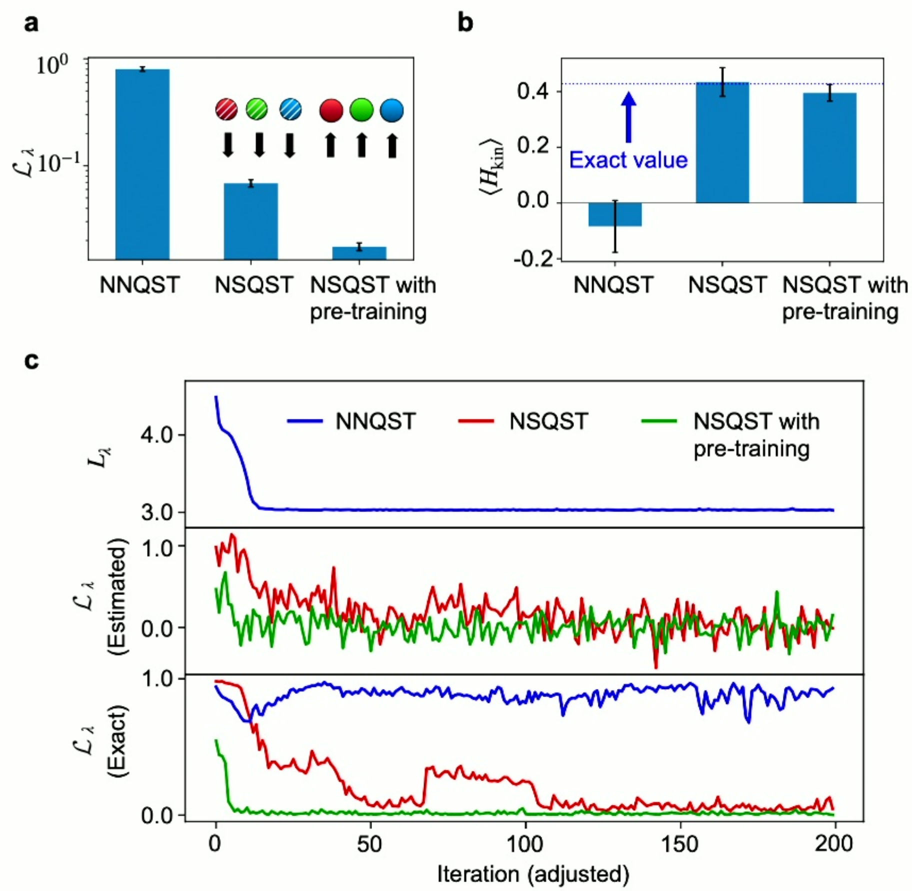
Advantages over NNQST for time-evolved states

- To find interesting examples beyond ground states, we chose a 6-qubit time-evolved state (after Trotterized time evolution).

- NNQST is trained on 21 nearly-diagonal measurement bases (512 shots per base).
- NSQST is trained on using 100 Clifford shadows per iteration.
- NSQST with pre-training, the pre-training stage uses NNQST's measurement resources, then trained using 100 Clifford shadows per iteration.

$$H_{kin} = -\frac{1}{2}(\sigma_1^+\sigma_2^z\sigma_3^z\sigma_4^- - \sigma_2^+\sigma_3^z\sigma_4^z\sigma_5^- + \sigma_3^+\sigma_4^z\sigma_5^z\sigma_6^- + \mathrm{H.c.})$$

Atas et al., Phys. Rev. Research **5**, 033184 (2023).



NSQST loss function

- We use an infidelity-based loss function, where the infidelity is approximated by a set of Clifford shadows.

$$
\begin{aligned}
\mathcal{L}_\lambda(\mathcal{E}) &:= 1 - |\langle\psi_\lambda|\Phi\rangle|^2 \\
&\approx 1 - \frac{1}{N}\sum_{i=1}^{N} \mathrm{Tr}(O_\lambda \hat{\rho}_i) \\
&= 1 - \frac{1}{N}\sum_{i=1}^{N} \langle\psi_\lambda|\, \hat{\rho}_i(\mathcal{E}, U_i, b_i)\, |\psi_\lambda\rangle
\end{aligned}
$$

- In the noiseless case, the Clifford shadow takes the expression $\hat{\rho}_i(U_i, b_i) := \mathcal{M}^{-1}(|\phi_i\rangle\langle\phi_i|) = (2^n+1)|\phi_i\rangle\langle\phi_i| - \mathbb{I}$

$$
\mathcal{L}_\lambda(\mathbb{I}) \approx 2 - \frac{2^n+1}{N}\sum_{i=1}^{N} |\langle\phi_i|\psi_\lambda\rangle|^2 \longrightarrow
$$

Cannot be computed exactly when system size is large, one of Clifford shadow's classical post-processing issue…

- Let's first approximate it using Monte Carlo method and try it on small systems. Just for now, a new set of Clifford shadows are measured in every iteration to avoid overfitting.

