



OPEN Universal compilation for quantum state tomography

Vu Tuan Hai^{1,2} & Le Bin Ho^{3,4,5}✉

Universal compilation is a training process that compiles a trainable unitary into a target unitary. It has vast potential applications from depth-circuit compressing to device benchmarking and quantum error mitigation. Here we propose a universal compilation algorithm for quantum state tomography in low-depth quantum circuits. We apply the Fubini-Study distance as a trainable cost function and employ various gradient-based optimizations. We evaluate the performance of various trainable unitary topologies and the trainability of different optimizers for getting high efficiency and reveal the crucial role of the circuit depth in robust fidelity. The results are comparable with the shadow tomography method, a similar fashion in the field. Our work expresses the adequate capability of the universal compilation algorithm to maximize the efficiency in the quantum state tomography. Further, it promises applications in quantum metrology and sensing and is applicable in the near-term quantum computers for various quantum computing tasks.

Quantum computers promise an excellent computational capacity that is intractable for classical computers to solve challenging problems, including materials science^{1–3}, information science^{4,5}, computer science^{6,7}, mathematical science^{8–10}, and others. However, there are two major challenges to bringing quantum computers to materialize²: (1) it is difficult to access full information from entangled systems because of the state collapse upon measurements, and (2) it is difficult to build, control, and measure quantum states with arbitrarily high accuracy. In this regard, even though the current state-of-the-art quantum computers rely on the noisy intermediate-scale (NISQ devices,) which usually prevent high efficiency¹¹, various hybrid quantum-classical algorithms were proposed and actively studied recently¹², and that could be promising for quantum speedup in the regime of NISQ devices. Massive applications including variational quantum eigensolvers^{13–17}, quantum approximate optimization algorithms¹⁸, new frontiers in quantum foundations^{19–22}, and others, were reported.

Beyond the actively studied VQAs, the universal compilation has drawn tremendous interest recently. Its core idea relies on a training process to transform a trainable unitary into a target unitary^{23,24}. It was demonstrated in different applications from gate optimization²³, to quantum-assisted compiling process²⁴, continuous-variable quantum learning²⁵, and robust quantum compilation²⁶. The future of universal quantum compiling could be circuits depth-compression, black-box compiling, error mitigation, gate-fidelity benchmarking, and efficient gate synthesis.

In another aspect, quantum state tomography (QST) is a measurement process performed on numerous identical copies of a system to extract its state's information²⁷. In general, for a given unknown quantum state $|\psi\rangle$ in a complex Hilbert space of d -dimension, it requires an exponentially growing $2^d - 1$ measurements on different bases to completely reproduce the state, which is intractable for large systems. Numerous methods were proposed for improving the standard QST in terms of efficiency^{28–31}, methodology^{32–38}, quantum dynamic^{39–41}, and so on. Recently, the quantum circuits-based QST has attracted significant attention owing to the incredible advantages of the quantum device^{42–44}, which allows to efficiently prepare quantum states with high confidence, fully control the Hamiltonian for the state evolution, and directly access the measurement results. A variational approach⁴⁵ and single-shot measurements^{46,47} to name a few, were investigated.

Despite recent achievements on the QST, it is still challenging to implement in the NISQ devices. In this work, we introduce a promising application of the universal compilation on the QST. Our main idea is to use a trainable unitary acting upon a known fiducial state to reconstruct an unknown state, which is created by using a Haar random target unitary acting upon the fiducial state. The advantage of this method is that it requires low-depth trainable unitaries and few measurements to realize the target state, which significantly reduces the

¹University of Information Technology, Ho Chi Minh City 700000, Vietnam. ²Vietnam National University, Ho Chi Minh City 700000, Vietnam. ³Ho Chi Minh City Institute of Physics, National Institute of Applied Mechanics and Informatics, Vietnam Academy of Science and Technology, Ho Chi Minh City 700000, Vietnam. ⁴Frontier Research Institute for Interdisciplinary Sciences, Tohoku University, Sendai 980-8578, Japan. ⁵Department of Applied Physics, Graduate School of Engineering, Tohoku University, Sendai 980-8579, Japan. ✉email: binho@fris.tohoku.ac.jp

complexity and allows for tractability of large systems. Furthermore, the flexibility of the trainable unitaries is more elevated than that of the target unitaries, resulting in a better fault-tolerant capacity and thus allowing high efficiency for the trainable quantum circuits.

Concretely, we first introduce the general framework of the universal compilation-based quantum state tomography (UC-QST). We also introduce several gradient-based optimizers, including the standard gradient descent (SGD), the Adam, and the quantum natural gradient descent (QNG). We discuss the numerical experiment results for a representative case of single-qubit tomography, and then evaluate the reconstructing efficiency of unknown Haar random states via various popular circuit ansatzes. We find that the circuit depth plays a crucial role in the robust fidelity, i.e., by choosing a proper circuit depth via the number of layers in the quantum circuit, we get high fidelity at any qubit numbers. We finally compare the results with the shadow tomography method^{48,49}, a similar fashion in the field.

The study reveals that the accuracy mainly relies on (1) the ansatz topologies with the optimal circuit depth and (2) the significant impact of different optimizers. Our study can further promise applications in quantum metrology and sensing, and new frontier foundation aspects. Moreover, it is possible to implement the algorithm on near-term quantum computers, and thus it could be a valuable technique for verifying the fidelity of quantum circuits and studying various quantum computing tasks. These are also benefits that overcome the standard QST, which requires the set-up of traditional experiments, consumes heavy post-processing calculations to reproduce the quantum state, and the accuracy depends on the estimators, such as the Maximum-Likelihood and Least-Squares²⁷.

Results

We introduce a universal compilation scheme^{23–26} to translate a given state into another one and apply it to quantum state tomography.

Universal compilation-based quantum state tomography (UC-QST). A universal compilation scheme consists of a quantum part and a classical part, as shown in Fig. 1a. The quantum part is a circuit with parameterizable ansatzes. Let U is a fixed target unitary and $V^\dagger(\theta)$ is a trainable unitary ansatz (sets of quantum gates with some parameters θ) that act sequentially onto the circuit and transform an initial state $|\psi_0\rangle$ into a final state $|\psi_f\rangle$ as

$$|\psi_f\rangle = V^\dagger(\theta)U|\psi_0\rangle. \quad (1)$$

The transition probability yields

$$p(\psi_0 \rightarrow \psi_f) = |\langle\psi_0|\psi_f\rangle|^2 = |\langle\psi_0|V^\dagger(\theta)U|\psi_0\rangle|^2. \quad (2)$$

Our task is to maximize the transition probability $p_{\max}(\psi_0 \rightarrow \psi_f)$, such that a state $|\psi\rangle \equiv U|\psi_0\rangle$ is compiled to $|\phi(\theta)\rangle \equiv V(\theta)|\psi_0\rangle$. The maximization of the transition probability, i.e., $p(\psi_0 \rightarrow \psi_f) = 1$, implies $|\psi\rangle = |\phi(\theta)\rangle$, which can be applied to the QST as we will describe below.

