Efficient classical learning surrogates for quantum circuits at scale

Yuxuan Du

CCDS & SPMS NTU

2025.12.11

One-page summary

Article | Open access | Published: 22 April 2025

Efficient learning for linear properties of bounded-gate quantum circuits

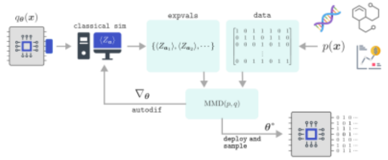
Yuxuan Du, Min-Hsiu Hsieh & Dacheng Tao

Nature Communications 16, Article number: 3790 (2025) | Cite this article


Min-Hsiu Hsieh   Dacheng Tao

Classical **learning** surrogates

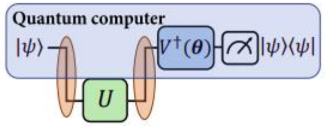
Train C; Deploy Q



[Joseph Bowles]

Train Q; Deploy Q

QNNs



Train C; Deploy C

No advantages

Train Q; Deploy C

Classical learning surrogate

[Provable Efficiency]

Demonstration of Efficient Predictive Surrogates for Large-scale Quantum Processors

Wei-You Liao[1,*] Yuxuan Du[2,*,†] Xinbiao Wang[2,*] Tian-Ci Tian[1], Yong Luo[3], Bo Du[3], Dacheng Tao[2,†] and He-Liang Huang[1†]

[1] *Henan Key Laboratory of Quantum Information and Cryptography, Zhengzhou, Henan 450000, China*
[2] *College of Computing and Data Science, Nanyang Technological University, Singapore 639798, Singapore and*
[3] *Institute of Artificial Intelligence, School of Computer Science, Wuhan University, Wuhan 430072, China*

arXiv:2507.1747 (noisy scenario; 20-qubit SC QC)

Sample-efficient quantum error mitigation via classical learning surrogates

Wei-You Liao,[1] Ge Yan,[2] Yujin Song,[2] Tian-Ci Tian,[1] Wei-Ming Zhu,[1] De-Tao Jiang,[1] Yuxuan Du,[2,3,*] and He-Liang Huang[1,†]

[1] *Henan Key Laboratory of Quantum Information and Cryptography, Zhengzhou, Henan 450000, China*
[2] *College of Computing and Data Science, Nanyang Technological University, Singapore, Singapore*
[3] *School of Physical and Mathematical Science, Nanyang Technological University, Singapore, Singapore*

arXiv:2511.07092 (noisy scenario; 100-qubit ZNE)

Artificial intelligence for representing and characterizing quantum systems

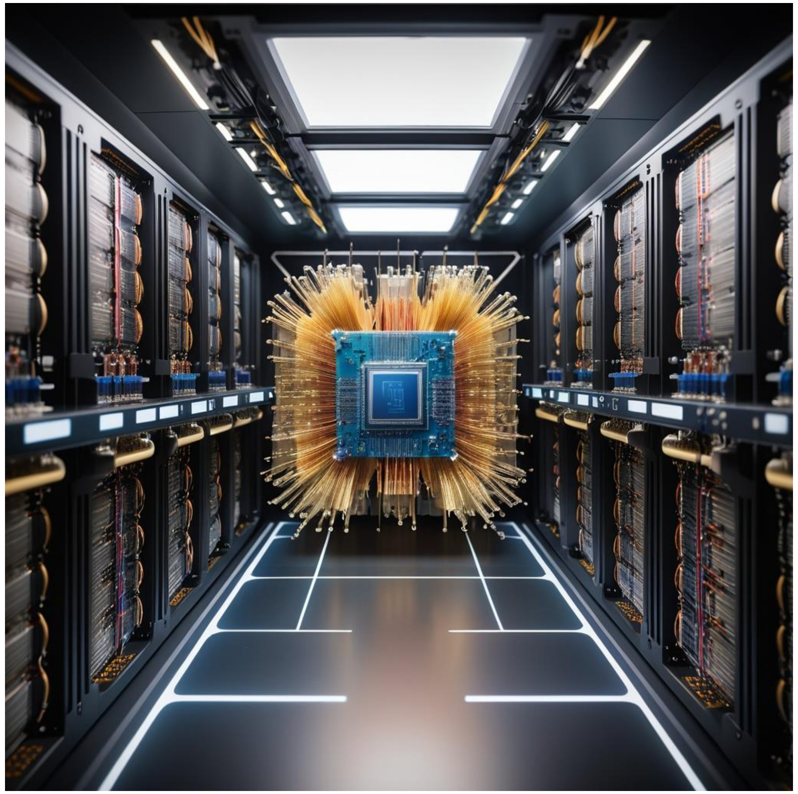
Yuxuan Du,[1] Yan Zhu,[2] Yuan-Hang Zhang,[3] Min-Hsiu Hsieh,[4] Patrick Rebentrost,[5,6] Weibo Gao,[7,5] Ya-Dong Wu,[8,*] Jens Eisert,[9,10] Giulio Chiribella,[2,11,12] Dacheng Tao,[1] and Barry C. Sanders[13]

[1] *College of Computing and Data Science, Nanyang Technological University, Singapore 639798, Singapore*
[2] *QICI Quantum Information and Computation Initiative, Department of Computer Science, The University of Hong Kong, Pokfulam Road, Hong Kong*
[3] *Department of Physics, University of California, San Diego, La Jolla, CA 92093, USA*
[4] *Hon Hai (Foxconn) Research Institute, Taipei, Taiwan*

arXiv:2509.04923 (survey; classical AI for Q)

Motivations & Backgrounds

How to understand quantum computers at scale?

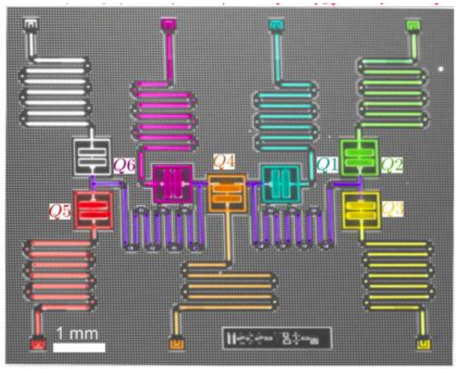


Motivations & Backgrounds

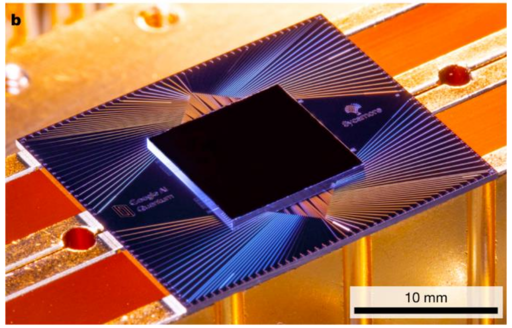
- What are *large-scale* quantum computers?

🎈Era of large-scale quantum computing

2017
IBM-Q
6 qubits

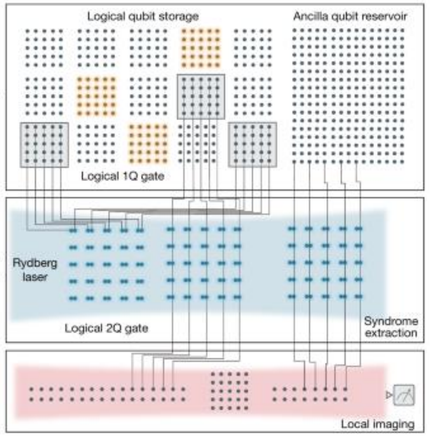


2019
Quantum supremacy
53 qubits



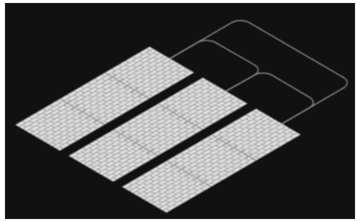
2023
IBM Condor
1121 qubits

2024
Reconfigurable atom arrays
256 qubits



2025
Google
(101 qubits)

Pasqal
(1000 atoms)

