DISSERTATION | DOCTORAL THESIS

Titel | Title

Experimental Photonic Quantum Machine Learning

verfasst von | submitted by

Zhenghao Yin

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of

Doktor der Naturwissenschaften (Dr.rer.nat.)

Wien | Vienna, 2025

Studienkennzahl lt. Studienblatt | Degree programme code as it appears on the student record sheet:	UA 796 605 411
Dissertationsgebiet lt. Studienblatt | Field of study as it appears on the student record sheet:	Physik
Betreut von | Supervisor:	Univ.-Prof. Dipl.-Ing. Dr. Philip Walther

Abstract

In the noisy intermediate-scale quantum era, quantum machine learning demonstrates significant potential advantages in many quantum applications, such as solving linear equations, variational quantum algorithms, and quantum chemistry. Currently, most quantum machine learning schemes rely on universal quantum computing architectures, while research on photonic systems remains relatively limited. In contrast, photons offer advantages such as long coherence times, room-temperature operation, and large-scale quantum interference, as showcased in boson sampling experiments. However, how to utilize single photons for scalable and robust quantum machine learning tasks remains an open question.

Building on this foundation, and inspired by quantum interference and photon indistinguishability, we introduce a quantum kernel method and analyze scaling laws in photonic quantum machine learning. The quantum kernel method encodes classical data into high-dimensional photon number states, thereby expanding the feature space and enhancing data representation capabilities. We also benchmark the quantum kernel method against various classical kernel approaches and experimentally verify quantum-enhanced performance on representative datasets. Furthermore, we theoretically establish a photon number scaling law, showing that increasing the photon number improves learning capacity and data efficiency in learning tasks. This result is preliminarily validated through experiments on unitary matrix learning and metric learning. Finally, we investigate the application of photonic quantum machine learning to natural language processing and discuss the prospects of quantum edge computing, highlighting their significance for future space-based applications.

In summary, the work presented in this thesis provides a theoretical foundation and experimental evidence for the scalable implementation of photonic quantum machine learning, revealing its potential advantages and application prospects in practical tasks.

Kurzfassung

Im Zeitalter der noisy intermediate-scale quantum Ära (NISQ) zeigt quantum machine learning bedeutende potenzielle Vorteile in vielen Quantenanwendungen, wie der Lösung linearen Gleichungen, variationalen Quantenalgorithmen und Quantenchemie. Derzeit stützen sich die meisten Verfahren des quantenmaschinellen Lernens auf universelle Quantencomputer-Architekturen, während die Forschung zu photonischen Systemen relativ begrenzt bleibt. Im Gegensatz dazu bieten Photonen Vorteile wie lange Kohärenzzeiten, Raumtemperaturbetrieb und großskalige Quanteninterferenz, wie in Bosonensampling-Experimenten gezeigt. Wie jedoch Einzelphotonen für skalierbare und robuste quantenmaschinelle Lernaufgaben genutzt werden können, bleibt weiterhin eine offene Frage.

Aufbauend auf dieser Grundlage und inspiriert von Quanteninterferenz und Photonen-Ununterscheidbarkeit führen wir eine Quanten-Kernel-Methode ein und analysieren Skalierungsgesetze im photonischen quantenmaschinellen Lernen. Die Quanten-Kernel-Methode kodiert klassische Daten in hochdimensionale Fock-Zustand (auch Photonenzahl genannt) und erweitert dadurch den Merkmalsraum sowie die Datenrepräsentationsfähigkeiten. Wir vergleichen die Quanten-Kernel-Methode auch mit verschiedenen klassischen Kernel-Ansätzen und zeigen experimentell, dass es mit der ersten Methode einen Leistungsboost bei den repräsentativen Datensätze gibt. Darüber hinaus etablieren wir hier ein theoretisches Photonenzahl-Skalierungsgesetz, welches zeigt, dass die Erhöhung der Photonenzahl die Lernkapazität und Dateneffizienz in Lernaufgaben verbessert. Dieses Ergebnis wird vorläufig durch Experimente zum Lernen unitärer Matrizen und Metriklernen validiert. Schließlich untersuchen wir die Anwendung des photonischen quantenmaschinellen Lernens auf die natürliche Sprachverarbeitung und diskutieren die Aussichten des Quanten-Edge-Computing, wobei wir ihre Bedeutung für zukünftige weltraumbasierte Anwendungen hervorheben.

Zusammenfassend liefert die in dieser Arbeit präsentierte Forschung eine theoretische Grundlage und experimentelle Evidenz für die skalierbare Implementierung des photonischen quantum machine learning und offenbart dessen potenzielle Vorteile und Anwendungsaussichten in praktischen Aufgaben.

Contents

List of Figures

1 Introduction

The Church-Turing thesis provided a foundational answer to the general computability problem, ushering in the era of modern computer science and technology [1, 2]. Turing's thesis introduced the concept of the Turing machinea theoretical device that uses a paper tape to represent the machine's state and performs computations by applying a sequence of operations over time. This abstraction laid the groundwork for nearly a century of progress in computation, spanning from tape-based machines to the ENIAC, from the invention of the transistor to the microprocessor, and ultimately from classical computing to quantum computing.

For decades, the advancement and scalability of classical computing have been driven by Moore's Law, which predicts the exponential growth in the number of transistors on integrated circuits. Today, the semiconductor industry routinely fabricates chips containing hundreds of trillions of transistors, a number that continues to rise. However, as cutting-edge complementary metal-oxide-semiconductor (CMOS) technology approaches the nanometer scale [3, 4], the end of Moore's Law is becoming inevitable. At these scales, quantum effects such as electron tunneling undermine the on-off mechanism of transistors, threatening the reliability of classical devices.

Moreover, shrinking integrated circuits leads to increased power consumption [5]. As components reach nanometer dimensions, leakage currents become a significant contributor to power dissipation, complicating heat management and device reliability. Ultimately, the benefits of scaling will diminish, and classical computing will face fundamental limitations in hardware improvement.

Photonic quantum computing

Although developed independently across different continents, the two theoriesquantum mechanics and computation theoryhave had successes in their respective fields. Almost during the same century and somewhat earlier, quantum mechanics was developed to solve micro-scale problems such as the hydrogen atom and black-body radiation. In later years, it has been widely accepted and adopted in modern physics research. Based on quantum mechanics, in the 20th century, knowing the statistical distribution and natural behavior of particles like photons and electrons, physicists and engineers have invented many technologies, such as the laser, the transistor, the superconductor, and so on. Nevertheless, quantum mechanics not only provides the theoretical toolbox for understanding tiny particles and devices, but also offers a new computing

paradigm, which shows exponential speedup in some specific tasks. Approaching the end of the aforementioned Moore's Law, interdisciplinary researchers have been seeking new computing machines and quantum computing is one of the most promising candidates.

Quantum computing emerged to solve these problems through its quantum properties, such as superposition and entanglement [6]. Rather than classical bits which only record bit 0 or 1, quantum bits (qubits) can be in a superposition of state $|0\rangle$ or state $|1\rangle$. This can be implemented by using the ground and excited states of an atom, or any other artificial two-level systems. In this way, considering the information entropy, a single qubit can encode more than one bit of information, and multiple qubits can then encode an exponential amount of information. From the 1990s onwards, the quantum computing community has been growing rapidly, and many quantum algorithms have been proposed to solve the problems that are hard for classical computers. Overall, the two most famous algorithms in the beginning of quantum computing theory are Shor's algorithm [7] and Grover's algorithm [8], which can factorize a number and search an unsorted database in polynomial time, respectively.

From the experimental perspective, to realize and manipulate these quantum properties, physicists have developed various quantum computing platforms, such as photons [9, 10], trapped ions [11–13], superconducting qubits [14–17], spin qubits [18–21], and others. In each of these platforms, the most challenging aspect is maintaining the coherence of physical qubits, so that quantum information can be preserved, controlled and measured. Unfortunately, qubit coherence is very fragile and easily disturbed by environmental noise, which causes decoherence of the quantum information.

So far, all these platforms feature different properties and advantages, and have demonstrated the capability for universal quantum computing. Among all quantum computing platforms, photonic quantum computing features long coherence times and room-temperature operation, and has demonstrated the capability for universal quantum computing using linear optics and single photons [9, 10]. Photonic systems pioneered quantum information processing, first demonstrating violation of Bell's inequality in the 1970s [22] and quantum teleportation in the 1990s [23], which paved the way for basic quantum information protocols. Considering the degrees of freedom (DoF) of photons, researchers have also proposed every possible qubit encoding, such as polarization, time-bin, spatial mode, orbital angular momentum, frequency, and others. Although all these degrees of freedom exhibit different physical phenomena, they are all based on photon interference and linear optics.

Meanwhile, Aaronson and Arkhipov proposed the boson sampling problem as a demonstrative task of quantum supremacy, which means humans cannot simulate this problem using classical computers if the problem scale is large enough [24]. Afterwards, groups around the world have experimentally realized the boson sampling problem using photons [25–27] and the Gaussian boson

sampling problem using squeezed states. Finally, the Gaussian boson sampling experiment showing quantum supremacy was demonstrated by the USTC group in China [28], which used 100 optical modes and up to 76 photons, with a counting rate of more than 10 million per second. Along with the random circuit sampling problem, which is usually implemented using superconducting qubits in a universal quantum circuit, researchers have shown that quantum machines can outperform classical computers for the first time in 2019.

Photonic integrated circuits

Meanwhile, from the development of electronic devices mentioned above, essentially the evolution of integrated circuits, the lesson we have learned is that Moore's Law is not only about scaling, but also about device integration. Considering conventional photonic devices, such as single photon sources, beam splitters, electro-optic modulators, and photon detectors, they are all bulky and suffer from loss and misalignment. For example, in the latest large-scale Gaussian boson sampling experiment [28] and related experiments [29, 30], the optical components including photon sources and detectors are all bulky optics, which are expensive and difficult to align. Integrated photonic circuits (PICs) provide a solution to these problems by integrating all optical components on a single chip, where light can be trapped and manipulated in waveguides. To fabricate such PICs, the most conventional method is to use CMOS technology, which is available and well-commercialized for conventional electronic integrated circuits. Typical PICs use silicon as the main material and feature millimeter footprints, making the scaling of optical modes much easier than with bulky optics.

Therefore, there is a natural tendency to implement photon-based quantum computing on PICs, from basic quantum gates to universal quantum circuits [31–34]. Recent research has shown that PICs can implement various protocols such as circuit-based quantum computing [35, 36] and measurement-based quantum computing [37–39]. The largest PIC so far contains 32 optical modes and generated 4 photons across all modes [39], showcasing the capability for generating large cluster states for measurement-based quantum computing. In addition, PICs can also serve as transmitters for quantum communication, such as quantum key distribution (QKD) and quantum teleportation [40–44]. For boson sampling, earlier experiments using PICs have demonstrated few modes and few photons, as coupling losses have limited the counting rate [26, 27, 45]. However, bulky optics have not been able to reach the full functionality of a universal interferometer network, and PICs remain very promising for implementing boson sampling within intermediate-scale devices.

Quantum machine learning

Nevertheless, it remains challenging to realize large-scale universal quantum computers for all physical systems, not only to build more qubits, but also to keep all qubits coherent and coupled. In 2019, Preskill proposed the noisy intermediate-scale quantum (NISQ) era [46], which means current quantum devices are not error-corrected and contain only dozens to hundreds of qubits. In this era, quantum machine learning (QML) is one of the most promising applications, combining the power of quantum computing with machine learning theory [47–49].

Meanwhile, classical photonic systems also demonstrate capability in machine learning tasks, such as photonic neural networks [50–54]. In these systems, photonics benefits from high-speed processing and low power consumption. In particular, matrix-vector multiplication, which is the core operation in neural network training, can be implemented using linear optics. Here, linear optics are used to implement programmable matrix multiplication, which was first proposed in 1989 by Reck et al. [55] and extended by Clements et al. [56].

Inspired by these efforts, research has begun deploying PICs to implement quantum machine learning (QML) algorithms, representing a novel combination of photonic quantum computing and machine learning advantages. For example, considering photon number states, the Hilbert space scales polynomially with photon number and optical modes, which is much more favorable than qubit-based quantum computing. In this way, one can encode classical data into high-dimensional quantum states, thereby enhancing data representation capability and quantum circuit expressivity. Moreover, the flying qubit nature of photons enables quantum machine learning implementations that are faster than conventional systems, without requiring cryogenic systems or high power consumption.

In Vienna, our team has demonstrated several pioneering works in photonic quantum machine learning. The first quantum reinforcement learning (RL) algorithm was implemented on PICs [57], which utilizes programmable Mach-Zehnder interferometers to implement quantum gates and trains the circuit to either gain reward or not. In this experiment, the quantum advantage is associated with Grover's search algorithm. Another example is the quantum memristor [58]. A classical memristor is an electronic device that can store the history of current, showing non-linear and non-volatile properties. The photonic memristor behaves similarly, as one can tune photon transmission using previous photon flux via photon counting thus showing memory effects.

Another direction that has recently been considered as part of quantum machine learning is variational quantum algorithms (VQAs). First demonstrated on photonic systems [59], VQAs leverage the principles of quantum superposition and entanglement to optimize complex functions, making them particularly suitable for tasks like feature selection and model training in machine learning [60, 61].

In summary, this dissertation targets to explore the interdisciplinary field of photonic quantum machine learning (PQML), which lies at the intersection of quantum computing, machine learning, and photonics.

As illustrated in Fig. 1.1, the overlap of any two of these three domains has already led to innovative approaches: quantum machine learning leverages quantum computing to enhance machine learning tasks in the noisy intermediate-scale quantum (NISQ) era; photonic quantum computing exploits the unique properties of photons, such as multiple degrees of freedom and long coherence times, to implement quantum algorithms; and photonic neural networks integrate neural network architectures within photonic circuits, enabling efficient matrix-vector multiplication, a core operation in neural network training.

PQML aims to realize quantum machine learning on photonic platforms, particularly PICs. This thesis is organized into three parts: Part I introduces foundational concepts and motivation; Part II presents rigorous theoretical protocols and their experimental demonstrations; and Part III discusses specific practical applications of PQML.

Part I comprises four introductory chapters. These chapters provide the motivation for this research, covering the background of linear optics quantum computing (LOQC), an overview of photonic integrated circuits, and foundational concepts in quantum machine learning.

Chapter 1 reviews the historical development of quantum computing, photonics, and machine learning, highlighting their convergence and the emergence of PQML as a new interdisciplinary field. Chapter 2 delves into quantum computing, with a focus on photonic platforms. It introduces the circuit model, the principles of LOQC, and the boson sampling problem, which demonstrates quantum advantage over classical computation. Chapter 3 discusses photonic integrated circuits (PICs), emphasizing their scalability, integration capabilities, and advantages over bulk optics for quantum machine learning applications. It also addresses the challenges of implementing and calibrating arbitrary linear optical circuits. Chapter 4 introduces the basic concepts of quantum machine learning, including both classical and quantum approaches. It then presents the theoretical framework for machine learning and parameterized quantum circuits (PQCs) as a universal model for quantum machine learning.

Part II focuses on two general and rigorous PQML frameworks. Chapter 5 benchmarks the performance of kernel-based PQML, demonstrating the advantage of using indistinguishable photons in machine learning tasks. This approach highlights quantum advantages in data expressivity and classification performance. Chapter 6 investigates the scaling properties of PQML. By introducing multiple photon states into linear optical circuits, it explores how this leads to polynomial scaling of the accessible Hilbert space and enhances learning capacity for complex machine learning problems. We showcased two tasks, unitary learning and metric learning, both of which are demonstrated

theoretically and experimentally.

Part III highlights specific PQML applications, recognizing that quantum advantage is not universal across all machine learning tasks. Chapter 7 presents the first experimental demonstration of quantum natural language processing (QNLP) using photonic systems, bridging quantum technologies and practical machine learning. Chapter 8 offers a forward-looking perspective on PQML, including the deployment of quantum machine learning payloads in space, with a payload developed during this project that was launched in the summer of 2025 via a D-Orbit ION carrier.

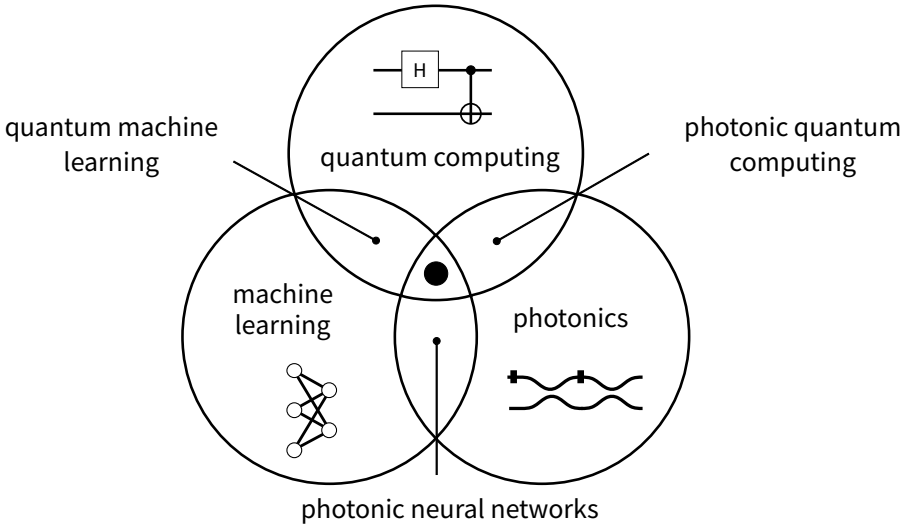

Figure 1.1: The core concept of photonic quantum machine learning. Photonic quantum machine learning (PQML) lies at the intersection of quantum computing, machine learning, and photonics. Each pairwise overlap has led to innovations: quantum machine learning uses quantum computers for advanced learning tasks; photonic quantum computing exploits photons for quantum algorithms; and photonic neural networks implement efficient matrix operations in photonic circuits. PQML unites these fields to enable quantum-enhanced machine learning.

2 Photonic Quantum Computing

In this chapter, we introduce the fundamental theory of linear optical quantum computing and boson sampling. We begin by reviewing the universal quantum computing model, focusing on qubits and quantum gates. Next, we present the linear optics quantum computing framework, where quantum information is encoded in photons and manipulated using linear optical elements. We then discuss measurement-based quantum computing, an alternative paradigm particularly well-suited to photonic systems. Finally, we explore the boson sampling problem, which generalizes linear optical systems for quantum computing and highlights the computational complexity of simulating bosonic systems.

2.1 Universal quantum computing

In quantum computing, the qubit is the fundamental unit as it exists in a superposition of only two states, $|\psi\rangle = \alpha\,|0\rangle + \beta\,|1\rangle$, where α and β are complex numbers normalized to unity. Compared with classical bits that represent either 0 or 1, qubits can exist in a superposition of both *basis states*. Considering the degrees of freedom of a single qubit: first, both α and β are complex numbers. Second, the normalization constraint reduces one degree of freedom. Finally, the global phase factor common to both α and β can be factored out, removing another degree of freedom. Thus, a single qubit has only two true degrees of freedom. Mathematically, it is common to represent a qubit state as a point on the Bloch sphere, which is a unit sphere:

$$|\psi\rangle = \cos(\theta/2)\,|0\rangle + e^{i\phi}\sin(\theta/2)\,|1\rangle \tag{2.1}$$

where θ is the polar angle and ϕ is the azimuthal angle. Note that in principle, both θ and ϕ are in the range $[0, 2\pi)$, but in practice, θ is usually restricted to $[0, \pi]$ and ϕ to $[0, 2\pi)$. For $\theta > \pi$, the state is equivalent to replacing θ with $2\pi - \theta$ and ϕ with $\phi + \pi$, as illustrated clearly in Fig. 2.1.

In quantum computing, another way to describe qubits is through the *density matrix*. For pure states, the density matrix is given by the outer product of the qubit state, $\rho = |\psi\rangle\langle\psi|$. For mixed states, the density matrix is given by a classical probability distribution, $\rho = \sum_i p_i\,|\psi_i\rangle\langle\psi_i|$, where p_i is the probability of state $|\psi_i\rangle$. The density matrix is Hermitian, positive semi-definite, and trace-preserving, which means $\mathrm{Tr}(\rho) = 1$:

$$\rho = \rho^\dagger, \quad \forall x,\ x^\dagger \rho x \geqslant 0, \quad \mathrm{Tr}(\rho) = 1. \tag{2.2}$$

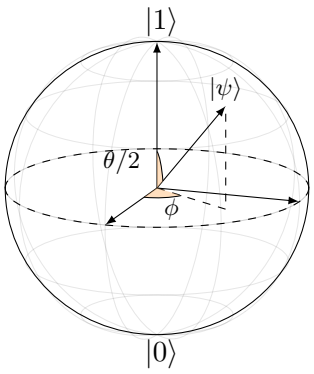

Figure 2.1: The Bloch sphere representation of a single qubit state $|\psi\rangle = \cos(\theta/2)\,|0\rangle + e^{i\phi}\sin(\theta/2)\,|1\rangle$. Here, θ is the polar angle and ϕ is the azimuthal angle.

For multiple qubits, the density matrix is given by the tensor product of the individual qubit density matrices, $\rho = \rho_1 \otimes \rho_2 \otimes \cdots \otimes \rho_n$. Regarding the positive semi-definite property, there exists a spectral decomposition of the density matrix, $\rho = \sum_i \lambda_i \,|\varphi_i\rangle\,\langle\varphi_i|$, where λ_i are the eigenvalues and $\lambda_i \geqslant 0$.

To perform quantum computing, one needs quantum gates to manipulate qubit states, such as Clifford gates and T-gates. The Clifford gates comprise a set of quantum gates that includes the Hadamard gate, the X-rotation gate, the Y-rotation gate, the Z-rotation gate, and the CNOT gate. In the single-qubit case, local gates like the Hadamard gate and rotation gates correspond to reflections or rotations of the qubit vector on the Bloch sphere. The Gottesman-Knill theorem proved that any quantum circuit can be decomposed into a sequence of Clifford gates and T-gates, which constitutes a universal gate set [62]. In this way, the topology and parameters of a quantum circuit define a specific quantum algorithm or quantum process. Hence, a complete-positive trace-preserving (CPTP) map is a quantum process that can be represented by a quantum circuit:

$$\mathcal{E}(\rho) = \sum_i K_i \rho K_i^\dagger, \quad \sum_i K_i^\dagger K_i = I \tag{2.3}$$

In the Schrödinger picture, the CPTP map is a unitary matrix U that acts on the qubit state $|\psi\rangle$:

$$\mathcal{E}(\rho) = U\rho U^\dagger \tag{2.4}$$

For applications like quantum simulation, which studies quantum processes usually represented by a Hamiltonian $\mathcal{H}$, the Hamiltonian serves as the Lie algebra generator of the unitary matrix U in group theory:

$$U = e^{-i\mathcal{H}t} \tag{2.5}$$

2.2 Linear optics quantum computing

To implement quantum computing in optical systems, we must first identify a two-level system. Like many other physical systems, photons can be superposed across two or more modes. Considering only one photon superposed across two modes, it is natural to represent the basis of a single qubit. The question then becomes finding the proper optical modes to operate any $\mathrm{SU}(2)$ unitary transformation. Fortunately, there are many natural optical modes that can be used, such as the polarization of a single photon or the path in a Mach-Zehnder interferometer (MZI). More generally, qubits can be represented by frequency, time-bin, and orbital angular momentum. Next, we need to ensure that the basis is unbiased. Polarization is a good example, as the operation of half-waveplates and quarter-waveplates can transform the quantum state into other orthogonal bases, which correspond to the rotation-X and rotation-Y operators. The same principle applies to path coding. In the following, we focus on path coding of a single photon across two optical modes, which is called dual-rail encoding [9].

For single qubit gates, an MZI with two phase shifters is equivalent to one Hadamard gate, one Z-rotation gate with phase θ, another Hadamard gate, and another Z-rotation gate with phase ϕ [63]. [1]

$$\begin{aligned}\mathrm{MZI}(\theta,\phi) &= \begin{pmatrix}1 & i\\ i & 1\end{pmatrix}\begin{pmatrix}e^{i\theta} & 0\\ 0 & 1\end{pmatrix}\begin{pmatrix}1 & i\\ i & 1\end{pmatrix}\frac{1}{2}\begin{pmatrix}e^{i\phi} & 0\\ 0 & 1\end{pmatrix} \\ &= ie^{i\theta/2}\begin{pmatrix}e^{i\phi}\sin\theta/2 & \cos\theta/2\\ e^{i\phi}\cos\theta/2 & -\sin\theta/2\end{pmatrix}\end{aligned} \tag{2.6}$$

If we consider the product of the rightmost three matrices, the product is equivalent to an X-rotation gate with phase θ. Thus, the total product of the four matrices composes a Z-rotation gate with phase ϕ and an X-rotation gate with phase θ, up to a global phase factor $ie^{i\theta/2}$.

However, the main challenge in realizing photonic quantum computing is the implementation of two-qubit gates, typically the controlled-NOT (CNOT) gate. Here, either the control qubit or the target qubit can perform single qubit gates on the corresponding optical modes, but they lack the photon-photon interaction required to perform the following operations: $|00\rangle \to |00\rangle$, $|01\rangle \to |01\rangle$, $|10\rangle \to |11\rangle$, and $|11\rangle \to |10\rangle$. Alternatively, we can use creation operators to describe the process, which is more intuitive in the Fock space:

$$a_1^\dagger a_2^\dagger \to a_1^\dagger a_2^\dagger, \quad a_1^\dagger a_3^\dagger \to a_1^\dagger a_4^\dagger, \quad a_2^\dagger a_3^\dagger \to a_2^\dagger a_4^\dagger, \quad a_2^\dagger a_4^\dagger \to a_2^\dagger a_3^\dagger. \tag{2.7}$$

[1] Rigorously, the beam splitter is not identical to a Hadamard gate, but it is equivalent to a Hadamard gate with a phase factor i.

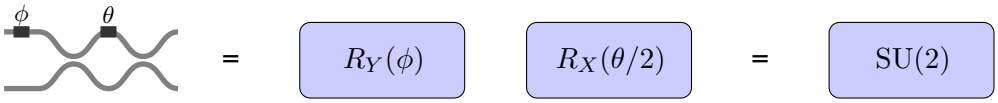

Figure 2.2: The MZI operates as a special unitary matrix $\mathrm{SU}(2)$ on the input field. The two phase shifters perform Z-rotation gates, and the beam splitters perform Hadamard gates, which is equivalent to the product of one X-rotation gate with phase θ and another Z-rotation gate with phase ϕ.

Here we denote the basis of the control qubit as $a_1^\dagger$ and $a_2^\dagger$, and the target qubit as $a_3^\dagger$ and $a_4^\dagger$.

Clearly, for the first two terms, the transition is not allowed simply because the photon number operator is conserved. For the latter two terms, the target qubit photon must flip according to the control qubit, which is not allowed in linear optics.

There exists a much simpler way to encode two-qubit gates using only one photon across four optical modes. In this case, 2^N optical modes are needed to represent N qubits. However, this scaling is exponential in the number of optical modes, which is impractical for large-scale quantum computing.

This challenge has led to several solutions for implementing two-qubit gates, which constitutes the main difficulty of linear optics quantum computing. The first solution uses entangled photon pairs to perform a non-deterministic nonlinear phase shift that succeeds with probability $1/4$, known as the KLM scheme [64]. The second solution uses two ancilla vacuum modes to post-select a subspace of the two-qubit gates, which is not deterministic but can be achieved with high success probability [65, 66]. The first experimental demonstration was in 2005 by O'Brien et al., which implemented a non-deterministic CNOT gate using two-photon interference with a success probability of $1/9$.

Another approach uses atomic systems or other artificial atoms, including quantum dots, atoms, and spins, to assist the photon-photon interaction [68–70]. For example, a single atom can mediate the interaction between two photons, effectively creating a two-qubit gate [69]. This approach relies on strong coupling between the artificial atom and the photons, usually achieved in an optical cavity or waveguide [70, 71]. However, this method is experimentally challenging due to the requirement of strong coupling and precise control of the atom-photon interactions.

2.3 Measurement-based quantum computing

Apart from the aforementioned probabilistic linear optics approach, we also discuss the concept of measurement-based quantum computing (MBQC), also

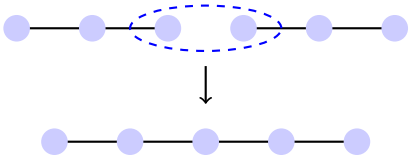

Figure 2.3: The fusion gate for measurement-based quantum computing. The fusion gate heralds the successful generation of a cluster state by measuring one photon from a large number of photons.

known as one-way quantum computing [72–74]. It is proven to be equivalent to the circuit model of quantum computing [73, 75]. In this framework, rather than representing each qubit by a single photon, it uses a large set of entangled states, called graph states or cluster states:

$$|G\rangle = \prod_{(a,b)\in E} \mathrm{CZ}_{a,b} \bigotimes_{v\in V} |+\rangle_v \tag{2.8}$$

where $G = (V, E)$ is a graph with vertices V and edges E, $|+\rangle = (|0\rangle + |1\rangle)/\sqrt{2}$ is the plus state, and $\mathrm{CZ}_{a,b}$ is the controlled-Z operation between qubits a and b.

To generate such entangled states in photonic systems, one typically uses parametric processes such as SPDC or SFWM to produce entangled photon pairs, which are then combined into larger entangled states using so-called fusion gates [76–78]. This method is particularly well-suited to photonic systems, as high-quality entangled photon pairs can be reliably generated using these processes [79–81].

A central operation in this approach is the fusion gate, as illustrated in Fig. 2.3. Here, two cluster states are fused into a larger cluster state by performing a Bell state measurement (BSM), which is typically implemented using a beam splitter followed by single-photon detectors. When one of the detectors registers exactly one photon, the measurement projects the input photons onto an entangled Bell basis, thereby fusing the two cluster states. This process is also called post-selection, as the successful fusion is heralded by the detection event. However, this process succeeds with a probability of 50% and requires sacrificing one photon [82].

2.4 Fock space, linear optics and boson sampling

From the above discussion, it is clear that there exists a gap between the qubit basis and photon number states across multiple optical modes. As bosonic particles, photons can occupy multiple modes with an arbitrary number of photons. Consider a quantum system with m modes, with n photons occupied

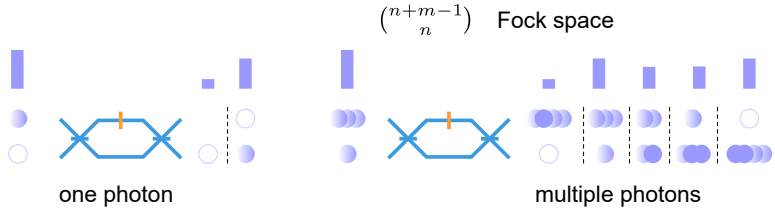

Figure 2.4: Fock space representation of photonic quantum states.

in total. Each basis can be represented by a *photon number state*:

$$|n\rangle = |n_1, n_2, \ldots, n_m\rangle = \frac{(a_1^\dagger)^{n_1}(a_2^\dagger)^{n_2}\cdots(a_m^\dagger)^{n_m}}{\sqrt{n_1! n_2! \cdots n_m!}} |0\rangle^{\otimes m} \tag{2.9}$$

Here, n_i is the number of photons in the i-th mode, and $\sum_{i=1}^{m} n_i = n$. The creation and annihilation operators satisfy the bosonic commutation relations.

It is straightforward to calculate the number of states allowed in this photon number space Φ_{nm}, which is a photon-number-conserved subspace truncated from the Fock space $\Phi_{nm} \subsetneq \mathcal{F}$. The dimension of this space is given by:

$$M = \dim \Phi_{nm} = \binom{m+n-1}{n} \tag{2.10}$$

Using the Stirling approximation, the dimension of the Fock space grows exponentially with the number of modes m and photons n. For large m and n, this scaling is much faster than the 2^n growth of the qubit Hilbert space, highlighting the computational complexity inherent in simulating bosonic systems. For example, when the number of photons equals the number of modes $m = n$, it scales faster than the qubit space, which is 2^n.

Inspired by this observation, in 2011, Aaronson and Arkhipov introduced the Boson Sampling problem, which generalizes linear optics and photon number states.

Theorem 2.1. *Given a unitary matrix u with m modes, and an input photon number state $|s\rangle = |s_1, s_2, \ldots, s_m\rangle$ containing n photons, the probability of detecting a specific output state $|t\rangle = |t_1, t_2, \ldots, t_m\rangle$ is given by the permanent of the matrix $\tilde{u}$:*

$$\langle t| \varphi(u) |s\rangle = \frac{\mathrm{Perm}(\tilde{u})}{\sqrt{s_1! \cdots s_m! t_1! \cdots t_m!}} \tag{2.11}$$

Here $\tilde{u}$ is an $n \times n$ matrix, which is generated from the unitary matrix u by repeating the i-th column s_i times, and the j-th row t_j times.

There are several interesting points to note here. First, there are three matrices present: the unitary $u \in \mathbb{C}^{m\times m}$, the matrix $\tilde{u} \in \mathbb{C}^{n\times n}$, and the unitary $\varphi(u) \in \mathbb{C}^{D\times D}$. Second, despite the depth of the linear optics circuit, the unitary u is always $m \times m$. For example, for a circuit with L layers of MZIs, the total unitary is given by the product of the L unitaries, $u = u_L u_{L-1} \cdots u_1$, where each u_i is always an $m \times m$ unitary matrix. Third, the matrix $\tilde{u}$ is not always unitary as it repeats the rows and columns of u. Finally, the unitary $\varphi(u)$ is the actual quantum process that acts on the Fock space $\mathcal{F}$, which transforms the input state $|s\rangle$ to the output state $|t\rangle$. Therefore, if one could sample all the input states and output states, it would be equivalent to knowing the entire quantum process $\varphi(u)$. Clearly, the dimension of $\varphi(u)$ is M, which grows exponentially with m and n.

Hardness and approximation of Boson Sampling

In Ref. [24], it is proven that Boson Sampling is #P-hard to simulate classically. This means that no polynomial-time classical algorithm can solve this problem unless the polynomial hierarchy collapses to the third level, which is considered highly unlikely in computational complexity theory. The core of the proof relies on the fact that calculating the permanent of a matrix is #P-hard, and the output probabilities of Boson Sampling are directly related to the permanents of $\tilde{u}$ [83].

However, subsequent works have developed approximation methods for Boson Sampling, which are known as Gurvits' approximation [84]. First, Glynn's formula exists to calculate the permanent of a matrix:

Theorem 2.2. *Given a matrix X, the approximation of the permanent of X is given by:*

$$\mathrm{Perm}(X) = \mathbb{E}_{\boldsymbol{x}\sim\mathcal{N}(0,1)}\left[x_1 x_2 \dots x_n \prod_{i=1}^{n}(a_{i,1}x_1 + a_{i,2}x_2 + \cdots + a_{i,n}x_n)\right] \tag{2.12}$$

where $\mathcal{N}$ is the normalized distribution of vector $\boldsymbol{x} = (x_1, x_2, \dots, x_n)^T \in \mathbb{R}^n$.

Therefore, in Ref. [85], it is proven that to estimate the permanent of a matrix, one runs $O(n^2/\epsilon^2)$ iterations to achieve $\pm\epsilon||A||^n$ additive error, where $||A||$ is the operator norm of matrix A. This concludes that the Boson Sampling problem can be estimated efficiently within a small error in polynomial time.

Lie algebra and Lie group in linear optics

Similarly, in Ref. [24], one insightful observation is that the map φ, connecting two unitary groups $U(m)$ and $U(M)$, is a homomorphism map, $\varphi : \mathbb{C}^{n\times n} \mapsto \mathbb{C}^{D\times D}$. Note that $U(m)$ is the unitary group of $m\times m$ unitary matrices, which acts on the classical optical modes. Meanwhile, $U(M)$ is the unitary group of $M \times M$ unitary

matrices, which describes the actual quantum process in the Fock space. Since φ is a homomorphic map, one can study many properties in the exponential Fock space through the m-dimensional $U(m)$ group.

A very useful property of this homomorphism is that:

$$\varphi(AB) = \varphi(A)\varphi(B), \quad \forall A, B \in u(m). \tag{2.13}$$

This means for the Boson Sampling problem, sampling the product matrix $\varphi(AB)$ is equivalent to sampling the product of the two matrices $\varphi(A)$ and $\varphi(B)$.

Furthermore, in Ref. [86], the commutation relation between the classical unitary $u(m)$ and Fock space unitary $U(M)$ is explained more clearly, as illustrated in Fig. 2.5.

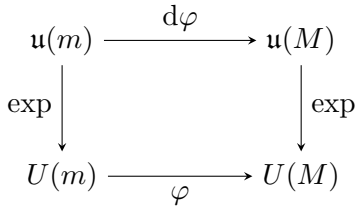

Figure 2.5: The relation between Lie algebra and Lie group in linear optics. $\mathfrak{u}(m)$ and $\mathfrak{u}(M)$ are the Lie algebras of $U(m)$ and $U(M)$ respectively. The exponential map connects the Lie algebra to the Lie group, while the homomorphism $\mathrm{d}\varphi$ connects the two Lie algebras.

Here, $\mathfrak{u}(m)$ is the Lie algebra of $U(m)$, which is the set of all $m \times m$ Hermitian matrices. $\mathfrak{u}(M)$ is the Lie algebra of $U(M)$, which is the set of all $M \times M$ Hermitian matrices, corresponding to the Hamiltonian of the quantum process in Fock space. One can understand both Lie algebras as the tangent space of the Lie group, which means that in both spaces, the Hermitian matrices generate the unitary matrices through the exponential map.

In summary, linear optical quantum computing has demonstrated the capability in the following experiments. First, it can implement arbitrary single qubit gates using MZIs. Second, it can implement non-deterministic two-qubit gates using ancilla photons. Based on these two building blocks, it has succeeded and is leading in demonstrating small-scale quantum algorithms, such as factoring [87], one-way computing [88], quantum walk [89, 90] and quantum simulation [91]. Third, it shows the capability of generating entangled photon pairs using SPDC up to twelve photons [92]. In the scheme of boson sampling, it was first demonstrated with a few photons [25–27, 93]. Furthermore, extending the photon number states into weak squeezed light, it was demonstrated up to 76 output photon clicks [94], which demonstrates quantum supremacy over classical computers.

Nevertheless, looking toward the future of photonic quantum computing, there are still many challenges to overcome. First, the photon sources are still not perfect single-photon sources, suffering from either probabilistic generation or multiphoton emission. This is especially crucial for SPDC sources. For ideal single-photon sources, such as quantum dots, they can be deterministic but still suffer from probabilistic coupling and low indistinguishability. Second, photon loss remains a significant issue for both free-space and integrated photonics, including coupling loss, propagation loss, and detection loss. Third, current research mainly focuses on linear optics, but nonlinear effects such as Kerr gates are still not practical due to low efficiency. To overcome these challenges, future research should focus on many aspects, including the development of high-quality photon sources, low-loss photonic components, and efficient nonlinear optical processes. By solving these challenges, photonic quantum computing can continue to advance and realize its full potential in quantum information processing.

3 Photonic Integrated Circuits

Referring to the previous chapter, free-space optical components have been widely used in quantum optics experiments for decades. However, as the complexity of quantum optics experiments increases, free-space optics face several challenges, including large footprints, instability, and high costs. The core idea of photonic integrated circuits (PIC) is to solve the above issues by integrating all free-space optical components into a single chip [95–99]. This integration aims to: 1) reduce the footprint of optical systems, 2) improve system stability and reliability, 3) reduce mass production costs, 4) enhance system scalability, and 5) decrease energy consumption.

In this chapter, we first summarize the materials related to the silicon family—silicon (Si), silicon nitride (SiN), and fused silica—which are the most CMOS-compatible materials from a quantum optics perspective. Next, we introduce the basic components used in PIC for photon manipulation, in perspective of both classical and quantum optics experiments. Furthermore, we discuss universal circuits for linear optics quantum computing, focusing on their composition and decomposition into interferometer arrays, as well as circuit calibration methods.

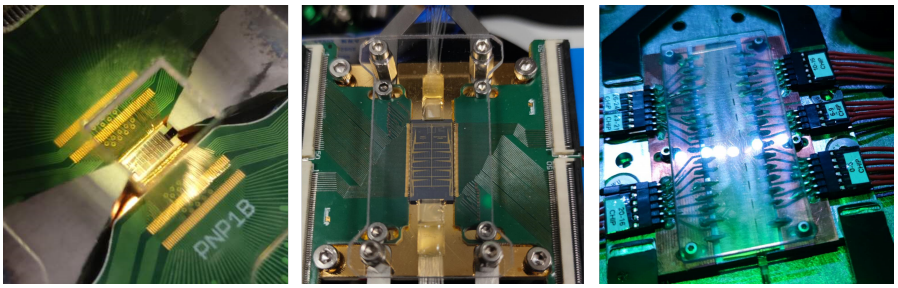

Figure 3.1: Integrated photonic circuits made of several different materials (silicon, SiN, silica, from left to right) manufactured by AFM, CNR (start-up as Ephos now), and Lionix.

3.1 Material Platforms

Over the past half century, the development of electronic integrated circuits has relied on the CMOS process using silicon as the primary material, demonstrating the capability to integrate millions or billions of transistors within a millimeter-scale footprint. This has also made silicon the most popular material for PIC

due to its CMOS compatibility. However, as a semiconductor with an indirect band gap, silicon cannot emit light directly, which presents the main challenge for optoelectronic integration.

For quantum optics applications, single photon sources based on parametric down-conversion nonlinear processes do not require direct light emission from silicon, but rather rely on optical nonlinearity. Although silicon lacks second-order optical nonlinearity due to the symmetry of its crystal lattice, it does possess third-order optical nonlinearity. Unlike conventional spontaneous parametric down-conversion using second-order nonlinearity, third-order nonlinearity enables spontaneous four-wave mixing (SFWM), which has been reported to generate parametric down-conversion photons over the last decade [100].

Meanwhile, silicon can detect single photons at nanometer resolution through the avalanche effect in photodiodes [101–103]. However, this effect only works for short-wave infrared light around 800 nm, while SFWM generates single photons in the telecom band around 1550 nm. Therefore, superconducting nanowire single photon detection, either off-chip or through on-chip integration, has been developed as an alternative solution [104, 105].

Similar to silicon, SiN is another material that has recently attracted significant attention in both industrial and academic communities. SiN has a lower refractive index than silicon, which typically results in a larger device footprint. The lower Kerr nonlinearity of SiN also necessitates the use of resonator structures, such as micro-rings, to enhance the nonlinearity [106]. Another advantage of SiN is its transparency range from 400 nm to 2200 nm, which enables operation in both visible and near-infrared wavelength ranges.

Silica is another material used for PIC fabrication, which was first demonstrated in realizing a quantum CNOT gate [107]. General silica used in photonics encompasses several different material systems, including Hydex glass [108], boro-aluminosilicate glass, and conventional silicon dioxide. While waveguides can be fabricated using CMOS technology on doped silica, more recent advances focus on laser writing waveguide technology, where waveguides are written with femtosecond lasers in bulk silica substrates [109–112]. Subsequent procedures such as microtrench formation and metal ablation can be performed using laser processing, resulting in fully functional PIC.

The main advantage of silica is that the waveguides are much larger than silicon or SiN waveguides, enabling easy coupling to optical fibers. A significant drawback of silica waveguides is the low thermal conductivity of the substrate, which results in slow response times and high thermal crosstalk between neighboring phase shifters [109, 116].

In Table 3.1, we summarize the main properties of three circuits based on the aforementioned materials.

Materials	Si	SiN	fused silica
Refractive index (@1550 nm)	∼3.48	∼2.0	∼1.54
Spot size (μm)	∼0.5	∼1	10–20
Transparency range (nm)	>1000	400–2350	400–2200
Propagation loss (dB/cm)	<0.3–2.4	<0.01	0.1–1

Table 3.1: Comparison of different photonic integrated circuit platforms. For silicon, see Ref. [113, 114] For SiN, see Ref. [115]. For fused silica, see Ref. [109, 112]

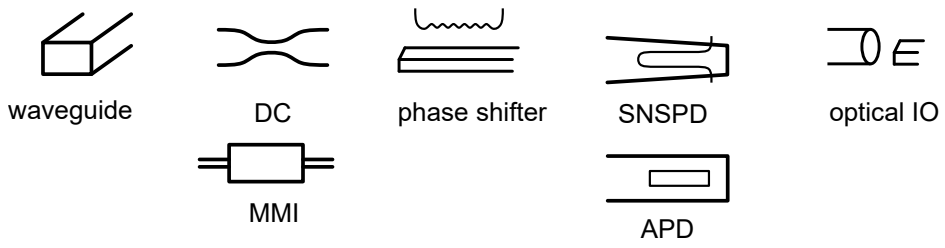

Figure 3.2: Basic components of photonic integrated circuits.

3.2 Basic components

Typical components of PIC include waveguides, optical input/output (optical I/O), beam splitters, phase shifters, modulators, and detectors to achieve the full functionality [31–34, 114, 117, 118]. These components can be used to construct either universal circuits for specific purposes, such as tunable interferometers, or full optical circuits that include single photon sources, interferometers, and detectors, which are shown in Fig. 3.2.

1. **Waveguide**

   The most critical parameter of an optical waveguide for quantum applications is the propagation loss of the fundamental mode, which typically consists of two components: material absorption and scattering loss due to sidewall roughness. Among the silicon family, silicon-on-insulator (SOI) is the most popular material for waveguide fabrication, but it suffers from two-photon absorption and Raman scattering at the telecom wavelength band [119]. SiN has gained increasing popularity over the past decade, particularly due to its similar fabrication process to silicon while avoiding the aforementioned limitations of silicon [108].

   Silica represents another material for waveguide fabrication, utilizing both CMOS processes and laser writing technology. The primary advantage of silica is that through multiple laser writing passes, the mode spot size

can be tuned to match that of an optical fiber, thereby reducing coupling loss to negligible levels [109–112]. However, this approach complicates the fabrication of metallic components, such as phase shifters, and poses challenges for signal routing as the geometry scales up.

2. **Beam splitter**

   Two main implementations of beam splitters exist in PIC: the directional coupler (DC) and the 2×2 multi-mode interferometer (MMI) [120].

   In the directional coupler approach, two waveguides separated by a few hundred nanometers share their evanescent fields, resulting in electromagnetic field coupling between the waveguides [121]. A typical directional coupler consists of two bending waveguides and is characterized by the -phase shift length, denoted as L_π. The advantage of this coupler type is that the coupling ratio can be continuously tuned from 0 to 1 by adjusting the coupler length. The drawback is its high sensitivity to wavelength variations and fabrication imperfections, particularly the gap between the two waveguides.

   The MMI approach is based on optical interference and can be understood through the optical *Talbot effect* [122, 123]. By designing a specific rectangular waveguide (typically larger than the fundamental mode), light can be split into N different paths with equal power distribution. The advantage of MMI is its reduced sensitivity to wavelength variations and fabrication imperfections due to its simpler geometry.

3. **Phase shifter**

   While the silicon photonics industry offers various phase shifters-including electro-optic phase shifters, thermo-optic modulators, and mechanical modulators-the thermo-optical phase shifter is most commonly used in quantum experiments due to its low insertion loss and straightforward fabrication [124].

   Compared to alternative phase shifters, the thermo-optical approach requires only a top metallic layer serving as the heater, which can be integrated with the wire routing layer [125]. Since this metal layer must be separated from the waveguide layer by several micrometers, the fabrication process does not affect the waveguide geometry, thus introducing no additional contamination or optical loss.

4. **Detector**

   While the ultimate goal of silicon photonics is to achieve fully integrated circuits, current technology still faces challenges in the heterogeneous integration of active semiconductors, such as germanium and other III-V materials. Due to lattice mismatch considerations, germanium is the most popular material for diode fabrication on silicon wafers.

However, for single photon detection applications, either avalanche photodiodes (APDs) [103] or superconducting nanowire single photon detectors (SNSPDs) must be employed. Significant efforts have been directed toward both approaches, particularly for SNSPDs. One of the main advantages of SNSPDs is the integration of multiple detectors on top of waveguides, enabling high number resolution and detection efficiency [126, 127]. However, silicon-based APDs can only operate with visible light, for which silicon is not transparent, while SNSPDs require a complete cryogenic environment for operation.

5. **Optical I/O**

   To couple light into and out of the PIC, typically from optical fibers, the optical mode must be transformed from the fiber dimensions (several micrometers) to the waveguide dimensions (hundreds of nanometers)-a mode size reduction of approximately ten times.

   The first structure employed is the inverse-tapered waveguide, where the spot size is gradually increased as the waveguide becomes progressively narrower [128]. The second structure is the mode converter, primarily used for Si waveguides, which typically consists of two waveguide layers, such as silicon oxynitride (SiON) and silicon [129]. Since SiON has a refractive index between that of silicon and the cladding material, light can be guided sequentially from the Si waveguide to the SiON waveguide, and finally to the fiber.

   Beyond these edge-coupling structures, another approach utilizes the grating coupler-a periodic conical structure etched at the waveguide ends. The period and duty cycle of the grating determine the operating wavelength.

Currently, the more popular way to produce the PIC is to use so-called fabless foundry, which only produces the chips based on the user's design. Previously, the user needs to design each functional component separately, and test unit by unit. The fabless foundry separates the design and fabrication by providing a photonic design kit (PDK), which includes the design rules, fundamental components, and even simulation and routing tools [96, 113, 130]. After submitting the layout design, called a tapeout, the foundry will merge the different designs in a single wafer, and then fabricate the chips in what is called a multi-project wafer (MPW).

3.3 Universal linear optical circuits

The nature of linear optics is that interference of multiple photon states happens in a linear optical system, which can be sufficiently described by classical light. Therefore, it is vital to construct a tunable linear optics circuit to implement

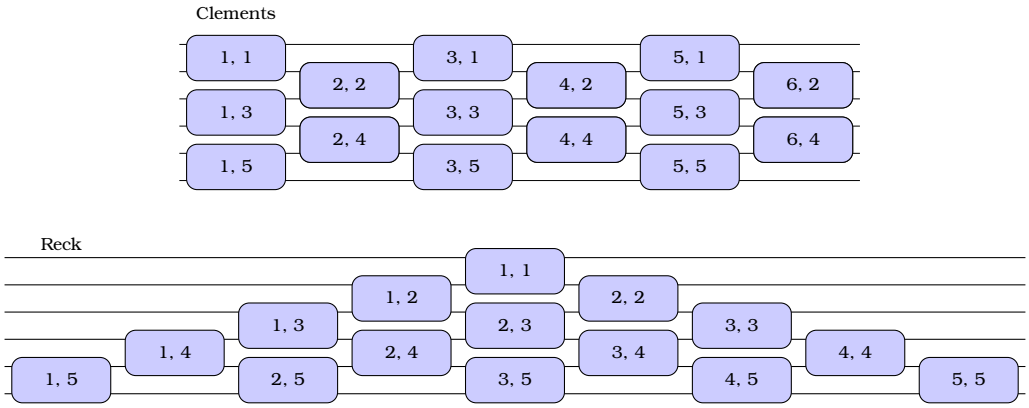

Figure 3.3: Decomposition of unitary matrix with six modes into Clements and Reck architectures.

arbitrary unitary transformations for quantum interference. Considering energy conservation and the unitarity of optical paths, linear optics must obey special unitary transformations $\mathrm{SU}(N)$. In experiments, particularly in PIC, classical light or single photons can only interfere within two neighboring modes in a Mach-Zehnder interferometer (MZI). Intuitively, one could presume that if the total unitary transformation could be represented as a product of a series of local interferometers, it can be represented as smaller 2×2 unitary matrices.

To construct an $N \times N$ unitary matrix, we first need $2N^2$ real parameters. Due to the unitarity constraint $UU^\dagger = I$, we introduce $N^2 + N$ constraints, where N^2 comes from the upper triangular part and N from the diagonal. This reduces the number of independent parameters to $N(N-1)$, which equals the number of phase shifters needed to construct a fully programmable linear optics circuit.

Reck et al. [55] first proposed a universal circuit architecture that decomposes an $N \times N$ arbitrary unitary matrix into a series of 2×2 unitaries in a triangular mesh, as shown in Fig. 3.3. Later in 2015, Clements et al. [56] proposed a new architecture similar to the Reck architecture but implemented in a rectangular mesh, also shown in Fig. 3.3. This new architecture features shorter circuit depth with the same number of phase shifters, where the loss is more balanced between different modes, see Fig. 3.4.

