Neural quantum state tomography, improvements and applications

Toward digital twins for quantum states

References:

- Bennewitz et al., Neural error mitigation of near-term quantum simulations. *Nature Machine Intelligence* 4.7 (2022): 618-624.
- Wei et al., Neural-Shadow Quantum State Tomography." *preprint arXiv:2305.01078* (2023).

Pooya Ronagh

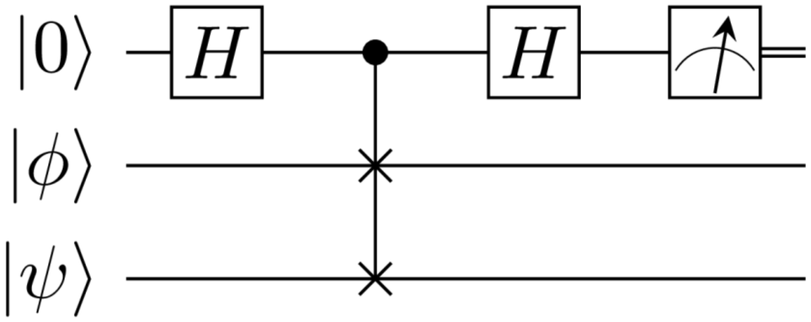
Research Assistant Professor, IQC, University of Waterloo, Perimeter Institute | CTO, 1QBit

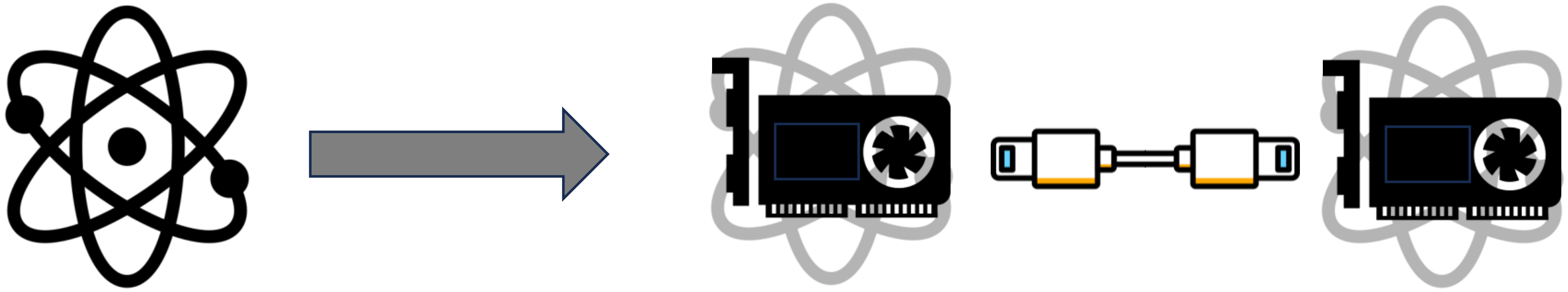
Motivation 1–Big quantum data!

- Quantum experiments are creating increasing amount of experimental data.
- The amount of classical memory required to express quantum states grow exponentially.
- We’ll have a huge amount of data to post-process, analyze, and collect statistics from.
- BUT! Quantum states are difficult to express classically?
- Goal: Bend the curve of exponential classical memory required for expressing quantum states.
- ML has been very successful in doing the same for classical big data: turning big data into AI.
- Let’s do the same to quantum data to achieve **operational access to quantum data** instead of storing exponentially large tables and sweeping over them in every query.
- Applications:
  - Cleaning up the state! Imposing purity, symmetries, etc. of the target state.
  - Manipulating the state: decreasing its energy “further” variationally.
  - Observable estimation at the cost of classical inference from a model, rather than sweeping over exponentially large raw data.

Motivation 2–Classical cloning is cheap!

- Unlike quantum states, classical memory is easy to replicate.
- Calculating overlaps of digital twins is much easier than performing swap operations:
  - Fidelity estimation,
  - Entanglement entropy estimation.
- Applications:
  - Verification of quantum computation,
  - Cross platform benchmarks.





Motivation 3–Quantum computing is expensive!

- Quantum experiments are expensive to do, repeat, and make widely accessible.
- Fault-tolerant quantum computers will be large sophisticated facilities.
- At 1 USD / second / 1000 qubits, Shor's factorization will cost +500M.
- Need for efficient and standardized ways to make the results of experiments available to the community.

CERN data centre, wikimedia.org

IBM Launch IBM Quantum

Pay-As-You-Go Plan

via IBM Cloud

$1.60 USD / second

arXiv > quant-ph > arXiv:1905.09749

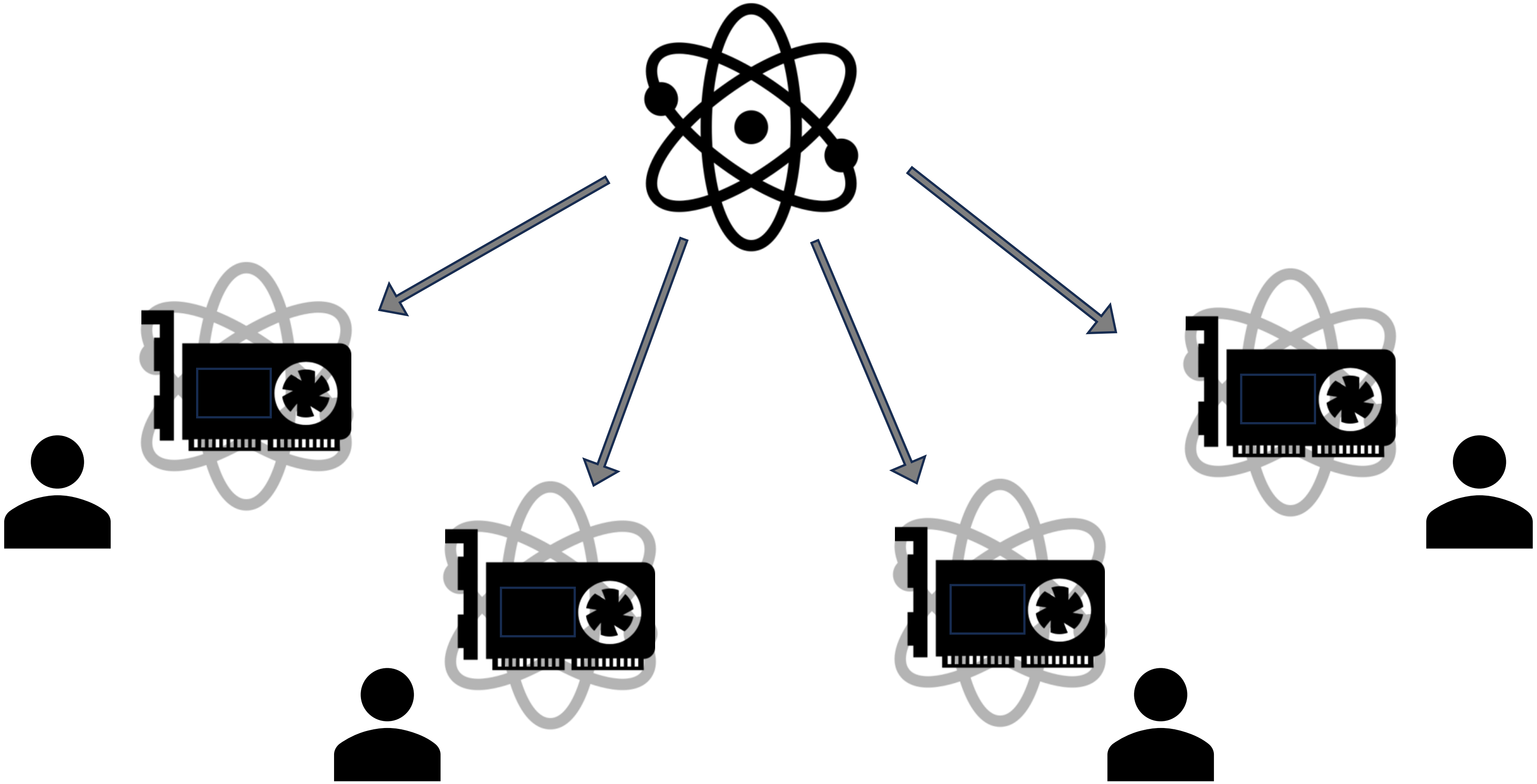
Quantum Physics

[Submitted on 23 May 2019 (v1), last revised 13 Apr 2021 (this version, v3)]

How to factor 2048 bit RSA integers in 8 hours using 20 million noisy qubits

Craig Gidney, Martin Ekerå

Motivation 3–Quantum computing is expensive!



Neural-network quantum state tomography

- Neural quantum state tomography aims at reconstructing a quantum state using a generative model.

