Fock State-enhanced Expressivity of Quantum Machine Learning Models

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

[1]Centre for Quantum Technologies, National University of Singapore, 3 Science Drive 2, Singapore 117543
[2]School of Electrical and Computer Engineering, Technical University of Crete, Chania, Greece 73100
*dimitris.angelakis@nus.edu.sg

I. 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 shed some light on the unique advantages offered by quantum photonics for enhancing 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.

II. Linear quantum photonic circuits (QPCs)

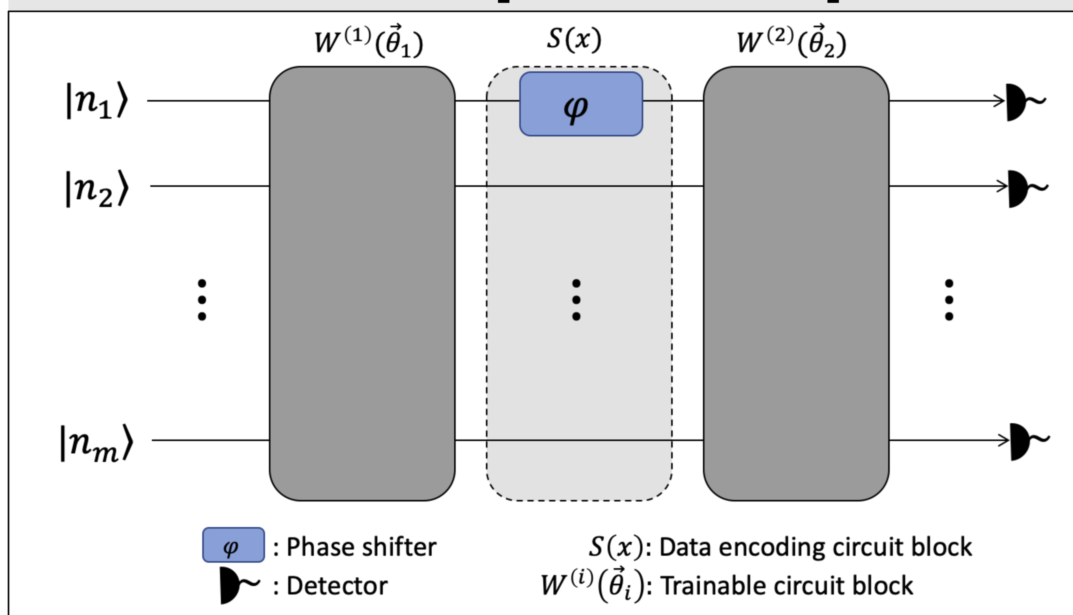


Fig 1: The parameterized linear quantum photonic circuits with m spatial modes consist of two trainable circuit blocks $W^{(i)}(\vec{\theta}_i)$ and one data encoding block $S(x)$ with a data encoding phase shifter at the top spatial mode. The multi-photon Fock states $|n_1, n_2, \ldots, n_m\rangle$ that passed through the circuits are detected by photon number resolving (PNR) or on-off detectors.

- Circuit output can be expressed as a Fourier series:

$$f^{(n)}(x, \vec{\Theta}, \vec{\lambda}) = \langle n_1, n_2, \ldots, n_m | \mathcal{U}^{\dagger}(x, \vec{\Theta}) \mathcal{M}(\vec{\lambda}) \mathcal{U}(x, \vec{\Theta}) | n_1, n_2, \ldots, n_m \rangle = \sum_{\omega \in \Omega_n} c_{\omega}(\vec{\Theta}, \vec{\lambda}) e^{-i\omega x}$$

where $\sum_i n_i = n$, $\mathcal{U}(x, \vec{\Theta}) = \mathcal{W}^{(2)}(\vec{\theta}_2) \mathcal{S}(x) \mathcal{W}^{(1)}(\vec{\theta}_1)$ with $\vec{\Theta} = (\vec{\theta}_1, \vec{\theta}_2)$, and $\mathcal{M}(\vec{\lambda})$ is a parameterized observable.

