IOE | Intelligent Opto-Electronics

Quantum embedding learning on variational photonic quantum circuits

Guanheng Ren, Donglai Wang, Lizheng Yan, Yuancheng Zhan, Yiming Ma, Zhanshan Wang, Hui Zhang and Xinbin Cheng

Citation: Ren GH, Wang DL, Yan LZ, et al. Quantum embedding learning on variational photonic quantum circuits. *Intelligent Opto-Electronics* **1**, 250010(2025).

https://doi.org/10.29026/ioe.2025.250010

Received: 3 October 2025; Accepted: 16 December 2025; Published online: 25 December 2025

IOE | Intelligent Opto-Electronics
www.oejournal.org/ioe

Article
December 2025, Vol. 1, No. 2

DOI: 10.29026/ioe.2025.250010

Open Access

Quantum embedding learning on variational photonic quantum circuits

Guanheng Ren[1,2,3,4†], Donglai Wang[1,2,3,4†], Lizheng Yan[1,2,3,4], Yuancheng Zhan[5], Yiming Ma[6,7*], Zhanshan Wang[1,2,3,4], Hui Zhang[1,2,3,4*] and Xinbin Cheng[1,2,3,4*]

Abstract: Quantum embedding is a key technique in quantum machine learning, mapping classical data into a high-dimensional Hilbert space through parameterized quantum circuits to enhance data separability. While effective on general quantum devices, its implementation on photonic platforms faces severe challenges, as entangling gates rely on post-selection and lead to exponentially decreasing success probabilities when cascaded. To address this, we propose a quantum embedding learning approach tailored for photonic quantum circuits, which incorporates multi-photon post-selection into a variational ansatz, enabling the efficient generation of entanglement while maintaining circuit scalability on NISQ devices. By training the transfer matrices, our method achieves the desired photon distribution probabilities and input-output correspondence. We demonstrate its versatility through three tasks: Bell-state projection, quantum support vector machines (qSVM), and quantum clustering of *K*-means. The experimental results indicate enhanced performance in state discrimination, classification, and clustering, showing that the method provides a practical route to implement quantum embedding in integrated photonic circuits.

Keywords: photonic quantum circuits; quantum embedding learning; variational quantum circuits; photonic integrated circuits; integrated quantum photonics

Introduction

In quantum machine learning tasks involving classical data, a critical initial step is quantum embedding, which maps classical data into a high-dimensional Hilbert space via quantum circuits, thereby enhancing data separability[1–3]. Quantum embedding can be implemented via various encoding schemes, including angle, basis-state, and amplitude embedding, where parameterized quantum circuits map input data into quantum states $|x\rangle$, and quantum measurements enable classification or other learning tasks[4–6]. Resonating with metric learning[7,8], quantum embedding is central to harnessing the learning potential of quantum computing, as it determines the expressivity and trainability of quantum models[9–12].

Existing studies on quantum embedding primarily focus on efficiently mapping classical data within quantum circuits, treating the embedding process as an optimization problem. Representative works include directly training feature mappings via quantum metric learning to enhance data separability[13]; leveraging quantum random access codes to increase the expressive power of embeddings[14,15]; and introducing trainable quantum embedding kernels that combine kernel–target alignment with noise mitigation strategies to achieve robust classification[16–18]. In addition, some studies integrate quantum embeddings with classical kernel methods, demonstrating their potential applicability within conventional machine learning frameworks[19,20].

From an implementation standpoint, integrated quantum photonic circuits represent a vital platform for quantum computing. Characterized by scalability, stability, and programmability[21–24], integrated quantum photonic circuits have demonstrated significant potential in chip-scale

Received: 3 October 2025
Accepted: 16 December 2025
Published online: 25 December 2025

[1]Optical Engineering, School of Physics Science and Engineering, Tongji University, Shanghai 200092, China; [2]MOE Key Laboratory of Advanced Micro-Structured Materials, Shanghai 200092, China; [3]Shanghai Institute of Intelligent Science and Technology, Tongji University, Shanghai 200092, China; [4]Shanghai Frontiers Science Center of Digital Optics, Shanghai 200092, China; [5]Centre for Quantum Technologies, National University of Singapore, Singapore 999002, Singapore; [6]School of Microelectronics, Shanghai University, Shanghai 201800, China; [7]Shanghai Collaborative Innovation Center of Intelligent Sensing Chip Technology, Shanghai University, Shanghai 201800, China.
[†]These authors contributed equally to this work.
*Correspondence: YM Ma, E-mail: yimingma@shu.edu.cn; H Zhang, E-mail: jovie_huizhang@tongji.edu.cn; XB Cheng, E-mail: chengxb@tongji.edu.cn

quantum communication[25,26], quantum information processing[27–29], Boson sampling[30–32], and the quantum simulation of chemical and physical systems[33,34]. Moreover, owing to their programmability, integrated photonic circuits also show promising applications in emerging areas such as quantum machine learning (QML)[35]. However, conducting quantum embedding research on photonic platforms faces distinct challenges. The primary difficulty arises because the implementation of entangling gates typically relies on post-selection, rendering the operations inherently probabilistic. Consequently, directly employing parameterized ansatzes of the conventional circuit-model leads to a drastic decrease in the probability of success when cascading multiple entangling gates[36–38]. In the noisy intermediate-scale Quantum (NISQ) era, this issue is exacerbated by the lack of mature fault-tolerance mechanisms, imposing substantial limitations on the achievable depth of photonic quantum circuits[39].

While NISQ devices are constrained by noise and limited depth, they offer significant potential for quantum machine learning, particularly through variational quantum embeddings that map classical data into high-dimensional Hilbert spaces. However, realizing such embeddings on integrated photonic platforms faces a critical hurdle: the probabilistic post-selection required for non-linearity cannot be directly cascaded in a modular fashion without exponentially diminishing success rates. To resolve this, we reformulate the embedding process not as a sequence of independent layers, but as a unified trainable non-unitary mapping matrix $\boldsymbol{W}(\vec{\theta})$. By treating the circuit evolution and post-selection as a holistic operator constrained by normalization, we effectively bypass the cascading bottleneck, ensuring a physically realizable and efficient implementation on realistic hardware.

Here, we propose a quantum embedding learning approach based on photonic quantum circuits. It introduces a multi-photon post-selection mechanism within a variational ansatz, enabling the generation of complex entanglement while preserving the dual-rail structure. By parameterizing and optimizing the ansatz weight matrix W, we efficiently approximate the target quantum output photon distribution on finite-size NISQ platforms, thereby enabling optimized learning for specific quantum tasks. To demonstrate the versatility of this approach, we apply it to three representative tasks: i) designing a projector capable of simultaneously distinguishing all four Bell states, highlighting the potential of multi-photon post-selection circuits in quantum measurement; ii) constructing a quantum support vector machine (qSVM) for supervised learning, leveraging quantum embedding to enhance classification capability in a high-dimensional Hilbert space; and iii) implementing quantum K-means clustering for unsupervised learning, improving clustering performance through similarity estimation between quantum states. Our experiments demonstrate the feasibility of applying the proposed variational embedding framework to diverse tasks. These results highlight the potential of implementing quantum embedding on photonic quantum circuit platforms, providing an exploratory pathway toward practical quantum machine learning applications.