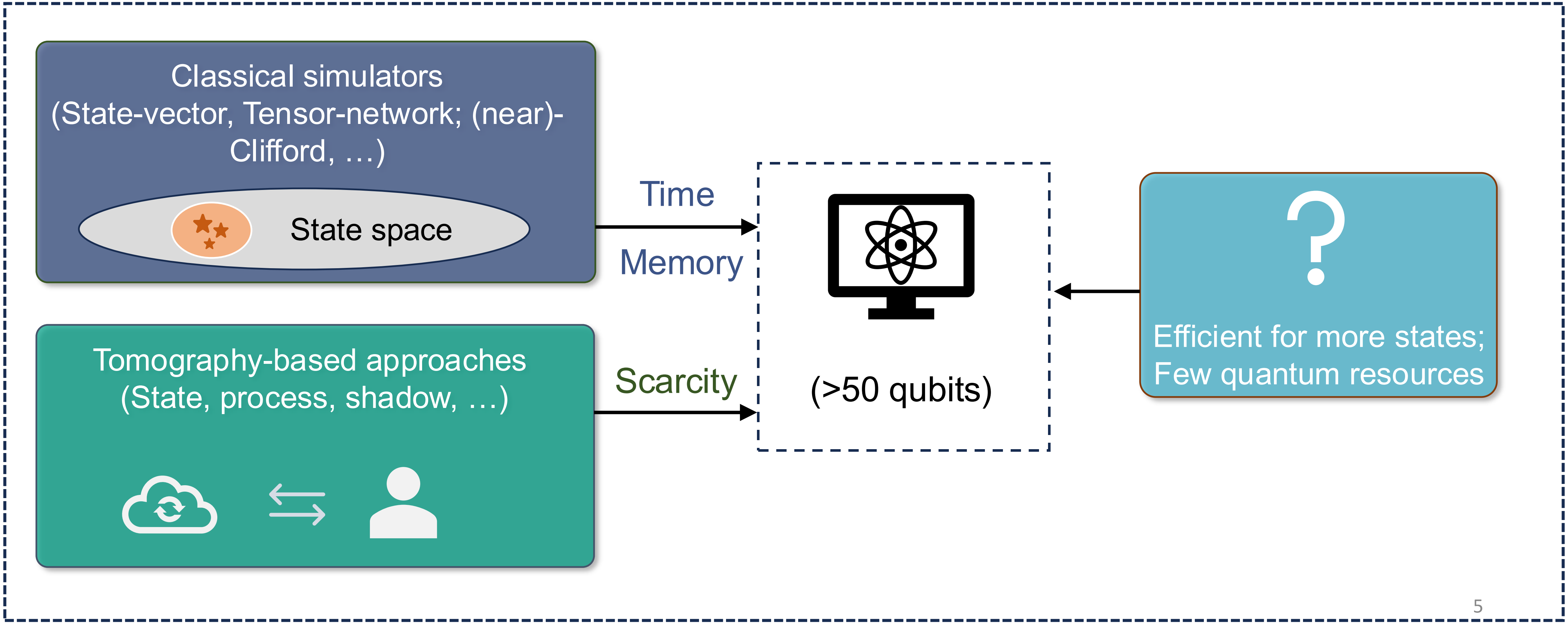


Images from nature, 549(7671), 242-246; Nature 574.7779 (2019): 505-510; Nature 626.7997 (2024): 58-65; https://www.nature.com/articles/d41586-023-03854-1

Motivations & Backgrounds

Challenges in Understanding Large-qubit Quantum Circuits:

Conventional approaches become **expensive** or **incapable!**



Question: Is there any efficient classical ML model that can well predict the linear properties (mean values) of large-qubit quantum circuits?

Question: Is there any efficient classical ML model … large-qubit quantum circuits?

Yes, it is classical ***learning*** surrogate (or agent, emulator, learner …)

Article | Open access | Published: 22 April 2025

Efficient learning for linear properties of bounded-gate quantum circuits

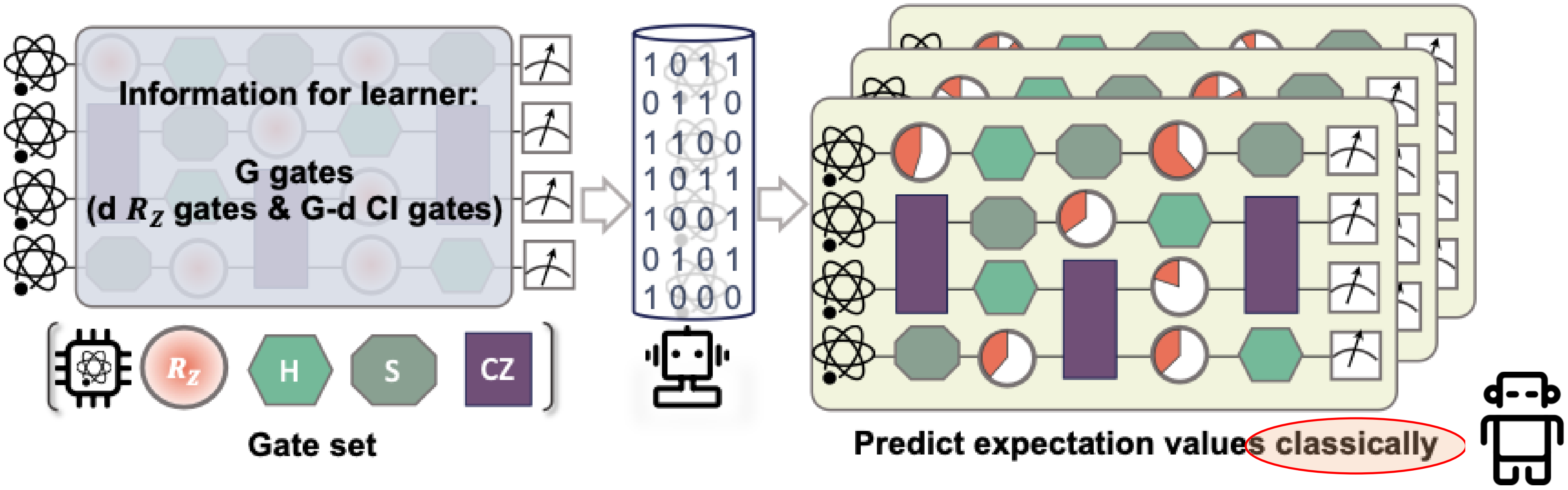
Yuxuan Du, Min-Hsiu Hsieh & Dacheng Tao

Nature Communications **16**, Article number: 3790 (2025) | Cite this article

3234 Accesses | **3** Citations | **20** Altmetric | Metrics

Problem Setup

Focus: The explored quantum circuits are formed by RZ + Clifford gates (universality)



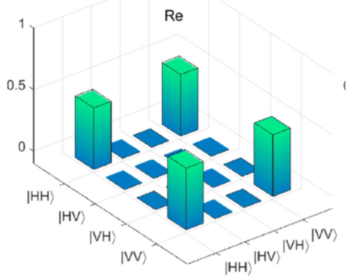
Quantum circuit: Arbitrary *N*-qubit input states, RZ + Clifford gates, incoherent measurements

Efficient classical learner: Computational complexity of training and inference polynomially scales with *N and d*

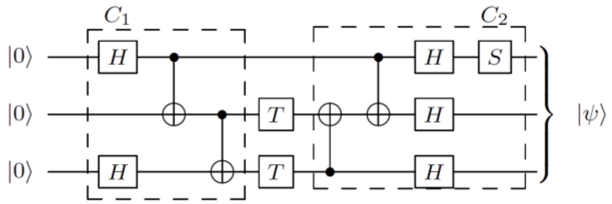
Efficient classical learning surrogates have many crucial implications

Enrich quantum learning theory

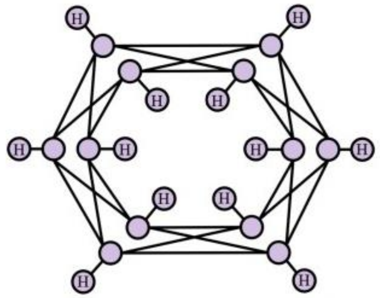
Classical Learner (Train Q; Deploy C)



