

Variational data encoding and correlations in quantum-enhanced machine learning

Ming-Hao Wang(王明浩) and Hua Lü(吕桦)

Citation: Chin. Phys. B, 2024, 33 (**9**): 090307. DOI: 10.1088/1674-1056/ad5c3b
Journal homepage: http://cpb.iphy.ac.cn; http://iopscience.iop.org/cpb

What follows is a list of articles you may be interested in

Analysis of learnability of a novel hybrid quantum—classical convolutional neural network in image classification

Tao Cheng(程涛), Run-Sheng Zhao(赵润盛), Shuang Wang(王爽), Rui Wang(王睿), and Hong-Yang Ma(马鸿洋)
Chin. Phys. B, 2024, 33 (**4**): 040303. DOI: 10.1088/1674-1056/ad1926

Variational quantum semi-supervised classifier based on label propagation

Yan-Yan Hou(侯艳艳), Jian Li(李剑), Xiu-Bo Chen(陈秀波), and Chong-Qiang Ye(叶崇强)
Chin. Phys. B, 2023, 32 (**7**): 070309. DOI: 10.1088/1674-1056/acb9fb

A new method of constructing adversarial examples for quantum variational circuits

Jinge Yan(颜金歌), Lili Yan(闫丽丽), and Shibin Zhang(张仕斌)
Chin. Phys. B, 2023, 32 (**7**): 070304. DOI: 10.1088/1674-1056/ac9b32

Quantum algorithm for neighborhood preserving embedding

Shi-Jie Pan(潘世杰), Lin-Chun Wan(万林春), Hai-Ling Liu(刘海玲), Yu-Sen Wu(吴宇森), Su-Juan Qin(秦素娟), Qiao-Yan Wen(温巧燕), and Fei Gao(高飞)
Chin. Phys. B, 2022, 31 (**6**): 060304. DOI: 10.1088/1674-1056/ac523a

Quantum partial least squares regression algorithm for multiple correlation problem

Yan-Yan Hou(侯艳艳), Jian Li(李剑), Xiu-Bo Chen(陈秀波), and Yuan Tian(田源)
Chin. Phys. B, 2022, 31 (**3**): 030304. DOI: 10.1088/1674-1056/ac1b84

Variational data encoding and correlations in quantum-enhanced machine learning

Ming-Hao Wang(王明浩)[1,†] and Hua Lü(吕桦)[2,‡]

[1] *School of Physics, Hubei University, Wuhan* 430062, *China*
[2] *School of Science, Hubei University of Technology, Wuhan* 430068, *China*

(Received 16 March 2024; revised manuscript received 23 June 2024; accepted manuscript online 27 June 2024)

Leveraging the extraordinary phenomena of quantum superposition and quantum correlation, quantum computing offers unprecedented potential for addressing challenges beyond the reach of classical computers. This paper tackles two pivotal challenges in the realm of quantum computing: firstly, the development of an effective encoding protocol for translating classical data into quantum states, a critical step for any quantum computation. Different encoding strategies can significantly influence quantum computer performance. Secondly, we address the need to counteract the inevitable noise that can hinder quantum acceleration. Our primary contribution is the introduction of a novel variational data encoding method, grounded in quantum regression algorithm models. By adapting the learning concept from machine learning, we render data encoding a learnable process. This allowed us to study the role of quantum correlation in data encoding. Through numerical simulations of various regression tasks, we demonstrate the efficacy of our variational data encoding, particularly post-learning from instructional data. Moreover, we delve into the role of quantum correlation in enhancing task performance, especially in noisy environments. Our findings underscore the critical role of quantum correlation in not only bolstering performance but also in mitigating noise interference, thus advancing the frontier of quantum computing.

Keywords: quantum machine learning, variational data encoding, quantum correlation

PACS: 03.67.–a, 03.67.Ac, 03.65.Ud **DOI:** 10.1088/1674-1056/ad5c3b

1. Introduction

Quantum information processing stands at the forefront of next-generation information technology, offering the potential for exponential speedups over classical counterparts.[1] Landmark algorithms like Shor's for large number factoring and the Harrow–Hassidim–Lloyd algorithm for linear systems exemplify this potential within the standard gate-based quantum computation model.[2,3] However, the current landscape is dominated by noisy-intermediate-scale quantum (NISQ) devices, characterized by their hundreds of noisy qubits and the consequent limitations in achieving large-scale, fault-tolerant quantum computing.[4] This reality steers contemporary research towards designing algorithms suitable for NISQ devices that still exploit quantum advantages.[5,6]

In this context, hybrid quantum-classical algorithms (HQCAs) have emerged as a promising approach, demonstrating success in various applications ranging from calculating eigenstates of physical Hamiltonians to optimization and classification tasks.[7–13] HQCAs, akin to machine learning algorithms, involve training computers to recognize patterns and minimize cost functions.[14] They utilize parameterized quantum circuits (PQCs) or quantum neural networks, widely adopted in quantum machine learning (QML) for tasks like classification and generative modeling.[15–22] The potential synergy between machine learning and quantum computing is a burgeoning area of interest.[23–29]

Central to HQCAs is the encoding of classical information into quantum states, a process known as quantum feature mapping (QFM).[30,31] This encoding leverages the expressiveness of large Hilbert spaces to enhance data processing for quantum advantage.[24] However, the choice of encoding schemes is critical, as it can significantly impact the performance of quantum algorithms. Many encoding protocols have been proposed such as basis encoding,[2] amplitude encoding,[32] tensor encoding.[33] These encoding schemes are fixed and cannot be changed during training. Training is embodied in the processing of encoded data, such as the training of parameters in neural networks. Thus changing the encoding schemes means training from scratch. Several variants have also been proposed such as the alternating execution of encoding and processing, which greatly enhances the expression ability of encoding and is also expected to reduce the complexity of subsequent computing.[34–36]

Along this line of thinking, this paper delves into the concept of variational data encoding (VDE). With VDE, we investigate quantum regression algorithms. We explore the trainability of VDE using PQCs, which are adjustable in quantum-classical optimization loops. Utilizing the trainability of VDE, our investigation extends to the role of quantum correlation in training, suggesting its potential to enhance capacity and resist local noise. This provides a new perspective for studying the

†Corresponding author. E-mail: wangmh@hubu.edu.cn
‡Corresponding author. E-mail: lvhuahg@163.com

© 2024 Chinese Physical Society and IOP Publishing Ltd

role of quantum correlation in quantum computing.

