

# Quantum circuit architecture search on a superconducting processor

Kehuan Linghu,<sup>1,\*</sup> Yang Qian,<sup>2,3,\*</sup> Ruixia Wang,<sup>1,†</sup> Meng-Jun Hu,<sup>1</sup> Zhiyuan Li,<sup>1</sup> Xuegang Li,<sup>1</sup> Huikai Xu,<sup>1</sup> Jingning Zhang,<sup>1</sup> Teng Ma,<sup>1</sup> Peng Zhao,<sup>1</sup> Dong E. Liu,<sup>4,1</sup> Min-Hsiu Hsieh,<sup>5</sup> Xingyao Wu,<sup>3,‡</sup> Yuxuan Du,<sup>3,§</sup> Dacheng Tao,<sup>3,¶</sup> Yirong Jin,<sup>1</sup> and Haifeng Yu<sup>1</sup>

<sup>1</sup>*Beijing Academy of Quantum Information Sciences, Beijing 100193, China*

<sup>2</sup>*School of Computer Science, Faculty of Engineering, University of Sydney, Australia*

<sup>3</sup>*JD Explore Academy, China*

<sup>4</sup>*State Key Laboratory of Low Dimensional Quantum Physics,  
Department of Physics, Tsinghua University, Beijing, 100084, China*

<sup>5</sup>*Centre for Quantum Software and Information,  
Faculty of Engineering and Information Technology, University of Technology Sydney, Australia*

(Dated: January 5, 2022)

Variational quantum algorithms (VQAs) have shown strong evidences to gain provable computational advantages for diverse fields such as finance, machine learning, and chemistry. However, the heuristic ansatz exploited in modern VQAs is incapable of balancing the tradeoff between expressivity and trainability, which may lead to the degraded performance when executed on the noisy intermediate-scale quantum (NISQ) machines. To address this issue, here we demonstrate the first proof-of-principle experiment of applying an efficient automatic ansatz design technique, i.e., quantum architecture search (QAS), to enhance VQAs on an 8-qubit superconducting quantum processor. In particular, we apply QAS to tailor the hardware-efficient ansatz towards classification tasks. Compared with the heuristic ansätze, the ansatz designed by QAS improves test accuracy from 31% to 98%. We further explain this superior performance by visualizing the loss landscape and analyzing effective parameters of all ansätze. Our work provides concrete guidance for developing variable ansätze to tackle various large-scale quantum learning problems with advantages.

The successful exhibition of random quantum circuits sampling and Boson sampling over fifty qubits [1–4] evidences the potential of using current quantum hardware to address classically challenging problems. A leading strategy towards this goal is variational quantum algorithms (VQAs) [5, 6], which leverage classical optimizers to train an *ansatz* that can be implemented on noisy intermediate-scale quantum (NISQ) devices [7]. In the past years, a growing number of theoretical studies has shown the computational superiority of VQAs in the regime of machine learning [8–16], quantum many body physics [17–20], and quantum information processing [21–23]. On par with the achievements, recent studies have recognized some flaws of current VQAs through the lens of the tradeoff between the expressivity and learning performance [14, 24]. That is, an ansatz with very high expressivity may encounter the barren plateau issues [25–28], while an ansatz with low expressivity could fail to fit the optimal solution [29]. With this regard, designing a problem-specific and hardware-oriented ansatz is of great importance to guarantee good learning performance of VQAs and the precondition of pursuing quantum advantages.

Pioneered experimental explorations have validated the crucial role of ansatz when applying VQAs to accomplish tasks in different fields such as machine learning [30–33],

quantum chemistry [19, 34–38], and combinatorial optimization [39–42]. On the one side, envisioned by the no-free lunch theorem [43, 44], there does not exist a universal ansatz that can solve all learning tasks with the optimal performance. To this end, myriad handcraft ansätze have been designed to address different learning problems [45–47]. For instance, the unitary coupled cluster ansatz and its variants attain superior performance in the task of estimating molecular energies [48–51]. Besides devising the problem-specific ansätze, another indispensable factor to enhance the performance of VQAs is the compatibility between the exploited ansatz and the employed quantum hardware, especially in the NISQ scenario [39]. Concretely, when the circuit layout of ansatz mismatches with the qubit connectivity, additional quantum resources, e.g., SWAP gates, are essential to complete the compilation. Nevertheless, these extra quantum resources may inhibit the performance of VQAs, because of the limited coherence time and inevitable gate noise of NISQ machines. Considering that there are countless learning problems and diverse architectures of quantum devices [52–54], it is impractical to manually design problem-specific and hardware-oriented ansätze.

To enhance the capability of VQAs, initial studies have been carried out to seek feasible strategies of *automatically designing* a problem-specific and hardware-oriented ansatz with both good trainability and sufficient expressivity. Conceptually, the corresponding proposals exploit random search [55], evolutionary algorithms [56, 57], deep learning techniques [58–64], and adaptive strategies [65–67] to tailor a hardware-efficient ansatz [19], i.e., inserting or removing gates, to decrease the cost function. In contrasts

\* These two authors contributed equally

† wangrx@baqis.ac.cn

‡ wu.x.yao@gmail.com

§ duyuxuan123@gmail.com

¶ dacheng.tao@gmail.com

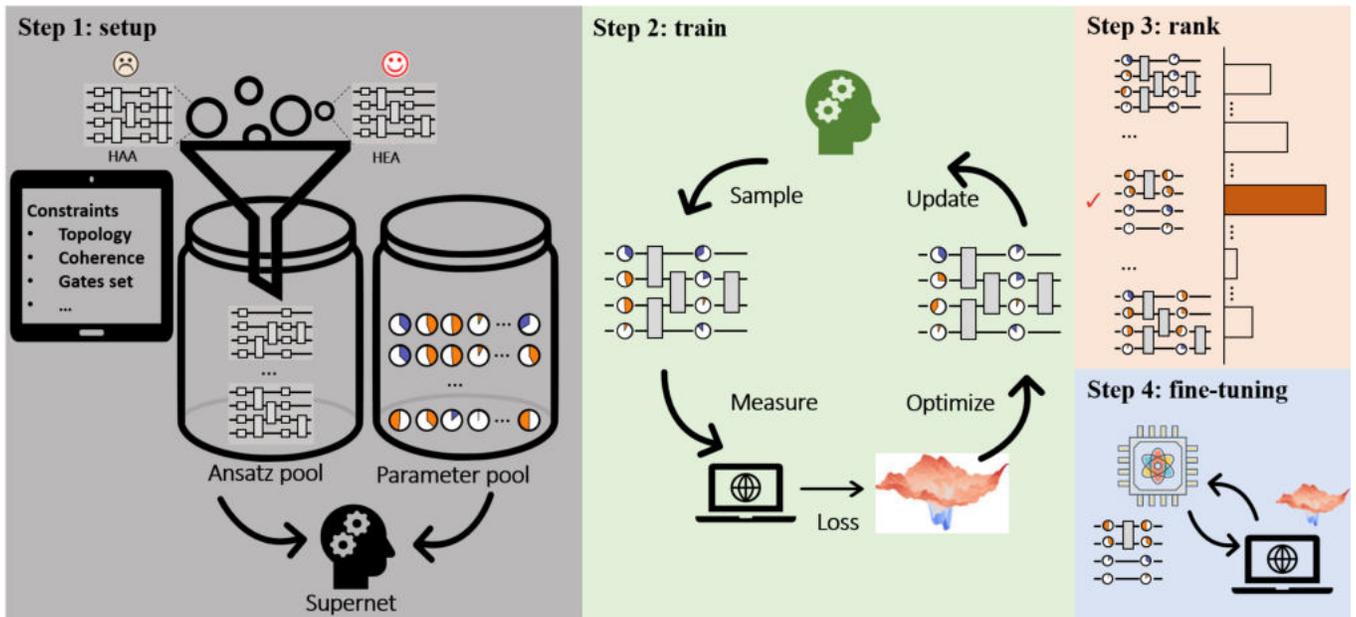


FIG. 1: **Experimental implementation of QAS.** The first step is to construct the ansätze pool  $\mathcal{S}$  which contains all candidate ansätze satisfying the hardware and physical constraints, such as hardware topology and maximal decoherence time. Meantime, the parameter pool for all candidate ansätze is initialized in a layer-by-layer manner. The gate arrangement together with corresponding parameters constitutes the supernet. The second step is to sample ansatz from the supernet, measure the observable, calculate the loss, optimize the corresponding parameters based on the objective function and update the parameters in the supernet. Repeat the above process until reaching the maximal number of iterations. Once the ansätze pool is well trained, the following steps are searching in  $\mathcal{S}$ , ranking according to the performance and selecting the optimal ansatz for fine tuning.

with conventional VQAs that only adjust parameters, optimizing both parameters and circuit layouts enable the enhanced learning performance of VQAs. Meanwhile, the automatic nature endows the power of these approaches to address broad learning problems. Despite the prospects, little is known about the effectiveness of these approaches executed on the real quantum devices.