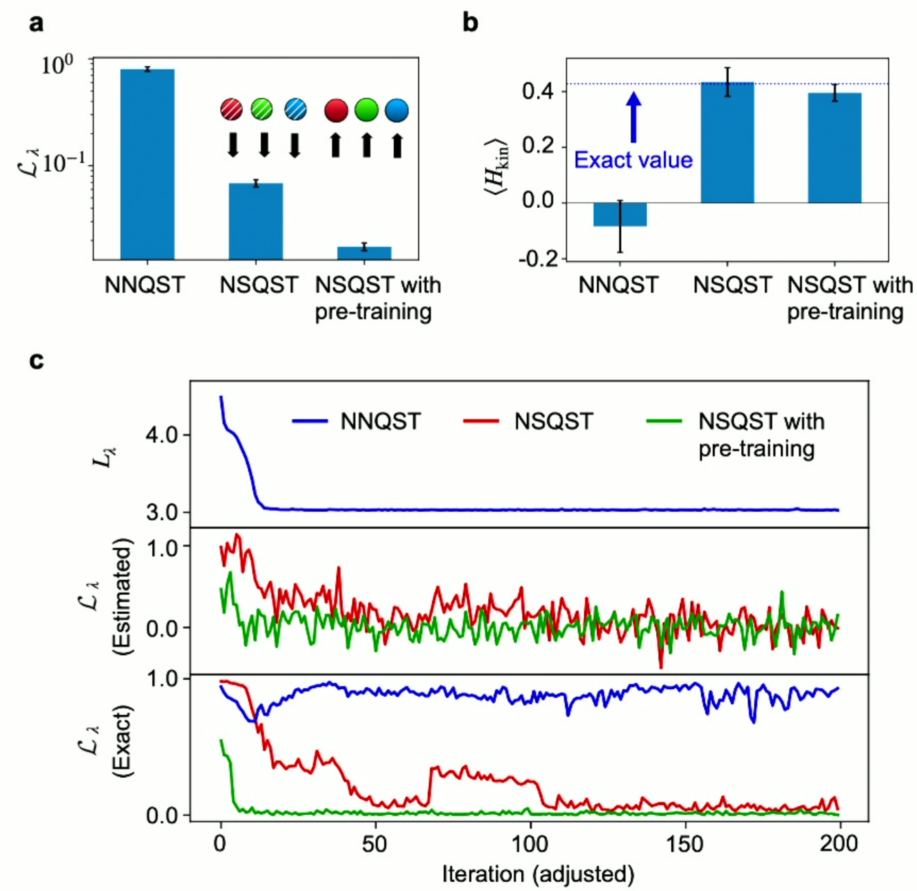
Advantages over NNQST for time-evolved states

- To find interesting examples beyond ground states, we chose a 6-qubit time-evolved state (after Trotterized time evolution).

- NNQST is trained on 21 nearly-diagonal measurement bases (512 shots per base).
- NSQST is trained on using 100 Clifford shadows per iteration.
- NSQST with pre-training, the pre-training stage uses NNQST's measurement resources, then trained using 100 Clifford shadows per iteration.

$$H_{kin} = -\frac{1}{2}(\sigma_1^+ \sigma_2^z \sigma_3^z \sigma_4^- - \sigma_2^+ \sigma_3^z \sigma_4^z \sigma_5^- + \sigma_3^+ \sigma_4^z \sigma_5^z \sigma_6^- + \mathrm{H.c.})$$

Atas et al., Phys. Rev. Research **5**, 033184 (2023).