Design and architecture

Quantum embedding learning

Quantum embedding refers to mapping classical data into quantum states, enabling subsequent quantum algorithms to operate within the Hilbert space. In our framework, we first employ dual-rail encoding at the physical level, where a single qubit is represented by a photon distributed across two spatial modes $(\hat{a}_{2i-1}, \hat{a}_{2i})$. If a photon enters the $\hat{a}_{2i-1}$ path, it is encoded as $|0\rangle_i$; conversely, if it enters the $\hat{a}_{2i}$ path, it is encoded as $|1\rangle_i$. This process is abstracted into a feature map where classical data x are embedded into the Hilbert space by applying data-dependent unitary operations $\boldsymbol{U}(x)$, such as rotation gates (R_x, R_y), to an initial reference state, yielding quantum states of the form

$$|\phi(x)\rangle = \boldsymbol{U}(x)|0\rangle \ . \tag{1}$$

To further enhance expressivity, we extend this mapping to

$$|\phi(x)\rangle = \boldsymbol{W}(\vec{\theta})\boldsymbol{U}(x)|0\rangle \ , \tag{2}$$

where $\boldsymbol{W}(\vec{\theta})$ corresponds to a physically implemented non-linear transformation. The vector $\vec{\theta}$ represents trainable variational parameters. The overall architecture is shown in Fig. 1, where two groups of classical data are exemplified using triangles and squares. Subsequently, variational photonic circuits are employed to construct a hyperplane via quantum embedding learning, enabling the separation of the two data classes.

A schematic of the entire embedding process is provided in Fig. 1(a). In practice, the weight matrix $\boldsymbol{W}(\vec{\theta})$ is optimized in the variational stage to realize the desired quantum algorithm. A cost function is first defined according to the task objective, after which the input data are encoded onto quantum modes via unitary transformations $\boldsymbol{U}(x)$. These states are then processed by the parametrized circuit $\boldsymbol{W}(\vec{\theta})$, where $\vec{\theta}$ is optimized. The physical implementation is illustrated in Fig. 1(b). Measurement outcomes are obtained via post-selection, with final detection performed by single-photon detectors (Fig. 1(c)). Detected signals are fed back to the optimization routine, where gradient-based algorithms iteratively update $\vec{\theta}$ to minimize the cost function. The resulting quantum states are represented in the Hilbert space, and the visualization of the Bloch sphere confirms that the two data classes are successfully distinguished.

Variational photonic circuits

We present a framework to realize an n-qubit quantum

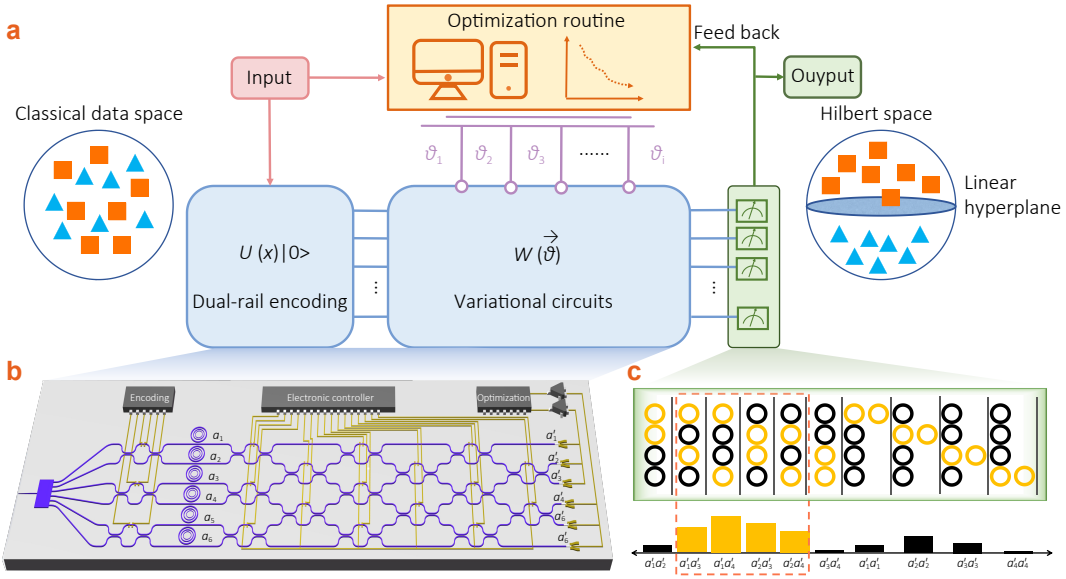

Fig. 1 | Schematic diagram of the quantum embedding learning architecture based on variational photonic quantum circuits. (**a**) Two classes of classical data in the input space are represented by triangles and squares. Through quantum embedding learning, the data are mapped into the Hilbert space, where a trained hyperplane separates the two classes. The workflow involves constructing a cost function based on the task. Original data are mapped by $U(x)$ and encoded via dual-rail encoding, then evolved through a weight matrix $W(\vec{\theta})$ with trainable parameters $\vec{\theta}$. The evolved states are measured via post-selection, yielding probabilities on the chosen basis. These results are fed back into the optimization loop to update $\vec{\theta}$. (**b**) Physical implementation of $U(x)W(\vec{\theta})$ through photonic waveguide circuits. (**c**) environment realized by single-photon detectors used to acquire output probabilities from the photonic circuits. Yellow circles represent detected photons: a single yellow circle indicates one photon signal in the corresponding channel, two circles indicate two detection events, while black indicates no photon detection.

gate or input-output process using n photons and a $2n \times 2n$ linear optical network circuit, as shown in Fig. 1(b). Input photons are encoded in dual-rail format, where a pair of adjacent waveguide path modes $(\hat{a}_{2i-1}, \hat{a}_{2i})$ constitutes a qubit. The linear optical circuit is described by $\boldsymbol{W} = (w_{i,j})$, an arbitrary complex-valued weight matrix enabling the implementation of post-selected entanglement. The integrated circuit generates the Heisenberg evolution of mode operators

$$\hat{a}_i \to \sum_j w_{i,j}\hat{a}'_j \,. \tag{3}$$

The input state can be any individual basis state or a weighted combination of the n-qubit computational basis. In contrast, the output can be any of the C^n_{3n-1} possible outcomes from the probabilistic event where n photons are distributed across $2n$ waveguides. Typically, we only consider computational output states (Fig. 1(c)), where photons simultaneously appearing on both waveguides of the same qubit are considered invalid. Similarly, cases where multiple photons appear on the same waveguide are invalid (and are indistinguishable unless photon-number-resolving detectors are used). Given an input state, the amplitude of each output state can be computed by the permanent of $\boldsymbol{R}$ according to the indistinguishability of bosons, where $\boldsymbol{R} = (r_{i,j})$ is an $n \times n$ submatrix of $\boldsymbol{W}$.