In this study, we demonstrate the first proof-of-principle experiment of applying an efficient automatic ansatz design technique, i.e. quantum architecture search (QAS) scheme [68], to enhance VQAs on an 8-qubit superconducting quantum processor. In particular, we focus on data classification tasks and utilize QAS to pursue a better classification accuracy. To our best knowledge, this is the first experimental study towards multi-class learning. Moreover, to understand the noise-resilient property of QAS, we fabricate a controllable dephasing noisy channel and integrate it into our quantum processor. Assisted by this technique, we experimentally demonstrate that the ansatz designed by QAS is compatible with the topology of the employed quantum hardware and attains much better performance than hardware-efficient ansatz [19] when the system noise becomes large. Experimental results indicate that under a certain level of noise, the ansatz designed by QAS achieves the highest test accuracy (95.6%), while other heuristic ansätze only reach 90% accuracy. Additional analyses of loss landscape further explain the advantage of the QAS-based ansatz in both optimization

and effective parameter space. These gains in performance suggest the significance of developing QAS and other automatic ansatz design techniques to enhance the learning performance of VQAs.

## Result

**The mechanism of QAS.** The underlying principle of QAS is optimizing the quantum circuit architecture and the trainable parameters *simultaneously* to minimize an objective function. For elucidating, in the following, we elaborate on how to apply QAS to tailor the hardware-efficient ansatz (HEA). Mathematically, an  $N$ -qubit HEA  $U(\boldsymbol{\theta}) = \prod_{l=1}^L U_l(\boldsymbol{\theta}) \in SU(2^N)$  yields a multi-layer structure, where the circuit layout of all blocks is identical, the  $l$ -th block  $U_l(\boldsymbol{\theta})$  consists of a sequence of parameterized single-qubit and two-qubits gates, and  $L$  denotes the block number. Note that our method can be generalized to prune other ansätze such as the unitary coupled cluster ansatz [49] and the quantum approximate optimization ansatz [69].

QAS is composed of four steps to tailor HEA and output a problem-dependent and hardware-oriented ansatz, as shown in Fig. 1. The first step is specifying the ansätze pool  $\mathcal{S}$  collecting all candidate ansätze. Suppose that  $U_l(\boldsymbol{\theta})$  for  $\forall l \in [L]$  can be formed by three types of parameterized single-qubit gates, i.e., rotational gates along three axis, and one type of two-qubits gates, i.e., CNOT gates. When the layout of different blocks can be varied

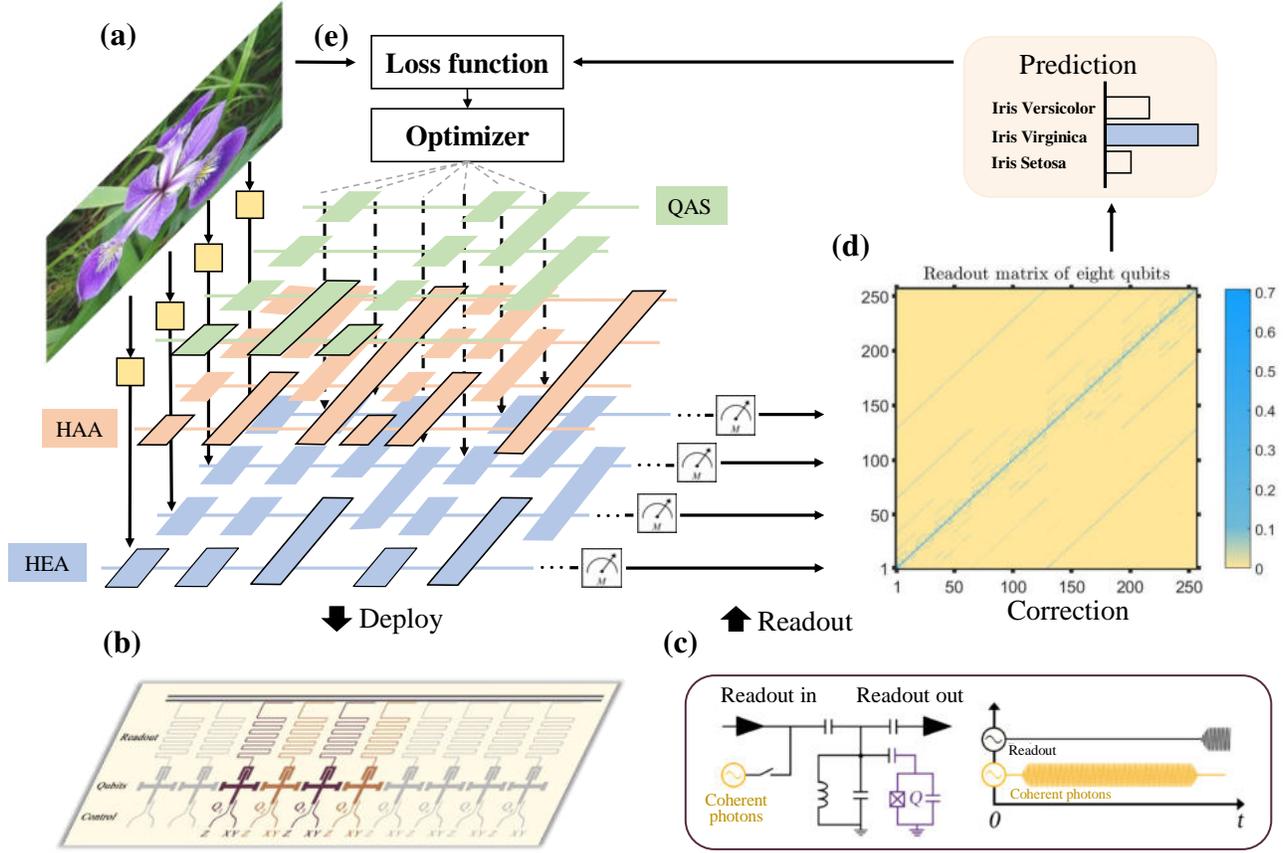


FIG. 2: **Experimental setups.** (a) The construction of quantum classifiers with the exploited three different ansätze, i.e., hardware-agnostic ansatz (HAA), hardware-efficient ansatz (HEA), and the ansatz searched by QAS, towards the Iris dataset. For all classifiers, the gate encoding method is adopted to embed the classical feature vector  $\mathbf{x}_i$  into the quantum state  $\rho_i$ . After interacting  $\rho_i$  with the ansatz  $U(\boldsymbol{\theta})$ , the generated state is measured by a fixed operator  $\Pi$  to obtain the prediction  $\text{Tr}(\Pi U(\boldsymbol{\theta})\rho_i U(\boldsymbol{\theta})^\dagger)$ . (b) All three quantum classifiers are deployed on a 8-qubit superconducting processor with the chain topology. The activated qubits are highlighted by the purple color. (c)-(d) To suppress the system noise, error mitigation techniques of measurements are used in our quantum hardware. Namely, the collected measurement results are operated with a correction matrix to estimate the ideal results. Refer to Method for details. (e) A classical optimizer continuously updating the parameters in  $U(\boldsymbol{\theta})$  to minimize the discrepancy between the predictions of quantum classifiers and ground-truth labels indicated by the objective function.

by replacing single-qubit gates or removing two-qubits gates, the ansätze pool  $\mathcal{S}$  includes in total  $O((3^N + 2^N)L)$  ansatz. Denote the input data as  $\mathcal{D}$  and an objective function as  $\mathcal{L}$ . The goal of QAS is finding the best candidate ansatz  $\mathbf{a} \in \mathcal{S}$  and its corresponding optimal parameters  $\boldsymbol{\theta}_{\mathbf{a}}^*$ , i.e.,

$$(\boldsymbol{\theta}_{\mathbf{a}}^*, \mathbf{a}^*) = \arg \min_{\boldsymbol{\theta}_{\mathbf{a}} \in \mathcal{C}, \mathbf{a} \in \mathcal{S}} \mathcal{L}(\boldsymbol{\theta}_{\mathbf{a}}, \mathbf{a}, \mathcal{D}, \mathcal{E}_{\mathbf{a}}), \quad (1)$$

where the quantum channel  $\mathcal{E}_{\mathbf{a}}$  simulates the quantum system noise induced by  $\mathbf{a}$ .