Here we explain both methods in detail. First, we consider the inverse of a real MZI without global phase, which is also a 2×2 unitary matrix with free parameters $\tilde{\theta}$ and $\tilde{\phi}$:

$$T(\tilde{\theta}, \tilde{\phi}) = \begin{pmatrix} e^{-i\tilde{\phi}} \sin \tilde{\theta}/2 & e^{-i\tilde{\phi}} \cos \tilde{\theta}/2 \\ \cos \tilde{\theta}/2 & -\sin \tilde{\theta}/2 \end{pmatrix}. \tag{3.1}$$

Second, if T is right-multiplied with an existing 2×2 submatrix U, the left bottom element in the product matrix can be nulled out by choosing the proper $\tilde{\theta}$ and $\tilde{\phi}$

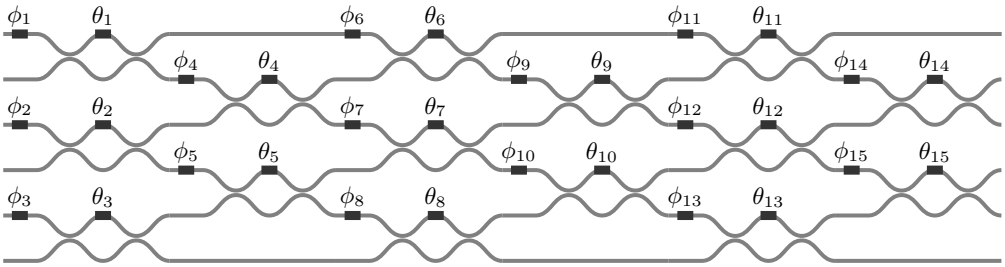

Figure 3.4: Universal circuits for linear optics quantum computing. Here we show a 6-mode universal circuit made of 15 MZIs and 30 phase shifters.

as follows:

$$\begin{aligned} UT &= \begin{pmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{pmatrix} \begin{pmatrix} e^{-i\tilde{\phi}} \sin\tilde{\theta}/2 & e^{-i\tilde{\phi}} \cos\tilde{\theta}/2 \\ \cos\tilde{\theta}/2 & -\sin\tilde{\theta}/2 \end{pmatrix} \\ &= \begin{pmatrix} U_{11}e^{-i\tilde{\phi}} \sin\tilde{\theta}/2 + U_{12}\cos\tilde{\theta}/2 & U_{11}e^{-i\tilde{\phi}} \cos\tilde{\theta}/2 - U_{12}\sin\tilde{\theta}/2 \\ U_{21}e^{-i\tilde{\phi}} \sin\tilde{\theta}/2 + U_{22}\cos\tilde{\theta}/2 & U_{21}e^{-i\tilde{\phi}} \cos\tilde{\theta}/2 - U_{22}\sin\tilde{\theta}/2 \end{pmatrix} \end{aligned} \tag{3.2}$$

where the left bottom element becomes zero if and only if

$$U_{21}e^{-i\tilde{\phi}} \sin\tilde{\theta}/2 + U_{22}\cos\tilde{\theta}/2 = 0. \tag{3.3}$$

The solution to this equation is

$$\tilde{\theta} = 2\arctan\left|\frac{U_{22}}{U_{21}}\right|, \qquad \tilde{\phi} = -\arg\frac{U_{21}}{U_{22}} \tag{3.4}$$

The same procedure works for right multiplication. Note that after applying the inverse MZI, the left bottom element is zero, but all other three elements are changed as well.

From the above derivations, this indicates that by finding the set of inverse MZIs in a specific order, the product of all inverse MZIs and the original unitary matrix can be diagonalized:

$$U\prod_{s\in S} T_s(\tilde{\theta}_s, \tilde{\phi}_s) = D \tag{3.5}$$

where D is a diagonal matrix and S is the set of the modes in the circuit. Considering that each submatrix is represented as an MZI on neighboring modes, each mode set will be like $\{(i, i+1)|1 \leq i \leq N-1\}$. In fact, Reck et al. [55] and Clements et al. [56] choose different orders of applying the inverse MZIs as shown in Fig. 3.3. The former chooses the order continuously for each diagonal line, while the latter chooses the order by each row.

Beyond both architectures, Bell et al. [131] proposed a new method by absorbing the external phase shifter ϕ inside the beam splitters, which further reduces the chip length. There are two phase shifters between two beam splitters, which can also control the amplitude and phase of the input light.

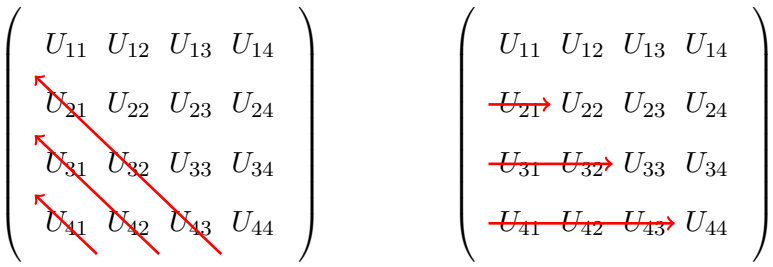

Figure 3.5: Decomposition order in rectangular mesh (left) and triangular mesh (right). In the rectangular mesh, the decomposition starts from mode (1,2) and vanishing U_{41}. The second step is to vanishing U_{42}, U_{31}, then U_{43} and so on. Note the element label in the matrix is not the same as the mode label during the multiplication and decomposition, because left multiplication of block matrix effects the corresponding columns, while right multiplication effects the corresponding rows.

3.4 Characterization

The PIC system involves not only optical components, but also relies on multiple-signal electronic control systems and dedicated thermal management to achieve full functionality. Unfortunately, the design of a PIC does not always guarantee that the perfect layout and structure are reproduced in fabrication as expected. Thus, the characterization of the PIC is urgently needed to qualify and calibrate its performance. In this section, we will first discuss the reasons causing deviation from ideal performance, and second, the characterization methods to optimize the performance.

3.4.1 Error analysis

Regarding the errors that could occur in the PIC, we can classify them into three categories: passive error, active error, and quantum error. The previous two exist in any experiment using classical light, which of course would also affect quantum optics experiments. The quantum error, which mainly comes from the shot noise of the single photon source, is specific to quantum optics experiments.

Passive Error

For passive error, it originates from the imperfection of the fabrication process, wavelength dependence, waveguide birefringence, uniformity, and lossy channels. Depending on the material and fabrication, some of the errors can be negligible, while others can be significant and thus need to be considered in the design and calibration.

- **Biased beam splitter** ($|U_{11}| \neq |U_{12}|$)
  Beam splitters may deviate from the ideal 50:50 splitting ratio due to fabrication imperfections, resulting in biased interference. In the Clements architecture, one could start from a single MZI in the corner, for example the top left one. Due to fabrication imperfection, the beam splitters can be biased with a small ratio rather than 50:50, such as 51:49. In the worst case, the imperfect MZI can be represented as [63]

$$\begin{aligned} \mathrm{MZI}_{\mathrm{bias}}(\theta,\phi) &= \begin{pmatrix} 1 & 0 \\ 0 & e^{i\phi} \end{pmatrix} \begin{pmatrix} \cos(\pi/4+\alpha) & i\sin(\pi/4+\alpha) \\ i\sin(\pi/4+\alpha) & \cos(\pi/4+\alpha) \end{pmatrix} \\ &\times \begin{pmatrix} 1 & 0 \\ 0 & e^{i\theta} \end{pmatrix} \begin{pmatrix} \cos(\pi/4+\beta) & i\sin(\pi/4+\beta) \\ i\sin(\pi/4+\beta) & \cos(\pi/4+\beta) \end{pmatrix} \\ &= \begin{pmatrix} \cos\beta & i\sin\beta \\ i\sin\alpha & \cos\beta \end{pmatrix} \mathrm{MZI}(\theta,\phi) \begin{pmatrix} \cos\alpha & ie^{-i\phi}\sin\alpha \\ ie^{i\phi}\sin\alpha & \cos\alpha \end{pmatrix} \end{aligned} \tag{3.6}$$

  In this way, the bias added to two beam splitters can be understood as two additional rotation operations, and the second rotation is adjoint with another phase rotation ϕ. In the case $\alpha = \beta$, the rotation operations are cancelled. The extinction ratio of the MZI is determined by the bias α and β:

$$\begin{aligned} \mathrm{ER}_{\mathrm{top}} &= \frac{\cos^2(\alpha-\beta)}{\sin^2(\alpha+\beta)}, \\ \mathrm{ER}_{\mathrm{bottom}} &= \frac{\cos^2(\alpha+\beta)}{\sin^2(\alpha-\beta)} \end{aligned} \tag{3.7}$$

  Here top and bottom refer to switching the input light from the top port to output at the same top port or the bottom port. We can see that the extinction ratio for the bottom is possible to reach an absolutely higher value, as long as $\alpha = \beta$.

- **Nonlinear phase shifters** ($P \neq VI$)
  The relationship between applied power and induced phase shift may be nonlinear, affecting calibration accuracy. For a single phase shifter, the phase change is propagated from the applied voltage or current to the heating power, and then to the refractive index change, and finally to the phase change.

$$\phi \propto \Delta n \propto P \propto V^2 \propto I^2 \tag{3.8}$$

  In the real case, each proportionality above is not consistent. For example, the resistance of the heater may change with the electric power as heat is accumulated.

- **Non-uniformity** ($U_{12} \neq U_{34}$)
  Variations in device parameters across different elements can lead to non-uniform circuit behavior. For a chip made by the CMOS process, the

fabrication will vary slightly across the chip. For example, in the directional coupler, the gap between two waveguides is critical for the coupling ratio. Thus even in the same design, fabrication imperfection or low uniformity will cause the coupling ratio across the chip to be different. For each physical quantity and device specification, one must consider the uniformity in the design stage. Good uniformity can be achieved by setting the critical parameters at values with high tolerance. Sometimes, if the process does not reproduce the same uniformity, one should consider sweeping the critical parameters several times to find the best performance.

- **Loss channel** ($UU^\dagger \neq I$)
  Optical losses cause the implemented transformation to deviate from unitary. Consider a diagonal matrix representing the loss at each mode, $D = \mathrm{diag}(\eta_1, \eta_2, \ldots, \eta_N)$. The product of the ideal unitary matrix U and the loss matrix D holds $UD = DU$, since D is diagonal. This indicates that although the loss channel is local, it can be propagated through the whole circuit and merged into a single diagonal matrix. Therefore, the loss matrix can be measured separately for each channel, along with the input and output optics. At the end, the loss matrix has to be included in any calibration during the experiment.

Active Error

Despite the passive errors caused by imperfection, crosstalk is another important factor to qualify the performance of mesh-based circuits. It includes crosstalk such as electric crosstalk, optical crosstalk, and thermal crosstalk.

- **Heat dissipation** ($Q = Pt$)
  Thermal conductivity of silicon is larger than that of SiN, which is larger than that of silica. Thus the heat generated in phase shifters, especially in thermo-optical phase shifters, should be dissipated from the device as fast as possible. While the total heat dissipation is not only determined by material thermal conductivity, but also by the geometry of the device. Thus the heat capacity of the substrate should be larger than the heat generated during operation. A significant difference in heat dissipation between silicon or SiN and silica is that the former two are usually fabricated on a silicon wafer, which has much larger thermal conductivity than a silica substrate. A typical operation speed of thermal phase shifters on silicon or SiN can be 1kHz, while silica waveguides can only operate slower than 1Hz.

- **Electronic crosstalk** (mV/V)
  Electrical signals may interfere between channels, causing unintended phase shifts. Due to the crosstalk mentioned above, the control task of

all phase shifters is applying tens of powers independently, covering at least 2π phase tuning range simultaneously. Commercially, the control is done by a current-driving power supply, which is not only expensive but also not scalable for complex circuits [132]. During the design stage, due to the schematic routing of multiple phase shifters, it is common to have signals of different phase shifters routed from the same ground pin. Even though the typical width of metallic wire can be larger than several tens of micrometers, which has several ohm resistance, the ground pin can still be significant as it is in series with other pins. For example, the common ground pin is connected to N channels of metallic resistors with resistance R, the total resistance in load is

$$R_L = R_G + \frac{R}{N} \tag{3.9}$$

A typical value of R_G is 10-20 Ω, and R is 100-200 Ω, which is determined by the heater and ground wires design. The above formula shows that the total resistance is contributed by the ground pin more as the number of phase shifters increases.

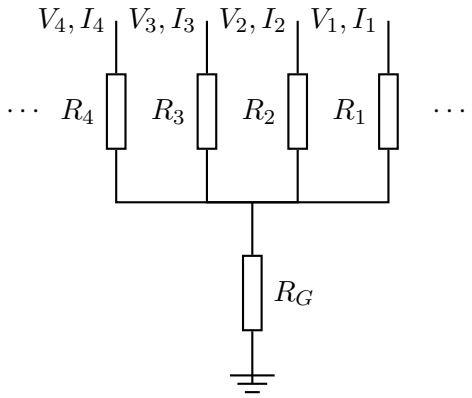

Figure 3.6: Effective circuit model of resistor network: R_G in series with n parallel resistors R.

- **Thermal crosstalk** (20%–50%)
  Heating one phase shifter can affect neighboring elements, leading to correlated errors. In the case of dense schematics, a resistor serving as a phase shifter can warm up the neighboring area, which leads to phase changes at other phase shifters. This indicates a crosstalk matrix X, which describes how much phase change is induced at one phase shifter when heating another phase shifter. Hence, the real phase change is the sum of the in-situ phase change and the crosstalk from neighboring phase

shifters:

$$\begin{pmatrix} \hat{\phi_1} \\ \hat{\phi_2} \\ \vdots \\ \hat{\phi_N} \end{pmatrix} = \begin{pmatrix} X_{11} & X_{12} & \cdots & X_{1N} \\ X_{21} & X_{22} & \cdots & X_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ X_{N1} & X_{N2} & \cdots & X_{NN} \end{pmatrix} \begin{pmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_N \end{pmatrix} = X \begin{pmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_N \end{pmatrix} \tag{3.10}$$

Here, X is usually a symmetric matrix and the diagonal elements indicate the efficiency of phase change at each phase shifter. All elements X_{ij} should be smaller than 1. Due to the chip design, some phase shifters are closer, which results in a block matrix structure of X. Thus to drive the phase in demand, it is equivalent to finding the solution of

$$\boldsymbol{\phi} = X^{-1}\hat{\boldsymbol{\phi}} \tag{3.11}$$

The equation above only holds when the crosstalk is linear; for the nonlinear case, the characterization is more complicated because

$$\hat{\boldsymbol{\phi}} = \boldsymbol{\phi} + X\boldsymbol{\phi} + X^2\boldsymbol{\phi} + O(X^n\boldsymbol{\phi}) \tag{3.12}$$

- **Optical crosstalk**
  Optical crosstalk occurs when light from one channel leaks into another. This can happen due to two close waveguides, where the light can transition from one waveguide to another. Another cause can be misalignment of the optical input/output for example during the scattering from a fiber array into the waveguide. By designing the waveguide with a larger buffer distance (> 10x the waveguide width), the crosstalk can be reduced.

Quantum Error

All errors above display the same characteristics in an experiment using classical light, but in quantum optics experiments, some errors are specific due to statistics and other quantum effects.

- **Shot noise** ($\Delta n = \sqrt{\bar{n}}$)
  Typical photon sources, such as spontaneous parametric down-conversion, pumped by a coherent laser, feature Poisson statistics. The standard deviation of the photon number is correlated with the mean photon number, $\Delta n = \sqrt{\bar{n}}$.

- **Dark counts** ($C(n = 0) > 0$)
  While the chip is working at room temperature, the thermal phase shifters can be heated up to some temperature locally, which emit infrared thermal photons. Usually due to the sensitivity of the detector, those photons can be ignorable for silicon APDs, but can affect nanowire detectors as they are sensitive to infrared photons. A typical solution is to add band-pass filters to block the infrared photons before entering the detectors.

3.4.2 Calibration and control

The calibration relies only on the optical power change one can measure at the output ports. As illustrated in Fig. 3.4, many phase shifters are located in the middle of a circuit, which means that the phase change in one specific MZI will change the light on two specific output ports. For both ports, one can fit the output power as a function of the phase change:

$$P(\phi) = A + B\cos(\phi + C), \tag{3.13}$$

Here, A, B, C and ϕ_0 are the fitting parameters. In particular, C is the phase offset, namely the original phase difference between two arms. A and B determine the extinction ratio of the MZI. Assuming the phase has the following relationship with the applied sweeping current:

$$\phi = DI^2 + EI, \tag{3.14}$$

If the phase shifter is more nonlinear, one can use a higher order polynomial to fit the phase change. Next, one can substitute the above equation into the power fitting function,

$$P(I) = A + B\cos\Big(DI^2 + EI + C\Big). \tag{3.15}$$

However, both ports are connected to other MZIs, which means the light on the two output ports will spread into all other output ports. Even more, if the circuit is deep enough, the two output ports will interfere with each other again, with respect to the circuit coming after the MZI. Therefore, the key to calibrating the MZI is to guide the light from one MZI to the output port, with as little noise as possible.

Here, a common way to calibrate the circuit is to follow the so-called *diagonal principle* to calibrate first the internal phase shifters, and then the external phase shifters [133]. We show the example of a 6-mode circuit in three steps as follows.

Single-MZI calibration

The inner phase determines the phase difference between two arms of a single MZI, thus the output power should reflect the phase change. As illustrated in Fig. 3.7, one could first choose the MZIs from top left to bottom right, namely sending light from the top left port and measuring the output power at the bottom right port. The reason is that even if each MZI is not perfectly balanced, the light leaked will not enter other output ports. For each phase shifter, one can ramp the current from 0 to some value, measure the output power, and fit the power using the function in 3.15. Once the previous phase shifter is calibrated, it is better to set the phase to zero, i.e., $\theta_i = 2\pi$. In this way, the phase

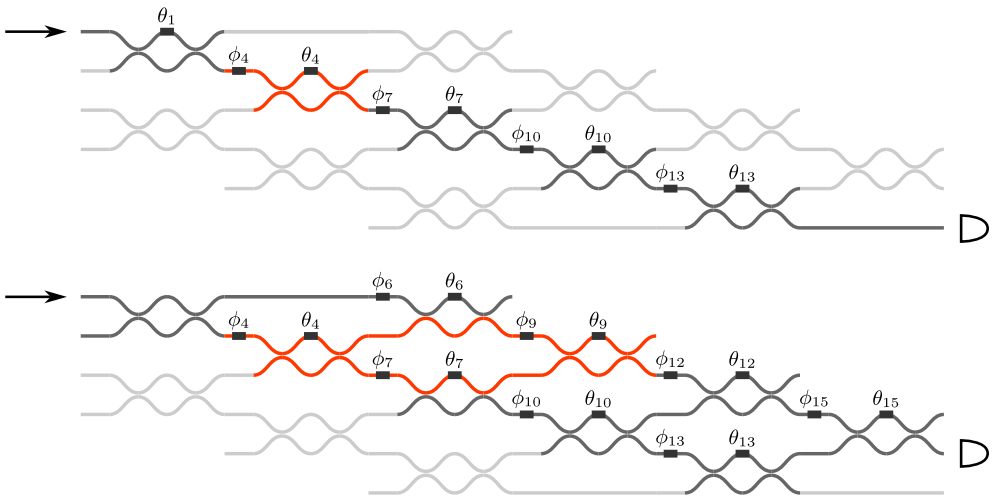

Figure 3.7: Calibration of a universal circuit.

shifters $\theta_1, \theta_4, \theta_7, ...\theta_{13}$ can be calibrated. After the first N inner phase shifters, we could move to off-diagonal MZIs, for example sending light from the third port and measuring the output power at the sixth port as well. Here, since the light MZI in the first diagonal is already calibrated, one can set the phase here as $\theta_{13} = \pi$ to make sure the light is sent to the bottom right port. By repeating the above process, one can calibrate all inner phase shifters.

Meta-MZI calibration

For the rest of the phase shifters ϕ_i, since they do not control optical power directly in each MZI, it is necessary to construct a larger MZI, which is called a *meta-MZI*. As shown in the lower part in Fig. 3.7, the meta-MZI is constructed by four MZIs. Two on the top and bottom act as identity by applying phase as π, and the two on the left and right act as beam splitters by setting the phase as $\pi/2$. After this, we follow the same procedure as above, sending light from the top left port and measuring the output power at the end of the following MZIs. For example, the laser is connected to the top left port, and the output power is measured at the second bottom right port. Here, by tuning each phase shifter ϕ_7 or ϕ_9, one could measure the output power and fit the power. Doing the same procedure and shifting to the next diagonal MZI, one can calibrate all external phase shifters ϕ_i.

Crosstalk compensation

To lower the electric crosstalk, it is achievable to use current driving rather than voltage driving. A more effective way is to use separate ground pins for each phase shifter, but this challenges the routing of the circuit. As discussed above, to compensate the crosstalk, one needs to measure the crosstalk matrix X and then invert it. The challenge lies not only in that the crosstalk can be nonlinear, but also that integer multiples of 2π can be added to the solution in Eq. 3.11.

Recently, there is a new tendency to utilize machine learning methods to calibrate PIC [134]. By training a neural network with current or voltage as input, and the output phase or optical power as output, the neural network can learn the relationship between the input and output, thus to represent the crosstalk and error implicitly.

To summarize, using PIC for both classical optical and quantum computing is a promising approach to save energy costs and increase scalability for future novel computing architectures. Many efforts can be explored not only to design and fabricate large scale and high-performance circuits [39, 114] but also to develop specific and advanced protocols to operate on PIC. Beyond silicon photonics, other material platforms, such as lithium niobate, aluminum nitride, and indium phosphide, are also being actively investigated for PIC applications and integrated to silicon photonics as hybrid integrated photonics [33, 34]. By combining PIC with other quantum technologies, such as on-chip photon sources, quantum computing using PIC will be a promising direction for future quantum information processing [31, 32, 117, 118].

4 Quantum Machine Learning

Quantum machine learning (QML) is an interdisciplinary field that combines quantum computing and machine learning [135–137]. This field has emerged to explore the potential advantages that quantum mechanics might offer for machine learning tasks, particularly in the current era of noisy intermediate-scale quantum (NISQ) devices with limited qubit counts [138, 139]. The fundamental question driving QML research is whether quantum computers can provide computational advantages over classical computers for learning tasks, either through faster training, better generalization, or the ability to solve problems that are intractable for classical methods [140–143]. Furthermore, QML offers new perspectives on understanding quantum advantage and the nature of learning itself, for example by exploring sampling times and estimation errors in quantum systems [144–146].

Before delving into specific QML algorithms, it is essential to clarify the scope of this field. Since many subjects involving either machine learning or artificial intelligence [147], and quantum computing, can be included in the field of QML, it is common to classify this field from two perspectives: the method approach and the data or the problem to be solved, as shown in Fig. 4.1. The method approach distinguishes between classical and quantum algorithms, while the data perspective considers whether the input data is classical (such as images or text) or quantum (such as quantum states or quantum channels). This classification framework helps organize the diverse research directions within QML and clarifies the specific advantages and challenges associated with each approach.

For example, neural networks can be trained to solve many-body problems, such as solving the ground states of a many-body system [148]. Of course, this approach relies on existing solutions of a given quantum system. The model can then be deployed on classical computers, such as GPU clusters, and trained based on known data [149, 150]. Another example is quantum state tomography, which can be solved using neural networks to represent quantum states and training the network to fit measurement data [151]. Another distinct approach is to use few local random measurements to reconstruct global properties of quantum states, such as state fidelity or entanglement entropy [152, 153], which significantly reduces the training cost.

Second, one can adopt quantum models to solve classical machine learning tasks. For example, one might use a quantum circuit to classify classical data. Such research focuses on two main aspects: first, how to map classical data into quantum states; second, how to use quantum computers to accelerate

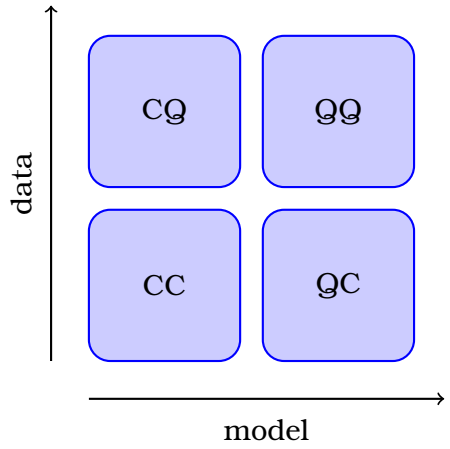

Figure 4.1: Classification of quantum machine learning according to method used and data type. The model can be either classical or quantum, and the data can be either classical or quantum. We target the approach of using quantum models to solve both classical and quantum data in this dissertation.

training. In particular, this involves benchmarking performance against existing classical models, such as distinguishing handwritten digits by encoding pixels into phases of quantum gates and training the phases based on measurement results [154–160]. However, to the best of our knowledge, there is no evidence that quantum models can outperform existing classical models on classical data. Nevertheless, exploring the potential of quantum models remains an interesting research topic [161, 162].

Third, both the model and the data can be quantum [152, 163]. This means that rather than using conventional quantum algorithms, one can choose an ansatz as the initial solution and optimize the parameters during an iterative training process that converges to the optimal solution. For example, one could use a variational quantum algorithm (VQA) to solve quantum problems, such as finding the ground state of a many-body system [59, 164]. This direction is promising and attractive since several problems in quantum physics have been proven to be advantageous when using a machine learning approach, such as learning fixed properties of quantum systems [144, 165]. Of course, one can also use classical machine learning methods to solve classical problems, but this is not the focus of this dissertation.

In this dissertation, we focus on the approach of using quantum models to solve problems involving both classical and quantum data. Here we introduce the basic concepts of classical machine learning and variational quantum algorithms, which form the foundation of quantum machine learning.

4.1 Classical machine learning

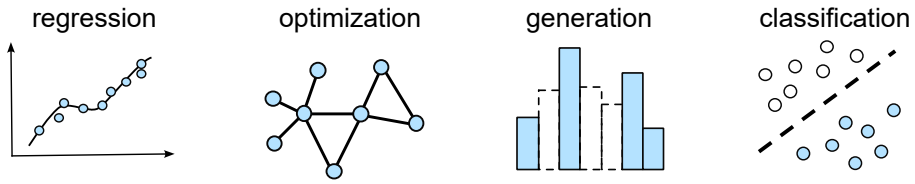

Figure 4.2: Various machine learning tasks. The core is to learn the statistics of all forms of data using a specific model based on a training strategy, thus enabling prediction of outputs for unseen data.

The core of machine learning is to learn the statistics of data using a specific model based on a training strategy. Given the variance in data, the model must be *intelligent* enough to fit the data distribution. Depending on whether the data has labels, i.e., whether the output is known, learning tasks can be classified into two categories: supervised learning and unsupervised learning.

In supervised learning, the data is associated with given labels. For example, the data points might be pixels of handwritten digits with the corresponding digits as labels. In unsupervised learning, the data is not initially associated with labels; there may exist some unknown number of classes. For the model, it must express the data distribution and determine the labels independently. In this dissertation, we focus on supervised learning, as it is more general and widely used in practice.

For example, we can simplify the learning process as learning a function $f : \boldsymbol{x} \mapsto \boldsymbol{y}$ from a set of known data $\{\boldsymbol{x}_i, \boldsymbol{y}_i\}_{i=1}^{N}$, which is called the *training set*.

$$T = \{(\boldsymbol{x}_1, \boldsymbol{y}_1), (\boldsymbol{x}_2, \boldsymbol{y}_2), \ldots, (\boldsymbol{x}_N, \boldsymbol{y}_N)\}^T. \tag{4.1}$$

Here, each data point $\boldsymbol{x}_i \in \mathcal{X}$ is a vector of d dimensions or d features, $\boldsymbol{x}_i = (x_i^{(1)}, x_i^{(2)}, \ldots, x_i^{(d)})$. The same applies to the label $\boldsymbol{y}_i \in \mathcal{Y}$. We call $\boldsymbol{x}_i$ the input data points and $\boldsymbol{y}_i$ the corresponding labels or outputs.

If there exists a function f determined by parameters $\boldsymbol{\theta} = \{\theta_1, \theta_2, \ldots, \theta_n\}$, which gives the output corresponding to each input data point $\boldsymbol{x}_i$:

$$\tilde{\boldsymbol{y}}_i = f_{\boldsymbol{\theta}}(\boldsymbol{x}_i). \tag{4.2}$$

If the output $\tilde{\boldsymbol{y}}_i$ is sufficiently close to each label $\boldsymbol{y}_i$, we say the function $f_{\boldsymbol{\theta}}$ has learned the data distribution. Note that the learning function $f_{\boldsymbol{\theta}}$ is a variational function determined by the parameters $\boldsymbol{\theta}$.

Loss function

To quantify learning performance, we need a function to measure the difference between the predicted output $\tilde{\boldsymbol{y}}_i$ and the true output $\boldsymbol{y}_i$. Since the learning

target is to minimize the difference between the predicted output $f_{\boldsymbol{\theta}}(\boldsymbol{x}_i)$ and the true output $\boldsymbol{y}_i$, we define the *loss function* as

$$\mathcal{L}(\boldsymbol{\theta}) = \frac{1}{N}\sum_{i=1}^{N} |f_{\boldsymbol{\theta}}(\boldsymbol{x}_i) - \boldsymbol{y}_i|^2. \tag{4.3}$$

Sometimes, this is called *empirical risk*, $\mathcal{L}_{\text{emp}}$, indicating that the loss function is calculated based on the training set and does not always accurately reflect the true data distribution. Here, the difference is measured by the squared error, also called the L_2 distance. There are many other forms of loss functions, such as cross-entropy loss or Kullback-Leibler divergence. Once the loss function is defined, the learning process aims to minimize the loss function with respect to the parameters $\boldsymbol{\theta}$, i.e., to find a global minimum of the loss function.

To minimize the loss function, we design a strategy to update the parameters $\boldsymbol{\theta}$. Most commonly, the parameters $\boldsymbol{\theta}$ are optimized using gradient descent methods, such as stochastic gradient descent (SGD). In SGD, the parameters are updated iteratively by the following rule:

$$\boldsymbol{\theta}_{t+1} \leftarrow \boldsymbol{\theta}_t - \eta \nabla_{\boldsymbol{\theta}} \mathcal{L}, \tag{4.4}$$

where η is the learning rate, which controls the step size of the update, and $\nabla_{\boldsymbol{\theta}}\mathcal{L}$ is the gradient of the loss function with respect to the parameters $\boldsymbol{\theta}$.

However, since the parameters are usually high-dimensional, calculating the gradient is non-trivial. The gradient is defined by

$$\nabla_{\boldsymbol{\theta}}\mathcal{L} = \left(\frac{\partial \mathcal{L}}{\partial \theta_1}, \frac{\partial \mathcal{L}}{\partial \theta_2}, \ldots, \frac{\partial \mathcal{L}}{\partial \theta_n}\right), \tag{4.5}$$

where $\nabla_{\boldsymbol{\theta}}$ is the gradient operator with respect to the parameters $\boldsymbol{\theta}$. This process is called *backpropagation*. If one can calculate the gradient efficiently, the training process becomes much more efficient and faster.

The training process is repeated until the loss function converges, i.e., reaches the global minimum or at least a local minimum. For example, when the loss function falls below a certain threshold ϵ, the training process is stopped. The optimal parameters $\boldsymbol{\theta}^*$ are found by minimizing the empirical risk:

$$\boldsymbol{\theta}^* = \arg\min_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta}). \tag{4.6}$$

If the training set is not very large, one can use the entire training set to calculate the loss function and update the parameters. Otherwise, the training process can be performed on a subset of the training set, called *batch training*. The size of the mini-batch is a hyperparameter that can be tuned to achieve better performance.

After training, the learned function is used to predict the output of previously unseen data, called the *test set*. To verify whether the learned function

generalizes well, the *test loss* is calculated by

$$\mathcal{L}_{\text{test}}(\boldsymbol{\theta}^*) = \frac{1}{M}\sum_{j=1}^{M} |f_{\boldsymbol{\theta}^*}(\boldsymbol{x}_j) - \boldsymbol{y}_j|^2, \tag{4.7}$$

where $\{\boldsymbol{x}_j, \boldsymbol{y}_j\}_{j=1}^{M}$ is data from the test set.

Both the training loss and test loss depend on the data selection, meaning they do not always reflect the true data distribution. Thus, the *expected risk* is defined as the average loss over all possible data points:

$$\mathcal{L}_{\text{exp}}(\boldsymbol{\theta}) = \mathbb{E}_{\boldsymbol{x}\sim\mathcal{X},\boldsymbol{y}\sim\mathcal{Y}}|f_{\boldsymbol{\theta}}(\boldsymbol{x}) - \boldsymbol{y}|^2. \tag{4.8}$$

The expected risk is sometimes used to estimate the *generalization error*, which indicates how well the learned function $f_{\boldsymbol{\theta}}$ generalizes to unseen data. For example, if the training loss is very small but the test loss is very large, we say the model does not generalize well and suffers from overfitting.

Models

Conventional machine learning methods include a wide range of algorithms, such as linear regression, logistic regression, support vector machines (SVM), decision trees, and neural networks [166]. A simple way to classify these methods is into two categories: linear and nonlinear methods. For example, SVM and linear regression are linear methods, while neural networks are nonlinear methods.

We now take SVM as an example of linear methods. As a binary classifier, SVM aims to find a hyperplane to separate data points into two classes using a sign function:

$$f(\boldsymbol{x}) = \text{sign}(\boldsymbol{\omega}^T\boldsymbol{x} + b), \tag{4.9}$$

where $\boldsymbol{\omega}$ is the weight vector and b is the bias. To solve for the parameters $\boldsymbol{\omega}$ and b, this is equivalent to solving a linear optimization problem with constraints. More details can be found in Chapter 5.

Nevertheless, linear models have limitations as they may not capture complex patterns in the data. For example, they cannot handle data that is separated by a hypersphere rather than a hyperplane. To overcome this limitation, one can use the kernel trick to map input data into a higher-dimensional space, where the data may become linearly separable [167, 168].

Another direction is to use more complex models. Neural networks have been very successful in this regard. Initially, perceptrons were proposed as simple models of artificial neurons [169]. Later, recurrent neural networks (RNNs) were introduced to handle sequential data, such as time series or natural language [170–174]. For example, a two-layer feedforward neural network can be expressed as

$$\begin{aligned} \boldsymbol{y} &= \sigma(\boldsymbol{w}_1\boldsymbol{x} + \boldsymbol{b}_1), \\ \boldsymbol{z} &= \sigma(\boldsymbol{w}_2\boldsymbol{y} + \boldsymbol{b}_2), \end{aligned} \tag{4.10}$$

where σ is a nonlinear activation function, such as the sigmoid function or the rectified linear unit (ReLU) function. $\boldsymbol{w}_1$ and $\boldsymbol{w}_2$ are weight matrices, and $\boldsymbol{b}_1$ and $\boldsymbol{b}_2$ are bias vectors at each layer. This model can be extended to more layers; for example, the output at the l-th layer is given by a nested function

$$\boldsymbol{z} = \boldsymbol{y}_l(\boldsymbol{y}_{l-1}(\dots \boldsymbol{y}_2(\boldsymbol{y}_1(\boldsymbol{x})))). \tag{4.11}$$

One of the most important properties of neural networks is that one can derive the gradient of the loss function with respect to the parameters using backpropagation [172]. A significant advantage of neural networks is that the gradient of the loss function can be calculated efficiently using the chain rule. Each term in the gradient is given by

$$\frac{\partial \mathcal{L}}{\partial \theta_i} = \sum_i \frac{\partial \mathcal{L}}{\partial z_i} \cdot \frac{\partial z_i}{\partial \theta_i} = \sum_i \sum_j \frac{\partial \mathcal{L}}{\partial z_i} \cdot \frac{\partial z_i}{\partial y_j} \cdot \frac{\partial y_j}{\partial \theta_i}, \tag{4.12}$$

where $\boldsymbol{\theta}$ is the set of parameters, including the weights and biases of the neural network. Using backpropagation, the gradient calculation can be performed efficiently at each layer and multiplied together, significantly reducing computational cost.

4.2 Variational quantum algorithms

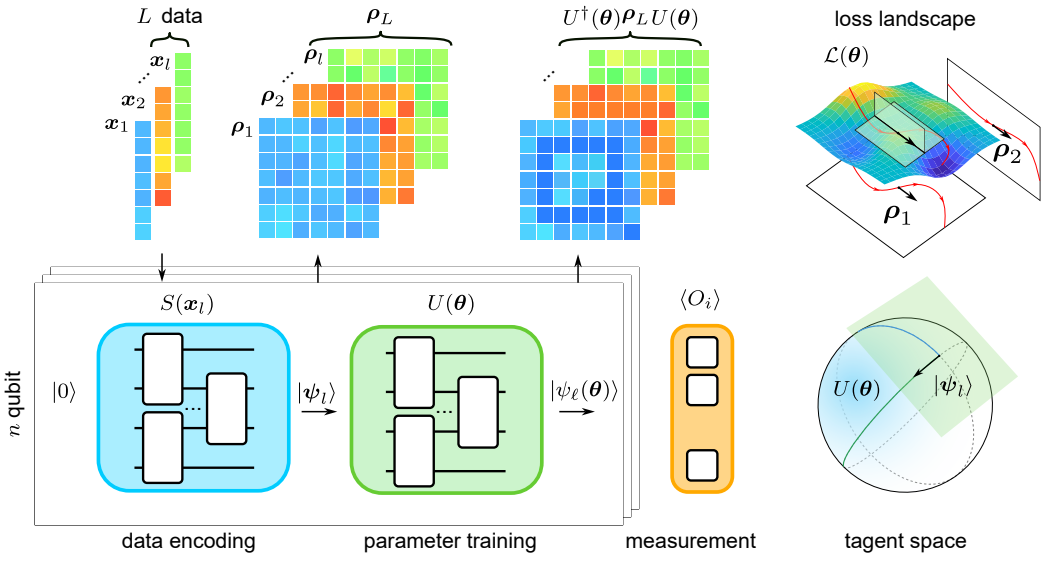

Figure 4.3: A quantum circuit to perform a variational quantum algorithm.

Before introducing QML, we note that the conventional approach to solving quantum problems is to design specific quantum circuits that specify the input

qubits and quantum gates. This approach has been successful in many quantum algorithms. For example, Grover's search algorithm [8] requires $O(\sqrt{N})$ queries on an oracle and reflection operations to find the marked item from an unsorted database. In Shor's algorithm [7] for integer factorization, the quantum Fourier transform (QFT) is used to locate the period of a function, which is key to this algorithm. These algorithms are designed for specific problems, meaning both the quantum circuit and the parameters in the circuit are generally fixed.

Parameterized quantum circuits

Apart from classical models, quantum circuits can be treated as models or functions for machine learning, similar to neural networks [164, 175–178]. First, they have input and output, namely quantum states before and after the circuit. Second, they are parameterized by adjustable classical parameters and can be connected using specific architectures, such as entangling gates.

Consider a quantum circuit with n qubits, where the given circuit ansatz contains K layers of quantum gates:

$$U(\boldsymbol{\theta}) = \prod_{k=1}^{K} U_k(\boldsymbol{\theta}_k), \tag{4.13}$$

where $U_k(\boldsymbol{\theta}_k)$ is the unitary operator of the k-th layer, parameterized by $\boldsymbol{\theta}_k$. For example, it can be a sequence of single-qubit rotations $R_x(\theta), R_z(\theta)$ and two-qubit entangling gates such as CNOT gates. More generally, the unitary operator can be expressed as a sequence of two-qubit arbitrary gates, namely $\mathrm{SU}(4)$ gates.

In particular, if the data is classical, one must encode the classical data $\boldsymbol{x}_\ell$ into quantum states first [158, 179]. Thus the circuit becomes the product of an encoding circuit $S(\boldsymbol{x})$ and a parameterized circuit $U(\boldsymbol{\theta})$, as illustrated in Fig. 4.3.

$$|\psi_\ell(\boldsymbol{\theta})\rangle = U(\boldsymbol{\theta}) S(\boldsymbol{x}_\ell) |0\rangle \tag{4.14}$$

In the case of a pure quantum state, the output state of the quantum circuit can be given by

$$|\psi_\ell(\boldsymbol{\theta})\rangle = U(\boldsymbol{\theta}) |\psi_\ell\rangle \tag{4.15}$$

Here, without loss of generality, we assume the input state is the ground state $|0\rangle^{\otimes n}$, i.e., all qubits are initialized to the state $|0\rangle$. In density matrix form, the output state can be expressed as

$$\rho_\ell(\boldsymbol{\theta}) = |\psi_\ell(\boldsymbol{\theta})\rangle \langle\psi_\ell(\boldsymbol{\theta})| = U(\boldsymbol{\theta}) |\psi_\ell\rangle \langle\psi_\ell| U^\dagger(\boldsymbol{\theta}). \tag{4.16}$$

An intuition for the expressivity of parameterized quantum circuits is that the circuit can be seen as a unitary operator acting on the Hilbert space of n qubits, which is $2^n \times 2^n$-dimensional. Since one can perform arbitrary unitary $\mathrm{SU}(4)$ operations on each pair of qubits (requiring 16 real parameters), the circuit

depth must reach $O(2^n)$ to achieve full expressivity of the Hilbert space. Thus, a major motivation for using quantum circuits is to find the most efficient ansatz that can approximate the target function with reasonable circuit depth.

Measurement and cost function

Now we consider the measurement set $O = \{O_1, O_2, \ldots, O_m\}$, which is a set of observables that can be measured on the output state. The expectation value of each observable O_i is given by

$$\begin{aligned} \langle O_i \rangle &= \mathrm{Tr}[O_i \rho_\ell(\boldsymbol{\theta})] \\ &= \mathrm{Tr}\Big[O_i U(\boldsymbol{\theta}) \left|\psi_\ell\right\rangle \left\langle\psi_\ell\right| U^\dagger(\boldsymbol{\theta})\Big] \end{aligned} \tag{4.17}$$

where $\rho_\ell(\boldsymbol{\theta})$ is the density matrix of the output state.

The VQA algorithm aims to minimize a cost function $\mathcal{L}(\boldsymbol{\theta}, \langle O_i \rangle)$ that depends on the expectation values of observables O_i with respect to parameters $\boldsymbol{\theta}$. The choice of cost function depends on the specific task and can be very problem-specific.

A typical cost function using a single observable O is the weighted sum of expectation values:

$$\begin{aligned} \mathcal{L}(\boldsymbol{\theta}) &= \frac{1}{L} \sum_\ell \mathrm{Tr}\Big[OU(\boldsymbol{\theta}) \left|\psi_\ell\right\rangle \left\langle\psi_\ell\right| U^\dagger(\boldsymbol{\theta})\Big] \\ &= \frac{1}{L} \sum_\ell \mathrm{Tr}\left[OU(\boldsymbol{\theta}) \sum_\ell \left|\psi_\ell\right\rangle \left\langle\psi_\ell\right| U^\dagger(\boldsymbol{\theta})\right] \\ &= \mathrm{Tr}\Big[OU(\boldsymbol{\theta}) \rho_L U^\dagger(\boldsymbol{\theta})\Big], \end{aligned} \tag{4.18}$$

where $\rho_L = \frac{1}{L} \sum_\ell \left|\psi_\ell\right\rangle \left\langle\psi_\ell\right|$ is the density matrix of the training set.

Gradient descent

As in classical machine learning, the parameters $\boldsymbol{\theta}$ can be optimized using gradient descent methods:

$$\boldsymbol{\theta}_{t+1} \leftarrow \boldsymbol{\theta}_t - \eta \nabla_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta}, \langle O_i \rangle). \tag{4.19}$$

However, unlike classical neural networks, there is no explicit general analytical expression for the gradient in quantum circuits, except for certain cases [180, 181]. Thus, a numerical differentiation method is often used to calculate the gradient. For example, the finite difference method gives the partial derivative with respect to parameter θ_i of the cost function $\mathcal{L}$ as

$$\frac{\partial \mathcal{L}}{\partial \theta_i} = \frac{\mathcal{L}(\theta_+) - \mathcal{L}(\theta_-)}{2\lambda}, \tag{4.20}$$

where λ is the finite difference step and $\theta_{\pm} = \theta_i \pm \lambda$ correspondingly. This approach requires estimating each partial derivative separately, requiring $2K + 1$ circuit evaluations for K parameters. In practice, estimating just one gradient can take considerable time and computational resources.

Therefore, gradient-free methods are often more favorable to avoid the high cost of estimating gradients. A typical method is *Simultaneous perturbation stochastic approximation* (SPSA) [182]. Rather than using the gradient estimated by the finite difference method, SPSA estimates the gradient using a random direction, reducing the cost of estimating the gradient to just 2 circuit evaluations (one for the positive perturbation and one for the negative perturbation). On the other hand, it does not provide accurate information about the loss landscape and consequently does not guarantee convergence to the optimal solution. The convergence rate is also typically slower than standard gradient descent methods. Overall, the runtime of a VQA algorithm is determined by the number of data points N, the number of parameters K, and the number of iterations I, scaling as $O(NKI)$.

5 Experimental Quantum-Enhanced Kernel-Based Machine Learning On A Photonic Processor

Recently, machine learning had a remarkable impact, from scientific to everyday-life applications. However, complex tasks often imply unfeasible energy and computational power consumption. Quantum computation might lower such requirements, although it is unclear whether enhancements are reachable by current technologies. Here, we demonstrate a kernel method on a photonic integrated processor to perform a binary classification. We show that our protocol outperforms state-of-the-art kernel methods including gaussian and neural tangent kernels, exploiting quantum interference, and brings a smaller improvement also by single photon coherence. Our scheme does not require entangling gates and can modify the system dimension through additional modes and injected photons. This result opens to more efficient algorithms and to formulating tasks where quantum effects improve standard methods.

The past decades have witnessed a swift development of technologies based on quantum mechanical phenomena, which have opened up new perspectives in a wide spectrum of applications. These range from the realization of a global-scale quantum communication network, the Quantum Internet, to the simulation of quantum systems, to quantum computing [183, 184]. In particular, the interest towards the last field has been fueled by some milestone discoveries, such as Shor's factorization and Grover's search algorithm [7, 8], which have promised that quantum processors can outperform their classical counterparts. However, a clear advantage of quantum computation has been experimentally demonstrated only recently and on different computational tasks, boson sampling [24, 94, 185] and random circuit sampling [186], which do not have clear practical applications.

Given these premises, our goal is to investigate the tasks in which quantum computing can enhance the operation of classical computers for practically relevant tasks. Moreover, the question is whether this can be achieved for

Contribution statement: This chapter is a reprint of the published paper: Zhenghao Yin, et. al "Experimental Quantum-Enhanced Kernel-Based Machine Learning On A Photonic Processor", *Nature Photonics* 19, 1020-1027 (2025). In this work, I am the main contributor to the theoretical model, the data analysis, the experimental design, and the writing of the manuscript.

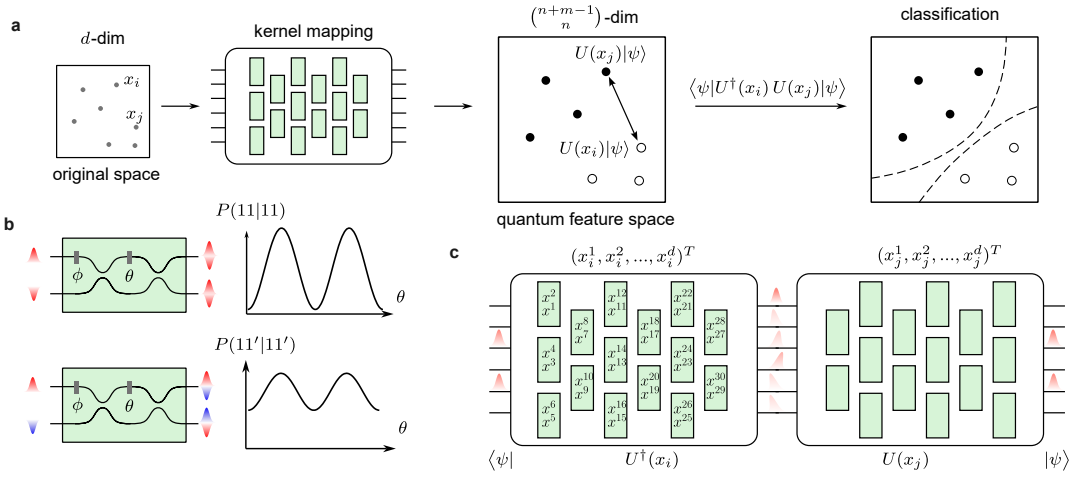

Figure 5.1: **Photonic quantum kernel estimation. a**. The photonic quantum kernel maps each data point x_i to be classified from a d-dimensional space into a quantum state $|\Phi\rangle_i$, living in a Hilbert feature space. In detail, the classical data x_i is encoded into a unitary evolution $U(x_i)$ applied on a fixed input state $|\psi\rangle$. This implies $|\Phi\rangle_i = U(x_i)|\psi\rangle$. After mapping all the data points in the dataset, from the inner pairwise products, we perform the classification finding the hyperplane best separating the classes, i.e. through a support vector machine (SVM), according to Eq. (5.3). **b.** Pairs of indistinguishable photons and distinguishable photons show a different behaviour when injected in a Mach-Zehnder interferometer (MZI). Here, input states 11 and $11'$ indicate, respectively, two indistinguishable and distinguishable photons being injected in the circuit and being detected at the output modes. **c.** Estimation of the inner product of two data points x_i and x_j by encoding them in two unitaries $U(x_i)$ and $U(x_j)$. The inner product $\langle\phi_j|\phi_i\rangle$ amounts to $\langle\psi|U^\dagger(x_i)U(x_j)|\psi\rangle$. This is equivalent to projecting the evolved state $U^\dagger(x_j)U(x_i)|\psi\rangle$ onto $|\psi\rangle$. Each box represents a programmable MZI with two free parameters (namely a beam splitter with tunable reflectivity and phase), as shown in **b**.

problems that are now within the reach of state-of-art technology, where only noisy intermediate-scale quantum computers are available [46]. In this context, a flurry of interest has been devoted to the open question of whether the new paradigm of quantum computing can have an impact on machine learning [49], which has revolutionized classical computation, granting new possibilities and changing our everyday lives, from email filtering to artificial intelligence. The two main directions that have been investigated until now are, on one side, whether quantum computation could improve the efficiency of the learning process,

allowing us to find better optima with the need of a lower number of inquiries [57, 187–189] and, on the other, how quantum behaviours can enhance the expressivity of the input encoding, exploiting correlations between variables that are hard to reproduce through classical computation [190, 191].

In this context, a straightforward application of quantum computing on kernel models has become evident. Kernel methods are widely used tools in machine learning [168, 192], that base their functioning on the fact that patterns for data points, which are hard to recognize in their original space, can become easy to identify once mapped nonlinearly to a *feature space*. When the suitable mapping is performed, it is possible to identify the hyperplane which best separates the classes of feature data points, through a support vector machine [193] (SVM), according to the inner product of the mapped data. Let us note that the only part of the model that is trained is the SVM, which is efficient, once the inner products are available. Hence, an interesting question is whether using a quantum apparatus to perform the data mapping and evaluate the inner products can enhance the overall performance of the algorithm, benefitting from feature maps exploiting the evolution of quantum systems and outsourcing the hardest part of the computation to the quantum hardware. This question was theoretically and rigorously answered in the affirmative by [194], although the implementation of the proposed task is out of reach for state-of-the-art technologies. Hence, it is still an open question, whether some improvement in the performance of such a method can be achieved through a quantum feature map, for tasks that can be implemented on current quantum platforms. In this context, a moderately-sized quantum feature space can be proven to be more suitable to preserve the similarity among data that belongs to same class. The reason for that is that the advantage of a high-dimensional feature space, allowing to improve the expressivity of the model, is canceled out by the dimension increase per se that typically decreases the overlap between different states, such that they become almost orthogonal [145]. This is due to the fact that the expected overlap value of two random states amounts to the inverse of the space dimension and, in high-dimensional metric spaces, almost all probability measures concentrate around the mean or median of a function (*concentration of measure* [195]). The effect is then that the inner products among the feature data points will tend to vanish and a higher precision will be required to properly measure them. This ultimately leads to the so-called *exponential concentration* [196, 197], where it is not possible to capture the information about the similarities between states, making the model, for all intents and purposes, useless for any type of classification.

In this work, we investigate this topic and experimentally demonstrate a quantum kernel estimation, where feature data points are evaluated through the unitary evolution of two-boson Fock states (see Fig. 5.1). Such encoding, even for relatively small dimensions, proves suitable, as it allows to achieve high classification accuracies.

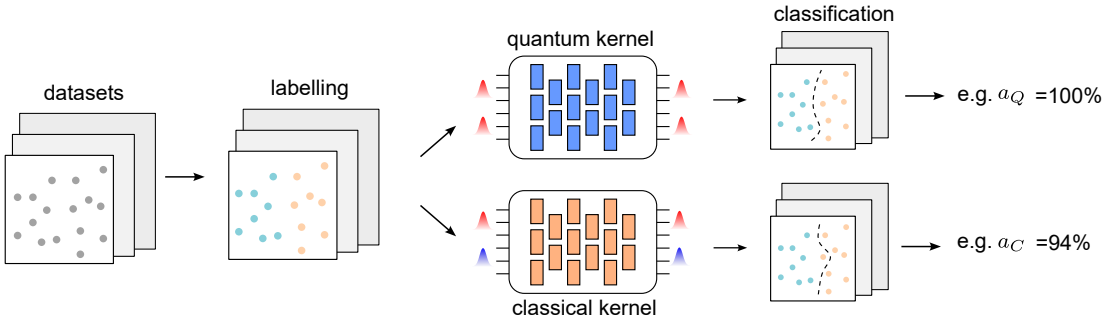

Figure 5.2: **Classification tasks for photonic kernel methods.** The datasets are randomly generated and consist in d- dimensional vectors, with entries between 0 and 1. Then, we randomly assign labels to each point as belonging to class +1 or -1 and we test the ability of our photonic kernels, displaying and not displaying quantum interference (respectively indicated as *quantum kernel* and *classical kernel*), to correctly classify the data. This is quantified by the accuracy of our models, which we indicate as a_Q and a_C.

Furthermore, we show that for given tasks, this algorithm leads to an enhancement in the performance of quantum kernels with respect to their classical counterparts. These tasks are selected by maximizing the so-called *geometric difference*, which measures the separation in performance between a pair of kernels [145]. In particular, we separate between quantum and classical kernels, that is, photonic kernels that do or do not exhibit quantum interference.

In our implementation, we exploit a photonic platform based on an integrated photonic processor [112] where we inject two-boson Fock states to map the data to be classified (see Fig. 5.1a). To estimate quantum and classical kernels, we inject indistinguishable and distinguishable photons, respectively. This photonic platform is particularly suitable for this task, as it allows us to encode and manipulate our input data with high fidelity.

