






Quantum circuit autoencoder

Jun Wu ^{1,*} Hao Fu ¹ Mingzheng Zhu ¹ Haiyue Zhang ² Wei Xie ^{1,†} and Xiang-Yang Li^{1,3,‡}

¹*School of Computer Science and Technology, University of Science and Technology of China, Hefei 230027, China*

²*Department of Computer Science, University of Illinois Urbana-Champaign, Urbana-Champaign, Illinois 61801, USA*

³*Hefei National Laboratory, University of Science and Technology of China, Hefei 230088, China*



(Received 29 October 2023; revised 3 January 2024; accepted 5 March 2024; published 26 March 2024)

A quantum autoencoder is a quantum neural network model for compressing information stored in quantum states. However, one needs to process information stored in quantum circuits for many tasks in the emerging quantum information technology. In this work, generalizing the ideas of classical and quantum autoencoders, we introduce the quantum circuit autoencoder (QCAE) model to compress and encode information within quantum circuits. We provide a comprehensive protocol for QCAE and design a variational quantum algorithm, *varQCAE*, for its implementation. We theoretically analyze this model by deriving conditions for lossless compression and establishing both upper and lower bounds on its recovery fidelity. Finally, we apply *varQCAE* to three practical tasks, and numerical results show that it can effectively (1) compress the information within quantum circuits, (2) detect anomalies in quantum circuits, and (3) mitigate the depolarizing noise in quantum circuits. These suggest that our algorithm is potentially applicable to other information processing tasks for quantum circuits.

DOI: [10.1103/PhysRevA.109.032623](https://doi.org/10.1103/PhysRevA.109.032623)

I. INTRODUCTION

Quantum technologies have witnessed substantial advancements in recent years, exemplified by noteworthy experiments in quantum supremacy [1–3] on noisy intermediate-scale quantum (NISQ) devices [4]. Despite these advancements, NISQ devices have grappled with substantial noises and limited qubit numbers. To unlock the full potential of NISQ devices, it becomes imperative to deploy techniques for quantum information compression.

Autoencoders are a prevalent artificial neural network approach for compressing and encoding information [5]. In Fig. 1(a) a typical autoencoder framework is depicted, showcasing the primary concept of information compression through a bottleneck while preserving data reconstruction fidelity. Notably, in the field of quantum computing, a quantum autoencoder (QAE) has been introduced [6]. The QAE methodology involves information compression by discarding the “trash” system during the encoding step, followed by state reconstruction aided by a “reference” state. Figure 1(b) illustrates the typical diagram of a quantum autoencoder, showcasing its process for efficient quantum information compression and reconstruction. References [7,8] proposed different QAE models applied to quantum machine learning. Reference [9] analyzed the relationship between the QAE trainability and the cost function, resulting in the proposal of a local cost function version of QAE. Reference [10] investigated the reconstruction fidelity upper bound and introduced a noise QAE to enhance compression performance.

In addition to quantum states, the information processing within quantum circuits is also a common practice [11,12]. Quantum machine learning tasks often involve converting classical information into quantum information using parameterized encoding circuits. For instance, Ref. [13] utilized four-qubit circuits to transform the Iris data set (a public image data set) [14] into quantum states, with information stored in both quantum states and circuits. Reference [15] proposed a QAE-based circuit compression strategy to reduce communication costs in quantum cloud computing. While processing information within a quantum circuit in a noisy environment, the objective may shift to transforming the information into an unknown quantum channel. However, it is important to note that QAE cannot be directly applied to compressing information stored within a quantum circuit.

Considering the issues above, there is a need for an elaborate study on quantum circuit autoencoder. The quantum circuit autoencoder can also act as a generalization of QAE. For example, it can subsume QAE in some cases, such as the purified quantum query access model.

Reference [16] proposed a gate compression model that uses two unitary operators to reduce the input gate’s dimension and another two unitary operators to reconstruct the original gate. The authors also provided a method to achieve exponential reduction in dimension. Reference [15] applied a quantum autoencoder on quantum cloud computing and proposed a quantum gate autoencoder to reduce communication qubit resources. These two models can be considered prototypes of the quantum circuit autoencoder. However, they consider only quantum circuits consisting of single-qubit gates in the form of IID and a family of parameterized quantum circuits. In contrast, general quantum circuits may consist of multiple qubits and not just single-qubit gates.

*jun_wu@mail.ustc.edu.cn

†Corresponding author: xxieww@ustc.edu.cn

‡Corresponding author: xiangyangli@ustc.edu.cn

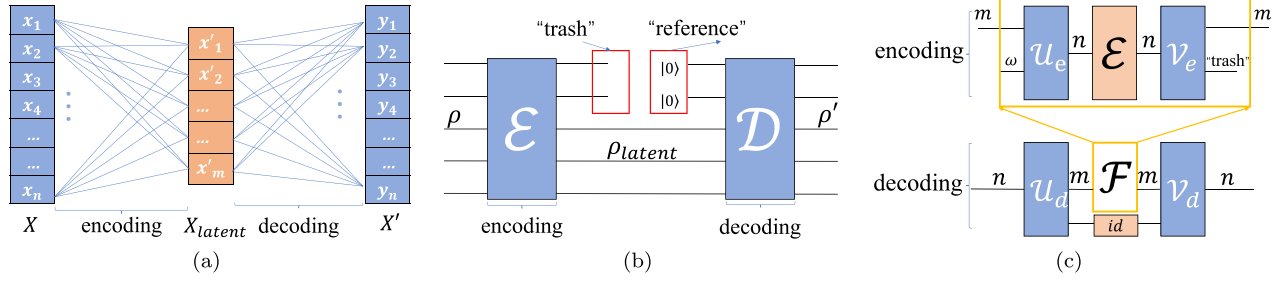


FIG. 1. Diagrams of three different autoencoders. (a) The standard *autoencoder* encodes an n -dimensional input data X into a lower dimensional representation X' of dimensionality m , which it then decodes to reconstruct the original n -dimensional data as Y . (b) The *quantum autoencoder* takes an input n -qubits state ρ , transforms it into a lower-dimensional (m -qubits) state ρ_{latent} , and then decodes this state to a n -dimensional state ρ' . (c) The *quantum circuit autoencoder* encodes a D -dimensional quantum circuit into a d -dimensional circuit and reconstructs the original D -dimensional quantum circuit through decoding.

In this paper, we propose a quantum circuit autoencoder (QCAE) model as depicted in Fig. 1(c). For a quantum channel \mathcal{E} acting on n qubits, we construct encoders \mathcal{U}_e and \mathcal{V}_e to obtain $\mathcal{F} = \text{tr}_{\text{trash}}[\mathcal{V}_e \circ \mathcal{E} \circ \mathcal{U}_e]$ acting on an m -qubit system ($m < n$), where \mathcal{U}_e , \mathcal{V}_e , and a partial trace operation constitute a supermap [17] that maps an n -qubit channel to an m -qubit channel. The goal is to maximize the reconstruction fidelity between $\tilde{\mathcal{E}} = \mathcal{V}_d \circ [\mathcal{F} \otimes \text{id}] \circ \mathcal{U}_d$ and \mathcal{E} , where id is the identity channel.

To implement the QCAE on NISQ devices, we design a variational quantum algorithm (VQA) [18], referred to as *varQCAE*. By setting the encoders and decoders as the parameterized quantum circuits (PQCs) [19], we use the classical optimizer to find optimal parameters for the quantum circuit autoencoder, obtaining executable sequences of local gates suitable for NISQ devices. A VQA consists of PQCs, loss function, and optimizer, and an inevitable issue is the Barren Plateau (BP) [20]. We use the hardware efficient ansatz [21] as the PQCs in *varQCAE*. We propose a perfect compression condition to help design the loss function and decrease computation cost. A local cost function, inspired by Ref. [9], is also designed to reduce the impact of BP. Furthermore, we analyze the fidelity bound of *varQCAE*, including an upper bound for general channels and a lower bound for special cases.

Conventional autoencoders have diverse applications, such as dimension reduction [22], anomaly detection [23], and denoising [24]. Our work employs *varQCAE* for quantum circuit tasks, including information compression, anomaly detection, and denoising on quantum circuits. We evaluate the performance of *varQCAE* on IBM qiskit [25] and Mindquantum [26]. In our demonstrations, the *varQCAE* can compress the information within parameterized quantum circuits with a reconstruction error of approximately 0.05. Moreover, the distributions of anomalous scores of “normal” and “abnormal” quantum circuits data sets are significantly different, in which we use two different ways to generate circuits in these two data sets. As for denoising, *varQCAE* can reduce the impact of depolarizing error on circuits. In summary, these results indicate that *varQCAE* can be applied to these applications.

Our main contributions: First, we have designed a protocol for the quantum circuit autoencoder. The encoding and decoding components of the autoencoder are illustrated in Fig. 1(c). Second, we have theoretically analyzed the protocol

and established the conditions necessary for the faithful recovery of quantum circuits, which is crucial for designing the loss function. Third, we have enhanced the prototype protocol from various perspectives, such as designing a variational quantum algorithm called *varQCAE* and proving its reconstruction fidelity bound. Finally, we have numerically applied *varQCAE* to several applications, including information compression, anomaly detection, and quantum circuit denoising. We evaluated the performance on IBM qiskit [25] and Mindquantum [26].

II. PRELIMINARIES

A quantum system A corresponds to a Hilbert space \mathcal{H}_A . The *quantum state* of system A is described by a density operator on \mathcal{H}_A , a positive semidefinite operator with trace one. A quantum state ρ is called *pure* if it has rank one and is called *mixed* otherwise.

In this work, we denote the *maximally mixed state* as $\omega = \mathbb{1}/d$ for a d -dimensional system and the *maximally entangled state* as $\phi^+ = (1/d) \sum_{i,j=1}^d |i\rangle\langle j| \otimes |i\rangle\langle j|$ for a d^2 -dimensional bipartite system. The *fidelity* between two quantum states ρ and σ is defined as

$$F(\rho, \sigma) := \|\sqrt{\rho}\sqrt{\sigma}\|_1^2 = (\text{tr} \sqrt{\sqrt{\rho}\sigma\sqrt{\rho}})^2, \quad (1)$$

with a special case $F(\rho, |\psi\rangle\langle\psi|) = \langle\psi|\rho|\psi\rangle$.

A *quantum operation* (or *quantum channel*) $\mathcal{E}_{A \rightarrow B}$ with input system A and output system B is a completely positive, trace-preserving linear map from the linear operators on \mathcal{H}_A to the linear operators on \mathcal{H}_B . We use id to denote the identity quantum channel, which means $\text{id}(\rho) = \rho$ for any state ρ . The *mixed-unitary quantum channel* is the convex combination of unitary operations. For a series of quantum circuits U_1, U_2, \dots , we can utilize a controlled circuit to implement a mixed-unitary channel in practice [27].

In this work, subscripts indicate the input and output systems, and we omit the identity operator $\mathbb{1}$ when it does not introduce ambiguity. For instance, $X_A Y_B \equiv Y_B X_A \equiv X_A \otimes Y_B$ denotes applying $X_A \otimes Y_B$ to the composite system AB . We write $X_{AB} Y_{BC} \equiv (X_{AB} \otimes \mathbb{1}_C)(\mathbb{1}_A \otimes Y_{BC})$ and $\mathcal{E}_{B \rightarrow C}(X_{AB}) \equiv (\text{id}_A \otimes \mathcal{E}_{B \rightarrow C})X_{AB}$. We also write the partial trace of a multipartite operator by directly omitting the subscript the partial trace takes on, for example, $X_B := \text{Tr}_A(X_{AB})$.

