Commun. Theor. Phys. **77** (2025) 035102 (9pp) https://doi.org/10.1088/1572-9494/ad89ac

Variational quantum algorithm for designing quantum information maskers*

Jin-Ze Li, Ming-Hao Wang and Bin Zhou****

School of Physics, Hubei University, Wuhan 430062, China

E-mail: wangmh@hubu.edu.cn and binzhou@hubu.edu.cn

Received 20 July 2024, revised 18 October 2024
Accepted for publication 21 October 2024
Published 12 December 2024

Abstract

Since the concept of quantum information masking was proposed by Modi *et al* (2018 *Phys. Rev. Lett.* 120, 230 501), many interesting and significant results have been reported, both theoretically and experimentally. However, designing a quantum information masker is not an easy task, especially for larger systems. In this paper, we propose a variational quantum algorithm to resolve this problem. Specifically, our algorithm is a hybrid quantum–classical model, where the quantum device with adjustable parameters tries to mask quantum information and the classical device evaluates the performance of the quantum device and optimizes its parameters. After optimization, the quantum device behaves as an optimal masker. The loss value during optimization can be used to characterize the performance of the masker. In particular, if the loss value converges to zero, we obtain a perfect masker that completely masks the quantum information generated by the quantum information source, otherwise, the perfect masker does not exist and the subsystems always contain the original information. Nevertheless, these resulting maskers are still optimal. Quantum parallelism is utilized to reduce quantum state preparations and measurements. Our study paves the way for wide application of quantum information masking, and some of the techniques used in this study may have potential applications in quantum information processing.

Keywords: variational quantum algorithm, quantum information masking, quantum parallelism

(Some figures may appear in colour only in the online journal)

1. Introduction

Quantum information technology exploits various quantum properties, such as quantum superposition, quantum entanglement, etc, to provide more powerful information processing capabilities [1]. Quantum algorithms, such as Shor's factoring algorithm [2, 3], Grover's searching algorithm [4] and the Harrow–Hassidim–Lloyd quantum algorithm for solving linear equations [5] display better quantum speed-up than their classical counterparts. In addition, various quantum communication protocols have been designed based on quantum entanglement including quantum teleportation [6, 7], quantum dense coding [8], quantum key distribution [9, 10] and quantum secret sharing [11, 12]. These protocols provide us with more flexible and secure communication methods [13]. Some of them have been commercialized and begun to provide services to individuals and institutions [14, 15].

While quantum theory has advantages in processing information, it also has some limitations, which are characterized by a branch of no-go theorems such as the no-cloning theorem [16–18], no-broadcasting theorem [19], no-deleting theorem [20] and no-hiding theorem [21]. Recently, Modi *et al* proposed the concept of quantum information masking (QIM) and thus discovered a new no-go theorem called the no-masking theorem, which claims that it is impossible to mask an arbitrary quantum state into a bipartite quantum entangled system [22].

QIM has attracted extensive research interest. It is found that the maximal set of maskable states is hyperdisks when

* Supported by the National Natural Science Foundation of China (under Grant Nos. 12105090 and 12074107), the Program of Outstanding Young and Middle-aged Scientific and Technological Innovation Team of Colleges and Universities in Hubei Province of China (under Grant No. T2020001), and the Innovation Group Project of the Natural Science Foundation of Hubei Province of China (under Grant No. 2022CFA012).
** Authors to whom any correspondence should be addressed.

masking quantum states into a bipartite quantum system [22–24]. Different from many no-go theorems, it is proved that an arbitrary set of quantum states can be masked into a multipartite quantum system [25]. To avoid the collusion that some participants would reveal the information about the encoded quantum states, a more generalized definition of k-uniform masking was proposed [26]. Similar to quantum cloning, the probabilistic or approximate masking that allows for some imperfections of masking are also investigated [27, 28]. Beyond that, it has been shown that QIM has a close relationship with other quantum information processing tasks such as quantum secret sharing [11, 12, 26], quantum error-correction codes [26, 29] and quantum bit commitment [22, 30–33]. Complementing these theoretical advancements, experimental demonstrations of quantum information masking have materialized, with Liu *et al* executing masking operations using photonic systems and validating key theoretical predictions [34]. In parallel, Zhang *et al* pioneered the experimental realization of masking in high-dimensional systems [35], propelling the field forward with tangible evidence of masking's practical feasibility. These collective efforts have significantly bolstered the theoretical and experimental foundations of quantum information masking, underscoring its pivotal role in the broader landscape of quantum information processing.

To use QIM, we first need to design maskers that meet the requirements for different tasks. Routinely, analytical algebra methods have been developed to obtain quantum maskers [29, 36]. However, these methods become difficult and even unsolvable as the system grows, due to the 'exponential explosion' on the dimension of the system. To circumvent the 'exponential explosion', we propose a variational quantum algorithm to design quantum maskers. Our algorithm is a hybrid quantum–classical model, where the quantum device is parameterized to realize quantum maskers and the classical device is responsible for evaluating and optimizing the quantum device. By carefully designing the loss function, it can be used to characterize the performance of quantum maskers. Our algorithm will return a perfect masker if it exists, otherwise, the masker returned is still optimal. To speed up the design process and reduce quantum resources on quantum state preparations and measurements, quantum parallelism is utilized to evaluate the loss function.

2. Preliminary

QIM aims to encode quantum information into a multipartite system so that the original quantum information is completely unknown to local subsystems. Li and Wang generalized the definition of masking to multipartite scenarios and proved that an arbitrary quantum state can be masked when more participants are allowed [25]. To avoid collusion of some participants, Shi *et al* further generalized QIM to k-uniform QIM [26]. Let us briefly review and rephrase it [37].

First, let us denote a d-dimension Hilbert space corresponding to a quantum system X by $\mathcal{H}_X$. We use $\mathcal{QS} = \{(p_i, |\psi_i\rangle)\}_{i=1}^n$ to denote a quantum information source that generates a pure state $|\psi_i\rangle$ with probability p_i. We also use S to denote the state set $\{|\psi_i\rangle\}_{i=1}^n$. For brevity, we use $[N]$ to denote the set $\{1, 2, \cdots, N\}$. Then, QIM is defined as follows.

Definition 1. Given $\mathcal{QS} = \{(p_i, |\psi_i\rangle)\}_{i=1}^n$ labeled A, if there exists a quantum operation M that maps any pure state $|\psi_i\rangle \in S \subset \mathcal{H}_A$ to a state $|\Psi_i\rangle \in \otimes_{l=1}^N \mathcal{H}_{B_l}$ so that all the marginal states of arbitrary k subsystems are identical, i.e. for all $l_1 < l_2 < \cdots < l_k \in [N]$,

$$\sigma_i^{B_{l_1 l_2 \cdots l_k}} \equiv \mathrm{Tr}_{B^c_{l_1 l_2 \cdots l_k}}(|\Psi_i\rangle_B \langle\Psi_i|_B), \tag{1}$$

are independent of $|\psi_i\rangle$, and we say that the quantum information contained in $|\psi_i\rangle$ is k-uniformly masked into N subsystems by M, and M is a 1-to-N and k-uniform masker denoted by (N,k) masker. Here, $B^c_{l_1 l_2 \cdots l_k}$ denotes the complementary set of $\{B_{l_1}, B_{l_2}, \cdots B_{l_k}\}$ over $\{B_1, B_2, \cdots B_N\}$. We say M is a universal masker if it can mask all pure states in $\mathcal{H}_A$.

