RESEARCH   **Open Access**

Fock state-enhanced expressivity of quantum machine learning models

Beng Yee Gan[1*], Daniel Leykam[1] and Dimitris G. Angelakis[1,2,3*]

*Correspondence:
gan.bengyee@u.nus.edu;
dimitris.angelakis@gmail.com
[1] Centre for Quantum Technologies,
National University of Singapore, 3
Science Drive 2, Singapore 117543,
Singapore
[2] School of Electrical and Computer
Engineering, Technical University of
Crete, Chania, 73100, Greece
Full list of author information is
available at the end of the article

Abstract

The data-embedding process is one of the bottlenecks of quantum machine learning, potentially negating any quantum speedups. In light of this, more effective data-encoding strategies are necessary. We propose a photonic-based bosonic data-encoding scheme that embeds classical data points using fewer encoding layers and circumventing the need for nonlinear optical components by mapping the data points into the high-dimensional Fock space. The expressive power of the circuit can be controlled via the number of input photons. Our work sheds some light on the unique advantages offered by quantum photonics on the expressive power of quantum machine learning models. By leveraging the photon-number dependent expressive power, we propose three different noisy intermediate-scale quantum-compatible binary classification methods with different scaling of required resources suitable for different supervised classification tasks.

Keywords: Quantum physics; Machine learning; Quantum photonics

1 Introduction

Machine learning approaches such as artificial neural networks are powerful tools for solving a wide range of problems including image classification and regression. However, the scalability of machine learning implemented using general-purpose electronic circuits is limited by their high power consumption and the end of Moore's law. These issues motivate the pursuit of dedicated hardware for machine learning including photonic neural networks [1–4] and quantum circuits [5–11].

The combination of ideas from the photonic and quantum machine learning communities may enable further speed-ups and novel functionalities [12–19]. For example, both classical and quantum photonic neural networks are presently limited by the difficulty of incorporating nonlinear activation functions. This challenge can be circumvented using the kernel trick, in which the input data is mapped into a high-dimensional feature space where simple linear models become effective [13, 20–22]. The simplest quantum feature map based on repeated application of data-dependent single qubit rotations is already sufficient to serve as a universal function approximator [23–25].

Despite progress in various aspects of near-term quantum machine learning algorithms [26] including experimental realizations [16, 27–30], proposals for various platforms
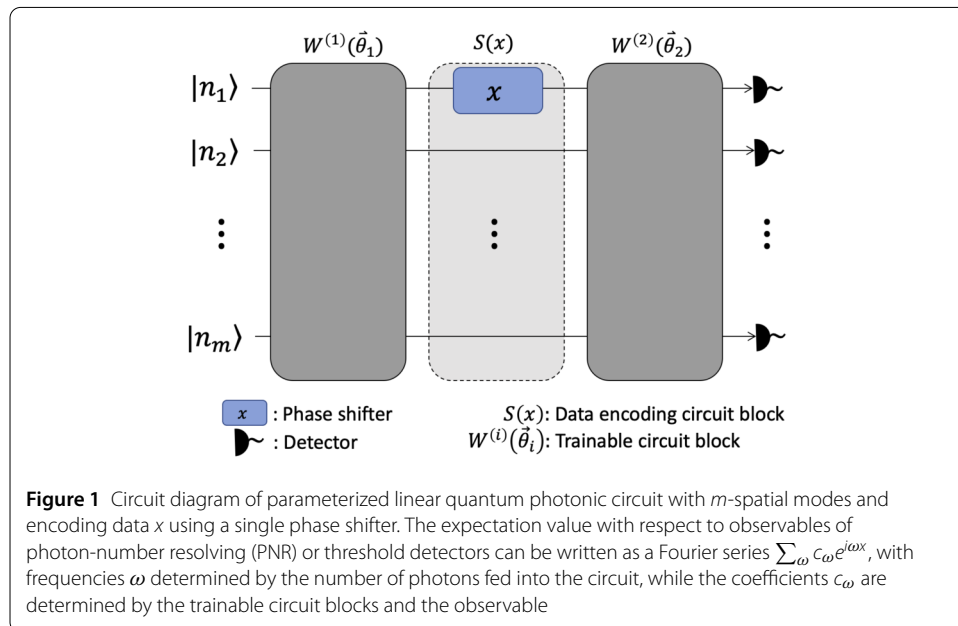
🌱 Springer

[28, 31] and studies of statistical properties of quantum machine learning models [32–34], the encoding of input data is still a significant bottleneck for (quantum) photonic machine learning hardware. For example, the expressive power of quantum circuits based on parameterized single qubit rotations is limited by the number of encoding gates used [23, 24]. Similarly, some existing quantum machine learning algorithms with proven speedups for future fault-tolerant quantum computers assume the existence of quantum-random access memory (RAM) [35] that can provide the input data in a quantum superposition with no overhead [36–42]. Yet, the sources of the speedup of these algorithms are still under active debate [43, 44]. Thus, a pressing goal is to develop machine learning algorithms that avoid encoding large input datasets [45–47] or more efficient data-encoding methods. This article addresses the latter problem.

Specifically, we generalize the qubit-based circuit architecture analyzed in Refs. [23, 24] to quantum photonic circuits (QPCs) constructed using linear optical components such as beam splitters and phase shifters, photon detectors, and Fock state inputs. We consider parameterized linear QPCs [Fig. 1(a)] consisting of two trainable circuit blocks with one data encoding block sandwiched between them. We show that for a fixed number of encoding phase shifters, the expressive power of the parameterized quantum circuit is improved by embedding the classical data into the higher-dimensional Fock space. This enables the approximation of classical functions using fewer encoding layers while circumventing the need for nonlinear components. The origin of this improved encoding efficiency is that each phase shifter simultaneously uploads the input data onto multiple Fock basis states simultaneously.

Similar to Ref. [24], n-photon quantum machine learning models can be expressed as a Fourier series

$$f^{(n)}(x, \mathbf{\Theta}, \boldsymbol{\lambda}) = \sum_{\omega \in \Omega_n} c_\omega(\mathbf{\Theta}, \boldsymbol{\lambda}) e^{i\omega x}, \tag{1}$$

Figure 1 Circuit diagram of parameterized linear quantum photonic circuit with m-spatial modes and encoding data x using a single phase shifter. The expectation value with respect to observables of photon-number resolving (PNR) or threshold detectors can be written as a Fourier series $\sum_\omega c_\omega e^{i\omega x}$, with frequencies ω determined by the number of photons fed into the circuit, while the coefficients c_ω are determined by the trainable circuit blocks and the observable

where $\Omega_n \in \mathbb{N}$ is the frequency spectrum and $\{c_\omega\}$ are the Fourier coefficients that depend on trainable circuit block's parameters $\boldsymbol{\Theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ and observable's parameters $\boldsymbol{\lambda}$. The expressive power of the Fourier series is determined by two components: the spectrum of frequencies ω, and the Fourier coefficients c_ω. We show that the frequency spectrum of the circuit can be controlled by the number of input photons. Thus, a rich frequency spectrum can be generated by providing sufficient number of input photons to linear QPCs with a constant number of spatial modes. In contrast, qubit-based circuits require deeper or wider circuits to increase the size of their frequency spectrum. When generalized to arbitrary input states and observables the QPCs can also generate arbitrary set of Fourier coefficients that combine the frequency dependent basis functions e^{iwx}, allowing them to approximate any square-integrable function on a finite interval to arbitrary precision [24, 48, 49].

As an application of the parameterized linear quantum photonic circuits, we consider three different machine learning approaches for supervised data classification: (1) A variational classifier based on minimizing a cost function by training the circuit parameters. (2) Kernel methods, which employ fixed circuits, with training carried out on observables only. (3) Random kitchen sinks, which use a set of random circuits to approximate a desired kernel function. Each of these methods has different scaling with the dimension of the data and number of training points used, and so each is better-suited to different types of supervised learning problems.

The outline of this paper is as follows. Section 2 introduces our proposed linear quantum photonic circuit architecture and analyzes how its expressive power depends on the number of spatial modes and input photons. Next, Sect. 3 illustrates the photon number-dependent performance of the circuit for supervised classification problems. Section 4 concludes the paper.

2 Parametrized linear quantum photonic circuit model

To demonstrate the Fock state-enhanced expressive power of linear quantum photonic circuits, in this Section we consider the encoding of univariable functions onto circuit's output. For simplicity we consider the circuit architecture illustrated schematically in Fig. 1, consisting of a single data-dependent encoding layer $\mathcal{S}$ sandwiched between two trainable beam splitter meshes $\mathcal{W}^{(1,2)}$, described by the unitary transformation

$$\mathcal{U}(x, \boldsymbol{\Theta}) = \mathcal{W}^{(2)}(\boldsymbol{\theta}_2)\mathcal{S}(x)\mathcal{W}^{(1)}(\boldsymbol{\theta}_1), \tag{2}$$

where $\boldsymbol{\Theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ parameterizes transformations applied by trainable beam splitter meshes and x is the input data. The n-photon quantum model (circuit's output) is defined as the expectation value of some observable $\mathcal{M}(\boldsymbol{\lambda})$ with respect to a state prepared via the parameterised linear QPC,

$$f^{(n)}(x, \boldsymbol{\Theta}, \boldsymbol{\lambda}) = \langle \mathbf{n}^{(i)} | \mathcal{U}^\dagger(x, \boldsymbol{\Theta}) \mathcal{M}(\boldsymbol{\lambda}) \mathcal{U}(x, \boldsymbol{\Theta}) | \mathbf{n}^{(i)} \rangle, \tag{3}$$

where $|\mathbf{n}^{(i)}\rangle = |n_1^{(i)}, n_2^{(i)}, \ldots, n_m^{(i)}\rangle$ is the input n-photon Fock state with $n = \sum_i^m n_i^{(i)}$ and $\boldsymbol{\lambda}$ parameterizes the observable. We consider measurements made using either photon number-resolving (PNR) detectors or single photon (threshold) detectors, corresponding to $\mathcal{M}$ being diagonal in the Fock state basis with d or d' distinct parameterized eigenvalues $\{\lambda_j \in \boldsymbol{\lambda}\}_{j=1}^{d^{(\prime)}}$, respectively.

The multi-mode Fock state unitary transformation $\mathcal{U}(x, \boldsymbol{\Theta})$ is constructed from permanents of submatrices of the m-mode linear transformation matrix $U(x, \boldsymbol{\Theta}) = W^{(2)}(\boldsymbol{\theta}_2)S(x)W^{(1)}(\boldsymbol{\theta}_1)$ using the scheme of Ref. [50] with $W^{(i)}$ as the programmable transfer matrix, describing the universal multiport interferometer that realizes arbitrary linear optical input-output transformations [51–53]. Each trainable unitary $W^{(i)}(\boldsymbol{\theta}_i)$ is parameterized by a vector $\boldsymbol{\theta}_i$ of $m(m-1)$ phase shifter and beam splitter angles constructed using the encoding of Reck et al. [51]. The data encoding block $S(x)$ employs a single tunable phase shifter placed at the first spatial mode.