A quantum channel can be represented by a *Choi state* [28,29]. The *Choi state* of a quantum operation $\mathcal{E}_{A \rightarrow B}$ is defined as

$$\begin{aligned} J^{\mathcal{E}} &= (\text{id}_{\tilde{A} \rightarrow A} \otimes \mathcal{E}_{A \rightarrow B}) \phi_{\tilde{A}A}^+ \\ &= \frac{1}{d_A} \sum_{i,j=0}^{d_A-1} |i\rangle\langle j| \otimes \mathcal{E}(|i\rangle\langle j|), \end{aligned} \quad (2)$$

where $\phi_{\tilde{A}A}^+ = \frac{1}{d_A} \sum_{i,j=0}^{d_A-1} |i\rangle\langle j|_{\tilde{A}} \otimes |i\rangle\langle j|_A$ is an maximally entangled state, \tilde{A}, A are isomorphic systems, and $\{|i\rangle\}$ is an orthogonal basis of the input space \mathcal{H}_A .

The output of the channel $\mathcal{E}_{A \rightarrow B}$ with input ρ_A can be recovered by

$$\mathcal{E}_{A \rightarrow B}(\rho_A) = d_A \text{Tr}_A [J^{\mathcal{E}} (\rho_A^T \otimes \mathbb{1}_B)]. \quad (3)$$

Consider a quantum channel $\Pi_{A_1 A_2 \rightarrow B_1 B_2}$ with a Choi state represented by J^Π . In this case, the input system of Π comprises composite systems A_1 and A_2 , while the output system involves composite systems B_1 and B_2 . We can define the reduced channel $\mathcal{N}_{A_1 \rightarrow B_1}$ as the channel with a Choi state obtained by tracing out systems A_2 and B_2 , denoted as $\text{tr}_{A_2 B_2}(J^\Pi)$.

In this work, the similarity between two quantum channels is characterized by the fidelity of their respective Choi states. To be specific, we define the fidelity of two quantum channels \mathcal{E}_1 and \mathcal{E}_2 ,

$$F(\mathcal{E}_1, \mathcal{E}_2) := F(J^{\mathcal{E}_1}, J^{\mathcal{E}_2}), \quad (4)$$

where $F(\cdot)$ is the fidelity defined as in Eq. (1).

III. METHOD

A. Sketch of our method

We present the diagram of our QCAE model. The goal is to find encoders and decoders to encode \mathcal{E} through a bottleneck and decode it to original circuits as faithfully as possible. We design the *varQCAE*, a variational quantum algorithm, to implement QCAE. Our algorithm uses the parameterized quantum circuits controlled by a set of parameters to represent the encoders and decoders. Therefore, *varQCAE* aims to find the optimal control parameters to maximize the similarity between original and reconstructed quantum channels.

The QCAE, as shown in Fig. 1(c), consists of two separate processes: encoding and decoding. During the encoding process, the training data set $\{U_i\}_{i=1}^{N_{\text{train}}}$ is encoded as a mixed-unitary quantum channel \mathcal{E} on an n -qubit system. For an arbitrary state ρ , the mixed-unitary quantum channel \mathcal{E} can be written as

$$\mathcal{E}(\rho) = \sum_{i=1}^{N_{\text{train}}} p_i U_i \rho U_i^\dagger. \quad (5)$$

Then the encoders $\mathcal{U}_e(\theta)$ and $\mathcal{V}_e(\theta)$ act on the channel \mathcal{E} and obtain the reduced channel \mathcal{F} by partially tracing the last $(n - m)$ qubits (i.e., “trash” systems). As a result, the encoders and partial trace together form an operator that supermaps a 2^n -dimensional channel \mathcal{E} to a 2^m -dimensional channel \mathcal{F} . In the decoding process, the decoders $\mathcal{U}_d(\theta)$ and $\mathcal{V}_d(\theta)$ are applied to the channel $\mathcal{F} \otimes \text{id}$ to yield a new quantum channel $\tilde{\mathcal{E}}$. Finally, we feed the similarity between \mathcal{E} and $\tilde{\mathcal{E}}$ to the

classical optimizer to update parameters θ . Repeat these steps until the loss is converged or other termination condition is satisfied.

In the *varQCAE*, we set the decoders in the decoding process to be the conjugate transpose of the encoders, i.e., $\mathcal{U}_d = \mathcal{U}_e^\dagger$ and $\mathcal{V}_d = \mathcal{V}_e^\dagger$. Therefore, we could consider only the encoding process and omit the subscript in encoders and decoders for convenience. We apply $\Pi := \mathcal{V}(\theta) \circ \mathcal{E} \circ \mathcal{U}(\theta)$ on a product state $\omega_{A'} \otimes \phi_{C_1 \tilde{C}_1}^+$, where $\omega_{A'}$ is the maximally mixed state on the latent subsystem A' and $\phi_{C_1 \tilde{C}_1}^+$ is the maximally entangled state on the “trash” subsystem C_1 and the isomorphic system \tilde{C}_1 of C_1 . The reason for using this initial state relates to the construction of the loss function, which will be explained in Sec. III B and Proposition 1. The resulting state is on the composite system $B' C_2 \tilde{C}_2$. Considering utilizing the decoding scheme as the validation, the decoders are applied on $\mathcal{F} \otimes \text{id}$, assuming that the channel $\Pi_{C_1 \tilde{C}_1 \rightarrow C_2 \tilde{C}_2}$ is the identity channel. As a consequence, the goal of *varQCAE* is to maximize the fidelity of $\phi_{C_1 \tilde{C}_1}^+$ and $\phi_{C_2 \tilde{C}_2}^+$. Then we use the classical optimizer to find the near-optimal control parameters. The PQC used for encoders is the hardware-efficient ansatz [21] in our variational algorithm, and the structure is shown in Fig. 2(b).

See Algorithm 1 and Fig. 2 for more details of *varQCAE*.

ALGORITHM 1. Main Algorithm: *varQCAE*

Input: Training data $\mathcal{D}_{\text{train}} = \{U_1, \dots, U_n\}$, and the circuit ansatzes $\mathcal{U}(\theta)$ and $\mathcal{V}(\theta)$ and the number of iterations *ITR*;

1: **Training Process:**

2: Set epoch $t = 0$ and initialize θ_0 randomly;

3: Let encoders $\mathcal{U}_e(\theta) \leftarrow \mathcal{U}(\theta)$ and $\mathcal{V}_e(\theta) \leftarrow \mathcal{V}(\theta)$;

4: **while** not converged or $t \leq \text{ITR}$ **do**

5: $t \leftarrow t + 1$;

6: Initialize loss $\mathcal{L}(\theta_t) = 0$;

7: **for** each U_i in $\mathcal{D}_{\text{train}}$ **do**

8: $\Pi \leftarrow \mathcal{V}(\theta_t) \circ U_i \circ \mathcal{U}(\theta_t)$;

9: $\psi_{B' C_2 \tilde{C}_2} \leftarrow \Pi(\omega_{A'} \otimes \phi_{C_1 \tilde{C}_1}^+)$;

10: $\mathcal{L}'(\theta_t) \leftarrow \text{tr}[O \text{tr}_{B'}[\psi_{B' C_2 \tilde{C}_2}]]$, where \mathcal{L} is the loss function in Eq. (6);

11: $\mathcal{L}(\theta_t) \leftarrow \mathcal{L}(\theta_t) + \frac{1}{n}[1 - \mathcal{L}'(\theta_t)]$;

12: **end for**

13: Update parameters θ_{t+1} of $\mathcal{L}(\cdot)$ using classical optimizer;

14: **end while**

15: **Output** the near-optimal parameters θ^* ;

16: **Reconstruction Validation Process:**

17: Initialize the reconstruction error $L_{\text{val}} = 0$;

18: Let decoders $\mathcal{U}_d(\theta) \leftarrow \mathcal{U}(\theta)^\dagger$ and $\mathcal{V}_d(\theta) \leftarrow \mathcal{V}(\theta)^\dagger$;

19: **for** each U_i in $\mathcal{D}_{\text{train}}$ **do**

20: Apply encoders to obtain

$\mathcal{F}_i \leftarrow \text{tr}_{C_1 \tilde{C}_1 C_2 \tilde{C}_2}[\mathcal{V}_e(\theta^*) \circ U_i \circ \mathcal{U}_e(\theta^*)]$;

21: $\tilde{U}_i \leftarrow \mathcal{V}(\theta^*)^\dagger \circ (\mathcal{F}_i \otimes \text{id}) \circ \mathcal{U}(\theta^*)^\dagger$;

22: Calculate the reconstruction fidelity $F(J^{U_i}, J^{\tilde{U}_i})$;

23: $L_{\text{val}} \leftarrow L_{\text{val}} + \frac{1}{n}(1 - F(J^{U_i}, J^{\tilde{U}_i}))$;

24: **end for**

25: **Output** The reconstruction error L_{val} ;

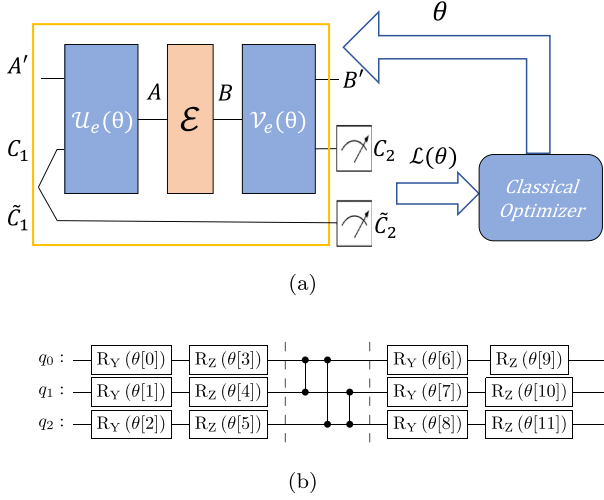


FIG. 2. (a) Diagram of the *varQCAE*. In the training process of *varQCAE*, we consider only the encoding process in the whole QCAE framework in Fig. 1(c), as we set the decoders as the dagger of the encoders. We apply the quantum channel $\Pi := \mathcal{V}_e \circ \mathcal{E} \circ \mathcal{U}_e$ on the initial state $\omega_{A'} \otimes \phi_{C_1\tilde{C}_1}^+$. The resulting state is on the composite system $B'C_2\tilde{C}_2$, and we compare the state fidelity between two states on the subsystems $C_1\tilde{C}_1$ and $C_2\tilde{C}_2$. The loss function estimates the state fidelity discussed in Sec. III B. (b) The hardware-efficient ansatz we used to construct the encoders.

