Optimized Quantum Autoencoder

Yibin Huang[1,2], Muchun Yang[1,2], and D. L. Zhou[1,2*]

[1]*Institute of Physics, Beijing National Laboratory for Condensed Matter Physics, Chinese Academy of Sciences, Beijing 100190, China*
[2]*School of Physical Sciences, University of Chinese Academy of Sciences, Beijing 100049, China*

(Received 15 September 2025; accepted manuscript online 5 January 2026)

Quantum autoencoder (QAE) compresses a bipartite quantum state into its subsystem using a self-checking mechanism. How to characterize and minimize the lost information in this process is essential for understanding the compression mechanism of QAE. Here, we investigate how to minimize the lost information in QAE for any input mixed state. We theoretically show that the lost information is the quantum mutual information between the remaining subsystem and the discarded one; the encoding unitary transformation is designed to minimize this mutual information. Furthermore, we show that the optimized unitary transformation can be decomposed as the product of a permutation unitary transformation and a disentanglement unitary transformation, and the permutation unitary transformation can be searched by a regular Young tableau algorithm. When the search can be made exhaustive in lower-dimensional systems, the lost information is minimized numerically, which is shown theoretically to be a global minimum. When the dimension of the system becomes larger such that an exhaustive search is impossible, we adopt an approximate search algorithm to numerically identify that our compression scheme gives lower lost information than that from the quantum variational circuit-based QAE.

DOI: 10.1088/0256-307X/43/2/020601 CSTR: 32039.14.0256-307X.43.2.020601

1. Introduction. Information compression is fundamental in classical and quantum information processing.[1–7] As an effective tool for classical information compression, classical autoencoders have wide applications in feature extraction,[8,9] denoising,[10,11] and data compression.[12,13] Quantum autoencoder (QAE), as its extension from the classical regime to the quantum regime, was proposed and successfully demonstrated by Romero *et al.* for compressing an ensemble of pure states.[14] The goal of QAE is to compress quantum data from a $d_A \times d_B$-dimensional Hilbert space to a d_B-dimensional Hilbert space using a unitary transformation U, which is selected by approximately recovering the quantum data from the compressed data through a decoder. QAEs have been implemented experimentally using photons[15,16] and superconducting qubits.[17]

Recently, QAE has served as a useful tool for numerous applications, such as quantum error correction,[18–20] quantum phase transitions,[21–23] information scrambling,[24,25] anomaly detection,[26–28] qubit readout,[29] complexity of quantum states,[30] and detecting families of quantum many-body scars.[31] In addition, reference [32] investigates QAE using entropy production. Reference [33] examines information loss, resource costs, and run time in the practical application of QAE.

The goal of QAE is to find an optimal unitary transformation U in a $d_A \times d_B$-dimensional Hilbert space. However, since the unitary group in $d_A \times d_B$-dimensional Hilbert space is a continuous group with an infinite number of elements, finding an optimal unitary transformation from this group is very challenging. Constrained by computing capacity, we often do not aim to precisely find the optimal U. Instead, we use the variational quantum circuit (VQC) method to identify an approximate solution. As our analytical and/or computational capabilities advance, we should begin to explore methods to precisely solve this problem. Previously, some works have attempted to solve this problem precisely. For instance, reference [34], by considering specific cases where the rank of the quantum state is smaller than the dimension of the compressed space, identified the optimal U for such quantum states. However, for general cases, there still does not exist a method to precisely solve this problem.

This Letter aims to present a method to determine the optimal unitary transformation U applicable to general quantum states.

To demonstrate our method, we first theoretically show that the lost information during the compression process is the quantum mutual information between the remaining subsystem and the discarded subsystem. The compression process of QAE effectively reduces the mutual information between these two parts. We prove that there exists an optimal unitary transformation which can be decomposed as the product of a permutation unitary transformation and a disentanglement unitary transformation. From this result, we find that the problem of seeking the optimal compression strategy for QAE essentially consists of two parts: a quantum problem and a classical one. The quantum

*Corresponding author. Email: zhoudl72@iphy.ac.cn
© 2026 Chinese Physical Society and IOP Publishing Ltd. All rights, including for text and data mining, AI training, and similar technologies, are reserved.

problem is essentially about how to diagonalize a quantum state in some product basis. The classical one is, given a set of numbers, finding a specific ordering that minimizes the objective function. For the diagonalization problem, related studies have been conducted in the past.[35] For the classical part of the problem, we can use the regular Young tableau algorithm proposed in this paper to solve it. After determining the optimal V_τ, we present a method to implement it using a quantum circuit in the Supplementary Materials (SM).