The second step is optimizing Eq. (1) with in total  $T$  iterations. As discussed in our technical companion paper [68], seeking the optimal solution  $(\boldsymbol{\theta}_{\mathbf{a}}^*, \mathbf{a}^*)$  is computationally hard, since the optimization of  $\mathbf{a}$  is discrete and the size of  $\mathcal{S}$  and  $\mathcal{C}$  exponentially scales with respect to  $N$  and  $L$ . To conquer this difficulty, QAS exploits the supernet and weight sharing strategy to ensure a

good estimation of  $(\boldsymbol{\theta}_{\mathbf{a}}^*, \mathbf{a}^*)$  within a reasonable computational cost. Concisely, weight sharing strategy correlates parameters among different ansätze in  $\mathcal{S}$  to reduce the parameter space  $\mathcal{C}$ . As for supernet, it plays two significant roles, i.e., configuring the ansätze pool  $\mathcal{S}$  and parameterizing an ansatz  $\mathbf{a} \in \mathcal{S}$  via the specified weight sharing strategy. In doing so, at each iteration  $t$ , QAS randomly samples an ansatz  $\mathbf{a}^{(t)} \in \mathcal{S}$  and updates its parameters with  $\boldsymbol{\theta}_{\mathbf{a}}^{(t+1)} = \boldsymbol{\theta}_{\mathbf{a}}^{(t)} - \eta \nabla \mathcal{L}(\boldsymbol{\theta}_{\mathbf{a}}^{(t)}, \mathbf{a}^{(t)}, \mathcal{D}, \mathcal{E}_{\mathbf{a}^{(t)}})$  and  $\eta$  being the learning rate. Due to the weight sharing strategy, the parameters of the unsampled ansätze are also updated.

The last two steps are ranking and fine tuning. Specifically, once the training is completed, QAS ranks a portion of the trained ansätze and chooses the one with the best performance. The ranking strategies are diverse, including random searching and evolutionary searching. Finally, QAS utilizes the selected ansatz to fine tune the optimized parameters with few iterations. Refer to Ref. [68] for the

omitted technical details of QAS.

**Experimental implementation.** We implement QAS on a quantum superconducting processor to accomplish the classification tasks for the Iris dataset. Namely, the Iris dataset  $\mathcal{D} = \{\mathbf{x}_i, y_i\}_{i=1}^{150}$  consists of three categories of flowers (i.e.,  $y_i \in \{0, 1, 2\}$ ) and each category includes 50 examples characterized by 4 features (i.e.,  $\mathbf{x}_i \in \mathbb{R}^4$ ). In our experiments, we split the Iris dataset into three parts, i.e., the training dataset  $\mathcal{D}_T = \{\mathbf{x}, y\}$ , the validating dataset  $\mathcal{D}_V$ , and the test dataset  $\mathcal{D}_E$  with  $\mathcal{D} = \mathcal{D}_T \cup \mathcal{D}_V \cup \mathcal{D}_E$ . The functionality of  $\mathcal{D}_T$ ,  $\mathcal{D}_V$ , and  $\mathcal{D}_E$  is estimating the optimal classifier, preventing the classifier to be over-fitted, and evaluating the generalization property of the trained classifier, respectively.

Our experiments are carried out on a quantum processor including 8 Xmon superconducting qubits with the one-dimensional chain structure. As shown in Fig. 2(b), the employed quantum device is fabricated by sputtering a Aluminium thin film onto a sapphire substrate. The single qubit rotation gate  $R_X$  ( $R_Y$ ) along X-axis (Y-axis) is implemented with microwave pulse, and the Z rotation gate  $R_Z$  is realized by virtual Z gate [70]. The construction of the CZ gate is completed by applying the avoided level crossing between the high level states  $|11\rangle$  and  $|02\rangle$  or  $|11\rangle$  and  $|20\rangle$ . The calibrated readout matrix is shown in Fig. 2(d) and the device parameters is summarized in Table II of Appendix A.

We fabricate the controllable dephasing noise as a measurable disturbance to the quantum evolution. The operators for the noise channel can be written as  $E_0 = \sqrt{1 - \alpha p}[1, 0; 0, 1]$  and  $E_1 = \sqrt{\alpha p}[1, 0; 0, -1]$ .  $\alpha$  is a constant and the value of  $p$  can be tuned in our experiment by changing the average number of the coherent photons on the readout cavity's steady state. The intensity of coherent photons is represented by the amplitude  $p$  of the curve shown on the AWGs.

The experimental implementation of the quantum classifiers is as follows. As illustrated in Fig. 2(a), the gate encoding method is exploited to load classical data into quantum states. The encoding circuit yields  $U_E(\mathbf{x}) = \otimes_{j=1}^4 R_Y(\mathbf{x}_{i,j})$ . To evaluate the effectiveness of QAS, three types of ansätze  $U(\boldsymbol{\theta})$  are used to construct the quantum classifier. The first two types are heuristic ansatz, which are hardware-agnostic ansatz (HAA) and hardware-efficient ansatz (HEA). As depicted in Fig. 2(a), HAA  $U_{\text{HAA}}(\boldsymbol{\theta})$  is designed for a general paradigm and ignores the topology of a specific quantum hardware platforms; HEA  $U_{\text{HEA}}(\boldsymbol{\theta})$  adapts the quantum hardware constraints, where all inefficient two-qubit operators that connect two physically nonadjacent qubits are forbidden. The third type of ansatz refers to the output of QAS, denoted as  $U_{\text{QAS}}(\boldsymbol{\theta})$ . The mean square error between the prediction and real labels is employed as the objective function for all quantum classifiers. The noise rate of the dephasing channel  $p$  is set as 0, 0.01 and 0.015. We benchmark the test accuracy of these three ansätze HAA, HEA and QAS, and explore whether QAS attains the highest test accuracy. Refer to Appendix B for more implementation

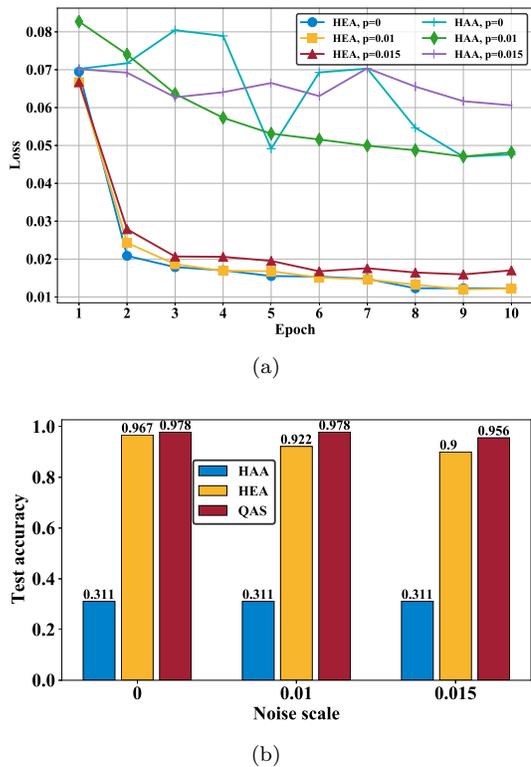


FIG. 3: **The performance of quantum classifiers.** (a) The training loss of quantum classifiers with the HAA and HEA ansätze under different noise settings. (b) The test accuracy achieved by HAA, HEA and the ansatz searched by QAS under different noise settings.

details.