2. Variational data encoding

This section provides a comprehensive overview of QFMs, a cornerstone in quantum computation. QFMs are instrumental in encoding an input vector x into a corresponding quantum state $|\psi(x)\rangle$. From a mathematical perspective, this encoding is represented as a mapping

$$\psi : x \mapsto |\psi(x)\rangle, \tag{1}$$

where x is an N-dimensional vector belonging to the real vector space $\mathbb{R}^N$, and $|\psi(x)\rangle$ denotes a quantum state within a d-dimensional Hilbert space $\mathcal{H}$.

Physically, QFMs are implemented through a data-dependent quantum circuit. This circuit, characterized by a unitary operation $E(x)$, transforms a standard initial state $|0\rangle$ into the desired quantum state:

$$|\psi(x)\rangle = E(x)|0\rangle. \tag{2}$$

Typically, QFMs exhibit nonlinearity and the dimensionality of the quantum states' Hilbert space often surpasses that of the input space. This characteristic allows for a more expressive representation of data in quantum computing.

QFMs are closely related to kernel methods in machine learning.[36–38] The trick for both is to map the input vectors into a higher-dimensional feature space in which the feature vectors are easier to analyze. The common quantum encoding schemes, such as basis encoding and amplitude encoding, fall into this category. We refer interested readers to the reference and its supplemental material for details.[31]

To make QFMs learnable, we further parameterize QFMs by introducing a parameter $\xi = (\xi_1, \xi_2, \ldots)$, denoting them as $E_\xi(x)$. We emphasize that ξ is fixed for a specific QFM. Changing it results in a new QFM. By contrast, x is the input vector to be encoded and it varies with different inputs. We call this quadratic parameterized quantum circuit variational data encoding. To find the right parameter ξ, a classical optimizer is needed. The optimization process consists of two parts. First, the quantum hardware is run and outputs measurements. Second, the classical optimizer calculates the loss function and its gradient based on these measurements, and updates the parameter ξ. After optimization, we hope to obtain a high-performance QFM such that the encoded states $|\psi(x)\rangle = E_\xi(x)|0\rangle$ is suitable for subsequent quantum computation to reveal quantum advantage. The optimization will be further detailed in Subsection 3.3.

3. Variational quantum regression algorithms

In the context of VDE, we introduce variational quantum regression algorithms (VQRAs) and provide a detailed exploration of their theoretical framework.

3.1. Theory

Consider a dataset, often referred to as training data in machine learning, consisting of M data points, denoted as $\{x^{(m)}, y^{(m)}\}_{m=1}^M$. This dataset is drawn from a specific input set $\mathcal{X}$, where each $x^{(m)} = \{x_1^m, x_2^m, \ldots, x_N^m\}^{\mathrm{T}}$ represents an N-feature vector in the real vector space $\mathbb{R}^N$, and $y^{(m)}$ corresponds to a target value in $\mathbb{R}$. The goal of a VQRA is to learn a functional mapping from x^m to y^m with as much accuracy as possible. This learned function is then utilized to predict the target value y for a new, previously unseen input point x.

Assuming each $x^{(m)}$ has been transformed into a quantum state $|\psi(x^{(m)})\rangle$ via a VDE circuit $E_\xi(x^{(m)})$, we have $|\psi(x^{(m)})\rangle = E_\xi(x^{(m)})|0\rangle$. This transformation facilitates the definition of a complex kernel as follows:

$$\kappa(x^{(m)}, x^{(n)}) = \langle\psi(x^{(m)})|\psi(x^{(n)})\rangle, \tag{3}$$

which subsequently establishes a reproducing kernel Hilbert space R_κ.[36–39] A function $f(\cdot)$ within R_κ can be expressed through the inner products of the encoded state $|\psi(x)\rangle$ and another state $|\Psi\rangle$ in $\mathcal{H}$, as

$$f(x) = \langle\Psi|\psi(x)\rangle. \tag{4}$$

Selecting an appropriate VDE and state $|\Psi\rangle$, $f(\cdot)$ can be employed to approximate the desired input–output function. It is important to note that quantum state inner products are typically complex. However, a real kernel can be realized by either constraining the amplitudes of quantum feature states to real values or by considering the absolute square of the inner product.[31] Additionally, the magnitude of $f(\cdot)$ is bounded between 0 and 1, necessitating suitable scaling when applicable.

To realize their full potential, VQRAs require training to identify the optimal parameters ξ. This optimization process is conducted using a classical optimizer, which leverages a carefully crafted loss function. A common choice in regression tasks is the mean square error (MSE), defined as

$$L = \frac{1}{M}\sum_{m=1}^{M}(f(x^{(m)}) - y^{(m)})^2 + \Omega(||f||), \tag{5}$$

where $\Omega(||f||)$ represents a regularization term aimed at reducing overfitting. This term is crucial in ensuring the generalizability of the model and maintaining a balance between fitting the training data and avoiding excessive complexity in the learned function.

In exploring the optimal function $\tilde{f}(x) = \langle\Psi|\psi(x)\rangle$, we assume the state $|\Psi\rangle$ can be decomposed as $|\Psi\rangle = \sum_{m=1}^M \alpha_m|\psi(x^{(m)})\rangle$. This allows us to express $\tilde{f}(x)$ as

$$\tilde{f}(x) = \sum_{m=1}^{M}\alpha_m^*\langle\psi(x^{(m)})|\psi(x)\rangle$$

$$= \sum_{m=1}^{M} \alpha_m^* \kappa(x^{(m)}, x). \tag{6}$$

If we define $\alpha_m^* = \beta_m y^{(m)}$ and substitute this into Eq. (6), it yields

$$\tilde{f}(x) = \sum_{m=1}^{M} \beta_m y^{(m)} \kappa(x^{(m)}, x). \tag{7}$$

We can interpret Eq. (7) as follows:

(i) The state $|\Psi\rangle = \sum_{m=1}^{M} \beta_m^* y^{(m)} |\psi(x^{(m)})\rangle$ encapsulates all the training data through quantum superposition.

(ii) The kernel function quantifies the similarity between pairs of data points in the input space.

(iii) Predictions are made based on the weighted similarity between new data points and the existing training data, where the weights are given by β_m.

Setting β_m as a constant across all m is a simplistic approach, implying equal weighting for all training data points. However, this method may overlook the nuances of the training data distribution and the inherent correlations among data points. Instead, optimizing these weightings could illuminate the underlying correlations, and this optimization can be effectively achieved using machine learning techniques. In Appendix A, we give a detailed discussion of the optimization of classical kernel methods. In the following text, we focus on the implementation and performance of quantum-enhanced machine learning, whose parameters are optimized by a classical optimizer.