2. Mixed-State Compression Model. In quantum information theory, a source is described by a mixed state. Traditional QAEs[14] focus on encoding the pure-state components of the source and aim to faithfully recover each pure-state branch through an encoding–decoding procedure. Within this framework, the reference state must be chosen as a pure state. However, such a model is only capable of perfectly recovering sources of low rank.[36] For general quantum sources, references [36,37] resolve this limitation by allowing the reference state to be a mixed state.

In this Letter, we also study the compression model for general quantum sources, as illustrated in Fig. 1. In this model, a bipartite quantum state σ_{AB} undergoes a unitary evolution U, the state becomes $\sigma_{AB}^U = U\sigma_{AB}U^\dagger$, whose reduced states for the subsystem A and the subsystem B are given respectively by $\sigma_A^U = \mathrm{Tr}_B\sigma_{AB}^U, \sigma_B^U = \mathrm{Tr}_A\sigma_{AB}^U$. Then the state of subsystem A is discarded, and the state of subsystem B is σ_B^U. To recover the initial state σ_{AB}, an auxiliary quantum state ρ_A on subsystem A is introduced and the inverse evolution of the original unitary transformation is applied, and the ultimate quantum state is $\sigma_{AB}^{\mathrm{out}} = U^\dagger[\rho_A \otimes \sigma_B^U]U$. Our aim is to find a unitary evolution U and a density matrix ρ_A to minimize the distance between σ_{AB} and $\sigma_{AB}^{\mathrm{out}}$. In this model, since the reference state is taken to be a mixed state, our optimization objective changes accordingly. Rather than aiming to perfectly recover every pure-state component of the source, we return to the standard framework of quantum information compression: we do not require each pure component to be perfectly compressed for finite copies. Instead, we require that, in the asymptotic limit, the source state can be compressed and recovered nearly perfectly within its typical subspace. Quantum relative entropy characterizes the difference between two quantum states in terms of their typical subspace structure in the asymptotic regime, and therefore, we use quantum relative entropy as the distance measure between quantum states.[38] Based on this, our objective becomes

$$\min_{U,\rho_A} S(\sigma_{AB}||\sigma_{AB}^{\mathrm{out}}). \tag{1}$$

The statistical interpretation of quantum relative entropy is: it tells us how difficult it is to distinguish the state ρ from σ.[38,39] Meanwhile, the relative entropy and the fidelity satisfy the following relation:

$$F(\rho,\sigma) \geq 1 - \frac{\sqrt{S(\rho||\sigma)}}{2}, \tag{2}$$

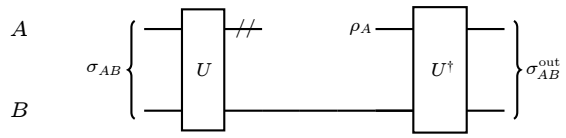

Fig. 1. In this model, the quantum state σ_{AB} undergoes unitary evolution U, followed by the removal of subsystem A to obtain the compressed quantum state σ_B. By taking the tensor product with a quantum state ρ_A on subsystem A and then applying the inverse evolution of the original unitary transformation, the quantum state is restored, we ultimately obtain the quantum state $\sigma_{AB}^{\mathrm{out}}$.

where $F(\rho,\sigma)$ denotes the fidelity between quantum states ρ and σ. This relation can be obtained directly from the Fuchs–van de Graaf inequalities as shown in Theorem 1 of Ref. [40] together with the quantum Pinsker inequality.[41] Thus, the quantity we optimize here essentially serves as a lower bound on the fidelity. On the other hand, the advantage of using relative entropy is that, as we will see below, the information lost during the QAE compression process consists of two parts: one part comes from discarding the local reduced density matrices, and the other part arises from the mutual information.

3. Minimization over the auxiliary state. We divide the minimization problem (1) into two steps. In the first step, we minimize over the auxiliary state ρ_A for any fixed unitary transformation U. In the second step, we minimize over the unitary transformation U using the optimized auxiliary state obtained in the first step. In this section, we focus on completing the first step by presenting the following intuitive yet important theorem.

Theorem 1. For any unitary transformation U,

$$\min_{\rho_A} S(\sigma_{AB}||\sigma_{AB}^{\mathrm{out}}) = S^U(A:B), \tag{3}$$

where $S^U(A:B)$ is the quantum mutual information of σ_{AB}^U, defined by $S^U(A:B) = S(\sigma_A^U) + S(\sigma_B^U) - S(\sigma_{AB}^U)$. The minimization is taken if and only if $\rho_A = \sigma_A^U$.