The submatrix $\boldsymbol{R}$ is constructed as follows: Suppose the waveguide index of the input state is $\mathbf{X} = (x_1, x_2, \cdots, x_n)$, where $x_i \in \{2i-1, 2i\}$. The submatrix for the output state index $\mathbf{Y}^j = (y_1, y_2, \cdots, y_n)$ (where$j = 1, 2, \cdots, C^n_{3n-1}$) is

$$\boldsymbol{R}_{\mathbf{X},\mathbf{Y}^j} = \begin{bmatrix} w_{y_1,x_1} & w_{y_1,x_2} & \cdots & w_{y_1,x_n} \\ w_{y_2,x_1} & w_{y_2,x_2} & \cdots & w_{y_2,x_n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{y_n,x_1} & w_{y_n,x_2} & \cdots & w_{y_n,x_n} \end{bmatrix} . \tag{4}$$

The amplitude of this output state is the permanent of $\boldsymbol{R}$

$$\mathrm{perm}(\boldsymbol{R}) = \sum_{\sigma \in S_n} \prod_{i=1}^{n} r_{\sigma(i),i} \,. \tag{5}$$

By expanding the input Fock state under this operator transformation, the output amplitudes for indistinguishable bosons can be expressed by the permanent of the corresponding submatrix $\boldsymbol{R}$. The sum extends over all elements σ of the symmetric group S_n (all permutations of $1, 2, \cdots, n$). The probability of the output state $\mathbf{Y}^j$ over the entire set is

$$P_{\mathbf{X}}(\mathbf{Y}^j) = \frac{\| \mathrm{perm}(\boldsymbol{R}_{\mathbf{X},\mathbf{Y}^j}) \|^2}{\sum_{j=1}^{C^n_{3n-1}} \| \mathrm{perm}(\boldsymbol{R}_{\mathbf{X},\mathbf{Y}^j}) \|^2} . \tag{6}$$

Because $\boldsymbol{W}$ can be non-unitary under loss or post-selection, we introduce spectral normalization to preserve bounded probabilities. If $\boldsymbol{W}$ is unitary, the denominator sums to 1. For a superposed or entangled initial input state composed of multiple computational input states

$\mathbf{X} = \sum_i \alpha_i \mathbf{X}^i$ (where $\sum_i \alpha_i^2 = 1$), the amplitude for output state Y^j is $\sum_i \alpha_i \,\mathrm{perm}(R_{\mathbf{X}^i,\mathbf{Y}^j})$, and the probability is

$$P_{\mathbf{X}}(\mathbf{Y}^j) = \frac{\| \sum_i \alpha_i \,\mathrm{perm}(\boldsymbol{R}_{\mathbf{X}^i,\mathbf{Y}^j})\|^2}{\sum_{j=1}^{C_{3n-1}^n} \| \sum_i \alpha_i \,\mathrm{perm}(\boldsymbol{R}_{\mathbf{X}^i,\mathbf{Y}^j})\|^2}. \tag{7}$$

If $\boldsymbol{W}$ is a random complex matrix ($\|W\| < 1$), the sum of probabilities will be less than 1, representing photon loss. Using singular value decomposition (SVD), an arbitrary complex-valued matrix $\boldsymbol{W}$ can be decomposed as $\boldsymbol{W} = \boldsymbol{R}_1 \Sigma \boldsymbol{R}_2^\dagger$, where $\boldsymbol{R}_1$ and $\boldsymbol{R}_2^\dagger$ are unitary matrices and $\boldsymbol{\Sigma} = \mathrm{diag}(\lambda_1, \cdots, \lambda_{2n})$ contains singular values $\lambda_1 \geq \ldots \geq \lambda_{2n}$. When $\lambda_1 \leq 1$, $\boldsymbol{\Sigma}$ can be realized by applying photon loss to each mode. We normalize $\boldsymbol{W}$ by dividing it by its spectral norm $\|\boldsymbol{W}\| = \lambda_1$, denoted as $\tilde{\boldsymbol{W}} = \boldsymbol{W}/\|\boldsymbol{W}\|$. The success probability of the post-selection-based transformation $\boldsymbol{W}$ is $1/\|\boldsymbol{W}\|^{2n}$[35]. This means that the success probability will decrease as the number of photons (quantum bits) n increases, and this decay is $1/\|\boldsymbol{W}\|^{2n}$.

Using this framework, we map the circuit design problem to an optimization problem. Given a desired gate or input-output relation and a trainable matrix $\boldsymbol{W}$, we can derive the output probability distribution for all input states. By training $\boldsymbol{W}$, we steer the output probability distributions (or the loss function derived from them) toward the target. Training can be performed offline using the mathematical model or online directly on the photonic circuits. Given the limited efficiency of multi-photon coincidence counting as photon numbers scale, an efficient strategy is to pre-train on the mathematical model to obtain a coarse solution for $\boldsymbol{W}$, followed by fine-tuning on the actual chip.

Results

Complete Bell projector

The Bell projector is an indispensable building block for teleporting qubits, swapping entanglement, extracting error syndromes, and driving measurement-based quantum computation. It is well established that conventional linear optical circuits without ancillas or squeezing cannot achieve complete Bell state discrimination: at most, only two of the four states can be unambiguously distinguished in a single-shot measurement. This limitation arises from the bosonic symmetry of photons and the linear nature of beam splitter transformations.Here, we leverage the proposed variational ansatz to design a Bell projector capable of distinguishing all four Bell states in a single-shot detection. The desired correspondence between the input state and output states (coincidence bases) is

$$\begin{cases} |\phi^+\rangle \to |00\rangle\,(a_1' a_3'), \, |\phi^-\rangle \to |01\rangle\,(a_1' a_4'), \\ |\psi^+\rangle \to |10\rangle\,(a_2' a_3'), \, |\psi^-\rangle \to |11\rangle\,(a_2' a_4'). \end{cases} \tag{8}$$

Consequently, the desired unitary operator in the Hilbert space is

$$\begin{aligned} \boldsymbol{U} &= |00\rangle\langle\phi^+| + |01\rangle\langle\phi^-| + |10\rangle\langle\psi^+| + |11\rangle\langle\psi^-| \\ &= \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & -1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & -1 & 0 \end{bmatrix}. \end{aligned} \tag{9}$$