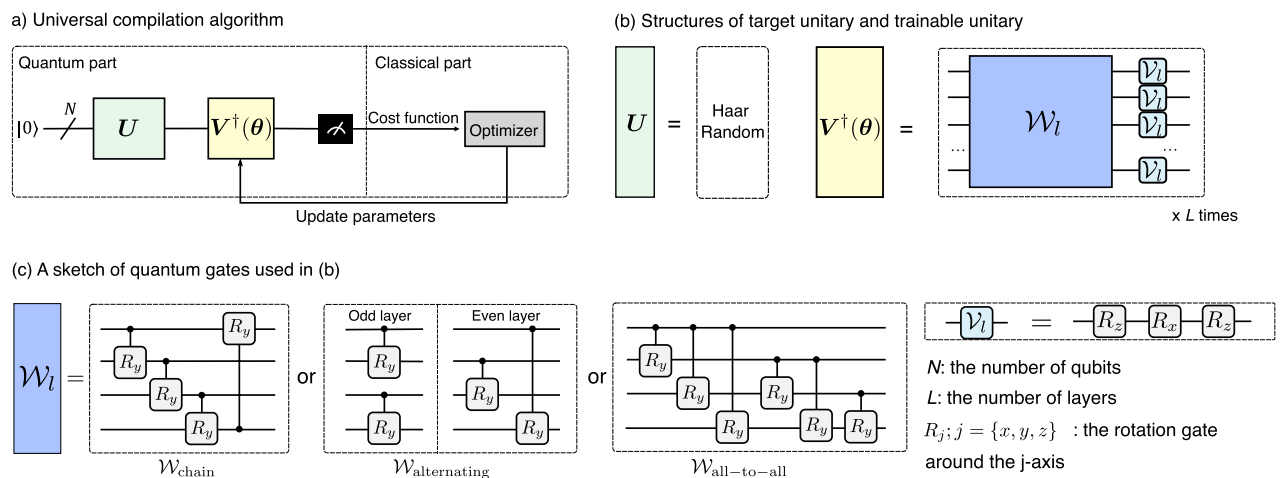


Figure 1. Universal compilation-based quantum state tomography. (a) A universal compilation algorithm consists of a quantum part and a classical part. In the quantum part, a final state is created by applying a set of quantum gates U followed by V^\dagger onto the initial circuit and then be measured. In the classical part, we compute the appropriate cost function, use an optimizer to compute new parameters, and update the scheme until it converges. (b) Structures of the target and trainable unitaries used in the QST. The unitary U is a Haar random generator while V^\dagger is parameterized into $V^\dagger(\theta)$ and broken out into entangled gates W and local rotation gates V with several structures as shown in the figure. (c) A sketch of some quantum gates used in (b). Other notations: N the number of qubits, L the number of layers, $R_j, j = \{x, y, z\}$ the rotation gate around the j axis.

Concretely for the QST, let $|\psi_0\rangle \equiv |0\rangle = |0\rangle^{\otimes N}$, where N is the number of qubits, we transform it into a random (unknown) quantum state $|\psi\rangle = U|0\rangle$, via a Haar random unitary U^{50} . To reconstruct this state, we apply a trainable unitary evolution $V^\dagger(\theta)$ that can learn the role of U , i.e., a reconstructed state $|\phi(\theta)\rangle = V(\theta)|0\rangle$ resembles to the unknown state $|\psi\rangle$, where $\theta = \{\theta_1, \theta_2, \dots, \theta_M\}$ can be adaptively updated during a training process, M is the number of trainable parameters. There is no free lunch for the choice of $V^\dagger(\theta)^{25}$. However, it can break out into a sequence of single-qubit and multi-qubit gates as

$$V^\dagger(\theta) = \prod_{l=1}^L \mathcal{V}_l(\theta_l) \mathcal{W}_l(\theta_l); \text{ with } \mathcal{V}_l = R_z R_x R_z, \quad (3)$$

as shown in Fig. 1b, wherein \mathcal{W}_l includes the chain, alternating, and all-to-all structures⁵¹ as shown in Fig. 1c. We emphasize that the entangled gates \mathcal{W} consist of two-qubit controlled y -rotation gates, which differs from previous works⁵¹. We refer to these gates as parameter-dependent entanglement gate. They are useful for preparing variational states metrology²¹ and rapid entangled circuits⁵², for testing of the expressibility and entangling capability⁵³, and so on.

To qualify how closed the two states are, we consider the Fubini-Study distance as⁵⁴

$$d(\psi, \phi(\theta)) = \sqrt{1 - |\langle \phi(\theta) | \psi \rangle|^2} = \sqrt{1 - p_0(\theta)}, \quad (4)$$

where $p_0(\theta) = |\langle \phi(\theta) | \psi \rangle|^2 = |\langle 0 | V^\dagger(\theta) U | 0 \rangle|^2$ is the probability for getting the outcome $|0\rangle$. In the quantum circuit, we apply a sequence of U followed by $V^\dagger(\theta)$ onto the initial state $|0\rangle$ to get the final state $V^\dagger(\theta)U|0\rangle$ and then measure a projective operator $P_0 = |0\rangle\langle 0|$, which yields the probability $p_0(\theta)$.

The variational (reconstructed) state becomes the target (unknown) state if the distance reaches zero. In the classical part, we thus use the Fubini-Study distance as a cost function that needs to minimize, i.e., $C(\theta) = d(\psi, \phi(\theta))$, such that

$$\theta^* = \underset{\theta}{\operatorname{argmin}} C(\theta). \quad (5)$$

By training the variational circuit until it converges, we obtain the optimal θ^* and the reconstructed state yields $|\phi(\theta^*)\rangle = V(\theta^*)|0\rangle$. This is a normalized pure state because $V(\theta^*)$ is a unitary ansatz, i.e., $V^\dagger(\theta^*)V(\theta^*) = I$.

For the training process, we apply gradient-based optimizations to iteratively update the parameters θ and minimize the cost function. We first compute the derivative $\partial_{\theta_j} C(\theta)$ for all $\theta_j \in \theta$ and then compute new parameters via various appropriate optimizers, including the Standard gradient descent (SGD), Adam gradient descent⁵⁵, and Quantum natural gradient (QNG)⁵⁶. See “Methods” section for details.

Numerical results. *Single-qubit QST.* We first consider reconstructing an abstract single-qubit state encodes in a quantum circuit as shown in the inset Fig. 2. We randomly generate an unknown quantum state $|\psi\rangle = U_3|0\rangle$, where

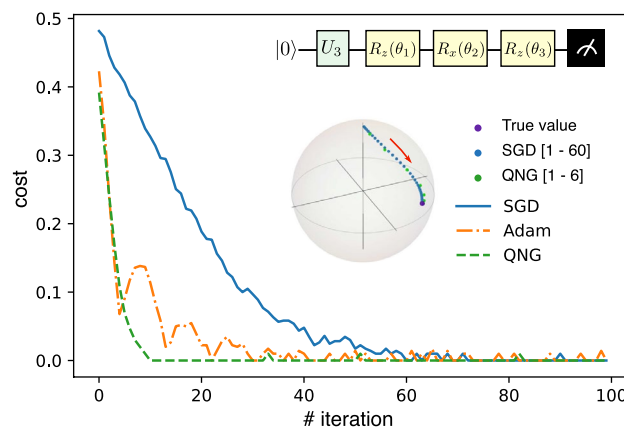


Figure 2. Single-qubit tomography. Cost function versus the number of iterations for difference optimizers: SGD (blue), Adam (orange), and QNG (green). Inset 1: quantum circuit for a single-qubit tomography where an unknown qubit state is generated by a random unitary $U_3(\theta, \phi, \lambda)$, and $V^\dagger(\theta)$ is made up of $R_z(\theta_1)$, $R_x(\theta_2)$, and $R_z(\theta_3)$ gates. Inset 2: Bloch sphere represents the qubit states: violet circle: the true state, blue circle: the trajectory of the reconstructed state under the SGD optimizer for the iterations run from 1 to 60, green circle: the trajectory under the QNG optimizer for the iteration runs from 1 to 6.

$$U_3(\theta, \phi, \lambda) = \begin{pmatrix} \cos \frac{\theta}{2} & -e^{i\lambda} \sin \frac{\theta}{2} \\ e^{i\phi} \sin \frac{\theta}{2} & e^{i(\phi+\theta)} \cos \frac{\theta}{2} \end{pmatrix}, \quad (6)$$

where we set random with Haar measure $\sin(\theta)/2$, ϕ , and λ . To reconstruct $|\psi\rangle$, we set the unitary $V^\dagger(\theta) = R_z(\theta_3)R_x(\theta_2)R_z(\theta_1)$. Indeed, a single-qubit rotation is $R_j(\theta) = \exp(-i\frac{\theta}{2}\sigma_j)$, $j \in \{x, y, z\}$, and σ_j is a Pauli matrix applied on the qubit. We train the scheme with 100 iterations using various optimizers and show the cost function versus iteration in the main (Fig. 2). Here, the QNG optimizer gives the best optimization. In the inset figure, we show the trajectory in the Bloch sphere of the reconstructed state $|\phi(\theta)\rangle$ under the updated of θ for two cases of SGD and QNG optimizers. The former needs around 60 iterations for the reconstructed state to reach the true state, while the latter only requires around 6 iterations to reach the same accuracy.