The proof of Theorem 1 can be found in the SM. The proof is straightforward, as the mutual information measures the minimal distance between the state and the set of non-correlated states (product states).[42–44] For non-correlated states, after discarding one subsystem, the original state can be fully recovered by tensoring with an appropriate quantum state.

It is worth noting that if we allow the use of general quantum channels, the compressed information can be recovered using the Petz recovery map method.[45] However, in this context, we impose a strong restriction on the recovery channel, allowing only the tensor product of a quantum state of system A and the application of inverse unitary operations.

Following Theorem 1, the minimization in Eq. (1) is equivalent to the minimization

$$\min_{U,\rho_A} S(\sigma_{AB}||\sigma_{AB}^{\mathrm{out}}) = \min_U S^U(A:B). \tag{4}$$

Thus, our task is to reduce the quantum mutual information by optimizing U.

4. Transforming Unitary Transformation into Permutation. In this section, we will come to the second step

of our minimization: find a unitary transformation U to minimize the quantum mutual information $S^U(A:B)$. We adopt the following strategy: first, find a disentanglement unitary transformation V_D such that the quantum state $\sigma_{AB}^{V_D}$ is diagonal in a certain product basis of the AB system; second, find a permutation unitary V_τ to minimize the classical correlation in the separable state $\sigma_{AB}^{V_D}$. Note that a similar strategy has been used in the study of the distribution of mutual entropy in a tripartite pure state by Yang *et al.*[46]

Before stating our theorem, we introduce two related concepts: a disentanglement unitary transformation and a permutation unitary transformation. To define a disentanglement unitary transformation, we start with the eigen decomposition of σ_{AB}, $\sigma_{AB} = \sum_{\alpha=1}^{d_A d_B} p_\alpha |\phi_\alpha\rangle \langle\phi_\alpha|$, where d_A (d_B) is the dimension of the Hilbert space of subsystem A (B), p_α and $|\phi_\alpha\rangle$ are eigenvalues and eigenvectors of the state σ_{AB}, respectively. Without loss of generality, we assume here that all eigenvalues of σ_{AB} are nonzero. Then, for a one-to-one map D that maps $\alpha \to (i,m)$, $D(\alpha) = (i,m)$, with $i \in \{1,2,\cdots,d_A\}$ and $m \in \{1,2,\cdots,d_B\}$, we define a disentanglement unitary transformation $V_D |\phi_\alpha\rangle = |i\rangle_A \otimes |m\rangle_B = |im\rangle$, where $\{|i\rangle_A\}$ and $\{|m\rangle_B\}$ are orthonormal bases of Hilbert space $\mathcal{H}_A$ and $\mathcal{H}_B$, respectively. We can see that V_D transforms σ_{AB} into a separable state: $\sigma_{AB}^{V_D} = V_D \sigma_{AB} V_D^\dagger = \sum_{im} p_{im} |im\rangle\langle im|$. In this sense, we call V_D the disentanglement unitary transformation. Note that the map D is not uniquely defined. In fact, any one-to-one map D' from $\alpha \to (i,m)$ can be related to D by $D' = \tau \circ D$, where τ is a permutation of the set $\{(i,m)\}$. It is easy to check that $V_{D'} = V_\tau V_D$ with the permutation unitary transformation $V_\tau |im\rangle = |\tau(im)\rangle$, and it is also a disentanglement unitary transformation.

In addition, because the mutual information $S^U(A:B) = S^{U^A \otimes U^B U}(A:B)$ holds for any local unitary transformation U^A and U^B, the minimization of this mutual information does not depend on the choice of the basis vectors of $\mathcal{H}_A$ and $\mathcal{H}_B$. In practice, the basis of $\mathcal{H}_A$ and $\mathcal{H}_B$ can be chosen according to specific needs. For instance, in a qubit system, the computational basis is an excellent choice. We are now ready to present our main theorem.

Theorem 2. The minimization of mutual information by a unitary transformation

$$\min_U S^U(A:B) = \min_{V_\tau} S^{V_\tau V_D}(A:B), \tag{5}$$

where V_τ is a permutation unitary transformation, and V_D is an arbitrary disentanglement unitary transformation.