B. Loss function

In this section, we define the loss function of the training process in *varQCAE*. Given the data set $\mathcal{D} := \{U_k\}_{k=1}^N$ and parameterized quantum circuits $U(\theta)$ and $V(\theta)$, the loss function is designed as

$$\mathcal{L}(\mathcal{D}, \theta) := 1 - \frac{1}{N} \sum_{k=1}^N \left[\text{tr} \left(O \text{tr}_{B'} \left[(\mathcal{V}(\theta) \circ U_k \circ \mathcal{U}(\theta)) \left(\omega_{A'} \otimes \phi_{C_1\tilde{C}_1}^+ \right) \right] \right) \right], \quad (6)$$

where O is a local observable and

$$O = \sum_{k=0}^{n-m-1} \phi_{k,k+(n-m)}^+ \otimes \mathbb{1}_{\overline{k,k+(n-m)}}, \quad (7)$$

where n and m are the number of system qubits of the original and latent circuits, respectively. Here $\phi_{k,k+(n-m)}^+$ is the maximally entangled state on the k and $k+(n-m)$ subsystems, which can be written as

$$\begin{aligned} \phi_{k,l}^+ &= \frac{1}{2} [|00\rangle\langle 00| + |00\rangle\langle 11| + |11\rangle\langle 00| + |11\rangle\langle 11|]_{k,l} \\ &= \frac{1}{2} [I + Z_k Z_l + X_k X_l - Y_k Y_l]. \end{aligned} \quad (8)$$

The goal of QCAE is to maximize the reconstruction fidelity between the original channel \mathcal{E} and the reconstruction channel $\tilde{\mathcal{E}}$, and a straightforward method is to calculate the reconstruction error

$$\begin{aligned} \text{error}_1 &= 1 - F(\mathcal{E}, \tilde{\mathcal{E}}) \\ &= 1 - \text{tr} [J^\mathcal{E} J^{\tilde{\mathcal{E}}}]. \end{aligned} \quad (9)$$

However, this method falls short as a proper loss function for several reasons. First, error_1 computes fidelity between two 4^n -dimensional quantum states, incurring prohibitively high computational costs. Second, the observable in this function remains neither fixed nor explicit. Due to these limitations, we opt the function as

$$\begin{aligned} \text{error}_2 &= 1 - F(\phi_{C_1\tilde{C}_1}^+, \psi_{C_2\tilde{C}_2}) \\ &= 1 - \text{tr} [\phi_{C_1\tilde{C}_1}^+ \psi_{C_2\tilde{C}_2}], \end{aligned} \quad (10)$$

which is to compare two quantum states on the input subsystem $C_1\tilde{C}_1$ and output subsystem $C_2\tilde{C}_2$. These two quantum states' dimensions are $4^{(n-m)}$. As a result, this method reduces the computational cost, and the observable in this method is the maximally entangled state. In addition, we proposed the *perfect compression condition* in Proposition 1 to claim them both equal to 0 when the perfect compression is achieved.

Furthermore, there is an inevitable issue that the gradient exponential vanishes in a variational quantum algorithm. This issue is called the Barren Plateau (BP) [20] problem and has been studied in many works, such as Refs. [9,30]. In Ref. [9] the relationship between cost function and BP has been discussed. In addition, the authors demonstrated that the local cost function can reduce the adverse effects of BP. Inspired by this idea, we design the loss function in Eq. (6) to reduce the impact of BP in *varQCAE*.

In Ref. [9] the reconstruction error function, which directly compares the fidelity of two quantum states, is defined as the *global cost function*; the loss function, such as the \mathcal{L} in Eq. (6), which is the summation of the expectations of local observables, is defined as the *local cost function*. The authors prove that the global cost function leads to exponentially vanishing gradient even though the ansatz is shallow and that the local cost function leads to, at worst, polynomially vanishing gradients.

Given the *varQCAE*, we briefly discuss analytic gradient and the BP issues that the expectation gradient is approximate to zero exponentially. The ansatz we use is a two-local parameterized quantum circuit deployed as the sequence of single-qubit rotations and two-qubit gates. Therefore, we can use the parameter-shift rule [31] to obtain the partial derivative:

$$\frac{\partial \mathcal{L}}{\partial \theta_j} = \mathcal{L}(\theta_+) - \mathcal{L}(\theta_-), \quad (11)$$

where θ_+ and θ_- are different from θ only at the j th parameter: $\theta_j \rightarrow \theta_j \pm \frac{\pi}{4}$. As a result, if we have access to a quantum computer, we can utilize it to estimate the gradients in polynomial time.

Regarding the BP issue that plagues many variational quantum algorithms, we would like to highlight two critical aspects in the context of our *varQCAE*. First, our approach involves performing $(n-m)$ -qubit measurements and utilizing the outcomes as the basis for the loss function. Second, the observable in the loss function [refer to Eq. (6)] is a summation of several two-qubit observables. These two considerations render the loss function local, and it has been demonstrated that a shallow parameterized quantum circuit (PQC) with a local loss function, at worst, exhibits a polynomially vanishing gradient [9]. Consequently, our *varQCAE* may mitigate the

barren plateau issue, particularly when the number of layers L is $O(\log(n))$. Figure 6 in Appendix D provides a visualization of the landscape of *varQCAE*.

In addition, we analyze the algorithm cost from several aspects. *The number of the control parameters*: we employ a hardware-efficient ansatz which entails a total of $8np$ control parameters in *varQCAE*, where n is the system size and p is the repetition times of the ansatz used in the algorithm. *The number of gates*: There are $4np$ single rotation gates and $n(n-1)p$ CNOT gates.

C. Reconstruction validation

After completing the training process for *varQCAE*, it is crucial to develop a strategy for evaluating the performance of the training results. During training, our focus is exclusively on the encoding part. This focus must be extended to the entire scheme in the validation process.

After training, we obtained the near-optimal parameters θ^* and the reduced channel \mathcal{F} from Π , $\mathcal{F} = \text{tr}_{\tilde{C}_1\tilde{C}_2\tilde{C}_3}[\mathcal{V}(\theta^*) \circ \mathcal{E} \circ \mathcal{U}(\theta^*)]$. Finally, $\tilde{\mathcal{E}} = \mathcal{V}^\dagger(\theta^*) \circ (\mathcal{F} \otimes \text{id}) \circ \mathcal{U}^\dagger(\theta^*)$ is obtained in the decoding process. In the reconstruction validation part, we will evaluate the infidelity between two quantum channels \mathcal{E} and $\tilde{\mathcal{E}}$. A simple way is to calculate the infidelity between their Choi states, as shown in Eq. (9). Two issues must be addressed in the validation process. One is to obtain the reduced channel \mathcal{F} , and another is to obtain the product channel $\mathcal{F} \otimes \text{id}$.

To address the first issue, we propose an equivalence problem: extracting the Choi state $J^{\mathcal{F}}$ of \mathcal{F} from J^Π . After preparing the Choi state J^Π for channel Π , the reduced Choi state $J^{\mathcal{F}} = \text{tr}_{[m+1,n],[n+m+1,2n]} J^\Pi$ is obtained by performing a partial trace on the subsystems from the m th to n th qubits and from $(n+m)$ -th to $2n$ -th qubits. As a result, the reduced Choi state of $J^{\mathcal{F}}$ is obtained from Π .

The second issue is efficiently constructing the Choi state $J^{\mathcal{F} \otimes \text{id}}$. We obtained the reduced Choi state $J^{\mathcal{F}}$ from the previous problem and calculated the Choi state of the identity channel, J^{id} . A direct strategy involves obtaining the state $J^\Pi \otimes J^{\text{id}}$ and applying the swap operator to adjust the subsystems, yielding $J^{\mathcal{F} \otimes \text{id}}$. The swap operator swaps the $[m+1, 2m]$ qubit subsystem with the $[2m+1, m+n]$ qubit subsystem. The issue of designing the circuit consisting of swap circuits is equivalent to a permutation problem. For this circuit design problem, we propose a strategy, the details of which are presented in Appendix G.

IV. THEORETICAL ANALYSIS

In this section we present the critical theoretical findings, including the perfect compression condition and the fidelity bound associated with *varQCAE*. The perfect compression condition can justify our choice of loss function \mathcal{L} . Furthermore, the upper bound on reconstruction fidelity implies that the efficacy of our method is constrained by the rank of the input quantum channel, a parameter intricately tied to the quantity of input quantum circuits. Additionally, the lower bound on reconstruction fidelity, which is under the consideration of the input channel as the depolarizing channel, serves as a performance guarantee for our algorithm.

An important step of the loss function design is changing the reconstruction error estimation from Eq. (9) to Eq. (10). The change is driven by the objective of downsizing the measurement system from n qubits to $(n-m)$ qubits, a critical step in reducing computational costs. The crucial observation facilitating the transformation is that both estimations converge to 0 when the input channel can be perfectly recovered after compression. The subsequent proposition provides the analytical insight into this transformation:

Proposition 1. (Perfect compression condition) The channel \mathcal{E} can be recovered from \mathcal{F} by the recovery scheme illustrated in Fig. 1(c) if and only if

$$\text{tr}_{B'} \Pi(\omega_{A'} \otimes \phi_{C_1\tilde{C}_1}^+) = \phi_{C_2\tilde{C}_2}^+, \quad (12)$$

where ϕ^+ is the maximally entangled state, ω denotes the maximally mixed state, and $\Pi = \mathcal{V} \circ \mathcal{E} \circ \mathcal{U}$ is the channel obtained by applying encoders to \mathcal{E} .

The proof is shown in Appendix A. Proposition 1 indicates that the recovery of a quantum channel after compression is feasible if the origin channel can be processed as a product of a compressed channel and an identity channel under the influence of two unitary operators. This proposition implies the feasibility of achieving the learning task, namely, finding the optimal \mathcal{U} and \mathcal{V} , by training solely on the “trash” state. This concept underpins the design of the reconstruction error (10).

As an information compression method, it is imperative to assess its performance in terms of recovery. We provide the upper and lower bounds on reconstruction fidelity for *varQCAE*, with the lower bound derived under the assumption that the input channel is depolarizing.

Lemma 2. Consider quantum states ρ and σ , with r being the rank of σ . The fidelity between ρ and σ is bounded above by the sum of the largest r eigenvalues of ρ . This bound is attained if and only if $\rho = \sigma$.

Lemma 2 provides an upper bound on the fidelity between any two quantum states, a result instrumental in proving the following proposition.

Proposition 3. Consider $\tilde{\mathcal{E}}$ as the recovered quantum channel from $\mathcal{V} \circ \mathcal{E} \circ \mathcal{U}$, the recovery fidelity $F(\tilde{\mathcal{E}}, \mathcal{E})$ is bounded above by the sum of the largest d^2 eigenvalues of the Choi state of \mathcal{E} , where d is the dimension of the reduced quantum channel $\mathcal{F} = \text{tr}_{\text{trash}}[\mathcal{V} \circ \mathcal{E} \circ \mathcal{U}]$.

The proofs for these two results are detailed in Appendix B. Drawing inspiration from this proposition, it becomes evident that the reconstruction fidelity via *varQCAE* may not be optimal when the rank exceeds d^2 . For instance, consider the completely depolarizing quantum channel Δ with input and output dimensions D and compress it to d dimensions. The Choi state of Δ is $\text{diag}(\frac{1}{D^2}, \dots, \frac{1}{D^2})$. According to the proposition mentioned above, even under the best-case training scenario, the fidelity of reconstruction remains bounded by $\frac{d^2}{D^2}$.

While upper bounds provide valuable insights, lower bounds are also crucial as they offer a performance guarantee, at least for some special cases. In this study we explore the lower bound of reconstruction fidelity when the input channel is the depolarizing channel, as outlined in the following proposition.