2.1 n-photon quantum models as Fourier series

In this Section we will show how to express the n-photon quantum models as a Fourier series. For simplicity, we consider arbitrary unitary operations $\mathcal{W}(\boldsymbol{\theta}) = \mathcal{W}$, an arbitrary parameterized observable obtained using PNR detectors $\mathcal{M}(\boldsymbol{\lambda}) = \mathcal{M}$, and $|\boldsymbol{n}^{(i)}\rangle = |n, 0, \ldots, 0\rangle$ as the initial Fock state. The component of the output quantum state with photon numbers $\boldsymbol{n}^{(f)}$ can be written as [24]

$$\langle \boldsymbol{n}^{(f)} | \mathcal{U}(x) | \boldsymbol{n}^{(i)} \rangle = \sum_{\boldsymbol{n}'} \mathcal{W}^{(2)}_{\boldsymbol{n},\boldsymbol{n}'} e^{in'_1 x} \mathcal{W}^{(1)}_{\boldsymbol{n}',\boldsymbol{n}^{(i)}}, \tag{4}$$

where the summation runs over the basis of $d = \binom{n+m-1}{n}$ Fock states corresponding to different combinations of n photons in the m spatial modes. The data encoding block imposes a phase shift proportional to the number of photons in the first mode following the first beam splitter mesh.

The output of full model Eq. (3) is obtained by taking the modulus square of Eq. (4), multiplying by the corresponding observable weight, and then summing over all output Fock basis states. This yields an expression of the form

$$f^{(n)}(x) = \sum_{\boldsymbol{n}'',\boldsymbol{n}'} a_{\boldsymbol{n}'',\boldsymbol{n}'} e^{i(n'_1 - n''_1)x}, \tag{5}$$

where $a_{\boldsymbol{n}'',\boldsymbol{n}'}$ contain the matrix elements from the unitaries $\mathcal{W}^{(i)}$ and measurement's observable $\mathcal{M}$,

$$a_{\boldsymbol{n}'',\boldsymbol{n}'} = \sum_{\boldsymbol{n}} \mathcal{W}^{*(1)}_{\boldsymbol{n}^{(i)},\boldsymbol{n}''} \mathcal{W}^{*(2)}_{\boldsymbol{n}'',\boldsymbol{n}} \mathcal{M}_{\boldsymbol{n},\boldsymbol{n}} \mathcal{W}^{(2)}_{\boldsymbol{n},\boldsymbol{n}'} \mathcal{W}^{(1)}_{\boldsymbol{n}',\boldsymbol{n}^{(i)}}. \tag{6}$$

This expression can be simplified by grouping the basis function with the same frequency $\omega = n'_1 - n''_1$. This gives

$$f^{(n)}(x) = \sum_{\omega \in \Omega_n} c_\omega e^{i\omega x}, \tag{7}$$

where the coefficients c_ω are obtained by summing over all $a_{\boldsymbol{n}'',\boldsymbol{n}'}$ contributing to the same frequency

$$c_\omega = \sum_{\substack{\boldsymbol{n}'',\boldsymbol{n}' \\ n'_1 - n''_1 = \omega}} a_{\boldsymbol{n}'',\boldsymbol{n}'}, \tag{8}$$

with $c_\omega = c_{-\omega}^*$ and Eq. (7) is a real-value function, as it should be. The frequency spectrum $\Omega_n = \{n_1' - n_1'' | n_1', n_1'' \in [0, n]\}$ contains all frequencies that are accessible to the n-photon quantum model. For general trainable circuit blocks $\mathcal{W}^{(i)}(\boldsymbol{\theta}_i)$, measurement observable $\mathcal{M}(\boldsymbol{\lambda})$ and n-photon Fock state $|\boldsymbol{n}^{(i)}\rangle = |n_1^{(i)}, n_2^{(i)}, \ldots, n_m^{(i)}\rangle$, the n-photon quantum model reads

$$f^{(n)}(x, \boldsymbol{\Theta}, \boldsymbol{\lambda}) = \sum_{\omega \in \Omega_n} c_\omega(\boldsymbol{\Theta}, \boldsymbol{\lambda}) e^{i\omega x}. \tag{9}$$

2.2 Expressive power and trainability of linear quantum photonic circuits

Since the n-photon quantum model can be represented by a Fourier series, its expressive power can be studied via two properties: its frequency spectrum and Fourier coefficients. The former tells us which functions the quantum model has access to, while the latter determines how the accessible functions can be combined [24].

2.2.1 Photon-number dependent frequency spectrum

The frequency spectrum can be easily shown to be

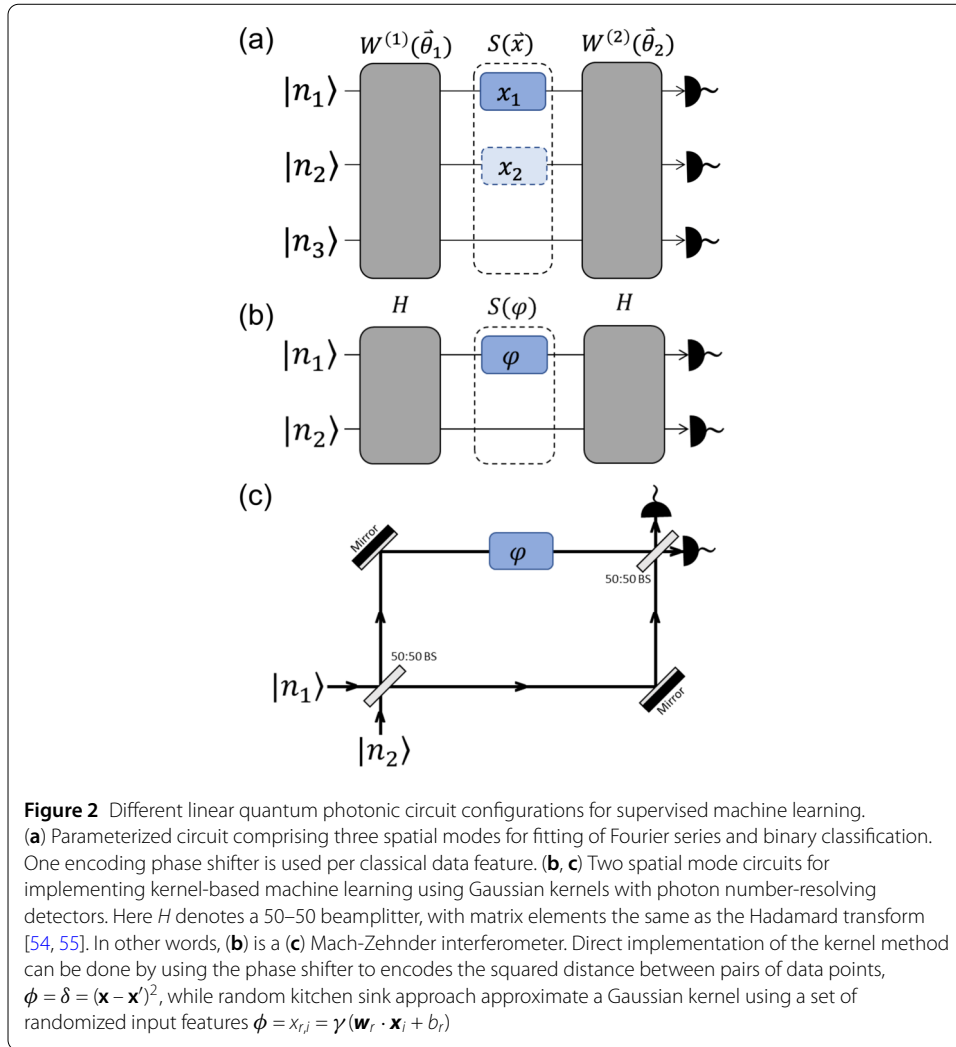$$\Omega_n = \{-n, -n+1, \ldots, n-1, n\}, \tag{10}$$

which is solely determined by the number of photons fed into the circuit. It always contains the zero frequency, i.e: $0 \in \Omega_n$, while the non-zero frequencies occur in pairs, i.e: $\omega, -\omega \in \Omega_n$. This motivates us to define the size of the frequency spectrum as $D_n = (|\Omega_n| - 1)/2 = n$ to quantify the number of independent non-zero frequencies the n-photon quantum model has access to. In comparison to Ref. [24], where the size of frequency spectrum is determined by the spectrum of the data encoding Hamiltonian, here the size of the frequency spectrum can be increased by feeding more photons into the circuit, while keeping the number of spatial modes and encoding phase shifters constant.

This implies that n-photon quantum models with more input photons can be more expressive, because they have access to more basis functions, and hence can learn Fourier series with higher frequencies. In the limit of $n \to \infty$, i.e: continuous variable quantum systems, n-photon quantum models can support the frequency spectrum $\Omega_\infty = \{-\infty, \ldots, -1, 0, 1, \ldots, \infty\}$ of a full Fourier series, in agreement with Ref. [24]. For a fixed number of input photons, the frequency spectrum can be broadened further using multiple encoding phase shifters, either in parallel or sequentially [23, 24] (see Appendix A).

As an example, we consider training a linear QPC with three spatial modes shown in Fig. 2(a) using a regularized squared loss cost function. The cost function $C(\boldsymbol{\Theta}, \boldsymbol{\lambda})$ is constructed using the measurement results and a training set of N desired input/output pairs $\{x_i \to g(x_i)\}_{i=1}^N$
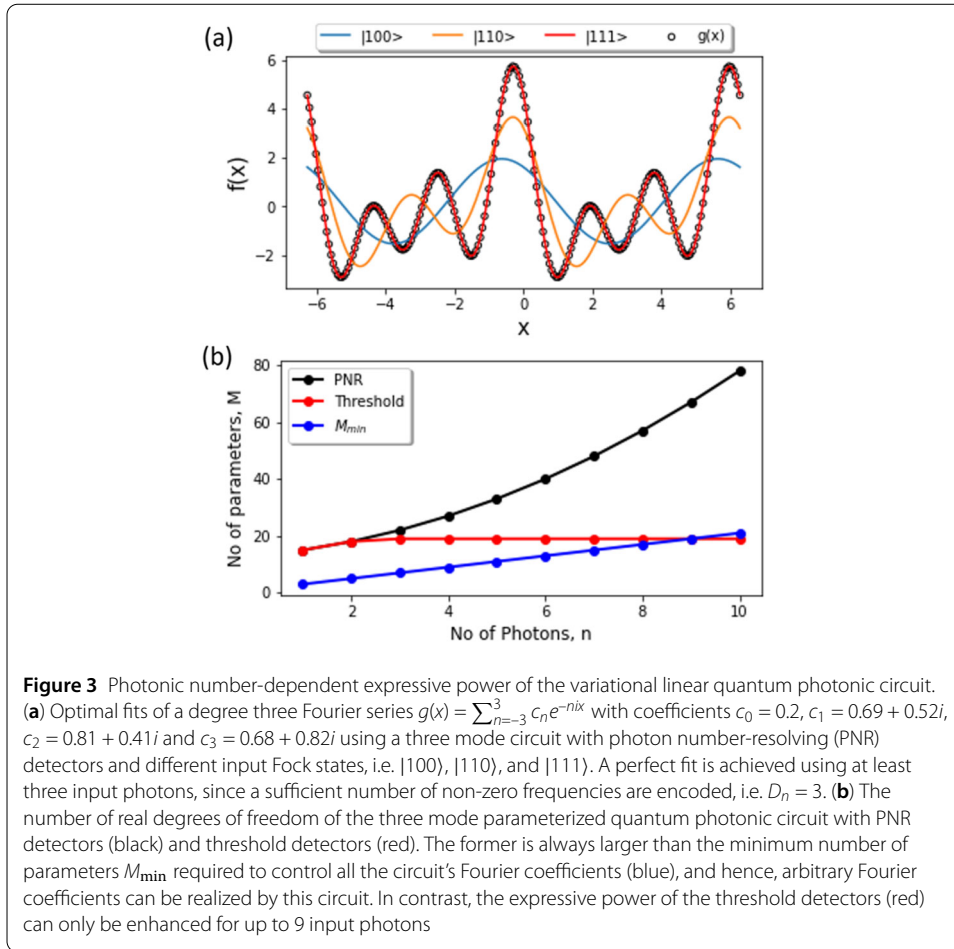
$$C(\boldsymbol{\Theta}, \boldsymbol{\lambda}) = \frac{1}{2N} \sum_{i=1}^N \big(g(x_i) - f^{(n)}(x_i, \boldsymbol{\Theta}, \boldsymbol{\lambda})\big)^2 + \alpha \boldsymbol{\lambda} \cdot \boldsymbol{\lambda}, \tag{11}$$

that is variationally minimized over $\boldsymbol{\Theta}$ and $\boldsymbol{\lambda}$ to learn the function $g(x)$. Here, $f(x, \boldsymbol{\Theta}, \boldsymbol{\lambda})$ is the n-photon quantum model in Eq. (9), while $\boldsymbol{\lambda} \cdot \boldsymbol{\lambda} = \sum_i \lambda_i^2$ is the sum of squared observable parameters, forming a regularization term with weight α. The regularization term