The objective is to find a linear transformation $\boldsymbol{W}$ (on waveguides) to realize $\boldsymbol{U}$ via post-selected entanglement. This is achieved by solving the 16 equations described by

$$\mathrm{perm}\left(\begin{bmatrix} w_{s_i(1),s_j(1)} & w_{s_i(1),s_j(2)} \\ w_{s_i(2),s_j(1)} & w_{s_i(2),s_j(2)} \end{bmatrix} \right) = U_{i,j}, \, (i,j \in \{1,2,3,4\}), \tag{10}$$

where $s_1 = (1,3)$, $s_2 = (1,4)$, $s_3 = (2,3)$, and $s_4 = (2,4)$. The set of equations can be unfolded as

$$\begin{aligned} perm\left(\begin{bmatrix} w_{1,1} & w_{1,3} \\ w_{3,1} & w_{3,3} \end{bmatrix} \right) &= U_{1,1} \\ perm\left(\begin{bmatrix} w_{1,1} & w_{1,4} \\ w_{3,1} & w_{3,4} \end{bmatrix} \right) &= U_{1,2} \\ &\vdots \\ perm\left(\begin{bmatrix} w_{2,2} & w_{2,3} \\ w_{4,2} & w_{4,3} \end{bmatrix} \right) &= U_{4,3} \\ perm\left(\begin{bmatrix} w_{2,2} & w_{2,4} \\ w_{4,2} & w_{4,4} \end{bmatrix} \right) &= U_{4,4} \end{aligned}. \tag{11}$$

The task can be interpreted as finding an optimal solution for $\boldsymbol{W}$ that minimizes the l_2 loss between the terms in Eq. 11. Additionally, we optimize the success probability by incorporating an objective to maximize $1/\|\boldsymbol{W}\|^4$. During training, we record $f_k (k = 1, \ldots, 4)$ to measure the likelihood of obtaining the desired output state for each Bell basis

$$\begin{cases} f_1 = \dfrac{P_{\phi^+}[s_1]}{\sum_k P_{\phi^+}[s_k]}, & f_2 = \dfrac{P_{\phi^-}[s_2]}{\sum_k P_{\phi^-}[s_k]}, \\ f_3 = \dfrac{P_{\psi^+}[s_3]}{\sum_k P_{\psi^+}[s_k]}, & f_4 = \dfrac{P_{\psi^-}[s_4]}{\sum_k P_{\psi^-}[s_k]}. \end{cases} \tag{12}$$

The training dynamics, specifically the l_2 loss and the recorded f_k, are shown in Fig. 2(a,b) . With the trained $\boldsymbol{W}$, the four Bell states exhibit nearly uniform conditional success probabilities $f_1 \approx f_2 \approx f_3 \approx f_4 \approx 1/4$ within the valid subspace, corresponding to the balanced $\sim 25\%$ distribution shown in Fig. 2(d). The resulting matrix $\boldsymbol{W}_2$ is

$$\boldsymbol{W}_2 = \begin{bmatrix} -1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ \dfrac{1}{\sqrt{2}} & \dfrac{1}{\sqrt{2}} & -\dfrac{1}{\sqrt{2}} & \dfrac{1}{\sqrt{2}} \\ \dfrac{1}{\sqrt{2}} & -\dfrac{1}{\sqrt{2}} & -\dfrac{1}{\sqrt{2}} & \dfrac{1}{\sqrt{2}} \end{bmatrix}. \tag{13}$$

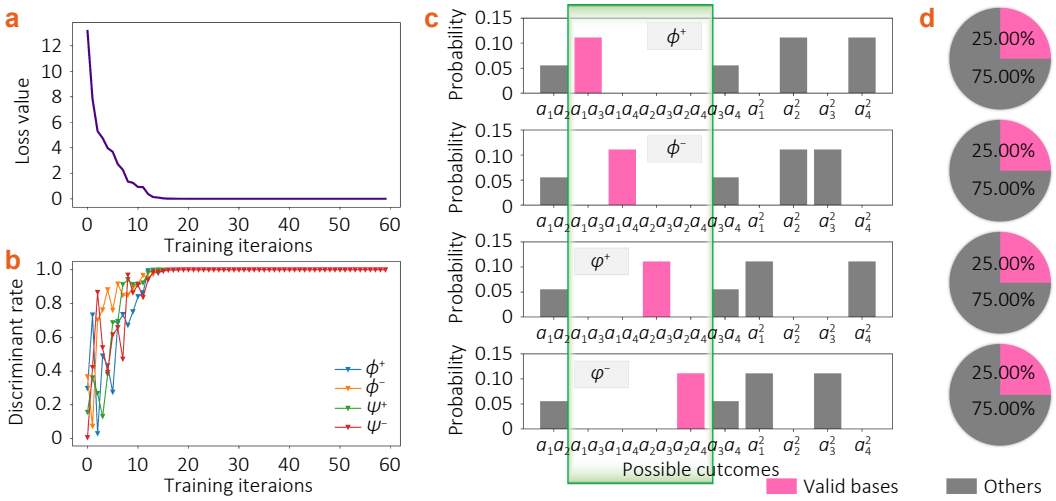

Fig. 2 | Results of the complete Bell projector. (**a**) The l_2 loss of the U-gate implementation. (**b**) The evolution of Bell state distinguishability with training iterations. (**c**) The probability of all 10 possible outcomes for 2 photons on 4 waveguides. (**d**) The ratio of valid dual-rail bases over the entire set.

Since $\boldsymbol{W}$ is non-unitary, the success probability of this complete Bell projector is 1/9. Quantitative analysis confirms that the trained circuit converges to a success probability of $P_{\text{succ}} = 1/\|\boldsymbol{W}\|^{2n}$, aligning exactly with the theoretical value of 1/9. Furthermore, the conditional discrimination precision, evaluated over successful post-selected trials, reaches nearly 100% (Fig. 2(c,d)), validating the high fidelity of the learned operation.

Quantum support vector machine

The quantum support vector machine (qSVM) addresses complex classification problems by embedding classical data into a quantum-enhanced feature space where nonlinear boundaries can be distinguished. Unlike classical SVMs that rely on explicit feature mapping, the qSVM naturally constructs the quantum kernel function through the inner product of quantum states.

The qSVM is particularly suited for handling one-dimensional datasets that are non-overlapping but not linearly separable (Fig. 3(a)). Its workflow is summarized in Fig. 3(b). Classical input features are encoded into quantum states using R_X and R_Y rotation gates. A parameterized