Proposition 4. For a given depolarizing channel \mathcal{E}_p with dimension D , utilizing *var*QCAE to compress it to a d -dimensional quantum channel and subsequently recover it to D -dimensional, the lower bound on the reconstruction fidelity is given by

$$F(\mathcal{E}_p, \tilde{\mathcal{E}}_p) \geq \left[\sqrt{\left(\frac{p}{D^2} + 1 - p\right)\left(\frac{p}{d^2} + 1 - p\right)} + (d^2 - 1)\frac{p}{Dd} \right]^2. \quad (13)$$

The proof is presented in Appendix C.

V. APPLICATIONS AND DEMONSTRATIONS

In this section we delve into practical applications of *var*QCAE, with a primary focus on quantum circuit information compression, the fundamental objective motivating the proposal of *var*QCAE. Additionally, we explore two other applications: *var*QCAE-based anomaly detection and denoising for quantum circuits.

In our demonstrations, we applied compression and reconstruction techniques to multiple quantum circuits, achieving remarkably low reconstruction error rates, with approximately 0.05. Moreover, *var*QCAE has demonstrated remarkable efficacy in detecting “abnormal” data from “normal” data and mitigating the noise on quantum circuits.

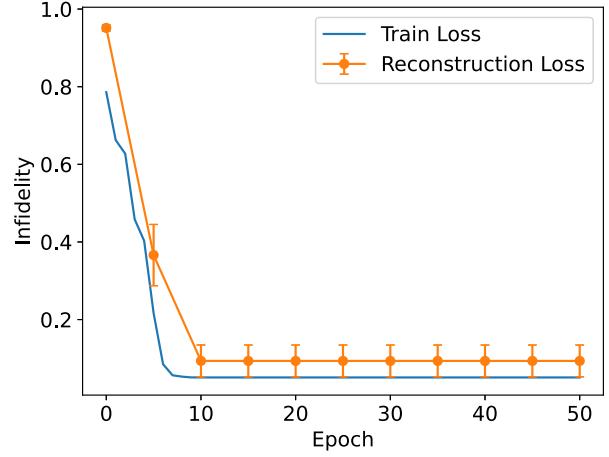
We utilized the quantum platforms Qiskit 0.45.2 and Mindquantum 0.9.0 in our demonstrations. The optimizer used in this work is the L-BFGS-B optimizer in Scipy 1.12.0. The updated code is available at the GitHub repository [32].

A. Quantum circuit information compression

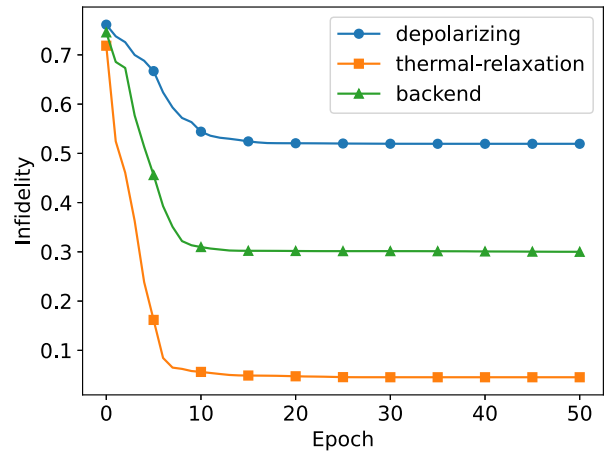
This section investigates the capability of leveraging *var*QCAE for information compression on quantum circuits. As illustrated in Sec. III, we consider a set of quantum circuits $\{U_i\}_{i=1}^N$ with a dimension D , constructing these circuits as a mixed quantum channel \mathcal{E} . The encoding process involves finding a supermap to map the D -dimensional quantum channel \mathcal{E} to a d -dimensional channel \mathcal{F} . The encoding process compresses the information within the quantum circuits. A critical metric for evaluating compression performance is the reconstruction fidelity between the original channel \mathcal{E} and the reconstructed channel $\tilde{\mathcal{E}}$ which is recovered from \mathcal{F} .

Our demonstrations focused on compressing information within parameterized quantum circuits (PQCs). PQCs serve as widely used encoding tools for translating classical information into quantum information in quantum neural networks (QNNs), owing to their solid expressive power. Specifically, we target the RealAmplitudes from the qiskit circuit library as the PQCs for compression. The parameters required for these circuits are independently generated using a normal distribution.

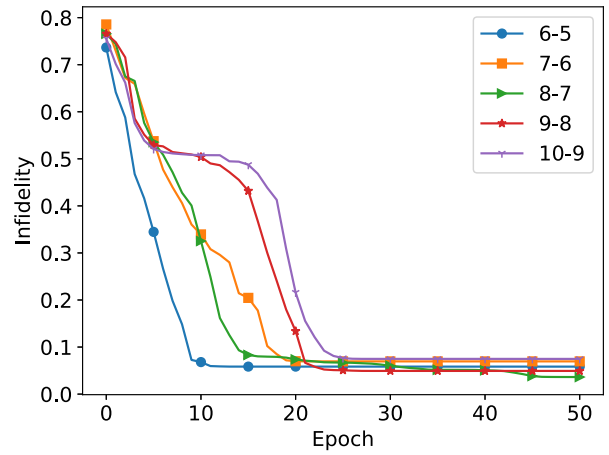
Figure 3 presents numerical results for evaluating the performance of *var*QCAE on quantum circuit information compression. There are three key demonstrations conducted:



(a)



(b)



(c)

FIG. 3. Training process of *var*QCAE for circuit information compression. (a) Changes of loss function and reconstruction errors in the training process on a five-qubit system. (b) Changes of loss function in the training process under the different noises. (c) Changes of loss function in the training process with the input circuits’ qubits are arranged from 6 to 10; the labels are the qubit pairs that are the qubits number of the input channel and the latent channel.

(1) In Fig. 3(a) the changes in the loss function (blue) and reconstruction error (orange) are shown during the training process. The input circuits consist of five qubits using the RealAmplitudes construction. The number of input quantum circuits is 10, and the latent circuits in *varQCAE* are set to four. The quantum simulator used in this demonstration is noiseless. The results of this demonstration validate the rationality of constructing the loss function \mathcal{L} as described in Eq. (6).

(2) Figure 3(b) illustrates the change in the loss function during the training process, subject to three different types of noise, including depolarizing error, thermal-relaxation, and the FakeTokyo noise in qiskit. The input circuits in this demonstration consist of four qubits using the RealAmplitudes construction. The number of input quantum circuits is 10, and the latent circuits in *varQCAE* set to three.

(3) In Fig. 3(c) the change in the loss function during the training process is depicted. Here the labels represent the qubit pair representation of the qubit numbers for both the input and latent circuits. This demonstration showcases that *varQCAE* is capable of convergence even with relatively large system sizes.

Overall, these demonstrations provide valuable insights into the performance and capabilities of *varQCAE* in quantum circuit information compression.

We selected L-BFGS-B as the classical optimizer and set the training epoch to 50. The super parameters' sensitivity analysis is shown in Appendix E.

B. Anomaly detection

In this section we apply *varQCAE* to identify anomalies in quantum circuits. Considering the scenario of chip anomaly detection, the objective is to identify abnormal chips within a collection of quantum chips. Classical data anomaly detection method may not be seamlessly applicable in this scenario. The *varQCAE*, leveraging variational algorithms, offers a solution tailored to the intricacies of quantum circuit data.

Conventional autoencoder can be used for anomaly detection [33]. After thorough training, the autoencoder learns the pattern of “normal” data. Subsequently, when presented with new data, the autoencoder compresses and reconstructs the input based on the learned patterns. “Normal” data conforming to the learned distribution tend to yield lower reconstruction errors, while anomalous data that deviate from the distribution result in higher reconstruction errors.

The specific framework is the following: The input is the “normal” data set X , anomalous data set $\{x^{(i)}, i = 1, \dots, N\}$ and a threshold α . Then design an autoencoder network and train it using the “normal” data set X . Next, for each data $x^{(i)}$ in the anomalous data set, we use the trained autoencoder to obtain the reconstruction $error(i)$. Finally, make the decision, and label the i th data $x^{(i)}$ as “abnormal” if $error(i) > \alpha$ and “normal” otherwise.

Similar to the conventional autoencoder, we investigate applying *varQCAE* to detect anomalous quantum circuit tasks. For the given “normal” quantum circuits set $\{U_{\text{normal}}\}$ and anomalous quantum circuits set, we train a quantum circuit autoencoder using $\{U_{\text{normal}}\}$. We also use the reconstruction

fidelity as anomalous scores for each circuit in the anomalous quantum circuits set. If the reconstruction fidelity is bigger than a given threshold, we label this circuit as “normal.” Otherwise, we label it as “abnormal.”

The following demonstrations demonstrate the potential of *varQCAE*-based quantum circuit anomaly detection.

Data preparation: To prepare the quantum circuit data set, we adopt the RealAmplitudes parameterized quantum circuit. Initially, we generate a quantum circuit with randomly generated parameters. Subsequently, we introduce Gaussian noise to the parameters multiple times, utilizing them to generate the original circuits' data set. The distinction between the “normal” and “abnormal” circuit data sets is achieved by evaluating the Choi state fidelity between the original circuit and each circuit in the original circuit data set. In the “normal” data set, every circuit exhibits a fidelity greater than 0.95, whereas in the “abnormal” data set, each circuit has a fidelity lower than 0.2. Anomaly detection aims to classify these two types of quantum circuits.

The results are shown in Fig. 4, providing that *varQCAE* can be highly effective in detecting abnormal data from normal data. Black bins show the distribution of the anomaly scores of the “normal” data, and orange bins show the distribution of the anomaly scores of the “abnormal” data. Figures 4(a) and 4(b) are the anomalous score distributions of our demonstrations with the noiseless and noise simulators, respectively. The noise we consider is the depolarizing error with error probability 0.01.

In Fig. 4(a) the input circuits are six qubits, and the latent circuits are five qubits. In Fig. 4(b) the input circuits are five qubits, and the latent circuits are four qubits. Moreover, the train data consist of 10 “normal” circuits; the test data consist of 40 “normal” and 40 “abnormal” circuits. We use the L-BFGS-B optimizer and set the training epoch as 50.

C. Quantum circuit denoising

In this section we consider applying *varQCAE* to denoise quantum circuits. In the NISQ era, circuit execution is limited by the effect of noise. An essential application of the conventional autoencoder is denoising data. The main idea is to extract the main character of data by autoencoder under the assumption that the noise in data is not the main feature. Reference [34] also proposed using quantum autoencoder denoise spin-flip errors and random unitary transformation errors concerning the GHZ state. In this work we consider denoising the depolarizing error on quantum circuits.

Depolarizing error: For an n -qubit quantum state ρ , the depolarizing channel error \mathcal{E}_p affects ρ according to

$$\mathcal{E}_p(\rho) = (1 - p)\rho + \sum_{\sigma^k \in \{I, X, Y, Z\}} \frac{p}{2^n} \left(\bigotimes_{i=1}^n \sigma_i^k \right) \rho \left(\bigotimes_{i=1}^n \sigma_i^k \right), \quad (14)$$

where p is the probability of being replaced, and $\sigma_i^k \in \{I, X, Y, Z\}$ is the Pauli operator acting on the i th qubit.