3.2. Implementation of parameterized quantum circuits

In this subsection, we delve into the implementation of PQCs for VQRAs, as illustrated in Fig. 1. The architecture comprises three main components: a memory circuit M_θ, a VDE circuit $E_\xi(x)$, and a swap-test circuit.[40] The memory circuit M_θ, a PQC itself, is characterized by an adjustable parameter set $\theta = (\theta_1, \theta_2, \ldots)$. Its primary role is to learn and prepare the state $|\Psi\rangle = M_\theta|0\rangle$, encapsulating the dataset information.

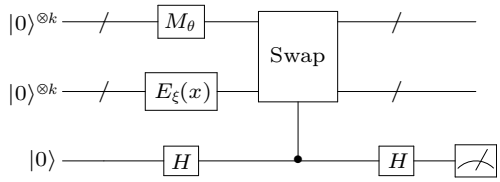


Fig. 1. Schematic representation of the quantum circuit utilized in VQRAs. It consists of three primary components: a memory circuit M_θ, a VDE circuit $E_\xi(x)$, and a swap-test circuit. The memory and VDE circuits are applied to k qubits initialized in state $|0\rangle^{\otimes k}$, while the swap-test circuit evaluates the overlap between the memory and encoded states.

The VDE circuit, another PQC, operates with two categories of parameters: trainable ξ and the input vector x. The circuit encodes x into a quantum state, where ξ is fine-tuned during the optimization process. Both the memory and encoder circuits act on k qubits initialized in the standard state $|0\rangle^{\otimes k}$.

The swap-test circuit's function is to evaluate the overlap between $|\Psi\rangle$ and $|\psi(x^{(m)})\rangle$, essentially computing the kernel function. According to quantum measurement theory, the probability of measuring '0' on the last qubit is given by $p(0) = \frac{1+|\langle\Psi|\psi(x^{(m)})\rangle|^2}{2}$,[40–42] providing a direct way to quantify the similarity between quantum states.

The construction of a PQC can vary significantly based on the specific research focus or the constraints of the quantum hardware.[12,43] Notably, PQCs exhibit promising advantages in expressive power, to the extent that even PQCs with a basic structure can outperform any classical neural network.[16] The ingenious design of ansatzes for PQCs lies outside the purview of this paper.[44,45] From the perspective of theoretical research, we enhance the expressive power of PQCs by augmenting their depths. For the memory circuit M_θ, we design it using D_M layers of a fundamental unit, each sharing the same structural framework. Each of these units is composed of single-qubit rotations $R_x(\theta_i)$ and controlled-NOT gates, as depicted in Fig. 2. Here, D_M represents the depth of M_θ, and increasing D_M correlates with enhanced expressivity of the memory circuit, allowing for more complex quantum state representations.[16,35]

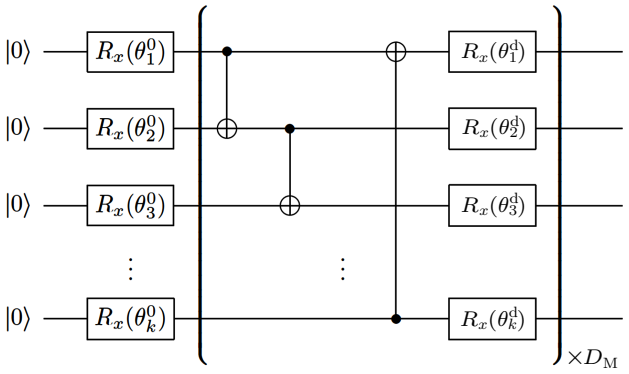


Fig. 2. Schematic representation of the memory circuit M_θ utilized in VQRAs. The circuit consists of D_M layers, each comprising a sequence of single-qubit rotations $R_x(\theta)$ along the X axis and controlled-NOT gates. Here, $R_x(\theta_i^j)$ represents the rotation operator for the i-th qubit in the j-th layer, and the combination of these rotations and entanglements forms the full structure of the circuit. This architecture is designed to enhance the expressivity and capacity of the memory circuit for encoding quantum information.

The construction of the VDE circuit follows a methodology similar to that of the memory circuit. Its detailed architecture is illustrated in Fig. 3, with D_E denoting the depth of the circuit $E_\xi(x)$. To foster entanglement within the circuit, we have designed the entanglement operations $E^{\mathrm{d},n}$ as follows:

$$\begin{aligned} E^{\mathrm{d},n} \;=\; & \mathrm{e}^{-\mathrm{i}X_kX_1\xi_{2k}^{\mathrm{d},n}}\,\mathrm{e}^{-\mathrm{i}X_{k-1}X_k\xi_{2k-1}^{\mathrm{d},n}}\cdots \\ & \times \mathrm{e}^{-\mathrm{i}X_2X_3\xi_{k+2}^{\mathrm{d},n}}\,\mathrm{e}^{-\mathrm{i}X_1X_2\xi_{k+1}^{\mathrm{d},n}}, \quad n = 1,2,\ldots,N, \end{aligned} \tag{8}$$

where X_i represents the Pauli-X operator acting on the i-th qubit. The depth D_E of the circuit $E_\xi(x)$ plays a critical role in determining its expressivity. Notably, setting $D_\mathrm{E} = 1$ results in a traditional encoding scheme. In contrast, a depth of $D_\mathrm{E} \geq 2$ corresponds to a data-reloading scheme, which has been shown to further enhance the expressivity of the circuit.[34]