For a (N,k) masker, it demands that all the marginal states of k subsystems should be independent of input states. This implies that a total of $\binom{N}{k} \equiv \frac{N!}{k!(N-k)!}$ marginal subsystems need to be evaluated. This is a huge amount of work, especially when N is large and k is approximately $N/2$.

Given $\mathcal{QS}$, N and k, a perfect (N,k) masker can completely mask the quantum source. However, in some cases, perfect maskers may not exist. For these cases, optimal maskers that mask original information as much as possible are practical and useful. Thus, our purpose is to design optimal maskers. For two quantum sources, even if they produce the same set of quantum states, the optimal maskers are usually different due to the different probabilities. This further increases the difficulty of designing quantum maskers.

3. Variational quantum algorithm for designing maskers

In this section, we introduce our variational quantum algorithm for designing maskers in detail. For simplicity, the rest of the discussion will be limited to qubit masking unless otherwise stated. For higher-dimensional cases, they can be simulated with qubits.

3.1. Parameterized quantum masker

First, in our algorithm, a quantum masker is implemented by a parameterized quantum circuit (PQC) $M(\theta)$ with adjustable parameters $\theta = (\theta_1, \theta_2, \cdots)$, as shown in figure 1. PQCs are widely used in variational quantum algorithms due to their expressivity, robustness and ease of implementation [38–42]. In the field of quantum machine learning, PQCs are an important way to realize quantum neural networks [43–45]. We call these kinds of maskers variational quantum maskers (VQMs).

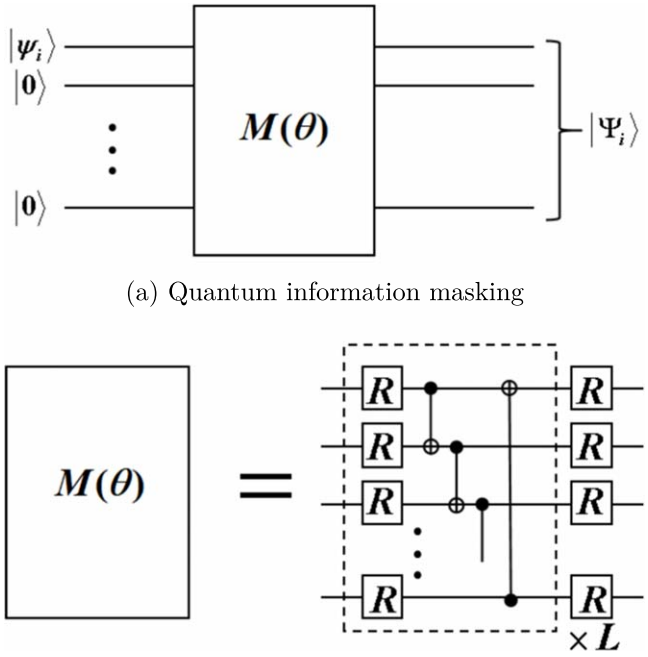


Figure 1. Illustration of quantum information masking and a parameterized quantum circuit model for a masker. R is a single-qubit rotation operation that can be parameterized by three real numbers that are not expressed explicitly.

VQMs can be constructed in various forms, depending on the specific research or the constraints of the quantum hardware [46, 47]. A well-designed ansatz, which is a structural implementation of a VQM, can significantly accelerate convergence to more accurate solutions. The ingenious design of ansatzes for PQCs lies outside the purview of this paper [48–50]. From the perspective of theoretical research, a common PQC scheme is shown in figure 1(b), which exhibits desirable properties [51]. Here, the VQM consists of L basic units, each of which is further composed of single-qubit operations $\{R\}$ and controlled-NOT gates. Each R is parameterized with three adjustable rotation angles α, β and γ, that is $R(\alpha, \beta, \gamma) = \mathrm{e}^{-\mathrm{i}Z\alpha/2}\mathrm{e}^{-\mathrm{i}Y\beta/2}\mathrm{e}^{-\mathrm{i}Z\gamma/2}$, where X, Y and Z are Pauli operators [38]. The role of controlled-NOT gates is to introduce quantum entanglement between qubits. Our task is to find the optimal parameters so that the PQC behaves as an optimal masker. This is addressed by a hybrid quantum–classical model.

3.2. Loss function

To optimize parameters, we design a loss function according to definition 1,

$$L(\theta) = \frac{1}{\mathcal{N}}\sum_{l}\sum_{i,j=1}^{n} p_i p_j S(\sigma_i^{B_l}, \sigma_j^{B_l}), \tag{2}$$

where $\mathcal{N} = 2\binom{N}{k}$ is the standardized coefficient. $S(\sigma_1, \sigma_2)$ is a similarity function that measures the similarity between σ_1 and σ_2. For ease of notation, here we use the composite index $l = \{l_1, l_2, \cdots, l_k\}$ to mark k subsystems. It is required that

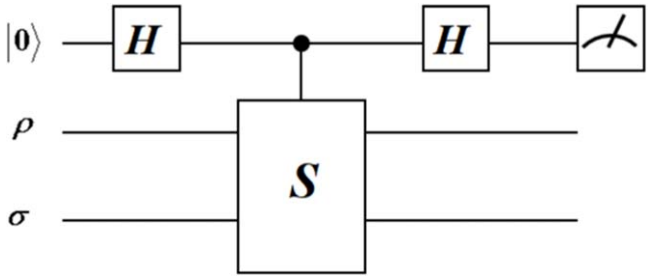


Figure 2. Diagram of the quantum circuit for the SWAP test. Here, H represents the Hadamard gate and S is the SWAP gate. By measuring the expectation value of the first qubit with respect to the Pauli-Z operator, we obtain the overlap between states ρ and σ.

$S(\sigma_1, \sigma_2) \geqslant 0$ holds for any σ_1 and σ_2. The equation is true if and only if $\sigma_1 = \sigma_2$. Thus, $L(\theta)$ is equal to the lower bound 0 when all states of k subsystems are identical (independent of input states), which implies the quantum information carried by input states is perfectly masked. Conversely, the greater the value of $L(\theta)$, the worse the performance of the masker. Therefore, loss value can be used as a measure of the performance of a masker.

The similarity measurement has a wide variety of definitions among math and machine learning practitioners [52]. In this study, we use the squared Hilbert–Schmidt distance as a measure of similarity [53, 54]. The Hilbert–Schmidt distance between σ_1 and σ_2 is defined as,

$$D_{\mathrm{H-S}}(\sigma_1, \sigma_2) = \sqrt{\mathrm{Tr}\left((\sigma_1 - \sigma_2)^{\dagger}(\sigma_1 - \sigma_2)\right)}. \tag{3}$$

Thus, the similarity function has the form of,

$$\begin{aligned} S(\sigma_1, \sigma_2) &= D^2_{\mathrm{H-S}}(\sigma_1, \sigma_2) \\ &= \mathrm{Tr}(\sigma_1^2) + \mathrm{Tr}(\sigma_2^2) - 2\,\mathrm{Tr}(\sigma_1\sigma_2). \end{aligned} \tag{4}$$

The quantity $\mathrm{Tr}(\sigma_1\sigma_2)$ is called the overlap between σ_1 and σ_2, which can be effectively estimated by SWAP tests [55, 56]. The quantum circuit of a SWAP test is shown in figure 2. See appendix A for a detailed discussion of SWAP tests.