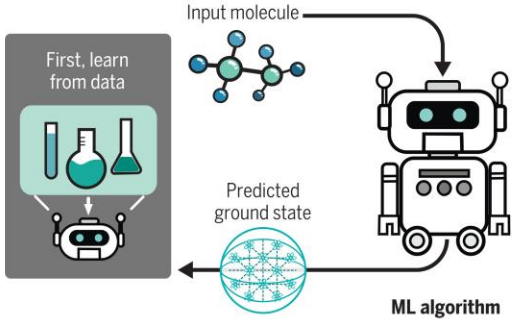
State learning



t-doped Clifford circuits

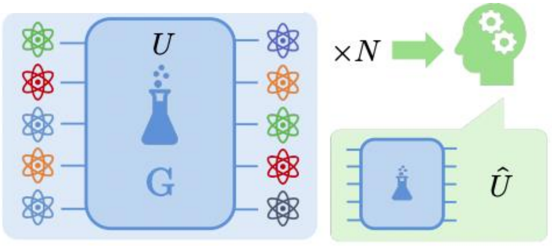


Graph states

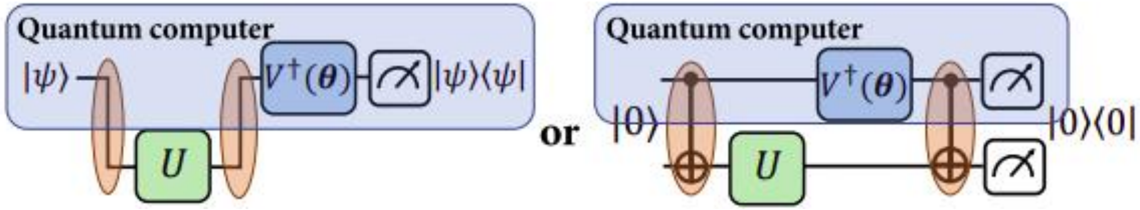


Ground states properties

Quantum Learner (Train Q; Deploy Q)



Dynamics of bounded-gate circuit

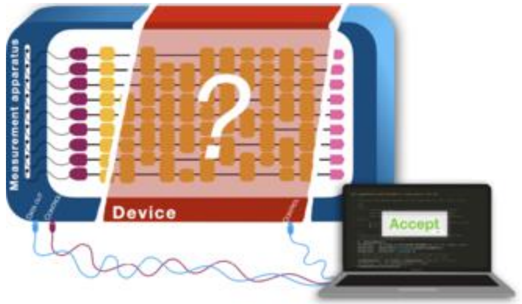


Dynamics of quantum circuit

Efficient classical learning surrogates have many crucial implications

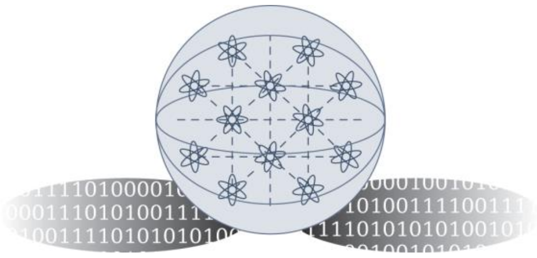
Deliver practical applications (resource reduction)

Quantum certification

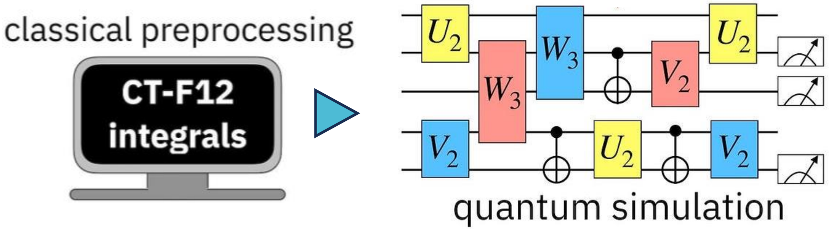


- Fidelity
- Purity
- Entropy
- …

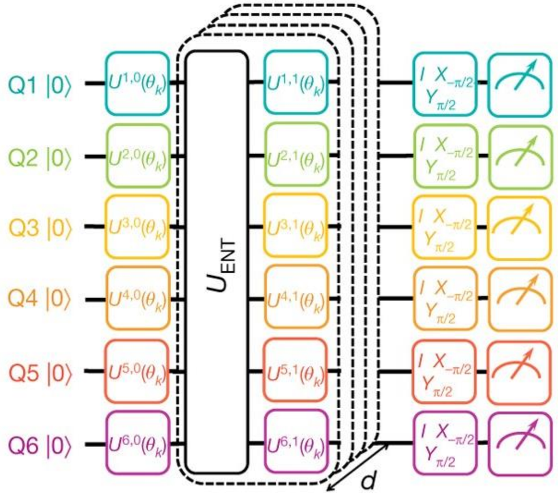
Quantum shadow estimation



Quantum simulation



VQAs



Images from *Nature Reviews Physics* 4.2 (2022): 81-81. https://medium.com/qiskit/a-tale-of-colliding-electrons-boosting-the-accuracy-of-chemical-simulations-on-quantum-computers-50a4b4ee5c64; Nature volume 549, pages242–246 (2017

Main Theoretical Results

Problem Setup: State Space

For an *N*-qubit parametrized quantum circuit, the concept class (d RZ gates and G-d Clifford gates) is

$$\mathcal{F} = \left\{ f(\boldsymbol{x}, O) = \mathrm{Tr}(\rho(\boldsymbol{x})O) \middle| U \in \mathsf{Arc}(U, d, G) \right\}$$

with $U(\boldsymbol{x}) = \prod_{l=1}^{d} \mathrm{RZ}(\boldsymbol{x}_l) u_e \in U(2^N)$ and $\rho(\boldsymbol{x}) = U(\boldsymbol{x}) \rho_0 U(\boldsymbol{x})^\top$

Central Aim: Complexity of classical learning surrogates to achieve ϵ-prediction error:

$$\mathbb{E}_{\boldsymbol{x} \sim [-\pi, \pi]^d} \left| h(\boldsymbol{x}, O) - f(\boldsymbol{x}, O) \right|^2 \leq \epsilon$$

Problem Setup: State Space

Central Aim: Complexity of classical learning surrogates to achieve ϵ-prediction error:

$$\mathbb{E}_{\boldsymbol{x}\sim[-\pi,\pi]^d} \left|h(\boldsymbol{x}, O) - f(\boldsymbol{x}, O)\right|^2 \leq \epsilon$$

Three subaims

1. What is the learnability of $\mathcal{F}$ in terms of the *sample* complexity?
2. What is the learnability of $\mathcal{F}$ in terms of the *computational* complexity?
3. How to design *a detailed* algorithm to match the above bounds?

Results for Sub-Aim 1 (sample complexity, Theorem 1)

A brief summary: $\Theta(d/\epsilon)$ training examples are sufficient and necessary to learn $\mathcal{F}$

$$\mathbb{E}_{\boldsymbol{x}\sim[-\pi,\pi]^d} \left|h(\boldsymbol{x}, O) - f(\boldsymbol{x}, O)\right|^2 \leq \epsilon$$

Technical tools

- Covering / Packing net
- Hypothesis testing
- Fano inequality

Lower bound

$n \sim \widetilde{\Omega}(d)$ is necessary

Upper bound

$n \sim \widetilde{O}(d)$ is sufficient

Results for Sub-Aim 2 (computational complexity)

A brief summary: there exists a class of circuits $\mathcal{F}$ that need an exponentially computational complexity to learn it.