Letter | Published: 26 February 2018

Neural-network quantum state tomography

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

Nature Physics **14**, 447–450 (2018) | Cite this article

30k Accesses | **435** Citations | **155** Altmetric | Metrics

- As a standalone algorithm, it is compared to Monte-Carlo algorithms for ground state preparation.
- As a quantum state tomography method, it is studied from the tomographic/information theoretic aspect (i.e., how well is the reconstruction, at what cost).
- Our goal: Turn NNQST into the quantum information scientist's daily R&D tools.
- This talk:
  - (a) Applications in ground state preparation,
  - (b) Improvements using the classical shadow formalism.

Machine learning introduction

- Generative models historically have emerged in ML from image processing tasks.
- A different collection of deep generative models have been developed motivated by natural language processing (NLP) tasks.
    - Auto-regressive models: $p(\sigma) = \prod_i p(\sigma_i|\sigma_1, \ldots, \sigma_{i-1})$,
    - Two common architectures in NLP:
        - Recurrent neural networks
        - Transformers
- RNNs:
    - Precursor to the more powerful SOTA transformers
    - Encoder-decoder mechanism
    - Sequence to sequence architecture

Article | Published: 11 March 2019

Reconstructing quantum states with generative models

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

Nature Machine Intelligence **1**, 155–161(2019) | Cite this article

2225 Accesses | **55** Citations | **33** Altmetric | Metrics

Sequence to Sequence Learning with Neural Networks

Ilya Sutskever Google ilyasu@google.com

Oriol Vinyals Google vinyals@google.com

Quoc V. Le Google qvl@google.com

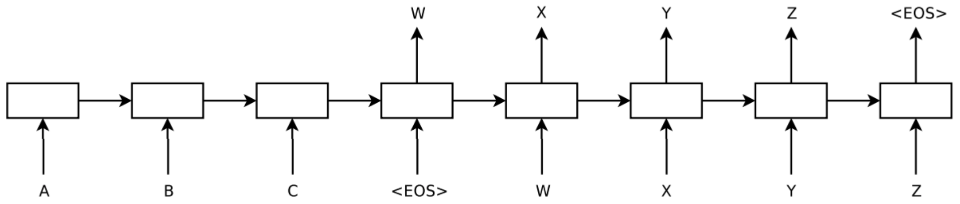


Figure 1: Our model reads an input sentence “ABC” and produces “WXYZ” as the output sentence. The model stops making predictions after outputting the end-of-sentence token. Note that the LSTM reads the input sentence in reverse, because doing so introduces many short term dependencies in the data that make the optimization problem much easier.

Attention

- As opposed to RNNs, all hidden states are available at the same time.
- Attention was then incorporated in a breakthrough model called the Transformer, which became a critical component of Dall-E, ChatGPT, Bard, etc.
- Perhaps better parametrized models for quantum data should be developed….

Cornell University

Computer Science > Computation and Language

arXiv:1706.03762 (cs)

[Submitted on 12 Jun 2017 (v1), last revised 2 Aug 2023 (this version, v7)]

Attention Is All You Need

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, Illia Polosukhin

arXiv Cornell University

Condensed Matter > Strongly Correlated Electrons

arXiv:1912.11052 (cond-mat)

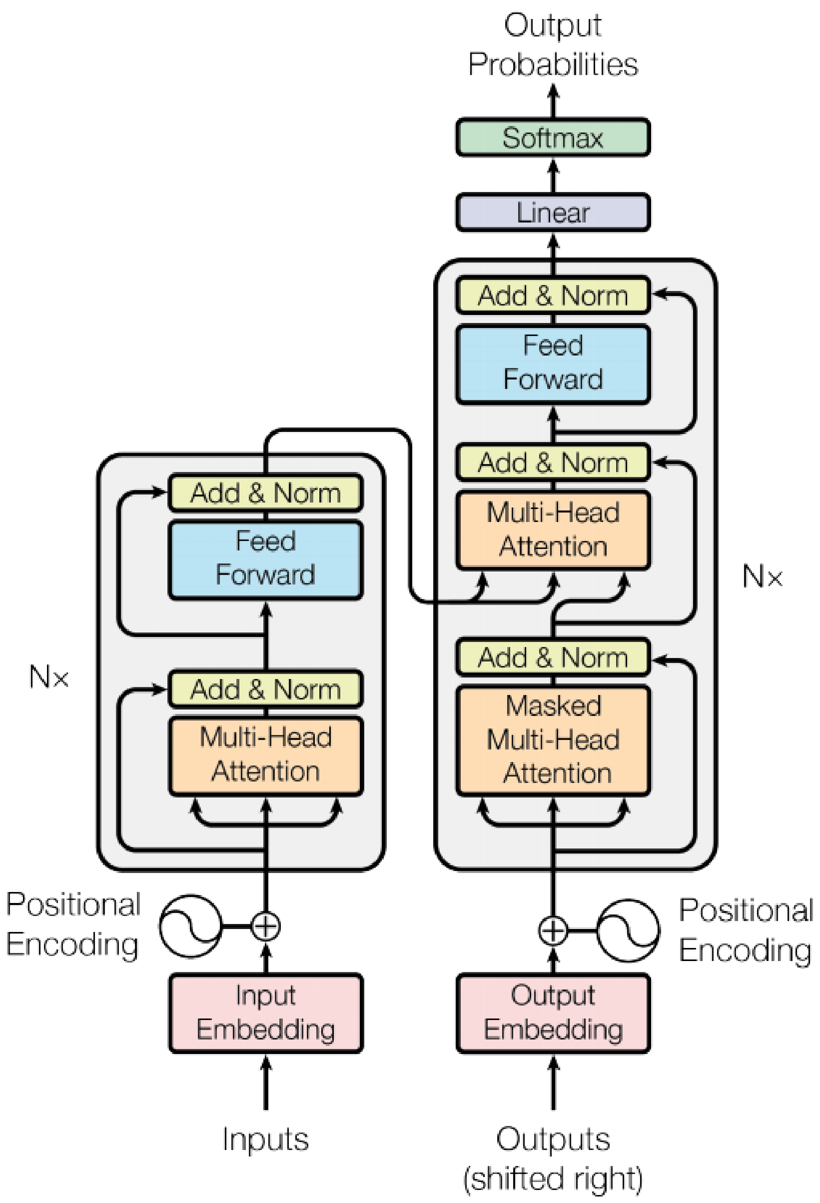
[Submitted on 23 Dec 2019]

Probabilistic Simulation of Quantum Circuits with the Transformer

Juan Carrasquilla, Di Luo, Felipe Pérez, Ashley Milsted, Bryan K. Clark, Maksims Volkovs, Leandro Aolita

Transformers

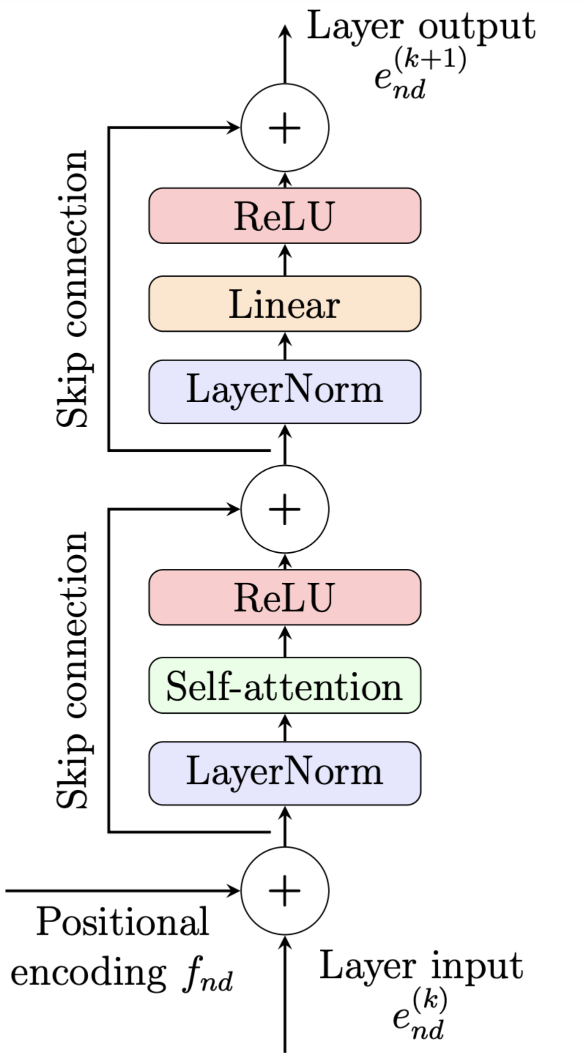
- "A model architecture eschewing recurrence and instead relying entirely on an attention mechanism to draw global dependencies between input and output." [Vaswani '17]
    - Avoid relying on temporal dependence of elements to each other.
    - Decide how important each element is with respect to all the other elements in the original sequence.
- The architecture includes an encoder and a decoder.
- Relies on two mechanism for attention:
    - A self-attention mechanism used in the encoder.
    - A cross-attention mechanism used in the decoder.
- The positional encoding can be used to keep track of the order in a sequence if needed (e.g., in machine translation).