Proof. If we denote $V = UV_D^{-1}$, then $U = VV_D$. Equation (5) can be written in a more direct form $\min_V S^{VV_D}(A:B) = \min_{V_\tau} S^{V_\tau V_D}(A:B)$. Because the von Neumman entropy is invariant under any unitary transformation, i.e., $S(\sigma_{AB}^{VV_D}) = S(\sigma_{AB}^{V_\tau V_D}) = S(\sigma_{AB})$, then we only need to prove

$$\begin{aligned} &\min_V S\left(\sigma_A^{VV_D}\right) + S\left(\sigma_B^{VV_D}\right) \\ &= \min_{V_\tau} S\left(\sigma_A^{V_\tau V_D}\right) + S\left(\sigma_B^{V_\tau V_D}\right). \end{aligned} \tag{6}$$

The state σ_{AB}^U can be written in the form $\sigma_{AB}^U = \sigma_{AB}^{VV_D} = \sum_{im} p_{im} V|im\rangle\langle im|V^\dagger$. Assuming that its reduced state σ_A^U is diagonal in the basis $\{|i^U\rangle,\ 1 \le i \le d_A\}$. and σ_B^U is diagonal in the basis $\{|m^U\rangle,\ 1 \le m \le d_B\}$, i.e., $\sigma_A^U = \sum_{j=1}^{d_A} p_{Aj}^U |j^U\rangle\langle j^U|$, $\sigma_B^U = \sum_{n=1}^{d_B} p_{Bn}^U |n^U\rangle\langle n^U|$, where $p_{Aj}^U = \sum_{imn} p_{im} p^U(jn|im)$, $p_{Bn}^U = \sum_{ijm} p_{im} p^U(jn|im)$, with $p^U(jn|im) = \langle j^U n^U | V |im\rangle \langle im| V^\dagger |j^U n^U\rangle$. Note that $p^U(jn|im)$ is a doubly stochastic matrix. Especially when $U = V_\tau V_D$, we can take $|j^U n^U\rangle$ to be $|jn\rangle$, and

$$p^{V_\tau V_D}(jn|im) = \begin{cases} 1, & \text{if } jn = \tau(im), \\ 0, & \text{otherwise.} \end{cases} \tag{7}$$

In other words, $p^{V_\tau V_D}$ is a permutation matrix. According to the Birkhoff's theorem,[47] the matrix $p^U(jn|im)$ can be written as $p^U(jn|im) = \sum_\tau \lambda_\tau^U p^{V_\tau V_D}(jn|im)$, where $\lambda_\tau^U \ge 0$ and $\sum_\tau \lambda_\tau^U = 1$. Hence $p_{Aj}^U = \sum_\tau \lambda_\tau^U p_{Aj}^{V_\tau V_D}$, $p_{Bn}^U = \sum_\tau \lambda_\tau^U p_{Bn}^{V_\tau V_D}$, which implies that $\sigma_A^U = \sum_\tau \lambda_\tau^U W_A^U \sigma_A^{V_\tau V_D} W_A^{U\,\dagger}$, $\sigma_B^U = \sum_\tau \lambda_\tau^U W_B^U \sigma_B^{V_\tau V_D} W_B^{U\,\dagger}$, where $W_A^U|i\rangle = |i^U\rangle$ and $W_B^U|m\rangle = |m^U\rangle$, which are local unitary transformations. According to the concavity of the von Neumman entropy,[1] we get

$$\begin{aligned} S(\sigma_A^U) &\ge \sum_\tau \lambda_\tau^U S(W_A^U \sigma_A^{V_\tau V_D} W_A^{U\,\dagger}) \\ &= \sum_\tau \lambda_\tau^U S(\sigma_A^{V_\tau V_D}), \end{aligned} \tag{8}$$

$$\begin{aligned} S(\sigma_B^U) &\ge \sum_\tau \lambda_\tau^U S(W_B^U \sigma_B^{V_\tau V_D} W_B^{U\,\dagger}) \\ &= \sum_\tau \lambda_\tau^U S(\sigma_B^{V_\tau V_D}), \end{aligned} \tag{9}$$

which implies that

$$\sum_\tau \lambda_\tau^U [S(\sigma_A^{V_\tau V_D}) + S(\sigma_B^{V_\tau V_D})] \le S(\sigma_A^U) + S(\sigma_B^U). \tag{10}$$

Therefore, for any unitary transformation U, there must exist a unitary transformation V_τ such that $S(\sigma_A^{V_\tau V_D}) + S(\sigma_B^{V_\tau V_D}) \le S(\sigma_A^U) + S(\sigma_B^U)$, which implies that the minimization (6) can be realized by some V_τ. □

Theorem 2 shows that the minimization of mutual information for the middle state can be divided into two steps: (i) completely destroying the bipartite entanglement of σ_{AB} by any disentanglement unitary transformation V_D; (ii) maximally destroying the bipartite classical correlation of $V_D \sigma_{AB} V_D^\dagger$ by a permutation unitary transformation V_τ. In particular, the number of the permutations is $(d_A d_B)!$, which means the minimization over V_τ is a discrete optimization problem. This converts a continuous optimization problem into a discrete optimization problem, reducing the complexity of the problem to some extent.