Haar random state QST. Now, we focus on a general random Haar state i.e., $|\psi\rangle = U_{\text{Haar}}|0\rangle$, as shown in Fig. 1b. To reconstruct the state, we use several ansatzes for the entangled gate \mathcal{W} in $V^\dagger(\theta)$, including the $\mathcal{W}_{\text{chain}}$, $\mathcal{W}_{\text{alternating}}$, and $\mathcal{W}_{\text{all-to-all}}$ structures. Refer Fig. 1c for the details of these structures, where we used the parameter-dependent controlled y -rotation gates to construct them. The circuit's depth for these structures are $(N+3)L$, $4L$, and $(N+2)L$, respectively. The trainable parameters are $M = 4NL$, $\lfloor NL/2 \rfloor + 3NL$, and $N(N+5)L/2$, respectively, which are grown linearly with N . This is suitable for NISQ devices even for the large number of qubits.

The results are shown in Fig. 3. Let us consider the fidelity between the true Haar state and the reconstructed state as

$$F(\psi, \phi(\theta)) = |\langle \phi(\theta) | \psi \rangle|^2, \quad (7)$$

which is the overlap between these two states. In Fig. 3a–c, we show the fidelities for different structures of \mathcal{W} . For each case, we fix $L = 2$ and examine the three optimizers SGD (blue circle), Adam (yellow square), and QNG (green triangle). We first observe that the SGD optimizer is not good for all \mathcal{W} structures and needs to choose an appropriate learning rate. The fidelities reduce with the increasing N and nearly vanish at $N = 6$. In contrast, the Adam optimizer exhibit high fidelities up to $N = 4$ for $\mathcal{W}_{\text{chain}}$ (a), $N = 5$ for $\mathcal{W}_{\text{all-to-all}}$ (c), and gradually reduces from $N = 2$ for $\mathcal{W}_{\text{alternating}}$ (b). Even though it is not stable near the optimal point, the Adam is remarkable for achieving high accuracy in the QST. Furthermore, the QNG optimizer also allows for getting such high accuracy up to $N = 4$ for $\mathcal{W}_{\text{chain}}$ (a) and even better than the Adam for $\mathcal{W}_{\text{alternating}}$ (b), while it gradually reduces for $\mathcal{W}_{\text{all-to-all}}$ (c). This observation can be explained by these own structures: the $\mathcal{W}_{\text{all-to-all}}$ contains the most number of parameters via the controlled y -rotation gates compared to the others, which results in the low accuracy. It is apparent that the QNG optimizer is sensitive to the controlled y -rotation gates, where the more controlled y -rotation gates, the less efficient QNG optimizer.

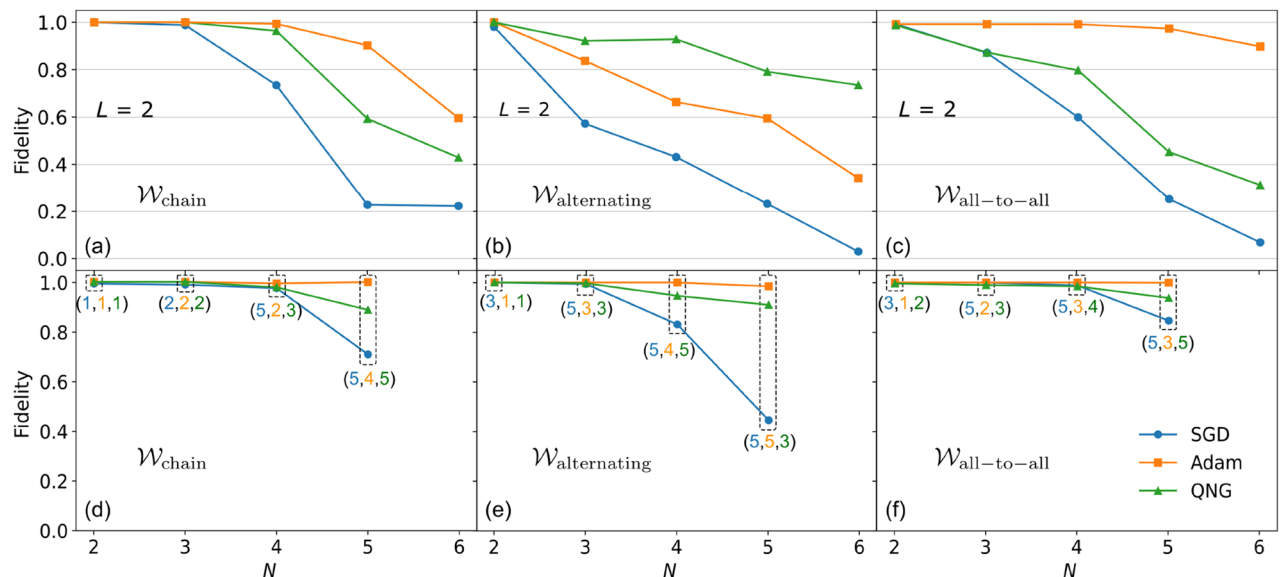


Figure 3. Numerical results for quantum state tomography. (a–c) Plot of the fidelity between an unknown Haar random state $|\psi\rangle$ and its reconstructed state $|\phi(\theta)\rangle$ for different \mathcal{W} structures: $\mathcal{W}_{\text{chain}}$ (a), $\mathcal{W}_{\text{alternating}}$ (b), and $\mathcal{W}_{\text{all-to-all}}$ (c). For each case, we show the results for different optimizers: SGD, Adam, and QNG. Here we fixed $L = 2$. (d–f) Plot of the fidelity similar as above for different L as shown in the colored parentheses (blue star, yellow dagger, green double dagger). Here, blue star is the optimal number of layers for the SGD, yellow dagger is the optimal number of layers for the Adam, and green double dagger is the optimal number of layers for the QNG. We choose the appropriate (blue star, yellow dagger, green double dagger) for each N and each \mathcal{W} structure so that the fidelity gets its (possible) highest accuracy.

Next, to achieve high accuracy for any qubit numbers N , we increase the number of layers L , while paying attention to the barren plateau^{57–61}, i.e., the accuracy of the training process reduces when increasing the parameters space. Figure 3d–f plot the fidelities versus N , where for each N , the corresponding L is shown in the colored parenthesis (blue star, yellow dagger, green double dagger), for the SGD, Adam, and QNG, respectively. The number of layers shown in the parenthesis is the smallest (optimal) L required for achieving such high accuracy before it goes down due to the barren plateau. As can be seen from the figure, the Adam method allows for reaching the maximum fidelity (results are shown up to $N = 5$ for all \mathcal{W} structures) with a suitable L as shown in the middle position of the parenthesis. Similarly, we can reach high accuracy with the QNG optimizer up to $N = 4$ when choosing an appropriate L as shown in the last position of the parenthesis. For the SGD, it is intractable for achieving high accuracy, such as for $\mathcal{W}_{\text{alternating}}$. Even though the relation between N and the required L is not clear, interestingly, we can see from the results up to $N = 5$, the required L is also around 5 (more L is redundancy or may reduce the accuracy due to the barren plateau, see details in “Methods” section).

We only simulate up to $N = 5$. However, for larger N , the scheme still works well. Evidently, in Fig. 3d–f, we enhance high fidelity with an appropriate optimizer for every N up to 5. Following the procedure in “Methods” section, we can entirely expand to a larger N while still maintaining high fidelity.