- The linear QPCs can be trained to perform function fitting, binary classification, and kernel approximation for machine learning (ML) algorithms (e.g: support vector machines) using its output observables.
- Study the expressivity of linear QPCs through
  - Frequency spectrum, Ω_n
  - Fourier coefficients, $\{c_{\omega}(\vec{\Theta}, \vec{\lambda})\}$

III. Expressivity of linear QPCs

Frequency spectrum, Ω_n:

- Informs us about which functions the quantum circuits can approximate.
- Depends only on the input photon number $\Omega_n = \{-n, -n+1, \ldots, n+1, n\}$
- Quantum circuits with higher input photon number are more expressive.

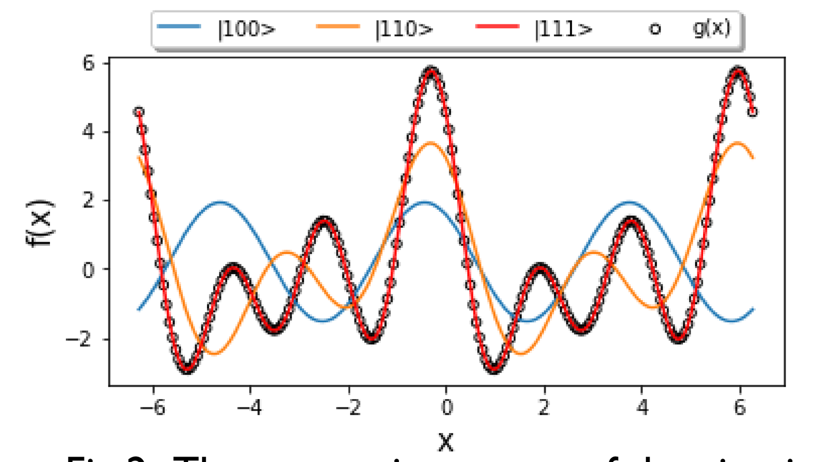


Fig 2: The expressive power of the circuits grows with the number of photons. For example, perfect fitting of a degree 3 Fourier series $g(x)$ is achieved using a 3-mode linear QPC with 3 input photons.

Fourier coefficients, $\{c_{\omega}(\vec{\Theta}, \vec{\lambda})\}$

- Determine how the accessible functions can be combined.
- Depend on the entire circuit, including observable parameters.
- Require at least $M_{min} = 2n + 1$ degrees of freedom to arbitrarily control n complex and one real Fourier coefficients.

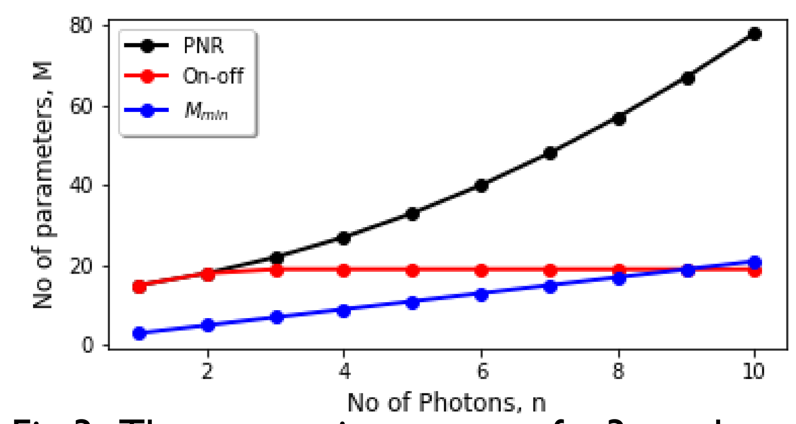


Fig 3: The expressive power of a 3-mode linear QPCs with on-off detectors (red) is only enhanced up to 9 input photons while the circuits with PNR detectors (black) may be trained to fit arbitrarily large frequencies by increasing the number of input photons.