Substituting equation (4) into (2), we obtain the concrete expression of the loss function as,

$$\begin{aligned} L(\theta) &= \frac{1}{\mathcal{N}}\sum_{l}\sum_{i,j} p_i p_j (\mathrm{Tr}(\sigma_i^{B_l}\sigma_i^{B_l}) + \mathrm{Tr}(\sigma_j^{B_l}\sigma_j^{B_l}) \\ &\quad - 2\,\mathrm{Tr}(\sigma_i^{B_l}\sigma_j^{B_l})) \\ &= \binom{N}{k}^{-1}\sum_{l}\left(\sum_{i} p_i\,\mathrm{Tr}(\sigma_i^{B_l}\sigma_i^{B_l}) - \sum_{i,j} p_i p_j\,\mathrm{Tr}(\sigma_i^{B_l}\sigma_j^{B_l})\right), \end{aligned} \tag{5}$$

where $\sigma_i^{B_l}$ is the marginal state of k subsystems labeled by composite index l (See equation (1)). Based on SWAP tests, each term in equation (5) can be evaluated, resulting in an evaluation of the loss function. As mentioned earlier, this method is extremely inefficient due to the large number of terms to be evaluated in equation (5), especially for large n, N and $k \approx \lfloor N/2 \rfloor$. Efficient methods need to be developed.

3.3. Parallel evaluation of loss function

Estimating loss functions directly requires repeating a large number of SWAP tests, consuming a lot of quantum resources and time. To speed up the process and save quantum resources, we take advantage of the technology of quantum parallelism [57]. The terms in equation (5) stem from two aspects, the states generated by the quantum information source and the combinations of k out of N. Let us explore how parallelism can be used to accelerate these two.

First, let us deal with the complexity that comes with quantum sources. We have shown in appendix A that $\sum_i p_i \operatorname{Tr}(\sigma_i^{B_l}\sigma_i^{B_l})$ can be evaluated by using the SWAP test only once with input state $\sigma^{B_l} = \sum_i p_i \sigma_i^{B_l} \otimes \sigma_i^{B_l}$, that is,

$$\sum_i p_i \operatorname{Tr}(\sigma_i^{B_l}\sigma_i^{B_l}) = \mathrm{ST}(\sigma^{B_l}), \tag{6}$$

where $\mathrm{ST}(\cdot)$ is a functional representation of the SWAP test. To clarify, 'running the SWAP test once' here refers to one evaluation of the overlap between two states, which requires multiple measurements, and the higher the accuracy requirement, the more measurements need to be performed. Similarly, $\sum_{i,j} p_i p_j \operatorname{Tr}(\sigma_i^{B_l}\sigma_j^{B_l})$ can be evaluated by inputting $\tilde{\sigma}^{B_l} = \sum_i p_i \sigma_i^{B_l} \otimes \sum_j p_j \sigma_j^{B_l}$, that is,

$$\sum_{i,j} p_i p_j \operatorname{Tr}(\sigma_i^{B_l}\sigma_j^{B_l}) = \mathrm{ST}(\tilde{\sigma}^{B_l}). \tag{7}$$

Thus, equation (5) can be rewritten as,

$$L(\theta) = \binom{N}{k}^{-1} \sum_l (\mathrm{ST}(\sigma^{B_l}) - \mathrm{ST}(\tilde{\sigma}^{B_l})). \tag{8}$$

Second, let us deal with the complexity that comes with combination. Due to the linearity of ST, equation (8) can be further rewritten as

$$L(\theta) = \mathrm{ST}(\sigma) - \mathrm{ST}(\tilde{\sigma}), \tag{9}$$

where $\sigma = \binom{N}{k}^{-1} \sum_l \sigma^{B_l}$ and $\tilde{\sigma} = \binom{N}{k}^{-1} \sum_l \tilde{\sigma}^{B_l}$. Therefore, the key to evaluating the loss function becomes the preparation of σ and $\tilde{\sigma}$. In appendix B, we give subroutines to prepare these two states deterministically.

3.4. Optimization

Once the loss function is effectively evaluated, the classical device can guide and update the adjustable parameters of the VQM. There are many optimization methods, including gradient-based methods and non-gradient-based methods [58]. Gradient-based methods are generally preferred due to their fast convergence speed and good precision when the parameter space is very large. However, to use gradient-based methods, the gradient information of the loss function needs to be obtained. There are different strategies to do so, and they may depend on the quantum device used. A common one is the finite difference method. Different from classical neural networks, the backpropagation algorithm is not suitable for PQCs. Fortunately, for a certain class of gradient-compatible PQCs, there are analytic methods to obtain their gradients such as 'parameter-shift rules' [59–61]. They express the gradient of a function as some combination of that function at two different points. However, unlike in the finite difference methods, those two points are not infinitesimally close together, but rather quite far apart. Once the gradient of loss function is obtained denoted by $\nabla L(\theta)$, θ is then updated using,

$$\theta \to \theta - \nabla L(\theta). \tag{10}$$

It should be noted that variational quantum algorithms are likely to encounter the barren plateau phenomenon, where the gradient vanishes [62–64]. Many fixes and workarounds have been proposed and investigated [41, 65–67]. This is beyond the scope of this paper.

3.5. Complete description of variational quantum algorithm

It is time to give a complete description of the variational algorithm for designing quantum maskers. We show the pseudocode in algorithm 1.

Algorithm 1. Variational quantum algorithm for designing quantum maskers

Input: Quantum source $\mathcal{QS}$, integers N and k, number of iterations ITR;
Output: Optimal $M(\theta)$
1 Select a circuit ansatz $M(\theta)$ to represent a parameterized masker;
2 Initialize parameters θ;
3 **for** $itr = 1, 2, \cdots, ITR$ **do**
4 Call subroutines to prepare states σ and $\tilde{\sigma}$;
5 Evaluate $\mathrm{ST}(\sigma)$ and $\mathrm{ST}(\tilde{\sigma})$ by SWAP tests, respectively;
6 Calculate $L(\boldsymbol{\theta}) = \mathrm{ST}(\sigma) - \mathrm{ST}(\tilde{\sigma})$;
7 Perform optimization for $L(\boldsymbol{\theta})$ and update the parameters $\boldsymbol{\theta}$;
8 **end**
9 **return** $M(\theta)$.

The subroutines of preparing states σ and $\tilde{\sigma}$ can be found in appendix B.

4. Consumption of quantum resources

In this section, we analyze the quantum resources required by our variational quantum algorithm in terms of the number of qubits, number of gates, and measurements. In appendix B, we have discussed that for a (N, k) quantum masker, it takes at most $2\log n + 3N$ qubits and $O(poly(kN, n\log^2 n))$ gates in our algorithm. Thus, here we focus on the measurements.