https://arxiv.org/pdf/1706.03762.pdf

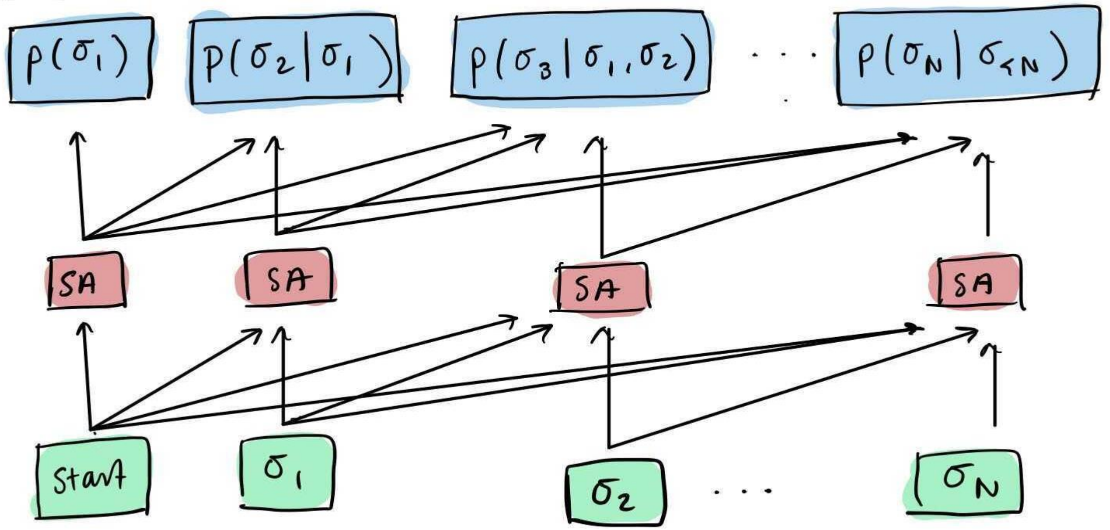
Queries, keys, and values

- Three types of vectors as learnable parameters:
  - Queries: $q = W_q\, x$
  - Keys: $k = W_k\, x$
  - Values: $v = W_v\, x$
- Keys and queries are of same dimension, but values may be of arbitrary dimension. We ignore this detail and simply write $Q, K, V \in \mathbb{R}^{d\times t}$.
- The attention weights are then generated via
$$a = [soft](arg)\max_{\beta}(K^T Q)\,.$$
- The output state is then generated using $\mathrm{A} = (a)$ and V:
$$H = VA \in \mathbb{R}^{d\times t}.$$



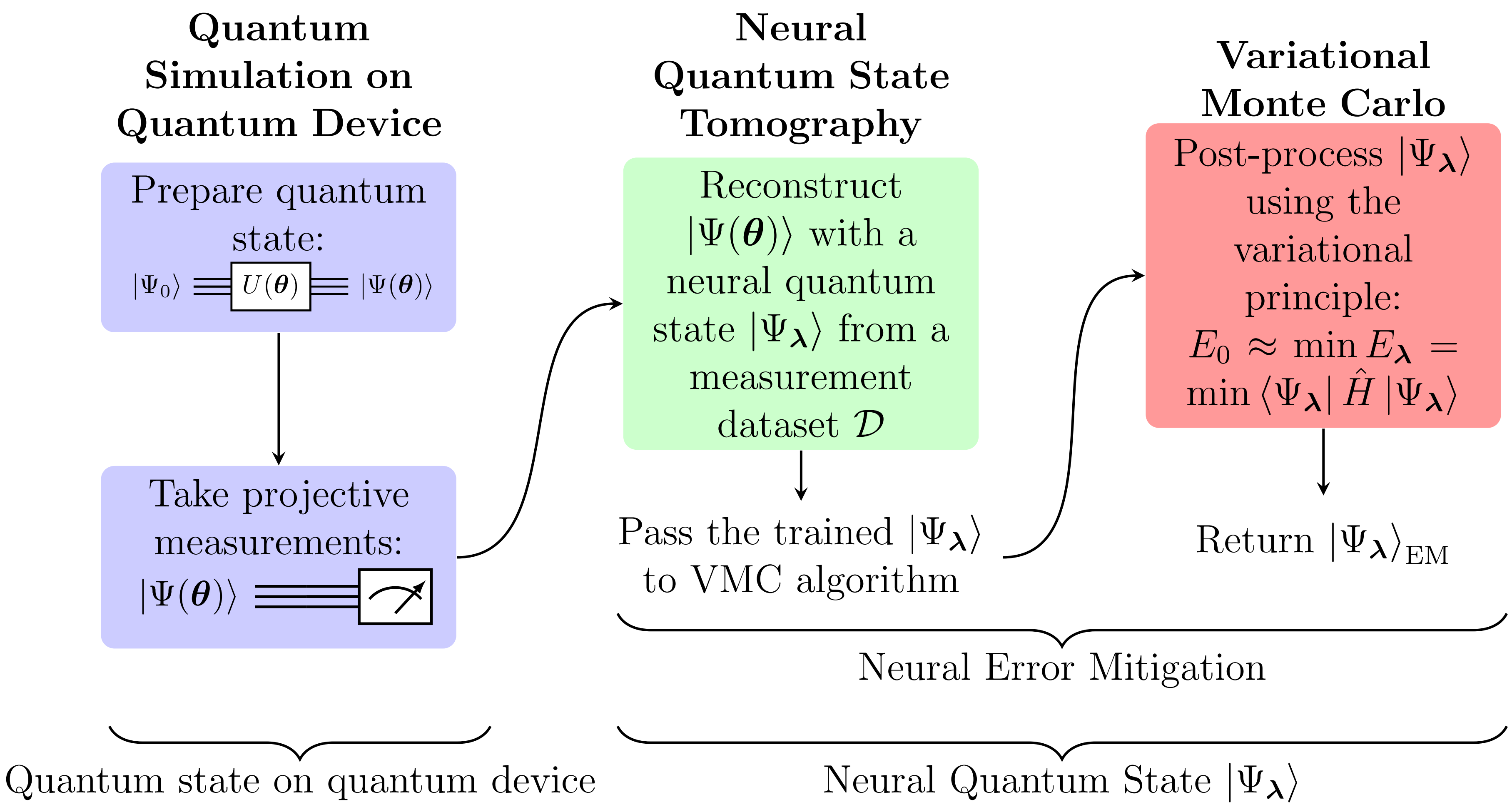
Neural error mitigation

- Error mitigation: post-processing step to alleviate errors affecting the output of a noisy quantum device.
- Many different creative ways to approach error mitigation
  - Average the results of circuits from a quasi-probability distribution (Temme et al., 2017)
  - Learn a scalable noise model by comparing noisy and noise-free circuits (Czarnik et al. 2020)
- Why stop there? We can clean up neural quantum states in other ways too (e.g., re-enforcing symmetries).
- Advantages:
  - Generally, does not require significant additional quantum resources;
  - Relevant for current and near-term quantum processors.



Schematics by E Bennewitz

Neural error mitigation



Neural error mitigation

- How do we represent $\phi_\lambda(\boldsymbol{\sigma})$ and $|\Psi_\lambda(\boldsymbol{\sigma})| = \sqrt{p_\lambda(\boldsymbol{\sigma})}$?

Exact	Approximate
$\|\Psi\rangle = \sum_\sigma e^{i\,\phi(\boldsymbol{\sigma})}\|\Psi(\boldsymbol{\sigma})\|\|\boldsymbol{\sigma}\rangle$	$\|\Psi_\lambda\rangle = \sum_\sigma e^{i\,\phi_\lambda(\boldsymbol{\sigma})}\|\Psi_\lambda(\boldsymbol{\sigma})\|\|\boldsymbol{\sigma}\rangle$

- Represent the probability amplitudes via the auto-regressive expansion $p_\lambda(\sigma) = \prod_{i=0}^{n} p(\sigma_i|\sigma_{<i})$ and sample from $p_\lambda(\sigma)$.
- Interpret the complex output of a Transformer as:

$$\ln(\Psi_\lambda(\boldsymbol{\sigma})) = i\,\phi_\lambda(\sigma) + \ln(|\Psi_\lambda(\sigma)|)\ ,$$

  - Real part: **log probability** $\frac{1}{2}\ln(p_\lambda(\boldsymbol{\sigma}))$, and
  - Imaginary part: **phase** $\phi_\lambda(\boldsymbol{\sigma})$**.**
- Optimize λ according to some cost function.
- To compute observables of interest $\langle O\rangle = \sum_\sigma p_\lambda(\sigma) O_{loc}$, where $O_{loc} = \sum_{\sigma'} O_{\sigma\sigma'} \frac{\Psi_\lambda(\sigma')}{\Psi_\lambda(\sigma)}$.