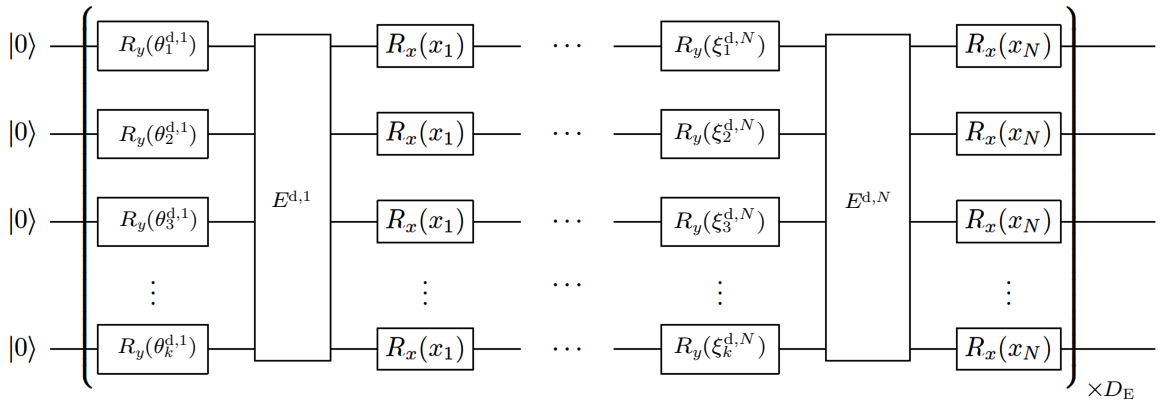


Fig. 3. Schematic representation of the VDE circuit $E_\xi(x)$. The circuit is composed of D_{E} layers, each containing a series of rotation gates $R_y(\xi_i^{\mathrm{d},j})$ and entanglement operations $E^{\mathrm{d},j}$, followed by $R_x(x_j)$ rotations. Here, ξ represents the set of trainable parameters defining the encoding scheme, while x denotes the input vector being encoded into the quantum state. This layered architecture allows for the flexible and expressive encoding of classical data into quantum states, catering to different encoding requirements and enhancing the circuit's overall capability.

3.3. Optimization

Training is a crucial step for VQRAs to effectively make predictions on new data. Initially, the parameters of VQRAs are set to random values. During each iteration of the training loop, the quantum device is tasked with state preparation, processing, and finally, performing measurements on designated qubits. These repeated operations yield estimates of $f(x^m)$. To determine the gradient of $f(x^m)$, it may be necessary to run variations of the circuit. These results, combined with a well-constructed loss function, enable a classical optimizer to refine the parameters within the quantum device. In our study, we employ MSE as the loss function, as detailed in Eq. (5). This choice of loss function aligns with the objective of minimizing the deviation between the predicted and actual values, thus enhancing the accuracy and reliability of the VQRA's predictions.

QML encompasses a variety of learning rules, which can be broadly classified into gradient-free and gradient-based methods.[33,46] The choice between these methods depends largely on the specific problem at hand, with each offering its own set of advantages and drawbacks. Gradient-based methods are often favored due to their rapid convergence and high precision, especially in scenarios involving a vast parameter space. To acquire gradient information, several techniques have been developed.[46] In our study, we opt for numerical differentiation as a straightforward approach to approximate gradients. This simplification allows us to utilize gradient-based optimization algorithms, such as Adam,[47] for updating the parameters in the PQCs. Employing such methods facilitates efficient and precise tuning of the circuit parameters, thereby enhancing the overall performance of the quantum machine learning model.

4. Numerical simulations

This section details the numerical simulations we conducted for VQRAs using the PennyLane software library. PennyLane, renowned for its versatility and open-source nature, serves as a comprehensive platform for quantum computing, quantum machine learning, and quantum chemistry. Its universal compatibility with various gate-based quantum computing platforms and simulators as backends makes it an ideal choice for a wide range of quantum algorithm implementations.[48] This flexibility is particularly valuable, not only for educational and research demonstrations but also for the future application of these algorithms to practical problems.[49]

In our simulations, we focused on assessing the performance of VQRAs, exploring the quantum correlations present in the quantum states, and examining the impacts of noise on the system. These simulations provide insights into the efficacy and robustness of VQRAs under various conditions, offering valuable data for their further development and application.

4.1. Performance

In evaluating the performance of VQRAs, we initially focus on their ability to fit various functions. Here, the quantum circuits are configured with $k = D_{\mathrm{M}} = 3$ and $D_{\mathrm{E}} = 6$, balancing complexity and computational feasibility.

Our simulations involved fitting tasks for several distinct functions: $f_1(x) = x^2$, $f_2(x) = \mathrm{e}^x/\mathrm{e}$, $f_3(x) = \sin^2(\pi x)$, and $f_4(x) = 1/(1+\mathrm{e}^{-10x})$, all within the domain $x \in [-1, 1]$. The ranges of all selected functions are all $[0, 1]$. Thus, we take the absolute value of overlap between encoded states and memory states as predicted values of the above functions, which is regarded as a classical post-processing on the complex kernel. The training data for each function were uniformly distributed across this range. To mimic realistic data conditions, we introduced small Gaussian noise with a standard deviation of $\sigma = 0.01$ to the training data.

The outcomes of these simulations, illustrated in Fig. 4, demonstrate that VQRAs, even with a limited number of qubits and shallow circuit depth, are capable of closely approximating all the tested functions. This success underscores the substantial expressive power inherent in the Hilbert space, even when harnessed by relatively simple quantum circuits.