Compare to the shadow tomography protocol. Finally, we address the merit of our UC-QST approach and the shadow tomography protocol^{48,49}, a recent promising method in this regime. A shadow tomography protocol is given as follows⁴⁹: (1) initially prepare a random unknown quantum state ρ , and the task ahead is to predict a target function underlying the state from its shadow, (2) randomly pick up a unitary U_k in a T -tuple \mathcal{U} , i.e., $\mathcal{U} = \{U_1, U_2, \dots, U_T\}$ then apply it to the initial state to transform $\rho \mapsto U_k \rho U_k^\dagger$, (3) measure the evolved state in the computational basis $|b\rangle = \{|0\rangle, |1\rangle\}^N$. Steps (2) and (3) are repeated for a certain number of measurements. For each measurement, we get a random classical snapshot

$$\sigma_{k,b} = U_k^\dagger |b\rangle \langle b| U_k. \quad (8)$$

We then define an invertible channel matrix

$$\mathcal{M}(\rho) = \mathbb{E}_k \sum_b \text{Tr}(\sigma_{k,b} \rho) \cdot \sigma_{k,b}, \quad (9)$$

where \mathbb{E}_k is the average over U_k , with a corresponding pick-up probability. Let \mathcal{M}^{-1} exists, and let p_k is the probability of picking up a unitary U_k , then we can reconstruct a (non-normalized) state as

$$\check{\rho} = \sum_k p_k \sum_b \text{Tr}(\sigma_{k,b} \rho) \cdot \mathcal{M}^{-1}(\sigma_{k,b}), \quad (10)$$

which is the classical shadow of the original unknown state ρ . For the transformation U belongs to a family of the global Clifford gates, i.e., $U \in \mathcal{U}_C = \{\text{CNOT}, \text{Hadamard}, \text{S_gate}, \text{T_gate}\}$, refer to Random Clifford measurements, the reconstructed state explicitly yields⁴⁹

$$\check{\rho} = (2^N + 1) U^\dagger |b\rangle \langle b| U - I. \quad (11)$$

For the transformation U belongs to the random Pauli gates, such as $U \in \mathcal{U}_P = \{\sigma_x, \sigma_y, \sigma_z, \dots\}$, refer to Random Pauli measurements, it straightforwardly yields⁴⁹

$$\check{\rho} = \bigotimes_{j=1}^N \left(3 U_j^\dagger |b_j\rangle \langle b_j| U_j - I \right), \quad (12)$$

for $b = (b_1, \dots, b_N) \in \{0, 1\}^N$.

For comparing the shadow tomography with the UC-QST scheme, we apply the Random Pauli measurements and consider the prediction of a linear function as a figure of merit for the accuracy. A global observable $Z \equiv \sigma_z^{\otimes N}$, gives the predicted (linear) expectation value as

$$\check{z} = \text{Tr}(Z \check{\rho}), \text{ that obeys } \mathbb{E}[\check{z}] = \text{Tr}(Z \rho). \quad (13)$$

The fluctuation (distribution around the true expectation value) of the predicted expectation value is given by the variance $\text{Var}[\check{z}]$ as

$$\text{Var}[\check{z}] = \mathbb{E}[(\check{z} - \mathbb{E}[\check{z}])^2] = [(\text{Tr}(Z \check{\rho}) - \text{Tr}(Z \rho))^2]. \quad (14)$$

In Fig. 4a, we show the variance $\text{Var}[\check{z}]$ as a function of the number of measurements for the shadow tomography. The variance slightly decreases when increasing the number of measurements from 10^2 to 10^5 . See the inset figure for the detailed zoom-in. The result is compared with the standard quantum limit (SQL), i.e., $\text{SQL} = 1/\text{\#measurement}$, and the Heisenberg limit (HL), i.e., $\text{HL} = 1/(\text{\#measurement})^2$. Here, the variance does not beat the SQL nor HL.

In Fig. 4b, we compare the variances obtained from the shadow tomography and the UC-QST for the different number of qubits N . For the shadow tomography, we fix $\text{\#measurement} = 10^5$. For the UC-QST, we consider the

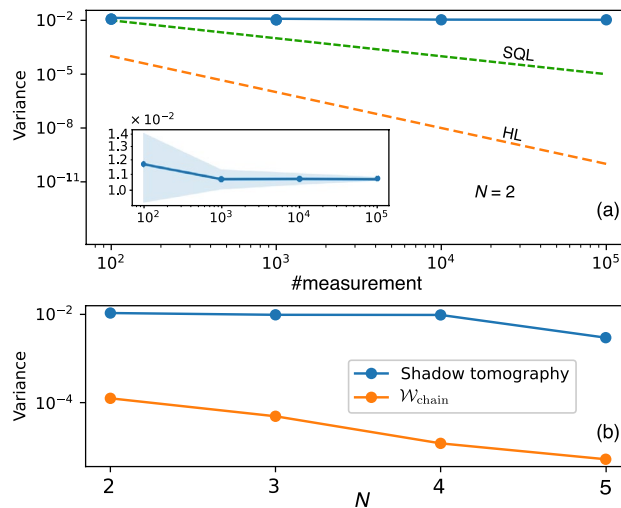


Figure 4. Comparison between the UC-QST approach and the shadow tomography method. **(a)** Log-log plot of the variance $\text{Var}[\hat{z}]$ versus the number of repeated measurements (#measurement) using the shadow tomography. The standard quantum limit (SQL) and Heisenberg limit (HL) are shown for comparison purposes. The number of qubits is fixed at $N = 2$. Inset: Zoom-in the variance, where the blue area is the standard deviation after talking ten runs. **(b)** Plot of variance for the shadow tomography (blue) and UC-QST (orange). For the shadow tomography, we fix #measurement = 10^5 , and for the UC-QST, we use the structure of $\mathcal{W}_{\text{chain}}$ and the Adam optimizer, the number of shots is 10^4 .

$\mathcal{W}_{\text{chain}}$ structure with the Adam optimizer as an example. The number of shots is fixed at 10^4 . It can be seen that the UC-QST offers a better result over 100 times than the resulting shadow tomography.

Discussion

We discuss some features of the UC-QST and shadow tomography in the previous section. The shadow tomography only allows predicting target functions, such as expectation values, entanglement entropies, correlation functions, and so on⁴⁹, while maintaining the precision. Whereas the UC-QST allows for reconstructing the entire quantum state up to a phase shift. Both schemes allow for predicting properties of quantum states or quantum states with fewer measurements compared to standard quantum tomography. Another remarkable feature is that the efficiency of the shadow tomography protocol depends on the random choice of the unitaries in an ensemble \mathcal{U} , while the efficiency of the UC-QST scheme relies on the choice of different ansatzes and optimizers. Finally, we emphasize that the comparison in this section only provides a very first glance about the two approaches. We need to further characterize these features in future works for more concrete evaluation.

Furthermore, the idea of UC-QST resembles the self-guided quantum tomography (SGQT)⁶² and single-shot measurement (SSM)^{46,47}. These previous approaches also iteratively search the estimated state to converge to the true state. However, the trainable unitary topologies and optimization algorithms are different. The SGQT uses a simultaneous perturbation stochastic approximation⁶³ to optimize the cost function, while the SSM trains a learning unitary to be a fiducial (known) state that converges to the true state. Here, we apply the universal compilation technique to train a learning unitary to be the target unitary.

Finally, we emphasize that the current method is suitable only for pure-state tomography and not for general mixed-state tomography.

Methods

Training process. The training process is a hybrid protocol as illustrated in Fig. 1a: a set of unitary gates U followed by V^\dagger are applied onto the circuit and the final state is measured afterwards. The results are sent to the classical counterpart to compute the corresponding cost function and then update new parameters θ using a suitable optimizer protocol until it reaches convergence.

We use gradient-based optimizations to iteratively update the parameters θ and minimize the cost function. To do that, we need to calculate the derivative $\partial C(\theta)/\partial \theta_j$ w.r.t θ_j in the j th gate for every $\theta_j \in \theta$. We compute two cases as follows. First, if the j th gate is a single-qubit rotation gate, i.e., $\exp(-i\theta_j \sigma_k/2)$, $k \in \{x, y, z\}$, then using the standard (two-term) parameter-shift rule^{64,65}, we have

$$\begin{aligned} \frac{\partial C(\theta)}{\partial \theta_j} &= -\frac{1}{2C(\theta)} \frac{\partial p_0(\theta)}{\partial \theta_j} \\ &= -\frac{1}{2C(\theta)} \frac{1}{2 \sin(s)} [p_0(\theta + s e_j) - p_0(\theta - s e_j)], \end{aligned} \quad (15)$$

where s denotes an arbitrary shift, and e_j is the j th unit vector, or in other words, we only add s to θ_j . Second, if the j th gate is a controlled rotation gate, i.e., $CR_y(\theta_j)$, then using the four-term parameter-shift rule⁶⁶, we partially compute

$$\frac{\partial p_0(\theta)}{\partial \theta_j} = d_+ \left[p_0(\theta + ae_j) - p_0(\theta - ae_j) \right] - d_- \left[p_0(\theta + be_j) - p_0(\theta - be_j) \right], \quad (16)$$

where $d_{\pm} = (\sqrt{2} \pm 1)/4\sqrt{2}$; $a = \pi/2$; $b = 3\pi/2$. Then, we get $\frac{\partial C(\theta)}{\partial \theta_j} = -\frac{1}{2C(\theta)} \frac{\partial p_0(\theta)}{\partial \theta_j}$.