Neural error mitigation

- Step 1 (neural quantum state tomography): Optimize λ with SGD according to cross entropy

$$L_\lambda = -\sum_{\sigma\in\{0,1\}^N} p_{VQE}(\sigma)\ln(p_\lambda(\sigma))$$

for which we estimate p_{VQE} using measurement samples, $D = \{(Z_1, 0), (Z_2, 1), (X_3, 1), \ldots\}$.

$$L_\lambda \approx -\frac{1}{|D|}\sum_{\sigma_{M\in D_M}} \ln\big(p_\lambda(\sigma_M)\big).$$

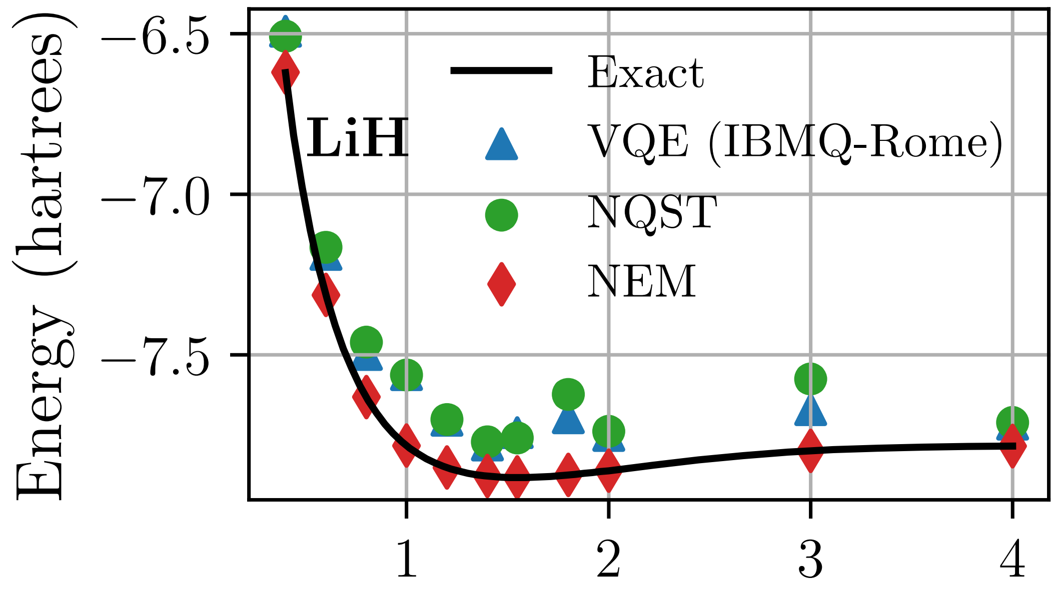
- Step 2 (variational Monte-Carlo): Optimize λ to obtain lower expected energy $\min E_\lambda = \min\langle\Psi_\lambda|\hat{H}|\Psi_\lambda\rangle$ according to

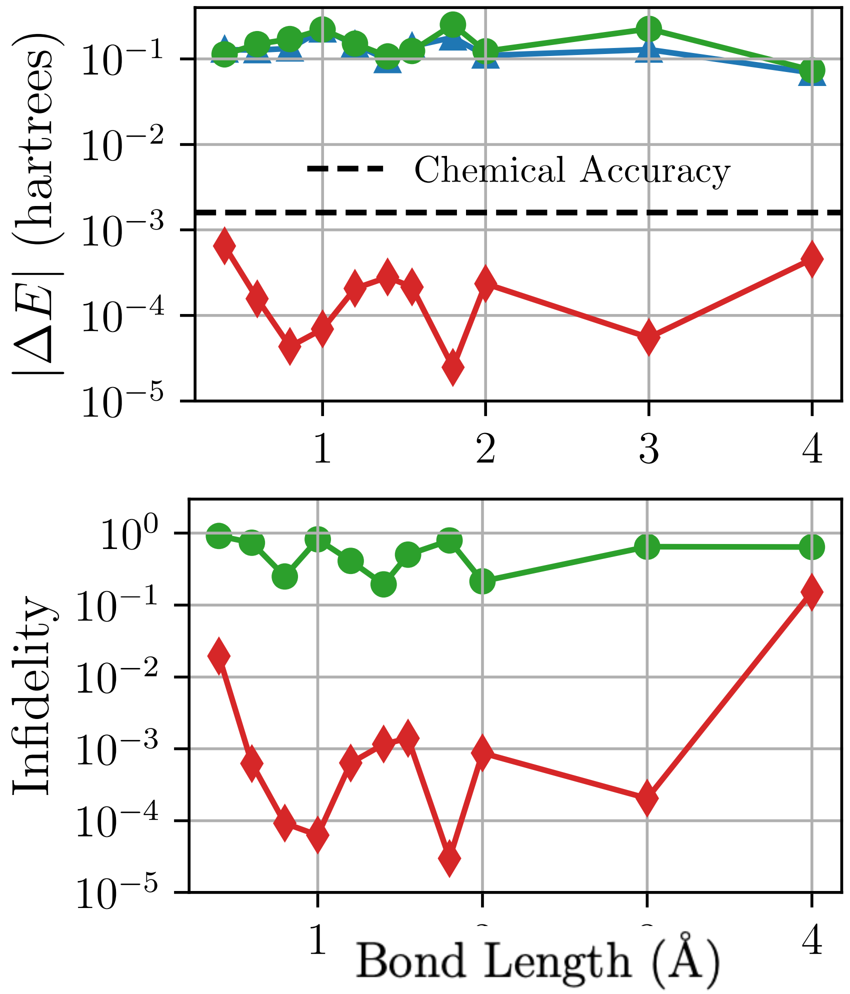
$$E_\lambda = \sum_\sigma p_\lambda(\sigma)\, E_{loc}(\sigma) \approx \frac{1}{N_s}\sum_{i=1}^{N_s} E_{loc}\left(\sigma_s^{(i)}\right) ,$$

where $E_{loc}(\sigma_s) = \sum_{\sigma\prime} H_{\sigma_s\sigma\prime}\, \frac{\Psi_\lambda(\sigma\prime)}{\Psi_\lambda(\sigma_s)}$.

Quantum chemistry (Example 1)

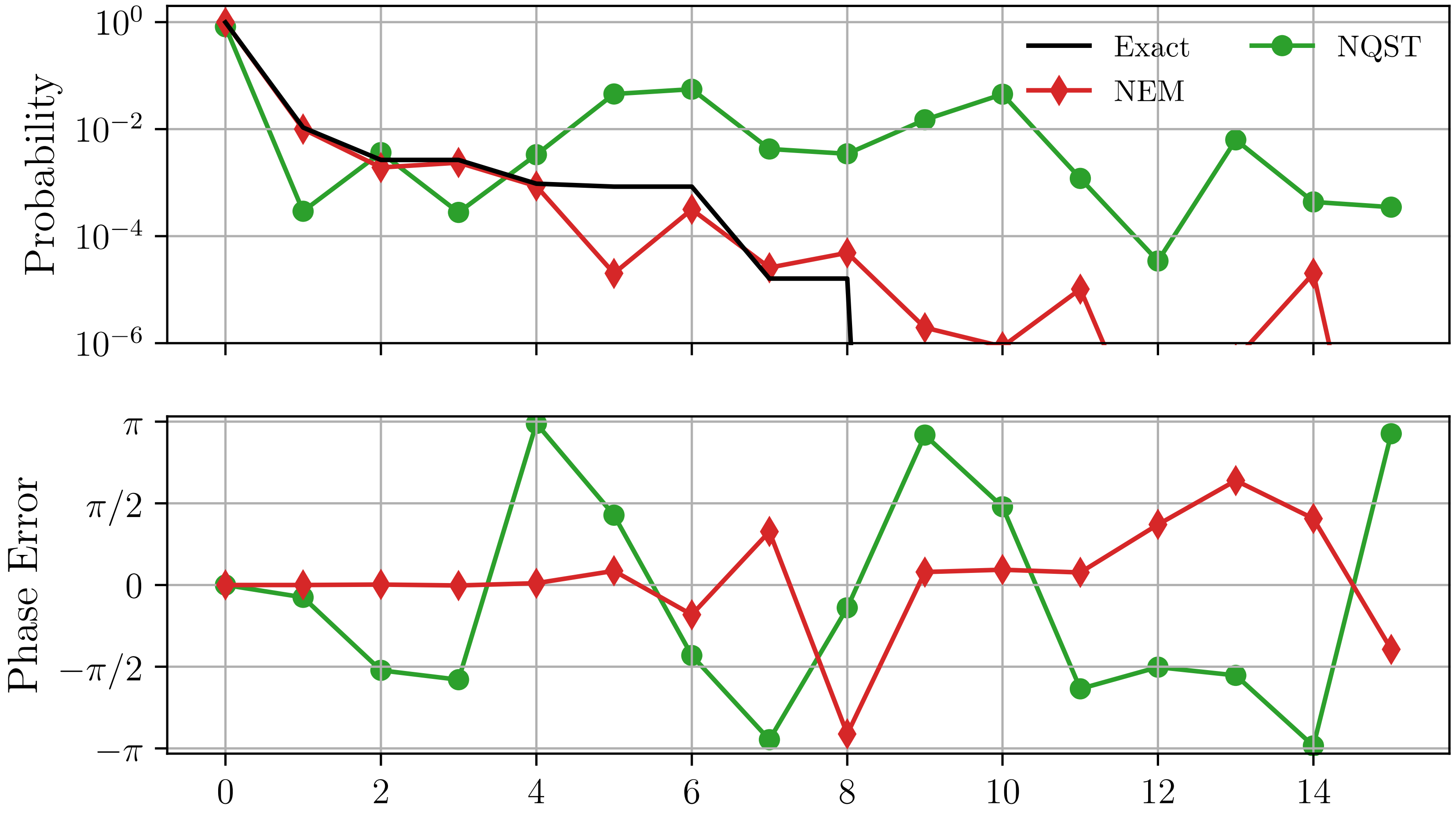
- Electronic structure Hamiltonian of LiH.
- Jordan-Wigner transformation: convert to a qubit-based Hamiltonian.
- VQE ansatz: the “hardware-efficient” ansatz of Kandala, et al. Nature (2017).