There are two subtle questions regarding the disentanglement unitary transformation V_D in Theorem 2 that need to be clarified. First, why can we use a disentanglement unitary transformation V_D, which is not uniquely defined, in Theorem 2? The key point is that the set $\{V_\tau V_D\}$ encompasses all disentanglement unitary transformations

and is independent of the choice of a specific V_D. Second, since in most practical cases in the context of QAE we do not know the density matrix of σ_{AB}, how do we construct a disentanglement unitary transformation V_D? To answer these questions, we observe that V_D is essentially to diagonalize the input state σ_{AB} in the product basis $\{|im\rangle\}$. Fortunately, the diagonalization of a density matrix has been investigated previously.[35] Note that any disentanglement unitary transformation discussed above can perform such diagonalization.

In practice, after constructing V_D, we measure the eigenvalues to obtain $\{p_{im}\}$, which will be input into a classical computer to solve this classical discrete optimization for searching a permutation τ.

In the next section, we develop a classical algorithm for finding such a permutation τ exactly or approximately.

5. Regular Young Tableau Traversal Algorithm. In this section, we introduce the procedure for finding the permutation τ. According to Eq. (7), we obtain

$$p^{\tau}_{jn} \equiv p^{V_\tau V_D}_{jn} = \sum_{im} p_{im} p^{V_\tau V_D}(jn|im) = p_{\tau^{-1}(jn)}, \quad (11)$$

which implies that the unitary transformation V_τ maps $p_{\tau^{-1}(jn)}$ to p^{τ}_{jn}. In general, permutations τ and matrices p^τ have a one-to-one correspondence, and the size of the searching space is $(d_A d_B)!$.

Before presenting a theorem to decrease the size of the searching space, we first define the concept of a decreasing matrix: A matrix M is said to be a decreasing matrix if and only if its element satisfy $M_{ij} \geq M_{ab}$ for all $i \leq a$ and $j \leq b$.

Theorem 3. For any matrix p^τ, we can always find a permutation τ' such that $p^{\tau' \circ \tau}$ is a decreasing matrix, and

$$S^{V_\tau V_D}(A:B) \geq S^{V_{\tau' \circ \tau} V_D}(A:B). \quad (12)$$

The proof of Theorem 3 can be found in the SM. A direct consequence of Theorem 3 is that the searching space consists of all the permutations for which p^τ is a decreasing matrix. Note that p^τ is a rearrangement of $\{p_\alpha\}$. Without loss of generality, we assume $p_1 \geq p_2 \geq \cdots \geq p_{d_A d_B}$, which allows us to create a definite map from the lower index α to its value p_α. So we only need to partition the index numbers into a $d_A \times d_B$ lattice, which forms a Young tableau. Theorem 3 implies that the searching space can be restricted to the permutations corresponding to regular Young tableaux, i.e., the Young tableaux whose rows and columns are increasing sequences.

The number of regular Young tableaux Y_{num} is given by the Hook length formula,[48] as numerically demonstrated in Table 1. We choose $(d_A, d_{B\max})$ such that Y_{num} is on the order of 10^8, a scale our personal computer can handle via the exhaustive method. Within this range of Hilbert space dimensions, we can generate all regular Young tableaux to find the optimal unitary transformation V_τ.

Table 1. We choose the order of magnitude of 10^8 as a standard, which can be computed using personal computers. With d_A fixed, within the computational power range, the maximum value of d_B is $d_{B\max}$. The corresponding numbers of regular Young tableaux is Y_{num}.