**Experimental results.** To comprehend the importance of the compatibility between quantum hardware and ansatz, we first examine the learning performance of the quantum classifiers with HAA and HEA under different noise rates. The achieved experimental results are demonstrated in Fig. 3(a). In particular, in the measure of training loss (i.e., the lower the better), the quantum classifier with the HEA significantly outperforms HAA for all noise settings. At the 10-th epoch, the training loss of the quantum classifier with HAA and HEA is 0.06 and 0.017 (0.049 and 0.012; 0.049 and 0.012) when  $p = 0.015$  ( $p = 0$ ;  $p = 0.01$ ), respectively. In addition, the optimization of the quantum classifier with HAA seems to be divergent when  $p = 0.015$ . We further evaluate the test accuracy to compare their learning performance. As shown in Fig. 3(b), there exists a manifest gap between the two ansätze, highlighted by the blue and yellow colors. For all noise settings, the test accuracy corresponding to HAA is only 31.1%, whereas the test accuracy corresponding to HEA is at least 95.6%. These observations signify the significance of reconciling the topology between the employed quantum hardware and ansatz, as the key motivation of this study.

We next experiment on QAS to quantify how it problem-

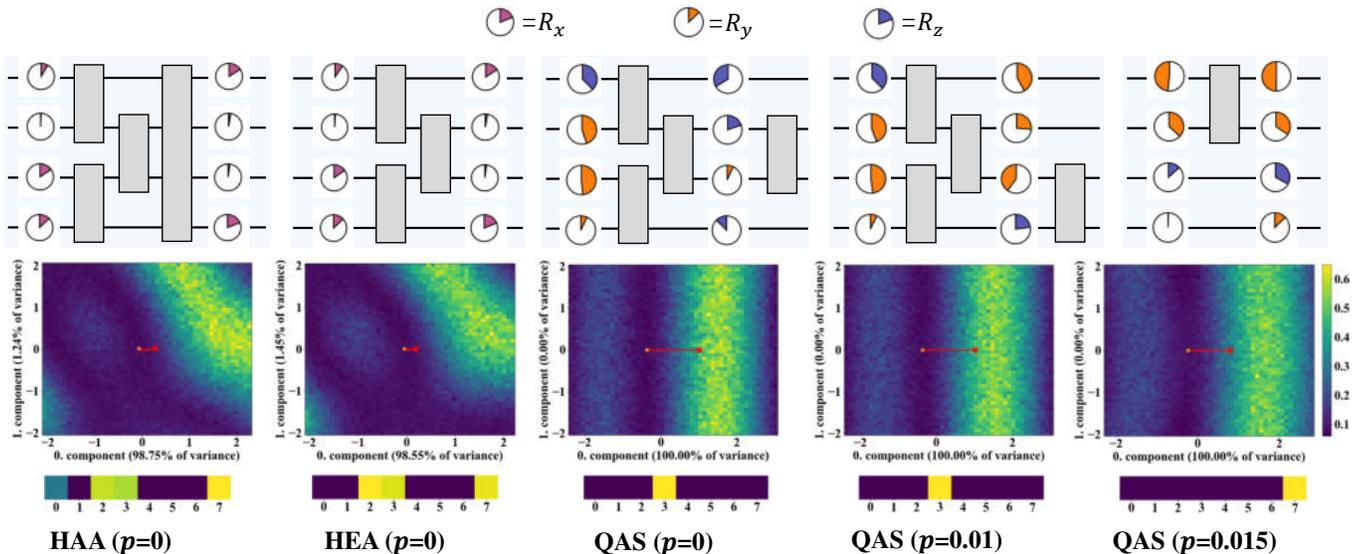


FIG. 4: **The circuit architecture and corresponding loss landscape.** The top row demonstrates the structure of HAA, HEA, and the ansatz searched by QAS under different noise settings. The color and angles refer to gate type and corresponding parameter. The middle row visualizes the loss landscape of each ansätze with respect to the trained parameters based on the technique developed in [71]. The red line tracks the optimization path of loss during the 50 epochs. Linear path indicates the loss landscape enjoys a simple structure and optimization is easy to converge. The bottom row shows the absolute value of the first row vector of PCA transform matrix, which reflects the contribution of each parameter to the first component. Number 0 – 7 denotes the parameter index.

specific and hardware-oriented designs to enhance the learning performance quantum classifiers. Concretely, as shown in Fig. 3(b), for all noise settings, the quantum classifier with the ansatz searched by QAS attains the best test accuracy than those of HAA and HEA. That is, when  $p = 0$  ( $p = 0.01$  and  $p = 0.015$ ), the test accuracy achieved by QAS is 97.8% (97.8% and 95.6%), which is higher than HEA with 96.7% (92.2% and 90.0%). Notably, although the test accuracy is slightly decreased for the increased system noise, the strength of QAS becomes evident over the other two ansätze. In other words, QAS shows the advantages to simultaneously alleviate the effect of quantum noise and search the optimal ansatz to achieve high accuracy. The superior performance validates the effectiveness of QAS towards classification tasks.

We last investigate the potential factors of ensuring the good performance of QAS from two perspectives, i.e., the circuit architecture and the corresponding loss landscape. The searched ansätze under three noise settings, HAA, and HEA are pictured in the top of Fig. 4. Compared with HEA and HAA, QAS reduces the number of CZ gates with respect to the increased level of noise. When  $p = 0.015$ , QAS chooses the ansatz containing only one CZ gate. This behavior indicates that QAS can adaptively control the number of quantum gates to balance the expressivity and learning performance. We plot the loss landscape of HAA, HEA, and the ansatz searched by QAS in the middle row of Fig. 4. To visualize the high-dimension loss landscape in a 2D plane, the dimension reduction technique, i.e., principal component analysis (PCA) [72] is applied to compress the parameter trajectory correspond-

ing to each optimization step. After dimension reduction, we choose the obtained first two principal components that explain most of variance as the landscape spanning vector. Refer to [71] and Appendix C for details. For HAA and HEA, the objective function is governed by both the 0-th component (98.75% of variance for HAA, 98.55% of variance for HEA) and 1-th component (1.24% of variance for HAA, 1.45% of variance for HEA). By contrast, for the ansätze searched by QAS, their loss landscapes totally depend on the 0-th component. Furthermore, the optimization path for QAS is exactly linear, while the optimization of HAA and HAA experiences a nonlinear curve. This difference reveals that QAS enables a more efficient optimization trajectory. As indicated by the bottom row of Fig. 4, there is a major parameter that contributes the most to the 0-th component in the three ansätze searched by QAS, while HAA and HEA have to consider multiple parameters to determine the 0-th component. This phenomenon reflects that ansätze searched by QAS are prone to have a smaller effective parameter space, which lead to less noise accumulation and further stronger noise robustness. These observations can be treated as the empirical evidence to explain the superiority of QAS.

## Discussion

Our experimental results provide the following insights. First, we experimentally verify the feasibility of applying automatically designing a problem-specific and hardware-oriented ansatz to improve the power of quantum classifiers. Second, the analysis related to the loss landscape

and the circuit architectures exhibits the potential of applying QAS and other variable ansatz construction techniques to compensate for the caveats incurred by executing variational quantum algorithms on NISQ machines.

Besides classification tasks, it is crucial to benchmark QAS and its variants towards other learning problems in quantum chemistry and quantum many-body physics. In these two areas, the employed ansatz is generally Hamiltonian dependent [48, 49, 73]. As a result, the way of constructing the ansätze pool should be carefully conceived. In addition, another important research direction is understanding the capabilities of QAS for large-scale problems. How to find the near-optimal ansatz among the exponential candidates is a challenging issue.

We note that although QAS can reconcile the imperfection of quantum systems, a central law to enhance the performance of variational quantum algorithms is promoting the quality of quantum processors. For this purpose, we will delve into carrying out QAS and its variants on more advanced quantum machines to accomplish real-world generation tasks with potential advantages.