To benchmark our enhanced performance, we compare classification accuracies between photonic kernels and state-of-the-art classical computational kernels. These include the standard gaussian kernels [168, 192], as well as the recently introduced neural tangent kernels [198], which simulate gradient descent over infinitely-wide neural networks. Our results show that photonic kernels outperform classical methods and that the accuracies are further enhanced in kernels displaying quantum interference.

5.1 Photonic quantum kernel estimation

The core of a kernel method is constituted by a function that maps non-linearly a set of N data points into a *feature space*. In its simplest version, these data points belong to two classes and the final goal of the algorithm is to separate them accordingly, by using only a linear model, i.e. a hyperplane. This is feasible since the map, being non-linear, will change the relative position between the data points, making the dataset (in general) easier to classify. For this reason, a support vector machine (SVM) can effectively find a hyperplane separating the two classes, with a high accuracy. From a more mathematical perspective, we can describe the map as a function that sends N input data points x_i, on which we wish to perform binary classification, from a space $\mathcal{X} \subseteq \mathbb{R}^d$ into a Hilbert space $\mathcal{H}$. Here, d is the dimension of each data point. This is done through a feature map $\Phi : \mathcal{X} \to \mathcal{H}$. Then, a SVM can be used to produce a *prediction* function $f_K : \mathcal{X} \to \mathbb{R}$ as

$$f_K(x) = \sum_i \alpha_i K(x, x_i), \tag{5.1}$$

where these α_i coefficients are obtained by solving a linear optimization problem. The inputs of the optimization are the labels y and the matrix obtained by computing the pairwise distances between data points is

$$K_{i,j} = K(x_i, x_j) = |\langle \Phi(x_i) | \Phi(x_j) \rangle|^2, \tag{5.2}$$

the so-called *Gram matrix* (see Supplementary Note 1 for further information).

In this work, we implement a quantum version of the kernel method, in which the aforementioned pairwise distances between data points, which belong to a class y taking values $+1$ or -1, are estimated by sampling from the output probability distribution arising from the unitary evolution of a Fock input state. This process is depicted in Fig. 5.1a. Therefore, our feature map plugs the data that needs to be classified into the free parameters defining a unitary evolution applied to a fixed Fock state of dimension m and whose sum of occupational numbers is n: $x \mapsto |\Phi(x)\rangle = U_x|\psi\rangle$. Here, $|\psi\rangle$ is the encoding state which is free to choose. Then, as shown in Fig. 5.1c, the pairwise inner products of the feature points are experimentally evaluated, as $|\langle \psi | U(x_i)^\dagger U(x_j) | \psi \rangle|^2$. Such unitaries can be effectively implemented by a programmable photonic circuit consisting of an array of Mach-Zehnder interferometers (MZIs)[56]. Hence, the dimension of the feature Hilbert space $\mathcal{H}$ will be $\binom{n+m-1}{n}$. At this point, the SVM finds the hyperplane separating the training data points through the aforementioned optimization process [199] and, the binary classification of unknown points x is given by the following relation:

$$y = \text{sign}\left(\sum_{i=1}^{N} \alpha_i y_i K(x, x_i) \right) \tag{5.3}$$

where α_i are the coefficients optimized in the training process and y_i is the class of the i-th point in the training. This model is defined *implicitly*, as the labels are assigned by weighted inner products of the encoded data points [157, 200–204].

If the Fock state contains indistinguishable bosons, they will exhibit quantum interference, as shown in Fig. 5.1b. In this case, the output probability distribution is given by the permanents of sub-matrices of the matrix representing the unitary evolution of the input [205]. More specifically, considering an input configuration s, the probability of detecting the output configuration t is given by $|\mathrm{Per}U_{s,t}|^2/\Pi_i^m s_i!\Pi_i^m t_i!$. Here, $\mathrm{Per}(\cdot)$ denotes the permanent matrix operation, s_i and t_i are the occupational numbers at the i-th mode and $U_{s,t}$ is the sub-matrix obtained by selecting the rows/columns corresponding to the occupied modes of the input/output Fock states. On the other hand, if the bosons are distinguishable, they will not exhibit quantum interference. In this case the probability will amount to $\mathrm{Per}|U_{s,t}|^2/\Pi_i^m s_i!\Pi_i^m t_i!$.

In the following, we will refer to a kernel implemented with indistinguishable bosons as a *quantum kernel*,

$$K_Q(x_i, x_j) = |\mathrm{Per}U_\psi(x_i, x_j)|^2/N' \tag{5.4}$$

and with distinguishable ones as a *classical kernel*,

$$K_C(x_i, x_j) = \mathrm{Per}|U_\psi(x_i, x_j)|^2/N' \tag{5.5}$$

Here, $U_\psi(x_i, x_j)$ is the matrix defined by data points x_i, x_j and selected encoding state ψ. N' is the coefficiency related to ψ. As long as there is at most one photon in each mode of ψ, $N' = 1$.

5.2 Classification task

At this point, we need to select a classification task that would benefit from the described model. Our strategy will then consist, first, in generating random data points and, second, in selecting the proper labeling, which would boost the performance of the quantum kernel (K_Q), with respect to the classical one (K_C), as depicted in Fig. 5.2. So, we need a quantifier which estimates the difference in the performance of the two models and, to this aim, we select the *model complexity*, which is defined as $s_K(y) = y^T K y$, where K is the kernel and y the labeling. This quantity is proportional to the upper bound on the classification error of a given model and it amounts to the number of features required to make accurate predictions (see the Supplementary Note 2). Hence, the potential advantage given by K_Q, with respect to K_C, is based on finding the largest possible separation between the two model complexities [145]. Therefore, our goal is to find a labelling y, such that

$$y^\star = \arg \min_{y \in \mathbb{R}^d} \left(\frac{s_{K_Q}(y)}{s_{K_C}(y)} \right) \tag{5.6}$$

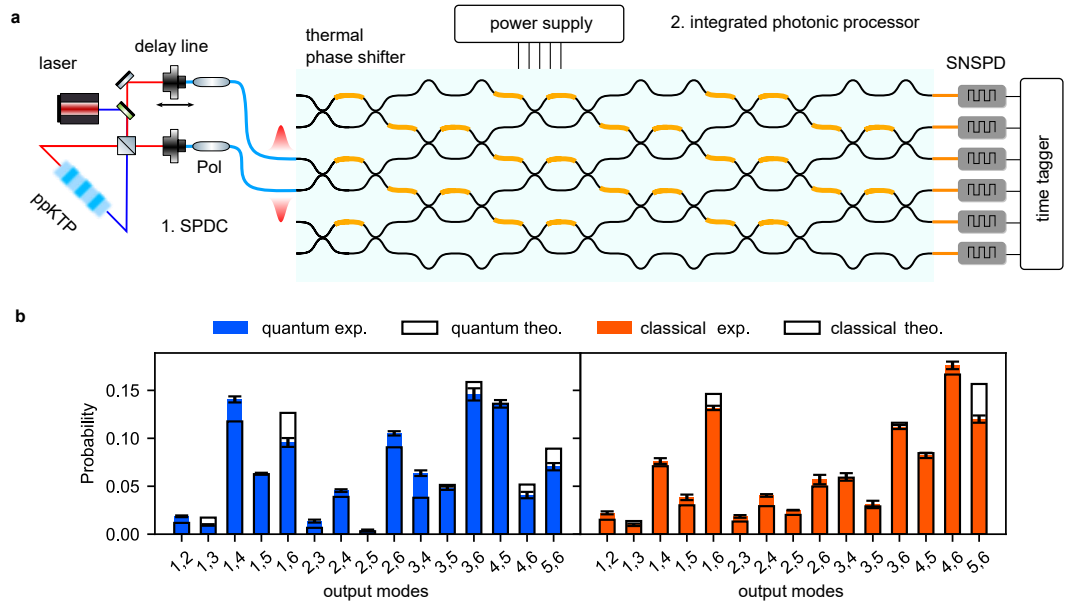

Figure 5.3: **Implementation of photonic quantum kernel estimation. a.** Experimental setup consisting of two parts, the off-chip single photon source and the programmable integrated photonic processor. The frequency degenerate photons are generated by a type-II spontaneous parametric down-conversion source. Afterwards, the two photons are made indistinguishable in their polarization and arrival time. Then, we inject these photons in two modes of an integrated photonic processor with six input/output modes [112]. Detection is performed by superconducting nanowire single-photon detectors (SNSPDs). The degree of indistinguishability can then be tuned through a delay line, changing their relative temporal delay. **b.** Probability distribution of photon detection events. We show two instances of the experimental photon detection probability, compared to the theoretical calculation. The quantum and classical kernel measurements are obtained respectively by injecting two indistinguishable and distinguishable photons into the third and fourth modes of the circuits, i.e. $|0,0,1,1,0,0\rangle$. The x axis shows all the circuit channels which output two photons simultaneously. Thus, all 15 possible photon detection configurations are accessible. The error bars shown in the plot were evaluated through a Monte Carlo simulation, considering an underlying Poissonian distribution of the counts. The measurements were performed with an integration time of 5 seconds, with a detection rate of 10 kHz.

Now, let us consider the following asymmetric distance g, denoted as *geometric difference*, between our two kernels, K_Q and K_C:

$$g_{CQ} = \sqrt{\left\| \sqrt{K_Q} \left(K_C\right)^{-1} \sqrt{K_Q} \right\|_\infty} \tag{5.7}$$

where $\|\cdot\|_\infty$ denotes the spectral norm. It can be shown that the following inequality holds [145]:

$$s_{K_C}(y) \quad \leq \quad g_{CQ}^2 \, s_{K_Q}(y) \tag{5.8}$$

Hence, if we maximize the geometric difference, we will have at least one task, which saturates inequality (5.8), for which the model complexity will be higher for the classical kernel, with respect to the quantum one. This indicates that the performance of K_Q will be competitive or better than K_C.

We can now use Eq. (5.7) to generate the classification task that, given two kernels K_Q, K_C and a set of data points $\{x_i\}$, produces the labels $\{y_i\}$ that maximise the difference in prediction error bound. This can be done through the following procedure: (i) evaluate the Gram matrices K_Q and K_C over a set of non-labelled data points $\{x_i\}$; (ii) compute the positive definite matrix $M = \sqrt{K_Q} \left(K_C\right)^{-1} \sqrt{K_Q}$; (iii) compute the eigenvalues and eigenvectors of M by spectral decomposition; (iv) find the maximum eigenvalue g and its corresponding eigenvector v; (v) assign the labels $y = \sqrt{K_Q} v$. From a practical point of view, we start with the two aforementioned kernels, K_C and K_Q, and then, by maximizing the geometric difference, we find the tasks for which the latter brings an enhanced accuracy of the classification. For more details regarding the algorithm to define the classification task, see the Supplementary Note 4. Let us note that the implemented tasks constitute instances of problems that can be naturally implemented with high accuracy on our quantum platform. As such they constitute a first stepping stone towards the identification of practical tasks for which quantum machine learning can enhance the performance of classical models.

5.3 Experiment

Our experimental setup consists of two parts, a single-photon source generating the input states and a programmable integrated photonic processor depicted in Fig. 5.3a. First, to generate the input state, we use a type II spontaneous parametric down-conversion source, which generates frequency degenerate single-photon pairs at 1546 nm in a periodically poled K-titanyl phosphate crystal. The two photons are then made indistinguishable in their polarization and arrival time, respectively, via wave retarders and a delay line, which we also use to tune the degree of indistinguishability of the generated photons.

For the implementation of photonic kernels, which map our input data to a feature space, we require an apparatus able to perform arbitrary unitary

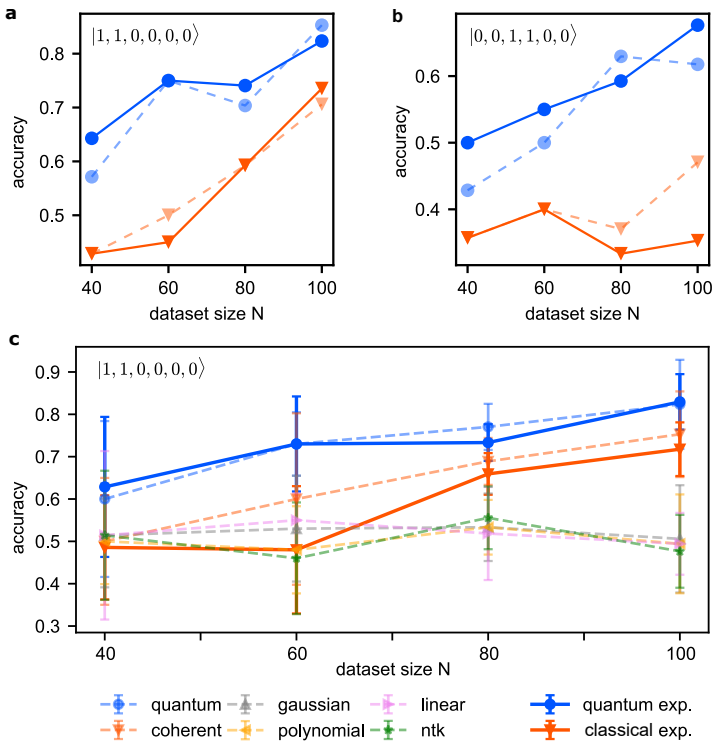

Figure 5.4: **Experimental classification accuracies. a-b.** We tested our method on datasets of different sizes (40, 60, 80, 100) and for two different input states ($|1,1,0,0,0,0\rangle$ and $|0,0,1,1,0,0\rangle$) respectively. For each dataset, $2/3$ of the data points were used for training the support vector machine (SVM) and $1/3$ for test. **c.** The average classification accuracies on 5 different sets for the quantum kernel (blue) and the classical (orange) kernel, along with the following other computational kernels: gaussian (grey), ntk (green), polynomial (yellow) and linear (purple). The dashed line indicates the results of numerical simulations, while the solid ones are the experimental results. The error bar shows the standard deviation of the classification accuracies on 5 random generated datasets classified by all the kernels.

transformations on a given input state. As mentioned before, our feature map sends each data point x_i onto the state resulting from the evolution $U(x_i)$ of a fixed input Fock state $|\psi\rangle$. Then, for the application of the SVM, which finds the best hyperplane separating the data, we need to evaluate the inner products between all of the points x_i, x_j in the feature space, which amounts to $\langle\psi|U(x_i)^\dagger U(x_j)|\psi\rangle$.

This implies that, if we take $|\psi\rangle$ as a Fock state of n photons over m modes, the inner product $\langle\Phi(x_i)|\Phi(x_j)\rangle$ is given by projecting the evolved state $U(x_i)^\dagger U(x_j)\,|\psi\rangle$ onto $|\psi\rangle$.

To this aim, we employ an integrated photonic processor [112] on a borosilicate glass substrate, in which optical waveguides are inscribed through femtosecond laser writing [109]. The circuit features six input/output modes [56], as depicted in Fig. 5.3a, where each interferometer is equipped with two thermal phase shifters [206], to provide tunable reflectivity and phase. Such arrangement allows us to perform any unitary transformation on the input photon states. Given this property, our device is also referred to as a universal photonic processor. Design, fabrication and calibration of the integrated photonic circuit are described in [112].

Specifically, the data were encoded in the values of the phase shifts, as follows:

$$x_i = (x_i^1, x_i^2, ..., x_i^{30}) \rightarrow \theta_i = (2\pi x_i^1, 2\pi x_i^2, ..., 2\pi x_i^{30}), \tag{5.9}$$

where θ_i are the phase shifts introduced by the phase shifters of a universal interferometer. Let us note that this encoding has the remarkable advantage that no extra processing is required on the input data, as they are directly plugged into the optical circuit parameters. Furthermore, in principle, we would need a sequence of two of such circuits (as in the scheme of Fig. 5.1c), to first implement $U^\dagger(x_i)$ and then $U(x_j)$ on our inputs. However, in our implementation, we adopt only one universal circuit and directly implement the unitary corresponding to the product $U(x_i)^\dagger U(x_j)$. This reduces the experimental complexity and the circuit propagation losses.

At the output, detection is performed by superconducting nanowire single-photon detectors (SNSPDs), where we post-select the output events to those featuring two detector clicks (coincidence counts, CC), i.e. *collision-free* events (see Supplementary Note 3). Thus, the elements of the Gram matrix of a given kernel can be estimated from the coincidence counting $K(x_i, x_j) = \mathrm{CC}_\psi^{ij} / \sum_{1\leq l<m\leq 6} \mathrm{CC}_{lm}^{ij}$. Here CC_{lm}^{ij} is the number of registered coincidence counts between channels l and m, when the implemented unitary is $U^\dagger(x_j)U(x_i)$ and ψ indicates the occupied modes of input state $|\psi\rangle$. To test the role of quantum interference in the accuracy of the classification, we tune the indistinguishability of the two photons by changing their relative temporal delay. An instance of the probability distribution of the same unitary is shown in Fig. 5.3b. The optimal classification task is chosen for each data set according to the algorithm explained in the previous section.

We test the performance of two photonic kernels in several different configurations. Firstly, we consider two different inputs, $|1,1,0,0,0,0\rangle$ and $|0,0,1,1,0,0\rangle$. This amounts to either injecting the photons into the first two modes or the central two modes. Secondly, we are able to control the indistinguishability of the bosons which allows us to implement, investigate a compare a quantum kernel with its related classical kernel. This is achieved by tuning the relative temporal

delay of the propagating photons. Consequently, at the detection, it is possible to (partially) distinguish which photon, out of the two input ones, is detected, by checking their arrival time. Perfect distinguishability is obtained when the temporal delay is longer than the coherence length of the photons. During the whole measurement, the maximal achieved indistinguishability between the photons is 0.9720 ± 0.0044, estimated through on-chip Hong-Ou-Mandel interference [207].

For both input states, we fix the encoding of each data point and vary datasets with four different sizes: 40, 60, 80 and 100. We use the setup depicted in Fig.5.3a to implement all pairwise products between the unitaries $U(x_i)^\dagger U(x_j)$. Hence, $|\langle\psi|U(x_i)^\dagger U(x_j)|\psi\rangle|^2$ is given by the probability of detecting the photons on the same modes from which they were injected. The rate of total post-selected coincidence counts amounts to 10 kHz and the measured probability distribution was averaged over 5 s for each unitary configuration. Let us point out that restricting to collision-free events is unlikely to bring to a decay in the rate of detected samples, when considering larger systems, since the bunching events become in general more rare for higher-dimensional Hilbert spaces, in absence of particular symmetry patterns in the matrix describing the unitary evolution of the input [24, 205].

For each size N, we perform $N(N-1)/2$ unitaries to compute the inner products. The distance between the unitaries experimentally realized and the target ones can be estimated as $\sum_i \sqrt{P_i^{\text{theo}} \cdot P_i^{\text{exp}}}$, where P_i^{exp} is the experimental detection frequency for the i-th output configuration, while P_i^{theo} is the one estimated based on the theory [205]. The mean fidelity of all datasets is 0.9816±0.0148 and 0.9934±0.0048, for the quantum kernel and classical kernel respectively. For each dataset, we use $2/3$ of the data points for training of the SVM, which yields the coefficients mentioned in 5.3. The remaining $1/3$ as the test dataset can be used to predict the classification accuracy. Accuracy is defined as the percentage of correctly classified points out of the total size of the test set. Let us note that values lower than 0.5 indicate that the model was not able to learn the features of the training set and generalize to unknown data.

In Fig. 5.4a and b, we show the test accuracies obtained by injecting two input states for four different dataset sizes, where the quantum kernel performs significantly better than the classical kernel at both experiments. In Fig. 5.4c, we report the average test accuracy obtained for five different datasets with the same size, varying the dataset sizes from 40 to 100 as well. Moreover, the results obtained with the quantum kernel (blue) and the classical kernel (orange) are compared with the following numerical kernels: neural tangent kernel (green) [198], gaussian (grey), polynomial (yellow) and linear (purple) [168, 192]. Here, the neural tangent kernel (ntk) adopts a infinite-width neural network to classify the data optimized by gradient descent. See the Supplementary Note 5 for more details. The dashed lines indicate the results of numerical simulations, while the solid lines indicate experimental results. Although the task is built only comparing the performance of the kernels based on indistinguishable and

distinguishable photons, the both obtained accurracies are higher than other classical kernels.

5.4 Discussion

In this work, we show the first experimental demonstration of quantum kernel estimation, based on the unitary evolution of Fock states through an integrated photonic processor. We map data into a feature space through the evolution of a fixed two-photon input state over six modes. To achieve this, we adopt an quantum processor realized by femtosecond laser writing in a borosilicate glass substrate [112]. The sampled output distribution is then fed into an SVM, performing the classification. Note that, although our apparatus only features tunable phase shifters and beam splitters, such encoding produces a sufficient non-linearity to achieve high classification accuracy of non-linearly separable datasets. This constitutes a difference of our method from alternative platforms, where entangling gates are typically needed [145, 208]. Furthermore, in our case, it is not necessary to increase the dimension of the feature Hilbert space to achieve a good accuracy. This circumvents the typical difficulty of quantum kernels whereby all data points are encoded in orthogonal states, leading to ineffective classification [145, 158, 196, 197]. Moreover, the fact that our model is effective for small dimensions is a crucial feature, since we require an approximation of the full probability distribution deriving from the evolution of our input state. Hence this study is relevant for medium-sized problems, because reaching high dimensions would imply the input/output combinations to grow exponentially, along with the number of required experimental shots to reach arbitrary accuracy.

The task we implement is designed by assigning binary labels to randomly generated data points, which we encode in the phase shifts of an optical circuit. This is done exploiting the so-called *geometric difference*, that selects the task for which the presence of quantum interference yields a better classification accuracy compared to cases where no interference is displayed. Interestingly, although the geometric difference compares the performance of a pair of kernels (in our case kernels implemented with indistinguishable/distinguishable bosons), both photonic kernels performed significantly better, for the selected tasks, than commonly used kernels, not only the gaussian, polyonomial and linear ones, but also the neural tangent kernel. This improvement in the performance, arising for both photonic kernels, is likely due to the fact that both the statistics have a similar dependence on the unitary evolution matrix, based on the permanent of sub-matrices (see Eqs. (5.4) and (5.5)), and are both complex enough to successfully tackle the selected tasks. In summary, our results indicate that a photonic kernel estimation can enhance the performance, even for medium-size problems, reachable by current quantum technologies. Moreover, using distinguishable bosons to have a (smaller) performance enhancement

represents an intriguing possibility, as it can prove convenient to reduce the impact of photon losses on the experimental time required to collect significant statistics.

Let us highlight that, although the interference of two photons in a six-mode unitary matrix (or more in general for medium-sized problems) can be estimated by classical computers, this does not affect the features of our protocol. First, because the interest in the investigation of the potentialities and limitations of a new computational method, like our proposed kernel, goes beyond the platform (quantum or classical) where it can be implemented, especially considering its particular suitability for given tasks. This is also the reason for which we do not use our photonic platform to reproduce classical kernels, as in [200, 209], but we focus on investigating those given by the natural evolution of bosons through a quantum circuit. Second, the approximation of permanents through classical algorithms, e.g. Gurvits one [84] has a slightly worse performance when compared to the direct sampling from an optical circuit, which naturally implements the studied kernel, for the case of inputs featuring product states [210]. In particular, the first scales as $O(n^2/e^2)$, whereas n is the number of photons and e is the required precision, while the direct sampling as $O(1/e^2)$ [85]. It is also noteworthy that the possibility of obtaining a quantum advantage on applications based on estimating expected values of unitary evolutions on linear optical circuits is still an open question [85, 210]. In our case, an advantage could emerge by considering the extension of our quantum kernel $K_Q(x_i, x_j) = |\langle\psi| U(x_i, x_j) |\psi\rangle|^2$ for entangled input states ψ, although there is no rigorous proof of it yet. Third, this protocol sheds light on alternative computational models, exploiting optical computation. This may be of particular importance when considering difficulties related to energy consumption, as it has been proved that partially optical networks can reduce energy requirements with respect to electronic ones [211].

Despite being overshadowed by deep neural networks, kernels are still widely used in a large number of tasks, due to their simplicity, and ability to learn from small datasets [212, 213]. Indeed, they have seen a recent revival in classical machine learning with the introduction of neural tangent kernels [198] and their use in the study of state-of-the-art neural network architectures such as transformers [214]. Another recent trend consists in merging neural networks and kernels, where notable examples are attention modules in natural language processing, and Hopfield layers [215]. Also from the quantum computing point of view, kernels have recently been getting a flurry of interest, both from the theoretical, as well as from experimental side [196, 209, 216, 217]. For the latter, it has been proven that the evolution of (photonic) quantum states through a circuit is complex enough to tackle image recognition tasks, both of simple toy models [209, 217], as well as satellite images [216]. Our method can therefore find a wide range of promising near-term applications in tasks such as information retrieval, natural language processing and medical image

classification [218–221], where kernels have been proposed as a keystone [222].

In particular, an interesting approach consists in pre-training a small (1-qubit) system to find the optimal encoding of classical data, which is then classified through a kernel method featuring a larger system, as proposed in [217]. This could be particularly beneficial on our photonic platform, since the pre-training could be performed through a classical light input over two spatial modes. This would imply outsourcing the computationally hardest part of the protocol to a fully optical platform, possibly bringing a significant reduction in energy consumption. Our experimental results also open the door to hybrid methods where photonic processors are used to enhance the performance of standard machine learning methods.

They also bring forward investigations on the non-linearities that can be achieved through photonic platforms [58, 223], in particular, for neuromorphic computation models, such as *reservoir computing* [224, 225]. In addition, we envisage the combination of this kind of non-linearity with those brought by the implementation of feedback loops, as in the case of quantum memristor [58] and the exploitation of quantum interference in the implementation of feature maps.

5.5 Methods

The two photon input states are generated by a type-II spontaneous parametric down-conversion source, which generates frequency degenerate single-photon pairs at 1546 nm via a 3-cm long periodically poled K-titanyl phosphate (pp-KTP) crystal. Afterwards, the two photons are made indistinguishable in their polarization, which is rotated through paddles, and in arrival time, through a delay line, which we use also to tune the degree of indistinguishability of the generated photons.

Then, we inject these photons into two selected modes of an integrated photonic processor with six input/output modes. This circuit features 27 thermal phase shifters and its architecture is universal, allowing to implement arbitrary unitary evolution on any input Fock state. Compared with the original architecture, the first three external phase shifters are omitted, as they will not effect the probability distribution. To apply accurate phases independently, each channel is supplied by a current source to avoid electrical crosstalk (4×Qontrol q8iv, 16bit DAC). In the end, detection is performed by superconducting nanowire single-photon detectors (SNSPDs) housed in a 1K cryostat. We post-select the detected events to the cases in which two detectors click simultaneously in a temporal window of 1 ns. A time tagger with a 15.63 ps resolution is used to process the real-time coincidence counting for all 15 post-selection patterns. The total coincidence counting is about 10 kHz, which varies due to the pump laser and the detection efficiency. For each unitary configuration, we integrate 5 s to estimate the probability distribution over 15 coincidence patterns.

5.6 Supplementary information

Support vector machines and kernel methods

For a dataset consisting of N data points with a ± 1 label, $D = \mathbb{S}(x_i, y_i)_{i=1}^N$, each data point is a d-length vector, $x_i = (x_i^1, x_i^2 ..., x_i^d)^T$, where $x_i \in \mathcal{X} \subseteq \mathbb{R}^d$ and x_i^j is the j-th feature of the input vector x_i. $y_i \in \mathcal{Y} = \mathbb{S}+1, -1$ is the binary label corresponding to the x_i. The classification task is to find a hyperplane $w \in \mathbb{R}^d$ and a bias parameter $b \in \mathbb{R}$, thus a linear *support vector machine* (SVM) is defined to predict the label of unknown data points x.

$$f(x) = \text{sign}(w \cdot x + b) \tag{5.10}$$

This is an optimization problem,

$$\begin{aligned} \min \quad & \frac{1}{2}||w||^2 \\ \text{s.t.} \quad & y_i(w \cdot x_i + b) - 1 \geq 0, \quad i = 1, 2, ..., N \end{aligned} \tag{5.11}$$

and it equals to solve the dual problem,

$$\begin{aligned} \min_{\alpha} \quad & \frac{1}{2} \sum_{i=1}^{N} \sum_{i=1}^{N} \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) - \sum_{i=1}^{N} \alpha_i \\ \text{s.t.} \quad & \sum_{i=1}^{N} \alpha_i y_i = 0 \\ & \alpha_i \geq 0, \quad i = 1, 2, ..., N \end{aligned} \tag{5.12}$$

The solution $\{\alpha_i\}$ are obtained by solving a linear optimization problem, given by the inner products of all input vectors $x_i \cdot x_j$ and the labels y_i.

Instead, rather than the direct linear inner products, the kernel function is used to transform the input vector from the original input space to another inner product *feature space*, $\Phi : \mathcal{X} \to \mathcal{H}$. The *Gram matrix* is a symmetric matrix where each element represents the pair-wise of inner product in a feature space $K_{ij} = \Phi(x_i) \cdot \Phi(x_j)$. The prediction function based on a kernel K is given as

$$f_K(x) = \sum_{i}^{N} \alpha_i K(x, x_i) \tag{5.13}$$

For some implicit kernel functions, it is easier to calculate the Gram matrices than the kernel functions. For example, if we define the feature space as *Fock space* $\mathcal{F}$, each original data point is mapped to a quantum state $x_i \mapsto |\Phi(x_i)\rangle = U(x_i)|\psi\rangle$ by a parameterized circuit $U(x_i)$ and Fock state $|\psi\rangle$.

Instead of performing quantum state tomography over every inner product, which is an expensive task considering the size of feature spaces, [226], it is

straightforward to measure the outcome probability Gram matrix

$$K(x_i, x_j) = |\langle\Phi(x_i)|\Phi(x_j)|\Phi(x_i)|\Phi(x_j)\rangle|^2 \\ = \left|\left\langle\psi|U(x_i)^\dagger U(x_j)|\psi\right|\psi|U(x_i)^\dagger U(x_j)|\psi\right\rangle\right|^2 \tag{5.14}$$

Thus, in our experiment, the task is to estimate each element of Gram matrices for each dataset which is equivalent to estimate the outcome probability of In Fig. 5.11, we show an example of experimentally reconstructed Gram matrices, as well as the ones calculated by the theory.

Geometric difference

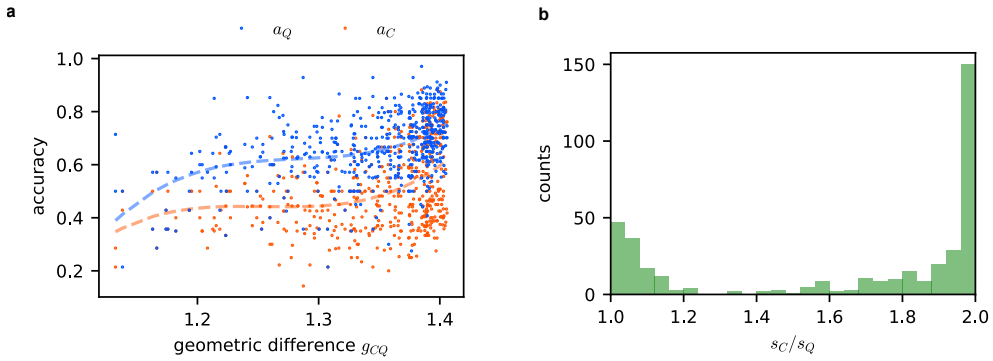

Figure 5.5: **Correlation between geometric difference, classification accuracy, and model complexity.** By varying the number of data points, the width of the unitary ansatz, and the photon number, we explore the correlation between the geometric difference and the classification accuracy and model complexity. **a**. The classification accuracy of the quantum kernel and classical kernel are plotted against the geometric difference. Each pair of points (red and blue) corresponding to a given geometric different refers to the same data set. The blue dots indicate the numerical accuracy achieved on test datasets of several dimensions exploiting the quantum kernel, varying the unitary ansatz width, encoding state and dataset sizes. The red ones, instead, refer to the classical kernel. The figure shows that, on average, the geometric difference is higher than 1, which implies that the quantum kernel has higher model complexity than the classical one. **b**. We plot the histogram of s_C/s_Q, which is related to the geometric difference as in Eq. (5.19). It can be seen that the majority of the tested datasets display a higher model complexity for the classical kernel, with respect to the quantum one.

For a given kernel method and a dataset $D = \mathbb{S}(x_i, y_i)_{i=1}^N$, the *model complexity*

of the triple (K, N, y) is given by:

$$s_K(y) = \sum_i \sum_j (K^{-1})_{ij} y_i y_j = y^T K^{-1} y \tag{5.15}$$

As shown in [145], this quantity can be used to bound the prediction error of a kernel method K on the dataset:

$$\mathbb{E}|f_K(x) - f(x)| \quad \leq \quad c \cdot \sqrt{\frac{s_K(y)}{N}} \tag{5.16}$$

Here, $x \in \mathcal{X}$, c is a constant and N is the number of data points.

The concept *geometric difference* is used to separate the complexity of two kernels. In this work, we target two kernels K_Q and K_C, which are the transform from original space to the Hilbert space given by the evolution of indistinguishable photons and the one given by distinguishable photons, respectively. Therefore, the optimal labelling is attained by solving the optimization problem by minimizing the ratio of the model complexity s_{K_Q} to s_{K_C}:

$$y^\star = \arg\min_{y \in \mathbb{R}^N} \frac{s_{K_Q}(y)}{s_{K_C}(y)} \tag{5.17}$$

The solution is $y = \sqrt{K_Q}\mathbf{v}$, where $\mathbf{v}$ is the eigenvector of matrix $\sqrt{K_Q}(K^C)^{-1}\sqrt{K_Q}$ with the eigenvalue noted as g^2_{CQ} [145]. This solution holds the inequality:

$$s_{K_C}(y) \quad \leq \quad g^2_{CQ}\, s_{K_Q}(y) \tag{5.18}$$

This inequality saturates when g^2 equals the spectral norm, as follows

$$g_{CQ} = \sqrt{\left\| \sqrt{K_Q}(K_C)^{-1}\sqrt{K_Q} \right\|_\infty} \tag{5.19}$$

In practice, a regularization parameter λ is used in the model complexity s_K, thus the geometric difference is alternated as

$$g_{CQ} = \sqrt{\left\| \sqrt{K_Q}(K_C + \lambda I)^{-1}\sqrt{K_Q} \right\|_\infty} \tag{5.20}$$

In the following simulations, we use $\lambda = 0.02$ without loss of generality.

Based on this model, we performed several simulations of randomly generated datasets, varying the number of data points, dimensions of unitaries, and photon numbers. To explore the correlation between geometric difference and the accuracy enhancement brought by quantum kernels, we plot test accuracies with respect to the the geometric difference in Fig. 5.5a. In general, a higher geometrical difference implies a better performance of the quantum kernel, with respect to the classical one. This is also reflected by Fig. 5.5b, where we show that, out of the sets we tested, the majority display a high ratio between the model complexity of the classical kernel and quantum one.

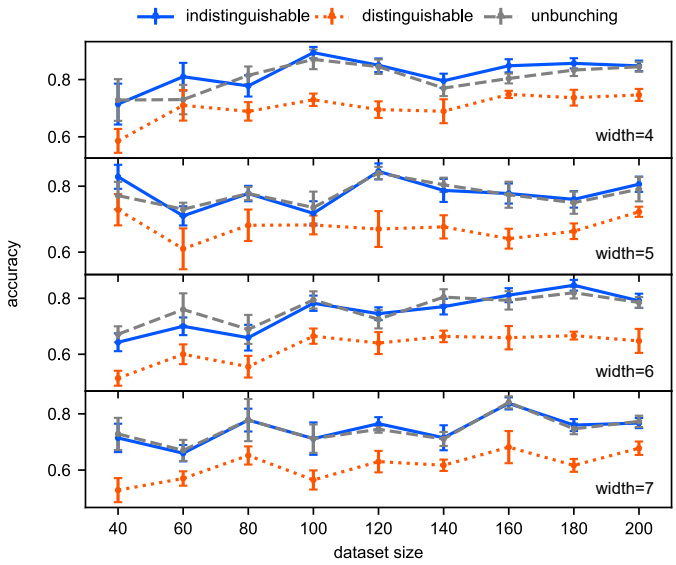

Figure 5.6: **Classification accuracies under approximation using unbunching photons.** Simulation of the classification accuracies using quantum kernels, both considering and not considering bunching events and classical kernels, varying different circuit widths. The error bars show the standard deviation of the test accuracies, obtained over 5 repetitions of the simulations with different random seeds.

Approximation of bosonic kernels with post selection

The inner product in bosonic kernels K_Q and K_C can be measured experimentally by collecting the probability over all output photon states $\mathbb{S}\phi_{m,n}$, which corresponds to the normalization of coincidence counting for all possible configuration. Here m is the mode number of the quantum circuit $U(\theta)$ with parameters θ and n is the photon number. For example, for an input state s and a output state t,

$$P_\theta(t|s) = \frac{\mathrm{CC}(t)}{\sum_{k\in\mathbb{S}\phi_{m,n}} \mathrm{CC}(k)} \tag{5.21}$$

including both unbunching (collision-free) photon states $\mathbb{U}$ and bunching photon states $\mathbb{B}$, $\mathbb{S}\phi_{m,n} = \mathbb{U} \oplus \mathbb{B}$, which requires photon number resolving detectors for each output mode. Since we are unable to detect those bunching events, a common solution to this issue is to only keep the statistics of unbunching events [27], when all photons output from the circuits are detected in separate modes.

We normalize the probabilities of unbunching states $\mathbb{U} \subsetneq \mathbb{S}\phi_{m,n}$:

$$P_\theta(t|s) \approx P_\theta^{\mathbb{U}}(t|s) = \frac{\mathrm{CC}(t)}{\sum_{k\in\mathbb{U}} \mathrm{CC}(k)} \tag{5.22}$$

[24, subSection 13] As shown in (7) the unbunching probabilities are a good estimate of the ideal bosonic statistics when $m >> n^2$. We can use the these statistics to define the *unbunching kernel*:

$$K(x_i, x_j) \approx K_U(x_i, x_j) = P_\theta^{\mathbb{U}}(\psi|\psi) = \frac{\mathrm{CC}_\psi}{\sum_{1\leq i<j\leq 6} \mathrm{CC}_{ij}} \tag{5.23}$$

for some state $\psi \in \mathbb{U}$. Under this approximation, the unbunching kernel is not a positive definite kernel since the estimated probability is not real inner products in a Hilbert space. In practice however, the function can still be used in a support vector machine and our simulations show that the resulting leaning algorithm performs similarly to the bosonic kernel, see Fig 5.6.

In this case, the labeling is still optimized for the quantum kernel K_Q with respect to the classical kernel K_C. We observe that the unbunching kernel performs similarly to the case where all of the output events are considered and consistently better than classical particles. This indicates this approximation is good enough to reproduce the quantum kernel by detecting only the collision-free photons.

Intermediate degrees of photon indistinguishability

To test the role of photon indistinguishability, we change the relative temporal delay between the input photons, making them distinguishable in their arrival times. This implies that, for delays that are lower than the coherence time of photons, we are sending the following mixture in the chip: $r\,|1,1,0,0,0,0\rangle + (1-r)\,|1,1',0,0,0,0\rangle$, with $0 < r < 1$. Here, r is defined as the *degree of indistinguishability* [227]. We report this transition, with the corresponding test accuracies for a data set of 40 elements in Fig. 5.7. These values are obtained for datasets of 40 elements, where $2/3$ are used as training and $1/3$ as test set, with input state $|0,0,1,1,0,0\rangle$.

Numerical experiments

Dataset generation algorithm

In the Algorithm 1, we report the algorithm used for all simulation mentioned in this work.

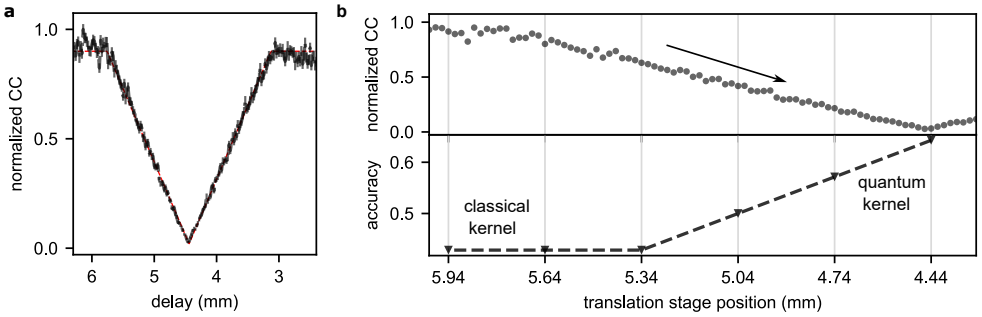

Figure 5.7: **Classification test accuracy with respect to degree of photon indistinguishability. a**. On-chip Hong-Ou-Mandel interference. The maximal visibility achieved amounts to 0.9720±0.0044. The noise in the plot comes from the translation stage and photon number counting in a single experiment. The red line is the obtained fitting triangular function. The measurements were taken with an integration time of 5 s, with a detection rate of 10 kHz. **b**. Transition from fully distinguishable to fully indistinguishable photon inputs.

Benchmark algorithm

To benchmark the photonic kernels, we choose the following four kernels to compare the classification accuracies, gaussian kernel, polynomial kernel, liner kernel and neural tangent kernel. First, for the gaussian kernel, the Gram matrices are calculated as

$$K_G(x_i, x_j) = \exp\left(\gamma |\, x_j - x_i \,|^2\right) \tag{5.24}$$

Here γ is a hyper parameter and x_i, x_j are two different data points. Second, the polynomial kernel is defined as

$$K_P(x_i, x_j) = (\gamma x_i \cdot x_j + r)^d \tag{5.25}$$

where there three hyper-parameters $\gamma r, d$. In both gaussian kernel and polynomial kernel, we run the grid search to find the optimal hyper-parameters. Next, for the linear kernel, we take the simple linear inner product as

$$K_L = x_i \cdot x_j \tag{5.26}$$

For the above three kernels, the Gram matrices are processed by a classical support vector machine to predict the data points in the test datasets.

Last but not least, to implement the neural tangent kernel, we adopt the open source proGram [228] which trains an ensemble of infinite width neural networks using gradient descent. The network consists two one input layer, two hidden layers and one output layer. Without loss of generality, the two hidden

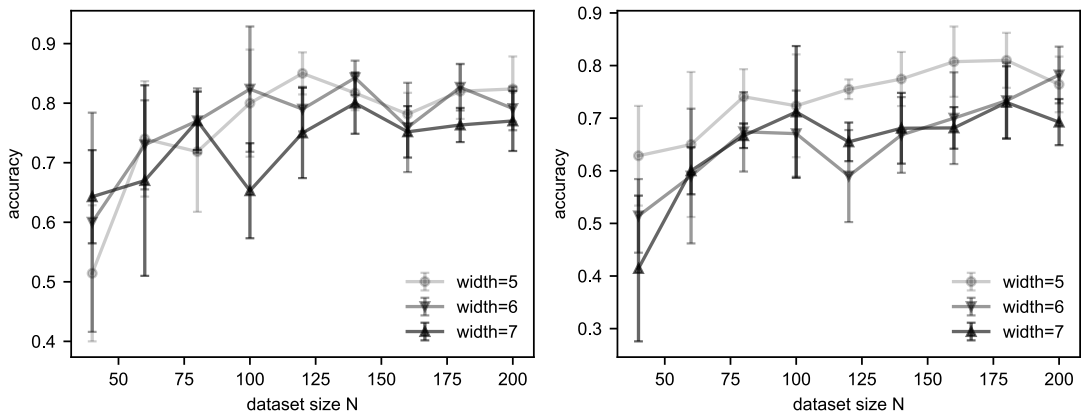

Figure 5.8: **Accuracy of the quantum kernel K_Q against the number of data points for several widths of the integrated circuit.** The two plots refer to two different input states, i.e. on the left, we consider the state $|1,1,0,0,0,0\rangle$ and, on the right, we consider $|0,0,1,1,0,0\rangle$. Each point is the mean accuracy on 5 iterations, each starting with randomly generated data. The standard error on the mean is shown by error bars. For each width, we observe a gradual increase in classification accuracy with number of data points. Typically, lower widths perform better for the same number of data points. We hypothesise that this is because the number of parameters available to the system increases with width, and so too many parameters on too little data leads to data being sparse in the space

layers both include 30 neurons, as the input data points are 30 dimensional. The sign function is applied on the output to identify the class. In fact, the hyper parameters in this infinite work does not affect the accuracy as the labeling is very specific to separate the photonic kernels.

Simulations

We conduct the following numerical simulations to substantiate the feasibility of our method. The classification accuracy by quantum, classical, gaussian, neural tangent, polynomial and linear kernels are labeled as $a_Q, a_C, a_G, a_N, a_P, a_L$ below.

Task1 As a first simulation, we fix photon number n, input state s and dataset size N, and get 3D plots of width m, depth k and accuracy $\mathbb{S}a_Q, a_C, a_G$. Results indicate that width and depth should be linearly related to maximise separation. In subsequent experiments we used $m = k$. We used two types of initial states,

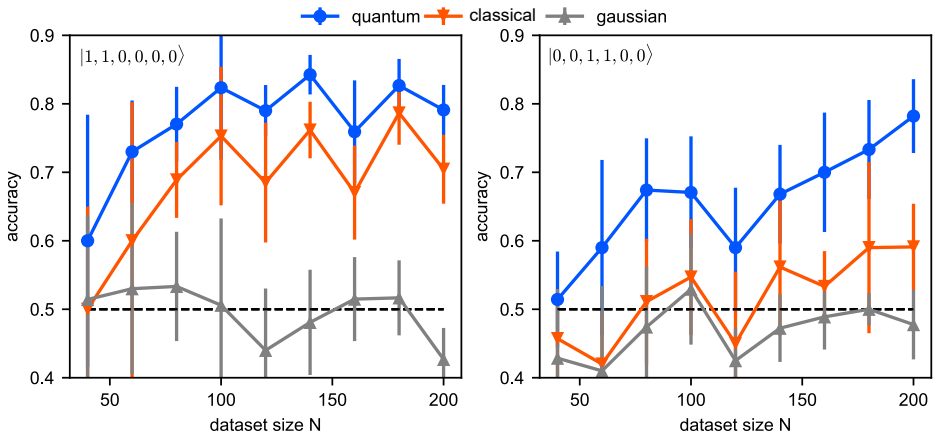

Figure 5.9: **Classification accuracies of photonic and Gaussian kernels.** We show the numerically estimated average test accuracies over 5 iterations for quantum kernel, classical kernel and Gaussian kernels as the number of data points grow. The black dash lines indicate the random guessing accuracy of 50%. The data for the quantum kernel comes directly from Figure 5.8 – the width that produced the highest accuracy is shown. The standard error on the mean is shown by error bars by five random dataset generation. We observe separation between the classification accuracy of the quantum and classical kernels. This is more prominent for the center initial state (right-hand figure). Both perform considerably better than the Gaussian kernel on the same data, which has an average accuracy of less than 60%.

Left state $|\psi_L\rangle = |1,1,0,0,0,0\rangle$ where photons are input of one side of the chip and Cent state $|\psi_C\rangle = |0,0,1,1,0,0\rangle$ where they are input in the middle.

Task 2 In our second simulation, we fix photon number n and the encoding state s, set $m = k$ and plot the quantum accuracy a_Q and classical accuracy against the number of data points N. The results are given in Fig. 5.8. We can observe this point experimentally by checking that $s_Q = g_{CQ}^2 s_C$. When this is violated the Gram matrix K_Q is not invertible, there are too many linear dependencies between the columns, i.e. too many data points in a space of dimension $\binom{n+m-1}{n}$.

Task 3 Now we have a way of scaling width m, depth k and number of data points N together, by taking the best chip size $m = k$ for a given number of data points. We plot the accuracies for quantum, classical and gaussian kernels as the system size grows. The results are given in Figure 5.9.

Algorithm 1 Dataset Generation

input circuit width m, depth k of the chip, number of input photons n, initial state $s \in \Phi_{m,n}$, number of data points N
generate N random data points $\{x_i\}_{i=1}^{N}, x_i \in \mathbb{R}^d$ // *Note d is proportional to mk*
$U(x_i) \leftarrow \prod_{j,k \in C} \mathrm{SU}_{MZI}(x_i^{(j)}, x_i^{(k)})$ // *C is the index in Clements coding*
// *SU_{MZI} is the matrix of Mach-Zehnder interferometer.*
while $0 < i \leq N$ **do** // *calculate the Gram matrix*
  while $0 < j \leq i$ **do**
    $K_Q(x_i, x_j) \leftarrow |\mathrm{per}(U(x_i, x_j))|^2$
    $K_C(x_i, x_j) \leftarrow \mathrm{per}(|U(x_i, x_j)|^2)$
    $j \leftarrow j + 1$
  end while
  $i \leftarrow i + 1$
end while
$S \leftarrow \sqrt{K_Q}(K_C + \lambda I)^{-1}\sqrt{K_Q}$
$\lambda \leftarrow \max$ eigenvalue of $S, \mathbf{v} \leftarrow \max$ eigenvector of S
$g_{CQ} \leftarrow \sqrt{\lambda}$
$y \leftarrow \mathrm{sign}(\mathbf{v})$ // *apply sign function to get binary labels*
output $D = \{(x_i, y_i)\}_{i=1}^{N}$.

Task 4 We also performed simulations to test unbunching kernels on the datasets resulting from separation between quantum and classical Gram matrices, see Figure 5.6. Separating directly between unbunching and distinguishable kernels is often not possible. This is because the separation method requires to compute the square root of the Gram matrix, but the one obtained from unbunching statistics is often not completely positive.

Experiment methods

As described in the main text, the experimental statistics of was post processed to calculate each element of the Gram matrices. Since the Gram matrices are symmetric, $K_{ij} = K_{ji}$, thus for N data points, we run $N(N-1)/2$ times to get the Gram matrices. In Figure 3**c**, for each dataset size varying from 40 to 100, we have 5 independent iterations, this indicates 53,300 running of unitary for both quantum and classical kernels. The total experiments lasted hundreds of hours.

The fidelities obtained from the implementation of unitaries are reported in Figure 5.10. The mean fidelity is 0.9816ś0.0148 and 0.9934ś0.0048, for the quantum kernel and classical kernel respectively. To illustrate the difference between two kernels, an example of estimated Gram matrices is shown in Figure 5.11.

To understand the impact of phase noise and imperfections in our experiments,

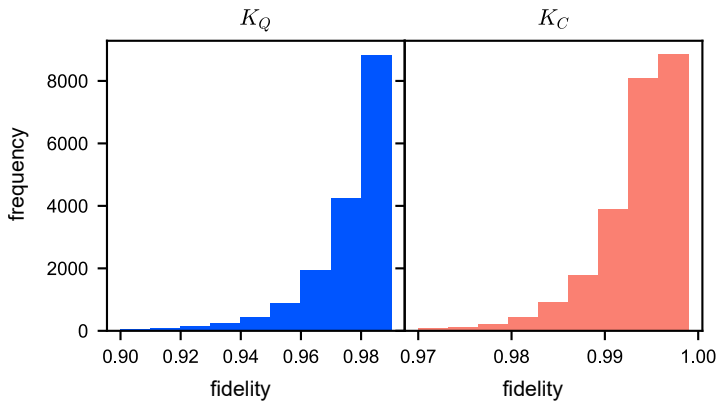

Figure 5.10: **Experimental probability fidelity of the unitaries.** The experimental fidelity for all data points performed by the quantum kernel (blue, left) and the classical kernel (orange, right). .