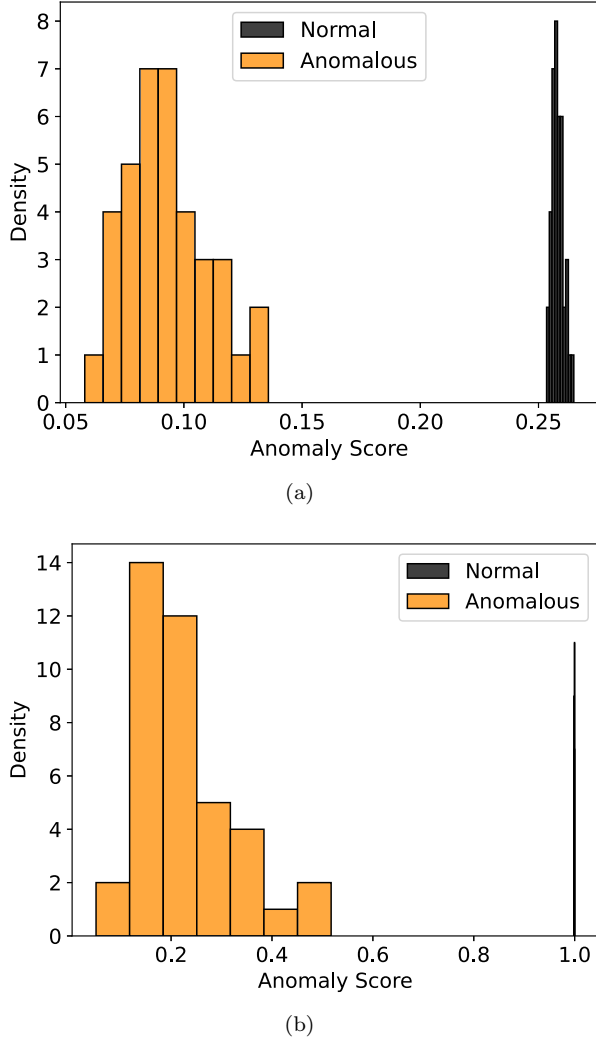


FIG. 4. Results of quantum circuits anomaly detection. Both figures are the anomaly scores distributions. (a) Anomalous score distribution without noise. We consider using the noiseless simulator; the input circuits use six qubits, and the latent circuits use five qubits. (b) Anomalous score distribution with noise. We consider using the simulator under the depolarizing noise with the error probability error 0.01, the input circuits using five qubits, and the latent circuits using four qubits. In demonstrations, we randomly choose 10 “normal” circuits to train *varQCAE* and choose 40 “normal” circuits in the rest of the “normal” data set and 40 “abnormal” circuits for estimating the reconstruction error. The blue bins show the anomaly scores of the “normal” test data set, and the orange bins show the anomaly scores of the “abnormal” test data set. The x label is the reconstruction fidelities and is set as the anomaly scores, and the y label is the density of the circuits with the same anomaly scores.

For a quantum circuit \mathcal{U} , \mathcal{U} is affected by the depolarizing channel error \mathcal{E}_p by

$$\mathcal{E}_p \circ \mathcal{U}(\rho) = (1-p)\mathcal{U}(\rho) + \sum_{\sigma^k \in \{I, X, Y, Z\}} \frac{p}{2^n} \left(\bigotimes_{i=1}^n \sigma_i^k \right) \times \mathcal{U}(\rho) \left(\bigotimes_{i=1}^n \sigma_i^k \right). \quad (15)$$

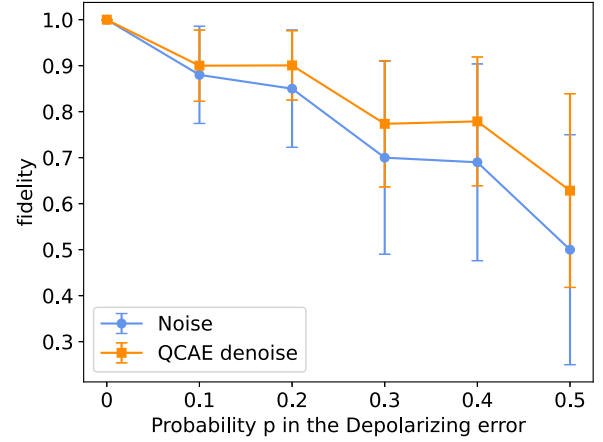


FIG. 5. Quantum circuit AutoEncoder denoises the parameterized quantum circuit under the depolarizing error. We show the average fidelity of noisy test states with the three-qubit amplitude PQC with parameters under the norm distribution $N(0, 0.6)$ before denoising (blue dots) and after denoising (orange squares). Error bars display standard deviations. 100 noisy training pairs, 100 training rounds, and the L-BFGS-B optimizer. The number of qubits in original and latent circuits is three and one.

The main components for quantum circuit denoising are as follows.

Data preparation: For a given quantum circuit \mathcal{U} , we sample a set \mathcal{U}_i from \mathcal{U} under the depolarizing channel \mathcal{E}_p . More specifically, Eq. (15) shows that the depolarizing channel is a weighted operator summation; each weight is the probability of adding the operator to the input circuit. In our demonstration, we sample an operator O_i with probability and set $\mathcal{U}_i = O_i \circ \mathcal{U}$. We can obtain the training set $\{\mathcal{U}_i\}$ by repeating this process.

*Denoising based on *varQCAE*:* In this step, we use the training data set $\{\mathcal{U}_i\}$ to train a *varQCAE* model. The training detail is shown in Sec. III A. After the training, we can use the *varQCAE* model to obtain the reconstruction data set $\{\tilde{\mathcal{U}}_i\}$.

Validation: A vital issue is evaluating the performance of the circuit denoise. We compute two indices for evaluation. One index is the sample impact, which reflects the similarity between the training set $\{\mathcal{U}_i\}$ and \mathcal{U} . We give the mean and variance of the similarities. Another index is the reconstruction impact or denoise performance, which reflects the similarity between the reconstruction set $\{\tilde{\mathcal{U}}_i\}$ and \mathcal{U} . We also calculate the mean and variance values.

Figure 5 is the result of denoising the parameterized quantum circuit under the depolarizing error. The original quantum circuit is the RealAmplitude circuit with the parameters generated by the distribution $N(0, 0.6)$. We sample 100 noise circuits and train the 100 epochs by *varQCAE*. The number of qubits of original circuits is three and one for the latent circuits. The blue dots are the mean fidelity of sample data. The orange squares are the mean fidelity after denoise. It shows that the *varQCAE* can mitigate the noise impact.

We also consider implementing the *varQCAE*-based circuit denoising on a noisy simulator, shown in Appendix F.

VI. CONCLUSION AND DISCUSSION

In this work, we introduce the quantum circuit autoencoder model and design a variational quantum algorithm for its implementation, referred to as *var*QCAE. Subsequently, we proposed the theoretical analysis to determine the condition for faithful compression, aiding in constructing the local loss function of *var*QCAE. Additionally, we establish an upper bound on the reconstruction fidelity of *var*QCAE, and calculate the fidelity lower bound for cases involving the depolarizing channel as the input channel. Moreover, we demonstrated the application of *var*QCAE in various toy scenarios, such as information compression, anomaly detection, and denoising for quantum circuits. Finally, we performed demonstrations and implemented *var*QCAE applications using the Qiskit and Mindquantum platforms.

There is much potential for further progress. (1) Determining tasks suitable for *var*QCAE. On the one hand, quantum circuit autoencoders have applications in data generation and feature extraction for information within quantum circuits. On the other hand, investigating practical applications rather than toy demonstrations in this work is also crucial work. In addition, finding more practical tasks beyond anomaly detection using Parameterized quantum circuits (PQCs) with different parameters and distributions are also appealing. (2) The reconstruction fidelity of *var*QCAE is bounded in Proposition 3. Using the noise-assisted channel to overcome the fidelity limited in *var*QCAE as discussed in Ref. [10], which uses a noise-assisted channel to overcome the fidelity limited in QAE. Reference [10] constructs the reference state by the measurement results from the “trash” subsystem after implementing the encoder, which means adding a noise channel on the “trash” subsystems. In *var*QCAE, it would be effective when using the noise channel rather than the identity channel in the reconstruction process. (3) In *var*QCAE, we use the PQCs as the encoders and decoders in QCAE. It might be more powerful to substitute the PQCs with the parameterized quantum channels. (4) This work considers only the lower bound on reconstruction fidelity for special cases, and it would be an interesting question to consider the general cases.

ACKNOWLEDGMENTS

We thank the anonymous reviewers for their valuable comments and feedback on our work. This work was partially supported by the National Natural Science Foundation of China (Grant No. 62102388), Innovation Program for Quantum Science and Technology (Grant No. 2021ZD0302901), and Anhui Initiative in Quantum Information Technologies (Grant No. AHY150100).

APPENDIX A: PROOF OF THE PERFECT COMPRESSION CONDITION

Proposition 1. (Perfect compression condition) The channel \mathcal{E} can be recovered from \mathcal{F} by the recovery scheme illustrated in Fig. 1(c) if and only if

$$\text{tr}_{B'} \Pi(\omega_{A'} \otimes \phi_{C_1\tilde{C}_1}^+) = \phi_{C_2\tilde{C}_2}^+, \quad (\text{A1})$$

where ϕ^+ is the maximally entangled state, ω denotes the maximally mixed state, and $\Pi = \mathcal{V} \circ \mathcal{E} \circ \mathcal{U}$ is the channel obtained by applying encoders to \mathcal{E} .

Proof. The condition is sufficient: If \mathcal{E} can be recovered from \mathcal{F} by the decoding scheme in Fig. 1(c) faithfully, we can get

$$\mathcal{E} = \mathcal{V}^\dagger \circ (\mathcal{F} \otimes \mathbb{1}) \circ \mathcal{U}^\dagger, \quad (\text{A2})$$

and this means

$$\Pi = \mathcal{V} \circ \mathcal{E} \circ \mathcal{V} = \mathcal{F}_{A' \rightarrow B'} \otimes \mathbb{1}_{C_1\tilde{C}_1 \rightarrow C_2\tilde{C}_2}, \quad (\text{A3})$$

which means that the channel Π is a product channel, and the subchannel on the subsystem C_1 is identity, that is, the result state on subsystems C_2 and \tilde{C}_2 is $\psi_{C_2\tilde{C}_2} = \phi^+$.