## Methods

**Noise setup.** Due to the ac Stark effect, photon number fluctuations from the readout cavity can cause qubit dephasing [74]. We implement a pure dephasing noisy channel in our device. To every qubit, the noise photons is generated by a coherent source with a Lorentzian-shaped spectrum, which are centered at the frequency of  $\omega_c$ .  $\omega_c$  is the center frequency of the readout cavity, which is over-coupled to the feedline at the input and output port, and capacitively coupled to the Xmon qubit. The Hamiltonian of system including the readout cavity and the qubit can be written as

$$H/\hbar = \omega_c a^\dagger a + \frac{\omega_q}{2} \sigma_z + g_r (a^\dagger \sigma_- + a \sigma_+) \quad (2)$$

where  $\sigma_\pm = \sigma_x \pm i\sigma_y$ ,  $\sigma_j$  ( $j = x, y, z$ ) is the Pauli operator for the X-mon qubit.  $a^\dagger$  ( $a$ ) is the cavity photon creation (annihilation) operator.  $\omega_q$  is the frequency between the ground and the first excited states of the qubit and  $g_r$  is the coupling strength between the qubit and the readout cavity.

By continuously sending the coherent photons to drive the readout cavity to maintain a coherent state, a noisy environment can be engineered. The noise channel can be described as the depolarization in the x-y plane of the Bloch sphere. The noise intensity can be tuned by changing the average number of the coherent photons on the readout cavity's steady state. The average number

of photons is represented by the amplitude of the curve shown in the AWGs. Under different noise settings, the values of  $T_2^*$  is shown in Table I.

**Readout correction.** The experimentally measured results of the final state for the eight qubits were corrected with a calibration matrix, which can be got in an experimentally calibration process. The reconstruction process for readout results is based on Bayes' rule. The colored schematic diagram for calibration matrix is shown in figure 2(d). Assume that,  $p_{ij}$  stands for the probability of getting a measured population  $|i\rangle$  when preparing a basis state  $|j\rangle$ . The calibration matrix is

$$F = \begin{pmatrix} p_{11} & p_{12} & \dots & p_{12^n} \\ p_{21} & p_{22} & \dots & p_{22^n} \\ \dots & \dots & \dots & \dots \\ p_{2^{n-1}1} & p_{2^{n-1}2} & \dots & p_{2^{n-1}2^n} \end{pmatrix}. \quad (3)$$

If we prepare a state  $|\psi\rangle$  on  $n$  qubits, and the probability distribution of the prepared state in  $2^n$  basis is  $P = [P_1, P_2, \dots, P_{2^n}]^T$ , then we will get a measured state probability distribution as  $\tilde{P}$  in experiment, the relationship between the two probability distribution is

$$\tilde{P} = FP. \quad (4)$$

Solving for  $P$ , we have

$$P = F^{-1}\tilde{P}. \quad (5)$$

## Acknowledgments

We appreciate the helpful discussion with Weiyang Liu and Guangming Xue. This work was supported by the NSF of Beijing (Grant No. Z190012), the NSFC of China (Grants No. 11890704, No. 12004042, No. 12104055, No. 12104056), and the Key-Area Research and Development Program of Guang Dong Province (Grant No. 2018B030326001).

**Author contributions.** Y.-X. D. and H.-F. Y. conceived the research. K.-H. L.-H. and Y. Q. designed and performed the experiment. Y. Q., and Y.-X. D. performed numerical simulations. Y. Q., X.-Y. W., R.-X. W., M.-J. H., and D. L. analyzed the results. All authors contributed to discussions of the results and the development of the manuscript. Y. Q., R.-X. W., Y.-X. D. and K.-H. L.-H. wrote the manuscript with input from all co-authors. Y.-X. D., X.-Y. W., D.-C. T. and R.-X. W. supervised the whole project.

- 
- [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* **574**, 505 (2019).  
 [2] 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.*, *Physical review letters* **127**, 180501 (2021).  
 [3] 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.*,

| Parameter                    | Q1   | Q2  | Q3   | Q4  | Q5   | Q6  | Q7   | Q8  |
|------------------------------|------|-----|------|-----|------|-----|------|-----|
| $T_2^*(p=0)$ ( $\mu s$ )     | 11.7 | 2.0 | 15.2 | 1.9 | 15.9 | 1.8 | 12.6 | 1.6 |
| $T_2^*(p=0.005)$ ( $\mu s$ ) | 10.6 | 1.9 | 15.3 | 1.8 | 15.0 | 1.6 | 10.7 | 1.5 |
| $T_2^*(p=0.01)$ ( $\mu s$ )  | 7.8  | 1.6 | 12.5 | 1.7 | 14.7 | 1.6 | 13.0 | 1.6 |
| $T_2^*(p=0.015)$ ( $\mu s$ ) | 0.2  | 1.2 | 12.7 | 1.6 | 14.4 | 1.6 | 12.1 | 1.5 |

TABLE I: The transverse relaxation time  $T_2^*$  under different noise settings.

- Science **370**, 1460 (2020).
- [4] Q. Zhu, S. Cao, F. Chen, M.-C. Chen, X. Chen, T.-H. Chung, H. Deng, Y. Du, D. Fan, M. Gong, *et al.*, Science Bulletin (2021).
- [5] 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.*, arXiv preprint arXiv:2101.08448 (2021).
- [6] M. Cerezo, A. Arrasmith, R. Babbush, S. C. Benjamin, S. Endo, K. Fujii, J. R. McClean, K. Mitarai, X. Yuan, L. Cincio, *et al.*, Nature Reviews Physics **3**, 625 (2021).
- [7] J. Preskill, Quantum **2**, 79 (2018).
- [8] A. Abbas, D. Sutter, C. Zoufal, A. Lucchi, A. Figalli, and S. Woerner, Nature Computational Science **1**, 403 (2021).
- [9] L. Bianchi, J. Pereira, and S. Pirandola, arXiv preprint arXiv:2102.08991 (2021).
- [10] K. Bu, D. E. Koh, L. Li, Q. Luo, and Y. Zhang, arXiv preprint arXiv:2101.06154 (2021).
- [11] M. C. Caro and I. Datta, Quantum Machine Intelligence **2**, 1 (2020).
- [12] M. C. Caro, H.-Y. Huang, M. Cerezo, K. Sharma, A. Sornborger, L. Cincio, and P. J. Coles, arXiv preprint arXiv:2111.05292 (2021).
- [13] Y. Du, M.-H. Hsieh, T. Liu, S. You, and D. Tao, PRX Quantum **2**, 040337 (2021).
- [14] Y. Du, Z. Tu, X. Yuan, and D. Tao, arXiv preprint arXiv:2104.09961 (2021).
- [15] H.-Y. Huang, R. Kueng, and J. Preskill, Physical Review Letters **126**, 190505 (2021).
- [16] H.-Y. Huang, M. Broughton, M. Mohseni, R. Babbush, S. Boixo, H. Neven, and J. R. McClean, Nature communications **12**, 1 (2021).
- [17] H.-Y. Huang, R. Kueng, G. Torlai, V. V. Albert, and J. Preskill, arXiv preprint arXiv:2106.12627 (2021).
- [18] S. Endo, J. Sun, Y. Li, S. C. Benjamin, and X. Yuan, Physical Review Letters **125**, 010501 (2020).
- [19] A. Kandala, A. Mezzacapo, K. Temme, M. Takita, M. Brink, J. M. Chow, and J. M. Gambetta, Nature **549**, 242 (2017).
- [20] G. Pagano, A. Bapat, P. Becker, K. S. Collins, A. De, P. W. Hess, H. B. Kaplan, A. Kyprianidis, W. L. Tan, C. Baldwin, *et al.*, Proceedings of the National Academy of Sciences **117**, 25396 (2020).
- [21] M. Cerezo, A. Poremba, L. Cincio, and P. J. Coles, Quantum **4**, 248 (2020).
- [22] Y. Du and D. Tao, arXiv preprint arXiv:2106.15432 (2021).
- [23] J. Carolan, M. Mohseni, J. P. Olson, M. Prabhu, C. Chen, D. Bunandar, M. Y. Niu, N. C. Harris, F. N. Wong, M. Hochberg, *et al.*, Nature Physics **16**, 322 (2020).
- [24] Z. Holmes, K. Sharma, M. Cerezo, and P. J. Coles, arXiv preprint arXiv:2101.02138 (2021).
- [25] J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Babbush, and H. Neven, Nature communications **9**, 1 (2018).
- [26] M. Cerezo, A. Sone, T. Volkoff, L. Cincio, and P. J. Coles, Nature communications **12**, 1 (2021).
- [27] A. Pesah, M. Cerezo, S. Wang, T. Volkoff, A. T. Sornborger, and P. J. Coles, Phys. Rev. X **11**, 041011 (2021).
- [28] E. Grant, L. Wossnig, M. Ostaszewski, and M. Benedetti, Quantum **3**, 214 (2019).
- [29] S. Bravyi, A. Kliesch, R. Koenig, and E. Tang, Physical Review Letters **125**, 260505 (2020).
- [30] V. Havlíček, A. D. Córcoles, K. Temme, A. W. Harrow, A. Kandala, J. M. Chow, and J. M. Gambetta, Nature **567**, 209 (2019).
- [31] H.-L. Huang, Y. Du, M. Gong, Y. Zhao, Y. Wu, C. Wang, S. Li, F. Liang, J. Lin, Y. Xu, *et al.*, Physical Review Applied **16**, 024051 (2021).
- [32] E. Peters, J. Caldeira, A. Ho, S. Leichenauer, M. Mohseni, H. Neven, P. Spentzouris, D. Strain, and G. N. Perdue, arXiv preprint arXiv:2101.09581 (2021).
- [33] M. S. Rudolph, N. B. Toussaint, A. Katabarwa, S. Johri, B. Peropadre, and A. Perdomo-Ortiz, arXiv preprint arXiv:2012.03924 (2020).
- [34] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, S. Boixo, M. Broughton, B. B. Buckley, D. A. Buell, *et al.*, Science **369**, 1084 (2020).
- [35] A. Robert, P. K. Barkoutsos, S. Woerner, and I. Tavernelli, npj Quantum Information **7**, 1 (2021).
- [36] S. Kais, Quantum Information and Computation for Chemistry, 1 (2014).
- [37] D. Wecker, M. B. Hastings, N. Wiebe, B. K. Clark, C. Nayak, and M. Troyer, Physical Review A **92**, 062318 (2015).
- [38] X. Cai, W.-H. Fang, H. Fan, and Z. Li, Physical Review Research **2**, 033324 (2020).
- [39] M. P. Harrigan, K. J. Sung, M. Neeley, K. J. Satzinger, F. Arute, K. Arya, J. Atalaya, J. C. Bardin, R. Barends, S. Boixo, *et al.*, Nature Physics **17**, 332 (2021).
- [40] N. Lacroix, C. Hellings, C. K. Andersen, A. Di Paolo, A. Remm, S. Lazar, S. Krinner, G. J. Norris, M. Gabureac, J. Heinsoo, *et al.*, PRX Quantum **1**, 110304 (2020).
- [41] L. Zhou, S.-T. Wang, S. Choi, H. Pichler, and M. D. Lukin, Physical Review X **10**, 021067 (2020).
- [42] S. Hadfield, Z. Wang, B. O’Gorman, E. G. Rieffel, D. Venturelli, and R. Biswas, Algorithms **12**, 34 (2019).
- [43] D. H. Wolpert and W. G. Macready, IEEE transactions on evolutionary computation **1**, 67 (1997).
- [44] K. Poland, K. Beer, and T. J. Osborne, arXiv preprint arXiv:2003.14103 (2020).
- [45] B. T. Gard, L. Zhu, G. S. Barron, N. J. Mayhall, S. E. Economou, and E. Barnes, npj Quantum Information **6**, 1 (2020).
- [46] M. Ganzhorn, D. J. Egger, P. Barkoutsos, P. Ollitrault, G. Salis, N. Moll, M. Roth, A. Fuhrer, P. Mueller, S. Woerner, *et al.*, Physical Review Applied **11**, 044092 (2019).