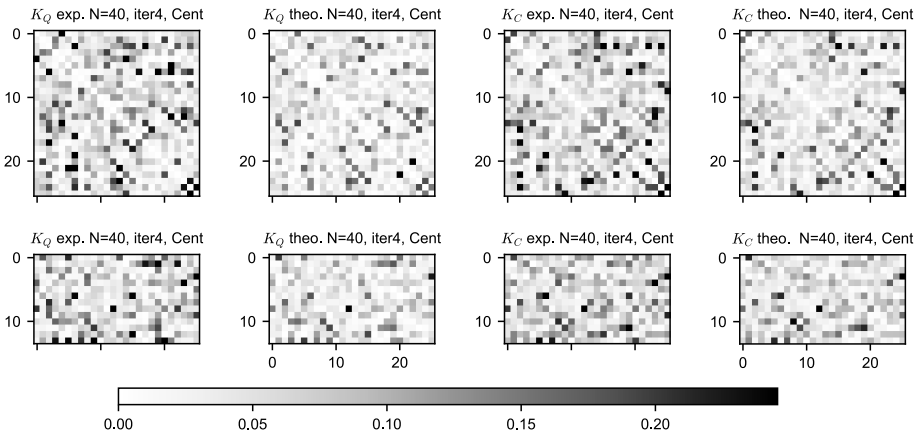

Figure 5.11: **Experimentally estimated Gram matrices.** We show the Gram matrices corresponding to the quantum kernel (left) and the ones referring to the classical kernel (right), compared to the theoretical predictions. The number of data points is 40, and the encoding state is $|0,0,1,1,0,0\rangle$. The top row shows the inner products of training Gram matrices and the bottom row shows the train-test Gram matrices.

we also performed the following simulations to verify the robustness of fidelity estimated used in the experiment. For all the simulations we run 500 random

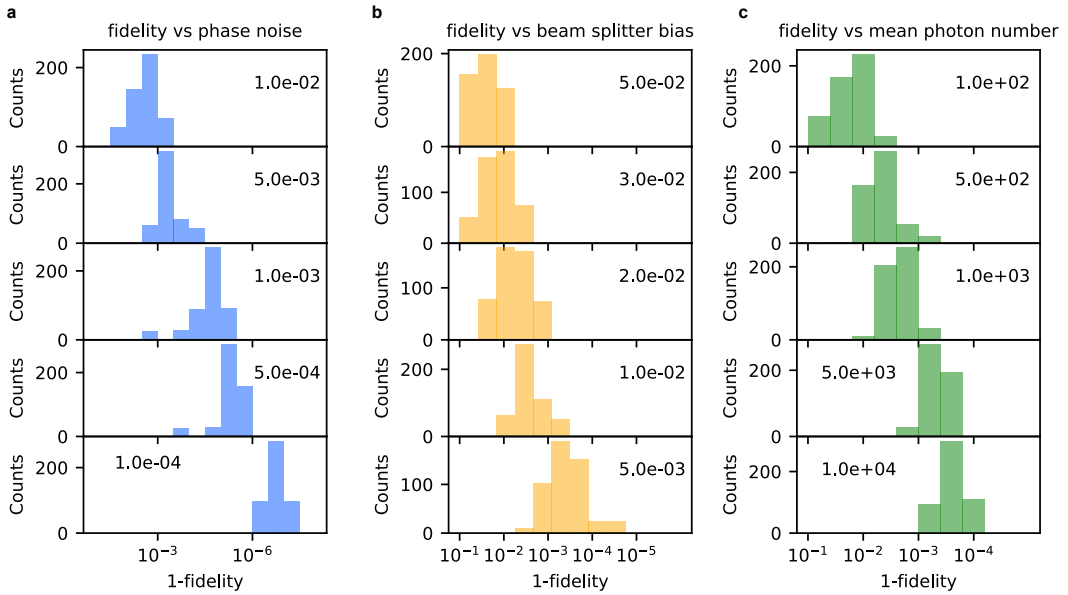

Figure 5.12: **Robustness of fidelity estimation**. The fidelity estimation of the implemented unitaries under three different noise types: phase noise in the MZI shown in **a**, beam splitter bias shown in **b**, and photon statistics shown in **c**. Each graph plots the histogram of fidelity between containing 500 random unitaries with noise (obtained sampling from a Gaussian distribution centered at the ideal value and with the standard deviation written in the plots) and without noise.

unitaries to estimate the fidelity, between the original unitary and the one perturbed by the noise, including the phase noise, bias of beam splitter and photon statistics respectively.

1. Phase noise in the MZI. In our setup, we use a 16 bit current output to drive every thermal phase shifter. The current output range is 24 mA and the corresponding current resolution is 0.003 mA. Considering that the average current to drive a 2π phase is around 12 mA, the phase resolution is 0.0015 rad. Moreover, let us note that, in our experiment, we do not use active temperature control and the temperature fluctuation is around 0.1 degree. To account for this kind of noise, we sample random phases from a normal distribution, around the correct value, with a standard deviation of 0.01 0.005 0.001 0.0005 and 0.0001 rad, respectively. The results are shown in Fig. 5.12a.

2. Bias of beam splitter. The ideal beam splitter, making up each MZI, should be 50:50, but the real devices used in our chip are directional couplers

with large curvature. To account for such imperfection, we consider our directional couplers as tunable Mach-Zehnders whose internal phase has an offset and amounts to $\pi/4 + \theta_{\text{bias}}$, rather than $\pi/4$.

$$U_{\text{BS with bias}} = \begin{pmatrix} \cos(\pi/4 + \theta_{\text{bias}}) & -\sin(\pi/4 + \theta_{\text{bias}}) \\ \sin(\pi/4 + \theta_{\text{bias}}) & \cos(\pi/4 + \theta_{\text{bias}}) \end{pmatrix} \tag{5.27}$$

Similarly to before, we select the bias, drawing them from a Gaussian distribution, centered at $\theta = 0$ and with a standard deviation of 0.05, 0.03, 0.02, 0.01, 0.005 rad. The results are shown in Fig. 5.12b.

3. Finite photon statistics. Considering that the underlying distribution for photon counting is Poissonian, every term in Eq. 5.21 has standard deviation $\Delta N_{\text{CC}} = \sqrt{N_{\text{CC}}}$. To estimate the impact of such noise, we performed a Monte Carlo simulation, considering different total coincidence countings, i.e. 100, 500, 1000, 5000, 10000, and created random matrices to sample the coincidence counting per channel. The obtained fidelity estimations are shown in Fig. 5.12c.

6 A scalable advantage in multi-photon quantum machine learning

Photons are promising candidates for quantum information technology due to their high robustness and long coherence time at room temperature [10, 64, 229]. Inspired by the prosperous development of photonic computing techniques, recent research has turned attention to performing quantum machine learning on photonic platforms [57, 230, 231]. Although photons possess a high-dimensional quantum feature space suitable for computation, a general understanding of how to harness it for learning tasks remains blank. Here, we establish both theoretically and experimentally a scalable advantage in quantum machine learning with multi-photon states. Firstly, we prove that the learning capacity of linear optical circuits scales polynomially with the photon number, enabling generalization from smaller training datasets and yielding lower test loss values. Moreover, we experimentally corroborate these findings through unitary learning and metric learning tasks, by performing online training on a fully programmable photonic integrated platform. Our work highlights the potential of photonic quantum machine learning and paves the way for achieving quantum enhancement in practical machine learning applications.

6.1 Introduction

The interplay of quantum computing and machine learning, known as quantum machine learning (QML), has emerged as an area of significant research interest in recent years [46, 137, 138]. With the anticipation of introducing substantial speedup or enhancement to machine learning problems and data science, certain QML protocols have demonstrated the potential to surpass their classical counterparts, namely reinforcement learning, quantum state learning, and quantum kernel methods [57, 194, 232]. Moreover, the noise resilience and ability of learning tasks make QML a suitable candidate for principle validation of quantum computing techniques on near-term noisy quantum hardware [46, 233].

Contribution statement: This chapter is a reprint of an unpublished manuscript under peer review. In this work, together with Dr. Yong Wang, I am one of the main contributors to the theoretical model, the data analysis, the experimental design, and the writing of the manuscript.

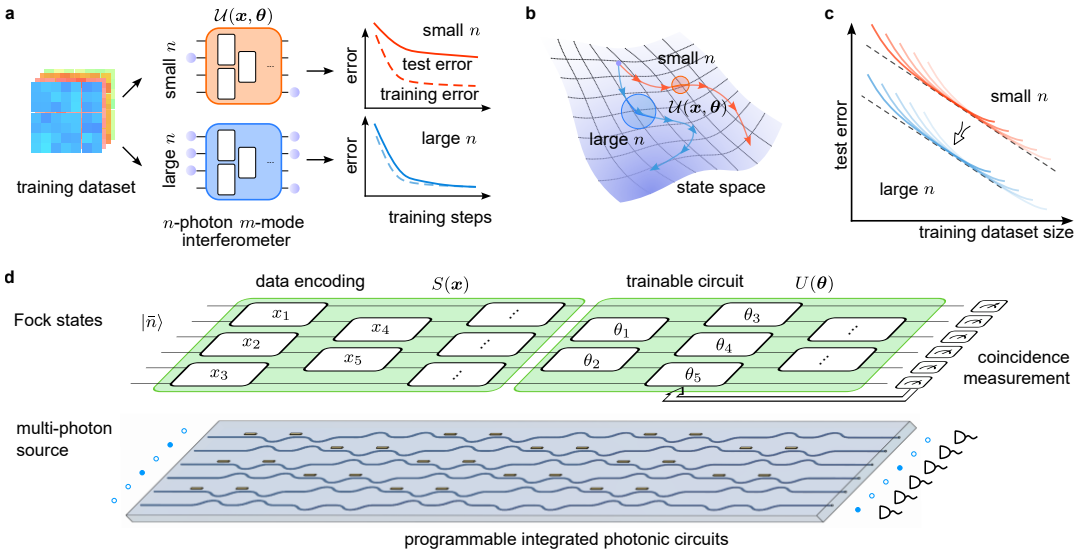

Figure 6.1: **Architecture of multi-photon quantum machine learning. a**. Our model consists of multi-photon states with n photons (shown as dots) propagating through a parameterized linear optical circuit $\mathcal{U}(\boldsymbol{x},\boldsymbol{\theta})$ with m modes, which encodes feature vectors $\boldsymbol{x}$ and trainable parameters $\boldsymbol{\theta}$. Photons are measured at the output modes, where the number of possible outcomes increases with n and leads to different learning results. **b**. Illustration of the accessible quantum state space during training with respect to $\boldsymbol{\theta}$. The size of the learnable state space increases with n, enhancing the expressivity and trainability of the QML model. **c**. The scaling trend of a photonic QML model describes the relationship between test error and the training dataset size. A higher photon number n corresponds to a larger learning capacity of the model, leading to lower test error and less training data required for learning. **d**. Our multi-photon QML setup using a programmable integrated photonic circuit. Identical photons are generated by multi-photon sources and injected into a photonic chip. The chip consists of tunable optical interferometers which imprint either feature vector data $\boldsymbol{x}$ or trainable parameters $\boldsymbol{\theta}$ as phase shifts onto the photon state. Each block in the circuit represents a 2×2 Mach-Zehnder interferometer containing two tunable phase shifters. The final states are measured via coincidence counting of photons in each mode, and used to update the parameters during the learning process.

As a fundamental quantum particle, photons are characterized by their robustness against decoherence, scalability potential, and room-temperature operability. Benefiting from these ideal properties, photonic platforms have been

extensively explored and utilized in quantum information processing, including quantum simulation [88, 91] and quantum computing [10, 64, 229, 230, 234]. These successes highlight the experimental feasibility of performing QML with photons, with proof-of-pricinple demonstrations including reinforcement learning [57], quantum-enhanced kernel methods [231], variational quantum algorithms [209, 235, 236], and photonic nonlinearity [237].

Moreover, photons are also theoretically anticipated as a powerful framework for computing tasks. The classical simulation complexity and potential computational advantage of photonic systems usually depend on the number of photons propagating within the quantum systems. As the Hilbert space of multi-photon states expands exponentially with the number of photons and modes [238, 239], specialized tasks such as boson sampling are known to become classically hard for sufficiently high photon numbers [239–241]. However, whether scaling up photonic systems can introduce advantages for practical machine learning problems or QML applications has not been explored from both theoretical and experimental perspectives.

In this study, we address this gap by demonstrating a scalable advantage in quantum machine learning with multi-photon states. We show that the learning capacity of general parametrized linear optical circuits increases polynomially with the photon number, corresponding to an enlarged set of trainable directions in the quantum state space. We theoretically and experimentally verify that this scalability directly translates into improved learning performance in terms of training dataset size and test error (Fig. 6.1a-c). Specifically, (1) in unitary learning, multi-photon states reduce the required training data from a linear to a constant scale for generalization; (2) in metric learning, higher test accuracy can be achieved with multi-photon states under the same variational ansatz and training conditions. The learning processes of these two tasks are realized by online training on a programmable integrated photonic circuit platform [242, 243] based on femtosecond laser waveguide writing [109], further underscoring the experimental viability and promise of multi-photon QML (Fig. 6.1d).

6.2 Multi-photon enhanced quantum machine learning

Our setup for supervised learning in multi-photon QML is shown in Fig. 6.2a. We have a set of L different training data $T_L = \{|\phi_\ell\rangle, y_\ell\}_{\ell=1}^{L}$, with each pure training state given by

$$|\phi_\ell\rangle = S(\boldsymbol{x}_\ell)\,|\bar{n}\rangle\ , \tag{6.1}$$

where $S(\boldsymbol{x}_\ell)$ is an m-mode unitary that encodes the feature vector $\boldsymbol{x}_\ell$, y_ℓ is the corresponding label, and $|\bar{n}\rangle$ is an n-photon input state [231]. Here, we restrict ourselves to Fock states with at most one photon per mode, i.e.,

$$|\bar{n}\rangle = |n_1, n_2, \cdots, n_m\rangle\,,\ n_i \in \{0, 1\}\,, \tag{6.2}$$

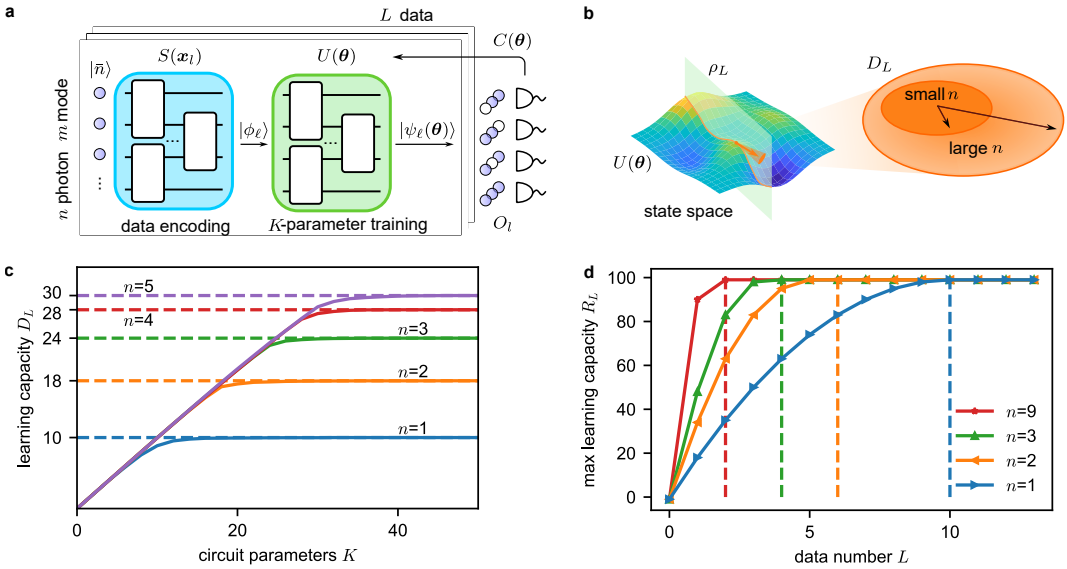

Figure 6.2: **Learning capacity of linear optical circuits. a**. A photonic QML model is trained on L training data, consisting of n-photon states $\{|\phi_\ell\rangle \equiv |\phi(\boldsymbol{x}_\ell)\rangle\}_{\ell=1}^{L}$ generated by a data-encoding unitary $S(\boldsymbol{x}_\ell)$ carrying feature vector $\boldsymbol{x}_\ell$. A parameterized unitary $U(\boldsymbol{\theta})$ with K trainable circuit parameters $\boldsymbol{\theta}$ is optimized with respect to the loss function $C(\boldsymbol{\theta})$. **b**. Sketch of the state space that can be accessed by $U(\boldsymbol{\theta})$ constrained to L training data. The size of this space is characterized by the learning capacity $D_L(n)$ (the rank of DQFIM $\mathcal{Q}_{ij}$), which increases with n and correlates with learning performance. **c**. Learning capacity D_L as a function of the number of trainable parameters K for different photon numbers n with $m = 6$ and $L = 1$. D_L increases with K, reaching the theoretical maximum value predicted by Eq. 6.6 as horizontal dashed lines. **d**. Maximal learning capacity $R_L(n)$ as a function of the training dataset size L for various n with $m = 10$. $R_L(n)$ increases with L and saturates at the critical dataset size L_c, indicated by vertical dashed lines.

where n_i denotes the number of photons in the ith mode and $n = \sum_i n_i$. The training data is then processed by a quantum neural network

$$|\psi_\ell(\boldsymbol{\theta})\rangle = U(\boldsymbol{\theta})\,|\phi_\ell\rangle\ , \tag{6.3}$$

which is given by a linear optical circuit $U(\boldsymbol{\theta})$ with K trainable circuit parameters $\boldsymbol{\theta} = \{\theta_1, \cdots, \theta_K\}$. The quantum neural network is optimized using the training data T_L with respect to a loss function $C(\boldsymbol{\theta})$, and then applied to test data that was not seen during training.

The power of machine learning models to learn and generalize can be understood from their complexity [140, 244–248]. Here, we use a quantum geometric

measure, known as the data quantum Fisher information matrix (DQFIM) [247, 249], which is based on the Fubini-Study metric [250]. Intuitively, the DQFIM quantifies the size of the accessible state space depending on the chosen dataset. This identifies the overparameterization transition where training finds good local minima [251], as well as when sufficient data is available for generalization [247]. While previous works were concentrated on qubit-based quantum neural networks, we now apply the DQFIM to multi-photon linear optics. It is given by the $K \times K$ matrix [247, 249]

$$\begin{aligned}\mathcal{Q}_{ij}(\boldsymbol{\theta}) = 4\,\mathrm{Re}[&\mathrm{tr}\Big(\partial_i U(\boldsymbol{\theta})\rho_L \partial_j U^\dagger(\boldsymbol{\theta})\Big) \\ &- \mathrm{tr}\Big(\partial_i U(\boldsymbol{\theta})\rho_L U^\dagger(\boldsymbol{\theta})\Big)\,\mathrm{tr}\Big(U(\boldsymbol{\theta})\rho_L \partial_j U^\dagger(\boldsymbol{\theta})\Big)]\,. \end{aligned} \tag{6.4}$$

Here, $\partial_i = \partial/\partial\theta_i$ and the mixture of training data $\rho_L = L^{-1}\sum_\ell \rho_\ell$. The rank of the DQFIM $D_L = \mathrm{rank}(\mathcal{Q}_{ij})$ characterizes the *learning capacity*, i.e., the number of independent directions in the parameter space $\boldsymbol{\theta}$ that can affect the model [247, 249, 251]. As illustrated in Fig. 6.2c, $D_L(K)$ increases with the number of circuit parameters K, until saturating at a critical number $K \geq K_\mathrm{c}$. At this point, the model becomes overparameterized and achieves the *maximal learning capacity* for a given number of training data L

$$R_L(n) = \max_K \mathrm{rank}(\mathcal{Q}_{ij})\,. \tag{6.5}$$

Such overparameterized models have a good optimization landscape with local minima close to the optimal solution and thus can be easily trained [251]. Note that R_L is also the upper bound on the rank of the classical Fisher information metric, which characterizes the complexity of neural networks [245].

For linear optical circuits, we can compute $R_L(n,m)$ analytically

$$\begin{aligned} R_L(n,m) &\leq 2mnL - n^2L^2 - 1 - (n-1)\delta_{L,1} && (nL \leq m)\,, \\ R_L(n,m) &= m^2 - 1 - (m-1)\delta_{L,1} && (nL > m)\,. \end{aligned} \tag{6.6}$$

where $\delta_{L,1} = 1$ for $L = 1$ and zero otherwise. A complete proof is provided in Supplementary Note 2.2. We highlight that $R_L(n)$ scales polynomially with photon number n. In particular, for $L = \mathrm{const}$, we have $R_L(n=1) \propto m$, while $R_L(n \propto m) \propto m^2$. Moreover, R_L is related to the minimal number of training states needed to generalize [247], as R_L grows with L and reaches its maximal value at a critical data number L_c. For $L \geq L_\mathrm{c}$, the dataset is overcomplete, i.e., one has sufficient data to fully characterize the model. In this regime, assuming a noise-free dataset, the model can generalize with low test error [247]. We find that L_c depends on photon number n as

$$L_\mathrm{c}(n,m) = \lceil (m-1)/n \rceil + 1\,. \tag{6.7}$$

We calculate R_L as a function of L for different values of n in Fig. 6.2d. A higher photon number yields larger learning capacity R_L, leading to a significantly lower

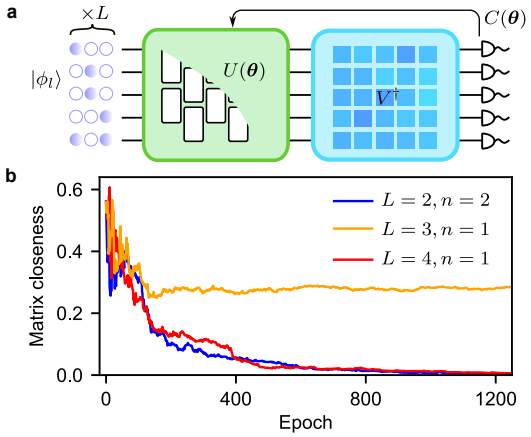

Figure 6.3: **Experimental unitary learning with multi-photon states. a**. We aim to learn an unknown unitary V by training a parameterized unitary $U(\boldsymbol{\theta})$ such that $U(\boldsymbol{\theta}) = V$. **b**. Experimental results of learning a 5×5 unitary V. We plot the matrix closeness $C_{\mathrm{M}}(U(\boldsymbol{\theta}), V)$ up to local phases and permutations. Perfect generalization corresponds to $C_{\mathrm{M}} = 0$. For $n = 2$ photons, we achieve nearly optimal closeness $C_{\mathrm{M}} \approx 0$ using only $L = 2$ training data (blue curve). In contrast, $n = 1$ photon requires at least $L \geq 4$ training data (red curve), and fails with $C_{\mathrm{M}} \approx 0.3$ for $L = 3$ (orange curve). With one additional training state, we can learn the full unitary V including local phases and mode permutations, as experimentally verified in Supplementary Note 4.

number of training states L_{c}. In fact, using $n = 1$ single-photon states (which also applies to coherent states produced by lasers, see Supplementary Note 2.2 for details), one requires $L_{\mathrm{c}} = m$ training data to generalize, while $L_{\mathrm{c}} = 2$ for $n = m - 1$. Therefore, multi-photon states allow for improved learning performance and generalization capability, such as reducing the amount of training data or achieving higher test accuracy.

6.3 Learning unitaries with multi-photon states

Now, we study the task of variationally learning a black-box unitary [252]. Conventional quantum tomography of quantum process unitaries usually involves complete bases of both input and output states [253, 254]. Meanwhile, with limited input states, one could use post-selection to learn the unitary adaptively, as demonstrated by variational unsampling and adaptive boson sampling [209, 236]. Here we show that by using multi-photon states, it is possible to learn the full unitary using only a constant number of training states, without the need

of complete bases and post-selection.

We assume that we have oracle access to an unknown $m \times m$ unitary V, without any prior knowledge of its internal structure. The goal is to learn a parameterized unitary $U(\theta)$ such that it represents V. To do so, we are given L arbitrary n-photon training data $|\phi_\ell\rangle$, along with their corresponding labels $V|\phi_\ell\rangle$, forming the training dataset $T_L = \{|\phi_\ell\rangle, V|\phi_\ell\rangle\}_{\ell=1}^{L}$. The parameterized ansatz $U(\boldsymbol{\theta})$ is applied to each training state, project onto the label state $V|\phi_\ell\rangle$, and minimize the loss function $C_{\text{train}}(\boldsymbol{\theta}) = 1 - L^{-1}\sum_{\ell=1}^{L}\left|\langle\phi_\ell|V^\dagger U(\boldsymbol{\theta})|\phi_\ell\rangle\right|^2$. To simplify the presentation, we focus on learning V up to local phase shifts and mode permutations; the full protocol that includes learning these components is described in Supplementary Note 3.

We report the experiment of learning a 5×5 unitary with different photon numbers n and training dataset sizes L in Fig. 6.3b. Our setup adapts a programmable integrated photonic circuit and an optical switch to implement the learning task for single-photon or two-photon input states. Pseudo photon-number resolution is used to detect bunching events in the two-photon case. The parameterized unitary $U(\boldsymbol{\theta})$ is trained from the same random initialization for all datasets by optimizing the parameters of the photonic chip using simultaneous perturbation stochastic approximation (SPSA) [255]. To determine whether the correct unitary has been learned, we compute the matrix closeness $C_{\text{M}}(U(\boldsymbol{\theta}), V)$, which quantifies the difference between $U(\boldsymbol{\theta})$ and V. In particular, C_{M} equals zero when the two unitaries match up to local phases and mode permutations (see Methods).

We consider three different training datasets: one set with size $L = 2$ of two-photon states ($n = 2$), and another two larger sets with $L = 3$ or $L = 4$ of single-photon states ($n = 1$). We find that to minimize matrix closeness and thereby achieve generalization, four single-photon training states ($L = 4$) are required, while three ($L = 3$) are insufficient. In contrast, two-photon states generalize successfully with only $L = 2$, requiring half as much training data as the single-photon case. This behaviour can be related to the learning capacity $R_L(n)$, which saturates at a critical number of training data that scales approximately as $L_{\text{c}} \propto 1/n$. Indeed, the experimentally observed values of L required for generalization have matched our theoretical prediction exactly, confirming that more photons reduce the data needed for generalization (Supplementary Note 3). We also experimentally demonstrate that the full unitary, including local phases and mode permutations, can be learned with one additional training state (Supplementary Note 4).

6.4 Quantum metric learning with multi-photon states

Metric learning [256] is a widely used supervised learning approach with important applications including face recognition [257] and vision language mod-

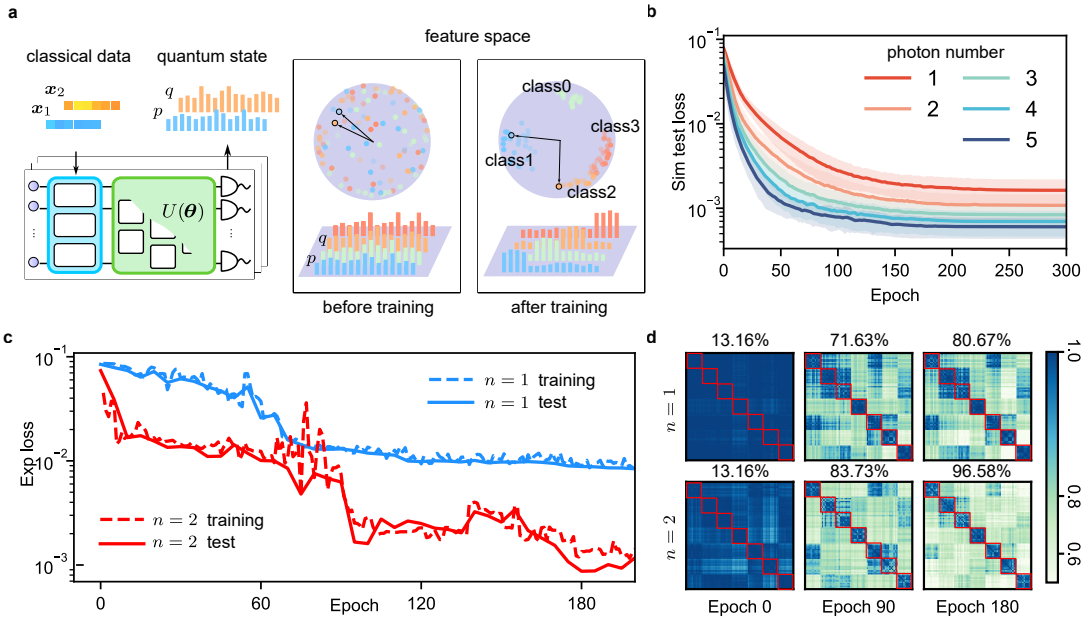

Figure 6.4: **Experimental quantum metric learning with multi-photon states.** **a**. Two input data points $\boldsymbol{x}_1, \boldsymbol{x}_2$ are encoded into the circuit and processed by a trainable parameterized unitary $U(\boldsymbol{\theta})$. The output probability distributions of photon configurations are measured in the Fock space as p and q, respectively. We use a cost function based on the cosine similarity $S_{\mathrm{C}}(p,q) = \sum_i \sqrt{p_i q_i}$ to guide the learning process, as detailed in Supplementary Note 5. After training, data points from the same class are close to each other on the hypersphere, and points from different classes are well separated. **b**. Simulation of test loss for different photon numbers $n = 1, 2, 3, 4, 5$. For each n, the model is trained until convergence with 100 random seeds, and the average loss on the test data is plotted with standard deviation (shaded region). **c**. Experimentally measured training and test loss for single-photon and two-photon cases during the training process. **d**. Measured Gram matrices on the test data at selected epochs (0, 90 and 180), where darker blue implies higher similarity S_{C}. The top and bottom rows show the results for single-photon and two-photon experiments, respectively. Groups of data from the same classes are highlighted in red boxes. Above each matrix we show the pairwise accuracy, which is defined as the probability of correctly identifying whether a pair of test data belongs to the same or different classes.

els [258]. The goal of metric learning is to learn a distance function or similarity measure in a high-dimensional embedding space, such that similar instances are mapped close together, while dissimilar ones are pushed farther apart.

The schematic diagram of quantum metric learning with multi-photon states is shown in Fig. 6.4a. Input data is encoded into quantum states and transformed by a quantum neural network. The resulting output state is measured in the Fock space, yielding a probability distribution $p_\ell(z) = |\langle z|\psi_\ell(\boldsymbol{\theta})\rangle|^2$, where $|z\rangle$ denotes a Fock basis state. Given two output probability distributions $\{p_i\}$ and $\{q_i\}$ corresponding to two data points, we use the angular distance between the unit vectors $\sqrt{p_i}$ and $\sqrt{q_i}$ on the hypersphere as the distance measure for metric learning, which is equivalent to computing the cosine similarity $S_{\mathrm{C}}(p,q) = \sum_i \sqrt{p_i q_i}$. The training objective is to minimize the angular distance (i.e., maximize the cosine similarity) between data points from the same class, and vice versa, and is independent of the number of photons or modes (Supplementary Note 5).

We apply our model to a vowel recognition dataset [259, 260] consisting of vowel-formant frequency vectors from seven classes (see Methods), which we encode into a variational photonic circuit with $m = 6$ photon modes. First, we conduct numerical simulations for photon numbers $n = 1, 2, \ldots, 5$. For each configuration, the model is trained with 100 random seeds until convergence, and the average loss values on the test data are plotted in Fig. 6.4b. We observe that test loss values decrease with larger n, which is consistent with the growth in learning capacity $R_L(n)$ as a function of n, as seen in Eq. 6.6 and Fig. 6.2c.

Next, we experimentally implement quantum metric learning using single-photon ($n = 1$) and two-photon input states ($n = 2$), where in both cases the chip parameters are optimized through online training with the SPSA method from the same initialization. As shown in Fig. 6.4c, the training and test losses are substantially lower for two-photon states compared to the single-photon case. The Gram matrices, which depict pairwise distances between test data, are shown in Fig. 6.4d for different training epochs. In the plots, test samples are sorted by class. For the two-photon case ($n = 2$), the matrices exhibit an approximate block-diagonal structure after training: data points from the same class are closer to each other and appear as darker blue blocks (highlighted in red boxes), whereas the off-diagonal blocks correspond to much larger distances between data from different classes and appear in lighter colours. It is worth noting that for $n = 2$, both the average accuracy and class separation improve markedly compared to $n = 1$ (Fig. 6.4d), which is in agreement with both the numerical simulation results (Fig. 6.4b) and the higher learning capacity of more photons. More experimental results and implementation details are provided in Supplementary Note 5.

6.5 Discussion

In this work, we prove that multi-photon states in linear optical circuits provide a fundamentally larger learning capacity, which characterizes the accessible quantum state space and serves as a hallmark for the training and generaliza-

tion capabilities of QML models [245, 247, 249, 251]. In particular, the unitary learning task pointed out that single-photon states (or coherent states) need $O(m)$ training data, while multi-photon states can accomplish the same tasks with only a constant number of data $O(1)$. For metric learning, an important machine learning task with broad practical relevance, multi-photon states generalize significantly better by achieving lower loss values, higher test accuracies, and clearer class separation. Our experiments, featuring multi-photon online training of two tasks, have shown the feasibility of QML applications on photonic hardware.

Contrasting with existing studies, our work provides a rigorous theoretical understanding for scaling photons in QML, focusing on its impact on training data requirements and generalization performance. Our experimental approach is based on readily available and routinely used photonic platforms, involving linear optical circuits, multi-photon states, and single-photon detectors. Notably, programmable integrated photonic circuits have been commonly applied in classical optical machine learning [260, 261]. Our proposal inspires potential performance gain by upgrading classical instruments such as lasers and photodiodes with multi-photon sources and photon detectors, without altering the internal logic of photonic circuits.

Furthermore, this work not only paves the way for a wide range of photonic QML applications, including variational quantum estimation and quantum simulation [91, 209, 262], but also illuminates quantum computing platforms beyond photons. Other bosonic systems such as cold atoms [263], microwave photons [264], and phonons [265], offer controllable numbers of particles that could be engineered for QML tasks.

Finally, we foresee several possible directions for future exploration building on our study. Firstly, quantum natural gradient methods [266] could be investigated to reduce training time by lowering the number of required epochs, as they are inherently compatible with our quantum geometric framework. Secondly, our protocols can be extended beyond linear optical circuits and photon-number states. For instance, incorporating optical nonlinearity [237, 238] or combining with continuous-variable states in the non-Gaussian domain [267] could expand the accessible state space and increase the upper bound of the learning capacity R_L [247, 251]. Thirdly, while our framework currently targets product states as input, it can be extended to non-separable states, such as those involving entanglement and graph states. In conclusion, the full potential of photonic QML remains largely untapped, and we believe that leveraging multi-photon states in large-scale programmable photonic systems represents a promising pathway for advancing machine learning.

6.6 Methods

Details of the experimental apparatus

We generate photon pairs at 1550 nm via a spontaneous parametric down-conversion source with a periodically poled potassium titanyl phosphate (ppKTP) crystal which is pumped at 775 nm. A tunable delay line is used to adjust the time delay on a single side, compensating for any temporal differences before the photons interfere on the chip. The integrated photonic circuits utilized in this study are fabricated on a laser-written glass waveguide platform [109, 111, 112, 206]. This circuit features 6 input and output ports and 15 Mach-Zehnder interferometers (MZIs), capable of performing any arbitrary SU(6) unitary operation. Each MZI is equipped with two thermo-optic resistive phase shifters for independent phase tuning from 0 to 2π. Electrical cross-talk between phase shifters is minimized by using a multi-channel current source to control the phase shifters. In total, 30 phase shifters are required; however, those acting on phase terms placed directly at the output ports (which do not influence quantum interference processes and are not measurable in our apparatus) are not operated in the present experiments. Superconducting nanowire single-photon detectors are employed to measure single-photon counting events, and a time tagger is used to record coincidence events with a resolution of 15.63 ps. In the unitary learning task where various learning states are required, we use a programmable optical switch to control the input ports of the photons.

Online training on quantum hardware

We adopt the simultaneous perturbation stochastic approximation (SPSA) method for online circuit parameter optimization on our photonic chip [182]. SPSA has desirable theoretical properties for noisy intermediate-scale quantum hardware implementations, and has proved successful in experimentally training superconducting quantum platforms [208, 268]. By estimating the gradients with only two objective function evaluations per iteration, SPSA avoids linearly increasing computational costs where parameters are varied one at a time. Moreover, the gradient approximation process of SPSA is intrinsically noisy, and thus robust to the intractable stochastic fluctuations on noisy quantum hardware.

Specifically, suppose that we want to find the optimal parameter $\boldsymbol{x}^*$ to minimize the objective function $f(\boldsymbol{x})$. For each iteration, we generate a random vector Δ, which is usually drawn from the Rademacher distribution, a discrete probability distribution with a fifty-fifty chance of being ± 1. Then the random vector is added and subtracted to the trainable parameters, and $f(\boldsymbol{x}+c\Delta)$ and $f(\boldsymbol{x}-c\Delta)$ are evaluated through experimental hardware measurements, where $c>0$ is the perturbation amplitude. The approximate gradient for $\boldsymbol{x}$ is thus given by

$$\boldsymbol{\delta}_{\text{approx}} = \frac{f(\boldsymbol{x}+c\Delta) - f(\boldsymbol{x}-c\Delta)}{2c\Delta}.$$

The parameters $\boldsymbol{x}$ can be subsequently updated to $\boldsymbol{x} - a\boldsymbol{\delta}_{\text{approx}}$, where $a > 0$ is the learning rate.

The SPSA hyperparameters including c and a, can be either estimated by tentative hardware runs or determined by digital simulations. We choose $c = 0.4, a = 3$ for the task of unitary learning, and $c = 0.4, a = 150$ for quantum metric learning. These values achieve good performance in digital simulations and show similar convergence behavior in physical experiments.

Learning unitaries with multi-photon states

To characterize that $U(\boldsymbol{\theta})$ correctly learned the unitary V up to local phases and permutations, we define the matrix closeness $C_{\text{M}}(U(\boldsymbol{\theta}), V)$

$$C_{\text{M}}(U, V) = \min_{F, P_\sigma}[1 - \frac{1}{m}\text{tr}(V^\dagger F P_\sigma U)] \tag{6.8}$$

where we minimize over the set of local phase unitaries $F(\boldsymbol{\phi}) = \sum_{j=1}^{m} e^{-i\phi_j} |j\rangle \langle j|$ with angles ϕ_j on the jth mode and mode permutations $\{P_\sigma\}_\sigma$. Here, P_σ are operators that permute modes according to permutation function $\sigma(j) = k$, with $j, k \in \{1, 2, \ldots, m\}$. For example, for $m = 3$, the permutation function with $\sigma(1) = 3$, $\sigma(2) = 1$ and $\sigma(3) = 2$ is a cyclic permutation of modes $(1, 2, 3) \to (3, 1, 2)$. The matrix closeness has $C_{\text{M}}(U, V) = 0$ only when $U = V$ up to local phases and mode permutations, indicating that we have learned V correctly.

Quantum metric learning with multi-photon states

The vowel dataset in this study has been previously used by multiple studies to demonstrate training novel machine learning hardware setups [260, 269]. The dataset consists of 12-dimensional data vectors of formant frequencies extracted from audio recordings, with 7 classes of 'ae', 'ah', 'aw', 'er', 'ih', 'iy' and 'uw'. Each class contains 37 samples, and the total size of the dataset is $37 \times 7 = 259$. We randomly split the dataset into training and test data with a ratio of $70/30$ and keep the same sample number for all classes.

In both digital simulations and hardware experiments, we choose a typical variational ansatz where the feature vectors are normalized and encoded onto the phase shifters of the photonic chip (Supplementary Fig. S3). For the digital simulations of one to five photons in Fig. 6.4b, we train using the Adam optimizer [270] with 100 different random seeds and plot the average simulated test loss curve. The learning rate is initially set to 0.1, and is decreased by a factor of 10 when the training loss has stopped decreasing for 10 epochs. For physical experiments with single-photon and two-photon states, we use SPSA optimization as introduced above. All relevant implementation details are provided in Supplementary Note 5.

6.7 Supplementary information

Preliminaries

We consider linear optical systems with Fock states, i.e., states with fixed photon numbers. Our terminology follows the conventions developed in boson sampling [24]. The Hilbert space of a system is spanned by Fock states of the form

$$|n_1, n_2, \cdots, n_m\rangle = \frac{1}{\sqrt{n_1! n_2! \ldots n_m!}} (\hat{a}_1^\dagger)^{n_1} (\hat{a}_2^\dagger)^{n_2} \cdots (\hat{a}_m^\dagger)^{n_m} |\text{vac}\rangle \ , \tag{6.9}$$

where $\hat{a}_i^\dagger$ is the creation operator of the ith mode, n_i is the photon number in the ith mode, m is the number of modes, and $|\text{vac}\rangle$ is the vacuum state. The system evolution is described by an $m \times m$ unitary matrix U parameterized by K tunable real parameters $\boldsymbol{\theta} = \{\theta_s\}_{s=1}^K$ as

$$U(\boldsymbol{\theta}) = \prod_{s=1}^{K} u_s(\theta_s). \tag{6.10}$$

In linear optics, the parameters are typically realized as phase shifts within Mach-Zehnder interferometers (MZIs). Note that the parameterization of U is not unique, and there exist many different equivalent configurations. The minimum number of real parameters to realize arbitrary unitary transformations is $m^2 - 1$, which can be efficiently implemented using MZI-based architectures [55, 56].

For an input Fock state $|\phi\rangle$ and unitary matrix U, the probability of measuring a specific output Fock state $|\psi\rangle$ is determined by the permanent of a matrix A, which is constructed from the input state $|\phi\rangle = |n_1, n_2, \cdots, n_m\rangle$, the measured output state $|\psi\rangle = |n_1', n_2', \cdots, n_m'\rangle$ and U[24]

$$\langle\psi| \tilde{U} |\phi\rangle = \frac{\text{perm}(A)}{\prod_i n_i! \prod_i n_i'!}. \tag{6.11}$$

Here, $\tilde{U}$ is a $d \times d$ unitary matrix determined by U, where $d = \binom{n+m-1}{n}$ is the dimension of the Fock space, corresponding to the number of photon occupation configurations for n photons in m modes. For pure input states, the evolution can also be described in the Schrödinger picture using the density matrix formalism $\tilde{\rho} = \tilde{U} \rho \tilde{U}^\dagger$, where ρ and $\tilde{\rho}$ denote the density matrices of the input and output states, respectively.

Quantum Fisher information matrix

Brief introduction to quantum Fisher information matrix

What is the capability of a trainable parameterized quantum circuit for quantum machine learning (QML) tasks? This question has been explored through the

concept of quantum effective dimension [249], which describes the ability of a given model to express quantum states. While previous studies have primarily focused on qubits, here we apply these measures to linear optical circuits with Fock states.

Consider a parameterized quantum state $U(\boldsymbol{\theta})\,|\phi\rangle \equiv |\psi(\boldsymbol{\theta})\rangle$, which is generated by applying a parameterized unitary transformation $U(\boldsymbol{\theta})$ to a fixed input state $|\phi\rangle$. The expressive power or capacity of such states can be characterized using the quantum Fisher information matrix (QFIM) [249]. For pure states the QFIM is defined as [250]

$$\mathcal{F}_{ij} = 4\mathrm{Re}(\langle\partial_i\psi|\partial_j\psi\rangle - \langle\partial_i\psi|\psi\rangle\,\langle\psi|\partial_j\psi\rangle)\,, \tag{6.12}$$

where $|\partial_i\psi\rangle$ denotes the derivative of $|\psi\rangle$ with respect to the ith component of $\boldsymbol{\theta}$. The QFIM can be interpreted as a metric that quantifies how a small perturbation in $\boldsymbol{\theta}$ affects the resulting quantum state as follows

$$|\,\langle\psi(\boldsymbol{\theta})|\psi(\boldsymbol{\theta}+\mathrm{d}\boldsymbol{\theta})\rangle\,|^2 = 1 - \frac{1}{4}\mathcal{F}_{ij}\mathrm{d}\boldsymbol{\theta}^i\mathrm{d}\boldsymbol{\theta}^j\,. \tag{6.13}$$

The quantum effective dimension $D_1 = \mathrm{rank}(\mathcal{F})$ is given by the rank of the QFIM, which captures the number of independent directions in the state space that can be explored by $|\psi(\boldsymbol{\theta})\rangle$ [249]. When the rank reaches its maximum and no longer increases with the addition of more circuit parameters, the circuit is said to be overparameterized, indicating that it is capable of exploring all accessible directions.

While the QFIM characterizes the capacity of a single parameterized state, QML tasks typically involve a set of training states. To properly describe such scenarios, one can generalize the QFIM to the data quantum Fisher information matrix (DQFIM), which describes the space of $U(\boldsymbol{\theta})$ that can be learned using multiple training states [247]. Specifically, consider a parameterized unitary $U(\boldsymbol{\theta}) \equiv U$ acting on a set of L training states $T_L = \{|\psi_\ell\rangle\}_{\ell=1}^L$, i.e., $U(\boldsymbol{\theta})\,|\psi_\ell\rangle$. We define the data state as

$$\rho_L = \frac{1}{L}\sum_{\ell=1}^{L}|\psi_\ell\rangle\,\langle\psi_\ell|\,, \tag{6.14}$$

which is the mixture of all training states. The data quantum Fisher information metric (DQFIM) is given by [247]

$$\mathcal{Q}_{ij}(\rho_L) = 4\mathrm{Re}(\mathrm{tr}(\partial_i U\rho_L\partial_j U^\dagger) - \mathrm{tr}(\partial_i U\rho_L U^\dagger)\mathrm{tr}(U\rho_L\partial_j U^\dagger))\,. \tag{6.15}$$

The rank of the DQFIM is given by

$$D_L(n) = \mathrm{rank}(\mathcal{Q})\,, \tag{6.16}$$

which denotes the number of learnable independent parameters of U that can be learned using dataset T_L with a circuit of K parameters. Increasing the number

of circuit parameters K enhances the expressiveness of the model, however only up to a limit. The maximal rank of the DQFIM, where one maximizes over the number of circuit parameters K,

$$R_L(n) = \max_K \mathbf{rank}(\mathcal{Q}) \tag{6.17}$$

measures the maximal possible number of independent parameters of U that can be learned via the training set T_L.

Now, we ask what is the maximal effective dimension for linear optical circuits. Consider a system of n indistinguishable photons distributed over m modes. Naively, the full Hilbert space can be written as $\mathbb{C}^{(n+1)^m}$, assuming each mode can host up to n photons. However, a fixed total number of photons n reduces the accessible Hilbert space to Fock basis states of the form $|n_1, n_2, \ldots, n_m\rangle = |n_1\rangle\,|n_2\rangle \ldots |n_m\rangle$ with $\sum_{i=1}^m n_i = n$ and $n_i \geq 0$. A simple combinatorial argument shows that the number of possible basis states is $\binom{n+m-1}{n}$, corresponding to the number of ways to choosing $m-1$ dividers to split n into m non-negative integers. Alternatively, in first quantization, the state of the system can be expressed as a tensor product of n photons $\otimes_{b=1}^n |\psi_b\rangle$ with wavefunction $|\psi_b\rangle$ of the bth photon. To construct the full many-body wavefunction, one has to take all symmetric combinations of the state.

However, since we apply only a linear optical unitary U acting on m modes, rather than a general unitary over the full Hilbert space, the set of accessible states is confined to a polynomially sized subspace within the exponentially large Hilbert space. While a straightforward parameterization of U appears to have $2m^2$ real parameters, the actual number of independent parameters R_∞ is lower due to constraints on U. First, the unitarity condition $\mathrm{tr}(UU^\dagger) = I$ imposes m^2 constraints. Second, the unitary is invariant under global phase rotations $U \to e^{i\phi}U$, introducing one additional constraint. Together, these reduce the number of independent real parameters to $R_\infty(n) = m^2 - 1$, which corresponds to the maximum number of degrees of freedom for a linear optical unitary acting on m modes, regardless of the photon number n. For linear optics, U acts on each photon individually, so it is sufficient to consider the transformation of individual photons when analyzing the system evolution.

Rank of QFIM for Fock states

We now want to understand the QFIM for a parameterized linear optical unitary U applied to a single Fock state, i.e., one training state with $L = 1$. A linear optical unitary can be written as an $m \times m$ dimensional matrix with entries u_{ks} written in the basis of optical modes $\{|s\rangle\}_{s=1}^m$

$$U = \sum_{k,s=1}^{m} u_{ks}\,|k\rangle\,\langle s|\ , \tag{6.18}$$

and is applied to a Fock state $|n_1, n_2, \ldots, n_m\rangle$. First, we consider the simplest case $n = 1$ with state $|n_1 = 1\rangle$, where only mode $s = 1$ is occupied and all other modes are empty. Without loss of generality, we assume that the photon is initially in the first mode. In this case, we have $|\psi_1(U)\rangle = U\,|n_1 = 1\rangle = \sum_{k=1}^{m} u_{k1}\,|n_k = 1\rangle$. Note that our approach is general for any single-photon state, as arbitrary state $\sum_k \alpha_k\,|n_k = 1\rangle$ can be transformed into this form via some fixed unitary V such that $|n_1 = 1\rangle = V \sum_k \alpha_k\,|n_k = 1\rangle$, and the total transformation becomes $U' = UV^\dagger$.

How many independent parameters of U can we learn by performing measurements on the state $|\psi(U)\rangle$? This can be solved through a simple counting argument [271]. The state

$$|\psi_1(U)\rangle = \sum_{k=1}^{m} u_{k1}\,|n_k = 1\rangle \tag{6.19}$$

depends only on the first column vector of U. Let us write $U = (u_1, \ldots, u_m)$, where $u_1 = (u_{11}, u_{21}, \ldots, u_{m1})^{\mathrm{T}}$ is the first column vector. By performing measurements on $|\psi_1\rangle$ and varying the parameters, one can learn about the column vector u_1. It has $2m$ real parameters as we can parameterize its coefficients as $u_{k1} = a_{k1} + ib_{k1}$. However, the parameters that can be learned are actually fewer because of the constraints on the state. In particular, the normalization condition $\langle\psi_1|\psi_1\rangle = 1$ implies $u_1^\dagger u_1 = 1$ and reduces the learnable parameters by one. Additionally, we have invariance under phase rotations, i.e., $u_1 \to e^{i\phi}u_1$ is not observable. Thus, from $|\psi_1\rangle$ we can learn

$$R_1(n = 1) = 2m - 2 \tag{6.20}$$

parameters of the column vector u_1.

Similarly, coherent states (such as those produced by lasers) without squeezing transformed by linear optical circuits are subject to the same bound as single photons, i.e.,

$$R_1^{\text{coherent}} = R_1(n = 1) = 2m - 2\,, \tag{6.21}$$

as we now proceed to show. Coherent states in m modes are written as [272]

$$|\psi_{\mathrm{c}}(\alpha)\rangle = \exp\Big(-|\alpha|^2/2\Big) \exp\left(\sum_{i=1}^{m} \alpha_i \hat{a}_i^\dagger\right) |\mathrm{vac}\rangle\ , \tag{6.22}$$

where $|\mathrm{vac}\rangle$ is the vacuum state, $\alpha_i \in \mathbb{C}$ are the complex amplitudes, and $\hat{a}_i^\dagger$ is the photon creation operator for mode i. A linear optical unitary U simply transforms the creation and annihilation operators as $U\hat{a}_i^\dagger \to \sum_k u_{ki}^* \hat{a}_k^\dagger$. Without loss of generality, let us assume that only the first mode of the input coherent state is occupied, i.e., $e^{-|\alpha|^2/2} e^{\alpha \hat{a}_1^\dagger}\,|\mathrm{vac}\rangle$. Applying the unitary transformation U yields

$$|\psi_{\mathrm{c}}(\alpha)\rangle = \exp\Big(-|\alpha|^2/2\Big) \exp\left(\alpha \sum_{k=1}^{m} u_{k1}^* \hat{a}_k^\dagger\right) |\mathrm{vac}\rangle\ . \tag{6.23}$$

Thus, the transformed state depends only on the first column of U, and the number of learnable parameters is the same as in the single-photon case.

Next, we consider the case of $n = 2$ where the photons are in two different modes, $n_1 = 1$ and $n_2 = 1$. From the photon in mode $s = 1$, we learn $2m - 2$ independent parameters of the column vector u_1. The second photon, occupying mode $s = 2$, reveals the state (written in first quantization)

$$|\psi_2\rangle = \sum_{k=1}^{m} u_{k2} |k\rangle \ , \tag{6.24}$$

which depends on the second column vector u_2 of U, also containing $2m$ real parameters. Due to normalization, we loose one parameter from the constraint $u_2^\dagger u_2 = 1$. Further, the overall phase of u_2 is unobservable, eliminating another parameter. Finally, unitarity requires that the first and second column vectors to be orthogonal, which implies $u_1^\dagger u_2 = 0$. This removes two parameters as the real and complex parts must be both zero. Thus, the number of additional learnable parameters from u_2 is $R_1(n = 2) - R_1(n = 1) = 2m - 4$. Combining u_1 and u_2, we get

$$R_1(n = 2) = 4m - 6 \tag{6.25}$$

free parameters that can be learned using two photons in orthogonal modes.

We now consider the incremental gain in learnable parameters when increasing the photon number from n to $n + 1$. Assume we begin with the Fock state $|F_n\rangle = |11 \dots 1100 \dots 0\rangle$, where n photons occupy the first n modes. This allows us to learn $R_1(n)$ independent real parameters from the corresponding column vectors $u_1, \dots, u_n$ of the unitary U. If we include an additional photon in the $(n+1)$th mode, the output state for this photon is $|\psi_{n+1}\rangle = \sum_{k=1}^{m} u_{k,(n+1)} |k\rangle$, which depends on the column vector u_{n+1}. Similarly, this column vector contains $2m$ parameters, of which two are removed due to normalization and invariance under overall phase. The new vector u_{n+1} must be orthogonal to all previous column vectors $u_1, \dots, u_n$, yielding $2n$ constraints. In total, we can learn

$$\Delta_n = R_1(n + 1) - R_1(n) = 2m - 2n - 2 \tag{6.26}$$

additional independent parameters.

Let us now sum the incremental gains and get the explicit expression for $R_1(n)$

$$R_1(n) = \sum_{k=0}^{n-1} \Delta_k = n(2m - n - 1) \, . \tag{6.27}$$

The growth of $R_1(n)$ stops when $\Delta_{m-1} = 0$, which occurs at $n = m - 1$. This means that with $n = m - 1$ photons, the QFIM reaches its maximum rank, and no additional learnable parameters can be obtained by adding more photons. For $n \geq m - 1$, the number of learnable parameters using Fock states satuates

at $R_1(m-1) = m^2 - m$. Hence, we can express the QFIM rank $R_1(n,m)$ as

$$\begin{aligned} R_1(n,m) &\leq 2mn - n^2 - n \quad (n \leq m-1)\,, \\ R_1(n,m) &= m^2 - m \qquad\quad (n \geq m-1)\,. \end{aligned} \tag{6.28}$$

This result shows that the number of learnable independent parameters for a single Fock state grows linearly with the number of modes m, but only sub-linearly with the number of photons n.