To obtain the evaluation of the loss function, measurement must be performed repeatedly. A finite number of measurements inevitably introduces errors. The more times the measurement is repeated, the smaller the error in the estimate. Using SWAP tests, we need to measure $O(\frac{1}{\epsilon^2})$ times to guarantee that our estimate of overlap is within a precision $\epsilon > 0$. Thus, in our algorithm we need $O(\frac{1}{\epsilon^2})$ measurements to evaluate $L(\theta)$ with an error $\epsilon > 0$. However, it takes $O\left(\binom{N}{k}\frac{n^2}{\epsilon^2}\right)$ measurements if we directly evaluate the terms in

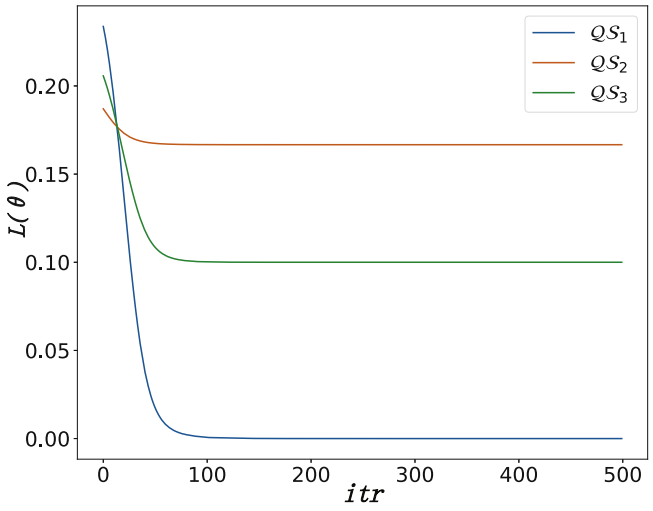


Figure 3. Comparative analysis of loss curves across different quantum sources. Here, the VQM consists of two basic units, that is $L = 2$. Numerical results show that the loss values decrease rapidly with iterations and converge to 0, $1/6$ and 0.1, respectively.

equation (5). This reveals the accelerating effect of quantum parallelism.

5. Numerical simulation

To verify the feasibility of our algorithm, in this section we present the numerical simulation results. Given the challenges associated with simulating quantum algorithms on classical computers, our focus here is on the simulation of designing (2, 1) maskers. This simplified case sufficiently demonstrates the effectiveness of our algorithm. Second, for the sake of simplicity, the quantum sources to be masked are chosen as,

$$\mathcal{QS}_1 = \left\{\left(\frac{1}{4}, |+\rangle\right), \left(\frac{1}{4}, |-\rangle\right), \left(\frac{1}{4}, |+\mathrm{i}\rangle\right), \left(\frac{1}{4}, |-\mathrm{i}\rangle\right)\right\}, \tag{11}$$

$$\mathcal{QS}_2 = \left\{\left(\frac{1}{6}, |+\rangle\right), \left(\frac{1}{6}, |-\rangle\right), \left(\frac{1}{6}, |+\mathrm{i}\rangle\right), \left(\frac{1}{6}, |-\mathrm{i}\rangle\right), \left(\frac{1}{6}, |0\rangle\right), \left(\frac{1}{6}, |1\rangle\right)\right\}, \tag{12}$$

$$\mathcal{QS}_3 = \left\{\left(\frac{2}{10}, |+\rangle\right), \left(\frac{2}{10}, |-\rangle\right), \left(\frac{2}{10}, |+\mathrm{i}\rangle\right), \left(\frac{2}{10}, |-\mathrm{i}\rangle\right), \left(\frac{1}{10}, |0\rangle\right), \left(\frac{1}{10}, |1\rangle\right)\right\}, \tag{13}$$

where $|\pm\rangle = \frac{1}{\sqrt{2}}(|0\rangle \pm |1\rangle)$, $|\pm \mathrm{i}\rangle = \frac{1}{\sqrt{2}}(|0\rangle \pm \mathrm{i}|1\rangle)$.

The numerical simulations are performed using the PennyLane software library. Renowned for its versatility and open-source nature, PennyLane serves as a comprehensive platform for quantum computing, quantum machine learning and quantum chemistry [68].

Figure 3 shows the loss curves across different quantum sources. The loss value decreases rapidly with iterations. However, the convergence values differ for different quantum sources. For $\mathcal{QS}_1$, the loss converges to zero, indicating that $\mathcal{QS}_1$ can be perfectly masked. In contrast, for $\mathcal{QS}_2$, the loss converges to $1/6$. This indicates that $\mathcal{QS}_2$ cannot be perfectly masked, a result consistent with the no-masking theorem. For $\mathcal{QS}_3$, the quantum states have a higher probability distribution along the equator of the Bloch sphere. Although perfect maskers do not exist, our algorithm returns an optimal solution, achieving a loss value of 0.1. The key strategy is that the VQM prioritizes masking quantum states with higher probabilities, thereby reducing information leakage. This strategy can be seen more intuitively in figure 4.

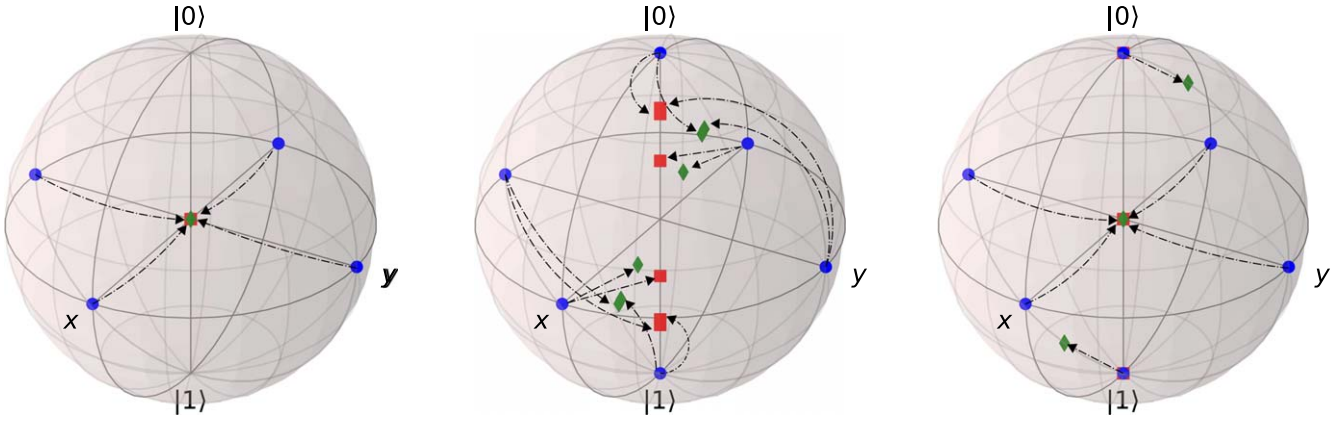
Figure 4 intuitively demonstrates the transformation of the maskers on the input qubits. Blue points represent the input qubits (to be masked), red squares represent the mixed states of subsystem B_1 and green diamonds represent the mixed states of subsystem B_2. For $\mathcal{QS}_1$, the input qubits are uniformly situated on the equator of the Bloch sphere. Figure 4(a) shows that the output states of the masker are both situated at the origin, ensuring perfect information masking. For $\mathcal{QS}_2$, the input qubits are uniformly situated at the vertices of a regular octahedron. It has been proved that these qubits cannot be perfectly masked [23]. Figure 4(b) illustrates the one-to-one correspondence between the input and the output qubits. The primary difference between $\mathcal{QS}_2$ and $\mathcal{QS}_3$ lies in the probability distribution. Figure 4(c) shows that the optimal masker prioritizes masking these qubits with higher probabilities to the origin, thereby maximizing the extent to which quantum information is masked.