Quantum chemistry (Example 1)

- Statistics of each computational basis state at bond length 1.4.
- Trick: maximize the L1 norm $\sum_{\sigma\in\{0,1\}^N} |\Psi_\lambda(\sigma)|$.



Institute for Quantum Computing and Department of Physics & Astronomy, University of Waterloo | Perimeter Institute Quantum Intelligence Lab | 1QBit

Lattice gauge theory (Example 2)

- Simulation of Lattice Schwinger model (an abelian lattice gauge theory, toy model for quantum electrodynamics in 1D) following the ansatz of Kokail, et al. Nature 569.7756 (2019).

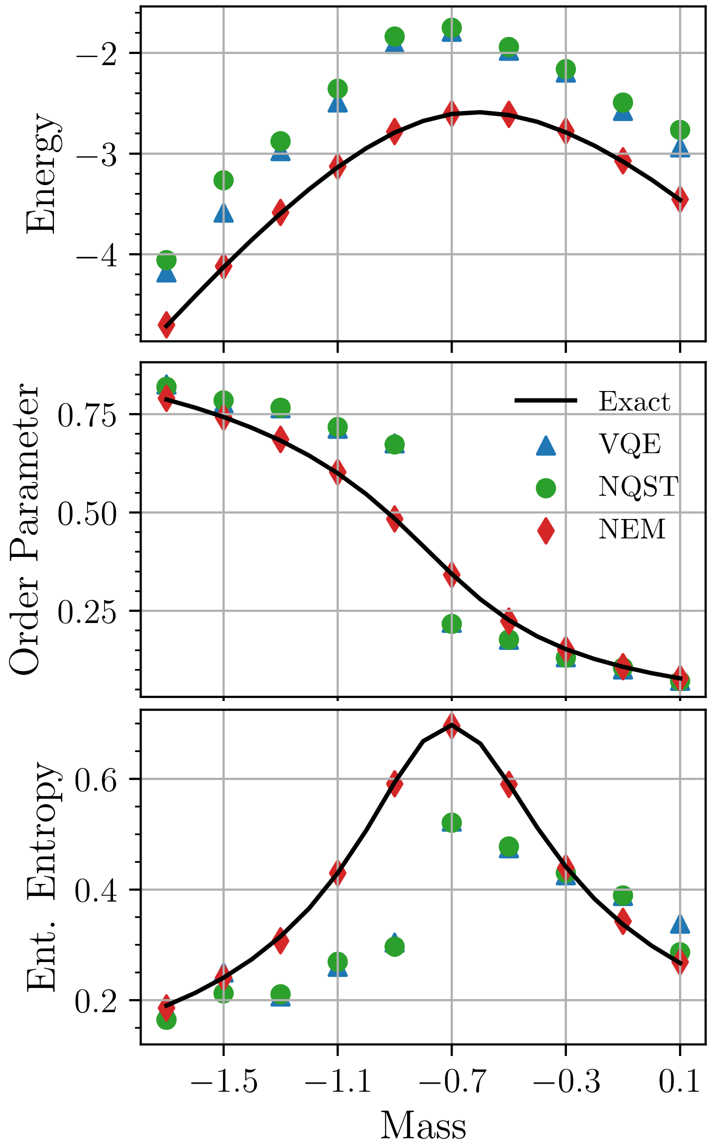
Bare electron mass term

$$\hat{H} = \frac{w}{2}\sum_{j=1}^{N-1}\left(\hat{X}_j\hat{X}_{j+1} + \hat{Y}_j\hat{Y}_{j+1}\right) + \frac{m}{2}\sum_{j=1}^{N}(-1)^j\hat{Z}_j + \bar{g}\sum_{j=1}^{N}\hat{L}_j^2.$$