Finally, we consider the general case of multiple training states $L > 1$ to compute $R_L(n)$. Here, we count the degrees of freedom of U that can be learned by using a dataset of L training states $\{|\psi_\ell\rangle\}_{\ell=1}^{L}$. Each $|\psi_\ell\rangle$ is an arbitrary n-photon Fock state, and is able to reveal at most n columns of U. Thus, across the full dataset, up to nL columns can be accessed. One can now compute R_L as the number of independent parameters of nL columns of the $m \times m$ unitary U. This has been derived in Ref. [247], where the (tight) upper bond is given by

$$\begin{aligned} R_L(n,m) &\leq 2mnL - n^2L^2 - 1 - \delta_{L,1}(n-1) \qquad (nL \leq m)\,, \\ R_L(n,m) &= m^2 - 1 - \delta_{L,1}(m-1) \qquad\quad (nL > m)\,. \end{aligned} \tag{6.29}$$

Notably, the case $L = 1$ as derived in (6.28) is a special case, where the number of parameters is reduced due to the fact that a single Fock state cannot learn any local phases. In contrast, $L > 1$ matches Ref. [247].

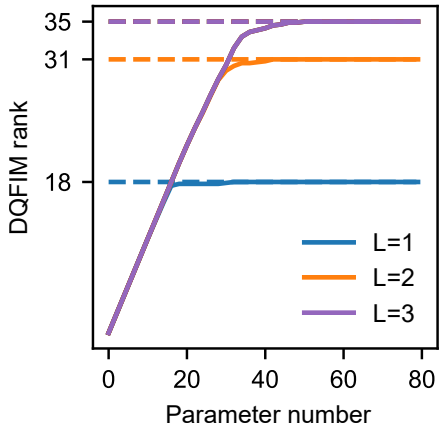

Figure 6.5: **Simulation of the DQFIM rank D_L with multiple training states.** D_L is shown against number of circuit parameters K for different numbers of training states $L = 1, 2, 3$ and a linear optical circuit with $m = 6$ modes and $n = 2$ photons.

Eq. (6.28) is verified in Fig. 2c in the main text for the case $m = 6$ and $n = 1, 2, 3, 4, 5$. Here, we further verify the general expression in Eq. (6.29) through numerical simulations with $m = 6$, $n = 2$ and varying numbers of L.

The results are presented in Fig. 6.5. As the number of trainable parameters increases, the circuit becomes overparameterized and the DQFIM rank saturates at 18, 31, and 35 for different values of L, which are in agreement with the values given by (6.29).

Efficient learning of unitaries with Fock states

Theoretical framework

We now study the problem of learning an unknown unitary V of a linear optical system. We assume that we only have oracle access to V: the internal structure of V is unknown, and we are only allowed to apply V to a set of training states of fixed dimension. The goal is to learn V by applying it to these training states, performing measurements, and inferring from the measurement outcomes.

To achieve this, we optimize a parameterized linear optical unitary $U(\boldsymbol{\theta})$, and we aim to find optimal parameters $\boldsymbol{\theta}^*$ such that

$$U(\boldsymbol{\theta}^*) = V \tag{6.30}$$

up to an irrelevant global phase. In general, to faithfully represent an m-mode unitary V, $U(\boldsymbol{\theta})$ must have at least $m^2 - 1$ real parameters. In this case, the model is said to be overparameterized. Overparameterized models are capable of fully exploring the underlying quantum state space, and as a result, they usually converge to the global minimum during training [247, 251, 273].

To train the model, we use a dataset of L input quantum states defined as

$$T_L^n = \{|\psi_\ell\rangle = W_\ell |\bar{n}\rangle\}_{\ell=1}^L , \tag{6.31}$$

where W_ℓ is a unitary for preparing the training state $|\psi_\ell\rangle$, and $|\bar{n}\rangle$ is a Fock state of n photons with at most one photon per mode. The learning process optimizes the parameters $\boldsymbol{\theta}$ such that

$$U(\boldsymbol{\theta}) |\psi_\ell\rangle = V |\psi_\ell\rangle , \forall \ell = 1, \ldots, L . \tag{6.32}$$

In practice, this is achieved by maximizing the cost function $C = L^{-1} \sum_{\ell=1}^L C_\ell(\boldsymbol{\theta})$ with

$$C_\ell(\boldsymbol{\theta}) = |\langle\psi_\ell| V^\dagger U(\boldsymbol{\theta}) |\psi_\ell\rangle|^2 = |\langle\bar{n}| W_\ell^\dagger V^\dagger U(\boldsymbol{\theta}) W_\ell |\bar{n}\rangle|^2 \tag{6.33}$$

for all ℓ.

We now ask: what is the minimal number of training states $L_{\mathrm{c}}(n, m)$ required to fully reconstruct an unknown unitary V as a function of the photon number n of the states in the training set and the number of modes m of V? We show that this is possible when the DQFIM rank reaches its maximal value $R_L = m^2 - 1$, leading to the bound

$$L_{\mathrm{c}}(n, m) = \mathrm{ceil}\left(\frac{m-1}{n}\right) + 1 , \tag{6.34}$$

where ceil(x) denotes the smallest integer greater than or equal to x. In particular, for $n = 1$ we have $L_{\mathrm{c}} = m$, while for $n \geq m - 1$ we have $L_{\mathrm{c}} = 2$ only.

Additionally, we find similar bounds for a slightly relaxed problem as studied in the main text, where we learn V only up to local phase rotations and mode permutations. That is, the learned unitary may differ from V by a diagonal phase unitary $F(\boldsymbol{\phi}) = \sum_{j=1}^{m} e^{-i\phi_j} |j\rangle \langle j|$ as well as a permutation operator P_σ which permute modes, e.g., $P_\sigma |10\rangle = |01\rangle$. Generally for quantum interference experiments, a unitary of this kind could be applied (as multiplication) to the right or to the left of unknown unitary, without altering the experiment. In this case, the number of required training states is

$$L_{\mathrm{c}}^{\mathrm{p}}(n, m) = \mathrm{ceil}\left(\frac{m-1}{n}\right), \tag{6.35}$$

and the maximal DQFIM rank is lower, given by $R_L = m^2 - m$.

This demonstrates a fundamental advantage in QML in terms of the number of training states required for learning. Specifically, single-photon states (as well as coherent light such as laser fields) require a number of training states $L \propto m$ which scales with the number of modes m. In contrast, Fock states with $n \propto m$ require only a constant number of training states $L = \mathrm{const}$.

Single-photon learning

We start by describing the training algorithm for the case $n = 1$, i.e., training with single-photon Fock states. Note that the results also apply to coherent states such as laser light, which behave similarly under linear optical transformations.

For the training set T_L^1, we select L state preparation unitaries W_ℓ, each applied on a single-photon state $|\bar{n} = 1\rangle$. To learn the target unitary V, one maximizes the cost function

$$C(\boldsymbol{\theta}) = \frac{1}{L} \sum_{\ell=1}^{L} C_\ell(\boldsymbol{\theta}) \,. \tag{6.36}$$

As shown in Ref. [247], to fully learn a general m-dimensional unitary, one requires at least $L = m$ training states. This result follows from the rank of the DQFIM, which reaches its maximum only when $L \geq m$. The reason is intuitive: to fully reconstruct V, we must be able to learn each of its coefficients v_{ij}. Let us consider the training state $|10\ldots0\rangle$, i.e., a single photon in the first mode. After applying V, one gets

$$V |10\ldots0\rangle = \sum_{i=1}^{m} v_{i1} |i\rangle \,, \tag{6.37}$$

where $|i\rangle$ denotes a single photon in the ith mode. Thus, this state provides information only about the first column of V. To fully learn V, we need at least $L \geq m$ linear independent training states to learn all columns. If we choose

$L < m$, not all columns of V are revealed, implying that we cannot completely learn V.

Note that if we want to learn V only up to local phases, $m-1$ training states are sufficient. This is because the final column of V does not need to be uniquely determined in this case.

Multi-photon learning

We now extend our analysis to learning with Fock states containing n photons. By the same reasoning as in the single-photon case, each training state now provides access to n columns of V. Consequently, fewer training states are required. For example, consider a training state $|110\dots0\rangle$ with $n=2$ photons. This state can be expressed in terms of bosonic creation operators as $|110\dots0\rangle = a_1^\dagger a_2^\dagger |\text{vac}\rangle$, where $a_j^\dagger$ creates a photon in mode j, and $|\text{vac}\rangle$ denotes the vacuum state. Under the action of a linear optical unitary V, the creation operators transform as $V a_j^\dagger = \sum_{j=1}^m v_{ij} a_i^\dagger$ [274]. Applying V yields

$$V|110\dots0\rangle = (\sum_{i=1}^m v_{i1} a_i^\dagger)(\sum_{j=1}^m v_{j2} a_j^\dagger)|\text{vac}\rangle \ . \tag{6.38}$$

In particular, we see that the transformed state now depends on the first and second columns of V, allowing one to learn both of them simultaneously.

However, there are two issues to consider when learning with multi-photon Fock states. Firstly, Fock states do not reveal information about local phase rotations. Consider the unitary consisting of mode-wise phase shifts $F(\boldsymbol{\phi}) = \sum_{j=1}^m e^{-i\phi_j} |j\rangle\langle j|$ with phase ϕ_j on the jth mode. This unitary leaves any Fock state invariant, e.g., for $n=m=4$

$$F(\boldsymbol{\phi})|1111\rangle = |1111\rangle \ . \tag{6.39}$$

As a result, to learn local phases, one requires a training state that is not a Fock state, but a coherent superposition of different modes such that $F(\boldsymbol{\phi})$ acts non-trivially on the state.

Secondly, consider learning a unitary that implements a permutation operation P_σ, which permute modes according to some permutation function $\sigma(j)=k$, with $j,k \in \{1,2,\dots,m\}$. For example, for $m=3$, the permutation function with $\sigma(1)=3$, $\sigma(2)=1$, and $\sigma(3)=2$ is a cyclic permutation of modes $(1,2,3)\to(3,1,2)$.

Crucially, permutations that only affect modes occupied by photons leave Fock states invariant. For example, consider P_σ that swaps the first and second modes $\sigma(1)=2$ and $\sigma(2)=1$. Applying this to the Fock state $|1100\rangle$, we find $P_\sigma|1100\rangle = |1100\rangle$, indicating that the state is unchanged. As the unitary P_σ has no effect on the Fock state, it cannot be distinguished from the identity operation, and thus cannot be correctly learned. This implies that when training with multi-photon Fock states, the learned unitary $U(\boldsymbol{\theta})$ can learn the target

unitary V only up to a permutation of the occupied modes. In practice, we observe that these permutation symmetries cause training to converge to local minima, even when $U(\boldsymbol{\theta})$ is overparameterized. That is, the model may become trapped in a solution associated with the wrong permutation symmetry. As a consequence, using the same training protocol as in the single-photon case becomes ineffective for the multi-photon regime.

To mitigate these problems, we now propose an efficient learning algorithm for m-mode unitaries using n-photon Fock states, which unifies the single-photon and multi-photon cases within a common framework. We represent Fock states as $|\boldsymbol{s}\rangle = |s_1, s_2, \ldots, s_m\rangle$, where s_j is the number of photons in the jth mode and $\sum_j s_j = n$. This protocol requires a total of $L(n,m) = \text{ceil}\left(\frac{m-1}{n}\right)+1$ training states, consisting of $L-1$ multi-photon Fock states with $n \geq 1$ and one single-photon state. It proceeds in three steps:

1. Train the parameterized unitary $U(\boldsymbol{\theta})$ using a training set consisting of $L-1$ n-photon Fock states as $T_{L-1} = \{\left|\boldsymbol{s}^{(\ell)}\right\rangle\}_{\ell=1}^{L-1}$. To ensure that V can be learned correctly (up to local phases and mode permutations), we require that for each mode $j = 1, 2, \ldots, m-1$, there must be at least one photon across the training set, i.e., $\sum_{\ell=1}^{L-1} s_j^{(\ell)} > 0$. We then minimize the loss function

$$C(\boldsymbol{\theta}) = 1 - \frac{1}{L}\sum_{\ell=1}^{L-1} |\langle \boldsymbol{s}^{(\ell)}|V^\dagger U(\boldsymbol{\theta})|\boldsymbol{s}^{(\ell)}\rangle|^2 \,. \tag{6.40}$$

   For $L-1 \geq \text{ceil}(\frac{m-1}{n})$, the union of the training states occupies at least $m-1$ modes. This condition is sufficient to saturate the rank of the DQFIM at m^2-m, which is the number of independent parameters of an $m \times m$ unitary up to local phases. If the learning task only requires identifying V up to local phase shifts and mode permutations, the algorithm can stop here. In that case, the minimum number of required training states is $L_{\text{c}}^{\text{p}}(n,m) = \text{ceil}\left(\frac{m-1}{n}\right)$.

   In the next two steps, an additional single-photon training state is used to learn the local phase and mode permutation, completing the full learning of V. With this additional state, the DQFIM reaches its theoretical maximum rank of m^2-1, implying that all independent parameters of the unknown unitary V can be learned.

2. We now describe how to resolve the mode permutation ambiguity. Notably, this step does not require variational learning and can be performed efficiently by measuring a single-photon state $|\psi_L\rangle = W|\bar{n}=1\rangle$, with W drawn from the Haar measure. We measure the occupation probabilities across all output modes for both the trained unitary $U(\boldsymbol{\theta}^*)$ and the target unitary

V. These are defined as

$$v_j = |\langle j| U(\boldsymbol{\theta}^*) W |\bar{n}=1\rangle|^2 , \qquad (6.41)$$
$$v'_j = |\langle j| VW |\bar{n}=1\rangle|^2 .$$

By comparing the probabilities, we can obtain the permutation P_σ (a bijective function which maps each mode $j \to j'$) such that distributions v_j and v'_j match, i.e., $v'_j = P_\sigma v_j$ for all $j = 1, \ldots, m$. P_σ can be implemented with linear optical circuits as a sequence of beam splitters which swap modes.

3. Finally, we determine the correct local phases $\boldsymbol{\phi}$ associated with a diagonal phase unitary

$$F(\boldsymbol{\phi}) = \sum_{j=1}^{m} e^{-i\phi_j} |j\rangle\langle j| . \qquad (6.42)$$

$F(\boldsymbol{\phi})$ is optimized with the same single-photon state that was used to find the mode permutation, by maximizing the cost function

$$C(\boldsymbol{\phi}) = |\langle \bar{n}=1| W^\dagger V^\dagger F(\boldsymbol{\phi}) P_\sigma U(\boldsymbol{\theta}^*) W |\bar{n}=1\rangle|^2 , \qquad (6.43)$$

where the optimization is carried out only over $\boldsymbol{\phi}$, and $\boldsymbol{\theta}^*$ are fixed to the values obtained in the first step. Once complete, we have the final result $F(\boldsymbol{\phi}^*) P_\sigma U(\boldsymbol{\theta}^*) = V$ up to a global phase.

In summary, the learning process is divided into two stages. In the first stage, training is performed using bare Fock states (e.g., $|1100\rangle$), which are sufficient to learn the target unitary up to local phase shifts and mode permutations. Empirically, we find that training reliably converges to the global minimum whenever $U(\boldsymbol{\theta})$ is overparameterized. While training with multi-photon Fock states ($n > 1$) rotated by random unitaries W_i, i.e., states of the form $W_i |1100\rangle$ (as shown in 6.31), can in principle encode local phase information, this approach often fails in practice. The training process becomes stuck in bad local minima, a failure we attribute to the permutation symmetry of Fock states. This symmetry leads to a poor optimization landscape when optimizing over linear optical unitaries. In contrast, our three-step protocol effectively resolves this issue without increasing the required number of training states while also ensuring convergence to the correct solution. Importantly, when the number of photons scales with the number of modes ($n \propto m$), the total number of required training states becomes constant. In the optimal case where $n = m - 1$, only two training states are needed to fully reconstruct the unitary. While for the single-photon case ($n = 1$), one requires $L = m$ training states, consistent with the results presented earlier in Section 6.7.

We simulate our three-step protocol for learning a unitary with $m = 5$ modes, including both local phase shifts and mode permutations. The simulation is

performed using different photon numbers n and varying numbers of training data L, where each training set consists of $L-1$ Fock states with n photons and one single-photon state to learn local phases and mode permutations. Fig. 6.6 shows the test error C_{test} after training. As predicted, larger photon numbers require fewer training states, following the theoretical bound of $L_c = \text{ceil}(\frac{m-1}{n}) + 1$. To evaluate generalization performance, the test error is computed as the infidelity between the outputs of the true unitary and the trained unitary applied to a Haar-random test state $|\psi\rangle$ as

$$C_{\text{test}} = 1 - |\langle\psi| V^\dagger F(\boldsymbol{\phi}^*) P_\sigma U(\boldsymbol{\theta}^*) |\psi\rangle|^2 . \tag{6.44}$$

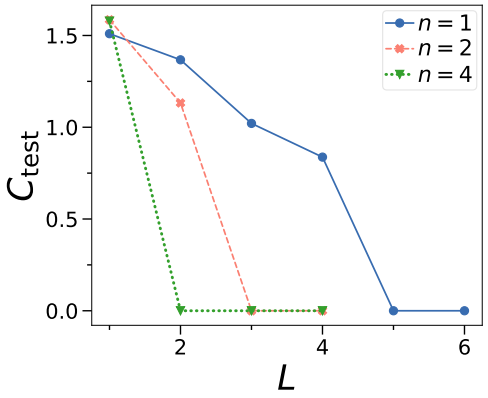

Figure 6.6: **Simulation of test error for learning unitaries using the three-step protocol.** The test error is shown for different numbers of training states L and photon numbers n, for learning a target unitary with $m = 5$ modes.

Hardware experiments for learning unitaries

Brief introduction of SPSA

To demonstrate the feasibility of photonic QML, we perform online training on our photonic quantum hardware. Starting from a random initial guess, all hardware parameters which are implemented as phase shifts in MZIs, are iteratively updated to minimize (or maximize) a given objective function.

Among various optimization strategies, gradient descent serves as the most widely-used and effective method. Gradient descent relies on gradient information, which can be obtained either by analytical expressions or finite difference approximation. Finite-difference methods are generally discouraged in quantum settings, as they demand a large number of measurements to estimate gradients with sufficient precision, making them inefficient for experimental use. In terms of analytical solutions, recent work has generalized the parameter-shift rule,

originally developed for qubit-based quantum circuits, to photonic systems [262, 275, 276]. These methods allow for the exact computation of gradients, but they still scale linearly with the number of parameters. As the parameter space grows, this scaling becomes a potential computational bottleneck [277].

In this work, we adopt a pseudo gradient-based optimization method known as simultaneous perturbation stochastic approximation (SPSA) [182]. SPSA estimates the gradients with only two objective function evaluations per iteration by applying a random perturbation vector to all parameters simultaneously, regardless of the number of trainable parameters. This random vector is usually drawn from the Rademacher distribution, where each component independently takes the value ± 1 with equal probability.

Specifically, suppose we aim to minimize an objective function $f(\boldsymbol{x})$ over a parameter vector $\boldsymbol{x}$. For each iteration, we generate a random vector $\boldsymbol{\Delta}$, and evaluate $f(\boldsymbol{x} + c\boldsymbol{\Delta})$ and $f(\boldsymbol{x} - c\boldsymbol{\Delta})$ through hardware runs and measurements. The approximate gradient for $\boldsymbol{x}$ is given by

$$\delta_{\text{approx}} = \frac{f(\boldsymbol{x} + c\boldsymbol{\Delta}) - f(\boldsymbol{x} - c\boldsymbol{\Delta})}{2c\boldsymbol{\Delta}}, \tag{6.45}$$

where $c > 0$ is the perturbation amplitude. δ_{approx} is then used to refine the parameters $\boldsymbol{x}$.

SPSA exhibits several desirable properties for hardware-based implementations. In terms of efficiency, it avoids the linearly increasing computational cost associated with varying each parameter individually. By perturbing all parameters simultaneously, SPSA maintains a constant number of objective function evaluations per iteration, making it especially suitable for high-dimensional parameter spaces. Moreover, the stochastic nature of gradient approximation in SPSA inherently provides robustness to noise, which is a critical feature for practical applications on noisy quantum hardware. The random fluctuations introduced by SPSA mirror the stochastic noise present in hardware measurements, allowing it to operate effectively even when gradients cannot be estimated precisely. While gradient-free methods such as the NelderMead algorithm have been employed to train photonic quantum neural networks[278], these methods tend to be more sensitive to hardware noise and often demonstrate inferior performance compared to SPSA [279].

In light of the above reasons, we choose the SPSA method for all physical experiments conducted in this study. To the best of our knowledge, this work represents the first application of SPSA for training photonic quantum hardware, which has only been implemented on qubit-based superconducting platforms before [208, 268].

Implementation details

As described in the main text, we test our protocol by learning a random 5×5 unitary matrix acting on five optical modes ($m = 5$), and compare the cases $n = 1$

and $n = 2$. The experiments are implemented on a six-mode programmable photonic chip, where modes $1 - 5$ are used to implement the unitary, and mode 6 is reserved for heralding when necessary.

According to the discussion in Supplementary Note 6.7, the first stage of learning requires at least four single-photon states ($n = 1$) and two two-photon states ($n = 2$). For $n = 1$, one photon is injected into modes 1, 2, 3, and 4 sequentially, corresponding to the Fock states $|10000\rangle, |01000\rangle, |00100\rangle$, and $|00010\rangle$. This is realized by a programmable optical switch, along with a tunable delay line to compensate for temporal variances across different optical paths. During each training epoch, each of the four states is inputted once and the output probability distributions are collected. Following the convention, we define the loss function as $\mathcal{L}(\theta) = C(\theta)$ according to 6.40, so that the loss value remains non-negative and converges to zero upon successful training. For $n = 2$, two identical photons are injected into mode pairs (1,2) and (3,4) in sequence, corresponding to the Fock states $|11000\rangle$ and $|00110\rangle$. We perform pseudo photon-number-resolving detection with beam splitters to measure the probability of photon bunching events where two photons reach the same output mode.

The complete experimental workflow is summarized in Algorithm 2.

Algorithm 2 Unitary learning

Input: $m \times m$ target unitary matrix V, training set containing $L - 1$ n-photon Fock states $\{|\boldsymbol{s}^{(\ell)}\rangle\}_{\ell=1}^{L-1}$, loss function $\mathcal{L}$, SPSA hyperparameters c and a.

Output: Optimized unitary matrix $U(\boldsymbol{\theta}) \approx V$, $\boldsymbol{\theta} \in \mathbb{R}^q$.

1: Select a random seed and initialize values for $\boldsymbol{\theta}$.
2: **repeat**
3: Generate a random vector $\boldsymbol{\Delta} \in \{-1, 1\}^q$ with the same dimension as $\boldsymbol{\theta}$.
4: **for** the ℓth input Fock state $|\boldsymbol{s}^{(\ell)}\rangle$ **do**
5: Set the MZI phases of the chip according to V and $\boldsymbol{\theta} + c\boldsymbol{\Delta}$.
6: Perform measurement on the quantum device and get the probability vector $\boldsymbol{p}_+^{(\ell)}$.
7: Set the MZI phases of the chip according to V and $\boldsymbol{\theta} - c\boldsymbol{\Delta}$.
8: Perform measurement on the quantum device and get the probability vector $\boldsymbol{p}_-^{(\ell)}$.
9: **end for**
10: Calculate the approximate gradient vector $\boldsymbol{\delta} \leftarrow \dfrac{\sum_\ell \mathcal{L}(\boldsymbol{p}_+^{(\ell)}) - \sum_\ell \mathcal{L}(\boldsymbol{p}_-^{(\ell)})}{2c\boldsymbol{\Delta}}$.
11: Update the trainable parameters $\boldsymbol{\theta} \leftarrow \boldsymbol{\theta} - a\boldsymbol{\delta}$.
12: **until** convergence

Regarding SPSA hyperparameters, the perturbation and learning rate are initially set as $c = 0.3$ and $a = 4$, and exponentially decay with the epoch number. We adopt decay coefficients recommended in prior work[208, 268, 280], with $\alpha = 0.602$ and $\gamma = 0.101$. These hyperparameters show ideal convergence

performance in numerical simulations for both $n = 1$ and $n = 2$ cases, and are further applied in the hardware experiments.

The result of the first training round has been presented in Fig. 3b in the main text, where the target unitary V is learned up to local phases and mode permutations. Here, we complete the learning process and implement the remaining steps as described in Supplementary Note 6.7. Specifically, a single-photon state is prepared to identify the permutation P_σ and learn the local phase unitary $F(\phi)$. To monitor the learning process, we compute the matrix closeness as $1 - \left|\frac{1}{m}\mathrm{tr}(V^\dagger F(\boldsymbol{\phi}) P_\sigma U(\boldsymbol{\theta}^*))\right|$, consistent with the definition used in the main text. It quantifies the overlap between the learned and target unitaries up to an unobservable global phase. A perfect learning result indicates that $V^\dagger F(\boldsymbol{\phi}) P_\sigma U(\boldsymbol{\theta}^*)$ is an identity matrix up to some global phase, so the matrix closeness should be close to zero.

The experimental results of second-round learning are shown in Fig. 6.7. Both $n = 1$ and $n = 2$ cases exhibit rapid convergence in approximately 150 training epochs, and the final similarity values are very close to one within the noise limits of the hardware. These results provide experimental support for the feasibility and robustness of our theoretical three-step learning protocol.

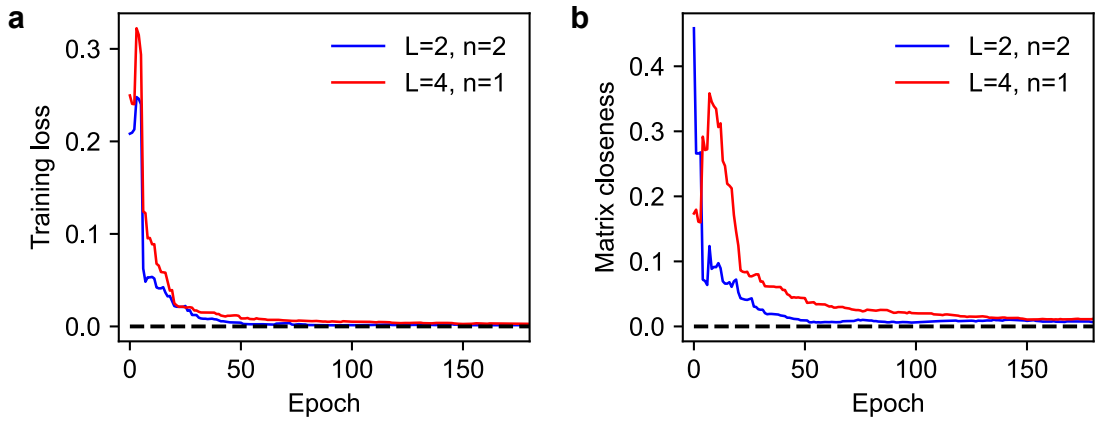

Figure 6.7: **Learning local phases of a unitary in the second round.** (a) Training loss and (b) matrix closeness for photon numbers $n = 1$ (red curves) and $n = 2$ (blue curves). In both cases, local phases are successfully learned, starting from the trained unitaries obtained in the first round of learning.

Quantum metric learning with multi-photon states

Theoretical framework

The goal of metric learning is to learn a representation in which similar data points are mapped closer together, while dissimilar ones are mapped farther apart. For an in-depth overview of metric learning techniques, we refer interested readers to comprehensive survey papers such as [256].

Here, we show that the mathematical structure of linear optical circuits is naturally compatible with metric learning. In many representative studies of metric learning [281–283], a neural network is trained to map each input to a normalized feature vector, such that all outputs lie on the surface of a unit hypersphere. This geometric constraint facilitates angular-based loss functions, which are widely used to enhance class separability. In our quantum setting, a variational linear optical circuit ansatz encodes each data point onto its parameters through some mapping scheme, and produces output probability distributions over Fock basis states, denoted by $\{p_i\}$ and $\{q_i\}$ for two respective data points. Since these are valid probability distributions $\sum_i p_i = \sum_i q_i = 1$, their square roots $\{\sqrt{p_i}\}$ and $\{\sqrt{q_i}\}$ naturally form unit vectors in a high-dimensional vector space.

A widely-used similarity measure for probability distributions is the Bhattacharyya coefficient $S_{\mathrm{C}}(p, q) = \sum_i \sqrt{p_i q_i}$, which is mathematically equivalent to the cosine similarity between $\{\sqrt{p_i}\}$ and $\{\sqrt{q_i}\}$. This enables us to interpret the output probability vectors of the photonic circuit as hyperspherical feature embeddings, analogous to those used in classical deep metric learning. We therefore define a loss function that encourages small angular distances (i.e., high cosine similarity) between embeddings of similar inputs, and large angular distances between dissimilar inputs.

A key hyperparameter in metric learning is the margin, which defines the minimum required separation between positive and negative pairs in the embedding space. The choice of margin significantly affects the difficulty of the learning task and the overall performance of the model. A larger margin imposes a stricter constraint by requiring that negative pairs be more widely separated, which makes optimization more challenging. However, this encourages the model to learn more discriminative features, thereby improving its ability to generalize and distinguish between different classes.

In this study, we integrate core concepts of metric learning into the framework of linear quantum optics. We introduce a hyperparameter t, referred to as the angular distance margin, to investigate different learning difficulty scenarios. A loss function $\mathcal{L}$ regularizes the training process as follows: for *positive* pairs (data points with the same label), the cosine similarity between the output distributions should be larger than t, while for *negative* pairs (data points with different labels), the cosine similarity should be smaller than t. Specifically, $\mathcal{L}$ is defined as

$$\mathcal{L} = \frac{1}{N_{\text{total}}}\Bigg[\sum_{\substack{(m,n)\\ y^{(m)}=y^{(n)}}} \text{ReLU}\Big(t - \sum_i \sqrt{p_i^{(m)} p_i^{(n)}}\Big) + \sum_{\substack{(m,n)\\ y^{(m)}\neq y^{(n)}}} \text{ReLU}\Big(\sum_i \sqrt{p_i^{(m)} p_i^{(n)}} - t\Big)\Bigg], \tag{6.46}$$

where $\text{ReLU}(\cdot) = \max(0, \cdot)$ is the standard rectified linear unit activation function, N_{total} is the number of all sample pairs within a training batch, $y^{(n)}$ is the class label of the nth data point, and $p^{(n)}$ is the measured real-valued probability vector for the nth data point, respectively. We fix $t = 0.9$ throughout all simulations and physical experiments presented in the main text.

Experimental implementation details

The vowel dataset in our work has been employed in multiple studies to benchmark novel machine learning hardware platforms [269, 284, 285]. The dataset consists of 12-dimensional data vectors of formant frequencies extracted from audio recordings, categorized into seven vowel classes: 'ae', 'ah', 'aw', 'er', 'ih', 'iy' and 'uw'. Each class contains 37 samples. We randomly split the dataset into training and test sets by a $70/30$ ratio, while maintaining class balance across both sets. As a result, the training and test sets contain $26 \times 7 = 182$ and $11 \times 7 = 77$ data points, respectively.

The schematic diagram of the variational circuit used for learning is shown in Fig. 6.8. To maximize the number of variational parameters on our square mesh of MZIs [56], we adopt the trainable linear scaling scheme proposed in [269] for data encoding. Specifically, we first perform min-max normalization across the whole dataset, so that all input components x_{1-12} lie in $[0, 1]$. For each component x_i, we assign two trainable parameters k_i and b_i, and compute the value of $\hat{x}_i = k_i x_i + b_i$, which is then mapped to the corresponding phase shifter in the circuit. The remaining MZIs carry free trainable parameters $(\varphi_{1-4}, \theta_{1-4})$. Because the external phases of the front MZIs do not affect the measured probabilities, we fix them to zero, and only the internal phases (θ_{5-8}) are trainable.

Similar to the unitary learning task, we employ SPSA method for online training. At each training step, we randomly sample five instances per class, resulting in $5 \times 7 = 35$ data points per mini-batch. The training loss is calculated and optimized over all pairwise combinations in the mini-batch. The experimental pipeline is shown in Algorithm 3.

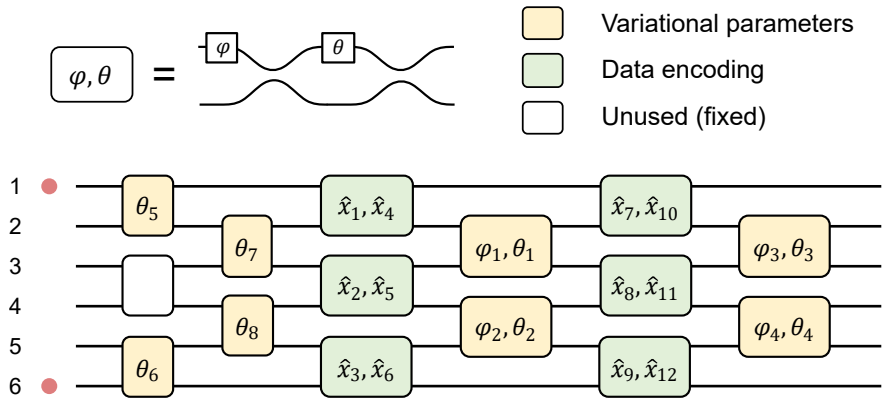

Figure 6.8: **Schematic diagram of the variational circuit used for quantum metric learning.** The phase shifters of the six-port square mesh of MZIs are configured either to encode input data components or to carry trainable parameters used for optimization.

Algorithm 3 Quantum metric learning

Input: Training and test set of data points $\{\boldsymbol{x}\}$ consisting of 12-dimensional real-valued vectors, class labels $\{y\}$, loss function $\mathcal{L}$, SPSA hyperparameters c and a, batch size N.

Output: Optimized parameters $\boldsymbol{\lambda} \in \mathbb{R}^q$, including $\varphi_{1-4}, \theta_{1-8}, k_{1-12}, b_{1-12}$ as defined above.

Select a random seed and initialize values for $\boldsymbol{\lambda}$.
repeat
  // *Training stage*
  Randomly select a mini-batch of N data points from the training set.
  Generate a random vector $\boldsymbol{\Delta} \in \{-1, 1\}^q$ with the same dimension as $\boldsymbol{\lambda}$.
  for the nth data point $\boldsymbol{x}^{(n)}$ **do**
    Set the corresponding MZI phases of the chip according to $\boldsymbol{x}^{(n)}$ and $\boldsymbol{\lambda} + c\boldsymbol{\Delta}$.
    Perform measurement on the quantum device and get the probability vector $\boldsymbol{p}_+^{(n)}$.
    Set the corresponding MZI phases of the chip according to $\boldsymbol{x}^{(n)}$ and $\boldsymbol{\lambda} - c\boldsymbol{\Delta}$.
    Perform measurement on the quantum device and get the probability vector $\boldsymbol{p}_-^{(n)}$.
  end for
  Calculate the approximate gradient vectors $\boldsymbol{\delta}_i \leftarrow \dfrac{\sum_n \mathcal{L}(\boldsymbol{p}_+^{(n)}, y^{(n)}) - \sum_n \mathcal{L}(\boldsymbol{p}_-^{(n)}, y^{(n)})}{2c\boldsymbol{\Delta}}$.
  Update the trainable parameters $\boldsymbol{\lambda} \leftarrow \boldsymbol{\lambda} - a\boldsymbol{\delta}$.
  // *Evaluation stage*
  for the nth data point $\boldsymbol{x}^{(n)}$ in the test set **do**
    Set the corresponding MZI phases of the chip according to $\boldsymbol{x}^{(n)}$ and $\boldsymbol{\lambda}$.
    Perform measurement on the quantum device and get the probability vector $\boldsymbol{p}^{(n)}$.
  end for
  Calculate the average loss value on the test set.
until convergence

Regarding the hardware setup, for the $n = 2$ case, two identical photons are generated by a spontaneous parametric down-conversion (SPDC) source and injected into ports 1 and 6. For the $n = 1$ case, only one of the two SPDC photons passes through a beam splitter whose outputs are coupled to ports 1 and 6, and the other photon is used for heralding. This configuration guarantees that the same optical paths and phase encoding structure are involved in both settings, enabling a fair comparison.

The experimental results are shown in Fig. 4c and 4d in the main text. Here, we further visualize the Gram matrix on the test set for $n = 2$ with a diverging colour map (Fig. 6.9). The matrix exhibits a clear 7×7 block structure, corresponding to the seven classes of the dataset. Although the angular distance margin was set to be $t = 0.9$ during training, we observe that most inter-class distances (off-diagonal blocks) fall well below 0.9. This suggests that the model has successfully learned a distance measure that clusters positive pairs closely and separates negative pairs, independently of the specific numerical value of t.

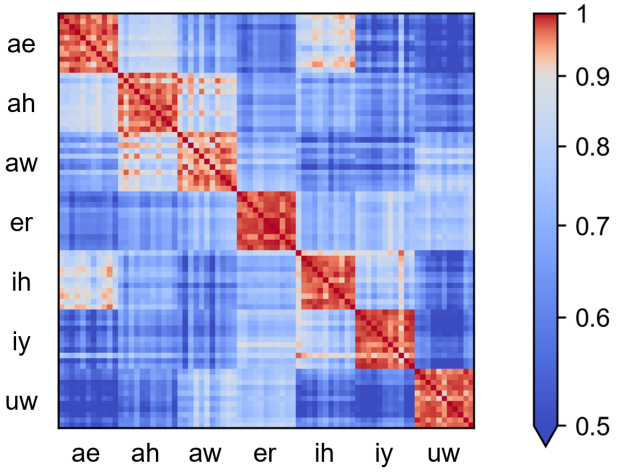

Figure 6.9: **Experimental Gram matrix on the test set for photon number $n = 2$ and margin $t = 0.9$.** This result corresponds to Fig. 4d in the main text.

To further demonstrate the robustness of online training on our photonic quantum hardware, we performed additional experiments with a different margin value $t = 0.95$, as shown in Fig. 6.10. This value approaches the average fidelity between theoretical and experimentally measured probability distributions of our photonic chip (around 0.98), making the training task more sensitive to noise and hence more challenging. Despite this, similar results are obtained as in the case of $t = 0.9$, where only a slight confusion between the classes 'ah' and 'aw' remains. This is consistent with the results of previous research [269, 285], as the instances of these two classes are intrinsically similar.

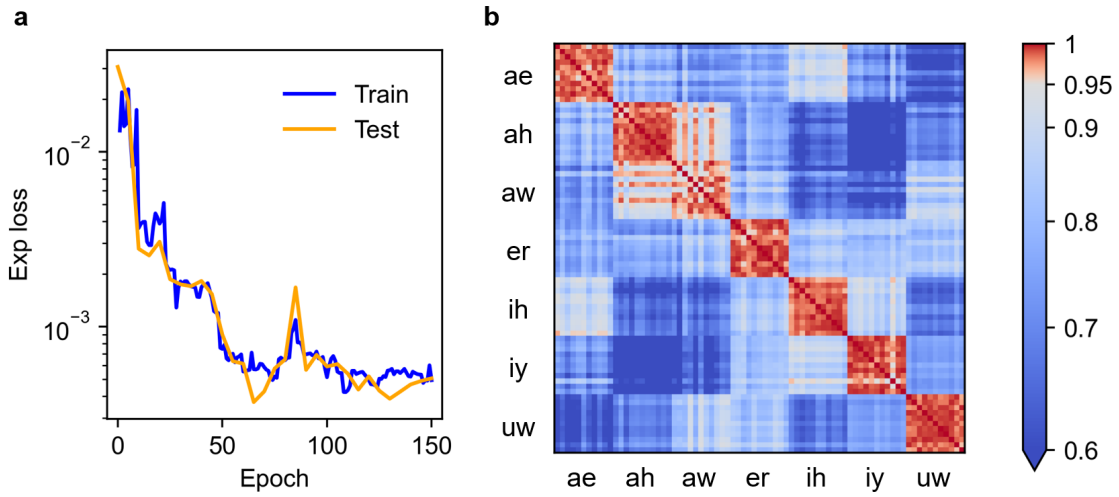

Figure 6.10: **Experimental result of quantum metric learning for photon number $n = 2$ and margin $t = 0.95$.** (a) Experimental training and test loss over epochs. (b) Gram matrix on the test set after training, which reveals inter-class and intra-class similarities.

Numerical simulations with larger photon numbers

Metric learning provides a good opportunity for comparing the model capability and trainability of different photonic configurations. In conventional classification tasks, fully-connected layers are often required to map model outputs to discrete class labels. However, since different combinations of photon and mode numbers yield output probability vectors of varying dimensions, fully-connected layers introduce varying numbers of trainable parameters, which complicates fair comparisons across models. In contrast, metric learning focuses on the inner products between probability vectors rather than the vectors themselves, thereby circumventing the issue of varying output dimensions arising from different photon and mode numbers.

In the main text (Fig. 4b), we display the performance scaling from photon number $n = 1$ to $n = 5$ under the margin hyperparameter $t = 0.9$. Here, we extend the analysis by training models with alternative values of $t = 0.5, 0.6, 0.7, 0.8$ while keeping all other settings identical. The final loss values on the test set are shown in Fig. 6.11. As the value of t decreases, the model is required to enforce larger angular distances between negative pairs, making the learning task progressively more difficult. This is reflected in the increasing loss values from $t = 0.9$ to $t = 0.5$. At the same time, we observe that the marginal improvement in loss diminishes with increasing photon number, a trend that holds across different values of t. This behavior is consistent with the sublinear scaling of the QFIM rank with respect to the photon number n. The only exceptions occur for $n = 5$ at $t = 0.5$ and $t = 0.6$. This may be attributed to the fact that, for larger

photon numbers, the trainable parameters become more highly correlated in the output probabilities, and the gradient landscape becomes more complex due to the increased difficulty of computing matrix permanents, which affects the optimization performance.

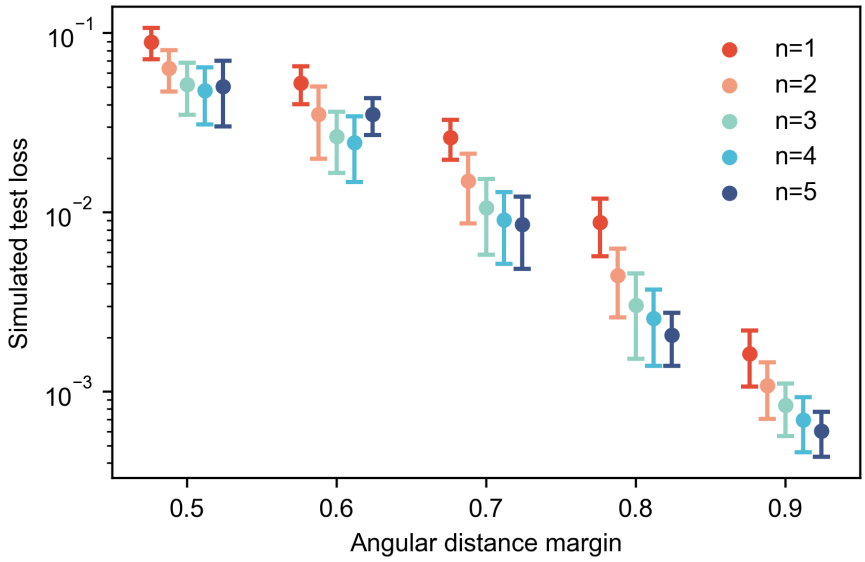

Figure 6.11: **Simulated test loss for different photon numbers** $n = 1, 2, 3, 4, 5$ **under varying hyperparameter values** $t = 0.5, 0.6, 0.7, 0.8, 0.9$**.** The results illustrate how decreasing t increases the learning difficulty and highlight the diminishing performance gain with larger n.

7 Quantum Natural Language Processing on a Photonic Quantum Computer

7.1 Introduction

Understanding human languages is a long-standing challenge for the area of artificial intelligence and machine learning [286, 287]. In the era of deep learning [174], natural language processing has delivered impressive achievements, such as machine translation, text summarization, and sentiment analysis, where words are represented in the form of trainable real-valued vectors, termed word embeddings [288]. By leveraging the power of computing hardware and algorithms, researchers have been able to develop increasingly sophisticated models for understanding and generating human language, for example, large language models [289–291]. However, the training burden of these models is also growing extremely heavy as the parameters needed in the models increase to billions, which requires massive computational resources and energy consumption.

Along the same lines, quantum computing promises the potential to alleviate some of these challenges by providing new paradigms for information processing and representation, which triggers the development of quantum machine learning (QML) [47, 155]. Utilizing quantum properties such as superposition and entanglement, QML aims to enhance the performance of machine learning tasks, such as lowering the training cost, offering higher accuracies, and providing new insights into the underlying data structures [145, 155, 292]. Among various QML applications, quantum natural language processing (QNLP) embeds words from a sentence into phases of parameterized quantum circuits, and uses the output probability distributions to perform various linguistic tasks. However, the physical implementation of QNLP tasks on quantum hardware, to the best of our knowledge, has not been demonstrated yet.

In this study, we provide the first implementation of QNLP tasks on a programmable photonic quantum computer. For proof-of-principle validation on NISQ photonic chips, we assign two trainable parameters for each word embedding,

Contribution statement: This chapter is based on an unpublished paper in preparation. In this work, I am the main contributor to the theoretical model, the data analysis, the experimental design, and the writing of the manuscript.

and different phase encoding schemes to distinguish adjectives, nouns and verbs. This is in line with previous qubit-based studies [293], where different ansätze are chosen for different parts of speech to preserve syntax information. Benefiting from the high tunability and scalability of integrated photonic circuits, we are able to implement in-situ training of 100 sentences with a total of 17 words, and achieve a classification accuracy near 100% on the test set.

7.2 Quantum natural language processing with single photons

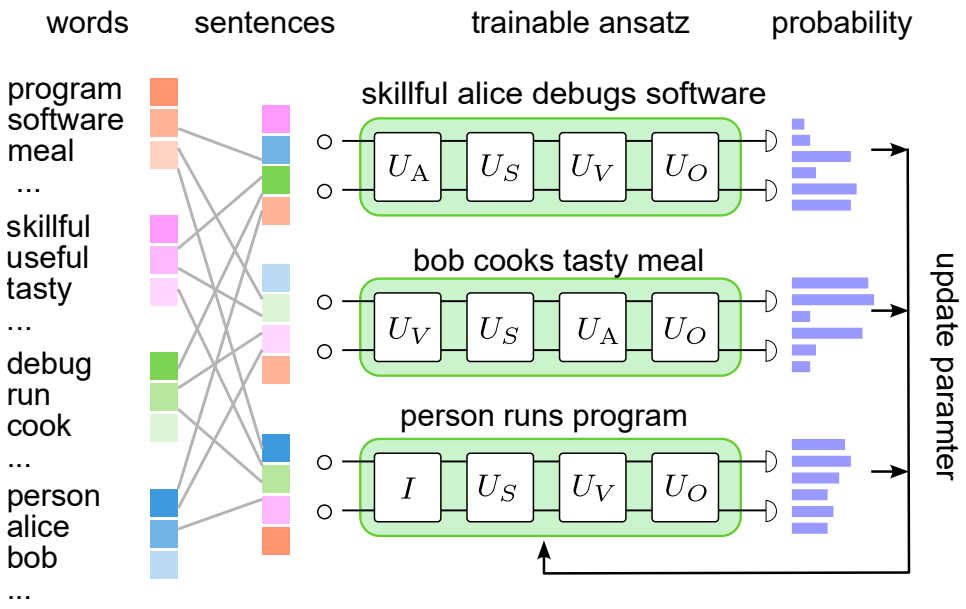

Figure 7.1: **Quantum Natural Language Processing (QNLP) with programmable photonic circuits.** For all words composing a sentence, we assign different encoding blocks for word embeddings of nouns, verbs and adjectives. Each of grammatical components is encoded as the phases of the MZIs, and output probability distributions are measured in the Fock space.

The schematic diagram of the QNLP task is shown in Fig. 7.1. For a given dataset, namely a *corpus* containing a set of sentences, one can encode each word in the sentences as several phases in an interferometer network. Thus, each word in the dataset is assigned a trainable unitary, and different sentences form distinct ansätze by sequentially concatenating the corresponding unitaries of the words that appear in the sentences. For example, the sentence "skillful alice debugs software" includes four words, "skillful", "alice", "debugs" and "software", which can be encoded as four unitaries. Note that every unitary is associated

with a specific word and the sentence is encoded as a sequence of corresponding unitaries. Here the word embeddings are parameterized but the grammatical structure of the sentence is preserved, which is crucial for the QNLP task. Of course, for complicated sentences, the ansatz takes a longer circuit depth. On the other hand, for datasets with more words, more parameters are needed for each corresponding unitary.

By injecting photons into the parameterized unitaries and measuring the output photon counting, we can obtain the output probability distribution of each sentence determined by the word embeddings. The probability distributions of different output port combinations are utilized to update the trainable parameters during the learning process.

7.3 Experimental configurations

To perform online training of QNLP tasks, we perform the text classification task and adopt the QNLP dataset originally introduced in [293] where a BNN model is trained to classify whether a given sentence is related to food (class 0) or information technology (class 1). It consists of 130 sentences and 17 words in total, with a partition of 70/30/30 into training, validation (termed as "development" in the original publication [293]) and test sets. All sentences share the same structure of "adjective (ADJ1) + noun (N1) + transitive verb (TV) + adjective (ADJ2) + noun (N2)", while the adjectives only appear in part of the sentences as illustrated in Fig. 7.2a. Specifically, the adjective is encoded on the external phases of two MZIs, while the corresponding noun is encoded on the internal phases of the same two MZIs to keep the attributive relations between the adjective and the noun. The verb is encoded on the two phases of a single MZI. If there is no adjective at some positions, the corresponding external phases are set to zero.

In order to perform the experiments on our six-port chip, we set the irrelevant MZI gates to implement identical matrices (white blocks in Fig. 7.2b). This ensures that photons will only propagate in the first four modes and relevant MZI gates. For the front three MZI gates of the chip, the external phases are set to zero (or any arbitrary values) because they do not affect the output probabilities, while the internal phases ($\theta_1, \theta_2, \theta_3$) are set to be trainable. The reason for including $\theta_1, \theta_2, \theta_3$ as trainable parameters is to let the photons diffuse into four modes and be operated by the subsequent MZI gates; otherwise the two parameters for the first adjective, $x^{(1)}_{\text{ADJ1}}$ and $x^{(2)}_{\text{ADJ2}}$, will not have influence on the output probabilities and it will make no sense to train them. This can also be seen as training an optimal input state for each sentence.

In terms of hardware setup, we select mode number as four and photon number as two. Specifically, two identical photons are injected into port 1 and port 4, and only no-collision output configurations are detected. Specifically, for four modes and two photons, we denote the $\binom{4}{2} = 6$ configurations as 12,

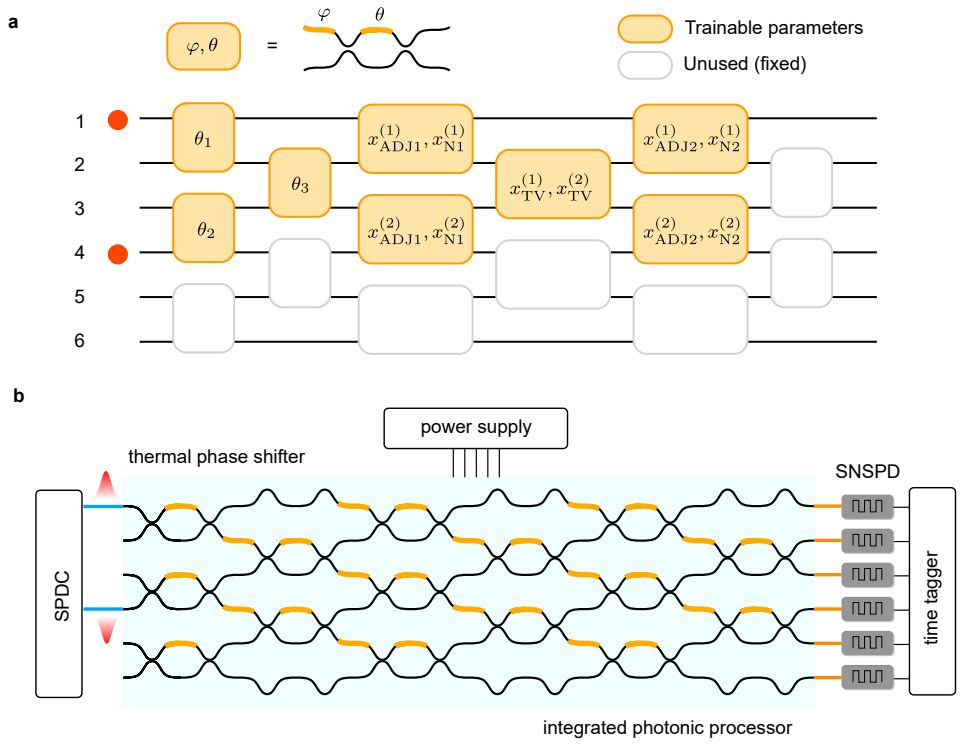

Figure 7.2: **Experimental configurations of QNLP. a**. Illustration of mapping the word embeddings onto the phases of a Clements-style six-port photonic chip in physical experiments. **b**. Experimental setup of the QNLP task, where two photons are injected into the first and forth modes of a six-port programmable photonic chip.

13, 14, 23, 24, 34. For instance, 12 indicates that two photons leave the chip from port 1 and port 2. Therefore, the output probability distribution is thus a 6-dimensional normalized probability vector $(p_{12}, p_{13}, p_{14}, p_{23}, p_{24}, p_{34})$, corresponding to which two out of four ports where two photons are output. For optimization, we define $p = p_{12} + p_{13} + p_{14}$ as the probability that a sentence belongs to class 0, so the training objective is to minimize p for sentences in class 0 and maximize p for sentences in class 1.