6. Conclusion

Designing optimal maskers is a prerequisite for the application of quantum masking in related quantum information processing tasks. This task belongs to optimal quantum state transformation [69]. In this paper, we present a variational quantum algorithm for designing QIM. This quantum–classical hybrid algorithm can be regarded as a quantum unsupervised learning. The loss value characterizes the performance of VQMs. If the loss value converges to zero, we obtain a perfect masker; otherwise, the perfect masker does not exist. However, we still obtain an optimal masker, which may still be useful in some situations.

To reduce state preparations and measurements, quantum parallelism is utilized to evaluate the loss function. Although we investigate the topic of quantum information masking, some techniques in this paper have the potential for wider applications such as the preparation of multipartite entanglement, and quantum secret sharing.

Finally, we point out that both empirical and theoretical results exhibit that the deployed ansatz heavily affects the performance of variational quantum algorithms, and VQMs are no exception. Research on designing ansatzes is underway [50]. Quantum information masking can provide specific research objects for the study on ansatzes.



(a) $\mathcal{QS}_1$ (b) $\mathcal{QS}_2$ (c) $\mathcal{QS}_3$

Figure 4. Illustration of optimal maskers. Panels (a)–(c) depict the transformation of the optimal maskers for different quantum sources, respectively. Blue points represent the input qubits, red squares represent the mixed states of subsystem B_1 and green diamonds represent the mixed states of subsystem B_2. Dotted arrows illustrate the correspondence between input and output qubits.

Appendix A. SWAP test

Here, we provide some discussion on SWAP tests. The quantum circuit of a SWAP test is shown in figure 2. Suppose the input state is in the form of $\rho \otimes \sigma$. Calculation shows that the probabilities of obtaining 0 and 1 by measuring the first qubit are,

$$p_0 = \frac{1}{2}(1 + \mathrm{Tr}(\rho\sigma)), \tag{A1}$$

$$p_1 = \frac{1}{2}(1 - \mathrm{Tr}(\rho\sigma)). \tag{A2}$$

Thus, the overlap between ρ and σ can be obtained by subtraction, that is,

$$\mathrm{Tr}(\rho\sigma) = p_0 - p_1 = \langle Z \rangle. \tag{A3}$$

Let us regard the SWAP test as a function and denote it by $\mathrm{ST}(\cdot)$. Equation (A3) can be reformulated as,

$$\mathrm{Tr}(\rho\sigma) \equiv \mathrm{ST}(\rho \otimes \sigma). \tag{A4}$$

Note that measurement must be repeated enough times to evaluate $\mathrm{Tr}(\rho\sigma)$.

In our main text, we need to obtain the evaluations of more complex quantities such as $\sum_{i=1}^{n} p_i \mathrm{Tr}(\rho_i\sigma_i)$. A trivial method is to evaluate each overlap $\mathrm{Tr}(\rho_i\sigma_i)$ and calculate the weighted mean. This method needs to repeat SWAP tests $O(n)$ times. A better method that only runs the SWAP test once is as follows.

Our approach is to prepare a state in the form of $\sum_{i=1}^{n} p_i \rho_i \otimes \sigma_i$ and input it into the SWAP test. Due to the linearity of quantum circuits, the output state is in the form of,

$$\begin{aligned} &|0\rangle\langle 0| \otimes \frac{1}{4}\sum_i p_i[\rho_i \otimes \sigma_i + (\rho_i \otimes \sigma_i)S \\ &+S(\rho_i \otimes \sigma_i) + \sigma_i \otimes \rho_i] \\ &+|1|\langle 1| \otimes \frac{1}{4}\sum_i p_i[\rho_i \otimes \sigma_i - (\rho_i \otimes \sigma_i)S \\ &-S(\rho_i \otimes \sigma_i) + \sigma_i \otimes \rho_i] \\ &+\cdots, \end{aligned} \tag{A5}$$

where S is the SWAP operator. Thus, the probabilities of obtaining 0 and 1 by measuring the first qubit are,

$$p_0 = \frac{1}{2} + \frac{1}{2}\sum_i p_i \mathrm{Tr}(\rho_i\sigma_i), \tag{A6}$$

$$p_1 = \frac{1}{2} - \frac{1}{2}\sum_i p_i \mathrm{Tr}(\rho_i\sigma_i). \tag{A7}$$

Finally, we obtain the evaluation of $\sum_i p_i \mathrm{Tr}(\rho_i\sigma_i)$ by subtracting p_0 and p_1 as before. This indicates the linearity of $\mathrm{ST}(\cdot)$. We come to the following conclusions from the SWAP test:

1. If we input $\rho \otimes \sigma$, then it outputs $\mathrm{ST}(\rho \otimes \sigma) = \mathrm{Tr}(\rho\sigma)$;
2. If we input $\sum_{i=1}^{n} p_i \rho_i \otimes \sigma_i$, then it outputs

$$\mathrm{ST}(\sum_{i=1}^{n} p_i \rho_i \otimes \sigma_i) = \sum_{i=1}^{n} p_i \mathrm{Tr}(\rho_i\sigma_i); \tag{A8}$$

3. If we input $\sum_{i=1}^{n} p_i \rho_i \otimes \sum_{j=1}^{m} p_j \sigma_j$, then it outputs

$$\mathrm{ST}(\sum_{i=1}^{n} p_i \rho_i \otimes \sum_{j=1}^{m} p_j \sigma_j) = \sum_{i=1}^{n}\sum_{j=1}^{m} p_i p_j \mathrm{Tr}(\rho_i\sigma_j). \tag{A9}$$