[MGD2024, arXiv: 2405.02027v2] Limitations from complexity theory

$$\mathcal{F}_{\text{Hard}} = \left\{ f(\boldsymbol{x}^{\perp}) = \operatorname{Tr}\left(U \left|\boldsymbol{x}^{\perp}\right\rangle \left\langle \boldsymbol{x}^{\perp}\right| U^{\dagger} O\right) \middle| \boldsymbol{x}^{\perp} \in \{-1, 1\}^{N}, O \sim \mathbb{D}_{O} \right\}$$

$\text{BQP} \subsetneq \text{P/Poly}$ → No alg that can solely use the collected measure-out data to learn $f(\boldsymbol{x}^{\perp})$ in a **polynomial** time

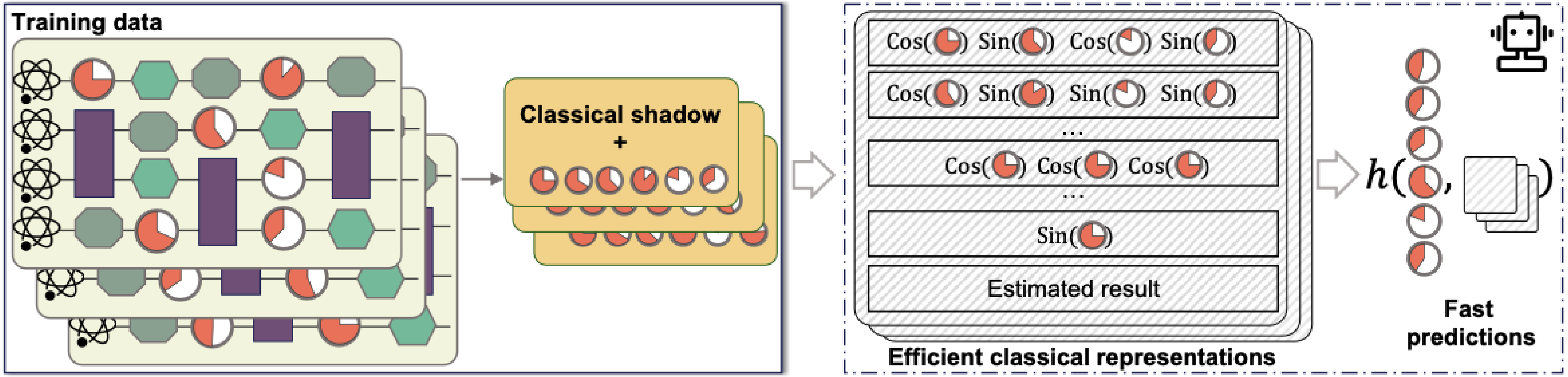
Question: Under some practical conditions, is there any efficient classical ML model that can well predict the linear properties of large-qubit quantum circuits?

Sample efficient & May not computationally efficient

But …

Results for Sub-Aim 3 (algorithmic implementation)

Here we devise an efficient classical learning surrogate to learn mean values for arbitrary observables.

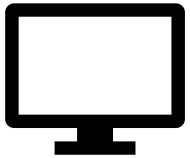


(I. Data collection, interaction with Q)

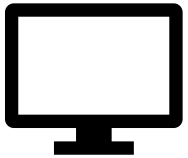
$\{\boldsymbol{x}^{(i)}, \tilde{\rho}_T(\boldsymbol{x}^{(i)})\}_{i=1}^{n}$

T Pauli-based snapshots

(II. Build kernel, No Q)



(III. Prediction, No Q)



Results for Sub-Aim 3 (algorithmic implementation)

The proposed kernel-based classical learning surrogate yields

$$h_s(\boldsymbol{x}, O) = \frac{1}{n}\sum_{i=1}^{n} \kappa_\Lambda\left(\boldsymbol{x}, \boldsymbol{x}^{(i)}\right) g\left(\boldsymbol{x}^{(i)}, O\right)$$

$\kappa_\Lambda(\boldsymbol{x}, \boldsymbol{x}^{(i)})$: a truncated trigonometric monomial kernel with

$$\kappa_\Lambda\left(\boldsymbol{x}, \boldsymbol{x}^{(i)}\right) = \sum_{\boldsymbol{\omega}, \|\boldsymbol{\omega}\|_0 \le \Lambda} 2^{\|\boldsymbol{\omega}\|_0} \Phi_{\boldsymbol{\omega}}(\boldsymbol{x}) \Phi_{\boldsymbol{\omega}}\left(\boldsymbol{x}^{(i)}\right) \in \mathbb{R}$$

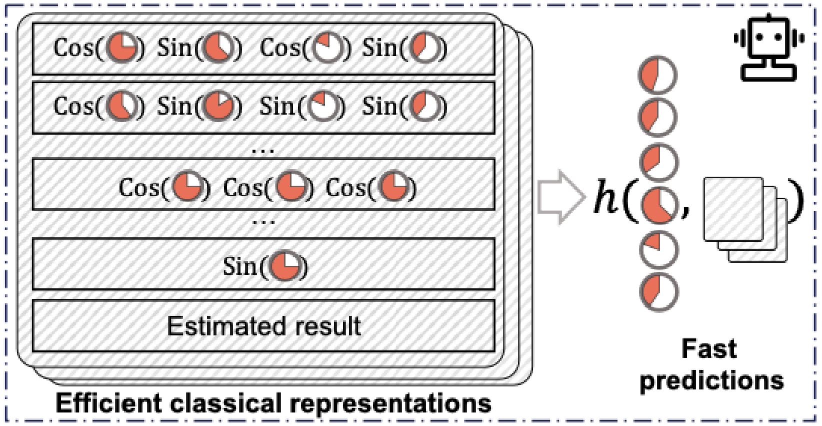
$\Phi_{\boldsymbol{\omega}}(\boldsymbol{x})$ with $\boldsymbol{\omega} \in \{0,1,-1\}^d$ is the trigonometric monomial basis

$$\Phi_{\boldsymbol{\omega}}(\boldsymbol{x}) = \prod_{i=1}^{d} \begin{cases} 1 & \text{if } \boldsymbol{\omega}_i = 0 \\ \cos(\boldsymbol{x}_i) & \text{if } \boldsymbol{\omega}_i = 1 \\ \sin(\boldsymbol{x}_i) & \text{if } \boldsymbol{\omega}_i = -1 \end{cases}.$$

$g(\boldsymbol{x}^{(i)}, O) = \mathrm{Tr}(\tilde{\rho}_T(\boldsymbol{x}^{(i)})O)$: shadow estimation

$\rho(\boldsymbol{x}) = \sum_{\boldsymbol{\omega}} \Phi_{\boldsymbol{\omega}}(\boldsymbol{x})\rho_{\boldsymbol{\omega}}$ by Pauli Transfer Matrix

(Smoothness) $\mathbb{E}_{\boldsymbol{x}\sim[-\pi,\pi]^d} \left\|\nabla_{\boldsymbol{x}} \mathrm{Tr}(\rho(\boldsymbol{x})O)\right\|_2 \le C.$



Results for Sub-Aim 3 (proof sketch)

The proof idea is separately bounding the estimation and truncation error:

$$\mathbb{E}_{\boldsymbol{x}\sim[-\pi,\pi]^d}\left[\left|\mathrm{Tr}(O\hat{\sigma}_n^{(1)}(\boldsymbol{x})) - \mathrm{Tr}(O\rho(\boldsymbol{x}))\right|^2\right]$$

$$\leq \left(\sqrt{\mathbb{E}_{\boldsymbol{x}\sim[-\pi,\pi]^d}\left[\left|\mathrm{Tr}(O\rho_\Lambda^{(1)}(\boldsymbol{x})) - \mathrm{Tr}(O\rho(\boldsymbol{x}))\right|^2\right]} + \sqrt{\mathbb{E}_{\boldsymbol{x}\sim[-\pi,\pi]^d}\left[\left|\mathrm{Tr}(O\sigma_n^{(1)}(\boldsymbol{x})) - \mathrm{Tr}(O\rho_\Lambda(\boldsymbol{x}))\right|^2\right]}\right)^2$$

Truncation error

(Smoothness) $\mathbb{E}_{x\sim[-\pi,\pi]^d}\left|\left|\nabla_x \mathrm{Tr}(\rho(x)O)\right|\right|_2 \leq C.$

▽

$$\mathbb{E}_{\boldsymbol{x}\sim[-\pi,\pi]^d}\left|\mathrm{Tr}(O\rho_\Lambda(\boldsymbol{x}) - \mathrm{Tr}(O\rho(\boldsymbol{x}))\right|^2 \leq \frac{C}{\Lambda}.$$

Estimation error

(Pauli-based Shadow error) $\|O\|^2_{\mathrm{shadow}} = 3^k$

▽

$$\mathbb{E}_{\boldsymbol{x}\sim[-\pi,\pi]^d}\left[\left|\mathrm{Tr}(O\hat{\sigma}_n(\boldsymbol{x})) - \mathrm{Tr}(O\rho_\Lambda(\boldsymbol{x}))\right|^2\right] \leq |\mathfrak{C}(\Lambda)|\frac{1}{2n}B^2 9^K \log\left(\frac{2\cdot|\mathfrak{C}(\Lambda)|}{\delta}\right)$$