We perform online training on our photonic quantum hardware from three different initializations. For a given sentence, we load the word embedding values onto the corresponding phase shifters of the chip. We use the simultaneous perturbation stochastic approximation (SPSA) method to perform online training, optimizing the hardware parameters based on a binary cross-entropy

loss function:

$$\mathcal{L}_{\text{pos}} = \frac{1}{N}\sum_{n}\mathcal{L}(\boldsymbol{p}_{+}^{(n)}, y^{(n)}),$$
$$\mathcal{L}_{\text{neg}} = \frac{1}{N}\sum_{n}\mathcal{L}(\boldsymbol{p}_{-}^{(n)}, y^{(n)}). \tag{7.1}$$

The approximate gradient vectors are then calculated as

$$\boldsymbol{\delta}_i = \frac{\mathcal{L}_{\text{pos}} - \mathcal{L}_{\text{neg}}}{2\boldsymbol{\Delta}_i} \tag{7.2}$$

Subsequently, we update the vector embeddings $\boldsymbol{x}_i \leftarrow \boldsymbol{x}_i - a\boldsymbol{\delta}_i$ for all words in the mini-batch.

The complete pipeline is shown in Algorithm 4. Regarding the hyperparameters, we set the batch size to 5, resulting in 70/5=14 mini-batches per epoch. The perturbation and learning rate for the SPSA method are set to constant values, with $c = 0.25$ and $a = 1$.

In our physical experiments, two photons are generated from a spontaneous parametric down-conversion source and injected into the first and fourth modes of a six-port integrated photonic chip based on laser-written fused silica waveguides. All phase shifters are thermo-optical and can be tuned independently by applying a current. As photons are injected into four modes of the photonic chip, only no-collision events are detected.

7.4 Results

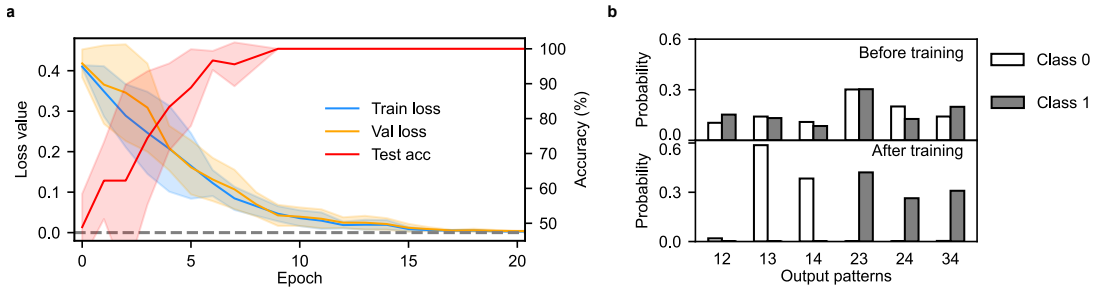

Figure 7.3: **Training results of quantum natural language processing. a**. The loss values of the training and validation sets during the training process. The loss values gradually converge to zero, and all sentences in the test set are correctly classified. **b**. The average output probability distributions on the test set before and after training. The probability distributions are randomly distributed before training, while clear patterns emerge after training.

Algorithm 4 Quantum Natural Language Processing (QNLP)

Input: Training and validation set of sentences consisting of M words, binary class labels $\{y\}$, SPSA hyperparameters c and a, loss function $\mathcal{L}$, batch size N.

Output: Vector embeddings $\{\boldsymbol{x}_i\}_{i=1}^{M} \in \mathbb{R}^q$ for all M words.

Select a random seed and randomly initialize values for $\{\boldsymbol{x}_i\}_{i=1}^{M}$.
repeat
  // *Training stage*
  for all mini-batches in the training set **do**
    Randomly select a mini-batch of N sentences from the training set.
    Generate random vectors $\boldsymbol{\Delta} \in \{-1, 1\}^q$ for all included words in the mini-batch.
    for the nth sentence **do**
      Set the corresponding MZI phases of the chip to $\{\boldsymbol{x}_i + c\boldsymbol{\Delta}_i\}$ for all words.
      Perform measurement on the quantum device and get the result $\boldsymbol{p}_{+}^{(n)}$.
      Set the corresponding MZI phases of the chip to $\{\boldsymbol{x}_i - c\boldsymbol{\Delta}_i\}$ for all words.
      Perform measurement on the quantum device and get the result $\boldsymbol{p}_{-}^{(n)}$.
    end for
    Calculate the average SPSA loss values for the mini-batch $\mathcal{L}_{\text{pos}}, \mathcal{L}_{\text{neg}}$.
    Calculate the approximate gradient vectors $\boldsymbol{\delta}_i$.
    Update the vector embeddings
  end for
  // *Evaluation stage*
  for the nth sentence in the validation set **do**
    Set the corresponding MZI phases of the chip to $\{\boldsymbol{x}_i\}$ for all words.
    Perform measurement on the quantum device and get the result $\boldsymbol{p}^{(n)}$.
  end for
  Calculate the average loss value on the validation set.
until convergence

We perform three independent runs for the hardware experiment with different random seeds (thus different parameter initializations), and the results are shown in Fig. 7.4. The dark blue and light blue curves correspond to $\mathcal{L}_{\text{pos}}$ and $\mathcal{L}_{\text{neg}}$ in Algorithm 4, which are calculated based on the experimental hardware probability outputs for each mini-batch.

The loss value on the validation set (orange lines) and the test accuracy on the test set (red lines) are evaluated at the end of each epoch when all training samples are traversed. For every run, the test accuracy reaches 100% before 10 epochs, while the loss value gradually decreases until it reaches the minimum. This demonstrates the feasibility and robustness of online training on our photonic quantum hardware.

Figure 7.3a indicates successful training outcomes for three independent runs, with all sentences in the test set correctly classified. This is further validated by plotting the average output probability vectors over the test set (Figure 7.3b). The probabilities of different port combinations exhibit desired patterns that are consistent with the training objective after optimization.

7.5 Discussion

In this work, we have demonstrated the first implementation of quantum natural language processing tasks on a programmable photonic quantum computer. By encoding the word embeddings of a sentence into the phases of a photonic chip, we can perform online training of the word embeddings and achieve a classification accuracy approaching 100% on the test set. The training is performed on a six-port integrated photonic chip, which is capable of performing the QNLP task with two photons and four modes. The results demonstrate that the photonic quantum computer can be used to perform QNLP tasks, and the training process is robust and efficient.

Looking toward future applications, we expect that the QNLP task can be performed on larger photonic chips with more modes and photons, which will allow us to encode more complex sentences, such as including clauses and phrases, and to perform more sophisticated tasks, such as machine translation and text summarization. We also expect that the QNLP task can be enhanced by adopting photonic nonlinearity, which will allow us to perform more complex operations on the quantum states and achieve better performance on these tasks.

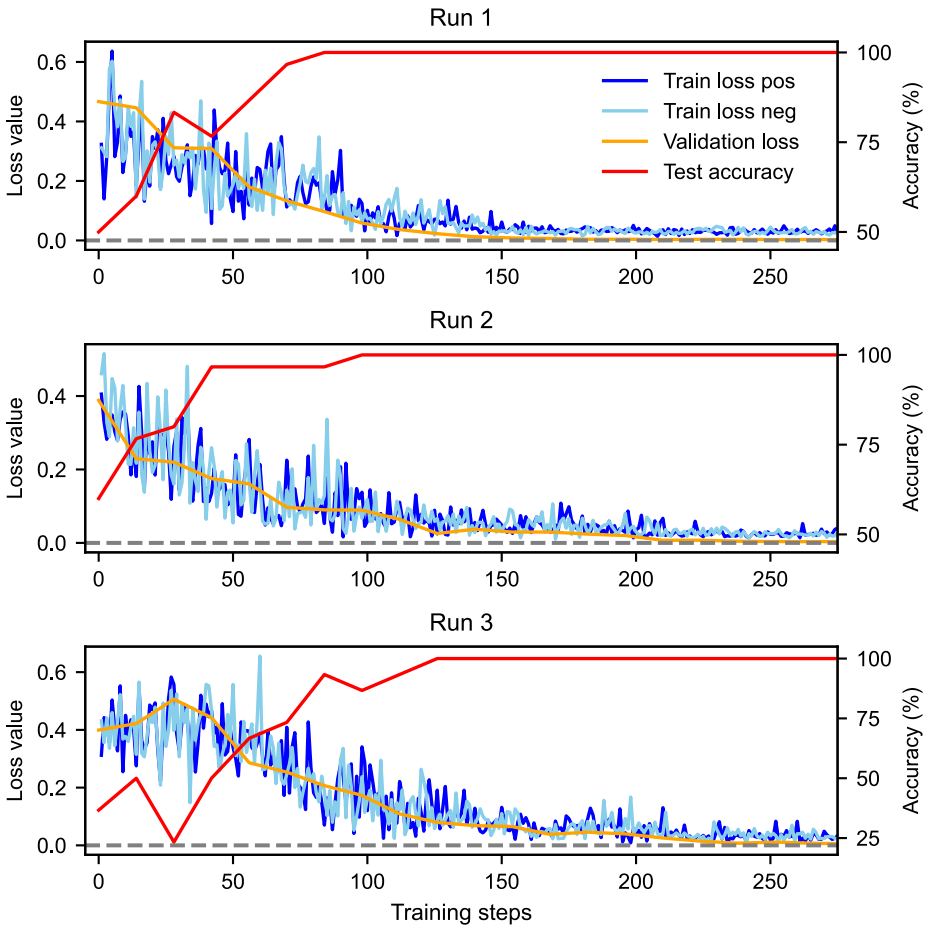

Figure 7.4: **Convergence plot of hardware experiments for quantum natural language processing.** The hardware experiments are performed with three independent runs and different initializations. Dark blue and light blue curves represent the average loss values for the mini-batch during SPSA optimization, with parameters perturbed in opposite directions. Orange curve represents the average loss evaluated on the validation set at the end of each epoch. Gray dashed line shows the zero loss value. Red curve represents the classification accuracy evaluated on the test set at the end of each epoch. One epoch corresponds to 14 training steps.

8 Towards Quantum Edge Computing in Space

From earth observation to space exploration, humanity is leveraging space technology to deepen our understanding of both the universe and our own planet. To achieve this, we are integrating advanced technologies developed on the ground into space missions, providing us with invaluable data and insights. Whether sending probes to the farthest reaches of the solar system or launching satellites to monitor the earth's atmosphere, oceans, and land, our efforts are expanding the frontiers of knowledge.

The rapid advancement of space technologies has led to an unprecedented increase in the volume and variety of data generated in orbit. This data includes satellite imagery, sensor measurements, and results from scientific experiments conducted in space. However, processing and analyzing such data directly onboard satellites remains a significant challenge due to the harsh space environment. For instance, while field-programmable gate arrays (FPGAs) are widely used in spaceborne computing units, standard FPGAs are vulnerable to radiation [294]. Radiation-hardened FPGAs offer improved reliability but are costly and often limited to one-time programming. In addition to computational and energy constraints, satellites must also contend with extreme temperatures, vacuum, and radiation. As a result, large volumes of data must be transmitted back to ground stations for further analysis, incurring high latency and substantial bandwidth requirements [295, 296].

Meanwhile, quantum technologies are progressing rapidly, offering the potential to solve complex computational problems beyond the reach of classical computers and to secure communications with fundamentally unbreakable encryption. This progress has sparked growing interest in applying quantum technologies to space applications, both in academia and industry [297]. A notable example is the *Micius* satellite, launched by China in 2016, which successfully demonstrated quantum key distribution (QKD) over long distances, paving the way for secure global communications. In this mission, entangled photon pairs were generated onboard and distributed to two ground stations, enabling demonstrations of quantum key distribution, quantum teleportation, and entanglement swapping [298–300]. The same team also developed a CubeSat-based QKD system, significantly reducing the cost and size of the satellite [301]. Looking ahead, the *QEYSSAT* mission, proposed by Thomas Jenewein et al., aims to demonstrate quantum communication in low Earth orbit (LEO) and establish a QKD network between satellites and ground stations [302, 303]. Another

example is the *SpooQy-1* experiment, which seeks to demonstrate the generation and detection of entangled photons in space using a CubeSat platform [304].

Despite these advances, most space-based quantum missions to date have focused on quantum communication, with relatively little attention given to quantum computation in space. Bringing quantum computational capabilities directly onboard satellites would enable in-situ data processing and analysis, reducing both latency and bandwidth requirements. For example, quantum algorithms could be used to precompress data before transmission to ground stations, or quantum states could be encoded in single photons and sent to Earth, offering greater robustness against noise and loss.

In this chapter, we introduce our project, the *Reconfigurable Lower Orbit Quantum Computer for Earth observation Technology* (ROQUET), which aims to demonstrate quantum edge computing in space. By integrating quantum computation with satellite platforms, ROQUET seeks to enable efficient, low-latency data processing for Earth observation and other space applications.

8.1 Experimental payload

To harness the power of photonic quantum technology, it is necessary to generate, manipulate, and detect single photons with high accuracy. As illustrated in Fig. 8.1, the core quantum experiment includes a single photon source, a programmable integrated photonic chip, and customized multi-channel single photon avalanche detectors (SPAD). Fiber components connect all these modules. To control all signals, including analog, digital, and radio frequency signals, we have designed several printed circuit boards to control phase shifters on the chip, regulate the temperature of the nonlinear crystal in the SPDC source, and distribute power at different voltages.

Here we introduce each component across several categories.

Optics

The general optical modules include beam splitters, wave plates, and mirrors for manipulating the polarization and path of the photons.

1. **Laser** We use a laser diode at 405 nm to pump the nonlinear crystal and generate photon pairs at 810 nm. The laser is pigtailed with a PM fiber to ensure polarization and mode quality. A simple communication interface is provided to control the laser switch and temperature.

2. **Single photon source** We use a ppKTP crystal to generate photon pairs at 810 nm via spontaneous parametric down-conversion (SPDC). To ensure the degeneracy and indistinguishability of the photons, we control the temperature of the crystal with a Peltier element. The pump laser and output photons are coupled in fibers using a triplet of lenses.

3. **Delay line** Based on the phase matching condition in the SPDC process, it is critical to maintain the fiber length difference within the two photons' coherence length, which is approximately one millimeter. To achieve this, we designed a free-space delay line to compensate for the fiber length difference. During assembly, we can adjust the position of the mirror and fix both collimators with adhesive.

4. **Programmable PIC** The core of our quantum experiment is a programmable photonic integrated circuit, which is fabricated with glass waveguides using femtosecond laser writing technique. The chip features 6 input and 6 output modes; both input and output are coupled with PM fiber arrays.

5. **SPAD** Single photon detection is realized with a custom-designed multi-channel single photon avalanche detector (SPAD) array. From the 6 fiber outputs of the PIC, we split each channel into collimators which are mounted in front of the single avalanche photodiodes. Each channel is equipped with a dedicated readout circuit to ensure high timing resolution and low dark count rate. To read out the electrical signal from the SPAD, we designed a passive quenching circuit for each channel. High voltage is supplied to the SPAD pin to set inverse bias above the breakdown voltage.

Electronics

The general electronics modules include current controllers, temperature controllers, power system, and on-board computer. All these modules can be turned on/off separately to save power consumption.

1. **Chip current controller** The chip current controller is designed to provide stable and precise current to the laser diode and other components. It features a 16-bit DAC current output for up to 29 channels and is capable of fast modulation for applications requiring high-speed operation. Built-in voltage comparators are included for measuring the voltage at each channel.

2. **Temperature controller** The temperature controller (TEC) is responsible for maintaining the optimal operating temperature of the laser diode and the nonlinear crystal. It utilizes a Peltier element for active cooling and heating, with a feedback loop to ensure precise temperature control.

3. **Power system** The power system is designed to convert the input power from the satellite bus to the required voltage levels for each component. It includes DC-DC converters and voltage regulators to ensure stable and efficient power delivery.

4. **On-board computer** The on-board computer features a space-qualified system-on-chip processor, AMD Zynq 7020. Along with a processing system

(PS) in this chip, a main advantage of this processor is that it includes programmable logic (PL), namely an FPGA core. We separate the hardware control parts and the application part into PL and PS respectively. The real-time PS parts run a customized version of Linux, including C/C++ and Python environments. The PL implements the time-tagger to record detection events from SPADs and send the data to PS via the AXI interface. Other on/off control signals are also implemented in PL.

5. **Camera** The camera is a commercial off-the-shelf product, specifically the Alvium G1 series developed by Allied Vision. The image sensor is Sony IMX249 with a resolution of 1936 (H) $\times$ 1216 (V) pixels (2.4 MP). It is capable of capturing RGB images of Earth's surface with a resolution of approximately 35 m/px. The interface to the camera is provided through an Ethernet connection, which can be directly connected to the on-board computer.

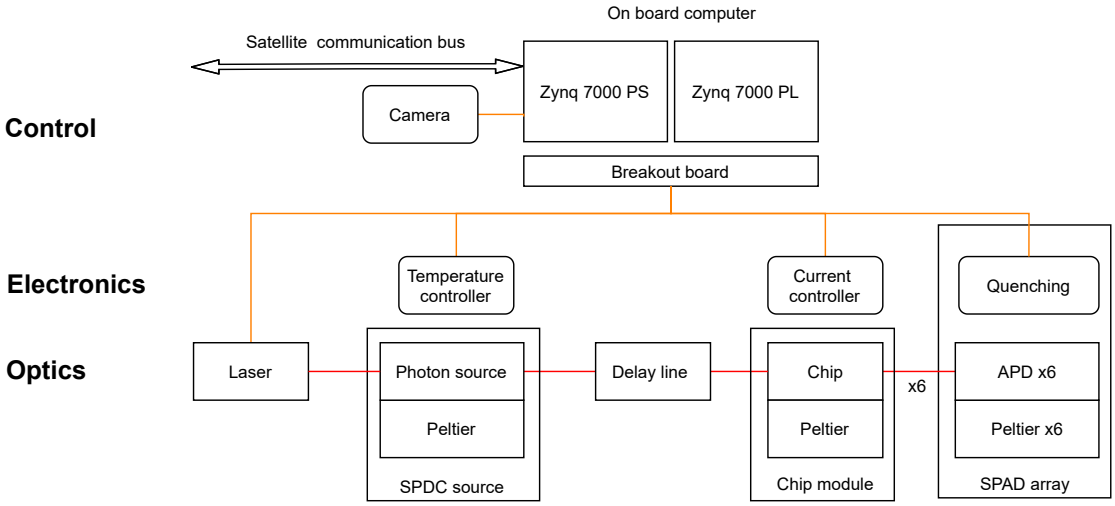

Figure 8.1: Experimental apparatus scheme. The setup contains three levels of experiments: optical quantum experiments, electronics control, and operating system. The SPDC source generates photon pairs, which are processed by the photonic chip and detected by SPADs. To control the temperature of the nonlinear crystal, chip, and SPADs, dedicated temperature controllers are designed and integrated into one PCB. The current controllers operate the phase shifters on the photonic chip. All signals are first processed by the FPGA-level firmware, then connected to the on-board computer for further programming and analysis.

Beyond this, a dedicated mechanical structure is required to assemble all components. The entire payload involves many physical interfaces, including

optical fibers, electrical connections, and thermal management systems.

Dimension	4 Units (3U optical setup, 1U camera system)
Weight	∼10 kg
Power	Camera: 3.2W On Board Computer (incl. FPGA): 2W Optical Setup: Max. 15W during boot of the experiment Operational mode: ∼5W
Operating Temperature	-20°C to 50°C
Voltage	5V, 12V for power supply and 3.3V for digital control
Communication	To Satellite On-Board Computer: Ethernet, SPI, I2C, Serial
Connectors	Fiber: FC/APC and splicing Electrical: Harwin M80, Coaxial
Materials	Carbon fiber, Titanium, Silica, Aluminium, Glass

Table 8.1: Overview of the payload specifications

8.2 Proposed experiment

With the experiments described below, we aim to demonstrate positive quantum effects for machine learning applications typical for remote sensing in Earth sciences.

1. **On-Board Calibration** In principle, the optical chip can operate with both quantum and classical light. Since we do not integrate a classical light source and detectors in the payload, we will use the single photon source and the SPADs to perform calibration in demand. This enables precise calibration of the chip's settings directly onboard the satellitea capability not previously demonstrated in this environment.

2. **Single Photon Interference** We will perform the first satellite-based experiment demonstrating single photon interference using a source that generates two indistinguishable photons. By directing these photons through the optical quantum processor, we can observe quantum interference effects such as the Hong-Ou-Mandel (HOM) dip through tuning the temperature of the crystal. This phenomenon has not yet been experimentally realized in space.

3. **Quantum Machine Learning for Pixel Classification** The objective is to classify each pixel in satellite images into predefined categories. Using RGB bands, our experiment targets 13 classes, including wooded, agricultural, residential, industrial, commercial, recreation, airport, quarry, military,

desert sand, mountain rock, other natural, and none. Quantum machine learning algorithms implemented on the photonic processor are expected to improve classification accuracy compared to classical approaches, as supported by simulations and laboratory tests.

4. **Quantum Autoencoder for Image Compression** We will implement a hybrid neural network that combines classical layers with a quantum photonic processor as the final layer [305, 306]. The goal is to encode and compress image data efficiently, reducing the number of features while preserving essential information. This approach is anticipated to achieve higher compression rates for RGB images, leveraging quantum effects for enhanced performance.

8.3 Outlook

In June 2025, the ROQUET payload was successfully launched aboard the *ION* satellite by D-Orbit, reaching a low Earth orbit (LEO) at an altitude of approximately 500 km. ROQUET represents a pioneering step towards quantum edge computing in space, enabling direct onboard processing of satellite images using photonic quantum hardware. This approach promises to reduce latency and bandwidth demands, making it highly relevant for Earth observation applications such as climate monitoring and resource management. Key challenges remain, including ensuring mechanical robustness, further miniaturization, and optimizing energy efficiency. Continued advancements in these areas will be crucial for realizing the full potential of quantum computing in space and expanding its impact on remote sensing and other scientific domains.

Looking ahead, the integration of quantum machine learning with classical methods offers the potential for more efficient and accurate data analysis while also minimizing energy consumption. The compact, modular, and low-power photonic systems developed for ROQUET set a foundation for future space missions and sustainable computing platforms.

9 Reprint of "Experimental quantum-enhanced kernel-based machine learning on a photonic processor"

nature photonics

Article

https://doi.org/10.1038/s41566-025-01682-5

Experimental quantum-enhanced kernel-based machine learning on a photonic processor

Received: 13 August 2024

Accepted: 16 April 2025

Published online: 02 June 2025

Check for updates

Zhenghao Yin[1,2] ✉, Iris Agresti[1] ✉, Giovanni de Felice[3], Douglas Brown[3], Alexis Toumi[3], Ciro Pentangelo[4,5], Simone Piacentini[5], Andrea Crespi[4,5], Francesco Ceccarelli[5], Roberto Osellame[5], Bob Coecke[3] & Philip Walther[1,6,7] ✉

Recently, machine learning has had remarkable impact in scientific to everyday-life applications. However, complex tasks often require the consumption of unfeasible amounts of energy and computational power. Quantum computation may lower such requirements, although it is unclear whether enhancements are reachable with current technologies. Here we demonstrate a kernel method on a photonic integrated processor to perform a binary classification task. We show that our protocol outperforms state-of-the-art kernel methods such as gaussian and neural tangent kernels by exploiting quantum interference, and provides further improvements in accuracy by offering single-photon coherence. Our scheme does not require entangling gates and can modify the system dimension through additional modes and injected photons. This result gives access to more efficient algorithms and to formulating tasks where quantum effects improve standard methods.

The past decades have witnessed a swift development of technologies based on quantum mechanical phenomena, opening up new perspectives in a wide spectrum of applications. These range from the realization of a global-scale quantum communication network—the quantum internet—to the simulation of quantum systems and to quantum computing[1,2]. In particular, the interest towards quantum computing has been fuelled by some milestone discoveries, such as Shor's[3] factorization and Grover's search algorithm[4], which have indicated that quantum processors can outperform their classical counterparts. However, a clear advantage of quantum computation has been experimentally demonstrated only recently and on different computational tasks, such as boson sampling[5–7] and random circuit sampling[8], which do not have clear practical applications.

Given these premises, our goal is to investigate the tasks in which quantum computing can enhance the operation of classical computers for practically relevant tasks. Moreover, the question is whether this can be achieved for problems that are now within the reach of state-of-art technology, where only noisy intermediate-scale quantum computers are available[9]. In this context, a flurry of interest has been devoted to the open question of whether the new paradigm of quantum computing can have an impact on machine learning[10], which has revolutionized classical computation, granting new possibilities and changing our everyday lives, from email filtering to artificial intelligence. The two main directions that have been investigated so far are, on the one hand, whether quantum computation could improve the efficiency of the learning process, enabling us to find better optima with the need of a

[1]University of Vienna, Faculty of Physics, Vienna Center for Quantum Science and Technology (VCQ), Vienna, Austria. [2]University of Vienna, Faculty of Physics, Vienna Doctoral School in Physics (VDSP), Vienna, Austria. [3]Quantinuum, Oxford, UK. [4]Dipartimento di Fisica, Politecnico di Milano, Milan, Italy. [5]Istituto di Fotonica e Nanotecnologie, Consiglio Nazionale delle Ricerche (IFN-CNR), Milan, Italy. [6]Christian Doppler Laboratory for Photonic Quantum Computer, Faculty of Physics, University of Vienna, Vienna, Austria. [7]Institute for Quantum Optics and Quantum Information (IQOQI) Vienna, Austrian Academy of Sciences, Vienna, Austria. ✉e-mail: zhenghao.yin@univie.ac.at; iris.agresti@univie.ac.at; philip.walther@univie.ac.at

https://doi.org/10.1038/s41566-025-01682-5

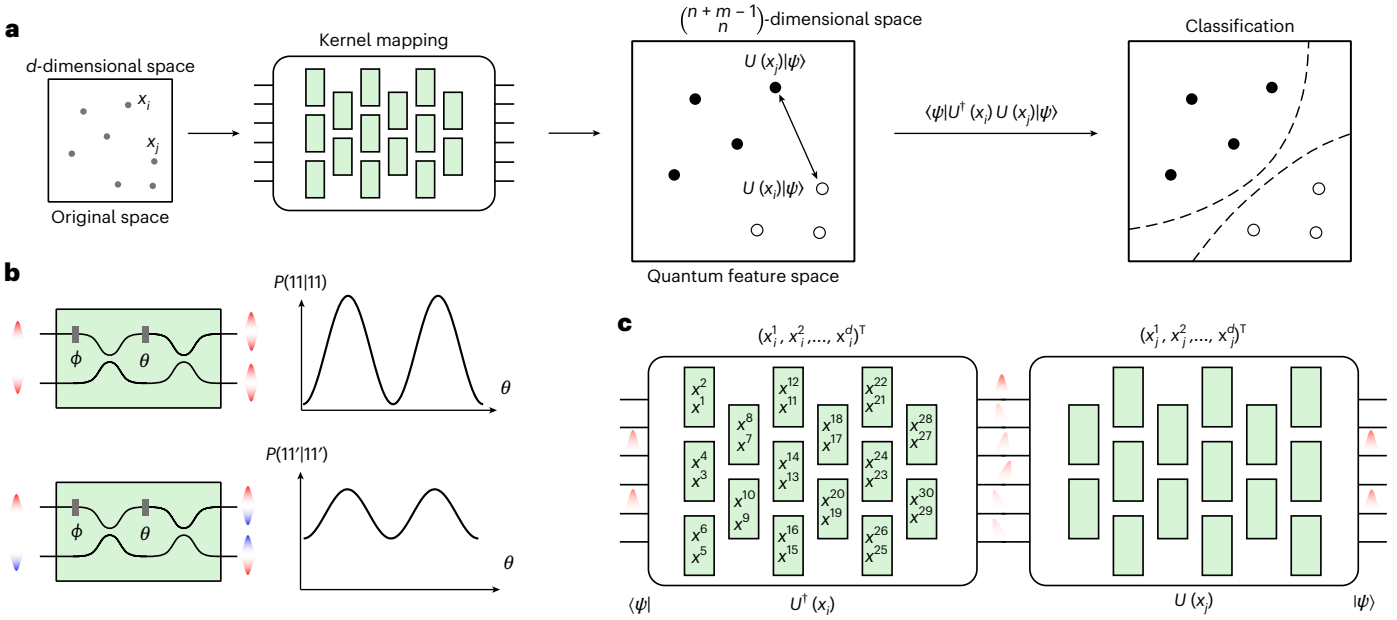

Fig. 1 | Photonic quantum kernel estimation. a, The photonic quantum kernel maps each data point x_i to be classified from a d-dimensional space into a quantum state $|\Phi\rangle_i$, living in a Hilbert feature space. In detail, the classical data x_i are encoded into a unitary evolution $U(x_i)$ applied on a fixed input state $|\psi\rangle$. This implies $|\Phi\rangle_i = U(x_i)\,|\psi\rangle$. After mapping all of the data points in the dataset, from the inner pairwise products, given by $\langle\psi|U(x_i)^\dagger U(x_j)|\psi\rangle$, we perform the classification, finding the hyperplane that best separates the classes, that is, through an SVM, according to equation (1). **b**, Pairs of indistinguishable photons (top) and distinguishable photons (bottom) show different behaviour when injected into a Mach–Zehnder interferometer (MZI), with external phase Φ and internal phase θ. Here, input states 11 and 11′ indicate, respectively, two indistinguishable and distinguishable photons being injected into the circuit and being detected at the output modes. $P(o_1, o_2|\, i_1, i_2)$ is the probability of registering o_1 and o_2 photons, respectively, on the first and second output of the Mach–Zehnder interferometer, conditioned on having i_1 and i_2 photons on the input modes. **c**, Estimation of the inner product of two data points x_i and x_j by encoding them in two unitaries $U(x_i)$ and $U(x_j)$. The inner product $\langle\phi_j|\phi_i\rangle$ amounts to $\langle\psi|U^\dagger(x_i)U(x_j)|\psi\rangle$. This is equivalent to projecting the evolved state $U^\dagger(x_j)U(x_i)|\psi\rangle$ onto $|\psi\rangle$. Each green-shaded box represents a programmable MZI with two free parameters (namely, a beamsplitter with tunable reflectivity and phase), as shown in **b**.

lower number of inquiries[11–14] and, on the other, how quantum behaviours can enhance the expressivity of the input encoding, exploiting correlations between variables that are hard to reproduce through classical computation[15,16].

In this context, a straightforward application of quantum computing on kernel models has become evident. Kernel methods are widely used tools in machine learning[17,18] that base their functioning on the fact that patterns for data points, which are hard to recognize in their original space, can become easy to identify once mapped nonlinearly to a feature space. When the suitable mapping is performed, it is possible to identify the hyperplane that best separates the classes of feature data points, through a support vector machine[19] (SVM), according to the inner product of the mapped data. Let us note that the only part of the model that is trained is the SVM, which is efficient, once the inner products are available. Hence, an interesting question is whether using a quantum apparatus to perform the data mapping and evaluate the inner products can enhance the overall performance of the algorithm, benefitting from feature maps exploiting the evolution of quantum systems and outsourcing the hardest part of the computation to the quantum hardware. This question was theoretically and rigorously answered in the affirmative by Liu and colleagues[20], although implementation of the proposed task is out of reach for state-of-the-art technologies. Hence, it is still an open question, whether some improvement in the performance of such a method can be achieved through a quantum feature map, for tasks that can be implemented on current quantum platforms. In this context, a moderately sized quantum feature space can be proved to be more suitable for preserving the similarity among data that belong to the same class. The reason for this is that the advantage of a high-dimensional feature space, enabling the expressivity of the model to be improved, is cancelled out by the dimension increase per se that typically decreases the overlap between different states, such that they become almost orthogonal[21]. This is due to the fact that the expected overlap value of two random states amounts to the inverse of the space dimension and, in high-dimensional metric spaces, almost all probability measures concentrate around the mean or median of a function (concentration of measure[22]). The effect is then that the inner products among the feature data points will tend to vanish and a higher precision will be required to properly measure them. This ultimately leads to the so-called exponential concentration[23,24], where it is not possible to capture the information about the similarities between states, making the model—to all intents and purposes—useless for any type of classification.

Here we investigate this topic and experimentally demonstrate a quantum kernel estimation, where feature data points are evaluated through the unitary evolution of two-boson Fock states (Fig. 1). Such encoding, even for relatively small dimensions, proves suitable as it enables high classification accuracies to be achieved.

Furthermore, we show that, for given tasks, this algorithm leads to an enhancement in the performance of quantum kernels with respect to their classical counterparts. These tasks are selected by maximizing the so-called geometric difference, which measures the separation in performance between a pair of kernels[21]. In particular, we separate between quantum and classical kernels, that is, photonic kernels that do or do not exhibit quantum interference, respectively.

In our implementation, we exploit a photonic platform based on an integrated photonic processor[25] where we inject two-boson Fock states to map the data to be classified (Fig. 1a). To estimate quantum and classical kernels, we inject indistinguishable and distinguishable photons, respectively. This photonic platform is particularly suitable for this task, as it enables us to encode and manipulate our input data with high fidelity.

https://doi.org/10.1038/s41566-025-01682-5

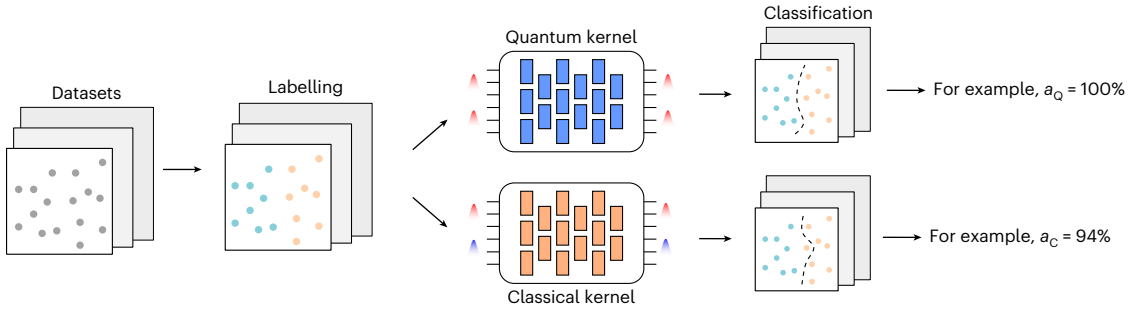

Fig. 2 | Classification tasks for photonic kernel methods. The datasets are randomly generated and consist of *d*-dimensional vectors, with entries between 0 and 1. Then, we randomly assign labels to each point as belonging to class +1 or −1 and we test the ability of our photonic kernels, displaying and not displaying quantum interference (indicated, respectively, as quantum kernel and classical kernel), to correctly classify the data. This is quantified by the accuracy of our models, which we indicate as a_Q and a_C.

To benchmark our enhanced performance, we compare classification accuracies between photonic kernels and state-of-the-art classical computational kernels. These include the standard gaussian kernels[17,18] and the recently introduced neural tangent kernels[26], which simulate gradient descent over infinitely-wide neural networks. Our results show that photonic kernels outperform classical methods and that the accuracies are further enhanced in kernels displaying quantum interference.

Photonic quantum kernel estimation

The core of a kernel method is constituted by a function that maps nonlinearly a set of N data points into a feature space. In its simplest version, these data points belong to two classes, and the final goal of the algorithm is to separate them accordingly, using only a linear model, that is, a hyperplane. This is feasible as the map, being nonlinear, will change the relative position between the data points, making the dataset (in general) easier to classify. For this reason, an SVM can effectively find a hyperplane that separates the two classes, with a high accuracy. From a more mathematical perspective, we can describe the map as a function that sends N input data points x_i, on which we wish to perform binary classification, from a space $\mathcal{X} \subseteq \mathbb{R}^d$ into a Hilbert space $\mathcal{H}$. Here, d is the dimension of each data point. This is done through a feature map $\Phi : \mathcal{X} \to \mathcal{H}$. Then, an SVM can be used to produce a prediction function $f_K : \mathcal{X} \to \mathbb{R}$ as $f_K(x) = \sum_i \alpha_i K(x, x_i)$, where these α_i coefficients are obtained by solving a linear optimization problem. The inputs of the optimization are the labels y, and the matrix obtained by computing the pairwise distances between data points is $K_{i,j} = K(x_i, x_j) = |\langle \Phi(x_i) | \Phi(x_j) \rangle|^2$, the so-called Gram matrix (see Supplementary Note 1 for further information).

In this work, we implement a quantum version of the kernel method, in which the aforementioned pairwise distances between data points, which belong to a class y taking values +1 or −1, are estimated by sampling from the output probability distribution that arises from the unitary evolution of a Fock input state. This process is depicted in Fig. 1a. Therefore, our feature map plugs the data that need to be classified into the free parameters defining a unitary evolution applied to a fixed Fock state of dimension m and whose sum of occupational numbers is n: $x \mapsto |\Phi(x)\rangle = U_x|\psi\rangle$. Here, $|\psi\rangle$ is the encoding state which is free to choose. Then, as shown in Fig. 1c, the pairwise inner products of the feature points are experimentally evaluated, as $|\langle \psi | U(x_i)^\dagger U(x_j) | \psi \rangle|^2$. Such unitaries can be effectively implemented by a programmable photonic circuit that consists of an array of Mach–Zehnder interferometers[27]. Hence, the dimension of the feature Hilbert space $\mathcal{H}$ will be $\binom{n+m-1}{n}$. At this point, the SVM finds the hyperplane separating the training data points through the aforementioned optimization process[28], and the binary classification of unknown points x is given by the following relation:

$$y = \mathrm{sgn}\left(\sum_{i=1}^{N} \alpha_i \, y_i K(x, x_i)\right) \tag{1}$$

where α_i are the coefficients optimized in the training process and y_i is the class of the ith point in the training. This model is defined implicitly, as the labels are assigned by weighted inner products of the encoded data points[29–34].

If the Fock state contains indistinguishable bosons, they will exhibit quantum interference, as shown in Fig. 1b. In this case, the output probability distribution is given by the permanents of submatrices of the matrix that represents the unitary evolution of the input[35]. More specifically, considering an input configuration s, the probability of detecting the output configuration t is given by $|\mathrm{Per} U_{s,t}|^2 / \Pi_i^m s_i! \Pi_i^m t_i!$. Here, $\mathrm{Per}(\cdot)$ denotes the permanent matrix operation, Π stands for the product notation, s_i and t_i are the occupational numbers at the ith mode and $U_{s,t}$ is the submatrix obtained by selecting the rows/columns corresponding to the occupied modes of the input/output Fock states. On the other hand, if the bosons are distinguishable, they will not exhibit quantum interference. In this case, the probability will amount to $\mathrm{Per}|U_{s,t}|^2 / \Pi_i^m s_i! \Pi_i^m t_i!$.

In the following, we will refer to a kernel implemented with indistinguishable bosons as a quantum kernel:

$$K_Q(x_i, x_j) = |\mathrm{Per} U_\psi(x_i, x_j)|^2 / N', \tag{2}$$

and with distinguishable ones as a classical kernel:

$$K_C(x_i, x_j) = \mathrm{Per}|U_\psi(x_i, x_j)|^2 / N'. \tag{3}$$

Here, $U_\psi(x_i, x_j)$ is the matrix defined by data points x_i and x_j and the selected input state ψ, and N' is the denominator $\Pi_i^m s_i! \Pi_i^m t_i!$ given by the occupational numbers of ψ. As long as there is at most one photon in each mode of ψ, $N' = 1$.

Classification task

At this point, we need to select a classification task that will benefit from the described model. Our strategy will therefore consist, first, of generating random data points and, second, selecting the proper labelling to boost the performance of the quantum kernel (K_Q) with respect to the classical kernel (K_C), as depicted in Fig. 2. Thus, we need a quantifier that estimates the difference in the performance of the two models and, to this end, we select the model complexity, which is defined as $s_K(y) = y^T K y$, where K is the kernel and y the labelling. This quantity is proportional to the upper bound on the classification error of a given model, and it amounts to the number of features required to make accurate predictions (Supplementary Note 2). Hence, the potential advantage given by K_Q, with respect to K_C, is based on finding

https://doi.org/10.1038/s41566-025-01682-5

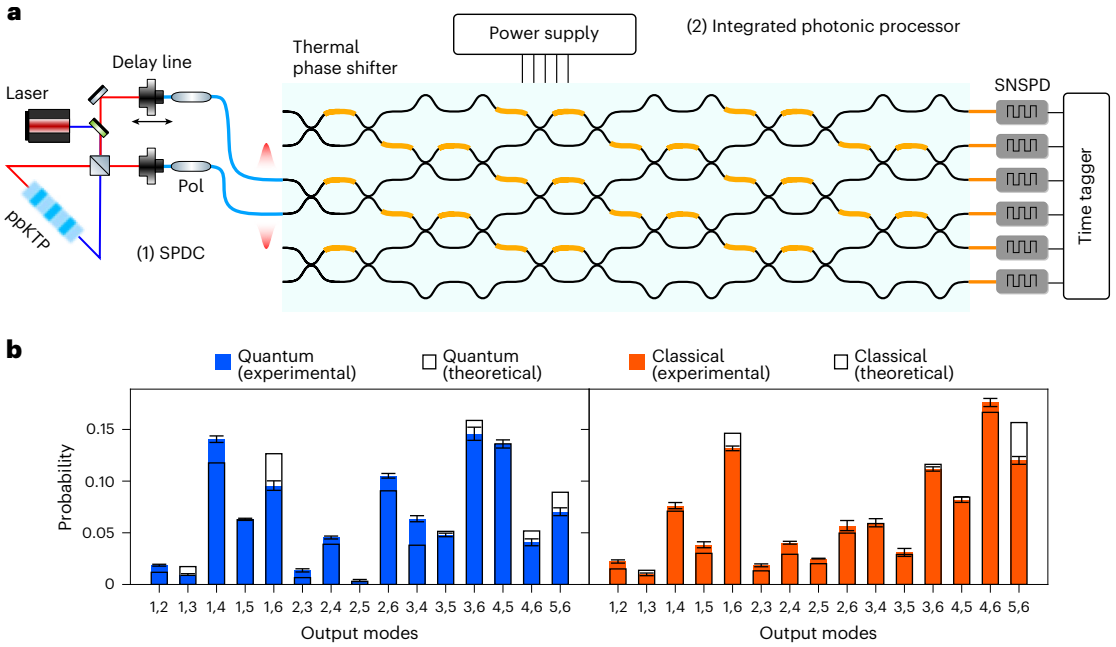

Fig. 3 | Implementation of photonic quantum kernel estimation. **a**, The experimental set-up consists of two parts: (1) the off-chip single-photon source and (2) the programmable integrated photonic processor. The frequency-degenerate photons are generated using a type-II SPDC source. Afterwards, the two photons are made indistinguishable in their polarization and arrival time. To enhance the quality of interference, two in-fiber polarizers (Pol) are placed at the two arms of the source. Then, we inject the generated photons in two modes of an integrated photonic processor with six input/output modes[25]. Detection is performed by SNSPDs. The degree of indistinguishability can then be tuned through a delay line, changing their relative temporal delay. **b**, Probability distribution of photon detection events. We show two instances of the experimental photon detection probability compared with the theoretical calculation. The quantum and classical kernel measurements are obtained, respectively, by injecting two indistinguishable and distinguishable photons into the third and fourth modes of the circuits, that is, $|0, 0, 1, 1, 0, 0\rangle$. The x axis shows all of the circuit channels that output two photons simultaneously. Thus, all 15 possible photon detection configurations are accessible. The error bars shown in the plot were evaluated through a Monte Carlo simulation, considering an underlying Poissonian distribution of the counts, and represent one standard deviation. The measurements were performed with an integration time of 5 s, at a detection rate of 10 kHz.

the largest possible separation between the two model complexities[21]. Therefore, our goal is to find the optimal labelling y, such that

$$y = \arg\min_{y\in\mathbb{R}^d}\left(\frac{s_{K_Q}(y)}{s_{K_C}(y)}\right). \tag{4}$$

Now, let us consider the following asymmetric distance g, denoted as geometric difference, between our two kernels, K_Q and K_C:

$$g_{CQ} = \sqrt{\left\|\sqrt{K_Q}(K_C)^{-1}\sqrt{K_Q}\right\|_\infty}, \tag{5}$$

where $\|\cdot\|_\infty$ denotes the spectral norm. It can be shown that the following inequality holds[21]:

$$s_{K_C}(y) \le g_{CQ}^2 s_{K_Q}(y). \tag{6}$$

Hence, if we maximize the geometric difference, we will have at least one task that saturates inequality (6), for which the model complexity will be higher for the classical kernel with respect to the quantum kernel. This indicates that the performance of K_Q will be competitive or better than K_C.

We can now use equation (5) to generate the classification task that, given two kernels K_Q and K_C and a set of data points $\{x_i\}$, produces the labels $\{y_i\}$ that maximise the difference in prediction error bound. This can be done through the following procedure: (1) evaluate the Gram matrices K_Q and K_C over a set of non-labelled data points $\{x_i\}$; (2) compute the positive definite matrix $M = \sqrt{K_Q}(K_C)^{-1}\sqrt{K_Q}$; (3) compute the eigenvalues and eigenvectors of M via spectral decomposition; (4) find the maximum eigenvalue g and its corresponding eigenvector v; and (5) assign the labels $y = \sqrt{K_Q}v$. From a practical point of view, we start with the two aforementioned kernels, K_C and K_Q, and then, by maximizing the geometric difference, we find the tasks for which the latter brings an enhanced accuracy of the classification. For more details regarding the algorithm to define the classification task, see Supplementary Note 4. Let us note that the implemented tasks constitute instances of problems that can be naturally implemented with high accuracy on our quantum platform. As such, they constitute a first stepping stone towards the identification of practical tasks for which quantum machine learning can enhance the performance of classical models.

Experiment

Our experimental set-up consists of two parts, a single-photon source generating the input states and a programmable integrated photonic processor, as depicted in Fig. 3a. First, to generate the input state, we use a type-II spontaneous parametric down-conversion (SPDC) source, which generates frequency-degenerate single-photon pairs at 1,546 nm in a periodically poled K-titanyl phosphate (ppKTP) crystal. The two photons are then made indistinguishable in their polarization and arrival time, respectively, via wave retarders and a delay line, which we also use to tune the degree of indistinguishability of the generated photons.

For the implementation of photonic kernels, which map our input data to a feature space, we require an apparatus that is able to perform arbitrary unitary transformations on a given input state. As mentioned above, our feature map sends each data point x_i onto the state resulting from the evolution $U(x_i)$ of a fixed input Fock state $|\psi\rangle$. Then, for

https://doi.org/10.1038/s41566-025-01682-5

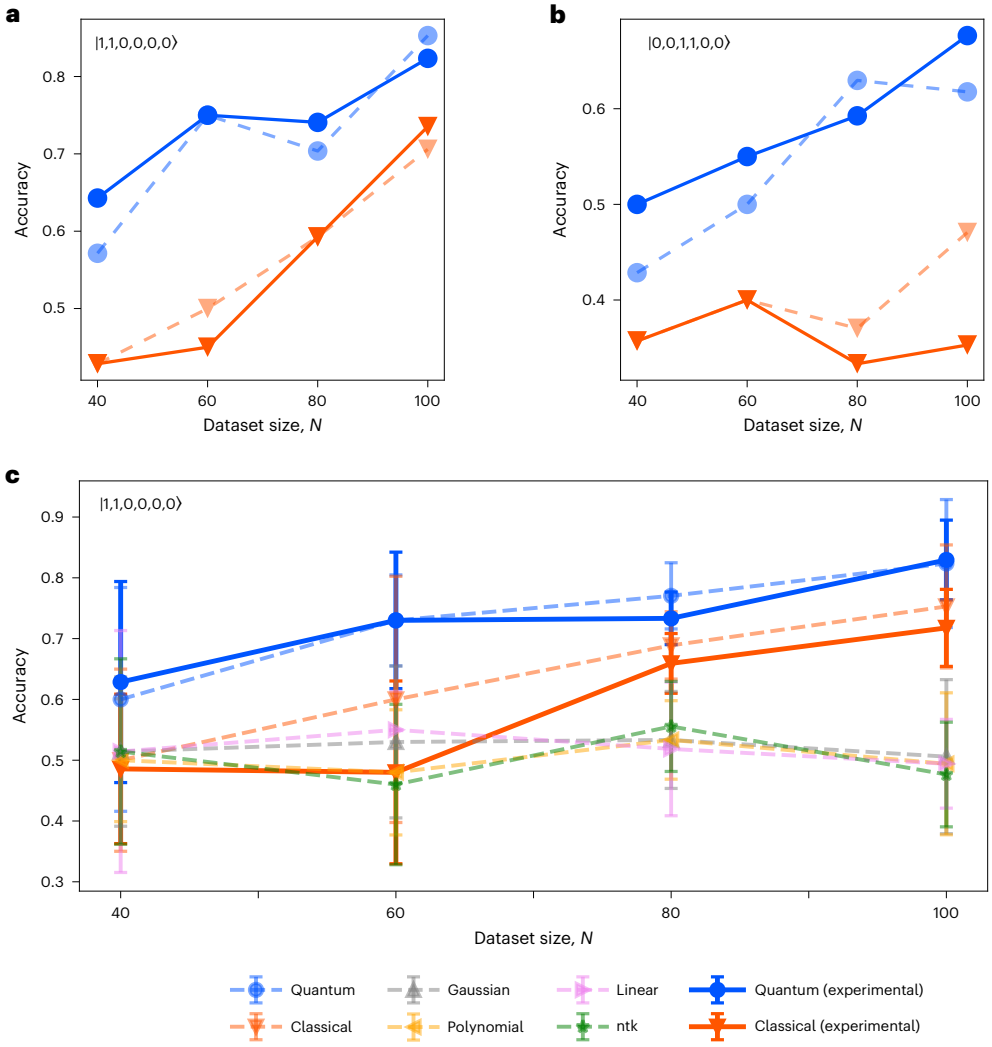

Fig. 4 | Experimental classification accuracies. a,**b**, We tested our method on datasets of different sizes (40, 60, 80 and 100) and for two different input states of $|1,1,0,0,0,0\rangle$ (**a**) and $|0,0,1,1,0,0\rangle$ (**b**). For each dataset, two-thirds of the data points were used for training the SVM, and the remaining one-third were used for testing. **c**, The average classification accuracies on five different sets for the quantum kernel (blue) and the classical kernel (orange), along with the following other computational kernels: gaussian (grey), ntk (green), polynomial (yellow) and linear (purple). The dashed lines indicate the results of numerical simulations, whereas the solid lines denote the experimental results. The error bars show the standard deviation of the classification accuracies on five random generated datasets classified by all of the kernels.

application of the SVM, which finds the best hyperplane separating the data, we need to evaluate the inner products between all of the points x_i, x_j in the feature space, which amounts to $\langle\psi|U(x_i)^\dagger U(x_j)|\psi\rangle$. This implies that, if we take $|\psi\rangle$ as a Fock state of n photons over m modes, the inner product $\langle\Phi(x_i)|\Phi(x_j)\rangle$ is given by projecting the evolved state $U(x_i)^\dagger U(x_j)\,|\psi\rangle$ onto $|\psi\rangle$.

To this end, we use an integrated photonic processor[25] on a borosilicate glass substrate, in which optical waveguides are inscribed through femtosecond laser writing[36]. The circuit features six input/output modes[27] (as depicted in Fig. 3a) where each interferometer is equipped with two thermal phase shifters[37], to provide tunable reflectivity and phase. Such an arrangement enables any unitary transformation to be performed on the input photon states. Given this property, our device is also referred to as a universal photonic processor. The design, fabrication and calibration of the integrated photonic circuit are described in ref. 25.

Specifically, the data were encoded in the values of the phase shifts as follows: $x_i = (x_i^1, x_i^2, \ldots, x_i^{30}) \rightarrow \theta_i = (2\pi x_i^1, 2\pi x_i^2, \ldots, 2\pi x_i^{30})$, where θ_i are the phase shifts introduced by the phase shifters of a universal interferometer. Let us note that this encoding has the remarkable advantage that no extra processing is required on the input data, as they are plugged directly into the optical circuit parameters. Furthermore, in principle, we would need a sequence of two of such circuits (as in the scheme of Fig. 1c), to first implement $U^\dagger(x_i)$ and then $U(x_j)$ on our inputs. However, in our implementation, we adopt only one universal circuit and directly implement the unitary corresponding to the product $U(x_i)^\dagger U(x_j)$. This reduces the experimental complexity and the circuit propagation losses.

At the output, detection is performed using superconducting nanowire single-photon detectors (SNSPDs), where we post-select the output events to those featuring two detector clicks (coincidence counts (CC)), that is, collision-free events (Supplementary Note 3). Thus, the elements of the Gram matrix of a given kernel can be estimated from the coincidence counting $K(x_i, x_j) = \mathrm{CC}^{ij}_{\psi} / \sum_{1\le l<m\le 6} \mathrm{CC}^{ij}_{lm}$. Here, CC^{ij}_{lm} is the number of registered coincidence counts between channels l and m, when the implemented unitary is $U^\dagger(x_j)U(x_i)$ and ψ indicates the occupied modes of input state $|\psi\rangle$. To test the role of quantum interference in the accuracy of the classification, we tune the indistinguishability of the two photons by changing their relative temporal delay. An instance of the probability distribution of the

same unitary is shown in Fig. 3b. The optimal classification task is chosen for each dataset according to the algorithm as explained in the previous section.

Results

We tested the performance of two photonic kernels in several different configurations. First, we considered two different inputs, $|1,1,0,0,0,0\rangle$ and $|0,0,1,1,0,0\rangle$. This amounts to either injecting the photons into the first two modes or the central two modes. Second, we were able to control the indistinguishability of the bosons, which allows us to implement, investigate a compare a quantum kernel with its related classical kernel. This is achieved by tuning the relative temporal delay of the propagating photons. Consequently, at the detection, it is possible to (partially) distinguish which photon—out of the two input photons—is detected, by checking their arrival time. Perfect distinguishability is obtained when the temporal delay is longer than the coherence length of the photons. During the whole measurement, the maximal achieved indistinguishability between the photons was 0.9720 ± 0.0044, estimated via on-chip Hong–Ou–Mandel interference[38].