Figure 2 Different linear quantum photonic circuit configurations for supervised machine learning. (**a**) Parameterized circuit comprising three spatial modes for fitting of Fourier series and binary classification. One encoding phase shifter is used per classical data feature. (**b**, **c**) Two spatial mode circuits for implementing kernel-based machine learning using Gaussian kernels with photon number-resolving detectors. Here H denotes a 50–50 beamplitter, with matrix elements the same as the Hadamard transform [54, 55]. In other words, (**b**) is a (**c**) Mach-Zehnder interferometer. Direct implementation of the kernel method can be done by using the phase shifter to encodes the squared distance between pairs of data points, $\phi = \delta = (\mathbf{x} - \mathbf{x}')^2$, while random kitchen sink approach approximate a Gaussian kernel using a set of randomized input features $\phi = x_{r,i} = \gamma(\mathbf{w}_r \cdot \mathbf{x}_i + b_r)$

has a two-fold role: it prevents model over-fits, and ensures that the model prediction is not based on output photon combinations that occur with very low probability. The latter is important for QPC-based machine learning models, because the number of measurements required to obtain all of the required expectation values scales with the number of spatial modes and photons.

We train the three mode linear QPC using the gradient-free algorithm in the *NLopt* nonlinear-optimization python package [56], i.e: BOBYQA algorithm [57] to fit a degree three Fourier series $g(x)$ with $x \in [-3\pi, 3\pi]$ using input states of up to 3 photons. We consider input states for which each spatial mode contains at most one photon, i.e. $|\boldsymbol{n}^{(i)}\rangle = |100\rangle$, $|110\rangle$, or $|111\rangle$. Figure 3(a) shows how the number of observable frequency components and hence the expressive power of the circuit grows with the number of input photons. Perfect fitting is achieved with three photons. In contrast, the frequency spectrum of similar qubit-based architectures cannot fit a degree two Fourier series using a single encoding block, requiring either a deeper or wider circuits with multiple encoding gates [23, 24].

Figure 3 Photonic number-dependent expressive power of the variational linear quantum photonic circuit. (**a**) Optimal fits of a degree three Fourier series $g(x) = \sum_{n=-3}^{3} c_n e^{-nix}$ with coefficients $c_0 = 0.2$, $c_1 = 0.69 + 0.52i$, $c_2 = 0.81 + 0.41i$ and $c_3 = 0.68 + 0.82i$ using a three mode circuit with photon number-resolving (PNR) detectors and different input Fock states, i.e. $|100\rangle$, $|110\rangle$, and $|111\rangle$. A perfect fit is achieved using at least three input photons, since a sufficient number of non-zero frequencies are encoded, i.e. $D_n = 3$. (**b**) The number of real degrees of freedom of the three mode parameterized quantum photonic circuit with PNR detectors (black) and threshold detectors (red). The former is always larger than the minimum number of parameters $M_{\mathbf{min}}$ required to control all the circuit's Fourier coefficients (blue), and hence, arbitrary Fourier coefficients can be realized by this circuit. In contrast, the expressive power of the threshold detectors (red) can only be enhanced for up to 9 input photons

2.2.2 Trainability of Fourier coefficients

Even if the parametrized QPC can generate the frequency spectrum required to fit the desired function, this does not necessarily imply that the optimal Fourier coefficients are accessible [24]; the linear circuits we consider cannot perform arbitrary Fock state transformations. However, we do not need to generate arbitrary Fock states and only require control over one real and D_n complex Fourier coefficients $\{c_\omega\}$. For n input photons and taking $D_n = n$, this requires at least $M_{min} = 2D_n + 1$ real degrees of freedom.

Each trainable circuit blocks has $m(m - 1)$ controllable parameters [51], while the number of controllable degrees of freedom of the parameterized observable depends on type of detector. For photon number resolving (PNR) detectors the number of degrees of freedom is

$$M_{\mathrm{PNR}} = 2m(m - 1) + \frac{(n + m - 1)!}{n!(m - 1)!}, \tag{12}$$

while threshold detectors have

$$M_{\mathrm{THR}} = 2m(m - 1) + \sum_{k=1}^{\min(n,m)} \frac{m!}{k!(k - m)!} \tag{13}$$

degrees of freedom.

For a fixed number of spatial modes and photons, threshold detectors have fewer controllable degrees of freedom compared to PNR detectors, and hence their expressive power saturates beyond a certain number of input photons. For example, Fig. 3(b) illustrates the expressive power of a circuit with three spatial modes. Using threshold detectors the expressive power is only enhanced by increasing the number of photons up to nine; beyond this, the number of controllable degrees of freedom is less than $M_{\min}$. On the other hand, using PNR detectors the circuit may in principle be trained to fit arbitrarily large frequencies by increasing the number of input photons. Of course, in practice the range of frequencies accessible using a single encoding gate will be limited by sensitivity to losses, which grows exponentially with the photon number.

2.2.3 Universality of the linear quantum photonic circuit

It is well known that a Fourier series can approximate any square-integrable function $g(x)$ on a finite interval to arbitrary precision [48, 49]. Thus, expressing the n-photon quantum model in term of a Fourier series allows us to demonstrate the universality of the quantum model by studying its ability to realise arbitrary Fourier series. Universality of a Fourier series is determined by two important ingredients: a sufficiently-broad frequency spectrum and arbitrary realizable Fourier coefficients. The analysis in Sect. 2.2.1 implies that the frequency spectrum Ω_n accessible by n-photon quantum models asymptotically contains any integer frequency if enough input photons are used, satisfying one of the criteria to achieve universality.

To realize arbitrary set of Fourier coefficients, at least $M \geq M_{\min} = 2n + 1$ degrees of freedom in the linear QPC are required. Here, we consider a linear QPC with PNR detectors. The PNR detectors are used because the expressive power of threshold detectors saturated beyond some threshold number of photons. One of the unique advantages of photonic system is the exponentially growing dimension of the Fock space with number of spatial modes and photons. For a linear QPC with constant number of spatial modes m, the dimension of the Fock space and M_{PNR} scales in the order of $O(n^{m-1})$, hence contributing $O(n^{m-1})$ of degrees of freedom. On the other hand, the degrees of freedom from the trainable circuit blocks scale with $O(m^2)$, which is negligible when $n \gg m$. By exploiting this advantage, it can be seen that M_{PNR} is always larger than $M_{\min}$ as the size of the frequency spectrum D_n and $M_{\min}$ scale linearly with photon number, i.e: $O(n)$. This is a necessary condition for the n-photon quantum model being able to realize arbitrary set of Fourier coefficients, which in the examples we consider also seems to be sufficient. More rigorously, following the arguments in Ref. [24] a universal function approximator may be obtained by generalizing our circuits to arbitrary (entangled) input states and observables by incorporating nonlinear elements into the circuits.[1]

As an example, we consider a linear QPC with 3 spatial modes. In this case, Eq. (12) is reduced to

$$M_{PNR} = 12 + \frac{(n+2)(n+1)}{2!}, \tag{14}$$

---

[1]The nonlinear optical elements enables the realization of two qubit entangling gates, while arbitrary single qubit rotations can be realized using linear optics such as beam splitter and phase shifters. By considering the dual-rail photonic qubit, our circuit could perform universal quantum computation, and the Fourier coefficients arguments follows from Ref. [24].

which is always larger than $M_{\min} = 2n + 1$ for $n \in \mathbb{N}$, as shown in Fig. 3(b). Hence, the *n*-photon quantum model with 3 spatial modes and a single phase shifter can act as a universal function approximator. In contrast, the qubit-type variational quantum circuits require deep or wide circuits and many encoding gates to ensure a rich frequency spectrum, and arbitrary global unitaries to realize arbitrary sets of Fourier coefficients [24].

2.2.4 Effect of noise on the expressive power of linear quantum photonic circuits

For noiseless linear quantum photonic circuits, we have shown that its expressive power will improve with the increasing number of photons and spatial modes. In this section, we will discuss the role of optical losses on the expressive power of linear quantum photonic ciruits. For real quantum photonic hardware, the optical loss sensitivity will grows exponentially with the circuit depth and number of input photons. The typical noise sources are (1) inefficient collection optics, (2) losses in the optical components due to absorption, scattering, or reflections from the surfaces, (3) inefficiency in the detection process due to using detectors with imperfect quantum efficiency, and they can be modelled using beam splitters [58]. These noises will obviously affect the frequency spectrum, where the higher frequency term cannot be distinguished from the lower frequency term, hence reducing the size of the frequency spectrum. We anticipate the noises to have a minimum impact on the Fourier coefficients, as they depend only on the physical components such as linear optics in trainable blocks and the detectors. Therefore, the output observables can still be written as Fourier series, just with reduced expressivity. This will place a practical limit on the complexity of the QML models using this scheme, unless one can include some kind of error correction scheme. When the losses are low enough, the detectors should have a sufficiently high signal to noise ratio that other noise sources can be neglected. Apart from the error correction scheme, the regularization term in the cost function should be able to help to minimize the detrimental influence of noise. It penalizes models including coefficients with huge weights, hence no particular output state should have a huge weight, reducing the model's sensitivity of noise in final prediction. Finally, the photonic circuit considered here are based on variational approach, therefore, they are robust against variations in the beam splitter ratios, tuning, and etc. The quantitative noise modelling of the linear photonic quantum circuits will be a subject for future research.

3 Supervised learning using linear quantum photonic circuits

As an application of the trainable linear QPCs we now consider different strategies for binary classification. In the first strategy the linear QPCs are directly used as variational quantum classifiers, classifying data directly on the high-dimensional Fock space by optimizing a regularized squared loss cost function. In this case, as the circuit becomes more expressive it becomes harder to train. Second, we consider kernel methods as a means of avoiding the costly circuit optimization step. We show how linear circuits can be used to implement Gaussian kernels either directly or using the random kitchen sinks algorithm, sampling kernels with different resolutions in parallel. Note that we are mainly interested in what kinds of kernel functions can be efficiently implemented using linear QPCs, instead of providing quantum kernels that might offer quantum advantages, motivated by ongoing interest in classical photonic circuits for machine learning [1–4].

3.1 Linear quantum photonic circuit as variational quantum classifiers

We perform binary classification of two-dimensional classical data. Each data dimension is encoded using a single phase shifter, as shown in Fig. 2(a). The mapping of the data into the high-dimensional Fock space is nonlinear, circumventing the need for nonlinear optical elements.