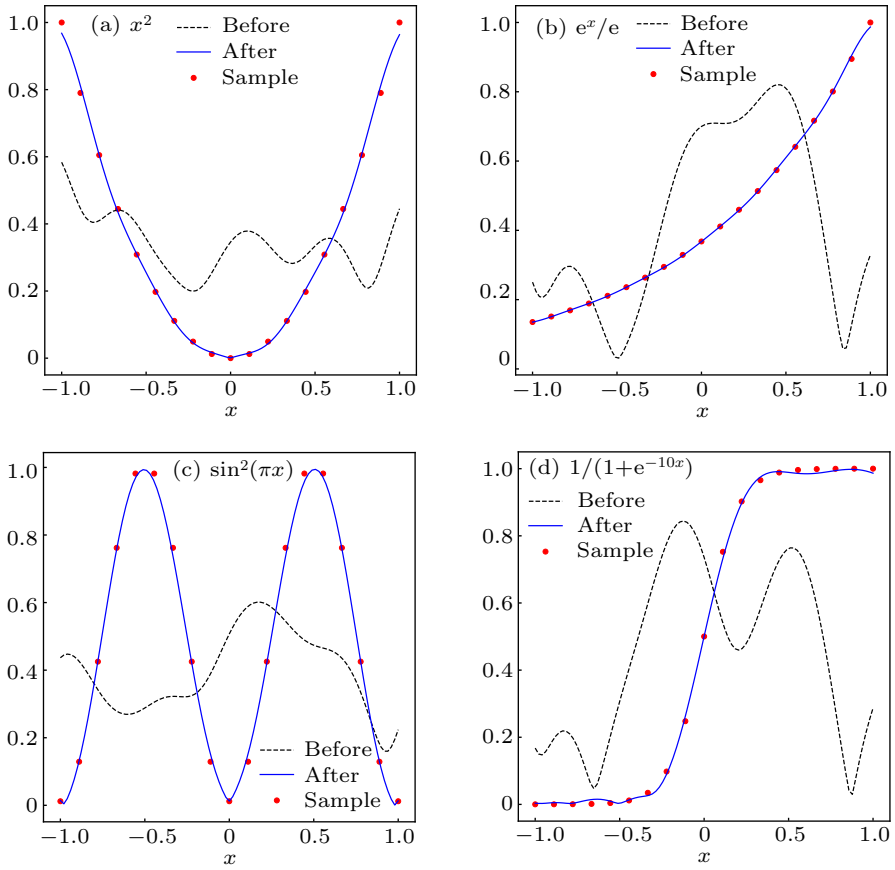


Fig. 4. Demonstration of VQRAs in fitting various single-variable functions. Panels (a)–(d) depict the fitting results for x^2, e^x/e, $\sin^2(\pi x)$, and $1/(1+e^{-10x})$ respectively. The red dots represent the training samples. The dashed lines represent the model's performance on sample fitting before training, while the solid lines represent the performance after training. The quantum circuit parameters were configured with $k = D_M = 3$ and $D_E = 6$. Each function was approximated by training the VQRA model through 2000 iterations. These results highlight the algorithm's capability in accurately modeling different mathematical functions, showcasing its adaptability and precision.

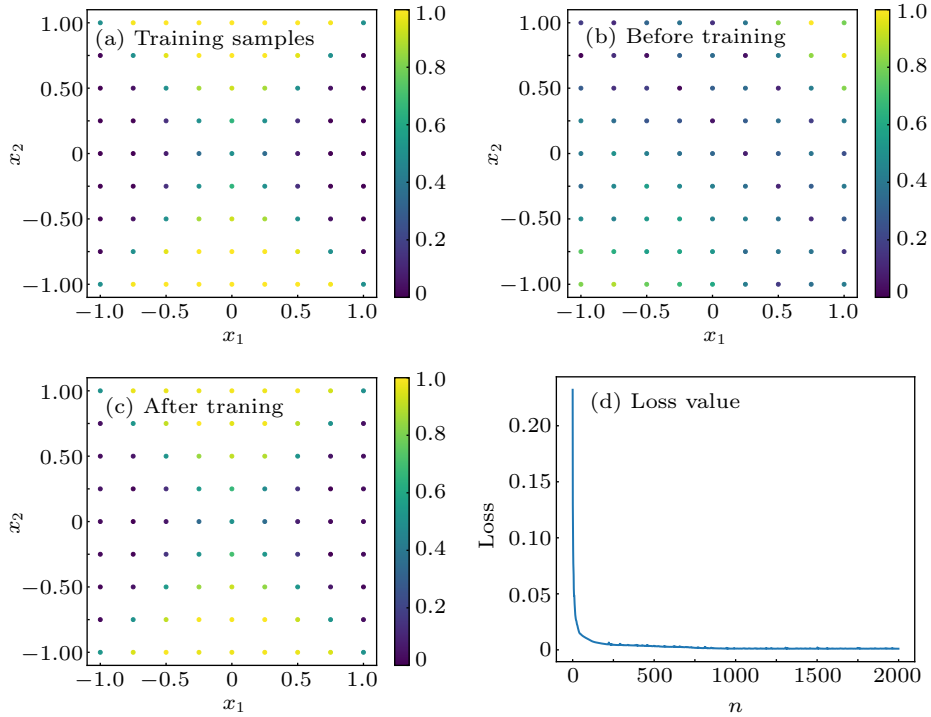


Fig. 5. Illustration of VQRA's efficacy in fitting the multivariate function $f_5(x_1,x_2) = 1/\left[1+e^{10(x_1^2-x_2^2)}\right]$. The circuit parameters were set to $k = D_M = 3$ and $D_E = 6$. Panel (a) displays the training data samples. Panel (b) shows the initial model performance before training. Panel (c) demonstrates the improved fitting accuracy after 2000 training iterations. Panel (d) depicts the progression of the loss function throughout the training process. This sequence of images highlights the significant improvement in model performance post-training, showcasing the adaptability and learning capacity of VQRA for complex multivariate functions.

In addition to single-variable functions, we extended our investigation to multivariate functions using VQRAs. An exemplary case is presented in Fig. 5, where we demonstrate the fitting of the function $f_5(x_1,x_2) = 1/\left[1+\mathrm{e}^{10(x_1^2-x_2^2)}\right]$ within the domain $x_1,x_2 \in [-1,1]$. This result exemplifies the capability of VQRAs to accurately model functions involving multiple variables, highlighting their versatility and effectiveness in handling complex datasets.

Furthermore, the successful fitting of both f_4 and f_5 functions suggests that VQRAs can be adapted for quantum classification tasks. By setting appropriate classification boundaries, these algorithms can be trained to distinguish between different classes of data, as supported by existing studies in quantum classification.[19,30] This versatility makes VQRAs a promising tool not only for regression but also for classification problems in quantum machine learning.

4.2. Quantum correlations in VQRAs

Quantum correlations are a cornerstone of quantum information processing, with their unique properties offering significant advantages over classical systems.[50–53] A deep understanding of these correlations, encompassing both their characterization and quantification, is essential for harnessing the full potential of quantum physics in practical applications. In the context of VQRAs, quantum correlations manifest within both the memory and encoded states. While a universal metric for quantifying quantum correlations in multibody quantum systems remains elusive,[54,55] our study takes a qualitative approach to examine the influence of quantum correlations on VQRAs. We aim to elucidate how these correlations contribute to the performance and efficacy of the algorithm, thereby providing insights into the intricate interplay between quantum mechanics and machine learning.

Initially, we examined VQRAs in an idealized, noise-free environment, focusing specifically on the fitting of function f_4. To delve deeper into the role of quantum correlations, we analyzed four distinct configurations of quantum circuits, each allowing for different levels of correlation. The configurations were as follows:

(i) **Configuration 1:** All entanglement gates were removed from both the memory and the VDE circuits.

(ii) **Configuration 2:** Entanglement gates in the memory circuit were retained, while those in the VDE circuit were removed.

(iii) **Configuration 3:** Entanglement gates in the memory circuit were removed, but those in the VDE circuit were kept.

(iv) **Configuration 4:** Both the memory and the VDE circuits retained their entanglement gates.

For each configuration, the depth D_E was varied from 1 to 6. We conducted 10 training rounds for each circuit, with 2000 iterations per round. Post-training, we computed the average values and standard deviations of the loss function. The results, depicted in Fig. 6, indicate that for shallow circuits, where quantum correlation is limited, configuration (Config.) 4 consistently outperforms the other configurations. However, as the circuit depth increases, this advantage diminishes. Deep circuits without correlations already exhibit satisfactory performance, obscuring the benefits of quantum correlations. The impact of quantum correlations is more pronounced in medium-depth circuits, where increased leveraging of correlations corresponds with improved fitting accuracy. For instance, in our simulations, we observed that $l_1 > l_2 > l_3 \approx l_4$ when D_E was set to 2, 3, 4, and 5, where l_i represents the loss value.