IV. Classification using linear QPCs

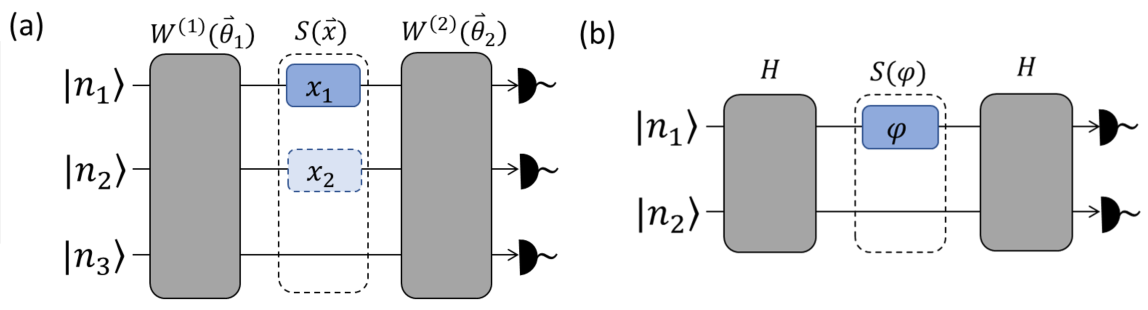


Fig 4: Linear QPCs with (a) 3 spatial modes for binary classification of two-dimensional input data. Each dimension of input data is encoded using one phase shifter and the class of data points is assigned by the sign of the circuit's output observables. (b) 2 spatial modes with fixed $W^{(i)}$ (50:50 beam splitters) for quantum-assisted kernel methods (Data encoded: Distances between two data points, $\varphi = \delta = (\vec{x}_i - \vec{x}_j)^2$ [See ②]) and quantum-enhanced random kitchen sinks (Data encoded: $\varphi = x_{r,i} = \gamma(\vec{w}_r \cdot \vec{x}_i + b_r)$ [See ③]). Classification is performed by input the approximated kernel values to kernel machines.

① Variational Quantum Classifier (VQC)