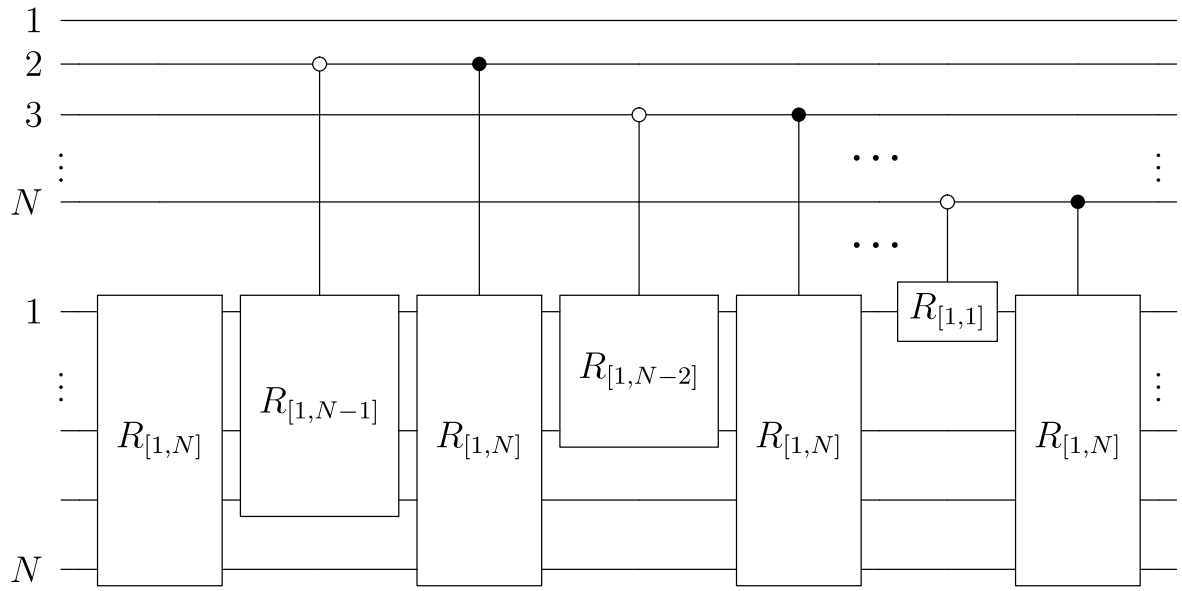


Figure 5. Diagram of quantum circuit for *C-R*.

Appendix B. Preparation of σ and $\tilde{\sigma}$

In our algorithms, a key step is to prepare σ and $\tilde{\sigma}$. Here, we introduce subroutines to prepare these states.

Suppose an arbitrary state $|\psi_i\rangle \in S$ is generated by applying a known unitary operator U_i on the standard state $|0\rangle$, that is,

$$|\psi_i\rangle = U_i|0\rangle. \tag{B1}$$

Based on U_i, we construct a controlled unitary operator:

$$C\text{-}U = \sum_{i=1}^{n} |i\rangle\langle i| \otimes U_i, \tag{B2}$$

where $\{|i\rangle\}$ forms a set of orthonormal bases. *C-U* can be decomposed into n multi-qubit controlled operations [1].

Given an initial state $\sum_{i=1}^{n}\sqrt{p_i}|i\rangle_I|0\rangle_A$, by applying *C-U* on I and A, we obtain:

$$|\psi\rangle = \sum_{i=1}^{n}\sqrt{p_i}|i\rangle_I|\psi_i\rangle_A. \tag{B3}$$

In general, preparing an arbitrary state is difficult. However, efficient algorithms exist for the states to be of the form $\sum_{i=1}^{n}\sqrt{p_i}|i\rangle$. For example, Grover and Rudolph proposed a scheme that requires a linear number of operations regarding the number of qubits [70]. Applying $M(\theta)$ on $|\psi\rangle$ gives us:

$$|\Psi\rangle = \sum_{i}^{n}\sqrt{p_i}|i\rangle_I|\Psi_i\rangle_B. \tag{B4}$$

If we focus on the k subsystems marked by l, we obtain $\sum_i p_i\sigma_i^{B_l}$. Making another one and putting them together, we obtain $\tilde{\sigma}^{B_l} = \sum_i p_i\sigma_i^{B_l}\otimes\sum_j p_j\sigma_j^{B_l}$.

Note that our purpose is to prepare $\tilde{\sigma} = \binom{N}{k}^{-1}\sum_l \tilde{\sigma}^{B_l}$. To this end, let us associate $l = \{l_1, l_2, \cdots, l_k\}$ with an n-qubit quantum state defined as,

$$|l\rangle = \cdots|0\rangle_{l_1-1}|1\rangle_{l_1}\cdots|1\rangle_{l_2}\cdots|1\rangle_{l_k}\cdots, \tag{B5}$$

that is the l_ith qubit is set at $|1\rangle$ and the rest at $|0\rangle$. For example, if $l = \{1, 3, 4\}$ and $N = 5$, the corresponding state is $|l\rangle = |1\rangle_1|0\rangle_2|1\rangle_3|1\rangle_4|0\rangle_5$. Based on this, we introduce Dicke state $|D_k^n\rangle$ defined as,

$$|D_k^N\rangle = \binom{N}{k}^{-\frac{1}{2}}\sum_l |l\rangle, \tag{B6}$$

that is the equal superposition of all N-qubit states $|l\rangle$ with Hamming weight k. Dicke states have garnered widespread attention for tasks in quantum information and as starting states for combinatorial optimization problems [71–74]. Then, we define the cyclic shift operation $R_{[i,j]}$ as below:

$$R_{[i,j]}|x_i\rangle_i|x_{i+1}\rangle_{i+1}\cdots|x_j\rangle_j = |x_{i+1}\rangle_i\cdots|x_j\rangle_{j-1}|x_i\rangle_j, \tag{B7}$$

which can be constructed by SWAP gates:

$$R_{[i,j]} = \prod_{r=0}^{j-i-1} S(j-r, j-r-1). \tag{B8}$$

Based on $R_{[i,j]}$, the controlled cyclic shift operations can be constructed:

$$C_0\text{-}R_{[i,j]} = |0\rangle\langle 0|\otimes R_{[i,j]} + |1\rangle\langle 1|\otimes I, \tag{B9}$$

$$C_1\text{-}R_{[i,j]} = |1\rangle\rangle 1|\otimes R_{[i,j]} + |0\rangle\langle 0|\otimes I. \tag{B10}$$

Next, we construct a controlled operation *C-R* defined as,

$$C\text{-}R = \prod_{i=0}^{N-2}\left(C_1^{N-i} - R_{[1,N]}\cdot C_0^{N-i} - R_{[1,1+i]}\right)\cdot R_{[1,N]}, \tag{B11}$$

where the superscript of C indicates the control qubit. The quantum circuit of *C-R* is shown in figure 5. Note that $R_{[1,1]}$ denotes identity operation. We reserve it here for clarity. It can be easily verified that *C-R* can determinedly shift target k subsystems into the last k subsystems according to $|l\rangle$ encoded into the control qubits.

For example, if we input $|l\rangle|\Psi_i\rangle$, the last k subsystems of system B output the state $\sigma_i^{B_l}$.

Now, it is time to give the subroutine to prepare $\tilde{\sigma}$. We show the subroutine in algorithm 2.

Algorithm 2. Prepare state $\tilde{\sigma}$

Input: $\mathcal{QS}$, N, k, M
Output: $\tilde{\sigma}$
1 Prepare $\sum_{i=1}^{n}\sqrt{p_i}|i\rangle_{I_1}|\psi_i\rangle_A$;
2 Apply M on systems A and R, where R refers to the necessary auxiliary system;
3 Discard system I_1, obtaining $\sum_{i=1}^{n}p_i|\Psi_i\rangle_B\langle\Psi_i|_B$;
4 Repeat steps 1-3, obtaining another copy of $\sum_{j=1}^{n}p_j|\Psi_j\rangle_{B'}\langle\Psi_j|_{B'}$;
5 Prepare Dicke state $|D_k^N\rangle_{I_2}$;
6 Apply C-R on systems I_2 and B;
7 Apply C-R on systems I_2 and B';
8 Reserve systems $B_{N-k+1,N-k+2,\cdots,N}$ and ${B'}_{N-k+1,N-k+2,\cdots,N}$, obtaining $\tilde{\sigma}$;
9 **return** $\tilde{\sigma}$.