To compute new parameters, we use several optimizers in all experiments: Standard gradient descent (SGD), Adam gradient descent⁵⁵, and Quantum natural gradient (QNG)⁵⁶.

The formula for SGD reads

$$\theta^{t+1} = \theta^t - \alpha \nabla_{\theta} C(\theta), \quad (17)$$

where $\nabla_{\theta} C(\theta) = (\partial_{\theta_1} C(\theta), \partial_{\theta_2} C(\theta), \dots, \partial_{\theta_M} C(\theta))^T$ for M training parameters, and α is the learning rate. In comparison, Adam is a non-local averaging optimizer that allows adapting the learning rate but requires more steps than the SGD

$$\theta^{t+1} = \theta^t - \alpha \frac{\hat{m}_t}{\sqrt{\hat{v}_t + \epsilon}}, \quad (18)$$

where $m_t = \beta_1 m_{t-1} + (1 - \beta_1) \nabla_{\theta} C(\theta)$, $v_t = \beta_2 v_{t-1} + (1 - \beta_2) \nabla_{\theta}^2 C(\theta)$, $\hat{m}_t = m_t / (1 - \beta_1^t)$, $\hat{v}_t = v_t / (1 - \beta_2^t)$, with the hyper-parameters are chosen as $\alpha = 0.2$, $\beta_1 = 0.8$, $\beta_2 = 0.999$ and $\epsilon = 10^{-8}$. Finally, the QNG is defined by

$$\theta^{t+1} = \theta^t - \alpha g^+ \nabla_{\theta} C(\theta), \quad (19)$$

where g^+ is the pseudo-inverse of a Fubini-Study metric tensor⁶⁷. Assume that we can group θ into \mathcal{L} layers, i.e., $\theta = \theta^{(1)} \oplus \theta^{(2)} \oplus \dots \oplus \theta^{(\mathcal{L})}$, so that in each layer $\theta^{(\ell)} = \{\theta_1^{(\ell)}, \theta_2^{(\ell)}, \dots, \theta_{M^{(\ell)}}^{(\ell)} \mid \sum_{\ell} M^{(\ell)} = M\}$, any two of unitaries satisfy $[G_i^{(\ell)}, G_j^{(\ell)}] = \delta_{ij}$. Then, the metric tensor g gives⁶⁸

$$g = \begin{pmatrix} [g^{(1)}] & & & \mathbf{0} \\ & [g^{(2)}] & & \\ & & \ddots & \\ \mathbf{0} & & & [g^{(\mathcal{L})}] \end{pmatrix} \quad (20)$$

where an element $g_{ij}^{(\ell)}$ of $g^{(\ell)}$ reads

$$g_{ij}^{(\ell)} = \text{Re}[\langle \partial_i \psi_{\ell} | \partial_j \psi_{\ell} \rangle - \langle \partial_i \psi_{\ell} | \psi_{\ell} \rangle \langle \partial_j \psi_{\ell} | \psi_{\ell} \rangle], \quad (21)$$

where $|\psi_{\ell}\rangle$ is the quantum state at the ℓ th layer. For unitary $G_i^{(\ell)} = e^{-i\theta_i^{(\ell)} K_i^{(\ell)}}$, e.g., a rotation gate, such that $[G_i^{(\ell)}, K_i^{(\ell)}] = 0$, then $g_{ij}^{(\ell)}$ is recast as⁶⁸

$$g_{ij}^{(\ell)} = \text{Re}[\langle \psi_{\ell-1} | K_i K_j | \psi_{\ell-1} \rangle - \langle \psi_{\ell-1} | K_i | \psi_{\ell-1} \rangle \langle \psi_{\ell-1} | K_j | \psi_{\ell-1} \rangle]. \quad (22)$$

See a detailed example of computing a tensor metric g below.

Each optimizer has its own pros and cons: (1) the SGD is simple but low coverage, one must choose a proper learning rate to achieve the best result, (2) the Adam allows to automatically adapt the learning rate and fast coverage but it is noisy near the optimal point, and (3) the QNG is better than other optimizers but also requires more computational cost regards to quantum circuits. While the SGD and Adam do not depend on quantum states and work for any classical data types, including the probabilities, the QNG optimizes the parameters towards the geometry of evolved quantum states and is thus expected to offer better and faster optimization. We conduct these optimizers based on their advantages and disadvantages and compare the results. They also serve as a test bed and reference for future works.

This work implements the numerical experiments using various configurations described above to train the variational models and compare them together. The numerical results are executed by Qiskit open-source package, version 0.24.0, which is available to run on all platforms. For each experiment, to get the probability p_0 we execute 10^4 shots using the *qasm* simulator backend. The number of iterations for every training process is fixed at 400, except for others shown in the text. It is sufficient for the cost function to converge for all data shown in the text. The experiments are then scaled up to 6 qubits for quantum state tomography to demonstrate the scalability. Furthermore, after the training process, we can reproduce the unknown state by applying $V(\theta^*)$ into the initial state $|\mathbf{0}\rangle$, and use it for further applications and other statistical computations.

Complexity. In terms of complexity, to execute the parameter-shift rule in Eq. (15), the quantum circuit executes $2M + 1$ times, one M times to compute $p_0(\theta + se_i)$, one M times to compute $p_0(\theta - se_i)$, and one time to compute $p_0(\theta)$. Furthermore, a single evaluation requires executing the circuit for a constant number of shots to reach a certain precision, and each execution involves around G gate operations. So, the complexity of each iteration is $\mathcal{O}[(2M + 1)G]$. Similarly, the complexity for an iteration with four-term parameter-shift rule is $\mathcal{O}[(4M + 1)G]$.

Ideally, after each step, the cost function will decrease with a linear or logarithmic speed regarding the number of iterations. However, the variational circuit always offers a lower bound of the cost function during the training process. In particular, this bound increases by the number of qubits N , which means the problem will be harder according to the size of the system

$$C(\theta) \geq \text{poly}(N). \quad (23)$$

The complexity of the ansatz $V(\theta)$ is another challenge. Its current structure is fixed into the chain, alternating, and all-to-all. However, the structure also needs to optimize in future works, e.g., using Genetic Algorithms for generating a compressed ansatz $V(\theta)$ that can work well on the current NISQ devices for the large number of qubits.

Fubini-Study tensor metric. We provide a practical example of how to compute a Fubini-Study tensor metric. Let us consider a concrete circuit as shown in Fig. 5. It consists of $R_x = \exp(-i\frac{\theta_x}{2}\sigma_x)$, $R_z = \exp(-i\frac{\theta_z}{2}\sigma_z)$, and $CR_y = |0\rangle\langle 0| \otimes I_2 + |1\rangle\langle 1| \otimes \exp(-i\frac{\theta_y}{2}\sigma_y)$. Since $[R_x, R_z] = 0$ (because they act on different qubits), we can group them into one layer (layer 1), with $\theta^{(1)} = \{\theta_0^{(1)}, \theta_1^{(1)}\} = \{\theta_x, \theta_z\}$, and put CR_y into another layer (layer 2), with $\theta^{(2)} = \{\theta_0^{(2)}\} = \{\theta_y\}$. The tensor metric g explicitly yields

$$g = \begin{pmatrix} g_{xx}^{(1)} & g_{xz}^{(1)} & 0 \\ g_{zx}^{(1)} & g_{zz}^{(1)} & 0 \\ 0 & 0 & g_{yy}^{(2)} \end{pmatrix}. \quad (24)$$