The condition is necessary: If Eq. (A1) is satisfied. Let $J_{A'C_1\tilde{C}_1B'C_2\tilde{C}_2}^\Pi$ be the Choi state of $\Pi_{A'C_1\tilde{C}_1 \rightarrow B'C_2\tilde{C}_2}$, we can deduce from Eq. (3), the result state after apply Π to initial state $\omega_{A'} \otimes \phi_{C_1\tilde{C}_1}^+$:

$$\begin{aligned} & \Pi_{A'C_1\tilde{C}_1 \rightarrow B'C_2\tilde{C}_2}(\omega_{A'} \otimes \psi_{C_1\tilde{C}_1}) \\ &= D \text{tr}_{A'C_1\tilde{C}_1} J_{A'C_1\tilde{C}_1B'C_2\tilde{C}_2}^\Pi(\omega_{A'} \otimes \phi_{C_1\tilde{C}_1}^+ \otimes \mathbb{1}_{B'C_2\tilde{C}_2}). \end{aligned} \quad (\text{A4})$$

Since $\text{tr}_{B'} \Pi(\omega_{A'} \otimes \phi_{C_1\tilde{C}_1}^+) = \phi_{C_2\tilde{C}_2}^+$, we have

$$\begin{aligned} & \text{tr}_{B'} \Pi(\omega_{A'} \otimes \phi_{C_1\tilde{C}_1}^+) \\ &= D \text{tr}_{A'C_1\tilde{C}_1B'} (J_{A'C_1\tilde{C}_1B'C_2\tilde{C}_2}^\Pi(\omega_{A'} \otimes \phi_{C_1\tilde{C}_1}^+ \otimes \mathbb{1}_{B'C_2\tilde{C}_2})) \\ &= \phi_{C_2\tilde{C}_2}^+. \end{aligned} \quad (\text{A5})$$

Since

$$\text{tr}[M_{AB}(\rho_A \otimes \mathbb{1}_B)] = \text{tr}[(\text{tr}_B M_{AB})\rho_A], \quad (\text{A6})$$

where M_{AB} is a quantum operation (or channel) on a composite system $\mathcal{H}_A \otimes \mathcal{H}_B$ and ρ_A is a density operator on \mathcal{H}_A , we have

$$\begin{aligned} \phi_{C_2\tilde{C}_2}^+ &= D \text{tr}_{A'C_1\tilde{C}_1B'} (J_{A'C_1\tilde{C}_1B'C_2\tilde{C}_2}^\Pi(\omega_{A'} \otimes \phi_{C_1\tilde{C}_1}^+ \otimes \mathbb{1}_{B'C_2\tilde{C}_2})) \\ &= \frac{D}{d} \text{tr}_{C_1\tilde{C}_1} (J_{C_1\tilde{C}_1C_2\tilde{C}_2}^\Pi(\phi_{C_1\tilde{C}_1}^+ \otimes \mathbb{1}_{C_2\tilde{C}_2})) \\ &= \Pi_{C_1\tilde{C}_1 \rightarrow C_2\tilde{C}_2}(\phi_{C_1\tilde{C}_1}^+), \end{aligned} \quad (\text{A7})$$

which implies that the reduced quantum channel $\Pi_{C_1\tilde{C}_1 \rightarrow C_2\tilde{C}_2}$ of Π is an identity channel, so the quantum channel \mathcal{F} can be deemed a d -dimensional channel with input system A' and output system B' .

As the state $\phi_{C_2\tilde{C}_2}^+$ is a maximally entangled state, so the state after we apply Π to the initial state $\omega_{A'} \otimes \phi_{C_1\tilde{C}_1}^+$ is the product state $\omega_{B'} \otimes \phi_{C_2\tilde{C}_2}^+$. This means Π can be written in the form of $\mathcal{F} \otimes \mathbb{1}$, which means that we can recover the quantum channel by

$$\begin{aligned} \mathcal{V}^\dagger \circ (\mathcal{F} \otimes \mathbb{1}) \circ \mathcal{U}^\dagger &= \mathcal{V}^\dagger \circ \Pi \circ \mathcal{U}^\dagger \\ &= \mathcal{V}^\dagger \circ \mathcal{V} \circ \mathcal{E} \circ \mathcal{U} \circ \mathcal{U}^\dagger \\ &= \tilde{\mathcal{E}}, \end{aligned} \quad (\text{A8})$$

as depicted in the decoding process of Fig. 1(c). ■

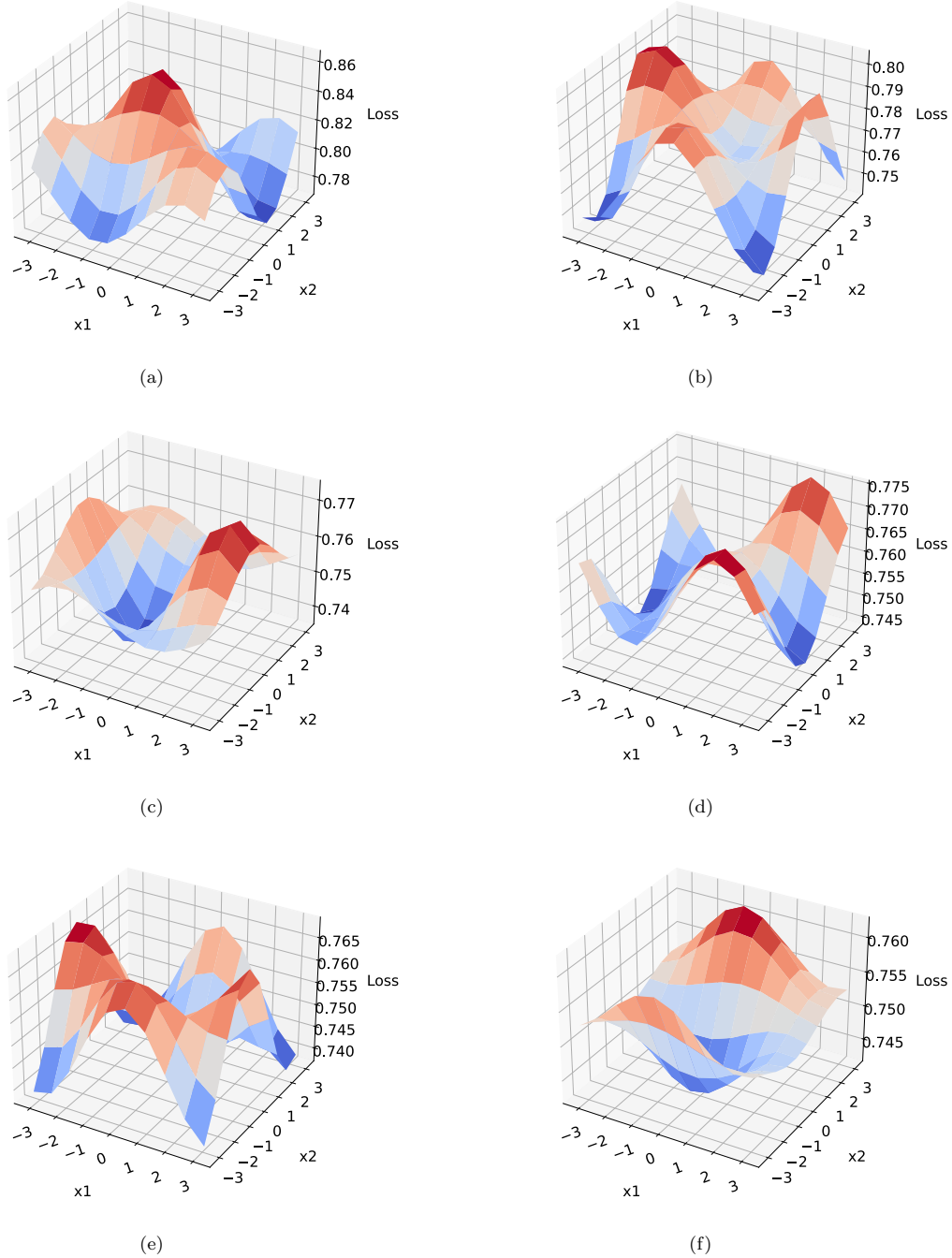


FIG. 6. Slice of loss landscape with respect to the first two circuit parameters by changing the input channels' size n and latent channel size m . Here the binary list represents (n, m) . (a) (3,2), (b) (4,3), (c) (5,4), (d) (4,2), (e) (5,3), (f) (6,4).

The inductions from Eq. (A5) to Eq. (A7) and from Eq. (A7) to Eq. (A8) use the fact that the initial state is a product state of ω and ϕ^+ .

APPENDIX B: PROOF OF THE UPPER BOUND OF varQCAE

Lemma 2. Consider quantum states ρ and σ , with r being the rank of σ . The fidelity between ρ and σ is bounded above by the sum of the largest r eigenvalues of ρ . This bound is attained if and only if $\rho = \sigma$.

Proof.

$$\begin{aligned}
 F(\rho, \sigma) &= F(\Pi_{\text{im}(\sigma)}\rho\Pi_{\text{im}(\sigma)}, \sigma) \\
 &\leq \text{tr}(\Pi_{\text{im}(\sigma)}\rho\Pi_{\text{im}(\sigma)}) \text{tr}(\sigma) \\
 &= \text{tr}(\Pi_{\text{im}(\sigma)}\rho) = \sum_{i=1}^r \lambda_i, \quad (\text{B1})
 \end{aligned}$$

where $\Pi_{\text{im}(\sigma)}$ denotes the projection onto the image of σ , $\lambda = (\lambda_1, \dots, \lambda_r)$ is the eigenvalues of ρ and $\lambda_1 \geq \dots \geq \lambda_r$, and the inequality in (B1) is due to the Proposition 3.12 in [35]. ■

Lemma 2 gives an upper bound on the fidelity between any two quantum state, and it can be used to prove the following proposition.

Proposition 3. Consider $\tilde{\mathcal{E}}$ as the recovered quantum channel from $\mathcal{V} \circ \mathcal{E} \circ \mathcal{U}$, the recovery fidelity $F(\tilde{\mathcal{E}}, \mathcal{E})$ is bounded above by the sum of the largest d^2 eigenvalues of the Choi state of \mathcal{E} , where d is the dimension of the reduced quantum channel $\mathcal{F} = \text{tr}_{\text{trash}}[\mathcal{V} \circ \mathcal{E} \circ \mathcal{U}]$.

Proof. Let $J_{A'B'}^{\mathcal{F}}$ and $J_{A'C_1B'C_2}^{\Pi}$ be the Choi state of \mathcal{F} and \mathcal{E} , respectively:

$$\begin{aligned} F(\tilde{\mathcal{E}}, \mathcal{E}) &= F(\mathcal{V}^\dagger \circ ((\mathcal{V} \circ \mathcal{E} \circ \mathcal{U})_{A' \rightarrow B'} \otimes \text{id}_{C_1\tilde{C}_1 \rightarrow C_2\tilde{C}_2}) \circ \mathcal{U}^\dagger, \mathcal{E}) \\ &= F((\mathcal{V} \circ \mathcal{E} \circ \mathcal{U})_{A' \rightarrow B'} \otimes \text{id}_{C_1\tilde{C}_1 \rightarrow C_2\tilde{C}_2}, \mathcal{V} \circ \mathcal{E} \circ \mathcal{U}) \\ &= F(J_{A'B'}^{\mathcal{F}} \otimes \phi_{C_1\tilde{C}_1, C_2\tilde{C}_2}^+ , J_{A'C_1\tilde{C}_1B'C_2\tilde{C}_2}^{\Pi}). \end{aligned} \quad (\text{B2})$$

It is easy to show that

$$\text{rank}(J_{A'B'}^{\mathcal{F}} \otimes \phi_{C_1, C_2}^+) \leq d^2. \quad (\text{B3})$$

By Lemma 2 we can get

$$\begin{aligned} F(J_{A'B'}^{\mathcal{F}} \otimes \phi_{C_1\tilde{C}_1, C_2\tilde{C}_2}^+ , J_{A'C_1\tilde{C}_1B'C_2\tilde{C}_2}^{\Pi}) \\ \leq \sum_{i=1}^{d^2} \lambda_i, \end{aligned} \quad (\text{B4})$$

where $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_{d^2})$ is the eigenvalues of $J^{\mathcal{E}}$ with $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_{d^2}$. ■