▽

Sample and runtime efficient $n \geq \tilde{O}\left(\left|\mathfrak{C}\left(\frac{4C}{\epsilon}\right)\right| 2B^2 9^K \epsilon^{-1}\right)$ ➡ $\mathbb{E}_{\boldsymbol{x}\sim[-\pi,\pi]^d}\left|h(\boldsymbol{x},O) - f(\boldsymbol{x},O)\right|^2 \leq \epsilon$

Results for Sub-Aim 3 (theoretical guarantee, Theorem 2)

Generalize to *d* Rot-Pauli + Clifford + T gates

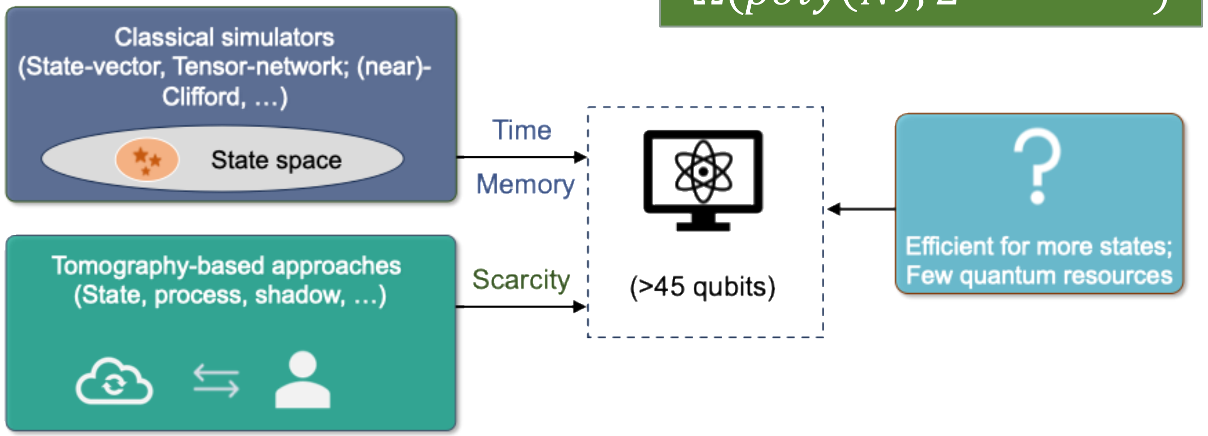
When $U(\boldsymbol{x})$ is compose of RZ, H, T gates, CNOT, etc

$$\mathbb{E}_{\boldsymbol{x}\sim[-\pi,\pi]^d,\,\text{shadow}}[\hat{\sigma}_n(\boldsymbol{x})] = \rho_\Lambda(\boldsymbol{x}).$$

$$n \geq \tilde{O}\left(\left|\mathfrak{C}\left(\frac{4C}{\epsilon}\right)\right| 2B^2 9^K \epsilon^{-1}\right)$$

Beyond near-Clifford simulators; sparse Pauli dynamics

$\Omega(poly(N), 2^{\#\text{ of T gates}})$



Surrogates are purely classical at the inference stage

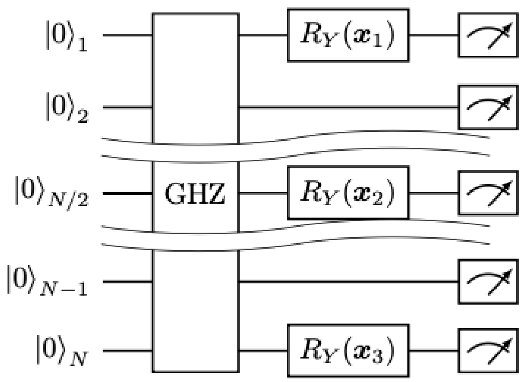
Numerical results

Numerical Results: Two-point Correlation of 60-qubit Rotational GHZ States

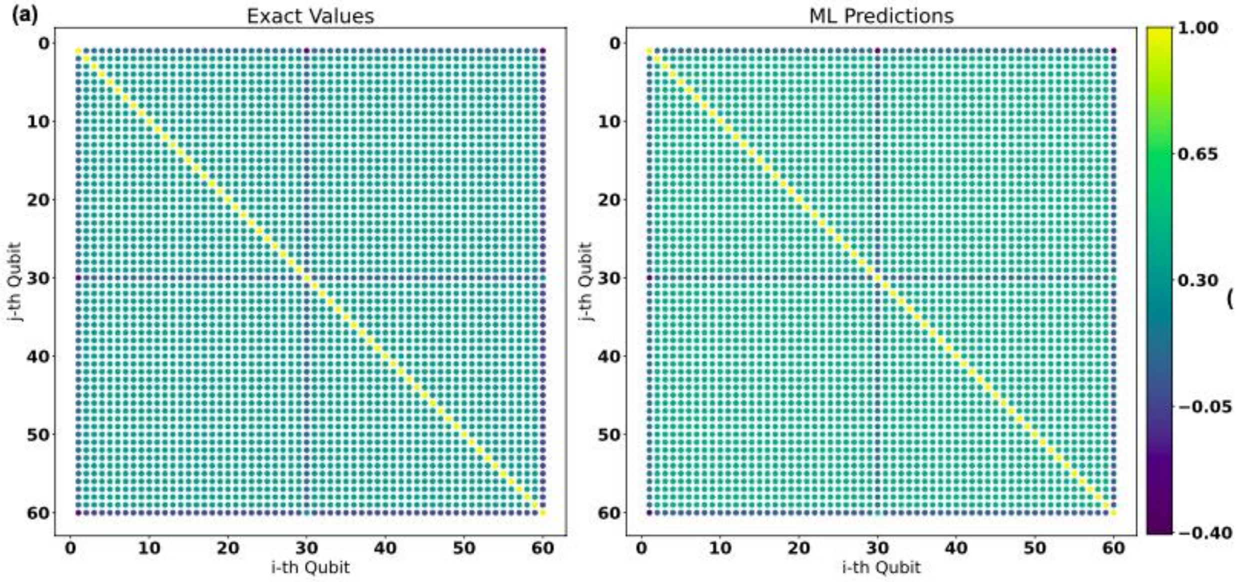
- Dataset construction of N-qubit rotational GHZ states:

$$|\mathrm{GHZ}(\boldsymbol{x})\rangle = (\mathrm{RY}_1(\boldsymbol{x}_1) \otimes \mathrm{RY}_{N/2}(\boldsymbol{x}_2) \otimes \mathrm{RY}_N(\boldsymbol{x}_3)) \frac{|0\cdots0\rangle + |1\cdots1\rangle}{\sqrt{2}}.$$

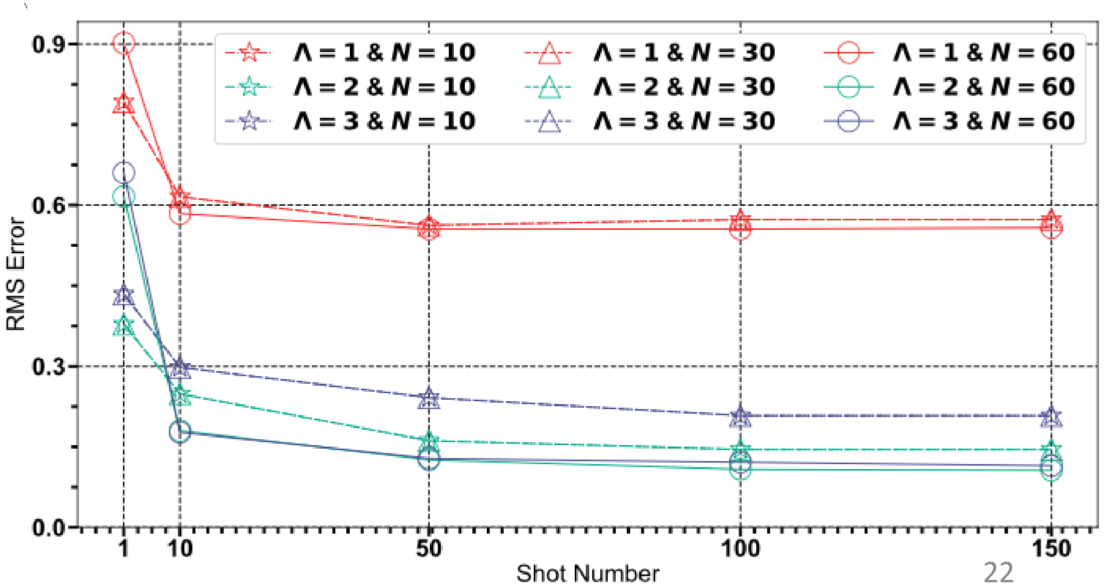
- Two-point Correlation: $C_{ij} = (X_iX_j + Y_iY_j + Z_iZ_j)/3$