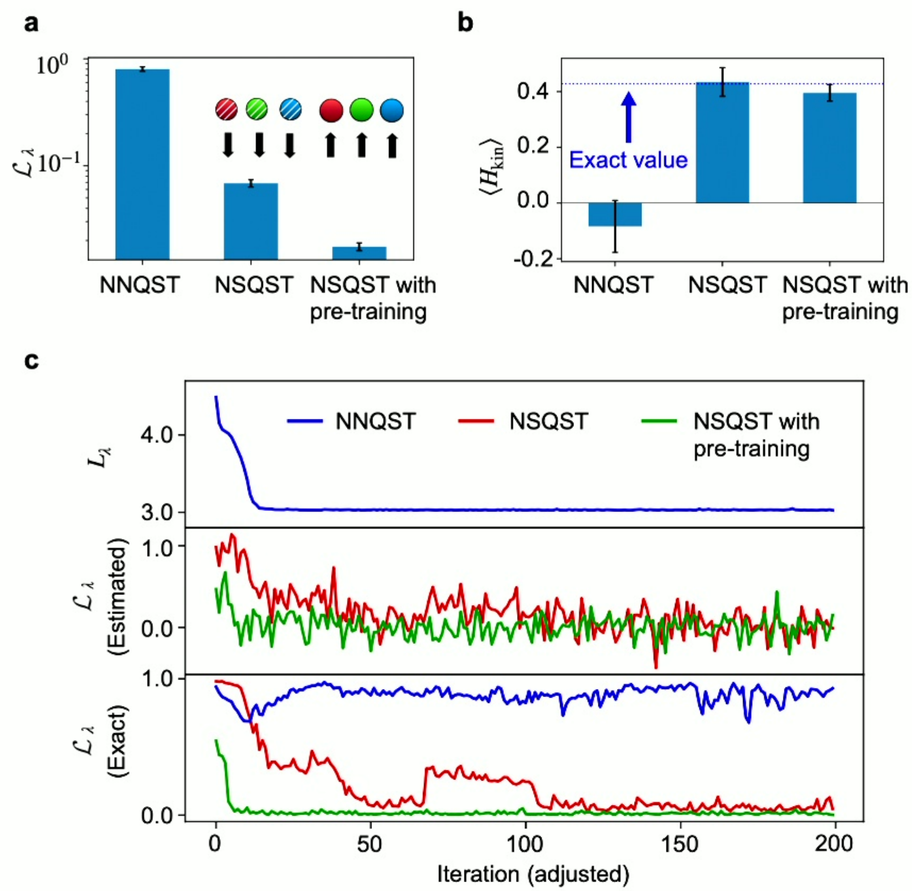
Advantages over NNQST for time-evolved states

- To find interesting examples beyond ground states, we chose a 6-qubit time-evolved state (after Trotterized time evolution).

- NNQST is trained on 21 nearly-diagonal measurement bases (512 shots per base).
- NSQST is trained on using 100 Clifford shadows per iteration.
- NSQST with pre-training, the pre-training stage uses NNQST's measurement resources, then trained using 100 Clifford shadows per iteration.