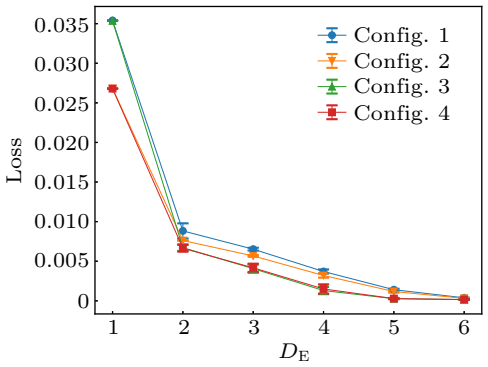


Fig. 6. Comparative analysis of loss curves across different configurations and encoding depths in VQRAs. Here, the quantum circuit's structural parameters were configured with $k = D_\mathrm{M} = 3$. To investigate the influence of circuit depth on performance, we varied D_E from 1 to 6 for each configuration. This figure highlights how different levels of quantum entanglement, influenced by the depth of the circuits, affect the loss and, consequently, the efficacy of the VQRA model. The results provide insights into the optimal balance between circuit depth and quantum correlation for efficient quantum information processing.

In practical scenarios, encoding processes in VQRAs cannot be completely isolated from environmental interactions, leading to inevitable errors caused by noise. Such the noise can be mathematically represented by noisy quantum channels with Kraus operators, defined as

$$\varepsilon(\rho) = \sum_k E_k \rho E_k^\dagger,$$

where each E_k is a Kraus operator satisfying $\sum_k E_k^\dagger E_k = I$. Given the myriad types of noisy quantum channels, we assume local and identical noise across all qubits for simplicity. This assumption is reasonable in scenarios where qubits are physically well-separated and uniformly affected by similar environmental factors.

We particularly focus on three prevalent types of noisy quantum channels for a single qubit: amplitude damping, phase damping, and symmetric depolarizing channels. These channels are characteristic of non-Markovian processes, leading to continuous loss of quantum information to the environment. Among them, symmetric depolarizing channels are noteworthy for their severe impact, replacing the qubit state with a completely mixed state at a probability p.[56] The channel is mathematically modeled as

$$\begin{aligned}\varepsilon(\rho) &= \left(1-\frac{3p}{4}\right)\rho + \frac{p}{4}(X\rho X + Y\rho Y + Z\rho Z) \\ &= \frac{pI}{2} + (1-p)\rho, \end{aligned} \tag{9}$$

where X, Y, Z denote Pauli operators, and $p \in [0,1]$ represents the noise strength.

To simplify our analysis, we introduced noise channels immediately following the VDEs, resulting in the noisy encoded states:

$$\tilde{\rho}_{x^{(m)}} = \left(\otimes_{i=1}^{k} \varepsilon_i\right)(\rho_{x^{(m)}}), \tag{10}$$

where $\rho_{x^{(m)}} = |\psi(x^{(m)})\rangle\langle\psi(x^{(m)})|$ is the noise-free encoded state for the input vector $x^{(m)}$, and ε_i represents the single-qubit noisy channel acting on the i-th qubit.

We conducted a detailed examination of the effects of noise on the effectiveness of VQRAs using the configurations outlined earlier. The structural parameters for these simulations were set to $k = D_{\mathrm{M}} = 3$ and $D_{\mathrm{E}} = 6$. We incrementally increased the intensity of the noise in each scenario. For each noise level, the circuit was subjected to 10 training rounds, with each round consisting of 2000 training iterations.

As depicted in Fig. 7, a consistent pattern emerged across all quantum circuit configurations: the loss value escalated as the noise intensity increased. This trend can be attributed to the noise masking the information encoded within the quantum states, thereby diminishing the algorithm's ability to accurately capture the intended data patterns. Interestingly, however, it was observed that configurations allowing for greater quantum correlations demonstrated relatively better resilience to noise. This suggests that the presence of quantum correlations within the circuits contributes to maintaining the VQRA's performance even in noisy conditions, highlighting the robustness of quantum correlations in preserving information integrity.

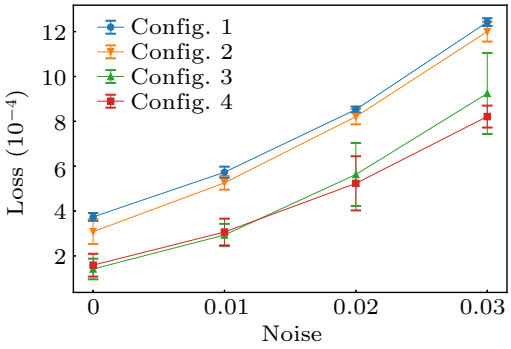


Fig. 7. Analysis of VQRA performance under varying noise levels for different circuit configurations. This figure presents the loss curves as a function of increasing noise intensity, illustrating how the performance of VQRAs is impacted in noisy environments. Each curve represents a different configuration of quantum entanglement within the circuit, providing insights into the resilience of VQRAs against environmental noise and the potential role of quantum correlations in mitigating noise effects.

An interesting observation from our study is that the slope of the loss curve in Fig. 7 can serve as an indicator of a quantum circuit's resistance to noise. A flatter slope signifies stronger resilience against noise disturbances. In scenarios with low noise intensity ($p < 0.03$), the differences in noise resistance across configurations become evident. In configuration 1, where there is an absence of quantum correlations among qubits, the information is localized within individual qubits. This localization makes the information highly susceptible to noise, leading to a sharp increase in the loss curve's slope. This is clearly observed in Fig. 7 as a rapid incline for configuration 1. Contrastingly, configuration 4, which allows quantum correlations among qubits, demonstrates a more robust response to noise. Here, information is not only stored in individual qubits but also in the correlations between them. During noisy conditions, the VDE learns to encode information into these quantum correlations, providing a buffer against local noise disruptions. This is reflected in Fig. 7 as a relatively gentler slope for configuration 4. However, it is important to note that in high-noise environments, where noise levels are substantial, even non-local quantum correlations are adversely affected, leading to a steep increase in the loss curve's slope. Under such conditions, the VDE struggles to find a secure means of encoding information, and the system's resistance to noise significantly weakens.