For both input states, we fixed the encoding of each data point and varied the datasets to have four different sizes: 40, 60, 80 and 100. We used the set-up depicted in Fig. 3a to implement all pairwise products between the unitaries $U(x_i)^\dagger U(x_j)$. Hence, $|\langle\psi|U(x_i)^\dagger U(x_j)|\psi\rangle|^2$ is given by the probability of detecting the photons on the same modes from which they were injected. The rate of total post-selected coincidence counts amounts to 10 kHz, and the measured probability distribution was averaged over 5 s for each unitary configuration. Let us point out that restricting to collision-free events is unlikely to bring to a decay in the rate of detected samples when considering larger systems, as the bunching events become in general more rare for higher-dimensional Hilbert spaces, in the absence of particular symmetry patterns in the matrix describing the unitary evolution of the input[5,35].

For each size N, we performed $N(N-1)/2$ unitaries to compute the inner products. The distance between the unitaries can be experimentally realized and the target ones can be estimated as $\sum_i \sqrt{P_i^{\text{theo}} P_i^{\text{exp}}}$, where P_i^{exp} is the experimental detection frequency for the ith output configuration, whereas P_i^{theo} is the detection frequency for the ith output configuration estimated on the basis of theory[35]. For the quantum kernel and classical kernel, the mean fidelity of all datasets is 0.9816 ± 0.0148 and 0.9934 ± 0.0048, respectively. For each dataset, we use two-thirds of the data points for training the SVM, which yields the coefficients mentioned in equation (1). The remaining one-third as the test dataset can be used to predict the classification accuracy. The accuracy is defined as the percentage of correctly classified points out of the total size of the test set. Let us note that values lower than 0.5 indicate that the model was not able to learn the features of the training set and generalize to unknown data.

In Fig. 4a,b, we show the test accuracies obtained by injecting two input states for four different dataset sizes, where the quantum kernel performs better than the classical kernel in both experiments. In Fig. 4c, we report the average test accuracy obtained for five different datasets of the same size in addition to varying the dataset size from 40 to 100. Moreover, the results obtained with the quantum kernel (blue) and the classical kernel (orange) are compared with the following numerical kernels: the neural tangent kernel (ntk, green)[26], gaussian kernel (grey), polynomial kernel (yellow) and linear kernel (purple)[17,18]. Here, the neural tangent kernel adopts an infinite-width neural network to classify the data optimized by gradient descent—see Supplementary Note 5 for more details. The dashed lines indicate the results of numerical simulations, whereas the solid lines indicate experimental results. Although the task is built comparing only the performance of the kernels based on indistinguishable and distinguishable photons, the obtained accuracies of both are higher than the other classical kernels.

Discussion

In this work we present the experimental demonstration of a quantum kernel estimation, based on the unitary evolution of Fock states through an integrated photonic processor. We map data into a feature space through the evolution of a fixed two-photon input state over six modes. To achieve this, we adopt an quantum processor realized via femtosecond laser writing on a borosilicate glass substrate[25]. The sampled output distribution is then fed into an SVM, performing the classification. It is of note that, although our apparatus features only tunable phase shifters and beamsplitters, such encoding produces a sufficient nonlinearity to achieve a high classification accuracy of nonlinearly separable datasets. This constitutes a difference between our method and alternative platforms, where entangling gates are typically needed[21,39]. Furthermore, in our case, it is not necessary to increase the dimension of the feature Hilbert space to achieve a good accuracy. This circumvents the typical difficulty of quantum kernels in which all data points are encoded in orthogonal states, leading to ineffective classification[21,23,24,40]. Moreover, the fact that our model is effective for small dimensions is a crucial feature, as we require an approximation of the full probability distribution that is derived from the evolution of our input state. Hence, this study is relevant for medium-sized problems, because reaching high dimensions will imply the input/output combinations to grow exponentially, along with the number of required experimental shots to reach arbitrary accuracy.

The task we implement is designed by assigning binary labels to randomly generated data points, which we encode in the phase shifts of an optical circuit. This is done by exploiting the so-called geometric difference, which selects the task for which the presence of quantum interference yields a better classification accuracy compared with cases where no interference is displayed. Although the geometric difference compares the performance of a pair of kernels (in our case kernels implemented with indistinguishable/distinguishable bosons), both photonic kernels performed significantly better for the selected tasks than commonly used kernels, not only gaussian, polynomial and linear kernels but also the neural tangent kernel. This improvement in the performance that is seen for both photonic kernels is probably due to the fact that both the statistics have a similar dependence on the unitary evolution matrix, which is based on the permanent of submatrices (equations (2) and (3)), and are both complex enough to successfully tackle the selected tasks. In summary, our results indicate that a photonic kernel estimation can enhance the performance—even for medium-sized problems—and is reachable by current quantum technologies. Moreover, using distinguishable bosons to have a (smaller) performance enhancement represents an intriguing possibility as it can prove convenient to reduce the impact of photon losses on the experimental time required to collect significant statistics.

Let us highlight that, although the interference of two photons in a six-mode unitary matrix (or more in general for medium-sized problems) can be estimated using classical computers, this does not affect the features of our protocol. First, this is because the interest in investigating the potentialities and limitations of a new computational method, such as our proposed kernel, goes beyond the platform (quantum or classical) where it can be implemented, especially considering its particular suitability for given tasks. This is also the reason why we do not use our photonic platform to reproduce classical kernels, as done in refs. 30,41, but focus on investigating those given by the natural evolution of bosons through a quantum circuit. Second, the approximation of permanents through classical algorithms, for example, the algorithm of Gurvits[42], has a slightly worse performance when compared with direct sampling from an optical circuit, which naturally implements the studied kernel, for the case of inputs featuring product states[43]. In particular, the first scales as $O(n^2/e^2)$, where n is the number of photons, e is the required precision and O stands for the asymptotic time complexity of the algorithms, indicating how fast

their execution time grows with n and e, whereas the direct sampling as $O(1/e^2)$ (ref. 44). It is also noteworthy that the possibility of obtaining a quantum advantage on applications based on estimating expected values of unitary evolutions on linear optical circuits remains an open question[43,44]. In our case, an advantage could emerge by considering the extension of our quantum kernel $K_Q(x_i, x_j) = |\langle \psi | U(x_i, x_j) | \psi \rangle|^2$ for entangled input states ψ, although there is no rigorous proof of that yet. Third, this protocol sheds light on alternative computational models, exploiting optical computation. This may be of particular importance when considering difficulties related to energy consumption, as it has been proved that partially optical networks can reduce energy requirements with respect to electronic ones[45].

Despite being overshadowed by deep neural networks, kernels are still widely used in a large number of tasks due to their simplicity and ability to learn from small datasets[46,47]. Indeed, they have seen a recent revival in classical machine learning with the introduction of neural tangent kernels[26] and their use in the study of state-of-the-art neural network architectures such as transformers[48]. Another recent trend consists of merging neural networks and kernels, where notable examples are attention modules in natural language processing and Hopfield layers[49]. Also from the quantum computing point of view, kernels have recently gained a flurry of interest, both from the theoretical as well as from the experimental side[23,41,50,51]. For the latter, it has been proved that the evolution of (photonic) quantum states through a circuit is complex enough to tackle image recognition tasks, both for simple toy models[41,51] and satellite images[50]. Our method can therefore find a wide range of promising near-term applications in tasks such as information retrieval, natural language processing and medical image classification[52–55], where kernels have been proposed as a keystone[56].

In particular, an interesting approach involves pre-training a small (single-qubit) system to find the optimal encoding of classical data, which is then classified through a kernel method featuring a larger system, as proposed in ref. 51. This could be particularly beneficial on our photonic platform as the pre-training could be performed through a classical light input over two spatial modes. This would imply outsourcing the computationally hardest part of the protocol to a fully optical platform, possibly bringing a significant reduction in energy consumption. Our experimental results also open the door to hybrid methods in which photonic processors are used to enhance the performance of standard machine learning methods. They also bring forward investigations on the nonlinearities that can be achieved through photonic platforms[57,58], in particular, for neuromorphic computation models, such as reservoir computing[59,60]. In addition, we envisage the combination of this kind of nonlinearity with that brought by the implementation of feedback loops, as in the case of a quantum memristor[58] and the exploitation of quantum interference in the implementation of feature maps.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41566-025-01682-5.

References

1. Wehner, S., Elkouss, D. & Hanson, R. Quantum internet: a vision for the road ahead. *Science* **362**, eaam9288 (2018).
2. Georgescu, I. M., Ashhab, S. & Nori, F. Quantum simulation. *Rev. Mod. Phys.* **86**, 153–185 (2014).
3. Shor, P. W. Algorithms for quantum computation: discrete logarithms and factoring. In *Proc. 35th Annual Symposium on Foundations of Computer Science* (ed. Johnson, D. S.) 124–134 (IEEE, 1994).
4. Grover, L. K. A fast quantum mechanical algorithm for database search. In *STOC '96: Proc. 28th Annual ACM Symposium on Theory of Computing* (ed. Miller, G. L.) 212–219 (Association for Computing Machinery, 1996).
5. Aaronson, S. & Arkhipov, A. The computational complexity of linear optics. In *STOC '11: Proc. 43rd Annual ACM Symposium on Theory of Computing* (eds Fortnow, L. & Vadhan, S. P.) 333–342 (Association for Computing Machinery, 2011).
6. Zhong, H.-S. et al. Quantum computational advantage using photons. *Science* **370**, 1460–1463 (2020).
7. Madsen, L. S. et al. Quantum computational advantage with a programmable photonic processor. *Nature* **606**, 75–81 (2022).
8. Arute, F. et al. Quantum supremacy using a programmable superconducting processor. *Nature* **574**, 505–510 (2019).
9. Preskill, J. Quantum computing in the NISQ era and beyond. *Quantum* **2**, 79 (2018).
10. Dunjko, V. & Briegel, H. J. Machine learning & artificial intelligence in the quantum domain: a review of recent progress. *Rep. Prog. Phys.* **81**, 074001 (2018).
11. Neven, H., Denchev, V. S., Rose, G. & Macready, W. G. Qboost: large scale classifier training withadiabatic quantum optimization. *J. Mach. Learn. Res.* **25**, 333–348 (2012).
12. Rebentrost, P., Mohseni, M. & Lloyd, S. Quantum support vector machine for big data classification. *Phys. Rev. Lett.* **113**, 130503 (2014).
13. Leifer, M. S. & Poulin, D. Quantum graphical models and belief propagation. *Ann. Phys.* **323**, 1899–1946 (2008).
14. Saggio, V. et al. Experimental quantum speed-up in reinforcement learning agents. *Nature* **591**, 229–233 (2021).
15. Boixo, S. et al. Characterizing quantum supremacy in near-term devices. *Nat. Phys.* **14**, 595–600 (2018).
16. Gan, B. Y., Leykam, D. & Angelakis, D. G. Fock state-enhanced expressivity of quantum machine learning models. *EPJ Quantum Technol.* **9**, 16 (2022).
17. Shawe-Taylor, J. & Cristianini, N. *Kernel Methods for Pattern Analysis* (Cambridge Univ. Press, 2004)
18. Hofmann, T., Schölkopf, B. & Smola, A. J. Kernel methods in machine learning. *Ann. Stat.* **36**, 1171–1220 (2008).
19. Cortes, C. & Vapnik, V. Support-vector networks. *Mach. Learn.* **20**, 273–297 (1995).
20. Liu, Y., Arunachalam, S. & Temme, K. A rigorous and robust quantum speed-up in supervised machine learning. *Nat. Phys.* **17**, 1013–1017 (2021).
21. Huang, H.-Y. et al. Power of data in quantum machine learning. *Nat. Commun.* **12**, 2631 (2021).
22. Talagrand, M. Concentration of measure and isoperimetric inequalities in product spaces. *Publ. Math. Inst. Hautes Études Sci.* **81**, 73–205 (1995).
23. Thanasilp, S., Wang, S., Cerezo, M. & Holmes, Z. Exponential concentration in quantum kernel methods. *Nat. Commun.* **15**, 5200 (2024).
24. Kübler, J., Buchholz, S. & Schölkopf, B. The inductive bias of quantum kernels. In *Advances in Neural Information Processing Systems* (eds Ranzato, M. et al.) 12661–12673 (Curran Associates, 2021).
25. Pentangelo, C. et al. High-fidelity and polarization-insensitive universal photonic processors fabricated by femtosecond laser writing. *Nanophotonics* **13**, 2259–2270 (2024).
26. Jacot, A., Gabriel, F. & Hongler, C. Neural tangent kernel: convergence and generalization in neural networks. In *Advances in Neural Information Processing Systems 31: 32nd Conference on Neural Information Processing Systems (NeurIPS 2018)* (eds Bengio, S. et al.) 8571–8580 (Neural Information Processing Systems Foundation, 2018).

27. Clements, W. R., Humphreys, P. C., Metcalf, B. J., Kolthammer, W. S. & Walsmley, I. A. Optimal design for universal multiport interferometers. *Optica* **3**, 1460–1465 (2016).
28. Boser, B. E., Guyon, I. M. & Vapnik, V. N. A training algorithm for optimal margin classifiers. In *COLT '92: Proc. 5th Annual Workshop on Computational Learning Theory* (ed. Haussler, D.) 144–152 (Association for Computing Machinery, 1992).
29. Lloyd, S., Schuld, M., Ijaz, A., Izaac, J. & Killoran, N. Quantum embeddings for machine learning. Preprint at https://arxiv.org/abs/2001.03622 (2020)
30. Bartkiewicz, K. et al. Experimental kernel-based quantum machine learning in finite feature space. *Sci. Rep.* **10**, 12356 (2020).
31. Huang, H.-Y., Kueng, R. & Preskill, J. Information-theoretic bounds on quantum advantage in machine learning. *Phys. Rev. Lett.* **126**, 190505 (2021).
32. Kusumoto, T., Mitarai, K., Fujii, K., Kitagawa, M. & Negoro, M. Experimental quantum kernel trick with nuclear spins in a solid. *NPJ Quantum Inf.* **7**, 94 (2021).
33. Schölkopf, B. & Smola, A. J. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond* (MIT Press, 2002)
34. Jerbi, S. et al. Quantum machine learning beyond kernel methods. *Nat. Commun.* **14**, 517 (2023).
35. Tichy, M. C. Interference of identical particles from entanglement to boson-sampling. *J. Phys. B* **47**, 103001 (2014).
36. Corrielli, G., Crespi, A. & Osellame, R. Femtosecond laser micromachining for integrated quantum photonics. *Nanophotonics* **10**, 3789–3812 (2021).
37. Ceccarelli, F. et al. Low power reconfigurability and reduced crosstalk in integrated photonic circuits fabricated by femtosecond laser micromachining. *Laser Photon. Rev.* **14**, 2000024 (2020).
38. Hong, C.-K., Ou, Z.-Y. & Mandel, L. Measurement of subpicosecond time intervals between two photons by interference. *Phys. Rev. Lett.* **59**, 2044–2046 (1987).
39. Havlíček, V. et al. Supervised learning with quantum-enhanced feature spaces. *Nature* **567**, 209–212 (2019).
40. Schuld, M. & Killoran, N. Quantum machine learning in feature Hilbert spaces. *Phys. Rev. Lett.* **122**, 040504 (2019).
41. Hoch, F. et al. Quantum machine learning with adaptive boson sampling via post-selection. *Nat. Commun.* **16**, 902 (2025).
42. Gurvits, L. On the complexity of mixed discriminants and related problems. In *Mathematical Foundations of Computer Science 2005: 30th International Symposium, MFCS 2005* (eds Jędrzejowicz, J. & Szepietowski, A.) 447–458 (Springer, 2005)
43. Lim, Y. & Oh, C. Efficient classical algorithms for linear optical circuits. Preprint at https://arxiv.org/abs/2502.12882 (2025)
44. Aaronson, S. & Hance, T. Generalizing and derandomizing Gurvits's approximation algorithm for the permanent. *Quantum Inf. Comput.* **14**, 541–559 (2014).
45. Hamerly, R., Bernstein, L., Sludds, A., Soljačić, M. & Englund, D. Large-scale optical neural networks based on photoelectric multiplication. *Phys. Rev. X* **9**, 021032 (2019).
46. Lee, J. et al. Finite versus infinite neural networks: an empirical study. In *Advances in Neural Information Processing Systems 33: 34th Conference on Neural Information Processing Systems (NeurIPS 2020)* (eds Larochelle, H. et al.) 15156–15172 (Neural Information Processing Systems Foundation, 2020).
47. Radhakrishnan, A., Ruiz Luyten, M., Prasad, N. & Uhler, C. Transfer learning with kernel methods. *Nat. Commun.* **14**, 5570 (2023).
48. Tsai, Y.-H. H., Bai, S., Yamada, M., Morency, L.-P. & Salakhutdinov, R. Transformer dissection: an unified understanding for transformer's attention via the lens of kernel. In *Proc. 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)* (eds Inui, K. et al.) 4343–4352 (Association for Computational Linguistics, 2019); https://doi.org/10.18653/v1/D19-1443
49. Ramsauer, H. et al. *Proc. 9th International Conference on Learning Representations* (ICLR, 2021).
50. Rodriguez-Grasa, P., Farzan-Rodríguez, R., Novelli, G., Ban, Y. & Sanz, M. Satellite image classification with neural quantum kernels. *Mach. Learn. Sci. Technol.* **6**, 015043 (2025).
51. Rodriguez-Grasa, P., Ban, Y. & Sanz, M. Neural quantum kernels: training quantum kernels with quantum neural networks. Preprint at https://arxiv.org/abs/2401.04642 (2024).
52. Wang, X., Du, Y., Luo, Y. & Tao, D. Towards understanding the power of quantum kernels in the NISQ era. *Quantum* **5**, 531 (2021).
53. Yu, C.-H., Gao, F., Wang, Q.-L. & Wen, Q.-Y. Quantum algorithm for association rules mining. *Phys. Rev. A* **94**, 042311 (2016).
54. Lorenz, R., Pearson, A., Meichanetzidis, K., Kartsaklis, D. & Coecke, B. QNLP in practice: running compositional models of meaning on a quantum computer. *J. Artif. Intell. Res.* **76**, 1305–1342 (2023).
55. Landman, J. et al. Quantum methods for neural networks and application to medical image classification. *Quantum* **6**, 881 (2022).
56. Schuld, M. & Petruccione, F. in *Machine Learning with Quantum Computers* (eds Schuld, M. & Petruccione, F.) 217–245 (Springer, 2021); https://doi.org/10.1007/978-3-030-83098-4_6
57. Denis, Z., Favero, I. & Ciuti, C. Photonic kernel machine learning for ultrafast spectral analysis. *Phys. Rev. Appl.* **17**, 034077 (2022).
58. Spagnolo, M. et al. Experimental photonic quantum memristor. *Nat. Photon.* **16**, 318–323 (2022).
59. Govia, L. C. G., Ribeill, G. J., Rowlands, G. E. & Ohki, T. A. Nonlinear input transformations are ubiquitous in quantum reservoir computing. *Neuromorphic Comput. Eng.* **2**, 014008 (2022).
60. Innocenti, L. et al. Potential and limitations of quantum extreme learning machines. *Commun. Phys.* **6**, 118 (2023).

Methods

The two-photon input states are generated by a type-II SPDC source, which generates frequency-degenerate single-photon pairs at 1,546 nm via a 3-cm-long ppKTP crystal. Afterwards, the two photons are made indistinguishable in their polarization (which is rotated through paddles) and in arrival time (through a delay line, which we also use to tune the degree of indistinguishability of the generated photons).

We then inject these photons into two selected modes of an integrated photonic processor with six input/output modes. This circuit features 27 thermal phase shifters and its architecture is universal, enabling us to implement arbitrary unitary evolution on any input Fock state. Compared with the original Clements architecture, the first three external phase shifters are not implemented as they will not affect the probability distribution, in the case of inputs not featuring coherent superpositions of Fock states. To apply accurate phases independently, each channel is supplied by a current source to avoid electrical crosstalk (4 × Qontrol Q8iv driver modules, 16-bit DAC). At the end, detection is performed by SNSPDs housed in a 1K cryostat. We post-select the detected events to the cases in which two detectors click simultaneously in a temporal window of 1 ns. A time tagger with a 15.63 ps resolution is used to process the real-time coincidence counting for all 15 post-selection patterns. The total coincidence counting is about 10 kHz, which varies due to the pump laser and the detection efficiency. For each unitary configuration, we integrate for 5 s to estimate the probability distribution over 15 coincidence patterns.

Data availability

The data that support the findings of this study and detailed explanations of the datasets are available from the project page via GitHub at https://github.com/dapingQ/PhoQuKs.

Code availability

The code scripts used in this study are available from the project page via GitHub at https://github.com/dapingQ/PhoQuKs.

Acknowledgements

We would like to thank A. Pearson for her active involvement in the first stages of this project. This research was funded in whole or in part by the Austrian Science Fund (FWF): [10.55776/ESP205] (PREQUrSOR) granted to I.A., and [10.55776/F71] (BeyondC), [10.55776/FG5] (Research Group 5) and [10.55776/I6002] (PhoMemtor) granted to P.W. For open access purposes, the author has applied a CC BY public copyright license to any author accepted manuscript version arising from this submission. This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 899368 (EPIQUS), granted to P.W.; the Marie Skłodowska-Curie grant agreement no. 956071 (AppQInfo), granted to P.W.; and grant agreement no. 101017733 (QuantERA II Programme, project PhoMemtor), granted to P.W. and R.O. Views and opinions expressed are, however, those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them. The financial support by the Austrian Federal Ministry of Labour and Economy, the National Foundation for Research, Technology and Development and the Christian Doppler Research Association is gratefully acknowledged. G.d.F would like to thank K. Meichanetzidis and T. Salvatori for insightful discussions on the project. The integrated photonic processor was partially fabricated at PoliFAB, the micro- and nanofabrication facility of Politecnico di Milano (https://www.polifab.polimi.it/). C.P., F.C. and R.O. wish to thank the PoliFAB staff for their valuable technical support. R.O. acknowledges financial support by the ICSC National Research Centre for High Performance Computing, Big Data and Quantum Computing, funded by European Union—NextGenerationEU.

Author contributions

Z.Y. and I.A. designed and conducted the experiments, and G.d.F., D.B. and A.T. developed the theory and algorithm. C.P., S.P., A.C. and F.C. conducted the design, fabrication and calibration of the integrated photonic processor. Z.Y., I.A. and G.d.F. wrote the first draft of the paper. R.O., B.C. and P.W. supervised the whole project. All authors discussed the results and reviewed the paper.

Funding

Open access funding provided by University of Vienna.

Competing interests

F.C. and R.O. are cofounders of the company Ephos. The other authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41566-025-01682-5.

Correspondence and requests for materials should be addressed to Zhenghao Yin, Iris Agresti or Philip Walther.

Peer review information *Nature Photonics* thanks Nicholas Harris and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

References

[1] A. M. Turing. 'On Computable Numbers, with an Application to the Entscheidungsproblem'. In: *Proceedings of the London Mathematical Society* s2-42.1 [1937], pp. 230–265. ISSN: 1460-244X. DOI: 10.1112/plms/s2-42.1.230. [Visited on 05/05/2025].

[2] A. M. Turing. 'I.—COMPUTING MACHINERY AND INTELLIGENCE'. In: *Mind* LIX.236 [1st Oct. 1950], pp. 433–460. ISSN: 1460-2113, 0026-4423. DOI: 10.1093/mind/LIX.236.433. [Visited on 05/05/2025].

[3] Intel. *Intel 18A | See Our Biggest Process Innovation*. Intel. 2025. [Visited on 05/05/2025].

[4] TSMC. *A16 Technology - Taiwan Semiconductor Manufacturing Company Limited*. A16 Technology - Taiwan Semiconductor Manufacturing Company Limited. 2025. [Visited on 05/05/2025].

[5] Mark Horowitz. '1.1 Computing's Energy Problem (and What We Can Do about It)'. In: *2014 IEEE International Solid-State Circuits Conference Digest of Technical Papers (ISSCC)*. 2014 IEEE International Solid-State Circuits Conference Digest of Technical Papers (ISSCC). Feb. 2014, pp. 10–14. DOI: 10.1109/ISSCC.2014.6757323. [Visited on 05/05/2025].

[6] Paul Benioff. 'The Computer as a Physical System: A Microscopic Quantum Mechanical Hamiltonian Model of Computers as Represented by Turing Machines'. In: *Journal of Statistical Physics* 22.5 [5 1st May 1980], pp. 563–591. ISSN: 1572-9613. DOI: 10.1007/BF01011339. [Visited on 21/05/2025].

[7] Peter W. Shor. 'Polynomial-Time Algorithms for Prime Factorization and Discrete Logarithms on a Quantum Computer'. In: *SIAM Journal on Computing* 26.5 [Oct. 1997], pp. 1484–1509. ISSN: 0097-5397. DOI: 10.1137/S0097539795293172. [Visited on 13/05/2025].

[8] Lov K. Grover. 'A Fast Quantum Mechanical Algorithm for Database Search'. In: *Proceedings of the Twenty-Eighth Annual ACM Symposium on Theory of Computing*. STOC '96. New York, NY, USA: Association for Computing Machinery, 1st July 1996, pp. 212–219. ISBN: 978-0-89791-785-8. DOI: 10.1145/237814.237866. [Visited on 13/05/2025].

[9] Pieter Kok et al. 'Linear Optical Quantum Computing with Photonic Qubits'. In: *Reviews of Modern Physics* 79.1 [2007], pp. 135–174. ISSN: 00346861. DOI: 10.1103/RevModPhys.79.135. arXiv: quant-ph/0512071.

[10] Jeremy L. O'Brien, Akira Furusawa and Jelena Vukovi. 'Photonic Quantum Technologies'. In: *Nature Photonics* 3.12 [2009], pp. 687–695. ISSN: 17494885. DOI: 10.1038/nphoton.2009.229. arXiv: 1003.3928.

[11] Wentao Chen et al. 'Scalable and Programmable Phononic Network with Trapped Ions'. In: *Nature Physics* 19.6 [June 2023], pp. 877–883. ISSN: 1745-2481. DOI: 10.1038/s41567-023-01952-5. [Visited on 15/04/2024].

[12] Karan K. Mehta et al. 'Integrated Optical Multi-Ion Quantum Logic'. In: *Nature* 586.7830 [2020], pp. 533–537. ISSN: 14764687. DOI: 10.1038/s41586-020-2823-6. PMID: 33087915.

[13] J. M. Pino et al. 'Demonstration of the Trapped-Ion Quantum CCD Computer Architecture'. In: *Nature* 592.7853 [8th Apr. 2021], pp. 209–213. ISSN: 0028-0836. DOI: 10.1038/s41586-021-03318-4. arXiv: 2003.01293.

[14] Steve M. Girvin. 'Superconducting Qubits and Circuits : Artificial Atoms Coupled to Microwave Photons'. In: *Lecture notes in Les Houches School of PHysics* 35.10 [1978], pp. 1265–1267. ISSN: 15383636. DOI: 10.1001/archpsyc.1978.01770340115013. PMID: 697541.

[15] R. J. Schoelkopf and S. M. Girvin. 'Wiring up Quantum Systems'. In: *Nature* 451.7179 [2008], pp. 664–669. ISSN: 14764687. DOI: 10.1038/451664a.

[16] Morten Kjaergaard et al. 'Superconducting Qubits: Current State of Play'. In: *Annual Review of Condensed Matter Physics* 11 [Volume 11, 2020 10th Mar. 2020], pp. 369–395. ISSN: 1947-5454, 1947-5462. DOI: 10.1146/annurev-conmatphys-031119-050605. [Visited on 12/10/2025].

[17] P. Krantz et al. 'A Quantum Engineer's Guide to Superconducting Qubits'. In: *Applied Physics Reviews* 6.2 [17th June 2019], p. 021318. ISSN: 1931-9401. DOI: 10.1063/1.5089550. [Visited on 12/10/2025].

[18] Guido Burkard. 'Semiconductor Spin Qubits'. In: *Reviews of Modern Physics* 95.2 [2023]. DOI: 10.1103/RevModPhys.95.025003.

[19] Anasua Chatterjee et al. 'Semiconductor Qubits in Practice'. In: *Nature Reviews Physics* 3.3 [Mar. 2021], pp. 157–177. ISSN: 2522-5820. DOI: 10.1038/s42254-021-00283-9. [Visited on 12/10/2025].

[20] R. Maurand et al. 'A CMOS Silicon Spin Qubit'. In: *Nature Communications* 7.1 [24th Nov. 2016], p. 13575. ISSN: 2041-1723. DOI: 10.1038/ncomms13575. [Visited on 12/10/2025].

[21] Jarryd J. Pla et al. 'A Single-Atom Electron Spin Qubit in Silicon'. In: *Nature* 489.7417 [Sept. 2012], pp. 541–545. ISSN: 1476-4687. DOI: 10.1038/nature11449. [Visited on 12/10/2025].

[22] John F. Clauser and Michael A. Horne. 'Experimental Consequences of Objective Local Theories'. In: *Physical Review D* 10.2 [1974], pp. 526–535. ISSN: 05562821. DOI: 10.1103/PhysRevD.10.526.

[23] Dik Bouwmeester et al. 'Experimental Quantum Teleportation'. In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 356.1743 [1998], pp. 1733–1737. ISSN: 1364503X. DOI: 10.1098/rsta.1998.0245.

[24] Scott Aaronson and Alex Arkhipov. 'The Computational Complexity of Linear Optics'. In: *Proceedings of the 43rd annual ACM symposium on Theory of computing - STOC '11* 9.4 [14th Nov. 2010], p. 333. ISSN: 07378017. DOI: 10.1145/1993636.1993682. arXiv: 1011.3245.

[25] Matthew A. Broome et al. 'Photonic Boson Sampling in a Tunable Circuit'. In: *Science* 339.6121 [15th Feb. 2013], pp. 794–798. DOI: 10.1126/science.1231440. [Visited on 12/04/2023].

[26] Justin B. Spring et al. 'Boson Sampling on a Photonic Chip'. In: *Science* 339.6121 [15th Feb. 2013], pp. 798–801. DOI: 10.1126/science.1231692. [Visited on 03/07/2023].

[27] Max Tillmann et al. 'Experimental Boson Sampling'. In: *Nature Photonics* 7.7 [2013], pp. 540–544. ISSN: 17494885. DOI: 10.1038/nphoton.2013.102. arXiv: 1212.2240.

[28] Han-sen Zhong et al. 'Experimental Gaussian Boson Sampling'. In: *Science Bulletin* 64.8 [Apr. 2019], pp. 511–515. ISSN: 20959273. DOI: 10.1016/j.scib.2019.04.007.

[29] Yu-Hao Deng et al. 'Gaussian Boson Sampling with Pseudo-Photon-Number-Resolving Detectors and Quantum Computational Advantage'. In: *Physical Review Letters* 131.15 [10th Oct. 2023], p. 150601. DOI: 10.1103/PhysRevLett.131.150601. [Visited on 07/07/2025].

[30] Han Sen Zhong et al. 'Phase-Programmable Gaussian Boson Sampling Using Stimulated Squeezed Light'. In: *Physical Review Letters* 127.18 [2021], p. 180502. ISSN: 10797114. DOI: 10.1103/PhysRevLett.127.180502. PMID: 34767431.

[31] Jianwei Wang et al. 'Integrated Photonic Quantum Technologies'. In: *Nature Photonics* 14 [May 2019]. ISSN: 17494893. DOI: 10.1038/s41566-019-0532-1.

[32] Damien Bonneau, Joshua W. Silverstone and Mark G. Thompson. 'Silicon Quantum Photonics'. In: *Topics in Applied Physics* 122 [2016], pp. 41–82. ISSN: 14370859. DOI: 10.1007/978-3-642-10503-6_2. arXiv: 1707.02334.

[33] Ali W. Elshaari et al. 'Hybrid Integrated Quantum Photonic Circuits'. In: *Nature Photonics* 14.5 [13th May 2020], pp. 285–298. ISSN: 1749-4885. DOI: 10.1038/s41566-020-0609-x.

[34] Jehyung Kim et al. 'Hybrid Integration Methods for On-Chip Quantum Photonics'. In: *Optica* 7.4 [20th Apr. 2020], p. 291. ISSN: 2334-2536. DOI: 10.1364/OPTICA.384118. arXiv: 1911.12756.

[35] Jacques Carolan et al. 'Universal Linear Optics'. In: *Science* 349.6249 [14th Aug. 2015], pp. 711–716. ISSN: 0036-8075. DOI: 10.1126/science.aab3642. arXiv: 1505.01182.

[36] R. Santagati et al. 'Silicon Photonic Processor of Two-Qubit Entangling Quantum Logic'. In: *Journal of Optics* 19.11 [1st Nov. 2017], p. 114006. ISSN: 2040-8978. DOI: 10.1088/2040-8986/aa8d56.

[37] Jianwei Wang et al. 'Multidimensional Quantum Entanglement with Large-Scale Integrated Optics'. In: *Science* 360.6386 [2018], pp. 285–291. ISSN: 10959203. DOI: 10.1126/science.aar7053. PMID: 29519918.

[38] Jeremy C. Adcock et al. 'Programmable Four-Photon Graph States on a Silicon Chip'. In: *Nature Communications* 10.1 [1 6th Aug. 2019], p. 3528. ISSN: 2041-1723. DOI: 10.1038/s41467-019-11489-y. [Visited on 15/06/2023].

[39] Jueming Bao et al. 'Very-Large-Scale Integrated Quantum Graph Photonics'. In: *Nature Photonics* [6th Apr. 2023], pp. 1–9. ISSN: 1749-4893. DOI: 10.1038/s41566-023-01187-z. [Visited on 19/06/2023].

[40] Benjamin J. Metcalf et al. 'Quantum Teleportation on a Photonic Chip'. In: *Nature Photonics* 8.10 [2014], pp. 770–774. ISSN: 17494893. DOI: 10.1038/nphoton.2014.217. arXiv: 1409.4267.

[41] Taofiq K. Paraïso et al. 'A Photonic Integrated Quantum Secure Communication System'. In: *Nature Photonics* 15.11 [2021], pp. 850–856. ISSN: 17494893. DOI: 10.1038/s41566-021-00873-0.

[42] P. Sibson et al. 'Chip-Based Quantum Key Distribution'. In: *Nature Communications* 8 [May 2016 2017], p. 13984. ISSN: 20411723. DOI: 10.1038/ncomms13984. arXiv: 1509.00768.

[43] Yuan Li et al. 'Experimental Quantum Homomorphic Encryption Using a Quantum Photonic Chip'. In: *Physical Review Letters* 132.20 [14th May 2024], p. 200801. ISSN: 0031-9007, 1079-7114. DOI: 10.1103/PhysRevLett.132.200801. [Visited on 14/06/2024].

[44] G. Zhang et al. 'An Integrated Silicon Photonic Chip Platform for Continuous-Variable Quantum Key Distribution'. In: *Nature Photonics* 13.12 [12th Dec. 2019], pp. 839–842. ISSN: 17494893. DOI: 10.1038/s41566-019-0504-5.

[45] Max Tillmann et al. 'Generalized Multiphoton Quantum Interference'. In: *Physical Review X* 5.4 [27th Oct. 2015], p. 041015. ISSN: 2160-3308. DOI: 10.1103/PhysRevX.5.041015. [Visited on 04/04/2024].

[46] John Preskill. 'Quantum Computing in the NISQ Era and Beyond'. In: *Quantum* 2 [July 2018], pp. 1–20. ISSN: 2521327X. DOI: 10.22331/q-2018-08-06-79. arXiv: 1801.00862.

[47] Jacob Biamonte et al. 'Quantum Machine Learning'. In: *Nature* 549.7671 [2017], pp. 195–202. ISSN: 14764687. DOI: 10.1038/nature23474. PMID: 28905917.

[48] Kishor Bharti et al. 'Noisy Intermediate-Scale Quantum Algorithms'. In: *Reviews of Modern Physics* 94.1 [15th Feb. 2022], p. 015004. DOI: 10.1103/RevModPhys.94.015004. [Visited on 06/11/2023].

[49] Vedran Dunjko and Hans J Briegel. 'Machine Learning & Artificial Intelligence in the Quantum Domain: A Review of Recent Progress'. In: *Reports on Progress in Physics* 81.7 [June 2018], p. 074001. ISSN: 0034-4885. DOI: 10.1088/1361-6633/aab406. [Visited on 08/10/2025].

[50] Yichen Shen et al. 'Deep Learning with Coherent Nanophotonic Circuits'. In: *Nature Photonics* 11.7 [12th July 2017], pp. 441–446. ISSN: 1749-4885. DOI: 10.1038/nphoton.2017.93. PMID: 20181947.

[51] Nicholas C. Harris et al. 'Linear Programmable Nanophotonic Processors'. In: *Optica* 5.12 [2018], p. 1623. ISSN: 2334-2536. DOI: 10.1364/optica.5.001623.

[52] Sufi R. Ahmed et al. 'Universal Photonic Artificial Intelligence Acceleration'. In: *Nature* 640.8058 [Apr. 2025], pp. 368–374. ISSN: 1476-4687. DOI: 10.1038/s41586-025-08854-x. [Visited on 18/04/2025].

[53] Shiyue Hua et al. 'An Integrated Large-Scale Photonic Accelerator with Ultralow Latency'. In: *Nature* 640.8058 [Apr. 2025], pp. 361–367. ISSN: 1476-4687. DOI: 10.1038/s41586-025-08786-6. [Visited on 18/04/2025].

[54] Sunil Pai et al. 'Experimentally Realized in Situ Backpropagation for Deep Learning in Photonic Neural Networks'. In: *Science* 380.6643 [28th Apr. 2023], pp. 398–404. DOI: 10.1126/science.ade8450. [Visited on 30/05/2023].

[55] Michael Reck et al. 'Experimental Realization of Any Discrete Unitary Operator'. In: *Physical Review Letters* 73.1 [4th July 1994], pp. 58–61. DOI: 10.1103/PhysRevLett.73.58. [Visited on 10/06/2023].

[56] William R. Clements et al. 'Optimal Design for Universal Multiport Interferometers'. In: *Optica* 3.12 [2016], p. 1460. ISSN: 2334-2536. DOI: 10.1364/optica.3.001460. arXiv: 1603.08788.

[57] V Saggio et al. 'Experimental Quantum Speed-up in Reinforcement Learning Agents'. In: *Nature* 591 [March 11th Mar. 2021], pp. 1–12. ISSN: 1476-4687. DOI: 10.1038/s41586-021-03242-7.

[58] Michele Spagnolo et al. 'Experimental Photonic Quantum Memristor'. In: *Nature Photonics* 16.4 [2022], pp. 318–323. ISSN: 17494893. DOI: 10.1038/s41566-022-00973-5.

[59] Alberto Peruzzo et al. 'A Variational Eigenvalue Solver on a Photonic Quantum Processor'. In: *Nature Communications* 5.1 [23rd Sept. 2014], p. 4213. ISSN: 2041-1723. DOI: 10.1038/ncomms5213.

[60] Francesco Hoch et al. *Variational Quantum Cloning Machine on a Photonic Integrated Interferometer*. 8th July 2024. DOI: 10.48550/arXiv.2407.06026. arXiv: 2407.06026. [Visited on 21/11/2024]. Pre-published.

[61] Valeria Cimini et al. 'Variational Quantum Algorithm for Experimental Photonic Multiparameter Estimation'. In: *npj Quantum Information* 10.1 [28th Feb. 2024], pp. 1–9. ISSN: 2056-6387. DOI: 10.1038/s41534-024-00821-0. [Visited on 04/03/2024].

[62] Daniel Gottesman and Isaac L. Chuang. 'Demonstrating the Viability of Universal Quantum Computation Using Teleportation and Single-Qubit Operations'. In: *Nature* 402.6760 [1999], pp. 390–393. ISSN: 00280836. DOI: 10.1038/46503. PMID: 12721623.

[63] Saumil Bandyopadhyay, Ryan Hamerly and Dirk Englund. 'Hardware Error Correction for Programmable Photonics'. In: *Optica* 8.10 [20th Oct. 2021], p. 1247. ISSN: 2334-2536. DOI: 10.1364/OPTICA.424052. arXiv: 2103.04993.

[64] E. Knill, R. Laflamme and G. J. Milburn. 'A Scheme for Efficient Quantum Computation with Linear Optics'. In: *Nature* 409.6816 [2001], pp. 46–52. ISSN: 00280836. DOI: 10.1038/35051009. PMID: 11343107.

[65] T. C. Ralph et al. 'Linear Optical CNOT Gate in the Coincidence Basis'. In: *Physical Review A* 65.6 [20th June 2002], p. 062324. ISSN: 1050-2947, 1094-1622. DOI: 10.1103/PhysRevA.65.062324. arXiv: quant-ph/0112088. [Visited on 27/02/2023].

[66] Holger F. Hofmann and Shigeki Takeuchi. 'Quantum Phase Gate for Photonic Qubits Using Only Beam Splitters and Postselection'. In: *Physical Review A* 66.2 [8th Aug. 2002], p. 024308. DOI: 10.1103/PhysRevA.66.024308. [Visited on 22/08/2023].

[67] J. L. O'Brien et al. 'Demonstration of an All-Optical Quantum Controlled-NOT Gate'. In: *Nature* 426.6964 [6964 Nov. 2003], pp. 264–267. ISSN: 1476-4687. DOI: 10.1038/nature02054. [Visited on 22/08/2023].

[68] K. Igeta and Y. Yamamoto. 'Quantum Mechanical Computers with Single Atom and Photon Fields'. In: *International Conference on Quantum Electronics (1988), Paper TuI4*. International Quantum Electronics Conference. Optica Publishing Group, 18th July 1988, TuI4. [Visited on 09/10/2025].

[69] Bastian Hacker et al. 'A Photon–Photon Quantum Gate Based on a Single Atom in an Optical Resonator'. In: *Nature* 536.7615 [Aug. 2016], pp. 193–196. ISSN: 1476-4687. DOI: 10.1038/nature18592. [Visited on 09/10/2025].

[70] Darrick E. Chang, Vladan Vuleti and Mikhail D. Lukin. 'Quantum Nonlinear Optics — Photon by Photon'. In: *Nature Photonics* 8.9 [Sept. 2014], pp. 685–694. ISSN: 1749-4893. DOI: 10.1038/nphoton.2014.192. [Visited on 09/10/2025].

[71] Wenlan Chen et al. 'All-Optical Switch and Transistor Gated by One Stored Photon'. In: *Science* 341.6147 [16th Aug. 2013], pp. 768–770. DOI: 10.1126/science.1238169. [Visited on 27/10/2025].

[72] Robert Raussendorf. 'A One-Way Quantum Computer'. In: *Physical Review Letters* 86.22 [2001], pp. 5188–5191. DOI: 10.1103/PhysRevLett.86.5188.

[73] Robert Raussendorf, Daniel E. Browne and Hans J. Briegel. 'Measurement-Based Quantum Computation on Cluster States'. In: *Physical Review A* 68.2 [25th Aug. 2003], p. 022312. ISSN: 1050-2947, 1094-1622. DOI: 10.1103/PhysRevA.68.022312. [Visited on 26/05/2025].

[74] R Raussendorf, J Harrington and K Goyal. 'Topological Fault-Tolerance in Cluster State Quantum Computation'. In: *New Journal of Physics* 9.6 [June 2007], p. 199. ISSN: 1367-2630. DOI: 10.1088/1367-2630/9/6/199. [Visited on 12/10/2025].

[75] H. J. Briegel et al. 'Measurement-Based Quantum Computation'. In: *Nature Physics* 5.1 [2009], pp. 19–36. ISSN: 17452481. DOI: 10.1038/nphys1157. arXiv: 0910.1116.

[76] Gerald Gilbert. 'Efficient Construction of Photonic Quantum-Computational Clusters'. In: *Physical Review A* 73.6 [2006]. DOI: 10.1103/PhysRevA.73.064303.

[77] Sara Bartolucci et al. 'Fusion-Based Quantum Computation'. 22nd Jan. 2021. arXiv: 2101.09310.

[78] Mercedes Gimeno-Segovia et al. 'From Three-Photon Greenberger-Horne-Zeilinger States to Ballistic Universal Quantum Computation'. In: *Physical Review Letters* 115.2 [2015], pp. 1–5. ISSN: 10797114. DOI: 10.1103/PhysRevLett.115.020502.

[79] Dik Bouwmeester. 'Observation of Three-Photon Greenberger-Horne-Zeilinger Entanglement'. In: *Physical Review Letters* 82.7 [1999], pp. 1345–1349. DOI: 10.1103/PhysRevLett.82.1345.

[80] Chao-Yang Lu et al. 'Experimental Entanglement of Six Photons in Graph States'. In: *Nature Physics* 3.2 [Feb. 2007], pp. 91–95. ISSN: 1745-2481. DOI: 10.1038/nphys507. [Visited on 12/10/2025].

[81] Xi Lin Wang et al. '18-Qubit Entanglement with Six Photons' Three Degrees of Freedom'. In: *Physical Review Letters* 120.26 [2018]. ISSN: 10797114. DOI: 10.1103/PhysRevLett.120.260502. arXiv: 1801.04043.

[82] An-Ning Zhang. 'Experimental Construction of Optical Multiqubit Cluster States from Bell States'. In: *Physical Review A* 73.2 [2006]. DOI: 10.1103/PhysRevA.73.022330.

[83] L. G. Valiant. 'The Complexity of Computing the Permanent'. In: *Theoretical Computer Science* 8.2 [1st Jan. 1979], pp. 189–201. ISSN: 0304-3975. DOI: 10.1016/0304-3975(79)90044-6. [Visited on 11/10/2025].

[84] Leonid Gurvits. 'On the Complexity of Mixed Discriminants and Related Problems'. In: *Mathematical Foundations of Computer Science 2005*. Ed. by Joanna Jdrzejowicz and Andrzej Szepietowski. Red. by David Hutchison et al. Vol. 3618. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005, pp. 447–458. ISBN: 978-3-540-28702-5 978-3-540-31867-5. DOI: 10.1007/11549345_39. [Visited on 05/06/2024].

[85] Scott Aaronson and Travis Hance. *Generalizing and Derandomizing Gurvits's Approximation Algorithm for the Permanent*. 4th Dec. 2012. arXiv: 1212.0025 [quant-ph]. [Visited on 05/06/2024]. Pre-published.

[86] Juan Carlos Garcia-Escartin, Vicent Gimeno and Julio José Moyano-Fernández. *Multiple Photon Hamiltonian in Linear Quantum Optical Networks*. 3rd Oct. 2023. DOI: 10.1016/j.optcom.2018.08.082. arXiv: 1605.02653 [quant-ph]. [Visited on 25/02/2025]. Pre-published.

[87] Alberto Politi, Jonathan C.F. Matthews and Jeremy L. O'Brien. 'Shor's Quantum Factoring Algorithm on a Photonic Chip'. In: *Science* 325.5945 [2009], p. 1221. ISSN: 00368075. DOI: 10.1126/science.1173731. PMID: 19729649.

[88] P Walther et al. 'Experimental One-Way Quantum Computing'. In: *Nature* 434.7030 [Mar. 2005], pp. 169–176. ISSN: 0028-0836. DOI: 10.1038/nature03347.

[89] Hagai B. Perets. 'Realization of Quantum Walks with Negligible Decoherence in Waveguide Lattices'. In: *Physical Review Letters* 100.17 [2008]. DOI: 10.1103/PhysRevLett.100.170506.

[90] Linda Sansoni et al. 'Two-Particle Bosonic-Fermionic Quantum Walk via Integrated Photonics'. In: *Physical Review Letters* 108.1 [5th Jan. 2012], p. 010502. DOI: 10.1103/PhysRevLett.108.010502. [Visited on 19/06/2023].

[91] Xiao-song Ma et al. 'Quantum Simulation of the Wavefunction to Probe Frustrated Heisenberg Spin Systems'. In: *Nature Physics* 7.5 [5 May 2011], pp. 399–405. ISSN: 1745-2481. DOI: 10.1038/nphys1919. [Visited on 07/08/2023].

[92] Han-Sen Zhong et al. '12-Photon Entanglement and Scalable Scattershot Boson Sampling with Optimal Entangled-Photon Pairs from Parametric Down-Conversion'. In: *Physical Review Letters* 121.25 [21st Dec. 2018], p. 250505. ISSN: 0031-9007. DOI: 10.1103/PhysRevLett.121.250505. PMID: 30608840.

[93] Nicolò Spagnolo et al. 'Experimental Validation of Photonic Boson Sampling'. In: *Nature Photonics* 8.8 [8 Aug. 2014], pp. 615–620. ISSN: 1749-4893. DOI: 10.1038/nphoton.2014.135. [Visited on 19/06/2023].

[94] Han-Sen Zhong et al. 'Quantum Computational Advantage Using Photons'. In: *Science* 370.6523 [18th Dec. 2020], pp. 1460–1463. DOI: 10.1126/science.abe8770. [Visited on 25/06/2024].

[95] Sasan Fathpour. 'Silicon Photonics'. In: *Handbook of Optoelectronics, Second Edition: Concepts, Devices, and Techniques* 1.12 [2017], pp. 759–784. ISSN: 07338724. DOI: 10.1201/9781315157009.

[96] Michael Hochberg and Tom Baehr-Jones. 'Towards Fabless Silicon Photonics'. In: *Nature Photonics* 4.8 [2010], pp. 492–494. ISSN: 1749-4885. DOI: 10.1038/nphoton.2010.172.

[97] Shawn Yohanes Siew et al. 'Review of Silicon Photonics Technology and Platform Development'. In: *Journal of Lightwave Technology* [2021], pp. 1–1. ISSN: 0733-8724. DOI: 10.1109/jlt.2021.3066203.

[98] Richard Soref. 'The Past, Present, and Future of Silicon Photonics'. In: *IEEE Journal on Selected Topics in Quantum Electronics* 12.6 [2006], pp. 1678–1687. ISSN: 1077260X. DOI: 10.1109/JSTQE.2006.883151. PMID: 18542597.

[99] David Thomson et al. 'Roadmap on Silicon Photonics'. In: *Journal of Optics (United Kingdom)* 18.7 [2016], pp. 0–20. ISSN: 20408986. DOI: 10.1088/2040-8978/18/7/073003.

[100] J. W. Silverstone et al. 'On-Chip Quantum Interference between Silicon Photon-Pair Sources'. In: *Nature Photonics* 8.2 [2014], pp. 104–108. ISSN: 17494885. DOI: 10.1038/nphoton.2013.339. PMID: 25382491.

[101] Brian F. Aull et al. 'Large-Format Geiger-Mode Avalanche Photodiode Arrays and Readout Circuits'. In: *IEEE Journal of Selected Topics in Quantum Electronics* 24.2 [Mar. 2018], pp. 1–10. ISSN: 1077-260X, 1558-4542. DOI: 10.1109/JSTQE.2017.2736440. [Visited on 23/10/2023].

[102] Gur Lubin et al. 'Quantum Correlation Measurement with Single Photon Avalanche Diode Arrays'. In: *arXiv* 27.23 [2019], pp. 32863–32882. ISSN: 23318422. DOI: 10.1364/oe.27.032863. PMID: 31878363.

[103] Salih Yanikgonul et al. 'Integrated Avalanche Photodetectors for Visible Light'. In: *Nature Communications* 12.1 [23rd Mar. 2021], p. 1834. ISSN: 2041-1723. DOI: 10.1038/s41467-021-22046-x. [Visited on 02/06/2025].

[104] Itamar Holzman and Yachin Ivry. 'Superconducting Nanowires for Single-Photon Detection: Progress, Challenges, and Opportunities'. In: *Advanced Quantum Technologies* 2.3–4 [2019], p. 1800058. ISSN: 2511-9044. DOI: 10.1002/qute.201800058. [Visited on 14/03/2025].

[105] Chandra M. Natarajan, Michael G. Tanner and Robert H. Hadfield. 'Superconducting Nanowire Single-Photon Detectors: Physics and Applications'. In: *Superconductor Science and Technology* 25.6 [2012], p. 063001. ISSN: 09532048. DOI: 10.1088/0953-2048/25/6/063001. arXiv: 1204.5560.

[106] Zhenghao Yin et al. 'Frequency Correlated Photon Generation at Telecom Band Using Silicon Nitride Ring Cavities'. In: *Optics Express* 29.4 [15th Feb. 2021], p. 4821. ISSN: 1094-4087. DOI: 10.1364/OE.416165.

[107] Alberto Politi et al. 'Silica-on-Silicon Waveguide Quantum Circuits'. In: *Science* 320.5876 [2008], pp. 646–649. ISSN: 00368075. DOI: 10.1126/science.1155441. arXiv: 0802.0136.

[108] David J. Moss et al. 'New CMOS-compatible Platforms Based on Silicon Nitride and Hydex for Nonlinear Optics'. In: *Nature Photonics* 7.8 [1st Aug. 2013], pp. 597–607. ISSN: 17494885. DOI: 10.1038/nphoton.2013.183.

[109] Giacomo Corrielli, Andrea Crespi and Roberto Osellame. 'Femtosecond Laser Micromachining for Integrated Quantum Photonics'. In: *Nanophotonics* 10.15 [2nd Nov. 2021], pp. 3789–3812. ISSN: 2192-8614. DOI: 10.1515/nanoph-2021-0419. [Visited on 25/01/2024].

[110] Roberto Osellame et al. 'Femtosecond Writing of Active Optical Waveguides with Astigmatically Shaped Beams'. In: *JOSA B* 20.7 [1st July 2003], pp. 1559–1567. ISSN: 1520-8540. DOI: 10.1364/JOSAB.20.001559. [Visited on 19/06/2023].