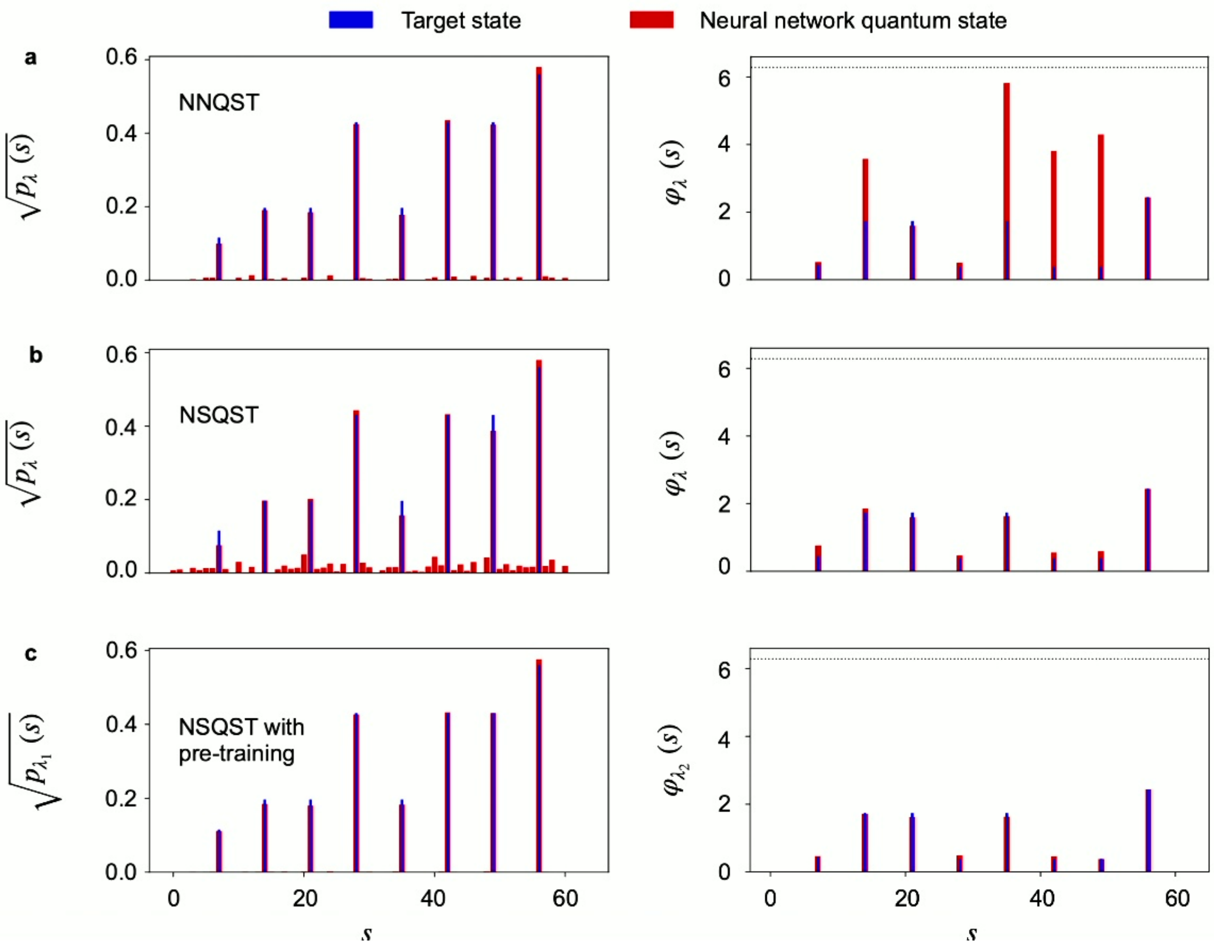
$$H_{kin} = -\frac{1}{2}(\sigma_1^+\sigma_2^z\sigma_3^z\sigma_4^- - \sigma_2^+\sigma_3^z\sigma_4^z\sigma_5^- + \sigma_3^+\sigma_4^z\sigma_5^z\sigma_6^- + \mathrm{H.c.})$$

Atas et al., Phys. Rev. Research **5**, 033184 (2023).



Why did NNQST fail?

- Relative phase are very important for predicting **non-diagonal** observables!
- The kinetic Hamiltonian is clearly non-diagonal, so incorrect relative phases will lead to incorrect observable prediction.



Advantages over direct shadow estimation (*new)

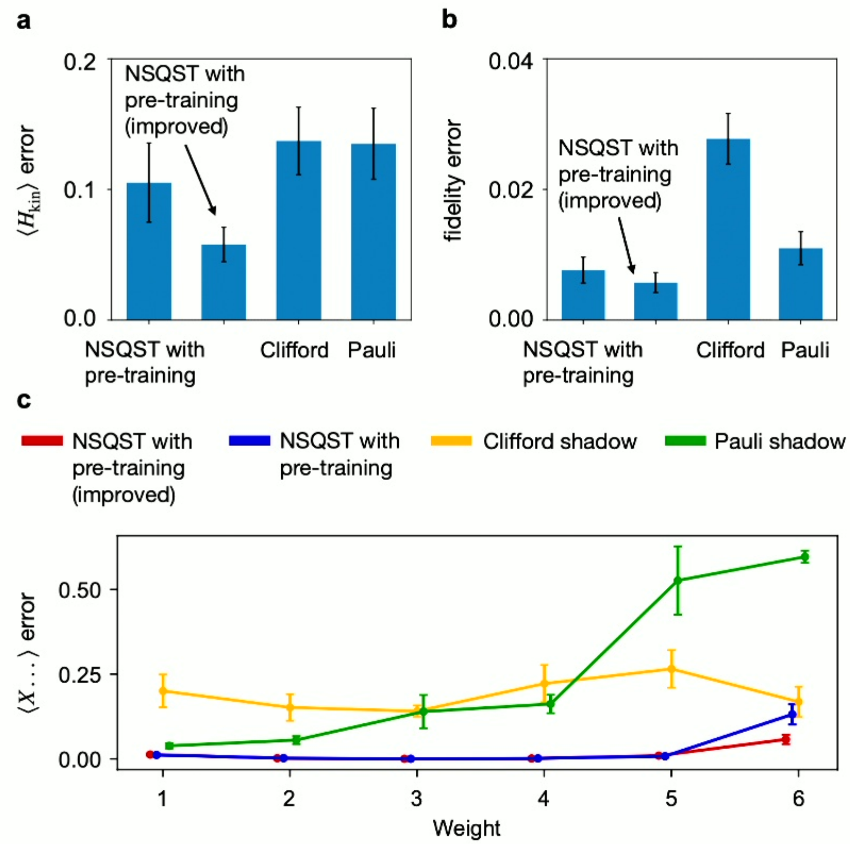
- Does NSQST offer any advantages over direct shadow estimation? In other words, does a model-based approach offer any advantages over using data alone?