d_A	2	3	4	5
$d_{B\max}$	18	9	6	5
Y_{num}	5×10^8	4×10^8	1×10^8	3×10^8

We use the example of compressing 4 qubits into 2 qubits to demonstrate the significance of the traversal algorithm (TA). We take the variational algorithm (VA) as the counterpart to the TA. To find the optimal U in minimization (4), VA constructs U in the following way:

$$U(\alpha) = e^{i \sum_j \alpha_j P_j}, \quad (13)$$

where P is the set of 4-qubit Pauli matrices, which has 256 elements in total, and $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_{256})$ is a real vector containing 256 components. VA typically uses the gradient descent method to optimize the parameter vector α in order to find the optimal U. However, VA inevitably encounters the problem of getting stuck in a local minimum while optimizing the parameters. This makes it impossible to determine whether the found U is optimal. In contrast, TA can find the optimal U and provide the minimum value of mutual information. Figure 2 shows the effectiveness of the two methods in finding the optimal U under two different cases. In the first case, we randomly generated 100 completely recoverable states which can be written as $U^i(\sigma^i_A \otimes \sigma^i_B)U^{i\dagger}$. For these states, if the optimal U is found, the mutual information can be reduced to zero. Figure 2(a) shows that for these states, TA identifies the optimal U, while VA fails to do so. In the second case,

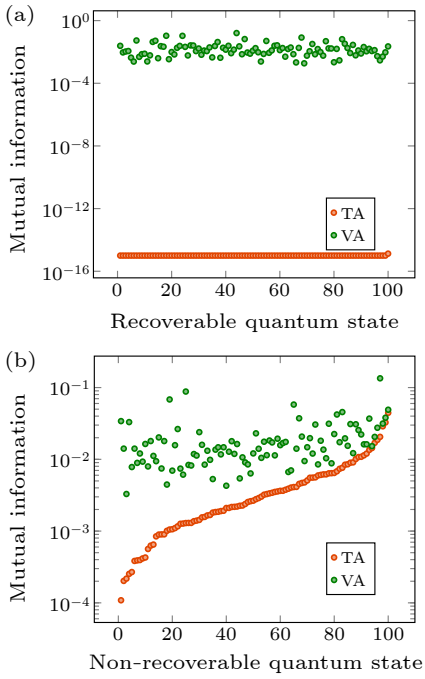

Fig. 2. We use TA and VA to reduce the mutual information between the two subsystems of a quantum state. (a) We randomly generated 100 recoverable quantum states. For these states, TA successfully finds the optimal unitary transformation, whereas VA does not. (b) We randomly generated 100 non-recoverable quantum states and observed that VA fails to find the optimal unitary transformation. In both figures, the quantum states are rearranged in ascending order according to the mutual information obtained from the TA.

we randomly generated 100 non-recoverable quantum states. For these quantum states, even if the optimal U is found, we cannot reduce the mutual information to zero. Figure 2(b) shows the results of the two methods. Since TA identifies the optimal U, using its result as a benchmark, we can see that VA fails to find the optimal U and consequently cannot provide the minimal mutual information. Our algorithm solves problems that could not be solved before: (i) determining the optimal U; and (ii) indicating the minimum mutual information achievable for a quantum state using the QAE method.

6. Search Algorithms to Reduce Mutual Information. As the number of qubits increases, the number of regular Young tableaux to consider grows exponentially, and it becomes challenging to accurately identify the optimal regular Young tableau. Fortunately, Theorem 3 indicates that the unitary operation corresponding to any regular Young tableau is superior to most other unitary operations. Therefore, even if we cannot find the exact solution, the approximate solution obtained by seeking a relatively optimal regular Young tableau is better than the majority of unitary operations within the overall unitary group. In this section, we introduce the search algorithm which consists of two parts to find a relatively optimal regular Young tableau.

The first part is called the breadth-first search algorithm, which includes randomly generating N_1 regular Young tableaux and selecting the N_2 tableaux with the lowest mutual information.

The second part is called the depth-first search algorithm, which begins with the N_2 regular Young tableaux obtained from the breadth-first search algorithm. Each tableau is treated as a seed. For each seed, we fine-tune the numbers in the tableau to generate a new set of regular Young tableaux. From this set, we identify the relatively optimal tableau to serve as the seed for the next round of iteration. This process is repeated for N_D iterations. The algorithm's code is provided in the SM.

We compare the search algorithm with the variational quantum circuit algorithm (VQA). Considering the case where $d_A = d_B = 8$, to find the optimal U in minimization (4), VQA uses VQCs to construct U. Here, we choose two circuit structures that are used in Ref. [14] and Ref. [36], respectively.

We compare the recovery performance of the two algorithms on two types of quantum states—recoverable and non-recoverable—and present the results in Fig. 3.

As depicted in Fig. 3(a), our algorithm consistently outperforms the other two algorithms in 100 optimization runs. On average, our search algorithm reduces the mutual information to 3.3×10^{-5}, while the other algorithms' average was only 2.2×10^{-2}. As depicted in Fig. 3(b), our algorithm consistently outperforms the other two algorithms in 100 optimization runs.