- [47] A. Choquette, A. Di Paolo, P. K. Barkoutsos, D. Sénéchal, I. Tavernelli, and A. Blais, *Physical Review Research* **3**, 023092 (2021).
- [48] Y. Cao, J. Romero, J. P. Olson, M. Degroote, P. D. Johnson, M. Kieferová, I. D. Kivlichan, T. Menke, B. Peropadre, N. P. Sawaya, *et al.*, *Chemical reviews* **119**, 10856 (2019).
- [49] J. Romero, R. Babbush, J. R. McClean, C. Hempel, P. J. Love, and A. Aspuru-Guzik, *Quantum Science and Technology* **4**, 014008 (2018).
- [50] A. Cervera-Lierta, J. S. Kottmann, and A. Aspuru-Guzik, *PRX Quantum* **2**, 020329 (2021).
- [51] R. M. Parrish, E. G. Hohenstein, P. L. McMahon, and T. J. Martínez, *Physical review letters* **122**, 230401 (2019).
- [52] L. Petit, H. Eenink, M. Russ, W. Lawrie, N. Hendrickx, S. Philips, J. Clarke, L. Vandersypen, and M. Veldhorst, *Nature* **580**, 355 (2020).
- [53] D. P. DiVincenzo, *Fortschritte der Physik: Progress of Physics* **48**, 771 (2000).
- [54] M. H. Devoret and R. J. Schoelkopf, *Science* **339**, 1169 (2013).
- [55] L. Cincio, K. Rudinger, M. Sarovar, and P. J. Coles, *PRX Quantum* **2**, 010324 (2021).
- [56] D. Chivilikhin, A. Samarin, V. Ulyantsev, I. Iorsh, A. Oganov, and O. Kyriienko, *arXiv preprint arXiv:2007.04424* (2020).
- [57] A. G. Rattew, S. Hu, M. Pistoia, R. Chen, and S. Wood, *arXiv preprint arXiv:1910.09694* (2019).
- [58] C. Chen, Z. He, L. Li, S. Zheng, and H. Situ, *arXiv preprint arXiv:2106.06248* (2021).
- [59] F.-X. Meng, Z.-T. Li, X.-T. Yu, and Z.-C. Zhang, *IEEE Transactions on Quantum Engineering* **2**, 1 (2021).
- [60] E.-J. Kuo, Y.-L. L. Fang, and S. Y.-C. Chen, *arXiv preprint arXiv:2104.07715* (2021).
- [61] S.-X. Zhang, C.-Y. Hsieh, S. Zhang, and H. Yao, *arXiv preprint arXiv:2010.08561* (2020).
- [62] S.-X. Zhang, C.-Y. Hsieh, S. Zhang, and H. Yao, *arXiv preprint arXiv:2103.06524* (2021).
- [63] M. Ostaszewski, L. M. Trenkwalder, W. Masarczyk, E. Scerri, and V. Dunjko, *arXiv preprint arXiv:2103.16089* (2021).
- [64] M. Pirhooshyaran and T. Terlaky, *Quantum Machine Intelligence* **3**, 1 (2021).
- [65] M. Bilkis, M. Cerezo, G. Verdon, P. J. Coles, and L. Cincio, *arXiv preprint arXiv:2103.06712* (2021).
- [66] H. R. Grimsley, S. E. Economou, E. Barnes, and N. J. Mayhall, *Nature communications* **10**, 1 (2019).
- [67] H. L. Tang, V. Shkolnikov, G. S. Barron, H. R. Grimsley, N. J. Mayhall, E. Barnes, and S. E. Economou, *PRX Quantum* **2**, 020310 (2021).
- [68] Y. Du, T. Huang, S. You, M.-H. Hsieh, and D. Tao, *arXiv preprint arXiv:2010.10217* (2020).
- [69] E. Farhi, J. Goldstone, and S. Gutmann, *arXiv preprint arXiv:1411.4028* (2014).
- [70] D. C. McKay, C. J. Wood, S. Sheldon, J. M. Chow, and J. M. Gambetta, *Physical Review A* **96**, 022330 (2017).
- [71] M. S. Rudolph, S. Sim, A. Raza, M. Stechly, J. R. McClean, E. R. Anschuetz, L. Serrano, and A. Perdomo-Ortiz, *arXiv preprint arXiv:2111.04695* (2021).
- [72] K. Pearson, *The London, Edinburgh, and Dublin philosophical magazine and journal of science* **2**, 559 (1901).
- [73] A. Peruzzo, J. McClean, P. Shadbolt, M.-H. Yung, X.-Q. Zhou, P. J. Love, A. Aspuru-Guzik, and J. L. O'Brien, *Nature communications* **5**, 4213 (2014).
- [74] F. Yan, D. Campbell, P. Krantz, M. Kjaergaard, D. Kim, J. L. Yoder, D. Hover, A. Sears, A. J. Kerman, T. P. Orlando, *et al.*, *Physical Review Letters* **120**, 260504 (2018).
- [75] R. A. Fisher, *Annals of eugenics* **7**, 179 (1936).
- [76] J. Kiefer and J. Wolfowitz, *The Annals of Mathematical Statistics*, 462 (1952).
- [77] K. Mitarai, M. Negoro, M. Kitagawa, and K. Fujii, *Physical Review A* **98**, 032309 (2018).