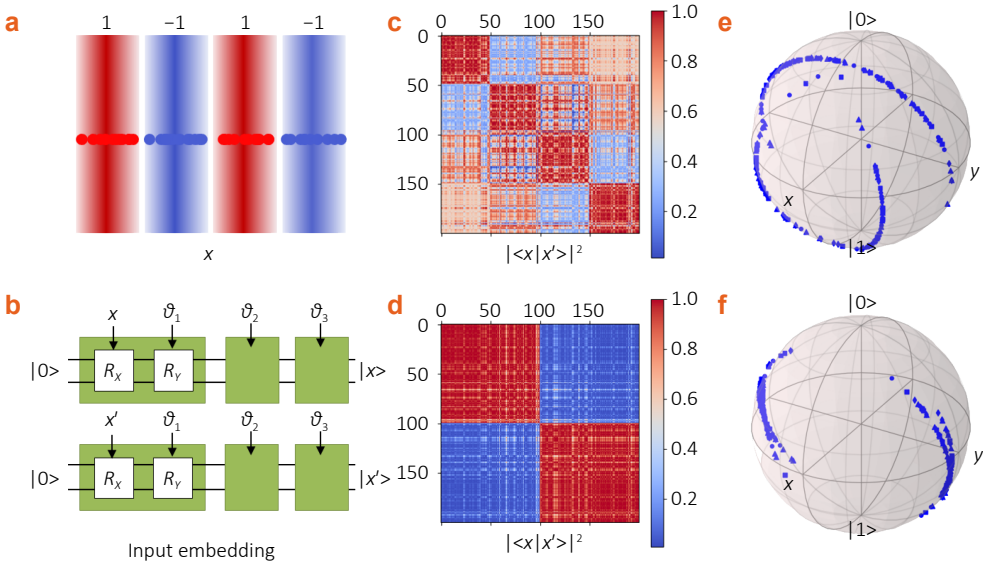

Fig. 3 | Principles and results of the quantum support vector machine (qSVM). (**a**) One-dimensional dataset with non-overlapping but not linearly separable data points. (**b**) Workflow of the qSVM: classical inputs are encoded into quantum states using R_X and R_Y gates, followed by a parameterized embedding module $U(x)W(\vec{\theta})$ and additional unitary layers. The system is then measured in the computational basis with post-selection. (**c**, **d**) Kernel matrix heatmaps before and after training. Optimization significantly improves data distinguishability; red areas indicate higher similarity, while blue areas represent lower similarity. (**e**, **f**) Bloch-sphere mappings: before training, quantum state trajectories of different classes overlap; after training, the trajectories are clearly separated, demonstrating that embedding reshapes the data geometry.

embedding module, $\boldsymbol{U}(x)\boldsymbol{W}(\vec{\theta})$, expands the feature space, and additional unitary layers evolve different inputs along distinct trajectories.

The system is measured in the computational basis, yielding probabilities over bases $\{00, 01, 10, 11\}$. Each data point i is represented by a four-dimensional probability vector $\vec{x}_i$. The similarity between the i-th and j-th inputs is quantified by the inner product of their corresponding vectors

$$k(\vec{x}_i, \vec{x}_j) = \left|\langle x_i | x_j \rangle\right|^2, \tag{14}$$

producing a kernel value within $[0, 1]$. By evaluating all pairwise inner products, the kernel matrix $\boldsymbol{K}$ is constructed. The kernel matrix can be visualized as a heatmap. Before training, the heatmap shows that data points from different classes are largely indistinguishable (Fig. 3(c)). To incorporate supervision, we define a cost function based on kernel-target alignment (KTA). The label matrix is $\boldsymbol{T}_{ij} = y_i y_j$ with $y \in \{-1, +1\}$. The alignment is expressed as the Frobenius inner product between $\boldsymbol{K}$ and $\boldsymbol{T}$, normalized by their norms. The cost function is

$$\mathcal{L} = -\frac{\langle \boldsymbol{K}, \boldsymbol{T} \rangle_F}{\|\boldsymbol{K}\|_F \|\boldsymbol{T}\|_F}. \tag{15}$$

Minimizing $\mathcal{L}$ encourages the quantum kernel to align with the label structure. After optimization, the kernel matrix heatmap reveals a clear block structure (Fig. 3(d)), demonstrating that the qSVM effectively reshapes the data geometry. This is further confirmed by Bloch-sphere mappings (Fig. 3(e,f), where trajectories corresponding to different classes become clearly separated.

The extension to a two-qubit circuit, illustrated in Fig. 4(a) incorporates a matrix to generate entanglement, enhancing the model's expressive power. This design maps data into a high-dimensional Hilbert space (Fig. 4(b,c)), where a linear decision surface projects back to a complex, non-linear boundary in the original feature space, forming the foundation for enhanced classification. This theoretical advantage is confirmed by the quantitative results in Table 1. Our photonic qSVM delivers strong accuracies of 0.69 and 0.88 on the single- and two-bit datasets, substantially exceeding the classical linear SVM (0.63, 0.83) and approaching the peak accuracy of the more complex PEFM model (0.72, 0.91). Crucially, our model's lower variance ($\pm$ 0.03 vs. PEFM's $\pm$ 0.05 and $\pm$ 0.03) signifies greater stability. A key strength of our shallow architecture is its inherent noise resilience, as quantified in Table 2. Under depolarizing noise, our qSVM sustained only a minimal

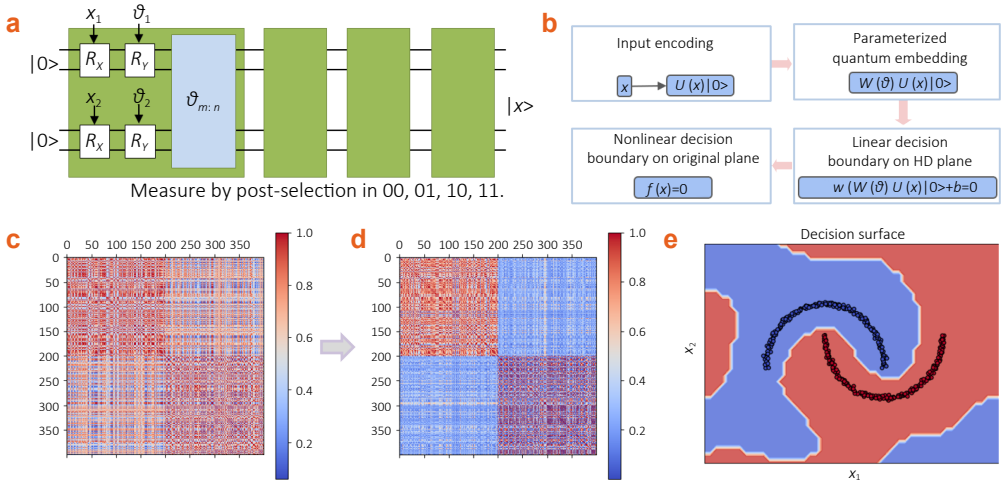

Fig. 4 | Results of the two-qubit quantum embedding and classification. (**a**) Two-qubit embedding circuit introducing a θ matrix to generate entanglement, with the parameterized variational block highlighted. (**b**) Principle of hyperplane learning: data mapped into a high-dimensional Hilbert space, where linear separation corresponds to a nonlinear boundary in the original space. (**c**) Decision function of the trained quantum model. Red and blue dots indicate samples from the two categories, with shaded regions showing the model's predicted class assignments and the white contour marking the decision boundary. The curved separation arises from a linear decision surface in the high-dimensional Hilbert space, which appears non-linear when projected onto this 2D plot. (**d**,**e**) Kernel matrices before and after training, showing that optimization enables clear class separation and improved classification performance.

Table 1 | Classification accuracy comparison of classical and quantum SVM variants. Results are averaged over 10 random splits (mean ± std).