Through comparison, we find that for the 6-qubit case, although the VQA performs well, there is still at least a two-order-of-magnitude gap between its results and the theoretically optimal solution.

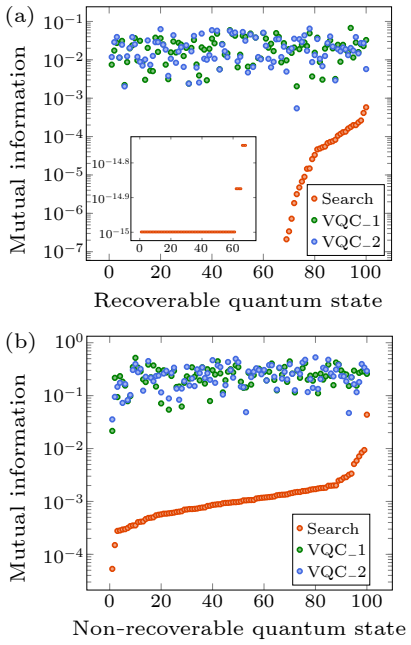

Fig. 3. (a) We randomly generate 100 recoverable states. We utilize these three methods for optimization. On average, the search algorithm reduces the mutual information to 3.3×10^{-5}, while the other algorithms average only 2.2×10^{-2}. (b) We randomly generate 100 non-recoverable quantum states on the AB system and optimize their mutual information using two VQCs (VQC_1, VQC_2) and the search algorithm, respectively. Among those states, the search algorithm consistently outperforms the other two algorithms. In both figures, the quantum states are rearranged in ascending order according to the mutual information obtained from the search algorithm.

7. Summary. In this Letter, we theoretically show that the lost information in QAE is the quantum mutual information between the remaining subsystem and the discarded one, and prove that there exists an optimal unitary transformation which can be decomposed into the product of disentanglement and permutation to minimize mutual information. For any given initial state, the aimed state after the quantum disentanglement can be given explicitly up to a permutation. The classical search for the optimal permutation is discrete, and in principle, the global optimum can be obtained by exhaustive search. However, with increasing number of qubits, the number of Young tableaux increases exponentially. We prove that the classical search can be simplified to a search over regular Young tableaux, enabling us to obtain the global optimal QAE for any mixed state in lower-dimensional Hilbert spaces. When exhaustive search is infeasible, we further present an approximation search algorithm. Compared to VQA, the search algorithm numerically reduces the error of QAE by two orders of magnitude on average.

In practice, the errors for our algorithm come mainly from experimental errors, such as the implementation of a known unitary operation. In addition, since the number of regular Young tableaux still increases exponentially with system size, more efficient classical search algorithms are still needed. To find the optimal solution of QAE for larger systems, future development may focus on the

following two directions: (i) How to more accurately diagonalize unknown quantum states in some specific basis? (ii) Finding more effective algorithms to solve the classical search problem.

Acknowledgements. This work was supported by the Science Challenge Project (Grant No. TZ2025017), and the National Key Research and Development Program of China (Grant Nos. 2021YFA0718302 and 2021YFA1402104).

References

[1] Nielsen M A and Chuang I L 2010 *Quantum Computation and Quantum Information: 10th Anniversary Edition* (Cambridge, UK: Cambridge University Press)

[2] Wilde M M 2017 *Quantum Information Theory* (Cambridge: Cambridge University Press, 2nd edition)

[3] Shannon C E 1948 *Bell Syst. Tech. J.* **27** 379

[4] Devetak I, Harrow A W, and Winter A 2004 *Phys. Rev. Lett.* **93** 230504

[5] Schumacher B 1995 *Phys. Rev. A* **51** 2738

[6] Jozsa R and Schumacher B 1994 *J. Mod. Opt.* **41** 2343

[7] Hayashi M and Matsumoto K 2002 *Phys. Rev. A* **66** 022311

[8] Meng Q, Catchpoole D, Skillicom D, and Kennedy P J 2017 *International Joint Conference on Neural Networks (IJCNN)* p. 364

[9] Vincent P, Larochelle H, Bengio Y, and Manzagol P A 2008 *Proceedings of the 25th Int. Conf. on Machine Learning (ICML)* p. 1096

[10] Fogelman-Soulié F and Le Cun Y 1987 *Intellectica* **2** 114

[11] Vincent P, Larochelle H, Lajoie I, Bengio Y, and Manzagol P A 2010 *J. Mach. Learn. Res.* **11** 3371