RMS error for all qubit pairs $(\Lambda = 3, T = 1000, n = 30)$



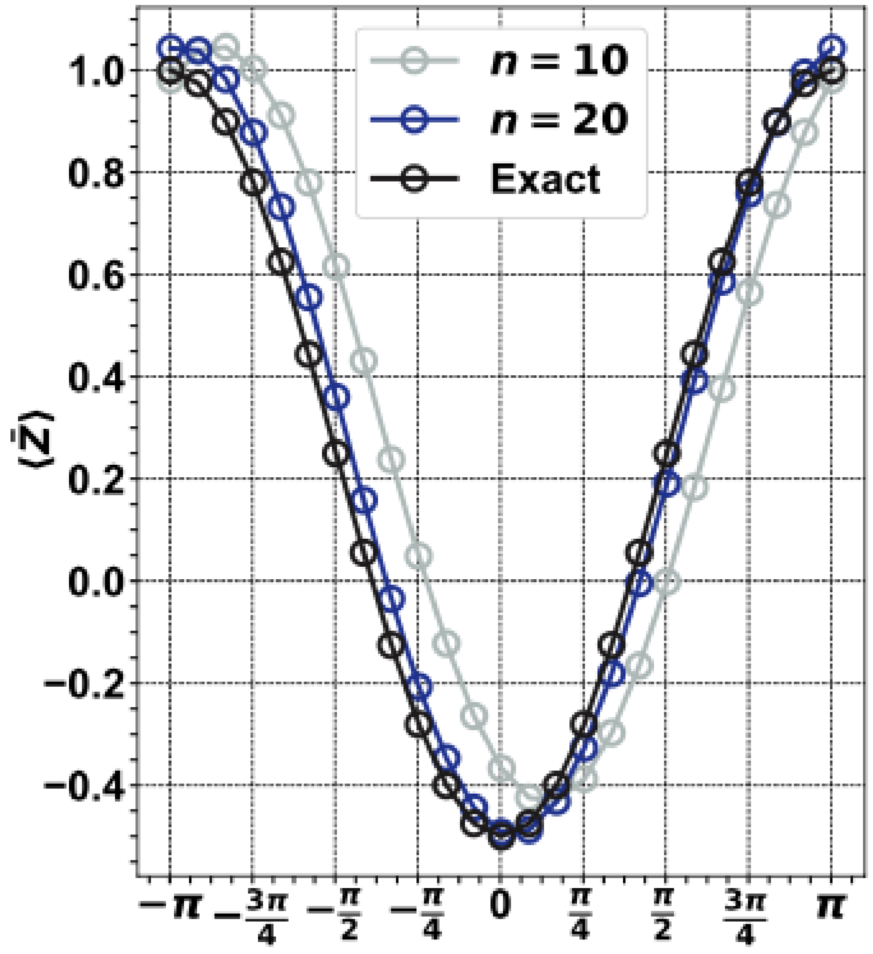
RMS error VS N, Λ, T $(n = 500)$



Numerical Results: Quantum Simulation & Pre-training VQAs

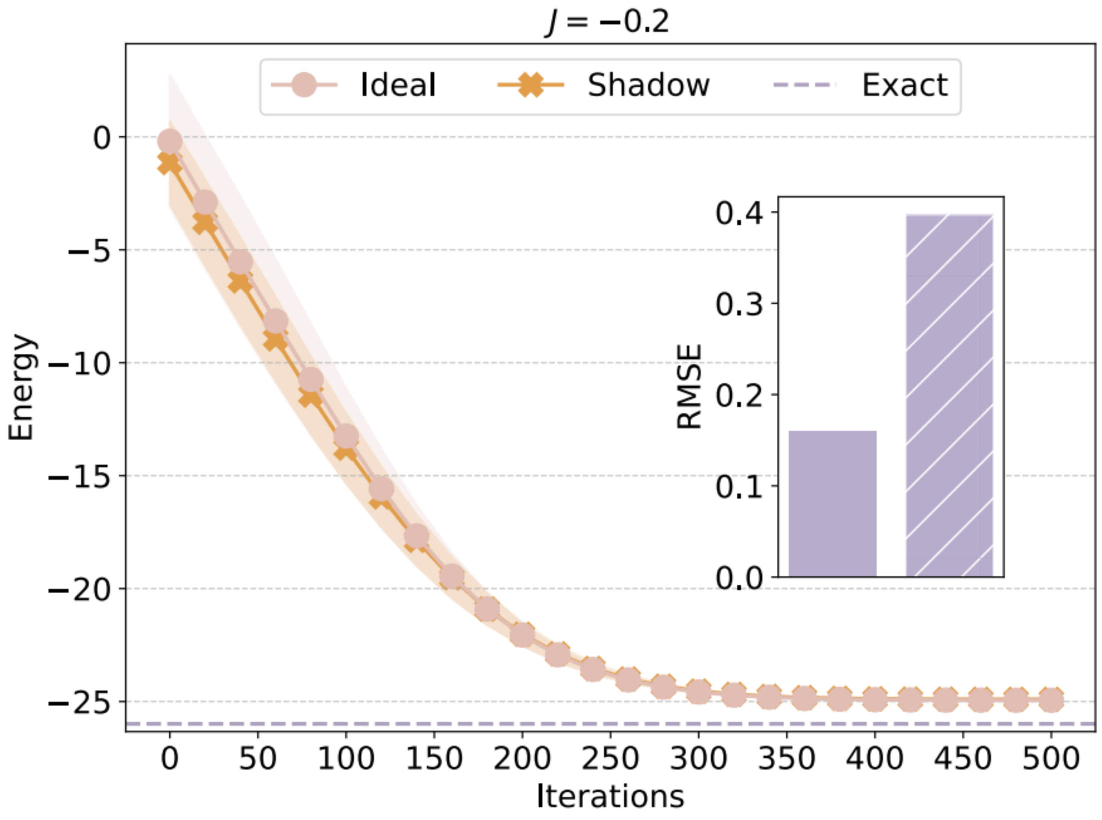
60-qubit global Hamiltonian simulation

$$U(\boldsymbol{x}) = \prod_{j=1}^{d}(e^{-\imath \boldsymbol{x}_j \otimes_{i=1}^{N} Z_i} \otimes_{i=1}^{N} \mathrm{RX}(\pi/3))$$