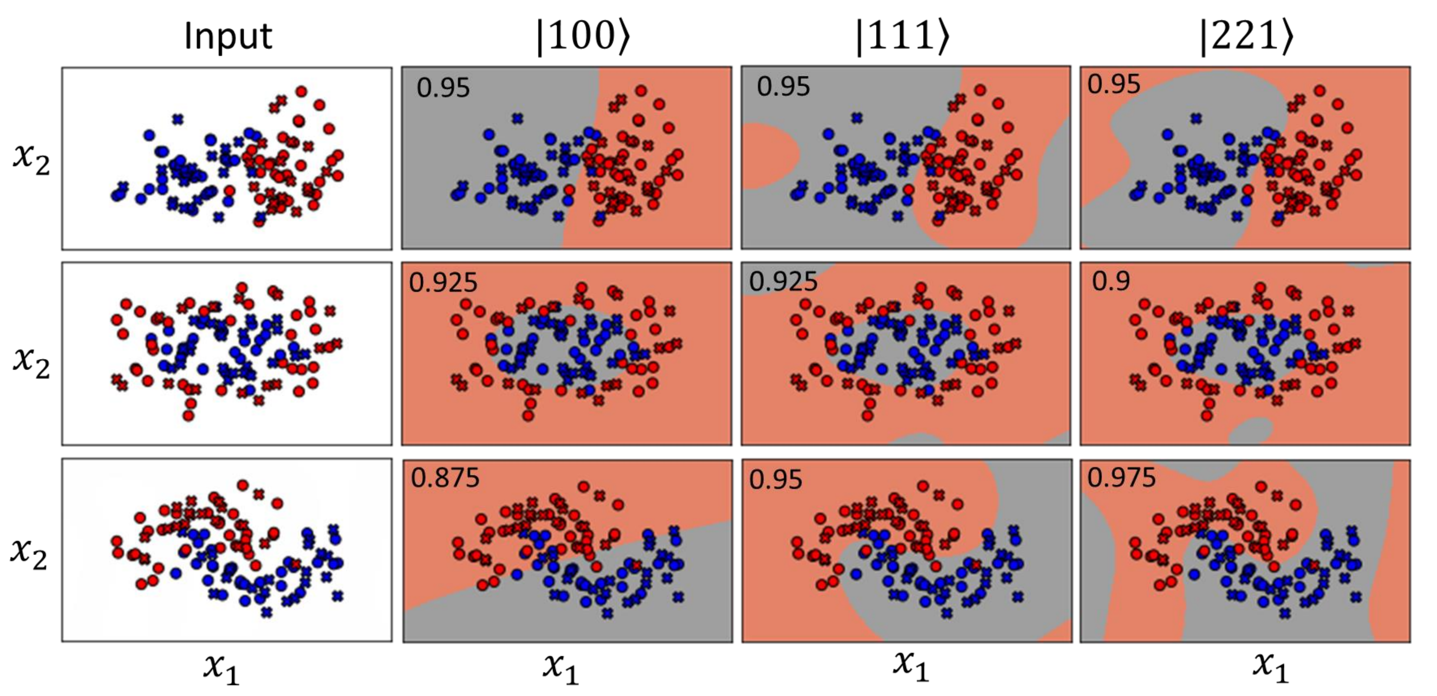
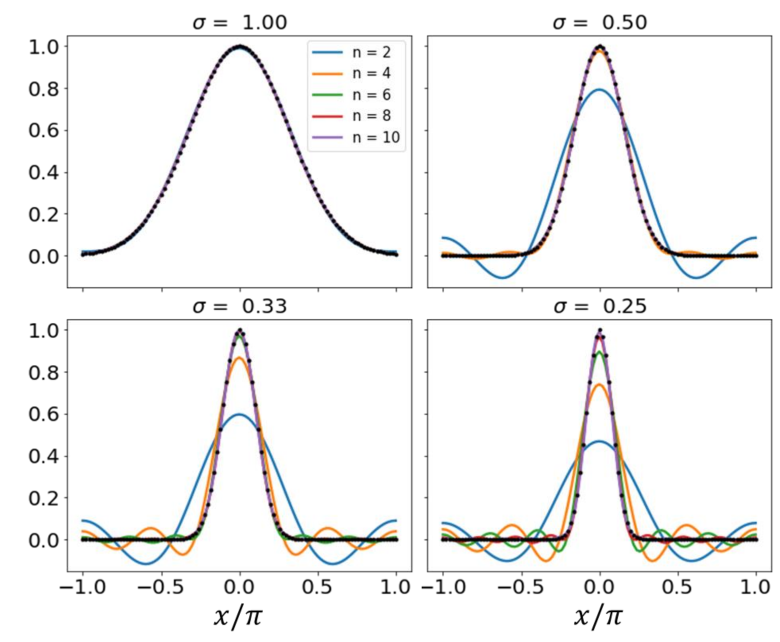


Fig 5: The training results on 3 different datasets, i.e: linear (first row), circle (second row) and moon (third row) datasets of 60 samples (red and blue solid circles) for different input photon numbers, i.e: one, three and five photons. The performance on test set (red and blue solid crosses) of 40 points is given in the upper left corner of each respective subplot. The contours show that the classification boundaries become more complicated with more input photons.

- Linear dataset: Can be separated by a linear decision boundary, hence, easily classified by quantum circuits with one photon.
- Circle dataset: The decision boundary is too constrained – high expressivity might lead to overfitting.
- Moon dataset: Performance improved with input photon number – a clear illustration of the impact of a higher expressive power on classification tasks.