[12] Bourlard H and Kamp Y 1988 *Biol. Cybern.* **59** 291

[13] Hinton G E and Zemel R S 1993 *Neural Information Processing Systems*

[14] Romero J, Olson J P, and Aspuru-Guzik A 2017 *Quantum Sci. Technol.* **2** 045001

[15] Pepper A, Tischler N, and Pryde G J 2019 *Phys. Rev. Lett.* **122** 060501

[16] Huang C J, Ma H L, Yin Q, Tang J F, Dong D Y, Chen C L, Xiang G Y, Li C F, and Guo G C 2020 *Phys. Rev. A* **102** 032412

[17] Ding Y C, Lamata L, Sanz M, Chen X, and Solano E 2019 *Adv. Quantum Technol.* **2** 1800065

[18] Locher D F, Cardarelli L, and Müller M 2023 *Quantum* **7** 942

[19] Zhang X M, Kong W C, Farooq M U, Yung M H, Guo G P, and Wang X 2021 *Phys. Rev. A* **103** L040403

[20] Bondarenko D and Feldmann P 2020 *Phys. Rev. Lett.* **124** 130502

[21] Baul A, Walker N, Moreno J, and Tam K M 2023 *Phys. Rev. E* **107** 045301

[22] Ch'ng K, Vazquez N, and Khatami E 2018 *Phys. Rev. E* **97** 013306

[23] Eberz D, Link M, Kell A, Breyer M, Gao K, and Köhl M 2023 *Phys. Rev. A* **108** 063303

[24] Kharkov Y A, Sotskov V E, Karazeev A A, Kiktenko E O, and Fedorov A K 2020 *Phys. Rev. B* **101** 064406

[25] Wu Y, Zhang P, and Zhai H 2021 *Phys. Rev. Res.* **3** L032057

[26] Ghosh K J B and Ghosh S 2023 *Phys. Rev. B* **108** 165408

[27] Kottmann K, Metz F, Fraxanet J, and Baldelli N 2021 *Phys. Rev. Res.* **3** 043184

[28] Ngairangbam V S, Spannowsky M, and Takeuchi M 2022 *Phys. Rev. D* **105** 095004

[29] Luchi P, Trevisanutto P E, Roggero A, DuBois J L, Rosen Y J, Turro F, Amitrano V, and Pederiva F 2023 *Phys. Rev. Appl.* **20** 014045

[30] Schmitt M and Lenarčić Z 2022 *Phys. Rev. B* **106** L041110

[31] Szo ldra T, Sierant P, Lewenstein M, and Zakrzewski J 2022 *Phys. Rev. B* **105** 224205

[32] Sone A, Yamamoto N, Holdsworth T, and Narang P 2023 *Phys. Rev. Res.* **5** 023039

[33] Patel S, Collis B, Duong W, Koch D, Cutugno M, Wessing L, and Alsing P 2023 *Physica Scripta* **98** 045111

[34] Ma H, Huang C J, Chen C, Dong D, Wang Y, Wu R, and Xiang G Y 2023 *Automatica* **147** 110659

[35] Cerezo M, Sharma K, Arrasmith A, and Coles P J 2022 *npj Quantum Inf.* **8** 113

[36] Cao C F and Wang X 2021 *Phys. Rev. Appl.* **15** 054012

[37] Ma H, Mooney G J, Petersen I R, Hollenberg L C L, and Dong D 2024 *npj Quantum Inf.* **10** 86

[38] Hiai F and Petz D 1991 *Commun. Math. Phys.* **143** 99

[39] Vedral V 2002 *Rev. Mod. Phys.* **74** 197

[40] Fuchs C A and van de Graaf J 1999 *IEEE Trans. Inf. Theory* **45** 1216

[41] Csiszár I and Körner J 2011 *Information Theory: Coding Theorems for Discrete Memoryless Systems* (Cambridge: Cambridge University Press, 2nd edition)

[42] Groisman B, Popescu S, and Winter A 2005 *Phys. Rev. A* **72** 032317

[43] Berta M and Majenz C 2018 *Phys. Rev. Lett.* **121** 190503

[44] Zhou D L 2009 *Phys. Rev. A* **80** 022113

[45] Petz D 1986 *Commun. Math. Phys.* **105** 123

[46] Yang M, Xu C Q, and Zhou D L 2023 *Phys. Rev. A* **108** 052402

[47] Bhatia R 1997 *Matrix Analysis* (New York, NY: Springer New York)

[48] Andrews G E 1984 *Encyclopedia of Mathematics and its Applications* (Cambridge University Press)