APPENDIX C: PROOF OF THE FIDELITY LOWER BOUND OF varQCAE ON COMPRESS DEPolarizing CHANNEL

Proposition 4. For a given depolarizing channel \mathcal{E}_p with dimension D , using the varQCAE to compress it to a d -dimensional quantum channel, and recover it to D -dimensional, the lower bound of the reconstruction fidelity is

$$\begin{aligned} F(\mathcal{E}_p, \tilde{\mathcal{E}}_p) \\ \geq \left[\sqrt{\left(\frac{p}{D^2} + 1 - p \right) \left(\frac{p}{d^2} + 1 - p \right)} + (d^2 - 1) \frac{p}{Dd} \right]^2. \end{aligned} \quad (\text{C1})$$

Proof. For an arbitrary quantum channel \mathcal{E} , let $\Pi_{A'C_1\tilde{C}_1 \rightarrow B'C_2\tilde{C}_2} := \mathcal{V} \circ \mathcal{E} \circ \mathcal{U}$ and $\mathcal{F}_{A' \rightarrow B'} = \text{tr}_{C_1\tilde{C}_1, C_2\tilde{C}_2} \Pi$, the recovery channel $\tilde{\mathcal{E}} = \mathcal{V}^\dagger \circ (\mathcal{F}_{A' \rightarrow B'} \otimes \text{id}_{C_1\tilde{C}_1 \rightarrow C_2\tilde{C}_2}) \circ \mathcal{U}^\dagger$. The states $J_{AB}^{\mathcal{E}}$, $J_{A'B'}^{\mathcal{F}}$, $J_{A'C_1\tilde{C}_1B'C_2\tilde{C}_2}^{\Pi}$ and $J_{AB}^{\tilde{\mathcal{E}}}$ are the Choi states of \mathcal{E} , \mathcal{F} , \mathcal{E} and $\tilde{\mathcal{E}}$, respectively.

Define the reconstruction fidelity as

$$\begin{aligned} F(\mathcal{E}, \tilde{\mathcal{E}}) &= \max_{\mathcal{U}, \mathcal{V}} F(\mathcal{E}, \mathcal{V}^\dagger \circ (\mathcal{F}_{A' \rightarrow B'} \otimes \text{id}_{C_1\tilde{C}_1 \rightarrow C_2\tilde{C}_2}) \circ \mathcal{U}^\dagger) \\ &= \max_{\mathcal{U}, \mathcal{V}} F(\mathcal{V} \circ \mathcal{E} \circ \mathcal{U}, \mathcal{F}_{A' \rightarrow B'} \otimes \text{id}_{C_1\tilde{C}_1 \rightarrow C_2\tilde{C}_2}) \\ &= \max_{\mathcal{U}, \mathcal{V}} F(\Pi_{A'C_1\tilde{C}_1 \rightarrow B'C_2\tilde{C}_2}, \mathcal{F}_{A' \rightarrow B'} \otimes \text{id}_{C_1\tilde{C}_1 \rightarrow C_2\tilde{C}_2}) \\ &= \max_{\mathcal{U}, \mathcal{V}} F(J_{A'C_1\tilde{C}_1 \rightarrow B'C_2\tilde{C}_2}^{\Pi}, J_{A' \rightarrow B'}^{\mathcal{F}} \otimes \phi_{C_1\tilde{C}_1, C_2\tilde{C}_2}^+). \end{aligned} \quad (\text{C2})$$

Setting $\mathcal{U} = \mathcal{V} = \text{id}$ yields a lower bound as follows:

$$F(\mathcal{E}, \tilde{\mathcal{E}}) \geq F_{\{\mathcal{U}=\mathcal{V}=\text{id}\}}(J_{A'C_1\tilde{C}_1 \rightarrow B'C_2\tilde{C}_2}^{\Pi}, J_{A' \rightarrow B'}^{\mathcal{F}} \otimes \phi_{C_1\tilde{C}_1, C_2\tilde{C}_2}^+). \quad (\text{C3})$$

Equation (C3) means that the reconstruction fidelity when the encoders and decoders are all is identity is a lower bound.

For the given depolarizing channel \mathcal{E}_p ,

$$J^{\mathcal{E}_p} = p\omega_{D \times D} + (1-p)\phi_D^+, \quad (\text{C4})$$

where $\omega_{D \times D} = \frac{1}{D^2}$ is the maximally entangled state and $\phi_D^+ = \sum_{i,j=0}^{D-1} |i\rangle\langle j| \otimes |i\rangle\langle j|$:

$$\begin{aligned} F_{\{\mathcal{U}=\mathcal{V}=\text{id}\}}(J_{A'C_1\tilde{C}_1 \rightarrow B'C_2\tilde{C}_2}^{\mathcal{E}_p}, J_{A' \rightarrow B'}^{\mathcal{F}} \otimes \phi_{C_1\tilde{C}_1, C_2\tilde{C}_2}^+) \\ = F\left(p\omega_{D \times D} + (1-p)\phi_D^+, p \frac{\mathbb{1}_{d \times d}}{d^2} \otimes \phi_{D/d}^+ + (1-p)\phi_D^+\right). \end{aligned} \quad (\text{C5})$$

Let $|\psi_1\rangle, \dots, |\psi_{D^2}\rangle$ be an orthogonal basis of the D^2 Hilbert space. The basis satisfies that

$$\phi_D^+ = |\psi_1\rangle\langle\psi_1| = \phi_d^+ \otimes \phi_{D/d}^+ = |\psi'_1\rangle\langle\psi'_1| \otimes |\psi''_1\rangle\langle\psi''_1| \quad (\text{C6})$$

and

$$|\psi_i\rangle = |\psi'_i\rangle \otimes |\psi''_i\rangle, \quad i \in [2, \dots, d^2], \quad (\text{C7})$$

where $\{|\psi'_i\rangle\}$ is an orthogonal basis of the d^2 Hilbert space.

The spectral decomposition of the two quantum state in the fidelity function in Eq. (C5) is

$$\begin{aligned} p\omega_{D \times D} + (1-p)\phi_D^+ &= \sum_{i=1}^{D^2} \lambda_i |\psi_i\rangle\langle\psi_i|, \\ \lambda_1 &= \frac{p}{D^2} + 1 - p, \\ \lambda_2 = \dots = \lambda_{D^2} &= \frac{p}{D^2}, \end{aligned} \quad (\text{C8})$$

and

$$\begin{aligned} p \frac{\mathbb{1}_{d \times d}}{d^2} \otimes \phi_{D/d}^+ + (1-p)\phi_D^+ &= \sum_{i=1}^{D^2} \mu_i |\psi_i\rangle\langle\psi_i|, \\ \mu_1 &= \frac{p}{d^2} + 1 - p, \\ \mu_2 = \dots = \mu_{d^2} &= \frac{p}{d^2}, \\ \mu_{d^2+1} = \dots = \mu_{D^2} &= 0. \end{aligned} \quad (\text{C9})$$

So the result in Eq. (C5) is

$$\begin{aligned} F\left(p\omega_{D \times D} + (1-p)\phi_D^+, p \frac{\mathbb{1}_{d \times d}}{d^2} \otimes \phi_{D/d}^+ + (1-p)\phi_D^+\right) \\ = \left[\text{tr} \left(\sqrt{\left(\sqrt{\sum_{i=1}^{D^2} \lambda_i |\psi_i\rangle\langle\psi_i|} \sum_{i=1}^{D^2} \mu_i |\psi_i\rangle\langle\psi_i| \sqrt{\sum_{i=1}^{D^2} \lambda_i |\psi_i\rangle\langle\psi_i|} \right)} \right) \right]^2 \end{aligned}$$

$$\begin{aligned}
 &= \left[\text{tr} \left(\sum_{i=1}^{D^2} \sqrt{\lambda_i \mu_i} |\psi_i\rangle\langle\psi_i| \right) \right]^2 \\
 &= \left[\sum_{i=1}^{D^2} \sqrt{\lambda_i \mu_i} \right]^2 \\
 &= \left[\sqrt{\left(\frac{p}{D^2} + 1 - p \right) \left(\frac{p}{d^2} + 1 - p \right) + (d^2 - 1) \frac{p}{Dd}} \right]^2.
 \end{aligned} \tag{C10}$$

APPENDIX D: THE VISUALIZATION OF THE LANDSCAPE

Figure 6 provides a visualization of the landscape of *var*QCAE, depicting a target channel formed by the convex combination of ten PQCs. As illustrated in the figure, the impact of the barren plateau is alleviated across various settings of the [original, latent] qubit pairs.

APPENDIX E: SENSITIVITY ANALYSIS OF *var*QCAE

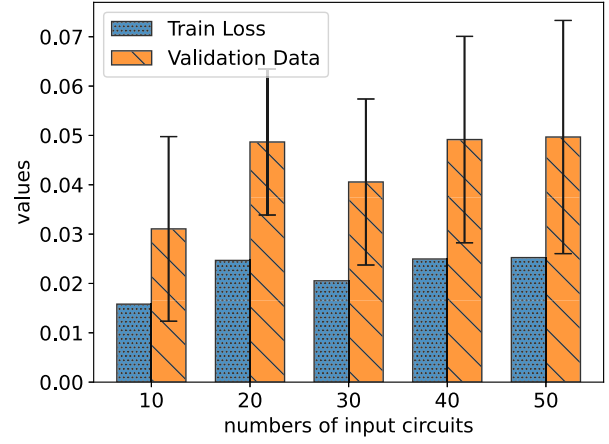
In this Appendix we experimentally analyze the impact of superparameters in *var*QCAE for compressing information within quantum circuits. Figure 7 illustrates the performance obtained when changing some settings, including the number of input circuits, the layers of the ansatzes, and the parameters' distribution.

The demonstration shows that the loss function and validation values increase with the number of input circuits. The reason is that the mixed channel's rank increases as the number of input circuits increases. In Fig. 7(a) the loss and validation are still under 0.07 when the number of input circuits increased to 50. In this demonstration, we set the number of original and latent circuits as four and three, use a one-layer ansatz to construct the input circuits, and the distribution to generate control parameters is $N(0, 0.1)$.

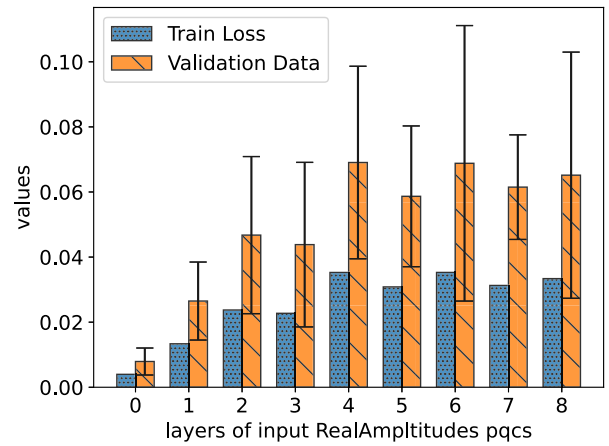
In Fig. 7(b) the results show that the loss function and validation values increase with the increase of the layers of ansatzes used in input circuits. The reason is also the increase of the rank of the mixed channel. In this demonstration, we set the number of original and latent circuits as four and three. The distribution to generate control parameters is $N(0, 0.1)$. In this demonstration, we set the number of original and latent circuits as four and three, use 20 circuits as the input, and the distribution to generate control parameters is $N(0, 0.1)$.