The n-photon supervised classification model for two-dimensional data $f^{(n)}(\boldsymbol{x}, \boldsymbol{\Theta}, \boldsymbol{\lambda})$ is defined as

$$f^{(n)}(\boldsymbol{x}, \boldsymbol{\Theta}, \boldsymbol{\lambda}) = \sum_{\boldsymbol{\omega} \in \Omega_n} c_{\boldsymbol{\omega}}(\boldsymbol{\Theta}, \boldsymbol{\lambda}) e^{i \boldsymbol{\omega} \cdot \boldsymbol{x}}, \tag{15}$$

where $\boldsymbol{x} = (x_1, x_2)$ is the 2-dimensional data feature, $\boldsymbol{\omega} = (\omega_1, \omega_2)$ is the 2-dimensional frequency vector, and $\Omega_n = \{-\boldsymbol{\omega}, 0, \boldsymbol{\omega} | \omega_1, \omega_2 \in [0, n], \omega_1 + \omega_2 \leq n\}$ is the frequency spectrum of the model. This encoding scheme will not generate a full frequency spectrum for multi-dimensional Fourier series but it suffice for the example considered here. See Appendix B for schemes that generate full frequency spectrum for multi-dimensional Fourier series.

The model is trained by minimizing the cost function

$$C(\boldsymbol{\Theta}, \boldsymbol{\lambda}) = \frac{1}{2N} \sum_{i=1}^{N} \big(g(\boldsymbol{x}_i) - f^{(n)}(\boldsymbol{x}_i, \boldsymbol{\Theta}, \boldsymbol{\lambda}) \big)^2 + \alpha \boldsymbol{\lambda} \cdot \boldsymbol{\lambda} \tag{16}$$

using the BOBYQA algorithm, with the decision boundary defined as

$$f_{\text{sgn}}^{(n)}(\boldsymbol{x}) = \text{sgn}\big[f^{(n)}(\boldsymbol{x}, \boldsymbol{\Theta}_{\text{opt}}, \boldsymbol{\lambda}_{\text{opt}}) \big], \tag{17}$$

where $\boldsymbol{\Theta}_{\text{opt}}$ and $\boldsymbol{\lambda}_{\text{opt}}$ are the optimized circuit's and observable's parameters and sgn is the sign function. Thus, the class of the data points is assigned by the sign of circuit output.

As an example, we trained the linear QPC to classify three different types of datasets from the *scikit-learn* machine learning library [59]: linear, circle, and moon. Figure 4 illustrates the trained models. The contour plots show that n-photon supervised classification models with higher photon number have more complicated classification boundaries, in agreement with previous analysis on the expressive power of quantum models. Since the linear data set can be separated by a linear decision boundary, unsurprisingly a single model photon is sufficient to learn the classification boundary. On the other hand, over-fitting can occur when the model expressive power is too large, as can be seen for the degraded performance for the circle dataset for the $|221\rangle$ input state. The classification performance for the more complicated moon dataset improves with the number of input photons. These examples illustrate the impact of a higher expressive power on classification using linear QPCs.

3.2 Linear quantum photonic circuits as Gaussian kernel samplers

Similar to the standard noisy and large scale variational circuits, the variational machine learning approach becomes more difficult to train as the dimension of the Fock space increases, likely due to the issue of vanishing cost function gradients [60–64], requiring exponentially-growing precision to optimize the circuit parameters in-situ [65]. In addition, it is expensive to train the quantum gates (in this case the tunable beam splitter meshes) in the noisy-intermediate scale quantum (NISQ) era as it is time-consuming to

Figure 4 Binary classification using the three mode linear quantum photonic circuit of Fig. 2(a) with different input Fock states |100⟩, |111⟩, and |221⟩, training using 60 points with a regularization weight $\alpha = 0.2$. First row: linearly-separable dataset. Middle row: circle dataset. Bottom row: moon dataset. The performance on a test set (red and blue solid cross) of 40 points is given in the upper left corner of each respective subplot. The classification boundaries for all datasets become more complicated as the number of input photons increases, illustrating the increasing expressive power. Increase of expressive power does not affect the trainability of the linear dataset, since a linear classifier suffices. The performance for the circle dataset degrades for larger input photons due to over-fitting, demonstrating that a larger expressive power is not necessarily better. On the other hand, a higher expressive power is necessary in order to accurately classify the moon dataset

reprogram quantum circuits [46]. Due to these limitations, it is more efficient to use NISQ devices as sub-routines for machine learning algorithms, e.g. to sample quantities that are useful for classical machine learning models but time-consuming to compute. In particular, variational quantum circuits can be used to approximate kernel functions for classical kernel models such as support vector machines [10, 13, 16, 20, 66]. Here, we show how the linear QPCs can be designed to approximate Gaussian kernels with a range of resolutions determined by the number of input photons.

3.2.1 Kernel methods

Kernel methods allow one to apply linear classification algorithms to datasets with nonlinear decision boundaries [67, 68]. The idea is to leverage feature maps $\phi(\boldsymbol{x})$ that map the nonlinear dataset from its original space into a higher dimensional feauture space in which a linear decision boundary can be found, enabling classification via a linear regression

$$f(\boldsymbol{x}) = \boldsymbol{w} \cdot \phi(\boldsymbol{x}),$$

using suitably-trained weights $\boldsymbol{w}$. Instead of computing and storing the high-dimensional feature vector ϕ, the kernel trick [68, 69] is employed by introducing a kernel function $k(\boldsymbol{x}, \boldsymbol{x}')$, which measures the pairwise similarity between the data points in the feature space. Formally, the kernel functions is defined as the inner product of two feature vectors

$$k(\boldsymbol{x}, \boldsymbol{x}') = \phi(\boldsymbol{x}) \cdot \phi(\boldsymbol{x}'). \tag{18}$$

According to representer theorem [70], the solution to the decision boundary can then be expressed in term of the kernel functions as

$$f(\mathbf{x}) = \sum_{i=1}^{N} \beta_i k(\mathbf{x}_i, \mathbf{x}), \tag{19}$$

converting the optimization problem into a convex optimization problem of finding the parameters β_i. In the case of a regularized squared loss cost function such as Eq. (16) the optimal parameters β_i can be obtained analytically [71, 72] as

$$\boldsymbol{\beta} = (K + \alpha \mathbf{I})^{-1} \mathbf{y}, \tag{20}$$

where N is the number of training data, K is the $N \times N$ kernel matrix with matrix elements $K_{ij} = k(\mathbf{x}_i, \mathbf{x}_j)$, I is the N-dimensional identity matrix, α is the regularization parameter, and $\mathbf{y}$ is the $N \times 1$ vector of the training data labels.

Although we currently have an example that shows rigorous performance guarantees of quantum kernel methods on artificial dataset [47], it is still unclear whether quantum machine learning models can achieve improved performance compared to classical machine learning approaches in practical problems by sampling from kernels that are hard to compute classically [13, 21, 73, 74]. Even in the absence of a rigorous quantum advantage, special-purpose electronic and photonic machine learning circuits are being pursued in order to increase the speed and energy-efficiency of well-established classical machine learning models [1–4]. Therefore, here we will focus on implementing the widely-used Gaussian kernel

$$k(\mathbf{x}, \mathbf{x}') = k(\mathbf{x} - \mathbf{x}') = e^{-\frac{1}{2\sigma^2}(\mathbf{x} - \mathbf{x}')^2}$$

with controllable resolution σ. The Gaussian kernel is a universal, infinite-dimensional kernel that can learn any continuous function in a compact space [75].

3.2.2 Linear quantum photonic circuits as sub-routine of kernel methods

We approximate the Gaussian kernel using the two mode QPC shown in Fig. 2(b), where 50-50 beamsplitters $\mathcal{H}$ are used for both trainable circuit blocks and the squared Euclidean distance between pairs of data points $\delta = (\mathbf{x} - \mathbf{x}')^2$ is encoded using a single phase shifter. The output of this circuit can be written as

$$
\begin{aligned}
f^{(n)}(\delta, \boldsymbol{\lambda}) &= \langle n, 0 | \mathcal{U}^{\dagger}(\delta) \mathcal{M}(\boldsymbol{\lambda}) \mathcal{U}(\delta) | n, 0 \rangle \\
&= \sum_{ij} \lambda_{ij} \big| \langle n_i, n_j | \mathcal{U}(\delta) | n, 0 \rangle \big|^2 \\
&= c_0^{(n)}(\boldsymbol{\lambda}) + 2 \sum_{k=1}^{n} c_k^{(N)}(\boldsymbol{\lambda}) \cos(k\delta),
\end{aligned}
$$

where $\mathcal{U}(\delta) = \mathcal{H} \mathcal{S}(\delta) \mathcal{H}$ and the trainable observable $\mathcal{M}(\boldsymbol{\lambda}) = \sum_{ij} \lambda_{ij} | n_i, n_j \rangle \langle n_i, n_j |$ with $n_i + n_j = n$. Similar to Sect. 3.1, the observable is trained to approximate the Gaussian kernel of resolution σ by minimizing the squared loss cost function using the BOBYQA algorithm

$$f^{(n)}(\delta, \boldsymbol{\lambda}^{(\sigma)}) \approx e^{-\frac{\delta}{2\sigma^2}}.$$

Figure 5 Approximating Gaussian kernels with different resolutions $\sigma = 0.25, 0.33, 0.50, 1.00$ using the two mode linear quantum photonic circuit of Fig. 2(b) with different numbers of input photons $n = 2, 4, 6, 8, 10$. Two photons are sufficient to approximate a low resolution kernel with $\sigma = 1.00$ (curves for all photon numbers overlap), while higher resolutions require more photons to approximate. For example, a circuit with four photons can fit Gaussian kernels with $\sigma = 0.50$, but not $\sigma = 0.33$ or 0.25

This approach has two advantages: Different kernel resolutions can be accessed using the same photon detection statistics $|\langle n_i, n_j | \mathcal{U}(\delta) | n, 0\rangle|^2$ by taking different linear combinations of the output observables $\mathcal{M}(\lambda^{(\sigma)})$. Second, this training only needs to be performed once; the tunable circuit blocks do not need to be reconfigured if the training data set changes.

We note that the domain of the input data, in this case the norm squared distances between pairs of data points, must lie within the interval that defines the circuit's Fourier series. This imposes an upper bound on the kernel resolution that the linear QPC has access to. The circuit with higher expressive power, i.e: higher number of input photons, can more precisely approximate kernels with higher resolution σ. Kernels with lower resolution can already be well-approximated by a circuit with only two input photons. Figure 5 shows the kernel training result for different desired resolutions and input photon numbers. Once the kernel has been trained, classification can be performed by feeding the measured similarity matrix into a classical machine learning model such as a support vector machine [76].

3.3 Quantum-enhanced random kitchen sinks

One limitation of kernel methods is their poor (quadratic) scaling with the size of the training data set. To circumvent this issue, the random kitchen sinks (RKS) algorithm was developed, which uses randomly-sampled feature maps in order to controllably approximate desired kernels and more efficiently train classical machine learning models [77–79]. In particular, sampling from random Fourier features enables approximation of the Gaussian kernel. This motivates us to propose a quantum-enhanced RKS algorithm, where the subroutine of RKS algorithm, i.e: the random feature sampler is replaced by linear QPCs. The linear QPCs can simultaneously sample random Fourier features of different frequencies, providing a unique advantage compare to the qubit-based architecture. Our

approach differs from previous proposals for quantum random kitchen sinks by directly constructing the kernel functions using the random Fourier features, instead of performing linear regression with random feature bit strings sampled from variational quantum circuits [80, 81].