The quantum states are explicitly expressed as

$$|\psi_0\rangle = |00\rangle, \quad (25)$$

$$|\psi_1\rangle = e^{-i\frac{\theta_x}{2}\sigma_x \otimes I_2} e^{-i\frac{\theta_z}{2}I_2 \otimes \sigma_z} |\psi_0\rangle, \quad (26)$$

$$|\psi_2\rangle = [|0\rangle\langle 0| \otimes I_2 + |1\rangle\langle 1| \otimes e^{-i\frac{\theta_y}{2}\sigma_y}] |\psi_1\rangle. \quad (27)$$

The elements $g_{ij}^{(1)}$ is given through Eq. (22) as

$$\begin{aligned} g_{xx}^{(1)} &= \langle \psi_0 | K_x^2 | \psi_0 \rangle - \langle \psi_0 | K_x | \psi_0 \rangle^2 = \frac{1}{4}, \\ g_{xz}^{(1)} &= \langle \psi_0 | K_x K_z | \psi_0 \rangle - \langle \psi_0 | K_x | \psi_0 \rangle \langle \psi_0 | K_z | \psi_0 \rangle = 0, \\ g_{zx}^{(1)} &= \langle \psi_0 | K_z K_x | \psi_0 \rangle - \langle \psi_0 | K_z | \psi_0 \rangle \langle \psi_0 | K_x | \psi_0 \rangle = 0, \\ g_{zz}^{(1)} &= \langle \psi_0 | K_z^2 | \psi_0 \rangle - \langle \psi_0 | K_z | \psi_0 \rangle^2 = 0, \end{aligned}$$

where $K_x = \frac{\sigma_x \otimes I_2}{2}$ and $K_z = \frac{I_2 \otimes \sigma_z}{2}$.

Next, we calculate $g_{yy}^{(2)}$. Starting from Eq. (21) in the main text, we derive

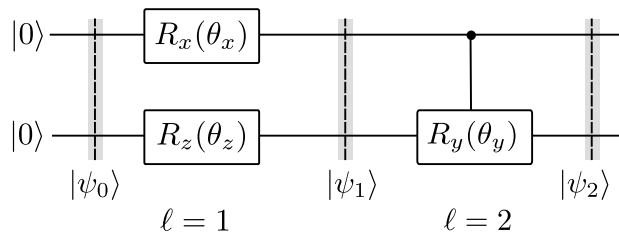


Figure 5. An example quantum circuit for evaluating the Fubini-Study tensor metric. The circuit starts with the initial state $|\psi_0\rangle = |00\rangle$ and evolves under a set of gates $R_x(\theta_x)$, $R_z(\theta_z)$ in layer 1, and $CR_y(\theta_y)$ in layer 2. Detailed of the Fubini-Study tensor metric evaluation for this circuit is provided in the text.

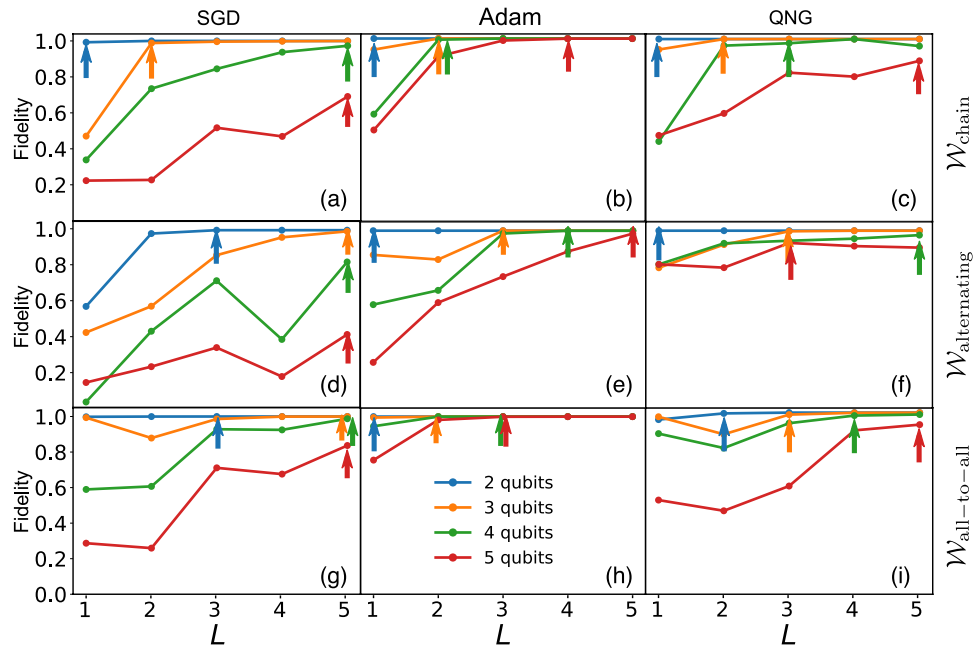


Figure 6. Plot of fidelity as a function of L for different \mathcal{W} structures and different optimizers. (a) ($\mathcal{W}_{\text{chain}}$, SGD), (b) ($\mathcal{W}_{\text{chain}}$, Adam), (c) ($\mathcal{W}_{\text{chain}}$, QNG), (d) ($\mathcal{W}_{\text{alternating}}$, SGD), (e) ($\mathcal{W}_{\text{alternating}}$, Adam), (f) ($\mathcal{W}_{\text{all-to-all}}$, QNG), (g) ($\mathcal{W}_{\text{all-to-all}}$, SGD), (h) ($\mathcal{W}_{\text{all-to-all}}$, Adam), (i) ($\mathcal{W}_{\text{all-to-all}}$, QNG).

$$|\partial_{\theta_y} \psi_2\rangle = -i|1\rangle \left\langle 1 \otimes \frac{\sigma_y}{2} e^{-i\frac{\theta_y}{2} \sigma_y} |\psi_1\rangle \right\rangle. \quad (28)$$

Then, we get

$$\begin{aligned} g_{yy}^{(2)} &= \langle \psi_1 | K_y^2 | \psi_1 \rangle - \langle \psi_1 | K_y | \psi_1 \rangle^2 \\ &= \frac{1}{4} \sin^2 \left(\frac{\theta_x}{2} \right), \end{aligned} \quad (29)$$

where $K_y = |1\rangle\langle 1| \otimes \frac{\sigma_y}{2}$. To derive expectation values in Eq. (29), we prepare $|\psi_1\rangle$ as in Fig. 5, then measure $\langle \psi_1 | K_y^2 | \psi_1 \rangle = \frac{1}{4} \langle \psi_1 | (|1\rangle\langle 1| \otimes \mathbf{I}_2) | \psi_1 \rangle$ and $\langle \psi_1 | K_y | \psi_1 \rangle = \frac{1}{2} \langle \psi_1 | (|1\rangle\langle 1| \otimes \sigma_y) | \psi_1 \rangle$. Finally, we obtain the tensor metric g

$$g = \begin{pmatrix} \frac{1}{4} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \frac{1}{4} \sin^2 \left(\frac{\theta_x}{2} \right) \end{pmatrix}. \quad (30)$$

Supported data for QST. We discuss more data supporting the results in Fig. 3d–f in the main text. As we discussed above, the accuracy can be improved when increasing the number of layers L . However, we cannot increase L arbitrarily large and need to stop at an optimal point. We define the optimal L as the smallest number of layers that, at the next layer, the accuracy saturates or starts to reduce. In Fig. 6 below, we discuss the optimal L for various cases, where we mark the optimal L with colored arrows. See also Table 1 below.

From the results here, we trace out the optimal L as shown in Fig. 3 in the main text.