Creation and annihilation of electron--positron pairs

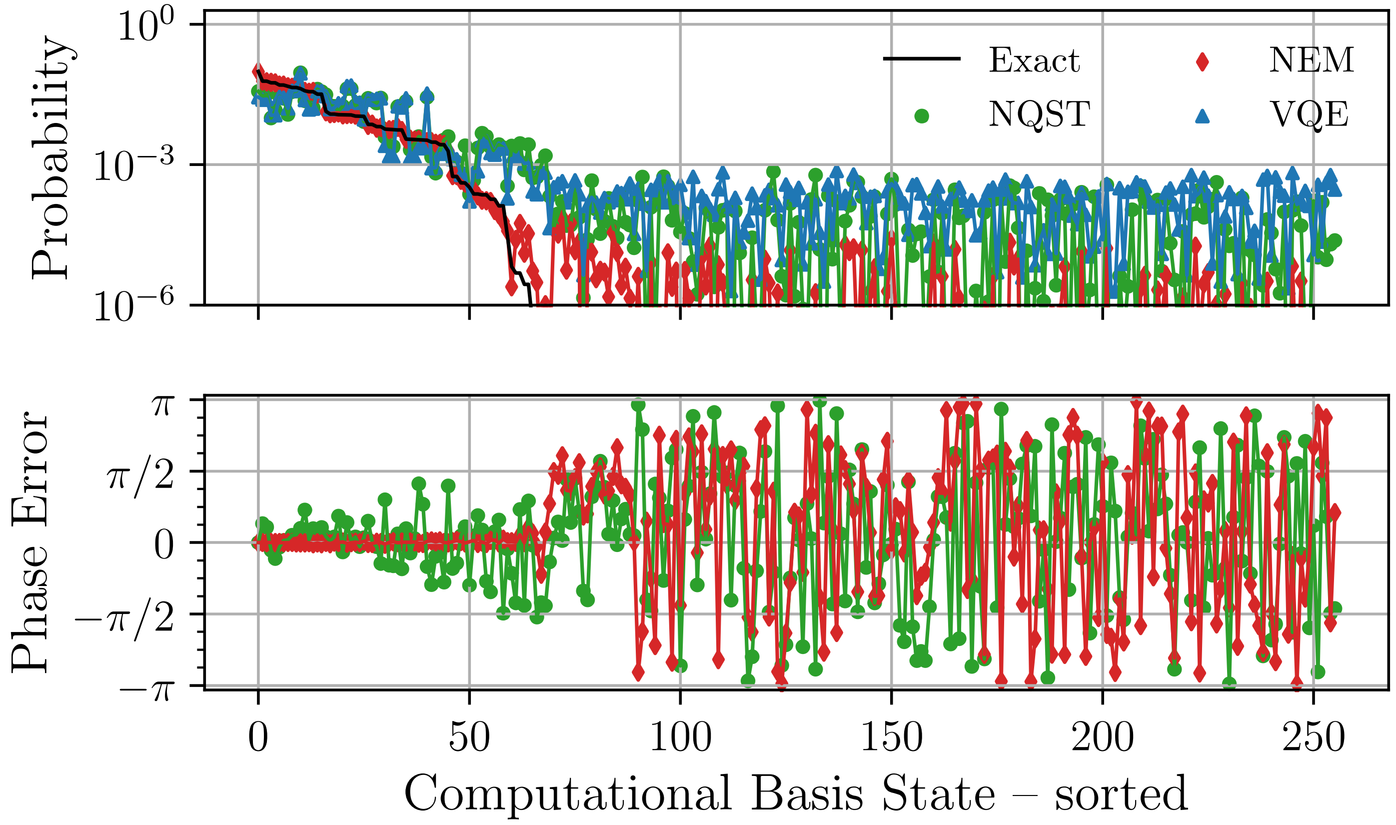
Coupling to electric field

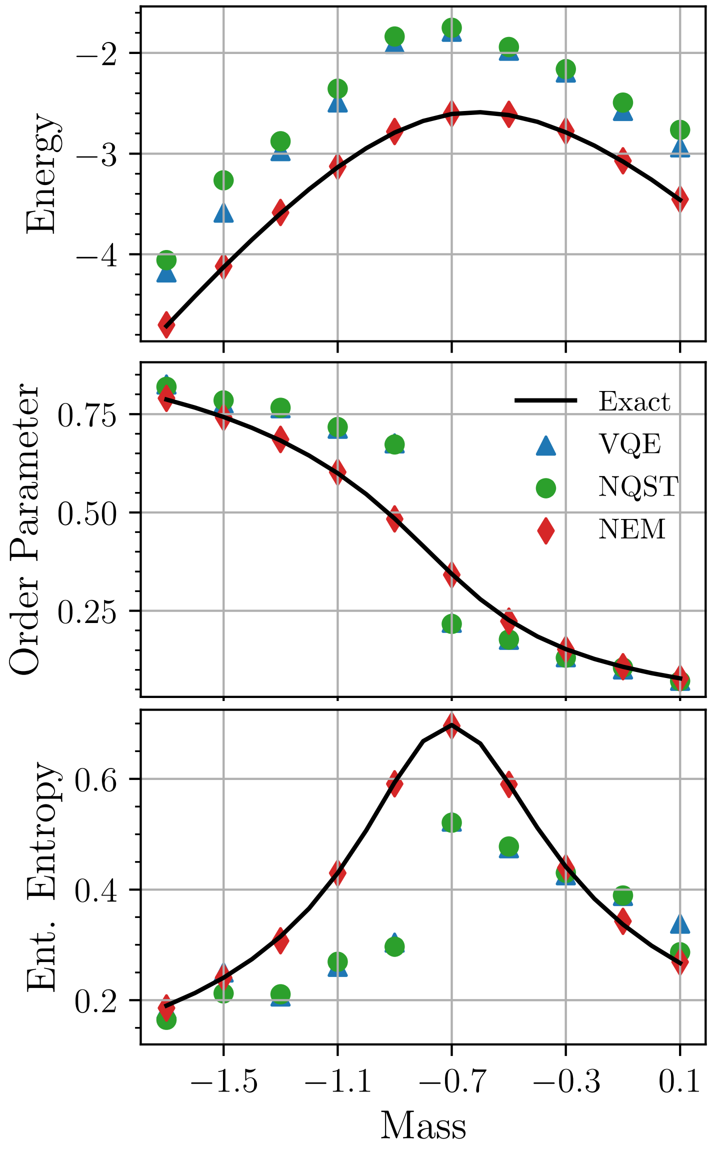
- The goal is to study a phase transition at about m= -0.7.



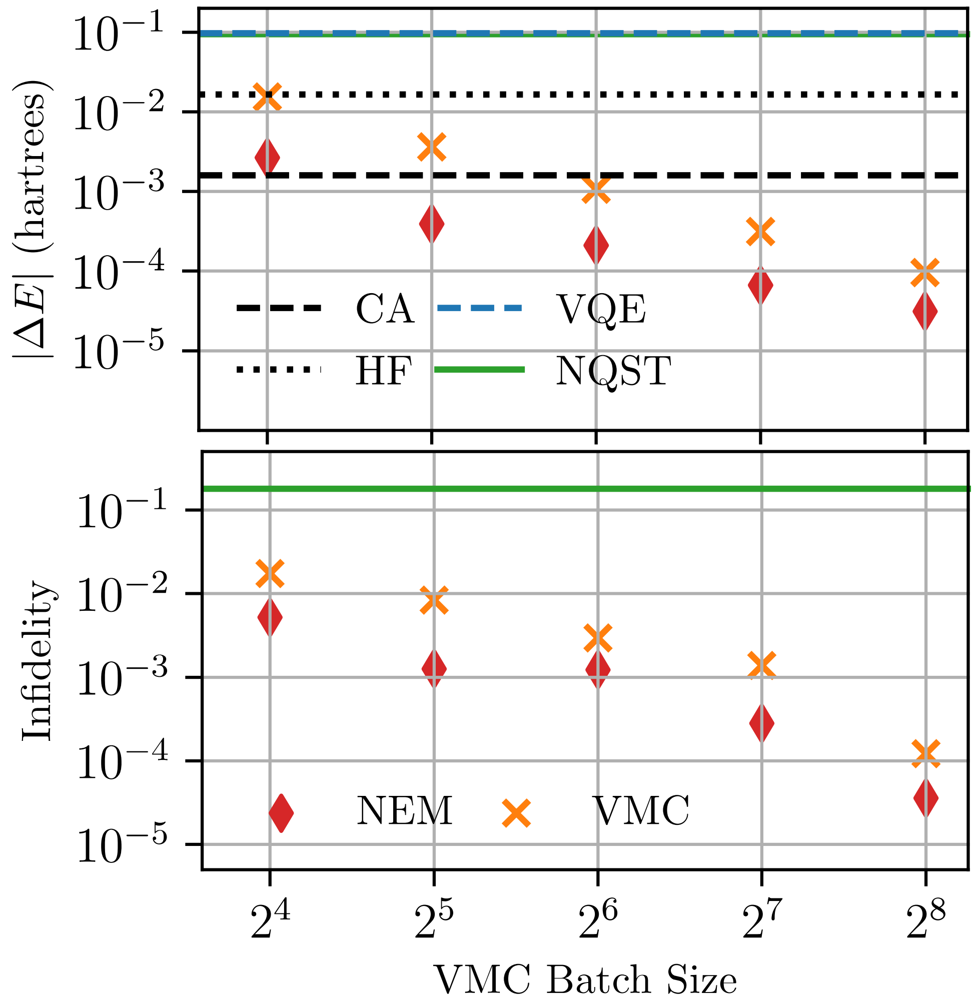
Institute for Quantum Computing and Department of Physics & Astronomy, University of Waterloo | Perimeter Institute Quantum Intelligence Lab | 1QBit

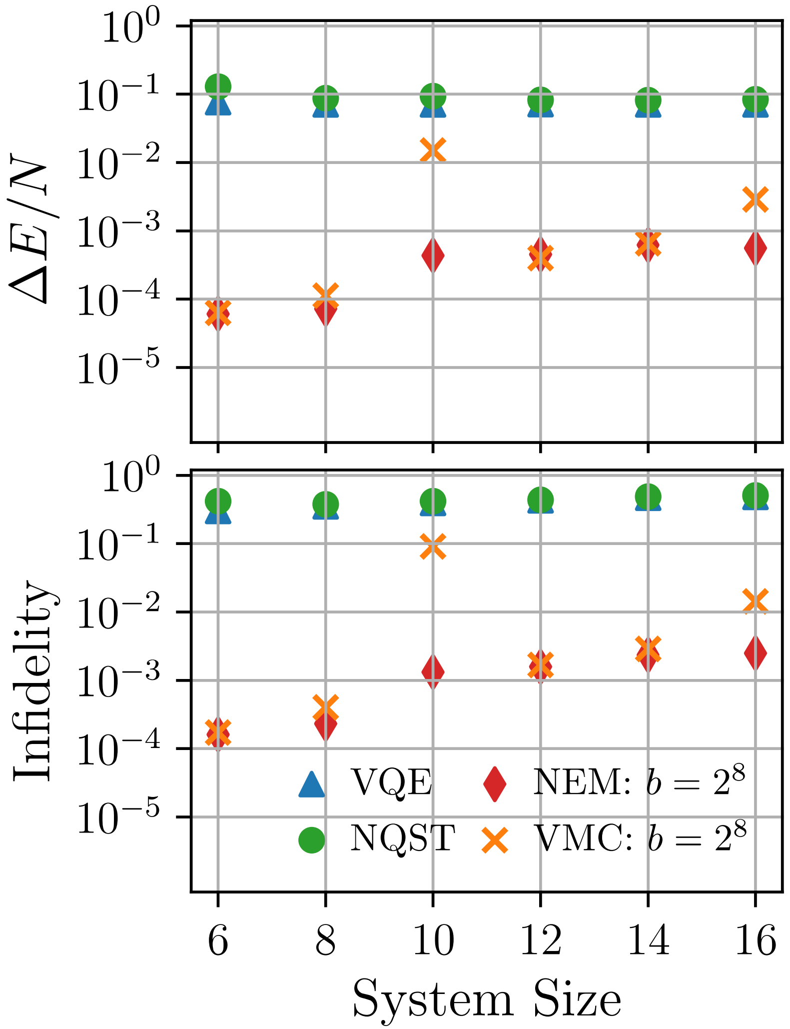
Lattice gauge theory (Example 2)