3.3.1 Random kitchen sinks

The randomized R-dimensional vectors known as the random Fourier features are defined as

$$
\mathbf{z}(\mathbf{x}) = \frac{1}{\sqrt{R}} \begin{pmatrix} z_{\mathbf{w}_1}(\mathbf{x}) \\ z_{\mathbf{w}_2}(\mathbf{x}) \\ \dots \\ z_{\mathbf{w}_R}(\mathbf{x}) \end{pmatrix},
\tag{21}
$$

where each $z_{\mathbf{w}_r}(\mathbf{x})$ is a randomized cosine function

$$
z_{\mathbf{w}_r}(\mathbf{x}) = \sqrt{2}\cos\left(\gamma\left[\mathbf{w}_r \cdot \mathbf{x} + b_r\right]\right),
\tag{22}
$$

$\mathbf{x}$ is the D-dimensional input data, $\mathbf{w}_r$ are D-dimensional random vector sampled from a spherical Gaussian distribution, and b_r are random scalars sampled from a uniform distribution,

$$
\mathbf{w} \sim \mathcal{N}_D(0, \mathbf{I});
$$

$$
b \sim \mathrm{Uniform}(0, 2\pi).
$$

The random Fourier features approximate the Gaussian kernel [77, 78]

$$
\mathbf{z}(\mathbf{x}) \cdot \mathbf{z}(\mathbf{x}') \approx k(\mathbf{x}, \mathbf{x}') = e^{-\frac{\gamma^2}{2}(\mathbf{x}-\mathbf{x}')^2}
\tag{23}
$$

with γ acting as a hyperparameter that controls the kernel resolution. Note that other commonly-used kernels including Laplacian and Cauchy kernels can be approximated using the RKS algorithm using different sampling distributions [77].

Substituting Eq. (23) into Eq. (19) yields

$$
f(\mathbf{x}) \approx \sum_{i=1}^{N} \beta_i \mathbf{z}(\mathbf{x}_i) \cdot \mathbf{z}(\mathbf{x}) = \mathbf{c} \cdot \mathbf{z}(\mathbf{x}),
\tag{24}
$$

where $\mathbf{c} = \sum_i \beta_i \mathbf{z}(\mathbf{x}_i)$. The optimal solution for a supervised learning problem using training data $\{(\mathbf{x}_j, y_j)\}_{j=1}^{N}$ that minimizes the regularized squared loss cost function [71, 72, 82] in Eq. (24) is

$$
f^*(\mathbf{x}) = \mathbf{c}_{\mathrm{opt}} \cdot \mathbf{z}(\mathbf{x}),
\tag{25}
$$

where $\boldsymbol{c}_{\mathrm{opt}} = (\boldsymbol{z}(\boldsymbol{X})^T \boldsymbol{z}(\boldsymbol{X}) + \alpha I_R)^{-1} \boldsymbol{z}(\boldsymbol{X})^T \boldsymbol{y}$, $\boldsymbol{z}(\boldsymbol{X})$ is an $N \times R$ matrix of the training data

$$\boldsymbol{z}(\boldsymbol{X}) = \begin{pmatrix} \boldsymbol{z}(\boldsymbol{x}_1)^T \\ \boldsymbol{z}(\boldsymbol{x}_2)^T \\ \vdots \\ \boldsymbol{z}(\boldsymbol{x}_N)^T \end{pmatrix}, \tag{26}$$

and I_R is the R-dimensional identity matrix.

The RKS algorithm addresses the poor scaling of kernel methods with the number of data points by mapping the data into a randomized low-dimensional feature space, turning Eq. (19) into a linear model on the R-dimensional vectors $\boldsymbol{z}(\boldsymbol{x})$. The complexity of finding the analytical solution for the coefficients is reduced from $O(N^3)$ to $O(R^3)$, saving enormous amounts of resources when $R \ll N$ while maintaining model performance comparable to standard classification methods [77, 78].

3.3.2 Linear quantum photonic circuits as random Fourier feature samplers
Using the same circuit as in Sect. 3.2.2 with a randomized input encoding [Fig. 2(b)], i.e. $x_{r,i} = \gamma(\boldsymbol{w}_r \cdot \boldsymbol{x}_i + b_r)$, the circuit output becomes

$$f^{(n)}(x_{r,i}, \boldsymbol{\lambda}) = c_0^{(n)}(\boldsymbol{\lambda}) + 2 \sum_{k=1}^{n} c_k^{(n)}(\boldsymbol{\lambda}) \cos(k x_{r,i}). \tag{27}$$

Constructing different observables $\mathcal{M}(\boldsymbol{\lambda}^{(k)})$ from the same photon detection statistics $|\langle n_i, n_j | \mathcal{U}(x_{r,i}) | n, 0 \rangle|^2$ allows one to isolate cosine functions with different frequencies k

$$f^{(n)}(x_{r,i}, \boldsymbol{\lambda}^{(k)}) = \sqrt{2} \cos(k\gamma [\boldsymbol{w}_r \cdot \boldsymbol{x}_i + b_r]), \tag{28}$$

which has the same structure as Eq. (22) with $\gamma \to k\gamma$. Thus, constructing the random Fourier features in Eq. (21) with the randomized cosine functions Eq. (28) enables us to approximate the kernel

$$k(\boldsymbol{x}, \boldsymbol{x}') = e^{-\frac{k^2 \gamma^2}{2}(\boldsymbol{x} - \boldsymbol{x}')^2} \tag{29}$$

with resolution $\sigma = \frac{1}{k\gamma}$. In other words, Gaussian kernels with different resolutions can be accessed using a single QPC and the same set of measurements by considering different observables. The number of kernel resolutions accessible by the circuit is equal to the size of the frequency spectrum, i.e: circuits with more input photons have access to more resolutions. Here, the photon-number dependent expressive power of the linear QPCs is leveraged to produce a linear combination of cosine functions of different frequencies, simultaneously producing multiple random Fourier features that approximate Gaussian kernels of different resolutions.

Figure 6 illustrates the performance of moon dataset classifiers using circuits with 10 input photons and random Fourier features of different dimensions, i.e: $R = 1, 10, 100$ and the same decision boundary as in Eq. (17). The circuit with input 10 photons can probe a range of kernel resolutions within one order of magnitude, e.g: for $\gamma = 1$, the accessible resolutions are $\sigma = \{1/n | 1 \leq n \leq 10\}$; six of these are shown in Fig. 6 to illustrate the working
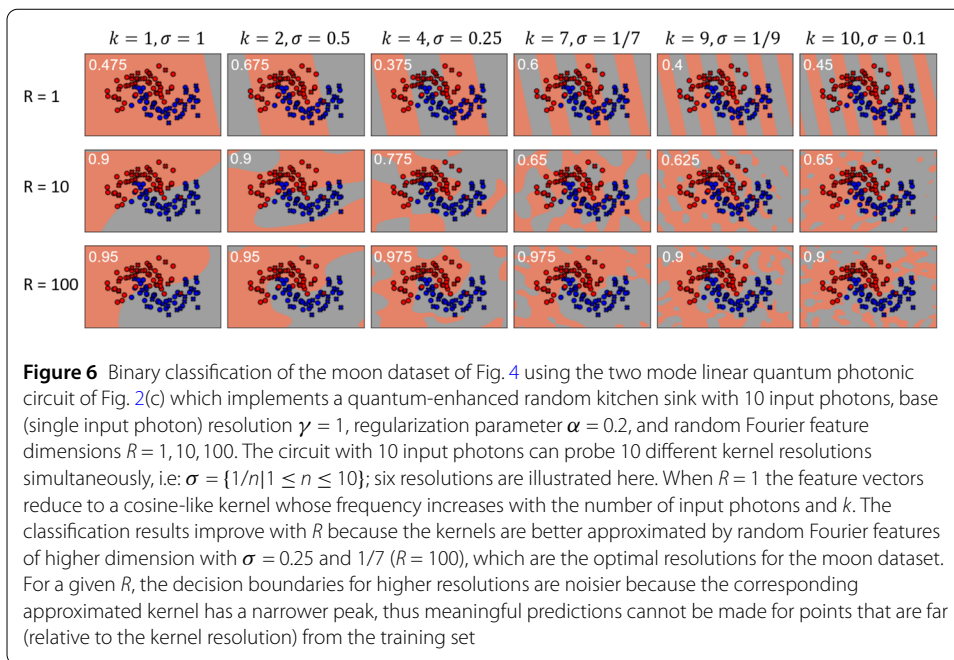
Figure 6 Binary classification of the moon dataset of Fig. 4 using the two mode linear quantum photonic circuit of Fig. 2(c) which implements a quantum-enhanced random kitchen sink with 10 input photons, base (single input photon) resolution $\gamma = 1$, regularization parameter $\alpha = 0.2$, and random Fourier feature dimensions $R = 1, 10, 100$. The circuit with 10 input photons can probe 10 different kernel resolutions simultaneously, i.e: $\sigma = \{1/n | 1 \leq n \leq 10\}$; six resolutions are illustrated here. When $R = 1$ the feature vectors reduce to a cosine-like kernel whose frequency increases with the number of input photons and k. The classification results improve with R because the kernels are better approximated by random Fourier features of higher dimension with $\sigma = 0.25$ and $1/7$ ($R = 100$), which are the optimal resolutions for the moon dataset. For a given R, the decision boundaries for higher resolutions are noisier because the corresponding approximated kernel has a narrower peak, thus meaningful predictions cannot be made for points that are far (relative to the kernel resolution) from the training set

Table 1 *Quantum resource requirements for different schemes.* The resource requirements for performing training and prediction are defined in term of the number of linear optical element (i.e. beam splitters and phase shifters) settings required, where N is the number of training data, D is the dimension of the data features, M is the number of trainable circuit and observable parameters, and R is the number of random Fourier features

Schemes	Training	Prediction
Variational methods	$O(NDM)$	D
Kernel methods	$O(N^2)$	N
Random kitchen sinks	$O(NR)$	R

principle of quantum-enhanced RKS. The decision boundary of smaller σ is considerably noisier than for larger σ. This is because the kernel with smaller resolution has a narrower peak, and hence, predictions far away from the training data points cannot be made. The random Fourier features with higher dimensionality provides a better approximation to the kernel, thus suppressing the noise around training data points while improving the classification accuracy. The optimal resolution for the moon dataset is $\sigma = 0.25$ and $1/7$ for $R = 100$.

3.4 Resource requirements for each scheme

Each of the classification methods has different strengths and limitations in terms of the resource requirements, i.e: number of distinct circuit evaluations for performing training and predictions, summarized in Table 1. Here, we are concerned only with number of distinct circuit evaluations required to perform training and prediction, since the quantum resources are much more precious than the classical resources in the NISQ era. For the variational circuit, the data features are directly encoded, but the beam splitters and phase shifters in the trainable circuit blocks need to be optimized in the training step. Hence, the training resource per optimization loop is $O(NDM)$, where N, D, and M are the number of training data, the dimension of the data features, and the number of trainable circuit and observable parameters, respectively. More Fourier frequencies can be obtained with

larger photon numbers but require larger M for universality, increasing the training time. Prediction requires reconfiguration of the D encoding phase shifters, which can however be performed in parallel.

Kernel methods, on the other hand, encode the differences between data inputs using one phase shifter and the training is outsourced to a classical computer, therefore the resources for training scale only with the number of training data, i.e: $O(N^2)$. The Gaussian kernel with different resolutions can be accessed with a fixed circuit by considering different observables; Gaussian kernels with higher resolutions are better approximated for circuits with larger numbers of input photons. In contrast to the variational methods, N different phase shifter settings are required to make predictions on new data. Random kitchen sinks have similar advantages to kernel methods, i.e: fixed circuit and different resolutions can be accessed by different observables, but have a better scaling $O(NR)$ with number of input data points, where R is the number of random features chosen. The predictions require R circuit settings regardless of the dimension of the data features.