Similarly, σ can be prepared in a slightly different way and we show it in algorithm 3.

Algorithm 3. Prepare state σ

Input: $\mathcal{QS}$, N, k, M
Output: σ
1 Prepare $\sum_{i=1}^{n}\sqrt{p_i}|i\rangle_{I_1}|\psi_i\rangle_A|\psi_i\rangle_{A'}$;
2 Apply M on systems A and R;
3 Apply M on systems A' and R';
4 Discard system I_1, obtaining $\sum_{i=1}^{n}p_i|\Psi_i\rangle_B\langle\Psi_i|_B\otimes|\Psi_i\rangle_{B'}\langle\Psi_i|_{B'}$;
5 Prepare Dicke state $|D_k^N\rangle_{I_2}$;
6 Apply C-R on systems I_2 and B;
7 Apply C-R on systems I_2 and B';
8 Reserve systems $B_{N-k+1,N-k+2,\cdots,N}$ and ${B'}_{N-k+1,N-k+2,\cdots,N}$, obtaining σ;
9 **return** σ.

From algorithm 2 and 3, it can be seen that the width of our algorithms is at most $3N+2\log n$. Suppose U_i are single-qubit unitary operations, it has been proved that an r-qubit controlled single-qubit unitary gate can be decomposed into a circuit with $O(r^2)$ single-qubit and CNOT gates [75]. Thus, C-U can be implemented using $O(n\log^2 n)$ gates. In addition, the preparation of $\sum_{i=1}^{n}\sqrt{p_i}|i\rangle$ and $|D_k^N\rangle_{I_3}$ requires $O(\log n)$ and $O(kN)$ gates, respectively [70, 76]. We further suppose VQM can be implemented with a polynomial number of operations. Then, the total number of gates of preparing $\tilde{\sigma}$ or σ is $O(poly(kN, n\log^2 n))$.

References

[1] Nielsen M A and Chuang I L 2012 *Quantum Computation and Quantum Information: X Anniversary Edition* (Cambridge University Press)
[2] Shor P W 1994 Algorithms for quantum computation: discrete logarithms and factoring *Proceedings 35th Annual Symposium on Foundations of Computer Science* p 124
[3] Shor P W 1999 Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer *SIAM Rev.* **41** 303
[4] Grover L K 1997 Quantum mechanics helps in searching for a needle in a haystack *Phys. Rev. Lett.* **79** 325
[5] Harrow A, Hassidim A and Lloyd S 2009 Quantum algorithm for linear systems of equations *Phys. Rev. Lett.* **103** 150502
[6] Bennett C H, Brassard G, Crépeau C, Jozsa R, Peres A and Wootters W K 1993 Teleporting an unknown quantum state via dual classical and Einstein–Podolsky–Rosen channels *Phys. Rev. Lett.* **70** 1895
[7] Bouwmeester D, Pan J W, Mattle K, Eibl M, Weinfurter H and Zeilinger A 1997 Experimental quantum teleportation *Nature* **390** 575
[8] Bennett C H and Wiesner S J 1992 Communication via one- and two-particle operators on Einstein–Podolsky–Rosen states *Phys. Rev. Lett.* **69** 2881
[9] Bennett C H 1992 Quantum cryptography using any two nonorthogonal states *Phys. Rev. Lett.* **68** 3121
[10] Gisin N, Ribordy G, Tittel W and Zbinden H 2002 Quantum cryptography *Rev. Mod. Phys.* **74** 145
[11] Hillery M, Bužek V and Berthiaume A 1999 Quantum secret sharing *Phys. Rev.* A **59** 1829
[12] Cleve R, Gottesman D and Lo H K 1999 How to share a quantum secret *Phys. Rev. Lett.* **83** 648
[13] Gisin N and Thew R 2007 Quantum communication *Nat. Photonics* **1** 165
[14] Chen Y A *et al* 2021 Satellite-to-ground quantum key distribution *Nature* **589** 214
[15] Lu C Y, Cao Y, Peng C Z and Pan J W 2022 Micius quantum experiments in space *Rev. Mod. Phys.* **94** 035001
[16] Wootters W K and Zurek W H 1982 A single quantum cannot be cloned *Nature* **299** 802
[17] Gisin N and Massar S 1997 Optimal quantum cloning machines *Phys. Rev. Lett.* **79** 2153
[18] Lamas-Linares A, Simon C, Howell J C and Bouwmeester D 2002 Experimental quantum cloning of single photons *Science* **296** 712
[19] Barnum H, Caves C M, Fuchs C A, Jozsa R and Schumacher B 1996 Noncommuting mixed states cannot be broadcast *Phys. Rev. Lett.* **76** 2818
[20] Pati A K and Braunstein S L 2000 Impossibility of deleting an unknown quantum state *Nature* **404** 164
[21] Braunstein S L and Pati A K 2007 Quantum information cannot be completely hidden in correlations: implications for the black-hole information paradox *Phys. Rev. Lett.* **98** 080502
[22] Modi K, Pati A K, Sen(De) A and Sen U 2018 Masking quantum information is impossible *Phys. Rev. Lett.* **120** 230501
[23] Liang X B, Li B and Fei S M 2019 Complete characterization of qubit masking *Phys. Rev.* A **100** 030304
[24] Ding F and Hu X 2020 Masking quantum information on hyperdisks *Phys. Rev.* A **102** 042404
[25] Li M S and Wang Y L 2018 Masking quantum information in multipartite scenario *Phys. Rev.* A **98** 062306
[26] Shi F, Li M S, Chen L and Zhang X 2021 k-uniform quantum information masking *Phys. Rev.* A **104** 032601
[27] Li B, Jiang S H, Liang X B, Li-Jost X, Fan H and Fei S M 2019 Deterministic versus probabilistic quantum information masking *Phys. Rev.* A **99** 052343
[28] Li M S and Modi K 2020 Probabilistic and approximate masking of quantum information *Phys. Rev.* A **102** 022418
[29] Han K, Guo Z, Cao H, Du Y and Yang C 2020 Quantum multipartite maskers vs. quantum error-correcting codes *Europhys. Lett.* **131** 30005