Figure 7(c) changes the values of σ in the distribution $N(\mu, \sigma)$; we can find that as the σ increases, the train and validation performance dramatically falls. This observation reveals that the rank of the input channel goes full as the number of circuits increases when the input circuits are all random unitaries. In this demonstration, we set the number of original and latent circuits as four and three, use 20 circuits as the input, use a one-layer ansatz to construct the input circuits, and the distribution to generate control parameters is $N(0, \sigma)$.

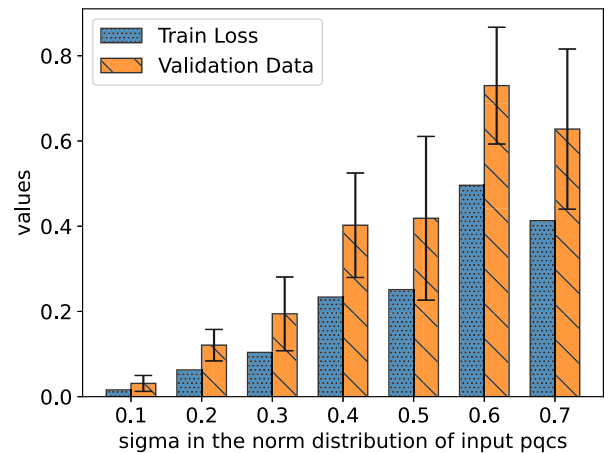
All the demonstrations show that the performance will be influenced dramatically as the rank of the input channel



(a)



(b)



(c)

FIG. 7. Sensitivity analysis on different parameters. We analyze the impacts of change in different input circuit data sets, including (a) the number of circuits, (b) the layers of each circuit, and (c) the sigma in the normal distribution to generate control parameters in input circuits.

increases, which meets the description of the upper bound of *var*QCAE in Proposition 3.

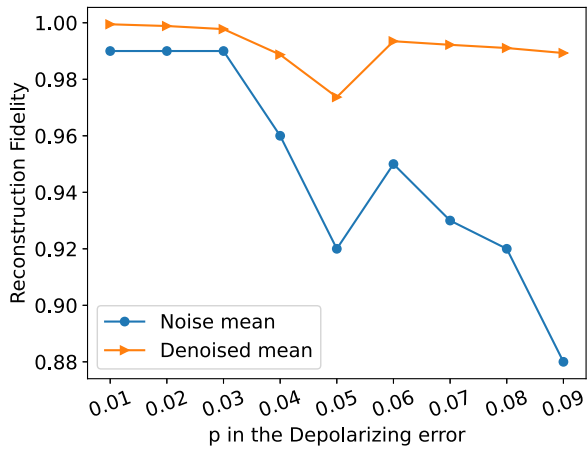


FIG. 8. Denoising performance under depolarizing errors. The input circuit consists of four qubits, while the latent circuit has three qubits. Depolarizing error probabilities range from 0.01 to 0.09, corresponding to error rates observed in NISQ devices.

APPENDIX F: QUANTUM CIRCUIT DENOISING UNDER NOISE

This section focuses on applying quantum circuit denoising using a noise simulator. Specifically, the parameterized quantum circuits within *var*QCAE are subjected to the same noise as the input circuits.

In Sec. VC we prepare the circuit data $\{U_i\}$. Multiple sampling from depolarizing noise are performed to obtain operators O_i ; These operators are then composed with the input quantum circuit U , resulting in U_i . In this section, we generate a random quantum circuit U and execute it on a simulator with noise \mathcal{N} . Their composite is $\mathcal{E} = U \circ \mathcal{N}$, which serves as an equivalent representation of the circuit data set $\{U_i\}$. Subsequently, we utilize *var*QCAE to compress and recover the circuit U on a simulator with depolarizing error.

The results of this process are illustrated in Fig. 8. The noise of the simulator is the depolarizing noise, and the noise strengths range from 0.01 to 0.09, which is the same error level as the NISQ device. The input circuit is 4 qubits and the latent circuit in *var*QCAE is three qubits. The optimizer is the L-BFGS-B in Scipy, and the training epoch is 50.

APPENDIX G: THE DIAGRAM OF DESIGNING SWAP CIRCUIT

The swap circuit construction issue, proposed in the Sec. III C, is a crucial technique to obtain the channel $\mathcal{F} \otimes \text{id}$. For a $2n + 2m$ qubit system, swap the $[n + 1, 2n]$ subsystem with the $[2n + 1, 2n + m]$ subsystem. This problem can be transformed to an equivalent permutation problem: For a number list $L_1 := [a_1, \dots, a_{2m}, b_1, \dots, b_{2n}]$, find a permutation sequence $\{[x_i, y_j] | x_i, y_j \in L_1, x_i \neq y_j\}$ to get the number list $L_2 := [a_1, \dots, a_m, b_1, \dots, b_n, a_{m+1}, \dots, a_{2m}, b_{n+1}, \dots, b_{2n}]$.

The swap circuit equivalent permutation problem can be solved by executing the SWAP operations as follow scheme:

$$\begin{aligned}
 & a_1, \dots, a_{2m}, b_1, \dots, b_{2n} \\
 & \downarrow \text{SWAP } a_{2m} \text{ with } \{b_1, \dots, b_n\} \\
 & a_1, \dots, a_{2m-1}, b_1, \dots, b_n, a_{2m}, b_{n+1}, \dots, b_{2n} \\
 & \downarrow \text{SWAP } a_{2m-1} \text{ with } \{b_1, \dots, b_n\} \\
 & a_1, \dots, a_{2m-2}, b_1, \dots, b_n, a_{2m-1}, a_{2m}, b_{n+1}, \dots, b_{2n} \\
 & \downarrow \\
 & \dots \\
 & \downarrow \text{SWAP } a_{m+1} \text{ with } \{b_1, \dots, b_n\} \\
 & a_1, \dots, a_m, b_1, \dots, b_n, a_{m+1}, \dots, a_{2m}, b_{n+1}, \dots, b_{2n}.
 \end{aligned} \tag{G1}$$

[1] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, R. Biswas, S. Boixo, F. G. Brandao, D. A. Buell *et al.*, *Nature (London)* **574**, 505 (2019).

[2] H.-S. Zhong, H. Wang, Y.-H. Deng, M.-C. Chen, L.-C. Peng, Y.-H. Luo, J. Qin, D. Wu, X. Ding, Y. Hu *et al.*, *Science* **370**, 1460 (2020).

[3] Y. Wu, W.-S. Bao, S. Cao, F. Chen, M.-C. Chen, X. Chen, T.-H. Chung, H. Deng, Y. Du, D. Fan *et al.*, *Phys. Rev. Lett.* **127**, 180501 (2021).

[4] J. Preskill, *Quantum* **2**, 79 (2018).

[5] C.-Y. Liou, J.-C. Huang, and W.-C. Yang, *Neurocomputing* **71**, 3150 (2008).

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

[7] K. H. Wan, O. Dahlsten, H. Kristjánsson, R. Gardner, and M. Kim, *npj Quantum Inf.* **3**, 36 (2017).

[8] G. Verdon, J. Pye, and M. Broughton, [arXiv:1806.09729](https://arxiv.org/abs/1806.09729).

[9] M. Cerezo, A. Sone, T. Volkoff, L. Cincio, and P. J. Coles, *Nat. Commun.* **12**, 1791 (2021).

[10] C. Cao and X. Wang, *Phys. Rev. Appl.* **15**, 054012 (2021).

[11] V. Giovannetti, S. Lloyd, and L. Maccone, *Phys. Rev. Lett.* **100**, 160501 (2008).

[12] K. Bharti, A. Cervera-Lierta, T. H. Kyaw, T. Haug, S. Alperin-Lea, A. Anand, M. Degroote, H. Heimonen, J. S. Kottmann, T. Menke *et al.*, *Rev. Mod. Phys.* **94**, 015004 (2022).

[13] E. Grant, M. Benedetti, S. Cao, A. Hallam, J. Lockhart, V. Stojevic, A. G. Green, and S. Severini, *npj Quantum Inf.* **4**, 65 (2018).

[14] R. A. Fisher, *Ann. Eugenics* **7**, 179 (1936).

[15] Y. Zhu, G. Bai, Y. Wang, T. Li, and G. Chiribella, *Quantum Mach. Intell.* **5**, 27 (2023).

[16] G. Chiribella, Y. Yang, and C. Huang, *Phys. Rev. Lett.* **114**, 120504 (2015).

[17] G. Chiribella, G. M. D’Ariano, and P. Perinotti, *Europhys. Lett.* **83**, 30004 (2008).

[18] M. Cerezo, A. Arrasmith, R. Babbush, S. C. Benjamin, S. Endo, K. Fujii, J. R. McClean, K. Mitarai, X. Yuan, L. Cincio *et al.*, *Nat. Rev. Phys.* **3**, 625 (2021).

[19] M. Benedetti, E. Lloyd, S. Sack, and M. Fiorentini, *Quantum Sci. Technol.* **4**, 043001 (2019).

- [20] J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Babbush, and H. Neven, *Nat. Commun.* **9**, 4812 (2018).
- [21] A. Kandala, A. Mezzacapo, K. Temme, M. Takita, M. Brink, J. M. Chow, and J. M. Gambetta, *Nature (London)* **549**, 242 (2017).
- [22] Y. Wang, H. Yao, and S. Zhao, *Neurocomputing* **184**, 232 (2016).
- [23] R. Chalapathy and S. Chawla, [arXiv:1901.03407](https://arxiv.org/abs/1901.03407).
- [24] L. Gondara, in *2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW)* (IEEE, New York, 2016), pp. 241–246.
- [25] Qiskit contributors, Qiskit: An open-source framework for quantum computing (2023), [10.5281/zenodo.2573505](https://zenodo.org/record/2573505).
- [26] MindQuantum Developer, Mindquantum, version 0.6.0 (2021), <https://gitee.com/mindsore/mindquantum>.
- [27] S.-J. Wei, T. Xin, and G.-L. Long, *Sci. China Phys. Mech. Astron.* **61**, 70311 (2018).
- [28] A. Jamiolkowski, *Rep. Math. Phys.* **3**, 275 (1972).
- [29] M.-D. Choi, *Linear Algebra Appl.* **10**, 285 (1975).
- [30] M. Cerezo and P. J. Coles, *Quantum Sci. Tech.* **6**, 035006 (2021).
- [31] M. Schuld, V. Bergholm, C. Gogolin, J. Izaac, and N. Killoran, *Phys. Rev. A* **99**, 032331 (2019).
- [32] <https://github.com/linke-quantum/QCAE-master>.
- [33] J. An and S. Cho, *Variational Autoencoder based Anomaly Detection using Reconstruction Probability* (2015).
- [34] D. Bondarenko and P. Feldmann, *Phys. Rev. Lett.* **124**, 130502 (2020).
- [35] J. Watrous, *The Theory of Quantum Information* (Cambridge University Press, Cambridge, 2018).