Pre-training VQE for 50-qubit TFIMs

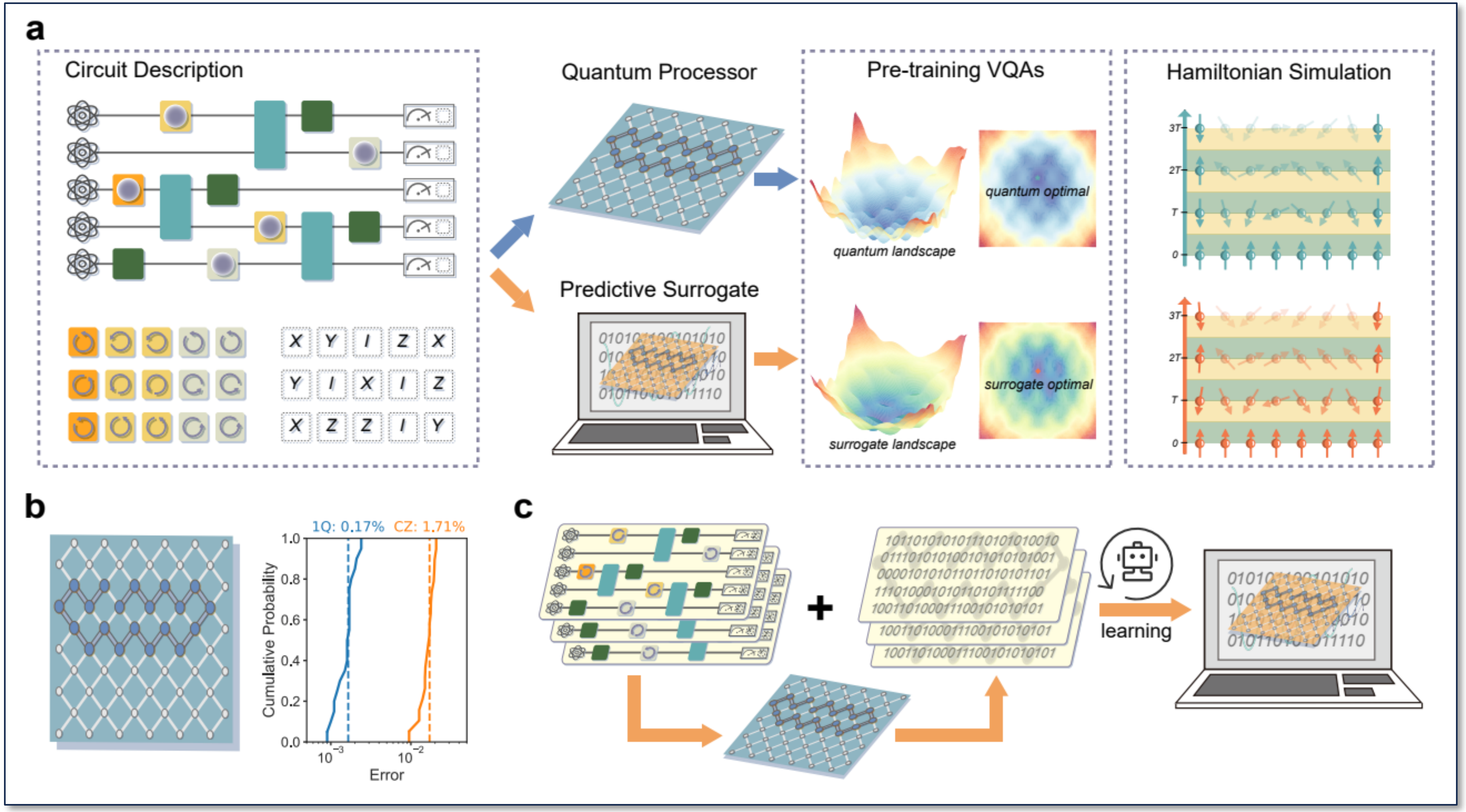
HVA with d=99; n=1500; T=300



Further results about learning surrogates

Beyond kernel-based classical surrogates

We propose two new classical surrogates to predict the behaviors of noisy quantum circuits



Experimental demonstration on a 20-qubits superconducting processor

Beyond kernel-based classical surrogates

We propose two new classical surrogates to predict the behaviors of noisy quantum circuits

Surrogate 1: kernel-based method on noisy processors

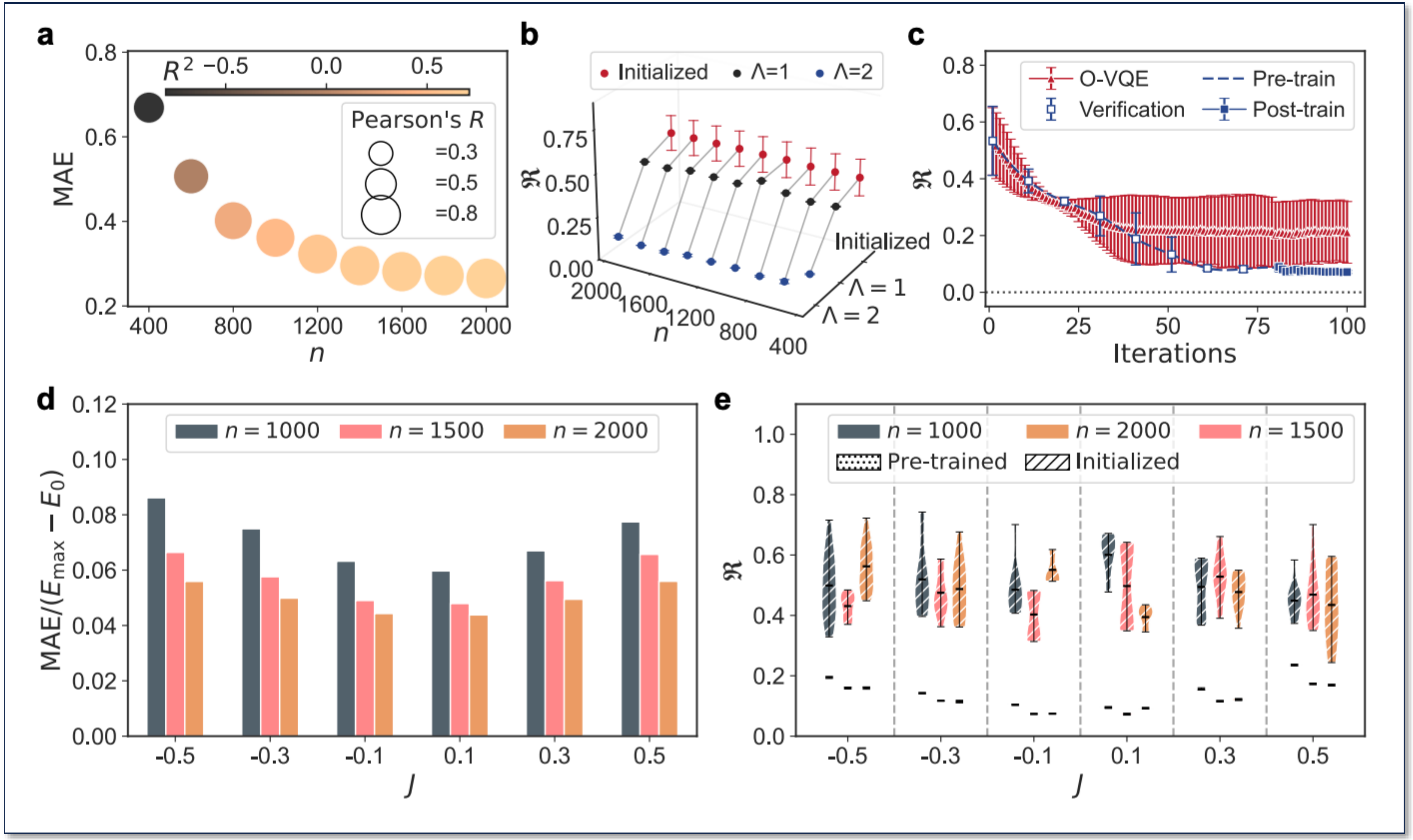
$$h_{\mathsf{cs}}(\boldsymbol{x}', O) = \frac{1}{n}\sum_{i=1}^{n} \kappa_{\Lambda}\left(\boldsymbol{x}', \boldsymbol{x}^{(i)}\right) g(\boldsymbol{x}^{(i)}, O) \quad \Rightarrow \quad n = \tilde{\Omega}\Big(\Big|\mathfrak{C}\Big(\min\Big\{\frac{4C}{\epsilon}, \frac{1}{2(p+p_Z)}\log\Big(\frac{2B}{\sqrt{\epsilon}}\Big)\Big\}\Big)\Big|\frac{2B^2 9^K}{\epsilon}\Big)$$

Surrogate 2: regression-based method on noisy processors [correlated inputs; arbitrary data distribution]

$$h_{\mathsf{qs}}(\boldsymbol{x}, \widehat{\mathbf{w}}) = \langle \boldsymbol{\Phi}_{\mathfrak{C}(\Lambda)}(\boldsymbol{x}), \widehat{\mathbf{w}} \rangle \quad \Rightarrow \quad n = \left(\frac{1}{q(1+R)}\right)^{4deq(1+R)} \cdot \frac{\log(1/\delta)}{9}$$