Structure	Optimizer	N			
		2 (♠)	3 (♠)	4 (♠)	5 (♠)
$\mathcal{W}_{\text{chain}}$	SGD	$L = 1$	$L = 2$	$L = 5$	$L = 5$
	Adam	$L = 1$	$L = 2$	$L = 2$	$L = 4$
	QNG	$L = 1$	$L = 2$	$L = 3$	$L = 5$
	in Fig 3d	(1 1 1)	(2 2 2)	(5 2 3)	(5 4 5)
$\mathcal{W}_{\text{alternating}}$	SGD	$L = 3$	$L = 5$	$L = 5$	$L = 5$
	Adam	$L = 1$	$L = 3$	$L = 4$	$L = 5$
	QNG	$L = 1$	$L = 3$	$L = 5$	$L = 3$
	In Fig. 3e	(3 1 1)	(5 3 3)	(5 4 5)	(5 5 3)
$\mathcal{W}_{\text{all-to-all}}$	SGD	$L = 3$	$L = 5$	$L = 5$	$L = 5$
	Adam	$L = 1$	$L = 2$	$L = 3$	$L = 3$
	QNG	$L = 2$	$L = 3$	$L = 4$	$L = 5$
	In Fig. 3f	(3 1 2)	(5 2 3)	(5 3 4)	(5 3 5)

Table 1. Number of optimal layers L taking from Fig. 6.

Data availability

Data are available from the corresponding authors upon reasonable request.

Code availability

All codes used to produce the findings of this study are available at: <https://github.com/vutuanhai237/UC-VQA>.

Received: 3 October 2022; Accepted: 3 March 2023

Published online: 06 March 2023

References

- de Leon, N. P. *et al.* Materials challenges and opportunities for quantum computing hardware. *Science* **372**, eabb2823. <https://doi.org/10.1126/science.abb2823> (2021).
- Alexeev, Y. *et al.* Quantum computer systems for scientific discovery. *PRX Quantum* **2**, 017001. <https://doi.org/10.1103/PRXQuantum.2.017001> (2021).
- Ebadi, S. *et al.* Quantum phases of matter on a 256-atom programmable quantum simulator. *Nature* **595**, 227–232. <https://doi.org/10.1038/s41586-021-03582-4> (2021).
- Pirandola, S., Eisert, J., Weedbrook, C., Furusawa, A. & Braunstein, S. L. Advances in quantum teleportation. *Nat. Photon.* **9**, 641–652. <https://doi.org/10.1038/nphoton.2015.154> (2015).
- Spiller, T. P. Quantum information technology. *Mater. Today* **6**, 30–36. [https://doi.org/10.1016/S1369-7021\(03\)00130-5](https://doi.org/10.1016/S1369-7021(03)00130-5) (2003).
- Shor, P. Algorithms for quantum computation: Discrete logarithms and factoring. In *Proceedings 35th Annual Symposium on Foundations of Computer Science*, 124–134. <https://doi.org/10.1109/SFCS.1994.365700> (1994).
- Grover, L. K. A fast quantum mechanical algorithm for database search. In *Proceedings of the Twenty-Eighth Annual ACM Symposium on Theory of Computing*, 212–219 (1996).
- Harrow, A. W., Hassidim, A. & Lloyd, S. Quantum algorithm for linear systems of equations. *Phys. Rev. Lett.* **103**, 150502. <https://doi.org/10.1103/PhysRevLett.103.150502> (2009).
- Xu, X., Benjamin, S. C. & Yuan, X. Variational circuit compiler for quantum error correction. *Phys. Rev. Appl.* **15**, 034068. <https://doi.org/10.1103/PhysRevApplied.15.034068> (2021).
- Lubasch, M., Joo, J., Moinier, P., Kiffner, M. & Jaksch, D. Variational quantum algorithms for nonlinear problems. *Phys. Rev. A* **101**, 010301. <https://doi.org/10.1103/PhysRevA.101.010301> (2020).
- Preskill, J. Quantum computing in the NISQ era and beyond. *Quantum* **2**, 79. <https://doi.org/10.22331/q-2018-08-06-79> (2018).
- Cerezo, M. *et al.* Variational quantum algorithms. *Nat. Rev. Phys.* **3**, 625–644. <https://doi.org/10.1038/s42254-021-00348-9> (2021).
- Peruzzo, A. *et al.* A variational eigenvalue solver on a photonic quantum processor. *Nat. Commun.* **5**, 4213. <https://doi.org/10.1038/ncomms5213> (2014).
- Nakanishi, K. M., Mitarai, K. & Fujii, K. Subspace-search variational quantum eigensolver for excited states. *Phys. Rev. Res.* **1**, 033062. <https://doi.org/10.1103/PhysRevResearch.1.033062> (2019).
- Kirby, W. M., Tranter, A. & Love, P. J. Contextual subspace variational quantum eigensolver. *Quantum* **5**, 456. <https://doi.org/10.22331/q-2021-05-14-456> (2021).
- Gard, B. T. *et al.* Efficient symmetry-preserving state preparation circuits for the variational quantum eigensolver algorithm. *NPJ Quantum Inf.* **6**, 10. <https://doi.org/10.1038/s41534-019-0240-1> (2020).
- Tkachenko, N. V. *et al.* Correlation-informed permutation of qubits for reducing ansatz depth in the variational quantum eigensolver. *PRX Quantum* **2**, 020337. <https://doi.org/10.1103/PRXQuantum.2.020337> (2021).
- Zhou, L., Wang, S.-T., Choi, S., Pichler, H. & Lukin, M. D. Quantum approximate optimization algorithm: Performance, mechanism, and implementation on near-term devices. *Phys. Rev. X* **10**, 021067. <https://doi.org/10.1103/PhysRevX.10.021067> (2020).
- Arrasmith, A., Cincio, L., Sornborger, A. T., Zurek, W. H. & Coles, P. J. Variational consistent histories as a hybrid algorithm for quantum foundations. *Nat. Commun.* **10**, 3438. <https://doi.org/10.1038/s41467-019-11417-0> (2019).
- Kaibuegger, R. *et al.* Variational spin-squeezing algorithms on programmable quantum sensors. *Phys. Rev. Lett.* **123**, 260505. <https://doi.org/10.1103/PhysRevLett.123.260505> (2019).
- Koczor, B., Endo, S., Jones, T., Matsuzaki, Y. & Benjamin, S. C. Variational-state quantum metrology. *New J. Phys.* **22**, 083038. <https://doi.org/10.1088/1367-2630/ab965e> (2020).
- Meyer, J. J., Borregaard, J. & Eisert, J. A variational toolbox for quantum multi-parameter estimation. *NPJ Quantum Inf.* **7**, 89. <https://doi.org/10.1038/s41534-021-00425-y> (2021).
- Heya, K., Suzuki, Y., Nakamura, Y. & Fujii, K. Variational quantum gate optimization. *arXiv:1810.12745* (arXiv preprint) (2018).