Model	Single-bit (1D clusters)	Two-bit (moons)	Scalability	Noise sensitivity
Linear SVM	0.63 ± 0.04	0.83 ± 0.02	Low	Robust
Our-qSVM	0.69 ± 0.03	0.88 ± 0.03	Moderate--High	Low
Photonic encoding feature map(PEFM)	0.72 ± 0.05	0.91 ± 0.03	Limited (kernel concentration)	High
Parametrized quantum kernel(PQK)	0.67 ± 0.04	0.85 ± 0.04	High	Moderate

Table 2 | Comparative analysis of classification accuracy (mean ± standard deviation) for quantum and classical models with and without noise.

Model	Normal accuracy	Noisy accuracy
Linear SVM	0.83 ± 0.02	0.83 ± 0.02
Our-qSVM (2q)	0.88 ± 0.03	0.83 ± 0.03
PEFM	0.91 ± 0.03	0.76 ± 0.05
PQK	0.85 ± 0.04	0.80 ± 0.04

performance loss (0.88 to 0.83), while the fidelity-based PEFM suffered severe degradation (0.91 to 0.76). This robustness is a direct consequence of the low-depth, weakly entangled strategy embodied by our circuit design. The training dynamics that yield these results are visually attested by the kernel matrices in Fig. 4(d,e). The evolution from a noisy to a clean, block-diagonal structure demonstrates the successful optimization of the quantum feature space to maximize class separation. Collectively, the visual evidence and robust metrics establish our variational photonic qSVM as a solution offering a superior balance of accuracy, scalability, and noise robustness for NISQ applications.

Quantum K-means

Quantum K-means (QKMeans) performs unsupervised clustering by leveraging quantum-enhanced distance estimation between data points and cluster centroids. The

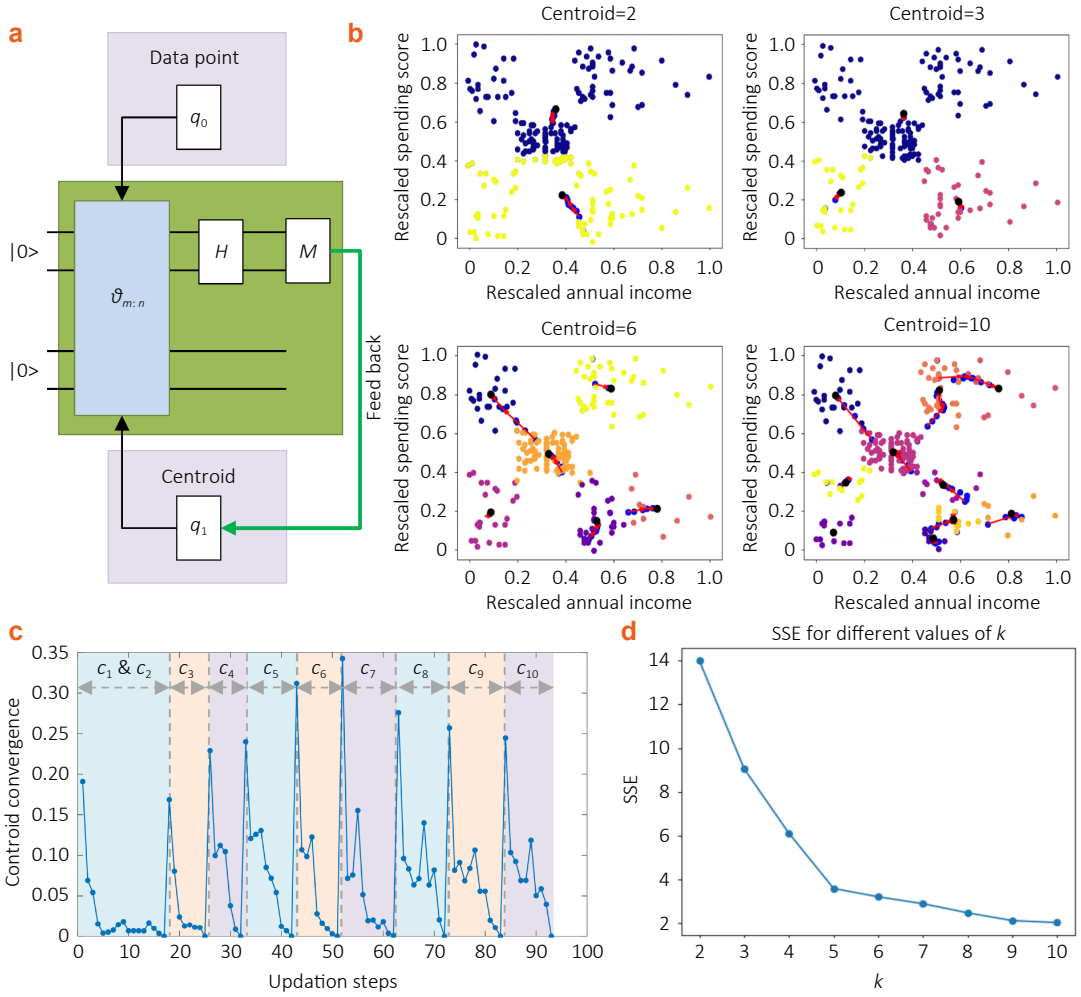

Fig. 5 | Quantum K-means (QKMeans) results. (**a**) Overall framework of QKMeans, where data points and centroids are dual-encoded into quantum states, processed by a variational circuit, and measured to estimate quantum distances. (**b**) Clustering outcomes with different centroid numbers, showing centroid migration and convergence behavior. (**c**) Cluster assignments for $K = 6 \sim 10$, demonstrating alignment with the underlying data distributions. Blue dots represent old centroids, black dots represent final centroids, and red arrows indicate migration. (**d**) Sum of squared errors (SSE) versus K, exhibiting the elbow pattern that indicates the optimal number of clusters.

framework is depicted in Fig. 5(a). Each classical data point and centroid is dual-encoded into quantum states (q_0 and q_1) and processed by a variational circuit parameterized by $\theta_{m:n}$. This block provides non-linear feature transformations. A Hadamard gate (H) and measurement operation (M) extract the overlap between encoded states, estimating the quantum distance

$$d(x_i, c_j) \propto 1 - \left|\langle q_0|q_1\rangle\right|^2 \tag{16}$$

Where q_0 and q_1 denote the quantum states encoding the i-th data point x_i and the j-th centroid c_j, respecti vely. The measurement result determines the cluster assignment, and the centroid register is updated iteratively until convergence.

Clustering results for different centroid numbers are presented in Fig. 5(b). For each K, the algorithm converges rapidly. When $K = 2$, clustering is unsatisfactory, as distant points are grouped. As K increases ($K = 6 \sim 10$), assignments align well with data distributions. The convergence curves (Fig. 5(c)) show rapid stabilization. The sum of squared errors (SSE) exhibits an elbow pattern at $K = 5$ (Fig. 5(d)), indicating the optimal number of clusters. These findings verify that quantum-enhanced distance estimation enables robust iterative clustering.