- Starting with small system size (6 qubits), we make two changes to reduce the number of measurements for NSQST with pre-training:

  1) re-use the Clifford shadows instead of re-measure in every iteration.

  2) Although still intractable for general states, we reduce the classical post-processing error in the loss function.

- 1000 direct measurements + 200 Clifford shadows
- 1200 Clifford/Pauli shadows

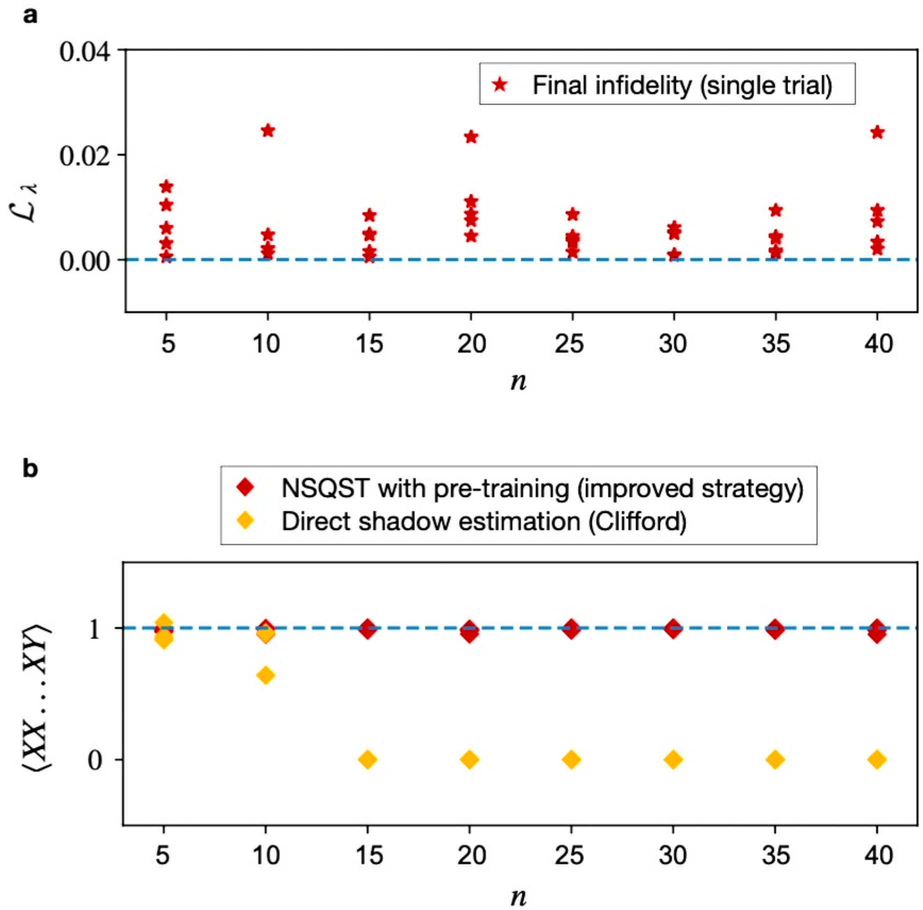


Scalable advantage?

- For a phase-shifted (by pi/2) GHZ state, we fix the number of measurements and check the final infidelity.

- We also predict one of the target state's stabilizers, as system size grows.