5. Discussion and conclusion

In this study, we explored variational data encoding within the framework of VQRAs and examined the pivotal role of quantum correlations in the encoded states. Utilizing PQCs, whose parameters are optimized through machine learning techniques, we successfully demonstrated the efficacy of quantum features in encoding comprehensive information from training data. Our findings suggest that the superposition states in the memory circuit not only store data but may also capture the global structure of the training dataset, offering avenues for quantum enhancement. In our VQRA models, the number of trainable parameters is $k(2ND_{\mathrm{E}} + D_{\mathrm{M}} + 1)$. The larger the hyper-parameters k, D_{E}, and D_{E}, the more expressive the circuits are. Suppose k, D_{E}, and D_{E} scales linearly with N because of the expressive power of quantum circuits.[16] Then the number of trainable parameters grows polynomially, which ensures the effectiveness of the training.

Our numerical simulations revealed that VQRAs can achieve remarkable performance even on quantum devices of a limited scale. This showcases the potential of VQRAs in quantum machine learning applications, especially considering the current capabilities of available quantum technology. Furthermore, we investigated the impact of noise, an inherent characteristic of NISQ devices, on the performance of VQRAs. While noise inevitably impairs the effectiveness of these algorithms, our studies indicate that quantum correlations can significantly bolster the robustness of encoded states against local noise, such as symmetric depolarizing disturbances. This resilience stems from the dual nature of information storage in VQRAs: both in local quantum states and in non-local quantum correlations. Our learning-based encoding schemes adaptively shift towards utilizing quantum correlations in noisy environments, thereby safeguarding the encoded information.

However, our study faced challenges in quantitatively assessing the role of quantum correlations due to the absence of

a universally accepted metric for multi-body quantum correlations. Future research endeavors could focus on identifying a suitable measure for quantum correlation and investigating how it varies during training under different noise conditions.

In conclusion, our findings underscore the importance of encoding information into quantum correlations for enhancing quantum machine learning algorithms. Quantum correlations not only provide an expanded space for information storage, potentially conserving quantum resources, but also offer a means to combat local noise. These characteristics are particularly promising for quantum algorithms designed for NISQ devices, suggesting a bright future for quantum-enhanced machine learning.

Appendix A: Derivation of optimization in classical kernel method

Here, we give the derivation of optimization problems in classical kernel methods. In classical kernel methods, the kernel functions are usually selected in advance based on experience. Consider a nonlinear feature mapping

$$\psi : x \in \mathbb{R}^N \mapsto \psi(x) \in \mathbb{R}^Q \tag{A1}$$

that maps the nonlinear relations into linear ones such that a real-valued and nonlinear function $f(x)$ can be approximated in the form of

$$f(x) = \sum_{i=1}^{Q} w_i \psi_i(x), \tag{A2}$$

where $w = (w_1, w_2, \ldots, w_Q)^{\mathrm{T}}$ is the undetermined weighting. The weighting w, for instance, can be determined by minimization the MSE

$$L = \sum_{m=1}^{M} (f(x^{(m)}) - y^{(m)})^2 + \xi \sum_{i=1}^{Q} w_i^2, \tag{A3}$$

where the ξ term is a regularization term to avoid overfitting.

Substituting Eq. (A2) into Eq. (A3), we obtain

$$\begin{aligned} L &= \sum_{m=1}^{M} (\sum_{i=1}^{Q} w_i \psi_i(x^{(m)}) - y^{(m)})^2 + \xi \sum_{i=1}^{Q} w_i^2 \\ &= (\psi w - y)^{\mathrm{T}} (\psi w - y) + \xi w^{\mathrm{T}} w. \end{aligned} \tag{A4}$$

Here ψ is the matrix with elements $\psi_{mi} = \psi_i(x^{(m)})$, and $y = (y^{(1)}, y^{(2)}, \ldots, y^{(M)})^{\mathrm{T}}$. Taking the derivative with respect to w and setting it to zero, we obtain

$$\psi^{\mathrm{T}}(\psi w - y) + \xi w = 0. \tag{A5}$$

the solution turns to be

$$w = (\psi^{\mathrm{T}} \psi + \xi I)^{+} \psi^{\mathrm{T}} y, \tag{A6}$$

where A^{+} represents the Moore–Penrose pseudoinverse of the matrix A. We rewrite Eq. (A5) to obtain

$$w = \xi^{-1} \psi^{\mathrm{T}} (y - \psi w) = \psi^{\mathrm{T}} \alpha. \tag{A7}$$

This means w can be written as a linear combination of the training points in the feature space, with $\alpha = \xi^{-1}(y - \psi w)$. Thus, we have

$$\alpha = \xi^{-1}(y - \psi \psi^{\mathrm{T}} \alpha), \tag{A8}$$

resulting

$$\alpha = (\psi \psi^{\mathrm{T}} + \xi I)^{+} y. \tag{A9}$$

Substituting Eq. (A7) into Eq. (A2) gives us

$$\begin{aligned} f(x) &= \sum_{i=1}^{Q} \sum_{m=1}^{M} \psi_i(x^{(m)}) \alpha_m \psi_i(x) \\ &= \sum_{m=1}^{M} \alpha_m \langle \psi(x^{(m)}), \psi(x) \rangle. \end{aligned} \tag{A10}$$

If we define the kernel function as

$$k(x', x) = \langle \psi(x'), \psi(x) \rangle, \tag{A11}$$

we have

$$f(x) = \sum_{m=1}^{M} \alpha_m k(x^{(m)}, x), \tag{A12}$$

the same as shown in Eq. (6). The optimal α is

$$\alpha = (K + \xi I)^{+} y, \tag{A13}$$

where K is the kernel matrix with elements $K_{mn} = k(x^{(m)}, x^{(n)})$.

Acknowledgement

Project supported by the National Natural Science Foundation of China (Grant Nos. 12105090 and 12175057).