Conclusion

This work proposes a framework for quantum embedding learning on variational photonic circuits, integrating dual-rail encoding with a trainable, post-selection-enabled linear network $\boldsymbol{W}(\vec{\theta})$. By reformulating circuit synthesis as an optimization over permanent-derived probabilities, the method enables efficient offline and online training compatible with NISQ photonic platforms. Theoretically, this framework unifies gate synthesis and embedding within a single optimization paradigm, introduces a tunable embedding $\boldsymbol{U}(x)\boldsymbol{W}(\vec{\theta})$ that reshapes data geometry in the Hilbert space, and develops a quantum distance estimator based on state overlaps that links quantum similarity to iterative clustering dynamics.

Practically, we demonstrated three representative tasks with quantitative validation. First, we implemented a non-unitary complete Bell projector. The experimental results align precisely with theoretical predictions regarding success probability and achieve near-perfect conditional discrimination accuracy, confirming the high fidelity of our operation. Second, in the quantum SVM task, our shallow photonic architecture proved to be more robust against experimental noise compared to existing methods such as PEFM and PQK. Statistical analysis highlights that our approach achieves a superior balance between high mean accuracy and low variance, ensuring stable classification performance. Third, we demonstrated a quantum K-means algorithm where rapid convergence and clear clustering structure (indicated by the SSE curve) validate the effectiveness of our quantum-enhanced distance estimation.

While our current results are based on numerical simulations, the proposed framework is amenable to integrated photonic implementation. A feasible setup could employ a silicon or silicon nitride chip[40,41] incorporating on-chip spontaneous four-wave mixing sources for photon-pair generation[40,42], a programmable Mach–Zehnder interferometer mesh with thermo-optic phase shifters to realize the variational transformation $\boldsymbol{W}(\vec{\theta})$[43–45], and superconducting nanowire single-photon detectors for post-selective multi-photon coincidence counting[46,47]. An FPGA-based feedback system could further support automated parameter tuning and real-time optimization, building upon existing demonstrations of large-scale reconfigurable meshes[35,48].

In such physical realizations, performance may be affected by noise sources such as depolarizing noise, dephasing, photon loss, and measurement noise. Additional experiments on the two-qubit moons dataset confirm that under depolarizing noise, our shallow photonic qSVM retains robust accuracy (0.83 ± 0.03 vs. 0.88 ± 0.03), outperforming both PEFM (0.76 ± 0.05) and PQK (0.80 ± 0.04), which suffer greater degradation due to their reliance on global coherence or measurement statistics. These results validate the superior noise resilience of our low-depth, weakly entangled architecture in the NISQ regime.

Furthermore, the variational learning framework presented here is not limited to discrete-variable systems but can be naturally extended to the continuous-variable (CV) domain. In this setting, programmable quantum circuits are realized through continuous-variable quantum neural networks (CV-QNNs)[49,50], where Gaussian operations —specifically displacement, squeezing, and beamsplitter transformations—act as learnable affine transformations on the phase space. Crucially, to achieve universal representational power, these Gaussian layers are interleaved with non-Gaussian operations, such as the Kerr effect or cubic phase gates, acting as the continuous analog to nonlinear activation functions. This extension is particularly promising for CV quantum communication, where variational schemes can optimize modulation parameters and compensate for channel noise in CV-QKD protocols[51]. Moreover, in the context of fault-tolerant quantum computing, such variational approaches offer a viable pathway for synthesizing non-Gaussian resource states, including Gottesman–Kitaev–Preskill (GKP) states, which are essential for correcting errors in bosonic modes[52].

Current limitations stem from the probabilistic nature of post-selection and the scaling of multi-photon circuits. Future progress will rely on brighter photon sources, low-loss interferometers, number-resolving detectors, active feedforward, and noise-aware training, extending this optimization-based design to boson sampling, non-Gaussian state engineering (e.g., GKP states), and hybrid quantum-classical models toward practical quantum advantage[53].