- 3000 direct measurements + 200 Clifford shadows
- 3200 Clifford shadows



Noise robustness

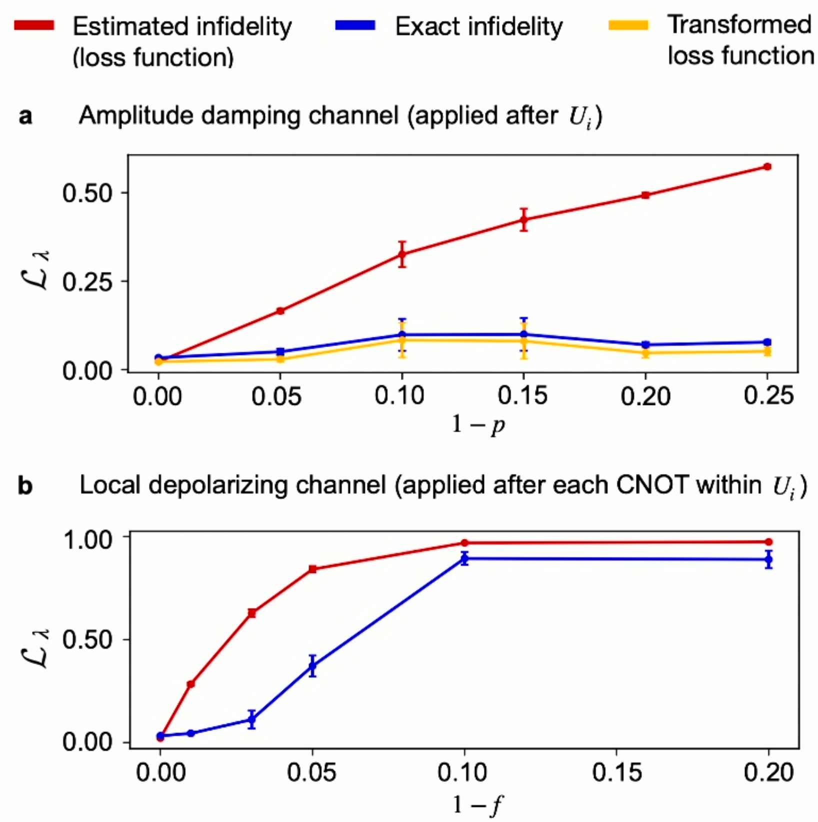
Going back to NSQST's loss function, if there is a noise channel **after** the Clifford unitary, we can simply modify the shadow expression to account for it. A stronger noise channel leads to larger variance in observable estimate.

The loss function is biased, unless we can estimate the strength of the noise channel with a calibration step.

But! The gradient remains unbiased, even if we know nothing about the noise channel!

Koh and Grewal. *Quantum* 6: 776 (2022).

$$\begin{aligned}\mathcal{L}_\lambda(\mathcal{E}) &:= 1 - |\langle\psi_\lambda|\Phi\rangle|^2 \\ &\approx 1 - \frac{1}{N}\sum_{i=1}^{N}\mathrm{Tr}(O_\lambda\hat{\rho}_i) \\ &= 1 - \frac{1}{N}\sum_{i=1}^{N}\langle\psi_\lambda|\,\hat{\rho}_i(\mathcal{E},U_i,b_i)\,|\psi_\lambda\rangle \\ &= 1 - \frac{1}{2^n}\left(1 - \frac{1}{f(\mathcal{E})}\right) - \frac{1}{Nf(\mathcal{E})}\sum_{i=1}^{N}|\langle\phi_i|\psi_\lambda\rangle|^2\end{aligned}$$



Future Prospects

1. Incorporate prior knowledge of the target state into the ansatz, such as symmetries.

Future Prospects

1. Incorporate prior knowledge of the target state into the ansatz, such as symmetries.
2. Reduce classical post-processing problem with more constrained ansatz.
3. Explore new hybrid training strategies with model-based approach.
4. Use other types of randomized measurements.

Future Prospects

1. Incorporate prior knowledge of the target state into the ansatz, such as symmetries.

2. Reduce classical post-processing problem with more constrained ansatz.

3. Explore new hybrid training strategies with model-based approach.

4. Use other types of randomized measurements.

5. Build a practical tool for experimentalists.

6. More and more…

Experimental feasibility: hardware-efficient shadows

So far, only Pauli shadows have been experimentally demonstrated, as no entangling gates are required.

A recent series of work explores the intermediate regime between Pauli and Clifford shadows.

Matrix product state methods are used to ensure efficient classical post-processing.

Hu et al. *Phys. Rev. Research* 5, 023027 (2023).