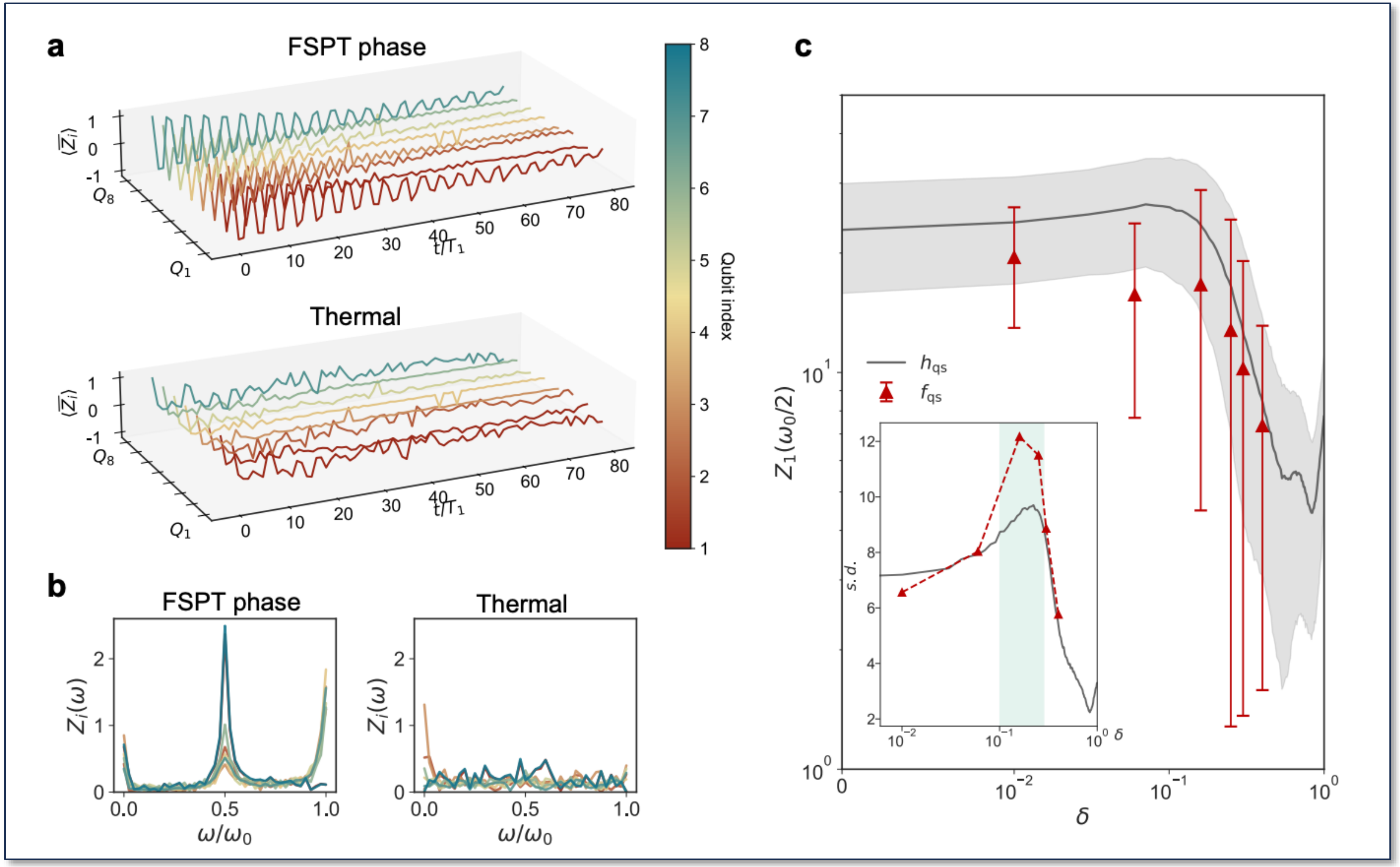
Beyond kernel-based classical surrogates

Task 1: Pre-train VQE for TFIM models. Outperform vanilla VQE with 0.023% measurements.



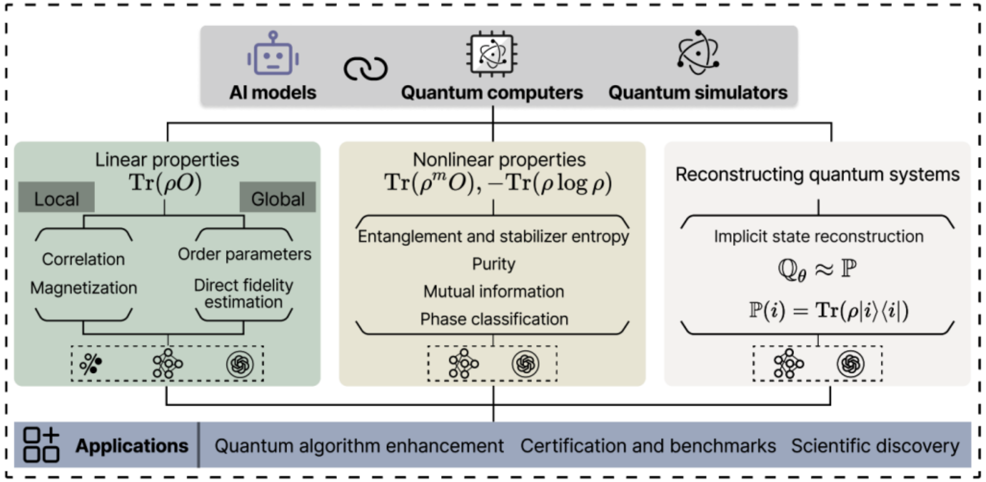
Beyond kernel-based classical surrogates

Task 2: Identification of non-equilibrium Floquet symmetry-protected topological phases.



Conclusion & Outlook

Conclusion & Outlook



Q1: Any provably efficient learning surrogate beyond PAC learning?

Q2: More applications of classical learning surrogates?

Q3: Any advanced classical learning surrogates beyond mean-value estimation?

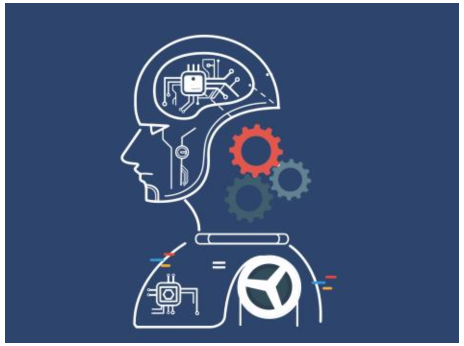
Artificial intelligence for representing and characterizing quantum systems

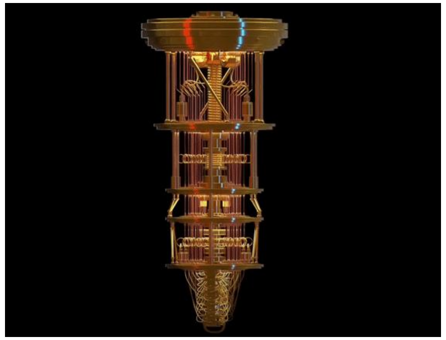
Yuxuan Du,[1] Yan Zhu,[2] Yuan-Hang Zhang,[3] Min-Hsiu Hsieh,[4] Patrick Rebentrost,[5,6] Weibo Gao,[7,5] Ya-Dong Wu,[8,*] Jens Eisert,[9,10] Giulio Chiribella,[2,11,12] Dacheng Tao,[1] and Barry C. Sanders[13]

[1]*College of Computing and Data Science, Nanyang Technological University, Singapore 639798, Singapore*
[2]*QICI Quantum Information and Computation Initiative, Department of Computer Science, The University of Hong Kong, Pokfulam Road, Hong Kong*
[3]*Department of Physics, University of California, San Diego, La Jolla, CA 92093, USA*
[4]*Hon Hai (Foxconn) Research Institute, Taipei, Taiwan*

Thank You for Listening!

AI





QC

SOTA AI Techniques — Quantum Computing

Email: yuxuan.du@ntu.edu.sg
Homepage: https://yuxuan-du.github.io/

Personal website