4 Conclusion

The data-embedding process is a bottleneck which must be addressed [46] in order to fully leverage the potential of quantum machine learning algorithms. In this paper, we addressed the data encoding problem by proposing a more gate-efficient bosonic encoding method. Our method has three potential advantages. First, it allows for a more efficient data encoding by modulating all Fock basis simultaneously using only one phase shifter, regardless of the input photon number. Second, the circuits employed a kernel-like trick, where nonlinearity is outsourced to quantum feature maps, i.e: the data-encoding phase shifter that encode the classical data into the high-dimensional Fock space [13, 20], avoiding the need of the experimental hard-to-implement nonlinear optical components. Subsequently, the expressive power of the circuit can be controlled by the number of input photons, while requiring fewer encoding layers compared to the qubit-based architecture [23, 24]. Finally, the circuits can be trained to implement commonly-used kernels with well-understood properties such as the Gaussian kernel.

Even though our photonic models are inspired by the BosonSampling circuits [83], we do not expect the arguments about the BosonSampling's classical non-simulability to hold for our circuits, for three reasons: (1) The model output is expectation values, not samples. (2) Our phoronic circuits are not sampled from the Haar random distribution. (3) The assumption of $m = n^2$ is relaxed, where m is the number of optical modes and n is the number of input photons. Even so, there exist other benefits of studying the use of this class of circuit as quantum machine learning models. Quantum machine learning is still in its infancy, and it is still unclear how to rigorously define a quantum advantage for generic machine learning problems [84]. In this work, we focused on a specific problem in this field, the data-encoding problem, showing using simple quantum machine learning models how bosonic circuits may enable more efficient data uploading. We expect our conclusions to be valid for other classes of quantum machine learning models which may be hard to classically simulate. In addition, we believe our photonic models will serve as primitive quantum machine learning model [84] that inspire researchers in the field to develop other photonic quantum machine learning algorithms that possess quantum advantages. Recently, the awareness of the importance of the energy aspect of quantum algorithms has been raised [85]. Although the energy aspect of our quantum circuits is

not studied, our models could inspire an application-oriented framework to compare the energy consumption of quantum machine learning based on different platforms.
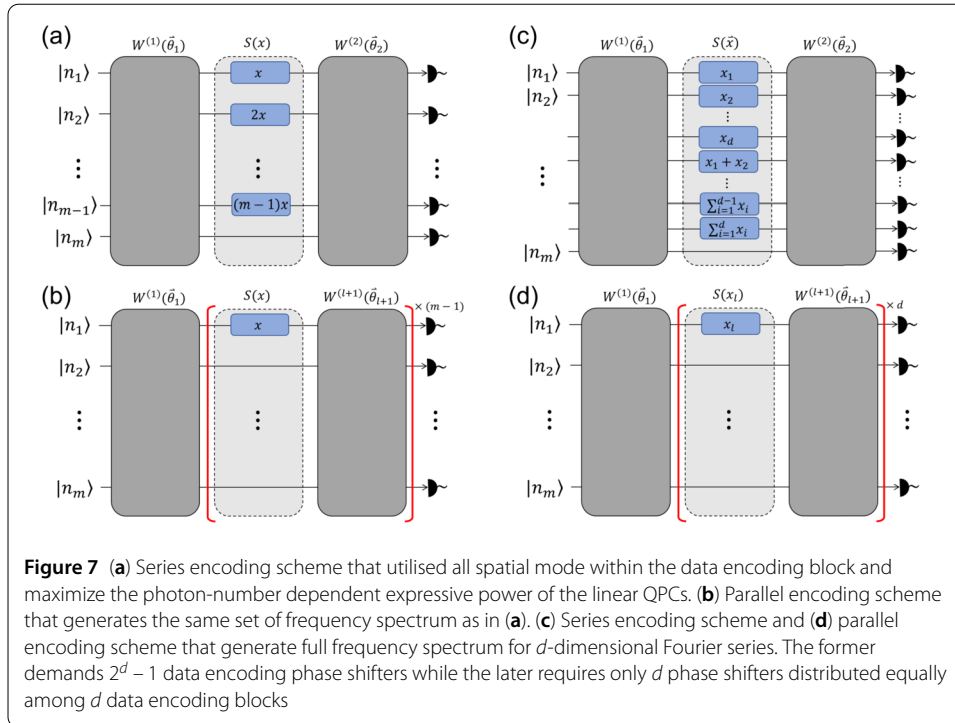
While the dimension of the Fock space grows (exponentially) with photon number and spatial modes, improving the expressive power, this is accompanied by higher sensitivity to optical losses and the need for more detection events in order to accurately sample all of the required output observables. Moreover, there exists a trade-off between the model's expressive power and ability to generalize, where circuits with higher expressive power can suffer from larger generalization errors, i.e. over-fitting [86, 87] and trainability issue [33]. One potential way to mitigates these issues is to define the quantum machine learning models in projected Fock spaces, which may lead to potential quantum advantages [73].

We proposed three different ways with different resource requirements to perform binary classification using linear quantum photonics circuits (QPCs). (1) Variational quantum classifiers that classify data points directly on the high-dimensional Fock space, while (2) and (3) implement Gaussian kernel for classical kernel machines directly or using the random kitchen sinks algorithm, sampling kernels with different resolutions in parallel. The random kitchen sink approach could be further improved by sampling the random features from a data-optimized distribution using fault-tolerant quantum computers [88, 89]. A linear QPC with three spatial modes and up to 10 input photons equipped with photon-number resolving detectors is sufficient to show a proof of concept experiment for all of the proposed approaches. Therefore, our proposed architecture can be implemented with current technology, such as integrated photonic circuits [90–92] or bulk optics [93] used for BosonSampling experiments [94]. Other experimental aspects such as the impact of the degree of multi-photon distinguishability and exponentially-scaling photon losses on expressive power are subject for future research.

While this article investigated the expressive power of the linear QPCs, the trainability and generalization power of the linear QPCs remains an open question. Apart from the gradient-free method used in this article for QML model training, gradient-based methods with analytical gradient [95–97] can potentially boost the training speed. However, current analytical gradient evaluation methods only apply to the photonic circuit with 1-photon input Fock state [98] or continuous variable quantum photonic systems [99–102]. Hence, more research needs to be done to find the analytical gradient for the quantum photonic circuits with general input Fock state, which requires the differentiation of the permanents of the transfer matrix. It will also be interesting to see the effect of different input states, i.e: coherent states and squeezed states on the expressive power, trainability, and generalization power of the linear QPCs [103]. It will be interesting to further explore the translation of ideas between classical and quantum photonic circuits for machine learning.

Appendix A: General encoding scheme for 1D frequency spectrum

In Sect. 2.2.1, we have derived the frequency spectrum for linear QPCs with single data encoding block that consists only one data encoding phase shifter. In this section, we broaden the frequency spectrum by adding more data encoding phase shifters into the same layer (series encoding) and different layers (parallel encoding). In the series encoding scheme [Fig. 7(a)], we consider linear QPCs with single data encoding block that consists of $m - 1$ phase shifters, i.e: the highest possible number of phase shifters that can be placed within

Figure 7 (**a**) Series encoding scheme that utilised all spatial mode within the data encoding block and maximize the photon-number dependent expressive power of the linear QPCs. (**b**) Parallel encoding scheme that generates the same set of frequency spectrum as in (**a**). (**c**) Series encoding scheme and (**d**) parallel encoding scheme that generate full frequency spectrum for d-dimensional Fourier series. The former demands $2^d - 1$ data encoding phase shifters while the later requires only d phase shifters distributed equally among d data encoding blocks

a data encoding block. For parallel encoding scheme [Fig. 7(b)], the $m - 1$ phase shifters are equally distributed among $m - 1$ data encoding blocks. One could consider different combinations of phase shifters in each layers and the expressive power will change accordingly.

As shown in Sect. 2.2.1, the size of the frequency spectrum of a m mode linear QPC with one data encoding phase shifter is given by $D_{(n,1,1)} = n$, where $D_{(n,L,q)}$ (two additional subscripts are added for clarity) denotes the size of frequency spectrum realizable by linear QPCs with n input photons and L data encoding blocks, each block consists of q data encoding phase shifters. For series encoding ($L = 1$), we can place one data encoding phase shifter per mode on the first $m - 1$ mode, each encodes phase proportional to its mode number [Fig. 7(a)], i.e: $i \cdot x$ phase shift with i denotes the mode number. The range of phases that could pick-up by n photon is $[0, (m - 1)n]$, where the lower (upper) bound is obtained when all photon passes through the last (second last) mode. Hence, the size of the frequency spectrum $D_{(n,1,m-1)}$ is $(m - 1)n$. Identical range of phases is also apply to the parallel encoding scheme, where the lower (upper) bound is achieved when none (all) of the photon passes through the first mode on each layer, thus, $D_{(n,m-1,1)} = (m - 1)n$.

Appendix B: Encoding scheme to generate full frequency spectrum for multi-dimensional Fourier series

In this section, we will introduce the series and parallel encoding schemes that can generate a full frequency spectrum for multi-dimensional Fourier series. For series encoding scheme [Fig. 7(c)], one would need $2^d - 1$ phase shifters to encode the positive phases of d-dimensional degree 1 Fourier series, i.e: $(\{\sum_{r=i}^{d} r_i x_i | r_1, r_2, \ldots, r_d \in \{0, 1\}\} \backslash \{0\})$. For example, one can use 7 phase shifters to encode $\{x_1, x_2, x_3, x_1 + x_2, x_1 + x_3, x_2 + x_3, x_1 + x_2 + x_3\}$. Then, the frequency spectrum of d dimensional degree n Fourier series, i.e: $\Omega_n^{(d)} = (-\boldsymbol{\omega}^{(n)}, 0, \boldsymbol{\omega}^{(n)})$

with $\boldsymbol{\omega}^{(n)} = (\omega_1^{(n)}, \omega_2^{(n)}, \ldots, \omega_d^{(n)})$ and $\omega_i^{(n)} \in \{0, 1, \ldots, n\}$ can be generated using n input photons. On the other hand, the same set of frequency spectrum can be generated using d data encoding blocks, each consists of one data encoding phase shifter that encodes one data feature. [Fig. 7(d)].

Abbreviations
RAM, random access memory; QPC, quantum photonic circuit; PNR, photon number-resolving; NISQ, noisy-intermediate scale quantum; RKS, random kitchen sinks.

Availability of data and materials
The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Declarations

Competing interests
Dimitris G. Angelakis is one of the editorial board members of EPJ Quantum Technology Journal.

Authors' contributions
BYG performed the calculations and wrote the first draft of the manuscript. DL and DGA supervised the project. All authors read and approved the final manuscript.

Author details
[1]Centre for Quantum Technologies, National University of Singapore, 3 Science Drive 2, Singapore 117543, Singapore.
[2]School of Electrical and Computer Engineering, Technical University of Crete, Chania, 73100, Greece. [3]AngelQ Quantum Computing, 531A Upper Cross Street, #04-95 Hong Lim Complex, Singapore 051531, Singapore.