References

1. Schuld M, Killoran N. Quantum machine learning in feature Hilbert spaces. *Phys Rev Lett* **122**, 040504 (2019).
2. Havlíček V, Córcoles AD, Temme K et al. Supervised learning with quantum-enhanced feature spaces. *Nature* **567**, 209–212 (2019).
3. Zhang H, Gu M, Jiang XD et al. An optical neural chip for implementing complex-valued neural network. *Nat Commun* **12**, 457 (2021).
4. Farhi E, Neven H. Classification with quantum neural networks on near term processors. *arXiv preprint* **1802**, 06002 (2018).
5. Schuld M, Bocharov A, Svore KM et al. Circuit-centric quantum classifiers. *Phys Rev A* **101**, 032308 (2020).
6. Benedetti M, Lloyd E, Sack S et al. Erratum: Parameterized quantum circuits as machine learning models (2019 *Quant. Sci. Tech.* 4 043001). *Quantum Sci Technol* **5**, 019601 (2020).
7. Mari A, Bromley TR, Izaac J et al. Transfer learning in hybrid classical-quantum neural networks. *Quantum* **4**, 340 (2020).
8. Benedetti M, Garcia-Pintos D, Perdomo O et al. A generative modeling approach for benchmarking and training shallow quantum circuits. *npj Quantum Inf* **5**, 45 (2019).
9. Harrow AW, Montanaro A. Quantum computational supremacy. *Nature* **549**, 203–209 (2017).
10. Schuld M, Sweke R, Meyer JJ. Effect of data encoding on the expressive power of variational quantum-machine-learning models. *Phys Rev A* **103**, 032430 (2021).
11. Caro MC, Gil-Fuster E, Meyer JJ et al. Encoding-dependent generalization bounds for parametrized quantum circuits. *Quantum* **5**, 582 (2021).
12. Thanasilp S, Wang S, Cerezo M et al. Exponential concentration in quantum kernel methods. *Nat Commun* **15**, 5200 (2024).
13. Lloyd S, Schuld M, Ijaz A et al. Quantum embeddings for machine learning (2020). arXiv: 2001.03622.
14. Thumwanit N, Lortaraprasert C, Yano H et al. Trainable discrete feature embeddings for quantum machine learning. In *2021 IEEE International Conference on Quantum Computing and Engineering (QCE)* 479–480 (IEEE, 2021). http://doi.org/10.1109/QCE52317.2021.00087.
15. Doriguello JF, Montanaro A. Quantum random access codes for Boolean functions. *Quantum* **5**, 402 (2021).
16. Hubregtsen T, Wierichs D, Gil-Fuster E et al. Training quantum embedding kernels on near-term quantum computers. *Phys Rev A* **106**, 042431 (2022).
17. Incudini M, Bosco DL, Martini F et al. Automatic and effectivediscovery of quantum kernels. *IEEE Trans Emerg Top Comput Intell* **106**, 1–10 (2024).
18. Sahin ME, Symons BCB, Pati P et al. Efficient parameter optimisation for quantum kernel alignment: a sub-sampling approach in variational training. *Quantum* **8**, 1502 (2024).
19. Schuld M. Supervised quantum machine learning models are kernel methods. *arXiv preprint* **2101**, 11020 (2021).
20. Zhang H, Wan LX, Haug T et al. Resource-efficient high-dimensional subspace teleportation with a quantum autoencoder. *Sci Adv* **8**, eabn9783 (2022).
21. Lim AEJ, Song JF, Fang Q et al. Review of silicon photonics foundry efforts. *IEEE J Sel Top Quantum Electron* **20**, 8300112 (2014).
22. Stern B, Ji XC, Okawachi Y et al. Battery-operated integrated frequency comb generator. *Nature* **562**, 401–405 (2018).
23. Guha B, Cardenas J, Lipson M. Athermal silicon microring resonators with titanium oxide cladding. *Opt Express* **21**, 26557 (2013).
24. Schwartz M, Schmidt E, Rengstl U et al. Fully on-chip single-photon Hanbury-Brown and Twiss experiment on a monolithic semiconductor–superconductor platform. *Nano Lett* **18**, 6892–6897 (2018).
25. Honjo T, Inoue K, Takahashi H. Differential-phase-shift quantum key distribution experiment with a planar light-wave circuit Mach–Zehnder interferometer. *Opt Lett* **29**, 2797 (2004).
26. Sibson P, Erven C, Godfrey M et al. Chip-based quantum key distribution. *Nat Commun* **8**, 13984 (2017).
27. Raussendorf R, Briegel HJ. A one-way quantum computer. *Phys Rev Lett* **86**, 5188–5191 (2001).
28. Nielsen MA. Optical quantum computation using cluster states. *Phys Rev Lett* **93**, 040503 (2004).
29. Zhang H, Wan LX, Paesani S et al. Encoding error correction in an integrated photonic chip. *PRX Quantum* **4**, 030340 (2023).
30. Aaronson S, Arkhipov A. The computational complexity of linear optics. In *Proceedings of the Forty-Third Annual ACM Symposium on Theory of Computing* 333–342 (ACM, 2011). http://doi.org/10.1145/1993636.1993682.
31. Crespi A, Osellame R, Ramponi R et al. Integrated multimode interferometers with arbitrary designs for photonic boson sampling. *Nat Photonics* **7**, 545–549 (2013).
32. Tillmann M, Dakić B, Heilmann R et al. Experimental boson sampling. *Nat Photonics* **7**, 540–544 (2013).
33. Aspuru-Guzik A, Walther P. Photonic quantum simulators. *Nat Phys* **8**, 285–291 (2012).
34. Paesani S, Gentile AA, Santagati R et al. Experimental Bayesian quantum phase estimation on a silicon photonic chip. *Phys Rev Lett* **118**, 100503 (2017).
35. Zhang H, Yang CR, Mok WK et al. Variational learning of integrated quantum photonic circuits via genetic algorithm. *Laser Photonics Rev* **19**, 2400359 (2025).
36. Ralph TC, White AG, Munro WJ et al. Simple scheme for efficient linear optics quantum gates. *Phys Rev A* **65**, 012314 (2001).
37. O'Brien JL, Pryde GJ, White AG et al. Demonstration of an all-optical quantum controlled-NOT gate. *Nature* **426**, 264–267 (2003).
38. Bharti K, Cervera-Lierta A, Kyaw TH et al. Noisy intermediate-scale quantum algorithms. *Rev Mod Phys* **94**, 015004 (2022).
39. Knill E, Laflamme R, Milburn GJ. A scheme for efficient quantum computation with linear optics. *Nature* **409**, 46–52 (2001).
40. Silverstone JW, Bonneau D, Ohira K et al. On-chip quantum interference between silicon photon-pair sources. *Nat Photonics* **8**, 104–108 (2014).
41. Pelucchi E, Fagas G, Aharonovich I et al. The potential and global outlook of integrated photonics for quantum technologies. *Nat Rev Phys* **4**, 194–208 (2022).
42. Paesani S, Borghi M, Signorini S et al. Near-ideal spontaneous photon sources in silicon quantum photonics. *Nat Commun* **11**, 2505 (2020).
43. Carolan J, Harrold C, Sparrow C et al. Universal linear optics. *Science* **349**, 711–716 (2015).
44. Harris NC, Steinbrecher GR, Prabhu M et al. Quantum transport simulations in a programmable nanophotonic processor. *Nat Photonics* **11**, 447–452 (2017).
45. Taballione C, Wolterink TAW, Lugani J et al. 8×8 reconfigurable quantum photonic processor based on silicon nitride waveguides. *Opt Express* **27**, 26842 (2019).
46. Najafi F, Mower J, Harris NC et al. On-chip detection of non-classical light by scalable integration of single-photon detectors. *Nat Commun* **6**, 5873 (2015).
47. Schuck C, Pernice WHP, Tang HX. NbTiN superconducting nanowire detectors for visible and telecom wavelengths single photon counting on Si3N4 photonic circuits. *Appl Phys Lett* **102**, 051101 (2013).
48. Zhang H, Thompson J, Gu ML et al. Efficient on-chip training of optical neural networks using genetic algorithm. *ACS Photonics* **8**, 1662–1672 (2021).
49. Killoran N, Bromley TR, Arrazola JM et al. Continuous-variable quantum neural networks. *Phys Rev Res* **1**, 033063 (2019).

50. Chang S Y, Vallecorsa S, Combarro E F, et al. Quantum generative adversarial networks in a continuous-variable architecture to simulate high energy physics detectors. *arXiv preprint* **2101**, 11132 (2021).
51. Liao Q, Xiao G, Zhong H et al. Multi-label learning for improving discretely-modulated continuous-variable quantum key distribution. *New J Phys* **22**, 083086 (2020).
52. Tzitrin I, Bourassa JE, Menicucci NC et al. Progress towards practical qubit computation using approximate Gottesman-Kitaev-Preskill codes. *Phys Rev A* **101**, 032315 (2020).
53. Zhang H, Wan LX, Ramos-Calderer S et al. Efficient option pricing with a unary-based photonic computing chip and generative adversarial learning. *Photonics Res* **11**, 1703 (2023).

Acknowledgements

Supported by National Natural Science Foundation of China (62505228, 62405173), Shanghai Pujiang Program (23PJ1413700), Chenguang Program of Shanghai Education Development Foundation and Shanghai Municipal Education Commission (24CGA19), and Fundamental Research Funds for the Central Universities.

Author contributions

Ren GH was responsible for code writing, experimental research, and the writing of the original paper. Wang DL was in charge of data organization, experimental research, and verification work. Yan LZ assisted in the drawing of pictures. Zhan YC also participated in the development of the algorithm. Zhang H proposed the research idea, obtained research funds, formulated the research methods and participated in the review and guidance of the manuscript. Ma YM, Wang ZS, and Chen XB participated in the research through providing financial support. All authors participated in the review and revision of the manuscript and approved the final version of the manuscript.

Competing interests

The authors declare no competing financial interests.