[111] Roberto Osellame, Giulio Cerullo and Roberta Ramponi, eds. *Femtosecond Laser Micromachining: Photonic and Microfluidic Devices in Transparent Materials*. Vol. 123. Topics in Applied Physics. Berlin, Heidelberg: Springer, 2012. ISBN: 978-3-642-23365-4 978-3-642-23366-1. DOI: 10.1007/978-3-642-23366-1. [Visited on 25/01/2024].

[112] Ciro Pentangelo et al. 'High-Fidelity and Polarization-Insensitive Universal Photonic Processors Fabricated by Femtosecond Laser Writing'. In: *Nanophotonics* [16th Jan. 2024]. ISSN: 2192-8614. DOI: 10.1515/nanoph-2023-0636. [Visited on 25/01/2024].

[113] S.K. Yeh et al. 'Silicon Photonics Platform for Next Generation Data Communication Technologies'. In: *2024 IEEE International Electron Devices Meeting (IEDM)*. 2024 IEEE International Electron Devices Meeting (IEDM). Dec. 2024, pp. 1–4. DOI: 10.1109/IEDM50854.2024.10873369. [Visited on 19/03/2025].

[114] Nicholas C. Harris et al. 'Large-Scale Quantum Photonic Circuits in Silicon'. In: *Nanophotonics* 5.3 [2016], pp. 456–468. ISSN: 21928614. DOI: 10.1515/nanoph-2015-0146.

[115] Junqiu Liu et al. 'High-Yield, Wafer-Scale Fabrication of Ultralow-Loss, Dispersion-Engineered Silicon Nitride Photonic Circuits'. In: *Nature Communications* 12.1 [16th Dec. 2021], p. 2236. ISSN: 2041-1723. DOI: 10.1038/s41467-021-21973-z.

[116] Ciro Pentangelo et al. 'Analytical Modeling of the Static and Dynamic Response of Thermally Actuated Optical Waveguide Circuits'. In: *Physical Review Research* 3.2 [4th May 2021], p. 023094. ISSN: 2643-1564. DOI: 10.1103/PhysRevResearch.3.023094. [Visited on 26/10/2025].

[117] Laurent Labonté et al. 'Integrated Photonics for Quantum Communications and Metrology'. In: *PRX Quantum* 5.1 [12th Feb. 2024], p. 010101. ISSN: 2691-3399. DOI: 10.1103/PRXQuantum.5.010101. [Visited on 13/02/2024].

[118] Emanuele Pelucchi et al. 'The Potential and Global Outlook of Integrated Photonics for Quantum Technologies'. In: *Nature Reviews Physics* 4.3 [2022], pp. 194–208. ISSN: 25225820. DOI: 10.1038/s42254-021-00398-z.

[119] Gary F. Sinclair et al. 'Temperature Dependence of the Kerr Nonlinearity and Two-Photon Absorption in a Silicon Waveguide at 1.55 μ m'. In: *Physical Review Applied* 11.4 [2019], p. 1. ISSN: 23317019. DOI: 10.1103/PhysRevApplied.11.044084.

[120] L.B. Soldano and E.C.M. Pennings. 'Optical Multi-Mode Interference Devices Based on Self-Imaging: Principles and Applications'. In: *Journal of Lightwave Technology* 13.4 [Apr. 1995], pp. 615–627. ISSN: 1558-2213. DOI: 10.1109/50.372474. [Visited on 26/10/2025].

[121] Katsunari Okamoto. *Fundamentals of Optical Waveguides*. 2006. ISBN: 978-0-12-525096-2. DOI: 10.1016/B978-0-12-525096-2.X5000-4.

[122] Gazi Mahamud Hasan et al. 'MMI Couplers and the Talbot Effect, Symmetries and Golden Ratio'. In: *Photonics* 12.3 [Mar. 2025], p. 229. ISSN: 2304-6732. DOI: 10.3390/photonics12030229. [Visited on 26/10/2025].

[123] Jianming Wen, Yong Zhang and Min Xiao. 'The Talbot Effect: Recent Advances in Classical Optics, Nonlinear Optics, and Quantum Optics'. In: *Advances in Optics and Photonics* 5.1 [31st Mar. 2013], pp. 83–130. ISSN: 1943-8206. DOI: 10.1364/AOP.5.000083. [Visited on 26/10/2025].

[124] Haoyang Sun et al. 'Silicon Photonic Phase Shifters and Their Applications: A Review'. In: *Micromachines* 13.9 [12th Sept. 2022], p. 1509. ISSN: 2072-666X. DOI: 10.3390/mi13091509. [Visited on 15/04/2023].

[125] C. M. Wilkes et al. '60 dB High-Extinction Auto-Configured Mach–Zehnder Interferometer'. In: *Optics Letters* 41.22 [15th Nov. 2016], p. 5318. ISSN: 0146-9592. DOI: 10.1364/OL.41.005318.

[126] W. H. P. Pernice et al. 'High-Speed and High-Efficiency Travelling Wave Single-Photon Detectors Embedded in Nanophotonic Circuits'. In: *Nature Communications* 3.1 [27th Dec. 2012], p. 1325. ISSN: 2041-1723. DOI: 10.1038/ncomms2307. [Visited on 26/10/2025].

[127] Risheng Cheng et al. 'A 100-Pixel Photon-Number-Resolving Detector Unveiling Photon Statistics'. In: *Nature Photonics* 17.1 [Jan. 2023], pp. 112–119. ISSN: 1749-4893. DOI: 10.1038/s41566-022-01119-3. [Visited on 29/10/2024].

[128] Juho Kerttula et al. 'Mode Evolution in Long Tapered Fibers with High Tapering Ratio'. In: *Optics Express* 20.23 [2012], p. 25461. DOI: 10.1364/oe.20.025461.

[129] T. Tsuchizawa et al. 'Microphotonics Devices Based on Silicon Microfabrication Technology'. In: *IEEE Journal of Selected Topics in Quantum Electronics* 11.1 [Jan. 2005], pp. 232–240. ISSN: 1558-4542. DOI: 10.1109/JSTQE.2004.841479. [Visited on 03/11/2025].

[130] Koen Alexander et al. 'A Manufacturable Platform for Photonic Quantum Computing'. 26th Apr. 2024. arXiv: 2404.17570 [physics, physics:quant-ph]. [Visited on 21/05/2024].

[131] Bryn A. Bell and Ian A. Walmsley. 'Further Compactifying Linear Optical Unitaries'. In: *APL Photonics* 6.7 [1st July 2021], p. 070804. ISSN: 2378-0967. DOI: 10.1063/5.0053421. arXiv: 2104.07561.

[132] Joshua Wimbridge Silverstone. 'Entangled Light in Silicon Waveguides'. Univeristy of Bristol, 2015.

[133] Nicholas Christopher Harris. 'Programmable Nanophotonics for Quantum Information Processing and Artificial Intelligence'. MIT, 2017.

[134] Andreas Fyrillas et al. 'Scalable Machine Learning-Assisted Clear-Box Characterization for Optimally Controlled Photonic Circuits'. In: *Optica* 11.3 [20th Mar. 2024], p. 427. ISSN: 2334-2536. DOI: 10.1364/OPTICA.512148. [Visited on 03/04/2024].

[135] Tobias Haug, Chris N Self and MS Kim. 'Quantum machine learning of large datasets using randomized measurements'. In: *arXiv preprint arXiv:2108.01039* [2021].

[136] Sofiene Jerbi et al. 'Quantum machine learning beyond kernel methods'. In: *Nature Communications* 14.1 [2023], p. 517.

[137] Jacob Biamonte et al. 'Quantum machine learning'. In: *Nature* 549.7671 [2017], pp. 195–202.

[138] Kishor Bharti et al. 'Noisy Intermediate-Scale Quantum Algorithms'. In: *Reviews of Modern Physics* 94.1 [2022], p. 015004. DOI: 10.1103/RevModPhys.94.015004.

[139] John Preskill. 'Quantum computing in the NISQ era and beyond'. In: *Quantum* 2.July [2018]. arXiv: 1801.00862, pp. 1–20. ISSN: 2521327X. DOI: 10.22331/q-2018-08-06-79.

[140] Matthias C. Caro et al. 'Generalization in Quantum Machine Learning from Few Training Data'. In: *Nature Communications* 13.1 [1 22nd Aug. 2022], p. 4919. ISSN: 2041-1723. DOI: 10.1038/s41467-022-32550-3. [Visited on 16/12/2023].

[141] Dirk Heimann et al. *Learning Capability of Parametrized Quantum Circuits*. 13th Mar. 2024. arXiv: 2209.10345 [quant-ph]. [Visited on 03/04/2024]. Pre-published.

[142] Sergio Boixo et al. 'Characterizing quantum supremacy in near-term devices'. In: *Nature Physics* 14.6 [2018], pp. 595–600.

[143] Sreetama Das, Stefano Martina and Filippo Caruso. *The Role of Data Embedding in Equivariant Quantum Convolutional Neural Networks*. 27th Jan. 2024. arXiv: 2312.13250 [quant-ph]. [Visited on 31/03/2024]. Pre-published.

[144] Hsin-Yuan Huang et al. 'Quantum Advantage in Learning from Experiments'. In: *Science* 376.6598 [10th June 2022], pp. 1182–1186. ISSN: 0036-8075. DOI: 10.1126/science.abn7293. arXiv: 2112.00778.

[145] Hsin Yuan Huang et al. 'Power of Data in Quantum Machine Learning'. In: *Nature Communications* 12.1 [2021]. ISSN: 20411723. DOI: 10.1038/s41467-021-22539-9. PMID: 33976136.

[146] Maria Schuld and Nathan Killoran. 'Is Quantum Advantage the Right Goal for Quantum Machine Learning?' 2022. arXiv: 2203.01340.

[147] Giuseppe Carleo et al. 'Machine Learning and the Physical Sciences'. In: *Reviews of Modern Physics* 91.4 [2019], p. 45002. ISSN: 0034-6861. DOI: 10.1103/revmodphys.91.045002. arXiv: 1903.10563.

[148] Giuseppe Carleo and Matthias Troyer. 'Solving the Quantum Many-Body Problem with Artificial Neural Networks'. In: *Science* 355.6325 [10th Feb. 2017], pp. 602–606. DOI: 10.1126/science.aag2302. [Visited on 25/09/2025].

[149] Alexandra Nagy. 'Variational Quantum Monte Carlo Method with a Neural-Network Ansatz for Open Quantum Systems'. In: *Physical Review Letters* 122.25 [2019]. DOI: 10.1103/PhysRevLett.122.250501.

[150] Filippo Vicentini. 'Variational Neural-Network Ansatz for Steady States in Open Quantum Systems'. In: *Physical Review Letters* 122.25 [2019]. DOI: 10.1103/PhysRevLett.122.250503.

[151] Giacomo Torlai et al. 'Neural-Network Quantum State Tomography'. In: *Nature Physics* 14.5 [May 2018], pp. 447–450. ISSN: 1745-2481. DOI: 10.1038/s41567-018-0048-5. [Visited on 27/10/2025].

[152] Andreas Elben et al. 'The Randomized Measurement Toolbox'. In: *Nature Reviews Physics* 5.1 [1 Jan. 2023], pp. 9–24. ISSN: 2522-5820. DOI: 10.1038/s42254-022-00535-2. [Visited on 17/08/2023].

[153] Hsin-Yuan Huang, Richard Kueng and John Preskill. 'Predicting Many Properties of a Quantum System from Very Few Measurements'. In: *Nature Physics* 16.10 [22nd Oct. 2020], pp. 1050–1057. ISSN: 1745-2473. DOI: 10.1038/s41567-020-0932-7. arXiv: 2002.08953.

[154] Maria Schuld and Francesco Petruccione. *Supervised Learning with Quantum Computers*. Quantum Science and Technology. Cham: Springer International Publishing, 2018. ISBN: 978-3-319-96423-2 978-3-319-96424-9. DOI: 10.1007/978-3-319-96424-9. [Visited on 14/07/2025].

[155] Maria Schuld and Francesco Petruccione. *Machine Learning with Quantum Computers*. Quantum Science and Technology. Cham: Springer International Publishing, 2021. ISBN: 978-3-030-83097-7 978-3-030-83098-4. DOI: 10.1007/978-3-030-83098-4. [Visited on 14/07/2025].

[156] Nathan Killoran et al. 'Continuous-Variable Quantum Neural Networks'. In: *Physical Review Research* 1.3 [2019], pp. 1–21. ISSN: 26431564. DOI: 10.1103/PhysRevResearch.1.033063. arXiv: 1806.06871.

[157] Seth Lloyd et al. *Quantum Embeddings for Machine Learning*. 2020. arXiv: 2001.03622. Pre-published.

[158] Maria Schuld and Nathan Killoran. 'Quantum Machine Learning in Feature Hilbert Spaces'. In: *Physical Review Letters* 122.4 [2019], p. 40504. ISSN: 10797114. DOI: 10.1103/PhysRevLett.122.040504. PMID: 30768345.

[159] Maria Schuld et al. 'Circuit-Centric Quantum Classifiers'. In: *Physical Review A* 101.3 [6th Mar. 2020], p. 032308. ISSN: 2469-9926, 2469-9934. DOI: 10.1103/PhysRevA.101.032308. arXiv: 1804.00633 [quant-ph]. [Visited on 14/03/2023].

[160] Maria Schuld. 'Supervised Quantum Machine Learning Models Are Kernel Methods'. 2021. arXiv: 2101.11020.

[161] Jeremy Adcock et al. 'Advances in Quantum Machine Learning'. 2015. arXiv: 1512.02900.

[162] Hayata Yamasaki, Natsuto Isogai and Mio Murao. *Advantage of Quantum Machine Learning from General Computational Advantages*. 5th Dec. 2023. arXiv: 2312.03057 [quant-ph, stat]. [Visited on 05/03/2024]. Pre-published.

[163] Hsin-Yuan Huang et al. 'Provably Efficient Machine Learning for Quantum Many-Body Problems'. In: *Science* 377.6613 [23rd Sept. 2022], eabk3333. DOI: 10.1126/science.abk3333. [Visited on 04/03/2024].

[164] M. Cerezo et al. 'Variational Quantum Algorithms'. In: *Nature Reviews Physics* 3.9 [9 Sept. 2021], pp. 625–644. ISSN: 2522-5820. DOI: 10.1038/s42254-021-00348-9. [Visited on 03/02/2023].

[165] Hong-Ye Hu et al. *Demonstration of Robust and Efficient Quantum Property Learning with Shallow Shadows*. 27th Feb. 2024. arXiv: 2402.17911 [cond-mat, physics:quant-ph]. [Visited on 29/08/2024]. Pre-published.

[166] Mehryar Mohri, Afshin Rostamizadeh and Ameet Talwalkar. *Foundations of Machine Learning*. Second edition. Adaptive Computation and Machine Learning. Cambridge, Massachusetts: The MIT Press, 2018. 486 pp. ISBN: 978-0-262-03940-6.

[167] Arthur Gretton et al. 'A Kernel Method for the Two-Sample-Problem'. In: *Advances in Neural Information Processing Systems 19*. Ed. by Bernhard Schölkopf, John Platt and Thomas Hofmann. The MIT Press, 7th Sept. 2007, pp. 513–520. ISBN: 978-0-262-25691-9. DOI: 10.7551/mitpress/7503.003.0069. [Visited on 28/10/2024].

[168] Thomas Hofmann, Bernhard Schölkopf and Alexander J. Smola. 'Kernel Methods in Machine Learning'. In: *The Annals of Statistics* 36.3 [June 2008], pp. 1171–1220. ISSN: 0090-5364, 2168-8966. DOI: 10.1214/009053607000000677. [Visited on 08/10/2025].

[169] F. Rosenblatt. 'The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain'. In: *Psychological Review* 65.6 [1958], pp. 386–408. ISSN: 1939-1471. DOI: 10.1037/h0042519.

[170] J J Hopfield. 'Neural Networks and Physical Systems with Emergent Collective Computational Abilities.' In: *Proceedings of the National Academy of Sciences* 79.8 [Apr. 1982], pp. 2554–2558. DOI: 10.1073/pnas.79.8.2554. [Visited on 28/09/2025].

[171] Kurt Hornik, Maxwell Stinchcombe and Halbert White. 'Multilayer Feedforward Networks Are Universal Approximators'. In: *Neural Networks* 2.5 [1st Jan. 1989], pp. 359–366. ISSN: 0893-6080. DOI: 10.1016/0893-6080(89)90020-8. [Visited on 28/09/2025].

[172] David E. Rumelhart, Geoffrey E. Hinton and Ronald J. Williams. 'Learning Representations by Back-Propagating Errors'. In: *Nature* 323.6088 [Oct. 1986], pp. 533–536. ISSN: 1476-4687. DOI: 10.1038/323533a0. [Visited on 25/09/2025].

[173] Y. Lecun et al. 'Gradient-Based Learning Applied to Document Recognition'. In: *Proceedings of the IEEE* 86.11 [Nov. 1998], pp. 2278–2324. ISSN: 1558-2256. DOI: 10.1109/5.726791. [Visited on 25/07/2025].

[174] Yann Lecun, Yoshua Bengio and Geoffrey Hinton. 'Deep Learning'. In: *Nature* 521.7553 [27th May 2015], pp. 436–444. ISSN: 14764687. DOI: 10.1038/nature14539. PMID: 26017442.

[175] Alice Barthe et al. *Parameterized Quantum Circuits as Universal Generative Models for Continuous Multivariate Distributions*. 29th May 2024. DOI: 10.48550/arXiv.2402.09848. arXiv: 2402.09848. [Visited on 05/12/2024]. Pre-published.

[176] Alberto Manzano et al. 'Parametrized Quantum Circuits and Their Approximation Capacities in the Context of Quantum Machine Learning'. In: [] DOI: 10.22331/q-2025-03-10-1658.

[177] Marcello Benedetti et al. 'Parameterized Quantum Circuits as Machine Learning Models'. In: *Quantum Science and Technology* 5.1 [4th Dec. 2019], p. 019601. ISSN: 2058-9565. DOI: 10.1088/2058-9565/ab5944. [Visited on 06/03/2024].

[178] Alistair Letcher, Stefan Woerner and Christa Zoufal. *Tight and Efficient Gradient Bounds for Parameterized Quantum Circuits*. 15th Feb. 2024. arXiv: 2309.12681 [quant-ph]. [Visited on 01/04/2024]. Pre-published.

[179] Kouhei Nakaji, Hiroyuki Tezuka and Naoki Yamamoto. *Deterministic and Random Features for Large-Scale Quantum Kernel Machine*. 5th Sept. 2022. arXiv: 2209.01958 [quant-ph]. [Visited on 03/02/2023]. Pre-published.

[180] Maria Schuld et al. 'Evaluating Analytic Gradients on Quantum Hardware'. In: *Physical Review A* 99.3 [21st Mar. 2019], p. 032331. ISSN: 2469-9926, 2469-9934. DOI: 10.1103/PhysRevA.99.032331. [Visited on 18/04/2023].

[181] Kosuke Mitarai et al. 'Quantum Circuit Learning'. In: *Physical Review A* 98.3 [10th Sept. 2018], p. 032309. ISSN: 2469-9926, 2469-9934. DOI: 10.1103/PhysRevA.98.032309. arXiv: 1803.00745 [quant-ph]. [Visited on 11/03/2023].

[182] J.C. Spall. 'Multivariate Stochastic Approximation Using a Simultaneous Perturbation Gradient Approximation'. In: *IEEE Transactions on Automatic Control* 37.3 [Mar. 1992], pp. 332–341. ISSN: 1558-2523. DOI: 10.1109/9.119632. [Visited on 28/09/2025].

[183] Stephanie Wehner, David Elkouss and Ronald Hanson. 'Quantum internet: A vision for the road ahead'. In: *Science* 362.6412 [2018], eaam9288.

[184] I. M. Georgescu, S. Ashhab and Franco Nori. 'Quantum Simulation'. In: *Reviews of Modern Physics* 86.1 [10th Mar. 2014], pp. 153–185. ISSN: 0034-6861, 1539-0756. DOI: 10.1103/RevModPhys.86.153. [Visited on 18/08/2023].

[185] Lars S. Madsen et al. 'Quantum Computational Advantage with a Programmable Photonic Processor'. In: *Nature* 606.7912 [2022], pp. 75–81. ISSN: 0028-0836. DOI: 10.1038/s41586-022-04725-x.

[186] Frank Arute et al. 'Quantum Supremacy Using a Programmable Superconducting Processor'. In: *Nature* 574.7779 [2019], pp. 505–510. ISSN: 14764687. DOI: 10.1038/s41586-019-1666-5. PMID: 31645734.

[187] Hartmut Neven et al. 'QBoost: Large Scale Classifier Training withAdiabatic Quantum Optimization'. In: *Proceedings of the Asian Conference on Machine Learning*. Asian Conference on Machine Learning. PMLR, 17th Nov. 2012, pp. 333–348. [Visited on 08/10/2025].

[188] Patrick Rebentrost, Masoud Mohseni and Seth Lloyd. 'Quantum Support Vector Machine for Big Data Classification'. In: *Physical Review Letters* 113.13 [25th Sept. 2014], p. 130503. ISSN: 0031-9007, 1079-7114. DOI: 10.1103/PhysRevLett.113.130503. arXiv: 1307.0471 [quant-ph]. [Visited on 09/08/2023].

[189] M. S. Leifer and D. Poulin. 'Quantum Graphical Models and Belief Propagation'. In: *Annals of Physics* 323.8 [1st Aug. 2008], pp. 1899–1946. ISSN: 0003-4916. DOI: 10.1016/j.aop.2007.10.001. [Visited on 08/10/2025].

[190] Sergio Boixo et al. 'Characterizing Quantum Supremacy in Near-Term Devices'. In: *Nature Physics* 14.6 [June 2018], pp. 595–600. ISSN: 1745-2481. DOI: 10.1038/s41567-018-0124-x. [Visited on 08/10/2025].

[191] Beng Yee Gan, Daniel Leykam and Dimitris G. Angelakis. 'Fock State-Enhanced Expressivity of Quantum Machine Learning Models'. In: *EPJ Quantum Technology* 9.1 [1 Dec. 2022], pp. 1–23. ISSN: 2196-0763. DOI: 10.1140/epjqt/s40507-022-00135-0. [Visited on 29/09/2023].

[192] John Shawe-Taylor and Nello Cristianini. *Kernel methods for pattern analysis*. Cambridge university press, 2004.

[193] Corinna Cortes and Vladimir Vapnik. 'Support-vector networks'. In: *Machine learning* 20 [1995], pp. 273–297.

[194] Yunchao Liu, Srinivasan Arunachalam and Kristan Temme. 'A Rigorous and Robust Quantum Speed-up in Supervised Machine Learning'. In: *Nature Physics* 17.9 [12th Sept. 2021], pp. 1013–1017. ISSN: 1745-2473. DOI: 10.1038/s41567-021-01287-z. arXiv: 2010.02174.

[195] Michel Talagrand. 'Concentration of Measure and Isoperimetric Inequalities in Product Spaces'. In: *Publications Mathématiques de l'Institut des Hautes Études Scientifiques* 81.1 [1st Dec. 1995], pp. 73–205. ISSN: 1618-1913. DOI: 10.1007/BF02699376. [Visited on 08/10/2025].

[196] Supanut Thanasilp et al. *Exponential Concentration in Quantum Kernel Methods*. 14th Apr. 2024. DOI: 10.48550/arXiv.2208.11060. arXiv: 2208.11060 [quant-ph, stat]. [Visited on 16/04/2024]. Pre-published.

[197] Jonas M. Kübler, Simon Buchholz and Bernhard Schölkopf. *The Inductive Bias of Quantum Kernels*. 9th Nov. 2021. arXiv: 2106.03747 [quant-ph, stat]. [Visited on 16/03/2023]. Pre-published.

[198] Arthur Jacot, Franck Gabriel and Clément Hongler. *Neural Tangent Kernel: Convergence and Generalization in Neural Networks*. 10th Feb. 2020. DOI: 10.48550/arXiv.1806.07572. arXiv: 1806.07572 [cs, math, stat]. [Visited on 26/03/2024]. Pre-published.

[199] Bernhard E Boser, Isabelle M Guyon and Vladimir N Vapnik. 'A training algorithm for optimal margin classifiers'. In: *Proceedings of the fifth annual workshop on Computational learning theory*. 1992, pp. 144–152.

[200] Karol Bartkiewicz et al. 'Experimental Kernel-Based Quantum Machine Learning in Finite Feature Space'. In: *Scientific Reports* 10.1 [2020], pp. 1–9. ISSN: 20452322. DOI: 10.1038/s41598-020-68911-5. PMID: 32704032.

[201] Hsin Yuan Huang, Richard Kueng and John Preskill. 'Information-Theoretic Bounds on Quantum Advantage in Machine Learning'. In: *Physical Review Letters* 126.19 [2021], p. 190505. ISSN: 10797114. DOI: 10.1103/PhysRevLett.126.190505. PMID: 34047595.

[202] Takeru Kusumoto et al. 'Experimental Quantum Kernel Trick with Nuclear Spins in a Solid'. In: *npj Quantum Information* 7.1 [1 8th June 2021], pp. 1–7. ISSN: 2056-6387. DOI: 10.1038/s41534-021-00423-0. [Visited on 21/02/2024].

[203] Bernhard Schölkopf and Alexander J Smola. *Learning with kernels: support vector machines, regularization, optimization, and beyond*. MIT press, 2002.

[204] Sofiene Jerbi et al. 'Quantum Machine Learning beyond Kernel Methods'. In: *arXiv.org* [25th Oct. 2021]. DOI: 10.1038/s41467-023-36159-y. [Visited on 27/11/2023].

[205] Malte C Tichy. 'Interference of Identical Particles from Entanglement to Boson-Sampling'. In: *Journal of Physics B: Atomic, Molecular and Optical Physics* 47.10 [May 2014], p. 103001. ISSN: 0953-4075. DOI: 10.1088/0953-4075/47/10/103001. [Visited on 08/10/2025].

[206] Francesco Ceccarelli et al. 'Low power reconfigurability and reduced crosstalk in integrated photonic circuits fabricated by femtosecond laser micromachining'. In: *Laser & Photonics Reviews* 14.10 [2020], p. 2000024.

[207] C. K. Hong, Z. Y. Ou and L. Mandel. 'Measurement of Subpicosecond Time Intervals between Two Photons by Interference'. In: *Physical Review Letters* 59.18 [2nd Nov. 1987], pp. 2044–2046. ISSN: 0031-9007. DOI: 10.1103/PhysRevLett.59.2044. [Visited on 12/01/2025].

[208] Vojtch Havlíek et al. 'Supervised Learning with Quantum-Enhanced Feature Spaces'. In: *Nature* 567.7747 [13th Mar. 2019], pp. 209–212. ISSN: 0028-0836. DOI: 10.1038/s41586-019-0980-2. PMID: 30867609.

[209] Francesco Hoch et al. 'Quantum machine learning with Adaptive Boson Sampling via post-selection'. In: *Nature Communications* 16.1 [2025], p. 902.

[210] Youngrong Lim and Changhun Oh. *Efficient Classical Algorithms for Linear Optical Circuits*. 18th Feb. 2025. DOI: 10.48550/arXiv.2502.12882. arXiv: 2502.12882 [quant-ph]. [Visited on 19/03/2025]. Pre-published.

[211] Ryan Hamerly et al. 'Large-scale optical neural networks based on photoelectric multiplication'. In: *Physical Review X* 9.2 [2019], p. 021032.

[212] Jaehoon Lee et al. 'Finite versus infinite neural networks: an empirical study'. In: *Advances in Neural Information Processing Systems* 33 [2020], pp. 15156–15172.

[213] Adityanarayanan Radhakrishnan et al. 'Transfer learning with kernel methods'. In: *Nature Communications* 14.1 [2023], p. 5570.

[214] Yao-Hung Hubert Tsai et al. 'Transformer Dissection: An Unified Understanding for Transformer's Attention via the Lens of Kernel'. In: *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, EMNLP-IJCNLP 2019, Hong Kong, China, November 3-7, 2019*. Ed. by Kentaro Inui et al. Association for Computational Linguistics, 2019, pp. 4343–4352. DOI: 10.18653/V1/D19-1443. URL: https://doi.org/10.18653/v1/D19-1443.

[215] Hubert Ramsauer et al. 'Hopfield networks is all you need'. In: *arXiv preprint arXiv:2008.02217* [2020].

[216] Pablo Rodriguez-Grasa et al. 'Satellite image classification with neural quantum kernels'. In: *Machine Learning: Science and Technology* [2024].

[217] Pablo Rodriguez-Grasa, Yue Ban and Mikel Sanz. *Training Embedding Quantum Kernels with Data Re-Uploading Quantum Neural Networks*. 9th Jan. 2024. DOI: 10.48550/arXiv.2401.04642. arXiv: 2401.04642 [quant-ph]. [Visited on 26/01/2024]. Pre-published.

[218] Xinbiao Wang et al. 'Towards Understanding the Power of Quantum Kernels in the NISQ Era'. In: *Quantum* 5 [Aug. 2021], p. 531. DOI: 10.22331/q-2021-08-30-531. URL: https://quantum-journal.org/papers/q-2021-08-30-531/ [visited on 15/02/2024].

[219] Chao-Hua Yu et al. 'Quantum Algorithm for Association Rules Mining'. In: *Physical Review A* 94.4 [Oct. 2016], p. 042311. DOI: 10.1103/PhysRevA.94.042311. URL: https://link.aps.org/doi/10.1103/PhysRevA.94.042311 [visited on 15/02/2024].

[220] Robin Lorenz et al. 'QNLP in Practice: Running Compositional Models of Meaning on a Quantum Computer'. In: *ArXiv e-prints* [Feb. 2021]. arXiv: 2102.12846. [Visited on 02/12/2021].

[221] Jonas Landman et al. 'Quantum Methods for Neural Networks and Application to Medical Image Classification'. In: *Quantum* 6 [Dec. 2022], p. 881. DOI: 10.22331/q-2022-12-22-881. URL: https://quantum-journal.org/papers/q-2022-12-22-881/ [visited on 15/02/2024].

[222] Maria Schuld and Francesco Petruccione. 'Quantum Models as Kernel Methods'. In: *Machine Learning with Quantum Computers*. Ed. by Maria Schuld and Francesco Petruccione. Quantum Science and Technology. Cham: Springer International Publishing, 2021, pp. 217–245. ISBN: 978-3-030-83098-4. DOI: 10.1007/978-3-030-83098-4_6. URL: https://doi.org/10.1007/978-3-030-83098-4_6 [visited on 15/02/2024].

[223] Zakari Denis. 'Photonic Kernel Machine Learning for Ultrafast Spectral Analysis'. In: *Physical Review Applied* 17.3 [2022]. DOI: 10.1103/PhysRevApplied.17.034077.

[224] LCG Govia et al. 'Nonlinear input transformations are ubiquitous in quantum reservoir computing'. In: *Neuromorphic Computing and Engineering* 2.1 [2022], p. 014008.

[225] Luca Innocenti et al. 'Potential and limitations of quantum extreme learning machines'. In: *Communications Physics* 6.1 [2023], p. 118.

[226] Leonardo Banchi, W. Steven Kolthammer and M. S. Kim. 'Multiphoton Tomography with Linear Optics and Photon Counting'. In: *Physical Review Letters* 121.25 [18th Dec. 2018], p. 250402. ISSN: 0031-9007, 1079-7114. DOI: 10.1103/PhysRevLett.121.250402. [Visited on 10/03/2023].

[227] J. J. Renema et al. 'Efficient Classical Algorithm for Boson Sampling with Partially Distinguishable Photons'. In: *Physical Review Letters* 120.22 [2018]. Publisher: American Physical Society, p. 220502. ISSN: 10797114. DOI: 10.1103/PhysRevLett.120.220502. URL: https://doi.org/10.1103/PhysRevLett.120.220502.

[228] Roman Novak et al. 'Neural Tangents: Fast and Easy Infinite Neural Networks in Python'. In: *International Conference on Learning Representations*. 2020. URL: https://github.com/google/neural-tangents.

[229] Jacques Carolan et al. 'Universal linear optics'. In: *Science* 349.6249 [2015], pp. 711–716.

[230] Nicolas Maring et al. 'A versatile single-photon-based quantum computing platform'. In: *Nature Photonics* 18.6 [2024], pp. 603–609.

[231] Zhenghao Yin et al. 'Experimental quantum-enhanced kernel-based machine learning on a photonic processor'. In: *Nature Photonics* [2025], pp. 1–8.

[232] Hsin-Yuan Huang et al. 'Provably Efficient Machine Learning for Quantum Many-Body Problems'. In: *Science* 377.6613 [23rd Sept. 2022], eabk3333. DOI: 10.1126/science.abk3333. [Visited on 04/03/2024].

[233] Marco Cerezo et al. 'Challenges and opportunities in quantum machine learning'. In: *Nature Computational Science* 2.9 [2022], pp. 567–576.

[234] PsiQuantum Team et al. 'A Manufacturable Platform for Photonic Quantum Computing'. In: *Nature* [Feb. 2025]. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/s41586-025-08820-7. [Visited on 26/03/2025].

[235] Hui Zhang et al. 'Resource-efficient high-dimensional subspace teleportation with a quantum autoencoder'. In: *Science advances* 8.40 [2022], eabn9783.

[236] Jacques Carolan et al. 'Variational Quantum Unsampling on a Quantum Photonic Processor'. In: *Nature Physics* 16.3 [2020], pp. 322–327. ISSN: 17452481. DOI: 10.1038/s41567-019-0747-6. arXiv: 1904.10463.

[237] Michele Spagnolo et al. 'Experimental Photonic Quantum Memristor'. In: *Nature Photonics* 16.4 [2022], pp. 318–323. ISSN: 17494893. DOI: 10.1038/s41566-022-00973-5.

[238] Gregory R. Steinbrecher et al. 'Quantum Optical Neural Networks'. In: *npj Quantum Information* 5.1 [17th Dec. 2019], p. 60. ISSN: 2056-6387. DOI: 10.1038/s41534-019-0174-7.

[239] Scott Aaronson and Alex Arkhipov. 'The computational complexity of linear optics'. In: *Proceedings of the forty-third annual ACM symposium on Theory of computing*. 2011, pp. 333–342.

[240] Han-Sen Zhong et al. 'Quantum computational advantage using photons'. In: *Science* 370.6523 [2020], pp. 1460–1463.

[241] Scott Aaronson and Lijie Chen. 'Complexity-theoretic foundations of quantum supremacy experiments'. In: *arXiv preprint arXiv:1612.05903* [2016].

[242] Ciro Pentangelo et al. 'High-fidelity and polarization-insensitive universal photonic processors fabricated by femtosecond laser writing'. en. In: *Nanophotonics* [Jan. 2024]. Publisher: De Gruyter. ISSN: 2192-8614. DOI: 10.1515/nanoph-2023-0636. URL: https://www.degruyter.com/document/doi/10.1515/nanoph-2023-0636/html [visited on 25/01/2024].

[243] Andrea Crespi et al. 'Integrated Multimode Interferometers with Arbitrary Designs for Photonic Boson Sampling'. In: *Nature Photonics* 7.7 [7 July 2013], pp. 545–549. ISSN: 1749-4893. DOI: 10.1038/nphoton.2013.112. [Visited on 19/06/2023].

[244] Yuxuan Du et al. 'Learnability of Quantum Neural Networks'. In: *PRX Quantum* 2.4 [17th Nov. 2021], p. 040337. DOI: 10.1103/PRXQuantum.2.040337. [Visited on 06/03/2024].

[245] Amira Abbas et al. 'The power of quantum neural networks'. In: *Nature Computational Science* 1.6 [2021], pp. 403–409.

[246] Leonardo Banchi, Jason Pereira and Stefano Pirandola. 'Generalization in quantum machine learning: A quantum information standpoint'. In: *PRX Quantum* 2.4 [2021], p. 040321.

[247] Tobias Haug and MS Kim. 'Generalization of quantum machine learning models using quantum fisher information metric'. In: *Physical Review Letters* 133.5 [2024], p. 050603.

[248] Elies Gil-Fuster, Jens Eisert and Carlos Bravo-Prieto. 'Understanding quantum machine learning also requires rethinking generalization'. In: *Nature Communications* 15.1 [2024], p. 2277.

[249] Tobias Haug, Kishor Bharti and MS Kim. 'Capacity and quantum geometry of parametrized quantum circuits'. In: *PRX Quantum* 2.4 [2021], p. 040309.

[250] Jing Liu et al. 'Quantum Fisher information matrix and multiparameter estimation'. In: *Journal of Physics A: Mathematical and Theoretical* 53.2 [2020], p. 023001.

[251] Martín Larocca et al. 'Theory of Overparametrization in Quantum Neural Networks'. In: *Nature Computational Science* 3.6 [6 June 2023], pp. 542–551. ISSN: 2662-8457. DOI: 10.1038/s43588-023-00467-6. [Visited on 04/07/2023].

[252] Sumeet Khatri et al. 'Quantum-assisted quantum compiling'. In: *Quantum* 3 [2019], p. 140.

[253] J. L. O'Brien et al. 'Quantum Process Tomography of a Controlled-NOT Gate'. In: *Physical Review Letters* 93.8 [20th Aug. 2004], p. 080502. DOI: 10.1103/PhysRevLett.93.080502. [Visited on 12/04/2023].

[254] M. Mohseni, A. T. Rezakhani and D. A. Lidar. 'Quantum-Process Tomography: Resource Analysis of Different Strategies'. In: *Physical Review A* 77.3 [Mar. 2008], p. 032322. ISSN: 1050-2947, 1094-1622. DOI: 10.1103/PhysRevA.77.032322. [Visited on 20/05/2025].

[255] James C Spall. 'Multivariate stochastic approximation using a simultaneous perturbation gradient approximation'. In: *IEEE transactions on automatic control* 37.3 [1992], pp. 332–341.

[256] Mahmut Kaya and Hasan akir Bilge. 'Deep metric learning: A survey'. In: *Symmetry* 11.9 [2019], p. 1066.

[257] Florian Schroff, Dmitry Kalenichenko and James Philbin. 'Facenet: A unified embedding for face recognition and clustering'. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015, pp. 815–823.

[258] Alec Radford et al. 'Learning transferable visual models from natural language supervision'. In: *International conference on machine learning*. PMLR. 2021, pp. 8748–8763.

[259] Logan G Wright et al. 'Deep physical neural networks trained with backpropagation'. In: *Nature* 601.7894 [2022], pp. 549–555.

[260] Saumil Bandyopadhyay et al. 'Single-chip photonic deep neural network with forward-only training'. In: *Nature Photonics* 18.12 [2024], pp. 1335–1343.

[261] Yichen Shen et al. 'Deep Learning with Coherent Nanophotonic Circuits'. In: *Nature Photonics* 11.7 [2017], pp. 441–446. ISSN: 1749-4885. DOI: 10.1038/nphoton.2017.93. pmid: 20181947.

[262] Valeria Cimini et al. 'Variational quantum algorithm for experimental photonic multiparameter estimation'. In: *npj Quantum Information* 10.1 [2024], p. 26.

[263] Aaron W. Young et al. 'An Atomic Boson Sampler'. In: *Nature* 629.8011 [May 2024], pp. 311–316. ISSN: 1476-4687. DOI: 10.1038/s41586-024-07304-4. [Visited on 28/01/2025].

[264] Xiaowei Deng et al. 'Quantum-enhanced metrology with large Fock states'. In: *Nature Physics* [2024], pp. 1–7.

[265] H. Qiao et al. 'Splitting Phonons: Building a Platform for Linear Mechanical Quantum Computing'. In: *Science* 380.6649 [9th June 2023], pp. 1030–1033. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.adg8715. [Visited on 15/06/2025].

[266] James Stokes et al. 'Quantum natural gradient'. In: *Quantum* 4 [2020], p. 269.

[267] Ulrik L. Andersen et al. 'Hybrid Discrete- and Continuous-Variable Quantum Information'. In: *Nature Physics* 11.9 [2015], pp. 713–719. ISSN: 17452481. DOI: 10.1038/nphys3410.

[268] Jennifer R. Glick et al. 'Covariant Quantum Kernels for Data with Group Structure'. In: *Nature Physics* [19th Jan. 2024]. ISSN: 1745-2473, 1745-2481. DOI: 10.1038/s41567-023-02340-9. [Visited on 02/02/2024].

[269] Logan G. Wright et al. 'Deep Physical Neural Networks Trained with Backpropagation'. In: *Nature* 601.7894 [Jan. 2022], pp. 549–555. ISSN: 1476-4687. DOI: 10.1038/s41586-021-04223-6. [Visited on 06/09/2024].

[270] Diederik P Kingma and Jimmy Ba. 'Adam: A method for stochastic optimization'. In: *arXiv preprint arXiv:1412.6980* [2014].

[271] John Polcari. *Representing unitary matrices by independent parameters*. Tech. rep. Working Paper, Rev 0, October, 2016, 2016.

[272] Jonatan Bohr Brask. 'Gaussian states and operations–a quick reference'. In: *arXiv preprint arXiv:2102.05748* [2021].

[273] Xuchen You, Shouvanik Chakrabarti and Xiaodi Wu. 'A convergence theory for over-parameterized variational quantum eigensolvers'. In: *arXiv preprint arXiv:2205.12481* [2022].

[274] Alexander Yu Vlasov. 'Permanents, bosons and linear optics'. In: *Laser Physics Letters* 14.10 [2017], p. 103001.

[275] Axel Pappalardo et al. 'A Photonic Parameter-shift Rule: Enabling Gradient Computation for Photonic Quantum Computers'. In: *arXiv preprint arXiv:2410.02726* [2024].

[276] Giorgio Facelli et al. 'Exact gradients for linear optics with single photons'. In: *arXiv preprint arXiv:2409.16369* [2024].

[277] Julien Gacon et al. 'Simultaneous perturbation stochastic approximation of the quantum fisher information'. In: *Quantum* 5 [2021], p. 567.

[278] Nicolas Maring et al. 'A Versatile Single-Photon-Based Quantum Computing Platform'. In: *Nature Photonics* [26th Mar. 2024], pp. 1–7. ISSN: 1749-4893. DOI: 10.1038/s41566-024-01403-4. [Visited on 17/05/2024].

[279] Aidan Pellow-Jarman et al. 'A comparison of various classical optimizers for a variational quantum linear solver'. In: *Quantum Information Processing* 20.6 [2021], p. 202.

[280] James C Spall. 'Implementation of the simultaneous perturbation algorithm for stochastic optimization'. In: *IEEE Transactions on aerospace and electronic systems* 34.3 [1998], pp. 817–823.

[281] Weiyang Liu et al. 'Sphereface: Deep hypersphere embedding for face recognition'. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017, pp. 212–220.

[282] Hao Wang et al. 'Cosface: Large margin cosine loss for deep face recognition'. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018, pp. 5265–5274.

[283] Jiankang Deng et al. 'Arcface: Additive angular margin loss for deep face recognition'. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2019, pp. 4690–4699.

[284] Miguel Romera et al. 'Vowel Recognition with Four Coupled Spin-Torque Nano-Oscillators'. In: *Nature* 563.7730 [Nov. 2018], pp. 230–234. ISSN: 1476-4687. DOI: 10.1038/s41586-018-0632-y. [Visited on 08/05/2024].

[285] Ali Momeni et al. 'Backpropagation-free training of deep physical neural networks'. In: *Science* 382.6676 [2023], pp. 1297–1303.

[286] Noam Chomsky. 'Language and Nature'. In: *Mind* 104.413 [1995], pp. 1–61. ISSN: 0026-4423. JSTOR: 2254605. [Visited on 25/07/2025].

[287] Noam Chomsky, Adriana Belletti and Luigi Rizzi. *On Nature and Language*. Cambridge: Cambridge University Press, 2002. 206 pp. ISBN: 978-0-521-81548-2. [Visited on 25/07/2025].

[288] Tomas Mikolov et al. 'Distributed Representations of Words and Phrases and Their Compositionality'. In: *Advances in Neural Information Processing Systems*. Vol. 26. Curran Associates, Inc., 2013. [Visited on 15/10/2025].

[289] Luciano Floridi and Massimo Chiriatti. 'GPT-3: Its Nature, Scope, Limits, and Consequences'. In: *Minds and Machines* 30.4 [1st Dec. 2020], pp. 681–694. ISSN: 1572-8641. DOI: 10.1007/s11023-020-09548-1. [Visited on 25/07/2025].

[290] Tom Brown et al. 'Language Models Are Few-Shot Learners'. In: *Advances in Neural Information Processing Systems*. Vol. 33. Curran Associates, Inc., 2020, pp. 1877–1901. [Visited on 25/07/2025].

[291] Alec Radford et al. 'Language Models Are Unsupervised Multitask Learners'. In: []

[292] Hsin-Yuan Huang. 'Learning in the Quantum Universe'.

[293] Robin Lorenz and Jonathan Barrett. 'Causal and Compositional Structure of Unitary Transformations'. In: *Quantum* 5 [2021], p. 511. DOI: 10.22331/q-2021-07-28-511. arXiv: 2001.07774.

[294] V. M. Placinta et al. 'Measurements of Radiation Effects in an Antifuse FPGA'. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 1055 [1st Oct. 2023], p. 168551. ISSN: 0168-9002. DOI: 10.1016/j.nima.2023.168551. [Visited on 03/10/2025].

[295] Oltjon Kodheli et al. 'Satellite Communications in the New Space Era: A Survey and Future Challenges'. In: *IEEE Communications Surveys & Tutorials* 23.1 [2021], pp. 70–109. ISSN: 1553-877X. DOI: 10.1109/COMST.2020.3028247. [Visited on 03/10/2025].

[296] Hemani Kaushal and Georges Kaddoum. 'Optical Communication in Space: Challenges and Mitigation Techniques'. In: *IEEE Communications Surveys & Tutorials* 19.1 [2017], pp. 57–96. ISSN: 1553-877X. DOI: 10.1109/COMST.2016.2603518. [Visited on 03/10/2025].

[297] Makan Mohageg et al. 'Towards Satellite Tests Combining General Relativity and Quantum Mechanics through Quantum Optical Interferometry: Progress on the Deep Space Quantum Link'. In: *EPJ Quantum Technology* 12.1 [20th June 2025], p. 78. ISSN: 2196-0763. DOI: 10.1140/epjqt/s40507-025-00370-1. [Visited on 02/10/2025].

[298] Juan Yin et al. 'Satellite-Based Entanglement Distribution over 1200 Kilometers'. In: *Science* 356.6343 [16th June 2017], pp. 1140–1144. DOI: 10.1126/science.aan3211. [Visited on 30/04/2025].

[299] Ji-Gang Ren et al. 'Ground-to-Satellite Quantum Teleportation'. In: *Nature* 549.7670 [Sept. 2017], pp. 70–73. ISSN: 1476-4687. DOI: 10.1038/nature23675. [Visited on 02/10/2025].

[300] Sheng-Kai Liao et al. 'Satellite-to-Ground Quantum Key Distribution'. In: *Nature* 549.7670 [Sept. 2017], pp. 43–47. ISSN: 1476-4687. DOI: 10.1038/nature23655. [Visited on 02/10/2025].

[301] Yang Li et al. 'Microsatellite-Based Real-Time Quantum Key Distribution'. In: *Nature* 640.8057 [Apr. 2025], pp. 47–54. ISSN: 1476-4687. DOI: 10.1038/s41586-025-08739-z. [Visited on 02/10/2025].

[302] Jean-Philippe Bourgoin et al. 'Free-Space Quantum Key Distribution to a Moving Receiver'. In: *Optics Express* 23.26 [28th Dec. 2015], pp. 33437–33447. ISSN: 1094-4087. DOI: 10.1364/OE.23.033437. [Visited on 02/10/2025].

[303] A. Scott et al. 'The QEYSSAT Mission: On-Orbit Demonstration of Secure Optical Communications Network Technologies'. In: *Environmental Effects on Light Propagation and Adaptive Systems III*. Environmental Effects on Light Propagation and Adaptive Systems III. Vol. 11532. SPIE, 20th Sept. 2020, pp. 71–76. DOI: 10.1117/12.2574154. [Visited on 02/10/2025].

[304] Aitor Villar et al. 'Entanglement Demonstration on Board a Nano-Satellite'. In: *Optica* 7.7 [20th July 2020], pp. 734–737. ISSN: 2334-2536. DOI: 10.1364/OPTICA.387306. [Visited on 30/04/2025].

[305] Anantha Rao et al. *Learning Hard Distributions with Quantum-Enhanced Variational Autoencoders*. 18th May 2023. arXiv: 2305.01592 [quant-ph]. [Visited on 22/11/2023]. Pre-published.

[306] Jonathan Romero, Jonathan P Olson and Alan Aspuru-Guzik. 'Quantum Autoencoders for Efficient Compression of Quantum Data'. In: *Quantum Science and Technology* 2.4 [1st Dec. 2017], p. 045001. ISSN: 2058-9565. DOI: 10.1088/2058-9565/aa8072. [Visited on 06/03/2024].

Acknowledgements

First and foremost, I am deeply grateful to my advisor, Prof. Philip Walther, for his unwavering support, guidance, and inspiration throughout my PhD journey. His mentorship has been invaluable, and I have learned so much from his expertise and dedication to research. I started to apply for PhD positions during the pandemic, and I am thankful to Philip for giving me the opportunity to join his group despite the challenging circumstances. His inspiration and encouragement on quantum technologies have profoundly triggered my interest in quantum machine learning.

I would also like to extend my sincere thanks to my colleagues, Valeria Saggio, Teodor Störmberg, Peter Schiansky and Michele Spagnolo, for their support to start my research work and for their help to settle down in Vienna. For Teo, I appreciate his patient guidance on the setup and calibration of our chip, which laid a solid foundation for my research. We also enjoyed many interesting and inspiring discussions about quantum information and quantum optics, especially those topics beyond our research field. For Peter, I am grateful for his endless help on everything in our lab, from the laser to the chip, from the electronics to the cryostat. He is always willing and eager to help and share his knowledge and experience, which made my research life much easier and joyful. For Michele and Valeria, I appreciate their collaboration and support on starting my first project. I also would like to thank Mirela Selimovic, Hubertus Lauterbacher, Martin Mauser, Elena Negro. We did not overlap on the same projects a lot, but their daily interactions and company in the lab have pleased me greatly. I want to thank Iris Agresti in particular, for her continuous support and collaboration, not only in the kernel project but also in many other experiments. We have shared many fruitful discussions and brainstorming sessions, which have greatly enriched my research experience.

In the ROQUET project, we have a great team including Tobias Guggemos, Antonius Scherer, Simon Steiner, Daniel Martinez Arias, Patrick Zahalka, Hans Küffner-McCauley, Riccardo Albiero, Francesco Ceccarelli as well as Peter and Iris mentioned above. I am thankful to work with all of them and learn a lot from their expertise, which is precious and valuable for my future research. Now our satellite is launched successfully, and I am excited to see the results of our work and wish you people all the best in the future.

During my time in the projects, the group led by Prof. Roberto Osellame at CNRS and Politecnico di Milano has been a great source of inspiration and collaboration. I sincerely appreciate their support and guidance throughout my research, from Prof. Osellame, Andrea Crespi, Ciro Pentangelo and Francesco

mentioned above. For Ciro, I deeply thank him for his time traveling and working with me, both in Vienna and Milano. I learned a lot how to fabricate, calibrate the laser-written chip and perform the experiments. Now they started up a new company named *Ephos*, and I wish them a great success in their future endeavors.

I am grateful to all current and former members of the Walther group for creating a stimulating and collaborative research environment. For Bob Peterson, I do appreciate his insightful and helpful discussions on electronics. I also want to thank Philip Jenke, Michael Antesberger and Lee Rozema for their maintenance of our cryostat and guidance for me. For Huan Cao, Haocun Yu, Patrik Sund, Yong Wang and Juan C. Loredo, I am grateful for their time and insights from their research.

Beyond our group, I would like to thank all members of the Vienna Center for Quantum Science and Technology (VCQ) and the Vienna Doctoral School in Physics (VDSP) for their support and for organizing various events and activities. I did not attend many activities due to the pandemic, but I still benefited a lot from the seminars and workshops they organized. For the retreat in 2022 and other events, I am thankful to meet many new friends and enjoy the beautiful nature in Austria. I also want to thank the administrative staff, in the quantum group, VCQ and VDSP, for their help and support on various matters.

Beyond my research life in Vienna, I am very grateful to all fellows in our MSCA Innovative Training Network (ITN) "AppQInfo", Gabriele Bressanini, Ewan Mer, Jernej Frank, Matthew Filipovich, Ilectra Karakosta-Amarantidou, Kannan Vijayadharan, Morteza Moradi, Valerii Kachin, Marco Robbio, Benoit Seron, Ioannis Georgakilas, Pietro Tassan, Laura dos Santos Martins and Lorenzo Stasi. I would like to thank my colleague Lennart Jehle and Lena Hansen specially, as he contributed significantly to organizing our ITN activities and events. During the time of all ITN events, we traveled and talked a lot together, and I am thankful for the friendships we built along the way. I also want to thank our supervisors and mentors, who provided valuable guidance and support throughout the ITN program.

Through my PhD journey, I have had the privilege of collaborating with many researchers from different institutions and backgrounds. I am grateful for their insights, support, and friendship, which have enriched my research experience. In my secondment at Imperial College London, I want to thank Prof. Ian Walmsley, Raj Patel, Shang Yu, Aonan Zhang and Zhenghao Li for their warm welcome and support. I also want to thank Tobias Haug for his collaboration and discussions on quantum machine learning. There are also many other researchers I have contacted and discussed with, we did not chance to establish a deep collaboration, but I still benefit a lot from their insights and suggestions.

Finally, I would like to express my deepest gratitude to my family and friends for their unwavering support and encouragement throughout my PhD journey.