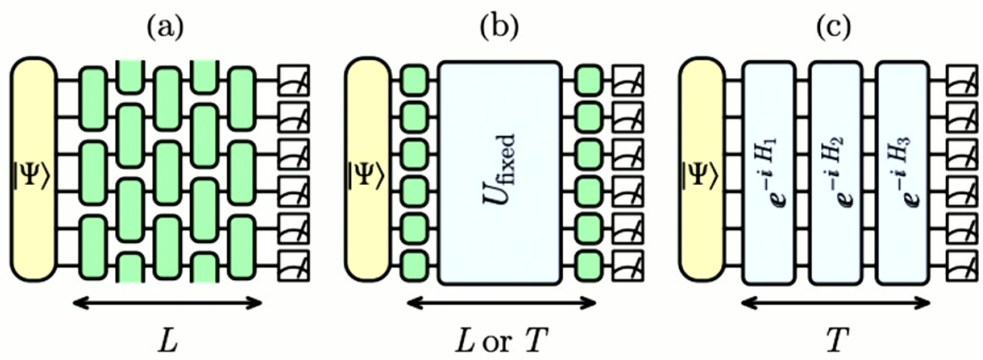
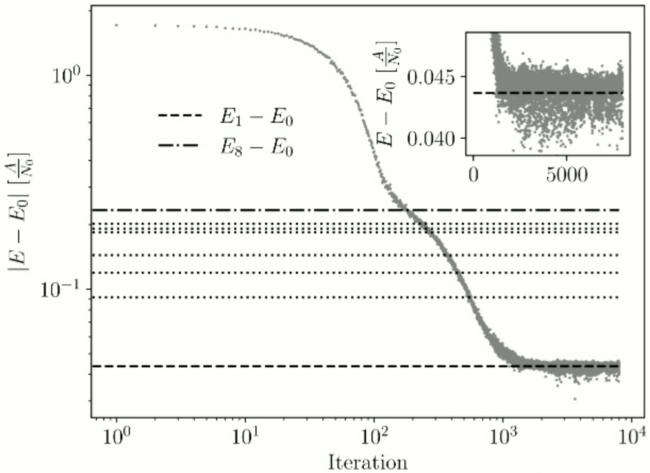
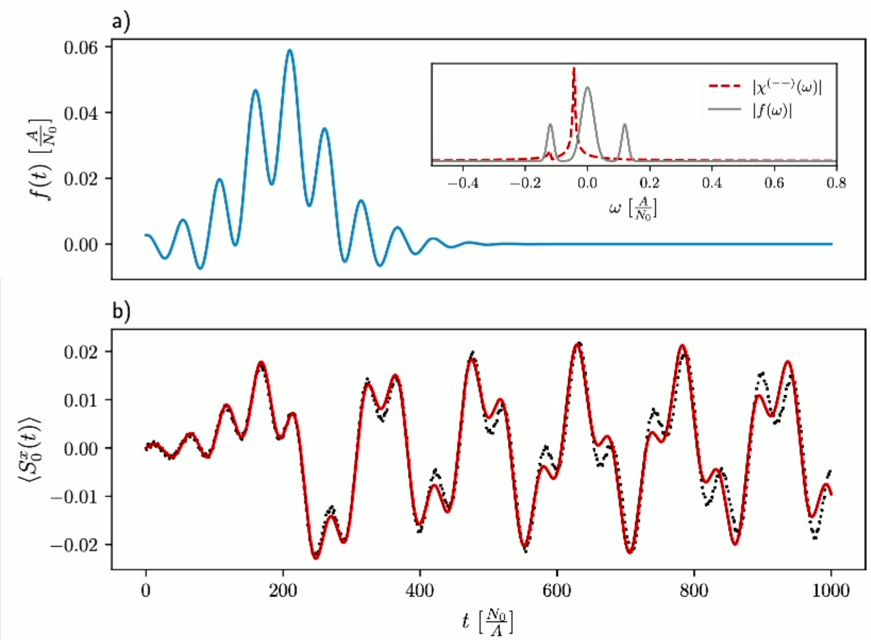


FIG. 3. Classical shadow tomography with (a) finite-depth random unitary/Clifford circuits (of L layers), (b) a fixed unitary twirled by single qubit random Clifford gates, and (c) discrete-time Hamiltonian dynamics (of T steps).

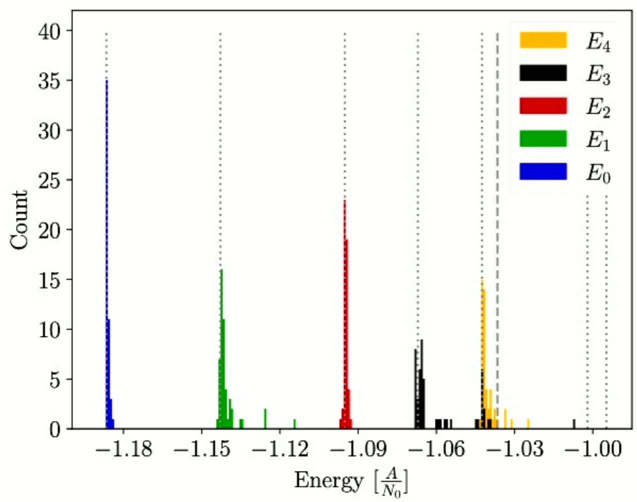
Ads: exploring excited states and low-lying dynamics



Wei et al. *Advanced Phys. Research* 2300078, (2023).