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References
1. De Marinis L, Cococcioni M, Castoldi P, Andriolli N. Photonic neural networks: a survey. IEEE Access. 2019;7:175827–41. https://doi.org/10.1109/ACCESS.2019.2957245.
2. Hamerly R, Bernstein L, Sludds A, Soljačć M, Englund D. Large-scale optical neural networks based on photoelectric multiplication. Phys Rev X. 2019;9:021032. https://doi.org/10.1103/PhysRevX.9.021032.
3. Roques-Carmes C, Shen Y, Zanoci C, Prabhu M, Atieh F, Jing L, Dubček T, Mao C, Johnson MR, Čeperić V et al. Heuristic recurrent algorithms for photonic Ising machines. Nat Commun. 2020;11:249. https://doi.org/10.1038/s41467-019-14096-z.
4. Shastri BJ, Tait AN, de Lima TF, Pernice WH, Bhaskaran H, Wright CD, Prucnal PR. Photonics for artificial intelligence and neuromorphic computing. Nat Photonics. 2021;15(2):102–14. https://doi.org/10.1038/s41566-020-00754-y.
5. Peruzzo A, McClean J, Shadbolt P, Yung M-H, Zhou X-Q, Love PJ, Aspuru-Guzik A, O'Brien JL. A variational eigenvalue solver on a photonic quantum processor. Nat Commun. 2014;5:4213. https://doi.org/10.1038/ncomms5213.
6. Mitarai K, Negoro M, Kitagawa M, Fujii K. Quantum circuit learning. Phys Rev A. 2018;98:032309. https://doi.org/10.1103/PhysRevA.98.032309.
7. Benedetti M, Lloyd E, Sack S, Fiorentini M. Parameterized quantum circuits as machine learning models. Quantum Sci Technol. 2019;4(4):043001. https://doi.org/10.1088/2058-9565/ab4eb5.
8. Schuld M, Bocharov A, Svore KM, Wiebe N. Circuit-centric quantum classifiers. Phys Rev A. 2020;101:032308. https://doi.org/10.1103/PhysRevA.101.032308.
9. Fujii K, Nakajima K. Quantum reservoir computing: a reservoir approach toward quantum machine learning on near-term quantum devices. Singapore: Springer; 2021. p. 423–50. https://doi.org/10.1007/978-981-13-1687-6_18.
10. Goto T, Tran QH, Nakajima K. Universal approximation property of quantum machine learning models in quantum-enhanced feature spaces. Phys Rev Lett. 2021;127(9):090506. https://doi.org/10.1103/PhysRevLett.127.090506.
11. Lloyd S, Schuld M, Ijaz A, Izaac J, Killoran N. Quantum embeddings for machine learning. 2020. https://doi.org/10.48550/arXiv.2001.03622. arXiv:2001.03622 [quant-ph].
12. Chatterjee R, Yu T. Generalized coherent states, reproducing kernels, and quantum support vector machines. 2016. https://doi.org/10.48550/arXiv.1612.03713. arXiv:1612.03713 [quant-ph].

13. Schuld M, Killoran N. Quantum machine learning in feature Hilbert spaces. Phys Rev Lett. 2019;122:040504. https://doi.org/10.1103/PhysRevLett.122.040504.

14. Steinbrecher GR, Olson JP, Englund D, Carolan J. Quantum optical neural networks. npj Quantum Inf. 2019;5:60. https://doi.org/10.1038/s41534-019-0174-7.

15. Killoran N, Bromley TR, Arrazola JM, Schuld M, Quesada N, Lloyd S. Continuous-variable quantum neural networks. Phys Rev Res. 2019;1:033063. https://doi.org/10.1103/PhysRevResearch.1.033063.

16. Bartkiewicz K, Gneiting C, Černoch A, Jiráková K, Lemr K, Nori F. Experimental kernel-based quantum machine learning in finite feature space. Sci Rep. 2020;10:12356. https://doi.org/10.1038/s41598-020-68911-5.

17. Taballione C, van der Meer R, Snijders HJ, Hooijschuur P, Epping JP, de Goede M, Kassenberg B, Venderbosch P, Toebes C, van den Vlekkert H et al. A universal fully reconfigurable 12-mode quantum photonic processor. Mater Quantum Technol. 2021;1:035002. https://doi.org/10.1088/2633-4356/ac168c.

18. Chabaud U, Markham D, Sohbi A. Quantum machine learning with adaptive linear optics. Quantum. 2021;5:496. https://doi.org/10.22331/q-2021-07-05-496.

19. Ghobadi R. Nonclassical kernels in continuous-variable systems. Phys Rev A. 2021;104(5):052403. https://doi.org/10.1103/PhysRevA.104.052403.

20. Havlíček V, Córcoles AD, Temme K, Harrow AW, Kandala A, Chow JM, Gambetta JM. Supervised learning with quantum-enhanced feature spaces. Nature. 2019;567(7747):209–12. https://doi.org/10.1038/s41586-019-0980-2.

21. Schuld M, Petruccione F. Machine learning with quantum computers. Switzerland: Springer; 2021. https://doi.org/10.1007/978-3-030-83098-4.

22. Schuld M. Supervised quantum machine learning models are kernel methods. 2021. https://doi.org/10.48550/arXiv.2101.11020. arXiv:2101.11020 [quant-ph].

23. Pérez-Salinas A, Cervera-Lierta A, Gil-Fuster E, Latorre JI. Data re-uploading for a universal quantum classifier. Quantum. 2020;4:226. https://doi.org/10.22331/q-2020-02-06-226.

24. Schuld M, Sweke R, Meyer JJ. Effect of data encoding on the expressive power of variational quantum-machine-learning models. Phys Rev A. 2021;103:032430. https://doi.org/10.1103/PhysRevA.103.032430.

25. Pérez-Salinas A, López-Núñez D, García-Sáez A, Forn-Díaz P, Latorre JI. One qubit as a universal approximant. Phys Rev A. 2021;104(1):012405. https://doi.org/10.1103/PhysRevA.104.012405.

26. Li W, Deng D-L. Recent advances for quantum classifiers. Sci China, Phys Mech Astron. 2022;65(2):1–23. https://doi.org/10.1007/s11433-021-1793-6.

27. Dutta T, Pérez-Salinas A, Cheng JPS, Latorre JI, Mukherjee M. Realization of an ion trap quantum classifier. 2021. https://doi.org/10.48550/arXiv.2106.14059. arXiv:2106.14059 [quant-ph].

28. Kusumoto T, Mitarai K, Fujii K, Kitagawa M, Negoro M. Experimental quantum kernel trick with nuclear spins in a solid. npj Quantum Inf. 2021;7(1):1–7. https://doi.org/10.1038/s41534-021-00423-0.

29. Peters E, Caldeira J, Ho A, Leichenauer S, Mohseni M, Neven H, Spentzouris P, Strain D, Perdue GN. Machine learning of high dimensional data on a noisy quantum processor. npj Quantum Inf. 2021;7(1):1–5. https://doi.org/10.1038/s41534-021-00498-9.

30. Ren W, Li W, Xu S, Wang K, Jiang W, Jin F, Zhu X, Chen J, Song Z, Zhang P, et al. Experimental quantum adversarial learning with programmable superconducting qubits. 2022. https://doi.org/10.48550/arXiv.2204.01738. arXiv:2204.01738 [quant-ph].

31. Tangpanitanon J, Thanasilp S, Dangniam N, Lemonde M-A, Angelakis DG. Expressibility and trainability of parametrized analog quantum systems for machine learning applications. Phys Rev Res. 2020;2(4):043364. https://doi.org/10.1103/PhysRevResearch.2.043364.

32. Abbas A, Sutter D, Zoufal C, Lucchi A, Figalli A, Woerner S. The power of quantum neural networks. Nat Comput Sci. 2021;1(6):403–9. https://doi.org/10.1038/s43588-021-00084-1.

33. Holmes Z, Sharma K, Cerezo M, Coles PJ. Connecting ansatz expressibility to gradient magnitudes and barren plateaus. PRX Quantum. 2022;3(1):010313. https://doi.org/10.1103/PRXQuantum.3.010313.

34. Caro MC, Huang H-Y, Cerezo M, Sharma K, Sornborger A, Cincio L, Coles PJ. Generalization in quantum machine learning from few training data. 2021. https://doi.org/10.48550/arXiv.2111.05292. arXiv:2111.05292 [quant-ph].

35. Giovannetti V, Lloyd S, Maccone L. Quantum random access memory. Phys Rev Lett. 2008;100:160501. https://doi.org/10.1103/PhysRevLett.100.160501.

36. Harrow AW, Hassidim A, Lloyd S. Quantum algorithm for linear systems of equations. Phys Rev Lett. 2009;103:150502. https://doi.org/10.1103/PhysRevLett.103.150502.

37. Wiebe N, Braun D, Lloyd S. Quantum algorithm for data fitting. Phys Rev Lett. 2012;109:050505. https://doi.org/10.1103/PhysRevLett.109.050505.

38. Lloyd S, Mohseni M, Rebentrost P. Quantum algorithms for supervised and unsupervised machine learning. 2013. https://doi.org/10.48550/arXiv.1307.0411. arXiv:1307.0411 [quant-ph].

39. Lloyd S, Mohseni M, Rebentrost P. Quantum principal component analysis. Nat Phys. 2014;10(9):631–3. https://doi.org/10.1038/nphys3029.

40. Rebentrost P, Mohseni M, Lloyd S. Quantum support vector machine for big data classification. Phys Rev Lett. 2014;113:130503. https://doi.org/10.1103/PhysRevLett.113.130503.

41. Biamonte J, Wittek P, Pancotti N, Rebentrost P, Wiebe N, Lloyd S. Quantum machine learning. Nature. 2017;549(7671):195–202. https://doi.org/10.1038/nature23474.

42. Dunjko V, Briegel HJ. Machine learning & artificial intelligence in the quantum domain: a review of recent progress. Rep Prog Phys. 2018;81(7):074001. https://doi.org/10.1088/1361-6633/aab406.

43. Tang E. Quantum principal component analysis only achieves an exponential speedup because of its state preparation assumptions. Phys Rev Lett. 2021;127(6):060503. https://doi.org/10.1103/PhysRevLett.127.060503.

44. Cotler J, Huang H-Y, McClean JR. Revisiting dequantization and quantum advantage in learning tasks. 2021. https://doi.org/10.48550/arXiv.2112.00811. arXiv:2112.00811 [quant-ph].

45. Lloyd S, Garnerone S, Zanardi P. Quantum algorithms for topological and geometric analysis of data. Nat Commun. 2016;7(1):1–7. https://doi.org/10.1038/ncomms10138.

46. Harrow AW. Small quantum computers and large classical data sets. 2020. https://doi.org/10.48550/arXiv.2004.00026. arXiv:2004.00026 [quant-ph].