| Parameter             | Q1     | Q2     | Q3     | Q4     | Q5     | Q6     | Q7     | Q8     |
|-----------------------|--------|--------|--------|--------|--------|--------|--------|--------|
| $\omega_i/2\pi$ (GHz) | 5.202  | 4.573  | 5.146  | 4.527  | 5.099  | 4.47   | 5.118  | 4.543  |
| $\alpha_i/2\pi$ (GHz) | -0.240 | -0.240 | -0.239 | -0.240 | -0.239 | -0.239 | -0.239 | -0.242 |
| $T_1$ ( $\mu s$ )     | 10.3   | 13.6   | 14.7   | 16.7   | 14.9   | 13.2   | 12.7   | 8.0    |
| $T_2^*$ ( $\mu s$ )   | 11.7   | 2.0    | 15.2   | 1.9    | 15.9   | 1.8    | 12.6   | 1.6    |
| $\bar{F}_s$           | 99.67% | 98.96% | 99.76% | 99.55% | 99.05% | 99.69% | 98.51% | 99.09% |
| $T_s$ (ns)            | 37     | 37     | 37     | 35     | 35     | 37     | 37     | 35     |
| $g_i/2\pi$ (MHz)      | 19.7   | 19.6   | 19.1   | 19.1   | 19.2   | 19.3   | 19.6   |        |
| $J_{z,i}/2\pi$ (MHz)  | 0.675  | 0.8    | 0.7    | 0.78   | 0.626  | 0.655  | 0.819  |        |
| $\bar{F}_{cz}$        | 97.23% | 96.11% | 97.48% | 93.54% | 93.72% | 94.95% | 95.54% |        |
| $T_{cz}$ (ns)         | 18.5   | 21.5   | 23     | 19.5   | 23.5   | 21.5   | 18.5   |        |

TABLE II: Device parameters.  $\omega_i$  and  $\alpha_i$  represent the qubit frequency and qubit anharmonicity respectively.  $T_1$  and  $T_2^*$  are the longitudinal and transverse relaxation time respectively.  $\bar{F}_s$  and  $T_s$  are the average fidelity and length of the single qubit gates.  $g_i$  is the coupling strength between nearby qubits, and  $J_{z,i}$  is the effective ZZ coupling strength.  $\bar{F}_{cz}$  are the fidelity of the CZ gates calibrated by quantum process tomography and  $T_{cz}$  are the length of the CZ gates.

## A. Experiment setup

### 1. Device parameters

The qubit parameters and length and fidelity for the single- and two-qubit gates of our device are summarized in Table II.

### 2. Electronics and control wiring

The device is installed at a cryogenic setup in a dilution refrigerator. The control and measurement electronics which is connected to the device is shown in figure M1. The electronic control module is divided by 6 area with different temperature. The superconducting quantum device is installed at the base plate with a cryogenic environment of 10 mK. For each qubit, the frequency is tuned by a flux control line with changing the magnetic flux through the SQUID loop, and the flux is controlled by a constant current which is generated by a voltage source and inductively coupled to the SQUID. Four attenuators are connected in the circuit in series to act as the thermal precipitator. The XY control for each qubit is achieved by up-converting the intermediate frequency signals with an analog IQ-mixer modules. The drive pulse is provided by the multichannel AWGs with a sample rate of 2 GSa/s.

The qubit readout is performed by a readout control system with sampling rate of 1 GSa/s. The readout pulse is upconverted to the frequency band of the readout cavity with the analog IQ-mixer and transmitted through the readout line with attenuators and low pass filters to the chip. At the output side, the constant and alternating currents are combined by a bias-tee, amplified by the parametric amplifier, and connected to the readout line through a circulator at 10 mK, as well as a high-electron mobility transistor (HEMT) at 4 K and two more amplifier at room temperature. The amplified signals finally are digitized by the Analog to Digital Converter.

## B. Implementation of quantum classifiers

In this section, we implement HAA, HEA and QAS for classification of Iris dataset on the 8-qubit superconducting quantum processors with controllable dephasing noise. A detailed description about the dataset and hyper-parameters configuration is given below.

### 1. Dataset

The classification data employed in this paper is the Iris dataset [75], which contains 150 instances characterized by 4 attributes and 3 categories. Each dimension of the feature vector is normalized to range  $[0, 1]$ . During training, the whole dataset is split into three parts, including training set (60 samples), validation set (45 samples) and test set (45 samples). The sample distribution is visualized by selecting the first two dimensions of feature vector. As shown in Fig. M2, samples of class 1 and 2 cannot be distinguished by a linear classifier. It means that nonlinearity should be introduced into the quantum classifier to achieve higher classification accuracy.

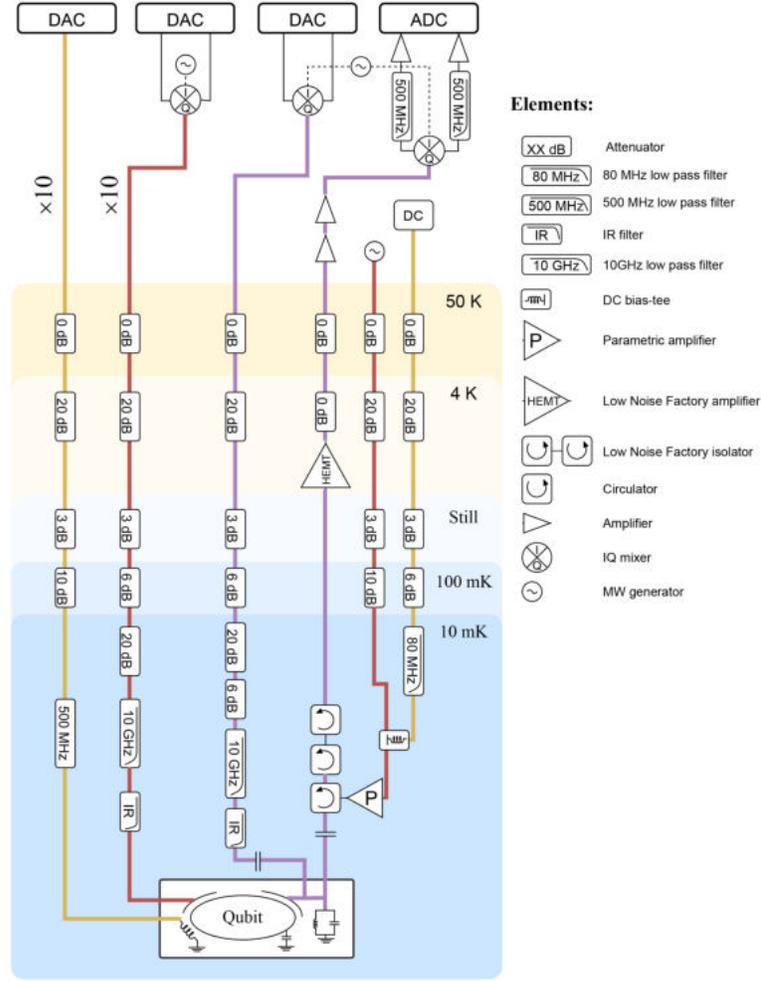


FIG. M1: Schematic diagram for the electronics and wiring setup for the superconducting quantum system.

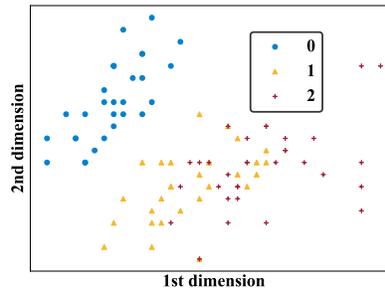


FIG. M2: The visualization of Iris dataset based on its first two dimensions.