Code available and compatible with Netket utilities.



Acknowledgement

Prof. Bill Coish

Prof. Pooya Ronagh

Prof. Christine Muschik

IQC Institute for Quantum Computing

PI PERIMETER INSTITUTE

McGill UNIVERSITY

Qi Quantum Interactions

UNIVERSITY OF WATERLOO CONCORDIA CUM VERITATE

Conferences

Machine Learning for Quantum Many-Body Systems, Perimeter Institute, Jun. 12th to Jun. 16th, 2023.

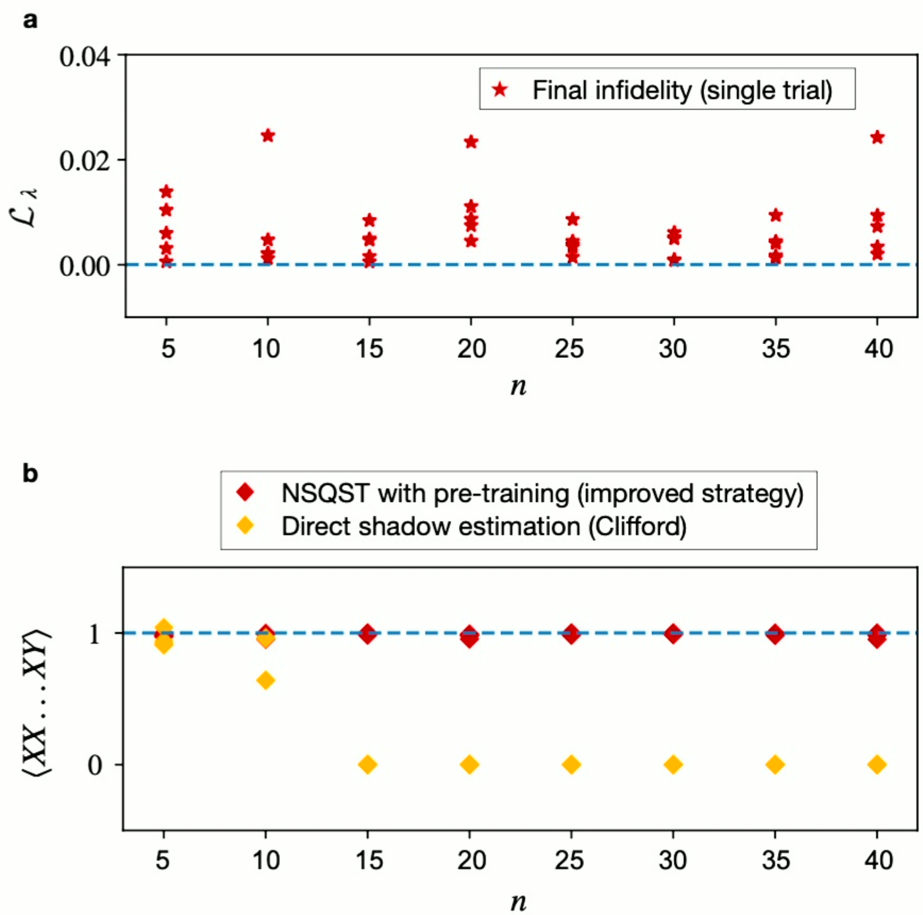
Coherent Quantum Dynamics, OIST, Sep. 26th to Oct. 5th, 2023.

Scalable advantage?

- For a phase-shifted (by pi/2) GHZ state, we fix the number of measurements and check the final infidelity.

- We also predict one of the target state's stabilizers, as system size grows.

- 3000 direct measurements + 200 Clifford shadows
- 3200 Clifford shadows



Future Prospects

1. Incorporate prior knowledge of the target state into the ansatz, such as symmetries.

2. Reduce classical post-processing problem with more constrained ansatz.

Scalable advantage?

- For a phase-shifted (by pi/2) GHZ state, we fix the number of measurements and check the final infidelity.

- We also predict one of the target state's stabilizers, as system size grows.

- 3000 direct measurements + 200 Clifford shadows
- 3200 Clifford shadows