47. Liu Y, Arunachalam S, Temme K. A rigorous and robust quantum speed-up in supervised machine learning. Nat Phys. 2021;17(9):1013–7. https://doi.org/10.1038/s41567-021-01287-z.
48. Carleson L. On convergence and growth of partial sums of Fourier series. Acta Math. 1966;116(1):135–57. https://doi.org/10.1007/BF02392815.
49. Weisz F. Summability of multi-dimensional trigonometric fourier series. 2012. https://doi.org/10.48550/arXiv.1206.1789. arXiv:1206.1789 [math.CA].
50. Scheel S. Permanents in linear optical networks. 2004. https://doi.org/10.48550/arXiv.quant-ph/0406127. arXiv:quant-ph/0406127.
51. Reck M, Zeilinger A, Bernstein HJ, Bertani P. Experimental realization of any discrete unitary operator. Phys Rev Lett. 1994;73:58–61. https://doi.org/10.1103/PhysRevLett.73.58.
52. Clements WR, Humphreys PC, Metcalf BJ, Kolthammer WS, Walmsley IA. Optimal design for universal multiport interferometers. Optica. 2016;3(12):1460–5. https://doi.org/10.1364/OPTICA.3.001460.
53. Bell BA, Walmsley IA. Further compactifying linear optical unitaries. APL Photonics. 2021;6:070804. https://doi.org/10.1063/5.0053421.
54. Motes KR, Olson JP, Rabeaux EJ, Dowling JP, Olson SJ, Rohde PP. Linear optical quantum metrology with single photons: exploiting spontaneously generated entanglement to beat the shot-noise limit. Phys Rev Lett. 2015;114:170802. https://doi.org/10.1103/PhysRevLett.114.170802.
55. Olson JP, Motes KR, Birchall PM, Studer NM, LaBorde M, Moulder T, Rohde PP, Dowling JP. Linear optical quantum metrology with single photons: experimental errors, resource counting, and quantum Cramér–Rao bounds. Phys Rev A. 2017;96:013810. https://doi.org/10.1103/PhysRevA.96.013810.
56. Johnson SG. The NLopt nonlinear-optimization package. 2014.
57. Powell MJ. The BOBYQA algorithm for bound constrained optimization without derivatives. Cambridge NA Report NA2009/06. Cambridge: University of Cambridge; 2009.
58. Fox AM. Quantum optics: an introduction. vol. 15. London: Oxford University Press; 2006.
59. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay E. Scikit-learn: machine learning in python. J Mach Learn Res. 2011;12(null):2825–30.
60. McClean JR, Boixo S, Smelyanskiy VN, Babbush R, Neven H. Barren plateaus in quantum neural network training landscapes. Nat Commun. 2018;9:4812. https://doi.org/10.1038/s41467-018-07090-4.
61. Wang S, Fontana E, Cerezo M, Sharma K, Sone A, Cincio L, Coles PJ. Noise-induced barren plateaus in variational quantum algorithms. Nat Commun. 2021;12(1):1–11. https://doi.org/10.1038/s41467-021-27045-6.
62. Marrero CO, Kieferová M, Wiebe N. Entanglement-induced barren plateaus. PRX Quantum. 2021;2:040316. https://doi.org/10.1103/PRXQuantum.2.040316.
63. Bittel L, Kliesch M. Training variational quantum algorithms is NP-hard – even for logarithmically many qubits and free fermionic systems. Phys Rev Lett. 2021;127:120502. https://doi.org/10.1103/PhysRevLett.127.120502.
64. Thanasilp S, Wang S, Nghiem NA, Coles PJ, Cerezo M. Subtleties in the trainability of quantum machine learning models. 2021. https://doi.org/10.48550/arXiv.2110.14753. arXiv:2110.14753 [quant-ph].
65. Arrasmith A, Cerezo M, Czarnik P, Cincio L, Coles PJ. Effect of barren plateaus on gradient-free optimization. Quantum. 2021;5:558. https://doi.org/10.22331/q-2021-10-05-558.
66. Haug T, Self CN, Kim M. Large-scale quantum machine learning. 2021. https://doi.org/10.48550/arXiv.2108.01039. arXiv:2108.01039 [quant-ph].
67. Schölkopf B, Smola A. Learning with kernels: support vector machines, regularization, optimization, and beyond. Adaptive computation and machine learning. Cambridge: MIT Press; 2002. p. 644.
68. Hofmann T, Schölkopf B, Smola AJ. Kernel methods in machine learning. Ann Stat. 2008;36(3):1171–220.
69. Mercer J. Functions of positive and negative type and their connection with the theory of integral equations. Philos Trans R Soc Lond A. 1909;209:415–46. https://doi.org/10.1098/rsta.1909.0016.
70. Schölkopf B, Herbrich R, Smola AJ. A generalized representer theorem. In: Helmbold D, Williamson B, editors. Computational learning theory. Berlin: Springer; 2001. p. 416–26. https://doi.org/10.1007/3-540-44581-1_27.
71. Bishop CM. Pattern recognition and machine learning (information science and statistics). Berlin: Springer; 2006.
72. Theodoridis S. Machine learning: a Bayesian and optimization perspective. 1st ed. San Diego: Academic Press; 2015.
73. Huang H-Y, Broughton M, Mohseni M, Babbush R, Boixo S, Neven H, McClean JR. Power of data in quantum machine learning. Nat Commun. 2021;12:2631. https://doi.org/10.1038/s41467-021-22539-9.
74. Wang X, Du Y, Luo Y, Tao D. Towards understanding the power of quantum kernels in the NISQ era. Quantum. 2021;5:531. https://doi.org/10.22331/q-2021-08-30-531.
75. Micchelli CA, Xu Y, Zhang H. Universal kernels. J Mach Learn Res. 2006;7:2651–67.
76. Steinwart I, Christmann A. Support vector machines. 1st ed. New York: Springer; 2008. https://doi.org/10.1007/978-0-387-77242-4.
77. Rahimi A, Recht B. Random features for large-scale kernel machines. In: Platt J, Koller D, Singer Y, Roweis S, editors. Advances in neural information processing systems. vol. 20. Red Hook: Curran Associates; 2007.
78. Rahimi A, Recht B. Uniform approximation of functions with random bases. In: 2008 46th annual allerton conference on communication, control, and computing. 2008. p. 555–61. https://doi.org/10.1109/ALLERTON.2008.4797607.
79. Rahimi A, Recht B. Weighted sums of random kitchen sinks: replacing minimization with randomization in learning. In: Proceedings of the 21st international conference on neural information processing systems. NIPS'08. Red Hook: Curran Associates; 2008. p. 1313–20.
80. Wilson C, Otterbach J, Tezak N, Smith R, Polloreno A, Karalekas PJ, Heidel S, Alam MS, Crooks G, da Silva M. Quantum kitchen sinks: an algorithm for machine learning on near-term quantum computers. 2018. https://doi.org/10.48550/arXiv.1806.08321. arXiv:1806.08321 [quant-ph].
81. Noori M, Vedaie SS, Singh I, Crawford D, Oberoi JS, Sanders BC, Zahedinejad E. Analog-quantum feature mapping for machine-learning applications. Phys Rev Appl. 2020;14:034034. https://doi.org/10.1103/PhysRevApplied.14.034034.
82. Hoerl AE, Kennard RW. Ridge regression: biased estimation for nonorthogonal problems. Technometrics. 1970;12(1):55–67. https://doi.org/10.1080/00401706.1970.10488634.
83. Aaronson S, Arkhipov A. The computational complexity of linear optics. In: Proceedings of the forty-third annual ACM symposium on theory of computing. 2011. p. 333–42. https://doi.org/10.1145/1993636.1993682.

84. Schuld M, Killoran N. Is quantum advantage the right goal for quantum machine learning? 2022. https://doi.org/10.48550/arXiv.2203.01340. arXiv:2203.01340 [quant-ph].

85. Auffeves A. Quantum technologies need a quantum energy initiative. 2021. https://doi.org/10.48550/arXiv.2111.09241. arXiv:2111.09241 [quant-ph].

86. Banchi L, Pereira J, Pirandola S. Generalization in quantum machine learning: a quantum information perspective. PRX Quantum. 2021;2:040321. https://doi.org/10.1103/PRXQuantum.2.040321.

87. Caro MC, Gil-Fuster E, Meyer JJ, Eisert J, Sweke R. Encoding-dependent generalization bounds for parametrized quantum circuits. Quantum. 2021;5:582. https://doi.org/10.22331/q-2021-11-17-582.

88. Yamasaki H, Subramanian S, Sonoda S, Koashi M. Learning with optimized random features: exponential speedup by quantum machine learning without sparsity and low-rank assumptions. In: Advances in neural information processing systems. vol. 33. Red Hook: Curran Associates; 2020. p. 13674–87.

89. Yamasaki H, Sonoda S. Exponential error convergence in data classification with optimized random features: acceleration by quantum machine learning. 2021. https://doi.org/10.48550/arXiv.2106.09028. arXiv:2106.09028 [quant-ph].

90. Carolan J, Harrold C, Sparrow C, Martín-López E, Russell NJ, Silverstone JW, Shadbolt PJ, Matsuda N, Oguma M, Itoh M et al. Universal linear optics. Science. 2015;349(6249):711–6. https://doi.org/10.1126/science.aab3642.

91. Zhong H-S, Li Y, Li W, Peng L-C, Su Z-E, Hu Y, He Y-M, Ding X, Zhang W, Li H, Zhang L, Wang Z, You L, Wang X-L, Jiang X, Li L, Chen Y-A, Liu N-L, Lu C-Y, Pan J-W. 12-photon entanglement and scalable scattershot boson sampling with optimal entangled-photon pairs from parametric down-conversion. Phys Rev Lett. 2018;121:250505. https://doi.org/10.1103/PhysRevLett.121.250505.

92. Hoch F, Piacentini S, Giordani T, Tian Z-N, Iuliano M, Esposito C, Camillini A, Carvacho G, Ceccarelli F, Spagnolo N, et al. Boson sampling in a reconfigurable continuously-coupled 3d photonic circuit. 2021. https://doi.org/10.48550/arXiv.2106.08260. arXiv:2106.08260 [quant-ph].

93. Wang H, Qin J, Ding X, Chen M-C, Chen S, You X, He Y-M, Jiang X, You L, Wang Z, Schneider C, Renema JJ, Höfling S, Lu C-Y, Pan J-W. Boson sampling with 20 input photons and a 60-mode interferometer in a 10^{14}-dimensional Hilbert space. Phys Rev Lett. 2019;123:250503. https://doi.org/10.1103/PhysRevLett.123.250503.

94. Brod DJ, Galvão EF, Crespi A, Osellame R, Spagnolo N, Sciarrino F. Photonic implementation of boson sampling: a review. Adv Photonics. 2019;1(3):034001. https://doi.org/10.1117/1.AP.1.3.034001.

95. Schuld M, Bergholm V, Gogolin C, Izaac J, Killoran N. Evaluating analytic gradients on quantum hardware. Phys Rev A. 2019;99(3):032331. https://doi.org/10.1103/PhysRevA.99.032331.

96. Banchi L, Crooks GE. Measuring analytic gradients of general quantum evolution with the stochastic parameter shift rule. Quantum. 2021;5:386. https://doi.org/10.22331/q-2021-01-25-386.

97. Wierichs D, Izaac J, Wang C, Lin CY-Y. General parameter-shift rules for quantum gradients. Quantum. 2022;6:677. https://doi.org/10.22331/q-2022-03-30-677.

98. Kerenidis I, Landman J, Mathur N. Classical and quantum algorithms for orthogonal neural networks. 2021. https://doi.org/10.48550/arXiv.2106.07198. arXiv:2106.07198 [quant-ph].

99. Banchi L, Quesada N, Arrazola JM. Training Gaussian boson sampling distributions. Phys Rev A. 2020;102(1):012417. https://doi.org/10.1103/PhysRevA.102.012417.

100. Miatto FM, Quesada N. Fast optimization of parametrized quantum optical circuits. Quantum. 2020;4:366. https://doi.org/10.22331/q-2020-11-30-366.

101. Yao Y, Miatto FM. Fast differentiable evolution of quantum states under gaussian transformations. 2021. https://doi.org/10.48550/arXiv.2102.05742. arXiv:2102.05742 [quant-ph].

102. Yao Y, Cussenot P, Wolf RA, Miatto F. Complex natural gradient optimization for optical quantum circuit design. Phys Rev A. 2022;105:052402. https://doi.org/10.1103/PhysRevA.105.052402.

103. Afek I, Ambar O, Silberberg Y. High-NOON states by mixing quantum and classical light. Science. 2010;328(5980):879–81. https://doi.org/10.1126/science.1188172.