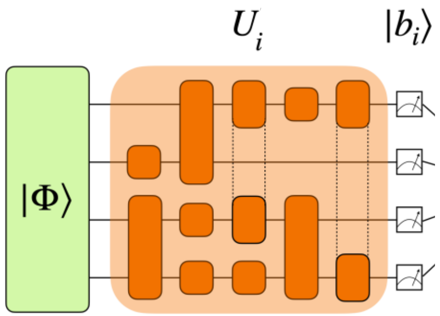
Comparison of NEM and standalone VMC





Learning the phases

- Perform random Clifford tails U_i and measure bit strings $|b_i\rangle$.
- Collect stabilizer states: $|\phi_i\rangle = U_i^\dagger |b_i\rangle$.
- Average effect of the Clifford twirling is a depolarizing noise channel $\mathcal{M}$ with strength $(2^n + 1)^{-1}$.
- Classical shadows: $\rho_i = \mathcal{M}^{-1}(|\phi_i\rangle\langle\phi_i|)$.
- Target state: $|\Phi\rangle\langle\Phi| = \mathbb{E}[\mathcal{M}^{-1}(|\phi_i\rangle\langle\phi_i|)]$.
- New loss:

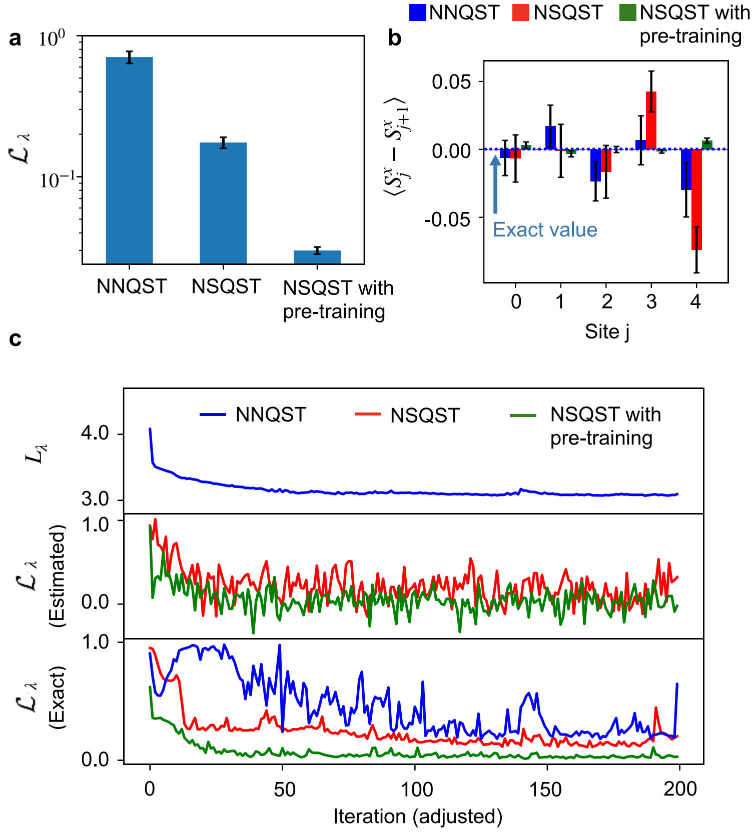


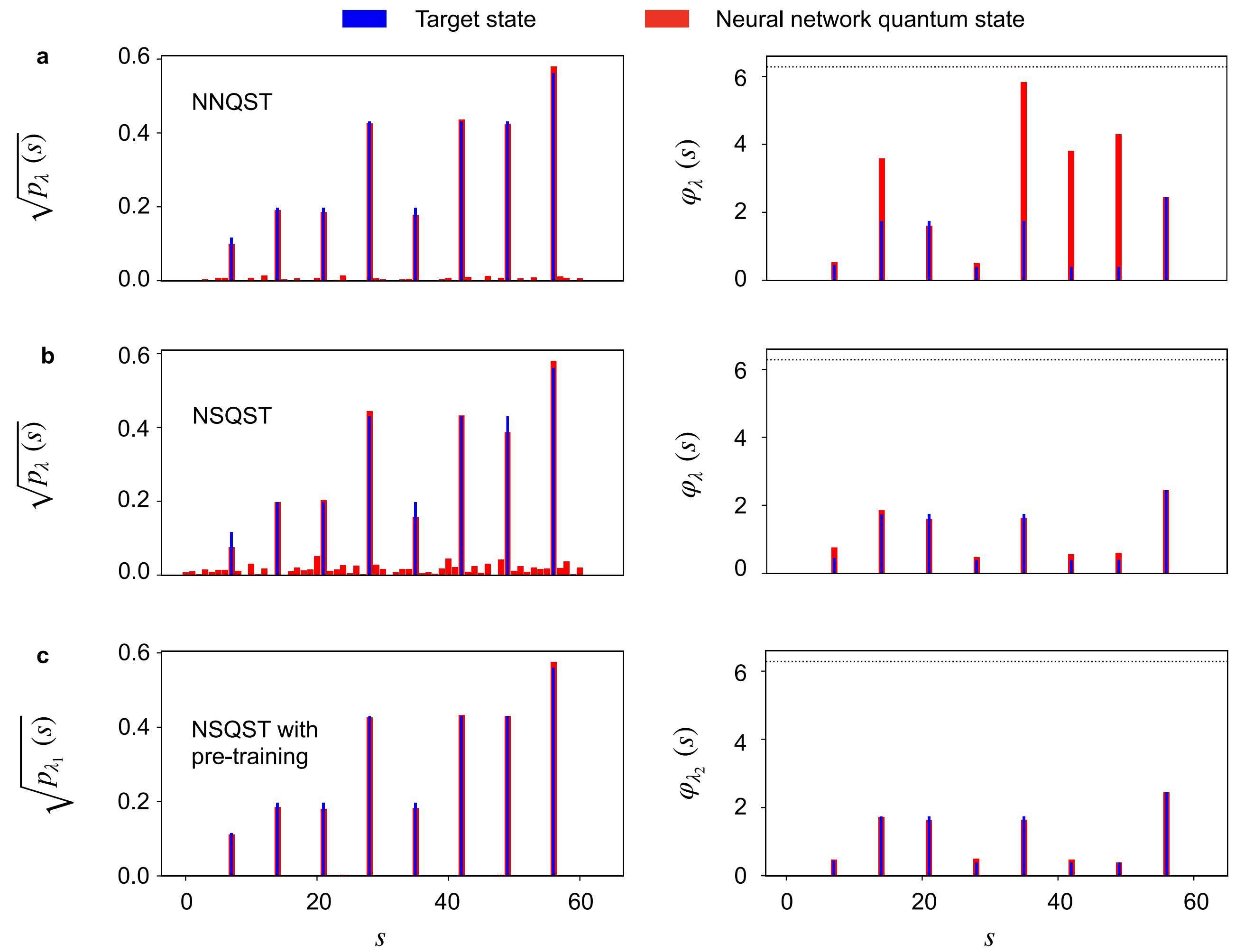
$$1 - |\langle\psi_\lambda|\Phi\rangle| \approx 1 - \frac{1}{N}\sum_i^N Tr(O_\lambda \rho_i) = 1 - \frac{1}{2^n}\left(1 - \frac{1}{f}\right) - \frac{1}{N}\sum_i^N |\langle\phi_i|\psi_\lambda\rangle|^2$$

Quantum material: Example 3

- 1D anti-ferromagnetic Heisenberg model two Trotter steps away from $|\uparrow\downarrow\uparrow\downarrow\uparrow\downarrow\rangle$.

$$H = \sum_{i=1}^{n-1} (X_i X_{i+1} + Y_i Y_{i+1} + Z_i Z_{i+1})$$





Robustness to noise

- Following D. E. Koh and S. Grewal, Quantum 6, 776 (2022):