References

[1] Nielsen M A and Chuang I L 2010 *Quantum computation and quantum information: 10th anniversary edition* (Cambridge University Press)
[2] Shor P W 1997 *SIAM J. Comput.* **26** 1484
[3] Harrow A W, Hassidim A and Lloyd S 2009 *Phys. Rev. Lett.* **103** 150502
[4] Preskill J 2018 *Quantum* **2** 79
[5] Bharti K, Cervera-Lierta A, Kyaw T H, Haug T,Alperin-Lea S, Anand A, Degroote M, Heimonen H, Kottmann J S, Menke T, Mok W K, Sim S, Kwek L and Aspuru-Guzik A 2022 *Rev. Mod. Phys.* **94** 015004
[6] Callison A and Chancellor N 2022 *Phys. Rev. A* **106** 010101
[7] Cerezo M, Arrasmith A, Babbush R, Benjamin S C, Endo S, Fujii K, McClean J R, Mitarai K, Yuan X, Cincio L and Coles P J 2021 *Nat. Rev. Phys.* **3** 625
[8] Zhang S X, Wan Z Q, Lee C K, Hsieh C Y, Zhang S and Yao H 2022 *Phys. Rev. Lett.* **128** 120502
[9] Jones T, Endo S, McArdle S, Yuan X and Benjamin S C 2019 *Phys. Rev. A* **99** 062304
[10] Liu H L, Wu Y S, Wan L C, Pan S J, Qin S J, Gao F and Wen Q Y 2021 *Phys. Rev. A* **104** 022418
[11] Xu X, Sun J, Endo S, Li Y, Benjamin S C and Yuan X 2021 *Sci. Bull.* **66** 2181
[12] Kandala A, Mezzacapo A, Temme K, Takita M, Brink M, Chow J M and Gambetta J M 2017 *Nature* **549** 242
[13] Chalumuri A, Kune R and Manoj B S 2021 *Quantum Inf. Process.* **20** 119
[14] Jordan M I and Mitchell T M 2015 *Science* **349** 255

[15] Benedetti M, Lloyd E, Sack S and Fiorentini M 2019 *Quantum Sci. Technol.* **4** 043001
[16] Du Y, Hsieh M H, Liu T and Tao D 2020 *Phys. Rev. Res.* 2 033125
[17] Nghiem N A, Chen S Y C and Wei T C 2021*Phys. Rev. Res.* **3** 033056
[18] LaRose R and Coyle B 2020 *Phys. Rev. A* **102** 032420
[19] Schuld M, Bocharov A, Svore K M and Wiebe N 2020 *Phys. Rev. A* **101** 032308
[20] Dallaire-Demers P L and Killoran N 2018 *Phys. Rev. A* **98** 012324
[21] Lloyd S and Weedbrook C 2018 *Phys. Rev. Lett.* **121** 040502
[22] Gao X, Anschuetz E R, Wang S T, Cirac J I and Lukin M D 2022 *Phys. Rev. X***12** 021037
[23] Biamonte J, Wittek P, Pancotti N, Rebentrost P, Wiebe N and Lloyd S 2017 *Nature* **549** 195
[24] Dunjko V, Taylor J M and Briegel H J. 2016 *Phys. Rev. Lett.* **117** 130501
[25] Dunjko V and Briegel H J 2018 *Rep. Prog. Phys* **81** 074001
[26] Ma Y G, Pang L G, Wang R and Zhou K. 2023 *Chin. Phys. Lett.* **40** 122101
[27] Bai S C, Tang Y C and Ran S J 2022 *Chin. Phys. Lett.* **39** 100701
[28] Xie J and Dong Y 2023 *Chin. Phys. B* **32** 120302
[29] Wang X Y, Dong C and Liu X 2024 *Chin. Phys. Lett.* **41** 031201
[30] Havlíček V, Córcoles A D, Temme K, Harrow A W, Kandala A, Chow J M and Gambetta J M 2019 *Nature* **567** 209
[31] Schuld M and Killoran N 2019 *Phys. Rev. Lett.* **122** 040504
[32] Giovannetti, V, Lloyd S and Maccone L 2008 *Phys. Rev. Lett.* **16** 160501
[33] Mitarai K, Negoro M, Kitagawa M and Fujii K 2018 *Phys. Rev. A* **98** 032309
[34] Pérez-Salinas A, Cervera-Lierta A, Gil-Fuster E and Latorre J I 2020 *Quantum* **4** 226
[35] Schuld M, Sweke R and Meyer J J 2021 *Phys. Rev. A* **103** 032430
[36] Jerbi S, Fiderer L J, Nautrup H P, Kübler J M, Briegel H J and Dunjko V 2023 *Nat. Commun.* **14** 517
[37] Hofmann T, Schölkopf B and Smola A J 2008 *Ann. Statist.* **36** 1171
[38] Lloyd S, Schuld M, Ijaz A, Izaac J and Killoran N 2020 arXiv: 2001.03622 [quant-ph]
[39] Hubregtsen T, Wierichs D, Gil-Fuster E, Derks P J H S, Faehrmann P K and Meyer J J 2022 *Phys. Rev. A* **106** 042431
[40] Buhrman H, Cleve R, Watrous J and Wolf R. 2001 *Phys. Rev. Lett.* **87** 167902
[41] Garcia-Escartin J C and Chamorro-Posada P 2013 *Phys. Rev. A* **87** 052330
[42] Zhao T H, Wang M H and Zhou B 2021 *Quantum Inf. Process.* **20** 1
[43] Tang H L, Shkolnikov V O, Barron G S, Grimsley H R, Mayhall N J, Barnes E and Economou S E 2021 *PRX Quantum* **2** 020310
[44] Du Y, Huang T, You S, *et al.* 2022 *npj Quantum Inf.* **8** 62
[45] Guo X, Muta T and Zhao J 2024 arXiv: 2405.05021 [quant-ph]
[46] Schuld M, Bergholm V, Gogolin C, Izaac J and Killoran N 2019 *Phys. Rev. A* **99** 032331
[47] Kingma D and Ba J 2014 arXiv: 1412.6980 [cs.LG]
[48] Bergholm V, Izaac J, Schuld M, *et al.* 2018 arXiv: 1811.04968 [quant-ph].
[49] Antipov A V, Kiktenko E O and Fedorov A K 2022 *PLOS One* **17** 1
[50] Horodecki R, Horodecki P, Horodecki M and Horodecki K 2009 *Rev. Mod. Phys.* **81** 865-
[51] Vedral V 2014 *Nat. Phys.* **10** 256
[52] Erhard M, Krenn M and Zeilinger A 2020 *Nat. Rev. Phys.* **2** 365
[53] Köhnke S, Agudelo E, Schünemann M, Schlettwein O, Vogel W, Sperling J and Hage B 2021 *Phys. Rev. Lett.* **126** 170404
[54] Modi K, Brodutch A, Cable H, Paterek T and Vedral V 2012 *Rev. Mod. Phys.* **84** 1655
[55] Adesso G, Bromley T R and Cianciaruso M 2016 *J. Phys. A: Math. Theor.* **49** 473001
[56] Fontana E, Fitzpatrick N, Ramo D M, Duncan R and Rungger I 2021 *Phys. Rev. A* **104** 022403