24. Khatri, S. *et al.* Quantum-assisted quantum compiling. *Quantum* **3**, 140. <https://doi.org/10.22331/q-2019-05-13-140> (2019).
25. Volkoff, T., Holmes, Z. & Sornborger, A. Universal compiling and (no-)free-lunch theorems for continuous-variable quantum learning. *PRX Quantum* **2**, 040327. <https://doi.org/10.1103/PRXQuantum.2.040327> (2021).
26. Jones, T. & Benjamin, S. C. Robust quantum compilation and circuit optimisation via energy minimisation. *Quantum* **6**, 628. <https://doi.org/10.22331/q-2022-01-24-628> (2022).
27. Paris, M. & Rehacek, J. *Quantum State Estimation* Vol. 649 (Springer, 2004).
28. Palmieri, A. M. *et al.* Experimental neural network enhanced quantum tomography. *NPJ Quantum Inf.* **6**, 20. <https://doi.org/10.1038/s41534-020-0248-6> (2020).
29. Cramer, M. *et al.* Efficient quantum state tomography. *Nat. Commun.* **1**, 149. <https://doi.org/10.1038/ncomms1147> (2010).
30. Jackson, C. & van Enk, S. J. Detecting correlated errors in state-preparation-and-measurement tomography. *Phys. Rev. A* **92**, 042312. <https://doi.org/10.1103/PhysRevA.92.042312> (2015).
31. Moroder, T. *et al.* Permutationally invariant state reconstruction. *New J. Phys.* **14**, 105001. <https://doi.org/10.1088/1367-2630/14/10/105001> (2012).
32. Ahmed, S., Sánchez Muñoz, C., Nori, F. & Kockum, A. F. Classification and reconstruction of optical quantum states with deep neural networks. *Phys. Rev. Res.* **3**, 033278. <https://doi.org/10.1103/PhysRevResearch.3.033278> (2021).
33. Tóth, G. *et al.* Permutationally invariant quantum tomography. *Phys. Rev. Lett.* **105**, 250403. <https://doi.org/10.1103/PhysRevLett.105.250403> (2010).
34. Torlai, G. *et al.* Neural-network quantum state tomography. *Nat. Phys.* **14**, 447–450 (2018).
35. Blume-Kohout, R. Optimal, reliable estimation of quantum states. *New J. Phys.* **12**, 043034. <https://doi.org/10.1088/1367-2630/12/4/043034> (2010).
36. Fiderer, L. J., Schuff, J. & Braun, D. Neural-network heuristics for adaptive Bayesian quantum estimation. *PRX Quantum* **2**, 020303. <https://doi.org/10.1103/PRXQuantum.2.020303> (2021).
37. Gross, D., Liu, Y.-K., Flammia, S. T., Becker, S. & Eisert, J. Quantum state tomography via compressed sensing. *Phys. Rev. Lett.* **105**, 150401. <https://doi.org/10.1103/PhysRevLett.105.150401> (2010).
38. Flammia, S. T., Gross, D., Liu, Y.-K. & Eisert, J. Quantum tomography via compressed sensing: Error bounds, sample complexity and efficient estimators. *New J. Phys.* **14**, 095022. <https://doi.org/10.1088/1367-2630/14/9/095022> (2012).
39. Czerwinski, A. Dynamic state reconstruction of quantum systems subject to pure decoherence. *Int. J. Theor. Phys.* **59**, 3646–3661. <https://doi.org/10.1007/s10773-020-04625-8> (2020).
40. Flurin, E., Martin, L. S., Hacoen-Gourgy, S. & Siddiqi, I. Using a recurrent neural network to reconstruct quantum dynamics of a superconducting qubit from physical observations. *Phys. Rev. X* **10**, 011006. <https://doi.org/10.1103/PhysRevX.10.011006> (2020).
41. Mäkinen, A., Ikonen, J., Partanen, M. & Möttönen, M. Reconstruction approach to quantum dynamics of bosonic systems. *Phys. Rev. A* **100**, 042109. <https://doi.org/10.1103/PhysRevA.100.042109> (2019).
42. Lvovsky, A. I. & Raymer, M. G. Continuous-variable optical quantum-state tomography. *Rev. Mod. Phys.* **81**, 299 (2009).
43. D'Ariano, G. M., De Laurentis, M., Paris, M. G., Porzio, A. & Solimeno, S. Quantum tomography as a tool for the characterization of optical devices. *J. Opt. B Quantum Semiclassical Opt.* **4**, S127 (2002).
44. Takeda, K. *et al.* Quantum tomography of an entangled three-qubit state in silicon. *Nat. Nanotechnol.* **20**, 1–5 (2021).
45. Liu, Y. *et al.* Variational quantum circuits for quantum state tomography. *Phys. Rev. A* **101**, 052316. <https://doi.org/10.1103/PhysRevA.101.052316> (2020).
46. Lee, S. M., Lee, J. & Bang, J. Learning unknown pure quantum states. *Phys. Rev. A* **98**, 052302. <https://doi.org/10.1103/PhysRevA.98.052302> (2018).
47. Lee, S. M., Park, H. S., Lee, J., Kim, J. & Bang, J. Quantum state learning via single-shot measurements. *Phys. Rev. Lett.* **126**, 170504. <https://doi.org/10.1103/PhysRevLett.126.170504> (2021).
48. Aaronson, S. Shadow tomography of quantum states. <https://doi.org/10.48550/ARXIV.1711.01053> (2017).
49. Huang, H.-Y., Kueng, R. & Preskill, J. Predicting many properties of a quantum system from very few measurements. *Nat. Phys.* **16**, 1050–1057. <https://doi.org/10.1038/s41567-020-0932-7> (2020).
50. White, C. D. & Wilson, J. H. Mana in haar-random states. [arXiv:2011.13937](https://arxiv.org/abs/2011.13937) (arXiv preprint) (2020).
51. Haug, T., Bharti, K. & Kim, M. Capacity and quantum geometry of parametrized quantum circuits. *PRX Quantum* **2**, 040309. <https://doi.org/10.1103/PRXQuantum.2.040309> (2021).
52. Schuld, M., Bocharov, A., Svore, K. M. & Wiebe, N. Circuit-centric quantum classifiers. *Phys. Rev. A* <https://doi.org/10.1103/physrva.101.032308> (2020).
53. Sim, S., Johnson, P. D. & Aspuru-Guzik, A. Expressibility and entangling capability of parameterized quantum circuits for hybrid quantum-classical algorithms. *Adv. Quantum Technol.* **2**, 1900070 (2019).
54. Kuzmak, A. R. Measuring distance between quantum states on a quantum computer. *Quantum Inf. Process.* **20**, 269. <https://doi.org/10.1007/s11128-021-03196-9> (2021).
55. Kingma, D. P. & Ba, J. Adam: A method for stochastic optimization. [arXiv:1412.6980](https://arxiv.org/abs/1412.6980) (arXiv preprint) (2014).
56. Stokes, J., Isaac, J., Killoran, N. & Carleo, G. Quantum natural gradient. *Quantum* **4**, 269 (2020).
57. McClean, J. R., Boixo, S., Smelyanskiy, V. N., Babbush, R. & Neven, H. Barren plateaus in quantum neural network training landscapes. *Nat. Commun.* **9**, 4812. <https://doi.org/10.1038/s41467-018-07090-4> (2018).
58. Cerezo, M., Sone, A., Volkoff, T., Cincio, L. & Coles, P. J. Cost function dependent barren plateaus in shallow parametrized quantum circuits. *Nat. Commun.* **12**, 1791. <https://doi.org/10.1038/s41467-021-21728-w> (2021).
59. Holmes, Z., Sharma, K., Cerezo, M. & Coles, P. J. Connecting ansatz expressibility to gradient magnitudes and barren plateaus. *PRX Quantum* **3**, 010313. <https://doi.org/10.1103/PRXQuantum.3.010313> (2022).
60. Ortiz Marrero, C., Kieferová, M. & Wiebe, N. Entanglement-induced barren plateaus. *PRX Quantum* **2**, 040316. <https://doi.org/10.1103/PRXQuantum.2.040316> (2021).
61. Wang, S. *et al.* Noise-induced barren plateaus in variational quantum algorithms. *Nat. Commun.* **12**, 6961. <https://doi.org/10.1038/s41467-021-27045-6> (2021).
62. Ferrie, C. Self-guided quantum tomography. *Phys. Rev. Lett.* **113**, 190404. <https://doi.org/10.1103/PhysRevLett.113.190404> (2014).
63. Spall, J. Multivariate stochastic approximation using a simultaneous perturbation gradient approximation. *IEEE Trans. Autom. Control* **37**, 332–341. <https://doi.org/10.1109/9.119632> (1992).
64. Mitarai, K., Negoro, M., Kitagawa, M. & Fujii, K. Quantum circuit learning. *Phys. Rev. A* **98**, 032309 (2018).
65. Schuld, M., Bergholm, V., Gogolin, C., Isaac, J. & Killoran, N. Evaluating analytic gradients on quantum hardware. *Phys. Rev. A* **99**, 032331 (2019).
66. Anselmetti, G.-L. R., Wierichs, D., Gogolin, C. & Parrish, R. M. Local, expressive, quantum-number-preserving VQE ansätze for fermionic systems. *New J. Phys.* **23**, 113010. <https://doi.org/10.1088/1367-2630/ac2cb3> (2021).
67. Harrow, A. W. & Napp, J. C. Low-depth gradient measurements can improve convergence in variational hybrid quantum-classical algorithms. *Phys. Rev. Lett.* **126**, 140502 (2021).
68. Stokes, J., Isaac, J., Killoran, N. & Carleo, G. Quantum natural gradient. *Quantum* **4**, 269. <https://doi.org/10.22331/q-2020-05-25-269> (2020).

Acknowledgements

This work is supported by the VNUHCM-University of Information Technology's Scientific Research Support Fund.

Author contributions

V.T.H. implemented the algorithms and performed numerical analysis. L.B.H. derived the theoretical framework. All authors wrote the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to L.B.H.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023