## 2. Objective function and accuracy measure

**Objective function.** We adopt the mean square error (MSE) as the objective function for all quantum classifiers, i.e.,

$$\mathcal{L}(\mathcal{D}, \theta) = \frac{1}{2n} \sum_{i=1}^n (\langle O \rangle_i - y_i)^2, \quad (\text{B1})$$

where  $\langle O \rangle_i = \langle 0|U_E(\mathbf{x}_i)^\dagger U(\boldsymbol{\theta})^\dagger O U(\boldsymbol{\theta}) U_E(\mathbf{x}_i)|0\rangle$ ,  $O$  refers to the observable,  $U_E(\mathbf{x}_i)$  denotes the unitary operator that embeds classical feature vector  $\mathbf{x}_i$  into quantum circuit, and  $U(\boldsymbol{\theta})$  is the variational quantum circuit with the trainable parameters  $\boldsymbol{\theta}$ .

**Definition of train, valid, and test accuracy.** Given an example  $\mathbf{x}_i$ , the quantum classifier predicts its label as

$$\tilde{y}_i = \begin{cases} 0, & \langle O \rangle_i \leq \frac{1}{6} \\ 1, & \frac{1}{6} < \langle O \rangle_i \leq \frac{1}{2} \\ 2, & \frac{1}{2} < \langle O \rangle_i \leq \frac{5}{6} \end{cases}. \quad (\text{B2})$$

The train (valid and test) accuracy aims to measure the differences between the predicted labels and true labels for examples in the train dataset  $\mathcal{D}_T$  (valid dataset  $\mathcal{D}_V$  and test dataset  $\mathcal{D}_E$ ), i.e.3,

$$\text{accuracy} = \frac{\sum_{(\mathbf{x}_i, y_i) \in \mathcal{D}} \mathbb{1}_{\tilde{y}_i = y_i}}{|\mathcal{D}|}, \mathcal{D} = \mathcal{D}_T \text{ or } \mathcal{D}_V \text{ or } \mathcal{D}_E, \quad (\text{B3})$$

where  $|\cdot|$  denotes the size of a set.

### 3. Training hyper-parameters

The trainable parameters for all ansätze are randomly initialized following the uniform distribution  $\mathcal{U}_{[-\pi, \pi]}$ . During training, the hyper-parameters are set as follows: the optimizer is stochastic gradient descent (SGD) [76], the batch size is 4 and the learning rate is fixed at 0.2. Specifically, parameter shift rule [77] is applied to compute the gradient of objective function with respect to single parameter.

For QAS, we train 5 candidate supernet for 40 epochs to fit the training set. During the search phase, we randomly sample 100 ansatz and rank them according to their accuracy on the validation set. Finally, the ansatz with the highest accuracy is selected as the target ansatz. The ansätze pool is constructed as follows. For the single-qubit gate, the candidate set is  $\{RY, RZ\}$ . For the two-qubit gate, QAS automatically determines whether applying  $CZ$  gates to the qubit pair  $(0, 1)$ ,  $(1, 2)$ ,  $(2, 3)$  or not, discarding all other combinations, such as  $(0, 2)$  and  $(0, 3)$ . These non-adjacent qubits connections require more gates when running on the superconducting processor of 1-D chain topology, leading to bigger noise accumulation.

#### C. More details of experimental results

##### 1. PCA used in visualization of loss landscape

To visualize the loss landscape of HAA, HEA and ansätze searched by QAS with respect to the parameter space, we apply principle component analysis (PCA) to the parameter trajectory collected in every optimization step and choose the first two components as the observation variable. To be concrete, given a sequence of trainable parameter vector along the optimization trajectory  $\{\boldsymbol{\theta}^{(1)}, \dots, \boldsymbol{\theta}^{(t)}, \dots, \boldsymbol{\theta}^{(T)}\}$  where  $T$  is the number of total optimization steps and  $\boldsymbol{\theta}^{(t)} \in \mathbb{R}^d$  denotes the parameter vector at the  $t$ -th step, we construct the matrix  $\Theta = [\boldsymbol{\theta}^{(1)}; \dots; \boldsymbol{\theta}^{(T)}] \in \mathbb{R}^{T \times d}$ . Once we apply PCA to  $\Theta$  and obtain the first two principal components  $E = [\mathbf{e}_0, \mathbf{e}_1]^T \in \mathbb{R}^{2 \times d}$ , the loss landscape with respect to trainable parameters can be visualized by performing a 2D scan for  $E\Theta^T$ . Simultaneously, the projection vector  $\mathbf{e}_i$  of each component indicates the contribution of each parameter to this component, implying how many parameters determine the value of objective function. Refer to [71] for details.

The optimization trajectory can provide certain information of the trainability and convergence of the employed ansatz in quantum classifiers. When the optimization path is exactly linear, it implies that the loss landscape is not intricate and the model can be easily optimized. On the contrary, the complicated nonlinear optimization curve indicates the difficulty of converging to the local minima.

##### 2. More experimental results

We conduct numerical experiments on classical computers to validate the effectiveness of QAS.

**Dephasing noise.** We simulate the dephasing noise channel as

$$\rho' = (1 - \bar{p})\rho + \bar{p}\sigma_z\rho\sigma_z, \quad (\text{C1})$$

where  $\rho$  and  $\rho'$  represent the ideal quantum state (density matrix) and noisy quantum state affected by dephasing channel,  $\sigma_z$  is the Pauli-Z operator, and  $\bar{p} = \alpha p$  is the noise strength, representing the probability that applying a Pauli-Z operator to the quantum state. In the experiments, the noise strength  $\bar{p}$  is set as  $\{0.05, 0.1, 0.15\}$ , and the circuit layer  $L$  is set as  $\{2, 4, 6\}$ . Each setting runs for 10 times to suppress the effects of randomness.

**Simulation results.** As shown in Fig. M3, QAS achieves the highest test accuracy over all noise and layer settings. When  $L = 2$  and  $p = 0.05$ , the performance gap between HEA (95.7%) and QAS (97.1%) is relatively small. With both the depth and noise strength increasing, HEA witnesses a rapid accuracy drop (60% for  $L = 6$  and  $p = 0.15$ ). By contrast, the test accuracy for QAS with  $L = 6$  and  $p = 0.15$  is 95%, which slightly decreases 2%. This behaviour accords with the results on the superconducting processor (the test accuracy of QAS running a superconducting device decreases from 97.8% to 95.6% when  $p$  increases from 0 to 0.015, refer to Fig. 3 for more details), further illustrating the advantage of QAS in error mitigation and model expressivity.

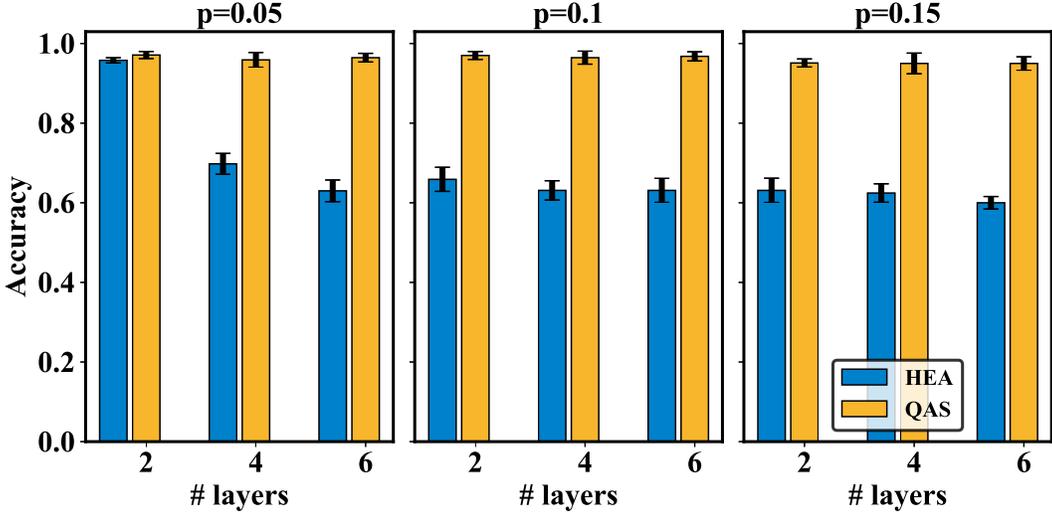


FIG. M3: The test accuracy achieved by HEA and QAS under various number of layers and noise strength when simulating on classical devices.