② Quantum-assisted kernel methods



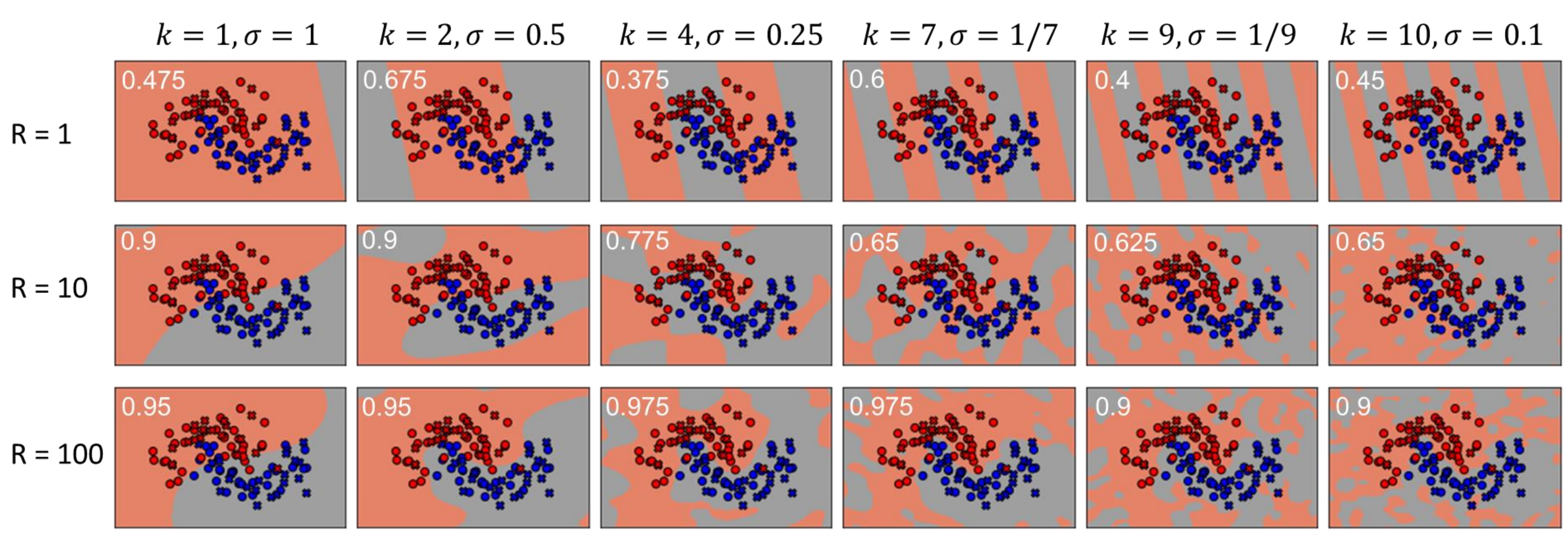
- Approximate Gaussian kernels with resolution σ

$$f^{(n)}(\delta, \vec{\lambda}^{(\sigma)}) \approx e^{\delta/2\sigma^2} = k(\vec{x}_i, \vec{x}_j)$$

using circuit output of Fig.4(b) with $\delta = (\vec{x}_i - \vec{x}_j)^2$.

- Kernels with finer resolution require more photon to approximate, i.e: circuit with 4 photons can fit kernels with $\sigma = 0.50$, but not $\sigma = 0.33/0.25$.

③ Quantum-enhanced Random Kitchen Sinks (RKS)



- Approximate Gaussian kernels with resolution $\sigma = 1/k\gamma$ by constructing random Fourier feature of different frequencies k

$$\vec{z}(\vec{x}_i) \cdot \vec{z}(\vec{x}_j) = e^{(\vec{x}_i - \vec{x}_j)^2/2\sigma^2}$$

with $\vec{z}(\vec{x}_i) = \left(f^{(n)}(x_{1,i}, \vec{\lambda}^{(k)}) \quad \cdots \quad f^{(n)}(x_{R,i}, \vec{\lambda}^{(k)})\right)^T / \sqrt{R}$ where $f^{(n)}(x_{r,i}, \vec{\lambda}^{(k)}) = \sqrt{2}\cos(k\gamma[\vec{w}_r \cdot \vec{x}_i + b_r])$, $\vec{w} \sim \mathcal{N}_D(0, I)$ and $b \sim \text{Uniform}(0, 2\pi)$.

- The circuit can access Gaussian kernel of different resolutions simultaneously by considering different observables $M(\vec{\lambda}^{(k)})$.
- Classify moon datasets using circuit with 10 input photons (six resolutions are illustrated here) and different dimensions of $\vec{z}$ $(R = 1, 10, 100)$; Optimal resolution: $\sigma = 0.25, 1/7$ $(R = 100)$.

V. Reference

Gan, B. Y. et al (2021). arXiv:2107.05224.
Schuld, M. et al. Phys. Rev. A 103 (3), 032430 (2020).

VI. Future works

- Consider different input states, i.e: coherent states and squeezed states.
- Study trainability and generalization power of linear QPCs.