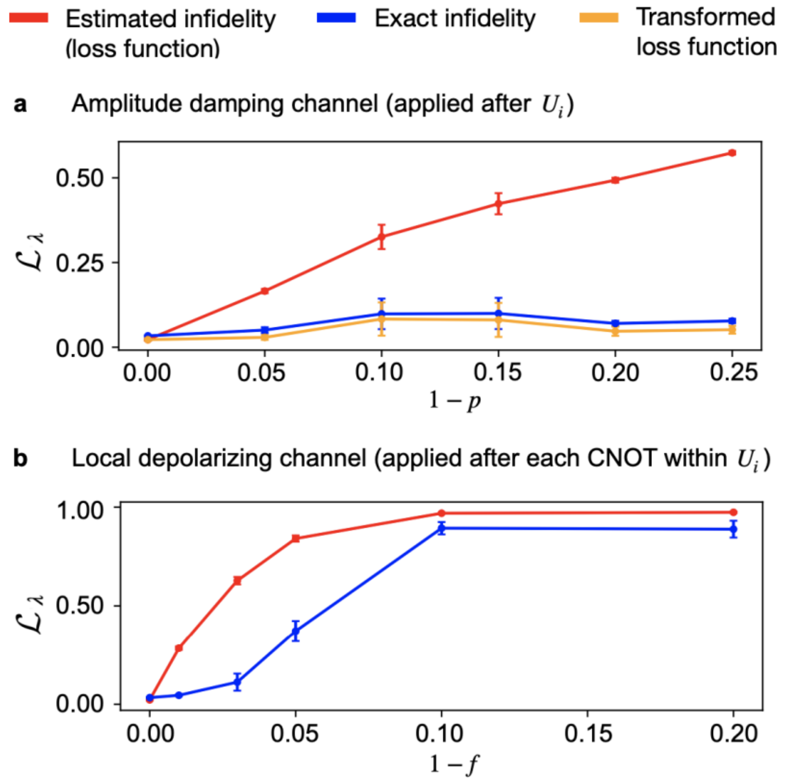
$$\hat{\rho}_i(\mathcal{E}, U_i, b_i) = \frac{1}{f(\mathcal{E})}|\phi_i\rangle\langle\phi_i| + \left(1 - \frac{1}{f(\mathcal{E})}\right)\frac{\mathbb{I}}{2^n} \qquad f(\mathcal{E}) = \frac{\mathrm{fid}(\mathcal{E}) - 1}{2^{2n} - 1}$$

- Scaled gradient:

$$\nabla_\lambda \mathcal{L}(\mathcal{E}) \approx \frac{-2}{N f(\mathcal{E})} \sum_{i=1}^{N} \Re\left[\left\langle \frac{\phi_i^*(s)}{\psi_\lambda^*(s)} D_\lambda(s) \right\rangle_{\psi_\lambda} \left\langle \frac{\phi_i(s)}{\psi_\lambda(s)} \right\rangle_{\psi_\lambda}\right]$$

- And the shifted loss function is

$$\mathcal{L}(\mathcal{E}) = \frac{1}{(2^n+1) f(\mathcal{E})}\mathcal{L}(\mathbb{I}) + \frac{(4^n - 1) f(\mathcal{E}) - 2^n + 1}{2^n(2^n+1) f(\mathcal{E})}$$



Comparison with direct shadow estimations

- Better generalization for non-local observations (e.g., long Pauli strings) in a QCD example:

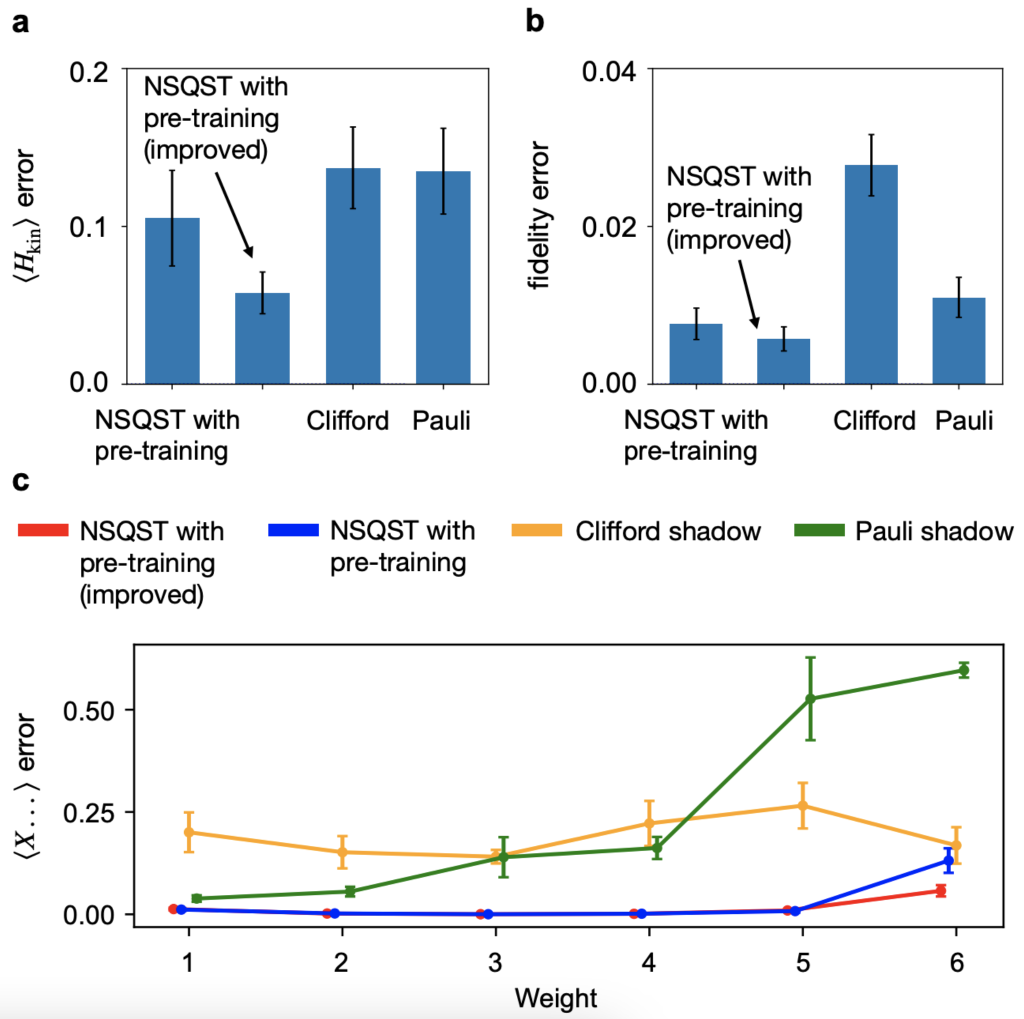
$$H_{SU(3)} = H_{kin} + \tilde{m} H_m + \frac{1}{2x} H_e,$$

where

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

$$H_m = \frac{1}{2}(6 - \sigma_1^z - \sigma_2^z - \sigma_3^z + \sigma_4^z + \sigma_5^z + \sigma_6^z),$$

$$H_e = \frac{1}{3}(3 - \sigma_1^z \sigma_2^z - \sigma_1^z \sigma_3^z - \sigma_2^z \sigma_3^z),$$



Concluding remarks

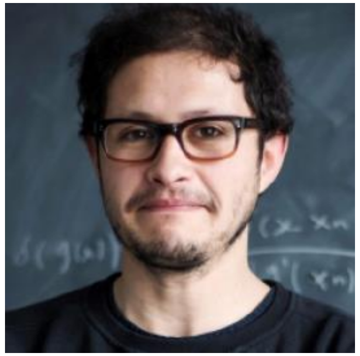
- Quantum experiments are very difficult to perform:
  - Expensive and long experimental cycles.
- “No cloning” has made us not think hard enough how to
  - Capture the experiments in re-usable fashion;
  - Make the results of the experiment available to the community.
- How can NSQST be better than just the list of measurements from a tomography scheme?
  - The same way GPT is much more useful than the entire corpus of text on the web.
  - NSQST provides an **operational representation** of the quantum state for
    - Other processes and applications
    - For interfacing between devices
- The neural network representation is a **digital twin** of the quantum state:
  - This shifts the value from quantum experiments to **quantum data**.
  - Inference from the digital twin is much **cheaper** than rerunning quantum experiments.
  - The digital twin is much **more malleable and easier to interface with** than the quantum computer.

Team

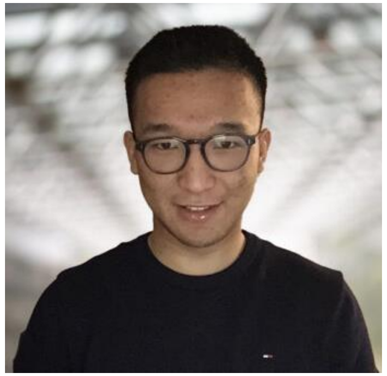
Elizabeth E Bennewitz
University of Maryland

Florian Hopfmueller
Nord Quantique

Bohdan Kulchytskyy
1QBit



Juan F Carrasquilla
Vector Institute, U Waterloo



Victor Wei
IQC, UW

Christine A. Muschik
IQC, UW, PI

W. A. Coish
McGill

Neural error mitigation of near-term quantum simulations. *Nature Machine Intelligence* 4.7 (2022): 618-624.

Neural-Shadow Quantum State Tomography." *preprint arXiv:2305.01078* (2023).

Pooya Ronagh
pooya.ronagh@uwaterloo.ca

1QBit
IQC Institute for Quantum Computing
Mitacs
PI Perimeter Institute
Vector Institute
CIFAR
Alfred P. Sloan Foundation
NSERC CRSNG
University of Waterloo
McGill
New Frontiers in Research Fund
Fonds Nouvelles frontières en recherche
Fonds de recherche Nature et technologies Québec
Canada First Research Excellence Fund
Innovation, Science and Economic Development Canada

Open positions:
Postdoc 1 (neural quantum states)
Postdoc 2 (quantum algorithms)
PhD (quantum algorithms)