[30] Mayers D 1997 Unconditionally secure quantum bit commitment is impossible *Phys. Rev. Lett.* **78** 3414
[31] Kent A 1999 Unconditionally secure bit commitment *Phys. Rev. Lett.* **83** 1447
[32] Danan A and Vaidman L 2011 Practical quantum bit commitment protocol *Quantum Inf. Process.* **11** 769
[33] Lie S H, Kwon H, Kim M S and Jeong H 2021 Quantum one-time tables for unconditionally secure qubit-commitment *Quantum* **5** 405
[34] Liu Z H *et al* 2021 Photonic implementation of quantum information masking *Phys. Rev. Lett.* **126** 170505
[35] Zhang R Q, Hou Z, Li Z, Zhu H, Xiang G Y, Li C F and Guo G C 2021 Experimental masking of real quantum states *Phys. Rev. Appl.* **16** 024052
[36] Du Y, Guo Z, Cao H, Han K and Yang C 2021 Masking quantum information encoded in pure and mixed states *Int. J. Theor. Phys.* **60** 2380
[37] Zhang S, Wang M and Zhou B 2023 Quantifying the information distribution of quantum information masking *Quantum Inf. Process.* **22** 284
[38] Mitarai K, Negoro M, Kitagawa M and Fujii K 2018 Quantum circuit learning *Phys. Rev.* A **98** 032309
[39] Benedetti M, Lloyd E, Sack S and Fiorentini M 2019 Parameterized quantum circuits as machine learning models *Quantum Sci. Technol.* **4** 043001
[40] Du Y, Hsieh M H, Liu T and Tao D 2020 Expressive power of parametrized quantum circuits *Phys. Rev. Res.* **2** 033125
[41] Cerezo M *et al* 2021 Variational quantum algorithms *Nat. Rev. Phys.* **3** 625
[42] Cerezo M and Coles P J 2021 Higher order derivatives of quantum neural networks with barren plateaus *Quantum Sci. Technol.* **6** 035006
[43] Beer K, Bondarenko D, Farrelly T, Osborne T J, Salzmann R, Scheiermann D and Wolf R 2020 Training deep quantum neural networks *Nat. Commun.* **11** 808
[44] Abbas A, Sutter D, Zoufal C, Lucchi A, Figalli A and Woerner S 2021 The power of quantum neural networks *Nat. Comput. Sci.* **1** 403
[45] Kwak Y, Yun W J, Jung S and Kim J 2021 Quantum neural networks: concepts, applications, and challenges arXiv:2108.01468
[46] Kandala A, Mezzacapo A and Temme K 2017 Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets *Nature* **549** 242
[47] Tang H L, Shkolnikov V O, Arron G B S, Grimsley H R, Mayhall N J, Barnes E and Economou S E 2021 Qubit-ADAPT-VQE: an adaptive algorithm for constructing hardware-efficient ansätze on a quantum processor *PRX Quantum* **2** 020310
[48] Du Y *et al* 2022 Quantum circuit architecture search for variational quantum algorithms *Npj Quantum Inf.* **8** 62
[49] Guo X, Muta T and Zhao J 2024 Quantum circuit ansatz: patterns of abstraction and reuse of quantum algorithm design arXiv:2405.05021
[50] Qin J 2023 Review of ansatz designing techniques for variational quantum algorithms *J. Phys.: Conf. Ser.* **2634** 012043
[51] Schuld M, Bocharov A, Svore K M and Wiebe N 2020 Circuit-centric quantum classifiers *Phys. Rev.* A **101** 032308
[52] Ontañón S 2020 An overview of distance and similarity functions for structured data *Artif. Intell. Rev.* **53** 5309
[53] Dajka J, Łuczka J and Hänggi P 2011 Distance between quantum states in the presence of initial qubit-environment correlations: a comparative study *Phys. Rev.* A **84** 032120
[54] Trávníček V, Bartkiewicz K, Černoch A and Lemr K 2019 Experimental measurement of the Hilbert–Schmidt distance between two-qubit states as a means for reducing the complexity of machine learning *Phys. Rev. Lett.* **123** 260501
[55] Buhrman H, Cleve R, Watrous J and Wolf R 2001 Quantum fingerprinting *Phys. Rev. Lett.* **87** 167902
[56] Garcia-Escartin J C and Chamorro-Posada P 2013 Swap test and Hong–Ou–Mandel effect are equivalent *Phys. Rev.* A **87** 052330
[57] LaPierre R 2021 Quantum parallelism and computational complexity *Introduction to Quantum Computing, The Materials Research Society Series* (Springer) p 139
[58] Schuld M and Petruccione F 2018 *Supervised Learning with Quantum Computers* (Springer)
[59] Li J, Yang X, Peng X and Sun C P 2017 Hybrid quantum-classical approach to quantum optimal control *Phys. Rev. Lett.* **118** 150503
[60] Mari A, Bromley T R and Killoran N 2021 Estimating the gradient and higher-order derivatives on quantum hardware *Phys. Rev.* A **103** 012405
[61] Wierichs D, Izaac J, Wang C and Lin C Y Y 2022 General parameter-shift rules for quantum gradients *Quantum* **6** 677
[62] McClean J R, Boixo S, Smelyanskiy V N, Babbush R and Neven H 2018 Barren plateaus in quantum neural network training landscapes *Nat. Commun.* **9** 4812
[63] Wang S, Fontana E, Cerezo M, Sharma K, Sone A, Cincio L and Coles P J 2021 Noise-induced barren plateaus in variational quantum algorithms *Nat. Commun.* **12** 6961
[64] Marrero C O, Kieferová M and Wiebe N 2021 Entanglement-induced barren plateaus *PRX Quantum* **2** 040316
[65] Sack S H, Medina R A, Michailidis A A, Kueng R and Serbyn M 2022 Avoiding barren plateaus using classical shadows *PRX Quantum* **3** 020365
[66] Pesah A, Cerezo M, Wang S, Volkoff T, Sornborger A T and Coles P J 2021 Absence of barren plateaus in quantum convolutional neural networks *Phys. Rev.* X **11** 041011
[67] Zhang H K, Liu S and Zhang S X 2024 Absence of barren plateaus in finite local-depth circuits with long-range entanglement *Phys. Rev. Lett.* **132** 150603
[68] Bergholm V *et al* 2018 PennyLane: automatic differentiation of hybrid quantum-classical computations arXiv:1811.04968
[69] Zhao T H, Wang M H and Zhou B 2021 Optimal quantum state transformations based on machine learning *Quantum Inf. Process.* **20** 212
[70] Grover L and Rudolph T 2002 Creating superpositions that correspond to efficiently integrable probability distributions arXiv:quant-ph/0208112
[71] Dicke R H 1954 Coherence in spontaneous radiation processes *Phys. Rev.* **93** 99
[72] Özdemir S K, Shimamura J and Imoto N 2007 A necessary and sufficient condition to play games in quantum mechanical settings *New J. Phys.* **9** 43
[73] Prevedel R, Cronenberg G, Tame M S, Paternostro M, Walther P, Kim M S and Zeilinger A 2009 Experimental realization of Dicke states of up to six qubits for multiparty quantum networking *Phys. Rev. Lett.* **103** 020503
[74] Tóth G 2012 Multipartite entanglement and high-precision metrology *Phys. Rev.* A **85** 022322
[75] da Silva A J and Park D K 2022 Linear-depth quantum circuits for multiqubit controlled gates *Phys. Rev.* A **106** 042602
[76] Bärtschi A and Eidenbenz S 2019 Deterministic preparation of Dicke states *Fundamentals of Computation Theory, Lecture Notes in Computer Science* vol 11651 (Springer) p 126